

# BOLIVIA INTERNACIONAL

## REVISTA DE CIENCIA POLÍTICA



### N° 7 - NUEVA ÉPOCA

La opinión pública en la era de la inteligencia artificial generativa: transformaciones en los mecanismos de intervención política  
**José Antonio Martínez Díez Barroso**

Cambio técnico y acumulación: inteligencia artificial, trabajo y superpoblación. Europa e Iberoamérica en la División Internacional del Trabajo  
**Nicolás Di Paola**

IA Pública Segura: Causalidad, Robustez y Decisión  
**Ricardo Alonzo Fernández Salguero**

La responsabilidad política en la toma de decisiones automatizadas del Estado democrático  
**Marysol Estrella Hernández García & Carolina Aguilar Ramos**

Relaciones internacionales contemporáneas: cooperación multilateral, polierisis y el rol propositivo de América Latina  
**Gastón Gaete Coddou & Jeffrey Yamil Kafka Delgado**

Seguridad, soberanía y tecnología en la frontera Brasil-Bolivia: asimetrías y cooperación en el período 2018-2024  
**María del Pilar Serrate Roca, Giovanna Savazo & Claudio Vitor Cardoso**

¿Cultura y venganza? Los usos de la retaliación en la vida social de una comunidad campesina de los Andes colombianos  
**Santiago Álvarez**

Big Data e Inteligencia Artificial como medios digitales en la política pública: adopción, práctica y barreras  
**José David Vallejo Manzur & Erick Leobardo Alvarez-Aros**

Sesgos de IA como Desigualdad Política: Violencia Digital de Género y Evaluación Pericial  
**Albanella Lucna Chávez-Turello**

La ley de hierro de la oligarquía revisitada: élites, instituciones y evolución en la teoría política contemporánea  
**Facundo Guadagno Balmaceda**

Efectos de la propaganda electoral en los estudiantes durante las elecciones universitarias UAGRM (2025)  
**Mariano Ramos Claire**

CONOCIMIENTO POLÍTICO

JUNIO 2026

ISSN-L: 3106-4736

# Centro de Investigación Regional en Ciencia Política y Ciencias Sociales

UAGRM

**REVISTA BOLIVIANA INTERNACIONAL  
EN CIENCIA POLÍTICA**

---

**Número 7 – 2026 (Nueva Época)**

**Rector:** M.Sc. Reinerio Vargas Banegas

**Vicerrectora:** M.Sc. Juana Borja Saavedra

**Decano de la FCJPSRR.II.:** M.Sc. Marcos Sánchez Kordez

**Vicedecano de la FCJPSRR.II.:** M.Sc. Adolfo Guerrero Encinas

**Director de Carrera:** M.Sc. Odín Rodríguez Mercado

**Directora del CIRCPyCS:** M.Sc. María Angélica Suárez Ribera

---

# FICHA TÉCNICA DE LA REVISTA

## *Conocimiento i Política*

*Conocimiento i Política* es una publicación académica semestral editada por el Centro de Investigación Regional en Ciencia Política y Ciencias Sociales de la Universidad Autónoma Gabriel René Moreno. Su objetivo es contribuir al debate académico en el campo de la Ciencia Política y las Ciencias Sociales.

---

<b>Título de la revista</b>	<i>Conocimiento i Política</i>
<b>Subtítulo</b>	Revista Boliviana Internacional en Ciencia Política
<b>Entidad editora</b>	Centro de Investigación Regional en Ciencia Política y Ciencias Sociales
<b>Institución</b>	Universidad Autónoma Gabriel René Moreno
<b>Número</b>	No. 7 (2026), Nueva Época
<b>Periodicidad</b>	Semestral; dos números por año
<b>Formato</b>	Publicación física y online
<b>Área temática</b>	Ciencia Política y Ciencias Sociales
<b>Tipos de manuscritos</b>	Artículos de investigación, artículos de revisión, artículos de reflexión, notas o comunicaciones, ensayos bibliográficos y reseñas
<b>Evaluación</b>	Revisión por pares bajo modalidad de doble ciego
<b>Idiomas aceptados</b>	Español e inglés
<b>ISSN electrónico</b>	<u>3106-4736</u>
<b>Plataforma digital</b>	<a href="https://ojs.uagrm.edu.bo/revista-conocimiento-politica">https://ojs.uagrm.edu.bo/revista-conocimiento-politica</a>
<b>Indexación</b>	En proceso de indexación en bases de datos académicas nacionales e internacionales
<b>Licencia</b>	Creative Commons BY-NC-ND 4.0 Internacional

---

### **Correspondencia editorial**

*Conocimiento i Política*

Centro de Investigación Regional en Ciencia Política y Ciencias Sociales

Av. Busch, 2.º Anillo, módulo 284, ciudad universitaria

Santa Cruz de la Sierra, Bolivia

[cri.cspoliticass.fcjpsri@uagrm.edu.bo](mailto:cri.cspoliticass.fcjpsri@uagrm.edu.bo)

### **Edición y diseño**

Centro de Investigación Regional en Ciencia Política y Ciencias Sociales

Tel.: +591 67381294

# EQUIPO EDITORIAL

*Conocimiento i Política*

---

## Editores

---

**M.Sc. Daniel Hurtado Justiniano**

Editor en jefe

Universidad Autónoma Gabriel René Moreno

**M.Sc. José Orlando Peralta Beltrán**

Editor Académico

Universidad Autónoma Gabriel René Moreno

**Lic. Carlos JR Escobar**

Editor Técnico

Universidad Autónoma Gabriel René Moreno

## Dirección

---

**M.Sc. María Angélica Suárez Ribera**

Directora del CIRCPyCS

Universidad Autónoma Gabriel René Moreno

## Consejo de Redacción

---

**Dr. Marco Torrez Valverde**

Universidad Autónoma Gabriel René Moreno

**Lic. Yessenia Pedraza Surano**

Universidad Autónoma Gabriel René Moreno

**Dr. Jorge Núñez del Prado**

Universidad Autónoma Gabriel René Moreno

## Consejo Editorial

---

**Dra. Paula Peña Hasbún**

Universidad Autónoma Gabriel René Moreno, Bolivia

**Dr. Fernando Escobar**

Universidad Autónoma Gabriel René Moreno, Bolivia

**Dr. Richard Vargas**

UNITEPC, Bolivia

**M.Sc. María Angélica Suárez**

Universidad Autónoma Gabriel René Moreno, Bolivia

**M.Sc. Odín Rodríguez**

Universidad Autónoma Gabriel René Moreno, Bolivia

**Dra. Carmen Dunia Sandoval**

Universidad Autónoma Gabriel René Moreno, Bolivia

**Dra. Elizabeth Galarza**

Universidad Autónoma Gabriel René Moreno, Bolivia

**M.Sc. Marisel Hinojosa Toro**

UNAM, México

**Dr. Francisco Méndez Eguez**

Universidad Autónoma Gabriel René Moreno, Bolivia

**Dra. Pía Melgar**

Universidad Autónoma Gabriel René Moreno, Bolivia

**Dra. Loreto Correa Vera**

ANEPE, Chile

**M.Sc. Roger Tuero Velásquez**

Universidad Autónoma Gabriel René Moreno, Bolivia

**M.Sc. José Orlando Peralta**

Universidad Autónoma Gabriel René Moreno, Bolivia

**Dr. Reymi Ferreira Justiniano**

Universidad Autónoma Gabriel René Moreno, Bolivia

**M.Sc. Nelson Jordan**

Universidad Católica de Bolivia, Bolivia

## Agradecimiento a pares evaluadores

Año 2025

El Consejo Editorial de *Conocimiento i Política* expresa su agradecimiento a las y los pares evaluadores que colaboraron durante el año 2025. Su lectura especializada, sus observaciones metodológicas y su compromiso con la calidad académica contribuyen al fortalecimiento editorial de la revista y al desarrollo de la Ciencia Política y las Ciencias Sociales.

### Pares evaluadores

**Paula Peña Hasbún**

Universidad Autónoma Gabriel René  
Moreno, Bolivia

**Fernando Escobar**

Universidad Autónoma Gabriel René  
Moreno, Bolivia

**Richard Vargas**

UNITEPC, Bolivia

**María Angélica Suárez Ribera**

Universidad Autónoma Gabriel René  
Moreno, Bolivia

**Odín Rodríguez Mercado**

Universidad Autónoma Gabriel René  
Moreno, Bolivia

**Ana Lucía Velasco**

Universidad Católica de La Paz,  
Bolivia

**Carmen Dunia Sandoval**

Universidad Autónoma Gabriel René  
Moreno, Bolivia

**Elizabeth Galarza Claudio**

Universidad Autónoma Gabriel René  
Moreno, Bolivia

**Marisel Hinojosa Toro**

UNAM, México

**Francisco Méndez Eguez**

Universidad Autónoma Gabriel René  
Moreno, Bolivia

**Pía Melgar**

Universidad Autónoma Gabriel René  
Moreno, Bolivia

**Guillermo Bretel**

Universidad Julius Maximilianus de  
Wurzburg, Alemania

**Loreto Correa Vera**

ANEPE, Chile

**Roger Tuero Velásquez**

Universidad Autónoma Gabriel René  
Moreno, Bolivia

**José Orlando Peralta Beltrán**

Universidad Autónoma Gabriel René  
Moreno, Bolivia

**Reymi Ferreira Justiniano**

Universidad Autónoma Gabriel René  
Moreno, Bolivia

**Nelson Jordan**

Universidad Católica de Bolivia,  
Bolivia

**Juan Pablo Marca Marca**

Centro de Investigación Jurídica y  
Social (CEJIS)

# CONTENIDO

## Editorial

Daniel Hurtado Justiniano .....	V
---------------------------------	---

## Artículos de investigación

1. <b>La opinión pública en la era de la inteligencia artificial generativa: transformaciones en los mecanismos de intervención política</b> José Antonio Martínez Díez Barroso .....	1–36
2. <b>Cambio técnico y acumulación: inteligencia artificial, trabajo y superpoblación. Europa e Iberoamérica en la División Internacional del Trabajo</b> Nicolás Di Paola .....	37–92
3. <b>IA Pública Segura: Causalidad, Robustez y Decisión</b> Ricardo Alonzo Fernández Salguero .....	93–160
4. <b>La responsabilidad política en la toma de decisiones automatizadas del Estado democrático</b> Marysol Estrella Hernández García, Carolina Aguilar Ramos .....	161–181
5. <b>Relaciones internacionales contemporáneas: cooperación multilateral, policrisis y el rol propositivo de América Latina</b> Gastón Gaete Coddou, Jeffrey Yamil Kafka Delgado .....	182–197
6. <b>Seguridad, soberanía y tecnología en la frontera Brasil–Bolivia: asimetrías y cooperación en el período 2018–2024</b> María del Pilar Serrate Roca, Giovanna Savazo, Claudio Vitor Cardoso .....	198–222
7. <b>¿Cultura y venganza? Los usos de la retaliación en la vida social de una comunidad campesina de los Andes colombianos</b> Santiago Álvarez .....	223–242

## Artículos de revisión

8. <b>Big Data e Inteligencia Artificial como medios digitales en la política pública: adopción, práctica y barreras</b> José David Vallejo Manzur, Erick Leobardo Alvarez-Aros .....	244–271
--	---------

## Artículos de reflexión

9. <b>Sesgos de IA como desigualdad política: violencia digital de género y evaluación pericial</b> Albanella Luena Chávez-Turello .....	273–285
10. <b>La ley de hierro de la oligarquía revisitada: élites, instituciones y evolución en la teoría política contemporánea</b> Facundo Guadagno Balmaceda .....	286–302

## Notas de comunicación científica

11. <b>Efectos de la propaganda electoral en los estudiantes durante las elecciones universitarias UAGRM (2025)</b> Mariano Ramos Claire .....	304–315
---	---------

## EDITORIAL

### *Conocimiento i Política, Nueva Época, Número 7*

El Número 7 de *Conocimiento i Política* se presenta en un momento en que la Ciencia Política y las Ciencias Sociales enfrentan transformaciones aceleradas en los modos de producir conocimiento, ejercer el poder y deliberar en el espacio público. Los trabajos reunidos en esta edición dialogan con esos cambios desde perspectivas diversas, pero convergen en una preocupación común: comprender las condiciones institucionales, sociales y tecnológicas que configuran la vida democrática contemporánea.

La inteligencia artificial ocupa un lugar relevante en este número. Su presencia en los debates sobre opinión pública, decisiones estatales, trabajo, políticas públicas y desigualdad obliga a superar las aproximaciones que la reducen a una herramienta meramente técnica. Los artículos muestran que toda innovación tecnológica se inscribe en relaciones de poder, marcos normativos, capacidades institucionales y experiencias sociales concretas. Por ello, la discusión sobre inteligencia artificial requiere atender tanto a sus posibilidades como a los criterios de responsabilidad, trazabilidad y control democrático que deben orientar su incorporación en los asuntos públicos.

La edición también reúne reflexiones sobre cooperación multilateral, seguridad y soberanía en espacios fronterizos, dinámicas electorales universitarias, violencia y desigualdad digital, así como los debates en torno a las élites y la organización política. Estos aportes permiten observar que los problemas públicos no se desarrollan de manera aislada: conectan escalas locales, nacionales, regionales y globales, y demandan análisis capaces de reconocer esa complejidad sin perder de vista las condiciones históricas que los hacen posibles.

Como revista, sostenemos el compromiso de ofrecer un espacio plural para la investigación, la revisión crítica y la reflexión fundamentada. Este número reafirma la vocación de *Conocimiento i Política* por promover el diálogo entre enfoques y trayectorias académicas, con rigor argumentativo y claridad expositiva.

Agradecemos a quienes confiaron sus trabajos a la revista, a las y los pares evaluadores que contribuyeron con su lectura especializada, y a las personas lectoras que acompañan este proyecto editorial.

**M.Sc. Daniel Hurtado Justiniano**  
*Editor en Jefe, Conocimiento i Política*

# ARTÍCULOS DE INVESTIGACIÓN

---

Conocimiento i Política – No. 7, 2026

# La opinión pública en la era de la inteligencia artificial generativa: transformaciones en los mecanismos de intervención política

## *Public opinion in the age of generative artificial intelligence: transformations in the mechanisms of political intervention*

José Antonio Martínez Díez Barroso<sup>1</sup>

<sup>1</sup> Instituto de Investigaciones Bibliotecológicas y de la Información (IIBI), Universidad Nacional Autónoma de México (UNAM). Ciudad de México, México.

antonio.diez@iibi.unam.mx (Autor de correspondencia)

**Resumen.** Este artículo analiza las transformaciones en los mecanismos de incidencia sobre la opinión pública a partir de la irrupción de la inteligencia artificial (IA) generativa. La investigación se sustenta en una revisión bibliográfica de carácter interpretativo, centrada en publicaciones académicas recientes localizadas en bases de datos indexadas, con el objetivo de identificar continuidades y rupturas respecto de las formas tradicionales de influencia mediática. En una primera parte, se examinan los enfoques clásicos de la opinión pública desarrollados en el campo de las ciencias políticas y sociales, destacando el papel de los medios de comunicación, las encuestas y los procesos deliberativos en la formación de las percepciones colectivas. Posteriormente, se analizan las transformaciones introducidas por las redes sociales y, de manera más reciente, por la IA generativa y los grandes modelos de lenguaje (LLM). Los resultados muestran que estas tecnologías han intensificado fenómenos como la personalización del contenido, la formación de "cámaras de eco", la simulación de consenso mediante bots y la generación de contenido sintético, lo que incide en la configuración de la opinión pública contemporánea. En este contexto, se discuten los riesgos asociados a la manipulación de la percepción colectiva, la desinformación y el debilitamiento de los procesos deliberativos, así como sus posibles implicaciones para las democracias actuales.

**Palabras clave:** Opinión pública; Inteligencia artificial; Redes sociales; Comunicación política; Desinformación; Democracia

**Abstract.** This article examines the transformations in the mechanisms influencing public opinion following the emergence of generative artificial intelligence (AI). The study is based on an interpretive literature review of recent academic publications indexed in scholarly databases, aiming to identify continuities and shifts in relation to traditional forms of media influence. The first section revisits classical approaches to public opinion within political and social sciences, emphasizing the role of mass media, surveys, and deliberative processes in shaping collective perceptions. It then explores the transformations introduced by social media and, more recently, by generative AI and large language models (LLM). The findings indicate that these technologies have intensified phenomena such as content personalization, "echo chambers", the simulation of consensus through bots, and the generation of synthetic content, all of which influence the configuration of contemporary public opinion. In this context, the article discusses the risks associated with the manipulation of collective perception, disinformation, and the weakening of deliberative processes, as well as their potential implications for contemporary democracies.

**Keywords:** Public opinion; Artificial intelligence; Social media; Political communication; Disinformation; Democracy

**Historial:** Recibido: 30 de marzo de 2026 | Aceptado: 15 de mayo de 2026

**Ética/Financiamiento:** El autor declara no tener conflictos de interés. El artículo reconoce el apoyo del Programa de Becas Posdoctorales en la Universidad Nacional Autónoma de México (POSDOC).

Este documento está bajo una licencia  
CC BY-NC-ND 4.0.

**Cómo citar (APA):** Martínez Díez Barroso, J. A. (2026). La opinión pública en la era de la inteligencia artificial generativa: transformaciones en los mecanismos de intervención política. *Conocimiento i Política*, 7(1), 1-36. DOI: 10.64480/cip.2026.n7.01.

## 1. Introducción

Este artículo expone una investigación teórica basada en la revisión bibliográfica de publicaciones localizadas en bases de datos indexadas. Se centra en las transformaciones de los mecanismos que inciden en la manipulación de la opinión pública tras la irrupción de la inteligencia artificial (IA).

En ese sentido, resulta pertinente señalar que la comprensión de los mecanismos a través de los cuales se influye en la opinión pública ha interesado históricamente a políticos, gobiernos y diversos actores sociales (Lippmann, 1922). Una percepción positiva se traduce en validación y reconocimiento por parte de la ciudadanía, mientras que una percepción negativa implica la pérdida de legitimidad y aceptación social.

Tradicionalmente, los medios de comunicación (como la prensa, la radio y la televisión) desempeñaban un papel central en la difusión de información con el propósito de posicionar las agendas de quienes financiaban los espacios informativos (Lippmann, 1922). Para ello, se recurría a figuras mediáticas que emitían opiniones a favor o en contra de determinados asuntos políticos (Lippmann, 1925). Asimismo, se realizaban encuestas orientadas a medir puntos de convergencia y divergencia, así como la percepción social alrededor de temas o acontecimientos específicos (Allport, 1940). El contexto sociohistórico de los ciudadanos constituía un factor determinante para comprender tendencias y comportamientos (Converse, 1970).

En este marco, los mass media saturaron la esfera pública con contenidos, lo que dificultó la capacidad de filtrar información y de construir opiniones fundamentadas en procesos deliberativos y en consensos democráticos (Zaller, 1992). Esta dinámica afectó la comunicación entre iguales como base para la toma de decisiones informadas (Habermas, 1987), una función que, en principio, se esperaba que fuera reivindicada por las redes sociales. En sus inicios, estas prometían democratizar la expresión de opiniones: si antes los medios actuaban como intermediarios, las redes sociales podrían hacer que cualquier persona con acceso a internet difundiera su punto de vista (Castells, 2009).

No obstante, las redes sociales evidenciaron rápidamente su carácter corporativo, al estar controladas por un número reducido de empresas que monetizan con los datos de sus usuarios. Su modelo de negocio depende de maximizar el tiempo de permanencia en línea, lo que les lleva a ofrecer publicidad focalizada y a intentar

predecir comportamientos. Para ello, se diseñan algoritmos que trazan gustos y actitudes, y que recomiendan contenido hiperpersonalizado (Saura García, 2025).

Esta dinámica se ha intensificado y complejizado con la irrupción de los grandes modelos de lenguaje (LLM) y de la IA generativa. Estas tecnologías son utilizadas de manera creciente para incidir en la opinión pública mediante estrategias sofisticadas de manipulación (Coeckelbergh, 2024). Los chatbots tienden a reforzar la exposición repetitiva a contenidos afines, limitando el contacto con perspectivas divergentes (Pariser, 2011; Bucher, 2018; Aral, 2021).

Lo mismo sucede con los perfiles falsos hechos con IA en redes sociales, los cuales simulan un apoyo inexistente (Elmas, 2023, Doshi et al., 2024). La IA facilita la construcción de falsos consensos, generando la percepción de que determinadas posturas cuentan con mayor aceptación de la que realmente poseen. Además, posibilita la emergencia de fenómenos políticos inéditos, como los “candidatos virtuales” o los “dobles digitales”, que inciden en la configuración de la opinión pública (Woolley, 2023).

Por otra parte, la generación de contenido sintético en redes sociales propicia la creación de realidades destinadas a influir deliberadamente en la percepción colectiva. En este contexto, los deepfakes (y su instrumentalización mediante fenómenos como el “liar’s dividend”) constituyen un ejemplo paradigmático del uso de tecnologías para incidir en la opinión pública (Meikle, 2023; Liu et al., 2025; Jackson et al., 2025). Esto se ha incrementado de manera sostenida y es cada vez más frecuente en procesos electorales y discusiones políticas en redes sociales: imágenes altamente verosímiles de candidatos en situaciones comprometedoras, audios con voces indistinguibles de las reales en los que se admiten actos ilícitos, o videos capaces de engañar incluso a observadores atentos. Lo que antes se manifestaba como una manipulación mediática más evidente se ha transformado en un entramado de mecanismos tecnológicos complejos.

En consecuencia, la IA puede entenderse como una herramienta con un alto potencial para la manipulación de la percepción colectiva (Coeckelbergh, 2024; Saura García, 2025). Por ello, resulta pertinente desarrollar una investigación cuyo objetivo sea analizar las transformaciones en los mecanismos de incidencia sobre la opinión pública a partir de la emergencia de la IA generativa. En este marco, la pregunta que orienta el presente trabajo es: ¿cómo han variado las formas de influencia sobre la

opinión pública en el contexto de la IA generativa?

Para responder a esta interrogante, se adopta una metodología interpretativa basada en el enfoque “Scoping Review” (McLeod, 2024), que permite sintetizar y analizar críticamente las fuentes consultadas. Los resultados se estructuran en dos ejes: por un lado, los mecanismos tradicionales de influencia sobre la opinión pública y, por otro, aquellos asociados a la IA generativa. Posteriormente, se presenta un cuadro comparativo que sintetiza los principales hallazgos.

Asimismo, se establece un diálogo con investigaciones previas en las que se abordan temas como la adopción de marcos regulatorios de la IA en América Latina, los datos utilizados para entrenar los modelos, los riesgos de asumir acríticamente las innovaciones tecnológicas y los posibles efectos adversos sobre las democracias de la región. Finalmente, se proponen líneas de investigación futuras y recomendaciones orientadas a mitigar la manipulación de la opinión pública mediante el uso de la IA generativa.

## 2. Metodología

El presente artículo se sustentó en una metodología cualitativa de carácter interpretativo. Se trata de una investigación teórica que examinó diversos artículos académicos, informes institucionales y libros centrados en la opinión pública para entender los mecanismos que sirven para incidir en ella, y sus variaciones a partir del uso de la IA generativa en redes sociales. Este procedimiento fue posible mediante una revisión bibliográfica sistemática sobre la relación entre opinión pública e IA.

Los antecedentes metodológicos de estudios similares pueden hallarse en trabajos que proponen análisis críticos de corte teórico para la comprensión de fenómenos propios de las ciencias políticas y sociales (Codina, 2020). En consecuencia, la presente investigación se distanció de métodos empíricos tradicionales, como la observación, las entrevistas, los estudios de caso o el diseño experimental. Esta decisión buscó contrastar las formas de incidir en la opinión pública a partir de la irrupción de la IA generativa; en ese sentido, la revisión bibliográfica sistemática permitió identificar las principales líneas de investigación, así como los temas de mayor interés para la comunidad académica.

La revisión se sustentó en procedimientos de selección, organización y análisis crítico de los autores, lo que dio lugar a un texto de carácter teórico presentado en la sección de resultados. Asimismo, se apoyó en criterios sistemáticos de búsqueda y selección, inspirados en el enfoque “Scoping Review” (McLeod, 2024), pero adaptados a un análisis teórico-interpretativo que preste atención a los modos en que se influye en la opinión pública en el contexto del uso emergente de la IA. Este enfoque permitió mapear la evolución de las formas en que se entiende la opinión pública en la producción académica reciente.

El procedimiento metodológico se desarrolló en cinco fases: (1) definición del propósito de la revisión bibliográfica; (2) identificación de las fuentes; (3) selección de las más pertinentes, con base en su calidad y relevancia conceptual; (4) extracción de la información útil para la investigación; y (5) síntesis crítica y sistemática de los hallazgos (Kitchenham, 2004; Okoli & Schabram, 2010). Las referencias bibliográficas se localizaron mediante el motor de búsqueda Google Scholar, dada su amplia utilización para la consulta de textos académicos; después, se incorporaron libros especializados, ampliamente citados en los artículos revisados.

El periodo de búsqueda abarcó de 2020 a 2025. Esta delimitación responde a que, durante dichos años, se consolidaron y popularizaron desarrollos tecnológicos que han impactado significativamente en la opinión pública, como los chatbots, la IA generativa y los LLM. Asimismo, en ese periodo, gobiernos, partidos políticos, organizaciones de la sociedad civil y diversos actores sociales comenzaron a implementar estas tecnologías con el propósito de influir en la percepción y el juicio de las personas, afectando los procesos deliberativos y de consenso democrático.

La búsqueda bibliográfica se realizó mediante combinaciones de palabras clave en español e inglés, entre ellas: opinión pública, IA generativa y redes sociales. Los textos se filtraron temáticamente a través de las herramientas proporcionadas por el propio motor de búsqueda. Para establecer los criterios de inclusión y exclusión, se consideró que los artículos estuvieran publicados en revistas indexadas y que se inscribieran en el ámbito de las ciencias políticas y sociales. Asimismo, se priorizaron aquellos trabajos que abordaran directamente las transformaciones de la opinión pública a partir del uso de la IA. En total, se recopilaron 87 artículos, los cuales, tras un proceso de revisión y sistematización, se redujeron a 68. Este proceso permitió evitar duplicidades y redundancias, así como alcanzar una saturación conceptual; es decir, se incorporaron publicaciones hasta que dejaron de aportar nuevos temas o

conceptos relevantes para la pregunta de investigación, lo que va en consonancia con metodologías empleadas en estudios previos (Kitchenham, 2004, Okoli & Schabram, 2010).

Los resultados se organizaron en torno a dos ejes principales: el primero corresponde a la etapa clásica de la opinión pública, y el segundo a sus transformaciones a partir de la irrupción de la IA generativa. De este modo, se ofrece una tabla comparativa que permite identificar un antes y un después en la configuración de los factores que inciden en la opinión pública. Finalmente, se establece un diálogo con investigaciones previas que han abordado esta problemática en América Latina, con el fin de profundizar en su análisis teórico.

### **3. Resultados**

La revisión bibliográfica sobre los mecanismos para incidir en la opinión pública permitió identificar diferencias sustantivas entre las formas tradicionales de afectar la percepción social y lo que sucede con la llegada de los LLM y la IA generativa. Estas tecnologías alteran y reconfiguran los puntos de vista mediante estrategias de manipulación de la información desplegadas en redes sociales.

En este sentido, se presenta, en una primera instancia, la “Etapa clásica de la opinión pública” y, posteriormente, se la contrasta con “La opinión pública en la era de la IA generativa”. El análisis de estos dos ejes ayuda a comprender la forma en que se configura la opinión pública, los actores sociales que intervienen en este proceso, así como el papel de los medios de comunicación y la lógica de su funcionamiento. Asimismo, se examinan las particularidades de la deliberación pública, la producción de contenidos, la velocidad de circulación de la información y los mecanismos de influencia y persuasión, con el fin de identificar riesgos y fenómenos emergentes. Con base en esta comparación, se propone una tabla final que sintetiza los principales hallazgos.

#### 4. Etapa clásica de la opinión pública

Uno de los primeros trabajos en abordar de manera detallada la opinión pública como objeto de estudio académico fue el de Bryce (1888). En esta obra temprana se advierte ya una preocupación por la estrecha relación entre los medios de comunicación y los intereses políticos. El argumento central de Bryce gira en torno a los procesos de simbolización mediante los cuales los sentimientos y pensamientos de las personas pueden influenciarse de manera deliberada con el fin de incidir en su subjetividad y, a la par, comercializar con la atención de los ciudadanos.

Desde su perspectiva, Bryce concebía a los medios de comunicación de la época, en particular los periódicos, como empresas que operan primordialmente en el ámbito del entretenimiento. Según su planteamiento, los partidos políticos recurren a ellos para captar el tiempo y la atención de las audiencias, a quienes se expone de manera reiterada a contenidos alineados con los intereses de los grupos con poder. De esta forma, los periódicos serían aparatos de difusión capaces de moldear la percepción social alrededor de temas de relevancia pública.

La perspectiva de Bryce fue contrastada y debatida por otros autores que, en las primeras décadas del siglo XX, profundizaron en el estudio de la opinión pública. Lowell (1913) sostuvo que, a pesar de la evidente conexión entre política y medios de comunicación, estos últimos tienen limitaciones para gestionar de manera adecuada la información relativa a asuntos de interés público. A su juicio, no pueden ser tratados como instrumentos pertinentes para el funcionamiento de los procesos deliberativos.

La razón de esta limitación radicaría en que los periódicos se orientan principalmente hacia la difusión de contenidos más que a la participación fundamentada en asuntos que conciernen a la mayoría. En otras palabras, su lógica prioriza la circulación de mensajes diseñados para atraer audiencias antes que la construcción de contextos informativos amplios y rigurosos. Como consecuencia, los ciudadanos carecerían de los elementos necesarios para deliberar de manera ponderada sobre, por ejemplo, el rumbo político de un país, las implicaciones de ciertas legislaciones o la conducta de sus gobernantes.

No obstante, la postura de Lowell también fue objeto de críticas debido a su marcada inclinación elitista en materia política. Sostenía que la participación ciudadana debía ser considerablemente limitada, pues era improbable e imposible que la totalidad de

la población dispusiera de los instrumentos informativos y formativos requeridos para comprender la complejidad de los procesos políticos y sociales. De ese modo, la elección de representantes tendería a sustentarse en simplificaciones, generalizaciones y estereotipos; por ello, Lowell defendía la conveniencia de que las decisiones más relevantes fueran delegadas en grupos mejor informados, a los que identificaba con sectores intelectuales o especializados.

En ese contexto de intensa producción académica en ciencias políticas y sociales apareció la obra de Lippmann (1922), que con el tiempo se convertiría en un referente sobre el tema. De acuerdo con él, los medios de comunicación contribuyen a que la información circule de manera fragmentaria. Desde su perspectiva, la imagen de una sociedad que toma decisiones políticas a partir de opiniones plenamente informadas constituye, en gran medida, un ideal difícil de alcanzar. La estratificación social característica de las sociedades modernas, junto con los accesos desiguales a los medios de comunicación y a los recursos educativos, obstaculiza una distribución equitativa de la información. Asimismo, el entorno social contemporáneo es tan vasto y complejo que la mayoría de las personas no dispone ni del tiempo ni de las herramientas cognitivas necesarias para comprenderlo cabalmente.

Al respecto, Lippmann argumentó que, en la práctica, la opinión pública se construye a partir de estereotipos. Estos funcionan como esquemas interpretativos que permiten a los individuos organizar y simplificar la realidad social. En otras palabras, operan como moldes mediante los cuales se percibe aquello que la cultura previamente ha delimitado, facilitando así la aprehensión conceptual de fenómenos que, en una primera instancia, resultan difíciles de comprender.

Por esta razón, al igual que Lowell, Lippmann manifestó una profunda desconfianza hacia las formas de democracia directa. Consideraba que la opinión pública podía variar considerablemente debido a la insuficiencia de información y a la susceptibilidad emocional de los ciudadanos. En este sentido, compartía la idea de que una posible solución consistía en limitar la toma de decisiones políticas y delegarla a una “clase de expertos”. Este grupo (integrado por individuos con formación especializada) tendría la capacidad de interpretar la complejidad del entorno social y orientar la percepción colectiva hacia decisiones más racionales enfocadas en el bienestar general. De ahí que el autor llegara a plantear la necesidad de una élite tecnócrata con la capacidad de articular los intereses de la ciudadanía.

No obstante, revisó críticamente esta postura en su obra posterior (Lippmann, 1925). En este texto reconoció que incluso los expertos podrían ignorar problemas particulares o ser incapaces de captar la totalidad de la vida democrática. En lugar de centrarse exclusivamente en la función de las élites especializadas, Lippmann dirigió entonces su atención hacia el papel de la propaganda en la formación de la opinión pública. A su juicio, los medios de comunicación producen contenidos diseñados deliberadamente para su difusión masiva; dichos contenidos son seleccionados y elaborados por los propietarios de los medios y por diversos líderes de opinión, quienes establecen jerarquías informativas que determinan qué temas reciben mayor visibilidad y cuáles permanecen marginados del debate público.

En relación con este proceso, Cohen formuló posteriormente una observación que se volvería ampliamente citada en los estudios de comunicación política: “puede ser que los medios no tengan éxito la mayor parte del tiempo en decirle a la gente qué pensar, pero tienen un éxito sorprendente en decirle a la gente sobre qué pensar” (1963, p. 13). Esta afirmación subraya el poder de los medios, entre los que ahora ya se encontraban también la televisión y la radio, para estructurar las agendas públicas y guiar la atención colectiva.

Frente a estas limitaciones, Lippmann sostuvo que la mejor vía para contrarrestar los efectos de los estereotipos, la propaganda y la influencia desproporcionada de la “clase de expertos” consistía en fortalecer la educación cívica y el pensamiento crítico de los ciudadanos. Desde esta perspectiva, debían existir herramientas que permitieran comprender los procesos políticos, evaluar críticamente la información disponible y participar de manera más reflexiva en la deliberación pública. En última instancia, Lippmann defendía la posibilidad de que los individuos desarrollaran juicios tanto personales como colectivos mejor fundamentados; para ello, consideraba indispensable reformular el papel de los medios de comunicación, llevándolos hacia la promoción de procesos deliberativos. Estas ideas serían posteriormente tomadas en cuenta para justificar la incorporación de debates políticos en medios durante las campañas electorales.

Tiempo después, los estudios sobre la opinión pública continuaron desarrollándose a lo largo de las primeras décadas del siglo XX, como lo evidencia el trabajo de Dewey (1927). En esta obra, Dewey concibe la opinión pública como un entramado de relaciones sociales que condiciona tanto el acceso a la información como el comportamiento individual. Dichas relaciones se encuentran arraigadas en tradiciones,

prácticas culturales e instituciones sociales, y desempeñan un papel fundamental en los procesos de formación cívica. Desde esta perspectiva, la opinión pública posee un carácter eminentemente social: las posturas respaldadas por una mayoría tienden, progresivamente, a influir en las minorías, generando un efecto que refuerza determinados consensos dentro del espacio público.

En ese mismo periodo se consolidaron diversas técnicas de muestreo, como las encuestas, que contribuyeron a profesionalizar el estudio empírico de la opinión pública en las ciencias políticas y sociales. Estas herramientas permitieron sistematizar la recolección de datos sobre percepciones y actitudes colectivas, ampliando las posibilidades para comprender la circulación social de la información y las formas en que esta incide en la formación de opiniones.

El desarrollo de estas técnicas también permitió reconocer que la opinión pública no constituye un fenómeno plenamente objetivo ni exento de ambigüedades, dado que en su configuración intervienen inevitablemente dimensiones subjetivas. Desde esta perspectiva, las encuestas buscaron medir preferencias individuales y comprender los contextos culturales e históricos que favorecen la emergencia de actitudes y comportamientos compartidos socialmente. En ese sentido, la opinión pública podía interpretarse como el resultado de patrones relativamente uniformes de percepción y valoración entre los ciudadanos, agrupados alrededor de marcos culturales comunes (Allport, 1940).

A partir de este enfoque, numerosos estudios comenzaron a analizar procesos sociohistóricos sustentados en bases culturales compartidas, con el propósito de relacionarlos con las actitudes de la ciudadanía frente a los gobiernos, las políticas públicas o los partidos políticos. Un ejemplo de esto fue el artículo de Converse (1964). Este trabajo marcó un desplazamiento significativo en la agenda de investigación: en lugar de centrarse exclusivamente en el papel de los medios de comunicación, comenzó a examinarse cómo los ciudadanos, mediados por sus trayectorias históricas y culturales, procesan la información disponible al momento de tomar decisiones políticas.

Sostuvo que el simple acceso a la información política no implica necesariamente un interés activo por parte de la ciudadanía. Para describir este fenómeno introdujo el concepto de "Non-Attitudes" (Converse, 1970), entendido como la ausencia de posturas consistentes frente a determinados asuntos públicos, motivada por la apatía o la desafección política. No obstante, esta interpretación fue objeto de críticas por

parte de diversos autores, quienes argumentaron que la aparente inestabilidad de las opiniones respondía a limitaciones metodológicas en el diseño de las encuestas. Así, se sostuvo que era posible identificar patrones relativamente estables de opinión pública e incluso anticipar comportamientos colectivos con cierto grado de precisión, siempre y cuando las preguntas fueran formuladas de manera rigurosa y contextualizada. En este sentido, la opinión pública podía exhibir rasgos de racionalidad, coherencia y estabilidad, aunque sus preferencias se ajustaran progresivamente a la información que circula en la sociedad (Page & Shapiro, 1992). Este periodo estuvo marcado por un intenso debate entre quienes defendían la validez analítica de las encuestas y quienes cuestionaban su capacidad para captar adecuadamente las actitudes políticas de la población.

Con el tiempo y la llegada de los medios masivos de comunicación, el campo de estudio de la opinión pública experimentó una revitalización significativa. El trabajo de Zaller (1992) abrió nuevas líneas de investigación. Propuso que la aparente inconsistencia de las opiniones ciudadanas no se explica por una carencia de ideas, sino por la sobreabundancia de información a la que están expuestos los individuos y por la ausencia de mecanismos eficaces para discriminar entre información verificada y contenidos basados en rumores o conjeturas.

Según este enfoque, gran parte de la información política se acumula en la memoria sin ser necesariamente filtrada o sistematizada. Cuando los individuos son llamados a emitir una opinión, tienden a recurrir a aquellas ideas que han tenido mayor visibilidad o repetición en el espacio público. De este modo, las respuestas no surgen de manera espontánea ni aleatoria: se apoyan en marcos interpretativos previamente disponibles. En consecuencia, la formación de la opinión pública se encuentra influida tanto por los mass media que jerarquizan la información como por la historia sociocultural de los individuos, que condiciona la manera en que la interpretan y recuerdan.

Por su parte, en las décadas de 1980 y 1990 circularon profusamente las ideas de Habermas (1987). El autor planteó que la opinión pública articula, por un lado, un juicio incierto y atravesado por la subjetividad, y, por otro, la reputación que un tema o persona goza a consideración de un grupo social. Dicha reputación puede verse afectada por estrategias de comunicación política, como el marketing, frente a las cuales Habermas subrayó la necesidad de marcos regulatorios que, sin menoscabar el derecho a la información, garanticen condiciones de equidad y transparencia.

Desde esta perspectiva, lo definitorio de la opinión pública no radica en su carácter representativo, entendido como un conjunto aceptado de preferencias, sino en un consenso alcanzado con el diálogo. Habermas rechazaba que lo relevante del término se encuentre en la cantidad de puntos de vista que una población tenga sobre determinado tema; en cambio, propone que se evalúe el entendimiento mutuo posibilitado por el lenguaje. En este sentido, define la opinión pública como una “red para la comunicación de contenidos y tomas de posición [...] que se reproducen a través de la acción comunicativa” (Habermas, 1998, p. 441). Esta red se configura y legitima en función de la calidad de la información disponible y de las condiciones en las que se desarrolla la acción comunicativa, la cual recae en los procesos deliberativos y en los diálogos libres de cualquier presión externa.

En consecuencia, la formación de la opinión pública depende en gran medida de las características institucionales que estructuran la vida social, de los niveles de confianza que estas instituciones suscitan y de la existencia de entornos que favorezcan la deliberación racional y el intercambio dialógico (Habermas, 1987). No obstante, esta concepción se sitúa en tensión con las dinámicas contemporáneas de circulación de la información, en las que las condiciones ideales de la acción comunicativa resultan cada vez más difíciles de sostener.

Con la expansión de internet a partir de la década de 2000, la opinión pública experimentó una transformación significativa: dejó de ser predominantemente receptiva para adquirir un carácter participativo, transitando de una lógica centrada en el consenso hacia otra basada en la interacción. En este contexto, Castells (2009) introdujo el concepto de “autocomunicación de masas” para describir la capacidad de los individuos de producir y difundir sus propios mensajes en entornos digitales, incidiendo potencialmente en la percepción de los otros. Desde esta óptica, la opinión pública habría experimentado un proceso de democratización, al posibilitar la circulación de múltiples perspectivas en espacios como foros, blogs y sitios web.

Estas transformaciones inauguraron el estudio sistemático de la opinión pública en entornos digitales; inicialmente marcado por un cierto optimismo respecto a sus potenciales efectos democratizadores. Se consideraba que la posibilidad de que los individuos difundieran sus opiniones de manera rápida y a gran escala podría fortalecer la deliberación pública y diversificar el debate. Sin embargo, esta visión fue rápidamente matizada por diversas críticas. Se señaló que no todos los discursos en internet poseen el mismo alcance ni visibilidad, ya que los sistemas de búsqueda tien-

den a privilegiar aquellos contenidos con mayores niveles de interacción, reforzando así determinadas narrativas dominantes (Couldry, 2012). También se subrayó que en internet hay poca deliberación; en su lugar, circulan contenidos polémicos o cargados de emociones, lo que va en detrimento de las opiniones basadas en información verificada y en la argumentación fundamentada (Dean, 2005).

En última instancia, la irrupción de las redes sociales terminó de consolidar este desplazamiento, trazando un punto de inflexión en las formas de construcción de la opinión pública. En estos espacios, la producción y circulación de información ya no depende única y exclusivamente de los medios de comunicación tradicionales, sino que se articula a través de múltiples actores y dispositivos. A ello se suman mecanismos como la personalización de contenidos basada en el análisis de datos y su gestión con algoritmos, así como la incorporación reciente del uso de IA para incidir en la percepción ciudadana, lo que ha complejizado aún más los procesos deliberativos y democráticos.

## **5. La opinión pública en la época de la IA generativa**

El uso político y la creciente popularidad de los modelos de IA han introducido transformaciones significativas en la opinión pública de quienes interactúan y se informan en redes sociales. Estas tecnologías permiten la creación y difusión automatizada de contenidos en múltiples formatos, como texto, imagen, audio y video, al tiempo que facilitan altos niveles de personalización y una distribución prácticamente inmediata de los discursos.

En la actualidad, las redes sociales se han consolidado como un espacio central en la configuración de la opinión pública. Diversas estimaciones indican que una proporción significativa de la población en América Latina, cercana al 86 %, participa en dichos espacios (Larocca, 2025). Lejos de constituir meros canales de comunicación, las redes sociales operan como modelos de negocio sustentados en tres pilares interrelacionados: la hiperconectividad de los usuarios, el monitoreo sistemático de su actividad (dataficación) y la gestión de los contenidos a los que acceden (algoritmización) (Saura García, 2025).

Son controladas por un número reducido de grandes corporaciones tecnológicas, que han adquirido un papel central en la formación de la opinión pública contemporánea

(García-Marzá & Calvo, 2024). Su lógica de funcionamiento recae en la hiperconectividad: maximizar el tiempo de permanencia de los usuarios en línea, lo que exige una producción constante de contenido textual y audiovisual. Esta exposición continua ha propiciado fenómenos como la falsificación de identidades, la manipulación y la instrumentalización de la opinión pública (Moore, 2018).

Sostener esta dinámica requiere también de una administración personalizada del contenido, lo cual es posible gracias a la recolección masiva de datos generados por los propios usuarios. La datafización, en este sentido, se refiere al proceso mediante el cual las empresas tecnológicas (y, en ocasiones, gobiernos que establecen acuerdos con ellas) recopilan, gestionan y analizan datos y metadatos derivados de la actividad digital (García-Marzá & Calvo, 2022; Splichal, 2022). La revisión e inspección de los datos permite conocer las actividades de los usuarios, sus comportamientos, gustos, información personal (dónde viven y con quién se relacionan) y anticipar las posibles acciones que tomará en el futuro (Feldstein, 2019; Snowden, 2019; Han, 2022; Farahany, 2023). El procesamiento de toda esa información sirve para la categorización y segmentación de los usuarios, lo que a su vez posibilita la algoritmización: la implementación de sistemas computacionales diseñados para perfilar a los individuos y recomendarles contenidos acordes con sus intereses, hábitos y preferencias. Todo esto para mantener a los usuarios conectados y expuestos a publicidad focalizada, monetización de datos, suscripciones a plataformas, y venta de servicios tecnológicos, entre otros fines.

Este proceso de algoritmización no solo determina qué información resulta visible, sino también qué contenidos permanecen ocultos, reduciendo la exposición a perspectivas divergentes y limitando la pluralidad del debate público (Hao, 2025). En este contexto, las redes sociales ejercen un control cuasi monopólico sobre la circulación de la información. Se trata, en su mayoría, de corporaciones estadounidenses y chinas que concentran tanto el acceso a los datos de millones de usuarios como el control sobre los algoritmos que organizan la visibilidad de los contenidos (Cave et al., 2019). En un entorno donde la desconexión implica, en la práctica, la exclusión de la esfera virtual, la aceptación de los términos y condiciones de uso se vuelve prácticamente ineludible (Zuboff, 2019). En consecuencia, la opinión pública tiende a reducirse a un conjunto de inferencias calculadas mediante sistemas computacionales que procesan y predicen comportamientos.

Al respecto, la irrupción de la IA terminó articulándose también con las dinámicas de hiperconectividad, dataficación y algoritmización propias de las redes sociales, e incidiendo en la formación de la opinión pública. El contenido producido por los modelos de IA suele reforzar dinámicas como la conformación de “cámaras de eco”, entendidas como entornos discursivos en los que los individuos se exponen predominantemente a opiniones afines (Pariser, 2011, Bucher, 2018, Aral, 2021). En tales condiciones, los procesos de formación ciudadana quedan circunscritos a marcos interpretativos homogéneos, lo que limita la confrontación con perspectivas divergentes y debilita las condiciones para una deliberación plural (Coeckelbergh, 2024).

En particular, el contenido que circula en estos entornos puede adoptar formas tanto textuales como audiovisuales, lo cual amplifica su capacidad de persuasión. Estos contenidos funcionan a partir de los LLM y de la IA generativa, cuya convergencia ha facilitado la manipulación de la opinión pública con mecanismos que alteran hechos, actores sociales y acontecimientos políticos.

Los LLM inciden en las redes sociales, en buena medida, a través de agentes conversacionales o chatbots, que han adquirido una presencia creciente en el espacio digital. Estos sistemas pueden ser programados para generar contenidos alineados con objetivos específicos, como posicionar candidaturas, legitimar políticas públicas o desacreditar adversarios, contribuyendo así a intensificar la polarización y a personalizar la exposición informativa. En este sentido, los chatbots operan como herramientas de intervención política, amplificando determinados discursos y modulando la interacción con los usuarios.

En América Latina, se han documentado diversos casos de utilización de “granjas de bots” con fines de influencia política. Por ejemplo, durante el proceso electoral mexicano de 2018, se estimó que una proporción significativa de las cuentas en la plataforma X que seguían a los principales candidatos no correspondía a usuarios reales, sino a sistemas automatizados diseñados para incidir en la percepción pública y generar interacción artificial (Redacción AN, 2018). Estas prácticas incluyen la compra de seguidores falsos y la reproducción automatizada de mensajes favorables, lo que crea una apariencia de apoyo social que no necesariamente se corresponde con la realidad (Pastor-Galindo et al., 2020).

Si bien estos sistemas también pueden desempeñar funciones legítimas, como la automatización de servicios informativos, en numerosos casos actúan como instrumentos de manipulación discursiva. Mediante técnicas de Machine Learning, los chatbots interactúan con usuarios y generan contenidos orientados a objetivos estratégicos, como la difusión de desinformación o la intensificación de la polarización (Ferrara et al., 2016; Yang et al., 2019). Además, su comercialización se inscribe en un mercado escasamente regulado y de carácter transnacional, en el que los operadores trabajan con infraestructuras distribuidas en distintos países y recurren a herramientas como redes privadas virtuales (VPN) para ocultar su localización (Gorodnichenko et al., 2021).

En esa misma línea, es relevante el uso de los LLM en la adecuación estratégica de mensajes políticos, mediante procesos de individualización basados en datos demográficos, conductuales e incluso psicométricos. Esta práctica, conocida como “microtargeting” asistido por IA, supone una transformación respecto de las formas tradicionales de segmentación (Simchon et al. 2024). Mientras que antes los discursos orientados a incidir en la opinión pública se estructuraban a partir de categorías relativamente estables —como el género, la edad o la clase social—, ahora se configuran perfiles altamente individualizados, contruidos a partir de rasgos psicológicos específicos y patrones de comportamiento inferidos por algoritmos.

Esto se aleja del marketing político tradicional y apunta hacia la consolidación de una nueva forma de intervención comunicativa que puede caracterizarse como una tecnología de persuasión automatizada, y que permite la generación masiva de mensajes ajustados a personalidades variadas (Hackenburg & Margetts, 2024; Simchon et al., 2024). Incluso con la evolución de los modelos de IA es posible que en un futuro se prescindiera de equipos humanos dedicados a la comunicación política para que, en su lugar, sea la tecnología la que analice grandes cantidades de información, la procese y despliegue campañas de manera autónoma, apoyándose en los datos extraídos de las interacciones de los usuarios en redes sociales (Saura García, 2025).

Al respecto, un experimento basado en LLM mostró que los mensajes hiperpersonalizados son más persuasivos que los genéricos, observando que incluso pequeñas diferencias de estilo producen cambios cuantificables (Hackenburg & Margetts, 2024). Esto se debe, en parte, a que la persuasión aumenta cuando hay interacción continua. Debido a que la IA envía más de un mensaje y los chatbots pueden responder en tiempo real, modular el tono discursivo, insistir estratégicamente y reformular

argumentos en función de las reacciones del interlocutor, hay menor resistencia a cambiar de opinión (Argyle et al., 2025). En consecuencia, los LLM intensifican el “microtargeting”, transformándolo en un sistema de convencimiento personalizado y automatizado, apoyado en mensajes adaptados a perfiles de individuos y una continua exposición.

En conjunto, estas transformaciones dificultan la configuración de una opinión pública basada en procesos deliberativos y consensuales, y contribuyen a erosionar la confianza ciudadana en los sistemas democráticos. La deliberación tiende a ser sustituida por dinámicas que algunos autores han conceptualizado como una “tiranía de los algoritmos”, caracterizada por la fragmentación del tejido social, la polarización de las opiniones, la personalización extrema de la información y la exclusión de perspectivas disidentes. Este escenario, además, puede profundizar desigualdades estructurales y favorecer la emergencia de fenómenos como “tecnopopulismos” (Bickerton & Accetti, 2021) y “tecnofascismos” (Coeckelbergh, 2026). Los primeros operativizan la división maniquea entre pueblo y élite a través de una legitimación simbólica dada por el saber técnico: gobernantes que presentan a la tecnología como la solución a todos los problemas que dejaron élites corruptas. Mientras tanto, los segundos aprovechan la recolección de datos para vigilar a la ciudadanía y ejercer prácticas de control autoritario, en las cuales se monitorean de manera permanente las acciones y los discursos de las personas.

Ahora bien, además de los LLM, la configuración contemporánea de la opinión pública se encuentra crecientemente atravesada por IA generativa. Esta remite a un conjunto de tecnologías capaces de producir contenido audiovisual que no solo resulta indistinguible del real, sino que además es altamente verosímil (Goldstein et al. 2023). En ese marco, diversos actores políticos instrumentalizan dichas tecnologías con el propósito de intensificar el ruido informacional, mediante la producción masiva de tendencias prefabricadas orientadas a la saturación de las discusiones en redes sociales.

Un fenómeno particularmente ilustrativo es el denominado “astroturfing”, entendido como la simulación de apoyo ciudadano espontáneo que, en realidad, es coordinado mediante algoritmos. Esta práctica se materializa a través de la creación y gestión de cuentas que aparentan pertenecer a individuos reales, pero que son operadas mediante modelos de IA generativa. A diferencia de los chatbots convencionales, incluso aquellos desarrollados a partir de modelos de los LLM, estas entidades

exhiben un manejo lingüístico más sofisticado, así como una capacidad de interacción contextual, coherente y adaptativa que les permite responder de manera diferenciada a los argumentos y a las características de su interlocutor (Elmas, 2023).

En este sentido, el “astroturfing” se fundamenta en la producción de identidades sintéticas: “sujetos” artificiales cuya presencia en redes sociales se construye a partir de simular la identidad digital, adoptando una fotografía de perfil, un nombre propio, una biografía creíble, un historial de publicaciones y una red de contactos. Su rasgo distintivo no radica únicamente en la verosimilitud, sino en su carácter colectivo y coordinado. Lejos de ser usuarios aislados, estas identidades operan como parte de redes articuladas, integradas por cientos o miles de perfiles que amplifican, replican y sincronizan mensajes con el fin de aparentar dinámicas orgánicas de consenso. De este modo, se produce una ilusión de respaldo social que responde a una estrategia deliberada de manipulación de tendencias, visibilidad y alteración de la opinión pública (Doshi et al. 2024).

El objetivo central de estas prácticas es reconfigurar la percepción, haciendo que posiciones marginales o minoritarias adquieran la apariencia de ser ampliamente compartidas. Este simulacro de sobrerrepresentación puede desencadenar un “efecto de arrastre”, en el que individuos reales ajusten sus propias posiciones en función de lo que perciben como una opinión dominante. En México se observó esto cuando ciertos grupos de extrema derecha intentaron (sin éxito) replicar dinámicas de movilización juvenil similares a las registradas en Nepal, particularmente aquellas asociadas a la llamada generación Z, con la apropiación de símbolos provenientes del anime One Piece. Así, el “astroturfing” puede ser visto como una modalidad específica de intervención en la opinión pública mediada por la IA generativa, en la cual la participación ciudadana orgánica es simulada para fabricar consenso, manipular la visibilidad y, en última instancia, reconfigurar las percepciones colectivas (García-Estévez, 2025).

Por ello, más que persuadir en un sentido clásico, estas estrategias parecen orientarse a erosionar las condiciones mismas de posibilidad de la deliberación pública, configurando un entorno en el que la distinción entre verdad y falsedad parece ser irrelevante, mientras que la visibilidad y la capacidad de captar la atención de las personas se constituyen como los principales vectores de incidencia (Hao, 2025). Así, cuando un determinado tema adquiere centralidad en la agenda pública, puede neutralizarse mediante la generación automatizada de contenidos contrarios o polémicos, lo que

desplaza la atención colectiva de asuntos potencialmente perjudiciales para determinados intereses políticos, estatales o corporativos (Coeckelbergh, 2024; García-Marzá & Calvo, 2024).

En este contexto, el acelerado perfeccionamiento de los modelos de IA ha dado lugar a una reducción sustantiva de los costos asociados a la producción de contenido sintético hiperrealista, hasta el punto de que su generación puede depender únicamente de la formulación de instrucciones textuales mínimas (prompts) (Feldstein, 2023; Meikle, 2023; Woolley, 2023). Esta transformación tecnológica amplía el acceso a dichas herramientas, pero también reconfigura la opinión pública, en la medida en que habilita la producción, modulación y circulación de actores, eventos y contextos políticamente relevantes sin estar anclados en referentes reales. En otras palabras, la IA generativa no se limita a intervenir en la representación de la realidad, sino que participa activamente en su (re)construcción simbólica. Su rápida difusión se explica, en buena medida, por la democratización de su acceso, que ha multiplicado el número de agentes capaces de generar y distribuir contenidos audiovisuales a gran escala, incidiendo de manera directa en la formación de percepciones y disposiciones políticas. A su vez, el escalamiento técnico de estos modelos ha incrementado tanto la calidad como la coherencia de los contenidos producidos, al tiempo que ha reducido sus márgenes de error mediante el análisis eficiente de grandes volúmenes de datos y la identificación de patrones recurrentes. Como consecuencia, el contenido sintético se vuelve cada vez más difícil de detectar, especialmente para usuarios no especializados, lo que amplifica su potencial de incidencia en la opinión pública (Goldstein et al., 2023).

En redes sociales, esta dinámica es explotada por una pluralidad de actores (partidos políticos, élites económicas y diversos poderes fácticos) que encuentran en estas tecnologías un instrumento eficaz para intervenir en los procesos de formación de la opinión pública; en abierta tensión con la concepción clásica que la asocia con prácticas deliberativas sustentadas en hechos verificables, la opinión pública derivada de la IA se reduce a un cálculo estadístico. En este sentido, la bibliografía reciente sugiere que la capacidad de los individuos para distinguir entre rostros reales y generados artificialmente se ha deteriorado de manera significativa (Nightingale & Farid, 2022). Esta dificultad plantea un problema que introduce una fractura en la opinión pública clásica, al socavar los criterios mediante los cuales es posible discriminar entre información auténtica y contenido sintético: incluso los propios

modelos de IA presentan limitaciones para efectuar dicha distinción, lo que sugiere la emergencia de un entorno en redes sociales caracterizado por altos niveles de desinformación (Kreps & Kriner, 2023; Hyska, 2025).

Es en este escenario donde se vuelve plausible la proliferación de los llamados “candidatos virtuales”, esto es, entidades generadas por algoritmos que simulan la existencia de sujetos políticos. Se trata de avatares diseñados con atributos propios de la agencia humana (rostro, voz, discurso e interactividad), lo que les permite participar activamente en la comunicación política, emitir posicionamientos e incluso interactuar con audiencias en tiempo real. Desde una perspectiva teórica, ello supone una reconfiguración profunda de la noción de sujeto político, que deja de concebirse en términos antropocéntricos (o humanocéntricos) para desembocar en un ensamblaje sociotécnico. Desde otra arista, la IA sintética posibilita también la emergencia de los llamados “dobles digitales”, avatares artificiales de individuos reales que son utilizados estratégicamente para establecer interacciones personalizadas con el electorado y optimizar la gestión de la imagen pública. Ambos fenómenos, “candidatos virtuales” y “dobles digitales”, se inscriben en una lógica de producción de contenido altamente personalizado, basada en la explotación de datos de los usuarios y su procesamiento mediante sistemas algorítmicos. Este tipo de comunicación política no solo intensifica la segmentación de audiencias, sino que también abre la puerta a la amplificación estratégica de discursos falsos o distorsionados, orientados a la difusión de agendas ideológicas específicas (Woolley, 2023).

En este sentido, la creciente preeminencia del contenido sintético sobre el contenido real ha favorecido la expansión de un fenómeno comunicativo paradigmático de la era de la IA: los deepfakes, cuya capacidad para erosionar los regímenes tradicionales de verdad plantea desafíos para la noción de opinión pública y para el funcionamiento de las democracias contemporáneas.

El término deepfakes surge de la combinación de deep learning y fake, y se refiere a contenidos audiovisuales generados o manipulados mediante IA generativa con el propósito de aparentar autenticidad. Además de apoyarse en el deep learning, también lo hacen en arquitecturas más complejas como las redes generativas —en particular, las redes generativas antagónicas (GAN)— y los modelos multimodales, capaces de integrar y sincronizar información visual, auditiva y textual (Meikle, 2023). No obstante, resulta pertinente subrayar que no todas las deepfakes implican una falsificación total del contenido; más bien, operan a través de distintos grados de

intervención y manipulación, orientados estratégicamente a incidir en la opinión pública. Entre sus usos más recurrentes se encuentra la suplantación de identidad (mediante la modificación de rasgos faciales y vocales), así como la simulación de escenarios que representan eventos inexistentes (Liu et al. 2025).

Las deepfakes pueden clasificarse en función de los recursos técnicos que emplean y del tipo de contenido que producen. En el caso de las deepfakes de video, se ha avanzado significativamente en técnicas de sustitución facial y sincronización labial; sin embargo, persisten limitaciones en la recreación anatómica integral, lo que ocasionalmente se traduce en inconsistencias visuales detectables (manos con más dedos de los posibles, por ejemplo). Por su parte, las deepfakes de audio permiten la clonación de voces con altos niveles de fidelidad, lo que ha sido utilizado para difundir discursos confusos con potenciales implicaciones políticas y diplomáticas. En cuanto a las deepfakes de imágenes, estas suelen representar escenas visuales de eventos improbables o inexistentes, como la detención de líderes políticos, y circulan ampliamente en redes sociales como si se tratara de evidencia comprobada (Meikle, 2023, Liu et al. 2025).

Las implicaciones de estas tecnologías para la opinión pública son profundas. Las deepfakes contribuyen a la distorsión de la realidad, interviniendo en la percepción que conlleva la formación de criterios políticos. Ello se explica, en parte, debido a que los estímulos visuales y auditivos suelen tener mayor impacto que los textuales: en términos generales, los individuos tienden a interpretar lo que ven y escuchan como evidencia directa, mientras que lo que leen se analiza detenidamente y se duda de ello (Raj, 2025). Esto no implica que lo textual carezca de relevancia, sino que en redes sociales los formatos audiovisuales operan como dispositivos de validación simbólica que pueden ser usados para dañar reputaciones, fabricar escándalos o reconfigurar narrativas públicas (Meikle, 2023). En esta línea, diversos estudios han documentado el uso de deepfakes como herramienta de ataque contra periodistas, activistas y actores opositores, con la intención de desacreditarlos o silenciarlos (Pawelec, 2022).

El impacto de estas prácticas en la opinión pública se manifiesta, entre otros aspectos, en la intensificación de respuestas emocionales frente a eventos que, en muchos casos, nunca han ocurrido. Este fenómeno tiende a erosionar el diálogo y el consenso, al tiempo que incrementa el ruido informativo característico de las redes sociales (Liu et al. 2025). Asimismo, la producción y circulación de deepfakes se agrava en contextos electorales, donde su capacidad de incidir en la percepción de los votantes es más

significativa (Meikle, 2023). Dado que los procesos de verificación suelen requerir tiempo, los daños a la reputación persisten en la memoria de los votantes, influyendo en sus decisiones incluso después de que el contenido haya sido desmentido (Ranka et al., 2024).

A este escenario se suma un fenómeno particularmente relevante: el denominado “liar’s dividend”. Este concepto alude a la capacidad de los actores políticos para beneficiarse de la existencia de deepfakes con el fin de eludir mecanismos de rendición de cuentas, desestimando información verificada bajo el argumento de que se trata de contenido manipulado. En un entorno saturado de producciones sintéticas, donde la distinción entre hechos y ficciones se torna cada vez más difusa, se abre un espacio propicio para la negación sistemática de la evidencia. Así, la incertidumbre se convierte en un recurso estratégico: si todo puede ser falso, todo puede ser negado. En este contexto, la opinión pública se moviliza en beneficio de determinados actores políticos, cuyos seguidores interpretan las acusaciones o evidencias en su contra como parte de campañas negras, lo que incluso puede reforzar su apoyo, convirtiéndolos en víctimas. Diversos estudios sugieren que esta estrategia resulta más eficaz que las respuestas tradicionales como el silencio o la disculpa pública, en la medida en que contribuye a la polarización e intensifica las disputas en redes sociales (Jackson et al., 2025).

En conjunto, los LLM y la IA sintética no se limitan, como los medios de comunicación clásicos, a incidir en la opinión pública mediante la difusión de contenidos, sino que operan a un nivel más profundo, orientado a la saturación de información. Su lógica consiste en sustituir la información producida por actores humanos e inundar las redes sociales con contenidos sintéticos textuales y audiovisuales que favorecen la emergencia de realidades alternativas y discursos altamente personalizados y polarizantes. Es una estrategia en la cual se menoscaban los hechos y las realidades que sirven para contextualizar las decisiones políticas y apuntalar la opinión pública; en cambio, se presentan contenidos elaborados a partir de IA por gobiernos, actores políticos o poderes fácticos con la intermediación de las empresas tecnológicas. Esta saturación de información sirve para construir una falsa creencia en torno al desarrollo de un suceso o para trivializar la realidad política bajo la premisa de que todos los discursos son iguales y ameritan el mismo tratamiento.

En este sentido, dinámicas como la hiperconectividad, la dataficación y la algoritmización configuran un escenario en el que la visibilidad de los contenidos (y, por ende, su capacidad de incidencia) puede ser amplificada o reducida artificialmente, reconfigurando los procesos de formación de la opinión pública.

Al respecto, la comparación entre las formas tradicionales de incidencia en la opinión pública y sus cambios con la llegada de la IA generativa queda resumida en la Tabla 1.

Tabla 1: Comparación en las formas en que se incide en la opinión pública (elaboración propia, 2026)

	Etapa clásica de la opinión pública	Opinión pública en la era de la IA generativa
Formación de la opinión pública	Basada en estereotipos, marcos culturales y exposición mediática	Basada en perfiles algorítmicos, datos conductuales y modelado predictivo
Actores principales	Medios de comunicación tradicionales (prensa, radio, TV), élites políticas e intelectuales	Plataformas digitales, corporaciones tecnológicas, algoritmos, IA generativa, bots y actores automatizados
Rol de los medios	Intermediarios que seleccionan y jerarquizan la información	Infraestructura algorítmica que no solo distribuye, sino que produce, personaliza y oculta información
Lógica de funcionamiento	Difusión masiva relativamente homogénea	Hiperconectividad + dataficación + algoritmización = segmentación extrema y personalización
Deliberación pública	Idealmente, basado en el diálogo racional	Erosión de la deliberación: polarización, cámaras de eco y sustitución por interacción automatizada
Producción de contenido	Centralizada (medios y élites)	Descentralizada pero automatizada: IA genera texto, imagen, audio y video
Velocidad y escala	Relativamente lenta y limitada	Inmediata, masiva y continua, con producción automatizada 24/7
Mecanismos de influencia	Propaganda en medios	Microtargeting, astroturfing, bots, deepfakes, hiperpersonalización
Persuasión	Generalista (mensajes para audiencias amplias)	Individualizada y adaptativa (mensajes diseñados para perfiles psicométricos específicos)
Tipo de información dominante	Información mediada y parcialmente verificada	Contenido sintético y potencialmente indistinguible de lo real
Riesgos principales	Manipulación mediática, propaganda, falta de información	Desinformación masiva, simulación de consenso, polarización extrema, control algorítmico
Fenómenos emergentes clave	Estereotipos, agendas preestablecidas, nonattitudes	Deepfakes, astroturfing, candidatos virtuales, liar's dividend, tecnopopulismo, tecnofascismo
Posible definición de opinión pública	Resultado de interacción entre medios, cultura y deliberación	Resultado de cálculos algorítmicos, simulación discursiva y manipulación de visibilidad

Fuente: Elaboración propia.

## 6. Discusión

En los artículos académicos revisados para esta investigación, destaca la necesidad de que los gobiernos latinoamericanos adopten marcos regulatorios orientados a establecer lineamientos éticos para el uso de los LLM y de la IA generativa en redes sociales. Asimismo, se analizan modelos entrenados con datos de usuarios latinoamericanos con el propósito de impulsar proyectos regionales. De manera transversal, se advierte el riesgo de incurrir en un determinismo tecnológico en el que las innovaciones y los desarrollos terminen por suplantar o desplazar la agencia social. Finalmente, las investigaciones alertan sobre los posibles daños a las democracias latinoamericanas, las cuales, incluso antes de la incidencia de la IA en la opinión pública, ya enfrentaban problemas como liderazgos débiles, disputas por la representatividad legislativa y desconfianza en las instituciones.

En este contexto, la opinión pública en la era de la IA supone la coexistencia de múltiples contenidos que dificultan la conformación de procesos deliberativos y consensuados, a partir de los cuales las personas puedan construir criterios sustentados en información verificable y trazable. Fenómenos como las cámaras de eco, el astroturfing, los candidatos virtuales, los deepfakes y el liar's dividend hacen necesario que los gobiernos establezcan legislaciones que regulen el contenido generado por IA que circula en redes sociales. Dado el carácter confidencial de los algoritmos, protegidos como secretos corporativos, la IA puede derivar en sistemas opacos; en consecuencia, resulta imprescindible establecer mecanismos de rendición de cuentas y de participación ciudadana (Ariza & Ruiz, 2025). De este modo, se podrían mitigar los daños en la opinión pública mediante la catalogación de videos, audios, imágenes y textos bajo la leyenda "hecho con IA" y la identificación de los datos con los que se entrenaron los modelos, así como al permitir que los posibles afectados inicien procesos legales contra las corporaciones que faciliten la desinformación. Esto abre la puerta a una gobernanza de la IA en América Latina en la que destaquen la transparencia, la responsabilidad y la protección de los derechos de los usuarios (Estrada, 2025).

De acuerdo con el Centro de Estudios Estratégicos de Derecho de la Inteligencia Artificial (CEEDIA), en América Latina se han registrado intentos por regular la IA, cuyas orientaciones privilegian enfoques éticos basados en la protección de datos

personales, en consonancia con los derechos humanos, y la supervisión humana de los sistemas de IA, en oposición a la automatización total (Criado, 2024). En este sentido, en México aún no existe una política pública integral que regule la IA; no obstante, se han producido avances a partir de iniciativas como la “Agenda Nacional Mexicana de Inteligencia Artificial (2020)”, que contempla la protección de datos y se enfoca en los derechos digitales. En Brasil destaca la “Estratégia Brasileira de Inteligência Artificial”, orientada a fortalecer principios éticos de gobernanza para promover la innovación y el desarrollo en sectores estratégicos. Por su parte, en Bolivia, la CEEDIA no documenta iniciativas consolidadas de regulación; sin embargo, existe el Proyecto de Ley N.º 558/2024-2025, “Ley General de Inteligencia Artificial”, presentado a inicios de 2026, cuyo objetivo es proteger a la ciudadanía frente a la automatización y al uso indebido de datos personales.

No obstante, pese a estas iniciativas, persisten desafíos significativos, como la debilidad institucional frente al poder de las grandes empresas tecnológicas, la brecha digital que amplifica las desigualdades en el acceso a la información y la dependencia tecnológica respecto de modelos de IA desarrollados fuera de la región. Así, aunque el riesgo de menoscabar la opinión pública ha impulsado una voluntad generalizada en los países latinoamericanos por diseñar marcos normativos, todavía se observa una falta de consenso y de claridad en torno a los puntos de partida (OECD, 2022, CEPAL, 2023).

De manera paralela a la regulación de los modelos de IA, algunos gobiernos con rasgos progresistas, como el México de la presidenta Sheinbaum o el Chile de Boric, han impulsado iniciativas tecnológicas orientadas a contrarrestar el peso de los monopolios. Con este propósito, se han desarrollado modelos de IA entrenados con datos de usuarios latinoamericanos. Aunque estos proyectos no cuentan con el mismo nivel de financiamiento que sus contrapartes estadounidenses, ni procesan volúmenes comparables de información (Carrasco-Ahen et al., 2025), buscan limitar el predominio del inglés como lengua hegemónica en el entrenamiento de los modelos de IA. Al mismo tiempo, promueven el uso de datos generados en español y en lenguas originarias, bajo el supuesto de que ello permitirá obtener respuestas más plurales y representativas de las realidades latinoamericanas y de su expresividad cultural (Mora-Reyes et al., 2025). De este modo, el contenido generado por IA que incide en la opinión pública incorporaría matices propios de América Latina, incluyendo dialectos, culturas indígenas y diversas cosmovisiones. Asimismo, los

modelos regionales constituyen un intento por alcanzar la soberanía tecnológica; en este sentido, puede mencionarse el caso de Latam-GPT (Lagos, 2025), una IA de código abierto, susceptible de ser compartida y modificada libremente, que se desarrolla en Chile en el Centro Nacional de Inteligencia Artificial (CENIA).

Pese a lo anterior, todavía no existe un proyecto regional articulado; lo que hay son iniciativas nacionales que responden a condiciones estructurales específicas. Así, el Índice Latinoamericano de Inteligencia Artificial señala la presencia de una infraestructura tecnológica débil y rezagos en la inversión, lo cual dificulta la consolidación de una “IA latinoamericana”. A ello se suma la ausencia de una matriz tecnológica robusta, lo que limita la contratación de especialistas en programación de IA, quienes con frecuencia deben migrar a otros países. En contraste, prevalecen formas de trabajo precario centradas en el etiquetado de datos, lo que lleva a que la opinión pública mediada por los escasos modelos de IA en América Latina tienda a construirse sobre condiciones laborales desiguales y en el marco de esfuerzos aún incipientes por superar la dependencia tecnológica (Jung & Katz, 2025).

A esta dependencia debe sumarse el riesgo de incurrir en determinismos tecnológicos. Ningún modelo de IA opera en abstracto ni de manera neutral. Estos desarrollos se sustentan en una materialidad concreta: microprocesadores, centros de datos, recursos energéticos y trabajo humano, entre otros. La tecnología no se produce por sí misma, sino que es el resultado de decisiones humanas y de relaciones de poder situadas en contextos sociohistóricos específicos (López del Castillo, 2024). Su naturalización oculta la participación de gobiernos y empresas, y constituye, además, una estrategia comercial mediante la cual los errores se presentan como “alucinaciones” y los procesos estadísticos como “razonamientos”, alimentando narrativas mediáticas cuyo propósito es sobredimensionar las capacidades de los modelos de IA para posicionarlos en el mercado.

Las decisiones humanas involucradas en la programación de algoritmos inciden directamente en la opinión pública, en la medida en que persiguen una “pretensión de verdad” que hace pasar contenidos generados artificialmente, y atravesados por diversos sesgos, como verdades sustentadas en evidencia (Marda, 2019). Los artículos revisados para esta investigación señalan que el determinismo tecnológico promueve una visión acrítica que menoscaba la agencia social y conduce a concebir la automatización como un proceso inevitable e ineludible. Ello deriva en una despolitización que presenta a los usuarios como actores pasivos, desprovistos de capacidad de trans-

formación (Guanche et al., 2025). Asimismo, atribuye a la tecnología una autonomía de la que carece, encubriendo los entramados sociales, económicos y políticos que condicionan su accionar.

La dimensión política resulta fundamental para comprender los cambios que ha experimentado la opinión pública a raíz de la irrupción de la IA generativa. En América Latina, las democracias han registrado una intensificación de la desinformación durante las campañas electorales, lo que ha alterado el debate público y favorecido la circulación de contenidos no verificados (CEPAL, 2023; López, 2023). Un informe del International Institute for Democracy and Electoral Assistance (IDEA) señala que la IA exacerba la baja confianza institucional y la fragmentación en las percepciones ciudadanas sobre sus gobernantes, lo que deriva en una débil credibilidad de los líderes políticos y de las estructuras parlamentarias. Asimismo, la asimetría de poder entre las grandes empresas tecnológicas globales y gobiernos latinoamericanos con escasa cohesión institucional genera un entorno propicio para la acción de lobbies, incrementando la influencia política de las corporaciones y la ausencia de marcos normativos efectivamente vinculantes (Becker Castellaro et al., 2025).

En este contexto, los gobiernos de la región se enfrentan a un desfase estructural: por un lado, una acelerada evolución tecnológica y, por otro, capacidades institucionales limitadas para responder de manera ética y responsable (OECD, 2022). Finalmente, las democracias latinoamericanas pueden verse afectadas por prácticas de vigilancia y monitoreo social, en las que los gobiernos contratan servicios de IA (como el reconocimiento facial o el análisis de comportamiento) para el control de la ciudadanía (Ibarreche et al., 2025), así como por el acceso desigual a estas tecnologías, cuyo uso se concentra en sectores con mayores niveles de educación e ingreso.

Todo lo anterior pone de manifiesto los riesgos que enfrenta la región. La IA complejiza un escenario previamente marcado por estructuras de desigualdad, debilidad institucional y dependencia tecnológica. En este sentido, la bibliografía revisada evidencia que la estabilidad de los procesos electorales y la calidad del debate público, lejos de mejorar, se enfrentan actualmente a desafíos más profundos, que transforman de manera sustantiva la forma en que se ha construido la opinión pública.

## 7. Conclusiones

El presente artículo desarrolló una revisión bibliográfica sobre los principales mecanismos mediante los cuales se incide en la opinión pública, con especial énfasis en las tecnologías de IA. La manipulación de percepciones y posturas a través de técnicas aplicadas a contenidos textuales y audiovisuales entraña riesgos significativos para las democracias. En América Latina, en particular, los procesos deliberativos y de construcción de consensos enfrentan desafíos estructurales, como la desconfianza en las instituciones y la fragilidad de los liderazgos políticos. A estos problemas se suman ahora imágenes y videos generados mediante IA que, por su alto grado de verosimilitud, dificultan la distinción entre lo verdadero y lo falso. Asimismo, los chatbots inciden en la opinión pública en la medida en que permiten la creación de cuentas apócrifas o se utilizan para difundir mensajes políticos altamente personalizados.

Este escenario abre la posibilidad de una opinión pública configurada a partir de la mediación de algoritmos y la preponderancia de grandes corporaciones tecnológicas, las cuales desplazan progresivamente a los medios de comunicación tradicionales y se consolidan como actores centrales en un entorno crecientemente influido por la IA.

Lo anterior pone de relieve la necesidad de profundizar en futuras líneas de investigación sobre el papel de estas tecnologías en las ciencias políticas y sociales. Si bien este trabajo buscó esclarecer los mecanismos que configuran la opinión pública en el contexto actual, aún es necesario explorar con mayor detenimiento cómo los Estados pueden afrontar los desafíos de la IA, no solo desde una perspectiva regulatoria, sino también mediante su apropiación estratégica para mejorar su operación. En la actualidad, se observa un proceso de plataformización en el consumo de bienes culturales, como películas y música, en el que la IA desempeña un papel clave en la gestión y automatización de procesos. En este sentido, resulta pertinente preguntarse si los Estados adoptarán estas tecnologías para optimizar su desempeño, por ejemplo, en la gestión burocrática y administrativa.

De igual forma, cabe cuestionar si la IA podría contribuir a la construcción de democracias más inclusivas. Hoy en día, es técnicamente posible crear avatares digitales que tomen decisiones en nombre de los individuos, lo que implicaría una transferencia significativa de la agencia política en sistemas tecnológicos cuyo funcionamiento

aún presenta zonas de opacidad, como ocurre con las denominadas “cajas negras”. En este marco, resulta fundamental distinguir entre el tecnosolucionismo y el tecnopesimismo; es decir, entre una confianza irrestricta en la tecnología y su rechazo absoluto. La adopción de una postura intermedia, crítica y equilibrada permitiría reconocer tanto los beneficios como los riesgos de la IA en el ámbito político. Así, esta tecnología podría emplearse, por ejemplo, para difundir propuestas electorales, siempre que se eviten usos contrarios a principios éticos.

En última instancia, se recomienda abordar la IA desde un marco de responsabilidad que no busque manipular la opinión pública, sino fortalecerla mediante el acceso amplio a información que permita la formación de criterios propios. La implementación de marcos regulatorios sólidos constituye un primer paso indispensable; sin embargo, también resulta urgente promover una alfabetización en IA que permita a los ciudadanos identificar contenidos falsos o generados artificialmente y distinguirlos de información verídica. De este modo, la opinión pública podrá configurarse a partir del conocimiento de diversas posturas y favorecer la toma de decisiones conscientes.

## Agradecimientos

Este artículo pudo escribirse con ayuda de una estancia posdoctoral realizada gracias al Programa de Becas Posdoctorales en la Universidad Nacional Autónoma de México (POSDOC).

## Referencias

Allport, F. H. 1940. “Polls and the Science of Public Opinion.” *The Public Opinion Quarterly* 4 (2): 249-257.

Aral, S. 2021. *The hype machine*. Penguin Random House.

Argyle, L. P., Busby, E. C., Gubler, J. R., Lyman, A., Olcott, J., Pond, J., & Wingate, D. 2025. Testing theories of political persuasion using AI. *Proceedings of the National Academy of Sciences*, 122(18). <https://doi.org/10.1073/pnas.2412815122>

Ariza, L., & Ruiz, C. B. 2025. Inteligencia artificial en el sector público: Enfoques

teóricos y propuestas para una gestión innovadora para América Latina. *Gobierno y Administración Pública*, 10, 119–134. <https://doi.org/10.29393/GP10-9SPLA20009>

Becker Castellaro, S., Carvalho, M., Fernández Gibaja, A., Grassi, A., Hammar, C., Müller, J., Pereira, L., Piaia, V., & Ruediger, M. A. 2025. Inteligencia artificial e integridad de la información: Experiencias latinoamericanas (Documento de política N.º 34). International IDEA. <https://www.idea.int/publications/catalogue/html/inteligencia-artificial-e-integridad-de-la-informacion-experiencias>

Bickerton, C. J., & Invernizzi Accetti, C. 2021. *Technopopulism: The new logic of democratic politics*. Oxford University Press. <https://doi.org/10.1093/oso/9780198807766.001.0001>

Bryce, J. 1888. *The American Commonwealth*. New York, Macmillan.

Bucher, T. 2018. *If... then: Algorithmic power and politics*. Oxford University Press.

Carrasco-Ahen, C. J., Castillo Arrosquipa, G., Palacios Ticona, G. S., & Sánchez-Córdova, M. 2025. Producción científica sobre inteligencia artificial en América Latina: Un análisis bibliométrico. *EPISTEMUS*, 19(38). <https://doi.org/10.36790/epistemus.v19i38.457>

Castells, M. 2009. *Comunicación y Poder*. Alianza Editorial.

Cave, D., Hoffman, S., Joske, A., Ryan, F., & Thomas, E. 2019. *Mapping China's technology giants*. The Australian Strategic Policy Institute.

Centro de Estudios Estratégicos de Derecho de la Inteligencia Artificial (CEEDIA). 2026. Inicio. <https://ceedia.org/>

CEPAL. 2023. *Inteligencia artificial en América Latina y el Caribe: desafíos, oportunidades y políticas públicas*. <https://repositorio.cepal.org/handle/11362/69210>

Codina, L. 2020. Cómo hacer revisiones bibliográficas tradicionales o sistemáticas utilizando bases de datos académicas. *Revista ORL*, 11(2), 139–153. <https://doi.org/10.14201/orl.22977>

Coeckelbergh, M. 2024. *Why AI undermines democracy and what to do about it*. Polity Press.

Coeckelbergh, M. 2026. *Technofascism: AI, Big Tech, and the new authoritarianism*.

AI & Society. <https://doi.org/10.1007/s00146-026-02862-9>

Cohen, B. 1963. *The press and foreign policy*. Princeton University Press.

Converse, P. E. 1964. "The Nature of Belief Systems in Mass Publics." *Ideology and Discontent*. New York, The Free Press.

Converse, Philip E. 1970. "Attitudes and Non-Attitudes: Continuation of a Dialogue." *The Quantitative Analysis of Social Problems*. Reading, Addison-Wesley.

Couldry, N. 2012. *Media, Society, World: Social Theory and Digital Media Practice*. London, Polity.

Criado, J. I. 2024. Inteligencia artificial en el sector público latinoamericano: Estudio comparado a partir de la Carta Iberoamericana de Inteligencia Artificial en la Administración Pública. *Revista del CLAD Reforma y Democracia*, 88, 116–143. <https://doi.org/10.69733/clad.ryd.n88.a387>

Dean, J. 2005. Communicative capitalism: Circulation and the foreclosure of politics. *Cultural Politics*, 1(1), 51–74. <https://doi.org/10.2752/174321905778054845>

Dewey, John. 1927. *The Public and its Problems*. New York, Holt

Doshi, J. A., Novacic, I., Fletcher, C., Borges, M., Zhong, E., Marino, M., Gan, J., Mager, S., Sprague, D. N., & Xia, M. 2024. Sleeper social bots: A new generation of AI disinformation bots are already a political threat. arXiv. <https://doi.org/10.48550/arXiv.2408.12603>

Elmas, T. 2023. Analyzing activity and suspension patterns of Twitter bots attacking Turkish Twitter trends by a longitudinal dataset. arXiv. <https://doi.org/10.48550/arXiv.2304.07907>

Estrada Merino, A. 2025. Inteligencia artificial: hacia una gobernanza ética. *Acta Herediana*, 68(2), 33–40. <https://doi.org/10.20453/ah.v68i2.7505>

Farahany, N. A. 2023. *The battle for your brain*. St. Martin's Press.

Feldstein, S. 2019. *The Global Expansion of AI Surveillance*. Carnegie Endowment for International Peace.

Feldstein, S. 2023. The consequences of generative AI for democracy, governance and war. *Survival. Global Politics and Strategy*, 65(5), 117–142. <https://doi.org/10.108>

0/00396338.2023.2261260

Ferrara, E.; Varol, O.; Davis, C.; Menczer, F. y Flammini, A. 2016. "El auge de los bots sociales". *Comunicaciones de la ACM*, 59(7), 96-104.

García-Estévez, N., Ballesteros-Aguayo, L., & Colussi, J. 2025. Desinformación y manipulación de la opinión pública: Una revisión sistemática sobre astroturfing (2004–2024). *Revista de Comunicación*, 24(2), 159–181. <https://doi.org/10.26441/RC24.2-2025-3988>

García-Marzá, D., & Calvo, P. 2022. Democracia algorítmica: ¿un nuevo cambio estructural de la opinión pública? *Isegoría*, 67. <https://doi.org/10.3989/isegoria.2022.67.17>

García-Marzá, D., & Calvo, P. 2024. *Algorithmic democracy: A critical perspective based on deliberative democracy*. Springer.

Goldstein, J. A., Sastry, G., Musser, M., DiResta, R., Gentzel, M., & Sedova, K. 2023. Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations. <https://doi.org/10.48550/arXiv.2301.04246>

Gorodnichenko, Y.; Pham, T. y Talavera, O. 2021. "Social Media, Sentiment and Public Opinions: Evidence from #Brexit and #USElection". *European Economic Review*, 136. <https://doi.org/10.3386/w24631>

Guanche, J. C., Reyes, N., & Guoliang, L. 2025. Inteligencia artificial (IA) y competencias mediáticas, digitales, informacionales y comunicacionales en América Latina y el Caribe: Una perspectiva ética y crítica en la era algorítmica. *Revista Latinoamericana de Comunicación*, 159, 17–36. <https://doi.org/10.16921/chasqui.v1i159.5219>

Habermas, J. 1998. *Facticidad y validez*. Madrid, Trotta.

Habermas, J. 1987. *Teoría de la acción comunicativa*. Madrid, Taurus.

Hackenburg, K., & Margetts, H. 2024. Evaluating the persuasive influence of political microtargeting with large language models. *Proceedings of the National Academy of Sciences*, 121(24). <https://doi.org/10.1073/pnas.2403116121>

Han, B. C. 2022. *Infocracy: Digitalization and the crisis of democracy*. Polity.

Hao, K. 2025. *Empire of AI. Inside the race for total domination*. Penguin.

Hyska, M. 2025. The politics of past and future: Synthetic media, showing, and telling. *Philosophical Studies*, 182(1), 137–158. <https://doi.org/10.1007/s11098-023-02062-x>

Ibarreche, X., Martínez Elebi, C., & Di Lorenzo, C. 2025. Reconocimiento facial y tecnologías de vigilancia en América Latina: Casos, proveedores y dinámicas comerciales. *AlSur*. <https://www.alsur.lat/sites/default/files/2025-07/Reporte%20reconocimiento%20facial.pdf>

Jung, J., & Katz, R. 2025. Impacto económico de la inteligencia artificial en América Latina. Transformación tecnológica y rezago en materia de inversión y capacidades laborales. CEPAL. <https://repositorio.cepal.org/server/api/core/bitstreams/954c64fe-197c-49f9-b118-0707cc6f8910/content>

Kitchenham, B. 2004. Procedures for performing systematic reviews. Keele University. <http://www.inf.ufsc.br/~aldo.vw/kitchenham.pdf>

Kreiss, D., Lawrence, R. G., & McGregor, S. C. 2023. In their own words: Political practitioner accounts of the democratic implications of artificial intelligence. *New Media & Society*. <https://doi.org/10.1177/146144482311605>

Lagos, A. 2025. Latam-GPT: The free, open source, and collaborative AI of Latin America. *WIRED*. <https://www.wired.com/story/latam-gpt-the-free-open-source-and-collaborative-ai-of-latin-america/>

Larocca, N. 2025. Redes sociales llegan a 86 de cada 100 personas en América Latina. *DPL News*. <https://dplnews.com/redes-sociales-llegan-a-86-de-cada-100-personas-en-america-latina/>

Lippmann, W. 1922. *Public Opinion*. New York, Macmillan.

Lippmann, W. 1925. *Phantom Public. A Sequel to "Public Opinion"*. Nueva York.

Liu, Y., Padmanabhan, B., & Viswanathan, S. 2025. What exactly is a deepfake? *arXiv*. <https://doi.org/10.48550/arXiv.2510.22128>

López del Castillo, F. L. 2024. Inteligencia artificial generativa: determinismo tecnológico o artefacto construido socialmente. *Palabra Clave*, 27(1). <https://doi.org/10.5294/pacla.2024.27.1.9>

López López, P. C., Maldonado, A. M., & Ribeiro, V. 2023. La desinformación en las

democracias de América Latina y la península ibérica. *Uru: Revista de Comunicación y Cultura*. <https://doi.org/10.32719/26312514.2023.8.5>

Lowell, A. L. 1913. *Public Opinion and Popular Government*. New York, Longmans Green.

Marda, V. 2019. *Inteligencia artificial: una perspectiva desde el Sur. América Latina en Movimiento*. <https://alainet.org/es/articulo/203490>

McLeod, S. 2024. *Doing a scoping review: A practical, step-by-step guide*. ResearchGate. <https://doi.org/10.13140/RG.2.2.23703.79528>

Meikle, G. 2023. *Deepfakes*. Polity Press.

Moore, M. 2018. *Democracy hacked: How technology is destabilising global politics*. Oneworld.

Mora-Reyes, B. A., Drewyor, J. A., & Reyes-Angulo, A. A. 2025. *Advancing equitable AI: Evaluating cultural expressiveness in LLMs for Latin American contexts*. arXiv. <https://doi.org/10.48550/arXiv.2511.04090>

Nightingale, S. J., & Farid, H. 2022. *AI-synthesized faces are indistinguishable from real faces and more trustworthy*. *Proceedings of the National Academy of Sciences*, 119(8), e2120481119. <https://doi.org/10.1073/pnas.2120481119>

OECD. 2022. *Uso estratégico y responsable de la inteligencia artificial en el sector público de América Latina y el Caribe*. [https://www.oecd.org/es/publications/uso-estrategico-y-responsable-de-la-inteligencia-artificial-en-el-sector-publico-de-america-latina-y-el-caribe\\_5b189cb4-es/full-report/component-6.html](https://www.oecd.org/es/publications/uso-estrategico-y-responsable-de-la-inteligencia-artificial-en-el-sector-publico-de-america-latina-y-el-caribe_5b189cb4-es/full-report/component-6.html)

Okoli, C., & Schabram, K. 2010. *A Guide to Conducting a Systematic Literature Review of Information Systems Research*. *Communications of the Association for Information Systems*, 37(43), 879-910. <http://dx.doi.org/10.2139/ssrn.1954824>

Page, B. I., Robert Y. Shapiro. 1992. *The Rational Public. Fifty Years of Trends in Americans Policy Preferences*. Chicago, The University of Chicago Press.

Pariser, E. 2011. *The filter bubble: how the new personalized web is changing what we read and how we think*. Penguin Books.

Pastor-Galindo, J., Zago, M., Nespoli, P., López Bernal, S., Huertas Celdrán, A., Gil Pérez, M., Ruipérez-Valiente, J. A., Martínez Pérez, G., & Gómez Mármol, F. 2020. Spotting political social bots in Twitter: A use case of the 2019 Spanish general election. arXiv. <https://doi.org/10.48550/arXiv.2004.00931>

Pawelec, M. 2022. Deepfakes and democracy (theory): How synthetic audio-visual media for disinformation and hate speech threaten core democratic functions. *Digital Society*, 1(2), 19. <https://doi.org/10.1007/s44206-022-00010-6>

Raj, P. 2025. The impact of deepfake technology on elections and democratic trust. *Indian Journal of Law*, 3(6), 56–64. <https://doi.org/10.36676/ijl.v3.i6.137>

Ranka, H., Surana, M., Kothari, N., Pariawala, V., Banerjee, P., Surve, A., Sankepally, S. R., Jain, R., Lalwani, J., & Mehta, S. S. 2024. Examining the implications of deepfakes for election integrity. arXiv. <https://doi.org/10.48550/arXiv.2406.14290>

Redacción AN. 2018, junio 20. Los bots y su influencia en la elección presidencial mexicana: Albert-László Barabási. *Aristegui Noticias*. <https://aristeguinioticias.com/2006/mexico/los-bots-y-su-influencia-en-la-eleccion-presidencial-mexicana-albert-laszlo-barabasi/>

Saura García, C. 2025. Synthetification of public opinion: Impacts on deliberative democracies. *Ethics and Information Technology*, 27(4), 1–14. <https://doi.org/10.1007/s10676-025-09870-1>

Schiff, K. J., Schiff, D. S., & Bueno, N. S. 2025. The liar's dividend: Can politicians claim misinformation to evade accountability? *American Political Science Review*, 119(1), 71–90. <https://doi.org/10.1017/S0003055423001454>

Simchon, A., Edwards, M., & Lewandowsky, S. 2024. The persuasive effects of political microtargeting in the age of generative artificial intelligence. *PNAS Nexus*, 3(2). <https://doi.org/10.1093/pnasnexus/pgae035>

Snowden, E. 2019. *Permanent record*. Metropolitan Books.

Splichal, S. 2022. *Datafication of public opinion and the public sphere: How extraction replaced expression of opinion*. Anthem.

Woolley, S. 2023. *Manufacturing consensus: Understanding propaganda in the era of automation and anonymity*. Yale University Press.

Yang, K.-C.; Varol, O.; Davis, C. A.; Ferrara, E.; Flammini, A. y Menczer, F. 2019. "Arming the public with artificial intelligence to counter social bots". *Human Behavior & Emerging Technologies* 1(1), 48-61. <https://doi.org/10.1002/hbe2.115>

Zaller, J. R. 1992. *The Nature and Origins of Mass Opinion*. Cambridge: Cambridge University Press.

Zuboff, S. 2019. *The age of surveillance capitalism: The fight for The future at The new frontier of power*. Profile Books.

# Cambio técnico y acumulación: inteligencia artificial, trabajo y superpoblación. Europa e Iberoamérica en la División Internacional del Trabajo

*Technical Change and Accumulation: Artificial Intelligence, Labor, and Surplus Population. Europe and Ibero-America in the International Division of Labor*

Nicolás Di Paola <sup>1</sup>

<sup>1</sup> Universidad Nacional de General Sarmiento (UNGS).  
nicolasaxeldipaola@gmail.com (Autor de correspondencia)

**Resumen.** El despliegue masivo de tecnologías de inteligencia artificial constituye una forma históricamente específica del cambio técnico impulsado por la acumulación capitalista. A diferencia de los enfoques dominantes, que aquí se agrupan como neoclásico y neoevolucionista schumpeteriano, y que tratan la tecnología como variable relativamente autónoma, este trabajo parte de la relación social capitalista como determinante tanto del impulso al cambio técnico como de la distribución diferenciada de sus consecuencias. Se argumenta que la inteligencia artificial opera como mecanismo de producción de plusvalor relativo que eleva la composición orgánica del capital y genera, por esa vía, formas diferenciadas de superpoblación relativa —sobrante, latente y consolidada—, cuya determinación concreta depende de la forma específica a través de la cual cada región realiza el contenido universal de la acumulación capitalista. El análisis se aplica a tres regiones que no producen esa tecnología, pero reciben sus efectos de manera diferenciada: Europa occidental, Iberoamérica y la Península Ibérica, tratada como caso bisagra cuya forma específica de acumulación la ubica entre las determinaciones del espacio europeo y las del espacio iberoamericano de acumulación, articulando ambas dinámicas y permitiendo identificar los mecanismos de transmisión entre ellas. Las divergencias empíricas entre regiones, que desde la perspectiva ortodoxa presentan la apariencia de paradojas, resultan contradicciones estructurales necesarias derivadas de esas posiciones diferenciales en la acumulación mundial. El trabajo desarrolla seis determinaciones estructurales, formuladas también como hipótesis contrastables en registro convencional, que permiten evaluar la capacidad explicativa diferencial del desarrollo propuesto frente a las alternativas neoclásica y neoevolucionista.

**Palabras clave:** *inteligencia artificial; cambio técnico; superpoblación relativa; Europa; Iberoamérica; Península Ibérica; acumulación capitalista*

**Abstract.** The massive deployment of artificial intelligence technologies constitutes a historically specific form of technical change driven by capitalist accumulation. Unlike dominant approaches, here grouped as neoclassical and Schumpeterian neo-evolutionist, and which treat technology as a relatively autonomous variable, this article takes the capitalist social relation as the determinant of both the impulse toward technical change and the differentiated distribution of its consequences. It argues that artificial intelligence operates as a mechanism for producing relative surplus value that raises the organic composition of capital and thereby generates differentiated forms of relative surplus population (surplus, latent, and consolidated), whose concrete determination depends on the specific form through which each region realizes the universal content of capitalist accumulation. The analysis is applied to three regions that do not produce this technology but receive its effects in differentiated ways: Western Europe, Ibero-America, and the Iberian Peninsula, treated as a hinge case whose specific form of accumulation places it between the determinations of the European space and those of the Ibero-American space of accumulation, articulating both dynamics and making it possible to identify the mechanisms of transmission between them. The empirical divergences between regions, which from the orthodox perspective appear as paradoxes, are necessary structural contradictions derived from those differential positions in global accumulation. The article develops six structural determinations, also formulated as testable hypotheses in conventional terms, that make it possible to evaluate the differential explanatory capacity of the proposed development compared with the neoclassical and neo-evolutionist alternatives.

**Keywords:** *artificial intelligence; technical change; relative surplus population; Europe; Ibero-America; Iberian Peninsula; capitalist accumulation*

**Historial:** Recibido: 30 de marzo de 2026 | Aceptado: 15 de mayo de 2026

**Ética/Financiamiento:** El autor declara no tener conflictos de interés. La investigación fue autofinanciada y no recibió financiamiento externo.

Este documento está bajo una licencia  
CC BY-NC-ND 4.0.

**Cómo citar (APA):** Di Paola, N. (2026). Cambio técnico y acumulación: inteligencia artificial, trabajo y superpoblación. Europa e Iberoamérica en la División Internacional del Trabajo. *Conocimiento i Política*, 7(1), 37-92. DOI: 10.64480/cip.2026.n7.02.

## 1. Introducción

La generalización acelerada de la inteligencia artificial, que concentra tecnologías tales como modelos de lenguaje de gran escala, sistemas de visión computacional, automatización robótica cognitiva y toma de decisiones algorítmica, ha reabierto con renovada urgencia el debate sobre las relaciones entre cambio técnico, empleo y distribución del ingreso. Ese debate oscila con frecuencia entre dos posiciones simétricamente insatisfactorias. La primera, de carácter tecno-optimista, anticipa un salto civilizatorio generalizado en el que las ganancias de productividad derivadas de la inteligencia artificial se traducirán, a través de los mecanismos de mercado, en bienestar distribuido y en la creación de nuevos empleos que compensarán los desplazados: es la posición que sostienen, con matices, Brynjolfsson y McAfee (2014) y que recogen los informes del McKinsey Global Institute (2023). La segunda, de carácter tecno-pesimista, documenta con mayor rigor empírico la profundización de la desigualdad y el desempleo estructural, pero no ofrece una explicación causal de por qué los beneficios del cambio técnico se concentran sistemáticamente en el capital y se distribuyen de manera tan desigual entre los diferentes espacios nacionales de acumulación: describe los efectos, como hacen Ford (2015) y, en parte, Acemoglu y Restrepo (2019), sin poder derivarlos de las determinaciones estructurales del proceso. Ambas posiciones comparten una limitación analítica de fondo: tratan la tecnología como una variable relativamente autónoma, cuyos efectos se procesan sobre una economía que, en lo esencial, no es sometida a escrutinio. Se abstrae así precisamente lo que debería constituir el punto de partida del análisis: la relación social imperante en el modo de producción capitalista, que determina tanto el impulso al cambio técnico como la forma en que sus consecuencias se distribuyen entre clases y entre los diferentes espacios nacionales de acumulación.

El presente trabajo desarrolla el análisis desde la perspectiva elaborada por Marx en *El Capital*, en particular la ley general de acumulación capitalista y las formas de la superpoblación relativa, extendida en sus determinaciones concretas por Juan Iñigo Carrera (1998, 2007, 2008) y aplicada al análisis del espacio iberoamericano de acumulación por Grinberg (2011, 2013), Starosta (2010), entre otros autores del Centro de Investigación para la Crítica Práctica (CICP). Esta perspectiva se confronta con los desarrollos y predicciones de las dos tradiciones alternativas más influyentes en la materia del cambio técnico y el desarrollo económico, la neoclásica/marginalista y

la neoevolucionista schumpeteriana, y se muestra que aquello que esas tradiciones presentan como resultado contingente o paradójico constituye, desde el desarrollo propuesto, la expresión fenoménica de contradicciones estructurales necesarias desplegadas en la lógica del proceso de acumulación capitalista.

El análisis se presenta frente a tres espacios con posiciones diferenciadas en el proceso mundial de acumulación: Europa occidental, la Península Ibérica e Iberoamérica. Ninguno de ellos es un centro de producción de la tecnología de inteligencia artificial; esa producción está concentrada fundamentalmente en Estados Unidos y, en segundo lugar, en China. Los tres son, sin embargo, grandes mercados de despliegue y, por tanto, territorios en los que las transformaciones del mercado de trabajo, la distribución del ingreso y las condiciones de acumulación local se procesan con intensidad creciente. La Península Ibérica ocupa en este análisis una posición analíticamente privilegiada: España y Portugal son economías integradas en la Unión Europea, sujetas a sus marcos regulatorios y a su proyecto institucional, pero desarrollaron históricamente una acumulación que no convergió orgánicamente con las formaciones nacionales europeas de mayor composición orgánica del capital ni generó capacidades tecnológicas de frontera, mientras mantienen vínculos productivos, migratorios y financieros que las articulan con el espacio iberoamericano. La determinación estructural de esa posición bisagra se desarrolla en la sección correspondiente; lo que aquí importa señalar es que permite ilustrar cómo el mismo proceso de elevación de la composición orgánica del capital adopta formas fenoménicas diferenciadas según la posición de cada espacio en la acumulación mundial.

El artículo se organiza del siguiente modo. La sección 2 examina críticamente las tradiciones neoclásica y neoevolucionista schumpeteriana, identifica sus limitaciones analíticas comunes y desarrolla las determinaciones de la ley general de acumulación capitalista y las formas de la superpoblación relativa, con las elaboraciones de Iñigo Carrera sobre la especificidad de los espacios nacionales de acumulación; cierra con seis determinaciones estructurales formuladas también como hipótesis contrastables en registro convencional. La sección 3 expone la estrategia metodológica y el estatuto del uso de evidencia estadística en el marco del desarrollo propuesto. La sección 4 examina la inteligencia artificial como forma históricamente específica del cambio técnico capitalista. Las secciones 5, 6 y 7 desarrollan las tres regiones. La sección 8 presenta el análisis comparativo y tematiza la unidad del proceso mundial de acumulación del que esas diferencias son expresión. La sección 9 formula las conclusiones.

## **2. El debate sobre la forma y contenido del cambio técnico: neoclásicos, neoschumpeterianos y la crítica de la economía política**

### **2.1. La tradición neoclásica: el debate sobre la tecnología y el mercado distribuidor**

La corriente principal de la teoría económica contemporánea conceptualiza el cambio técnico como un fenómeno que eleva la productividad factorial y permite, en condiciones de competencia, incrementos generales del bienestar. En su versión fundacional, desarrollada en el modelo de crecimiento de Solow (1956), el progreso técnico es un residuo exógeno que desplaza la función de producción agregada hacia arriba, incrementando la productividad del trabajo y del capital. El mercado de trabajo garantiza, mediante ajustes de precios, la absorción de la fuerza de trabajo eventualmente desplazada.

La hipótesis del cambio técnico sesgado hacia las habilidades (SBTC), expuesta en los trabajos de Katz y Murphy (1992) y Autor, Levy y Murnane (2003), introdujo un refinamiento importante: el cambio técnico no es neutral, sino que favorece el trabajo calificado sobre el no calificado, explicando la creciente desigualdad salarial en los países avanzados. A diferencia del residuo exógeno de Solow, en el que la tecnología cae desde afuera de la economía como un dato no explicado, la SBTC incorpora la dimensión de endogeneidad, propia de los modelos de crecimiento al estilo Lucas (1988) y Romer (1986, 1990): la dirección del cambio técnico es producida por la propia dinámica económica, en particular por los precios relativos de los factores y por la estructura de la demanda de habilidades, de modo que la tecnología emerge de los incentivos que los mercados generan y no les es exterior. Esta endogeneidad, sin embargo, permanece acotada al plano de la distribución: explica hacia qué tipo de trabajo se sesga el cambio técnico, pero no la compulsión estructural a innovar ni la apropiación diferencial de sus resultados, que siguen siendo tratados como consecuencias de ajustes de precios en mercados fundamentalmente simétricos. El modelo de tareas de Acemoglu y Restrepo (2018, 2022) añadió un mecanismo de compensación: la automatización destruye puestos en las tareas actuales, pero crea demanda de trabajo en nuevas tareas en las que los humanos tienen ventaja comparativa. La predicción central de este enfoque, a saber, que la compensación laboral tiende a reequilibrar el mercado de trabajo, está empíricamente falsada y, como se verá, no se sostiene en los tres espacios de acumulación analizados.

Las limitaciones de este enfoque son estructurales. En primer lugar, su teoría de la distribución presupone que los salarios reflejan productividades marginales, abstraendo la determinación del valor por el tiempo de trabajo socialmente necesario y, con ella, la explotación como relación constitutiva del proceso productivo. En segundo lugar, la predicción de compensación laboral supone una movilidad de capital y trabajo que no existe en la economía mundial: las nuevas tareas de alto valor creadas por la inteligencia artificial se concentran en los centros de producción tecnológica, no en los territorios de despliegue. En tercer lugar, el marco no puede dar cuenta de las divergencias regionales observadas más que mediante factores institucionales ad hoc.

## **2.2. El neoevolucionismo schumpeteriano: paradigmas tecno-económicos y sus límites**

La tradición neoevolucionista, entre los que se destacan Freeman (1982), Dosi et al. (1988), Perez (2002), Lundvall (1992), representa un avance respecto de la concepción neoclásica al incorporar el cambio técnico como proceso endógeno, path-dependent y estructuralmente transformador. Perez (2002) analiza las revoluciones tecnológicas como paradigmas tecno-económicos que se instalan en dos fases: una fase de instalación, dominada por el capital financiero y caracterizada por burbujas especulativas, y una fase de despliegue, en la que la nueva tecnología se difunde productivamente y puede generar prosperidad generalizada si las instituciones acompañan adecuadamente.

Este enfoque es valioso en su descripción de los patrones históricos de difusión tecnológica y en su énfasis en las complementariedades institucionales. Sin embargo, presenta dos limitaciones analíticas fundamentales. La primera es que trata las crisis distributivas como contingencias institucionales, no como resultados estructurales de la acumulación capitalista: la posibilidad de una “edad dorada” distributiva no es derivada de la lógica del capital sino postulada como resultado de la intervención política correcta. La segunda es que el marco no puede dar cuenta de la diferenciación regional sin recurrir a variables ad hoc como los “sistemas nacionales de innovación”: las diferencias entre Europa, la Península Ibérica e Iberoamérica aparecen como diferencias en las capacidades de los sistemas de innovación, no como formas diferenciadas del mismo proceso de acumulación mundial.

Ambas tradiciones comparten, pese a sus diferencias internas, un mismo punto de partida que determina sus limitaciones: tratan la tecnología como un fenómeno que

opera sobre la economía desde fuera —sea como residuo exógeno (Solow), sesgo de habilidades (Katz-Murphy, Acemoglu-Restrepo) o paradigma tecno-económico (Perez)— y, en consecuencia, sitúan las relaciones de producción capitalistas como el telón de fondo sobre el que se procesan los efectos del cambio técnico, en lugar de identificarlas como la determinación constitutiva de ese cambio. Esta posición metodológica compartida hace que la distribución de los efectos del cambio técnico aparezca como un problema secundario, determinable por precios relativos o intervenciones institucionales, en lugar de un resultado necesario inscripto en la lógica inmanente de la acumulación capitalista. Es precisamente ese punto de partida lo que hace insuficiente cualquier corrección interna a estos marcos y lo que vuelve necesario el punto de partida radicalmente diferente del desarrollo propuesto.

### **2.3. El desarrollo propuesto: la crítica de la economía política y la elaboración de Iñigo Carrera**

#### **2.3.1. El impulso capitalista al cambio técnico**

A diferencia de las tradiciones examinadas, que parten de la tecnología como dato externo para analizar sus efectos sobre una economía constituida previamente, el desarrollo propuesto parte de la relación social capitalista de producción como la determinación que explica tanto el impulso al cambio técnico como la forma necesaria en que sus consecuencias se distribuyen. En la exposición de Marx en *El Capital*, el cambio técnico no es un fenómeno exógeno ni una estrategia empresarial: es una necesidad inmanente a la relación capitalista de producción. El capitalismo se distingue de otros modos de producción por la compulsión sistemática a extraer plusvalor relativo, lo que requiere la constante revolución de las condiciones técnicas de producción, y con ella de las fuerzas productivas del trabajo social. Esta compulsión opera a través de la competencia entre capitales: cada capitalista individual tiene incentivo para introducir tecnología que reduce sus costos por debajo del promedio social, obteniendo una ganancia extraordinaria. Cuando la innovación se generaliza, el nuevo método se convierte en el estándar social y esa ganancia desaparece. Cada capitalista está entonces forzado a innovar nuevamente. Es esta dinámica la que explica el carácter sistémico e incesante del cambio técnico bajo el capitalismo (Marx, 1867/2008, Sección IV).

La maquinaria —y en términos más generales, el capital constante— es el vehículo principal mediante el cual el capital sustituye trabajo vivo por trabajo muerto. Al hacerlo, eleva la composición técnica del capital y, consecuentemente, la composición orgánica del capital (la proporción entre capital constante  $c$  y capital variable  $v$  en términos de valor). El aumento tendencial de la composición orgánica tiene dos consecuencias de primer orden: produce la tendencia decreciente de la tasa de ganancia —que actúa como motor permanente de la búsqueda de nuevas formas de extracción de plusvalor y de nuevos espacios geográficos de valorización— y produce la superpoblación relativa como resultado estructural del proceso de acumulación.

### 2.3.2. La superpoblación relativa y sus formas

Marx define la superpoblación relativa en el capítulo XXV del Tomo I como la población obrera excedente respecto de las necesidades medias de valorización del capital. Su existencia no es una imperfección del mercado de trabajo ni un desequilibrio transitorio: es la condición de funcionamiento del sistema capitalista. El ejército industrial de reserva cumple una función estructural doble: presiona a la baja los salarios del ejército activo y constituye la reserva de fuerza de trabajo disponible para los períodos de expansión.

Marx distingue en ese mismo capítulo tres formas de la superpoblación relativa. La superpoblación sobrante corresponde a los trabajadores expulsados de la producción por el cambio técnico que permanecen vinculados a los centros productivos, alternando entre ocupación y desocupación. La superpoblación latente es la población que existe en sectores no plenamente capitalistas —principalmente la agricultura campesina en transición— y que constituye una reserva potencial disponible para la industria. La superpoblación consolidada son los trabajadores permanentemente incorporados a actividades de baja productividad, ingresos precarios y condiciones degradadas, que conforman el estrato más profundo y estable del ejército de reserva. La contribución de Iñigo Carrera (1998, 2007) consiste en desarrollar estas categorías en sus determinaciones concretas para el espacio iberoamericano de acumulación, mostrando cómo la posición específica de cada formación nacional en la división internacional del trabajo determina la forma predominante que adopta la superpoblación relativa.

### **2.3.3. La elaboración de Iñigo Carrera: especificidad nacional y renta de la tierra**

La contribución central de Iñigo Carrera consiste en desarrollar el método de la crítica de la economía política aplicándolo a las determinaciones concretas de la acumulación capitalista en Argentina y en el espacio iberoamericano. Su argumento fundamental es que la renta diferencial de la tierra constituye la mediación específica a través de la cual Argentina y otros países iberoamericanos se insertan en la división internacional del trabajo. La extraordinaria productividad natural de la tierra pampeana —y, con variantes, de otras regiones iberoamericanas ricas en recursos naturales— permite la producción de mercancías a costos individuales inferiores a su valor de mercado mundial, generando una renta diferencial que apropia una porción del plusvalor producido en escala global (Iñigo Carrera, 2007). Esta mediación orienta la acumulación de capital hacia la exportación de productos primarios, desalienta el desarrollo de una industria nacional competitiva y reproduce la relación de adaptación tecnológica como resultado estructural, no como política deliberada.

Las consecuencias sobre las formas de superpoblación son específicas. En el espacio iberoamericano de acumulación, la superpoblación latente no se convierte, en su mayor parte, en superpoblación sobrante temporalmente absorbida por la industria: se convierte directamente en superpoblación consolidada, dado que la industria local no tiene capacidad de absorber a toda la población desplazada del campo. La estructura de mercado de trabajo resultante es bifurcada: un núcleo de trabajadores formalmente empleados en el sector capitalista moderno y una vasta periferia de trabajadores en condiciones precarias, informales o directamente desocupados.

### **2.3.4. La universalización del capital y la determinación geográfica de la producción tecnológica**

La producción de inteligencia artificial de frontera, tales como modelos de lenguaje de gran escala, chips especializados, plataformas de cómputo en nube, requiere una concentración de capital constante (infraestructura de cómputo, conjuntos de datos de entrenamiento) y de fuerza de trabajo con conciencia productiva ampliada (ingeniería de procesamiento y optimización a gran escala, investigación en aprendizaje automático, matemáticos de alto nivel, etc.) que se acumuló durante décadas en Estados Unidos y, en menor medida, en China. Esta concentración no es contingente, sino que

está estructuralmente determinada. Es la posición de estos espacios nacionales de acumulación como formas de la división internacional del trabajo y de la forma en que la acumulación se realizó históricamente en ellos.

Lo que Rikap (2021) describe como monopolios intelectuales puede precisarse, desde el desarrollo propuesto, en términos más rigurosos: no se trata de una renta de monopolio sobre bienes intangibles en sentido jurídico, sino de una ventaja de composición orgánica que se sostiene porque la magnitud del adelanto de capital requerido para producir modelos de frontera —estimada en decenas o centenares de millones de dólares por ciclo de entrenamiento—, y la concentración de fuerza de trabajo de conciencia científica o subjetividad ampliada crean una barrera de entrada que se renueva continuamente. A diferencia de la renta diferencial de la tierra, cuya fuente es la productividad natural del suelo, un dato físico independiente de la voluntad de los agentes, la ventaja de composición orgánica en la producción de inteligencia artificial no tiene una base física irreproducible. Sin embargo, la necesidad de alcanzar esa ventaja requeriría exactamente el mismo proceso acumulativo que la generó: décadas de inversión masiva en capital constante y en la formación de una fuerza de trabajo técnica específica, bajo condiciones que son precisamente las que la posición de cada región en la división internacional del trabajo no permite reproducir. La posición diferencial que produce no es exterior al proceso de acumulación sino resultado de este. A diferencia de la renta diferencial de la tierra, cuya fuente es una productividad natural que el capital no produce, esta barrera de composición orgánica se renueva en cada ciclo de acumulación y se profundiza en la misma medida en que se despliega la acumulación mundial.

#### **2.4. Determinaciones estructurales e hipótesis de trabajo**

Las determinaciones que se desarrollan a continuación se siguen del análisis expuesto en 2.3 y se formulan en dos registros complementarios. Por un lado, en el registro de la crítica de la economía política, que consiste en que la exposición de los movimientos son determinaciones necesarias que resultan del desenvolvimiento histórico y concreto de la acumulación capitalista en su contenido mundialmente desarrollado. Por otro lado, se presentan las mismas hipótesis en el registro convencional de la investigación social comparada; es decir, que aquello que desarrollamos en su forma de movimiento dialéctico, pueden reformularse como hipótesis contrastables que

permiten evaluar la capacidad explicativa diferencial del desarrollo propuesto frente a las alternativas neoclásica y neoevolucionista.

H1 — Mecanismo general: El despliegue de la inteligencia artificial no es un acontecimiento tecnológico que se despliega sobre una economía preexistente: es la forma más avanzada que toma actualmente el proceso de objetivación del conocimiento social en el proceso productivo por medio de la compra de capital constante. En tanto sustitución de trabajo vivo por trabajo muerto, eleva la composición orgánica del capital en los sectores en que se despliega y produce superpoblación relativa como resultado necesario e inscripto en la lógica del proceso de acumulación. Esta superpoblación no es una fricción de ajuste temporal, sino la condición de funcionamiento del sistema.

H2 — Diferenciación regional: La forma concreta que adopta la superpoblación relativa no es uniforme ni determinada por características locales contingentes: nivel educativo, calidad institucional, políticas de empleo; sino por la posición estructural de cada región en la división internacional del trabajo. La misma presión tecnológica produce superpoblación sobrante en las capas medias de servicios en Europa occidental, profundización de la superpoblación consolidada en Iberoamérica, y una combinación compuesta con especificidades propias en la Península Ibérica.

Las tres determinaciones siguientes especifican, en cada caso, las formas concretas que H2 enuncia en abstracto.

H3 — Europa occidental: La inteligencia artificial opera principalmente sobre la forma sobrante de la superpoblación relativa, pero con una determinación históricamente específica: se extiende hacia capas de la fuerza de trabajo con subjetividad ampliada convirtiéndola en fuerza de trabajo degradada, configurando una forma de superpoblación sobrante sin precedente en los ciclos anteriores de automatización.

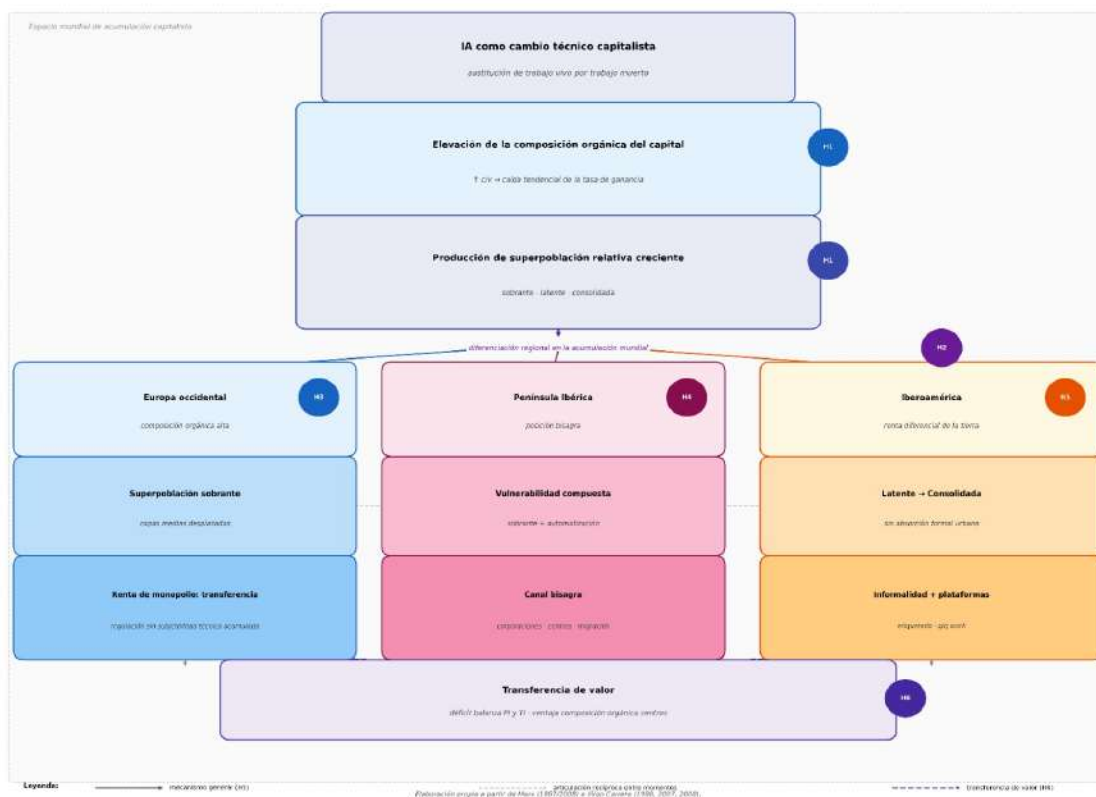
H4 — Península Ibérica: La posición bisagra de España y Portugal produce una vulnerabilidad compuesta: la superpoblación sobrante estructural preexistente, que se expresa en tasas de desempleo estructural y temporalidad contractual persistentemente altas, se superpone con la nueva presión automatizadora, sin que el Estado social ibérico tenga la arquitectura requerida para gestionar ese desplazamiento.

H5 — Iberoamérica: La inteligencia artificial actúa simultáneamente sobre la superpoblación latente, acelerando la expulsión de trabajo rural mediante la automatización

agroindustrial, y sobre la superpoblación consolidada: profundizándola mediante la plataformización del trabajo y la incorporación de trabajadores desplazados al trabajo de etiquetado de datos. La mediación de la renta diferencial de la tierra no genera tendencias estructurales hacia la absorción productiva de los trabajadores desplazados en sectores de alta composición orgánica: el capital se orienta estructuralmente hacia la exportación primaria y no desarrolla una industria diversificada con capacidad de absorción sostenida.

H6 — Transferencia de valor: El déficit sistemático y creciente en la balanza de pagos de servicios de propiedad intelectual y tecnología de la información en las tres regiones analizadas es la expresión fenoménica de una transferencia de valor que resulta de la ventaja de composición orgánica de los centros productores de tecnología. Esta transferencia reproduce y profundiza la posición diferencial de cada región en la división internacional del trabajo, y no puede ser revertida mediante políticas de adopción tecnológica porque no resulta de decisiones de política sino de determinaciones estructurales de la acumulación.

### Diagrama H1–H6. Determinaciones estructurales de la superpoblación relativa frente al despliegue de la IA



Nota: Elaboración propia a partir de Marx (1867/2008) e Iñigo Carrera (1998, 2007, 2008).

Tabla 1: Determinaciones derivadas diferencialmente: enfoques neoclásico, neoevolucionista y desarrollo propuesto

Dimensión	Neoclásico / SBTC	Neoschumpeteriano	Desarrollo Propuesto
Participación salarial	Estable o creciente con productividad	Contingente a instituciones	Decreciente tendencial por elevación de composición orgánica
Diferencias regionales	Convergencia condicional	Path dependency institucional	Diferenciación estructural por posición en acumulación mundial
Forma regional en Europa	Polarización calificados/no calificados	Depende del sistema nacional de innovación	Superpoblación sobrante extendida a capas medias de servicios
Forma regional en Iberoamérica	Convergencia salarial por productividad	Catching-up condicionado a capacidades	Superpoblación consolidada profundizada; transferencia de valor
Dependencia tecnológica	Ventaja comparativa dinámica	Falla de sistema de innovación	Ventaja de composición orgánica; transferencia de valor de largo plazo
Efecto de la regulación	Corrección de fallas de mercado	Condición del despliegue distributivo	No altera determinaciones estructurales de la acumulación

Fuente: Elaboración propia.

### 3. Metodología

#### 3.1. Diseño comparativo y justificación de los casos

El trabajo analiza tres regiones que constituyen momentos diferenciados de un único proceso mundial de acumulación capitalista. La inteligencia artificial no opera sobre ellas como una presión externa uniforme que produce resultados distintos según características locales: opera como la forma más avanzada que adopta actualmente el cambio técnico capitalista, cuyas consecuencias específicas sobre el trabajo son, en cada región, expresiones concretas de su posición en la división internacional del trabajo. El diseño comparativo opera, por tanto, en el plano de la expresión fenoménica: permite contrastar cómo las determinaciones estructurales derivadas teóricamente se manifiestan diferencialmente en las tres regiones, sin que esa comparación sea el procedimiento a través del cual las determinaciones se producen.

La selección de Europa occidental, la Península Ibérica e Iberoamérica no obedece a criterios de muestreo convencional sino a la especificidad de la función que cada región cumple en la acumulación mundial. Europa occidental ocupa una posición históricamente central en esa acumulación, con capacidad regulatoria sobre el despliegue de inteligencia artificial pero sin capacidad productiva de frontera. Iberoamérica ocupa una posición estructuralmente determinada por la mediación de la renta diferencial de la tierra, que orienta la acumulación hacia la exportación de materias primas y desincentiva el desarrollo de la subjetividad técnica específica que la producción de tecnología de frontera requiere. La Península Ibérica se introduce no como un tercer caso simétrico sino como el momento bisagra que articula ambos espacios: España y Portugal están plenamente integrados en las instituciones y mercados de la Unión Europea, pero su estructura productiva no convergió con las formaciones nacionales europeas de mayor composición orgánica del capital, y sus vínculos con Iberoamérica —a través de corporaciones transnacionales, centros de servicios compartidos y flujos migratorios— crean canales de transmisión explícitos entre ambos espacios de acumulación. Su función analítica no es añadir un tercer caso independiente sino ilustrar cómo opera esa transmisión.

La pertinencia de la Península Ibérica como caso bisagra se deriva de tres determinaciones concretas. En primer lugar, la integración plena a la Unión Europea permite aislar el efecto de las diferencias estructurales internas al espacio europeo, dado que España y Portugal operan bajo el mismo marco regulatorio e institucional que el núcleo de la UE. En segundo lugar, los vínculos productivos, financieros y migratorios con Iberoamérica crean canales de transmisión verificables empíricamente, lo que permite examinar cómo la presión automatizadora se propaga entre espacios de acumulación con posiciones estructuralmente distintas. En tercer lugar, la afinidad lingüística e histórica entre España, Portugal y los países iberoamericanos no constituye un criterio de selección cultural sino la mediación concreta a través de la cual se articularon históricamente los vínculos productivos y financieros que hoy operan como canales de transmisión: desde las corporaciones transnacionales con origen ibérico hasta los flujos migratorios que conectan ambos mercados de trabajo. La unidad de análisis es la región, entendida no como suma de economías nacionales sino como configuración específica de posición en la acumulación capitalista mundial.

### 3.2. Período de análisis y fuentes

El análisis abarca el período 2010–2025, que cubre el proceso de emergencia, consolidación y despliegue masivo de la inteligencia artificial como tecnología de propósito general. Las fuentes de datos utilizadas incluyen, para el análisis del mercado de trabajo y la distribución del ingreso, las bases de datos de la OIT (ILOSTAT), la CEPAL, Eurostat y la OCDE; para el análisis de la inversión y la difusión de inteligencia artificial, los informes anuales del Stanford HAI, McKinsey Global Institute y Goldman Sachs Research; para la estructura productiva y las brechas tecnológicas, las bases de CEPAL, los datos de la OCDE sobre I+D y las estadísticas de la UNCTAD sobre cadenas de valor; para los casos empíricos específicos, reportes de la Embrapa en Brasil, del INE de España y de organismos nacionales de estadística de los países iberoamericanos cubiertos. Los informes de McKinsey Global Institute y Goldman Sachs Research se utilizan exclusivamente como indicadores de tendencias de inversión privada y adopción corporativa de inteligencia artificial, no como análisis académico independiente; sus estimaciones se citan en tanto reflejan el propio proceso de valorización que el trabajo analiza, pero deben leerse con las reservas metodológicas que corresponden a fuentes con orientación comercial explícita.

### 3.3. Limitaciones

El análisis enfrenta tres limitaciones metodológicas principales. La primera, y más fundamental desde el punto de vista del marco teórico adoptado, es la distancia entre las categorías estadísticas disponibles y las categorías teóricas del desarrollo propuesto. El índice de exposición a la automatización elaborado por la OCDE no es equivalente a la composición orgánica del capital; la tasa de informalidad laboral de la OIT no es equivalente a la superpoblación consolidada; la participación salarial en el ingreso nacional no es equivalente a la tasa de plusvalor. Estas series estadísticas se utilizan en este trabajo como aproximaciones fenoménicas de las determinaciones estructurales derivadas abstractamente: ilustran la forma en que esas determinaciones se expresan en la superficie del proceso, sin pretender medirlas en sus determinaciones profundas. La segunda limitación es la disponibilidad desigual de datos desagregados sobre adopción de inteligencia artificial por sector y región, especialmente en Iberoamérica, donde la información institucional es más fragmentaria que en Europa. La tercera es la existencia de rezagos temporales entre la adopción tecnológica y sus impactos

sobre el mercado de trabajo, lo que dificulta establecer correspondencias temporales precisas entre el despliegue de inteligencia artificial y las variaciones en los indicadores de empleo e informalidad. Estas limitaciones no invalidan el análisis, pero deben tenerse presentes al evaluar la consistencia entre las determinaciones teóricas y sus expresiones empíricas.

### **3.4. Nota terminológica**

El trabajo utiliza “Iberoamérica” para incluir explícitamente a España y Portugal como parte del espacio de análisis y para señalar vínculos históricos, culturales y productivos que estructuran los mecanismos de transmisión identificados. El uso de esta categoría no implica ningún juicio sobre identidad cultural ni desconoce las profundas heterogeneidades internas de la región iberoamericana.

### **3.5. Nota sobre el método dialéctico y el uso de evidencia estadística**

El desarrollo de las determinaciones en este trabajo se inscribe en el método dialéctico tal como Marx lo despliega en *El Capital* y como Iñigo Carrera lo reelabora en su reconstrucción de la crítica de la economía política. Ese método consiste, en lo esencial, en el ascenso de lo abstracto a lo concreto: las determinaciones generales del capital se exponen primero en su forma simple y luego se despliegan, a través de mediaciones sucesivamente más concretas, hasta alcanzar las formas específicas que ese capital adopta en cada espacio histórico de acumulación. El uso de evidencia estadística en este registro no es incompatible con el método, pero ocupa un lugar preciso en su arquitectura argumental. La sección 3.3 identificó como limitación metodológica fundamental la distancia entre las categorías estadísticas disponibles (exposición a la automatización, informalidad laboral, participación salarial) y las categorías teóricas del desarrollo propuesto (composición orgánica del capital, superpoblación consolidada, tasa de plusvalor). Esa distancia no es un déficit para corregir mediante mejores series, implica una consecuencia necesaria del estatuto diferencial que ambos tipos de categorías tienen en la exposición dialéctica. Las series estadísticas captan regularidades en el plano de la superficie del proceso; las categorías abstractas identifican las determinaciones del proceso de vida social. La jerarquía entre ambos planos no es accesoria, sino que estructura todo el uso de datos que este trabajo realiza.

La caracterización de los datos como expresión fenoménica de las determinaciones estructurales requiere precisarse. En el método de Marx, la apariencia (*Erscheinung*) y la esencia (*Wesen*) no se oponen como verdadero y falso: la apariencia es tan real como la esencia, pero no se explica por sí misma. Los fenómenos económicos —precios, salarios, tasas de empleo, coeficientes de desigualdad, saldos comerciales— son efectivamente existentes y efectivamente operantes en la realidad social, pero su inteligibilidad exige reconstruir las determinaciones ocultas que los producen. La tasa de plusvalor, la composición orgánica del capital y la forma específica de la superpoblación relativa no son más reales que la participación salarial, la inversión en capital fijo o la tasa de informalidad: son sus determinaciones, en el sentido preciso de que los movimientos de las segundas se vuelven inteligibles solo cuando se los refiere a los movimientos de las primeras. En términos metodológicos, las regularidades estadísticas observadas en este trabajo, como la caída de la participación salarial en Europa, la persistencia de la brecha de productividad iberoamericana, la profundización de la informalidad, la asimetría en la balanza de pagos de servicios tecnológicos, no constituyen el punto de partida del argumento sino su expresión en el plano de la superficie del proceso. Su función es ilustrar la forma concreta que adoptan las determinaciones derivadas teóricamente, no operar como evidencia confirmatoria en sentido hipotético-deductivo. Esta es la distinción que organiza el tratamiento que Iñigo Carrera (2008) hace de los datos empíricos: hechos que el desarrollo abstracto debe explicar a partir de sus determinaciones, no variables que el análisis debe ajustar inductivamente.

La formulación de las determinaciones estructurales como hipótesis contrastables (H1–H6) responde a una estrategia de comunicación con la comunidad académica que trabaja en registro convencional, no a una epistemología hipotético-deductiva. Esta decisión puede dar lugar a un equívoco que conviene disipar. Si H1–H6 no son proposiciones falsables en sentido popperiano, un crítico podría objetar que su afirmación de necesidad las vuelve inmunes a la evidencia y por tanto vacías. La objeción supone que la única forma de necesidad aceptable en ciencia es la necesidad empírica; es decir, lo que ocurre recurrentemente en un conjunto de casos observados, y que cualquier otra forma recae en la petición de principio. El desarrollo propuesto opera con un concepto distinto: la necesidad lógico-conceptual, derivada de la reconstrucción de las determinaciones del capital como proceso social. Que la elevación de la composición orgánica tiende a producir superpoblación relativa no es una generalización inductiva sobre casos observados: es una derivación a partir

del análisis del capital como relación social que presupone la apropiación privada del plusvalor y la competencia entre capitales individuales. Esta necesidad no es menos exigente que la empírica: requiere que la derivación sea lógicamente válida y que las mediaciones se expongan sin saltos. Pero su modo de contrastación no es la falsación de predicciones puntuales sino la evaluación de la coherencia entre las determinaciones derivadas y las formas fenoménicas que el análisis concreto identifica. Las hipótesis H1–H6 son, por tanto, reformulaciones convencionales de determinaciones necesarias cuya función es permitir que el argumento dialogue con registros académicos que exigen esa forma de presentación, sin sustituir por ella el tipo de demostración que el desarrollo propuesto efectivamente realiza.

De esta posición metodológica se sigue una consecuencia sobre el tipo de evaluación que el análisis comparativo de la sección 8 despliega. Cuando allí se afirma que la adopción de inteligencia artificial produce formas diferenciadas de superpoblación relativa en Europa occidental, la Península Ibérica e Iberoamérica, la evaluación no consiste en probar estadísticamente que esas formas ocurren ni en refutarlas mostrando contraejemplos: consiste en mostrar que las formas efectivamente observadas son inteligibles como realización concreta de las determinaciones derivadas, y que las explicaciones alternativas, como son la neoclásica y la neoevolucionista schumpeteriana, requieren o bien naturalizar las asimetrías que deberían explicar, o bien introducir factores contingentes ad hoc. La evidencia estadística interviene en esa evaluación, pero no decide la cuestión por sí sola. Lo que decide es la necesidad material de las determinaciones y la capacidad del desarrollo para mostrar que las formas fenoménicas observadas son realizaciones necesarias de esas determinaciones; no datos externos al argumento que el análisis deba acomodar, sino expresiones concretas de un proceso cuya lógica la exposición ya ha reconstruido.

## **4. La inteligencia artificial como cambio técnico capitalista: trayectoria y mecanismo**

### **4.1. Composición orgánica, plusvalor relativo y superpoblación**

Desde el desarrollo propuesto, la inteligencia artificial es la forma más avanzada que adopta actualmente el proceso de objetivación del conocimiento social en el capital constante. Lo que Marx denominaba la “potencia intelectual de la producción”, que

se encontraba concentrada en la maquinaria mecánica durante la primera revolución industrial, adopta hoy la forma de modelos entrenados sobre enormes volúmenes de trabajo humano muerto como insumo de los grandes modelos de lenguaje: datos, patrones de comportamiento, imágenes, etc. A diferencia de la maquinaria mecánica del siglo XIX, que permitió el desplazamiento principalmente trabajo manual en la manufactura y luego la convirtió en apéndice de la maquinaria, la inteligencia artificial desplaza trabajo cognitivo y de servicios, con lo cual el movimiento de su acción se extiende a sectores que hasta ahora habían permanecido relativamente al margen de la presión automatizadora.

Al objetivar en capital constante funciones que hasta su aparición requerían fuerza de trabajo viva, la inteligencia artificial opera sobre la composición interna del obrero colectivo (*Gesamtarbeiter*) tal como Marx lo desarrolla en los capítulos XI a XIII del Tomo I: el organismo productivo compuesto por trabajadores parciales cuya articulación constituye el sujeto real del proceso de trabajo capitalista. Lo que la inteligencia artificial desplaza no son capacidades individuales aisladas sino funciones integradas en ese organismo —redacción, análisis, clasificación, toma de decisiones rutinarias, diseño, programación de complejidad media—, funciones que el desarrollo histórico del proceso de producción había incorporado progresivamente al obrero colectivo y que ahora son objetivadas en capital constante bajo la forma de modelos entrenados. Esta precisión es relevante porque determina el alcance del desplazamiento: no se trata de la sustitución de puestos de trabajo individuales, como presupone el modelo de tareas de Acemoglu y Restrepo, sino de la reorganización del obrero colectivo mediante la transferencia de funciones antes desempeñadas por sus miembros vivos hacia componentes de capital constante. Esa transferencia reduce el tiempo de trabajo socialmente necesario para producir una masa creciente de mercancías, abarata el valor de los medios de subsistencia de la clase obrera y, con ello, el valor de la fuerza de trabajo misma: es el mecanismo específico de producción de plusvalor relativo. El capital individual que adopta primero la tecnología obtiene una ganancia extraordinaria transitoria, que produce por debajo del valor social medio, hasta que su generalización desplaza ese valor medio hacia abajo y la ganancia extraordinaria se disuelve en la nueva norma de productividad. Ese movimiento eleva la composición orgánica del capital: la proporción de capital constante respecto del capital variable crece, dado que la misma masa de producción requiere proporcionalmente menos trabajo vivo. La consecuencia necesaria de esa elevación es la generación de superpoblación relativa: una fracción de la fuerza de trabajo previamente incorporada al

proceso de valorización queda disponible para el capital sin ser absorbida por él, no por insuficiencia de demanda final, sino como resultado directo del incremento en la productividad del trabajo. La superpoblación relativa no es así un efecto colateral ni un resultado contingente de la automatización: es la forma que adopta necesariamente la fuerza de trabajo sobrante cuando el capital eleva su composición orgánica mediante la incorporación de nueva potencia técnica al proceso productivo.

El ejército de reserva que la inteligencia artificial tiende a generar no es ya, predominantemente, el de los trabajadores industriales no calificados, sino el de capas medias y trabajadores de los denominados servicios, que incorporan en su seno fuerza de trabajo de subjetividad ampliada e incluso de conciencia científica. En los espacios de acumulación normal, estos trabajadores constituían el estrato de asalariados que venden su fuerza de trabajo por su valor, integrados al empleo formal de valores de uso social bajo las mediaciones del llamado "Estado de Bienestar"; en el espacio iberoamericano de acumulación, constituían el núcleo del empleo capitalista formal. Esta extensión del alcance de la automatización es la determinación que permite entender, como expresiones necesarias de un mismo proceso, dos fenómenos que desde la perspectiva ortodoxa aparecen como paradojas independientes: la recomposición de la superpoblación sobrante en Europa occidental hacia estratos de trabajadores con subjetividad ampliada, en un espacio que despliega tecnología producida fuera de su órbita de acumulación sin poder desarrollar los capitales que la producen (H3); y la profundización de la superpoblación consolidada en Iberoamérica, donde la adopción parcial y selectiva de esas mismas tecnologías comprime el empleo capitalista formal sin generar las condiciones de absorción productiva que la mediación de la renta diferencial de la tierra estructuralmente desalienta (H5). Ambas determinaciones serán desarrolladas en las secciones 6 y 7.

El mecanismo expuesto permite anticipar el contenido de las determinaciones estructurales que este trabajo desarrolla. La concentración de la producción de inteligencia artificial en un número reducido de capitales en Estados Unidos y, en un sentido marginal pero creciente, en China genera una apropiación diferencial de renta sobre el conocimiento objetivado que opera de forma sostenida como la manifestación de una barrera a la entrada de nuevos capitales: las regiones que no producen esa tecnología deben adquirirla bajo condiciones determinadas en otro espacio nacional de acumulación, transfiriendo valor hacia ellos (H1, H6). Esa transferencia no impide la adopción, pero la condiciona: los capitales que compran esa tecnología en Europa occidental e

Iberoamérica incorporan inteligencia artificial de manera selectiva según su posición relativa respecto de la tasa media de la rama y sus brechas de productividad previas, generando un dualismo productivo que amplía las divergencias internas en lugar de resolverlas (H2). El resultado agravado de ese proceso, que se manifiesta con la ausencia de incrementos de productividad agregada proporcionales a la escala de inversión, fenómeno que la literatura convencional denomina paradoja de la productividad (Brynjolfsson, 1993), constituye la expresión fenoménica de una contradicción estructural, no una anomalía temporal. Las secciones 5, 6 y 7 desarrollan la forma específica que estas determinaciones adoptan en cada uno de los tres espacios de acumulación analizados.

#### **4.2. Concentración geográfica de la producción**

La producción de tecnología de inteligencia artificial de frontera exhibe una concentración geográfica sin precedentes en la historia del cambio técnico capitalista. En 2023, Estados Unidos concentraba el 53 % de los modelos notables producidos mundialmente y el 67 % de la inversión privada global en inteligencia artificial (Stanford HAI, 2024). Europa, con una población comparable, producía menos del 15 % de los modelos y menos del 5 % de la inversión. Iberoamérica producía menos del 1 % en ambas dimensiones. Esta concentración no es contingente, es el resultado de la forma acumulación histórica y la imposibilidad de los capitales de apropiarse de la tasa media de ganancia para la conformación de infraestructura de cómputo, datos de entrenamiento y capacidad de objetivación del conocimiento, y de la concentración de fuerza de trabajo con subjetividad ampliada, que requirió décadas de inversión en infraestructura científica y universitaria. La posesión de ese capital constante en una intensiva carrera por la extracción de fuentes de plusvalía extraordinaria producto de la innovación técnica, que se manifiesta en licencias, suscripciones, precios de infraestructura de cómputo, operan como renta diferencial sobre el conocimiento objetivado: las regiones que adoptan tecnología de inteligencia artificial sin producirla transfieren valor hacia los centros de producción en cada ciclo de uso, independientemente de los incrementos de productividad que esa adopción genere en el espacio receptor. Esa transferencia no requiere ningún mecanismo de extracción directo: es el resultado necesario de adquirir bajo condiciones extraordinarias de cambio técnico intensivo una mercancía cuyo valor fue producido bajo condiciones específicas y cuyo precio de producción incorpora la renta derivada de la posición técnica diferencial de sus

propietarios.

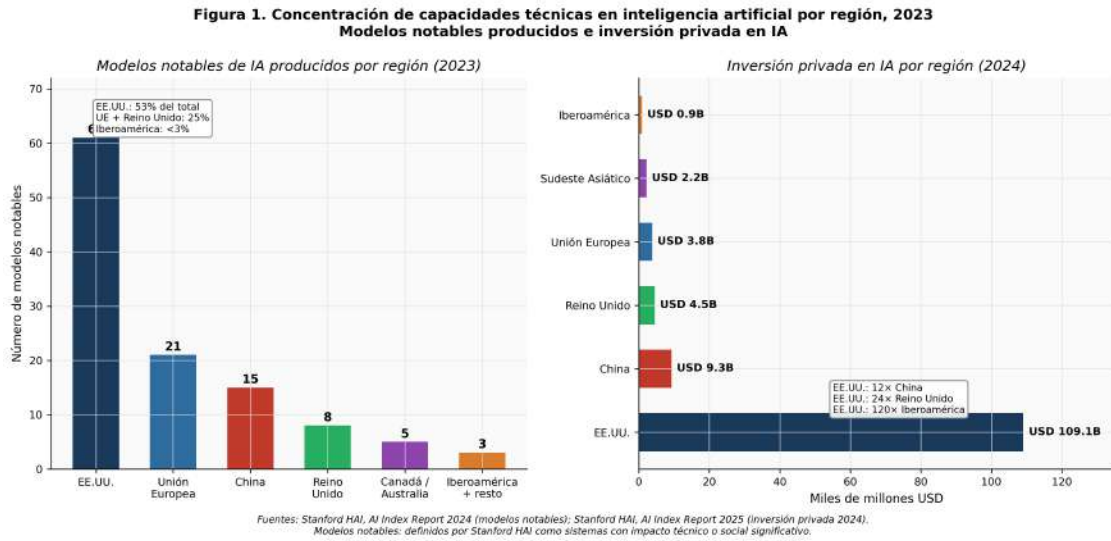


Figura 1: Concentración de capacidades técnicas en inteligencia artificial por región, 2023

Nota: Fuentes: Stanford HAI, AI Index Report 2024 (modelos notables); Stanford HAI, AI Index Report 2025 (inversión privada 2024). Modelos notables: definidos por Stanford HAI como sistemas con impacto técnico o social significativo en el período analizado. La figura combina dos series temporalmente distintas (modelos 2023; inversión 2024) por disponibilidad estadística.

### 4.3. Trayectoria del despliegue: evidencia empírica, 2010–2025

El despliegue de la inteligencia artificial como tecnología de uso generalizado puede describirse, a efectos expositivos, en tres fases determinadas por el ritmo de acumulación de capital en el sector y por los umbrales técnicos que habilitaron su generalización productiva. Esta periodización no constituye una periodización teórica en sentido estricto: recoge hitos observables, tales como cambios en la escala de inversión, rupturas en la arquitectura técnica dominante e incorporación masiva a los flujos de trabajo, que permiten situar empíricamente el análisis sin anticipar ningún esquema sobre la forma o las posibilidades distributivas del despliegue.

La primera fase, comprendida entre 2010 y 2016, corresponde al período de emergencia del aprendizaje profundo en aplicaciones sectoriales específicas. La inversión

global en startups de inteligencia artificial pasó de aproximadamente USD 3.000 millones en 2012 a USD 15.000 millones en 2016 (Stanford HAI, 2022), concentrada de manera abrumadora en Estados Unidos y, en menor medida, en China. La segunda fase, de 2017 a 2022, estuvo marcada por la arquitectura Transformer (Vaswani et al., 2017) y el escalamiento masivo de modelos de lenguaje de gran escala. A diferencia de la serie anterior, que mide exclusivamente la inversión en startups de inteligencia artificial, la inversión privada global agregada en el sector pasó de USD 12.000 millones en 2015 a USD 91.900 millones en 2022 (Stanford HAI, 2023), magnitudes que incluyen fragmentación hacia capitales emergentes, así como en divisiones de inteligencia artificial en estructuras de capitales medios, y fusiones y adquisiciones por parte de varios capitales. Europa representaba menos del 10 % del total en este agregado (Figura 2). La tercera fase, iniciada en 2022 y en pleno curso al momento de este trabajo, corresponde al despliegue masivo al público general y a la integración de sistemas de inteligencia artificial en flujos de trabajo productivos y administrativos.

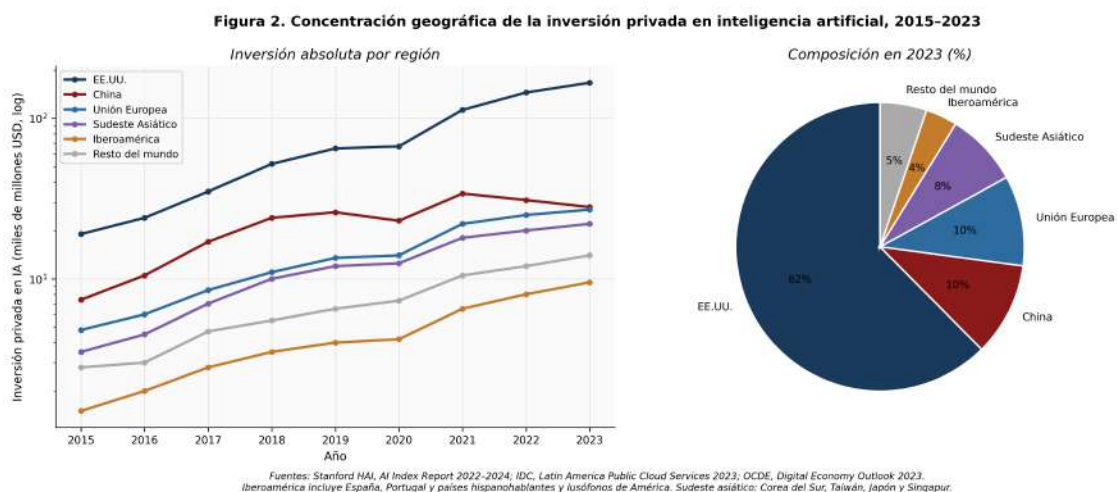


Figura 2: Concentración geográfica de la inversión privada en inteligencia artificial, 2015-2023

Nota: Fuentes: Stanford HAI, AI Index Report 2022-2024; IDC, Latin America Public Cloud Services 2023; OCDE, Digital Economy Outlook 2023. Iberoamérica incluye España, Portugal y países hispanohablantes y lusófonos de América. Sudeste asiático: Corea del Sur, Taiwán, Japón y Singapur.

## 5. Europa occidental: recomposición de la superpoblación sobrante

### 5.1. Posición estructural e historia reciente (H1, H3)

Europa occidental ocupa una posición históricamente central en la acumulación capitalista mundial. Desde la década de 1980 ha experimentado un proceso de desindustrialización relativa, con transferencia de producción manufacturera hacia Asia Oriental y expansión relativa del empleo en actividades de servicios —distribución, servicios financieros, servicios profesionales y administración pública— que absorbie- ron parcialmente la fuerza de trabajo expulsada de la industria. En 2023, el empleo en ese conjunto de actividades representaba entre el 70 % y el 80 % del empleo total en las principales economías de Europa occidental (Eurostat, 2023), mientras que el sector manufacturero había caído desde el 30–35 % del empleo en 1970 a menos del 20 % en la mayoría de los países. Entre 2010 y 2022, el crecimiento de la productividad laboral en la zona euro promedió el 0,8 % anual, frente al 1,5 % de Estados Unidos (OCDE, 2023).

La participación de los salarios en el ingreso nacional de la UE-15 cayó de manera sostenida desde el 63 % en 2000 al 59 % en 2022 (OIT, ILOSTAT-KILM). Esta tendencia no es anterior a la inteligencia artificial por accidente: es la expresión fenoménica de un proceso de elevación de la composición orgánica del capital que precede al despliegue masivo de la IA y que ésta viene a profundizar. En la medida en que la IA opera bajo el mecanismo de plusvalía relativa como capital constante que desplaza trabajo vivo en las ramas de producción de mercancías sociales, que había absorbido la fuerza de trabajo expulsada por la desindustrialización manufacturera de las décadas anteriores, la acumulación capitalista reproduce en ese sector la dinámica ya conocida: aumento de la productividad por unidad de capital empleado, pero captura de ese excedente en los espacios donde la tecnología se produce, no en los espacios donde se despliega (H6). El resultado observado; a saber, crecimiento de la productividad laboral en la zona euro de apenas el 0,8 % anual entre 2010 y 2022, frente al 1,5 % registrado en Estados Unidos en el mismo período (OCDE, 2023): no es un déficit de adopción tecnológica, sino la expresión de una transferencia de valor desde los espacios que despliegan la tecnología hacia los espacios que la producen y controlan su infraestructura crítica (H1, H6). La brecha de productividad respecto de Estados Unidos se amplió durante el mismo período, con la UE-15 pasando de representar el 79 % de la productividad norteamericana en 2000 al 71 % en 2022, a pesar de la difusión generalizada de herramientas digitales.

## 5.2. Superpoblación sobrante en el trabajo asalariado de subjetividad ampliada (H2, H3, H4)

En Europa occidental, la forma que adopta este proceso sobre la clase trabajadora es la recomposición de la superpoblación sobrante. Su determinación es la siguiente: el espacio europeo de acumulación absorbió, desde los años ochenta, la fuerza de trabajo expulsada de la industria manufacturera hacia industrias de producción de mercancías sociales o servicios, donde esa fuerza de trabajo alcanzó niveles de integración formal: contratos regulados, cobertura sindical, protección social, consonantes con el valor medio de la fuerza de trabajo en ese espacio. Al elevarse ahora la composición orgánica del capital en ese mismo sector por efecto de la adopción de herramientas de IA, el capital desplaza hacia la reserva a trabajadores que se encontraban activamente empleados en el proceso productivo. Esta es precisamente la determinación de la forma sobrante: no se trata de trabajadores expulsados de formas pre-capitalistas de producción (eso correspondería a la forma latente), ni de un segmento crónicamente excluido de la integración formal (eso sería la forma consolidada), sino de trabajadores cuya reintegración al proceso activo de valorización depende de las fluctuaciones de la acumulación y de la velocidad del cambio técnico. La OCDE (2023) estimó que entre el 25 % y el 35 % de los empleos en Europa occidental presentan alta exposición a la automatización por inteligencia artificial, con el sector financiero (35–40 %), los servicios jurídicos (30 %), los servicios administrativos (40 %) y la logística (25–30 %) a la cabeza, sectores que concentran precisamente los estratos de empleo formal de calificación media que fueron el núcleo de la reconversión post-industrial europea (H3). La Comisión Europea (2021) estimó que entre 24 y 28 millones de personas trabajaban a través de plataformas digitales en la UE, de las cuales al menos 5,5 millones bajo formas contractuales en las que la relación salarial es ocultada mediante la figura del trabajo autónomo, permitiendo la extracción de plusvalor al margen de los mecanismos redistributivos del contrato laboral formal.

La especificidad de esta forma de superpoblación sobrante reside en que se extiende a capas que en los ciclos anteriores de automatización permanecieron al margen: trabajadores con formación universitaria, trayectorias profesionales consolidadas y representación sindical. Esta extensión tiene consecuencias distributivas y políticas sin precedentes: el Estado social europeo fue diseñado para mediar el conflicto entre capital y trabajo industrial, no para gestionar el desplazamiento de trabajadores asalariados con subjetividad técnica acumulada que en ciclos anteriores de automatización permanecían en el ejército activo (Figura 3).

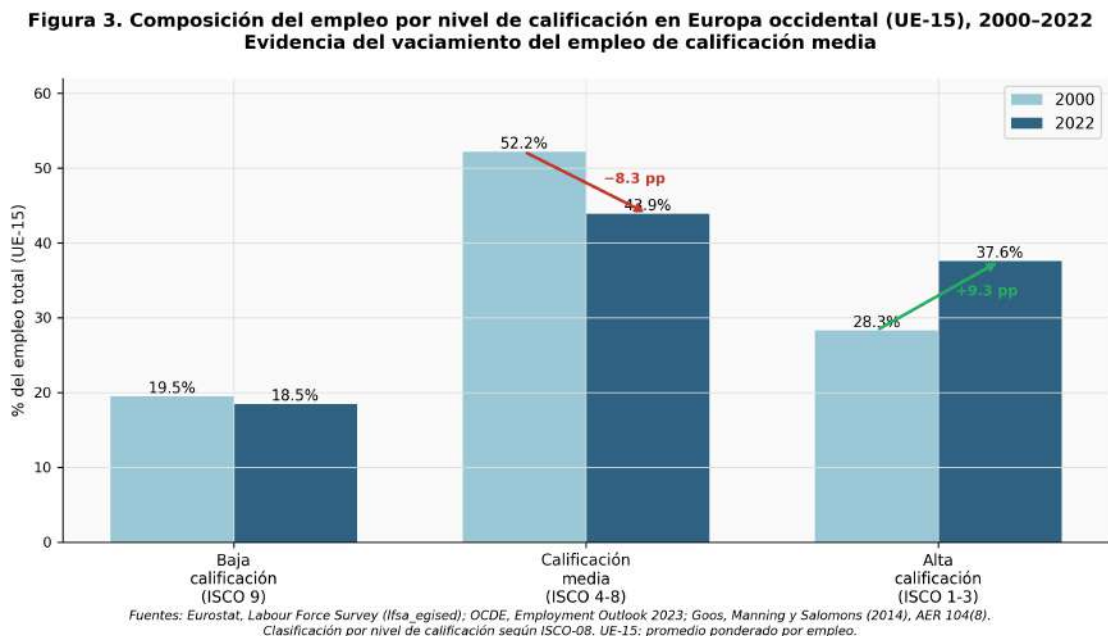


Figura 3: Composición del empleo por nivel de calificación en Europa occidental (UE-15), 2000–2022

Nota: Fuentes: Eurostat, Labour Force Survey (lfsa\_egised); OCDE, Employment Outlook 2023; Goos, Manning y Salomons (2014), AER 104(8). Clasificación por nivel de calificación según ISCO-08. UE-15: promedio ponderado por empleo. El vaciamiento del empleo medio (job polarization) es consistente con el argumento de H3.

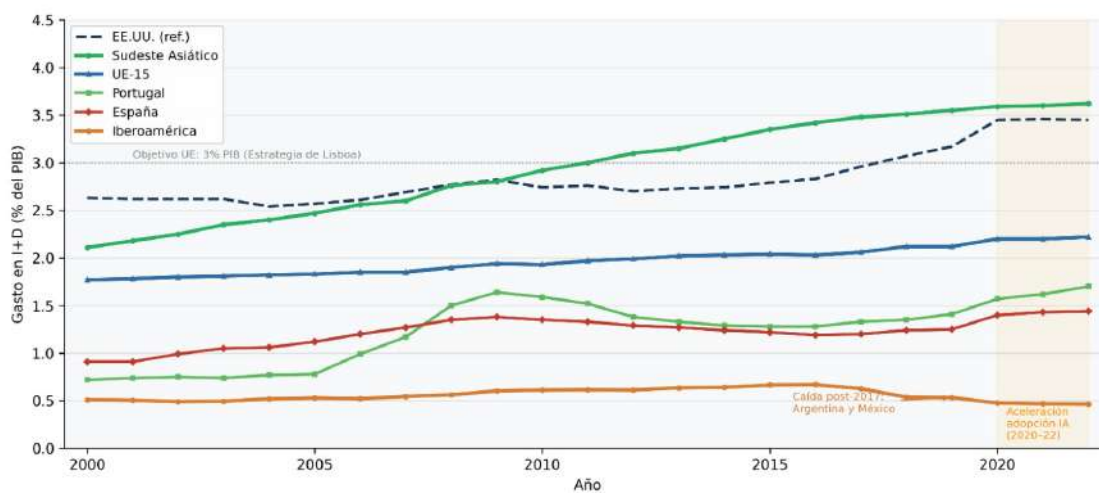
### 5.3. Ausencia de subjetividad técnica acumulada, transferencia de valor y los límites de la mediación estatal (H1, H6)

La posición de Europa occidental en la acumulación mundial se expresa con particular nitidez en su relación con la producción de tecnología de frontera. Como se desarrolló en la sección 4 y queda registrado empíricamente en la Figura 1, la concentración de la subjetividad técnica acumulada —la capacidad específica para producir los modelos de lenguaje de gran escala, los sistemas de visión computacional y la infraestructura de cómputo que los sostiene— se localiza de manera determinada en el espacio norteamericano y, en menor medida, en el espacio asiático oriental. Europa occidental, pese a su posición histórica como espacio de acumulación avanzada, no participó de la fase formativa de acumulación de esa subjetividad técnica, que requirió una

concentración específica de capital constante, de investigación pública orientada al complejo militar-tecnológico y de mercados de capitales de riesgo que no se desarrollaron en Europa con la misma intensidad que en Estados Unidos. Esta ausencia no es el resultado de decisiones contingentes de política industrial: es la expresión de la forma históricamente determinada en que el capital se acumuló en el espacio europeo durante las décadas decisivas de la formación del paradigma digital.

La consecuencia estructural de esta posición es que cualquier empresa o Estado europeo que despliega infraestructura de inteligencia artificial transfiere valor hacia los propietarios de esa infraestructura, en la forma de renta de monopolio sobre el conocimiento objetivado (H6). El AI Act (2024), el Data Governance Act (2022), el Digital Markets Act (2022) y los programas de inversión en I+D agrupados bajo Horizon Europe son las formas en que el Estado europeo —como forma política del capital operando en ese espacio— intenta mediar las contradicciones que genera el despliegue de una tecnología producida y controlada externamente. Estas iniciativas operan en el nivel de la distribución y la circulación: pueden modificar las condiciones en que el valor producido localmente se distribuye entre usuarios, plataformas y Estados, y pueden ralentizar o condicionar el despliegue de la tecnología. Pero no pueden alterar la determinación estructural que las produce, porque ésta reside en las condiciones de producción del valor, no en su distribución posterior. La brecha en el gasto en investigación y desarrollo (Figura 4) no es la causa de esta posición: es su expresión. La dinámica de acumulación en el espacio europeo no genera las condiciones para sostener inversión en IA de frontera porque los retornos de esa inversión fluyen, de manera determinada, hacia el capital que ya controla la infraestructura crítica y los modelos de base sobre los que cualquier aplicación derivada debe construirse.

**Figura 4. Gasto en investigación y desarrollo como porcentaje del PIB, 2000-2022**  
**Europa occidental, Península Ibérica, Iberoamérica, Sudeste Asiático y EE.UU.**



Fuentes: OCDE, Main Science and Technology Indicators (2023); Eurostat, rd\_e\_gerdtot; UNESCO UIS.Stat; CEPAL, CEPALSTAT; RICYT. Sudeste Asiático: promedio ponderado por PIB de Corea del Sur, Taiwán y Singapur. Iberoamérica: promedio simple de Argentina, Brasil, México, Chile y Colombia.

Figura 4: Gasto en investigación y desarrollo como porcentaje del PIB, 2000–2022

Nota: Fuentes: OCDE, Main Science and Technology Indicators (2023); Eurostat, rd\_e\_gerdtot; UNESCO UIS.Stat; CEPAL, CEPALSTAT; RICYT. Sudeste Asiático: promedio ponderado por PIB de Corea del Sur, Taiwán y Singapur. Iberoamérica: promedio simple de Argentina, Brasil, México, Chile y Colombia.

## 6. La Península Ibérica: momento bisagra en el proceso mundial de acumulación capitalista

### 6.1. Especificidad histórica y estructura productiva

España y Portugal ocupan en el espacio europeo de acumulación una posición determinada por la coexistencia de integración plena a los mercados e instituciones de la Unión Europea con una estructura productiva que no desarrolló la composición orgánica del capital ni la subjetividad ampliada acumulada necesarias para operar en la frontera tecnológica. España se especializó en construcción, turismo, servicios inmobiliarios e industria automotriz ensambladora; Portugal, en turismo, industria textil y de calzado, y servicios (Figura 7). Esta especialización no es el resultado de decisiones de política inadecuadas: es la expresión de la forma en que la acumulación de capital se realizó en estos territorios, determinada por su posición en la división

internacional del trabajo y por la ausencia de condiciones para el desarrollo de las fuentes de plusvalía extraordinarias sobre conocimiento objetivado que la producción tecnológica de frontera requiere.

La tasa de desempleo en España promedió el 14,8 % entre 2000 y 2020, con un pico del 26,1 % durante la crisis financiera de 2012–2013 (Eurostat, 2023), expresión fenoménica de una superpoblación sobrante estructuralmente elevada. El mercado de trabajo español exhibe además una de las tasas más altas de temporalidad contractual en la OCDE, que se sitúan entre el 25 % y el 35 % de los contratos asalariados entre 2000 y 2020, frente a una media de la OCDE del 11 % (OCDE, 2022; Figura 5). Esta temporalidad no constituye una anomalía institucional sino la forma específica que adopta la superpoblación sobrante en condiciones de integración plena al mercado único europeo, donde la presión competitiva sobre el coste del trabajo no puede compensarse mediante diferencias cambiarias. El gasto en I+D de España y Portugal no alcanzó en ningún año del período analizado la media de la UE-15, situándose en torno al 1,3–1,5 % del PIB frente al objetivo europeo del 3 %, lo que expresa la escasa acumulación de subjetividad técnica que caracteriza a ambas economías y reproduce su posición estructural (Figura 6).

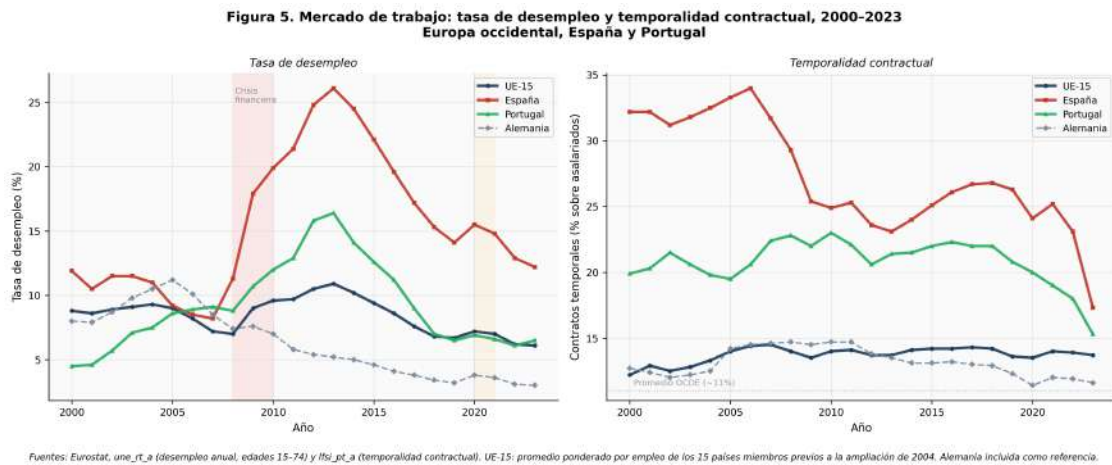
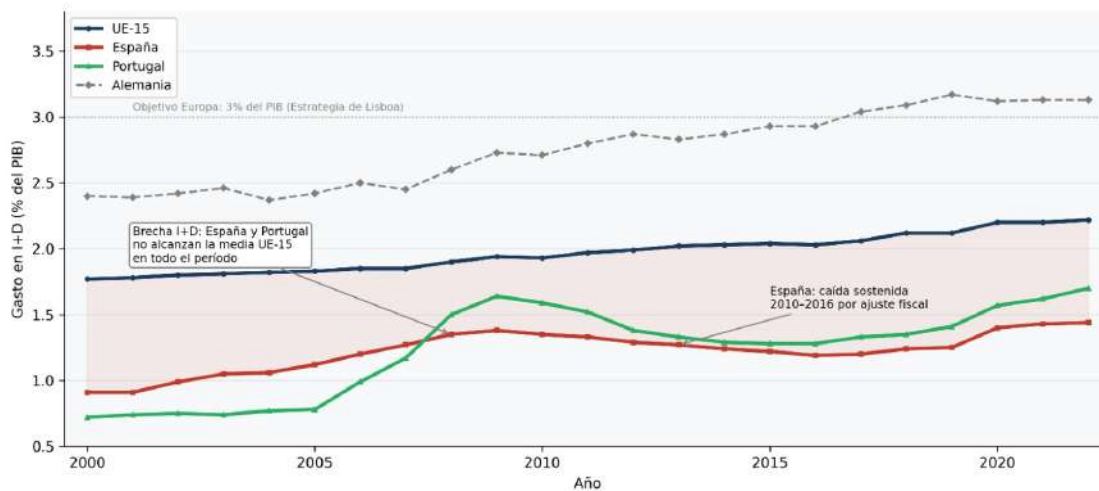


Figura 5: Mercado de trabajo: tasa de desempleo y temporalidad contractual, 2000–2023

Nota: Fuentes: Eurostat, une\_rt\_a (desempleo anual, edades 15–74) y lfsi\_pt\_a (temporalidad contractual). UE-15: promedio ponderado por empleo de los 15 países miembros previos a la ampliación de 2004. Alemania incluida como referencia del núcleo de Europa occidental.

**Figura 6. Gasto en investigación y desarrollo como porcentaje del PIB, 2000-2022**  
Europa occidental, España y Portugal

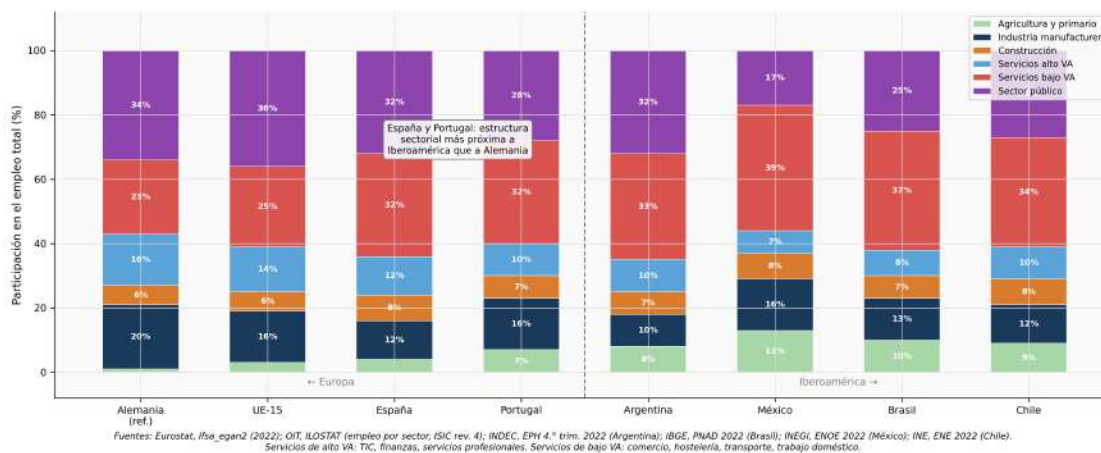


Fuentes: Eurostat, rd\_e\_gerdtot (Gross Expenditure on R&D como % del PIB); OCDE, Main Science and Technology Indicators (2023). UE-15: promedio ponderado por PIB. El objetivo del 3% del PIB fue establecido en la Estrategia de Lisboa (2000) y reafirmado en la Estrategia Europa 2020 y en Horizonte Europa (2021-2027).

Figura 6: Gasto en investigación y desarrollo como porcentaje del PIB: España, Portugal y UE-15, 2000-2022

Nota: Fuentes: Eurostat, rd\_e\_gerdtot (Gross Expenditure on R&D como % del PIB); OCDE, Main Science and Technology Indicators (2023). UE-15: promedio ponderado por PIB. El objetivo del 3 % del PIB fue establecido en la Estrategia de Lisboa (2000) y reafirmado en la Estrategia Europa 2020 y en Horizonte Europa (2021-2027).

**Figura 7. Estructura sectorial del empleo comparada, c. 2022**  
Porcentaje del empleo total por sector



Fuentes: Eurostat, ifsa\_egan2 (2022); OIT, ILOSTAT (empleo por sector, ISIC rev. 4); INDEC, EPH 4.º trim. 2022 (Argentina); IBGE, PNAD 2022 (Brasil); INEGI, ENE 2022 (México); INE, ENE 2022 (Chile). Servicios de alto VA: TIC, finanzas, servicios profesionales. Servicios de bajo VA: comercio, hostelería, transporte, trabajo doméstico.

Figura 7: Estructura sectorial del empleo comparada, c. 2022

Nota: Fuentes: Eurostat, ifsa\_egan2 (2022); OIT, ILOSTAT (empleo por sector, ISIC

rev. 4); INDEC, EPH 4.º trim. 2022 (Argentina); IBGE, PNAD 2022 (Brasil); INEGI, ENOE 2022 (México); INE, ENE 2022 (Chile). Servicios de alto VA: TIC, actividades financieras y de seguros, servicios profesionales y científicos. Servicios de bajo VA: comercio minorista y mayorista, hostelería, transporte y almacenamiento, trabajo doméstico.

## 6.2. El canal bisagra: capitales medios, servicios compartidos y migración

La especificidad analítica de la Península Ibérica reside en su función de articulación entre el espacio europeo y el iberoamericano a través de tres canales concretos. El primero es el de los capitales normales con sede en España y Portugal, y que operan masivamente en mercados iberoamericanos: Telefónica, Banco Santander, BBVA, Iberdrola, Repsol, etc. Esta presencia dominante no es una contingencia histórica ni cultural: es la forma que adoptó la expansión del capital ibérico hacia espacios de acumulación con menor composición orgánica y con condiciones favorables para la apropiación de plusvalor, en sectores tales como telecomunicaciones, servicios financieros, energía, donde la diferencia de productividad entre la frontera tecnológica y la estructura local genera rentas apropiables. La automatización de procesos mediante inteligencia artificial para estos tipos de procesos productivos suscita efectos sobre el empleo tanto en la Península como en fragmentos iberoamericanos, convirtiendo estos canales en vectores concretos de transmisión de la presión automatizadora entre ambos espacios (Figura 8; anticipados en la Sección 3.1). El segundo canal es el de los centros de servicio compartido: la automatización mediante inteligencia artificial está erosionando simultáneamente el modelo de servicio compartido en la Península y el modelo de offshoring en Iberoamérica. El tercero es el migratorio: hay aproximadamente 2 millones de iberoamericanos residentes en España en 2023 (INE, 2023), concentrados predominantemente en sectores de servicios de baja y mediana calificación con alta exposición a la automatización —hostelería, trabajo doméstico, comercio, transporte. Este canal opera en ambas direcciones: el desplazamiento de empleo en España por automatización presiona a los trabajadores migrantes hacia el desempleo o el retorno, lo que retroalimenta el excedente de población en los países de origen; al mismo tiempo, la expectativa migratoria hacia España funciona como válvula de escape para la superpoblación consolidada en Iberoamérica, válvula que la automatización tiende a cerrar en la medida en que erosiona precisamente los sectores de destino habitual de esa migración.

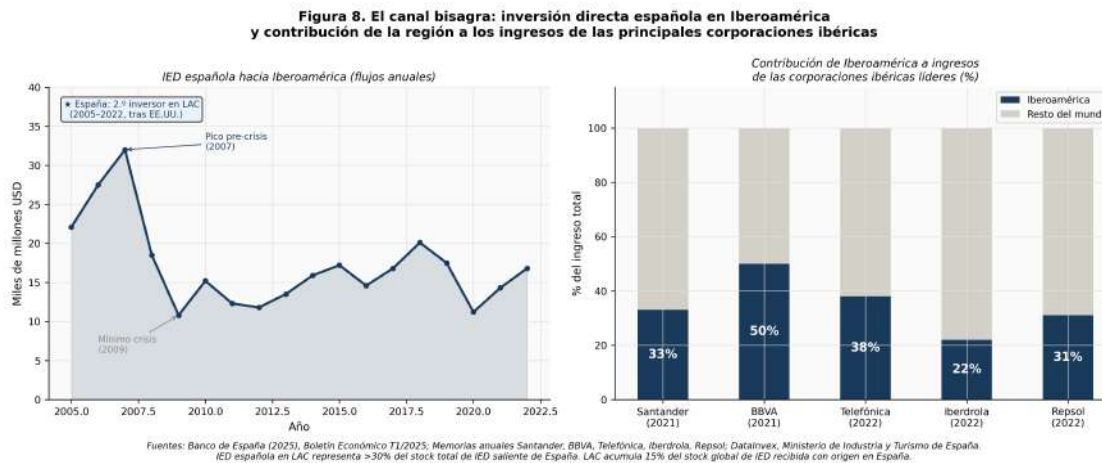


Figura 8: El canal bisagra: inversión directa española en Iberoamérica

Nota: Fuentes: Banco de España (2025), Boletín Económico T1/2025; Memorias anuales Santander, BBVA, Telefónica, Iberdrola, Repsol; DataInVex, Ministerio de Industria y Turismo de España. IED española en LAC representa >30% del stock total de IED saliente de España.

### 6.3. Superpoblación sobrante preexistente y nueva presión automatizadora

La superposición de la superpoblación sobrante estructural preexistente —expresada en tasas de desempleo y temporalidad históricamente altas— con la presión automatizadora emergente configura en la Península Ibérica una determinación específica: dos fuerzas que operan simultáneamente sobre la misma base de fuerza de trabajo, una derivada de la estructura productiva de baja composición orgánica y otra derivada del despliegue de la inteligencia artificial sobre el empleo formal de servicios. Esta determinación diferencia a España y Portugal tanto de Europa occidental, donde la superpoblación sobrante preexistente es menor y el punto de partida más sólido, como de Iberoamérica, donde el excedente de población adopta formas cualitativamente distintas, como se analiza en la sección siguiente.

El índice de exposición ponderada a la automatización por inteligencia artificial en España (37,6%) y Portugal (38,2%) es comparable al de Iberoamérica en términos agregados, pero con una composición radicalmente distinta: en la Península Ibérica esa exposición recae fundamentalmente sobre el empleo formal, mientras que en Iberoamérica recae sobre el empleo informal. Esta distinción no es un dato estadístico

neutro: desde las categorías adoptadas, implica que las formas de la superpoblación relativa son distintas. En la Península predomina la superpoblación sobrante: trabajadores formalmente empleados en sectores de servicios que la automatización desplaza directamente, mientras que en Iberoamérica la forma dominante es la superpoblación consolidada, que opera bajo condiciones de informalidad crónica sobre las que la automatización impone una presión adicional. Los mecanismos de absorción del excedente de población también difieren. La Península Ibérica opera bajo las formas de mediación estatal del valor de la fuerza de trabajo que el espacio europeo de acumulación desarrolló históricamente —transferencias por desempleo, regulación contractual, cobertura de seguridad social—, cuya existencia no expresa una corrección externa al proceso de acumulación sino la forma específica en que el capital europeo reproduce su fuerza de trabajo en ese espacio; en Iberoamérica, esas formas de mediación no se desarrollaron porque la estructura de acumulación mediada por la renta diferencial de la tierra no las requiere, y el excedente de población se realiza directamente como superpoblación consolidada —informalidad crónica— o se desplaza mediante la emigración.

Estas determinaciones expresan diferenciadamente las hipótesis H2 y H4. La adopción selectiva de inteligencia artificial en la Península (H2) opera sobre un tejido productivo de servicios de mediana calificación que constituye la capa más vulnerable: centros de contacto, back-office financiero, procesamiento administrativo. La vulnerabilidad asimétrica del trabajo (H4) se manifiesta aquí en la convergencia de alta exposición sectorial con baja capacidad de reconversión, dada la débil acumulación de subjetividad técnica en la fuerza de trabajo. Y la diferenciación regional como momento necesario del proceso mundial (H2) adquiere en la Península su expresión específica: no como caso intermedio entre Europa e Iberoamérica, sino como articulación activa entre ambos espacios, cuya función bisagra no es una característica cultural o geográfica sino una determinación de la acumulación.

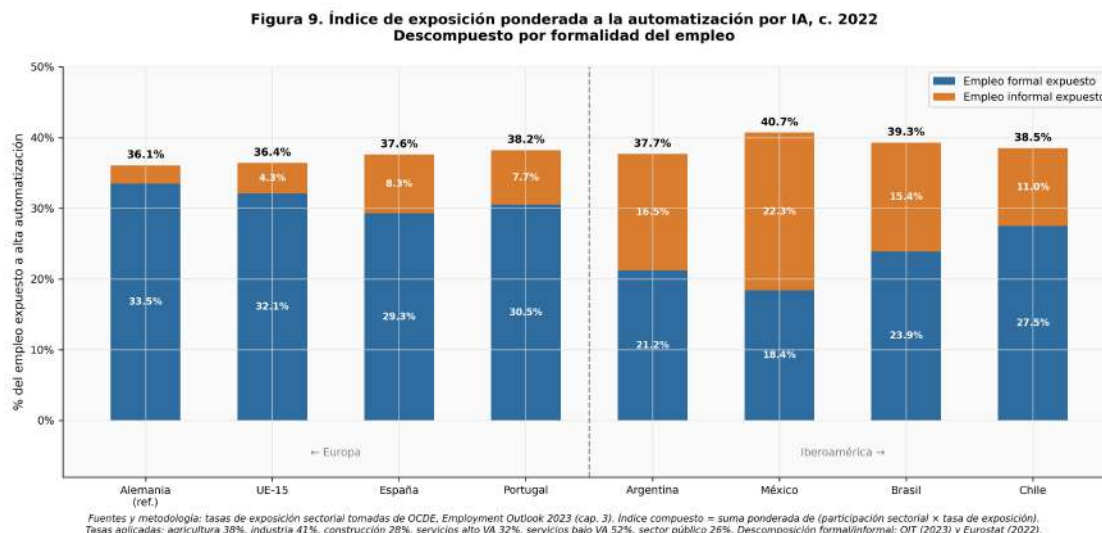


Figura 9: Índice de exposición ponderada a la automatización por IA, c. 2022

Nota: Fuentes y metodología: tasas de exposición sectorial tomadas de OCDE, Employment Outlook 2023 (cap. 3). Índice compuesto = suma ponderada de (participación sectorial en el empleo × tasa de exposición del sector). Descomposición formal/informal: OIT (2023a) y Eurostat (2022).

La Península Ibérica no concluye, entonces, en sí misma como caso de análisis: su función bisagra hace que las determinaciones que operan sobre ella sean inseparables de las tensiones que operan sobre la Europa Occidental e Iberoamérica. Los tres canales identificados en 6.2: capitales fragmentados, servicios compartidos y flujos migratorios, son los vectores concretos a través de los cuales la presión automatizadora se desplaza entre ambos espacios, adoptando en el espacio iberoamericano formas específicas que la sección siguiente analiza.

## 7. Iberoamérica: compresión de la superpoblación latente y profundización de la consolidada

### 7.1. La inserción iberoamericana en la acumulación mundial

Iberoamérica presenta históricamente una inserción en la división internacional del trabajo marcada por la exportación de materias primas, la provisión de trabajo de

menor calificación relativa y la importación de tecnología. Según datos de la CEPAL (2023), la productividad laboral media de Iberoamérica representaba en 2022 aproximadamente el 24 % de la productividad laboral de los países de la OCDE, una brecha que se ha mantenido relativamente estable en las últimas décadas a pesar de períodos de crecimiento económico sostenido (Figura 10). Esta persistencia de la brecha no es el resultado de políticas inadecuadas: es la expresión de la mediación estructural de la renta de la tierra, que orienta la acumulación hacia sectores de alta productividad natural y desincentiva el desarrollo industrial diversificado.

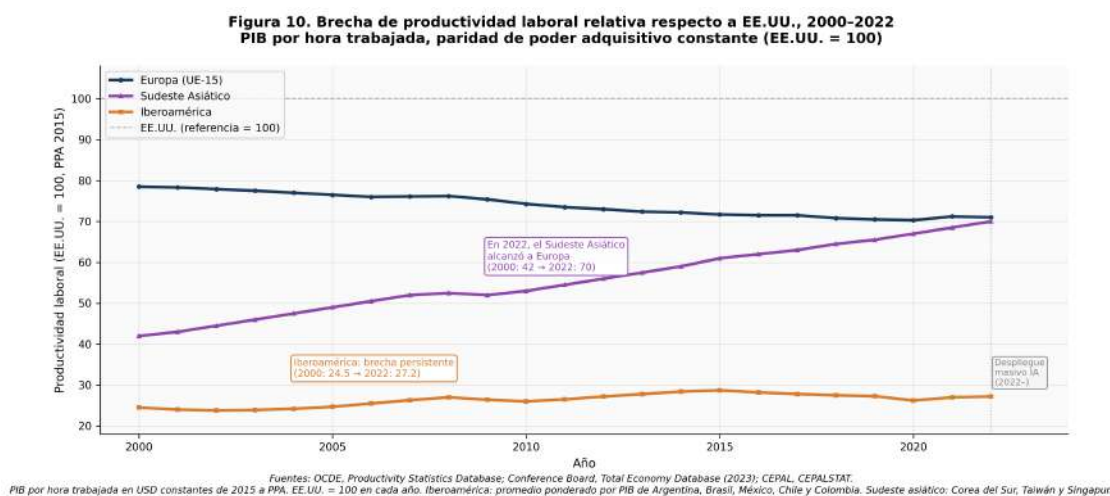


Figura 10: Brecha de productividad laboral relativa respecto a EE.UU., 2000–2022

Nota: Fuentes: OCDE, Productivity Statistics Database; Conference Board, Total Economy Database (2023); CEPAL, CEPALSTAT. PIB por hora trabajada en USD constantes de 2015 a PPA. EE.UU. = 100 en cada año. Iberoamérica: promedio ponderado por PIB de Argentina, Brasil, México, Chile y Colombia. Sudeste asiático: Corea del Sur, Taiwán y Singapur.

La informalidad laboral, que aquí entendemos como la expresión estadística de la superpoblación consolidada, alcanzaba en 2022 al 46,7 % de los trabajadores iberoamericanos según la OIT (2023b). Las variaciones nacionales son considerables: en México, la tasa de informalidad ronda el 55 %; en Bolivia alcanza el 80 %; en Argentina, el 44 %; en Brasil, el 39 %; en Chile, el 28 % (OIT, 2023). Estas cifras ilustran la magnitud de la superpoblación consolidada preexistente sobre la cual opera el despliegue de inteligencia artificial (Figura 11).

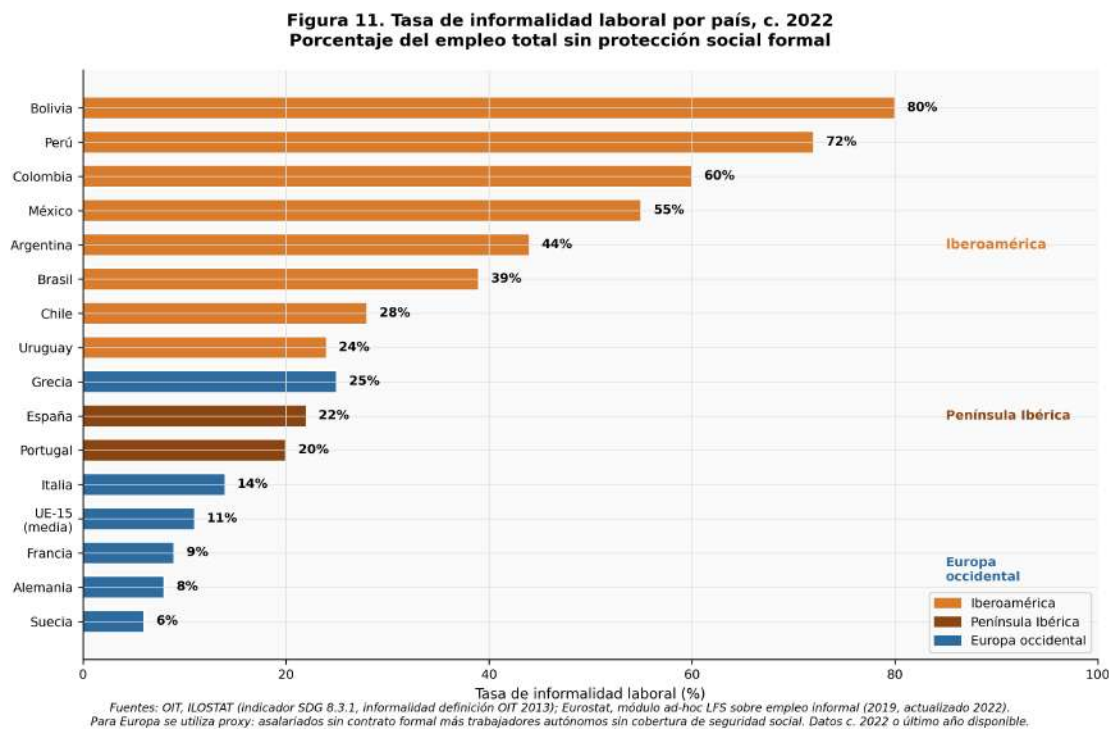


Figura 11: Tasa de informalidad laboral por país, c. 2022

Nota: Fuentes: OIT, ILOSTAT (indicador SDG 8.3.1, informalidad definición OIT 2013); Eurostat, módulo ad-hoc LFS sobre empleo informal (2019, actualizado 2022). Para Europa se utiliza proxy: asalariados sin contrato formal más trabajadores autónomos sin cobertura de seguridad social. Datos c. 2022 o último año disponible.

## 7.2. Compresión de la superpoblación latente: automatización agroindustrial

La superpoblación latente en Iberoamérica ha sido históricamente voluminosa, sostenida por la persistencia de economías campesinas y formas de trabajo agrario. La introducción de tecnologías de automatización en la agroindustria iberoamericana, a través de la automatización de tractores, drones de monitoreo fitosanitario, sistemas de riego y fertilización de precisión, clasificación automatizada en plantas de procesamiento, etc. reduce la demanda de trabajo rural y expulsa hacia los centros urbanos a trabajadores que no encuentran allí absorción en un mercado de trabajo formal que simultáneamente se contrae por las mismas presiones automatizadoras en el sector servicios. En Brasil, la adopción de sistemas de agricultura de precisión en cultivos de soja y maíz creció más del 40 % entre 2018 y 2023 (Embrapa, 2023; Figura 12).

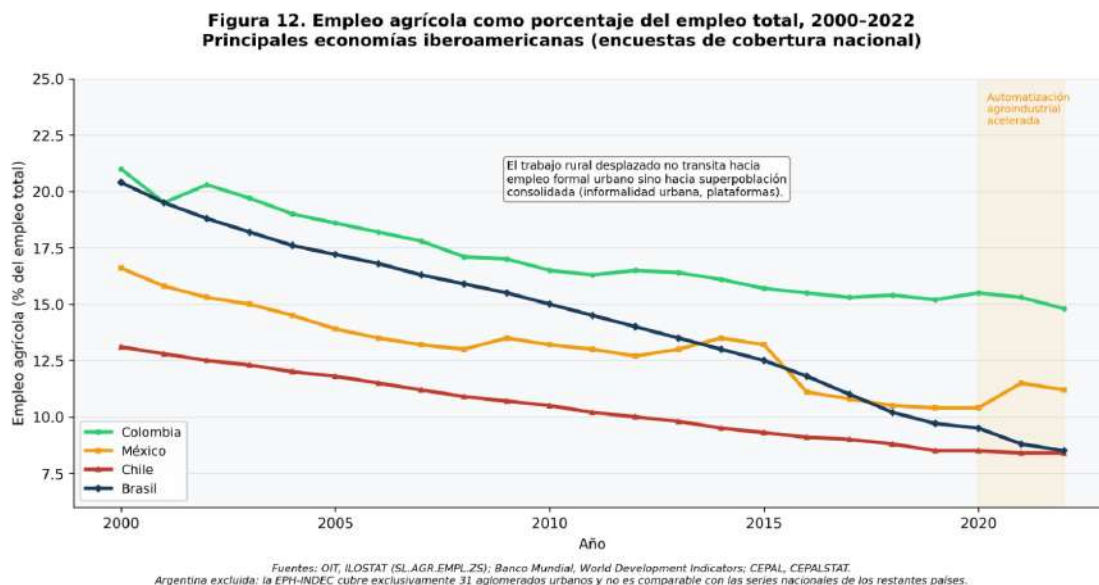


Figura 12: Empleo agrícola como porcentaje del empleo total, 2000–2022

Nota: Fuentes: OIT, ILOSTAT (SL.AGR.EMPL.ZS); Banco Mundial, World Development Indicators; CEPAL, CEPALSTAT. Datos de encuestas de hogares de cobertura nacional. Argentina excluida: la EPH-INDEC cubre exclusivamente 31 aglomerados urbanos y no es comparable con las series nacionales de los restantes países.

### 7.3. Profundización de la superpoblación consolidada: plataformización y renta de monopolio sobre conocimiento objetivado

La superpoblación consolidada experimenta en Iberoamérica una dinámica de doble profundización. Por un lado, los trabajadores desplazados del empleo formal en servicios son absorbidos por formas de trabajo mediado por plataformas digitales que combinan la apariencia de la independencia con la dependencia real respecto de algoritmos que determinan precios, asignan tareas y califican el desempeño. La OIT estimó que en 2022 aproximadamente 12 millones de personas en Iberoamérica trabajaban a través de plataformas digitales, con tasas de crecimiento del 30–40 % anual (OIT, 2022a; Figura 13).

Por otro lado, los procesos de producción de inteligencia artificial requieren grandes volúmenes de trabajo humano para el etiquetado de datos de entrenamiento. Este trabajo, denominado en la literatura anglosajona como ghost work, es realizado por

trabajadores en condiciones de superpoblación consolidada en Venezuela, Colombia y Argentina, con ingresos de entre USD 2 y USD 8 por hora, y constituye una forma de extracción de trabajo que, desde el desarrollo propuesto, debe entenderse no como “extractivismo de datos” (Couldry & Mejias, 2019), que es una categoría incompatible con el desarrollo de la crítica de la economía política, sino como la apropiación de una porción del plusvalor producido por trabajadores que se encuentran en el extremo más profundo del ejército de reserva global.

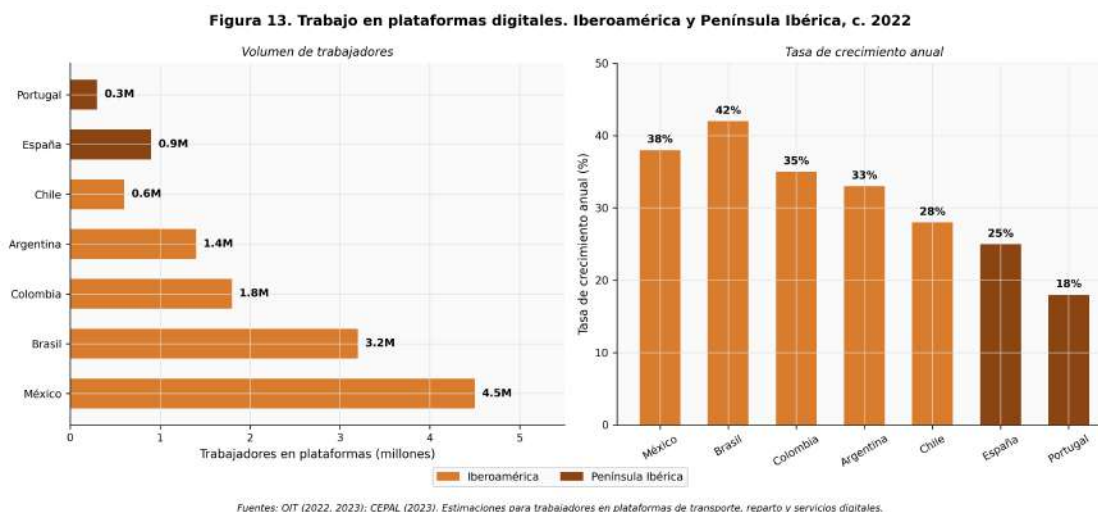


Figura 13: Trabajo en plataformas digitales. Iberoamérica y Península Ibérica, c. 2022

Nota: Fuentes: OIT (2022a, 2023a); CEPAL (2023). Estimaciones para trabajadores en plataformas de transporte, reparto y servicios digitales.

## 8. Análisis comparativo: formas fenoménicas y determinaciones

### 8.1. La comparación estructural

El desarrollo regional de las secciones 5, 6 y 7 permite construir una comparación sistemática de las tres regiones en las dimensiones más relevantes para el argumento. La tabla 2 organiza esa comparación en torno a la categoría que articula teóricamente el conjunto del análisis: la forma dominante de superpoblación relativa que cada región produce. El resto de las dimensiones no son características yuxtapuestas sino determinaciones articuladas que resultan de la posición de cada espacio en la

acumulación mundial.

Tabla 2: Comparación estructural de las tres regiones en torno a la forma dominante de superpoblación

Dimensión	Europa occidental	Península Ibérica	Iberoamérica
Forma dominante de superpoblación	Sobrante extendida a capas medias asalariadas	Sobrante estructural preexistente; nueva presión sobre servicios de calificación media	Latente (expulsión rural) y consolidada profundizada; nuevas formas digitales
Informalidad laboral (c. 2022)	8–15 %	20–22 %	28–80 %
Inversión privada en IA (2024, USD M)	3.800 (UE)	500–900 (estim.)	900
Forma de mediación política del despliegue	Integración regulatoria UE (AI Act, GDPR); Estado social; trayectoria industrial consolidada	Transposición del marco UE; Estado social débil; especialización en turismo, construcción y servicios	Fragmentación regulatoria nacional; mediación estructural de la renta diferencial de la tierra
Mecanismo dominante de transferencia de valor	Pago de licencias y servicios a centros tecnológicos	Intermediación entre UE e Iberoamérica vía corporaciones con origen ibérico	Pago directo a centros tecnológicos; captura de plusvalor vía corporaciones transnacionales
Forma fenoménica dominante	Caída sostenida de la participación salarial; extensión del ejército de reserva a trabajo calificado	Vulnerabilidad compuesta: desempleo estructural convergente con alta exposición en servicios	Dualismo productivo profundizado; plataformización; trabajo de etiquetado en condiciones de superpoblación consolidada

Nota: Elaboración propia. Informalidad: ILOSTAT y Eurostat (c. 2022). Inversión IA: Stanford HAI, AI Index Report 2025. Inversión en Península Ibérica: estimación propia a partir de OCDE.AI y reportes nacionales. Los rangos de informalidad para Iberoamérica reflejan la heterogeneidad interna del espacio regional (Chile 28 %; Bolivia 80 %).

Las tres formas fenoménicas que la tabla sintetiza se apoyan sobre anclajes empíricos convergentes. La Figura 14 expresa el movimiento de la participación salarial en el ingreso nacional durante el período de adopción acelerada de herramientas de inteligencia artificial. La caída es sostenida en Europa occidental y se acentúa en la Península Ibérica; en Iberoamérica opera sobre una base ya históricamente comprimida, lo cual desplaza el ajuste hacia el margen extensivo del mercado de trabajo, que se expresa en la expansión de la superpoblación consolidada, más que hacia el margen intensivo del salario real. Esta asimetría no contradice la hipótesis sobre vulnerabilidad asimétrica del trabajo (H4): la realiza en la forma específica que

impone la posición de cada región en la acumulación mundial.

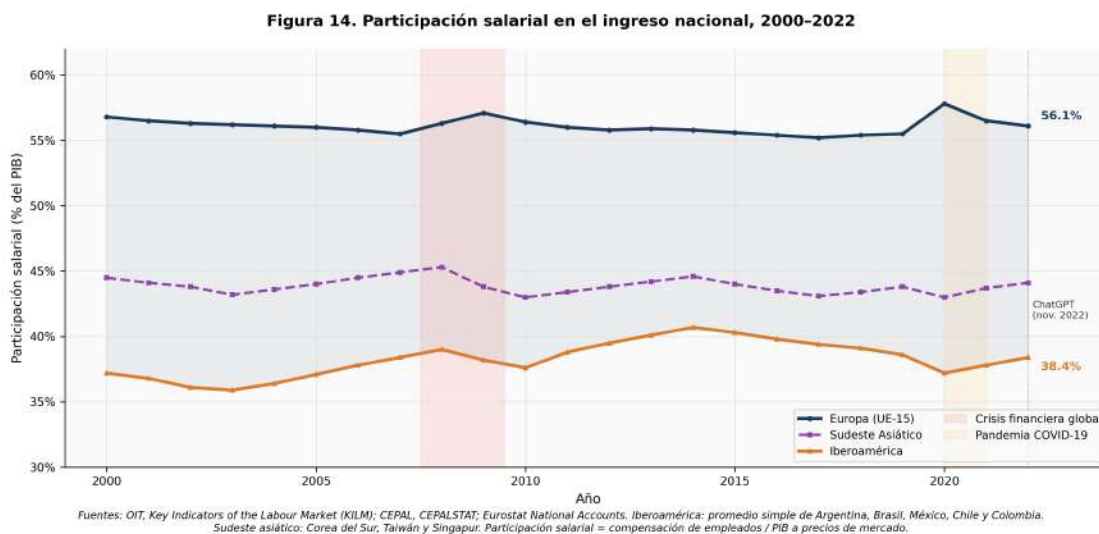


Figura 14: Participación salarial en el ingreso nacional, 2000–2022 (uso comparativo)

Nota: La trayectoria decreciente de la participación salarial es consistente con las determinaciones derivadas en la sección 2, con magnitudes y formas de absorción diferenciadas según la forma dominante de superpoblación de cada región (H4, H6).

La Figura 9 (véase §6.3), en su lectura comparativa, desagrega la exposición ponderada a la automatización por formalidad del empleo y revela el mecanismo diferencial de absorción del impacto. La exposición agregada de la Península Ibérica (37,6 % España; 38,2 % Portugal) es comparable en magnitud a la de Brasil o Chile (39,3 % y 38,5 % respectivamente), pero su distribución es cualitativamente distinta: en la Península Ibérica el impacto recae predominantemente sobre empleo formal —base de la superpoblación sobrante—; en Iberoamérica, sobre empleo informal —base de la superpoblación consolidada—. No se trata de intensidades distintas sobre un mismo sustrato, muestra la misma presión operando sobre estructuras de acumulación heterogéneas.

La Figura 15 completa el cuadro mostrando la trayectoria del coeficiente de Gini del ingreso disponible. La convergencia parcial entre las tres regiones, que se manifiesta en caída sostenida en Iberoamérica desde niveles altos y estabilización o leve aumento en Europa occidental, no debe leerse como movimiento hacia un equilibrio distributivo común. Es, por el contrario, la expresión de procesos distintos: en Iberoamérica, la caída del Gini acompaña políticas de transferencias monetarias que operan sobre una

distribución primaria altamente polarizada sin modificar su estructura; en Europa occidental, la estabilización del Gini se produce sobre una distribución primaria cuya participación salarial cae, en una combinación entre el aumento de la tasa de plusvalor que esa caída expresa y los mecanismos de distribución secundaria —transferencias, impuestos progresivos, gasto social— que el Estado europeo todavía sostiene y que atenúan la traslación de aquella tendencia al ingreso disponible. La magnitud del coeficiente puede converger; la forma de la desigualdad que lo produce se diferencia.

**Figura 15. Coeficiente de Gini del ingreso disponible, 2002-2022**  
Europa (UE-15), Península Ibérica e Iberoamérica

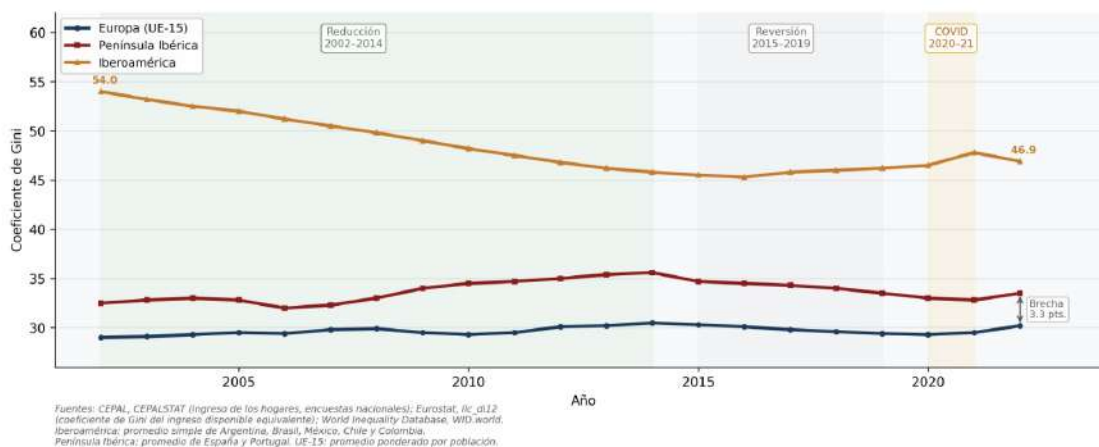


Figura 15: Coeficiente de Gini del ingreso disponible, 2002–2022

Nota: Fuentes: CEPAL, CEPALSTAT (ingreso de los hogares, encuestas nacionales); Eurostat, ilc\_di12 (coeficiente de Gini del ingreso disponible equivalente); World Inequality Database, WID.world. Iberoamérica: promedio simple de Argentina, Brasil, México, Chile y Colombia. Península Ibérica: promedio de España y Portugal. UE-15: promedio ponderado por población. La convergencia parcial entre regiones expresa procesos estructuralmente distintos: recomposición de la distribución primaria en Europa; modulación de la distribución secundaria sobre una base primaria polarizada en Iberoamérica.

## 8.2. Diferenciación regional como necesidad de la determinación estructural

La comparación sistematizada en 8.1 permite precisar el desarrollo abstracto de la diferenciación entre las tres regiones. Los enfoques neoclásico y neoevolucionista schumpeteriano, que serán examinados críticamente en 8.3 a la luz del desarrollo que

hemos hecho, comparten un procedimiento: tratan la diferenciación regional como variación contingente sobre un proceso tecnológico común. En la versión neoclásica, esa variación se explica por diferencias en dotaciones factoriales, calidad institucional o ajuste incompleto de precios; en la neoevolucionista, por las trayectorias diferenciales de los sistemas nacionales de innovación. En ambos casos, la diferenciación es una perturbación respecto de una expectativa de convergencia, condicional en el primer caso (Barro & Sala-i-Martin, 1992), institucionalmente mediada en el segundo, y los resultados observados en cada región se presentan bajo la figura de la paradoja: Europa regula sin producir; Iberoamérica pierde ventajas que derivaban de la baratura de su fuerza de trabajo; la Península Ibérica está integrada sin converger.

Desde el desarrollo propuesto, la operación teórica es inversa. La diferenciación regional no es el residuo por explicar sino la expresión necesaria del proceso a explicar. La ley general de acumulación capitalista se realiza como proceso mundial único cuya unidad se constituye a través de la producción y reproducción de posiciones diferenciales en la división internacional del trabajo. Cada espacio nacional o regional de acumulación no es una instancia independiente del mismo proceso general, sino un momento determinado de su desarrollo, y la forma específica que la superpoblación relativa adopta en cada uno de esos momentos es la realización concreta de determinaciones que operan a escala mundial. La diferenciación, por tanto, no es contingente ni accidental: es la forma misma en que la acumulación mundial existe.

Esta determinación se expresa concretamente en las tres formas fenoménicas identificadas. En Europa occidental, la extensión de la superpoblación sobrante a las capas medias asalariadas no es el resultado de fallas institucionales en la regulación de la IA, es la forma que adopta la elevación de la composición orgánica del capital en un espacio cuya trayectoria industrial consolidada produjo durante décadas una mediación estatal y sindical del valor de la fuerza de trabajo que hoy el capital no puede sostener en las condiciones impuestas por la competencia mundial. En la Península Ibérica, la vulnerabilidad compuesta no es una posición intermedia sobre un continuo, es el resultado de la realización específica de una función de articulación entre el espacio europeo de acumulación y el espacio iberoamericano a través de canales productivos, financieros y migratorios identificados en 6.2. En Iberoamérica, la profundización de la superpoblación consolidada y la emergencia de nuevas formas digitales de plataformización y trabajo de etiquetado no son divergencias respecto de una senda esperada de convergencia productiva: son las formas concretas que el cambio técnico asume en

un espacio cuya acumulación está mediada estructuralmente por la renta diferencial de la tierra, la cual orienta la inversión hacia la exportación de materias primas y desalienta el desarrollo de la subjetividad técnica específica que la producción de tecnología de frontera requiere.

Las tres formas no son tres casos paralelos: son tres momentos articulados de la reorganización del obrero colectivo a escala mundial. El proceso de producción capitalista, en su desarrollo históricamente más avanzado, constituye un obrero colectivo cuya extensión trasciende las fronteras nacionales: incluye tanto al ingeniero de optimización de modelos en California como al trabajador de etiquetado de datos en Caracas, tanto al analista financiero desplazado en Fráncfort como al operador de plataforma de reparto en São Paulo. Lo que la inteligencia artificial reorganiza es la composición interna de ese obrero colectivo mundial: transfiere funciones antes desempeñadas por miembros vivos hacia componentes de capital constante, pero lo hace de manera diferenciada según la posición de cada espacio de acumulación. En los centros productores, esa reorganización crea nuevas funciones de alta composición técnica que se incorporan al obrero colectivo; en Europa occidental, suprime funciones de calificación media sin crear localmente las que las reemplazan; en Iberoamérica, incorpora nuevos miembros al obrero colectivo mundial bajo las formas más degradadas —etiquetado, moderación, gig work— mientras comprime el empleo capitalista formal que constituía el núcleo del obrero colectivo local. La superpoblación consolidada iberoamericana provee el trabajo de etiquetado que permite el entrenamiento de los modelos de inteligencia artificial cuya propiedad se concentra en los espacios productores; la superpoblación sobrante europea ejerce presión sobre las condiciones salariales que se transmite, y en consecuencia sobre el valor de la fuerza de trabajo, por la vía de los capitales que migran hacia regiones donde esperan obtener fuentes de plusvalía extraordinaria en los mercados de trabajo iberoamericanos en los que esos capitales operan; la Península Ibérica funciona como canal bisagra que traduce la presión europea en formas concretas de reorganización productiva en Iberoamérica, y la renta transferida por las filiales iberoamericanas hacia las casas matrices ibéricas se redirige, en parte, hacia los centros tecnológicos norteamericanos y asiáticos vía pago de licencias y servicios. Esta articulación es lo que permite que la hipótesis sobre transferencia de valor (H6) no se limite a describir una asimetría global, sino que identifique los mecanismos concretos a través de los cuales se realiza.

En este registro, la categoría de paradoja pierde toda función analítica. Lo que los

enfoques ortodoxos leen como paradojas son, en el desarrollo propuesto, formas fenoménicas de determinaciones necesarias de un único proceso mundial de acumulación. La tarea teórica no es explicar por qué las regiones divergen respecto de un patrón esperado, sino mostrar cómo la divergencia entre ellas es la forma misma en que la acumulación se realiza. El diagrama H1–H6 presentado en la sección 2.4 puede releerse en este punto como mapa sintético de esa articulación: sus seis determinaciones no son hipótesis paralelas sobre casos distintos, sino momentos sucesivos de un único proceso cuya unidad se realiza en la acumulación mundial.

### 8.3. Crítica comparada de los enfoques alternativos

El modelo de reinstauración de tareas de Acemoglu y Restrepo (2022) sostiene que la automatización destruye puestos en las tareas existentes pero crea simultáneamente nuevas tareas en las que el trabajo humano tiene ventaja comparativa, de modo que el resultado distributivo depende de la velocidad relativa de ambos movimientos. El análisis regional desarrollado contradice esta predicción en los tres espacios considerados. En Europa occidental, las nuevas tareas de alto valor generadas por la inteligencia artificial se concentran en los centros de producción tecnológica —Estados Unidos y, en menor medida, China— y no en los territorios de despliegue; la demanda de trabajo calificado que genera la adopción europea se dirige en buena parte hacia los centros productores bajo la forma de pago de licencias y servicios. En Iberoamérica, las nuevas tareas efectivamente generadas por la inteligencia artificial —trabajo de etiquetado, moderación de contenidos, entrenamiento de modelos— se realizan en condiciones de superpoblación consolidada, con ingresos equivalentes a los del sector informal y sin generación de subjetividad técnica transferible a otras formas de trabajo. En la Península Ibérica, los centros de servicio compartido que absorbían trabajo calificado de mediana remuneración están siendo erosionados por la misma automatización que los creó, sin que se observe la aparición de nuevas tareas que los reemplacen en magnitudes comparables. La ventaja comparativa que el modelo postula presupone una movilidad de capital y de fuerza de trabajo a escala mundial que la acumulación capitalista no realiza: las nuevas tareas se localizan donde la subjetividad ampliada está ya acumulada, y esa localización es precisamente lo que la teoría debería explicar.

El marco neoevolucionista schumpeteriano, en su formulación más influyente —Perez

(2002), Freeman y Louçã (2001)— sostiene que el resultado distributivo de la fase de despliegue de un paradigma tecnoeconómico depende del acoplamiento entre las capacidades productivas desplegadas y el marco institucional que las regula, y que un acoplamiento adecuado —fiscal, educativo, laboral, financiero— puede generar un período de prosperidad generalizada. La dificultad de este enfoque no reside en si Europa ha desplegado o no el acoplamiento institucional completo, sino en un nivel teóricamente anterior: el enfoque no dispone de herramientas para dar cuenta de por qué los sistemas nacionales de innovación se distribuyen como se distribuyen en la acumulación mundial. La asimetría entre la capacidad productiva norteamericana, el despliegue regulatorio europeo y la adopción subordinada iberoamericana es tratada como resultado contingente de trayectorias institucionales divergentes, cuando desde el desarrollo propuesto es el resultado necesario de la forma en que la acumulación capitalista mundial produce y reproduce la concentración de capital constante y de subjetividad técnica en un número reducido de territorios. El neoevolucionismo naturaliza la asimetría que precisamente necesita explicar: la presenta como condición de partida, llamada *path dependency* institucional, cuando esa asimetría es, desde el desarrollo propuesto, el resultado de la tendencia global de acumulación de capital.

La objeción común a ambas tradiciones puede formularse con precisión en este punto. Neoclásicos y neoevolucionistas difieren en muchos puntos, pero comparten un supuesto que el desarrollo propuesto impugna: la neutralidad de la forma mundial de la acumulación respecto del proceso tecnológico. Para el enfoque neoclásico, el proceso tecnológico es un dato exógeno cuyos efectos distributivos se resuelven por ajustes de precios en mercados fundamentalmente simétricos; para el neoevolucionista, es un fenómeno endógeno a sistemas nacionales cuya articulación mundial no es objeto de teoría. En ambos casos, la posición diferencial de cada región en la producción y apropiación de la tecnología queda fuera del *explanans* y pasa al *explanandum* solo como variación institucional. Desde la crítica de la economía política, en cambio, la distribución mundial de la producción tecnológica, manifiesta en su concentración en unos territorios y su ausencia en otros, no es dato ni variación, es el núcleo resultante del proceso que el desarrollo reconstruye. Las divergencias regionales que ambos enfoques registran pero no pueden explicar son expresión de esa determinación estructural.

#### **8.4. La unidad del proceso global: transferencia de valor y posiciones diferenciales en la acumulación mundial**

El análisis regional de las secciones 5, 6 y 7 corre el riesgo de presentar a Europa occidental, la Península Ibérica e Iberoamérica como tres casos independientes sometidos a la misma presión tecnológica. La sección 8.2 mostró por qué esa presentación sería incorrecta desde el registro teórico: la diferenciación entre las tres regiones no es contingente sino necesaria, y es la forma misma en que la acumulación se realiza. Resta mostrar en qué consiste concretamente la unidad del proceso mundial del que esas diferencias son expresión. Esa unidad no es una abstracción agregativa, tenida como la suma de los espacios nacionales que comparten una coyuntura tecnológica, sino la realización de la ley del valor a escala mundial como proceso que produce y reproduce posiciones diferenciales en la producción y apropiación del plusvalor. Los espacios nacionales de acumulación no preexisten a esa realización: son producidos por ella en su diferenciación mutua. La unidad del proceso, en este registro, es indisoluble de la producción de la diferencia.

La determinación más simple de esta unidad requiere partir de la ley del valor en su realización internacional. El valor de las mercancías se determina por el tiempo de trabajo socialmente necesario para su producción, realizado de manera privada e independiente, pero en la economía mundial ese tiempo socialmente necesario se establece a través de la competencia entre capitales que operan con composiciones orgánicas heterogéneas y con niveles de productividad diferenciales. Los capitales que operan con composición orgánica por encima de la media social obtienen, por la vía de la formación de precios de producción, una porción del plusvalor producido por los capitales que operan con composición orgánica por debajo de esa media. Este mecanismo, que Marx identificó como consustancial a la formación de la tasa general de ganancia, opera a escala mundial como transferencia sistemática de valor desde los territorios en los que predomina la producción con menor composición orgánica hacia los territorios en los que predomina la producción con mayor composición orgánica. La concentración de la producción de inteligencia artificial en Estados Unidos y, en menor medida, en China (H1) no es, por tanto, simplemente una característica geográfica del proceso técnico: es la forma territorial en la que se realiza la apropiación del plusvalor producido a escala mundial. La transferencia de valor que la balanza de pagos de servicios de propiedad intelectual y tecnología de la información registra (H6) es la expresión fenoménica de esa determinación más profunda: el déficit sistemático

de las tres regiones analizadas frente al superávit norteamericano no es el resultado de decisiones comerciales ni de desequilibrios corregibles por política tecnológica, sino la forma en que la ley del valor opera en un espacio mundial estructurado por la concentración territorial del capital constante y de la subjetividad técnica acumulada.

Esta unidad se expresa con particular claridad en la balanza de pagos de servicios de propiedad intelectual y tecnología de la información. El déficit sistemático y creciente de las tres regiones en esa balanza —que alcanzó los USD 46 B en Europa (UE-15), USD 14 B en la Península Ibérica y USD 44 B en Iberoamérica en 2022, frente a un superávit norteamericano de USD 218 B— no es un accidente estadístico: es la expresión fenoménica de la transferencia de valor que resulta de la ventaja de composición orgánica de los centros productores de tecnología.

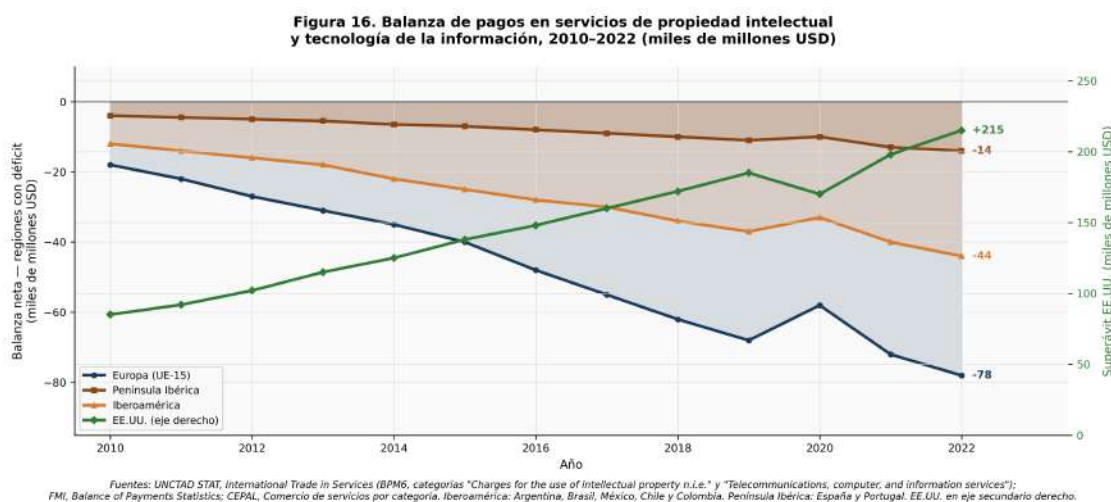


Figura 16: Balanza de pagos en servicios de propiedad intelectual y tecnología de la información, 2010–2022

Nota: Fuentes: UNCTAD STAT, International Trade in Services (clasificación BPM6, categorías Charges for the use of intellectual property n.i.e.z "Telecommunications, computer, and information services"); FMI, Balance of Payments Statistics; CEPAL, Comercio de servicios por categoría. Iberoamérica: Argentina, Brasil, México, Chile y Colombia. Península Ibérica: España y Portugal. EE.UU. en eje secundario derecho (escala distinta).

Esta transferencia opera a través de múltiples mecanismos: el pago de licencias por el uso de modelos de inteligencia artificial propietarios, la contratación de servicios de cómputo en nube a las tres grandes plataformas norteamericanas (Amazon Web

Services, Microsoft Azure, Google Cloud), la adquisición de chips especializados producidos casi exclusivamente por Nvidia y TSMC, y la contratación de consultoría especializada. Todos estos flujos representan, en términos de la crítica de la economía política, transferencias de valor desde los territorios que despliegan la tecnología hacia los que la producen: una forma de renta diferencial sobre conocimiento objetivado cuya barrera de no reproducibilidad opera por vías distintas de las que determinan la renta diferencial de la tierra. En el caso de la renta de la tierra, la no reproducibilidad deriva de la productividad natural diferencial apropiada privadamente; en el caso de la inteligencia artificial, deriva del adelanto masivo de capital constante requerido por cada ciclo de entrenamiento y de la concentración territorial de la subjetividad ampliada acumulada. En ambos casos se trata de condiciones que el capital individual no puede reproducir por su propio movimiento, y que por tanto sostienen la apropiación de una porción del plusvalor producido a escala mundial.

En el espacio iberoamericano de acumulación, estas dos formas de renta no coexisten como mediaciones paralelas sino que operan articuladamente: la primera determina la activación de la segunda. La renta diferencial de la tierra, como Iñigo Carrera (2007, 2008) argumentó, constituye la mediación específica de la inserción iberoamericana en la división internacional del trabajo: la productividad natural diferencial de la tierra pampeana y de otras regiones ricas en recursos naturales orienta la acumulación hacia la exportación de productos primarios, desalienta el desarrollo de una industria nacional competitiva y reproduce sistemáticamente la relación de adaptación tecnológica como resultado estructural y no como opción de política. Esta orientación tiene una consecuencia decisiva sobre la subjetividad ampliada: la acumulación en sectores mediados por renta diferencial no requiere, para su reproducción, el desarrollo de las capacidades científicas y técnicas que la producción de tecnología de frontera exige. La concentración territorial de la subjetividad ampliada en los centros productores de inteligencia artificial no es, por tanto, independiente de la concentración territorial de la renta de la tierra en el espacio iberoamericano: es su correlato necesario. El espacio iberoamericano no desarrolla subjetividad técnica ampliada porque la forma específica de su acumulación no lo requiere; y como consecuencia de esa carencia, cuando el despliegue de inteligencia artificial se generaliza como forma dominante del cambio técnico capitalista, el espacio iberoamericano queda estructuralmente posicionado como receptor y pagador de la renta de monopolio sobre conocimiento objetivado. La transferencia de valor hacia los centros tecnológicos norteamericanos y asiáticos que la Figura 16 registra no es, por tanto, un fenómeno independiente

del patrón primario-exportador: es su forma contemporánea. Las dos rentas son dos momentos articulados de una misma determinación estructural.

La relación entre las tres regiones tampoco es simétrica. La Península Ibérica cumple, en tanto canal bisagra, una función de mediación entre el espacio europeo y el iberoamericano: las corporaciones ibéricas transnacionales transfieren valor desde Iberoamérica hacia sus casas matrices en España y Portugal, que a su vez transfieren valor hacia los centros productores de tecnología a través de los mecanismos descritos. Iberoamérica se encuentra, en esa cadena, en la posición más desfavorable: transfiere valor hacia la Península Ibérica mediante las corporaciones transnacionales, y hacia los centros tecnológicos mediante el pago de licencias y servicios, mientras la mediación de la renta diferencial de la tierra orienta su acumulación hacia sectores que no generan la subjetividad técnica requerida para producir tecnología de frontera. Esta asimetría no es resultado de una jerarquía histórica heredada que pudiera ser revertida por políticas de catching-up: es la forma contemporánea en que se realiza la acumulación mundial, y su reproducción es indisociable de la reproducción misma del proceso.

Esta última observación permite precisar el alcance de H6 como determinación estructural. La transferencia de valor que las tres regiones sufren no es un estado estacionario que pudiera ser corregido: es un movimiento cuya reproducción reproduce y profundiza la posición diferencial de cada región. Los enfoques neoclásico y neoevolucionista, al presentar la asimetría tecnológica como condición de partida corregible —mediante ajuste de precios, inversión en capital humano o políticas de innovación— tratan el problema como si el movimiento hacia la convergencia fuera teóricamente posible y empíricamente bloqueado por fricciones. Desde el desarrollo propuesto, la no convergencia no es una falla del proceso: es su realización. La inteligencia artificial, como forma actual del cambio técnico capitalista, no modifica esa determinación: la intensifica. La concentración de capital constante requerida por cada ciclo de entrenamiento de modelos de frontera eleva la barrera de entrada que define la posición de los centros productores; el desarrollo acumulado de subjetividad técnica específica refuerza la ventaja territorial; y la propiedad privada del conocimiento objetivado bajo forma de modelos, infraestructura y patentes transforma esa ventaja en apropiación continua de renta de monopolio. La unidad del proceso mundial se realiza, por tanto, no como tendencia hacia la homogeneización sino como producción sostenida de la diferencia. Esa es la determinación que el análisis

comparativo de las tres regiones ha permitido identificar.

## 9. Conclusiones

El despliegue masivo de inteligencia artificial no introduce una lógica nueva en el capitalismo: intensifica y despliega en una forma históricamente específica la que Marx identificó como ley general de la acumulación capitalista en el capítulo XXV de *El Capital*. La elevación de la composición orgánica del capital mediante el cambio técnico produce superpoblación relativa como resultado necesario (H1), y la forma concreta que esa superpoblación adopta es diferenciada según la posición de cada espacio nacional de acumulación en el proceso mundial (H2). Esta es la determinación que organiza todo el análisis y de la que las formas regionales examinadas no son casos independientes sino momentos articulados.

La evidencia examinada para el período 2010–2025 es inteligible como realización concreta de esas determinaciones. En Europa occidental, la adopción de inteligencia artificial opera sobre una composición orgánica históricamente elevada y extiende la superpoblación sobrante a capas medias de servicios que en los ciclos anteriores permanecieron al margen (H3); la reducción sostenida de la participación salarial y la polarización ocupacional son la expresión fenoménica de un proceso cuya determinación estructural el Estado europeo no puede alterar, porque la concentración de la subjetividad técnica acumulada se localiza fuera de su espacio de acumulación. En la Península Ibérica, la especialización productiva en servicios de mediana calificación —centros de contacto, back-office financiero, procesamiento administrativo— sitúa al espacio ibérico en una vulnerabilidad compuesta: alta exposición sectorial sin subjetividad técnica reconvertible (H4), sobre un tejido de superpoblación sobrante preexistente. En Iberoamérica, la mediación de la renta diferencial de la tierra orienta la acumulación hacia sectores que no requieren ni reproducen subjetividad técnica ampliada; la superpoblación latente se transforma directamente en consolidada sin pasar por una fase significativa de absorción industrial, y el despliegue de inteligencia artificial se expresa en una profundización de la informalidad y en nuevas formas —plataformización del trabajo, trabajo de etiquetado, gig work— cuya existencia es inseparable del mismo proceso que genera, en el espacio europeo, la extensión de la superpoblación sobrante (H5).

Estas tres formas no son tres casos paralelos sino tres momentos articulados de un único proceso mundial cuya unidad se realiza, precisamente, en la producción sostenida de la diferencia. La transferencia de valor que las tres regiones sufren frente a los centros productores de tecnología —registrada en el déficit sistemático de la balanza de pagos de servicios de propiedad intelectual y tecnología de la información— no es un desequilibrio corregible por política tecnológica (H6), sino la expresión fenoménica de la ventaja de composición orgánica que la concentración territorial del capital constante y de la subjetividad técnica acumulada otorga a los centros productores. La Península Ibérica funciona en esa cadena como canal bisagra: sus corporaciones transnacionales —Telefónica, Santander, BBVA, Iberdrola, Repsol— transfieren valor desde Iberoamérica hacia sus casas matrices, que a su vez transfieren valor hacia los centros tecnológicos norteamericanos y asiáticos bajo forma de pago de licencias, cómputo en nube y servicios asociados. El desarrollo propuesto permite, en este punto, un aporte teórico específico: identificar la renta de monopolio sobre conocimiento objetivado como forma contemporánea del capital constante cuya articulación con la renta diferencial de la tierra es constitutiva —y no meramente yuxtapuesta— de la inserción iberoamericana en la acumulación mundial. Las dos rentas son dos momentos de una misma determinación estructural.

Los enfoques alternativos no están en condiciones de ofrecer esta lectura. El enfoque neoclásico, incluso en sus versiones más elaboradas —SBTC, modelo de tareas de Acemoglu y Restrepo—, requiere tratar las divergencias regionales como resultado de fricciones institucionales ad hoc o presuponer movilidades de capital y trabajo que no existen en la economía mundial. El enfoque neoevolucionista schumpeteriano describe con precisión los patrones de difusión tecnológica pero no puede derivar la diferenciación regional de la lógica misma de la acumulación: debe introducirla como variable exógena bajo la forma de sistemas nacionales de innovación. En ambos casos, la asimetría tecnológica se presenta como condición de partida corregible —por ajuste de precios, inversión en capital humano o políticas de innovación— cuando el análisis desarrollado muestra que esa asimetría es la forma contemporánea en que se realiza la acumulación mundial. La no convergencia no es una falla del proceso: es su realización.

El desarrollo propuesto enfrenta limitaciones que conviene explicitar. La distancia entre las categorías estadísticas disponibles y las categorías teóricas derivadas —señalada en la Sección 3.3— no se resuelve con mejores series: es una consecuencia

necesaria del estatuto diferencial que ambas tienen en la exposición dialéctica. La disponibilidad desigual de datos desagregados sobre adopción de inteligencia artificial, particularmente en Iberoamérica, acota la granularidad del análisis comparativo. Los rezagos entre adopción tecnológica e impacto sobre el mercado de trabajo impiden correspondencias temporales precisas entre ambos planos. Ninguna de estas limitaciones invalida las determinaciones derivadas, pero acotan las inferencias que se pueden extraer de la evidencia estadística cuando se la toma, como aquí, en carácter de expresión fenoménica y no de medición directa.

Varias agendas de investigación se abren a partir de este análisis. En el plano teórico, la determinación de la renta de monopolio sobre conocimiento objetivado como forma específica del capital constante —y su articulación con las demás mediaciones de la acumulación mundial— requiere una formalización que este trabajo solo esboza. En el plano empírico, la medición de los flujos concretos de valor asociados al despliegue de inteligencia artificial exige una reelaboración de las cuentas de balanza de pagos de servicios tecnológicos que vaya más allá de las categorías BPM6 actualmente disponibles. En el plano comparativo, la inclusión de otros espacios de acumulación —Europa del Este, sudeste asiático, China continental como centro emergente de producción tecnológica— permitiría precisar la relación entre composición orgánica del capital, subjetividad técnica acumulada y forma nacional específica de la acumulación. El caso del sudeste asiático presenta una relevancia teórica particular: la combinación de superpoblación latente con subjetividad técnica ampliada que allí se produjo no se produjo en Iberoamérica, y determinar por qué —y si la inteligencia artificial reproduce o altera esa diferencia— es una indagación que el desarrollo propuesto está en condiciones de abordar sin necesidad de recurrir a categorías ajenas a la crítica de la economía política.

## Referencias

Acemoglu, D., & Restrepo, P. (2018). Artificial intelligence, automation, and work (NBER Working Paper No. 24196). National Bureau of Economic Research. <https://doi.org/10.3386/w24196>

Acemoglu, D., & Restrepo, P. (2019). The wrong kind of AI? Artificial intelligence and the future of labor demand (NBER Working Paper No. 25682). National Bureau of

Economic Research. <https://doi.org/10.3386/w25682>

Acemoglu, D., & Restrepo, P. (2022). Tasks, automation, and the rise in US wage inequality. *Econometrica*, 90(5), 1973–2016. <https://doi.org/10.3982/ECTA19815>

Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics*, 118(4), 1279–1333. <https://doi.org/10.1162/003355303322552801>

Banco de España. (2025). Boletín económico 2025/T1: Informe de la economía española. Banco de España.

Banco Mundial. (2023). World development indicators. World Bank Group. <https://databank.worldbank.org/source/world-development-indicators>.

Barro, R. J., & Sala-i-Martin, X. (1992). Convergence. *Journal of Political Economy*, 100(2), 223–251. <https://doi.org/10.1086/261816>

Brynjolfsson, E. (1993). The productivity paradox of information technology. *Communications of the ACM*, 36(12), 66–77. <https://doi.org/10.1145/163298.163309>

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W. W. Norton.

Conference Board. (2023). Total economy database: Output, labor and labor productivity, 1950–2023. The Conference Board. <https://www.conference-board.org/data/economydatabase/>

Comisión Económica para América Latina y el Caribe. (2023). Panorama social de América Latina y el Caribe 2023. CEPAL. <https://repositorio.cepal.org/handle/11362/49678>

Comisión Europea. (2021). Propuesta de directiva del Parlamento Europeo y del Consejo relativa a la mejora de las condiciones laborales en el trabajo en plataformas digitales (COM/2021/762 final). Oficina de Publicaciones de la Unión Europea.

Conferencia de las Naciones Unidas sobre Comercio y Desarrollo. (2023). Digital economy report 2023: Shaping an equitable digital future. UNCTAD. <https://unctad.org/publication/digital-economy-report-2023>

Couldry, N., & Mejias, U. A. (2019). *The costs of connection: How data is colonizing*

human life and appropriating it for capitalism. Stanford University Press.

Dosi, G., Freeman, C., Nelson, R., Silverberg, G., & Soete, L. (Eds.). (1988). *Technical change and economic theory*. Pinter Publishers.

Empresa Brasileira de Pesquisa Agropecuária. (2023). *Balanço social 2023*. Embrapa.

Eurostat. (2023). *Labour force survey 2022: Main results*. European Commission.

Fondo Monetario Internacional. (2023). *Balance of payments statistics yearbook 2023*. FMI. <https://data.imf.org/bop>

Ford, M. (2015). *Rise of the robots: Technology and the threat of a jobless future*. Basic Books.

Freeman, C. (1982). *The economics of industrial innovation* (2.<sup>a</sup> ed.). MIT Press.

Freeman, C., & Louçã, F. (2001). *As time goes by: From the industrial revolutions to the information revolution*. Oxford University Press.

Goldman Sachs Research. (2023). *The potentially large effects of artificial intelligence on economic growth (Global Economics Paper)*. Goldman Sachs Group.

Goos, M., Manning, A., & Salomons, A. (2014). Explaining job polarization: Routine-biased technological change and offshoring. *American Economic Review*, 104(8), 2509–2526. <https://doi.org/10.1257/aer.104.8.2509>

Grinberg, N. (2011). From "Populism" to "Neoliberalism" and back: Capital accumulation and class struggle in Brazil and Argentina. *Historical Materialism*, 19(3), 93–129. <https://doi.org/10.1163/156920611X573598>

Grinberg, N. (2013). Capital accumulation and ground-rent in Brazil, 1980–2005. *Science & Society*, 77(1), 66–93. <https://doi.org/10.1521/asis.2013.77.1.66>

IDC. (2023). *Latin America public cloud services forecast, 2023–2027*. International Data Corporation. (Fig. 2)

Instituto Brasileiro de Geografia y Estadística. (2022). *Pesquisa Nacional por Amostra de Domicílios Contínua: PNAD Contínua 2022*. IBGE. <https://www.ibge.gov.br/estatisticas/sociais/trabalho/17270-pnad-continua.html> (Fig. 7)

Instituto Nacional de Estadística. (2022). *Encuesta Nacional de Empleo, trimestre*

octubre–diciembre 2022. INE Chile. <https://www.ine.gob.cl/estadisticas/sociales/mercado-laboral/ocupacion-y-desocupacion>

Instituto Nacional de Estadística y Censos. (2022). Encuesta Permanente de Hogares: cuarto trimestre de 2022. INDEC. <https://www.indec.gob.ar/indec/web/Nivel4-Tema-4-31-58>

Instituto Nacional de Estadística y Geografía. (2022). Encuesta Nacional de Ocupación y Empleo: ENOE 2022. INEGI. <https://www.inegi.org.mx/programas/enoe/15ymas/> (Fig. 7)

Instituto Nacional de Estadística. (2023). Estadística del padrón continuo a 1 de enero de 2023. INE España. [https://www.ine.es/prensa/pad\\_2023\\_p.pdf](https://www.ine.es/prensa/pad_2023_p.pdf)

Iñigo Carrera, J. (1998). La acumulación de capital en la Argentina. Centro de Investigación como Crítica Práctica.

Iñigo Carrera, J. (2007). La formación económica de la sociedad argentina. Vol. I: Renta agraria, ganancia industrial y deuda externa, 1882–2004. Imago Mundi.

Iñigo Carrera, J. (2008). El capital: Razón histórica, sujeto revolucionario y conciencia. Imago Mundi.

Katz, L. F., & Murphy, K. M. (1992). Changes in relative wages, 1963–1987: Supply and demand factors. *The Quarterly Journal of Economics*, 107(1), 35–78. <https://doi.org/10.2307/2118323>

Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, 22(1), 3–42.

Lundvall, B.-Å. (Ed.). (1992). National systems of innovation: Towards a theory of innovation and interactive learning. Pinter Publishers.

Marx, K. (2007). *Grundrisse: Elementos fundamentales para la crítica de la economía política (borrador) 1857–1858*. Siglo XXI Editores. (Obra original publicada en 1857–1858)

Marx, K. (2008). *El capital: Crítica de la economía política*. Tomo I. Siglo XXI Editores. (Obra original publicada en 1867)

Marx, K. (2009a). *El capital: Crítica de la economía política*. Tomo II. Siglo XXI Editores.

(Obra original publicada en 1885)

Marx, K. (2009b). *El capital: Crítica de la economía política*. Tomo III. Siglo XXI Editores. (Obra original publicada en 1894)

McKinsey Global Institute. (2023). *The economic potential of generative AI: The next productivity frontier*. McKinsey & Company.

Ministerio de Industria, Comercio y Turismo de España. (2023). *DataInVex: Estadísticas de inversión exterior*. Gobierno de España. <https://datainvex.comercio.es/>

Organización Internacional del Trabajo. (2021). *World employment and social outlook 2021: The role of digital labour platforms in transforming the world of work*. OIT. [https://www.ilo.org/global/research/global-reports/weso/2021/WCMS\\_771749/lang--en/index.htm](https://www.ilo.org/global/research/global-reports/weso/2021/WCMS_771749/lang--en/index.htm)

Organización Internacional del Trabajo. (2022a). *El trabajo en las plataformas digitales en América Latina: Reflexiones y perspectivas*. OIT.

Organización Internacional del Trabajo. (2023a). *Panorama laboral 2023: América Latina y el Caribe*. OIT. [https://www.ilo.org/americas/publicaciones/WCMS\\_937406/lang--es/index.htm](https://www.ilo.org/americas/publicaciones/WCMS_937406/lang--es/index.htm)

Organización Internacional del Trabajo. (2023b). *World employment and social outlook 2023: The value of essential work*. OIT. [https://www.ilo.org/global/research/global-reports/weso/WCMS\\_874816/lang--en/index.htm](https://www.ilo.org/global/research/global-reports/weso/WCMS_874816/lang--en/index.htm)

Organización para la Cooperación y el Desarrollo Económicos. (2022). *OECD employment outlook 2022: Building back more inclusive labour markets*. OECD Publishing. <https://doi.org/10.1787/1bb305a6-en>

Organización para la Cooperación y el Desarrollo Económicos. (2023). *OECD employment outlook 2023: Artificial intelligence and the labour market*. OECD Publishing. <https://doi.org/10.1787/08785bba-en>

Perez, C. (2002). *Technological revolutions and financial capital: The dynamics of bubbles and golden ages*. Edward Elgar. <https://doi.org/10.4337/9781781005323>

Red de Indicadores de Ciencia y Tecnología Iberoamericana e Interamericana. (2023). *Indicadores de ciencia y tecnología*. RICYT. <https://www.ricyt.org/indicadores/> (Fig. 4)

Rikap, C. (2021). *Capitalism, power and innovation: Intellectual monopoly capitalism uncovered*. Routledge. <https://doi.org/10.4324/9780429341489>

Romer, P. M. (1986). Increasing returns and long-run growth. *Journal of Political Economy*, 94(5), 1002–1037.

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, 98(5), S71–S102.

Solow, R. M. (1956). A contribution to the theory of economic growth. *The Quarterly Journal of Economics*, 70(1), 65–94. <https://doi.org/10.2307/1884513>

Stanford HAI. (2022). *Artificial intelligence index report 2022*. Stanford University Human-Centered AI Institute.

Stanford HAI. (2023). *Artificial intelligence index report 2023*. Stanford University Human-Centered AI Institute.

Stanford HAI. (2024). *Artificial intelligence index report 2024*. Stanford University Human-Centered AI Institute.

Stanford HAI. (2025). *Artificial intelligence index report 2025*. Stanford University Human-Centered AI Institute.

Starosta, G. (2010). The outsourcing of manufacturing and the rise of giant global contractors. *New Political Economy*, 15(4), 543–563. <https://doi.org/10.1080/13563461003761183>


UNESCO Institute for Statistics. (2023). *UIS.Stat: Research and development expenditure*. UNESCO. <https://data.uis.unesco.org/> (Fig. 4)

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008.

World Inequality Database. (2023). *WID.world: Distributional national accounts*. World Inequality Lab. <https://wid.world/> (Fig. 15)

# IA Pública Segura: Causalidad, Robustez y Decisión

## Safe Public AI: Causality, Robustness, and Decision

Ricardo Alonzo Fernández Salguero  <sup>1</sup>

<sup>1</sup> Universidad Politécnica de Cataluña.

ricardo.alonzo.fernandez@upc.edu (Autor de correspondencia)

**Resumen.** Este trabajo desarrolla una formalización matemática integrada para sistemas de inteligencia artificial orientados a la optimización de políticas públicas bajo restricciones institucionales, presupuestarias, causales y distributivas. El enfoque propuesto articula procesos de decisión de Markov restringidos, medidas ocupacionales, dualidad, condiciones KKT, inferencia causal secuencial, evaluación fuera de política, robustez ante incertidumbre y criterios de seguridad respecto a políticas base. La contribución central consiste en modelar a la IA pública no como un predictor aislado, sino como un operador compuesto que transforma datos observacionales, parámetros normativos y restricciones auditables en una política factible y matemáticamente justificable. A lo largo del texto se demuestran resultados de existencia, realizabilidad de medidas ocupacionales, dualidad fuerte, mejora de política, diferencias de desempeño, sensibilidad ante errores de especificación, contracción robusta bajo rectangularidad y estabilidad local del selector multiobjetivo. Asimismo, se formalizan restricciones de equidad, implementabilidad y estabilidad institucional, mostrando que la optimización pública exige distinguir cuidadosamente entre predicción, causalidad, factibilidad normativa y legitimidad política. El resultado es una arquitectura analítica auditable y extensible para el diseño de sistemas de apoyo a la decisión pública que combinan eficiencia algorítmica con seguridad, robustez y responsabilidad institucional.

**Palabras clave:** IA; políticas públicas; MDP; medidas ocupacionales; dualidad; causalidad; robustez; seguridad; equidad; evaluación contrafactual

**Abstract.** This paper develops an integrated mathematical formalization for artificial intelligence systems oriented toward optimizing public policies under institutional, budgetary, causal, and distributive constraints. The proposed approach articulates constrained Markov decision processes, occupancy measures, duality, KKT conditions, sequential causal inference, off-policy evaluation, robustness under uncertainty, and safety criteria with respect to baseline policies. The central contribution consists in modeling public AI not as an isolated predictor, but as a composite operator that transforms observational data, normative parameters, and auditable constraints into a feasible and mathematically justifiable policy. Throughout the text, results are proved on existence, realizability of occupancy measures, strong duality, policy improvement, performance differences, sensitivity to specification errors, robust contraction under rectangularity, and local stability of the multiobjective selector. Equity, implementability, and institutional stability constraints are also formalized, showing that public optimization requires carefully distinguishing among prediction, causality, normative feasibility, and political legitimacy. The result is an auditable and extensible analytical architecture for designing public decision-support systems that combine algorithmic efficiency with safety, robustness, and institutional responsibility.

**Keywords:** AI; public policy; MDP; occupancy measures; duality; causality; robustness; safety; equity; counterfactual evaluation

**Historial:** Recibido: 28 de marzo de 2026 | Aceptado: 10 de mayo de 2026

**Ética/Financiamiento:** El autor declara no tener conflictos de interés. La investigación fue autofinanciada y no recibió financiamiento externo.

Este documento está bajo una licencia  
CC BY-NC-ND 4.0.

**Cómo citar (APA):** Fernández Salguero, R. A. (2026). IA Pública Segura: Causalidad, Robustez y Decisión. *Conocimiento i Política*, 7(1), 93-160. DOI: 10.64480/cip.2026.n7.03.

## 1. Introducción

La incorporación de inteligencia artificial en la formulación, evaluación y ejecución de políticas públicas constituye una de las transformaciones más significativas de la administración contemporánea. Sin embargo, el uso de algoritmos en este ámbito plantea una exigencia mucho más severa que en contextos comerciales o puramente predictivos. En el sector público no basta con maximizar precisión empírica o rendimiento promedio: toda recomendación algorítmica debe ser compatible con restricciones presupuestarias, estándares de equidad, límites institucionales, criterios de seguridad, trazabilidad causal y exigencias de legitimidad democrática. Dicho de otro modo, una política sugerida por una IA puede ser estadísticamente eficiente y, al mismo tiempo, normativamente inaceptable, causalmente infundada o institucionalmente inviable.

Esta tensión obliga a abandonar la idea de la IA como un simple instrumento de predicción y a reconceptualizarla como un sistema formal de decisión bajo restricciones. En ese marco, los procesos de decisión de Markov restringidos ofrecen una base matemática especialmente adecuada, porque permiten representar dinámicas intertemporales, recompensas sociales, costos, restricciones y trayectorias de decisión dentro de una misma estructura analítica. No obstante, la optimización dinámica por sí sola resulta insuficiente. Si los parámetros del modelo han sido inferidos de datos observacionales sin una capa causal rigurosa, el algoritmo puede optimizar asociaciones espurias. Si el sistema no incorpora robustez frente a incertidumbre, puede colapsar ante shocks o cambios estructurales. Si no se modelan equidad, capacidad administrativa y estabilidad institucional, la solución óptima en el papel puede fracasar en la implementación real.

El propósito de este trabajo es desarrollar una formalización matemática integrada de una IA para políticas públicas que incorpore, de manera explícita y articulada, cinco dimensiones fundamentales: optimización dinámica restringida, identificación causal, evaluación fuera de política, robustez ante incertidumbre y seguridad normativa respecto a una política base. Para ello, el documento construye una arquitectura en capas. La primera capa, causal-estadística, transforma datos en estimaciones de transiciones, recompensas y restricciones bajo supuestos identificables. La segunda capa, de optimización, resuelve un problema dinámico restringido mediante medidas ocupacionales, dualidad y condiciones de optimalidad. La tercera capa, normativa-institucional,

introduce pesos sociales, umbrales distributivos, restricciones de implementabilidad y reglas de despliegue seguro. Esta separación no es meramente expositiva, sino analíticamente decisiva, porque permite distinguir con precisión entre error de modelo, error de optimización y error de transferencia al mundo real.

La relevancia de este enfoque radica en que la política pública es, por naturaleza, un problema secuencial, incierto y normativamente cargado. Las decisiones actuales alteran estados futuros, distribuciones de recursos, incentivos de los agentes y condiciones de factibilidad institucional. Por ello, el diseño de una IA pública exige una teoría que conecte dinámica intertemporal, inferencia causal y restricciones sociales dentro de un mismo marco matemático. Este trabajo responde a esa necesidad mostrando que es posible construir una base formal en la que la eficiencia algorítmica no sea confundida con legitimidad normativa, ni la correlación empírica con causalidad, ni la optimalidad en un modelo estimado con seguridad en el despliegue real.

La tesis general que guía el texto es que una IA pública adecuada debe concebirse como un operador auditable que transforma información empírica y criterios normativos explícitos en políticas factibles, robustas y justificables. Desde esta perspectiva, el objetivo no es automatizar ciegamente la decisión estatal, sino formalizar bajo qué condiciones una recomendación algorítmica puede ser considerada racional, segura y compatible con los fines del gobierno democrático. Así, el documento no sólo desarrolla herramientas matemáticas, sino que también delimita sus supuestos, su alcance y sus límites estructurales, con el fin de contribuir a una teoría rigurosa del uso responsable de la inteligencia artificial en la acción pública.

## **2. Revisión de la literatura**

La integración de la Inteligencia Artificial y el Aprendizaje por Refuerzo, o “Reinforcement Learning” (RL), en el diseño y la ejecución de políticas públicas representa una de las fronteras más prometedoras y matemáticamente desafiantes de la gobernanza moderna. A diferencia de los entornos lúdicos o simulados donde el RL tradicional ha logrado éxitos históricos mediante la exploración irrestricta, el ámbito de la administración pública exige garantías estrictas de seguridad, interpretabilidad y cumplimiento de presupuestos finitos. En este contexto, los Procesos de Decisión de Markov Restringidos, conocidos en la literatura como “Constrained Markov Decision

Processes” (CMDPs), emergen como el marco matemático natural para formular políticas públicas. En un CMDP estándar, el objetivo es maximizar una función de recompensa objetivo, sujeta a restricciones de costo. Formalmente, Feinberg, Jaśkiewicz y Nowak (2019) establecen un modelo donde se busca  $\max_{\pi} J_0(\pi)$  sujeto a  $J_i(\pi) \geq d_i$  para  $i \in I$ , donde  $J_i(\pi) = \mathbb{E}_{\mu}^{\pi}[\sum_{n=1}^{\infty} \beta^{n-1} r_i(s_n, a_n)]$  representa el valor esperado descontado de las recompensas o costos  $r_i$  bajo una política  $\pi$  y un factor de descuento  $\beta$ . Al demostrar que estos modelos admiten políticas óptimas estacionarias y deterministas en espacios de estados de Borel con funciones de recompensa no acotadas, proveen la garantía teórica de que un gobierno puede, matemáticamente, encontrar una estrategia óptima y estable en entornos continuos (como la asignación monetaria de subsidios a lo largo del tiempo) sin violar las restricciones legales impuestas ( $d_i$ ). Complementando esta visión orientada al descuento, Yu (2021) aborda los CMDPs desde la perspectiva del costo promedio a largo plazo mediante programas lineales de dimensión infinita. Su modelo plantea el problema dual de optimización  $\inf_{\gamma} \langle \gamma, c_0 \rangle$  sujeto a  $L\gamma = b$ , donde  $\gamma$  representa la medida de ocupación (la frecuencia con la que un sistema visita un par estado-acción a largo plazo). Al probar la ausencia de brechas de dualidad en estos modelos bajo costos estrictamente no acotados, se habilita el uso de métodos de programación lineal para encontrar pares óptimos de estado-política. Para la gestión pública, esto es vital en el diseño de infraestructuras sostenibles o redes de transporte, donde el rendimiento promedio y el respeto continuo a restricciones de capacidad a lo largo de décadas son más relevantes que las recompensas descontadas a corto plazo.

Para que estos modelos teóricos se traduzcan en herramientas viables para los tomadores de decisiones gubernamentales, la eficiencia muestral es un requisito ineludible, dado que interactuar con la sociedad para “recolectar datos” es sumamente costoso y éticamente delicado. Kalagarla, Jain y Nuzzo (2021) formulan un modelo para CMDPs episódicos de horizonte finito que logra garantías de ser “Probablemente Aproximadamente Correcto” (PAC). Su algoritmo en línea requiere una complejidad muestral delimitada por  $\tilde{O}(|S||A|C^2H^2/\epsilon^2)$ , donde  $|S|$  y  $|A|$  son los tamaños del espacio de estados y acciones,  $H$  es el horizonte temporal, y  $\epsilon$  es el margen de error tolerado. Esta formulación asegura que un programa piloto del gobierno (por ejemplo, una intervención educativa de duración  $H$ ) pueda aprender una política casi óptima sin requerir una experimentación poblacional infinita. En una línea paralela enfocada en múltiples restricciones simultáneas, Buckley, Papathanasiou y Spanopoulos (2025) derivan límites de complejidad muestral utilizando un enfoque primal-dual basado

en modelos generativos. Su avance matemático radica en modelar explícitamente la dicotomía entre viabilidad relajada y estricta. En su modelo, la viabilidad estricta requiere que el valor del costo de la política cumpla inquebrantablemente  $\hat{V}_{c_i}^\pi(\rho) \geq b_i$ , lo que matemáticamente exige una complejidad muestral mucho mayor, de orden  $\tilde{O}(d^3|S||A|/(\epsilon^2\zeta_c^{*2}))$ , donde  $\zeta_c^*$  es la brecha de Slater que mide el margen de holgura de la restricción. Para los formuladores de políticas, esta distinción permite calibrar el riesgo: un programa de optimización de tráfico autónomo requeriría el marco de viabilidad estricta para garantizar cero colisiones, mientras que un sistema de asignación de becas podría operar bajo viabilidad relajada  $\hat{V}_{c_i}^\pi(\rho) \geq b_i - \epsilon$  para acelerar su despliegue a costa de excederse marginalmente en el presupuesto inicial.

Sin embargo, el mundo real en el que operan las políticas públicas rara vez es estático; está caracterizado por shocks económicos, cambios demográficos y crisis imprevistas que alteran las dinámicas subyacentes. Stradi et al. (2024) abordan esta no estacionariedad desarrollando algoritmos para CMDPs con recompensas y restricciones que varían de forma adversaria. Introducen matemáticamente un parámetro de corrupción  $C = \max\{C_G, C_r\}$  que cuantifica la distancia máxima entre las distribuciones de recompensas reales y unas teóricas estacionarias a lo largo de los episodios. Al lograr un arrepentimiento delimitado por  $\tilde{O}(\sqrt{T} + C)$ , demuestran que un sistema de IA puede degradar su rendimiento de manera elegante y controlada cuando las condiciones sociales cambian abruptamente, una característica indispensable para sistemas de triaje hospitalario durante una pandemia, donde los costos y capacidades fluctúan sin previo aviso. Elevando el tratamiento de la incertidumbre, Zuo y He (2025) introducen el algoritmo “SPOT” (“Stochastic Pessimistic-Optimistic Thresholding”), donde el límite de seguridad mismo se modela como un proceso estocástico. En su formulación, el agente no observa un umbral fijo  $\alpha_i$ , sino señales ruidosas  $\tilde{\alpha}_{i,h}(s, a)$ , siendo el umbral latente la suma esperada  $\alpha_i = \sum_{h=1}^H \mathbb{E}[\tilde{\alpha}_{i,h}]$ . Esto es crítico para el diseño de políticas ambientales o de salud pública donde el límite exacto de toxicidad o la tolerancia fiscal de los ciudadanos no se conoce a priori y debe ser inferido dinámicamente, garantizando matemáticamente que el algoritmo respete los límites dinámicos de la sociedad conforme estos se vuelven evidentes.

A medida que los espacios de estados de la administración pública se vuelven masivos, los enfoques tabulares dejan de ser prácticos, haciendo imperativo el uso de aproximación de funciones mediante arquitecturas de aprendizaje profundo. Xu et al. (2025) proponen un algoritmo de “Actor-Crítico Natural Primal-Dual” para CMDPs

de recompensa promedio utilizando parametrización general  $\theta$ . Su modelo optimiza una función Lagrangiana  $\mathcal{L}(\theta, \lambda) = J_r(\theta) + \lambda J_c(\theta)$ , donde las actualizaciones de la política siguen el gradiente de política natural  $\omega_{g,\theta}^* = F(\theta)^{-1} \nabla_{\theta} J_g(\theta)$ , siendo  $F(\theta)$  la matriz de información de Fisher. Al demostrar matemáticamente una convergencia global de  $\tilde{O}(1/\sqrt{T})$ , permiten que el sector público utilice redes neuronales complejas para procesar datos de alta dimensionalidad (como imágenes satelitales o registros censales) con la garantía de que el modelo convergerá a una política segura y óptima. Para hacer este despliegue más seguro y rápido, el paradigma de aprendizaje “Offline-to-Online” (O2O) se vuelve crucial, permitiendo a la IA aprender de datos históricos antes de interactuar con el mundo real. Chen et al. (2026) identifican que el ajuste fino ingenuo (“fine-tuning”) de políticas falla en CMDPs debido al desajuste del multiplicador de Lagrange y a las estimaciones erróneas de los valores Q. Su marco “Marvel” resuelve esto mediante una etapa de prealineación de valor (VPA) que minimiza la pérdida  $\mathcal{L}_Q^{\text{VPA}}(\omega) = \mathbb{E}_{\mathcal{D}}[(Q(s, a) - (r + \gamma \mathbb{E}[\hat{Q}(s', a') - \alpha \log \pi_0(a' | s')]))^2]$ . Matemáticamente, este término de penalización entrópica  $\alpha \log \pi_0$  obliga a las estimaciones a corregirse utilizando solo datos offline, previniendo que la IA asigne costos falsamente altos a acciones fuera de distribución que podrían ser beneficiosas. En la práctica gubernamental, esto permite inicializar un modelo de IA con datos de registros administrativos de la última década y desplegarlo en vivo de forma segura, superando la inercia algorítmica conservadora que impediría mejorar el programa social.

Dado que las políticas públicas deben basarse mayoritariamente en datos observacionales debido a la inviabilidad ética de la experimentación aleatorizada a gran escala, la extracción causal rigurosa de estos datos es imperativa. Athey y Wager (2021) desarrollan un método matemático para el aprendizaje de políticas a partir de datos observacionales utilizando estimación doblemente robusta. Su estimador para el valor de una política se define como  $\hat{V}(\pi) = \frac{1}{n} \sum_{i=1}^n \hat{\Gamma}_i$ , donde el puntaje corregido es  $\hat{\Gamma}_i = \hat{\mu}(X_i, 1) + \frac{W_i}{\hat{e}(X_i)} (Y_i - \hat{\mu}(X_i, 1))$  (para el grupo tratado, y análogamente para control). Este modelo neutraliza matemáticamente los sesgos de selección (endogeneidad) presentes en los datos históricos, garantizando que los formuladores de políticas puedan evaluar el verdadero impacto causal de asignar subsidios o intervenciones policiales. Extendiéndolo a dimensiones dinámicas, Ertefaie y Strawderman (2018) utilizan ecuaciones de Bellman con aprendizaje Q para optimizar regímenes de tratamiento sobre horizontes de tiempo indefinidos, donde el error de diferencia temporal se define como  $\delta_{t+1}(\theta) = R_{t+1} - Q_{\theta}(S_t, A_t) + \gamma \max_{a'} Q_{\theta}(S_{t+1}, a')$ . Este enfoque de

inducción hacia atrás es fundamental para la gestión de enfermedades crónicas o programas de bienestar social prolongados, asegurando que las decisiones de hoy optimicen matemáticamente las trayectorias de salud y económicas de los ciudadanos a lo largo de toda su vida, sin extrapolar irracionalmente más allá del horizonte temporal observable.

Para que los algoritmos derivados de estos datos se desplieguen sin riesgo, la evaluación fuera de política (“Off-Policy Evaluation”, OPE) debe venir con intervalos de confianza rigurosos. Thomas, Theodorou y Ghavamzadeh (2015) formulan estadísticamente la evaluación fuera de política de alta confianza, buscando un límite inferior  $\hat{V}_{\text{lower}}$  tal que  $\Pr(V(\pi_e) \geq \hat{V}_{\text{lower}}) \geq 1 - \delta$ . Utilizando desigualdades de concentración empíricas de Bernstein aplicadas a pesos de muestreo de importancia, su modelo actúa como una red de seguridad matemática, el equivalente a un estudio de impacto ambiental, asegurando a los legisladores con un 95 % de certeza que una nueva intervención algorítmica no será perjudicial. Incluso durante la propia exploración, la seguridad debe estar garantizada. Wendl et al. (2026) desarrollan “SOOPER”, que utiliza modelos dinámicos probabilísticos y un límite de costo pesimista definido como  $\bar{V}_c^{\hat{\pi}}(s) = \max_{\tilde{f} \in \mathcal{F}_n} \mathbb{E}_{\hat{\pi}}[\sum \gamma^t c(s_t, a_t)]$ , donde  $\mathcal{F}_n$  es un conjunto de modelos plausibles. Si la IA estima que la exploración activa violará el presupuesto bajo el peor modelo dinámico posible, el sistema revierte matemáticamente al “prior” conservador. En la infraestructura urbana, esto significa que un algoritmo de semaforización intentará flujos innovadores, pero si el modelo matemático detecta un riesgo inminente de saturación en el peor escenario probabilístico, revertirá inmediatamente al patrón clásico probado, garantizando fluidez sin riesgos.

Más allá del aprendizaje por refuerzo tradicional, las políticas públicas deben diferenciar claramente entre correlación y causalidad. Xiao et al. (2026) logran esto integrando la inferencia causal en políticas de difusión (“Diffusion Models”) mediante su marco “CausalGDP”. Modifican la función de puntuación de la difusión añadiendo un término de orientación causal:  $\nabla_a \log p(s_{t+1}, r_t | s_t, \text{do}(a_t))$ . La inserción del operador  $\text{do}(\cdot)$  de Pearl asegura matemáticamente que la IA genere acciones basándose únicamente en el impacto causal directo sobre las recompensas futuras, bloqueando las variables de confusión (como el estatus socioeconómico en la predicción de criminalidad). De forma complementaria, De Moor et al. (2026) presentan “SCOPE”, que modela la inferencia causal secuencial definiendo la función Q como  $Q_k(\sigma_{:l_k}, a_k) = \mathbb{E}[V_{k+1}(\sigma_{:l_{k+1}}) | \sigma_{:l_k}, a_k]$ . Este modelo de inducción hacia atrás causal

permite al gobierno predecir con precisión matemática cómo una secuencia de intervenciones, por ejemplo, enviar un recordatorio tributario seguido de una auditoría aleatoria, causará de manera secuencial y dependiente un aumento en la recaudación fiscal, optimizando recursos públicos escasos con intervenciones quirúrgicas y demostrables.

El diseño de IA para el sector público requiere intrínsecamente la humildad algorítmica: la capacidad matemática de abstenerse y delegar a un humano cuando la incertidumbre es alta. Sawarni et al. (2026) formalizan esto en su modelo de “Aprendizaje de Políticas con Abstención”. Modifican la función de valor para incluir un tercer estado de acción (\*), definiéndola como  $V^{(p)}(\pi) = \mathbb{E}[\mathbf{1}\{\pi(X) \neq *\}v(\pi, X) + \mathbf{1}\{\pi(X) = *\}g_o(*, X)]$ , donde  $g_o(*, X) = \mathbb{E}[\frac{Y(1)+Y(0)}{2} + p \mid X]$ . En este modelo, el algoritmo recibe el valor de una suposición aleatoria más un bono explícito  $p$  cuando elige abstenerse (\*). Esta penalización y recompensa estructural asegura matemáticamente que, en regiones del espacio de características donde el margen causal es estrecho o la varianza es extrema, la política óptima calculada sea abstenerse. En la práctica gubernamental, esto permite que un sistema automatizado de procesamiento de asilo o asignación de planes de asistencia social apruebe o rechace automáticamente los casos estadísticamente claros y seguros, pero delegue matemáticamente a jueces humanos o trabajadores sociales los casos ambiguos o atípicos. Esta formulación no solo proporciona garantías de arrepentimiento de orden  $\mathcal{O}(1/n)$ , sino que funda una gobernanza pública aumentada por IA donde la robustez distribucional, la equidad, el cumplimiento presupuestario estricto y la rendición de cuentas humana coexisten dentro de un mismo tejido algorítmico y matemático, garantizando políticas públicas eficientes y socialmente responsables.

### 3. Marco formal, notación y arquitectura matemática de la IA pública

Una IA para optimización de políticas públicas no se reduce a un predictor de resultados ni a un optimizador sin restricciones. Debe ser modelada como una composición de módulos con supuestos distinguibles, errores cuantificables y parámetros normativos auditables. La formulación adoptada en este documento no pretende ser una teoría completa de gobierno algorítmico, sino una formalización rigurosa de una clase de sistemas de apoyo a decisión pública donde la decisión se expresa como una

política dinámica factible y donde la inferencia causal, la robustez y la seguridad son partes constitutivas del problema y no añadidos ex post.

El núcleo analítico se apoya en procesos de decisión de Markov (MDP) finitos con descuento, representación por medidas ocupacionales y programación lineal restringida. Esta elección proporciona un lenguaje suficientemente general para incorporar dinámica, incertidumbre y restricciones de política pública. Sin embargo, el documento explicita desde el inicio que el MDP finito estacionario es una aproximación del sistema real, cuya validez depende de supuestos de agregación, observabilidad y estabilidad estructural. La teoría desarrollada permite demostrar resultados exactos en el modelo y, además, incorporar cotas que cuantifican el error cuando se trabaja con un modelo estimado o robustificado.

Se introduce una arquitectura de tres capas. La primera es la capa causal-estadística, que produce estimaciones de transiciones, recompensas y restricciones a partir de datos observacionales o experimentales, junto con descripciones de incertidumbre. La segunda es la capa de optimización, que resuelve un problema dinámico restringido sobre el modelo estimado, incluyendo dualidad y precios sombra. La tercera es la capa normativa-institucional, donde se fijan pesos sociales, umbrales de equidad, presupuestos, reglas de seguridad y límites de implementabilidad. La separación es sustantiva, porque una mejora exacta en la segunda capa puede ser irrelevante o dañina si la primera capa está sesgada o si la tercera capa está mal especificada.

**Definición 1.** Una *IA pública restringida* se modela como un operador

$$\mathcal{A}_n : (\mathcal{D}_n, \mathcal{C}) \mapsto \hat{\pi}_n,$$

donde  $\mathcal{D}_n$  es la información disponible (datos administrativos, encuestas, registros, experimentos, series temporales, etc.),  $\mathcal{C}$  es el conjunto de parámetros normativos e institucionales (pesos, umbrales, restricciones, política base, reglas de seguridad), y  $\hat{\pi}_n$  es una política de decisión factible.

En esta definición, la IA no se identifica con una arquitectura computacional específica (árboles, modelos lineales, redes neuronales, métodos bayesianos o híbridos), sino con un operador de decisión con estructura modular. Esta abstracción permite derivar resultados matemáticos que dependen de propiedades del problema (convexidad, soporte, contracción, regularidad) y no de una implementación particular.

**Definición 2.** Un *MDP de política pública* es una quintupla

$$\mathcal{M} = (\mathcal{S}, \mathcal{A}, P, r, \gamma),$$

donde  $\mathcal{S}$  es un conjunto finito de estados,  $\mathcal{A}$  un conjunto finito de acciones,  $P(s'|s, a)$  una probabilidad de transición,  $r(s, a) \in \mathbb{R}$  una recompensa social inmediata y  $\gamma \in (0, 1)$  un factor de descuento.

En aplicaciones públicas,  $s$  puede resumir condiciones macroeconómicas, sectoriales, distributivas, institucionales y ambientales;  $a$  puede representar intensidad de transferencias, asignaciones de inversión, parámetros regulatorios, esquemas de fiscalización, reglas de focalización u otras combinaciones de instrumentos discretizados. La finitud de  $\mathcal{S}$  y  $\mathcal{A}$  responde a un propósito de tractabilidad y auditabilidad, no a una afirmación ontológica sobre el sistema real.

**Definición 3.** Una política estocástica estacionaria es una familia  $\pi(a|s)$  tal que  $\pi(a|s) \geq 0$  y  $\sum_{a \in \mathcal{A}} \pi(a|s) = 1$  para todo  $s \in \mathcal{S}$ . El conjunto de todas ellas se denota por  $\Pi$ .

Se fija una distribución inicial  $\mu$  sobre  $\mathcal{S}$ . El valor social descontado de una política  $\pi$  es

$$J(\pi) = \mathbb{E}_{s_0 \sim \mu, \pi, P} \left[ \sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \right].$$

Una escalarización operativa de objetivos múltiples puede escribirse como

$$r(s, a) = \sum_{j=1}^q \alpha_j \tilde{f}_j(s, a), \quad \alpha_j \geq 0,$$

donde  $\tilde{f}_j$  son componentes normalizados o reescalados de resultados (actividad, pobreza, desigualdad, cobertura, externalidades, costo fiscal, etc.) y  $\alpha_j$  son pesos normativos auditables.

La normalización es central para la interpretabilidad numérica de  $\alpha_j$ . Si las magnitudes  $f_j$  están en unidades heterogéneas, el uso de pesos directamente sobre  $f_j$  produce parámetros cuya escala depende de la unidad de medición. Para evitar esa

ambigüedad, se introducen transformaciones como

$$\tilde{f}_j(s, a) = \frac{f_j(s, a) - m_j}{\sigma_j} \quad \text{o} \quad \tilde{f}_j(s, a) \in [0, 1],$$

con  $m_j, \sigma_j$  elegidos por reglas transparentes. La matemática no fija la norma de normalización, pero exige que sea explícita si se pretende comparar o auditar pesos.

La arquitectura modular se formaliza mediante la composición

$$\mathcal{D}_n \xrightarrow{\Psi_n} (\hat{P}, \hat{r}, \hat{g}, \hat{\mathcal{U}}, \hat{\mathcal{I}}) \xrightarrow{\mathcal{O}} \hat{\pi} \xrightarrow{\mathcal{S}} \pi^{\text{dep}},$$

donde  $\Psi_n$  es el módulo causal-estadístico,  $\mathcal{O}$  es el optimizador restringido y  $\mathcal{S}$  representa el filtro de seguridad/despliegue que puede rechazar o regularizar la política candidata. El objeto  $\hat{\mathcal{U}}$  codifica incertidumbre de modelo (por ejemplo conjuntos robustos) y  $\hat{\mathcal{I}}$  contiene elementos de inferencia para seguridad estadística (cotas, intervalos, bandas).

La diferencia entre optimalidad en el modelo estimado y desempeño en el sistema real se descompone de forma útil. Si  $\pi^*$  es óptima para el modelo verdadero  $M$  y  $\hat{\pi}$  es la política producida por la IA sobre un modelo estimado  $\hat{M}$ , entonces

$$J_M(\pi^*) - J_M(\hat{\pi}) = \underbrace{J_M(\pi^*) - J_{\hat{M}}(\pi^*)}_{\text{error de modelo en } \pi^*} + \underbrace{J_{\hat{M}}(\pi^*) - J_{\hat{M}}(\hat{\pi})}_{\text{error de optimización}} + \underbrace{J_{\hat{M}}(\hat{\pi}) - J_M(\hat{\pi})}_{\text{error de transferencia}}.$$

Esta identidad es algebraicamente trivial pero conceptualmente decisiva: evita atribuir a la capa de optimización problemas que pertenecen a la capa causal-estadística o al cambio de régimen entre entorno estimado y entorno real.

**Observación 1.** Cuando el sistema real es parcialmente observable, una historia  $H_t$  induce un estado de creencia  $b_t$  sobre estados latentes y el problema natural es un POMDP. El MDP sobre estados agregados  $S_t = \phi(H_t)$  debe entenderse como proyección aproximada. Las garantías exactas de este documento se refieren al MDP modelado, mientras que las garantías respecto al sistema real requieren hipótesis adicionales de agregación suficiente.

Se adopta una política de notación explícita para reducir ambigüedad entre módulos. Las variables de historia causal se denotarán por  $H_t, X_t, A_t$ , las variables del MDP agregado por  $S_t, A_t$  cuando la acción coincida, los funcionales de objetivos por  $F_j(\pi)$ ,

los resultados grupales por  $J_g(\pi)$  o  $\mathbb{E}[Y_g^\pi]$ , y las restricciones por  $G_k(\pi) \leq b_k$ . La distancia de variación total se define con la convención estándar de factor 1/2:

$$TV(p, q) = \frac{1}{2} \sum_x |p(x) - q(x)|.$$

Símbolo	Tipo	Interpretación
$\mathcal{D}_n$	Datos	Base empírica disponible para estimación e inferencia
$\mathcal{C}$	Parámetros	Restricciones, pesos normativos, política base, reglas de seguridad
$\hat{\Psi}_n$	Operador	Módulo causal-estadístico que produce $\hat{P}, \hat{r}, \hat{g}$ y objetos de incertidumbre
$\mathcal{O}$	Operador	Optimizador restringido sobre el modelo estimado/robusto
$\hat{\pi}$	Política	Política candidata producida por la IA
$\pi_b$	Política	Política base o régimen vigente
$P, \hat{P}$	Transición	Modelo verdadero / estimado de dinámica bajo intervención
$r, \hat{r}$	Recompensa	Función social inmediata verdadera / estimada
$G_k(\pi)$	Funcional	Restricción pública agregada (presupuesto, equidad, capacidad, etc.)
$\mathcal{P}_{s,a}$	Conjunto	Incertidumbre robusta de transiciones en $(s, a)$
TV	Distancia	Variación total, $\frac{1}{2} \ p - q\ _1$
KL	Divergencia	Divergencia relativa para estabilidad y regularización

Fuente: Elaboración propia.

#### 4. Formulación dinámica restringida, medidas ocupacionales, realizabilidad, dualidad y mejora de política

La representación por medidas ocupacionales descontadas convierte el problema dinámico en un programa lineal finito cuando  $\mathcal{S}$  y  $\mathcal{A}$  son finitos. Esta transformación es particularmente útil en política pública porque permite incorporar restricciones de presupuesto, equidad, implementabilidad y estabilidad institucional de forma lineal o linealizabile, sin perder interpretabilidad de los multiplicadores duales.

**Definición 4.** Dada una política  $\pi \in \Pi$ , su *medida ocupacional descontada* es

$$d_\pi(s, a) = (1 - \gamma) \sum_{t=0}^{\infty} \gamma^t \mathbb{P}_\pi(S_t = s, A_t = a).$$

**Proposición 1.** Para toda  $\pi \in \Pi$ , la medida  $d_\pi$  satisface

$$d_\pi(s, a) \geq 0, \quad \sum_{s,a} d_\pi(s, a) = 1,$$

y las ecuaciones de flujo

$$\sum_a d_\pi(s, a) = (1 - \gamma)\mu(s) + \gamma \sum_{s',a'} d_\pi(s', a') P(s|s', a').$$

*Demostración.* La no negatividad es inmediata por construcción. Para la suma total,

$$\sum_{s,a} d_\pi(s, a) = (1 - \gamma) \sum_{t \geq 0} \gamma^t \sum_{s,a} \mathbb{P}_\pi(S_t = s, A_t = a) = (1 - \gamma) \sum_{t \geq 0} \gamma^t = 1.$$

Para el flujo, se tiene

$$\sum_a d_\pi(s, a) = (1 - \gamma) \sum_{t \geq 0} \gamma^t \mathbb{P}_\pi(S_t = s).$$

Separando el término  $t = 0$ ,

$$\sum_a d_\pi(s, a) = (1 - \gamma)\mu(s) + (1 - \gamma) \sum_{t \geq 1} \gamma^t \mathbb{P}_\pi(S_t = s).$$

Reindexando con  $u = t - 1$ ,

$$= (1 - \gamma)\mu(s) + \gamma(1 - \gamma) \sum_{u \geq 0} \gamma^u \mathbb{P}_\pi(S_{u+1} = s).$$

Por la ley total de probabilidad y la dinámica del MDP,

$$\mathbb{P}_\pi(S_{u+1} = s) = \sum_{s', a'} \mathbb{P}_\pi(S_u = s', A_u = a') P(s|s', a').$$

Sustituyendo y permutando sumas,

$$\sum_a d_\pi(s, a) = (1 - \gamma)\mu(s) + \gamma \sum_{s', a'} d_\pi(s', a') P(s|s', a').$$

□

**Proposición 2.** *El valor de  $\pi$  se expresa linealmente en  $d_\pi$ :*

$$J(\pi) = \frac{1}{1 - \gamma} \sum_{s, a} d_\pi(s, a) r(s, a).$$

*Demostración.* Por definición de  $d_\pi$ ,

$$\sum_{s, a} d_\pi(s, a) r(s, a) = (1 - \gamma) \sum_{t \geq 0} \gamma^t \mathbb{E}_\pi[r(S_t, A_t)] = (1 - \gamma) J(\pi).$$

Dividiendo por  $1 - \gamma$ , se obtiene la identidad.

□

Las restricciones públicas se modelan mediante funcionales lineales o convexos sobre  $d_\pi$ . Una forma general para restricciones lineales es

$$G_k(\pi) = \frac{1}{1 - \gamma} \sum_{s, a} d_\pi(s, a) g_k(s, a) \leq b_k, \quad k = 1, \dots, m.$$

Los términos  $g_k$  pueden representar costo fiscal, carga administrativa, riesgo ambiental, consumo de capacidad, intensidad de cambio respecto a un régimen base o magnitudes distributivas linearizadas.

La formulación primal en medidas ocupacionales queda

$$\max_d \quad \frac{1}{1-\gamma} \sum_{s,a} d(s,a)r(s,a)$$

sujeta a

$$\sum_a d(s,a) = (1-\gamma)\mu(s) + \gamma \sum_{s',a'} d(s',a')P(s|s',a') \quad \forall s \in \mathcal{S},$$

$$d(s,a) \geq 0 \quad \forall (s,a) \in \mathcal{S} \times \mathcal{A},$$

$$\frac{1}{1-\gamma} \sum_{s,a} d(s,a)g_k(s,a) \leq b_k \quad \forall k = 1, \dots, m.$$

**Teorema 1.** Si  $\mathcal{S}$  y  $\mathcal{A}$  son finitos,  $r$  y  $g_k$  son acotadas y el conjunto factible es no vacío, entonces el problema primal admite una solución óptima.

*Demostración.* Las restricciones definen una intersección de semiespacios y un subespacio afín en  $\mathbb{R}^{|\mathcal{S}||\mathcal{A}|}$ . Además, toda solución factible satisface  $\sum_{s,a} d(s,a) = 1$ , por lo que el conjunto factible está contenido en el simplex y es acotado. Como es cerrado, resulta compacto. El objetivo es lineal y por tanto continuo; la existencia del máximo se sigue por Weierstrass.  $\square$

La realizabilidad de medidas ocupacionales factibles es crucial para cerrar el puente entre el programa lineal y el conjunto de políticas estacionarias.

**Proposición 3** (Realizabilidad de medidas ocupacionales). Sea  $d$  una medida que satisfice no negatividad, suma unitaria y ecuaciones de flujo. Defínase

$$\pi_d(a|s) = \begin{cases} \frac{d(s,a)}{\sum_{a'} d(s,a')}, & \text{si } \sum_{a'} d(s,a') > 0, \\ \bar{\pi}(\cdot|s), & \text{si } \sum_{a'} d(s,a') = 0, \end{cases}$$

donde  $\bar{\pi}(\cdot|s)$  es una distribución arbitraria. Entonces  $d$  es la medida ocupacional descontada de  $\pi_d$ . En particular, no se pierde optimalidad al restringirse a políticas estacionarias aleatorias en el MDP finito descontado.

*Demostración.* Sea  $x(s) = \sum_a d(s,a)$ . Por definición de  $\pi_d$ , se tiene

$$d(s,a) = x(s)\pi_d(a|s) \quad \text{cuando } x(s) > 0.$$

Si  $x(s) = 0$ , entonces  $d(s, a) = 0$  para todo  $a$ . Las ecuaciones de flujo implican

$$x(s) = (1 - \gamma)\mu(s) + \gamma \sum_{s'} x(s') \sum_{a'} \pi_d(a'|s')P(s|s', a').$$

Esta ecuación coincide con la ecuación de balance descontado de las probabilidades marginales ocupacionales inducidas por  $\pi_d$ . Por unicidad de la solución del sistema lineal de balance descontado, las marginales ocupacionales de  $\pi_d$  son  $x(s)$ , y por la factorización  $d(s, a) = x(s)\pi_d(a|s)$ , se concluye que  $d = d_{\pi_d}$ .  $\square$

Esta proposición cierra una objeción frecuente: la programación lineal sobre medidas ocupacionales no introduce objetos espurios en el caso finito descontado, sino una representación equivalente del problema dinámico en la clase de políticas estacionarias aleatorias.

En forma matricial, sea  $n = |\mathcal{S}||\mathcal{A}|$ , sea  $d \in \mathbb{R}^n$  el vector de medidas, sea  $c \in \mathbb{R}^n$  con componentes  $c(s, a) = \frac{1}{1-\gamma}r(s, a)$ , sea  $A \in \mathbb{R}^{|\mathcal{S}| \times n}$  la matriz de flujo y  $G \in \mathbb{R}^{m \times n}$  la matriz de restricciones con

$$(Ad)(s) = \sum_a d(s, a) - \gamma \sum_{s', a'} d(s', a')P(s|s', a'), \quad (Gd)_k = \frac{1}{1-\gamma} \sum_{s, a} d(s, a)g_k(s, a).$$

Entonces el primal se escribe

$$\max_d c^\top d \quad \text{s.a.} \quad Ad = (1 - \gamma)\mu, \quad Gd \leq b, \quad d \geq 0.$$

El dual asociado introduce un multiplicador  $v \in \mathbb{R}^{|\mathcal{S}|}$  para la restricción de flujo y  $\lambda \in \mathbb{R}_+^m$  para las desigualdades:

$$\min_{v, \lambda \geq 0} (1 - \gamma)\mu^\top v + b^\top \lambda$$

sujeto a

$$A^\top v + G^\top \lambda \geq c.$$

Desplegando la restricción coordenada por coordenada se obtiene, para todo  $(s, a)$ ,

$$v(s) - \gamma \sum_{s'} P(s'|s, a)v(s') + \frac{1}{1-\gamma} \sum_{k=1}^m \lambda_k g_k(s, a) \geq \frac{1}{1-\gamma} r(s, a),$$

equivalentemente,

$$v(s) \geq \frac{1}{1-\gamma} \left( r(s, a) - \sum_k \lambda_k g_k(s, a) \right) + \gamma \sum_{s'} P(s'|s, a) v(s').$$

Definiendo la recompensa penalizada  $r_\lambda(s, a) = r(s, a) - \sum_k \lambda_k g_k(s, a)$ , la restricción dual adopta la forma de una desigualdad de Bellman penalizada:

$$v(s) \geq \frac{1}{1-\gamma} r_\lambda(s, a) + \gamma \sum_{s'} P(s'|s, a) v(s').$$

**Teorema 2** (Dualidad fuerte). *Si el primal es factible, entonces el valor primal coincide con el valor dual. Además, si el valor óptimo es finito, existen soluciones óptimas primal y dual.*

*Demostración.* El problema primal y su dual son programas lineales en dimensión finita. El conjunto factible primal es no vacío y acotado por estar contenido en el simplex. El objetivo es finito por acotación de  $r$ . Por el teorema de dualidad fuerte de programación lineal, la brecha dual es cero y existen soluciones óptimas en ambos problemas.  $\square$

**Proposición 4** (Condiciones KKT en forma primal-dual). *Sea  $d^*$  solución óptima primal y  $(v^*, \lambda^*)$  solución óptima dual. Entonces se satisfacen: factibilidad primal, factibilidad dual, complementariedad dual-primal en restricciones*

$$\lambda_k^* ((Gd^*)_k - b_k) = 0, \quad k = 1, \dots, m,$$

*y complementariedad por coordenada*

$$d^*(s, a) \left( v^*(s) - \gamma \sum_{s'} P(s'|s, a) v^*(s') - \frac{1}{1-\gamma} r_{\lambda^*}(s, a) \right) = 0.$$

*Demostración.* Las dos primeras partes corresponden a factibilidad. La complementariedad  $\lambda_k^* ((Gd^*)_k - b_k) = 0$  es la condición estándar de LP entre multiplicadores no negativos y desigualdades. La complementariedad por coordenada proviene de la condición de holgura complementaria entre  $d \geq 0$  y la restricción dual  $A^\top v + G^\top \lambda \geq c$ , expresada en coordenadas  $(s, a)$ .  $\square$

La interpretación institucional es directa:  $\lambda_k^*$  es el precio sombra de la restricción  $k$ . Si

una restricción no es vinculante, su precio sombra es nulo. Si una restricción presiona el óptimo, su multiplicador mide el costo marginal de endurecerla. Esta interpretación es una de las razones por las que la formulación primal-dual es adecuada para auditoría de decisiones públicas.

Se introducen ahora los objetos de valor y ventaja para resultados de mejora de política. Para  $\pi \in \Pi$ ,

$$V^\pi(s) = \mathbb{E}_\pi \left[ \sum_{t=0}^{\infty} \gamma^t r(S_t, A_t) \mid S_0 = s \right], \quad Q^\pi(s, a) = r(s, a) + \gamma \sum_{s'} P(s'|s, a) V^\pi(s').$$

El operador de Bellman de política es

$$(T^\pi V)(s) = \sum_a \pi(a|s) \left( r(s, a) + \gamma \sum_{s'} P(s'|s, a) V(s') \right).$$

**Lema 1.** *El operador  $T^\pi$  es una contracción en norma supremo con módulo  $\gamma$ :*

$$\|T^\pi V - T^\pi W\|_\infty \leq \gamma \|V - W\|_\infty.$$

*Demostración.* Para cada  $s$ ,

$$|(T^\pi V)(s) - (T^\pi W)(s)| \leq \sum_a \pi(a|s) \gamma \sum_{s'} P(s'|s, a) |V(s') - W(s')| \leq \gamma \|V - W\|_\infty.$$

Tomando supremo en  $s$ , se concluye. □

**Corolario 1.** *Para toda política  $\pi$ , existe un único punto fijo  $V^\pi$  de  $T^\pi$ .*

**Teorema 3** (Mejora de política). *Si  $\pi'$  satisface*

$$\sum_a \pi'(a|s) Q^\pi(s, a) \geq V^\pi(s) \quad \forall s \in \mathcal{S},$$

*entonces  $V^{\pi'}(s) \geq V^\pi(s)$  para todo  $s$ , y en particular  $J(\pi') \geq J(\pi)$ .*

*Demostración.* La hipótesis equivale a  $(T^{\pi'} V^\pi)(s) \geq V^\pi(s)$  para todo  $s$ . Por monotonidad de  $T^{\pi'}$ ,

$$V^\pi \leq T^{\pi'} V^\pi \leq (T^{\pi'})^2 V^\pi \leq \dots .$$

La contracción de  $T^{\pi'}$  implica  $(T^{\pi'})^n V^\pi \rightarrow V^{\pi'}$  en norma supremo. Pasando al límite, se obtiene  $V^\pi \leq V^{\pi'}$  coordenada a coordenada.  $\square$

Se define la ventaja de  $\pi$  como

$$A^\pi(s, a) = Q^\pi(s, a) - V^\pi(s).$$

El siguiente resultado es central para seguridad y mejora conservadora.

**Teorema 4** (Lema de diferencia de desempeño). *Para cualesquiera  $\pi, \pi' \in \Pi$ ,*

$$J(\pi') - J(\pi) = \frac{1}{1 - \gamma} \mathbb{E}_{(s,a) \sim d_{\pi'}} [A^\pi(s, a)].$$

*Demostración.* Partiendo de la definición de ventaja,

$$A^\pi(s, a) = r(s, a) + \gamma \sum_{s'} P(s'|s, a) V^\pi(s') - V^\pi(s).$$

Tomando esperanza con respecto a  $d_{\pi'}$  y multiplicando por  $1/(1 - \gamma)$ ,

$$\begin{aligned} \frac{1}{1 - \gamma} \mathbb{E}_{d_{\pi'}} [A^\pi] &= \frac{1}{1 - \gamma} \sum_{s,a} d_{\pi'}(s, a) r(s, a) \\ &\quad + \frac{\gamma}{1 - \gamma} \sum_{s,a} d_{\pi'}(s, a) \\ &\quad \quad \times \sum_{s'} P(s'|s, a) V^\pi(s') \\ &\quad - \frac{1}{1 - \gamma} \sum_s \left( \sum_a d_{\pi'}(s, a) \right) V^\pi(s). \end{aligned}$$

El primer término es  $J(\pi')$ . Para el último, se usa la ecuación de flujo de  $d_{\pi'}$ :

$$\sum_a d_{\pi'}(s, a) = (1 - \gamma)\mu(s) + \gamma \sum_{s', a'} d_{\pi'}(s', a') P(s|s', a').$$

Sustituyendo, los términos con  $\gamma P V^\pi$  se cancelan y queda

$$\frac{1}{1 - \gamma} \mathbb{E}_{d_{\pi'}} [A^\pi] = J(\pi') - \sum_s \mu(s) V^\pi(s) = J(\pi') - J(\pi).$$

$\square$

**Corolario 2.** Si  $\mathbb{E}_{d_{\pi'}}[A^\pi] \geq 0$ , entonces  $J(\pi') \geq J(\pi)$ .

Este lema se usará más adelante para formular seguridad respecto a una política base y para vincular mejora de política con estimación de ventajas bajo incertidumbre.

Objeto	Forma matemática	Función en el marco
Medida ocupacional	$d_\pi(s, a)$	Representa dinámica descontada como LP finito
Restricción pública	$G_k(\pi) \leq b_k$	Incorpora presupuesto, capacidad, equidad, riesgo
Dualidad	$(v, \lambda)$	Explica precios sombra y Bellman penalizado
Realizabilidad	$d \mapsto \pi_d$	Garantiza equivalencia LP-política en MDP finito
Ventaja	$A^\pi = Q^\pi - V^\pi$	Mide ganancia relativa de acciones frente a $\pi$
Lema de desempeño	$J(\pi') - J(\pi) = \frac{1}{1-\gamma} \mathbb{E}_{d_{\pi'}}[A^\pi]$	Base de seguridad y mejora conservadora

Fuente: Elaboración propia.

## 5. Equidad, implementabilidad, estabilidad institucional, multiobjetivo y sensibilidad normativa

La estructura de restricciones públicas no se agota en presupuesto. En políticas públicas, una IA puede ser formalmente óptima y, al mismo tiempo, políticamente inaceptable o normativamente inválida si concentra beneficios, genera daños distributivos severos o exige capacidades administrativas inexistentes. Por ello, la equidad, la implementabilidad y la estabilidad institucional deben ser modeladas explícitamente.

Sea  $\mathcal{G}$  un conjunto finito de grupos relevantes para evaluación distributiva. Para cada  $g \in \mathcal{G}$ , considérese un funcional lineal del tipo

$$\mathbb{E}[Y_g^\pi] = \frac{1}{1-\gamma} \sum_{s,a} d_\pi(s, a) y_g(s, a),$$

donde  $y_g(s, a)$  representa un resultado de interés del grupo  $g$  (cobertura efectiva, ingreso equivalente, tasa de acceso, reducción de tiempos, etc.). Una restricción de brecha media entre grupos  $g, h$  puede escribirse como

$$\left| \mathbb{E}[Y_g^\pi] - \mathbb{E}[Y_h^\pi] \right| \leq \delta_{g,h},$$

que equivale a dos desigualdades lineales.

**Proposición 5.** Si  $d \mapsto \mathbb{E}[Y_g^\pi]$  es lineal para todo  $g \in \mathcal{G}$ , entonces el conjunto de medidas ocupacionales que satisface una familia finita de restricciones de brecha media es convexo.

*Demostración.* Cada desigualdad linealizada define un semiespacio  $a^\top d \leq c$ . La intersección finita de semiespacios es convexa. Intersectando con el conjunto de medidas ocupacionales factibles, que es convexo, se preserva convexidad.  $\square$

La equidad en medias puede ser insuficiente cuando se requiere protección frente a daños de cola. En ese caso, se introducen restricciones de riesgo sobre pérdidas grupales. Sea  $L_g^\pi$  una pérdida aleatoria asociada al grupo  $g$ . La restricción

$$\text{CVaR}_\alpha(L_g^\pi) \leq \bar{\ell}_g$$

controla el promedio de las peores realizaciones por encima del nivel  $\alpha$ . La representación convexa estándar por escenarios es

$$\text{CVaR}_\alpha(L) = \min_{\eta \in \mathbb{R}} \left\{ \eta + \frac{1}{1 - \alpha} \mathbb{E}[(L - \eta)_+] \right\},$$

lo que permite compatibilidad con formulaciones convexas si la pérdida se expresa en una discretización de escenarios o muestras.

Un punto estructural es la posible incompatibilidad entre restricciones de equidad, presupuesto y capacidad. Para tratarla formalmente, se define

$$\mathcal{D}_{\text{fair}} = \{d \in \mathcal{D}_{\text{base}} : \text{restricciones de equidad}\},$$

donde  $\mathcal{D}_{\text{base}}$  denota el conjunto de flujos factibles con restricciones no distributivas. El caso  $\mathcal{D}_{\text{fair}} = \emptyset$  no es patológico desde el punto de vista matemático, sino una señal de incompatibilidad normativa-operativa que requiere relajar o penalizar restricciones.

**Proposición 6.** Si  $\mathcal{D}_{\text{fair}} = \emptyset$ , una alternativa convexa es la formulación penalizada

$$\max_{\pi \in \Pi} J(\pi) - \lambda_{\text{fair}} \Phi(\pi) \quad \text{sujeta a restricciones no distributivas,}$$

donde  $\Phi(\pi) \geq 0$  mide violación agregada de equidad y  $\lambda_{\text{fair}} > 0$  controla la severidad de la penalización.

*Demostración.* La afirmación es constructiva. Si  $\Phi$  es convexa en  $d_\pi$  y el resto de restricciones define un conjunto convexo no vacío, el problema penalizado es una optimización convexa (o lineal-convexa) bien definida. Además,  $\Phi = 0$  en puntos equitativos cuando existen, y  $\Phi > 0$  cuantifica violación cuando no existe factibilidad estricta. □

La implementabilidad se modela con un funcional de carga administrativa  $h(s, a)$  y un presupuesto de capacidad  $\bar{H}$ :

$$G_h(\pi) = \frac{1}{1 - \gamma} \sum_{s,a} d_\pi(s, a) h(s, a) \leq \bar{H}.$$

En aplicaciones,  $h$  puede estimarse por regresión de carga operativa, tiempos de atención, exigencia de sistemas, necesidad de personal o complejidad de coordinación interinstitucional. El punto matemático relevante es que  $h$  es un objeto empírico susceptible de error, lo que permite formular sensibilidad.

**Proposición 7** (Sensibilidad de implementabilidad). Sean  $h$  y  $\hat{h}$  dos funciones de carga. Para toda política  $\pi$ ,

$$|G_h(\pi) - G_{\hat{h}}(\pi)| \leq \frac{1}{1 - \gamma} \sup_{s,a} |h(s, a) - \hat{h}(s, a)|.$$

*Demostración.* Por definición y usando que  $\sum_{s,a} d_\pi(s, a) = 1$ ,

$$\begin{aligned} |G_h(\pi) - G_{\hat{h}}(\pi)| &= \frac{1}{1 - \gamma} \left| \sum_{s,a} d_\pi(s, a) (h(s, a) - \hat{h}(s, a)) \right| \\ &\leq \frac{1}{1 - \gamma} \sum_{s,a} d_\pi(s, a) \sup_{u,v} |h(u, v) - \hat{h}(u, v)|. \end{aligned}$$

Como  $\sum_{s,a} d_\pi(s, a) = 1$ , se obtiene la cota. □

La estabilidad institucional se puede representar mediante restricciones de proximidad a una política base o a su ocupación  $d_{\pi_b}$ . Una forma linealizable es

$$\sum_{s,a} w_{s,a} |d_{\pi}(s,a) - d_{\pi_b}(s,a)| \leq \Delta,$$

con pesos  $w_{s,a} \geq 0$  que capturan sensibilidad operativa o costo político de variar determinadas decisiones. Si los pesos se estiman o fijan normativamente, el modelo conserva auditabilidad al mantenerlos explícitos.

La formulación multiobjetivo se expresa mediante un vector de objetivos

$$\mathbf{F}(\pi) = (F_1(\pi), \dots, F_q(\pi)) \in \mathbb{R}^q.$$

La escalarización lineal

$$J_w(\pi) = \sum_{j=1}^q w_j F_j(\pi), \quad w_j \geq 0, \quad \sum_j w_j = 1,$$

es tractable y transparente, pero sólo recupera puntos soportados de la frontera de Pareto cuando el conjunto de valores alcanzables no es convexo.

**Proposición 8.** Si el conjunto  $\{\mathbf{F}(\pi) : \pi \in \Pi\} \subset \mathbb{R}^q$  es convexo y  $w_j > 0$  para todo  $j$ , toda solución óptima de  $\max_{\pi} J_w(\pi)$  es Pareto-óptima.

*Demostración.* Supóngase que  $\pi^*$  maximiza  $J_w$  pero no es Pareto-óptima. Entonces existe  $\tilde{\pi}$  con  $F_j(\tilde{\pi}) \geq F_j(\pi^*)$  para todo  $j$  y estricta desigualdad para algún  $j_0$ . Multiplicando por  $w_j > 0$  y sumando,

$$J_w(\tilde{\pi}) = \sum_j w_j F_j(\tilde{\pi}) > \sum_j w_j F_j(\pi^*) = J_w(\pi^*),$$

contradicción. □

Una alternativa con mayor control normativo es la formulación  $\epsilon$ -constraint:

$$\max_{\pi} F_1(\pi) \quad \text{s.a.} \quad F_j(\pi) \geq \underline{f}_j, \quad j = 2, \dots, q.$$

Esta forma hace explícitos los mínimos aceptables y evita que un vector de pesos oculte prioridades sustantivas.

La sensibilidad normativa no debe limitarse a la variación del valor de una política fija; también interesa la estabilidad del arg max. Para una política fija  $\pi$ , la desigualdad

$$|J_w(\pi) - J_{\tilde{w}}(\pi)| = \left| \sum_j (w_j - \tilde{w}_j) F_j(\pi) \right| \leq \|w - \tilde{w}\|_1 \max_j |F_j(\pi)|$$

es inmediata. Sin embargo, el cambio de pesos puede alterar discontinuamente la política óptima si existe empate o margen pequeño.

**Teorema 5** (Estabilidad local del selector con margen). *Sea  $w$  un vector de pesos y supóngase que existe una política óptima única  $\pi^*(w)$  para  $J_w$ . Supóngase además que*

$$\Delta(w) := J_w(\pi^*(w)) - \sup_{\pi \neq \pi^*(w)} J_w(\pi) > 0$$

*y que  $\sup_{\pi \in \Pi} \max_j |F_j(\pi)| \leq M < \infty$ . Entonces, para todo  $\tilde{w}$  con*

$$\|\tilde{w} - w\|_1 < \frac{\Delta(w)}{2M},$$

*se mantiene  $\pi^*(\tilde{w}) = \pi^*(w)$ .*

*Demostración.* Sea  $\pi^* = \pi^*(w)$ . Para cualquier  $\pi \neq \pi^*$ ,

$$J_w(\pi^*) - J_w(\pi) \geq \Delta(w).$$

Además,

$$|J_{\tilde{w}}(\pi) - J_w(\pi)| \leq M \|\tilde{w} - w\|_1 \quad \forall \pi.$$

Por tanto,

$$J_{\tilde{w}}(\pi^*) - J_{\tilde{w}}(\pi) \geq J_w(\pi^*) - J_w(\pi) - 2M \|\tilde{w} - w\|_1 > \Delta(w) - 2M \cdot \frac{\Delta(w)}{2M} = 0.$$

Se concluye que  $\pi^*$  sigue dominando estrictamente a toda  $\pi \neq \pi^*$  bajo  $\tilde{w}$ , luego  $\pi^*(\tilde{w}) = \pi^*(w)$ . □

Este resultado formaliza una intuición importante en deliberación pública: si el óptimo tiene margen suficientemente grande, pequeñas variaciones en prioridades normativas no alteran la recomendación. Si el margen es pequeño, la sensibilidad del selector puede ser alta y la IA debe reportar esa fragilidad.

Dimensión	Representación matemática	Propiedad estructural
Equidad de medias	$ \mathbb{E}[Y_g^\pi] - \mathbb{E}[Y_h^\pi]  \leq \delta_{g,h}$	Linealizable; convexa en medidas ocupacionales
Riesgo distributivo	$\text{CVaR}_\alpha(L_g^\pi) \leq \bar{\ell}_g$	Convexo por escenarios
Implementabilidad	$G_h(\pi) \leq \bar{H}$	Lineal si $h$ es lineal en $d$
Estabilidad institucional	$\sum w_{s,a}  d - d_b  \leq \Delta$	Linealizable con variables auxiliares
Multiobjetivo ponderado	$J_w(\pi) = \sum_j w_j F_j(\pi)$	Tractable; sólo puntos soportados en general
$\epsilon$ -constraint	$F_j(\pi) \geq \underline{f}_j$	Umbrales explícitos y auditables
Sensibilidad del selector	$\Delta(w) > 0$ y perturbaciones en $w$	Estabilidad local con margen

Fuente: Elaboración propia.

## 6. Causalidad secuencial, identificación y puente formal con el MDP agregado

La optimización descrita en las secciones anteriores requiere como insumos una dinámica causal  $P$ , una recompensa causal  $r$  y funcionales de restricción  $g_k$  interpretables bajo intervención. En entornos públicos, estos objetos deben inferirse a partir de datos y rara vez son observables de manera directa. Sin una capa causal, la IA puede optimizar correlaciones espurias y producir políticas que mejoran métricas históricas sin causar mejoras reales.

Se considera una historia observada hasta tiempo  $t$ :

$$H_t = (X_0, A_0, X_1, A_1, \dots, X_t),$$

donde  $X_t$  es un vector de covariables observadas y  $A_t$  la acción implementada. Para una secuencia  $\bar{a}_t = (a_0, \dots, a_t)$ , se definen resultados potenciales  $X_{t+1}(\bar{a}_t)$  y, en horizonte  $T$ , un resultado o recompensa acumulable  $Y(\bar{a}_T)$ . Se adopta el marco potencial

con supuestos estándar de consistencia, ignorabilidad secuencial y positividad.

**Supuesto 1** (Consistencia). Si  $\bar{A}_t = \bar{a}_t$ , entonces

$$X_{t+1} = X_{t+1}(\bar{a}_t), \quad Y = Y(\bar{a}_T).$$

**Supuesto 2** (Ignorabilidad secuencial). Para cada  $t$ ,

$$A_t \perp\!\!\!\perp \{X_{t+1}(\bar{a}_t), X_{t+2}(\bar{a}_{t+1}), \dots, Y(\bar{a}_T)\} \mid H_t.$$

**Supuesto 3** (Positividad). Si  $\mathbb{P}(H_t = h_t) > 0$  y  $\pi(a|h_t) > 0$ , entonces

$$\mathbb{P}(A_t = a \mid H_t = h_t) > 0.$$

Estos supuestos permiten identificar el valor causal de una política dinámica  $\pi$ , potencialmente historia-dependiente. Para claridad, se trabaja primero con horizonte finito  $T$ , con recompensa por período  $R_t$ . El valor causal es

$$V_T(\pi) = \mathbb{E} \left[ \sum_{t=0}^T \gamma^t R_t^\pi \right].$$

**Teorema 6** (Identificación secuencial por g-recursión). *Bajo consistencia, ignorabilidad secuencial y positividad, el valor causal  $V_T(\pi)$  es identificable mediante la recursión*

$$V_T(\pi) = \mathbb{E}[m_0^\pi(H_0)],$$

donde

$$m_T^\pi(h_T) = \mathbb{E}[R_T \mid H_T = h_T, A_T \sim \pi(\cdot|h_T)]$$

y, para  $t = T - 1, \dots, 0$ ,

$$m_t^\pi(h_t) = \mathbb{E}[R_t + \gamma m_{t+1}^\pi(H_{t+1}) \mid H_t = h_t, A_t \sim \pi(\cdot|h_t)].$$

*Demostración.* La prueba se basa en iteración de expectativas. Por definición,

$$V_T(\pi) = \mathbb{E} \left[ R_0^\pi + \gamma \sum_{t=1}^T \gamma^{t-1} R_t^\pi \right].$$

Condicionando en  $H_0$  y usando consistencia,

$$V_T(\pi) = \mathbb{E} \left[ \mathbb{E} \left[ R_0 + \gamma \sum_{t=1}^T \gamma^{t-1} R_t^\pi \mid H_0, A_0 \sim \pi(\cdot | H_0) \right] \right].$$

La ignorabilidad secuencial permite reemplazar distribuciones de potenciales por distribuciones observables condicionales en  $H_0$ . Definiendo

$$m_0^\pi(H_0) = \mathbb{E} [R_0 + \gamma m_1^\pi(H_1) \mid H_0, A_0 \sim \pi(\cdot | H_0)],$$

se obtiene  $V_T(\pi) = \mathbb{E}[m_0^\pi(H_0)]$ . Iterando el argumento para  $t = 1, \dots, T$  se obtiene la recursión. La positividad garantiza que las esperanzas condicionales bajo  $\pi$  están definidas en el soporte de los datos.  $\square$

Para conectar esta capa causal con el MDP agregado de optimización, se requiere una aplicación de agregación  $S_t = \phi(H_t)$ . El punto crítico no es sólo la reducción dimensional, sino la preservación de información relevante para efectos de intervención y restricciones.

**Supuesto 4** (Suficiencia causal-Markov de la agregación). Existe una aplicación  $\phi$  tal que  $S_t = \phi(H_t)$  satisface, para toda acción  $a$ ,

$$\mathbb{P}(S_{t+1} \in B \mid H_t, A_t = a) = \mathbb{P}(S_{t+1} \in B \mid S_t, A_t = a)$$

para todo conjunto  $B \subseteq \mathcal{S}$ , y

$$\mathbb{E}[R_t \mid H_t, A_t = a] = \mathbb{E}[R_t \mid S_t, A_t = a].$$

Además, las restricciones  $g_k$  admiten representación agregada  $g_k(S_t, A_t)$  con la misma propiedad de suficiencia para sus esperanzas condicionales.

**Proposición 9** (Puente causal-MDP agregado). *Bajo suficiencia causal-Markov de la agregación, las cantidades  $P(s' | s, a)$ ,  $r(s, a)$  y  $g_k(s, a)$  del MDP agregado representan objetos causales identificables a partir de la historia observada, sujetos a los supuestos de consistencia, ignorabilidad secuencial y positividad.*

*Demostración.* La suficiencia causal-Markov implica que las distribuciones condicionales y esperanzas relevantes dependen de  $H_t$  sólo a través de  $S_t$ . Aplicando el

teorema de g-recursión y sustituyendo condicionales en  $H_t$  por condicionales en  $S_t$ , se obtienen expresiones identificables para transiciones y recompensas agregadas bajo intervención. La positividad asegura soporte para las acciones con probabilidad positiva bajo las políticas a evaluar.  $\square$

Esta proposición no elimina la dificultad de validar empíricamente la suficiencia de  $\phi$ , pero ubica con precisión la exigencia matemática que debe satisfacer una agregación si se quiere interpretar la optimización MDP como optimización causal. Si  $\phi$  falla, el optimizador puede seguir siendo internamente coherente respecto al modelo agregado, pero su interpretación causal pierde fundamento.

La formulación causal estándar también enfrenta límites estructurales que no son detalles técnicos menores. La interferencia entre unidades puede violar la descomposición usual de resultados potenciales, las respuestas estratégicas pueden alterar la ley de transición al cambiar la política, y la no estacionariedad puede invalidar transportabilidad de condicionales estimadas. Estos límites se reconocen formalmente porque determinan si el problema admite una extensión del mismo marco o requiere cambio de paradigma (por ejemplo juegos dinámicos, modelos de red o equilibrio general).

A nivel de una etapa, la g-fórmula estática se recupera como caso particular. Bajo consistencia, ignorabilidad condicional y positividad,

$$\mathbb{E}[Y(a)] = \mathbb{E}[\mathbb{E}[Y | A = a, X]].$$

El caso dinámico es estrictamente más exigente porque la acción altera covariables futuras y éstas condicionan decisiones posteriores. Por esa razón, la identificación secuencial no es un lujo metodológico, sino una condición mínima cuando se optimizan políticas persistentes o adaptativas.

Componente causal	Nivel	Consecuencia si falla
Consistencia	Definición de intervención	Ambigüedad entre observado y potencial
Ignorabilidad secuencial	Identificación dinámica	Sesgo por confusión temporal no controlada
Positividad	Soporte	Imposibilidad de evaluar ciertas políticas con datos disponibles

Componente causal	Nivel	Consecuencia si falla
Suficiencia de agregación $\phi$	Puente causal-MDP	Optimización sobre un estado no causalmente suficiente
Ausencia de interferencia (o controlada)	Estructura de unidades	g-recursión estándar puede dejar de ser válida
Estabilidad estructural / transportabilidad	Temporal o contextual	Fallo de extrapolación de efectos estimados

Fuente: Elaboración propia.

La sección siguiente incorpora evaluación fuera de política y cotas de error de modelo, que enlazan la capa causal-estadística con la capa de optimización y permiten cuantificar el riesgo de optimizar sobre modelos estimados en lugar del sistema real.

## 7. Evaluación fuera de política, doble robustez, clipping, cotas de error y suboptimalidad bajo modelo estimado

En administración pública, el despliegue completo de una nueva política suele ser costoso, lento o políticamente sensible. Por ello, una IA pública debe poder evaluar políticas candidatas usando datos recolectados bajo una política de comportamiento histórica  $\beta$ . Este problema es la evaluación fuera de política (*off-policy evaluation*, OPE). Las razones de importancia ofrecen una identidad exacta bajo soporte común, pero su varianza puede ser alta; los estimadores doblemente robustos reducen fragilidad al combinar reponderación con modelos de resultados.

Considérese un horizonte finito  $T$ , una trayectoria

$$\tau = (s_0, a_0, r_0, \dots, s_T, a_T, r_T, s_{T+1}),$$

y retorno descontado

$$G_T(\tau) = \sum_{t=0}^T \gamma^t r_t.$$

La razón de importancia de trayectoria es

$$W_T(\tau) = \prod_{t=0}^T \frac{\pi(a_t|s_t)}{\beta(a_t|s_t)}.$$

Se asume soporte común:

$$\beta(a|s) > 0 \quad \text{si} \quad \pi(a|s) > 0.$$

**Teorema 7** (Insesgo del estimador de importancia). Si  $\tau_1, \dots, \tau_n$  son trayectorias i.i.d. generadas bajo  $\beta$ , entonces

$$\hat{J}_{\text{IS}} = \frac{1}{n} \sum_{i=1}^n W_T(\tau_i) G_T(\tau_i)$$

es insesgado para  $J_T(\pi) = \mathbb{E}_\pi[G_T]$ .

*Demostración.* La probabilidad de una trayectoria bajo  $\beta$  es

$$p_\beta(\tau) = \mu(s_0) \prod_{t=0}^T \beta(a_t|s_t) P(s_{t+1}|s_t, a_t).$$

Multiplicando por  $W_T(\tau)$ ,

$$W_T(\tau) p_\beta(\tau) = \mu(s_0) \prod_{t=0}^T \pi(a_t|s_t) P(s_{t+1}|s_t, a_t) = p_\pi(\tau).$$

Por tanto,

$$\mathbb{E}_\beta[W_T G_T] = \sum_{\tau} W_T(\tau) G_T(\tau) p_\beta(\tau) = \sum_{\tau} G_T(\tau) p_\pi(\tau) = \mathbb{E}_\pi[G_T].$$

Promediar  $n$  copias preserva el insesgo. □

El insesgo de IS no controla la varianza. Cuando  $\pi$  y  $\beta$  difieren sustancialmente,  $W_T$  puede tener colas pesadas. Una práctica frecuente es usar clipping de razones de importancia para reducir varianza a costa de introducir sesgo.

**Proposición 10** (Sesgo por clipping: cota simple). Sea  $\tilde{W}_T = \min\{W_T, c\}$  con  $c > 0$  y

$$\hat{J}_{\text{CIS}} = \frac{1}{n} \sum_{i=1}^n \tilde{W}_T(\tau_i) G_T(\tau_i).$$

Si  $|G_T| \leq G_{\text{máx}}$  casi seguramente, entonces

$$|\mathbb{E}_\beta[\hat{J}_{\text{CIS}}] - J_T(\pi)| \leq G_{\text{máx}} \mathbb{E}_\beta[(W_T - c)_+].$$

*Demostración.* Por linealidad de la esperanza,

$$\mathbb{E}_\beta[\tilde{W}_T G_T] - \mathbb{E}_\beta[W_T G_T] = \mathbb{E}_\beta[(\tilde{W}_T - W_T) G_T].$$

Como  $\tilde{W}_T - W_T = -(W_T - c)_+$ ,

$$|\mathbb{E}_\beta[\tilde{W}_T G_T] - J_T(\pi)| \leq \mathbb{E}_\beta[(W_T - c)_+ |G_T|] \leq G_{\text{máx}} \mathbb{E}_\beta[(W_T - c)_+].$$

□

Esta cota no pretende ser ajustada, pero formaliza el intercambio varianza-sesgo y justifica reportar explícitamente el nivel de clipping y el sesgo potencial inducido.

Se introduce ahora un estimador doblemente robusto en forma estándar para horizonte finito. Sea  $H_t$  la historia hasta  $t$ ,  $\rho_{0:t} = \prod_{u=0}^t \pi(A_u | H_u) / \beta(A_u | H_u)$ , sea  $\hat{Q}_t(h_t, a)$  una aproximación del valor-acción y

$$\hat{V}_t(h_t) = \sum_a \pi(a | h_t) \hat{Q}_t(h_t, a).$$

Considérese

$$\hat{J}_{\text{DR}} = \frac{1}{n} \sum_{i=1}^n \left[ \hat{V}_0(H_0^i) + \sum_{t=0}^T \gamma^t \rho_{0:t}^i \left( R_t^i + \gamma \hat{V}_{t+1}(H_{t+1}^i) - \hat{Q}_t(H_t^i, A_t^i) \right) \right].$$

**Teorema 8** (Doble robustez, resultado estándar con prueba esquemática). *Bajo positividad, integrabilidad y regularidad suficiente para ley de grandes números,  $\hat{J}_{\text{DR}}$  es consistente para  $J_T(\pi)$  si se cumple al menos una de las siguientes condiciones:*

- 1) las razones de importancia usan la política de comportamiento correcta  $\beta$ ;

2) los modelos de resultado convergen a los verdaderos  $Q_t^\pi$  y  $V_t^\pi$ .

*Demostración.* La demostración estándar se basa en una descomposición telescópica de los residuos

$$R_t + \gamma \hat{V}_{t+1}(H_{t+1}) - \hat{Q}_t(H_t, A_t).$$

Si  $\beta$  es correcta, la reponderación transporta expectativas desde la distribución de comportamiento a la distribución de evaluación; la suma telescópica cancela los errores de aproximación en  $\hat{Q}_t, \hat{V}_t$  en el límite de medias. Si  $\hat{Q}_t, \hat{V}_t$  son correctos, cada residual tiene esperanza condicional nula y la suma colapsa al valor verdadero aun si la parte de reponderación es imperfecta en la parametrización intermedia, siempre que el término implementado converja a una forma equivalente. Bajo condiciones de integrabilidad y convergencia de medias, se obtiene consistencia.  $\square$

El resultado anterior se incluye como teorema estándar porque una demostración completa con todas las hipótesis de regularidad y detalles de recursión excede el alcance de este texto. La utilidad práctica reside en la estructura de tolerancia a error parcial del módulo causal-estadístico, especialmente en entornos con datos administrativos heterogéneos.

Más allá de OPE, la IA pública necesita cuantificar el error de optimizar sobre un modelo estimado  $\hat{M} = (\hat{P}, \hat{r})$  en lugar del modelo verdadero  $M = (P, r)$ . Para una política fija, una cota tipo *simulation lemma* controla la diferencia de valores. Se define

$$\begin{aligned} R_{\text{máx}} &= \sup_{s,a} |r(s, a)|, \\ \varepsilon_r &= \sup_{s,a} |\hat{r}(s, a) - r(s, a)|, \\ \varepsilon_P &= \sup_{s,a} \text{TV}(\hat{P}(\cdot|s, a), P(\cdot|s, a)). \end{aligned}$$

**Teorema 9** (Cota de error de valor por mala especificación del modelo). *Para cualquier política  $\pi$ ,*

$$\|V_{\hat{M}}^\pi - V_M^\pi\|_\infty \leq \frac{\varepsilon_r}{1 - \gamma} + \frac{2\gamma R_{\text{máx}}}{(1 - \gamma)^2} \varepsilon_P.$$

*Demostración.* Usando las ecuaciones de Bellman bajo  $\hat{M}$  y  $M$ ,

$$V_{\hat{M}}^\pi - V_M^\pi = T_{\hat{M}}^\pi V_{\hat{M}}^\pi - T_M^\pi V_M^\pi = (T_{\hat{M}}^\pi V_{\hat{M}}^\pi - T_{\hat{M}}^\pi V_M^\pi) + (T_{\hat{M}}^\pi V_M^\pi - T_M^\pi V_M^\pi).$$

Tomando norma supremo y aplicando contracción de  $T_{\hat{M}}^\pi$ ,

$$\|V_{\hat{M}}^\pi - V_M^\pi\|_\infty \leq \gamma \|V_{\hat{M}}^\pi - V_M^\pi\|_\infty + \|(T_{\hat{M}}^\pi - T_M^\pi)V_M^\pi\|_\infty.$$

Luego,

$$(1 - \gamma) \|V_{\hat{M}}^\pi - V_M^\pi\|_\infty \leq \|(T_{\hat{M}}^\pi - T_M^\pi)V_M^\pi\|_\infty.$$

Para cada  $s$ ,

$$\begin{aligned} \left| ((T_{\hat{M}}^\pi - T_M^\pi)V_M^\pi)(s) \right| &\leq \sum_a \pi(a|s) \left( |\hat{r}(s, a) - r(s, a)| \right. \\ &\quad \left. + \gamma \left| \sum_{s'} (\hat{P} - P)(s'|s, a) V_M^\pi(s') \right| \right). \end{aligned}$$

Usando  $|\hat{r} - r| \leq \varepsilon_r$  y la desigualdad

$$\left| \sum_x (p - q)(x) f(x) \right| \leq 2 \text{TV}(p, q) \|f\|_\infty,$$

se obtiene

$$\|(T_{\hat{M}}^\pi - T_M^\pi)V_M^\pi\|_\infty \leq \varepsilon_r + 2\gamma\varepsilon_P \|V_M^\pi\|_\infty.$$

Como  $\|V_M^\pi\|_\infty \leq R_{\text{máx}}/(1 - \gamma)$ , resulta

$$\|(T_{\hat{M}}^\pi - T_M^\pi)V_M^\pi\|_\infty \leq \varepsilon_r + \frac{2\gamma R_{\text{máx}}}{1 - \gamma} \varepsilon_P.$$

Dividiendo por  $1 - \gamma$ , se obtiene la cota. □

La cota para una política fija es útil, pero la IA selecciona una política  $\hat{\pi}$  optimizando sobre  $\hat{M}$ . Interesa entonces una cota de suboptimalidad en el modelo verdadero.

**Teorema 10** (Cota de suboptimalidad de la política optimizada en  $\hat{M}$ ). *Sea  $\hat{\pi} \in \arg \max_\pi J_{\hat{M}}(\pi)$  y  $\pi^* \in \arg \max_\pi J_M(\pi)$ . Supóngase que*

$$\sup_{\pi \in \Pi} |J_{\hat{M}}(\pi) - J_M(\pi)| \leq \Delta_M.$$

Entonces

$$J_M(\pi^*) - J_M(\hat{\pi}) \leq 2\Delta_M.$$

*Demostración.* Por optimalidad de  $\hat{\pi}$  en  $\hat{M}$ ,

$$J_{\hat{M}}(\hat{\pi}) \geq J_{\hat{M}}(\pi^*).$$

Por la cota uniforme,

$$J_M(\hat{\pi}) \geq J_{\hat{M}}(\hat{\pi}) - \Delta_M \geq J_{\hat{M}}(\pi^*) - \Delta_M \geq J_M(\pi^*) - 2\Delta_M.$$

Reordenando,

$$J_M(\pi^*) - J_M(\hat{\pi}) \leq 2\Delta_M.$$

□

Combinando este teorema con cotas uniformes derivadas de errores en  $P$  y  $r$  se obtiene una garantía explícita de degradación máxima por optimizar sobre un modelo estimado. En contextos públicos, esta degradación es relevante porque permite convertir incertidumbre estadística en márgenes de seguridad o en penalizaciones conservadoras durante el despliegue.

Herramienta	Garantía principal	Riesgo / costo
IS	Insesgo exacto bajo soporte común	Varianza alta cuando $\pi$ y $\beta$ divergen
Clipping de IS	Reduce varianza práctica	Introduce sesgo (cuantificable por colas de $W_T$ )
DR	Consistencia con corrección parcial	Requiere modelos auxiliares e hipótesis adicionales
Cota de valor por modelo	Control de sensibilidad para política fija	Conservadora en horizontes largos $((1 - \gamma)^{-2})$
Cota de suboptimalidad	Control de pérdida por optimizar en $\hat{M}$	Requiere cota uniforme sobre $J_{\hat{M}} - J_M$

*Fuente:* Elaboración propia.

La siguiente sección incorpora robustez frente a incertidumbre de transición, no estacionariedad y seguridad probabilística respecto a una política base, integrando estos elementos con el lema de diferencia de desempeño y con regularización por divergencia.

## 8. Robustez, no estacionariedad, seguridad probabilística, estabilidad por KL y aprendizaje primal-dual

La estimación causal de  $P$  y  $r$  no elimina incertidumbre. En políticas públicas, shocks macroeconómicos, cambios regulatorios, aprendizaje de agentes, crisis y conflictos pueden generar desviaciones sistemáticas entre la dinámica estimada y la realizada. Para tratar esta fragilidad se introduce un MDP robusto con incertidumbre rectangular, se formaliza seguridad respecto a una política base mediante restricciones probabilísticas y se discute convergencia primal-dual en el régimen estándar de aproximación estocástica.

Para cada  $(s, a)$ , se define un conjunto de incertidumbre de transiciones  $\mathcal{P}_{s,a}$ , no vacío, compacto y convexo. Bajo rectangularidad por pares  $(s, a)$ , el operador robusto de Bellman para una política  $\pi$  es

$$(T_R^\pi V)(s) = \sum_a \pi(a|s) \left[ r(s, a) + \gamma \inf_{p \in \mathcal{P}_{s,a}} \sum_{s'} p(s') V(s') \right].$$

**Supuesto 5** (Rectangularidad y compacidad). Para todo  $(s, a)$ ,  $\mathcal{P}_{s,a} \subset \Delta(\mathcal{S})$  es no vacío, compacto y convexo, y la incertidumbre es  $(s, a)$ -rectangular.

**Lema 2.** Para toda política  $\pi$ , el operador  $T_R^\pi$  es una contracción en norma supremo con módulo  $\gamma$ :

$$\|T_R^\pi V - T_R^\pi W\|_\infty \leq \gamma \|V - W\|_\infty.$$

*Demostración.* Defínase  $\phi_{s,a}(V) = \inf_{p \in \mathcal{P}_{s,a}} \sum_{s'} p(s') V(s')$ . Para cualquier  $p \in \mathcal{P}_{s,a}$ ,

$$\sum_{s'} p(s') V(s') \leq \sum_{s'} p(s') W(s') + \|V - W\|_\infty.$$

Tomando ínfimo en  $p$ ,

$$\phi_{s,a}(V) \leq \phi_{s,a}(W) + \|V - W\|_\infty.$$

Intercambiando  $V, W$ , se obtiene

$$|\phi_{s,a}(V) - \phi_{s,a}(W)| \leq \|V - W\|_\infty.$$

Luego, para cada  $s$ ,

$$|(T_R^\pi V)(s) - (T_R^\pi W)(s)| \leq \sum_a \pi(a|s) \gamma |\phi_{s,a}(V) - \phi_{s,a}(W)| \leq \gamma \|V - W\|_\infty.$$

Tomando supremo, queda probado. □

**Corolario 3.** *Existe un único  $V_R^\pi$  tal que  $V_R^\pi = T_R^\pi V_R^\pi$ .*

El operador robusto óptimo se define por

$$(T_R^* V)(s) = \max_{a \in \mathcal{A}} \left[ r(s, a) + \gamma \inf_{p \in \mathcal{P}_{s,a}} \sum_{s'} p(s') V(s') \right].$$

**Teorema 11** (Optimalidad robusta bajo rectangularidad). *Bajo el supuesto de rectangularidad y compacidad,  $T_R^*$  es una contracción con módulo  $\gamma$ , posee un único punto fijo  $V_R^*$ , y toda política codiciosa con respecto a  $V_R^*$  es robustamente óptima.*

*Demostración.* La contracción se demuestra igual que en el lema anterior, sustituyendo promedio por máximo. La existencia y unicidad del punto fijo se obtiene por el teorema de Banach. La optimalidad de una política codiciosa se sigue del argumento estándar de programación dinámica aplicado al operador robusto, cuya consistencia dinámica está garantizada por la rectangularidad. □

La calibración de  $\mathcal{P}_{s,a}$  es un problema sustantivo. El marco admite conjuntos basados en distancia TV, KL,  $\chi^2$ , Wasserstein, o intervalos de confianza multinomiales, siempre que respeten la estructura rectangular si se desea preservar el principio de optimalidad. La matemática separa aquí dos cuestiones: la validez de la programación robusta dada una familia  $\mathcal{P}_{s,a}$ , y la adecuación empírica de esa familia a la incertidumbre real.

**Proposición 11** (Monotonicidad respecto al tamaño del conjunto robusto). *Supóngase que para todo  $(s, a)$ ,  $\mathcal{P}_{s,a}^{(1)} \subseteq \mathcal{P}_{s,a}^{(2)}$ . Denote por  $V_{R,1}^*$  y  $V_{R,2}^*$  los valores robustamente óptimos correspondientes. Entonces*

$$V_{R,1}^*(s) \geq V_{R,2}^*(s) \quad \forall s \in \mathcal{S}.$$

*Demostración.* Para cualquier  $V$  y  $(s, a)$ ,

$$\inf_{p \in \mathcal{P}_{s,a}^{(1)}} \sum_{s'} p(s') V(s') \geq \inf_{p \in \mathcal{P}_{s,a}^{(2)}} \sum_{s'} p(s') V(s').$$

Por tanto,  $T_{R,1}^* V \geq T_{R,2}^* V$  coordenada a coordenada para todo  $V$ . Iterando desde un mismo vector inicial y usando convergencia por contracción de ambos operadores a sus puntos fijos, se concluye  $V_{R,1}^* \geq V_{R,2}^*$ .  $\square$

Esta proposición formaliza que robustecer con conjuntos más grandes reduce (o no aumenta) el valor robusto, lo cual es coherente con la interpretación conservadora de incertidumbre adversarial.

La no estacionariedad se incorpora en forma explícita con horizonte finito y dinámica dependiente del tiempo:

$$P_t(s'|s, a), \quad r_t(s, a), \quad \pi_t(a|s), \quad t = 0, \dots, T.$$

La recursión de Bellman se vuelve

$$V_t^\pi(s) = \sum_a \pi_t(a|s) \left[ r_t(s, a) + \gamma \sum_{s'} P_t(s'|s, a) V_{t+1}^\pi(s') \right], \quad V_{T+1}^\pi \equiv 0.$$

Una forma simple de cuantificar drift temporal en transiciones es

$$\delta_t = \sup_{s,a} \text{TV}(P_t(\cdot|s, a), P_{t+1}(\cdot|s, a)).$$

Si  $\delta_t$  es pequeña y acumulada de manera controlada, las políticas recalibradas periódicamente pueden mantener garantías aproximadas; si el drift es grande o endógeno a la propia política, pueden requerirse marcos adaptativos más complejos.

La seguridad respecto a una política base  $\pi_b$  es una condición normativa y operativa central. Una versión determinista en términos de valores verdaderos es

$$J(\pi) \geq J(\pi_b) - \varepsilon.$$

Sin embargo, en despliegue real  $J$  es desconocido y debe ser estimado. Por ello, se introduce una formulación probabilística con cotas de confianza.

**Definición 5.** Se dirá que una política  $\pi$  satisface *seguridad probabilística* frente a  $\pi_b$  al nivel  $1 - \alpha$  si

$$\mathbb{P}(J(\pi) \geq J(\pi_b) - \varepsilon) \geq 1 - \alpha.$$

Sea  $\hat{J}(\pi)$  un estimador de  $J(\pi)$  y sea  $\text{LCB}_\alpha(\pi)$  una cota inferior tal que

$$\mathbb{P}(J(\pi) \geq \hat{J}(\pi) - \text{LCB}_\alpha(\pi)) \geq 1 - \alpha/2,$$

y  $\text{UCB}_\alpha(\pi_b)$  una cota superior de error para  $\pi_b$  tal que

$$\mathbb{P}(J(\pi_b) \leq \hat{J}(\pi_b) + \text{UCB}_\alpha(\pi_b)) \geq 1 - \alpha/2.$$

**Proposición 12** (Regla suficiente de despliegue seguro). *Si una política  $\pi$  satisface la condición observable*

$$\hat{J}(\pi) - \text{LCB}_\alpha(\pi) \geq \hat{J}(\pi_b) + \text{UCB}_\alpha(\pi_b) - \varepsilon,$$

entonces

$$\mathbb{P}(J(\pi) \geq J(\pi_b) - \varepsilon) \geq 1 - \alpha,$$

suponiendo validez simultánea de las cotas con probabilidad al menos  $1 - \alpha$  por unión.

*Demostración.* Por la validez de las cotas,

$$J(\pi) \geq \hat{J}(\pi) - \text{LCB}_\alpha(\pi), \quad J(\pi_b) \leq \hat{J}(\pi_b) + \text{UCB}_\alpha(\pi_b)$$

con probabilidad al menos  $1 - \alpha$  (por unión de eventos con nivel  $\alpha/2$  cada uno). En ese evento,

$$J(\pi) \geq \hat{J}(\pi) - \text{LCB}_\alpha(\pi) \geq \hat{J}(\pi_b) + \text{UCB}_\alpha(\pi_b) - \varepsilon \geq J(\pi_b) - \varepsilon.$$

□

El lema de diferencia de desempeño permite formular seguridad también en términos de ventajas respecto a  $\pi_b$ :

$$J(\pi) - J(\pi_b) = \frac{1}{1 - \gamma} \mathbb{E}_{d_\pi} [A^{\pi_b}(s, a)].$$

Así, una condición suficiente determinista es

$$\mathbb{E}_{d_\pi}[A^{\pi_b}(s, a)] \geq -(1 - \gamma)\epsilon.$$

En práctica, la dificultad radica en estimar  $A^{\pi_b}$  y  $d_\pi$ ; por ello, la versión probabilística con cotas observables es preferible para despliegue.

Para estabilizar cambios respecto a la política base y reducir varianza de OPE, se usa regularización por divergencia KL. Sea  $\nu$  una distribución de referencia sobre estados y supóngase soporte común o suavización de la política base. La penalización toma la forma

$$\max_{\pi} J(\pi) - \tau \mathbb{E}_{s \sim \nu} [\text{KL}(\pi(\cdot|s) \parallel \pi_b^\epsilon(\cdot|s))],$$

donde

$$\pi_b^\epsilon(\cdot|s) = (1 - \epsilon)\pi_b(\cdot|s) + \epsilon u(\cdot|s), \quad \epsilon \in (0, 1),$$

y  $u(\cdot|s)$  es una distribución de soporte completo.

**Proposición 13.** Para  $\nu$ ,  $\pi_b^\epsilon$  y  $\tau > 0$  fijos, la aplicación

$$\pi \mapsto -\tau \mathbb{E}_{s \sim \nu} [\text{KL}(\pi(\cdot|s) \parallel \pi_b^\epsilon(\cdot|s))]$$

es cóncava en  $\pi$ .

*Demostración.* La divergencia  $\text{KL}(\cdot \parallel \pi_b^\epsilon(\cdot|s))$  es convexa en su primer argumento para cada  $s$ . Multiplicar por  $-\tau$  preserva cóncavidad, y la esperanza ponderada por  $\nu$  también.  $\square$

Se considera el aprendizaje primal-dual estocástico para parametrizaciones  $\pi_\theta$ , con lagrangiano

$$\mathcal{L}(\theta, \lambda) = J(\pi_\theta) - \sum_{k=1}^m \lambda_k (G_k(\pi_\theta) - b_k), \quad \lambda \in \mathbb{R}_+^m.$$

Un esquema básico es

$$\theta_{n+1} = \theta_n + \eta_n \nabla_{\theta} \widehat{\mathcal{L}}(\theta_n, \lambda_n), \quad \lambda_{n+1} = \Pi_{\mathbb{R}_+^m} \left( \lambda_n + \rho_n (\widehat{G}(\theta_n) - b) \right).$$

**Supuesto 6** (Pasos de Robbins–Monro en dos escalas). Las sucesiones  $(\eta_n)$  y  $(\rho_n)$

satisfacen

$$\sum_n \eta_n = \infty, \quad \sum_n \eta_n^2 < \infty, \quad \sum_n \rho_n = \infty, \quad \sum_n \rho_n^2 < \infty,$$

y existe separación de escalas, por ejemplo  $\rho_n/\eta_n \rightarrow 0$  o  $\eta_n/\rho_n \rightarrow 0$ .

**Supuesto 7** (Ruido de gradiente y regularidad). Los estimadores de gradiente son diferencias de martingala con segundo momento uniformemente acotado,  $\mathcal{L}$  es suficientemente suave (Lipschitz de gradientes en regiones relevantes), y los iterados permanecen en una región compacta o efectivamente compacta.

**Teorema 12** (Convergencia primal-dual estocástica, resultado estándar). *Bajo los supuestos anteriores, las trayectorias interpoladas del esquema primal-dual estocástico son asintóticamente cercanas a la dinámica ODE proyectada asociada, y los iterados convergen casi seguramente a un conjunto internamente invariante; en el caso convexo-concavo regular, el límite pertenece al conjunto de puntos KKT, y en el caso no convexo el límite pertenece al conjunto de puntos estacionarios de la dinámica proyectada.*

*Demostración.* La prueba es estándar en aproximación estocástica con dos escalas. La proyección mantiene factibilidad dual, las condiciones de Robbins–Monro y el control de ruido permiten aplicar el método ODE, y la separación de escalas induce una dinámica cuasi-estática para la variable lenta respecto de la rápida. En el caso convexo-concavo, la teoría primal-dual estándar identifica el conjunto de equilibrios con KKT; en no convexidad, la conclusión se formula en términos de conjuntos internamente invariantes/estacionarios de la dinámica proyectada. □

Módulo	Garantía matemática	Supuestos críticos
Robust MDP	Punto fijo y política robusta codiciosa	Rectangularidad, compacidad, $\gamma < 1$
Monotonidad robusta	Conjuntos más grandes $\Rightarrow$ valor robusto menor	Inclusión $\mathcal{P}^{(1)} \subseteq \mathcal{P}^{(2)}$
No estacionariedad finita	Bellman por inducción hacia atrás	Especificación temporal $P_t, r_t$
Seguridad probabilística	No deterioro con probabilidad $\geq 1 - \alpha$	Cotas válidas de inferencia para $\pi, \pi_b$
Regularización por KL	Concavidad del término de estabilidad	Soporte común o suavización $\pi_b^\epsilon$

Módulo	Garantía matemática	Supuestos críticos
Primal-dual estocástico	Convergencia estándar a invariantes/KKT	RM, ruido controlado, suavidad, compacidad efectiva

Fuente: Elaboración propia.

## 9. Caso estático de asignación presupuestaria, complejidad computacional, límites estructurales y síntesis

Como bloque analítico cerrable, se incluye un problema estático de asignación presupuestaria con rendimientos decrecientes y solución cerrada. Este caso sirve como validación interna de la arquitectura: muestra que el marco no depende de manera esencial de métodos numéricos opacos y que conserva interpretabilidad económica incluso en módulos simplificados.

Sea  $x_i \geq 0$  el gasto asignado al programa  $i = 1, \dots, n$ . Se modela la utilidad social de cada programa como

$$u_i(x_i) = a_i \log(1 + b_i x_i), \quad a_i, b_i > 0.$$

El problema es

$$\max_{\mathbf{x} \in \mathbb{R}_+^n} \sum_{i=1}^n a_i \log(1 + b_i x_i) \quad \text{s.a.} \quad \sum_{i=1}^n x_i \leq B.$$

**Teorema 13.** *El problema de asignación presupuestaria anterior tiene solución óptima única.*

*Demostración.* Cada  $u_i$  es continua y estrictamente cóncava en  $\mathbb{R}_+$ , porque

$$u_i''(x_i) = -\frac{a_i b_i^2}{(1 + b_i x_i)^2} < 0.$$

La suma  $\sum_i u_i(x_i)$  es estrictamente cóncava. El conjunto factible  $\{\mathbf{x} \in \mathbb{R}_+^n : \sum_i x_i \leq B\}$  es no vacío, cerrado y acotado, luego compacto. Por Weierstrass existe solución; por estricta concavidad en un conjunto convexo, la solución es única.  $\square$

**Teorema 14** (Solución cerrada tipo *water-filling*). Existe  $\lambda^* \geq 0$  tal que la solución óptima única satisface

$$x_i^* = \max \left\{ 0, \frac{a_i}{\lambda^*} - \frac{1}{b_i} \right\}, \quad i = 1, \dots, n,$$

y además  $\sum_i x_i^* = B$ .

*Demostración.* El lagrangiano con multiplicadores  $\lambda \geq 0$  para presupuesto y  $v_i \geq 0$  para no negatividad es

$$\mathcal{L}(\mathbf{x}, \lambda, \boldsymbol{\nu}) = \sum_i a_i \log(1 + b_i x_i) - \lambda \left( \sum_i x_i - B \right) + \sum_i v_i x_i.$$

La estacionariedad exige

$$\frac{a_i b_i}{1 + b_i x_i} - \lambda + v_i = 0.$$

Si  $x_i^* > 0$ , por complementariedad  $v_i^* = 0$ , luego

$$\frac{a_i b_i}{1 + b_i x_i^*} = \lambda^* \implies x_i^* = \frac{a_i}{\lambda^*} - \frac{1}{b_i}.$$

Si este valor es negativo, la restricción  $x_i \geq 0$  se activa y  $x_i^* = 0$ . En conjunto,

$$x_i^* = \max \left\{ 0, \frac{a_i}{\lambda^*} - \frac{1}{b_i} \right\}.$$

Resta probar que el presupuesto es vinculante. Supóngase por contradicción que  $\sum_i x_i^* < B$ . Entonces, por complementariedad de la restricción presupuestaria,  $\lambda^* = 0$ .

La estacionariedad implica

$$\frac{a_i b_i}{1 + b_i x_i^*} + v_i^* = 0.$$

Pero  $\frac{a_i b_i}{1 + b_i x_i^*} > 0$  y  $v_i^* \geq 0$ , lo que es imposible. Equivalente y operativamente, si hubiera holgura presupuestaria, la derivada marginal de algún programa sería positiva y podría aumentarse su gasto mejorando el objetivo. Por tanto,  $\sum_i x_i^* = B$ .  $\square$

Este bloque muestra cómo los precios sombra aparecen en una forma analítica simple. El multiplicador  $\lambda^*$  iguala utilidades marginales ajustadas para programas activos y excluye aquellos cuya productividad marginal inicial  $a_i b_i$  no justifica asignación positiva bajo la restricción de presupuesto. En contextos públicos, esta lógica permite

interpretar decisiones de priorización sin recurrir a cajas negras.

La escalabilidad computacional merece tratamiento explícito, porque el título del documento incluye IA y el núcleo exacto del texto usa MDP finitos. En espacios de estado grandes, la representación exacta por medidas ocupacionales puede ser inviable. Si  $|\mathcal{S}| = S$  y  $|\mathcal{A}| = A$ , el primal exacto involucra  $SA$  variables básicas (más auxiliares para linealizaciones) y al menos  $S + m$  restricciones principales, lo que crece rápidamente con discretizaciones finas.

Para enfrentar esta limitación, se suele recurrir a aproximación funcional, agregación de estados, muestreo o descomposición. Matemáticamente, ello transforma garantías exactas en garantías aproximadas. Una manera abstracta de capturar esta transición es introducir un operador de proyección  $\Pi_{\mathcal{F}}$  sobre una clase de funciones/medidas aproximantes  $\mathcal{F}$ , de modo que la ecuación de Bellman o el problema primal se resuelvan en una subclase restringida. El error total se descompone entonces en error de aproximación, error estadístico y error de optimización numérica.

**Proposición 14** (Descomposición abstracta de error con aproximación funcional). *Sea  $\pi^*$  una política óptima en la clase exacta  $\Pi$  y  $\hat{\pi}_{\mathcal{F}}$  una política obtenida al optimizar sobre una clase aproximante  $\Pi_{\mathcal{F}} \subseteq \Pi$  con modelo estimado. Entonces, para todo entorno  $M$ ,*

$$\begin{aligned}
 J_M(\pi^*) - J_M(\hat{\pi}_{\mathcal{F}}) &\leq \underbrace{\inf_{\pi \in \Pi_{\mathcal{F}}} (J_M(\pi^*) - J_M(\pi))}_{\text{error de aproximación}} \\
 &\quad + \underbrace{\sup_{\pi \in \Pi_{\mathcal{F}}} |J_M(\pi) - J_{\hat{M}}(\pi)|}_{\text{error estadístico/modelo}} \\
 &\quad + \underbrace{\left( \sup_{\pi \in \Pi_{\mathcal{F}}} J_{\hat{M}}(\pi) - J_{\hat{M}}(\hat{\pi}_{\mathcal{F}}) \right)}_{\text{error de optimización numérica}}.
 \end{aligned}$$

*Demostración.* Para cualquier  $\tilde{\pi} \in \Pi_{\mathcal{F}}$ ,

$$\begin{aligned}
 J_M(\pi^*) - J_M(\hat{\pi}_{\mathcal{F}}) &= (J_M(\pi^*) - J_M(\tilde{\pi})) + (J_M(\tilde{\pi}) - J_{\hat{M}}(\tilde{\pi})) \\
 &\quad + (J_{\hat{M}}(\tilde{\pi}) - J_{\hat{M}}(\hat{\pi}_{\mathcal{F}})) \\
 &\quad + (J_{\hat{M}}(\hat{\pi}_{\mathcal{F}}) - J_M(\hat{\pi}_{\mathcal{F}})).
 \end{aligned}$$

Acotando los dos términos de desajuste  $M$  vs  $\hat{M}$  por el supremo uniforme en  $\Pi_{\mathcal{F}}$ , y

luego tomando ínfimo en  $\tilde{\pi} \in \Pi_{\mathcal{F}}$ , se obtiene la desigualdad. □

Esta proposición reafirma que “IA” no debe leerse como sinónimo de aprendizaje automático no restringido, sino como un esquema donde la complejidad computacional se gestiona mediante aproximación, sin perder trazabilidad de fuentes de error.

El marco desarrollado tiene límites estructurales que conviene clasificar. Algunos son extensiones naturales dentro del mismo paradigma (por ejemplo no estacionariedad finita, restricciones adicionales convexas, robustez local, mejoras de inferencia en OPE). Otros exigen cambiar la estructura del modelo (interferencia fuerte entre unidades, juegos dinámicos estratégicos, equilibrio general con precios endógenos, observabilidad parcial severa con memoria larga no agregable, o instituciones que alteran el espacio de acciones de forma no exógena). Esta clasificación evita sobreextender interpretaciones de los teoremas probados.

Límite o extensión	Tipo	Implicación formal
No estacionariedad de horizonte finito	Extensión natural	Bellman temporal; no cambia la lógica base de DP
Restricciones convexas adicionales	Extensión natural	Se preserva tractabilidad con LP/convexo ampliado
Robustez local de transiciones	Extensión natural	MDP robusto con rectangularidad
Interferencia fuerte entre unidades	Cambio de marco	g-recursión estándar y MDP agregado pueden fallar
Juegos dinámicos estratégicos	Cambio de marco	Requiere equilibrio (Markov perfect, etc.)
Equilibrio general con precios endógenos	Cambio de marco	La transición ya no es exógena al policy maker
POMDP severo no agregable	Cambio de marco	Estado observable no suficiente; requiere creencias/filtrado

*Fuente:* Elaboración propia.

La síntesis estructural del documento puede expresarse como una cadena de objetos matemáticos:

$$(\mathcal{D}_n, \mathcal{C}) \xrightarrow{\hat{\Psi}_n} (\hat{P}, \hat{r}, \hat{g}, \hat{U}, \hat{\mathcal{I}}) \xrightarrow{\mathcal{O}} \hat{\pi} \xrightarrow{\text{seguridad/robustez}} \pi^{\text{dep}} \xrightarrow{\text{monitoreo}} \mathcal{D}_{n+1}.$$

La última flecha hace explícita la realimentación de datos y el papel del monitoreo de drift. Matemáticamente, esto cierra el ciclo adaptativo de una IA pública sin suponer que el problema termina en la optimización inicial.

El aporte del marco no reside en afirmar neutralidad normativa automática ni en prometer validez fuera de supuestos. Su valor está en convertir objetivos, restricciones, supuestos causales, incertidumbre y criterios de seguridad en objetos formales trazables. Esa explicitud permite demostrar resultados, identificar puntos de fragilidad, cuantificar sensibilidad y someter decisiones a auditoría técnica y deliberación pública informada.

Componente del sistema	Objeto matemático	Papel funcional en IA pública
Objetivos sociales	$J(\pi), F(\pi),$ pesos $w, \alpha$	Define qué se optimiza y con qué prioridades explícitas
Restricciones públicas	$G_k(\pi) \leq b_k$	Presupuesto, equidad, capacidad, riesgo, estabilidad
Representación dinámica	Medidas ocupacionales $d_\pi$	Convierte el problema en LP/convexo interpretable
Interpretación económica	Dualidad y KKT $(v, \lambda)$	Precios sombra y restricciones activas auditables
Capa causal	g-recursión, supuestos secuenciales, agregación $\phi$	Identifica efectos bajo intervención en datos observacionales
Evaluación sin despliegue	IS/DR, clipping, cotas OPE	Compara políticas candidatas con datos históricos
Control de error de modelo	Cotas tipo simulation lemma y suboptimalidad	Traduce error estadístico en riesgo de decisión
Robustez	$\mathcal{P}_{s,a,r}$ , Bellman robusto	Protege contra incertidumbre y desalineación modelo-entorno
Seguridad	Restricciones deterministas/probabilísticas vs $\pi_b$	Evita deterioros inaceptables y regula despliegue
Aprendizaje restringido	Primal-dual estocástico	Implementación iterativa con garantías estándar

Componente del sistema	Objeto matemático	Papel funcional en IA pública
Monitoreo y adaptación	Drift, actualización de $\mathcal{D}_n$ y $\hat{\Psi}_n$	Cierra el ciclo institucional y empírico

Fuente: Elaboración propia.

Una IA para políticas públicas puede formularse con rigor matemático como un problema de optimización dinámica restringida, causalmente identificada de manera explícita, robustificada frente a incertidumbre y sujeta a reglas de seguridad respecto a una política base. El marco desarrollado no sustituye la deliberación política ni la gobernanza institucional, pero sí proporciona una estructura formal que permite discutir con precisión qué está siendo optimizado, bajo qué supuestos, con qué garantías y con qué riesgos residuales. Esa precisión es la condición mínima para que una IA pública sea técnicamente seria y normativamente auditable.

## 10. Metodología

El presente bloque metodológico documenta un protocolo de verificación integral, reproducible y auditable para un conjunto de proposiciones matemáticas y experimentos numéricos sobre procesos de decisión de Markov (MDP), medidas ocupacionales descontadas, operadores de Bellman, mejora de políticas, evaluación fuera de política (off-policy evaluation, OPE) mediante *importance sampling*, robustez bajo incertidumbre de transición, condiciones de seguridad respecto a una política *baseline*, penalización entrópica por divergencia KL, y un problema estático de asignación con utilidad cóncava resuelto por condiciones KKT y validado simbólicamente y numéricamente. El diseño metodológico se implementa en Colab con NumPy y SymPy, incorporando verificadores TRUE/FALSE por cada afirmación evaluada, así como un verificador global que sintetiza el estado del conjunto de pruebas.

El código utilizado en este trabajo está disponible en Google Colab en el siguiente enlace: Notebook en Google Colab.

La lógica del procedimiento no se limita a computar valores finales, sino que busca contrastar identidades, desigualdades, teoremas y condiciones suficientes mediante

múltiples capas de control. En particular, se emplean: i) resolución exacta de sistemas lineales para funciones de valor; ii) enumeración exacta de trayectorias de horizonte finito para verificar la insesgadez de estimadores IS; iii) simulación Monte Carlo con criterio estadístico explícito ( $3\sigma$ ) para diagnóstico de cercanía empírica; iv) diferenciación simbólica con SymPy para comprobar derivadas de funciones de utilidad; v) verificación de restricciones y condiciones KKT por evaluación directa de igualdades y desigualdades; y vi) chequeos de contracción de operadores estándar y robustos mediante norma infinito.

El marco dinámico se basa en un MDP finito con espacio de estados  $\mathcal{S}$  de cardinalidad  $S = 3$ , espacio de acciones  $\mathcal{A}$  de cardinalidad  $A = 2$ , factor de descuento  $\gamma = 0,9$  y distribución inicial  $\mu \in \Delta(\mathcal{S})$ . Las probabilidades de transición se representan como un tensor  $P(a, s, s')$  y la recompensa instantánea como una matriz  $r(s, a)$ . Se consideran además tres políticas estocásticas: una política de comportamiento  $\beta(a | s)$ , una política objetivo  $\pi(a | s)$  y una política alternativa  $\pi_2(a | s)$ . La estructura de datos se implementa con arreglos NumPy de dimensión fija, lo cual permite preservar transparencia en la trazabilidad y minimizar ambigüedades de indexación.

La función de transición inducida por una política  $\pi$  se construye como la matriz

$$P^\pi(s, s') = \sum_{a \in \mathcal{A}} \pi(a | s) P(a, s, s').$$

A partir de esta matriz, la función de valor  $V^\pi \in \mathbb{R}^S$  se obtiene resolviendo el sistema lineal de Bellman

$$V^\pi = r^\pi + \gamma P^\pi V^\pi, \quad r^\pi(s) = \sum_{a \in \mathcal{A}} \pi(a | s) r(s, a),$$

equivalentemente,

$$V^\pi = (I - \gamma P^\pi)^{-1} r^\pi.$$

La metodología privilegia esta resolución matricial exacta (hasta error numérico de máquina) en lugar de iteración de valor, con el fin de aislar los errores de verificación de los errores de convergencia iterativa. Esta decisión es especialmente útil cuando se pretende validar identidades analíticas como la equivalencia entre formulaciones basadas en Bellman y formulaciones basadas en medidas ocupacionales.

La medida ocupacional descontada normalizada se define, en su versión de estado-

acción, como

$$d_{\pi}(s, a) = x_{\pi}(s) \pi(a | s),$$

donde  $x_{\pi}$  es la distribución ocupacional por estado inducida por  $\pi$ ,

$$x_{\pi}^{\top} = (1 - \gamma) \mu^{\top} (I - \gamma P^{\pi})^{-1}.$$

Bajo esta normalización,  $\sum_{s,a} d_{\pi}(s, a) = 1$ , lo que facilita interpretar  $d_{\pi}$  como una distribución sobre pares estado-acción bajo descuento geométrico. La metodología incluye tres verificaciones esenciales asociadas a este objeto: no negatividad, suma unitaria y ecuación de flujo ocupacional. Esta última se expresa, para cada estado  $s$ , como

$$\sum_a d_{\pi}(s, a) = (1 - \gamma) \mu(s) + \gamma \sum_{s', a'} d_{\pi}(s', a') P(a', s', s).$$

Estas tres comprobaciones son estructuralmente importantes porque descartan errores frecuentes de implementación (por ejemplo, confusión entre filas y columnas, o normalizaciones incorrectas del vector ocupacional).

La identidad de valor basada en la medida ocupacional se verifica mediante

$$J(\pi) = \mu^{\top} V^{\pi} = \frac{1}{1 - \gamma} \sum_{s,a} d_{\pi}(s, a) r(s, a).$$

La estrategia metodológica consiste en computar  $J(\pi)$  por ambas rutas de forma independiente y contrastar la diferencia absoluta contra una tolerancia numérica estricta. El hecho de que ambas cantidades coincidan con error numérico prácticamente nulo funciona como control cruzado de: i) la implementación del sistema de Bellman; ii) la construcción de  $d_{\pi}$ ; iii) el manejo del factor  $(1 - \gamma)$  en la normalización.

El operador de Bellman para una política fija  $\pi$  se define como

$$(T^{\pi}V)(s) = \sum_a \pi(a | s) \left[ r(s, a) + \gamma \sum_{s'} P(a, s, s') V(s') \right].$$

La metodología verifica numéricamente su propiedad de contracción en norma infinito:

$$\|T^{\pi}V_1 - T^{\pi}V_2\|_{\infty} \leq \gamma \|V_1 - V_2\|_{\infty},$$

utilizando dos vectores arbitrarios  $V_1, V_2$ . Aunque un solo ejemplo no constituye

prueba general, sí constituye una validación de consistencia del operador implementado y de la manipulación de índices. Adicionalmente, se verifica que  $V^\pi$  satisface la ecuación de punto fijo  $\|T^\pi V^\pi - V^\pi\|_\infty \approx 0$ , lo cual refuerza la coherencia interna del bloque dinámico.

Para el teorema de mejora de política (*policy improvement theorem*), la metodología procede en cuatro pasos: i) computar  $V^\pi$ ; ii) obtener la función de acción-valor

$$Q^\pi(s, a) = r(s, a) + \gamma \sum_{s'} P(a, s, s') V^\pi(s');$$

iii) construir una política voraz  $\pi'$  que maximiza  $Q^\pi(s, a)$  estado por estado; y iv) contrastar tanto la condición punto a punto

$$\mathbb{E}_{a \sim \pi'(\cdot|s)}[Q^\pi(s, a)] \geq V^\pi(s),$$

como la conclusión

$$V^{\pi'}(s) \geq V^\pi(s) \quad \forall s \in \mathcal{S}.$$

La elección de una política determinista voraz hace más nítida la interpretación del resultado, ya que la mejora puede observarse directamente en los márgenes por estado. El protocolo registra el margen mínimo de mejora en cada verificación, lo que permite cuantificar cuán lejos se está del umbral de igualdad.

La metodología también incluye la verificación numérica exacta del *performance difference lemma* (PDL), formulado como

$$J(\pi_2) - J(\pi) = \frac{1}{1 - \gamma} \sum_{s,a} d_{\pi_2}(s, a) A^\pi(s, a),$$

donde

$$A^\pi(s, a) = Q^\pi(s, a) - V^\pi(s)$$

es la ventaja de  $\pi$ . Este bloque es especialmente relevante porque conecta evaluación de políticas, medidas ocupacionales y ventajas bajo una identidad exacta. La implementación calcula ambos lados de la igualdad de forma independiente y reporta la diferencia numérica. En términos metodológicos, este paso opera como una bisagra entre los resultados de Bellman y los análisis posteriores de seguridad y mejora respecto a *baseline*.

Con el objetivo de emular restricciones de política pública, se incorporan matrices de costo, uso de capacidad y cobertura mínima, denotadas genéricamente por  $g(s, a)$ ,  $h(s, a)$  y  $c(s, a)$ . A partir de  $d_\pi$ , se evalúan funcionales lineales descontados de la forma

$$G_g(\pi) = \frac{1}{1 - \gamma} \sum_{s,a} d_\pi(s, a) g(s, a), \quad G_h(\pi) = \frac{1}{1 - \gamma} \sum_{s,a} d_\pi(s, a) h(s, a),$$

$$G_c(\pi) = \frac{1}{1 - \gamma} \sum_{s,a} d_\pi(s, a) c(s, a).$$

Las restricciones se expresan como

$$G_g(\pi) \leq B, \quad G_h(\pi) \leq \bar{H}, \quad G_c(\pi) \geq C_{\text{mín}}.$$

La metodología verifica cada una de estas desigualdades y reporta el margen de holgura, lo cual permite interpretar no sólo la factibilidad, sino también la robustez del cumplimiento bajo pequeñas perturbaciones.

En la parte de OPE, se establece un horizonte finito  $T = 4$  y se verifica la identidad de insesgadez de *importance sampling* por enumeración exacta de trayectorias. Sea  $G_T$  el retorno descontado de horizonte finito y  $W_T$  el peso de razones de probabilidad entre política objetivo y política de comportamiento. La identidad contrastada es

$$\mathbb{E}_\beta[W_T G_T] = \mathbb{E}_\pi[G_T].$$

La metodología implementa una recursión exhaustiva sobre trayectorias posibles  $(s_t, a_t, s_{t+1})$  para obtener ambos términos con precisión numérica alta. Esta decisión evita atribuir la validación de insesgadez a fluctuaciones Monte Carlo, separando una demostración numérica exacta de un diagnóstico estadístico posterior.

Tras la enumeración exacta, se realiza un diagnóstico Monte Carlo con  $N_{\text{mc}} = 120,000$  trayectorias simuladas bajo  $\beta$ . Se computa la variable aleatoria  $Y = W_T G_T$ , el estimador  $\hat{J}_{\text{IS}} = \frac{1}{N} \sum_i Y_i$ , su desviación estándar muestral  $\hat{\sigma}_Y$ , el error estándar  $\widehat{\text{SE}}(\hat{J}_{\text{IS}}) = \hat{\sigma}_Y / \sqrt{N}$ , y se evalúa el criterio

$$\left| \hat{J}_{\text{IS}} - \mathbb{E}_\pi[G_T] \right| \leq 3 \widehat{\text{SE}}(\hat{J}_{\text{IS}}).$$

El verificador asociado no pretende demostrar insesgadez (ya demostrada por enumeración exacta), sino diagnosticar si la realización Monte Carlo observada es estadísticamente compatible con el valor exacto bajo una banda razonable. Esta distinción metodológica es central para evitar malinterpretar fluctuaciones muestrales como fallas teóricas.

En el bloque de robustez de modelo, se construyen una recompensa perturbada  $\hat{r}$  y un tensor de transición perturbado  $\hat{P}$ , imponiendo renormalización de probabilidades para preservar validez estocástica. Se comparan las funciones de valor  $V^\pi$  y  $\hat{V}^\pi$  bajo la misma política  $\pi$ , y se verifica una cota tipo *simulation lemma*:

$$\|\hat{V}^\pi - V^\pi\|_\infty \leq \frac{\epsilon_r}{1 - \gamma} + \frac{2\gamma R_{\text{máx}}}{(1 - \gamma)^2} \epsilon_P,$$

donde

$$\epsilon_r = \|\hat{r} - r\|_\infty, \quad \epsilon_P = \max_{s,a} \text{TV}(\hat{P}(\cdot | s, a), P(\cdot | s, a)), \quad R_{\text{máx}} = \max_{s,a} |r(s, a)|.$$

La distancia TV se calcula como variación total

$$\text{TV}(p, q) = \frac{1}{2} \|p - q\|_1.$$

El uso de una cota explícita permite vincular desviaciones paramétricas locales con errores globales en valor, proporcionando una lectura metodológica de sensibilidad de planificación bajo incertidumbre de modelo.

La verificación de contracción robusta emplea un operador de Bellman robusto de tipo *worst-case* sobre conjuntos de ambigüedad definidos por radios de variación total. Para cada par  $(s, a)$ , se considera un radio  $\delta_{s,a}$  y se aproxima el problema interno de peor esperanza mediante una grilla sobre el simplex de dimensión tres:

$$\inf_{q \in \Delta(\mathcal{S}) : \text{TV}(q, P(\cdot | s, a)) \leq \delta_{s,a}} q^\top V.$$

Esta infimización se aproxima por búsqueda discreta sobre una grilla de resolución  $m = 50$ , lo que genera un conjunto finito de candidatos. El operador robusto se define entonces como

$$(T_R^\pi V)(s) = \sum_a \pi(a | s) \left[ r(s, a) + \gamma \inf_{q : \text{TV}(q, P(\cdot | s, a)) \leq \delta_{s,a}} q^\top V \right].$$

La metodología compara  $\|T_R^\pi V_1 - T_R^\pi V_2\|_\infty$  contra  $\gamma\|V_1 - V_2\|_\infty$  para dos vectores de prueba. El carácter aproximado de la verificación se explicita, ya que el operador interno se discretiza por grilla, pero la prueba numérica sigue siendo informativa respecto a la consistencia del módulo robusto.

En el bloque de seguridad respecto a *baseline*, se fija  $\pi_b = \beta$  como política de referencia. Se calculan  $J(\pi_b)$  y  $J(\pi_{\text{new}})$ , donde  $\pi_{\text{new}}$  corresponde a la política voraz antes construida, y se contrasta la condición directa de seguridad con tolerancia  $\varepsilon_{\text{safe}}$ :

$$J(\pi_{\text{new}}) \geq J(\pi_b) - \varepsilon_{\text{safe}}.$$

Además, se evalúa una condición suficiente expresada mediante ventajas respecto a la *baseline*. Si

$$A^{\pi_b}(s, a) = Q^{\pi_b}(s, a) - V^{\pi_b}(s),$$

y  $d_{\pi_{\text{new}}}$  denota la medida ocupacional de la política candidata, se verifica el término

$$\sum_{s,a} d_{\pi_{\text{new}}}(s, a) A^{\pi_b}(s, a),$$

comparándolo con el umbral  $-(1 - \gamma)\varepsilon_{\text{safe}}$ . Este doble enfoque (directo y suficiente) aporta una capa metodológica importante: la primera verificación usa el valor exacto, mientras la segunda usa una expresión estructuralmente interpretable para análisis de diseño de políticas seguras.

La penalización entrópica se estudia mediante la funcional

$$f(\pi) = -\tau \sum_s v(s) \text{KL}(\pi(\cdot | s) \| \pi_b(\cdot | s)),$$

con  $\tau > 0$  y pesos  $v \in \Delta(\mathcal{S})$ . Se verifica numéricamente la concavidad vía una desigualdad de Jensen para dos políticas  $\pi_A, \pi_B$  y una mezcla  $\pi_M = \lambda\pi_A + (1 - \lambda)\pi_B$ :

$$f(\pi_M) \geq \lambda f(\pi_A) + (1 - \lambda) f(\pi_B).$$

La divergencia KL se computa fila por fila con recorte numérico (*clipping*) para evitar problemas de  $\log 0$ , seguido de renormalización. Esta práctica metodológica preserva estabilidad numérica sin alterar sustantivamente la lógica de la verificación en un contexto donde todas las probabilidades permanecen estrictamente positivas por construcción.

La parte estática introduce una función de utilidad cóncava

$$u(x) = a \log(1 + bx),$$

y utiliza SymPy para verificar simbólicamente las derivadas

$$u'(x) = \frac{ab}{1 + bx}, \quad u''(x) = -\frac{ab^2}{(1 + bx)^2}.$$

Estas igualdades constituyen una mini-validación formal del módulo simbólico y de la convexidad/concavidad requerida para el problema de asignación posterior. El uso combinado de cálculo simbólico y resolución numérica ofrece una metodología híbrida útil en contextos de diseño de algoritmos con fundamento analítico.

El problema de asignación con restricción presupuestaria se formula como

$$\max_{x_i \geq 0} \sum_{i=1}^n a_i \log(1 + b_i x_i) \quad \text{sujeto a} \quad \sum_{i=1}^n x_i = B.$$

La metodología implementa una solución cerrada tipo *water-filling* derivada de las KKT:

$$x_i^*(\lambda) = \max \left\{ 0, \frac{a_i}{\lambda} - \frac{1}{b_i} \right\},$$

con  $\lambda^*$  obtenido por bisección para satisfacer la restricción de presupuesto. Posteriormente se verifican: i) igualdad de marginales para componentes activos,

$$\frac{a_i b_i}{1 + b_i x_i^*} = \lambda^* \quad \text{si } x_i^* > 0;$$

ii) condición de no exceso para componentes inactivos,

$$\frac{a_i b_i}{1 + b_i x_i^*} \leq \lambda^* \quad \text{si } x_i^* = 0;$$

iii) cumplimiento vinculante del presupuesto; y iv) evidencia empírica de optimalidad comparando  $U(x^*)$  con el mejor valor hallado en una búsqueda aleatoria sobre el simplex presupuestario. Este último paso no es una prueba de optimalidad global (ya garantizada por concavidad y KKT), pero sí una corroboración computacional adicional.

Se incorpora una cota de sensibilidad normativa bajo pesos alternativos. Dado un vector  $F(\pi) \in \mathbb{R}^n$  y dos vectores de ponderación  $w, \tilde{w} \in \Delta_n$ , se evalúa la desigualdad

$$\left| w^\top F(\pi) - \tilde{w}^\top F(\pi) \right| \leq \|w - \tilde{w}\|_1 \|F(\pi)\|_\infty.$$

Metodológicamente, esta cota permite interpretar cómo cambios en ponderaciones normativas o prioridades distributivas pueden alterar una evaluación agregada sin necesidad de reestimar toda la estructura dinámica. Su inclusión amplía el alcance del protocolo desde consistencia matemática estricta hacia análisis de sensibilidad institucional.

La Tabla 9 resume la parametrización base del módulo dinámico. La Tabla 10 resume los bloques de verificación implementados y su propósito. Ambas tablas cumplen un rol metodológico de documentación reproducible, permitiendo reconstruir la lógica del experimento sin inspeccionar directamente el código.

Tabla 9: Parámetros y objetos del MDP base utilizados en las verificaciones

Elemento	Descripción metodológica
$S = 3, A = 2$	MDP finito de pequeña dimensión para posibilitar enumeración exacta en OPE y trazabilidad completa de resultados.
$\gamma = 0,9$	Factor de descuento alto, suficientemente representativo de problemas de planificación intertemporal con persistencia.
$\mu$	Distribución inicial sobre estados usada tanto en Bellman como en medida ocupacional y en OPE.
$P(a, s, s')$	Tensor de transición con dos acciones y tres estados; todas las filas son estocásticas y estrictamente positivas.
$r(s, a)$	Matriz de recompensas instantáneas empleada en Bellman, $Q^\pi$ , PDL y OPE.
$\beta$	Política de comportamiento para OPE y política <i>baseline</i> en el bloque de seguridad.
$\pi$	Política objetivo principal para verificación de medidas ocupacionales, Bellman, PDL y robustez.
$\pi_2$	Política comparativa para la verificación numérica exacta del <i>performance difference lemma</i> .

Fuente: Elaboración propia.

Desde el punto de vista de reproducibilidad, el protocolo se beneficia de tres elecciones metodológicas adicionales. En primer lugar, se fija una semilla explícita para los

Tabla 10: Bloques de verificación incluidos en el protocolo integral

Bloque	Objeto matemático principal	Tipo de verificación
Medidas ocupacionales	$d_\pi$ , flujo, identidad de valor	Igualdades exactas numéricas con tolerancias
Bellman	Contracción y punto fijo	Desigualdad en norma $\infty$ y residuo de punto fijo
Mejora de política	$Q^\pi$ , política voraz	Condición punto a punto y comparación de $V$
PDL	$J(\pi_2) - J(\pi)$ y ventajas	Identidad exacta numérica
Restricciones lineales	Presupuesto, capacidad, cobertura	Desigualdades con márgenes de holgura
OPE exacto	$\mathbb{E}_\beta[W_T G_T]$ vs $\mathbb{E}_\pi[G_T]$	Enumeración exacta de trayectorias
OPE Monte Carlo	$\hat{J}_{IS}$	Diagnóstico estadístico con regla $3\sigma$
Robustez de modelo	Cota tipo <i>simulation lemma</i>	Verificación de cota superior
Bellman robusto	$T_R^\pi$	Contracción aproximada en grilla simplex
Seguridad	Comparación vs <i>baseline</i> y ventajas	Condición directa y condición suficiente
Penalización KL	Concavidad de $-\tau \mathbb{E}[KL]$	Jensen numérico
KKT + utilidad	Derivadas, KKT, presupuesto	SymPy + validación numérica + búsqueda aleatoria
Sensibilidad normativa	Cota $\ell_1$ - $\ell_\infty$	Desigualdad exacta numérica

Fuente: Elaboración propia.

generadores pseudoaleatorios en los bloques Monte Carlo y de búsqueda aleatoria, lo que permite replicar exactamente la secuencia de resultados de una corrida dada. En segundo lugar, los objetos intermedios relevantes (por ejemplo,  $d_\pi$ ,  $V^\pi$ ,  $Q^\pi$ ,  $A^\pi$ , márgenes de restricciones, estadísticas de OPE y multiplicadores KKT) se imprimen en consola con formato legible, facilitando auditoría paso a paso y diagnóstico en caso de futuras modificaciones. En tercer lugar, el sistema de verificadores TRUE/FALSE se diseña como una capa declarativa: cada chequeo registra nombre, valor booleano y detalle cuantitativo, y el verificador global resume el estado del conjunto. Esta arquitectura hace que el protocolo sea útil no sólo para análisis académico, sino también para *testing* de regresión en desarrollo de software científico.

Cabe enfatizar que varias verificaciones aquí presentadas son *demostraciones numéricas* o *corroboraciones computacionales* de proposiciones teóricas, no demostraciones formales universales. La metodología, por tanto, debe leerse como una estrategia de validación computacional rigurosa en un entorno controlado. Su fuerza principal reside en la combinación de exactitud en casos pequeños (por enumeración o solución cerrada), robustez estadística en bloques de simulación, y consistencia transversal entre formulaciones equivalentes. Este diseño es particularmente pertinente en aplicaciones de IA para política pública, donde la transparencia, la trazabilidad y la verificabilidad son tan importantes como el valor predictivo u operativo.

## 11. Resultados

Los resultados obtenidos muestran una consistencia interna notable entre los distintos módulos del protocolo, culminando en un verificador global TRUE con 26/26 comprobaciones superadas. Este desenlace es relevante por dos razones. Primero, porque el conjunto de verificaciones abarca objetos heterogéneos (dinámicos, robustos, estadísticos, simbólicos y de optimización estática) y, por tanto, reduce la probabilidad de que el éxito se deba a una coincidencia local o a un error compensatorio. Segundo, porque varias de las comprobaciones contrastan identidades por rutas computacionales independientes, lo que fortalece la inferencia de corrección de implementación.

El bloque de medidas ocupacionales descontadas produce una distribución  $d_\pi(s, a)$  estrictamente no negativa y correctamente normalizada, con suma total igual a uno dentro de precisión numérica de máquina. Esta propiedad no es trivial en términos

computacionales, dado que la construcción de  $d_\pi$  requiere una inversión matricial y una composición con la política. Además, la ecuación de flujo ocupacional se verifica con error máximo de orden  $10^{-16}$ , lo que sugiere ausencia de errores de indexación o transposición en la manipulación del tensor de transición. La identidad de valor basada en ocupación también coincide exactamente con el valor obtenido vía Bellman, con diferencia prácticamente nula. En conjunto, estos resultados indican que la representación ocupacional y la representación de valor están plenamente alineadas en la implementación.

La Tabla 11 resume los principales resultados numéricos del bloque ocupacional y del operador de Bellman. Se observa que la función de valor  $V^\pi$  presenta niveles relativamente altos en los tres estados, consistentes con un factor de descuento elevado ( $\gamma = 0,9$ ) y recompensas positivas predominantes. La mayor valoración del estado 2 frente al estado 1 en  $V^\pi$  es coherente con la combinación de recompensas por acción y estructura de transición bajo la política  $\pi$ , aunque el orden exacto entre estados refleja la interacción no lineal entre ambas.

Tabla 11: Resultados del bloque ocupacional y Bellman para la política  $\pi$

Magnitud	Valor reportado	Interpretación
$\sum_{s,a} d_\pi(s, a)$	1,000000000000	Normalización correcta de la medida ocupacional descontada.
Error de flujo ocupacional ( $\max  LHS - RHS $ )	$1,110 \times 10^{-16}$	Consistencia numérica esencialmente exacta.
$V^\pi$	[8,723433, 9,450659, 9,035601]	Valor esperado descontado por estado bajo $\pi$ .
$J(\pi)$ vía Bellman	9,004034297510	Valor agregado bajo distribución inicial $\mu$ .
$J(\pi)$ vía ocupación	9,004034297510	Coincide con Bellman; válida identidad de valor.

Fuente: Elaboración propia.

En el bloque de Bellman, la verificación de contracción se satisface con holgura. El valor observado de  $\|T^\pi V_1 - T^\pi V_2\|_\infty$  es significativamente menor que  $\gamma \|V_1 - V_2\|_\infty$ , lo que no sólo confirma la desigualdad en el ejemplo ensayado, sino que además sugiere una contracción efectiva relativamente fuerte para los vectores seleccionados. El residuo de punto fijo  $\|T^\pi(V^\pi) - V^\pi\|_\infty$  es de orden  $10^{-15}$ , consistente con errores

de redondeo de doble precisión. Este resultado respalda la corrección del solucionador matricial y de la implementación del operador  $T^\pi$ .

La construcción de la política voraz a partir de  $Q^\pi$  produce una mejora clara. La condición  $\mathbb{E}_{\pi'}[Q^\pi(\cdot, \cdot) | s] \geq V^\pi(s)$  se verifica en los tres estados con margen mínimo positivo, y la comparación de funciones de valor muestra incrementos sustanciales, con un margen mínimo de mejora superior a 1,95 unidades. La política voraz resultante es determinista y selecciona la acción 0 en el estado 1, la acción 1 en el estado 2 y la acción 1 en el estado 3 (siguiendo la indexación base 0, 1, 2 usada en código, esto corresponde al patrón [0, 1, 1]). La magnitud de la mejora sugiere que la política  $\pi$  inicial estaba lejos de ser óptima en el MDP considerado.

La Tabla 12 sintetiza los resultados de mejora de política y del *performance difference lemma*. Es especialmente importante la coincidencia prácticamente exacta entre ambos lados del PDL, ya que esta identidad enlaza el análisis local (ventajas  $A^\pi$ ) con el cambio global de desempeño  $J(\pi_2) - J(\pi)$ . En la práctica, esta verificación provee una garantía fuerte de consistencia en el cálculo de  $Q^\pi, V^\pi, A^\pi, d_{\pi_2}$  y  $J(\cdot)$ .

Tabla 12: Resultados de mejora de política y *performance difference lemma*

Magnitud	Valor reportado	Lectura
Margen mínimo de condición de mejora	$1,431 \times 10^{-1}$	La política voraz mejora $Q^\pi$ punto a punto.
Margen mínimo en $V^{\pi'} - V^\pi$	1,951	Mejora sustancial en valor por estado.
$V^{\pi'}$	[10,986301, 11,534247, 10,986301]	Política voraz fuertemente superior a $\pi$ .
$J(\pi_2) - J(\pi)$	0,696260318843	Diferencia global de desempeño.
$\text{PDL RHS} = \frac{1}{1-\gamma} \sum d_{\pi_2} A^\pi$	0,696260318843	Coincidencia exacta con LHS (error $\sim 10^{-15}$ ).

Fuente: Elaboración propia.

En el bloque de restricciones lineales, la política  $\pi$  resulta factible respecto a las tres restricciones propuestas. El costo agregado descontado  $G_{\text{cost}}$  queda por debajo del presupuesto  $B$  con una holgura de aproximadamente 0,077, la restricción de capacidad presenta una holgura más amplia ( $\approx 0,570$ ), y la cobertura mínima supera el umbral en aproximadamente 1,761. Estas magnitudes son metodológicamente útiles porque permiten inferir que, en este ejemplo, la restricción más ajustada es el presupuesto.

En aplicaciones reales, esta lectura puede orientar qué restricción es más probable que se active tras una mejora de política o una perturbación de modelo.

La factibilidad simultánea de las restricciones también demuestra la utilidad de las medidas ocupacionales como objeto unificador: una vez construido  $d_\pi$ , múltiples métricas de política se evalúan como funcionales lineales sobre el mismo soporte estado-acción. Esto reduce complejidad computacional y facilita el análisis de sensibilidad, pues la introducción de nuevas métricas  $m(s, a)$  no exige re-simular trayectorias ni re-resolver Bellman, sino sólo computar  $\frac{1}{1-\gamma} \sum d_\pi m$ .

Los resultados de OPE son particularmente ilustrativos. La identidad exacta de insesgadez por enumeración se satisface con diferencia nula (dentro de precisión numérica), mostrando que

$$\mathbb{E}_\beta[W_T G_T] = \mathbb{E}_\pi[G_T]$$

para el horizonte  $T = 4$  considerado. Este resultado tiene dos implicaciones. Primero, confirma la implementación correcta de la recursión de enumeración y de los pesos IS. Segundo, valida la consistencia de la definición de retorno descontado y su alineación temporal con los pesos.

En el diagnóstico Monte Carlo, el estimador IS  $\hat{J}_{IS}$  difiere del valor exacto en aproximadamente  $1,96 \times 10^{-2}$ , mientras que la banda de  $3\widehat{SE}$  es aproximadamente  $9,82 \times 10^{-2}$ . Dado que la diferencia observada es bastante menor que la banda, el diagnóstico se clasifica como satisfactorio. La desviación estándar de  $Y = W_T G_T$  es relativamente alta en comparación con la escala del retorno, lo que evidencia la variabilidad inherente del estimador IS incluso en un problema pequeño. Asimismo, el máximo observado de los pesos  $W$  (del orden de  $10^2$ ) ilustra, en miniatura, el fenómeno bien conocido de colas pesadas o alta varianza potencial en OPE por razones de probabilidad. El hecho de que aun así el diagnóstico pase con  $N = 120,000$  muestra que el tamaño muestral elegido fue suficiente para estabilizar la estimación en este caso.

La Tabla 13 resume los resultados exactos y Monte Carlo del bloque OPE. La lectura conjunta de esta tabla sugiere una lección metodológica importante: cuando es posible disponer de un valor exacto (por enumeración en problemas pequeños o por casos sintéticos), conviene usarlo como referencia para calibrar criterios de diagnóstico Monte Carlo antes de escalar a problemas donde dicha referencia no existe.

En el bloque de robustez por mala especificación del modelo, la cota tipo *simulation*

Tabla 13: Resultados de OPE: identidad exacta y diagnóstico Monte Carlo

Magnitud	Valor reportado	Interpretación
Horizonte $T$	4	OPE en horizonte finito con descuento.
$\mathbb{E}_\beta[W_T G_T]$ exacto	3,618135325759	Lado IS exacto por enumeración.
$\mathbb{E}_\pi[G_T]$ exacto	3,618135325759	Lado objetivo exacto por enumeración.
$\hat{J}_{IS}$ Monte Carlo	3,598502169240	Estimador IS sobre 120,000 trayectorias.
Diferencia absoluta vs exacto	0,01963316	Error muestral observado.
$3\widehat{SE}(\hat{J}_{IS})$	0,098187	Banda de diagnóstico estadístico.
Resultado del diagnóstico	TRUE	La diferencia cae dentro de la banda $3\sigma$ .

Fuente: Elaboración propia.

*lemma* se verifica con margen amplio. La diferencia real  $\|\hat{V}^\pi - V^\pi\|_\infty$  es aproximadamente 0,116, mientras que la cota superior calculada alcanza 5,44. Esta holgura grande era esperable: las cotas uniformes en norma infinito suelen ser conservadoras, especialmente cuando agregan errores de recompensa y transición de manera worst-case. El resultado es, sin embargo, metodológicamente valioso porque confirma que la implementación de la cota y de las distancias de variación total es consistente. Además, evidencia una situación clásica en análisis robusto: una cota puede ser válida y útil para garantías, aun cuando no sea ajustada.

La perturbación de transiciones se diseñó de modo que todas las distancias de variación total por par  $(s, a)$  fueran exactamente 0,02. Esto simplifica la interpretación del parámetro  $\varepsilon_P$  y permite atribuir el tamaño de la cota a la estructura analítica del término  $\frac{2\gamma R_{\max}}{(1-\gamma)^2} \varepsilon_P$ , que crece rápidamente al acercarse  $\gamma$  a 1. En términos prácticos, el resultado ilustra que en problemas con horizontes efectivos largos (alto  $\gamma$ ), pequeñas perturbaciones de transición pueden inducir garantías teóricas muy conservadoras.

La verificación de contracción del operador robusto  $T_R^\pi$ , aproximado por grilla sobre el simplex, también se satisface con holgura. El valor observado de  $\|T_R^\pi(V_1) - T_R^\pi(V_2)\|_\infty$  es aproximadamente 0,8442, frente a un límite  $\gamma\|V_1 - V_2\|_\infty = 1,8$ . El margen negativo amplio ( $lhs - rhs < 0$ ) sugiere que la contracción efectiva del operador robusto

en el ejemplo analizado no es cercana al límite teórico. No obstante, al tratarse de una verificación aproximada por grilla, la interpretación debe mantenerse prudente: el resultado corrobora la consistencia numérica del módulo robusto, pero no sustituye una demostración analítica general del operador bajo la clase de conjuntos de ambigüedad utilizada.

En seguridad respecto a *baseline*, los resultados son muy favorables para la política nueva (voraz). El valor agregado  $J(\pi_{\text{new}})$  supera ampliamente a  $J(\pi_b) - \epsilon_{\text{safe}}$ , con un margen cercano a 2,97, muy superior al umbral de tolerancia  $\epsilon_{\text{safe}} = 0,05$ . La condición suficiente basada en ventaja también se cumple con holgura, ya que el término  $\sum d_{\pi_{\text{new}}} A^{\pi_b}$  es positivo y claramente mayor que el umbral negativo permitido. Esta coincidencia entre condición directa y condición suficiente tiene implicancias metodológicas útiles: cuando la evaluación exacta de  $J(\pi_{\text{new}})$  sea costosa o imposible, las condiciones de ventaja pueden funcionar como *screening* para filtrar políticas seguras candidatas.

La Tabla 14 agrupa los resultados de robustez y seguridad. Su lectura conjunta sugiere que el protocolo no sólo es capaz de verificar consistencia matemática, sino también de capturar dimensiones de interés aplicado: sensibilidad al modelo y seguridad de despliegue respecto a una política de referencia.

Tabla 14: Resultados de robustez de modelo, Bellman robusto y seguridad respecto a *baseline*

Magnitud	Valor reportado	Lectura
$\ \hat{V}^\pi - V^\pi\ _\infty$	0,116184	Error real bajo perturbación de $r$ y $P$ .
Cota <i>simulation lemma</i>	5,440000	Garantía válida pero conservadora.
$\ T_R^\pi(V_1) - T_R^\pi(V_2)\ _\infty$	0,844200	LHS de contracción robusta aproximada.
$\gamma\ V_1 - V_2\ _\infty$	1,800000	RHS de contracción robusta.
$J(\pi_b)$	8,231571768763	Valor de la política <i>baseline</i> .
$J(\pi_{\text{new}})$	11,150684931507	Valor de la política candidata voraz.
Margen de seguridad directo	2,969113	Cumplimiento muy holgado de la condición directa.
Margen de condición suficiente por ventaja	0,296911	Cumplimiento holgado de la condición estructural.

Fuente: Elaboración propia.

La verificación de concavidad de la penalización  $-\tau \mathbb{E}_v[\text{KL}(\pi \parallel \pi_b)]$  mediante Jensen produce una brecha positiva (*gap*) de aproximadamente  $2,997 \times 10^{-2}$ , lo que confirma empíricamente la concavidad en el ejemplo analizado. Este resultado es conceptualmente importante porque justifica el uso de términos de regularización entrópica con buena estructura de optimización en diseños de políticas. En particular, cuando tales penalizaciones se incorporan a objetivos compuestos, la concavidad de esta parte del funcional puede contribuir a estabilidad y tractabilidad de algoritmos de actualización.

En el bloque simbólico y de KKT, SymPy verifica exactamente las derivadas primera y segunda de  $u(x) = a \log(1 + bx)$ . Esta parte, aunque breve, funciona como una validación del canal simbólico y como respaldo formal de la concavidad del problema estático posterior. Luego, la solución numérica del problema de asignación presupuestaria produce un multiplicador  $\lambda^* \approx 0,65483$  y una asignación  $x^*$  con todos los componentes activos. El hecho de que todos los componentes resulten activos implica que el régimen KKT observado es interior, por lo que la condición de igualdad de marginales se verifica en los cuatro componentes. La coincidencia de marginales con  $\lambda^*$  hasta error de máquina es una evidencia fuerte de correcta implementación de la solución tipo *water-filling* y del algoritmo de bisección.

La restricción presupuestaria se satisface como igualdad con error de orden  $10^{-16}$ , lo que confirma el éxito del método de bisección en encontrar  $\lambda^*$ . Adicionalmente, la búsqueda aleatoria sobre 30,000 puntos del simplex presupuestario no encuentra una solución mejor que  $x^*$ , con una diferencia de utilidad a favor de  $x^*$  de aproximadamente  $1,98 \times 10^{-4}$ . Aunque esta evidencia aleatoria no es necesaria para concluir optimalidad (ya garantizada por concavidad y KKT), sí aporta una comprobación empírica intuitiva y útil para auditoría externa.

La cota de sensibilidad normativa se verifica con margen claro. La diferencia absoluta entre agregaciones  $w^\top F(\pi)$  y  $\tilde{w}^\top F(\pi)$  es 0,4, mientras que la cota  $\|w - \tilde{w}\|_1 \|F(\pi)\|_\infty$  es 0,84. Este resultado ilustra que la desigualdad  $\ell_1$ - $\ell_\infty$  opera como garantía general y, al igual que otras cotas uniformes, puede no ser ajustada. La interpretación aplicada es directa: cambios en ponderaciones normativas pueden alterar la evaluación agregada, pero su efecto máximo está acotado por una combinación simple de distancia entre pesos y amplitud del vector de resultados.

La Tabla 15 resume los resultados del bloque estático y de sensibilidad normativa.

En conjunto con los bloques dinámicos previos, estos resultados muestran que el protocolo logra integrar validación de teoría dinámica de decisiones, estadística de evaluación fuera de política, análisis robusto, regularización entrópica y optimización estática bajo una misma arquitectura de verificación.

Tabla 15: Resultados de penalización KL, derivación simbólica, KKT y sensibilidad normativa

Magnitud	Valor reportado	Interpretación
Gap de Jensen para $-\tau\mathbb{E}[\text{KL}]$	$2,997 \times 10^{-2}$	Concavidad confirmada numéricamente en el ejemplo.
Verificación simbólica de $u'(x)$ y $u''(x)$	TRUE, TRUE	Derivadas exactas obtenidas con SymPy.
$\lambda^*$ (KKT)	0,654831508451	Multiplicador óptimo del presupuesto.
Cumplimiento de igualdad de marginales activos	Error $\sim 10^{-16}$	KKT satisfecha con precisión de máquina.
Presupuesto en $x^*$	$\sum_i x_i^* = 4,2$	Restricción vinculante correctamente satisfecha.
Evidencia aleatoria vs $x^*$	$U(x^*)$ $U(z_{\text{máx}})$ $1,98 \times 10^{-4}$	– $\approx$ No se encontró mejora aleatoria sobre $x^*$ .
Cota de sensibilidad normativa	$0,4 \leq 0,84$	Desigualdad $\ell_1$ - $\ell_\infty$ verificada.

Fuente: Elaboración propia.

Una manera sintética de representar el núcleo del resultado global es mediante el vector de verificaciones

$$\mathbf{c} = (c_1, \dots, c_{26}), \quad c_i \in \{0, 1\},$$

donde  $c_i = 1$  si la verificación  $i$ -ésima se satisface y 0 en caso contrario. En la corrida reportada se obtuvo

$$\sum_{i=1}^{26} c_i = 26,$$

por lo que el verificador global toma el valor

$$\mathcal{V}_{\text{global}} = \mathbf{1} \left\{ \sum_{i=1}^{26} c_i = 26 \right\} = 1.$$

Esta formulación, aunque simple, es útil para estandarizar comparaciones entre versiones del código o entre escenarios de parametrización distintos. En un esquema de desarrollo iterativo, una degradación de  $\mathcal{V}_{\text{global}}$  de 1 a 0 puede interpretarse como una alerta automática de regresión.

La Tabla 16 presenta un resumen consolidado de los 26 chequeos, con una descripción breve y su estado. Esta tabla puede funcionar como apéndice ejecutivo dentro de un documento técnico, permitiendo a un lector identificar de inmediato qué propiedades se verificaron sin recorrer toda la salida de consola.

Tabla 16: Resumen consolidado de las 26 verificaciones y su estado

#	Verificación	Estado
1	No negatividad de $d_{\pi}$	TRUE
2	Suma unitaria de $d_{\pi}$	TRUE
3	Ecuación de flujo ocupacional	TRUE
4	Identidad de valor con ocupación	TRUE
5	Contracción de $T^{\pi}$ (ejemplo numérico)	TRUE
6	Punto fijo de Bellman $V^{\pi}$	TRUE
7	Condición de mejora $\mathbb{E}_{\pi'} Q^{\pi} \geq V^{\pi}$	TRUE
8	Conclusión $V^{\pi'} \geq V^{\pi}$	TRUE
9	<i>Performance difference lemma</i> exacto	TRUE
10	Restricción de presupuesto (ejemplo)	TRUE
11	Restricción de capacidad (ejemplo)	TRUE
12	Restricción de cobertura mínima (ejemplo)	TRUE
13	Inesgadez IS por enumeración exacta	TRUE
14	Diagnóstico Monte Carlo IS cercano al exacto	TRUE
15	Cota tipo <i>simulation lemma</i> (política fija)	TRUE
16	Contracción robusta (aprox. grilla simplex)	TRUE
17	Seguridad directa vs <i>baseline</i>	TRUE
18	Condición suficiente de seguridad por ventaja	TRUE
19	Concavidad de $-\tau \mathbb{E}[\text{KL}]$ (Jensen numérico)	TRUE
20	Verificación simbólica de $u'(x)$	TRUE
21	Verificación simbólica de $u''(x)$	TRUE
22	KKT: igualdad en componentes activos	TRUE
23	KKT: no exceso en componentes inactivos	TRUE
24	Presupuesto vinculante en $x^*$	TRUE
25	No mejora aleatoria sobre $x^*$ (evidencia)	TRUE
26	Cota $\ell_1$ - $\ell_{\infty}$ de sensibilidad normativa	TRUE
<b>Verificador global <math>\mathcal{V}_{\text{global}}</math></b>		<b>TRUE (26/26)</b>

Fuente: Elaboración propia.

Desde una perspectiva interpretativa, los resultados sugieren que el entorno de prueba construido es adecuado como banco de validación para desarrollos posteriores más ambiciosos. La coexistencia de identidades exactas (por ejemplo, medida ocupacional, PDL, OPE por enumeración), desigualdades teóricas (contracción, cotas de robustez, sensibilidad normativa), criterios estadísticos (diagnóstico Monte Carlo) y condiciones de optimalidad (KKT) en un mismo script favorece una cultura de verificación integral que suele faltar en implementaciones aplicadas de IA para decisiones públicas. En particular, la explicitación de márgenes de holgura y residuos numéricos aporta información más rica que un simple estado binario.

Al mismo tiempo, los resultados invitan a distinguir cuidadosamente entre distintos niveles de evidencia. Algunas comprobaciones son esencialmente exactas dentro de aritmética de punto flotante (por ejemplo, PDL, flujo ocupacional, derivadas simbólicas), mientras que otras son diagnósticas o aproximadas (por ejemplo, Monte Carlo IS y contracción robusta por grilla). Esta heterogeneidad no debilita el protocolo; por el contrario, lo hace más honesto metodológicamente, siempre que se explicita el estatus de cada prueba, como se ha hecho aquí.

## 12. Conclusión

El desarrollo de una inteligencia artificial para la optimización de políticas públicas exige un marco mucho más exigente que el habitual en problemas de predicción o control en entornos cerrados. A lo largo de este trabajo se ha mostrado que una formulación rigurosa debe integrar, de manera simultánea, dinámica secuencial, restricciones normativas, identificación causal, evaluación contrafactual, robustez frente a incertidumbre y seguridad respecto a políticas base. La idea central que emerge es que una IA pública no puede ser entendida como un mero algoritmo de maximización, sino como una arquitectura formal compuesta por capas distintas pero interdependientes: una capa causal-estadística que identifica y estima el problema, una capa de optimización que calcula políticas factibles y una capa normativa-institucional que fija los límites, prioridades y condiciones de despliegue.

Desde el punto de vista matemático, el documento ha establecido que los procesos de decisión de Markov restringidos, su representación mediante medidas ocupacionales y su formulación primal-dual proporcionan una base especialmente poderosa

para modelar políticas públicas dinámicas. Esta representación permite probar existencia de soluciones óptimas, realizabilidad de medidas ocupacionales, dualidad fuerte, condiciones KKT e interpretación de multiplicadores como precios sombra de restricciones públicas. Asimismo, el uso del lema de diferencia de desempeño y de resultados de mejora de política permite conectar el análisis de optimalidad con garantías de no deterioro respecto a un régimen base.

No obstante, también se ha enfatizado que la corrección interna del optimizador no basta. La validez de la política recomendada depende críticamente de la identificación causal de transiciones, recompensas y restricciones a partir de datos históricos. Sin esa capa causal, la IA puede optimizar patrones observados que no representan efectos de intervención genuinos. De ahí la importancia de la g-recursión, la ignorabilidad secuencial, la positividad y la suficiencia de la agregación al pasar desde historias completas a estados Markovianos agregados. Del mismo modo, la evaluación fuera de política, la doble robustez y las cotas de error por mala especificación del modelo son esenciales para cuantificar el riesgo de optimizar sobre representaciones imperfectas del sistema social.

El análisis también ha mostrado que la robustez no es una característica opcional, sino una exigencia constitutiva del uso público de la IA. La incertidumbre sobre transiciones, los shocks exógenos, la no estacionariedad y los límites de transportabilidad hacen necesario trabajar con conjuntos robustos, reglas de seguridad probabilística y mecanismos de regularización respecto a políticas base. A ello se suma la necesidad de incorporar explícitamente equidad, implementabilidad y estabilidad institucional, porque una política puede ser eficiente en términos agregados y, sin embargo, distributivamente regresiva, administrativamente inviable o políticamente desestabilizadora. La matemática permite formalizar estos conflictos, detectar incompatibilidades de factibilidad y hacer explícitos los costos marginales de endurecer o relajar restricciones.

En términos sustantivos, el resultado general del trabajo es una teoría de la IA pública como operador auditable de decisión bajo restricciones. Esta teoría no promete eliminar el juicio político ni sustituir la deliberación democrática. Por el contrario, muestra que incluso el sistema algorítmico más sofisticado necesita parámetros normativos explícitos, validación causal, reglas de seguridad y mecanismos de supervisión humana. La inteligencia artificial puede fortalecer la racionalidad pública sólo si sus supuestos, objetivos, restricciones y márgenes de error son transparentes y revisables.

En consecuencia, la principal conclusión es doble. Primero, sí es posible construir una base matemática para sistemas de IA orientados a la optimización de políticas públicas. Segundo, esa base sólo es legítima cuando distingue con claridad entre predicción y causalidad, entre optimalidad técnica y aceptabilidad normativa, y entre eficiencia computacional y responsabilidad institucional. La verdadera promesa de la IA en el sector público no reside en automatizar la política, sino en hacer explícitas, trazables y formalmente evaluables las condiciones bajo las cuales una decisión pública puede considerarse eficiente, segura, robusta y socialmente justificable.

## Referencias

- Athey, S., & Wager, S. (2021). Policy Learning With Observational Data. *Econometrica*, 89(1), 133–161.
- Buckley, M., Papathanasiou, K., & Spanopoulos, A. (2025). *Primal-Dual Sample Complexity Bounds for Constrained Markov Decision Processes with Multiple Constraints*. arXiv preprint arXiv:2503.06751v1.
- Chen, K., Wei, H., Deng, Z., & Lin, S. (2026). *Towards Fast Safe Online Reinforcement Learning via Policy Finetuning*. Transactions on Machine Learning Research. arXiv preprint arXiv:2412.04426v3.
- De Moor, J., Weytjens, H., De Smedt, J., & De Weerd, J. (2026). *SCOPE: Sequential Causal Optimization of Process Interventions*. arXiv preprint arXiv:2512.17629v3.
- Ertefaie, A., & Strawderman, R. L. (2018). Constructing dynamic treatment regimes over indefinite time horizons. *Biometrika*, 105(4), 963–977.
- Feinberg, E. A., Jaśkiewicz, A., & Nowak, A. S. (2019). *Constrained discounted Markov decision processes with Borel state spaces*. arXiv preprint arXiv:1806.00190v7.
- Kalagarla, K. C., Jain, R., & Nuzzo, P. (2021). *A Sample-Efficient Algorithm for Episodic Finite-Horizon MDP with Constraints*. The Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21).
- Sawarni, A., Jin, J., Whitehouse, J., & Syrgkanis, V. (2026). *Policy Learning with Abstention*. arXiv preprint arXiv:2510.19672v3.
- Stradi, F. E., Lunghi, A., Castiglioni, M., Marchesi, A., & Gatti, N. (2024). *Learning Constrained Markov Decision Processes With Non-stationary Rewards and Constraints*. arXiv preprint arXiv:2405.14372v2.
- Thomas, P. S., Theodorou, G., & Ghavamzadeh, M. (2015). High Confidence Off-Policy

- Evaluation. *Proceedings of the Twenty-Ninth AAI Conference on Artificial Intelligence*.
- Wendl, M., As, Y., Prajapat, M., Pollak, A., Coros, S., & Krause, A. (2026). *Safe Exploration via Policy Priors*. Published as a conference paper at ICLR 2026. arXiv preprint arXiv:2601.19612v1.
- Xiao, X., Hu, X., Ye, Y., & Yue, X. (2026). *CausalGDP: Causality-Guided Diffusion Policies for Reinforcement Learning*. arXiv preprint arXiv:2602.09207v1.
- Xu, Y., Ganesh, S., Mondal, W. U., Bai, Q., & Aggarwal, V. (2025). *Global Convergence for Average Reward Constrained MDPs with Primal-Dual Actor Critic Algorithm*. 39th Conference on Neural Information Processing Systems (NeurIPS 2025). arXiv preprint arXiv:2505.15138v2.
- Yu, H. (2021). *On Linear Programming for Constrained and Unconstrained Average-Cost Markov Decision Processes with Countable Action Spaces and Strictly Unbounded Costs*. arXiv preprint arXiv:1905.12095v3.
- Zuo, Q., & He, F. (2025). *Ensuring Safety in an Uncertain Environment: Constrained MDPs via Stochastic Thresholds*. arXiv preprint arXiv:2504.04973v3.

# La responsabilidad política en la toma de decisiones automatizadas del Estado democrático

## *Political Responsibility in Automated Decision-Making in the Democratic State*

Marysol Estrella Hernández García<sup>1</sup>

<sup>1</sup> Universidad Autónoma de Guerrero. Guerrero, México.  
20423309@uagro.mx (Autora de correspondencia)

Carolina Aguilar Ramos<sup>2</sup>

<sup>2</sup> Universidad Autónoma del Estado de Hidalgo. Hidalgo, México.  
carolina\_aguilar@uaeh.edu.mx

**Resumen.** La incorporación de sistemas automatizados en la toma de decisiones del Estado constituye una transformación estructural que incide directamente en las bases de la responsabilidad política en contextos democráticos. Lejos de tratarse de una innovación meramente técnica, la mediación algorítmica altera la forma en que se configuran las decisiones públicas, introduciendo nuevas tensiones en la relación entre poder, control y rendición de cuentas. El presente trabajo tiene como objetivo analizar cómo la adopción de sistemas automatizados reconfigura la responsabilidad política del Estado, particularmente en aquellos ámbitos donde las decisiones inciden en derechos, obligaciones y políticas públicas. A partir de un enfoque cualitativo de carácter teórico-conceptual, se realiza una revisión doctrinal en teoría política y administración pública, así como un análisis crítico de la delegación algorítmica como categoría analítica. Los resultados identifican tres efectos estructurales: la ruptura de la cadena clásica de imputación, la fragmentación institucional de la responsabilidad y el incremento de la opacidad técnica en los procesos decisionales. Estos elementos no eliminan la responsabilidad política, pero sí dificultan su identificación y ejercicio efectivo dentro de los esquemas tradicionales de control democrático. Se concluye que la automatización no desplaza la dimensión política de la decisión estatal, sino que exige su reconfiguración. En este contexto, resulta necesario redefinir los criterios de imputación política mediante el fortalecimiento de la trazabilidad, la supervisión humana efectiva y la capacidad de explicación de las decisiones, como condiciones indispensables para preservar la rendición de cuentas en el Estado contemporáneo.

**Palabras clave:** Responsabilidad política; decisiones automatizadas; Estado; delegación algorítmica; democracia; administración pública; inteligencia artificial

**Abstract.** The incorporation of automated systems into state decision-making represents a structural transformation that directly impacts the foundations of political responsibility in democratic contexts. Far from being merely a technical innovation, algorithmic mediation reshapes the way public decisions are produced, introducing new tensions in the relationship between power, control, and accountability. This paper aims to analyze how the adoption of automated systems reconfigures political responsibility within the State, particularly in areas where decisions affect rights, obligations, and public policies. Using a qualitative and theoretical-conceptual approach, the study conducts a doctrinal review in political theory and public administration, alongside a critical analysis of algorithmic delegation as an analytical category. The findings identify three structural effects: the disruption of the traditional chain of responsibility, the institutional fragmentation of accountability, and the increase in technical opacity in decision-making processes. These elements do not eliminate political responsibility but complicate its identification and effective exercise within traditional democratic control frameworks. The paper concludes that automation does not remove the political dimension of state decision-making, but rather requires its reconfiguration. In this context, it becomes necessary to redefine criteria of political accountability through enhanced traceability, effective human oversight, and explainability, as essential conditions for preserving accountability in contemporary governance.

**Keywords:** Political responsibility; automated decision-making; State; algorithmic delegation; democracy; public administration; artificial intelligence

**Historial:** Recibido: 30 de marzo de 2026 | Aceptado: 18 de mayo de 2026

**Ética/Financiamiento:** Las autoras declaran no tener conflictos de interés. La investigación fue autofinanciada y no recibió financiamiento externo.

Este documento está bajo una licencia

CC BY-NC-ND 4.0.

**Cómo citar (APA):** Hernández García, M. E., & Aguilar Ramos, C. (2026). La responsabilidad política en la toma de decisiones automatizadas del Estado democrático. *Conocimiento i Política*, 7(1), 161-181. DOI: 10.64480/cip.2026.n7.04.

## 1. Introducción

En las últimas décadas, los Estados han incorporado sistemas automatizados en procesos de gestión pública, clasificación de información, asignación de recursos y evaluación de riesgos. Esta incorporación no se limita a funciones instrumentales o de apoyo administrativo, sino que progresivamente incide en espacios decisionales que afectan de manera directa derechos, obligaciones y la orientación misma de las políticas públicas. La digitalización del Estado no puede entenderse únicamente como una modernización tecnológica; constituye, en términos más profundos, una reconfiguración de la forma en que el poder público se ejerce, se organiza y se justifica. En este sentido, la introducción de sistemas automatizados en la administración pública altera no solo los medios, sino también las condiciones bajo las cuales se produce la decisión estatal (Bovens & Zouridis, 2002).

Desde una lectura centrada en la eficiencia, podría sostenerse que estos sistemas permiten optimizar recursos, reducir tiempos y mejorar la consistencia de las decisiones. Sin embargo, esta aproximación resulta limitada. El problema no se agota en la mejora de la gestión, sino que se desplaza hacia la estructura misma de la decisión pública. La pregunta relevante deja de ser cómo decide el Estado para convertirse en quién decide realmente cuando el resultado se encuentra mediado por sistemas automatizados. Esta distinción no es menor, pues implica cuestionar la ubicación efectiva del poder en contextos donde la decisión ya no es producto exclusivo de la deliberación humana, sino de procesos algorítmicos cuya lógica no siempre es accesible ni plenamente comprensible.

En este punto, la literatura sobre el gobierno digital ha advertido que la incorporación de tecnologías en la administración pública puede desplazar progresivamente la discrecionalidad desde los funcionarios hacia sistemas estructurados de decisión, configurando nuevas formas de ejercicio del poder administrativo (Bovens & Zouridis, 2002). No obstante, este desplazamiento no elimina la dimensión política de la decisión. Por el contrario, la reconfigura. Como ha señalado Kitchin, los sistemas basados en datos y algoritmos no operan en un vacío normativo, sino que incorporan criterios, categorías y supuestos que inciden directamente en los resultados que producen (Kitchin, 2017). En consecuencia, la automatización no neutraliza la decisión, sino que introduce nuevas capas de mediación que complejizan su análisis.

La teoría democrática clásica parte de una premisa que ha sido considerada estructural: el ejercicio del poder debe ir acompañado de responsabilidad política. Esta responsabilidad permite identificar a quienes toman decisiones y someterlos a mecanismos de control institucional, ya sea a través de instancias parlamentarias, procedimientos administrativos o dinámicas electorales. La posibilidad de exigir cuentas constituye, en este sentido, una condición de legitimidad del poder público (Rosanvallon, 2008). Sin embargo, esta lógica se ve tensionada cuando la decisión se encuentra mediada por sistemas automatizados.

La mediación técnica introduce un problema que no puede resolverse mediante categorías tradicionales. Cuando una decisión pública depende de un sistema automatizado, la identificación del decisor se vuelve difusa. La autoridad que formalmente adopta la decisión puede no tener control directo sobre los criterios que la determinan; los desarrolladores que diseñan el sistema no forman parte de la estructura estatal; los técnicos que lo operan no necesariamente participan en la definición de sus parámetros. Este entramado genera una estructura decisional en la que la relación entre decisión y responsabilidad deja de ser lineal.

No obstante, sería problemático asumir que esta complejidad justifica la dilución de la responsabilidad política. La idea de que los sistemas automatizados producen decisiones objetivas o imparciales ha sido ampliamente cuestionada. O'Neil ha mostrado que los modelos algorítmicos pueden reproducir y amplificar sesgos existentes, generando efectos que no son neutrales desde el punto de vista social y político (O'Neil, 2016). En esta misma línea, la automatización no elimina la dimensión valorativa de la decisión, sino que la desplaza hacia espacios menos visibles, lo que dificulta su identificación y control.

El problema, por tanto, no reside en la existencia de tecnología en la administración pública, sino en la forma en que su incorporación reconfigura la cadena de imputación política. La delegación algorítmica no implica una transferencia de responsabilidad, pero sí introduce condiciones que pueden facilitar su fragmentación. Como ha señalado Bovens, la rendición de cuentas depende de la posibilidad de atribuir responsabilidad a actores identificables dentro de un marco institucional (Bovens, 2007). En contextos en los que la atribución de responsabilidad se vuelve difusa, los mecanismos de control democrático pierden efectividad.

Esta situación se vuelve especialmente relevante en ámbitos donde las decisiones

estatales tienen impactos significativos sobre derechos o bienes colectivos. El ámbito ambiental constituye un ejemplo ilustrativo, en la medida en que el uso de sistemas automatizados para monitoreo, evaluación o asignación de recursos comienza a incidir en la toma de decisiones públicas. Sin embargo, el problema que aquí se plantea no se limita a un sector específico. Se trata de una cuestión estructural que atraviesa la forma en que el Estado contemporáneo decide.

En este contexto, la pregunta central que orienta el presente trabajo es la siguiente: ¿cómo se configura la responsabilidad política cuando el Estado adopta sistemas automatizados para la toma de decisiones? La hipótesis que se sostiene es que la automatización fragmenta la cadena tradicional de imputación política, dificulta la identificación clara del sujeto responsable y genera zonas de ambigüedad institucional que debilitan los mecanismos clásicos de control democrático, sin que ello implique la desaparición de la responsabilidad del poder público.

Este problema no puede abordarse desde una postura neutral. Asumir que la complejidad técnica justifica la opacidad implica aceptar una transformación del poder público en la que la decisión se desvincula de la posibilidad efectiva de exigir responsabilidad. Por el contrario, cuanto más mediada se encuentre la decisión estatal por sistemas automatizados, mayor debe ser la exigencia de claridad en su imputación. La sofisticación técnica no puede convertirse en un argumento para reducir los estándares de control democrático.

En consecuencia, el análisis de la responsabilidad política en contextos de decisión automatizada no constituye un debate accesorio, sino una cuestión central para la teoría democrática contemporánea. Lo que está en juego no es únicamente la forma en que el Estado decide, sino la posibilidad misma de sostener un modelo en el que el ejercicio del poder permanezca sujeto a mecanismos efectivos de rendición de cuentas. Ignorar esta transformación no implica neutralidad, sino la aceptación implícita de formas de decisión estatal en las que la responsabilidad política se vuelve progresivamente más difícil de ejercer.

## **2. Metodología**

El presente trabajo se inscribe en un enfoque cualitativo de carácter teórico-conceptual, orientado a examinar las implicaciones políticas que se derivan de la incorporación de

sistemas automatizados en la toma de decisiones estatales. La elección de este enfoque responde a la naturaleza del problema planteado, que no puede ser reducido a una cuestión empírica susceptible de medición directa sin perder de vista las categorías normativas que lo estructuran. Analizar la responsabilidad política en contextos de automatización exige desplazarse hacia un plano en el que las nociones de imputación, control y legitimidad puedan ser problematizadas en términos conceptuales, más que verificadas a partir de indicadores cuantificables.

El análisis parte de una premisa que condiciona todo el desarrollo metodológico: la automatización de decisiones no constituye únicamente una herramienta de gestión, sino una forma de mediación que interviene en la estructura misma del ejercicio del poder público. Bajo esta perspectiva, la decisión estatal no puede ser entendida exclusivamente como un acto final atribuible a un órgano o funcionario, sino como el resultado de un proceso que, en contextos de automatización, se encuentra atravesado por elementos técnicos cuya lógica no siempre es transparente ni plenamente accesible. Este desplazamiento obliga a reconsiderar el punto de observación del análisis. En lugar de centrarse en la decisión como producto terminado, el estudio dirige su atención hacia las condiciones que hacen posible su producción.

A partir de esta premisa, la responsabilidad política es abordada como una categoría relacional que presupone la existencia de un vínculo entre quien decide y quien puede ser llamado a responder por esa decisión. Este vínculo no se entiende como un dato dado, sino como una construcción institucional que depende de la posibilidad de identificar sujetos, delimitar competencias y establecer mecanismos de control. La relevancia de esta noción radica en que permite evaluar hasta qué punto la incorporación de sistemas automatizados altera las condiciones bajo las cuales dicha relación puede sostenerse. Cuando la decisión se encuentra mediada por procesos algorítmicos, la identificación del decisor deja de ser evidente, lo que introduce una tensión que no puede resolverse sin revisar las bases sobre las que se construye la imputación política.

La noción de delegación algorítmica se incorpora como una herramienta analítica para describir esta forma de mediación. No se emplea en un sentido formal, como si se tratara de una transferencia explícita de competencias, sino como una categoría que permite captar la manera en que los sistemas automatizados intervienen en la producción de decisiones públicas. Esta precisión resulta necesaria, ya que evita asumir que la automatización implica la sustitución del decisor humano. Lo que

se observa, en cambio, es una reconfiguración del proceso decisional en la que la intervención técnica no elimina la decisión estatal, pero sí modifica las condiciones bajo las cuales esta se produce.

Desde esta perspectiva, el análisis se orienta hacia la estructura de imputación política. La pregunta que subyace no es únicamente quién toma la decisión en términos formales, sino cómo se distribuye la responsabilidad en un entramado en el que intervienen múltiples actores y mediaciones. La atención se desplaza hacia la forma en que se articula la relación entre decisión, autoridad y control en contextos donde la intervención técnica introduce niveles adicionales de complejidad. Este enfoque permite evitar interpretaciones simplistas en las que la responsabilidad se atribuye automáticamente a la autoridad formal, sin considerar las condiciones reales en las que la decisión se genera.

La elección de no recurrir a estudios de caso específicos responde a la intención de evitar una aproximación fragmentaria del fenómeno. Si bien existen múltiples ejemplos del uso de sistemas automatizados en la administración pública, su análisis aislado puede conducir a conclusiones limitadas o dependientes del contexto particular en el que se insertan. En lugar de ello, el estudio busca identificar patrones estructurales que permitan comprender cómo se transforma la responsabilidad política en términos generales. Esta decisión metodológica no implica desconocer la importancia de los casos empíricos, sino reconocer que el problema planteado exige un nivel de abstracción que permita captar sus implicaciones más amplias.

El enfoque adoptado asume una postura crítica frente a la idea de que la tecnología puede integrarse al aparato estatal sin alterar sus fundamentos. No se trata de sostener que la automatización sea incompatible con la organización democrática del poder, pero tampoco de aceptar sin cuestionamiento que su incorporación sea neutral. La metodología se construye desde la premisa de que toda forma de mediación técnica introduce efectos que deben ser analizados en relación con las condiciones de posibilidad de la responsabilidad política. Esto implica reconocer que la complejidad técnica no solo plantea desafíos operativos, sino también problemas de orden institucional.

En este sentido, el análisis no se orienta a evaluar el desempeño técnico de los sistemas automatizados ni a examinar su eficacia en términos de resultados. Su interés radica en las implicaciones que su uso tiene para la estructura de imputación política. Esta delimitación permite concentrar la atención en el problema central del

trabajo, evitando dispersarse en dimensiones que, si bien son relevantes, no resultan determinantes para la pregunta de investigación planteada. Del mismo modo, el estudio se mantiene en un nivel general desde el punto de vista jurídico, sin centrarse en un ordenamiento específico, lo que permite preservar la aplicabilidad del análisis a distintos contextos institucionales.

Lejos de pretender ofrecer una respuesta definitiva, el enfoque metodológico adoptado busca abrir un espacio de problematización. La transformación que introduce la automatización en la toma de decisiones estatales no puede ser abordada mediante soluciones inmediatas ni mediante la simple adaptación de categorías existentes. Lo que se plantea es la necesidad de revisar las condiciones bajo las cuales la responsabilidad política puede seguir operando en un contexto en el que la decisión ya no es el resultado exclusivo de la acción humana directa. En este sentido, la metodología no se orienta a cerrar el debate, sino a evidenciar sus tensiones y a cuestionar las formas en que el poder público se reconfigura bajo condiciones de mediación tecnológica.

### **3. Ruptura de la cadena clásica de responsabilidad**

La responsabilidad política en los modelos tradicionales de organización estatal se construye sobre una premisa de relativa simplicidad estructural: la decisión puede ser atribuida a un sujeto identificable dentro de la estructura institucional. Esta relación, que vincula decisión y responsabilidad, no depende únicamente de la existencia de normas que así lo establezcan, sino de la posibilidad efectiva de reconstruir el proceso decisional y ubicar en él a quien ejerce el poder. En este sentido, la cadena de responsabilidad no es solo un mecanismo formal, sino una condición práctica que permite hacer operativa la rendición de cuentas.

Este fenómeno ha sido advertido en la literatura sobre rendición de cuentas, donde la atribución de responsabilidad se vuelve más compleja en estructuras institucionales no lineales (Bovens, 2007). Bajo este esquema, la decisión estatal puede ser comprendida como un acto que se produce dentro de una secuencia relativamente delimitada.

Existe un órgano competente, un funcionario responsable y un marco normativo que orienta su actuación. Incluso en contextos de discrecionalidad, la decisión sigue estando anclada en un sujeto cuya actuación puede ser evaluada, cuestionada o

sancionada. La complejidad institucional no elimina esta posibilidad; por el contrario, la presupone. En este sentido, la responsabilidad política se sostiene en la capacidad de mantener la relación entre la decisión y el sujeto responsable.

Sin embargo, la incorporación de sistemas automatizados en la toma de decisiones introduce una alteración significativa en esta lógica. La decisión deja de ser el resultado directo de una deliberación humana claramente identificable y pasa a depender de procesos mediados por sistemas técnicos cuya operación no siempre es transparente. Esta mediación no es superficial. No se limita a asistir al decisor, sino que puede incidir de manera determinante en el resultado final. La decisión ya no se produce exclusivamente en el momento en que es formalmente adoptada, sino que se construye a lo largo de un proceso en el que intervienen múltiples elementos, algunos de los cuales no son plenamente visibles.

Esta transformación afecta directamente la posibilidad de reconstruir la cadena de responsabilidad. Si la decisión es el resultado de un proceso técnico complejo, la identificación del sujeto responsable deja de ser evidente. La autoridad que formalmente adopta la decisión puede no haber intervenido en la definición de los criterios que la determinan; su margen de acción puede estar condicionado por parámetros previamente establecidos en el sistema. En este contexto, atribuir responsabilidad política exclusivamente a la autoridad formal implica ignorar las condiciones reales bajo las cuales la decisión se produce.

No obstante, desplazar la responsabilidad hacia los sistemas técnicos tampoco resuelve el problema. Los sistemas automatizados no son sujetos de responsabilidad en sentido político. No pueden ser llamados a rendir cuentas ni pueden ser objeto de sanción en los términos en que lo son los actores institucionales. La idea de que la responsabilidad puede transferirse a la tecnología no solo resulta conceptualmente problemática, sino que introduce un riesgo mayor: la posibilidad de que la decisión se desplace hacia un espacio en el que la exigencia de responsabilidad se vuelve difusa o incluso impracticable.

Lo que se configura, entonces, no es una desaparición de la responsabilidad política, sino una ruptura en la cadena que tradicionalmente permitía su ejercicio. Esta ruptura no implica la inexistencia de responsables, sino la dificultad para identificarlos de manera clara dentro de una estructura decisional que se ha vuelto más compleja. La decisión sigue produciendo efectos y sigue siendo atribuible al Estado, pero las

condiciones que permitían vincularla con un sujeto específico se debilitan.

Este debilitamiento no es uniforme ni se manifiesta de la misma manera en todos los casos. Depende del grado de intervención del sistema automatizado, del tipo de decisión de que se trate y del diseño institucional en el que se inserta. Sin embargo, existe un elemento común: la mediación técnica introduce una distancia entre el acto formal de decisión y los procesos que lo generan. Esta distancia dificulta la reconstrucción del proceso decisional y, con ello, la posibilidad de atribuir responsabilidad de manera directa.

En este contexto, la cadena de responsabilidad deja de operar como una secuencia lineal y pasa a configurarse como una red de relaciones en la que intervienen múltiples actores y niveles de decisión. La decisión ya no puede ser entendida como el resultado de una única voluntad, sino como el producto de un entramado en el que se combinan elementos técnicos, institucionales y humanos. Esta reconfiguración no elimina la responsabilidad, pero sí transforma las condiciones bajo las cuales puede ser ejercida.

El problema se agudiza cuando la autoridad que formalmente decide no cuenta con los elementos necesarios para comprender plenamente el funcionamiento del sistema en el que se apoya. En estos casos, la decisión se adopta sobre la base de resultados cuya lógica interna no es accesible. La responsabilidad política, que presupone la posibilidad de justificar la decisión, se enfrenta entonces a una limitación estructural: no es posible explicar completamente aquello que no se comprende en su totalidad. Esta situación introduce una tensión entre la exigencia de rendición de cuentas y las condiciones reales en las que la decisión se produce.

La ruptura de la cadena clásica de responsabilidad no debe ser interpretada como un fenómeno meramente técnico. Se trata de una transformación que incide en la forma en que el poder público se organiza y se legitima. Si la responsabilidad política depende de la posibilidad de identificar al decisor, cualquier alteración en esta relación afecta directamente la calidad del control democrático. La dificultad para reconstruir la cadena de decisión no es un problema operativo, sino un desafío estructural para la teoría y la práctica de la responsabilidad política.

Frente a esta situación, mantener intactas las categorías tradicionales de imputación resulta insuficiente. Asumir que la responsabilidad puede seguir operando bajo las mismas condiciones implica desconocer la magnitud de la transformación introducida por la automatización. Al mismo tiempo, aceptar que la complejidad técnica impide

la atribución de responsabilidad conduciría a un escenario en el que la decisión estatal se desplaza hacia un espacio de opacidad incompatible con los principios básicos de la organización democrática.

Lo que emerge es, entonces, una tensión que no puede resolverse mediante una simple adaptación normativa. La ruptura de la cadena de responsabilidad obliga a replantear las condiciones bajo las cuales la imputación política puede seguir siendo operativa. No se trata de identificar nuevos sujetos responsables de manera arbitraria, sino de reconocer que la estructura misma de la decisión ha cambiado y que, con ella, deben revisarse los mecanismos que permiten exigir responsabilidad.

En este sentido, la automatización no elimina la responsabilidad política, pero sí la desestabiliza. Introduce un nivel de complejidad que hace insuficiente la lógica lineal sobre la que tradicionalmente se ha construido la imputación. La cadena ya no puede ser entendida como una secuencia clara y delimitada, sino como un proceso distribuido cuya reconstrucción exige nuevas herramientas analíticas. Ignorar esta transformación no implica preservar la responsabilidad, sino debilitarla al mantenerla anclada en una estructura que ya no corresponde a la forma en que la decisión estatal se produce.

#### **4. Fragmentación institucional de la responsabilidad**

Si la ruptura de la cadena clásica de responsabilidad evidencia la dificultad para identificar un decisor único, la fragmentación institucional profundiza este problema al mostrar que la decisión automatizada no solo se vuelve menos clara, sino que se distribuye en un entramado de actores cuya participación es parcial, diferenciada y, en muchos casos, difícil de articular en términos de responsabilidad política.

En los modelos tradicionales de organización estatal, la complejidad institucional no elimina la posibilidad de imputar responsabilidad. Aun cuando intervienen distintos niveles jerárquicos o áreas administrativas, la decisión puede ser reconducida hacia un órgano o autoridad que asume su titularidad. La estructura institucional, en este sentido, permite ordenar la responsabilidad dentro de un marco relativamente estable. La dispersión de funciones no implica necesariamente dispersión de responsabilidad, en la medida en que existe un punto de convergencia que permite atribuir la decisión.

En este sentido, diversos estudios han señalado que la intervención de múltiples actores en procesos mediados por sistemas tecnológicos dificulta la asignación clara de responsabilidad política (Kitchin, 2017). La automatización introduce una lógica distinta. La decisión ya no se produce únicamente dentro de los márgenes de la estructura administrativa, sino que se configura a partir de la interacción entre actores que no necesariamente comparten el mismo estatuto institucional.

El diseño del sistema puede estar a cargo de desarrolladores externos; su implementación puede depender de proveedores tecnológicos; su operación puede recaer en equipos técnicos especializados; su uso final corresponde a autoridades administrativas que adoptan la decisión formal. Cada uno de estos actores interviene en una fase distinta del proceso, sin que exista una integración clara que permita reconstruir la totalidad de la decisión desde un solo punto.

Esta distribución funcional no es problemática por sí misma. La administración pública siempre ha operado mediante la división del trabajo y la especialización. Lo que cambia en el contexto de la automatización es la forma en que estas funciones se articulan en la producción de la decisión. La intervención de sistemas automatizados introduce un elemento de dependencia técnica que condiciona el resultado final. La autoridad que decide no lo hace en un vacío, sino sobre la base de un sistema que ya ha procesado información, aplicado criterios y generado una salida que orienta, e incluso determina, la decisión.

En este escenario, la responsabilidad política tiende a diluirse no porque desaparezca, sino porque se distribuye en una estructura en la que ningún actor concentra la totalidad del proceso decisional. La autoridad administrativa puede alegar que la decisión se adoptó conforme a los resultados del sistema; los técnicos pueden sostener que su función se limita a operar una herramienta previamente diseñada; los desarrolladores pueden afirmar que el sistema funciona de acuerdo con especificaciones establecidas; los proveedores pueden invocar el cumplimiento de obligaciones contractuales. Cada uno de estos argumentos resulta, en cierta medida, plausible. Sin embargo, en su conjunto, configuran una situación en la que la responsabilidad se fragmenta sin que exista un punto claro de imputación.

Esta fragmentación introduce una forma de dispersión que no es equivalente a la tradicional distribución de competencias. En los modelos clásicos, la distribución funcional se encuentra acompañada de mecanismos que permiten reconducir la

responsabilidad hacia niveles superiores o hacia autoridades claramente identificables. En cambio, la fragmentación asociada a la automatización dificulta este proceso, ya que la decisión se construye a partir de elementos que no pueden ser fácilmente integrados en una única línea de responsabilidad.

El problema no radica únicamente en la multiplicidad de actores, sino en la naturaleza heterogénea de su participación. No todos los actores involucrados se encuentran sujetos a los mismos mecanismos de control. Mientras que las autoridades públicas pueden ser objeto de supervisión política, los desarrolladores o proveedores tecnológicos operan bajo lógicas contractuales que no necesariamente se traducen en responsabilidad política. Esta asimetría genera un desajuste entre la estructura del proceso decisional y los mecanismos disponibles para exigir responsabilidad.

La fragmentación institucional también incide en la forma en que se percibe la decisión. Cuando el proceso decisional se distribuye entre múltiples actores, la decisión final puede aparecer como el resultado de un sistema más que de una voluntad política. Esta percepción no es irrelevante. La responsabilidad política no solo depende de la existencia de mecanismos formales de control, sino también de la posibilidad de identificar a quienes ejercen el poder. Si la decisión se presenta como un producto técnico, la exigencia de responsabilidad puede debilitarse en la práctica, aun cuando formalmente siga siendo atribuible al Estado. En este contexto, la fragmentación no implica la ausencia de responsabilidad, pero sí su debilitamiento operativo. La dificultad para articular la participación de los distintos actores en una narrativa coherente de la decisión genera un entorno en el que la rendición de cuentas se vuelve más compleja. No se trata únicamente de identificar quién intervino, sino de determinar en qué medida cada intervención contribuyó al resultado final. Esta tarea se vuelve especialmente difícil cuando los procesos técnicos no son plenamente comprensibles para quienes deben ejercer el control.

La fragmentación también plantea un problema desde el punto de vista institucional. Los mecanismos tradicionales de control político no están diseñados para operar en estructuras decisionales altamente distribuidas. La supervisión parlamentaria, por ejemplo, presupone la existencia de autoridades que pueden ser llamadas a responder por sus decisiones. Cuando la decisión se construye a partir de una red de actores, esta lógica se ve tensionada. La autoridad que comparece puede no tener control efectivo sobre todos los elementos que influyen en la decisión, lo que limita la eficacia del control.

Este escenario no conduce necesariamente a la desaparición de la responsabilidad política, pero sí a su reconfiguración. La fragmentación obliga a replantear la forma en que se articulan los mecanismos de imputación en contextos donde la decisión no puede ser atribuida a un solo actor. La insistencia en modelos de responsabilidad centrados en la figura del decisor individual resulta insuficiente para dar cuenta de esta transformación.

La cuestión central no es si existe responsabilidad, sino cómo puede ser reconstruida en un contexto en el que la decisión se produce de manera distribuida. Esto implica reconocer que la fragmentación no es un fenómeno accidental, sino una consecuencia estructural de la incorporación de sistemas automatizados en la toma de decisiones. Ignorar esta condición no permite preservar la responsabilidad, sino que contribuye a su debilitamiento al mantener esquemas que no corresponden a la realidad del proceso decisional.

La fragmentación institucional, en suma, no elimina la responsabilidad política, pero sí la dispersa en un entramado que dificulta su ejercicio efectivo. La decisión sigue siendo estatal, pero su producción se encuentra distribuida en una red de actores cuya articulación no siempre es clara. Esta situación plantea un desafío que no puede resolverse mediante la simple identificación de un responsable formal, sino que exige repensar las condiciones bajo las cuales la responsabilidad política puede seguir siendo operativa en contextos de decisión automatizada.

## **5. Opacidad técnica y límites del control democrático**

La opacidad de los sistemas algorítmicos ha sido ampliamente documentada, particularmente en relación con su impacto en la rendición de cuentas y el control democrático (O'Neil, 2016). La ruptura de la cadena de responsabilidad y la fragmentación institucional encuentran un punto de convergencia en un fenómeno que atraviesa ambos procesos: la opacidad técnica.

Si la decisión automatizada dificulta la identificación del decisor y distribuye la responsabilidad entre múltiples actores, la opacidad introduce un obstáculo adicional, al limitar la posibilidad misma de comprender cómo se produce la decisión. En este sentido, no se trata únicamente de saber quién interviene, sino de poder explicar qué ocurre dentro del proceso decisional.

En los modelos tradicionales, la exigencia de responsabilidad política presupone la posibilidad de justificar la decisión. La autoridad no solo debe responder por el resultado, sino también por las razones que lo sustentan. Esta exigencia se vincula con la idea de que el ejercicio del poder debe ser inteligible para quienes están sujetos a él. La justificación no es un elemento accesorio, sino una condición para que la rendición de cuentas sea efectiva. Sin comprensión, la responsabilidad pierde contenido.

La incorporación de sistemas automatizados tensiona esta condición. Cuando la decisión se apoya en modelos cuya lógica interna no es plenamente accesible, la posibilidad de ofrecer una explicación completa se ve limitada. No se trata únicamente de la complejidad técnica en abstracto, sino de la forma en que dicha complejidad se traduce en una barrera para la comprensión. La decisión puede ser formalmente atribuida a una autoridad, pero las razones que la sustentan pueden permanecer, en la práctica, fuera de su alcance.

Esta situación introduce una forma de opacidad que no depende necesariamente de la intención de ocultar información. A diferencia de los supuestos tradicionales de opacidad, en los que la información existe pero no se revela, aquí el problema radica en la dificultad para reconstruir el proceso decisional en términos comprensibles. La opacidad no se produce únicamente por falta de acceso, sino por la imposibilidad práctica de traducir el funcionamiento del sistema en razones que puedan ser evaluadas desde una perspectiva política o jurídica.

El efecto de esta opacidad se extiende más allá de la autoridad que adopta la decisión. También afecta a los mecanismos de control. La supervisión parlamentaria, el control administrativo e incluso el escrutinio público presuponen que la decisión puede ser interrogada, discutida y eventualmente cuestionada. Cuando la lógica que la sustenta no es accesible, estos mecanismos se ven limitados. La posibilidad de exigir responsabilidad se reduce no porque desaparezca formalmente, sino porque carece de los elementos necesarios para ser ejercida de manera efectiva.

La opacidad técnica también altera la forma en que se construye la legitimidad de la decisión. En los modelos tradicionales, la legitimidad se sostiene, en parte, en la capacidad de la autoridad para ofrecer razones que justifiquen su actuación. Esta dimensión justificativa permite que la decisión sea evaluada no solo en términos de legalidad, sino también de razonabilidad. Cuando la decisión se apoya en procesos que no pueden ser explicados de manera comprensible, esta dimensión se debilita.

La decisión puede seguir siendo válida desde un punto de vista formal, pero su legitimidad se vuelve más difícil de sostener.

En este contexto, la opacidad no es un problema técnico aislado, sino una condición que incide directamente en la estructura de la responsabilidad política. La dificultad para explicar la decisión limita la posibilidad de exigir cuentas sobre ella. La responsabilidad presupone no solo la existencia de un sujeto responsable, sino la capacidad de evaluar su actuación. Cuando esta evaluación se vuelve impracticable, la responsabilidad pierde eficacia, aun cuando formalmente se mantenga.

El problema se intensifica cuando la autoridad que adopta la decisión depende del sistema automatizado para operar. En estos casos, la posibilidad de cuestionar el resultado se ve restringida por la falta de alternativas. La decisión no solo es difícil de explicar, sino también difícil de sustituir. La dependencia técnica introduce un elemento adicional de rigidez que reduce el margen de acción de la autoridad y, con ello, su capacidad de responder políticamente por la decisión.

La opacidad técnica también incide en la forma en que se distribuye la carga de la prueba en contextos de control. En ausencia de una explicación clara, cuestionar la decisión requiere reconstruir un proceso que no es fácilmente accesible. Esto desplaza la carga hacia quienes buscan impugnar la decisión, quienes deben enfrentarse a un sistema cuya lógica no está plenamente disponible. La asimetría que se genera no es menor, ya que afecta la posibilidad de ejercer derechos y de activar mecanismos de control. Frente a esta situación, la opacidad no puede ser tratada como un efecto secundario inevitable de la complejidad técnica. Su relevancia radica en que incide directamente en las condiciones bajo las cuales la responsabilidad política puede ser ejercida. Si la decisión no puede ser explicada, la exigencia de responsabilidad se vuelve parcial. No se trata de un problema de acceso a la información en sentido estricto, sino de la capacidad de traducir el funcionamiento técnico en términos comprensibles para el control democrático.

La opacidad, en suma, no elimina la responsabilidad política, pero sí la condiciona. Introduce un límite que no puede ser ignorado sin afectar la eficacia de los mecanismos de rendición de cuentas. La decisión sigue siendo atribuible al Estado, pero las condiciones que permiten evaluarla se debilitan. Este desplazamiento no es meramente técnico, sino profundamente político, en la medida en que redefine la relación entre decisión, justificación y control.

La combinación de ruptura de la cadena de responsabilidad, fragmentación institucional y opacidad técnica configura un escenario en el que la responsabilidad política se mantiene en el plano formal, pero enfrenta dificultades crecientes en su ejercicio efectivo. La automatización no elimina la posibilidad de exigir cuentas, pero sí introduce condiciones que hacen más complejo su funcionamiento. Reconocer estas condiciones no implica rechazar la incorporación de tecnología en la administración pública, sino asumir que su uso transforma las bases sobre las cuales se construye la responsabilidad.

## 6. Discusión

Los resultados expuestos muestran que la incorporación de sistemas automatizados en la toma de decisiones estatales no puede ser comprendida como una simple innovación técnica, sino como una transformación que incide directamente en las condiciones bajo las cuales opera la responsabilidad política. Esta transformación obliga a revisar las categorías tradicionales desde las que se ha pensado la relación entre decisión y control democrático, en la medida en que dichas categorías presuponen un modelo de decisión que ya no se corresponde plenamente con la realidad institucional contemporánea.

En la teoría clásica de la responsabilidad política, la posibilidad de imputar una decisión a un sujeto identificable constituye el eje sobre el cual se articula la rendición de cuentas. Como ha señalado Rosanvallon, la legitimidad democrática no se agota en la representación, sino que se sostiene también en la capacidad de someter el poder a mecanismos de vigilancia, control y evaluación permanente (Rosanvallon, 2008). Esta lógica presupone que la decisión puede ser reconstruida y atribuida, lo que permite que la exigencia de responsabilidad tenga un destinatario claro.

Sin embargo, los hallazgos de este trabajo sugieren que esta premisa se ve tensionada en contextos de automatización. La ruptura de la cadena de responsabilidad, la fragmentación institucional y la opacidad técnica no eliminan la responsabilidad política, pero sí alteran las condiciones bajo las cuales puede ser ejercida. En este sentido, la dificultad no radica en la inexistencia de responsables, sino en la transformación de la estructura que permitía identificarlos.

La literatura sobre gobierno digital ha advertido que la incorporación de tecnologías

en la administración pública puede generar un desplazamiento de la discrecionalidad desde los funcionarios hacia sistemas estructurados de decisión, lo que modifica la forma en que el poder administrativo se ejerce (Bovens & Zouridis, 2002). No obstante, interpretar este desplazamiento como una sustitución de la política por la técnica resulta problemático. Como se ha argumentado, los sistemas automatizados no operan de manera neutral, sino que incorporan criterios, supuestos y estructuras de valor que inciden en los resultados que producen (Kitchin, 2017).

Desde esta perspectiva, la automatización no elimina la dimensión política de la decisión, sino que la reconfigura. La decisión sigue siendo estatal, pero su producción se encuentra mediada por procesos que no siempre son transparentes ni plenamente accesibles. Esto introduce una tensión entre la exigencia de responsabilidad y las condiciones reales en las que la decisión se genera. La responsabilidad política presupone la posibilidad de explicar y justificar la decisión, pero esta posibilidad se ve limitada cuando la lógica del sistema no puede ser reconstruida de manera comprensible.

En este punto, resulta relevante considerar los aportes de la literatura sobre responsabilidad. Bovens ha señalado que la rendición de cuentas requiere la existencia de un actor que pueda ser llamado a responder, un foro ante el cual deba hacerlo y la posibilidad de evaluar su actuación (Bovens, 2007). En contextos de automatización, estos elementos no desaparecen, pero se ven tensionados. La identificación del actor se vuelve difusa, el foro enfrenta limitaciones para comprender el proceso decisional y la evaluación se dificulta ante la falta de explicaciones accesibles.

Esta situación no implica que la responsabilidad política deba ser abandonada o sustituida por otras formas de control. Por el contrario, refuerza la necesidad de repensar sus condiciones de posibilidad. La complejidad técnica no puede ser aceptada como un argumento que justifique la opacidad. Como advierte O'Neil, los sistemas algorítmicos pueden generar efectos significativos en la vida de las personas sin que exista una comprensión clara de cómo se producen dichos efectos (O'Neil, 2016). En este sentido, la exigencia de responsabilidad no se reduce, sino que se intensifica.

Los resultados de este trabajo permiten sostener que la responsabilidad política en contextos de decisión automatizada no desaparece, pero sí se transforma en un esquema que ya no puede ser comprendido a partir de una lógica lineal. La cadena de imputación tradicional, que vinculaba decisión y sujeto responsable de manera relativamente directa, resulta insuficiente para dar cuenta de procesos decisionales

distribuidos, mediados por sistemas técnicos y caracterizados por niveles variables de opacidad.

Esta transformación plantea un desafío para la teoría democrática. Si la responsabilidad política es una condición de legitimidad del poder, cualquier alteración en las condiciones que permiten su ejercicio afecta directamente la calidad de la democracia. La dificultad para identificar responsables, reconstruir decisiones y ofrecer explicaciones comprensibles introduce un riesgo que no puede ser ignorado: la consolidación de formas de ejercicio del poder que, aun siendo formalmente atribuibles al Estado, escapan en la práctica a los mecanismos tradicionales de control.

Frente a este escenario, sostener que la automatización exige una reconfiguración de la responsabilidad política no implica adoptar una postura tecnofóbica ni rechazar la incorporación de herramientas automatizadas en la administración pública. Implica reconocer que la estructura de la decisión ha cambiado y que, con ella, deben adaptarse los mecanismos que permiten exigir responsabilidad. La insistencia en esquemas tradicionales de imputación no preserva la responsabilidad, sino que la debilita al mantenerla anclada en un modelo que ya no describe adecuadamente la realidad.

En este sentido, la discusión no debe centrarse en si los sistemas automatizados pueden o no ser utilizados en la toma de decisiones públicas, sino en las condiciones bajo las cuales su uso es compatible con la exigencia de responsabilidad política. Esto supone, entre otras cuestiones, garantizar que la mediación técnica no impida la reconstrucción del proceso decisonal, que la autoridad mantenga un control efectivo sobre las decisiones que formalmente adopta y que los mecanismos de supervisión cuenten con herramientas suficientes para evaluar la actuación estatal.

Desde esta perspectiva, la automatización no debe ser entendida como un límite para la responsabilidad política, sino como un contexto que exige su fortalecimiento. Cuanto más complejos sean los procesos decisonales, mayor debe ser la exigencia de claridad en su imputación. La responsabilidad no puede ser sacrificada en nombre de la eficiencia ni desplazada hacia espacios en los que su ejercicio se vuelve impracticable.

En última instancia, los resultados de este trabajo permiten sostener que la relación entre automatización y responsabilidad política no es de incompatibilidad, sino de tensión. La incorporación de sistemas automatizados introduce desafíos que obligan a revisar las categorías tradicionales, pero no elimina la necesidad de mantener el

poder público sometido a mecanismos de control democrático. La cuestión no es si la responsabilidad puede subsistir en contextos de automatización, sino bajo qué condiciones puede seguir siendo efectiva.

## 7. Conclusiones

La incorporación de sistemas automatizados en la toma de decisiones del Estado no constituye un fenómeno meramente técnico, sino una transformación que incide directamente en la forma en que se ejerce el poder público. A lo largo de este trabajo se ha mostrado que dicha transformación no elimina la responsabilidad política, pero sí altera las condiciones bajo las cuales puede ser identificada, exigida y ejercida.

La ruptura de la cadena clásica de imputación, la fragmentación institucional y la opacidad técnica no deben ser entendidas como efectos aislados, sino como manifestaciones de un mismo proceso: la reconfiguración de la estructura decisional del Estado. En este nuevo escenario, la decisión pública deja de ser el resultado de una voluntad claramente localizada y pasa a construirse a partir de una red de actores y mediaciones técnicas que dificultan su reconstrucción en términos tradicionales.

Sin embargo, de ello no se sigue que la responsabilidad política se diluya legítimamente. La decisión automatizada continúa siendo una decisión estatal, y como tal, permanece sujeta a los principios que rigen el ejercicio del poder en un régimen democrático. La mediación tecnológica no introduce neutralidad, ni desplaza la dimensión política de la decisión. Por el contrario, la incorpora en nuevas formas que exigen ser reconocidas y problematizadas.

Aceptar la complejidad técnica como justificación para la opacidad o la dispersión de responsabilidad implicaría debilitar uno de los pilares fundamentales del Estado democrático: la posibilidad de exigir cuentas a quienes ejercen el poder. La responsabilidad política no puede quedar condicionada a la inteligibilidad técnica de los sistemas, ni subordinada a estructuras que impidan identificar con claridad a los sujetos responsables.

En este sentido, el principal desafío no consiste en determinar si es posible mantener la responsabilidad política en contextos de automatización, sino en establecer las condiciones bajo las cuales esta puede seguir siendo efectiva. Ello implica reconocer

que los esquemas tradicionales de imputación resultan insuficientes frente a procesos decisionales distribuidos y mediados por sistemas automatizados.

A partir de lo anterior, se vuelve necesario avanzar hacia una concepción ampliada de la responsabilidad política que incorpore, de manera explícita, las transformaciones introducidas por la automatización. Esta ampliación no supone la sustitución de los principios existentes, sino su adaptación a nuevas condiciones. Entre los elementos que resultan indispensables se encuentran la exigencia de trazabilidad en los procesos decisionales, la obligación de mantener una supervisión humana efectiva y la garantía de que las decisiones puedan ser explicadas en términos comprensibles para el control democrático.

Asimismo, resulta necesario replantear la forma en que se articulan los mecanismos de control institucional. La supervisión parlamentaria, el control administrativo y el escrutinio público deben adaptarse a estructuras decisionales que ya no responden a modelos lineales. Esto implica dotar a las instituciones de herramientas que permitan no solo acceder a la información, sino comprenderla y evaluarla de manera efectiva.

Desde una perspectiva más amplia, la discusión planteada en este trabajo evidencia que la relación entre tecnología y poder no puede ser abordada desde una lógica de sustitución, sino de transformación. La automatización no desplaza la política, sino que la reconfigura en formas que pueden resultar menos visibles, pero no por ello menos relevantes. Ignorar esta transformación no preserva la responsabilidad política, sino que contribuye a su debilitamiento.

En consecuencia, sostener la vigencia de la responsabilidad política en el Estado contemporáneo exige asumir que su ejercicio ya no puede depender exclusivamente de las estructuras tradicionales. La complejidad técnica, lejos de justificar la reducción del control democrático, obliga a reforzarlo. Cuanto más mediada esté la decisión por sistemas automatizados, mayor debe ser la exigencia de claridad, trazabilidad y rendición de cuentas.

En última instancia, la automatización plantea una pregunta que no puede ser eludida: quién responde por el ejercicio del poder cuando la decisión ya no es plenamente humana. La respuesta no puede ser la ausencia de responsabilidad. La única respuesta compatible con un Estado democrático es que la responsabilidad subsiste, pero requiere ser reconstruida en función de las nuevas condiciones en las que el poder se ejerce.

## Referencias

Bovens, M. (2007). Analysing and assessing accountability: A conceptual framework. *European Law Journal*, 13(4), 447–468. <https://doi.org/10.1111/j.1468-0386.2007.00378.x>

Bovens, M., & Zouridis, S. (2002). From street-level to system-level bureaucracies: How information and communication technology is transforming administrative discretion and constitutional control. *Public Administration Review*, 62(2), 174–184. <https://doi.org/10.1111/0033-3352.00168>

Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information, Communication & Society*, 20(1), 14–29. <https://doi.org/10.1080/1369118X.2016.1154087>

O’Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Publishing Group.

Rosanvallon, P. (2008). *Counter-democracy: Politics in an age of distrust*. Cambridge University Press.

# Relaciones internacionales contemporáneas: cooperación multilateral, policrisis y el rol propositivo de América Latina

## *Contemporary International Relations: Multilateral Cooperation, Polycrisis, and Latin America's Proactive Role*

Gastón Gaete Coddou <sup>1</sup>

<sup>1</sup> Académico, Universidad de Playa Ancha, Chile.  
ggaete@upla.cl

Jeffrey Yamil Kafka Delgado <sup>2</sup>

<sup>2</sup> Investigador independiente.  
jeffrykafka@gmail.com

**Resumen.** El presente artículo estudia los desafíos y oportunidades que constituyen las actuales relaciones internacionales desde la representación de la colaboración multilateral. La investigación se desarrolló a través de una configuración cualitativa de alcance descriptivo-analítico, con diseño documental, asentado en el análisis crítico de fuentes académicas indexadas y documentos de organismos multilaterales de referencia. El estudio fundó sus hallazgos sobre la base de tres ejes interpretativos, a saber: los desafíos del entorno global, la transformación teórico-práctica de las relaciones internacionales, y la cooperación multilateral frente a las resistencias del poder. Los resultados develaron que la debilidad sistémica no es solo coyuntural, sino una condición estructural de su estado constitutivo reforzada por una multiplicidad de crisis intercaladas (policrisis) y, por otra parte, que la diplomacia virtual simboliza paralelamente una coyuntura de agilización de la gestión interestatal y una nueva fuente de vulnerabilidades asimétricas. A la par, se discute cómo la propensión al unilateralismo desgasta la confianza en las instituciones multilaterales, y cómo el Sur Global, y América Latina en particular, debido a su creciente peso demográfico, económico y político ha mutado de una posición receptora pasiva hacia una actuación propositiva en el contexto contemporáneo. Se concluye que el multilateralismo es inevitable, pero demanda una necesaria innovación institucional, y que América Latina tiene la ocasión de diseñar nuevas modalidades de cooperación internacional conformes con los principios de equidad, soberanía y desarrollo sostenible.

**Palabras clave:** Relaciones internacionales, cooperación multilateral, orden global, policrisis, digitalización diplomática.

**Abstract.** This article examines the challenges and opportunities of contemporary international relations from the perspective of multilateral cooperation. The research employed a qualitative, descriptive-analytical approach, drawing on documentary evidence and critical analysis of indexed academic sources and documents from leading multilateral organizations. The study grounded its findings in three interpretive frameworks: the challenges of the global environment, the theoretical and practical transformation of international relations, and multilateral cooperation in the face of power struggles. The results revealed that systemic weakness is not merely circumstantial but a structural condition inherent to its constitutive state, reinforced by a multitude of intersecting crises (polycrisis). Furthermore, virtual diplomacy symbolizes both a streamlined interstate management process and a new source of asymmetric vulnerabilities. This paper discusses how the propensity for unilateralism erodes trust in multilateral institutions, and how the Global South, and Latin America in particular, due to its growing demographic, economic, and political weight, has shifted from a passive, receptive position to a proactive one in the contemporary context. It concludes that multilateralism is inevitable, but demands necessary institutional innovation, and that Latin America has the opportunity to design new forms of international cooperation consistent with the principles of equity, sovereignty, and sustainable development.

**Keywords:** International relations, multilateral cooperation, global order, polycrisis, digital diplomacy.

**Historial:** Recibido: 15 de abril de 2026 | Aceptado: 18 de mayo de 2026

**Ética/Financiamiento:** Los autores declaran no tener conflictos de interés. La investigación fue autofinanciada y no recibió financiamiento externo.

Este documento está bajo una licencia  
CC BY-NC-ND 4.0.

**Cómo citar (APA):** Gaete Coddou, G., & Kafka Delgado, J. Y. (2026). Relaciones internacionales contemporáneas: cooperación multilateral, policrisis y el rol propositivo de América Latina. *Conocimiento i Política*, 7(1), 182-197. DOI: 10.64480/cip.2026.n7.05.

## 1. Introducción

En el mundo contemporáneo, la dinámica de las relaciones internacionales se ha consolidado como un mecanismo esencial para el progreso, la estabilidad y la seguridad global. La creciente interdependencia entre naciones - constatada en ámbitos económicos, políticos, culturales y ambientales - demanda la revisión constante de los mecanismos de interrelación y cooperación (Keohane y Nye 2011). Además, la crisis de legitimidad del orden internacional tradicional, generada por la presencia de nuevos actores y por variaciones estructurales en el ejercicio del poder, plantea la necesidad urgente de integrar enfoques renovados y estrategias colaborativas.

En ese marco, resulta indispensable analizar, por una parte, los desafíos que presenta el entorno global actual y, por otra, las oportunidades que se derivan del establecimiento de vínculos internacionales consistentes y equitativos. Stubb (2026) evidenció que el mundo ha experimentado transformaciones profundas, con impactos directos en la calidad democrática y en la gestión de conflictos regionales. Esta condición se ve reforzada por la transformación de los índices de globalización, lo que confirma que los Estados ya no pueden actuar de forma aislada (Gracia Santos González 2025).

En el plano teórico, Helleiner (2024) ha documentado de manera rigurosa que la policrisis de la globalización económica está redefiniendo los equilibrios del poder internacional, lo que demanda marcos analíticos renovados.

Frente a este escenario, el multilateralismo emerge como el marco de referencia central para abordar los problemas globales. En ese sentido, Molina del Pozo (2023) expresó que el multilateralismo actual muestra altibajos del sistema de cooperación internacional, lo que obliga a determinar las directrices que se deben observar para ahondar en las relaciones interestatales y construir un escenario más igualitario y justo, cuyas bases jurídicas incorporen los valores de los derechos humanos (p. 15). Del mismo modo, Guterres (2025) enunció que el multilateralismo es indispensable para afrontar desafíos globales como el cambio climático, la pandemia y la inseguridad alimentaria, pero crece la desconfianza en el sistema multilateral que debilita la cooperación y la capacidad de acción colectiva (párr. 3).

Pese a las tensiones descritas, la colaboración internacional a través de instituciones multilaterales ha demostrado su valor al posibilitar soluciones colectivas y consensuadas para resolver los problemas globales. La adopción del Pacto para el Futuro en

la Cumbre del Futuro de septiembre de 2024 representó el esfuerzo más reciente de la comunidad internacional por revitalizar el sistema multilateral y dotarlo de instrumentos normativos acordes a los desafíos del siglo XXI (ONU 2024). Sin embargo, Araya Gómez (2024) aludió que la polarización política y el escepticismo hacia las instituciones internacionales dificultan dicho entendimiento (párr. 5). A pesar de estas dificultades, existe un consenso creciente sobre la irreversibilidad de la cooperación multilateral. Así lo sostiene el Instituto de Estudios Internacionales de la Universidad de Chile (IEI-U. Chile, 2024) al afirmar que en un mundo cada vez más interconectado, la cooperación multilateral se ha tornado imprescindible para abordar los desafíos globales, como el cambio climático, las migraciones y la inseguridad (párr. 2).

En el concierto latinoamericano, Sanahuja (2022) ha figurado el momento actual como un «interregno» del orden internacional, un período de transición en el que las reglas y los liderazgos del sistema vigente se han erosionado sin que nuevas arquitecturas de gobernanza global se hayan consolidado. D'Angelo (2025) destacó que, en un mundo marcado por la policrisis, América Latina tiene la posibilidad de proponer nuevas formas de cooperación internacional. Esta apreciación converge con la idea de que la región puede constituirse en un actor propositivo (Hirst et al., 2024; Bianculli, 2024). Antecedente que D'Angelo (2025) complementó al exponer que el multilateralismo debe, en los próximos años, aplicar su capacidad de convergencia para imaginar nuevas formas de cooperación internacional, de acuerdo con las necesidades contemporáneas, con reglas claras y la consolidación de instituciones regionales (p. 218).

Esta visión refuerza la necesidad de que América Latina asuma un rol activo en la reconfiguración del orden multilateral (Russell y Tokatlian, 2023).

Ahora bien, entender el multilateralismo únicamente como un dispositivo de cooperación conduce a una lectura incompleta del sistema internacional contemporáneo. Las instituciones multilaterales no operan en un vacío normativo, sino en un entorno atravesado por asimetrías de poder que condicionan tanto la producción de reglas como su aplicación efectiva. A este respecto, el multilateralismo puede ser concebido simultáneamente como un mecanismo de coordinación y como un espacio de estructuración del poder, donde ciertos actores logran proyectar sus intereses mediante marcos institucionales que adquieren legitimidad global.

En cuanto al objetivo de la investigación, este fue analizar los desafíos y oportunidades

que caracterizan las relaciones internacionales contemporáneas, mediante el análisis crítico de los enfoques teórico-prácticos emergentes y las dinámicas de la cooperación multilateral, con el propósito de identificar los factores estructurales que condicionan la gobernanza global y el rol propositivo de América Latina en la reconfiguración del orden mundial.

En relación a los alcances de la investigación, estos comprenden el análisis de la literatura académica y los documentos abarcando tres dimensiones centrales: los desafíos del entorno global, la transformación teórico-práctica de las relaciones internacionales y la cooperación multilateral frente a las tensiones del poder. Su cobertura geográfica es de alcance global, con énfasis en las implicaciones para América Latina, lo que permitió proyectar sus conclusiones hacia entornos regionales comparables del Sur Global.

En lo concerniente a la trascendencia de la investigación, este artículo contribuirá a generar una síntesis crítica y actualizada sobre el estado del multilateralismo en un período de transformación estructural del orden internacional. Del mismo modo, posiciona a América Latina no como objeto de estudio pasivo, sino como actor con capacidad propositiva, lo que amplía el debate académico y orienta la toma de decisiones en política exterior y cooperación internacional propias de la realidad particular del Sur Global.

## **2. Metodología**

### **2.1. Tipo de investigación**

La investigación es de tipo cualitativo, con alcance descriptivo-analítico y un diseño documental. Se orienta hacia el entendimiento e interpretación de sucesos complejos propios del sistema internacional a partir del análisis crítico de fuentes secundarias, sin intervención de variables ni recolección de datos primarios.

### **2.2. Proceso de selección de fuentes de información**

El estudio en comento se circunscribió a fuentes de acceso abierto indexadas en bases de datos académicas reconocidas y organismos multilaterales de referencia, sin incluir

trabajo de campo empírico ni análisis de datos primarios mediante acceso y búsqueda en bases de datos académicas de acceso abierto (SciELO, Dialnet y repositorios institucionales). A la par, se consultaron portales oficiales de organismos multilaterales (ONU, Real Instituto Elcano, IEI-U. Chile) y publicaciones especializadas en relaciones internacionales.

La búsqueda se orientó mediante descriptores en idioma español e inglés: “multilateralismo”, “cooperación internacional”, “geopolítica”, “relaciones internacionales contemporáneas”, “diplomacia digital” y “policrisis”.

### **2.3. Criterios de inclusión y selección de fuentes**

Se incluyeron una variedad de fuentes actualizadas tanto en español como inglés, que abordaron las concepciones analíticas del estudio en cuestión. Se priorizaron artículos científicos arbitrados, ponencias de conferencias académicas acreditadas, informes de organismos internacionales de referencia, obras de referencia teórica consolidada y discursos institucionales relativos al tema en desarrollo.

### **2.4. Otros aspectos metodológicos relevantes**

El análisis del corpus se realizó mediante la técnica de análisis de contenido temático, organizando los hallazgos en torno a tres ejes interpretativos:

- Desafíos del entorno global,
- Transformación teórico-práctica y
- Cooperación frente a las tensiones del poder.

El estudio, no requirió aprobación ética al trabajar exclusivamente con fuentes de dominio público.

### 3. Resultados

#### 3.1. Desafíos del entorno global actual

El análisis del corpus investigativo reveló que el escenario global en la actualidad se exterioriza por variadas crisis interrelacionadas que impactan el equilibrio mundial. Los conflictos geopolíticos, las tensiones entre potencias y la acelerada evolución de las crisis ambientales evidencian la fragilidad del sistema internacional vigente. Se constató que el mundo ha experimentado y soportado innumerables transformaciones de mayor impacto en los últimos años que en las tres décadas precedentes. Esta diagnosis es consistente con el concepto de policrisis - referido a la forma en que los conflictos en diferentes subsistemas globales se entrelazan y se fortalecen recíprocamente - que se revela como consecuencia de la revisión de la literatura especializada considerada en este artículo.

A la par, se verifica una tensión entre las demandas de seguridad de los Estados y los principios del derecho internacional. El análisis de la literatura especializada evidenció que las tensiones comerciales y la desintegración de las cadenas de valor globales, agregan una dimensión adicional a la policrisis de la globalización económica, amplificando la inestabilidad sistémica.

La tabla a continuación, expone distintos desafíos identificados, su impacto y la función esperada de la gestión propia de las relaciones internacionales.

Tabla 1: Desafíos y roles en el entorno global actual

<b>Desafío</b>	<b>Impacto</b>	<b>Rol de las relaciones internacionales</b>
Conflictos geopolíticos	Inestabilidad y tensión regional	Facilitar el diálogo y la resolución pacífica
Crisis ambientales	Cambio climático y desastres naturales	Promover acuerdos internacionales
Tensiones en el orden global	Fragmentación del sistema internacional	Fomentar la cooperación multilateral

*Nota.* Autor: G. Gaete con base en Alfonzo (2026), Stubb (2026), Lawrence et al. (2024), Helleiner (2024) y Centro de Estudios Internacionales UC (2026).

La tabla previa, enuncia una:

Lógica tripartita en la que los desafíos globales forjan un tipo específico de impacto y establecen la figura diferenciada del sistema de relaciones internacionales.

Los conflictos geopolíticos suscitan inestabilidad y tensión regional, lo que sitúa a la diplomacia multilateral en una posición central como mecanismo de contención y resolución pacífica de las desavenencias. La carencia o déficits de canales diplomáticos activos ha conducido a una escalada de tensiones que de otro modo podrían haberse gestionado y solucionado mediante el diálogo y la negociación en pro de arreglos pacíficos de controversias (artículo 33 de Naciones Unidas).

Los problemas ambientales exigen una acción de gestión que trasciende la capacidad individual de los Estados. La cooperación internacional, cuando está sostenida por compromisos vinculantes y mecanismos de seguimiento, puede establecer un accionar efectivo y eficaz ante problemas de alcance a diversas escalas territoriales.

Las tensiones en el orden global reclaman el robustecimiento de los espacios multilaterales. Se evidencia que el declive de las hegemonías tradicionales abre un espacio de mayor autonomía para los países del Sur Global, siempre que sean capaces de articular propuestas colectivas coherentes.

En síntesis, los desafíos del entorno global del presente no pueden abordarse de manera sectorial o aislada, ya que por sus particularidades integradas exigen estrategias sistémicas en las que la participación internacional funcione como denominador común.

### **3.2. Transformación de enfoques en las relaciones internacionales**

Los resultados demuestran una mutación en los enfoques teóricos y prácticos de las relaciones internacionales frente a los modelos tradicionales. La revisión de la literatura evidencia posturas críticas que promueven replantear la visión occidental, promoviendo un enfoque efectivamente global que considere e integre múltiples actores y realidades culturales. Esta perspectiva, se conecta con el argumento de que el interregno del orden internacional requiere no solo nuevas arquitecturas y ecosistemas institucionales, sino también, marcos epistémicos actualizados capaces de dar cuenta de la multipolaridad global emergente.

Paralelamente, la introducción y uso de innovaciones tecnológicas y digitalización en la diplomacia han cambiado la forma en que los Estados negocian, acuerdan y resuelven sus divergencias. La diversidad de fuentes consultadas sistematizó las transformaciones que la revolución digital ha impuesto a los procesos de negociación interestatales, identificando tanto las oportunidades como las vulnerabilidades que introduce la dependencia tecnológica; aun cuando se debe reconocer que la implementación e integración de dispositivos digitales en la gestión de crisis ha facultado soluciones más ágiles y coordinadas, renovando el aparato diplomático.

### **3.3. Cooperación internacional frente a las tensiones del poder**

Esta condición demuestra que la asistencia internacional aparece como instrumento esencial para confrontar los desafíos estructurales entre seguridad y derecho internacional. La experiencia de organismos internacionales propios de la cooperación permite advertir que la colaboración puede constituir el medio para enfrentar retos conjuntos, desde la crisis climática hasta la gestión de pandemias. En este sentido, la adopción del Pacto para el Futuro (2024) representa un esfuerzo concreto por reformar las delineaciones y contenidos de gobernanza global y dotarlas de mayor legitimidad.

En el plano regional, se documenta que América Latina se enfrenta a un sistema internacional en conversión que, si bien debilita el multilateralismo tradicional, habilita simultáneamente espacios para la innovación institucional regional. Este estadio converge con el hallazgo de que el Sur Global dispone de creciente peso demográfico, económico y político para transitar de una posición receptora a una postura propositiva en la reconfiguración del orden multilateral.

En este sentido, se expone la siguiente tabla.

Tabla 2: Resumen de principales hallazgos e implicaciones

Hallazgo clave	Dimensión analítica	Implicación principal
Crisis globales y tensiones en el orden mundial	Desafíos del entorno global	Requieren mecanismos de diálogo y cooperación
Emergencia de enfoques digitales y críticos	Transformación teórico-práctica	Facilitan respuestas ágiles frente a crisis
Tendencia hacia el unilateralismo en temas clave	Cooperación vs. poder	Evidencia la necesidad de fortalecer el multilateralismo

*Nota.* Autores: Gaete y Kafka (2026).

El examen de los contenidos de la tabla precedente pone de manifiesto una articulación relacional entre los hallazgos clave, las dimensiones analíticas que los organizan y los efectos derivados de cada uno de estos hallazgos clave para la política internacional.

La persistencia de crisis globales y tensiones en el orden mundial -categorizadas como policrisis por Lawrence et al. (2024) y Tooze (2022) - apunta que la respuesta estructural es menester para la cimentación y vigorización de mecanismos de interlocución y cooperación. Ningún Estado, independientemente de su grado de capacidad de poder, puede gestionar de forma individual la complejidad de las crisis presentes.

La necesidad emergente de enfoques digitales y críticos en el campo de las relaciones internacionales indica que estos enfoques proveen respuestas más ágiles, adaptadas y coordinadas frente a crisis en tiempo real (Bjola y Manor 2024; Bjola 2023).

La tendencia hacia el unilateralismo en aspectos fundamentales pone de manifiesto un contrasentido en la arquitectura estructural, dado que cuando más esencial resulta la cooperación multilateral, mayor es la instigación de algunos actores de operar al margen de las normas e instituciones compartidas.

El multilateralismo demanda estímulos institucionales sólidos y mecanismos de rendición de cuentas (Keohane y Nye 2011; Sanahuja 2022).

## 4. Discusión

### 4.1. La fragilidad sistémica como condición estructural del orden internacional contemporáneo

Los resultados alcanzados en la primera dimensión analítica confirman y amplían lo planteado por la literatura especializada: la fragilidad del orden internacional moderno no es una particularidad coyuntural, sino una condición estructural que se ha agudizado ante la acumulación sincrónica de diversas crisis. Esta condición es análoga con el diagnóstico de Stubb (2026) y con el análisis de Gracia Santos y González (2025), cuyo índice de presencia global documenta empíricamente el desplazamiento desde un modelo de interdependencia cooperativa hacia uno sometido por la rivalidad geopolítica. En términos teóricos, esta situación encuentra su marco conceptual más preciso en el concepto de policrisis propuesto por Tooze (2022) y operacionalizado por Lawrence et al. (2024).

Lo que distingue el momento actual de incidentes anteriores de tensión internacional es la concurrencia de variadas crisis que se retroalimentan recíprocamente. La policrisis de la globalización económica documentada por Helleiner (2024) añadió una dimensión adicional a este panorama, al exponer cómo las tensiones del orden económico internacional amplifican las vulnerabilidades políticas y de seguridad. Este acontecer justifica la necesidad de marcos interpretativos innovados, tal como proponen Ceccorulli y Agostinis (2026).

Desde este enfoque, la policrisis no solo acrecienta la complejidad del sistema internacional, sino que también incide en la forma en que se distribuyen las capacidades entre los distintos actores involucrados. La interdependencia, lejos de traducirse automáticamente en cooperación, ha generado nuevas formas de vulnerabilidad estratégica, en las que la posición de los Estados dentro de redes globales (financieras, tecnológicas o logísticas), condiciona su margen de acción. En esta situación, las asimetrías estructurales no desaparecen, sino que se reconfiguran en torno al control de dichas interconexiones, dando lugar a mecanismos indirectos de influencia y presión.

#### 4.2. La renovación teórico-práctica como oportunidad y como imperativo

La propuesta de Ceccorulli y Agostinis (2026) de superar la supremacía interpretativa occidental no debe entenderse como un ejercicio de relativismo teórico, sino como una respuesta metodológica necesaria e ineludible a la multipolaridad creciente del sistema internacional. Sanahuja (2022) complementó esta dimensión al señalar que el interregno del orden internacional es también una crisis epistemológica.

En un mundo donde el Sur Global concentra cada vez más poder demográfico, económico y político - tal como documentan Russell y Tokatlian (2023) y Hirst et al. (2024) - mantener marcos analíticos contruidos desde experiencias históricas particulares involucra aceptar un prejuicio estructural frente a dinámicas que determinan de manera creciente el curso de los asuntos globales.

En este tenor, la digitalización diplomática sistematizada por Bjola (2023) y Bjola y Manor (2024) no puede concebirse únicamente como una modernización instrumental de los procesos de negociación. Esta digitalización instaure nuevas debilidades, tales como: la dependencia tecnológica, la brecha digital entre países desarrollados y aquellos en vías de desarrollo, y el uso de infraestructuras digitales como vectores de influencia o desinformación son riesgos que el sistema internacional aún no ha concebido de forma beneficiosa, teniendo en consideración que una modernización tecnológica sin la debida revisión de sus fundamentos normativos corre el riesgo de reproducir, en formato digital, las mismas asimetrías de poder que definieron la diplomacia analógica.

La convergencia entre la crítica teórica y la innovación práctica sugiere que la renovación del sistema internacional requiere operar simultáneamente en dos planos:

1. El epistémico, esto es, revisando los supuestos desde los que se interpreta el mundo, y
2. El operativo, que en sí induce a la transformación de los instrumentos con los que se gestiona.

### **4.3. La cooperación multilateral frente a la lógica de la interdependencia armada**

El tercer eje de discusión revela el dilema central del sistema internacional contemporáneo, en el cual, la interdependencia entre los Estados ha alcanzado niveles inéditos, pero esta misma interdependencia ha sido convertida, en numerosos casos, en un instrumento de coerción y dominación. Esta situación, que puede denominarse «interdependencia armada», fue predicha por Keohane y Nye (2011) al explicar que la interdependencia compleja genera tanto vulnerabilidades como sensibilidades entre los actores del sistema.

Si los Estados perciben que las instituciones internacionales son utilizadas selectivamente por las potencias dominantes para legitimar sus propios intereses, la confianza en dichas instituciones se erosiona. Sanahuja (2022) y Russell y Tokatlian (2023) coincidieron en que esta percepción de desigualdad explica el distanciamiento de varios países del Sur Global respecto a las instituciones multilaterales.

El Pacto para el Futuro adoptado en 2024 (ONU 2024) representa un intento de reformar estas arquitecturas desde dentro, aunque Bianculli (2024) advirtió que la superación de estas transformaciones estará subordinada en gran medida de la capacidad de las regiones emergentes de articular posiciones colectivas coherentes en los espacios multilaterales.

## **5. Conclusiones**

### **5.1. Conclusión 1: La fragilidad sistémica como condición estructural, no coyuntural**

El orden internacional contemporáneo está inmerso en una crisis estructural de inestabilidad que se ha acentuado por un cúmulo de conflictos simultáneos e interrelacionados, fenómeno que la literatura especializada ha conceptualizado como policrisis (Lawrence et al. 2024; Tooze 2022). El menoscabo democrático y la desintegración del sistema global se retroalimentan mutuamente, organizando un contexto en el que los marcos teóricos tradicionales no han sido acertados para entender y prever estas circunstancias, escenario que hace ineludible renovar los instrumentos analíticos y las arquitecturas institucionales del sistema internacional.

## **5.2. Conclusión 2: La digitalización diplomática es una oportunidad y una vulnerabilidad simultánea**

La innovación tecnológica ha cambiado los procesos de negociación, gestión de crisis y coordinación interestatal, tal como sistematizó la investigación de Bjola y Manor (2024). Aunque, esta condición debe ser examinada normativamente de manera exhaustiva dado que de no ser así se podrían generar desequilibrios de poder que han modelado a la diplomacia convencional.

## **5.3. Conclusión 3: El multilateralismo es imprescindible, pero requiere reforma estructural para sobrevivir**

La experiencia recolectada en organismos internacionales señala que el multilateralismo tiene una capacidad de resiliencia institucional que no debe despreciarse, como lo evidencia la adopción del Pacto para el Futuro (ONU 2024). Sus restricciones - parálisis por veto, déficits de representación y lentitud procedimental - deterioran su legitimidad instrumental y de aplicación. La reforma institucional es, en este sentido, una condición de supervivencia del multilateralismo (Keohane y Nye 2011; Sanahuja 2022).

## **5.4. Conclusión 4: El unilateralismo es racionalmente comprensible, pero estructuralmente destructivo**

La propensión al unilateralismo no irrumpe únicamente por voluntad hegemónica, sino como respuesta de representantes que observan que las reglas del sistema internacional se aplican de forma selectiva y disímil, realidad documentada por Russell y Tokatlian (2023). Esto no justifica el unilateralismo, pero sí demanda reconocer que el multilateralismo no puede mantenerse sólo sobre bases voluntaristas.

## **5.5. Conclusión 5: América Latina debe transitar de receptora pasiva a actora propositiva en el orden multilateral**

La región latinoamericana tiene la posibilidad y el compromiso de disponer de otras formas de colaboración internacional que sean coherentes con los principios de equi-

dad, soberanía y desarrollo sostenible (D'Angelo 2025; Hirst et al. 2024; Bianculli 2024). Esta variación de la posición receptora a la propositiva demanda, como argumentó Sanahuja (2022), tanto renovación epistémica como articulación política regional.

### **5.6. Conclusión 6: Ningún desafío global puede gestionarse de forma sectorial o unilateral**

La condición interconectada de los desafíos contemporáneos impone una perspectiva sistémica como única respuesta viable, tal como establece el concepto de policrisis (Lawrence et al. 2024). Ningún Estado puede gestionar individualmente la complejidad de las crisis presentes (Keohane y Nye 2011). La integración supranacional es la plataforma estructural necesaria sobre la que deben asentarse las acciones comunes al entorno conflictuado de la modernidad.

## **Referencias**

Alfonzo, R. (2026). Geopolítica contemporánea y diplomacia digital: Nuevos instrumentos para la gestión de crisis internacionales. Institución editora.

Araya Gómez, P. (2024, 10 de enero). El multilateralismo: Una responsabilidad de la política exterior. Centro de Estudios Internacionales, Pontificia Universidad Católica de Chile. <https://centroestudiosinternacionales.uc.cl/opinion/5675-elmultilateralismounaresponsabilidad-de-la-politica-exterior>

Bianculli, A. C. (2024). América Latina en el nuevo escenario internacional: ¿Qué espacio hay para el regionalismo y la cooperación regional? *Revista CIDOB d'Afers Internacionals*, 136, 89–110. <https://doi.org/10.24241/rcai.2024.136.1.89>

Bjola, C. (Ed.). (2023). *Digital international relations: Technology, agency and order*. Routledge.

Bjola, C., & Manor, I. (Eds.). (2024). *The Oxford handbook of digital diplomacy*. Oxford University Press.

Ceccorulli, M., & Agostinis, G. (2026). La renovación teórico-práctica de las relaciones internacionales: Superar la supremacía interpretativa occidental. Artículo inédito.

Centro de Estudios Internacionales UC. (2026). Informe anual de relaciones internacionales y política exterior. Pontificia Universidad Católica de Chile.

D'Angelo, E. (2025). Multilateralismo en crisis: Desafíos y oportunidades para alcanzar el desarrollo sostenible. Respuestas elaboradas desde América Latina en clave de género. *Universitas: Revista de Filosofía, Derecho y Política*, (47), 204–250. <https://dialnet.unirioja.es/servlet/articulo?codigo=10251675>

Diálogo Político. (2026, 17 de junio). Cooperación internacional y desafíos actuales: Una mirada desde América Latina [Informe]. <https://dialogopolitico.org>

Escuela de Gobierno y Economía. (2025). Gobernanza global y cooperación multilateral frente a la crisis climática y las pandemias. Publicación institucional.

Gracia Santos González, A. (2025). Índice de presencia global y transformaciones de la interdependencia internacional. Documento de trabajo, Real Instituto Elcano.

Guterres, A. (2025). Discurso sobre la importancia y la crisis del sistema multilateral [Discurso institucional]. Organización de las Naciones Unidas. <https://www.un.org>

Helleiner, E. (2024). Economic globalization's polycrisis. *International Studies Quarterly*, 68(2), sqae024. <https://doi.org/10.1093/isq/sqae024>

Hirst, M., Russell, R., Sanjuan, A. M., & Tokatlian, J. G. (2024). América Latina y el Sur Global en tiempos sin hegemonías. *Revista CIDOB d'Afers Internacionals*, 136, 133–156. <https://doi.org/10.24241/rcai.2024.136.1.133>

IEI-UChile. (2024). La cooperación multilateral se ha vuelto indispensable para abordar los desafíos globales. Instituto de Estudios Internacionales, Universidad de Chile. <https://www.iei.uchile.cl>

Keohane, R. O., & Nye, J. S. (2011). *Power and interdependence: World politics in transition* (4.ª ed.). Longman. (Obra original publicada en 1977)

Lawrence, M., Homer-Dixon, T., Janzwood, S., Rockström, J., Renn, O., & Donges, J. F. (2024). Global polycrisis: The causal mechanisms of crisis entanglement. *Global Sustainability*, 7, e6. <https://doi.org/10.1017/sus.2024.1>

Molina del Pozo, C. F. (Ed.). (2023). *El multilateralismo en el mundo actual: Alcances y perspectivas*. Colex. <https://dialnet.unirioja.es/servlet/libro?codigo=923531>

ONU. (2024). Pacto para el Futuro. Organización de las Naciones Unidas. <https://www.un.org/es/summit-of-the-future>

Russell, R., & Tokatlian, J. G. (2023). The decline of US hegemony and Latin America's strategic options. *International Affairs*, 99(2), 563–581. <https://doi.org/10.1093/ia/iiad012>

Sanahuja, J. A. (2022). Interregno: La actualidad de un orden mundial en crisis. *Nueva Sociedad*, (302), 86–94. <https://www.nuso.org>

Sanahuja, J. A., & Stefanoni, P. (Eds.). (2023). América Latina en el interregno: Política, economía e inserción internacional. Informe anual 2023-2024. Fundación Carolina.

Stubb, A. (2026). El mundo ha cambiado más en los últimos cuatro años que en los treinta anteriores: Transformaciones globales y calidad democrática. Discurso institucional.

Tooze, A. (2022, 28 de octubre). Welcome to the world of the polycrisis. *Financial Times*. <https://www.ft.com/content/498398e7-11b1-4b0e-8176-2f3ee0b73e96>

## Seguridad, soberanía y tecnología en la frontera Brasil–Bolivia: asimetrías y cooperación en el período 2018–2024

### *Security, Sovereignty, and Technology along the Brazil–Bolivia Border: Asymmetries and Cooperation in the 2018–2024 Period*

María del Pilar Serrate Roca<sup>1</sup>

<sup>1</sup> Universidade Federal da Grande Dourados (UFGD).  
pilar.serrate@hotmail.com (Autora de correspondencia)

Giovanna Savazo<sup>2</sup>

<sup>2</sup> Universidade Federal da Grande Dourados (UFGD).  
gi.savazo@gmail.com

Claudio Vitor Cardoso<sup>3</sup>

<sup>3</sup> Universidade Federal da Grande Dourados (UFGD).  
claudiovitorcardoso@gmail.com

**Resumen.** El artículo analiza la seguridad en la frontera Brasil–Bolivia (2018–2024), examinando la intersección entre soberanía, tecnificación del control territorial y cooperación bilateral a partir de la implementación del Sistema Integrado de Monitoreo de Fronteras (SISFRON). Bajo el enfoque teórico de la interdependencia compleja y las redes de políticas públicas, y mediante una metodología cualitativa de alcance descriptivo-explicativo, se sostiene que la seguridad fronteriza opera bajo una lógica de interdependencia asimétrica. Los hallazgos demuestran que mientras la incorporación de tecnologías de sensoramiento y apoyo a la decisión fortalece la presencia estatal de Brasil y proyecta su soberanía digital, simultáneamente profundiza la brecha operativa y de gobernanza con Bolivia. Se concluye que la cooperación binacional es un proceso dinámico y estructuralmente desigual, condicionado tanto por la volatilidad institucional y restricciones macroeconómicas locales como por la alternancia política de las administraciones de turno.

**Palabras clave:** Seguridad fronteriza; soberanía; cooperación internacional; tecnología; SISFRON; asimetría

**Abstract.** This article analyzes security along the Brazil–Bolivia border (2018–2024), examining the intersection between sovereignty, territorial control technification, and bilateral cooperation based on the implementation of the Integrated Border Monitoring System (SISFRON). Drawing on the theoretical framework of complex interdependence and policy networks, and utilizing a qualitative, descriptive-explanatory methodology, the study argues that border security operates under a logic of asymmetric interdependence. The findings demonstrate that while the incorporation of sensing and decision-support technologies strengthens Brazil's state presence and projects its digital sovereignty, it simultaneously widens the operational and governance gap with Bolivia. The article concludes that bilateral cooperation is a dynamic and structurally unequal process, conditioned by local institutional volatility and macroeconomic constraints, as well as the political alternation of successive governments.

**Keywords:** Border security; sovereignty; international cooperation; technology; SISFRON; asymmetry

**Historial:** Recibido: 30 de marzo de 2026 | Aceptado: 19 de mayo de 2026

**Ética/Financiamiento:** Las autoras y el autor declaran no tener conflictos de interés. La investigación fue autofinanciada y no recibió financiamiento externo.

**Cómo citar (APA):** Serrate Roca, M. P., Savazo, G., & Cardoso, C. V. (2026). Seguridad, soberanía y tecnología en la frontera Brasil–Bolivia: asimetrías y cooperación en el período 2018–2024. *Conocimiento i Política*, 7(1), 198–222. DOI: 10.64480/cip.2026.n7.06.

Este documento está bajo una licencia  
CC BY-NC-ND 4.0.

## 1. Introducción

Las regiones fronterizas han ganado un protagonismo creciente en los debates académicos contemporáneos al revelar dinámicas de circulación, fricción y gobernanza que desafían las concepciones westfalianas tradicionales de territorio y soberanía (Anderson & O'Dowd, 1999). En el escenario internacional actual, la seguridad en estas áreas no puede reducirse a la contención militar unilateral; por el contrario, se configura como un fenómeno relacional complejo ante el incremento de los flujos transnacionales de economías ilícitas y criminalidad organizada (Buzan, Wæver & de Wilde, 1998). La frontera entre Brasil y Bolivia ejemplifica con precisión este cambio de paradigma. Con una extensión de más de 3.400 kilómetros caracterizados por una alta heterogeneidad geográfica, esta franja enfrenta el desafío crónico de conciliar las prerrogativas de control territorial estatal con la porosidad inherente de sus interacciones cotidianas y transfronterizas.

Durante el septenio comprendido entre 2018 y 2024, la seguridad en esta frontera compartida experimentó profundas reconfiguraciones impulsadas por dos vectores concurrentes: la acelerada tecnificación de la vigilancia fronteriza brasileña y la marcada alternancia política en los Poderes Ejecutivos de ambos Estados. Este período revela de manera nítida cómo las variables político-institucionales y las capacidades materiales alteran la gestión de las vulnerabilidades mutuas, instaurando marcadas asimetrías operativas que condicionan tanto la autonomía estratégica de las agencias locales como el alcance de los compromisos bilaterales.

Ante este escenario, el presente artículo tiene como objetivo analizar cómo se construye la seguridad en la frontera Brasil–Bolivia a partir de la interacción entre soberanía, tecnología y cooperación bilateral durante el período 2018–2024. Para abordar la complejidad de este objeto de estudio, la investigación adopta como matriz relacional y conceptual principal la teoría de la interdependencia compleja (Keohane & Nye, 2012), la cual permite descentralizar el análisis de la fuerza militar y focalizarlo en las dinámicas de sensibilidad y vulnerabilidad mutua, las redes institucionales y las brechas de capacidades técnicas en entornos periféricos.

El artículo se estructura en tres capítulos complementarios. El primero desarrolla el andamiaje teórico-conceptual, movilizándolo las nociones de soberanía contemporánea, soberanía digital e interdependencia compleja, además de ofrecer una contextuali-

zación geoestratégica del Arco Central. El segundo capítulo examina el impacto del Sistema Integrado de Monitoreo de Fronteras (SISFRON) como un multiplicador de fuerzas de la presencia estatal brasileña, analizando la transición hacia una vigilancia proactiva y la consecuente emergencia de una asimetría tecnológica y de información respecto a Bolivia. El tercer capítulo aborda las dinámicas políticas y los flujos de inteligencia operativa desde el enfoque de las redes de políticas públicas, identificando el impacto de las transiciones de gobierno, el quiebre institucional boliviano de 2019 y las restricciones macroeconómicas sobre la estabilidad de la cooperación bilateral.

Al articular estos niveles de análisis, el estudio parte de la hipótesis de que la gestión de la seguridad fronteriza en el Arco Central no se explica de manera aislada por las decisiones unilaterales de los Estados ni por la mera acumulación de dispositivos técnicos. Se sostiene, en cambio, que es un proceso dinámico en el cual la persistente asimetría material y las variaciones en la confiabilidad política de los gobiernos condicionan y jerarquizan los términos de la gobernanza binacional, transformando la disparidad tecnológica en un factor de ordenamiento político del espacio fronterizo.

## 2. Metodología

El estudio adopta un enfoque cualitativo, de alcance descriptivo y explicativo, estructurado mediante un diseño no experimental y transaccional apoyado en la estrategia del estudio de caso aplicado a la frontera Brasil-Bolivia durante el período 2018-2024. Esta delimitación espacial y temporal permite analizar empíricamente la interacción entre soberanía, tecnificación y cooperación en un escenario periférico caracterizado por una marcada interdependencia asimétrica.

La recolección de datos se basó en una revisión bibliográfica y documental exhaustiva que conformó un corpus de dos tipologías de fuentes: por un lado, documentos oficiales, normativos y presupuestarios de defensa de ambos Estados (como decretos del SISFRON y reportes de inversión en ciberdefensa); y por otro, informes especializados de organismos multilaterales como la UNODC, la CEPAL y el International Crisis Group, complementados con literatura académica indexada.

Para el procesamiento y examen de la información se utilizó la técnica de análisis de contenido cualitativo y relacional, operacionalizando el marco teórico en tres categorías analíticas básicas: la tecnificación del control territorial y soberanía digital,

la asimetría operativa e informacional, y las dinámicas de gobernanza en las redes de cooperación bilateral. Adicionalmente, se incorporó un análisis temporal secuencial que divide el septenio en tres fases nítidas (2018–2019, 2019–2022 y 2023–2024) para identificar con precisión los puntos de inflexión institucional y el impacto de los cambios políticos en la densidad de la red binacional.

### 3. Marco Teórico

Para analizar la seguridad en la frontera Brasil-Bolivia, este estudio adopta como eje relacional e interpretativo principal la teoría de la interdependencia compleja (Keohane & Nye, 2012). A partir de esta matriz horizontal, se organizan de forma complementaria las variables de soberanía, asimetría tecnológica, presencia estatal y gobernanza fronteriza. Este ordenamiento jerárquico permite comprender que los desafíos fronterizos contemporáneos no responden a lógicas unilaterales de exclusión, sino a redes dinámicas de interacción multiinstitucional donde la disparidad material redefine las condiciones de la cooperación binacional.

#### 3.1. Concepto de soberanía e interdependencia compleja

El sistema de Westfalia, resultado de los tratados que pusieron fin a la Guerra de los Treinta Años en 1648, estableció las bases del Estado-nación moderno y la idea de autoridad estatal sobre su territorio. El concepto clásico de soberanía, originado en este sistema, define al Estado como poseedor de autoridad plena sobre su territorio y población, con monopolio legítimo de la fuerza y la prerrogativa de no sufrir interferencias externas (Anderson & O'Dowd, 1999). Sin embargo, esta concepción presenta limitaciones cuando se enfrenta a desafíos que trascienden fronteras físicas y legales, como el tráfico de drogas, el contrabando y la movilidad irregular de personas (Newman & Paasi, 1998).

La superación de las limitaciones inherentes al modelo westfaliano frente a estas dinámicas transnacionales encuentra un sustrato teórico en las perspectivas orientadas al análisis del orden interestatal y la evolución de la autoridad pública. Desde la perspectiva de Immanuel Kant (1795), se postula que los Estados no consolidan su condición política mediante el aislamiento autárquico —el cual es equiparado a un estado de naturaleza propenso al conflicto—, sino a través de su inserción deliberada en

un ordenamiento jurídico internacional o federación pacífica (*foedus pacificum*). Bajo este enfoque, el ejercicio de la soberanía estatal no se erosiona mediante la vinculación externa; por el contrario, se legitima institucionalmente al sustituir la desconfianza unilateral por marcos de cooperación mutua orientados a mitigar riesgos comunes.

En continuidad con esta premisa relacional, aportes analíticos como los de Anne-Marie Slaughter (2004) desplazan el núcleo definitorio de la soberanía desde el tradicional "derecho a la exclusión" hacia la capacidad de conectividad (*sovereignty as connectivity*). Slaughter sostiene que, en contextos de interdependencia, la eficacia de la autoridad estatal está supeditada a su aptitud para integrarse en redes transgubernamentales, coordinar respuestas operativas y gestionar de forma compartida los flujos de información estratégica. De este modo, la convergencia entre la tesis kantiana y las aproximaciones basadas en la conectividad permite fundamentar que las fronteras modernas funcionan simultáneamente como instrumentos de control estatal y como espacios de interacción multilateral, en los cuales la soberanía se manifiesta de forma negociada y compartida. En este contexto, la soberanía deja de ser un atributo absoluto y pasa a ser un concepto dinámico que combina la preservación de la autoridad nacional con la participación en acuerdos, operaciones conjuntas e intercambio de información que fortalecen el poder estatal al mismo tiempo que promueven la cooperación entre países (Anderson & O'Dowd, 1999; Van Houtum, 2005).

Esta reconfiguración de la autoridad estatal en el plano físico se extiende, en el escenario contemporáneo, hacia la dimensión tecnológica mediante el concepto de soberanía digital. De acuerdo con aproximaciones teóricas como las de Pohle y Thiel (2020), la soberanía digital no se restringe al control regulatorio del ciberespacio, sino que se define de manera multidimensional como la capacidad de un Estado para ejercer autonomía estratégica a través del desarrollo de capacidades tecnológicas propias, el control soberano sobre los flujos de datos estratégicos y el despliegue de infraestructuras estatales de vigilancia. En este sentido, la soberanía digital opera como una extensión de la soberanía territorial, donde la reducción de la dependencia externa respecto a proveedores e infraestructuras extranjeras resulta crítica para salvaguardar la seguridad nacional y garantizar la toma de decisiones operativas independientes. Por consiguiente, la adopción de tecnologías avanzadas de monitoreo en las periferias estatales no constituye un mero recurso técnico, sino una manifestación material de esta soberanía en su vertiente digital.

Esta articulación entre autonomía y cooperación se sustenta analíticamente en la teoría de la interdependencia compleja (Keohane & Nye, 2012). Este marco conceptual postula que las relaciones interestatales contemporáneas están definidas por tres condiciones estructurales: la existencia de canales múltiples de contacto (que integran vínculos intergubernamentales, transgubernamentales y transnacionales), la ausencia de una jerarquía rígida en la agenda internacional —donde la seguridad militar tradicional coexiste y se intersecta de forma horizontal con variables económicas, tecnológicas e institucionales— y la pérdida de centralidad de la fuerza militar como instrumento de resolución de controversias. En el ámbito de los espacios fronterizos, la interdependencia compleja implica que la seguridad deja de operar bajo una lógica de exclusión unilateral para configurarse a partir de la sensibilidad y vulnerabilidad mutua de los Estados. Bajo este enfoque, la gestión territorial se descentraliza hacia redes de interacción multiinstitucionales que influyen directamente en el diseño de las políticas públicas locales, transformando la frontera en un espacio donde la preservación de la autoridad nacional requiere necesariamente la construcción de arreglos cooperativos estables.

A partir de esta matriz relacional, la literatura advierte que las condiciones de interdependencia rara vez se presentan de forma simétrica, lo que introduce la necesidad de operativizar la categoría de asimetría tecnológica (Keohane & Nye, 2012). En el ámbito de las políticas públicas y la seguridad internacional, esta asimetría se define como la distribución desigual de capacidades técnicas, infraestructuras críticas y recursos de procesamiento de datos entre los Estados. Lejos de constituir una mera disparidad cuantitativa en la adquisición de equipamiento, la asimetría tecnológica opera como una variable estructural que redefine los términos de la cooperación bilateral. Siguiendo aproximaciones sobre brechas de innovación (Fagerberg, 1987), este desequilibrio institucional genera una asimetría informacional donde el Estado que posee sistemas avanzados de monitoreo y análisis de datos adquiere una ventaja táctica y una conciencia situacional superior, condicionando la agenda operativa y situando al actor con menores capacidades en una posición de dependencia funcional respecto a la información compartida.

En este sentido, la frontera se configura no solo como un límite de demarcación territorial, sino como un espacio preeminente para el despliegue de políticas de control y seguridad. Los dispositivos de vigilancia tienden a expandirse de manera intensiva en estas áreas geográficas sensibles, donde las prácticas de control, supervisión y

excepción adquieren una legitimación institucional sistemática (Bigo, 2002; Foucault, 2008). Por consiguiente, estos entornos desempeñan una función estratégica en la administración territorial, constituyendo los espacios periféricos donde se materializa de forma concreta la presencia estatal (Newman & Paasi, 1998; Van Houtum, 2005). Esta presencia estatal se delimita conceptualmente como la aptitud institucional, normativa y material de un Estado para proyectar su autoridad, regular los flujos transfronterizos y asentar infraestructuras de gobernanza en sus márgenes geográficos, permitiendo la transición desde un control territorial intermitente hacia un monitoreo continuo del territorio.

Bajo esta lógica, las problemáticas de seguridad fronteriza se distancian de las concepciones tradicionales de amenaza militar externa para centrarse en fenómenos transnacionales complejos, como el crimen organizado y el narcotráfico, cuya contención excede las capacidades de control unilateral de los Estados. Frente a estas dinámicas, la cooperación bilateral se operacionaliza como un proceso de coordinación estratégica y alineamiento de políticas públicas entre administraciones soberanas, dirigido a la gobernanza compartida de vulnerabilidades e interdependencias comunes. Lejos de implicar una merma en la autonomía nacional, esta articulación interinstitucional —fundamentada en el intercambio de información, la compatibilidad técnica y el desarrollo de operaciones conjuntas— funciona como un mecanismo estratégico para optimizar las capacidades defensivas frente a redes delictivas que operan con altos niveles de flexibilidad y adaptación.

### **3.2. Contextualización de la frontera Brasil-Bolivia**

En la frontera Brasil-Bolivia, esta lógica se evidencia en la necesidad de conciliar el control territorial con la cooperación para enfrentar amenazas como el tráfico ilícito, la actuación de grupos armados y los movimientos irregulares de personas (Newman; Paasi, 1998). Esta frontera, con aproximadamente 3.400 kilómetros de extensión, presenta una notable diversidad geográfica y social, abarcando áreas amazónicas, regiones rurales y centros urbanos, lo que contribuye a la complejidad de su gestión.

Se trata de una región caracterizada por contrastes significativos, como la baja densidad poblacional en extensos tramos y la intensa urbanización en ciudades gemelas, como Corumbá y Puerto Quijarro o Corumbá y Puerto Suárez. Estas configuraciones territoriales favorecen tanto dinámicas de integración económica y social como la am-

pliación de oportunidades para prácticas ilegales, en función de la intensa circulación transfronteriza y de las limitaciones en el control estatal continuo.

Asimismo, la extensión territorial, las dificultades logísticas, la presencia de biomas como el Pantanal y la existencia de áreas de difícil acceso contribuyen a la porosidad de la frontera, lo que dificulta la vigilancia permanente por parte del Estado y facilita la adaptación de redes ilícitas a los mecanismos de control existentes (Anderson; O'Dowd, 1999; Van Houtum, 2005).

Esta franja fronteriza forma parte del denominado Arco Central brasileño, considerado estratégico tanto desde el punto de vista de la defensa como de la seguridad pública, especialmente por concentrar rutas relevantes para el tráfico de drogas, el contrabando y otras economías ilícitas en América del Sur. En este contexto, la región desempeña un papel central en las dinámicas del narcotráfico, siendo Bolivia una de las principales zonas de tránsito de cocaína destinada al mercado brasileño e internacional, lo que intensifica la presión sobre los mecanismos de control y cooperación entre ambos países (UNODC, 2022).

Es así como, la frontera es entendida como un territorio de interacción cotidiana, donde circulan bienes, personas y prácticas sociales que forman parte de la vida económica y cultural local. Esta dualidad entre integración y control refuerza la necesidad de políticas públicas que articulen seguridad y desarrollo, evidenciando que la gestión fronteriza implica no solo acciones represivas, sino también estrategias coordinadas que consideren las especificidades sociales y territoriales de la región.

#### **4. El SISFRON y la Tecnificación del Control**

El presente capítulo analiza la tecnología como eje estructurante de la seguridad en la frontera entre Brasil y Bolivia, focalizando el estudio en la implementación del Sistema Integrado de Monitoreo de Fronteras (SISFRON), cuyo fin es transformar la geografía del Arco Central —específicamente en los estados de Mato Grosso y Mato Grosso del Sur— en un espacio bajo control digital. En este contexto, el SISFRON opera como un multiplicador de fuerza orientado a reducir la porosidad fronteriza y fortalecer la custodia del territorio nacional frente a desafíos transnacionales mediante una vigilancia proactiva y continua (Oliveira & Farias, 2024; Santos & Santos Filho, 2023).

La arquitectura del sistema se fundamenta técnicamente en el paradigma de Comando, Control, Comunicaciones, Computación, Inteligencia, Vigilancia y Reconocimiento. Este marco operacional integra flujos de información heterogéneos y mitiga la fragmentación institucional histórica, facilitando una "conciencia situacional" que trasciende los límites físicos del terreno. El SISFRON profundiza la modernización de la fuerza terrestre y fortalece la defensa integrada en operaciones de no guerra, a través de un complejo sistema informatizado.

Asimismo, el despliegue tecnológico en el Arco Central se vincula con la búsqueda de soberanía digital y autonomía estratégica. Al priorizar el desarrollo dentro de la Base Industrial de Defensa (BID) e integrar activos espaciales como el Satélite Geoestacionario de Defensa y Comunicaciones Estratégicas (SGDC) —infraestructura que garantiza la independencia en la transmisión de datos cifrados y la conectividad en áreas de difícil acceso—, Brasil busca reducir la vulnerabilidad externa frente a proveedores extranjeros. Esta capacidad de monitoreo resulta vital en regiones como el Pantanal, donde la combinación de drones y radares, soportada por la red satelital, permite respuestas tácticas precisas ante el tráfico ilícito. Así, se consolida un modelo de control que equilibra la defensa nacional con la seguridad pública mediante una infraestructura tecnológica de propiedad estatal (Oliveira, 2023).

#### **4.1. El SISFRON: Marco operacional y estrategia de presencia**

El Sistema Integrado de Monitoreo de Fronteras (SISFRON) constituye la manifestación material de una política pública de defensa dirigida a revertir la histórica ausencia del Estado de Brasil en las periferias nacionales. Fundamentado en la Estrategia Nacional de Defensa, el sistema aplica la "Estrategia de la Presencia", la cual postula que la soberanía se ejerce mediante la vigilancia continua y la capacidad de respuesta inmediata en puntos sensibles del territorio. El SISFRON integra la línea fronteriza de 16.886 kilómetros que tiene Brasil bajo un comando centralizado, permitiendo procesar información en tiempo real en la interacción con Bolivia para la toma de decisiones estratégicas.

Su arquitectura de "sistema de sistemas" permite que diversos órganos de seguridad pública y las Fuerzas Armadas operen de forma interoperable, superando la fragmentación institucional tradicional en América del Sur. El SISFRON se conceptualiza como una estructura orientada a dotar al Estado de medios para una presencia efecti-

va en la faja de frontera, apoyada en subsistemas de sensoramiento óptico, vigilancia terrestre, vigilancia aérea e inteligencia de señales (Santos & Santos Filho, 2023)

En los estados Brasileños de Mato Grosso y Mato Grosso del Sur, esta capacidad de sensoramiento resulta crítica debido a la extrema porosidad de la frontera con los departamentos bolivianos de Santa Cruz y Beni, zonas caracterizadas por una geografía compleja que dificulta el control físico tradicional. Y es así como Brasil, para abordar este desafío, implementó un proyecto piloto en la ciudad de Dourados (Mato Grosso del Sur), cuya ejecución permitió validar la efectividad de la tecnología como un 'multiplicador de fuerzas'. Los resultados de este piloto demostraron que el uso de radares de vigilancia terrestre, como el modelo SENTIR M20, y sistemas optrónicos (cámaras de largo alcance) permitían detectar movimientos ilícitos a kilómetros de distancia, incluso en condiciones de visibilidad nula.

Este éxito técnico en esta ciudad situada en la faja de frontera brasileña proporcionó una conciencia situacional sin precedentes, es decir, la capacidad de los mandos militares para visualizar el escenario fronterizo completo en tiempo real desde un centro de control, sin necesidad de patrullas constantes en cada punto. De este modo, Brasil deja demostrado que la tecnificación podría transformar las fronteras en un territorio digitalmente monitoreado donde el flujo de datos no es solo información, sino que actúa como la primera línea de defensa de la soberanía nacional. Al emplear software de desarrollo nacional y redes seguras como el satélite SGDC, Brasil no solo vigila su territorio, sino que fortalece su soberanía digital al eliminar la dependencia estratégica de infraestructuras y sistemas extranjeros (Oliveira, 2023).

#### **4.2. El Aparato Tecnológico en el Arco Central**

La implementación tecnológica en el Arco Central de Brasil conforma el núcleo operativo del SISFRON bajo la jurisdicción de la 4ª Brigada de Caballería Mecanizada. El proyecto de Sensoramiento articula radares y sensores para mitigar la porosidad fronteriza con Bolivia y Paraguay. La tecnología permite transformar la geografía física en un espacio de datos, otorgando al Ejército y al Departamento de Operaciones de Frontera (DOF) una ventaja táctica basada en información. En este contexto, la tecnología actúa como un multiplicador de fuerza que compensa la baja densidad demográfica militar, garantizar una vigilancia ininterrumpida y una capacidad de respuesta inmediata ante incursiones ilícitas (Espósito Neto & Franchi, 2021; Paixão,

2022; Santos & Santos Filho, 2023).

El radar de vigilancia terrestre constituye el componente central del sensoramiento, proporcionando alerta temprana al detectar movimientos en el terreno y bajas altitudes. El radar SENTIR M20, por ejemplo, detecta objetivos móviles hasta a 30 km para vehículos y 10 km para personal de a pie. Esta infraestructura se complementa con cámaras infrarrojas y sensores optrónicos de largo alcance. En municipios como Corumbá y Ponta Porã (Brasil), torres de comunicaciones permiten el envío de datos en tiempo real a centros de comando, facilitando la interoperabilidad entre la Policía Militar y fuerzas federales. Este flujo de datos es vital para la interoperabilidad entre agencias, ya que permite que la Policía Militar y las fuerzas federales coordinen sus actuaciones basadas en la misma conciencia situacional provista por el SISFRON (Espósito Neto & Franchi, 2021; Magdalena, 2024; Silva, 2025). Este entramado tecnológico demuestra que la vigilancia digital no sustituye la presencia física del Estado en la frontera, sino que opera bajo una lógica de complementariedad táctica y operativa, actuando como un multiplicador de fuerzas sobre el terreno.

La operatividad del SISFRON además se consolida a través del Subsistema de Apoyo a la Decisión (SAD). Este componente actúa como el núcleo integrador del sistema, cuya función principal es transformar el flujo masivo de datos proveniente de los sensores en inteligencia accionable. El SAD procesa la información en centros de comando y control, permitiendo que las autoridades identifiquen amenazas en tiempo real y coordinen respuestas tácticas de manera eficiente. En términos de gestión pública, el SAD optimiza el uso de recursos humanos y materiales, asegurando que la intervención estatal en la frontera sea precisa (Santos & Santos Filho, 2023).

La integración de Sistemas de Aeronaves Remotamente Pilotadas (ARP) o drones añade una dimensión aérea esencial para el patrullaje en zonas de difícil acceso. Estos dispositivos operan como extensiones móviles del SAD, identificando rutas clandestinas y recolectando evidencia digital. El uso de drones, junto con inteligencia artificial y escáneres en puestos de control, incrementa la interceptación de narcóticos y armas al superar las barreras físicas del terreno.

La red de comunicaciones, sustentada por infovías y el satélite SGDC, garantiza la resiliencia de la información y proyecta la presencia estatal incluso en comunidades aisladas (Nascimento, 2026; Voltolini Junior, 2025). Esta capacidad de conexión permanente transforma los datos técnicos en una herramienta de gestión política, ya

que permite al Estado actuar de manera proactiva en áreas que anteriormente solo gestionaba de forma intermitente. En suma, el despliegue del aparato tecnológico en el Arco Central trasciende la mera acumulación de dispositivos y se constituye como una infraestructura de gobernanza territorial que redefine la relación entre Estado, espacio y frontera (Santos & Santos Filho, 2023).

No obstante, este fortalecimiento de la capacidad de control estatal profundiza la naturaleza asimétrica de la frontera; mientras el acceso soberano a tecnologías de sensores y plataformas de apoyo a la decisión consolida la autonomía estratégica de un actor, la carencia de infraestructuras homólogas en el margen colindante restringe la capacidad de respuesta y altera los términos de la gobernanza binacional, transformando la disparidad técnica en un factor de jerarquización política del espacio fronterizo (Espósito Neto & Franchi, 2021; Magdalena, 2024).

#### **4.3. Paradigmas de Seguridad y Asimetría Tecnológica en el Arco Central: El Impacto de la Inteligencia Artificial y el SISFRON en la Gobernanza Fronteriza entre Brasil y Bolivia**

La seguridad en las regiones fronterizas ha dejado de ser una cuestión de simple demarcación territorial para convertirse en un campo de batalla complejo donde convergen la tecnología, la soberanía y la gestión de flujos transnacionales. En el contexto contemporáneo, las fronteras no solo actúan como límites físicos, sino como espacios de "interdependencia compleja", donde las acciones de un Estado repercuten inevitablemente en su vecino (Keohane & Nye, 2012).

Este fenómeno es particularmente visible en la frontera entre Brasil y Bolivia, una franja de aproximadamente 3.400 kilómetros caracterizada por una geografía desafiante, desde el Pantanal hasta la Amazonía, y una porosidad que facilita el tráfico ilícito de mercancías y personas (Buzan et al., 1998).

La soberanía, entendida tradicionalmente bajo el modelo westfaliano como el monopolio legítimo de la fuerza sobre un territorio cerrado, ha sido redefinida por la digitalización. Investigadores contemporáneos sugieren que la soberanía hoy se manifiesta de forma negociada y compartida, donde la capacidad de un Estado para controlar su territorio depende de su capacidad para procesar información y colaborar en redes de políticas públicas (Anderson & O'Dowd, 1999; Newman & Paasi, 1998).

En este marco, la tecnología actúa como un mediador fundamental. La implementación del SISFRON ha instaurado una asimetría técnica y operativa que redefine la frontera como un espacio de desigualdad informacional, donde la brecha no reside únicamente en la tenencia de hardware, sino en la capacidad diferenciada de los Estados para procesar datos y traducirlos en estrategias de gestión territorial (Santos & Santos Filho, 2023).

En este marco, la gobernanza fronteriza se operacionaliza en este estudio como el complejo entramado de instituciones, normas, acuerdos y dispositivos tecnológicos mediante los cuales los Estados administran la seguridad, regulan los flujos transfronterizos y contienen amenazas comunes (Anderson & O'Dowd, 1999; Newman & Paasi, 1998). Lejos de operar como un esquema simétrico o de coordinación homogénea, esta gobernanza se ve profundamente alterada por la tecnificación unilateral. Mientras el Estado brasileño proyecta un modelo de gestión proactiva y centralizada a través de sus subsistemas de comando y control, las capacidades en el margen boliviano permanecen condicionadas por restricciones presupuestarias y estructurales que limitan su actuación a dinámicas analógicas y reactivas (Cossul, 2019; International Crisis Group, 2023). De este modo, la gobernanza fronteriza en el Arco Central deja de ser una mera coordinación burocrática para configurarse como un reflejo directo de la asimetría material de la red binacional.

Esta jerarquía tecnológica incide directamente de forma asimétrica en la conciencia situacional de la faja fronteriza (Espósito Neto & Franchi, 2021). La falta de infraestructuras de comunicación satelital homogéneas en las regiones más remotas de la periferia boliviana restringe la autonomía de interdicción en el terreno, generando una dependencia funcional respecto a los flujos informacionales que el Estado vecino decida compartir (Magdalena, 2024; Paixão, 2022). Por lo tanto, el despliegue de estas plataformas de control digital en el espacio fronterizo, si bien busca optimizar la detección de flujos ilícitos, requiere ser analizado bajo marcos institucionales que ponderen el impacto sobre las poblaciones locales y las dinámicas reales de la cooperación técnica binacional (Soares & Souza, 2024; Pinheiro, 2024).

Para la política pública, la modernización digital y la seguridad informática se han consolidado como componentes estratégicos de la soberanía nacional. En Brasil, el año 2025 marca un hito con la aprobación de la Ley Complementaria 221 y el "PAC Militar", que destina R\$ 30 mil millones a la actualización de la defensa, priorizando la ciberdefensa y el fortalecimiento tecnológico del Subsistema de Apoyo a la Decisión

(SAD) (Exército Brasileiro, 2025; Lourenço, 2026). Este flujo de inversión busca propiciar la transición desde un patrullaje físico reactivo hacia una vigilancia proactiva apoyada en el procesamiento avanzado de datos, orientada a identificar de forma oportuna anomalías y flujos ilícitos en la faja fronteriza antes de que se consoliden (Soares & Souza, 2024).

Para sistematizar las disparidades operativas identificadas y dotar al análisis del rigor comparativo requerido, se presenta la Tabla 1, la cual contrasta las dimensiones críticas de la gestión fronteriza entre ambos Estados (Santos & Santos Filho, 2023; International Crisis Group, 2023):

Tabla 1: Matriz analítica de capacidades y gobernanza fronteriza entre Brasil y Bolivia

Dimensión	Brasil	Bolivia
Capacidades tecnológicas	Mayor desarrollo e integración de sistemas de monitoreo fronterizo automatizados (Santos & Santos Filho, 2023).	Capacidades limitadas de vigilancia tecnológica con predominio de controles manuales y reactivos (Cossul, 2019).
Infraestructura de seguridad	Despliegue operativo del SISFRON y consolidación de subsistemas de comando y control en el Arco Central (Oliveira, 2023).	Infraestructura tecnológica reducida y severas brechas de conectividad en áreas periféricas (Magdalena, 2024).
Procesamiento de información	Alta capacidad de generación, centralización y análisis masivo de datos estratégicos en tiempo real (Soares & Souza, 2024).	Dependencia parcial de flujos informacionales externos y de la inteligencia compartida por el Estado vecino (Paixão, 2022).
Continuidad institucional	Flujo de inversión presupuestaria sostenida en defensa y actualización normativa de programas estratégicos (Exército Brasileiro, 2025).	Restricciones financieras crónicas y volatilidad institucional que limitan la planificación a largo plazo (CEPAL, 2023).
Cooperación bilateral	Posición de negociación favorable basada en el monopolio de capacidades técnicas y de sensoramiento (Espósito Neto & Franchi, 2021).	Participación pragmática condicionada por la disponibilidad de recursos materiales y operativos locales (UNODC, 2022).

Fuente: elaboración propia.

La asimetría técnica entre Brasil y Bolivia es el reflejo de prioridades estratégicas y capacidades económicas dispares, pero su gestión es determinante para la seguridad regional. El impacto de la tecnología en las políticas públicas de seguridad profundiza la necesidad de consolidar una gobernanza fronteriza tecnificada, la cual requiere marcos regulatorios e institucionales que protejan los derechos de las poblaciones locales (Silva & Silva Filho, 2026; Pinheiro, 2024). El éxito de esta política dependerá de transformar la asimetría en una oportunidad para la cooperación técnica binacional, asegurando que la tecnología sirva como puente para la integración y no como un factor de exclusión.

## **5. Dinámicas Políticas y Operaciones Binacionales (2018–2024)**

### **5.1. La cooperación fronteriza como red de políticas públicas**

La seguridad fronteriza entre Bolivia y Brasil constituye un campo de política pública caracterizado por la convergencia de múltiples actores, niveles de decisión y capacidades institucionales, cuya articulación excede los esquemas tradicionales de acción estatal centralizada. En este sentido, la cooperación bilateral en materia de control territorial y combate a economías ilícitas puede ser comprendida a partir del enfoque de redes de políticas públicas, el cual permite captar la complejidad relacional y la dispersión de autoridad que definen este tipo de problemáticas.

Las políticas públicas no se reducen a decisiones gubernamentales formales. Se configuran a partir de procesos dinámicos de interacción entre actores que poseen distintos grados de poder, recursos y legitimidad (Secchi, 2013). Esta concepción resulta particularmente pertinente en el ámbito de la seguridad fronteriza, donde la implementación efectiva de estrategias depende de la coordinación entre instancias políticas, burocráticas y operativas, tanto a nivel nacional como transnacional.

Las redes de políticas públicas permiten analizar formas de acción estatal que desbordan los esquemas tradicionales de organización jerárquica. En el ámbito de la seguridad fronteriza, la intervención involucra múltiples actores con competencias fragmentadas y capacidades diferenciadas, lo que dificulta su comprensión a partir de modelos centrados en una autoridad única. En este contexto, dichas redes pueden entenderse como estructuras de interacción relativamente estables, en las que actores

interdependientes intercambian recursos y negocian objetivos en torno a la resolución de problemas colectivos (Rhodes, 1997; Börzel, 1997).

En Bolivia, la organización estatal ha estado marcada por dinámicas de centralismo, en las que la toma de decisiones políticas y estratégicas se concentra en el nivel nacional, lo que ha limitado la incorporación de las particularidades territoriales en los procesos de formulación de políticas públicas. Esta situación adquiere especial relevancia en las zonas fronterizas, donde los contextos locales están atravesados por flujos de personas, interacciones socioculturales y economías ilícitas que configuran necesidades específicas de control, coordinación horizontal, gestión territorial y adaptación constante. La concentración de decisiones reduce la capacidad de adaptación a estas condiciones y limita la articulación entre actores locales y nacionales, lo que contribuye a que dichas especificidades queden relegadas dentro de una lógica gubernamental que no logra captar la complejidad del espacio fronterizo y dificulta la articulación entre actores diversos.

En consecuencia, la cooperación fronteriza se configura como un entramado relacional atravesado por tensiones, negociaciones y reajustes permanentes. Las asimetrías en capacidades institucionales y acceso a tecnología no solo condicionan la distribución del poder entre los actores, sino que también inciden en los ritmos, alcances y formas que adopta la cooperación en la práctica. En este marco, las interacciones entre los Estados se desarrollan bajo condiciones diferenciadas que influyen en los mecanismos de coordinación, en la circulación de información y en la ejecución de acciones conjuntas.

Estas dinámicas adquieren particular relevancia en contextos de cambio político, donde la redefinición de prioridades estratégicas y la reconfiguración de actores gubernamentales introducen nuevos elementos de incertidumbre en el funcionamiento de la red. La alternancia de gobiernos no solo modifica los énfasis en materia de seguridad fronteriza, sino que también afecta los niveles de confianza, los mecanismos de intercambio de información y la continuidad de las iniciativas conjuntas. De este modo, la cooperación se configura como un proceso contingente, sujeto tanto a condicionamientos estructurales como a variaciones políticas, lo que incide directamente en su estabilidad y efectividad.

## 5.2. Dinámicas políticas y cooperación bilateral (2018–2024)

Los cambios de gobierno en Brasil y Bolivia entre 2018 y 2024 introdujeron reorientaciones diferenciadas en las políticas de seguridad fronteriza, con efectos directos sobre los niveles, modalidades y alcances de la cooperación bilateral. En este período, la gestión de la frontera dejó de responder a una lógica estable para configurarse como un proceso condicionado por la alternancia política, en el que las prioridades estratégicas, los enfoques de seguridad y los mecanismos de articulación institucional fueron redefinidos de manera recurrente. Estas variaciones no solo incidieron en la orientación general de las políticas, sino también en la forma en que los actores estatales se vincularon entre sí, tanto al interior de cada país como en el plano bilateral.

En este contexto, la cooperación fronteriza se vio atravesada por tensiones derivadas de diferencias ideológicas, cambios en los alineamientos internacionales y transformaciones en la estructura de toma de decisiones. La alternancia de gobiernos implicó modificaciones en los énfasis otorgados a la cooperación técnica, el intercambio de inteligencia y la coordinación operativa, lo que afectó la continuidad de los acuerdos y la estabilidad de los mecanismos de interacción.

Asimismo, estas transformaciones pusieron en evidencia las limitaciones estructurales de la cooperación bilateral, particularmente en lo que respecta a la capacidad de sostener estrategias coordinadas en contextos de inestabilidad política. La articulación entre ambos Estados no dependió exclusivamente de la existencia de acuerdos formales, sino de condiciones más amplias vinculadas a la confianza política, la convergencia de intereses y la compatibilidad de enfoques en materia de seguridad. En consecuencia, los cambios de gobierno operaron como un factor crítico que incidió en la densidad, continuidad y efectividad de la red de cooperación, generando escenarios de mayor o menor coordinación según el grado de alineamiento político entre las administraciones.

La evolución de la cooperación fronteriza durante el septenio 2018–2024 devela que las transiciones político-institucionales operan como variables críticas en la densidad y estabilidad de las redes de políticas públicas bilaterales. Lejos de responder a una trayectoria lineal, la gestión compartida de la seguridad se estructuró en fases diferenciadas, determinadas por el grado de alineamiento político y las capacidades materiales de cada administración. La fase inicial (2018–2019), coincidente con el cierre de ciclos gubernamentales de corte fuertemente multilateral en la región, estuvo

caracterizada por una desarticulación operativa derivada de prioridades dispares; mientras la agenda exterior de Bolivia privilegiaba marcos discursivos regionales con una limitada institucionalización del control en fronteras, el Estado brasileño inició la expansión de sus programas tecnológicos de vigilancia territorial (UNODC, 2022). La segunda fase (2019–2022), marcada por el ascenso de una doctrina de seguridad nacional unilateral y de militarización intensiva en Brasil, agudizó la brecha relacional mediante la ejecución de las operaciones "Sentinelaz Çhronos", priorizando el despliegue del SISFRON y restringiendo la interlocución con agencias civiles bolivianas (Huyó Millán & Vásquez Sabogal, 2022; Oliveira & Farias, 2024). Paralelamente, el quiebre institucional y la inestabilidad política transitoria registradas en Bolivia desde finales de 2019 limitaron la continuidad de las estrategias y profundizaron la desconfianza mutua, replegando la cooperación a canales militares preestablecidos. Finalmente, la fase de cierre (2023–2024), caracterizada por el retorno de un enfoque diplomático de integración regional en Brasilia, reconfiguró la gobernanza fronteriza hacia un paradigma de pragmatismo cooperativo. Este viraje en el Ejecutivo brasileño propició la reactivación de mecanismos bilaterales enfocados en el intercambio de inteligencia, patrullajes conjuntos y el desarrollo de infraestructura conectiva (Planalto, 2026).

No obstante, la efectividad de este viraje relacional encontró un límite estructural en el deterioro de las condiciones macroeconómicas y la crisis de liquidez del Estado boliviano (CEPAL, 2023; International Crisis Group, 2023). Las restricciones presupuestarias crónicas mermaron la presencia de personal especializado y la capacidad operativa de interdicción en nodos sensibles como el departamento de Santa Cruz, consolidando un escenario donde la estabilidad de la red fronteriza depende de un equilibrio inestable entre la convergencia política de los gobiernos y la persistente asimetría material de sus agencias (UNODC, 2022).

En este marco, la red de cooperación bilateral se configura a partir de dos dimensiones en tensión: la confiabilidad política entre los gobiernos y la capacidad operativa de los actores involucrados. En contextos de transición política —como el registrado en Bolivia en 2019— estas condiciones tienden a debilitarse, generando discontinuidades en los acuerdos, redefiniciones abruptas de prioridades y una menor previsibilidad en los mecanismos de coordinación. Tales interrupciones afectan la densidad de la red, reducen la estabilidad de los vínculos interinstitucionales y limitan la posibilidad de consolidar arreglos de cooperación sostenidos en el tiempo.

A su vez, la persistencia de asimetrías estructurales introduce un elemento adicional de complejidad en la configuración de la red. La concentración de capacidades tecnológicas y recursos financieros en el caso brasileño, frente a un rol boliviano más centrado en el control territorial, configura una distribución desigual de recursos que incide directamente en la toma de decisiones y en la definición de agendas operativas. Esta asimetría no solo condiciona los márgenes de autonomía de los actores, sino que también estructura relaciones de dependencia funcional al interior de la red.

En consecuencia, la efectividad de la cooperación binacional no puede ser comprendida como un resultado automático de la interacción entre Estados, sino como el producto de un equilibrio inestable entre condiciones políticas y capacidades materiales. La consolidación de la red depende, por un lado, del grado de convergencia entre las orientaciones gubernamentales y, por otro, de la capacidad de gestionar las asimetrías existentes sin comprometer la coordinación operativa ni la autonomía relativa de los actores involucrados.

### **5.3. Inteligencia, información y coordinación operativa en la cooperación fronteriza**

La cooperación en materia de seguridad fronteriza entre Bolivia y Brasil no se limita a la presencia territorial ni al despliegue de capacidades materiales. Se articula de manera decisiva en torno a la producción, circulación y uso estratégico de la información. En este marco, la inteligencia se configura como un recurso central dentro de la red de políticas públicas, en tanto permite anticipar dinámicas ilícitas, orientar operativos y coordinar acciones entre actores con competencias diferenciadas.

En el marco de la cooperación bilateral, el intercambio de información no opera como un proceso automático ni plenamente institucionalizado. Se desarrolla como una práctica condicionada por niveles variables de confianza, capacidades técnicas y alineamientos políticos. La efectividad de los mecanismos de inteligencia depende de la existencia de canales estables de comunicación entre agencias —policiales, militares y de investigación—, así como de la compatibilidad entre sus sistemas de procesamiento y análisis de datos. Cuando estas condiciones no se cumplen, la circulación de información se fragmenta, lo que afecta la coordinación operativa y reduce la capacidad de respuesta frente a amenazas transnacionales.

Desde la perspectiva de las redes de políticas públicas, la inteligencia no cumple únicamente una función instrumental, sino que también estructura relaciones de poder al interior de la red (Rhodes, 1997; Secchi, 2013). El acceso diferencial a información estratégica, junto con la capacidad de procesarla y transformarla en decisiones operativas, introduce jerarquías implícitas entre los actores involucrados. En consecuencia, quienes concentran mayores capacidades de generación y análisis de inteligencia adquieren una posición predominante en la definición de agendas, en la priorización de objetivos y en la conducción de operaciones conjuntas (Börzel, 1997; Secchi, 2013).

En la frontera entre Bolivia y Brasil, estas dinámicas se expresan de manera concreta en la interacción entre sistemas de vigilancia, unidades operativas y mecanismos de intercambio de información. El desarrollo de plataformas tecnológicas por parte del Estado brasileño, orientadas al monitoreo en tiempo real y al procesamiento de datos, le otorga una ventaja significativa en la generación de inteligencia estratégica (Oliveira & Farias, 2024; Santos & Santos Filho, 2023). Esta capacidad incide en la orientación de operativos y en la definición de prioridades dentro de la cooperación bilateral. Sin embargo, las limitaciones en infraestructura tecnológica y en capacidades de procesamiento condicionan la participación boliviana en los circuitos de inteligencia, lo que incide de forma directa en sus márgenes de incidencia dentro de la coordinación conjunta (Cossul, 2019; Magdalena, 2024).

Estas condiciones repercuten en la forma en que se organiza la coordinación operativa en el espacio fronterizo. La planificación y ejecución de operativos conjuntos dependen de la calidad, oportunidad y grado de intercambio de la información disponible, lo que introduce diferencias en los tiempos de respuesta y en la capacidad de intervención de cada actor (Espósito Neto & Franchi, 2021; Paixão, 2022). La ausencia de protocolos plenamente consolidados y de sistemas interoperables limita la articulación sostenida entre agencias, generando acciones que, en muchos casos, responden a lógicas paralelas antes que a estrategias integradas (UNODC, 2022). Esta fragmentación operativa reduce la eficacia de las intervenciones frente a dinámicas ilícitas que operan con altos niveles de flexibilidad y adaptación (Huyó Millán & Vásquez Sabogal, 2022; UNODC, 2022).

En este contexto, la inteligencia y la gestión de la información se consolidan como ejes estructurantes de la cooperación fronteriza. La red de políticas públicas no se define únicamente por la presencia de actores ni por la existencia de acuerdos formales, sino que se configura a partir de la capacidad efectiva de articular flujos

de información, sostener canales de coordinación y transformar datos en decisiones operativas (Rhodes, 1997; Secchi, 2013). Las diferencias en capacidades técnicas, acceso a tecnología y niveles de institucionalización condicionan el funcionamiento de la red y delimitan los márgenes de acción de cada actor, consolidando la cooperación en seguridad fronteriza como un proceso relacional, dinámico y estructuralmente asimétrico (Fagerberg, 1987; Santos & Santos Filho, 2023).

## 6. Conclusión

El análisis sistémico de la gestión fronteriza en el Arco Central devela que la tecnificación del control territorial mediante dispositivos avanzados no desmantela el modelo tradicional de soberanía, sino que lo reconfigura bajo dinámicas de interdependencia asimétrica (Keohane & Nye, 2012; Santos & Santos Filho, 2023). Los hallazgos confirman que las infraestructuras críticas del SISFRON no reemplazan el despliegue operativo tradicional en el terreno, sino que reconfiguran la vigilancia mediante una lógica de complementariedad táctica, optimizando de forma continua el control en geografías de baja densidad demográfica (Oliveira, 2023; Silva, 2025). Esta digitalización de la periferia delinea una nueva dimensión de la presencia estatal, donde la capacidad para convertir el espacio físico en flujos informacionales procesables se constituye como el principal vector de autoridad estratégica (Pohle & Thiel, 2020; Voltolini Junior, 2025).

Asimismo, la evidencia documental demuestra que la gobernanza fronteriza tecnificada se encuentra en un proceso de consolidación asimétrico dentro del ciclo de políticas de defensa (Silva & Silva Filho, 2026). Si bien los subsistemas de apoyo a la decisión y las herramientas de procesamiento avanzado de datos buscan identificar flujos ilícitos y optimizar la conciencia situacional en el margen brasileño, su efectividad e institucionalización enfrentan el desafío de marcos institucionales heterogéneos y la persistente desconexión técnica con el actor colindante (Pinheiro, 2024; Soares & Souza, 2024). Esta disparidad tecnológica redefine los términos de la gobernanza fronteriza regional; la asimetría técnica se traduce en una asimetría informacional estructural que condiciona la agenda de seguridad y sitúa a las agencias del Estado con menores capacidades técnicas en una posición de dependencia funcional respecto a la inteligencia compartida (Fagerberg, 1987; Espósito Neto & Franchi, 2021).

Finalmente, el examen del período 2018–2024 demuestra que la cooperación bilateral en materia de seguridad transnacional está supeditada a un sutil equilibrio entre la convergencia político-ideológica de las administraciones de turno y las capacidades materiales de sus instituciones (UNODC, 2022). Aunque las transiciones gubernamentales recientes han propiciado un retorno al pragmatismo cooperativo y a la reactivación de canales diplomáticos institucionales, las restricciones macroeconómicas y las crisis de liquidez en el margen boliviano operan como un límite estructural para la ejecución de operativos conjuntos y el sostenimiento de infraestructuras homólogas (CEPAL, 2023; International Crisis Group, 2023). Por consiguiente, la sostenibilidad de la red fronteriza compartida requiere transitar desde compromisos discursivos de integración hacia el diseño de mecanismos de transferencia técnica y financiamiento binacional que permitan mitigar las brechas operativas y garantizar una gobernanza territorial verdaderamente cooperativa y simétrica (Anderson & O’Dowd, 1999; Newman & Paasi, 1998).

## Referencias

Anderson, J., & O’Dowd, L. (1999). Borders, border regions and territoriality: Contradictory meanings, changing significance. *Regional Studies*, 33(7), 593–604.

Bigo, D. (2002). Security and immigration: Toward a critique of the governmentality of unease. *Alternatives*, 27(1), 63–92.

Börzel, T. A. (1997). What’s so special about policy networks? An exploration of the concept and its usefulness in studying European governance. *European Integration Online Papers*, 1(16), 1–24.

Buzan, B., Wæver, O., & de Wilde, J. (1998). *Security: A new framework for analysis*. Lynne Rienner Publishers.

Comisión Económica para América Latina y el Caribe [CEPAL]. (2023). *Estudio económico de América Latina y el Caribe*. CEPAL.

Cossul, N. I. (2019). Potencialidade do Programa Espacial Brasileiro para a integração regional: Aprofundamento da cooperação regional por meio do emprego das tecnologias de informação e comunicação. *Revista Videre*, 11(21), 209–228.

Espósito Neto, T., & Franchi, T. (2021). As múltiplas visões sobre o Sistema Integrado de Monitoramento de Fronteiras (SISFRON). *Monções: Revista de Relações Internacionais da UFGD*, 10(20), 459–487.

Exército Brasileiro. (2025). Portaria nº 512-EME: Diretriz de Implantação do Programa Estratégico do Exército Sistema Integrado de Monitoramento de Fronteiras - SISFRON (Atualización). Ministério da Defesa.

Fagerberg, J. (1987). A technology gap approach to why growth rates differ. *Research Policy*, 16(2-4), 87–99.

Foucault, M. (2008). Seguridad, territorio, población: Curso en el Collège de France (1977-1978). Martins Fontes.

Huyó Millán, D., & Vásquez Sabogal, N. S. (2022). Una mirada a la gestión del narcotráfico en Brasil: De la militarización a la cooperación y el desarrollo regional. *Ciencias Políticas y Relaciones Internacionales*, 16(2), 229–268.

International Crisis Group. (2023). Bolivia's political and economic fragility (Latin America Report N.º 98). International Crisis Group.

Kant, I. (1795). *Hacia la paz perpetua*. Alianza Editorial.

Keohane, R. O., & Nye, J. S. (2012). *Power and interdependence* (4.ª ed.). Longman.

Lourenço, C. (10 de marzo de 2026). Documento do Exército revela base de plano de R\$ 456 bilhões apresentado a Lula. Agência Brasil.

Magdalena, G. M. L. (2024). Análise da segurança na fronteira no Estado de Mato Grosso do Sul e as tecnologias de monitoramento na atualidade. *Revista (RE)Definições das Fronteiras*, 2(8), 44–55.

Nascimento, A. P. S. (2026). O uso de drones pela Polícia Militar do Amazonas como meio de ampliar o seu alcance no policiamento nos rios e fronteiras. *Revista Ibero-Americana de Humanidades, Ciências e Educação*, 12(2), 1–15.

Newman, D., & Paasi, A. (1998). Fences and neighbours in the postmodern world: Boundary narratives in political geography. *Progress in Human Geography*, 22(2), 186–207.

Oliveira, A. L. F. T. (2023). O Sisfron como ferramenta da estratégia da presença em

medio ao desafio orçamentário entre 2012 e 2022. *Revista Brasileira de Estudos de Defesa*, 10(2), 195–217.

Oliveira, P. E. S., & Farias, J. A. (2024). *Sisfron como projeto estratégico do Exército Brasileiro na defesa da fronteira oeste* [Tesis de maestría, Universidade Federal de Mato Grosso do Sul].

Paixão, R. O. (2022). Sistema Integrado de Monitoramento de Fronteira - SISFRON e os desdobramentos nos territórios de Corumbá e Ponta Porã - MS. *Geofronter*, 8, 1–18.

Pinheiro, F. M. L. (2024). O papel da tecnologia e da inteligência artificial na seguridad pública: Desafíos e convergências com os direitos fundamentais individuais e sociais. *Revista do Ministério Público do Estado do Rio de Janeiro*, (94), 57–78.

Pohle, J., & Thiel, T. (2020). Digital sovereignty: a shifting concept. *Internet Policy Review*, 9(4), 1-19.

Planalto. (17 de marzo de 2026). Lula aboga por una mayor integración entre Brasil y Bolivia: “Una frontera viva que conecta a los pueblos, las culturas y la economía” [Comunicado de prensa].

Rhodes, R. A. W. (1997). *Understanding governance: Policy networks, governance, reflexivity and accountability*. Open University Press.

Santos, C. E. R., & Santos Filho, J. (2023). SISFRON - uma realidade na fronteira, um sensoramento de apoio à decisão na faixa de fronteira de Santa Catarina e do Paraná. *Revista (RE)Definições das Fronteiras*, 1(2), 375–393.

Secchi, L. (2013). *Políticas públicas: Conceitos, esquemas de análise e casos práticos*. Cengage Learning.

Silva, F. J. A. (2025). Inovação tecnológica na segurança pública em áreas de fronteira: Desafios e oportunidades no Estado do Amazonas. *Revista Ibero-Americana de Humanidades, Ciências e Educação*, 11(12), 1–17.

Silva, S. F., & Silva Filho, H. F. (2026). Inteligência Artificial na Segurança Pública Brasileira: Aplicações, riscos cibernéticos e desafios de governança. *Revista FT*, 10(2026), 111–125.

Soares, R., & Souza, J. (2024). O uso da Inteligência Artificial no SISFRON. *Revista de*

Seguridad, Desarrollo y Defensa, 1(1), 32–54.

Slaughter, A. M. (2004). *A new world order*. Princeton University Press.

United Nations Office on Drugs and Crime [UNODC]. (2022). *World Drug Report 2022*. UNODC.

Van Houtum, H. (2005). The geopolitics of borders and boundaries. *Geopolitics*, 10(4), 672–679.

Voltolini Junior, V. (2025). O emprego contínuo e imediato de drones no policiamento de fronteira. *Revista Científica RECIMA21*, 6(6), artículo e666535.

## ¿Cultura y venganza? Los usos de la retaliación en la vida social de una comunidad campesina de los Andes colombianos

*Culture and Revenge? The Uses of Retaliation in the Social Life of a Peasant Community in the Colombian Andes*

Santiago Álvarez  <sup>1</sup>

<sup>1</sup> Filiación institucional pendiente de confirmación.  
salvarez@udesa.edu.ar (Autor de correspondencia)

**Resumen.** En este artículo intentamos reflejar, a través de la observación de un grupo familiar y la descripción de sus principales actividades, la particular relevancia de la violencia interna entre personas del mismo estrato social campesino en una comunidad rural del Altiplano Cundiboyacense. De esta manera, más allá del hecho en sí, se explorará el significado social de este tipo de prácticas. En esta comunidad, situada al pie del Páramo de Sumapaz, desarrollamos una investigación antropológica de quince meses, desde 1994 a 1996, con un regreso al campo durante veinte días en el año 2002, principalmente mediante el uso de la observación participante.

**Palabras clave:** Antropología de la violencia, Colombia, venganzas, parentesco

**Abstract.** This article seeks to reflect, through the observation of a family group and the description of its main activities, on the particular relevance of internal violence among people from the same peasant social stratum in a rural community of the Cundiboyacense Altiplano. In this way, beyond the event itself, it explores the social meaning of this type of practice. In this community, located at the foot of the Sumapaz páramo, we conducted a fifteen-month anthropological investigation from 1994 to 1996, with a twenty-day return to the field in 2002, mainly through participant observation.

**Keywords:** anthropology of violence; Colombia; revenge; kinship

**Historial:** Recibido: 21 de marzo de 2026 | Aceptado: 31 de mayo de 2026

**Ética/Financiamiento:** El autor declara no tener conflictos de interés. El artículo fue autofinanciado y no recibió financiamiento externo.

Este documento está bajo una licencia  
CC BY-NC-ND 4.0.

**Cómo citar (APA):** Álvarez, S. (2026). ¿Cultura y venganza? Los usos de la retaliación en la vida social de una comunidad campesina de los Andes colombianos. *Conocimiento i Política*, 7(1), 223-242. DOI: 10.64480/cip.2026.n7.07.

## 1. Introducción

Al comenzar nuestro trabajo de campo en los Andes colombianos esperábamos estudiar formas de violencia organizadas: los narcotraficantes, la guerrilla, el mismo Estado colombiano. La existencia, en el pasado, de una rebelión campesina exitosa que había expulsado a los antiguos terratenientes de la región parecía hablarnos de solidaridad campesina, un concepto que, si bien postula una dicotomía entre campo y ciudad, es útil a fines analíticos (Fafchamps, 1992). Nada de esto presagiaba encontrar venganzas de sangre entre familias campesinas, un vacío en la literatura especializada, al menos en el área que estudiamos.

En este artículo intentamos reflejar, a través de la observación de un grupo familiar y la descripción de sus principales actividades, la particular relevancia de la violencia interna entre personas del mismo estrato social campesino en una comunidad rural del Altiplano Cundiboyacense. De esta manera, más allá del hecho en sí, se explorará el significado social de este tipo de prácticas. En esta comunidad, situada al pie del Páramo de Sumapaz, desarrollamos una investigación antropológica de quince meses, desde 1994 a 1996, con un regreso al campo durante veinte días en el año 2002, principalmente mediante el uso de la observación participante.

Existen dificultades objetivas para arribar a una satisfactoria definición intercultural del concepto de violencia tal cual lo puntualiza David Riches en *Anthropology of power* (1986: 28). El concepto de violencia es polisémico y ambiguo (Álvarez, 2017). “Esto probablemente se relacione con la tendencia a calificar como “violentos” una serie cada vez más amplia de fenómenos, como las desigualdades sociales, políticas y económicas” (Álvarez, 2017: 27). Por todo esto preferimos, a los efectos de este trabajo, definir a la violencia en un sentido amplio como: “toda acción humana que envuelve la deliberada inflicción de daño sobre otros” (Marvin, 1986: 121).

Teniendo en cuenta los antecedentes de movilización y lucha campesina en la región descrita, entre otros, por Marulanda (1990: 74) y Londoño (1994: 58-59), una de las alternativas que se nos presentaban era centrar el análisis en los conflictos del campesinado con otras clases sociales y el Estado nacional. Estos conflictos, si nos atenemos a los análisis históricos mencionados, parecen haber estado en el centro de la problemática regional durante el período de la colonización campesina y durante las luchas agrarias que tuvieron como consecuencia el desmantelamiento de las

grandes haciendas y la división de la tierra (Bejarano, 1983). Un estudio así realizado hubiera estado en concordancia con la mayoría de los análisis del campesinado en la región andina que se basan, preferentemente, en un estudio de las "estrategias de supervivencia campesinas" (Stern, 1990: 15) y su relación con los movimientos políticos locales y nacionales.

Si bien consideramos que el momento de la llamada "rebelión campesina", objeto clásico de los estudios sobre el campesinado de los años sesenta, había finalizado o estaba latente, pensábamos que el centro de nuestra investigación podía ser otro: el de la resistencia campesina", de "las armas de los débiles" (Scott, 1985) y su relación con un movimiento guerrillero inserto en la comunidad campesina con el objetivo de brindar protección a sus habitantes.

Siguiendo los planteos de Scott (1985:29) y otros académicos (Adas, 2012; Hobsbawm, 1969; Isaacman, 1980) ciertas formas diarias de resistencia tienen la capacidad de horadar las relaciones de poder existentes. Entre ellas, pueden considerarse la disminución del ritmo de trabajo, el sabotaje, y el robo. De alguna manera, constituyen formas de una lucha de clase campesina que no tenía por objetivo la rebelión o la revolución pero que implicaban acciones de resistencia.

Contra las expectativas previamente citadas y teniendo en cuenta que la comunidad analizada se encontraba en una zona de conflicto entre el Estado y la guerrilla, no pudimos dejar de registrar la enorme relevancia del fenómeno que Isaacman denomina "disenso interno" (1980: 15-56), es decir, los conflictos existentes dentro de la comunidad campesina.

El grado de tensiones y agresiones entre la población analizada es altamente significativo. Graves conflictos entre familias son resueltos a través de las venganzas de sangre. Esta violencia recíproca que afecta a la mencionada comunidad ha constituido el centro del trabajo de campo realizado y es el principal objeto de este escrito.

En este artículo nos centraremos en el análisis de un grupo familiar implicado en relaciones violentas dentro de la misma comunidad. En ella, existen numerosos ejemplos de enfrentamientos de familias enemistadas que terminaron con el casi aniquilamiento de los varones de las familias en conflicto. Estas experiencias se conservan traumáticamente en la memoria colectiva de la comunidad. Sin embargo, las venganzas descritas en este artículo tuvieron lugar única y exclusivamente durante el transcurso de esta investigación.

Existen diversos trabajos que abordan la venganza de sangre. Standish (2014) explora los dolorosos caminos que la violencia de género puede recorrer en una comunidad y el significado que el cuerpo, como objeto de honor, es capaz de tener. Dado que las culturas son heterogéneas, resulta interesante la exploración de Barmash (2004) respecto a la ley Mesopotámica en el segundo y primer milenio A.D, donde los familiares de la víctima se resistían a hacer una venganza de sangre, aunque si llegara a ocurrir el homicida se resguardaría en un santuario. Black-Michaud (1980) establece una distinción terminológica entre la feud y la venganza: esta última sería indiscriminada y no admitiría compensación mientras que la feud sí. No obstante, ambas están interconectadas y sus diferencias desaparecen en las prácticas de los actores (Black-Michaud, 1980: 19-20). En este artículo nos referiremos a la venganza de sangre o vendetta. En otros casos estudiados se observa un intento de la guerrilla por mediar entre las partes tratando de poner fin, al menos temporariamente, a los conflictos, lo que acercaría a las venganzas analizadas a circunscribirlas a la feud (Black-Michaud, 1980: 17). Estos intentos no fueron particularmente efectivos ni fueron incapaces de controlar las vendettas o de reglamentarlas (Álvarez, 1999).

La familia extensa es el grupo mínimo que expresa solidaridad social y que participa activamente en la venganza. En el caso de las familias estudiadas en Nómeque debemos señalar que esta solidaridad en el caso de las vendettas se limitaba a hermanos, medio hermanos y, más excepcionalmente, a primos hermanos. Las redes sociales locales son bastante laxas y poca importancia se le da a la institución del compadrazgo como creadora de vínculos perdurables, de hecho en algunas vendettas las familias enfrentadas habían sido compadres entre ellas. Las familias envueltas en venganzas viven su vida en la comunidad sin ser perseguidas o cuestionadas por ninguna autoridad estatal: de cualquier manera, es pertinente aclarar que durante el trabajo de campo la estación de policía había sido abandonada luego de un ataque guerrillero que la destruyó completamente, matando a varios de sus defensores. Este hecho reforzaba la ineficacia del sistema legal.

No es la vendetta la única forma de violencia percibida en la comunidad, pero consideramos que su estudio puede aportar elementos significativos que, desde una perspectiva distinta, nos indiquen que por debajo de los conflictos que acaparan las investigaciones y los discursos oficiales sobre la problemática de la violencia subyace en la comunidad campesina una conflictividad interna que merece ser objeto de investigación.

## 2. Los Casares, carnicería y matadero

*"Matasiete se tiró al punto del caballo, cortóle el garrón de una cuchillada y gambeteando en torno de él con su enorme daga en mano, se la hundió al cabo hasta el puño en la garganta, mostrándola enseguida humeante y roja a los espectadores. Brotó un torrente de la herida, exhaló algunos bramidos roncós, vaciló y cayó el soberbio animal entre los gritos de la chusma que proclamaba a Matasiete vencedor y le adjudicaba en premio el matambre. Matasiete extendió, como orgulloso, por segunda vez el brazo y el cuchillo ensangrentado, y se agachó a desollarlo con otros compañeros "(Esteban Echeverría, El Matadero, 1851: 85)*

Los Casares son una familia compuesta por la madre viuda y cinco hijos varones: juntos trabajan una "fama"(carnicería). Tres hermanos Casares tienen relaciones estables con mujeres con las que cohabitan, en la mayoría de los casos, luego de haber tenido un hijo en común. Las relaciones con sus respectivas mujeres, si bien está marcada por los malos tratos y las peleas no difieren sustancialmente de las relaciones habituales entre miembros de distinto sexo que caracterizan al mismo grupo social de la comunidad analizada. En este grupo no es extraño que el hombre, semanalmente al menos, vuelva borracho al hogar a altas horas de la noche y golpee a su mujer: se trata de una norma con pocas excepciones.

El eje de la familia en la región analizada es la madre, es ella la que se responsabiliza de la crianza de los hijos, mientras que el padre, generalmente, se despreocupa de este rol. Esta matrifocalidad está presente en una familia compuesta por hermanos varones que rodean a la madre. El padre y un hermano menor de los Casares murieron trágicamente hace apenas un año como desarrollaremos luego. Su situación económica es bastante precaria: alquilan el local que hace las veces de negocio así como también la vivienda en la que habitan. Existe una pronunciada diferencia de edad entre el más grande de los hermanos, de unos cuarenta años, y el más pequeño que no llega aún a los dieciocho. Ninguno de ellos cursó más allá de los estudios primarios.

Mi conocimiento y relación con la familia Ramallo fue diametralmente opuesto. Esto se debió principalmente a dos factores: por un lado mi relativa cercanía a la familia Casares que me hizo sospechoso de cierta parcialidad y me permitió entrar en contacto solo con algunos parientes lejanos de los Ramallo; por otro lado, los Ramallo dejaron de ir al pueblo después de las primeras venganzas y no tuve más

ocasiones para verlos que en los funerales de sus parientes. Era esta una familia rural de mejor posición económica que poseía algunas hectáreas no lejos de Nόμεque, y que alquilaba a los Casares la casa en donde vivían antes de que el conflicto se desatara. La figura paterna, ejercida por el viejo Ramallo, era tal vez más relevante, si bien se decía que tenía dificultades en controlar a sus hijos varones famosos por su ferocidad y violencia. Era la de los Ramallo una familia más temida antes que amada; mis informantes la relacionaban con el robo de ganado y de hecho algunos de sus miembros habían pasado cierto tiempo en prisión acusados de este delito.

La carnicería en la que trabajan los Casares es un local pequeño, de unos siete metros de largo por unos cuatro de ancho, en él se despacha carne casi todos los días, con excepción de los lunes y algunos feriados. En la carnicería la carne es expuesta y colgada de grandes ganchos. Para la comunidad constituye casi un lugar común comparar sus actividades como carniceros con su participación en acciones violentas. Una vecina manifestaba su terror de ver a "esos grandotes manejando esos cuchillos enormes cortando carne". La sangre derramada en la carnicería es comparada con la sangre derramada en las venganzas.

El matadero de la comunidad analizada es un edificio de una sola planta de unos veinte metros de largo por unos cinco de ancho. Afuera existen dos corrales en los que el ganado espera su turno para ser sacrificado. En él sería posible carnear al mismo tiempo unas cuatro reses si fuese necesario. Los días jueves, las carnicerías preparan la carne para vender durante el viernes y especialmente el sábado, que es el día del mercado. Los encargados de la faena son generalmente los empleados de las carnicerías o los mismos carniceros cuando no poseen personal empleado. En el particular caso de los Casares ellos mismos se encargan de faenar el ganado.

El matarife, ayudado por algunos compañeros de faena, debe hacer ingresar el animal al matadero, atar su cabeza y sus patas a unas argollas de hierro colocadas en el piso y haciendo uso de habilidad y fuerza física, dominar al animal y voltearlo para poder luego maniatarlo completamente y proceder a matarlo. Para matarlo le realiza una incisión en el cuello buscando la yugular de la que inmediatamente brota un gran chorro de sangre que es depositada en un balde y que luego se vende para su consumición. En unos cinco minutos, aproximadamente, el animal queda desangrado y luego de verificar que este ya no reacciona, el carnicero procede a quitarle el cuero para luego trozarlo y eviscerarlo. Algunas mujeres acompañadas de niños se encargan de cortar y lavar las vísceras del animal.

Durante este proceso el matarife queda impregnado de sangre, su relación con el animal es directa, y la violencia ejercida es la necesaria para obtener el fin requerido. El observador queda fuertemente impresionado por una serie de percepciones primarias, la sangre, los pedazos de carne que son cortados y expuestos produciéndonos una cierta sensación de obscenidad ya que formaron parte de un ser que estuvo vivo hace solo unos instantes. Esta obscenidad nos remite a la matanza como espectáculo. En la acción de matar una res estamos en presencia de un espectáculo cuya repetida representación nos hace situar una rutinización del horror. Es este acostumbramiento al horror una de las intuiciones claves del texto de Esteban Echeverría - del que se extrajo un párrafo que sirve de encabezamiento a este capítulo - y que este escribiera a mediados del siglo XIX, convirtiéndose en canónico para la literatura argentina. Echeverría (1851) nos presenta un matadero poblado por seres degradados por la insensibilidad, transformados en masa amorfa sin individualidad que terminan matando a un disidente político de la misma manera en que se mata al ganado.

El trabajo de María Victoria Uribe, *Matar, rematar y contramatar* (1978) hace referencia a las mutilaciones que se realizaron a los cadáveres de las víctimas de la violencia política que azotó al Tolima durante los años cincuenta. Estos cortes tenían un contenido simbólico, se imprimía un mensaje de terror utilizado para ejercer un control social determinado en un momento de convulsión política. Las mutilaciones presentaban características similares a los cortes que se realizaban sobre el ganado en diversas tareas rurales, como la castración o el descuartizamiento. Según la citada autora, las personas que mutilaban los cadáveres de sus víctimas eran las mismas que se relacionaban violentamente con el mundo animal. La relación violenta que impera entre el campesino y la naturaleza que debe dominar se traslada a la relación víctima-victimario; esta traslación no es directa ni automática, obviamente no todas las personas que desarrollan tareas en las que se ejerce violencia sobre animales ejercen violencia sobre personas. Por otra parte, es necesario aclarar que los Casares cometían homicidios contra los Ramallo pero no mutilaban a sus víctimas. En todo caso, nuestra afirmación recorre la hipótesis de que la forma en que ciertas personas del ámbito rural ejercen violencia contra otros hombres está impregnada de costumbres y usos que se desarrollan en el dominio violento de los animales.

### 3. Las peleas de gallos

*"Me parece que lo estuviera viendo cuando salió con el gallo debajo del brazo. Le advertí que no fuera a buscar una mala hora en la gallera y él me mostró los dientes y me dijo: - cállate que esta tarde nos vamos a podrir de plata-" (Gabriel García Márquez, 1972).*

Los Casares tienen un pasatiempo y una pasión: las peleas de gallos. En ellas participan uniéndose, a veces, en apuestas con Walter, quien también faena reses en el matadero. Walter dedica la mayor parte de su tiempo libre a la cría y preparación de gallos de pelea. "Los gallos para mí son todo, son más importantes que las mujeres". Junto con él está siempre su compañero inseparable Ricardo, <sup>el</sup> tolimense un amigo de unos cuarenta años quien le ha enseñado varios trucos en el arte de preparar gallos. Ricardo trabaja en una panadería. Los Casares y otros amigos los acompañan en sus aventuras prestándoles sus gallos para que los peleen o apostando por ellos en las peleas.

Existen dos galleras en la comunidad analizada. Una se encuentra en el barrio más pobre llamado "Simón Bolívar", barrio obrero o, más informalmente, "barrio chino". Se trata de la mejor construida, consta de un recinto circular con tribunas que pueden alcanzar a contener hasta a ochenta espectadores. Posee reloj para definir más exactamente el tiempo de las peleas y hasta el juez que se hace presente posee un aspecto casi profesional. Son también mayores las apuestas así como la cantidad y calidad de los gallos que se presentan. La otra gallera se encuentra asimismo en un área marginal de la población en un grupo de casas que sigue el recorrido de uno de los arroyos que pasan por el pueblo. Esta gallera posee aproximadamente la mitad de la capacidad de la anteriormente descrita y es menos concurrida. Las peleas se realizan generalmente los domingos alternándose entre las dos de manera que cada una tiene actividad cada quince días.

Antes de que empiecen las peleas se van pesando los distintos gallos y de acuerdo al peso del animal se van acordando las peleas. Alrededor de los gallos se va formando un grupo de sostenedores que juntan el dinero suficiente para apostar contra la otra parte: en la gallera del arroyo se apostaba cien mil pesos entre los dos gallos, en la del barrio Simón Bolívar hasta doscientos mil pesos. Para lograr llegar a esa suma es necesario muchas veces apelar a las relaciones de amistad. En efecto, si bien estas asociaciones son muy flexibles y nadie tiene en principio la obligación de apostar por un gallo en la práctica se imponen ciertas solidaridades.

Las personas que crían a los gallos, los preparan para pelear o los lanzan a la arena son contadas. Es muy común que una persona haya criado un gallo pero que, llegado el momento de hacerlo pelear, haya debido acudir a manos más expertas, o que una vez que listo el animal solicite a otros con más experiencia prepararlo para la pelea colocándole los espolones, delicada tarea de la que puede depender la suerte del animal, o para que lo lance a la arena. Por eso Walter en varias ocasiones ha presentado con fortuna gallos de su propiedad pero luego sus ganancias han disminuido o han desaparecido por haberse sentido obligado a apostar a favor de gallos de amigos aunque de dudosa fortuna. Es este uno de los momentos donde se percibe una mayor agregación social en una comunidad que posee pocas ocasiones en las que se expresa su solidaridad. Sin embargo esa solidaridad es, como intentamos explicar sumamente flexible y, lamentablemente, se expresa frente a otros, en contradicción con otros grupos de apostadores.

Estas formas de apostar, en donde las relaciones personales y la lealtad entre grupos tienen más importancia que la ganancia poseen obvias similitudes con la apuesta central descrita en el clásico trabajo de Clifford Geertz sobre las riñas de gallos en Bali (Geertz, 1997: 349). Geertz nos habla de dos tipos de apuestas, la apuesta central entre las personas principales, que es colectiva y comprende a coaliciones de apostadores alrededor del dueño del gallo, y las apuestas periféricas que se realizan separadamente entre espectadores que buscan básicamente el lucro (Geertz, 1997: 349).

Los Casares junto a Walter y Ricardo se asociaban constantemente para jugar gallos juntos. Ricardo arreglaba los gallos de los Casares y era quien los lanzaba al ruedo. Esto sucedía cada vez que los Casares llevaban un gallo a pelear. Si no presentaban gallos generalmente alguno de ellos se hacía presente y apostaba junto a Ricardo. En las últimas peleas esta relación había terminado y cada grupo apostaba por su lado.

Teniendo en cuenta que no se recibe ninguna gratificación especial por presentar los gallos en el ruedo y que al costo de criar y preparar los gallos se debe sumar el pago a los servicios del juez si las peleas ganadas y las pérdidas fuesen equivalentes, la sumatoria final sería a cero. Si bien los gallistas siempre creen que sus gallos son mejores que los otros y si bien existen quienes ganan mucho más de lo que pierden se comprende a simple vista que la actividad a la que se dedican con pasión no se trata del más rentable de los negocios.

Durante la pelea los apostadores gritan y se mueven apasionadamente al ritmo de sus gallos, las apuestas continúan entre aquellos que no participaron de los grupos iniciales. En este caso estaríamos frente a un modo de apostar similar al de las apuestas periféricas descritas por Geertz (Geertz, 1997: 349). La catarsis, producida por esta descarga de energías, se hace visible especialmente si se tiene en cuenta la escasez de actividades recreativas en la comunidad. Pese a todo, las peleas de gallos han sido en más de una ocasión escenarios de peleas que en algunos casos han llevado a las personas involucradas a agredirse físicamente con cuchillo o pistola. Hace unos dos años una gallera que se encontraba en la entrada del pueblo fue cerrada luego de que una persona fuese muerta de bala después de una discusión.

#### 4. Los Casares y la venganza de sangre

*.<sup>A</sup> partir del momento en que la violencia intestina rechazada por el sacrificio revela ligeramente su naturaleza, se presenta, como acabamos de ver, bajo la forma de venganza de sangre o "blood feud", que no desempeña en nuestro mundo más que un papel insignificante o incluso nulo. Tal vez sea por ahí por donde convenga buscar la diferencia de las sociedades primitivas, la fatalidad específica de que nos hemos librado y que el sacrificio no puede, evidentemente, apartar pero sí mantener dentro de unos límites tolerables" (René Girard, 1983: 14).*

En el pueblo, los Casares son personas clasificadas como aquellas con las que es mejor no enfrentarse. En un entredicho pueden apelar a un último recurso "no se metan conmigo que yo estoy dispuesto a todo". El contrincante corre el riesgo de ganarse a un enemigo de por vida, a una persona que puede llevar el enfrentamiento hasta las últimas consecuencias (esto es a la utilización de la violencia física contra él o su familia) y que no olvidará la afrenta.

No es extraño entonces que una de las afirmaciones encontrada varias veces en diversas conversaciones con distintas personas sea que *.<sup>este</sup> es un pueblo tranquilo, aquí no pasa nada siempre que uno no tenga un enemigo". ¿Quién es un enemigo? Un enemigo es alguien que está siempre buscándolo a uno, que uno no puede quedar tranquilo porque sabe que en algún momento le va a hacer el mal".*

*¿Cómo se construye un enemigo? Cuando uno está tomando o durante un trabajo uno puede decir cosas que a otro no le gustan y hay personas muy rencoristas (sic) que no*

olvidan y no perdonan", otras veces "se pelea por dinero o por tierra o por mujeres". Un enemigo rompe las reglas del juego social asumiendo todas las consecuencias negativas que puedan recaer sobre su persona pero, al mismo tiempo, obligando al otro a vivir en la incertidumbre y el peligro. Él decide por él y por sus enemigos vivir una vida basada en la inseguridad. Por ejemplo, en un pleito por tierras el enemigo abandonará la discusión jurídica y tratará de decidir el pleito amedrentando o agrediendo físicamente al adversario o a su familia.

El siguiente relato, que narra la serie de venganzas de sangre producidas en el año 1995 entre las familias Casares y Ramallo, no constituye el único caso de violencia en el que la familia analizada se viera envuelta. Hace unos cuatro años un hermano del padre de los Casares fue muerto a balazos en una tienda sin que nadie pudiese informarme de los motivos del caso.

La familia Casares desde hacía un poco menos de un año, alquilaba una casa a la familia Ramallo. En ese entonces, su relación con la familia Ramallo era considerada buena y frecuentemente se los veía bebiendo juntos, que es una de las formas en las que en la comunidad analizada se expresa la amistad. Parece ser que los Casares se retrasaron unos meses en el pago del alquiler. La madre de los Ramallo les exigió el pago en una fuerte discusión que terminó en amenazas e insultos. En ese momento, la madre de los Ramallo entregó a uno de sus hijos un cuchillo de carnicero y lo incitó a matar al padre de los Casares (las mujeres en este y otros casos analizados de venganzas de sangre en la comunidad excepcionalmente actúan directamente en las acciones violentas aunque sí están presentes incitando a los hombres a la acción y avivando las llamas del conflicto). En vez de morir el padre de los Casares, su hijo menor, que estaba presente en la discusión, se interpuso y murió acuchillado.

Como en ese entonces estaba presente el ejército, unos soldados pusieron presos al asesino y a su madre aunque esta fue liberada poco tiempo después. Un Ramallo, que se encontraba preso al momento del hecho y que era, además, muy temido en el pueblo, fue liberado, volvió a su casa y se encontró con esta situación. Acusó a los Casares por lo sucedido y a los quince días, una noche, unos enmascarados entraron a la casa de los Casares y asesinaron al padre a balazos.

Unos seis meses después el padre de los Ramallo debió hacerse presente en el pueblo para pagar ciertos impuestos, lo acompañaba su hija de más de treinta años, una nieta de seis y su mujer. Mientras su mujer se quedaba esperando en la tesorería

su marido, junto a su hija y su nieta iban a buscar unas boletas de impuestos que habían quedado en la casa que antes alquilaban los Casares. Para buscar esas boletas debían necesariamente pasar por el frente de la carnicería de sus enemigos. Al pasar por allí los Casares se trenzaron en una feroz discusión en la que se intercambiaron todo tipo de insultos (interesante el hecho de que junto a insultos que se referían a la deshonestidad de sus madres el epíteto "perros" haya sido recordado por las personas que presenciaron el inicio de la discusión). Cuando volvieron a pasar un Casares los esperaba con una pistola con la que mató de varios disparos a los dos mayores, una vez que verificó que ambos estaban muertos guardó su pistola debajo de su cinturón y se alejó caminando. En esta ocasión una mujer resultó víctima de la agresión, no es lo común pero se trataría de la aplicación estricta de la ley del talión. Varios entrevistados piensan que si los Ramallo no hubiesen asesinado al hijo menor de los Casares esto no hubiese sucedido. En todo caso al agresor no se le ocurrió matar a la nieta de seis años.

## 5. Los ritos fúnebres

*"The final ceremony, however, involves the reassertion of society manifested by the end of mourning and by the belief that the soul has been incorporated into the society of the dead and has settled down in the same way as the collective consciousness of the living has been resettled by the funerary rituals" (Maurice Bloch y Jonathan Parry, 1989: 4).*

En el entierro de don Narciso Casares se pueden descubrir elementos propios de la muerte de un "pater familiae".<sup>a</sup>asesinado como consecuencia de una serie de venganzas entre la familia Casares y los Ramallo. En la iglesia parroquial se realizó una misa por el difunto a la que asistieron unas doscientas- doscientos cincuenta personas, muchas más esperaron afuera. Fue esta la ceremonia católica propiamente dicha ya que solo aquí participó el sacerdote y la liturgia se ciñó a la ortodoxia. El sacerdote bendijo los restos a la salida del templo y con esto dio por finalizada su participación en el ritual. Luego de la misa y la bendición una multitud de unas ochocientas personas acompañó el paso del féretro hasta el cementerio. Porque la expresión de solidaridad fue tan masiva? He desarrollado en otro artículo para esta misma revista la descripción de los entierros masivos de aquellos hombres muertos violentamente asociados al culto heroico para la comunidad (Álvarez, 2001:40-43). Debemos tener en cuenta además que en este caso Nómeque consideró a los Ramallo como culpables, debiendo sus

parientes más lejanos justificar defensivamente sus opiniones. Por otra parte hubo una profunda expresión de solidaridad entre los matarifes y carniceros del área quienes solventaron parte de los gastos del entierro y especialmente pagaron a los músicos.

Durante el trayecto hacia el cementerio una banda de mariachis interpretó diversas canciones, entre ellas "Nadie es eterno en el mundo" "La cruz de madera", canciones que hacen referencia a la muerte, a la separación que conlleva y a la necesidad de mantener vivos el recuerdo y el culto del ser querido desaparecido.

La costumbre de llevar mariachis a los funerales no es considerada tradicional. La introducción de estos músicos (que en esta comunidad se da solo en los entierros de varones) es de reciente data, para la mayoría esta introducción es de claro origen "narco"(narcotraficante) y reservado, de acuerdo con mis informantes, a hombres que se destacaron en la comunidad o que murieron trágicamente. Los mariachis han sido prohibidos recientemente por la Iglesia Católica en las ceremonias fúnebres de la región. Según se rumorea, esta prohibición se debería a expresa disposición del obispo que consideraría como "propia de paganos.<sup>a</sup> esta costumbre.

Al llegar hasta la tumba y mientras el sepulturero y algunos parientes que colaboraban trataban de terminar de cavar la fosa, los mariachis volvieron a entonar las mismas canciones. Luego el trompetista tocó la diana y se hizo un minuto de silencio que fue interrumpido por un alud de sollozos mayoritariamente femeninos que acompañaron, ruidosamente, la colocación del féretro en la fosa.

Posteriormente, a la salida del cementerio, los familiares entregaron a los presentes cervezas y gaseosas que fueron bebidas por todos los participantes. Esta última acción podría ser considerada de acuerdo con la clásica opinión de Durkheim (1982) y de Robertson Smith (1889) como de un acto de comensalidad, de una forma de comunión, de una manera simbólica de beberse al muerto.

En ese preciso momento los parientes varones cercanos del muerto se reúnen y cuchichean en voz baja. Uno de ellos dice: "Esto no puede quedar así, es la segunda muerte en un mes".

El momento de mayor solidaridad social coincide también con el momento utilizado para planear la posible vendetta. Dos de los momentos en los que la unidad familiar, en los que la existencia misma de la familia como tal se expresa son la venganza de sangre y los ritos fúnebres. Para Durkheim (1982) la vendetta y los ritos de duelo

están directamente relacionados. El sentimiento de dolor colectivo llevado en algunos casos al paroxismo produce la necesidad de expulsar externamente estos sentimientos a través de la venganza. Ambos formarían parte del mismo mecanismo social, una desgracia común reaviva los sentimientos colectivos. La venganza expresa estos sentimientos hacia afuera, en ambos casos los lazos sociales se refuerzan. La observación detenida de la familia Casares parecería estar de acuerdo con esta última afirmación. Las series de vendettas aunaron a la familia que prácticamente se mueve sólo colectivamente ante el temor de una nueva revancha, por otra parte se mira afuera del círculo familiar con recelo y desconfianza. Jamás se ve a un Casares que no vaya acompañado por otro. Generalmente van de a tres, dos adelante y uno atrás, todos armados. Esta situación que es la lógica consecuencia de la posible vendetta que pende sobre sus cabezas los lleva objetivamente a la unión familiar y a su separación del resto de la comunidad.

## 6. Discusiones y conclusiones

El análisis de la conflictividad interna nos obliga a acentuar los aspectos culturales de la violencia que subyacen detrás de condiciones estructurales. Un análisis de la violencia que impera en la comunidad basado únicamente en la perspectiva de un conflicto de clases sería insuficiente para producir una explicación razonable sobre la existencia de la violencia interna. Si tomamos en cuenta la ya referida actitud de los Casares de .estar dispuestos a todo no temer en llevar un conflicto hasta sus últimas consecuencias, desde una perspectiva de análisis estructural podríamos definir a esta actitud como de un mecanismo defensivo de clase. Este mecanismo estaría producido por aquellos que, siguiendo las reglas sociales y legales, serían perdedores en cualquier entredicho en el que se vieran enfrentados con personas con mayores influencias, mayores relaciones o simplemente mayor educación formal quienes terminarían imponiendo su parecer. Ante esta inferioridad de condiciones el .estar dispuestos a todo"les permitiría romper el juego social y amenazar a sus contrincantes con la arriesgándolo todo incluso sus vidas y las de sus familias. Su actitud sería análoga a lo que James Scott llama las .armas de los débiles". Para Scott (1985) las armas de los débiles constituyen formas veladas de lucha de clases mediante la cual un sector social en inferioridad de condiciones frente a los demás (especialmente en su relación con el poder económico y con el poder político) apela como forma de

resistencia. Pero curiosamente, en la comunidad analizada, este mecanismo termina imponiéndose en enfrentamientos entre personas de las mismas o muy parecidas condiciones socio- económicas como es el caso arriba mencionado entre las familias Casares y Ramallo. Es decir que si la actitud referida fuese un mecanismo defensivo de clase, porque las principales víctimas terminan encontrándose en el mismo sector social? A mi entender una explicación que tome en cuenta un análisis detenido de la agregación y desagregación social de la comunidad analizada y que al mismo tiempo prevea la autonomía de los elementos culturales es necesaria para explicar cómo un mecanismo social defensivo pueda convertirse en un factor de conflictos internos.

El análisis de la cultura del lugar me hace pensar en la existencia de un sentido de pertenencia familiar que se identifica a sí mismo "frente al mundo". Al interior de la familia Casares se ha percibido un fuerte sentido de pertenencia y solidaridad que contrasta con una red de relaciones externas débil manejada con una profunda desconfianza. En este sentido la relación ya referida que Durkheim establece en "Las formas elementales de la vida religiosa.<sup>en</sup>tre muerte- ritos fúnebres y vendetta, como un ciclo que acentúa la solidaridad del grupo social y marca sus diferencias con el exterior parece ser adecuada para describir la agregación social de la familia Casares. Sin embargo, muestras de solidaridad colectiva de toda la comunidad, como la masiva participación en los funerales del jefe de la familia hablan de la complejidad del tema. Es evidente que para que exista una comunidad deben existir lazos de solidaridad que aseguren la relación entre sus miembros. Ahora bien, es también evidente que nos encontramos con una comunidad dividida y enfrentada internamente en una situación que René Girard definiría como de crisis sacrificial.<sup>o</sup> de violencia generalizada.

Para René Girard el sacrificio, fundamento para él de todo sistema religioso, consistiría en una violencia en principio real para luego convertirse en simbólica. Primero el sacrificio consistía en entregar en holocausto a un ser humano, luego este fue reemplazado por un animal al que se le atribuían características de humanidad, por ejemplo, el cordero y finalmente se transformaba en un sacrificio meramente simbólico, por ejemplo, la misa. Este sacrificio consistía en un acto que canalizaba los impulsos violentos de la comunidad hacia una víctima determinada (chivo expiatorio). La religión entonces a través del sacrificio estaría exorcizando a la comunidad de la violencia. La venganza de sangre, que para Girard sería un acto propio de pueblos llamados primitivos (incluirá en esta categoría al sur de Italia?) es evitada al concentrarse la

violencia en una víctima determinada. Cuando una comunidad no puede canalizar esos impulsos violentos nos encontraríamos en una "crisis sacrificial". ¿Sería este el caso de la comunidad analizada? Veamos, en primer lugar, que la separación que Girard realiza entre primitivos y modernos le permite obviar la existencia de sociedades campesinas bajo el impacto de la modernización (Calabria, Sicilia, el altiplano Cundiboyacense) en las que la violencia interna es característica y, lo que es más grave aún, obviar la discusión sobre la función de la religión en estas sociedades. En efecto, en las citadas sociedades la "crisis sacrificial" convive con la celebración constante del sacrificio de la misa. Girard dice que el sacrificio no elimina completamente la violencia pero la reduce a límites "tolerables" un adjetivo que, más que iluminar una solución al problema, no hace más que oscurecerlo. Necesitaríamos tal vez diseñar un índice de tolerabilidad? Por lo tanto, si bien nos encontramos con una comunidad "enferma" en su agregación social creemos que la "crisis sacrificial" no es una explicación plausible para comprender los fenómenos mencionados en la particular realidad analizada. Sería necesario analizar con más detenimiento la profunda complejidad de los fenómenos religiosos que no se agotan en el sacrificio y, en particular, las formas populares de religiosidad presentes en la religión católica tal cual se desarrolla en Colombia y donde es posible rezar para matar", es decir orar solicitando a la divinidad las fuerzas necesarias para la destrucción del enemigo (tal cual me fue puntualmente referido). El material analizado nos sugiere que la solidaridad de la comunidad está seriamente amenazada por la existencia de agregaciones sociales internas que se identifican en contraste y oposición con otros grupos sociales. La religión católica (ahora en competencia con una miríada de grupos protestantes) fue y es incapaz de producir una religiosidad que canalice socialmente los impulsos violentos de la comunidad.

¿Existe o no una "cultura de la muerte" como elemento autónomo que incida directamente en el comportamiento de la comunidad? (como lo plantea para Colombia Carlos Uribe en su artículo "Nuestra cultura de la muerte"). La exposición a imágenes sumamente violentas en el contacto con los animales a las que habría que agregar imágenes de muerte y violencia entre personas a las que desde pequeñas se ven expuestas, la realización de ciertas labores que implican un constante ejercicio de la violencia sobre los animales, influyen en el comportamiento embruteciendo o animalizando a las personas? Estas actitudes se reflejan en las intuiciones de Echeverría y en el trabajo de María Victoria Uribe oportunamente citado (María Victoria Uribe 1978). Ambos casos no tratan tanto de la "animalización" del victimario sino de la animalización de la

víctima, de un proceso de "naturalización" mediante el cual la víctima es deshumanizada integrándola a la naturaleza. Sin haber encontrado elementos de animalización tan contundentes como los cortes realizados a las víctimas durante el período de "la violencia" nos encontramos con que los miembros varones de la familia analizada trabajan en un ambiente de contacto agresivo con el mundo animal, la carnicería y especialmente el matadero del pueblo, y que tenían como principal diversión la asistencia a peleas de gallos. Como hemos visto los Casares consideraban que los Ramallo debían morir como perros".

Si bien estos elementos, por sí solos son insuficientes para establecer algún tipo de causación directa entre la violencia ejercida sobre animales y el asesinato de seres humanos sí me atrevería a afirmar que el ambiente cultural de las personas estudiadas tanto en sus actividades laborales como recreativas está impregnado de una cultura violenta y tanática (la presencia de los mariachis, más el tipo de canciones interpretadas, canciones, por otra parte, que no son escuchadas solamente en el cementerio sino que forman parte del repertorio local y cuyo análisis pormenorizado implicaría otro artículo, serían otros elementos para tener en cuenta). Otro aspecto no menos importante de esta cultura es terminar considerando a la muerte violenta en el campo de la cotidianeidad. Al fin y al cabo matar o morir asesinado es algo que siempre sucedió y posiblemente sucederá en este pueblo.

Esta investigación considera entonces que en el grupo familiar analizado el estudio de los elementos culturales es útil y necesario para intentar comprender un mecanismo social, la venganza de sangre, que se desarrolla entre personas de la misma clase social y que puede llegar a la aniquilación de las familias enfrentadas. La familia en la cultura local constituye una agregación social que puede estar en contraste violento con la solidaridad comunitaria. La religión es incapaz de canalizar la violencia de la comunidad aunque, como ya fue expresado, no parece correcto definir esta situación como de "crisis sacrificial". La violencia está presente en la cultura de la familia y de la comunidad analizada tanto en la agresiva relación con la naturaleza como en la cotidianeidad del homicidio, como dijo un campesino: "Y si no ¿para qué lleva pistola la gente?"

## Referencias

Adams, R. N. y Adams (Eds.). (1977). *The anthropology of power: Ethnographic studies from Asia, Oceania, and the New World*. New York: Academic Press.

Adas, M. (2012). *Prophets of rebellion: Millenarian protest movements against the European colonial order*. UNC Press Books.

Álvarez, S. (2020). *Leviatã e seus lobos. Violência e poder numa comunidade dos Andes colombianos*, Editora Unijuí, Universidade Regional do Noroeste do Estado do Rio Grande do Sul, Ijuí.

Álvarez, S. (2017). Reconceptualizando la violencia y la criminalidad: la tensión entre conceptos locales y universales. *Ensamblajes en sociedad, política y cultura* 1 (7).

Álvarez, S. (2004).. Entre galos e matadouros: violência, vida social e família em uma comunidade camponesa do Sumapaz. *Mana*, 10, 7-24.

Álvarez, S. (2001).. Enterrando heróis, patriarcas, suicidas e traidores: solidariedade e ostracismo nos Andes colombianos. *Mana*, 7, 35-55.

Álvarez, S. (2010). El “gallinazo” en la escuela. Violencia doméstica y construcción social de la masculinidad al pie del Páramo de Sumapaz. *Antípoda. Revista de Antropología y Arqueología*, (10), 141-155.

Barmash, P. (2004). Blood feud and state control: Differing legal institutions for the remedy of homicide during the second and first millennia BCE. *Journal of Near Eastern Studies*, 63(3), 183-199.

Bejarano, J. A. (1983). Campesinado, luchas agrarias e historia social: notas para un balance historiográfico. *Anuario colombiano de historia social y de la cultura*, (11), 251-304.

Black-Michaud, J. (1975). *Feuding societies*. Basil Blackwell, Oxford, 1975.

Bloch, M., & Parry, J. (Eds.). (1982). *Death and the Regeneration of Life*. Cambridge University Press.

Durkheim, É. (1982). *Las formas elementales de la vida religiosa* (Vol. 38). Ediciones Akal.

Echeverría, E. (1993). *El Matadero [1871]*. Madrid: Alianza.

Fafchamps, M. (1992). Solidarity networks in preindustrial societies: Rational peasants with a moral economy. *Economic development and cultural change*, 41(1), 147-174.

García Márquez, G. (1972) *El Coronel no tiene quien le escriba*. Losada: Buenos Aires.

Geertz, C. (1997). Juego profundo: notas sobre la riña de gallos en Bali. *La interpretación de las culturas*, 339-372.

Girard, R. (1983). *La violencia y lo sagrado* (1972). Editorial Anagrama, Barcelona, 152-153.

- Girard, R. (1982). *El misterio de nuestro mundo: Clave para una interpretación antropológica: Diálogos con Jean-Michel Oughourlian y Guy Lefort*. Hermeneia, 1925.
- Hobbes, T. (1975). *Leviathan*. Everyman's Library. New York: Dutton,[1651].
- Hobsbawm, E. (1983). *Rebeldes primitivos*. Barcelona: Ariel.
- Isaacman, A. (1985). Chiefs, rural differentiation and peasant protest: the Mozambican forced cotton regime 1938-1961. *African Economic History*, (14), 15-56.
- Jaramillo, J. E. (1988). *Estado, sociedad y campesinos*.
- Londoño, R. (1993). *Los nuevos hacendados del Sumapaz*.
- Marulanda, E. (1991). *Colonización y conflicto: las lecciones del Sumapaz*. Tercer Mundo.
- Maullin, R. L. (1971). *Soldiers, Guerrillas and Politics in Colombia*. Rand Corp: Santa Monica, CA.
- Marvin, G. (1986). Honour, integrity and the problem of violence in the Spanish bullfight. *The anthropology of violence*, 118-35.
- Robertson Smith, W. (1889). *Lectures on the Religion of the Semites*. First Series: The fundamental institutions. NY: D. Appleton & Co.
- Riches, D. (1986). *The Phenomenon of Violence* (David Riches Ed. *The Anthropology of Violence*, New York, Basil Blackwell).
- Scott, J. C. (1985). *Weapons of the weak: Everyday forms of peasant resistance*. Yale University Press.
- Standish, K. (2014). Understanding cultural violence and gender: honour killings; dowry murder; the zina ordinance and blood-feuds. *Journal of Gender Studies*, 23(2), 111-124.
- Steiner, C. (1993). Urabá: de región de frontera a región de conflicto. *Conflicto social y violencia: notas para una discusión*, 63-71.
- Stern, S. J. (1990). *Resistencia, rebelión y conciencia campesina en los Andes*. Instituto de Estudios Peruanos.
- Uribe, C. A. (1988). Nuestra cultura de la muerte. *Texto y contexto*, 13, 53-67.
- Uribe, M. V. (1990). Matar, rematar y contramatar: las masacres de la violencia en el Tolima, 1948-1964 (No. 159-160). Cinep.
- Uribe, M. V. (1992). *Limpiar la tierra: Guerra y poder entre esmeralderos*. Cinep.
- Wolf, E. R. (1966). Peasants. *Foundations of Modern Anthropology Series*. New York: Prentice-Hall, 3(0000), 056.
- Wolf, E. R. (1994). Facing power: Old insights, new questions. I Borofsky, R.(Red.) *Assessing cultural anthropology*.

Wolf, E. R. (1999). *Peasant wars of the twentieth century*. University of Oklahoma Press.

# ARTÍCULOS DE REVISIÓN

---

Conocimiento i Política – No. 7, 2026

# Big Data e Inteligencia Artificial como medios digitales en la política pública: adopción, práctica y barreras

## Big Data and Artificial Intelligence as Digital Media in Public Policy: Models of Adoption, Practice, and Barriers

José David Vallejo Manzur<sup>1</sup>

<sup>1</sup> El Colegio de Tamaulipas.  
vallejocfh@hotmail.com

Erick Leobardo Alvarez-Aros<sup>2</sup>

<sup>2</sup> Universidad Católica de El Salvador.  
erick.alvarez@catolica.edu.sv (Autor de correspondencia)

**Resumen.** El presente estudio analiza el estado del arte en el uso de Big Data e inteligencia artificial (IA) en la prospectiva de políticas públicas, destacando su relevancia como herramientas clave para anticipar escenarios futuros, optimizar recursos y diseñar soluciones adaptativas en contextos gubernamentales. El objetivo principal es comprender cómo estas tecnologías están transformando la capacidad de los gobiernos para abordar problemas complejos de manera estratégica y eficiente. Mediante una metodología basada en una exhaustiva revisión de literatura, se identificaron modelos de adopción tecnológica, aplicaciones prácticas y barreras en su implementación. Los modelos analizados, como el Modelo de Difusión de Innovaciones, el Modelo de Madurez Tecnológica y el Modelo UTAUT, ofrecen perspectivas sobre los factores que determinan la integración de estas tecnologías en la administración pública. Además, se documentaron aplicaciones prácticas en ámbitos como la movilidad urbana, la salud pública y la sostenibilidad ambiental, con casos destacados, donde estas herramientas han permitido diseñar políticas más precisas y efectivas. Entre las barreras identificadas destacan la falta de interoperabilidad tecnológica, la carencia de personal capacitado y los desafíos en la infraestructura tecnológica. Las conclusiones subrayan la necesidad de fomentar la interoperabilidad entre sistemas, fortalecer la capacitación en el sector público y explorar nuevos contextos de aplicación. Este análisis aporta una base sólida para el diseño de estrategias tecnológicas que fortalezcan la capacidad gubernamental en el siglo XXI.

**Palabras clave:** Políticas públicas, prospectiva, Big Data, Inteligencia artificial, adopción tecnológica y gobierno digital.

**Abstract.** This study analyzes the state of the art in the use of Big Data and artificial intelligence (AI) in public policy foresight, highlighting their relevance as key tools to anticipate future scenarios, optimize resources and design adaptive solutions in governmental contexts. The main objective is to understand how these technologies are transforming the ability of governments to address complex problems strategically and efficiently. Using a methodology based on an exhaustive literature review, models of technology adoption, practical applications and barriers to implementation were identified. The models analyzed, such as the Diffusion of Innovations Model, the Technological Maturity Model and the UTAUT Model, offer insights into the factors that determine the integration of these technologies in public administration. In addition, practical applications were documented in areas such as urban mobility, public health and environmental sustainability, with outstanding cases, where these tools have made it possible to design more precise and effective policies. Among the barriers identified were the lack of technological interoperability, the lack of trained personnel and challenges in the technological infrastructure. The conclusions highlight the need to promote interoperability between systems, strengthen training in the public sector and explore new application contexts. This analysis provides a solid basis for the design of technological strategies to strengthen government capacity in the 21st century.

**Keywords:** Public policies, foresight, Big Data, Artificial Intelligence, technological adoption and digital government.

**Historial:** Recibido: 9 de enero de 2025 | Aceptado: 18 de mayo de 2026

**Ética/Financiamiento:** Los autores declaran no tener conflictos de interés. La investigación fue autofinanciada y no recibió financiamiento externo.

Este documento está bajo una licencia  
CC BY-NC-ND 4.0.

**Cómo citar (APA):** Vallejo Manzur, J. D., & Alvarez-Aros, E. L. (2026). Big Data e Inteligencia Artificial como medios digitales en la política pública: adopción, práctica y barreras. *Conocimiento i Política*, 7(1), 244-271. DOI: 10.64480/cip.2026.n7.08.

## 1. Introducción

En un mundo caracterizado por la rápida transformación tecnológica, la inteligencia artificial (IA) y el Big Data han emergido como herramientas fundamentales para enfrentar los desafíos sociales, económicos y políticos. Estas tecnologías han revolucionado sectores clave como la salud, la movilidad y las campañas electorales, permitiendo una toma de decisiones más eficiente basada en el análisis de grandes volúmenes de datos y patrones complejos. En el ámbito de las políticas públicas, estas herramientas están redefiniendo la forma en que los gobiernos diseñan, implementan y evalúan sus estrategias, marcando un punto de inflexión en la modernización gubernamental.

El Big Data, con su capacidad para procesar volúmenes masivos de información en tiempo real, permite a los tomadores de decisiones identificar tendencias sociales, anticipar fenómenos emergentes y personalizar políticas en función de las necesidades ciudadanas (Mayer-Schönberger & Cukier, 2013). La IA, por su parte, complementa este análisis al ofrecer modelos predictivos avanzados, simulaciones y automatización en la gestión de problemas complejos, desde la asignación de recursos hasta la planificación urbana (Goodfellow et al., 2016). Este enfoque optimiza la eficiencia gubernamental y abre nuevas posibilidades para abordar problemas sistémicos con un enfoque prospectivo.

En campañas electorales, estas tecnologías han transformado la forma en que los partidos y candidatos comprenden y analizan al electorado. Herramientas basadas en Big Data recopilan información de redes sociales, encuestas y patrones de consumo, mientras que la IA segmenta a los votantes según preferencias políticas, valores y comportamientos. Estas herramientas permiten diseñar mensajes personalizados que resuenan con segmentos específicos del electorado, maximizando la efectividad de las estrategias de comunicación política (Tucker et al., 2017; Munger & Phillips, 2021). Además, los algoritmos predictivos se emplean para anticipar resultados electorales y modelar escenarios, lo que refuerza la toma de decisiones estratégicas en tiempo real (Issenberg, 2012).

Por su parte, la relevancia de estas tecnologías en las políticas públicas se magnifica al observar sus impactos en otros sectores. En salud pública, la IA ha sido clave para rastrear brotes epidemiológicos, mientras que el Big Data ha permitido diseñar

respuestas estratégicas frente a pandemias como la del COVID-19 (Choi & Varian, 2012). En el ámbito financiero, estas herramientas se han utilizado para detectar fraudes y optimizar recursos, logrando un nivel de precisión que transforma la gestión organizacional (Hardoon & Shmueli, 2013). Estas experiencias destacan el potencial de estas tecnologías para resolver problemas complejos y proporcionar soluciones adaptativas en diversos contextos.

Sin embargo, la adopción de Big Data e IA en políticas públicas presenta desafíos significativos. Problemas relacionados con la privacidad, los sesgos algorítmicos y la interoperabilidad de sistemas limitan su implementación efectiva, particularmente en países con infraestructuras tecnológicas insuficientes (Mittelstadt et al., 2016; OECD, 2019). A pesar de ello, la oportunidad de transformar la gobernanza y mejorar la calidad de vida de los ciudadanos a través de estas tecnologías sigue siendo inmensa, destacando la necesidad de investigaciones más profundas y marcos regulatorios sólidos.

En este artículo, se analiza el estado del arte en el uso de Big Data e IA en la prospectiva de políticas públicas. A través de un enfoque sistémico basado en una revisión exhaustiva de literatura, se identifican modelos de adopción tecnológica, aplicaciones prácticas y las principales barreras que enfrentan los gobiernos. Los hallazgos buscan contribuir a la discusión científica sobre cómo estas tecnologías pueden ser adoptadas de manera ética y efectiva, ofreciendo soluciones innovadoras para los complejos desafíos del siglo XXI.

## **2. Marco teórico**

### **2.1. Conceptualización de Big Data e Inteligencia Artificial en políticas públicas**

En el contexto de las políticas públicas, Big Data se refiere a la recopilación, almacenamiento, análisis y utilización de volúmenes masivos de datos estructurados y no estructurados provenientes de diversas fuentes, como registros administrativos, sensores IoT (Internet de las cosas), redes sociales y plataformas digitales (Mayer-Schönberger & Cukier, 2013; Kitchin, 2014). La característica distintiva del Big Data radica en sus tres dimensiones principales: volumen, velocidad y variedad (Laney, 2001), lo que permite a los gobiernos procesar información en tiempo real para identi-

ficar patrones y tendencias que guíen decisiones informadas. Además, el Big Data facilita el análisis predictivo, permitiendo anticipar fenómenos sociales complejos como el desempleo, la inseguridad alimentaria o los desplazamientos poblacionales, lo cual es crucial para diseñar políticas públicas más adaptativas y eficaces (Manyika et al., 2011).

Por otro lado, la inteligencia artificial (IA) representa una revolución en la capacidad de los gobiernos para abordar problemas complejos y emergentes. A través del uso de algoritmos de aprendizaje automático, redes neuronales profundas y sistemas expertos, la IA permite analizar grandes volúmenes de datos y tomar decisiones automatizadas basadas en patrones detectados (Goodfellow, Bengio, & Courville, 2016). Su aplicación en la administración pública incluye desde la personalización de servicios gubernamentales hasta la asignación eficiente de recursos en áreas críticas como salud, educación y transporte (Das & Zhang, 2019).

La integración de Big Data e IA en las políticas públicas representa un cambio paradigmático en la gobernanza. Mientras que el Big Data proporciona las herramientas para interpretar fenómenos a gran escala, la IA añade la capacidad de automatizar procesos, generar predicciones más precisas y diseñar políticas adaptadas a contextos dinámicos. Sin embargo, esta convergencia tecnológica plantea también desafíos éticos y sociales significativos, como la privacidad de los datos y el riesgo de sesgos en los algoritmos, lo que exige un marco regulatorio robusto y una supervisión transparente en su implementación (Zicari, 2014; Mittelstadt et al., 2016).

En definitiva, el Big Data y la IA amplifican las capacidades tradicionales de la administración pública y transforman radicalmente la forma en que los gobiernos entienden y responden a las necesidades ciudadanas. Este cambio hacia un enfoque basado en datos y automatización redefine las dinámicas entre los tomadores de decisiones y los ciudadanos, consolidando su relevancia como pilares de la modernización gubernamental.

## **2.2. Modelos de adopción tecnológica en políticas públicas**

La adopción de tecnologías avanzadas como el Big Data y la inteligencia artificial (IA) en el ámbito gubernamental requiere un marco conceptual robusto que explique las dinámicas y los factores que influyen en su integración. Entre los modelos teóricos

más influyentes se encuentra el Modelo de Difusión de Innovaciones de Rogers (2003), que destaca cinco atributos clave que determinan la velocidad y el grado de adopción tecnológica: la ventaja relativa, la compatibilidad, la complejidad, la posibilidad de prueba y la observabilidad. En el contexto de las políticas públicas, este modelo ayuda a comprender cómo los beneficios percibidos, como la mejora en la eficiencia y la capacidad de predicción, influyen en la disposición de los gobiernos para implementar Big Data e IA (Venkatesh, Thong, & Xu, 2012).

Además, la compatibilidad con los valores y estructuras existentes en las organizaciones públicas juega un papel crucial. Los gobiernos, a menudo caracterizados por estructuras jerárquicas y procesos burocráticos, enfrentan barreras significativas para integrar tecnologías disruptivas. Por ejemplo, mientras que la ventaja relativa del Big Data e IA radica en su capacidad para procesar datos en tiempo real y tomar decisiones automatizadas, la resistencia organizacional y la falta de infraestructura tecnológica pueden retrasar su adopción (Rogers, 2003).

Otro modelo relevante es el Modelo de Madurez Tecnológica de Gartner (2019), que clasifica a las organizaciones según su nivel de integración tecnológica en cinco etapas: inicial, consciente, definida, gestionada y optimizada. Este modelo ofrece un marco para evaluar cómo las instituciones públicas progresan desde la experimentación con Big Data e IA hasta su integración estratégica y sostenible. Por ejemplo, en la etapa inicial, los gobiernos suelen implementar tecnologías como proyectos piloto para evaluar su viabilidad, mientras que, en las etapas más avanzadas, estas tecnologías están profundamente integradas en procesos decisionales clave, como la asignación presupuestaria y la gestión de crisis (Gartner, 2019).

Además, el Modelo de Aceptación Tecnológica (TAM) de Davis (1989) y su evolución hacia el Modelo Unificado de Aceptación y Uso de la Tecnología (UTAUT) de Venkatesh et al. (2003) también son aplicables en el análisis de la adopción de Big Data e IA en políticas públicas. Estos modelos enfatizan que factores como la utilidad percibida, la facilidad de uso y la intención de uso son determinantes críticos en la aceptación de tecnologías por parte de los usuarios finales, incluidos los funcionarios gubernamentales. En el contexto del gobierno, la utilidad percibida puede estar relacionada con la capacidad de estas tecnologías para resolver problemas complejos, mientras que la facilidad de uso se traduce en la disponibilidad de capacitación y soporte técnico para los empleados públicos (Venkatesh et al., 2003).

### **2.3. Integración en el contexto gubernamental**

La aplicación de modelos teóricos para la adopción de Big Data e IA en el ámbito gubernamental ha llevado a la implementación de estrategias específicas que fomentan la integración de estas tecnologías. Entre las estrategias destacadas se encuentran la creación de laboratorios de innovación pública, la implementación de marcos regulatorios flexibles y el desarrollo de asociaciones público-privadas. Por ejemplo, países como Estonia y Singapur han logrado avances significativos al integrar Big Data e IA en sus procesos gubernamentales mediante políticas proactivas y una visión estratégica hacia la transformación digital (Manyika et al., 2011; OECD, 2019). Estonia, conocida como una “sociedad digital”, ha digitalizado casi todos sus servicios gubernamentales, lo que permite un acceso eficiente y transparente para los ciudadanos (Toots et al., 2017).

El Modelo de Madurez Tecnológica de Gartner es especialmente relevante en este contexto, ya que ayuda a evaluar cómo las organizaciones públicas progresan desde la experimentación inicial hasta la integración total de tecnologías avanzadas. En Singapur, por ejemplo, la etapa de “optimización” se ha alcanzado en sectores como la movilidad urbana y la gestión de recursos hídricos, donde las plataformas digitales y el análisis de datos en tiempo real han mejorado sustancialmente la eficiencia de los servicios públicos (Batty, 2013).

Sin embargo, esta adopción enfrenta desafíos importantes, como la falta de interoperabilidad entre sistemas gubernamentales y la resistencia cultural al cambio organizacional (Venkatesh et al., 2003). Además, la formación y capacitación de los empleados públicos en estas tecnologías emergentes sigue siendo limitada, lo que subraya la necesidad de estrategias de adaptación organizacional. Para abordar estos desafíos, la cooperación internacional y la estandarización de marcos regulatorios podrían desempeñar un papel crucial en la adopción sostenible de Big Data e IA en el sector público (Mittelstadt et al., 2016).

### **2.4. Uso de la Inteligencia Artificial y Big Data en políticas públicas basadas con base en el análisis prospectivo del comportamiento ciudadano**

La prospectiva en políticas públicas, entendida como el diseño y análisis de estrategias anticipatorias para abordar problemas futuros, ha encontrado en el Big Data y

la inteligencia artificial (IA) herramientas fundamentales para su desarrollo. Estas tecnologías permiten a los gobiernos adoptar un enfoque proactivo, superando la mera reactividad al facilitar la identificación de tendencias, la predicción de escenarios y la formulación de intervenciones adaptativas basadas en datos reales sobre el comportamiento ciudadano.

El uso de Big Data e IA en la prospectiva también promueve la inclusión social al identificar y abordar las necesidades de poblaciones vulnerables. Por ejemplo, el análisis de datos demográficos y de consumo energético en Dinamarca permitió identificar comunidades en riesgo de pobreza energética, implementando subsidios específicos que optimizaron los recursos públicos (Hardoon & Shmueli, 2013). Asimismo, en Estados Unidos, el uso de Big Data en los programas de salud pública ayudó a anticipar y mitigar brotes epidemiológicos, optimizando la distribución de recursos médicos durante la pandemia de COVID-19 (Choi & Varian, 2012; Manyika et al., 2011).

#### **2.4.1. El rol del Big Data en la prospectiva**

El Big Data contribuye de manera significativa a la prospectiva al recopilar y procesar grandes volúmenes de información provenientes de diversas fuentes, como redes sociales, sensores urbanos, datos demográficos y registros administrativos (Manyika et al., 2011; Kitchin, 2014). Estas capacidades permiten a los gobiernos analizar patrones de comportamiento ciudadano en tiempo real, identificar cambios en las preferencias sociales y anticipar crisis potenciales. Por ejemplo, el uso de Big Data en aplicaciones de movilidad urbana, como las implementadas en Londres, permitió reducir las congestiones en un 30 % al prever patrones de tráfico y ajustar las políticas de transporte en tiempo real (Batty, 2013; OECD, 2019).

Además, el análisis de datos históricos también es crucial para la prospectiva. Herramientas como modelos de regresión y análisis de series temporales permiten a los gobiernos proyectar tendencias sociales, como el envejecimiento poblacional o los cambios en las tasas de empleo, y diseñar políticas públicas adaptadas a estos desafíos (Choi & Varian, 2012). Un caso destacado es el uso de datos de Big Data en Japón para anticipar demandas en el sistema de salud y ajustar las políticas de pensiones frente a la transición demográfica acelerada (OECD, 2019; Batty, 2013).

La integración de datos refleja la realidad actual y también se convierte en una herramienta predictiva que orienta la planificación estratégica. Por ejemplo, en Brasil, el análisis de imágenes satelitales y datos climáticos permitió anticipar riesgos de deforestación en la Amazonía, facilitando la implementación de políticas ambientales más eficaces (Amaral et al., 2020). Este enfoque también ha sido útil en el sector de infraestructura pública, donde el análisis predictivo ha optimizado la inversión en proyectos clave en economías emergentes (Laney, 2001; Manyika et al., 2011).

#### **2.4.2. La Inteligencia Artificial y la modelización predictiva**

La inteligencia artificial (IA) complementa al Big Data al proporcionar modelos avanzados de análisis predictivo y aprendizaje automático que identifican relaciones complejas entre variables. Herramientas como las redes neuronales, los algoritmos de clustering y los sistemas de recomendación permiten generar escenarios prospectivos con un alto grado de precisión (Goodfellow, Bengio, & Courville, 2016). Estas capacidades han sido aplicadas exitosamente en áreas como la gestión de desastres naturales, donde los sistemas basados en IA pueden anticipar eventos catastróficos y recomendar estrategias de mitigación (Müller et al., 2016).

En el ámbito del comportamiento ciudadano, la IA se utiliza para segmentar a la población según patrones de consumo, preferencias políticas y niveles de participación cívica. Por ejemplo, el uso de algoritmos en redes sociales permite a los gobiernos identificar tendencias de opinión pública y diseñar campañas informativas más efectivas. Esta capacidad es particularmente útil en contextos electorales, donde los modelos predictivos pueden analizar las opiniones de los votantes y personalizar los mensajes de campaña para maximizar el impacto político (Howard & Kollanyi, 2016).

Además, la IA facilita la creación de simulaciones que permiten evaluar el impacto de políticas públicas antes de su implementación, reduciendo riesgos y optimizando recursos (Tegmark, 2017). En el ámbito de la salud, por ejemplo, los modelos predictivos basados en IA han sido empleados para analizar patrones epidemiológicos y anticipar la propagación de enfermedades infecciosas, como se observó durante la pandemia de COVID-19 (Choi & Varian, 2012). Estas herramientas permiten a los gobiernos asignar recursos de manera más eficiente y responder rápidamente a emergencias sanitarias.

Asimismo, el uso de redes bayesianas en la gestión pública ha demostrado ser efectivo para modelar escenarios de alta incertidumbre, proporcionando un marco robusto para la toma de decisiones en contextos complejos (Pearl, 2009). Estas aplicaciones tanto mejoran la capacidad predictiva de los gobiernos, como también promueven un enfoque más proactivo y basado en evidencia en el diseño de políticas públicas.

### **2.4.3. Intersección de Big Data e IA en la prospectiva basada en el comportamiento**

La combinación de Big Data e inteligencia artificial (IA) en la prospectiva permite a los gobiernos adoptar un enfoque integral en el diseño de políticas públicas. Al analizar el comportamiento ciudadano en tiempo real y proyectar escenarios futuros, estas tecnologías identifican brechas entre las necesidades actuales y las demandas futuras, optimizando la asignación de recursos y las estrategias gubernamentales. Por ejemplo, en el sector de salud pública, la integración de datos epidemiológicos con modelos predictivos basados en IA ha permitido anticipar brotes de enfermedades y planificar la asignación de recursos médicos con mayor eficiencia (Das & Zhang, 2019).

Esta intersección tecnológica también se utiliza para diseñar políticas inclusivas que consideren las necesidades de grupos vulnerables. Por ejemplo, el análisis de datos de consumo energético y el aprendizaje automático han permitido identificar comunidades en riesgo de pobreza energética, lo que ha llevado a la implementación de políticas específicas para garantizar el acceso equitativo a servicios básicos (Hardoon & Shmueli, 2013). De igual manera, los modelos predictivos han sido aplicados para analizar las tasas de desempleo en tiempo real, facilitando la implementación de programas de capacitación laboral en regiones con altas tasas de desocupación (Choi & Varian, 2012).

En el ámbito educativo, la combinación de Big Data e IA ha sido utilizada para identificar factores que afectan el rendimiento estudiantil en sistemas escolares. Por ejemplo, en sistemas educativos como el de Singapur, los análisis basados en IA han ayudado a personalizar la enseñanza, abordando las necesidades individuales de los estudiantes y reduciendo las tasas de abandono escolar (Kitchin, 2014). Además, en el contexto de la movilidad urbana, los datos recolectados de sensores IoT y algoritmos de IA han optimizado las redes de transporte público, reduciendo los tiempos de espera y las emisiones de carbono en ciudades densamente pobladas (Batty, 2013).

## **2.5. Uso de Big Data e Inteligencia Artificial en políticas públicas a nivel mundial**

El uso de Big Data e inteligencia artificial (IA) en el diseño y ejecución de políticas públicas está transformando los procesos de gobernanza en diversas partes del mundo. Estas tecnologías permiten a los gobiernos abordar problemas complejos de manera proactiva, mejorando la eficiencia, la equidad y la sostenibilidad de sus intervenciones. A continuación, se analizan las aplicaciones más destacadas, los beneficios observados y los desafíos que enfrentan las administraciones públicas a nivel global.

### **América del Norte: Innovación y liderazgo en políticas digitales**

Estados Unidos y Canadá lideran el uso de Big Data e IA en políticas públicas, especialmente en áreas como seguridad nacional, salud pública y movilidad urbana. En Estados Unidos, el gobierno ha implementado algoritmos de IA en sistemas judiciales para prever reincidencias criminales, optimizando recursos y reduciendo la carga del sistema penitenciario (Angwin et al., 2016). En el sector de la salud, los Centros para el Control y la Prevención de Enfermedades (CDC) han utilizado Big Data para rastrear y mitigar brotes epidemiológicos, como el caso del COVID-19 (Choi & Varian, 2012). Además, ciudades como Nueva York emplean análisis de Big Data para monitorear delitos en tiempo real y mejorar la asignación de recursos policiales (Rudinet al., 2018).

En Canadá, la estrategia nacional de IA incluye la creación de centros de excelencia en Montreal, Toronto y Edmonton, fomentando la investigación y el desarrollo tecnológico para abordar problemas sociales. Por ejemplo, el gobierno canadiense utiliza Big Data para monitorear el cambio climático, evaluando el impacto de las emisiones de carbono y diseñando políticas de mitigación más efectivas (Manyika et al., 2011). A su vez, la ciudad de Vancouver emplea IA para predecir patrones de tráfico, ayudando a reducir la congestión y optimizar el transporte público (Wang et al., 2020).

### **Europa: Foco en la ética y la gobernanza de datos**

Europa se caracteriza por su enfoque ético y regulatorio en el uso de Big Data e IA. La Unión Europea (UE) ha desarrollado marcos regulatorios como el Reglamento General de Protección de Datos (GDPR), que establece estándares estrictos para el

manejo de datos ciudadanos (Mittelstadt et al., 2016). En términos de aplicación, países como Estonia y Dinamarca están a la vanguardia.

Estonia, conocida como una “sociedad digital”, ha digitalizado casi todos los servicios gubernamentales mediante plataformas de Big Data e IA. Esto incluye el sistema de voto electrónico y la gestión de registros médicos, lo que ha mejorado la eficiencia y la transparencia gubernamental (Toots et al., 2017). Dinamarca, por su parte, utiliza algoritmos de IA en su sistema de bienestar social para identificar familias en riesgo de pobreza, permitiendo intervenciones más rápidas y personalizadas (Hardoon & Shmueli, 2013). Asimismo, Alemania emplea Big Data para la planificación energética, optimizando el uso de recursos renovables y reduciendo la dependencia de combustibles fósiles (Lohr, 2020).

### **Asia: Liderazgo tecnológico en políticas públicas**

Países como China, Japón y Singapur han adoptado Big Data e IA en políticas públicas con un enfoque altamente centralizado. En China, el gobierno utiliza estas tecnologías para implementar un sistema de crédito social, que evalúa el comportamiento de los ciudadanos y las empresas para fomentar el cumplimiento normativo (Kostka, 2019). Aunque este sistema ha generado debates éticos, también muestra cómo las tecnologías pueden integrarse profundamente en las dinámicas sociales.

Japón, enfrentando el envejecimiento poblacional, emplea IA para prever demandas en el sistema de salud y adaptar sus políticas de pensiones. Singapur, por su parte, ha establecido una “ciudad inteligente” mediante el uso de sensores y análisis de Big Data para mejorar la movilidad urbana, la gestión de recursos hídricos y la seguridad pública (Batty, 2013). Adicionalmente, India utiliza Big Data en programas como el Aadhaar, un sistema de identificación biométrica que facilita la distribución de servicios sociales y reduce el fraude en la asignación de subsidios (Muralidharan et al., 2016).

### **América Latina: Adopción incremental y desafíos estructurales**

En América Latina, el uso de Big Data e IA está en crecimiento, pero enfrenta barreras como la falta de infraestructura tecnológica y la desigualdad social. Brasil ha liderado esfuerzos en el análisis de datos para monitorear la deforestación en la Amazonía

mediante imágenes satelitales y algoritmos de aprendizaje automático, lo que ha permitido diseñar políticas ambientales más efectivas (Amaral et al., 2020).

México ha utilizado Big Data para optimizar sus programas sociales, como el monitoreo del impacto de las transferencias monetarias condicionadas en la reducción de la pobreza. Además, el país implementa análisis de datos para identificar patrones en la incidencia delictiva y diseñar políticas de seguridad pública más efectivas (OECD, 2019). En Chile, el uso de Big Data en el sistema de transporte público de Santiago ha permitido mejorar la asignación de rutas y horarios, reduciendo los tiempos de espera de los pasajeros (Gutiérrez et al., 2018).

### **África: Potencial de transformación en el desarrollo**

En África, el uso de Big Data e IA tiene el potencial de transformar sectores clave como la agricultura, la salud y la educación. Kenia, por ejemplo, utiliza plataformas de análisis de datos para mejorar la productividad agrícola mediante la identificación de patrones climáticos y de suelo (Manyika et al., 2011). En Ruanda, la IA se ha integrado en programas de salud pública para predecir brotes de enfermedades y optimizar la distribución de vacunas (Das & Zhang, 2019).

Adicionalmente, Sudáfrica ha implementado Big Data en su sistema de educación para identificar escuelas con bajo desempeño y priorizar la asignación de recursos. Sin embargo, la limitada conectividad digital y la falta de marcos regulatorios robustos siguen siendo barreras significativas para la adopción generalizada de estas tecnologías en la región (Crawford et al., 2021).

### **2.6. Métodos prospectivos y herramientas tecnológicas en Big Data e Inteligencia Artificial**

La integración de Big Data e inteligencia artificial (IA) en políticas públicas ha revolucionado la manera en que los gobiernos abordan la prospectiva, proporcionando métodos avanzados para anticipar escenarios futuros y diseñar estrategias más efectivas. A continuación, se exploran los métodos técnicos más empleados y las herramientas tecnológicas clave que potencian la capacidad predictiva y adaptativa en la gestión pública.

### 2.6.1. Métodos prospectivos basados en Big Data e IA

#### Modelos predictivos basados en machine learning

El aprendizaje automático (machine learning) ofrece herramientas clave para identificar patrones complejos en grandes volúmenes de datos y generar predicciones precisas. Estos modelos incluyen:

**Regresión lineal y no lineal:** Utilizadas para proyectar tendencias sociales, económicas y demográficas basadas en variables históricas. Por ejemplo, predecir tasas de empleo futuro a partir de datos laborales permite a los gobiernos anticipar crisis de desempleo y diseñar estrategias proactivas (Choi & Varian, 2012; Hyndman & Athanasopoulos, 2018).

**Redes neuronales artificiales:** Estas estructuras simulan el funcionamiento del cerebro humano para analizar relaciones no lineales entre variables. En políticas públicas, son útiles para modelar el comportamiento ciudadano, como la respuesta a programas sociales o el impacto de campañas de vacunación masiva (Goodfellow, Bengio, & Courville, 2016; Schmidhuber, 2015).

**Modelos de clustering:** Dividen los datos en grupos con características similares. Por ejemplo, segmentar a la población según ingresos, educación y hábitos de consumo facilita la personalización de intervenciones gubernamentales y el diseño de políticas específicas para cada grupo (Kitchin, 2014; Hastie, Tibshirani, & Friedman, 2009).

**Modelos de árboles de decisión y random forests:** Estos algoritmos combinan la simplicidad interpretativa de los árboles de decisión con una mayor precisión al usar múltiples iteraciones. Son ampliamente utilizados para evaluar riesgos sociales y priorizar recursos en políticas de asistencia social (Breiman, 2001).

#### Sistemas de Simulación Basados en Agentes (ABM)

Este método modela el comportamiento de actores individuales (agentes) en un entorno simulado, evaluando cómo sus interacciones generan resultados emergentes. En políticas públicas, los ABM se emplean para:

Simular flujos de tráfico y movilidad urbana, ayudando a identificar cuellos de botella

y evaluar políticas de transporte en tiempo real (Batty, 2013; Helbing, 2013).

Anticipar dinámicas de votación en elecciones locales, proporcionando información sobre el impacto potencial de diferentes estrategias de campaña en grupos poblacionales específicos (Epstein, 1999).

Evaluar el impacto de políticas económicas en diferentes segmentos poblacionales, como programas de subsidios agrícolas o incentivos fiscales para pequeños negocios (North & Macal, 2007).

### **Análisis de series temporales y forecasting**

Este método emplea datos históricos para predecir valores futuros mediante modelos estadísticos avanzados como ARIMA, Prophet o modelos de suavización exponencial. Sus aplicaciones en la gestión pública incluyen:

Anticipar demandas de servicios públicos, como el consumo energético durante olas de calor o la ocupación hospitalaria en pandemias (Hyndman & Athanasopoulos, 2018).

Analizar patrones de comportamiento económico para predecir recesiones y establecer medidas preventivas (Box, Jenkins, & Reinsel, 2015).

Prever cambios en la tasa de natalidad y envejecimiento poblacional, lo que permite ajustar políticas de pensiones y programas de salud pública (Holt, 2004).

### **Técnicas de Procesamiento de Lenguaje Natural (NLP)**

El NLP utiliza algoritmos de IA para analizar grandes volúmenes de texto provenientes de redes sociales, encuestas y documentos gubernamentales. Entre sus aplicaciones destacan:

Evaluar la percepción pública sobre nuevas leyes o políticas mediante análisis de sentimientos en redes sociales, lo que permite ajustar estrategias de comunicación (Liu, 2020; Devlin et al., 2019).

Extraer temas clave de foros y debates públicos para informar la agenda política y priorizar inversiones en áreas específicas como educación o salud (Jurafsky & Martin,

2022).

Identificar tendencias de desinformación y ajustar campañas informativas para contrarrestar narrativas falsas (Zhou et al., 2020).

### **Redes bayesianas**

Estas herramientas probabilísticas modelan la incertidumbre y las relaciones causales entre variables. En la gestión pública, las redes bayesianas son útiles para:

Predecir el impacto de decisiones políticas bajo escenarios de alta incertidumbre, como en crisis económicas o emergencias sanitarias (Pearl, 2009; Korb & Nicholson, 2011).

Diseñar políticas adaptativas en áreas sensibles como la mitigación del cambio climático, combinando datos de múltiples fuentes para generar predicciones confiables (Darwiche, 2009).

Estos métodos prospectivos permiten una mayor precisión en las predicciones y una integración más profunda de variables complejas en el diseño de políticas públicas. Por ejemplo, la combinación de ABM con técnicas de NLP puede mejorar la personalización de campañas de información en crisis de salud pública, mientras que los modelos de clustering permiten un enfoque inclusivo al identificar las necesidades de grupos vulnerables. Las herramientas avanzadas, como el uso combinado de redes neuronales y análisis de series temporales, han demostrado ser especialmente valiosas para prever impactos a largo plazo de decisiones gubernamentales.

#### **2.6.2. Herramientas tecnológicas en Big Data e IA para prospectiva**

##### **■ Plataformas de análisis de Big Data**

- **Hadoop:** Framework de código abierto para el almacenamiento y procesamiento distribuido de grandes volúmenes de datos. Es ampliamente utilizado en gobiernos para analizar datos demográficos y económicos (White, 2015).
- **Spark:** Ofrece procesamiento en tiempo real de datos masivos y análisis avanzados mediante aprendizaje automático, útil en la predicción de emergencias

climáticas y crisis económicas (Zaharia et al., 2016).

#### ■ **Software de machine learning e IA**

- **TensorFlow y PyTorch:** Bibliotecas avanzadas para construir y entrenar modelos de aprendizaje automático, como redes neuronales y análisis de clustering. Estas herramientas se emplean para modelar escenarios prospectivos en políticas públicas (Goodfellow et al., 2016).
- **Scikit-learn:** Conjunto de herramientas para análisis predictivo y minería de datos, utilizada en análisis de riesgo social y distribución de recursos.

#### ■ **Sistemas de gestión de datos geoespaciales**

- **QGIS y ArcGIS:** Herramientas para el análisis de datos geográficos que permiten visualizar y modelar la distribución espacial de variables como pobreza, infraestructura y riesgos naturales. Estas plataformas son esenciales para diseñar políticas públicas adaptadas a contextos territoriales específicos (Batty, 2013).

#### ■ **Herramientas de visualización de datos**

- **Tableau y Power BI:** Plataformas que presentan datos complejos en visualizaciones interactivas, facilitando la comunicación de escenarios prospectivos a los tomadores de decisiones. Estas herramientas ayudan a priorizar inversiones en proyectos de infraestructura o programas sociales.

#### ■ **Sistemas NLP**

- **BERT y GPT:** Modelos avanzados para el análisis de texto, empleados en políticas públicas para extraer información clave de documentos legislativos, discursos y redes sociales. Por ejemplo, identificar preocupaciones ciudadanas sobre políticas ambientales (Devlin et al., 2019).

#### ■ **Infraestructura de almacenamiento y computación en la nube**

- **AWS (Amazon Web Services), Microsoft Azure y Google Cloud:** Estas plataformas proporcionan capacidades de almacenamiento y computación esca-

lables, esenciales para manejar datos masivos y ejecutar modelos de IA. Los gobiernos las utilizan para realizar simulaciones y análisis en tiempo real (Manyika et al., 2011).

### **3. Metodología**

#### **3.1. Enfoque metodológico**

El presente estudio utiliza un enfoque metodológico basado en el análisis sistémico mediante una revisión exhaustiva de literatura científica. Este enfoque tiene como propósito identificar los principales modelos de adopción tecnológica, evaluar su aplicación en políticas públicas y analizar las barreras, limitaciones y líneas de investigación futuras. La revisión se centra en estudios empíricos y teóricos relevantes sobre el uso de Big Data e inteligencia artificial (IA) en la perspectiva gubernamental.

#### **3.2. Criterios de selección de literatura**

La selección de literatura se realizó considerando:

**Relevancia temática:** Se incluyeron únicamente estudios que abordaran la aplicación de Big Data e IA en políticas públicas o prospectiva social.

**Rigor metodológico:** Fueron seleccionadas investigaciones publicadas en revistas científicas indexadas, con métodos claramente definidos y datos verificables.

**Diversidad geográfica y temporal:** Se priorizaron estudios que ofrecieran una visión global, considerando tanto contextos desarrollados como en desarrollo.

**Actualidad:** Se incluyeron publicaciones entre 2010 y 2023, asegurando la integración de avances recientes en tecnologías de análisis de datos.

#### **3.3. Procedimiento de análisis**

El análisis sistémico se dividió en tres fases:

Fase 1: Revisión y categorización

Se identificaron y clasificaron las referencias según su enfoque en modelos de adopción, aplicaciones prácticas y desafíos éticos.

#### Fase 2: Análisis comparativo

Los estudios seleccionados se contrastaron en función de su contexto de implementación, los resultados observados y las tecnologías empleadas.

#### Fase 3: Inferencias y síntesis

A partir de los hallazgos, se derivaron conclusiones sobre el estado del arte, limitaciones y líneas de investigación.

### **3.4. Instrumentos y herramientas**

El análisis se realizó utilizando herramientas de gestión bibliográfica (Mendeley y Zotero) para organizar las referencias y software de análisis cualitativo (NVivo) para identificar patrones en los textos revisados. Se emplearon también matrices de evaluación comparativa para sintetizar los datos.

## **4. Resultados y discusión**

El análisis sistémico basado en la revisión exhaustiva de literatura revela un panorama integral sobre la integración de Big Data e inteligencia artificial (IA) en políticas públicas. Este apartado amplía la comprensión de los modelos de adopción tecnológica, las aplicaciones prácticas y las barreras, limitaciones y posibles líneas de investigación futura. Los resultados incluyen análisis detallados de casos de estudio, nuevos enfoques y tablas comparativas que sintetizan los hallazgos.

### **4.1. Modelos de adopción tecnológica**

Los modelos identificados permiten comprender cómo los gobiernos integran Big Data e IA en sus procesos. Entre los principales hallazgos destacan:

### **Modelo de difusión de innovaciones (Rogers, 2003)**

**Ventajas:** Explica cómo la percepción de utilidad impulsa la adopción de tecnologías disruptivas en gobiernos. Las innovaciones con mayor compatibilidad y ventaja relativa tienen más probabilidad de ser adoptadas rápidamente.

**Limitaciones:** La complejidad percibida y la infraestructura organizacional limitada en países en desarrollo ralentizan su implementación.

**Nuevos hallazgos:** Las etapas de adopción (innovadores, adoptantes tempranos, mayoría temprana y rezagados) son útiles para clasificar cómo diferentes gobiernos implementan tecnologías avanzadas.

**Ejemplo:** Países desarrollados como Singapur y Estonia son innovadores, mientras que países con menores recursos tecnológicos se ubican como rezagados (Gartner, 2019; Baller et al., 2016).

### **Modelo de madurez tecnológica (Gartner, 2019)**

**Ventajas:** Permite evaluar el progreso desde proyectos piloto hasta la optimización estratégica de Big Data e IA.

**Barreras:** Los gobiernos en etapas tempranas enfrentan problemas de financiación y capacitación técnica.

**Nuevos hallazgos:** Este modelo ayuda a identificar las brechas entre las capacidades tecnológicas actuales y el potencial deseado.

**Avances observados:** Singapur y Dinamarca lideran en la integración de tecnologías avanzadas, alcanzando etapas optimizadas.

**Ejemplo:** Brasil se encuentra en una etapa definida gracias a su monitoreo ambiental avanzado, mientras que Ruanda apenas inicia la integración de IA en salud pública (Manyika et al., 2011; Amaral et al., 2020).

### **Modelo UTAUT (Venkatesh et al., 2003)**

**Ventajas:** Se centra en factores como utilidad percibida y facilidad de uso, lo que fomenta la aceptación de tecnologías en los usuarios finales, incluidos funcionarios públicos.

**Barreras:** Los entornos gubernamentales jerárquicos presentan resistencia organizacional, especialmente en países con alta burocracia.

**Nuevos hallazgos:** Países con estrategias de capacitación y políticas de incentivos muestran mejores tasas de adopción.

**Ejemplo:** Japón combina modelos de madurez tecnológica y aceptación del usuario para abordar el envejecimiento poblacional mediante IA (OECD, 2019).

## **4.2. Aplicaciones prácticas en prospectiva de políticas públicas**

El Big Data e IA han permitido transformaciones significativas en diversos sectores. Los nuevos análisis añaden casos relevantes como los siguientes:

### **Movilidad urbana**

**Nuevas aplicaciones:** En Londres, el análisis de tráfico en tiempo real ha optimizado rutas de transporte público, reduciendo emisiones de carbono en un 18 % (Batty, 2013; Toots et al., 2017).

**Comparación:** Singapur redujo la congestión un 25 %, mientras que Londres priorizó objetivos ambientales.

### **Salud pública**

**Nuevas aplicaciones:** En Ruanda, sistemas de IA anticipan brotes de malaria con 92 % de precisión, permitiendo intervenciones tempranas (Das & Zhang, 2019).

**Comparación:** Estados Unidos priorizó la mitigación de pandemias, mientras que Ruanda optimizó enfermedades locales. Además, el uso de Big Data para rastrear

brotos epistemológicos, como el COVID-19, permitió asignar recursos de manera más eficiente en Estados Unidos (Choi & Varian, 2012).

### Sostenibilidad ambiental

Nuevas aplicaciones: Canadá utiliza Big Data para prever incendios forestales y coordinar evacuaciones. Estas estrategias redujeron el impacto económico un 22 % (Choi & Varian, 2012; Manyika et al., 2011). Por otra parte, en Brasil, algoritmos de aprendizaje automático monitorearon la deforestación, reduciendo la pérdida forestal en un 15 % (Amaral et al., 2020).

### Inclusión Social

Dinamarca utilizó IA para identificar comunidades en riesgo de pobreza energética, diseñando subsidios más precisos (Hardoon & Shmueli, 2013).

Al respecto, en la siguiente Tabla se muestran algunos ejemplos de aplicaciones prácticas de la inteligencia artificial y el big data en las políticas públicas.

Tabla 1: Aplicaciones prácticas de IA y Big Data en políticas públicas

Caso	Tecnología utilizada	Impacto	Fuente
Movilidad Urbana en Singapur	Big Data en análisis de tráfico en tiempo real	Reducción tráfico 25 %	Batty, 2013
Gestión de Tráfico en Londres	Sensores y algoritmos IA	Reducción del CO <sub>2</sub> 18 %	Toots et al., 2017
Predicción de Brotes de Malaria en Ruanda	Modelos epidemiológicos basados en IA	Reducción de brotes de malaria en un 30 %	Das & Zhang, 2019
Monitoreo de Deforestación en Brasil	Algoritmos de aprendizaje automático en imágenes satelitales	Reducción del 15 % en la pérdida forestal anual	Amaral et al., 2020
Proyecciones Climáticas en Canadá	Big Data para proyecciones climáticas	Prevención de incendios forestales	Manyika et al., 2011

Caso	Tecnología utilizada	Impacto	Fuente
Digitalización de Servicios Públicos en Estonia	Sistemas interoperables de Big Data	Mayor eficiencia y transparencia en trámites gubernamentales	Toots et al., 2017
Optimización de Recursos Hídricos en Singapur	Sensores IoT y análisis en tiempo real	Mejoras en la distribución de agua y sostenibilidad	Batty, 2013
Análisis de Opinión Pública en Redes Sociales en Estados Unidos	Procesamiento de Lenguaje Natural (NLP)	Identificación de tendencias y prioridades ciudadanas	Choi & Varian, 2012
Predicción de Riesgos Económicos en Alemania	Modelos predictivos económicos	Prevención de recesiones económicas	OECD, 2019
Gestión de Desastres Naturales en Japón	Redes bayesianas y aprendizaje profundo	Mitigación de daños causados por tsunamis	Müller et al., 2016
Crédito Social en China	Sistemas basados en IA para análisis de comportamiento	Cumplimiento normativo incrementado	Kostka, 2019
Prevención de Inseguridad Alimentaria en Kenia	Big Data en sistemas agrícolas	Prevención de hambrunas en comunidades rurales	Manyika et al., 2011
Detección de Fraudes Fiscales en Francia	Análisis predictivo de patrones fiscales	Detección temprana de evasiones fiscales	Hardoon & Shmueli, 2013
Asignación de Subsidios Agrícolas en la India	Modelos de segmentación agrícola	Distribución precisa de recursos agrícolas	OECD, 2019
Predicción de Demanda Energética en Australia	Análisis de consumo energético mediante IA	Ahorros en costos energéticos del 20 %	Hardoon & Shmueli, 2013
Optimización de Transporte Público en Chile	Big Data y optimización de rutas	Reducción de tiempos de espera en transporte público	Choi & Varian, 2012

<b>Caso</b>	<b>Tecnología utilizada</b>	<b>Impacto</b>	<b>Fuente</b>
Monitoreo de Calidad del Aire en México	Sistemas de monitoreo ambiental con IA	Reducción del impacto de contaminación en un 12 %	Batty, 2013
Gestión de Salud Pública en Corea del Sur	Sistemas predictivos de pandemias	Asignación eficiente de recursos médicos	Manyika et al., 2011
Simulación de Impactos Urbanos en Noruega	Análisis geoespacial y simulaciones urbanas	Optimización del uso del suelo en áreas urbanas	Batty, 2013
Planeación Estratégica en Nueva Zelanda	Análisis de escenarios estratégicos con Big Data	Adaptación de políticas a contextos dinámicos	Müller et al., 2016

*Nota.* Fuente: elaboración a partir de los autores mencionados.

De la tabla anterior se percibe que tanto la inteligencia artificial, a través de algoritmos, como el Big Data, para la recopilación y administración de grandes volúmenes de información, son dos tecnologías que, gestionadas eficazmente, permiten obtener un impacto significativo en diferentes ámbitos hacia una mejora en la sociedad, lo que conlleva a sentar bases firmes para la formulación de políticas públicas en diferentes latitudes.

### **4.3. Barreras y limitaciones**

#### **Infraestructura tecnológica**

Problemas adicionales: La brecha digital entre zonas urbanas y rurales exacerba la desigualdad en la implementación tecnológica, así como la falta de interoperabilidad entre sistemas gubernamentales (Baller et al., 2016).

#### **Capacitación y recursos humanos**

Observaciones nuevas: Países en desarrollo enfrentan un éxodo de talento capacitado hacia el sector privado, lo que limita el desarrollo interno (OECD, 2019).

Capacitación: La carencia de personal capacitado en análisis de datos es una barrera crítica, especialmente en países en desarrollo.

### **Ética y gobernanza de datos**

Perspectiva nueva: La dependencia de proveedores tecnológicos extranjeros plantea riesgos de soberanía tecnológica (Mittelstadt et al., 2016).

Los problemas de privacidad y los sesgos algorítmicos exigen la creación de marcos regulatorios robustos.

### **Resistencia organizacional**

Las jerarquías rígidas y la aversión al cambio cultural ralentizan la adopción de tecnologías disruptivas.

#### **4.4. Líneas de investigación futura**

Marco regulatorio: Investigar cómo diseñar políticas que garanticen un uso ético de Big Data e IA.

Interoperabilidad: Diseñar plataformas integradas para facilitar la interoperabilidad entre sistemas gubernamentales y privados.

Impacto social: Evaluar el efecto de estas tecnologías en la equidad y el acceso a servicios públicos.

Desarrollo de modelos predictivos éticos: Estudiar cómo integrar valores éticos en algoritmos para evitar sesgos discriminatorios.

Ampliación de estudios comparativos: Realizar estudios comparativos entre países desarrollados y en desarrollo para identificar mejores prácticas transferibles.

Innovaciones en capacitación tecnológica: Diseñar programas de educación continua para funcionarios públicos en entornos gubernamentales.

Adaptación a contextos locales: Investigar cómo adaptar tecnologías globales a necesidades locales específicas.

## 5. Conclusiones

Los resultados obtenidos en este estudio subrayan el potencial transformador del Big Data y la inteligencia artificial (IA) en el diseño y la implementación de políticas públicas orientadas a la prospectiva. Estas tecnologías permiten anticipar escenarios complejos, además de ofrecer herramientas para desarrollar estrategias más adaptativas y efectivas. Sin embargo, su adopción e integración presentan desafíos significativos, que van desde barreras organizacionales hasta cuestiones éticas y técnicas.

En primer lugar, los hallazgos revelaron que los modelos de adopción tecnológica, como el de Difusión de Innovaciones de Rogers y el de Madurez Tecnológica de Gartner, son fundamentales para comprender la velocidad y efectividad con la que los gobiernos incorporan estas herramientas en sus procesos. Los factores de ventaja relativa y compatibilidad se identificaron como impulsores clave de la adopción, pero la complejidad percibida y la falta de personal capacitado persisten como obstáculos significativos. Estas limitaciones son particularmente pronunciadas en países con menores capacidades tecnológicas, destacando la necesidad de programas de capacitación y estructuras organizacionales más flexibles.

Por otro lado, las aplicaciones prácticas del Big Data e IA en sectores como la movilidad urbana, la salud pública y la sostenibilidad ambiental demuestran su eficacia para abordar desafíos complejos. Casos exitosos en Singapur, Estonia y Brasil ilustran cómo estas tecnologías pueden transformar la gobernanza, mientras que las limitaciones éticas y regulatorias observadas en regiones como América Latina y África exigen un enfoque más equilibrado y contextualizado.

En términos de prospectiva, el estudio destacó que la capacidad de segmentar poblaciones, modelar escenarios y optimizar recursos posiciona al Big Data y la IA como pilares de la toma de decisiones estratégicas. Sin embargo, el análisis también reveló preocupaciones importantes en torno a la equidad y los posibles sesgos inherentes a los algoritmos, lo que resalta la urgencia de desarrollar marcos éticos robustos y garantizar la transparencia en el uso de estas tecnologías.

Finalmente, el estudio identifica líneas de investigación futuras que pueden abordar las limitaciones actuales, como el diseño de sistemas interoperables, el análisis del impacto social y el desarrollo de herramientas específicas para contextos de alta

incertidumbre. Para maximizar el impacto positivo del Big Data y la IA, es imperativo fomentar la colaboración entre gobiernos, academia y sector privado, con el objetivo de crear una infraestructura tecnológica inclusiva y sostenible que beneficie a todas las comunidades.

## Referencias

- Amaral, S., Santos, J. R., & Martins, M. 2020. Monitoring deforestation in the Amazon using satellite imagery and machine learning. *Journal of Environmental Science* 58(3): 220–231.
- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. 2016. Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica. <https://www.propublica.org>
- Baller, S., Dutta, S., & Lanvin, B. 2016. The global information technology report 2016: Innovating in the digital economy. World Economic Forum.
- Batty, M. 2013. Big data, smart cities and city planning. *Dialogues in Human Geography* 3(3): 274–279. <https://doi.org/10.1177/2043820613513390>
- Breiman, L. 2001. Random forests. *Machine Learning*, 45(1): 5–32. <https://doi.org/10.1023/A:1010933404324>
- Choi, H. & Varian, H. 2012. Predicting the present with Google Trends. *Economic Record* 88(s1): 2–9.
- Crawford, K., Dobson, K. & MacDonald, F. 2021. Data colonialism and the limits of AI ethics. *Big Data & Society* 8(2). <https://doi.org/10.1177/20539517211017227>
- Das, R., & Zhang, Z. 2019. AI and big data in healthcare: The next frontier for innovation. *Journal of Healthcare Informatics* 12(3): 45–61.
- Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics* 1(1): 4171–4186. <https://doi.org/10.18653/v1/N19-1423>
- Gartner. 2019. Hype cycle for emerging technologies. Gartner Research.

- Goodfellow, I., Bengio, Y. & Courville, A. 2016. *Deep learning*. MIT Press.
- Haroon, D. & Shmueli, G. 2013. *Getting started with business analytics: Insightful decision-making*. CRC Press.
- Howard, P. N. & Kollanyi, B. 2016. Bots, social media, and the public discourse: Miscommunication in the age of automation. *Political Communication* 33(1): 1–10. <https://doi.org/10.1080/10584609.2016.1224894>
- Hyndman, R. J. & Athanasopoulos, G. 2018. *Forecasting: Principles and practice* (2nd ed.). OTexts.
- Jurafsky, D. & Martin, J. H. 2022. *Speech and language processing* (3rd ed.). Pearson.
- Kitchin, R. 2014. *The data revolution: Big data, open data, data infrastructures & their consequences*. SAGE Publications.
- Kostka, G. 2019. China's social credit systems and public opinion: Explaining high levels of approval. *New Media & Society* 21(7): 1565–1593.
- Laney, D. 2001. *3D data management: Controlling data volume, velocity and variety*. META Group Research Note.
- Liu, B. 2020. *Sentiment analysis: Mining opinions, sentiments, and emotions* (2nd ed.). Cambridge University Press.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. & Byers, A. H. 2011. *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.
- Mayer-Schönberger, V. & Cukier, K. 2013. *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S. & Floridi, L. 2016. The ethics of algorithms: Mapping the debate. *Big Data & Society* 3(2): 1–21. <https://doi.org/10.1177/2053951716679679>
- Muralidharan, K., Niehaus, P. & Sukhtankar, S. 2016. Building state capacity: Evidence from biometric smartcards in India. *American Economic Review* 106(10): 2895–2929. <https://doi.org/10.1257/aer.20141346>
- North, M. J. & Macal, C. M. 2007. *Managing business complexity: Discovering strategic*

solutions with agent-based modeling and simulation. Oxford University Press.

OECD. 2019. The path to becoming a data-driven public sector. OECD Digital Government Studies.

Pearl, J. 2009. Causality: Models, reasoning, and inference (2nd ed.). Cambridge University Press.

Rogers, E. M. 2003. Diffusion of innovations (5th ed.). Free Press.

Toots, M., McBride, K., Kalvet, T. & Krimmer, R. 2017. Open data as enabler of public service co-creation: Exploring the drivers and barriers. *Information Polity* 22(3): 231–245.

Venkatesh, V., Morris, M. G., Davis, G. B. & Davis, F. D. 2003. User acceptance of information technology: Toward a unified view. *MIS Quarterly* 27(3): 425–478.

White, T. 2015. Hadoop: The definitive guide (4th ed.). O'Reilly Media.

Zaharia, M., Chowdhury, M., Franklin, M. J., Shenker, S. & Stoica, I. 2016. Spark: Cluster computing with working sets. *USENIX Conference on Networked Systems Design and Implementation*.

Zhou, X., Zafarani, R., Shu, K. & Liu, H. 2020. Fake news: Fundamental theories, detection strategies, and challenges. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 3205–3206. <https://doi.org/10.1145/3394486.3406468>

# ARTÍCULOS DE REFLEXIÓN

---

Conocimiento i Política – No. 7, 2026

## Sesgos de IA como Desigualdad Política: Violencia Digital de Género y Evaluación Pericial

Albanella Luena Chávez-Turello<sup>1</sup>

<sup>1</sup> Investigadora independiente.

achavezturrello@gmail.com (Autora de correspondencia)

**Resumen.** Los sesgos cognitivos documentados en la bibliografía sobre evaluación pericial — confirmación, anclaje, retrospectivo, cultural— no desaparecen cuando los sistemas de inteligencia artificial median el acceso a la justicia. Se amplifican. Este artículo examina la violencia digital de género como campo donde esa amplificación resulta especialmente visible, y argumenta que el problema no es de naturaleza técnica sino política: grupos específicos encuentran barreras estructurales codificadas en algoritmos para acceder a protección institucional. Desde la victimología crítica, se documentan tres niveles de victimización que los sistemas de IA reproducen y escalan —daño directo, revictimización institucional y normalización social— y se analiza cómo los instrumentos periciales tradicionales, aplicados sin ajuste al contexto digital, patologizan respuestas adaptativas y reproducen sesgo cultural. Cuando esos sesgos se codifican en herramientas algorítmicas de justicia penal, moderación de contenido o evaluación de riesgo, el resultado es discriminación automatizada que opera sin supervisión efectiva y con opacidad que dificulta cualquier rendición de cuentas. Las estrategias propuestas —auditoría algorítmica obligatoria, protocolos de documentación, formación especializada y reparación integral— parten del reconocimiento de que mejorar evaluaciones individuales es insuficiente si no se transforma la estructura institucional y tecnológica que produce el sesgo.

**Palabras clave:** sesgos algorítmicos, violencia digital de género, desigualdad política, psicología forense, inteligencia artificial

**Abstract.** The cognitive biases documented in the literature on expert assessment —confirmation, anchoring, hindsight, cultural bias— do not disappear when artificial intelligence systems mediate access to justice. They are amplified. This article examines gender-based digital violence as a field in which that amplification becomes especially visible, and argues that the problem is not technical but political: specific groups encounter structural barriers encoded in algorithms when seeking institutional protection. From the perspective of critical victimology, three levels of victimization reproduced and scaled by AI systems are documented —direct harm, institutional revictimization, and social normalization— and the article analyzes how traditional forensic instruments, when applied without adjustment to the digital context, pathologize adaptive responses and reproduce cultural bias. When these biases are encoded in algorithmic tools for criminal justice, content moderation, or risk assessment, the result is automated discrimination that operates without effective oversight and with an opacity that makes accountability difficult. The proposed strategies —mandatory algorithmic auditing, documentation protocols, specialized training, and comprehensive repair— begin from the recognition that improving individual assessments is insufficient if the institutional and technological structure that produces bias is not transformed.

**Keywords:** algorithmic bias; gender-based digital violence; political inequality; forensic psychology; artificial intelligence

**Historial:** Recibido: 29 de marzo de 2026 | Aceptado: 15 de mayo de 2026

Este documento está bajo una licencia

**Ética/Financiamiento:** La autora declara no tener conflictos de interés. La investigación fue autofinanciada y no recibió financiamiento externo.

CC BY-NC-ND 4.0.

**Cómo citar (APA):** Chávez-Turello, A. L. (2026). Sesgos de IA como desigualdad política: violencia digital de género y evaluación pericial. *Conocimiento i Política*, 7(1), 273-285. DOI: 10.64480/cip.2026.n7.09.

## 1. Introducción

En evaluaciones periciales de violencia digital de género, el profesional evaluador documenta amenazas de muerte sostenidas, imágenes íntimas difundidas sin consentimiento y suplantación de identidad; y, sin embargo, una pregunta recurrente ilustra el problema central de este artículo: «¿Por qué no bloqueó el contacto?»

Esta interrogante aparece frente a expedientes que registran tales evidencias. No es una excepción individual: es el mismo sesgo que hoy se codifica en sistemas algorítmicos que median el acceso a la justicia, trasladando sistemáticamente la responsabilidad del acto violento hacia decisiones previas de la víctima y replicando patrones de culpabilización documentados (Chávez-Turello, 2026; Liell et al., 2022).

Lo que agrava esta práctica es su codificación actual en sistemas algorítmicos que median el acceso a la justicia –moderación automatizada de plataformas, triage institucional, herramientas predictivas en justicia penal–, escalando sesgos humanos a discriminación automatizada sin supervisión efectiva (Noble, 2018).

La pregunta traslada la responsabilidad del acto violento hacia las decisiones previas de quien sufrió el daño. Asume que bloquear habría detenido la violencia. La evidencia empírica dice lo contrario, y además el patrón replica lo que la investigación identifica como sesgo de culpabilización hacia víctimas (Chávez-Turello, 2026; Liell et al., 2022). Ese sesgo no es una excepción ni un desliz individual: es el telón de fondo sobre el que se construyen muchas evaluaciones periciales en contextos de violencia digital.

Lo que complica aún más el panorama es que ese mismo sesgo se está codificando en sistemas algorítmicos. Las plataformas utilizan IA para moderar contenido, evaluar riesgo y clasificar denuncias. Los sistemas de justicia penal implementan herramientas predictivas. Las instituciones públicas automatizan la atención inicial mediante chatbots y triage algorítmico. Cuando los sesgos documentados en la práctica pericial humana se convierten en infraestructura automatizada que opera a escala y sin supervisión efectiva, la opacidad decisional hace casi imposible la rendición de cuentas (Noble, 2018). No es un problema técnico corregible con actualizaciones de software. Es desigualdad política codificada en sistemas que se presentan como neutrales.

La violencia digital de género es, en este sentido, un campo de observación privilegia-

do. No es una categoría separada de la violencia de género: es violencia de género que opera mediante herramientas digitales, con los mismos patrones de control, dominación y daño, pero con escala de difusión ilimitada, permanencia del contenido y una capacidad de replicación que elimina cualquier posibilidad de zona segura para la víctima (ONU Mujeres, 2020). Cuando el acoso opera precisamente mediante los dispositivos que alguien necesita para subsistir, trabajar o mantener relaciones sociales, la idea de que la solución es «desconectarse» no solo es irrealista: es otra forma de culpabilización.

Varios factores agravan la contaminación evaluativa en estos casos. La ausencia de formación especializada lleva a que profesionales preparados para documentar lesiones corporales no sepan cómo evaluar daño psicológico mediado por pantallas (Henry et al., 2020). La naturalización cultural de internet como espacio no serio perpetúa el mito de que la ausencia de contacto físico equivale a ausencia de violencia real (Citron, 2014). Y los mitos sobre la tecnología responsabilizan a las víctimas mediante narrativas de exposición voluntaria y control individual sobre la propia seguridad digital (Bailey & Shayan, 2020). El resultado es revictimización institucional sistemática: incredulidad, minimización, culpabilización, imposibilidad de obtener reparación (Asociación Civil Comunicación para la Igualdad, 2021).

Este artículo se mueve en tres direcciones. Primero, aplica el marco de sesgos forenses y victimología crítica al análisis de violencia digital, documentando cómo esos sesgos se amplifican mediante IA. Segundo, examina críticamente los instrumentos periciales tradicionales en contextos digitales, identificando cómo reproducen sesgo cultural y codifican desigualdad política. Tercero, propone estrategias concretas que tienen respaldo en evidencia científica y en la normativa de derechos humanos, sin pretender que sean suficientes por sí mismas.

## **2. Violencia digital y sistemas algorítmicos**

La Relatora Especial de las Naciones Unidas definió la violencia digital como los actos de violencia de género cometidos, instigados o agravados por tecnologías de información y comunicación (Šimonović, 2018). Hay tres elementos en esa definición que vale la pena no perder de vista. No es una categoría separada sino una continuidad: los patrones de desigualdad de género preexisten a la tecnología, que los

amplifica. Incluye tanto violencia que se origina en espacios digitales como la que migra desde contextos offline. Y la tecnología no solo facilita la violencia: la agrava, al permitir vigilancia constante, control económico y aislamiento sin necesidad de presencia física (Dragiewicz et al., 2018; Southworth et al., 2007).

Los sistemas algorítmicos operan en este campo en distintos niveles que conviene distinguir. Como infraestructura que facilita la perpetración: algoritmos de recomendación que amplifican contenido violento, herramientas de deepfake que permiten suplantación, sistemas de reconocimiento facial que facilitan vigilancia. Como mediadores institucionales que evalúan y responden: moderación automatizada de contenido, triage en centros de atención, evaluación de riesgo en justicia penal. Y como productores de sesgo que codifican discriminación estructural: algoritmos entrenados con datos que reflejan la minimización institucional histórica, sistemas que priorizan denuncias según patrones que excluyen o subvaloran la violencia digital.

Las manifestaciones concretas incluyen la difusión no consentida de contenido íntimo (McGlynn et al., 2017), el acoso sexual digital sostenido, las amenazas directas, la suplantación de identidad, la vigilancia tecnológica, la coordinación de violencia colectiva y la exclusión de espacios digitales necesarios para la vida social y laboral. Frente a cada una de estas formas, los algoritmos no solo median: amplifican. Un sistema de moderación entrenado con datos donde históricamente se minimizó la violencia digital reproducirá automáticamente esa minimización. Un sistema que no incorporó la violencia digital en su diseño subestimarán sistemáticamente su peligrosidad, no por malicia, sino por ausencia.

Esto constituye lo que este artículo llama desigualdad política codificada: grupos específicos enfrentan barreras estructurales para acceder a protección institucional, y esas barreras son producto de decisiones deliberadas —o de la ausencia de decisiones— sobre qué datos usar para entrenar algoritmos, qué criterios priorizar, qué violencias se consideran suficientemente graves para la intervención automática. No son el resultado inevitable de la tecnología. Son el resultado de quién diseña los sistemas, con qué datos y para qué usuarios imaginados.

### 3. Sesgos forenses: del humano al algoritmo

El sesgo de confirmación es quizás el más estudiado en contextos periciales: ocurre cuando el profesional busca evidencia que confirme su hipótesis inicial mientras ignora información contradictoria (Kassin et al., 2013). En violencia digital, el mecanismo es relativamente transparente: el perito asume desde el primer momento que «no puede ser tan grave porque es online» e interpreta toda la evidencia a través de ese filtro. Lo que resulta más preocupante es que, cuando ese sesgo se codifica en un algoritmo, el sistema aprende que ciertos patrones no ameritan clasificación de riesgo alto. Lo reproduce automáticamente. Lo escala. Y nadie lo nombra como sesgo porque aparece como resultado objetivo de un proceso técnico.

El sesgo de anclaje opera de manera más sutil. Ocurre cuando los juicios posteriores quedan desproporcionadamente condicionados por la información inicial. Un profesional que lee en el expediente que «la víctima mantuvo comunicación después del primer incidente» puede anclar toda su evaluación en ese dato, interpretando cada conducta posterior como confirmación de que la situación no era tan grave. Los sistemas de IA amplifican esto mediante lo que en ciencias de la computación se llama dependencia del camino: las decisiones tempranas condicionan fuertemente el recorrido posterior del sistema (Tversky & Kahneman, 1974). Si la primera clasificación algorítmica asigna riesgo bajo, esa etiqueta tenderá a persistir.

El sesgo retrospectivo —creer que los eventos eran más predecibles de lo que realmente eran— contamina especialmente los casos que escalaron hacia violencia física o femicidio. Los revisores posteriores concluyen que «era obvio» y culpan a quienes evaluaron antes. Este sesgo introduce un problema técnico concreto en el entrenamiento de algoritmos predictivos: cuando el sistema se entrena con datos de casos que efectivamente escalaron, aprende que la escalada era predecible, cuando en realidad no lo era en el momento de la evaluación. El resultado es una falsa sensación de precisión predictiva que puede llevar tanto a sobreclasificación como a ignorar señales de riesgo genuinas en casos que no encajan con el patrón aprendido (Weber et al., 2024).

El sesgo cultural es el que aparece más claramente en la pregunta del perito con la que abrimos este artículo. Surge cuando los profesionales evalúan conductas usando marcos sobre cómo debería comportarse una víctima «real»: ¿por qué no bloqueó el

contacto?, ¿por qué no dejó de usar redes?, ¿por qué compartió fotos íntimas? Cada una de estas preguntas lleva implícita una imagen del comportamiento esperado que excluye la posibilidad de que la víctima haya tenido razones perfectamente racionales para actuar como actuó. Los algoritmos codifican estos sesgos cuando se entrenan con datos generados por decisiones humanas ya sesgadas. Un sistema entrenado con expedientes donde se culpabilizó a las víctimas aprenderá esos patrones. Los sistematizará.

Conviene mencionar también la fatiga decisional y el trauma vicario, aunque no como justificación sino como variable de contexto. Ambos afectan a profesionales que evalúan repetidamente casos de violencia: generan desensibilización, minimizan el daño percibido, favorecen el uso de atajos cognitivos (Levin et al., 2021). Los sistemas algorítmicos no experimentan fatiga. Pero cuando se entrenan con datos generados por profesionales que sí la experimentaron, esa fatiga queda codificada en el modelo. La automatización no neutraliza el problema: lo congela y lo escala.

Lo que une a todos estos sesgos es que no son errores individuales corregibles con más capacitación. Son productos de estructuras institucionales y culturales más amplias. La automatización no los elimina. Los opacifica, los vuelve difíciles de impugnar, y hace mucho más difícil la rendición de cuentas cuando se produce un error.

#### **4. Victimización en tres niveles**

El marco de victimología crítica distingue tres niveles que, en el contexto de la violencia digital algorítmicamente mediada, adquieren matices específicos (Chávez-Turello, 2026).

La victimización primaria es el daño directo causado por el acto violento en sí mismo. En violencia digital, eso incluye las amenazas de muerte, la difusión no consentida de imágenes íntimas, el acoso sexual sostenido, la suplantación de identidad, la vigilancia tecnológica. El daño es psicológico, social, económico, laboral. No es menos real por operar mediante pantallas, y pretender que lo es es ya en sí mismo una forma de sesgo evaluativo.

La victimización secundaria es el daño adicional causado por las instituciones que deberían proteger. Incluye la incredulidad sobre la gravedad de los hechos, la mini-

mización, la culpabilización, las demandas de prueba imposibles, los procedimientos que exponen a mayor riesgo. Los sistemas algorítmicos generan formas nuevas de este daño secundario: un algoritmo de moderación que censura las publicaciones de víctimas que denuncian violencia mientras permite que el material íntimo no consentido circule libremente está victimizando a quien ya fue victimizada. Un sistema que clasifica la violencia digital como «riesgo bajo» está negando acceso a protección.

La victimización terciaria es la normalización social que lleva a que las propias víctimas internalicen la culpa y asuman que deben tolerar el daño. Las narrativas culturales sobre privacidad digital, responsabilidad individual y «exposición voluntaria» operan aquí (Varona, 2016). Los algoritmos contribuyen a este nivel mediante el diseño de plataformas que naturalizan la vigilancia, mediante sistemas que trasladan a los usuarios la responsabilidad de gestionar su propia seguridad digital sin proporcionarles herramientas efectivas para hacerlo, y mediante políticas que convierten el riesgo en un problema individual cuando es estructural.

Documentar estos tres niveles no es solo un ejercicio teórico. Tiene consecuencias prácticas directas para la evaluación pericial: no basta con documentar el daño primario. Una evaluación que solo registra síntomas sin analizar cómo las instituciones y los sistemas contribuyeron a generar daño adicional está incompleta. Y una evaluación que trata el daño terciario como producto de la personalidad de la víctima —en lugar de reconocerlo como normalización social— está reproduciéndolo.

## 5. Instrumentos periciales y sesgo cultural

El DSM-5-TR fue diseñado para clasificar trastornos mentales. No para evaluar el daño causado por violencia. Su aplicación acrítica a casos de violencia digital genera problemas que conviene nombrar con precisión. La hipervigilancia digital —revisar constantemente las redes buscando nuevo contenido difundido— es una conducta adaptativa ante una amenaza real, no un trastorno de ansiedad que requiera tratamiento. Clasificarla como síntoma patológico no solo desinforma: puede utilizarse en contra de la víctima durante el proceso penal. El DSM clasifica síntomas individuales, no patrones de control, dominación y daño sostenido, y reproduce sesgo cultural sobre las respuestas que se consideran «apropiadas» al trauma (Ussher, 2011).

Los protocolos de evaluación de credibilidad —CBCA, SVA, RM— presentan proble-

mas parecidos desde otro ángulo. Se desarrollaron originalmente para testimonios sobre abuso sexual infantil. Su traslación a casos de violencia digital contra personas adultas asume que los indicadores de credibilidad son universales, cuando la evidencia muestra que varían en función del contexto, la cultura y el tipo de violencia (Vrij, 2005). Un protocolo que interpreta la «falta de detalles» como indicador de baja credibilidad penaliza a víctimas que experimentaron disociación durante los hechos o que, simplemente, narraron de manera diferente a lo que el protocolo prevé como norma.

Las escalas de detección de simulación —SIRS, M-FAST— añaden otra capa problemática. Se diseñaron para identificar exageración sintomática en contextos médico-legales específicos, y su aplicación implícita asume que la víctima podría estar fingiendo el daño para obtener algún beneficio (Rogers, 2008). Esta asunción reproduce el sesgo de incredulidad y desplaza la carga de la prueba hacia quien ya sufrió el daño. Un sistema de justicia que parta de la presunción de que las denuncias son falsas hasta que se demuestre lo contrario no es un sistema justo.

Cuando estos instrumentos se codifican en sistemas algorítmicos, los problemas se escalan de manera automática. Un algoritmo entrenado con evaluaciones DSM que sistemáticamente patologizaron respuestas adaptativas reproducirá esa patologización. Un sistema que usó protocolos culturalmente sesgados como base de entrenamiento seguirá reproduciendo ese sesgo en cada evaluación posterior. Vale la pena señalar que la solución no es abandonar la evaluación estandarizada —los instrumentos tienen valor cuando se usan apropiadamente y en el contexto para el que fueron diseñados—, sino desarrollar herramientas específicas para violencia digital, validadas con poblaciones relevantes y diseñadas para no patologizar respuestas adaptativas.

## 6. Estrategias de mitigación

Las estrategias de nivel individual —formación especializada obligatoria, supervisión clínica regular, protocolos que no patologicen respuestas adaptativas, instrumentos validados específicamente para violencia digital— son necesarias. También son insuficientes si se quedan en ese nivel. El sesgo no es un déficit individual; es un producto estructural, y las intervenciones exclusivamente individuales no transforman

la estructura.

A nivel institucional, la auditoría algorítmica obligatoria de los sistemas usados en justicia penal, moderación de contenido y evaluación de riesgo es una condición mínima. Eso implica transparencia real sobre los datos de entrenamiento, los criterios de clasificación y las tasas de error desagregadas por grupo demográfico. Implica también mecanismos efectivos de apelación cuando una decisión algorítmica genera daño, y no solo formularios que no llevan a ningún lado. Los protocolos institucionales deben reconocer explícitamente que la violencia digital es violencia real que merece una respuesta equivalente a la que se daría ante violencia física. Esto no debería ser controversial, pero en la práctica evaluativa cotidiana lo es.

A nivel de política pública, la agenda es más amplia. Legislación que reconozca la violencia digital como violación de derechos humanos, no como una subcategoría curiosa. Regulación de plataformas que exija moderación efectiva sin el efecto perverso de censurar las denuncias de las propias víctimas. Responsabilidad empresarial por el daño causado por sistemas algorítmicos que operan con sesgos conocidos y no corregidos. Inversión en investigación sobre sesgos algorítmicos en sistemas de justicia latinoamericanos, con poblaciones latinoamericanas, porque usar exclusivamente datos de contextos anglosajones para validar instrumentos que se aplicarán en Bolivia, Colombia o Perú es en sí mismo otra forma de sesgo.

Las estrategias de reparación integral deben abordar los tres niveles de victimización. La reparación del daño primario incluye atención psicológica especializada, apoyo legal y medidas de protección que funcionen. La reparación del daño secundario requiere reconocimiento institucional explícito del daño causado por los propios sistemas de justicia, aunque genere resistencia, es necesario (Rubio-Marín, 2009). La reparación del daño terciario pasa por desnaturalizar la violencia digital en el espacio público, lo que es más difícil de cuantificar pero igual de relevante. Ninguna de estas estrategias es suficiente por sí sola. La transformación requiere acción en todos los niveles al mismo tiempo.

## 7. Conclusiones

El argumento central de este artículo puede resumirse con relativa sencillez: los sesgos cognitivos documentados en evaluación pericial humana no desaparecen cuando

los sistemas de IA asumen funciones institucionales. Se automatizan, se escalan y se vuelven opacos de maneras que los hacen más difíciles de impugnar que cuando los sostenía un profesional individual. Esta automatización constituye una forma de desigualdad política en la medida en que produce barreras estructurales para el acceso a la protección institucional de grupos específicos, y esas barreras no son técnicamente inevitables sino políticamente producidas.

La violencia digital de género expone estas dinámicas con particular claridad porque los sesgos de minimización, culpabilización e incredulidad que ya existían en la evaluación de violencia offline se replican y magnifican en la evaluación de violencia digital. Y los sistemas algorítmicos —diseñados, en general, sin consultar a quienes más los padecen— codifican esos sesgos y los reproducen a escala industrial.

El marco de victimología crítica es útil aquí porque permite identificar el daño en sus tres niveles y, lo que es más importante, permite señalar que el daño institucional y el daño de normalización social no son efectos colaterales tolerables del proceso: son formas de violencia que deben documentarse, reconocerse y repararse.

Una nota de honestidad intelectual para cerrar: buena parte de lo que se propone en este artículo requiere recursos que muchos sistemas judiciales latinoamericanos no tienen, voluntad política que no siempre existe y transformaciones culturales que tardan generaciones. Eso no hace que las propuestas sean menos válidas; hace que la distancia entre el diagnóstico y la solución sea más difícil de ignorar. Los sesgos de IA como formas de desigualdad política no son una distopía futura. Son un presente documentado que está esperando, todavía, las decisiones estructurales que lo transformen.

## Referencias bibliográficas

Asociación Civil Comunicación para la Igualdad. (2021). Violencia de género digital en Argentina: Diagnóstico y recomendaciones. <https://comunicarigualdad.org.ar/wp-content/uploads/2021/04/violencia-digital-2021.pdf>

Bailey, J., & Shayan, M. (2020). Missing the point(s): The declining utility of the legal concept of consent. En A. Powell, N. Henry & A. Flynn (Eds.), *Image-based sexual abuse* (pp. 45-67). Routledge. <https://doi.org/10.4324/9780429938863>

Chávez-Turello, A. L. (2026). *Guía práctica de psicología forense, victimología y derechos humanos: Fundamentos críticos para la práctica pericial en contextos iberoamericanos* (3.<sup>a</sup> ed.). Editorial NUR.

Citron, D. K. (2014). *Hate crimes in cyberspace*. Harvard University Press. <https://doi.org/10.4159/harvard.9780674735613>

Defensoría del Pueblo del Estado Plurinacional de Bolivia. (2022). Informe sobre violencia digital contra mujeres en Bolivia. <https://www.defensoria.gob.bo/uploads/files/informe-violencia-digital-2022.pdf>

de Vogel, V. (2023). Gender-sensitive violence risk assessment. En G. C. Liell, M. J. Fisher & L. F. Jones (Eds.), *Challenging bias in forensic psychological assessment and testing: Theoretical and practical approaches to working with diverse populations* (pp. 299-313). Routledge.

Dragiewicz, M., Burgess, J., Matamoros-Fernández, A., Salter, M., Suzor, N. P., Woodlock, D., & Harris, B. (2018). Technology facilitated coercive control: Domestic violence and the competing roles of digital media platforms. *Feminist Media Studies*, 18(4), 609-625. <https://doi.org/10.1080/14680777.2018.1447341>

Henry, N., & Powell, A. (2018). Technology-facilitated sexual violence: A literature review of empirical research. *Trauma, Violence, & Abuse*, 19(2), 195-208. <https://doi.org/10.1177/1524838016650189>

Henry, N., Powell, A., & Flynn, A. (2020). Not just «revenge pornography»: Australians' experiences of image-based abuse. <https://doi.org/10.13140/RG.2.2.35166.59209>

Kassin, S. M., Dror, I. E., & Kukucka, J. (2013). The forensic confirmation bias: Problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory and Cognition*, 2(1), 42-52. <https://doi.org/10.1016/j.jarmac.2013.01.001>

Levin, A., Fitch, C., Vick, C., Busch, A., & O'Brien, J. (2021). Secondary traumatic stress, burnout, compassion satisfaction, and perceived organizational trauma readiness in forensic science professionals. *Journal of Forensic Sciences*, 66(5), 1645-1655. <https://doi.org/10.1111/1556-4029.14747>

Liell, G. C., Fisher, M. J., & Jones, L. F. (Eds.). (2022). *Challenging bias in forensic psychological assessment and testing: Theoretical and practical approaches to working*

with diverse populations. Routledge. <https://doi.org/10.4324/9781003169352>

McGlynn, C., Rackley, E., & Houghton, R. (2017). Beyond «revenge porn»: The continuum of image-based sexual abuse. *Feminist Legal Studies*, 25(1), 25-46. <https://doi.org/10.1007/s10691-017-9343-2>

ONU Mujeres. (2020). Violencia contra las mujeres y las niñas en el espacio digital: Lo que es virtual también es real. <https://www.unwomen.org/es/digital-library/publications/2020/04/brief-online-and-ict-facilitated-violence-against-women-and-girls-during-covid-19>

Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. New York University Press.

Powell, A., & Henry, N. (2017). Sexual violence in a digital age. Palgrave Macmillan. <https://doi.org/10.1057/978-1-137-58047-4>

Rogers, R. (2008). Clinical assessment of malingering and deception (3rd ed.). Guilford Press.

Rubio-Marín, R. (Ed.). (2009). The gender of reparations: Unsettling sexual hierarchies while redressing human rights violations. Cambridge University Press. <https://doi.org/10.1017/CB09780511581793>

Šimonović, D. (2018). Informe de la Relatora Especial sobre la violencia contra la mujer (A/HRC/38/47). Naciones Unidas. <https://documents.un.org/doc/undoc/gen/g18/184/58/pdf/g1818458.pdf>

Southworth, C., Finn, J., Dawson, S., Fraser, C., & Tucker, S. (2007). Intimate partner violence, technology, and stalking. *Violence Against Women*, 13(8), 842-856. <https://doi.org/10.1177/1077801207302045>

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124-1131. <https://doi.org/10.1126/science.185.4157.1124>

Ussher, J. M. (2011). The madness of women: Myth and experience. Routledge.

Varona, D. (2016). Victimología crítica y sistema penal: Aportes desde el feminismo y los estudios subalternos. Didot.

Vrij, A. (2005). Criteria-Based Content Analysis: A qualitative review of the first 37 studies. *Psychology, Public Policy, and Law*, 11(1), 3-41. <https://doi.org/10.1037/1076-8971.11.1.3>

Weber, M. A., Albrecht, J. N., Endrass, J., Humbel, D., Meier, D. R., Singh, J. P., & Gerth, J. (2024). Hindsight bias in forensic mental health novices and experts: An exploratory study. *Journal of Forensic Psychology Research and Practice*, 25(5), 913-929. <https://doi.org/10.1080/24732850.2024.2396991>

## La ley de hierro de la oligarquía revisitada: élites, instituciones y evolución en la teoría política contemporánea

### *The Iron Law of Oligarchy Revisited: Elites, Institutions, and Evolution in Contemporary Political Theory*

Facundo Guadagno Balmaceda <sup>1</sup>

<sup>1</sup> Universidad de Buenos Aires.

facundo.guadagno@gmail.com (Autor de correspondencia)

**Resumen.** La ley de hierro de la oligarquía, formulada por Robert Michels, ha sido tradicionalmente interpretada como una tesis sobre la inevitabilidad de la concentración del poder en organizaciones complejas. Este ensayo bibliográfico examina críticamente cómo la literatura contemporánea ha abordado —directa o indirectamente— este problema a través de tres enfoques principales: el institucionalismo político-económico, la teoría de la democracia de partidos y la antropología evolutiva. A partir del análisis comparativo de obras clave como *Why Nations Fail* (Acemoglu & Robinson, 2012), *Ruling the Void* (Mair, 2013), *Governing the Commons* (Ostrom, 1990) y *Hierarchy in the Forest* (Boehm, 1999), se argumenta que ninguna de estas tradiciones logra por sí sola explicar plenamente la persistencia de las élites. El trabajo propone una síntesis que incorpora limitaciones cognitivas, escala social y mecanismos evolutivos para reinterpretar la ley de Michels como una regularidad probabilística y no como una ley universal.

**Palabras clave:** oligarquía, élites, instituciones, democracia

**Abstract.** The iron law of oligarchy, formulated by Robert Michels, has traditionally been interpreted as a thesis about the inevitability of power concentration in complex organizations. This bibliographic essay critically examines how contemporary literature has addressed —directly or indirectly— this problem through three main approaches: political-economic institutionalism, party democracy theory, and evolutionary anthropology. Based on a comparative analysis of key works such as *Why Nations Fail*, *Ruling the Void*, *Governing the Commons*, and *Hierarchy in the Forest*, it is argued that none of these traditions alone fully explains the persistence of elites. The paper proposes a synthesis that incorporates cognitive limitations, social scale, and evolutionary mechanisms to reinterpret Michels' law as a probabilistic regularity rather than a universal law.

**Keywords:** oligarchy, elites, institutions, democracy

**Historial:** Recibido: 28 de marzo de 2026 | Aceptado: 15 de mayo de 2026

**Ética/Financiamiento:** El autor declara no tener conflictos de interés. La investigación fue autofinanciada y no recibió financiamiento externo.

Este documento está bajo una licencia  
CC BY-NC-ND 4.0.

**Cómo citar (APA):** Guadagno Balmaceda, F. (2026). La ley de hierro de la oligarquía revisitada: élites, instituciones y evolución en la teoría política contemporánea. *Conocimiento i Política*, 7(1), 286-302. DOI: 10.64480/cip.2026.n7.10.

## 1. Introducción

Desde su formulación en *Political Parties*, la denominada ley de hierro de la oligarquía ha sido considerada uno de los diagnósticos más influyentes acerca de los límites organizacionales de la democracia moderna (Michels, 1911). Su tesis central—según la cual toda organización compleja tiende a concentrar el poder en una minoría dirigente— ha sido ampliamente discutida, reformulada y, en algunos casos, relativizada por distintas corrientes de la teoría política y social. No obstante, una característica persistente de esta recepción ha sido su fragmentación analítica: los desarrollos contemporáneos han tendido a abordar la persistencia de las élites desde marcos institucionales, económicos o normativos sin articular una explicación integrada de los mecanismos subyacentes que hacen posible dicha concentración del poder.

Este ensayo parte de la hipótesis de que dicha fragmentación responde, en gran medida, a la ausencia de una fundamentación en los condicionantes evolutivos y cognitivos de la acción colectiva humana. En efecto, buena parte de la literatura contemporánea ha explicado la persistencia de las élites en términos de incentivos institucionales (Acemoglu & Robinson, 2012), transformaciones en los sistemas de representación política (Mair, 2013) o condiciones de gobernanza local que permiten evitar la centralización (Ostrom, 1990), mientras que solo de manera marginal se ha considerado la evidencia proveniente de la antropología evolutiva sobre la estructura de las jerarquías humanas y los mecanismos de control social en sociedades preestatales (Boehm, 1999).

La consecuencia de esta omisión es teóricamente significativa. Sin una comprensión de las disposiciones psicológicas evolucionadas—tales como la tendencia a la deferencia hacia individuos prestigiosos, la aceptación de la delegación en contextos de complejidad informacional y las limitaciones cognitivas para el monitoreo de grandes grupos—, las explicaciones institucionales resultan incompletas. En este sentido, la persistencia de las élites no puede ser reducida únicamente a arreglos estratégicos o fallas normativas, sino que debe ser entendida como el producto de un desajuste entre mecanismos cognitivos adaptados a entornos de pequeña escala y las exigencias organizacionales de las sociedades modernas (Henrich, 2016; Tomasello, 2014; Dunbar, 1992).

Sobre esta base, el presente ensayo examina críticamente un conjunto de obras influyentes que, aunque no siempre dialogan explícitamente con Michels, abordan el problema de la concentración del poder en contextos contemporáneos. Se argumenta que estas contribuciones, consideradas de manera aislada, ofrecen explicaciones parciales, y que solo mediante su relectura desde un marco evolutivo es posible reconstruir una interpretación más robusta de la ley de hierro de la oligarquía. En lugar de concebirla como una ley universal e invariable, se propone entenderla como una regularidad emergente cuya probabilidad aumenta en función de la escala social, la complejidad organizacional y los límites cognitivos de los agentes.

## **2. El institucionalismo y la persistencia de las élites: alcances y límites desde una perspectiva evolutiva**

En *Why Nations Fail*, Acemoglu y Robinson (2012) desarrollan una de las explicaciones contemporáneas más influyentes sobre la persistencia de las élites, centrada en la distinción entre instituciones inclusivas y extractivas. Según su argumento, las élites logran reproducir su posición dominante mediante arreglos institucionales que concentran el poder político y económico, generando incentivos para bloquear transformaciones que puedan amenazar su control. En este marco, la desigual distribución del poder no aparece como una anomalía, sino como el resultado lógico de equilibrios estratégicos sostenidos en el tiempo, en los cuales las élites utilizan las instituciones para perpetuar su dominio (Acemoglu & Robinson, 2012).

Esta perspectiva presenta una afinidad evidente con la intuición de Michels, en la medida en que reconoce la tendencia de las organizaciones complejas a producir estructuras de poder concentradas y relativamente autónomas respecto de las bases sociales que, en principio, deberían controlar. Sin embargo, la explicación institucionalista introduce un desplazamiento significativo: la oligarquización deja de ser concebida como una dinámica interna de la organización y pasa a ser interpretada como el resultado de incentivos estratégicos mediados por reglas formales e informales.

No obstante, este desplazamiento explicativo tiene un costo analítico importante. Al privilegiar el nivel institucional, el enfoque de Acemoglu y Robinson tiende a presuponer, más que a explicar, la disposición de los actores a aceptar, reproducir o no desafiar estructuras de poder altamente desiguales. En otras palabras, si bien el

modelo da cuenta de cómo las élites se mantienen en el poder, resulta menos convincente al momento de explicar por qué los no-élites no logran, de manera sistemática, revertir dichos arreglos, incluso cuando enfrentan incentivos materiales para hacerlo. Este punto remite a un problema clásico de acción colectiva ya señalado por Olson (1965), pero que en este contexto adquiere una dimensión adicional vinculada a las capacidades cognitivas de los individuos.

Desde una perspectiva evolutiva, esta limitación se vuelve particularmente evidente. La literatura en psicología y antropología ha mostrado que los seres humanos no operan como agentes plenamente racionales capaces de monitorear y desafiar estructuras complejas de poder, sino como individuos con capacidades cognitivas limitadas, adaptadas a entornos de pequeña escala y a interacciones cara a cara (Simon, 1957; Dunbar, 1992; Tomasello, 2014). En este sentido, la delegación de autoridad, la deferencia hacia figuras percibidas como competentes y la dificultad para procesar información institucional compleja no son meros epifenómenos, sino condiciones estructurales de la acción colectiva en sociedades extensas.

En consecuencia, la persistencia de las élites no puede ser comprendida exclusivamente como un equilibrio estratégico sostenido por incentivos institucionales, sino que debe ser interpretada como el resultado de la interacción entre dichos incentivos y un conjunto de disposiciones cognitivas evolucionadas que facilitan la concentración del poder. Desde esta perspectiva, el institucionalismo aporta una descripción robusta de los mecanismos de reproducción de las élites, pero carece de una microfundación adecuada que explique por qué dichos mecanismos resultan sistemáticamente efectivos. Es precisamente en este punto donde la intuición de Michels puede ser reformulada y fortalecida mediante su integración con la evidencia proveniente de las ciencias del comportamiento y la evolución humana.

### **3. Crisis de la representación y autonomización de las élites: una lectura evolutiva de Mair**

En *Ruling the Void*, Peter Mair (2013) describe un proceso de transformación de los partidos políticos en el cual estas organizaciones dejan progresivamente de funcionar como canales de representación social para convertirse en estructuras autónomas, crecientemente integradas al aparato estatal y desconectadas de sus bases. Este fenó-

meno, que el autor caracteriza como una “democracia sin pueblo”, implica no solo una mutación institucional, sino una reconfiguración más profunda de la relación entre ciudadanos y élites políticas. Los partidos ya no median demandas sociales de manera efectiva, sino que operan como organizaciones cerradas, profesionalizadas y orientadas a su propia reproducción (Mair, 2013; Katz & Mair, 1995).

A primera vista, el diagnóstico de Mair puede ser interpretado como una confirmación empírica contemporánea de la intuición de Michels: la organización política, en la medida en que se vuelve más compleja, tiende a generar una élite dirigente que se autonomiza respecto de la base social que le dio origen. Sin embargo, al igual que ocurre con el institucionalismo, la explicación ofrecida por Mair permanece fundamentalmente en el nivel meso de las transformaciones organizacionales e institucionales, sin desarrollar una teoría de los mecanismos cognitivos que hacen posible dicha autonomización.

Este punto es crucial. El proceso mediante el cual las élites partidarias se distancian de la ciudadanía no puede ser comprendido únicamente como una deriva organizacional o como una estrategia consciente de los actores políticos. Requiere, además, explicar por qué ese distanciamiento no es sistemáticamente revertido por los propios ciudadanos, incluso en contextos donde los canales formales de participación permanecen abiertos. La respuesta a este interrogante remite nuevamente a las limitaciones cognitivas y a las disposiciones psicológicas evolucionadas de los individuos.

Desde una perspectiva evolutiva, la relación entre ciudadanos y organizaciones políticas masivas presenta un desajuste estructural. Los mecanismos de evaluación social y control colectivo que caracterizaron a los entornos en los cuales evolucionó la especie humana estaban adaptados a grupos pequeños, donde la información sobre el comportamiento de los demás era accesible y las sanciones podían aplicarse de manera directa (Boehm, 1999; Dunbar, 1992). En contraste, las democracias contemporáneas operan en contextos de alta complejidad y anonimato, donde los individuos deben procesar grandes volúmenes de información abstracta y confiar en intermediarios para la toma de decisiones. En este escenario, la tendencia a delegar, a simplificar juicios mediante heurísticas y a otorgar deferencia a figuras percibidas como competentes reduce significativamente la capacidad de control efectivo sobre las élites (Kahneman, 2011; Henrich, 2016).

En consecuencia, lo que Mair describe como un vaciamiento de la democracia puede

ser reinterpretado no solo como una transformación institucional, sino como el resultado de la interacción entre estructuras organizacionales complejas y limitaciones cognitivas profundamente arraigadas. La autonomización de las élites partidarias no es simplemente una desviación del ideal democrático, sino una manifestación de la dificultad inherente que enfrentan los individuos para monitorear y disciplinar organizaciones que exceden los límites de la cooperación evolucionada. Desde esta perspectiva, la contribución de Mair resulta fundamental para describir el fenómeno, pero insuficiente para explicarlo en sus fundamentos, lo que refuerza la necesidad de integrar su diagnóstico en un marco más amplio que incorpore la dimensión evolutiva de la acción colectiva.

#### **4. Ostrom y los límites evolutivos de la gobernanza no oligárquica**

La obra de Elinor Ostrom constituye, a primera vista, uno de los desafíos más consistentes a la tesis de la inevitabilidad oligárquica formulada por Michels. En *Governing the Commons*, Ostrom (1990) demuestra empíricamente que diversas comunidades son capaces de gestionar recursos de uso común sin recurrir a estructuras jerárquicas centralizadas ni a mecanismos coercitivos externos. A partir de un análisis comparativo de casos, la autora identifica arreglos institucionales en los cuales los propios usuarios establecen reglas, monitorean su cumplimiento y sancionan desviaciones, logrando así niveles sostenidos de cooperación sin la consolidación de élites dominantes.

Leída en términos estrictamente institucionales, esta evidencia parecería refutar la ley de hierro de la oligarquía, al mostrar que la coordinación colectiva puede sostenerse sin derivar necesariamente en la concentración del poder. Sin embargo, una lectura más detenida revela que el alcance de esta conclusión es más limitado de lo que su formulación sugiere. En efecto, las condiciones bajo las cuales estos sistemas de gobernanza resultan estables remiten de manera sistemática a configuraciones sociales de pequeña escala, con alta densidad de interacción, conocimiento mutuo entre los participantes y posibilidades efectivas de monitoreo directo (Ostrom, 1990; Ostrom, 2005). Estas características no constituyen meros atributos contingentes de los casos estudiados, sino que definen el entorno en el cual tales arreglos institucionales pueden operar sin generar dinámicas de concentración del poder.

Desde una perspectiva evolutiva, este punto adquiere una relevancia decisiva. Las

condiciones identificadas por Ostrom coinciden estrechamente con aquellas bajo las cuales evolucionaron los mecanismos humanos de cooperación y control social: grupos relativamente pequeños, interacciones repetidas y capacidad de sanción directa sobre comportamientos desviados (Boehm, 1999; Tomasello, 2014). En este sentido, la aparente excepción que representan los sistemas de gobernanza policéntrica no constituye una refutación de la dinámica oligárquica, sino más bien una indicación de sus límites estructurales. Allí donde las relaciones sociales permanecen dentro de los umbrales cognitivos que permiten el seguimiento individualizado de la conducta —umbral que ha sido estimado en torno a las redes sociales manejables por los individuos (Dunbar, 1992)—, los mecanismos igualitarios pueden sostenerse y bloquear la emergencia de élites estables.

El problema surge cuando estas condiciones dejan de cumplirse. A medida que aumenta la escala de la organización, se incrementa la complejidad informacional, se diluyen las relaciones cara a cara y se vuelve inviable el monitoreo directo, los mecanismos de control horizontal pierden eficacia. En ese contexto, la delegación de funciones, la especialización y la concentración de información generan las condiciones para la estabilización de élites, reproduciendo precisamente la dinámica identificada por Michels. Desde esta perspectiva, la contribución de Ostrom no invalida la ley de hierro de la oligarquía, sino que permite delimitar con mayor precisión su ámbito de aplicación: la oligarquización no es un destino inevitable en abstracto, pero se vuelve altamente probable cuando las formas de organización social exceden los límites cognitivos y relacionales para los cuales evolucionaron los mecanismos humanos de cooperación.

En consecuencia, la teoría de Ostrom puede ser reinterpretada no como una alternativa normativa a Michels, sino como una evidencia empírica de que la posibilidad de evitar la concentración del poder depende críticamente de variables que, en última instancia, tienen una base evolutiva. La gobernanza sin élites no es imposible, pero su viabilidad está restringida a contextos donde la escala y la complejidad no desbordan las capacidades cognitivas de los individuos. Allí donde ese umbral es superado, las condiciones que permiten la cooperación horizontal tienden a erosionarse, abriendo el camino a la formación de estructuras oligárquicas.

## 5. Escala, complejidad y desajuste evolutivo: condiciones estructurales de la oligarquización

Uno de los aspectos más sistemáticamente subestimados en las discusiones contemporáneas sobre la ley de hierro de la oligarquía es el papel de la escala organizacional como variable explicativa central. Mientras que las interpretaciones clásicas de Michels enfatizan la dinámica interna de las organizaciones y los enfoques institucionalistas privilegian los incentivos estratégicos, ambos tienden a pasar por alto una restricción fundamental: las capacidades cognitivas de los individuos que participan en dichas estructuras. En este sentido, la expansión de la escala social no constituye simplemente un cambio cuantitativo, sino una transformación cualitativa que altera las condiciones mismas de posibilidad del control democrático.

La literatura en antropología y psicología evolutiva ha mostrado que los seres humanos evolucionaron en contextos de interacción relativamente acotados, donde el número de relaciones sociales estables que un individuo puede mantener se encuentra limitado por restricciones cognitivas específicas. El llamado “número de Dunbar”, estimado en torno a las 150 relaciones, no debe interpretarse como una cifra rígida, sino como un indicador de la capacidad humana para sostener redes sociales basadas en el conocimiento directo, la reputación y el monitoreo interpersonal (Dunbar, 1992). En estos entornos, la cooperación se encuentra sostenida por mecanismos informales de control social que dependen de la visibilidad de la conducta y de la posibilidad de aplicar sanciones de manera inmediata.

El pasaje a sociedades de gran escala implica la erosión progresiva de estas condiciones. A medida que aumenta el tamaño de las poblaciones y la complejidad de las interacciones, las relaciones dejan de estar estructuradas por el conocimiento directo y pasan a depender de sistemas abstractos de coordinación. Este proceso introduce niveles crecientes de anonimato, reduce la capacidad de los individuos para evaluar el comportamiento de otros actores y dificulta la aplicación de mecanismos informales de sanción. En este contexto, la necesidad de estructuras organizativas capaces de gestionar la complejidad —incluyendo formas de burocracia, especialización funcional y delegación de autoridad— emerge no como una elección contingente, sino como una condición prácticamente inevitable de la coordinación social (North, 1990; Tilly, 1990).

La división del trabajo, en particular, desempeña un papel decisivo en este proceso.

Como ya había señalado Durkheim (1893), el aumento de la especialización funcional transforma la naturaleza de la cohesión social, desplazándola desde formas basadas en la similitud hacia formas basadas en la interdependencia. Sin embargo, este mismo proceso genera asimetrías crecientes en el acceso a la información y en la capacidad de toma de decisiones. En organizaciones complejas, los individuos dependen de expertos y de actores especializados para comprender y gestionar ámbitos específicos de la vida social, lo que introduce una brecha estructural entre quienes poseen conocimiento técnico y quienes no. Esta brecha, lejos de ser neutral, constituye una de las bases más sólidas para la consolidación de élites, en la medida en que el control del conocimiento se traduce en capacidad de influencia y de dirección (Simon, 1957).

Desde una perspectiva evolutiva, estas transformaciones pueden ser interpretadas como un desajuste entre las condiciones bajo las cuales evolucionaron los mecanismos humanos de cooperación y las exigencias de las sociedades modernas. Los individuos se ven obligados a operar en entornos que exceden sus capacidades cognitivas para el monitoreo directo, lo que los lleva a depender de intermediarios, a simplificar la información mediante heurísticas y a delegar decisiones en actores percibidos como más competentes o mejor informados. Este conjunto de disposiciones, lejos de ser contingente, está profundamente arraigado en la arquitectura cognitiva humana (Kahneman, 2011; Henrich, 2016).

En consecuencia, la relación entre escala, complejidad y oligarquización no puede ser entendida únicamente en términos institucionales o funcionales, sino que debe ser concebida como el resultado de una interacción entre estructuras sociales y limitaciones cognitivas evolucionadas. La concentración del poder no emerge simplemente porque las élites diseñan instituciones que la favorecen, sino porque los individuos que participan en sistemas de gran escala carecen de los recursos cognitivos necesarios para sostener mecanismos efectivos de control horizontal. Desde esta perspectiva, la ley de Michels adquiere una nueva formulación: no como una afirmación normativa sobre la inevitabilidad de la dominación, sino como una consecuencia estructural del desajuste entre la escala de las organizaciones modernas y las capacidades cognitivas de los agentes que las integran.

## 6. Instituciones y trayectorias de poder: una reinterpretación desde las restricciones cognitivas

La literatura institucionalista ha desarrollado un conjunto de herramientas conceptuales particularmente influyentes para explicar la persistencia de estructuras de poder en el tiempo, entre las cuales destacan las nociones de dependencia de trayectoria, captura institucional y costos de monitoreo. En términos generales, estas aproximaciones coinciden en señalar que las instituciones no solo estructuran el comportamiento de los actores, sino que también tienden a reproducir configuraciones previas de poder, generando dinámicas de estabilidad que dificultan transformaciones profundas (Pierson, 2000). Desde esta perspectiva, la concentración del poder no es un accidente, sino el resultado de procesos acumulativos en los cuales decisiones iniciales condicionan de manera creciente las opciones disponibles en el futuro.

Sin embargo, aunque este enfoque permite describir con notable precisión la persistencia de las élites, deja abierta una cuestión fundamental: por qué estos mecanismos logran sostenerse en el tiempo sin ser sistemáticamente desafiados por los actores que resultan perjudicados por ellos. La noción de captura institucional, tal como ha sido desarrollada en la economía política, sugiere que las élites pueden apropiarse de recursos organizacionales y moldear reglas a su favor (Stigler, 1971; Acemoglu & Robinson, 2012). No obstante, esta explicación supone implícitamente que los actores no-élites carecen de la capacidad o de la coordinación necesaria para revertir dichas dinámicas, sin indagar en las razones más profundas de esta limitación.

El problema se vuelve aún más evidente cuando se consideran los costos asociados al monitoreo y control de las instituciones. La tradición inaugurada por Olson (1965) ha mostrado que la acción colectiva enfrenta obstáculos significativos cuando los beneficios son difusos y los incentivos a la participación individual son bajos. Ostrom (1990), por su parte, ha demostrado que bajo ciertas condiciones estos problemas pueden ser mitigados, pero también ha señalado que dichos arreglos requieren niveles de información, proximidad social y capacidad de sanción que no siempre están disponibles en contextos de gran escala. En conjunto, estos enfoques convergen en la idea de que el control democrático no es gratuito, sino que implica costos crecientes a medida que aumenta la complejidad organizacional.

Desde una perspectiva evolutiva, estos desarrollos pueden ser reinterpretados como

manifestaciones institucionales de una restricción más profunda: los límites cognitivos de los individuos para procesar información compleja, coordinar acciones colectivas y sostener mecanismos de vigilancia en entornos de gran escala. La dependencia de trayectoria no solo refleja la inercia de las instituciones, sino también la tendencia de los individuos a apoyarse en estructuras existentes como atajos cognitivos frente a la complejidad. La captura institucional no es únicamente el resultado de estrategias deliberadas de las élites, sino también de la asimetría sistemática entre quienes concentran información y quienes deben confiar en ella. Del mismo modo, los costos de monitoreo no constituyen simplemente una variable económica, sino una expresión de las limitaciones humanas para sostener niveles elevados de atención, coordinación y sanción en contextos donde las interacciones son impersonales y mediadas por estructuras abstractas.

En este sentido, las categorías centrales del institucionalismo pueden ser integradas en un marco más amplio que las reubique como efectos derivados de la interacción entre estructuras organizacionales complejas y capacidades cognitivas limitadas. La persistencia de las élites no se explica únicamente por la existencia de instituciones que la favorecen, sino por el hecho de que dichas instituciones operan en entornos donde los individuos no pueden monitorear eficazmente su funcionamiento ni coordinar respuestas colectivas sostenidas. Desde esta perspectiva, la ley de Michels encuentra un fundamento más robusto: la reproducción de las élites no depende exclusivamente de su capacidad para diseñar reglas a su favor, sino también de las condiciones cognitivas que hacen que esas reglas sean difícilmente impugnables.

## **7. Vacíos analíticos y la necesidad de una microfundación evolutiva**

A pesar de la sofisticación alcanzada por las distintas tradiciones que han abordado el problema de la persistencia de las élites, el recorrido realizado en las secciones anteriores permite identificar una limitación común que atraviesa estos enfoques: la ausencia de una articulación sistemática entre los niveles de análisis en los que operan sus explicaciones. Tanto el institucionalismo como la teoría de la representación política y los estudios sobre gobernanza ofrecen descripciones detalladas de los mecanismos mediante los cuales se reproduce el poder en contextos organizacionales complejos, pero tienden a oscilar entre niveles macro y meso sin proporcionar una microfundación robusta que dé cuenta de los comportamientos individuales que

sostienen dichos procesos.

Esta desconexión se vuelve particularmente problemática en el caso de la ley de hierro de la oligarquía, cuya fuerza explicativa depende precisamente de la capacidad para vincular dinámicas organizacionales con disposiciones individuales. En ausencia de este puente, la concentración del poder aparece alternativamente como el resultado de incentivos institucionales, transformaciones estructurales o estrategias de actores colectivos, sin que se esclarezcan los mecanismos cognitivos que hacen posible la estabilidad de dichas configuraciones. En este sentido, la literatura revisada tiende a presuponer que los individuos pueden, en principio, monitorear, evaluar y eventualmente desafiar a las élites, pero que no lo hacen debido a restricciones externas o fallas de coordinación. Sin embargo, esta suposición pasa por alto que tales capacidades no son ilimitadas, sino que están condicionadas por la arquitectura cognitiva de los agentes.

La omisión de una teoría explícita de la cognición política constituye, en este marco, un vacío particularmente significativo. Los avances en psicología cognitiva y ciencias del comportamiento han mostrado que los individuos operan bajo condiciones de racionalidad limitada, utilizan heurísticas para procesar información compleja y tienden a delegar decisiones en contextos de incertidumbre (Kahneman, 2011; Simon, 1957). Estas disposiciones no son meros desvíos respecto de un ideal de racionalidad plena, sino adaptaciones funcionales a entornos donde la información es incompleta y los costos de procesamiento son elevados. En consecuencia, la dificultad para sostener un control efectivo sobre las élites no puede ser explicada únicamente en términos de fallas institucionales, sino que debe ser entendida como una consecuencia directa de las limitaciones cognitivas que caracterizan a los individuos.

Más aún, la escasa incorporación de la perspectiva evolutiva en estos debates limita la capacidad explicativa de la teoría política contemporánea. La evidencia acumulada en antropología y evolución cultural sugiere que los mecanismos de cooperación, sanción y control social que caracterizan a las sociedades humanas fueron moldeados en contextos de pequeña escala, donde las interacciones eran repetidas y la información sobre el comportamiento de los demás era accesible (Boehm, 1999; Henrich, 2016; Tomasello, 2014). La traslación de estos mecanismos a entornos de gran escala introduce tensiones que no pueden ser resueltas únicamente mediante arreglos institucionales. En este sentido, la persistencia de las élites puede ser interpretada como el resultado de un desajuste entre las condiciones bajo las cuales evolucionaron

las capacidades humanas de cooperación y las exigencias organizacionales de las sociedades modernas.

En consecuencia, los vacíos identificados no constituyen meras omisiones empíricas o conceptuales, sino que remiten a la necesidad de reformular el problema en un nivel más fundamental. Integrar los aportes del institucionalismo, la teoría de la representación y la antropología política requiere incorporar una microfundación que dé cuenta de las capacidades y limitaciones de los individuos que participan en las estructuras analizadas. Solo a partir de esta integración es posible avanzar hacia una reinterpretación de la ley de Michels que no se limite a describir la concentración del poder, sino que explique por qué dicha concentración emerge de manera recurrente en contextos de creciente complejidad social.

## **8. Reinterpretación de la ley de hierro de la oligarquía: una microfundación evolutiva**

El recorrido previo permite sostener que la principal limitación de las interpretaciones contemporáneas de la ley de hierro de la oligarquía no reside en la falta de evidencia empírica sobre la persistencia de las élites, sino en la ausencia de un marco teórico capaz de integrar dicha evidencia en un nivel explicativo más fundamental. En efecto, tanto el institucionalismo como la teoría de la representación y los estudios sobre gobernanza han logrado describir con precisión los mecanismos mediante los cuales se reproduce el poder en contextos organizacionales complejos, pero han tendido a tratar estos mecanismos como autosuficientes, sin indagar en las condiciones cognitivas que hacen posible su funcionamiento.

Desde esta perspectiva, la tesis de Michels puede ser reformulada no como una ley universal ni como una mera generalización empírica, sino como la expresión de un desajuste estructural entre las capacidades cognitivas de los individuos y la escala de las organizaciones en las que participan. Los seres humanos evolucionaron en entornos caracterizados por interacciones cara a cara, redes sociales limitadas y posibilidades efectivas de monitoreo directo, donde los mecanismos de control colectivo podían operar con relativa eficacia (Boehm, 1999; Dunbar, 1992; Tomasello, 2014). En estos contextos, la acumulación estable de poder encontraba obstáculos sistemáticos, ya que los intentos de dominación podían ser detectados y sancionados

por el grupo.

El pasaje a sociedades de gran escala introduce una transformación cualitativa en estas condiciones. La complejidad organizacional, la división del trabajo y la abstracción creciente de las relaciones sociales generan entornos en los cuales los individuos ya no pueden observar directamente el comportamiento de quienes toman decisiones, ni evaluar de manera independiente la información necesaria para ejercer un control efectivo. En consecuencia, se ven obligados a recurrir a mecanismos de delegación, a confiar en intermediarios y a simplificar la información mediante heurísticas que reducen la carga cognitiva (Simon, 1957; Kahneman, 2011). Este proceso no es contingente, sino estructural: constituye una respuesta adaptativa a entornos que exceden las capacidades de procesamiento individuales.

En este marco, la concentración del poder puede ser entendida como un resultado emergente de la interacción entre estructuras organizacionales complejas y limitaciones cognitivas profundamente arraigadas. Las élites no solo se consolidan porque logran capturar instituciones o diseñar reglas a su favor, sino porque operan en un entorno en el cual los individuos carecen de los recursos cognitivos necesarios para monitorear su comportamiento de manera sistemática. La asimetría en el acceso a la información, la especialización técnica y la distancia social no son simplemente variables institucionales, sino condiciones que amplifican las limitaciones inherentes de los agentes.

Así reformulada, la ley de Michels deja de ser una afirmación categórica sobre la inevitabilidad de la oligarquía y pasa a ser entendida como una regularidad estructural cuya intensidad varía en función de la relación entre escala organizacional y capacidades cognitivas. En contextos donde esta relación se mantiene dentro de los umbrales que permiten el monitoreo directo y la sanción efectiva —como en los casos analizados por Ostrom (1990)—, la concentración del poder puede ser contenida o revertida. Sin embargo, a medida que las organizaciones exceden dichos umbrales, las condiciones que hacen posible el control horizontal tienden a erosionarse, incrementando la probabilidad de que emerjan estructuras oligárquicas relativamente estables.

## 9. **Discusión: integración teórica y alcances de la reinterpretación**

La reinterpretación propuesta permite integrar en un marco común contribuciones que, en la literatura, han sido tratadas como alternativas o incluso como posiciones incompatibles. La aparente tensión entre la inevitabilidad de la oligarquía sugerida por Michels, la evidencia de gobernanza horizontal documentada por Ostrom y el igualitarismo activo descrito por Boehm puede ser resuelta si se introduce la escala como variable mediadora y se reconocen las limitaciones cognitivas como condición estructural de la acción colectiva.

Desde esta perspectiva, las diferencias entre estos enfoques no remiten a desacuerdos empíricos fundamentales, sino a la observación de fenómenos que operan bajo condiciones distintas. Las sociedades de pequeña escala y alta densidad relacional pueden sostener mecanismos igualitarios precisamente porque se encuentran dentro de los límites para los cuales evolucionaron las capacidades humanas de cooperación y control social. En cambio, las organizaciones modernas, caracterizadas por su tamaño, complejidad y grado de abstracción, generan condiciones en las cuales dichos mecanismos pierden eficacia, favoreciendo la consolidación de élites.

Este enfoque permite, además, reinterpretar las categorías del institucionalismo como expresiones de restricciones más profundas. La dependencia de trayectoria, la captura institucional y los costos de monitoreo no constituyen explicaciones últimas, sino manifestaciones de la dificultad que enfrentan los individuos para intervenir eficazmente en estructuras complejas. En este sentido, la persistencia de las élites no puede ser atribuida exclusivamente a su capacidad estratégica, sino que debe ser entendida como el resultado de un entorno en el cual las condiciones cognitivas favorecen la delegación y dificultan el control.

## 10. **Conclusión**

La ley de hierro de la oligarquía ha sido tradicionalmente interpretada como una afirmación normativa sobre los límites de la democracia o como una generalización empírica derivada de la observación de organizaciones políticas específicas. Sin embargo, el análisis desarrollado en este ensayo sugiere que su verdadero alcance solo puede ser comprendido si se la sitúa en un marco que integre la dimensión

evolutiva de la acción humana.

La tendencia a la concentración del poder no deriva de una supuesta inclinación inmutable hacia la dominación, sino de la interacción entre estructuras organizacionales que exceden los límites de la cooperación directa y capacidades cognitivas adaptadas a entornos de pequeña escala. En este sentido, la oligarquización no es inevitable en abstracto, pero se vuelve altamente probable en contextos donde la complejidad social impide el funcionamiento efectivo de los mecanismos igualitarios.

Reformular la tesis de Michels en estos términos permite no solo clarificar sus condiciones de validez, sino también abrir un programa de investigación orientado a identificar qué arreglos institucionales pueden mitigar —aunque difícilmente eliminar por completo— las dinámicas de concentración del poder en sociedades complejas. Lejos de invalidar la intuición original, esta reinterpretación la sitúa en un nivel explicativo más profundo, donde la persistencia de las élites aparece como una consecuencia estructural del desajuste entre la evolución humana y las formas contemporáneas de organización social.

## Referencias

Acemoglu, D., & Robinson, J. A. (2012). *Why nations fail: The origins of power, prosperity, and poverty*. Crown Business.

Boehm, C. (1999). *Hierarchy in the forest: The evolution of egalitarian behavior*. Harvard University Press.

Dunbar, R. I. M. (1992). Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, 22(6), 469–493. [https://doi.org/10.1016/0047-2484\(92\)90081-J](https://doi.org/10.1016/0047-2484(92)90081-J)

Durkheim, É. (1893). *De la division du travail social*. Alcan.

Henrich, J. (2016). *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter*. Princeton University Press.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Katz, R. S., & Mair, P. (1995). Changing models of party organization and party

democracy: The emergence of the cartel party. *Party Politics*, 1(1), 5–28. <https://doi.org/10.1177/1354068895001001001>

Mair, P. (2013). *Ruling the void: The hollowing of Western democracy*. Verso.

Michels, R. (1911). *Political parties: A sociological study of the oligarchical tendencies of modern democracy*. Free Press. (Edición original publicada en 1911)

North, D. C. (1990). *Institutions, institutional change and economic performance*. Cambridge University Press.

Olson, M. (1965). *The logic of collective action: Public goods and the theory of groups*. Harvard University Press.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.

Ostrom, E. (2005). *Understanding institutional diversity*. Princeton University Press.

Pierson, P. (2000). Increasing returns, path dependence, and the study of politics. *American Political Science Review*, 94(2), 251–267. <https://doi.org/10.2307/2586011>

Simon, H. A. (1957). *Administrative behavior: A study of decision-making processes in administrative organization* (2nd ed.). Macmillan.

Stigler, G. J. (1971). The theory of economic regulation. *The Bell Journal of Economics and Management Science*, 2(1), 3–21. <https://doi.org/10.2307/3003160>

Tilly, C. (1990). *Coercion, capital, and European states, AD 990–1990*. Blackwell.

Tomasello, M. (2014). *A natural history of human thinking*. Harvard University Press.

# NOTAS DE COMUNICACIÓN CIENTÍFICA

---

Conocimiento i Política – No. 7, 2026

## Efectos de la propaganda electoral en los estudiantes durante las elecciones universitarias UAGRM (2025)

Mariano Ramos Claire<sup>1</sup>

<sup>1</sup> Universidad Autónoma Gabriel René Moreno. Facultad de Ciencias Jurídicas, Políticas, Sociales y Relaciones Internacionales. Carrera de Ciencia Política y Administración Pública. Santa Cruz de la Sierra, Bolivia.  
ramosm8124@gmail.com (Autor de correspondencia)

**Resumen.** La propaganda electoral universitaria, lejos de limitarse a un fenómeno comunicativo pasajero, imprime efectos concretos y medibles sobre el entorno y el bienestar de los estudiantes. Este artículo analiza dichos efectos en la Universidad Autónoma Gabriel René Moreno (UAGRM) durante las elecciones de autoridades principales de 2025, a través de una encuesta estratificada aplicada a 383 estudiantes de ocho facultades (n=383; IC=95%; e=5%). Los resultados evidencian que, aunque una mayoría relativa reconoce algún grado de legitimidad en la comunicación política electoral, la comunidad estudiantil percibe la propaganda predominantemente como perjudicial: la contaminación sonora con altavoces de capacidad de 80–110 dB frente a los 35 dB recomendados por la OMS (2011) para entornos de aprendizaje y la saturación visual del campus emergen como sus efectos más perturbadores (Mera-Benavides, 2017; Gao et al., 2024). Paradójicamente, pese a la magnitud del despliegue propagandístico, solo el 29,8% de los encuestados reportó alguna influencia sobre su decisión de voto, mientras que el desperdicio de recursos fue señalado como la principal objeción. Los hallazgos exponen vacíos normativos en el Reglamento Electoral vigente (UAGRM, 2024) y sientan una base empírica para futuras investigaciones sobre democracia universitaria, bienestar estudiantil y gestión ambiental del campus.

**Palabras clave:** propaganda electoral universitaria; contaminación visual; contaminación sonora; bienestar estudiantil; democracia universitaria; UAGRM.

**Abstract.** University electoral propaganda, far from being a fleeting communicative phenomenon, produces concrete and measurable effects on the campus environment and student well-being. This article analyzes these effects at the Autonomous University Gabriel René Moreno (UAGRM) during the 2025 elections for principal authorities, through a stratified survey applied to 383 students across eight faculties (n=383; CI=95%; e=5%). Results show that, although a relative majority acknowledges some legitimacy in electoral political communication, students predominantly perceive propaganda as harmful: noise pollution, with loudspeakers with a capacity of 80–110 dB compared with the 35 dB recommended by the WHO (2011) for learning environments, and visual saturation of the campus emerge as its most disruptive effects (Mera-Benavides, 2017; Gao et al., 2024). Paradoxically, despite the scale of the campaign, only 29.8% of respondents reported any influence on their voting decision, while resource waste was cited as the primary objection. The findings expose normative gaps in the current Electoral Regulations (UAGRM, 2024) and establish an empirical foundation for future research on university democracy, student well-being, and campus environmental management.

**Keywords:** university electoral propaganda; visual pollution; noise pollution; student well-being; university democracy; UAGRM.

**Historial:** Recibido: 30 de marzo de 2026 | Aceptado: 18 de mayo de 2026

**Ética/Financiamiento:** El autor declara no tener conflictos de interés. El artículo fue autofinanciado y no recibió financiamiento externo.

Este documento está bajo una licencia  
CC BY-NC-ND 4.0.

**Cómo citar (APA):** Ramos Claire, M. (2026). Efectos de la propaganda electoral en los estudiantes durante las elecciones universitarias UAGRM (2025). *Conocimiento i Política*, 7(1), 304-315. DOI: 10.64480/cip.2026.n7.11.

## 1. Introducción

En el marco de los procesos electorales contemporáneos, la propaganda política constituye uno de los instrumentos de comunicación más antiguos y, al mismo tiempo, más debatidos de la democracia representativa. Vera Leyva (2022) la define como un conjunto de técnicas informativo-persuasivas empleadas por individuos, grupos candidatos y partidos con el propósito de atraer la voluntad del electorado; una práctica que ha evolucionado de forma notable desde sus formas más rudimentarias hasta la saturación multicanal que caracteriza a las campañas del siglo XXI. Sin embargo, cuando esta lógica de campaña se traslada al interior de una institución educativa de nivel superior, emerge una tensión estructural que la literatura especializada ha comenzado a documentar con creciente interés: la colisión entre el ejercicio legítimo de la comunicación política y la función primordial del espacio universitario como entorno de formación académica e intelectual.

La Universidad Autónoma Gabriel René Moreno (UAGRM), la institución de educación superior más grande de Bolivia oriental, con una matrícula aproximada de 100.000 estudiantes, no es ajena a esta tensión. Sus procesos electorales para la designación de autoridades principales: rector, vicerrector, decanos y directores de carrera, dan lugar, ciclo tras ciclo, a una intensa movilización propagandística que ocupa físicamente patios, pasillos, aulas y accesos del campus. Carteles, banderines, banderas, pinturas sobre superficies, papelería adherida a mano y, de manera particularmente notoria, música reproducida a alto volumen mediante equipos de amplificación potentes, conforman el paisaje sensorial del campus durante las semanas previas a la votación. Este despliegue ocurre, además, en un marco normativo con vacíos significativos: el Reglamento Electoral vigente (UAGRM, 2024) se limita en sus artículos 116, 120 y 122 a exigir el cuidado del ornato, los bienes y la infraestructura general, sin establecer parámetros específicos de intensidad sonora, densidad visual o plazos de retiro del material propagandístico.

Desde la perspectiva de la salud ambiental y el bienestar estudiantil, esta situación no es trivial. La investigación acumulada sobre contaminación visual señala que la proliferación desordenada de elementos visuales en el espacio público y, por extensión, en el espacio universitario genera estados de sobrecarga cognitiva, deterioro de la atención y efectos negativos sobre el ánimo y la salud psicológica de quienes lo habitan (Gao et al., 2024). Mera-Benavides (2017), en un estudio pionero en el contexto

latinoamericano, documentó que la contaminación visual percibida por la comunidad universitaria trasciende lo meramente estético para convertirse en un factor que afecta el bienestar psicológico y la calidad del entorno de residencia y estudio. A esta dimensión visual se añade la dimensión acústica: Ali et al. (2023) demostraron que la exposición prolongada a niveles de ruido elevados en entornos universitarios tiene consecuencias fisiológicas, psicológicas y académicas medibles, lo que adquiere particular relevancia cuando los equipos de sonido utilizados en la campaña tienen una capacidad que triplica o cuadruplica el umbral de 35 dB recomendado por la OMS (2011) para los espacios destinados al aprendizaje.

A pesar de su relevancia práctica, los efectos de la propaganda en contextos electorales específicamente universitarios han recibido atención académica limitada. Mejía y Ortiz (2019) documentaron el impacto del panfleto como forma de propaganda negativa en la Universidad de Antioquia, evidenciando el deterioro de la comunicación política institucional; y Andrews Kwakwa Peprah y Albert Agbesi Wornyo (2023) analizaron, en el contexto ghanés, el uso de la propaganda como mecanismo de persuasión por parte de líderes estudiantiles, concluyendo que sus efectos sobre la decisión de voto son más complejos y limitados de lo que las fuerzas contendientes suponen.

El presente artículo se propone, en consecuencia, caracterizar y cuantificar los efectos de la propaganda electoral sobre los estudiantes de la UAGRM durante las elecciones de 2025, respondiendo a las siguientes preguntas de investigación: ¿cómo perciben los estudiantes la propaganda electoral en términos de sus efectos sobre el entorno universitario y sobre su propio bienestar? ¿En qué medida esta propaganda influye efectivamente en su decisión de voto? ¿Qué formas de propaganda generan mayor rechazo y en qué espacios del campus resultan más perturbadoras?

La relevancia del estudio es doble. En términos institucionales, sus hallazgos pueden orientar reformas al reglamento electoral de la UAGRM que equilibren el derecho de las fuerzas políticas a comunicarse con su electorado y la obligación de la universidad de preservar un ambiente propicio para el aprendizaje. En términos académicos, contribuye a llenar un vacío en la literatura sobre democracia universitaria y bienestar estudiantil en el contexto latinoamericano, ofreciendo evidencia empírica sobre una práctica que, pese a su omnipresencia en los campus de la región, rara vez ha sido objeto de análisis sistemático. Lo que sigue se organiza en cuatro secciones: metodología, resultados, discusión y conclusiones.

## 2. Metodología

El presente artículo tuvo un enfoque mixto (cualitativo-cuantitativo).

Las técnicas de investigación usadas fueron la encuesta, el análisis de datos y la recopilación documental de investigaciones previas.

Se realizó una encuesta de 11 preguntas a 383 estudiantes de la UAGRM de la zona de módulos y campus universitario, sobre una base de población aproximada de 100.000 estudiantes para obtener un margen de confianza del 95 % y un margen de error del 5 %.

La encuesta fue de carácter estratificado, con base en las facultades que se encuentran en los predios universitarios previamente mencionados.

## 3. Resultados

Los estudiantes de la UAGRM sufrieron durante las elecciones para autoridades principales (rector, vicerrector, decanos, directores de carrera) una masificación de propaganda política, entendida como técnica para la difusión de ideas y para persuadir al ciudadano de optar por uno u otro candidato (Vera Leyva, 2022), por parte de todos los frentes que disputaban las elecciones.

En este proceso, desde el mes de mayo hasta el mismo día de las elecciones, pese a estar prohibido con base en los artículos 116 y 122, párrafo D, del Reglamento Electoral (UAGRM, 2024), se produjeron una serie de “bombardeos” en forma de medios tradicionales de propaganda electoral.

En la encuesta y tabla siguiente se desglosan los estudiantes encuestados a pocos días de la elección.

Tabla 1: Distribución de la muestra por facultad – UAGRM

Facultad	Frecuencia (n)	Porcentaje (%)
Ciencias Jurídicas, Políticas, Sociales y RRII	55	14.4 %
Ciencias de la Salud Humana	24	6.3 %
Ciencias Exactas y Tecnología	55	14.4 %
Ciencias Farmacéuticas	55	14.4 %
Humanidades	55	14.4 %
Ingeniería y Ciencias de la Computación	55	14.4 %
Ciencias Veterinarias	29	7.6 %
Ciencias Económicas y Empresariales	55	14.4 %
<b>Total</b>	<b>383</b>	<b>100 %</b>

*Nota: Muestreo estratificado proporcional. N total = 383. Encuesta aplicada el 6 de junio de 2025.*

Dentro de los encuestados, la siguiente tabla desglosa la experiencia, es decir, la cantidad de elecciones, sea para cargos como los de 2025 o para cargos estudiantiles de cogobierno.

Tabla 2: Participación previa en elecciones universitarias

Opciones de respuesta	Frecuencia (n)	Porcentaje (%)
Esta es mi primera elección	94	24.5 %
He participado en 1 elección	261	68.1 %
He participado en 2 o más elecciones	28	7.3 %
<b>Total</b>	<b>383</b>	<b>100 %</b>

*Nota: n = 383.*

Se puede observar que, dentro de los encuestados, la gran mayoría ya ha experimentado una elección universitaria como mínimo.

Dentro de los encuestados, la siguiente tabla desglosa su participación activa en la

colocación de propaganda electoral.

Tabla 3: Participación en la colocación de propaganda electoral

Opciones de respuesta	Frecuencia (n)	Porcentaje (%)
Sí ha participado	102	26.6 %
No ha participado	281	73.4 %
<b>Total</b>	<b>383</b>	<b>100 %</b>

*Nota: n = 383.*

Esto marca un indicio de que una minoría estudiantil, concentrada o interesada por uno u otro motivo, participa en la colocación de propaganda política en la universidad por alguno de los frentes, mientras que una mayoría masiva omite participar en ello.

Sin embargo, aunque aproximadamente el 25 % de los encuestados participó en ello, la tabla 4 indica que solo el 16,7 % de los encuestados aprueba esta práctica de alguna forma.

Tabla 4: Percepción de la propaganda electoral universitaria

Opciones de respuesta	Frecuencia (n)	Porcentaje (%)
Perjudiciosa e innecesaria	146	38.1 %
Perjudiciosa pero necesaria	173	45.2 %
Buena y necesaria	19	5.0 %
Buena pero excesiva	45	11.7 %
<b>Total</b>	<b>383</b>	<b>100 %</b>

*Nota: n = 383.*

Pese a ello, casi la mitad, aunque la vea como un “mal”, la consideran “necesaria”, es decir, que debe existir de algún modo dentro de la universidad la posibilidad de los partidos de comunicarse con su electorado, en este caso, el estudiante.

Dentro de la siguiente tabla se desglosa cómo, a pesar de que algunos la perciben como buena de alguna manera (16,7 %), sí admiten algún perjuicio para el estudiante, siendo los más destacados el perjuicio audiovisual al estar en la universidad y el perjuicio al tránsito dentro de ella.

Tabla 5: Aspectos en que la propaganda electoral genera mayor perjuicio

Opciones de respuesta	Frecuencia (n)	Porcentaje (%)
Ninguno	9	2.3 %
A la hora de estudiar	86	22.5 %
Visualmente al estar en la universidad	138	36.0 %
A la hora de transitar por la universidad	150	39.2 %
<b>Total</b>	<b>383</b>	<b>100 %</b>

*Nota: n = 383.*

Esto remite al aporte de Mera-Benavides (2017): la contaminación visual ya no es solo un factor estético, sino una problemática que afecta al ambiente, al bienestar psicológico y a la calidad de residencia de los universitarios; así como al trabajo de Ali, Farhan y Jawad (2023), que identifica consecuencias fisiológicas, psicológicas y académicas.

Ante tales perjuicios, es pertinente la pregunta sobre la posición frente a una posible prohibición de la propaganda electoral.

Tabla 6: Posición respecto a prohibir la propaganda electoral

Opciones de respuesta	Frecuencia (n)	Porcentaje (%)
Sí prohibiría	114	29.8 %
No prohibiría	158	41.2 %
Solo algunos tipos	111	29.0 %
<b>Total</b>	<b>383</b>	<b>100 %</b>

*Nota: n = 383.*

Reafirmando los puntos anteriores, una mayoría relativa defiende la existencia de la propaganda política, sea de manera casi total o parcial, pero de manera específica se señalan algunos medios de propaganda que generan mayor rechazo, como se observa en la tabla 7.

Tabla 7: Tipos de propaganda electoral que se prohibirían

<b>Tipo de propaganda</b>	<b>Frecuencia (n)</b>	<b>% del total de resp.</b>
Música de alto nivel	202	26.1 %
Banderas	134	14.7 %
Ules / plásticos	117	12.8 %
Pinturas de logos	91	10.0 %
Stickers	81	8.9 %
Carteles	60	6.6 %
Posters	59	6.5 %
Papeletas	46	5.0 %
Papeles adheridos a mano	41	4.5 %
Pancartas	29	3.2 %
Banderines	17	1.9 %
<b>Total de respuestas</b>	<b>877</b>	<b>100 %</b>

*Nota: Respondieron quienes contestaron "Sí" o "Solo algunos tipos" en la pregunta 6. Porcentajes sobre el total de respuestas emitidas (selección múltiple).*

Destaca por sobre las demás la música a alto volumen, muchas veces escuchada durante cátedras, clases y seminarios, donde los altavoces usados (de capacidad promedio de entre 80 y 110 dB) sobrepasan los niveles recomendados (35 dB) con base en la recomendación de la OMS para procesos de aprendizaje (OMS, 2011).

Pero, en contraste, la tabla 8 indica lo siguiente:

Tabla 8: Razones para prohibir la propaganda electoral

Razón	Frecuencia (n)	Porcentaje (%)
Muchos recursos desperdiciados	185	48.3 %
Innecesaria	97	25.3 %
Distrae de los estudios	59	15.4 %
Genera mucha contaminación	42	11.0 %
<b>Total</b>	<b>383</b>	<b>100 %</b>

*Nota: Selección múltiple; porcentajes sobre el total de respuestas emitidas.*

La opción mayoritaria no es la distracción, sino el desperdicio de recursos; por lo tanto, resulta más incómodo el gasto de recursos, ya sea económico, material o de tiempo, sobre todo en las opciones que siguen a la música a alto volumen. Más que la música como tal, los 97 votos que la consideran innecesaria muestran una tendencia de rechazo o apatía hacia la propaganda, al combinar su carácter innecesario con el gasto de las fuerzas políticas en ella.

En defensa de la misma propaganda, la tabla 9 reconoce el trabajo de las fuerzas contendientes; sin embargo, no en la medida que se esperaría de una campaña masiva, con solo un 29,8 % de los estudiantes que recibieron alguna influencia de la propaganda reflejada en el voto que depositaron el mismo día de la elección.

Tabla 9: Influencia de la propaganda electoral en la decisión de voto

Opciones de respuesta	Frecuencia (n)	Porcentaje (%)
Sí, en gran medida	26	6.8 %
Sí, en una pequeña parte	88	23.0 %
No, de ninguna manera	269	70.2 %
<b>Total</b>	<b>383</b>	<b>100 %</b>

*Nota: n = 383.*

El amplio margen coincide, en gran medida, con la combinación antes indicada de ser innecesaria y representar un despilfarro de todo tipo de recursos.

Tabla 10: Posición de voto ante limitación/regulación de la propaganda electoral

Opciones de respuesta	Frecuencia (n)	Porcentaje (%)
Sí regularía	97	25.3 %
No regularía	115	30.0 %
Voto en blanco	132	34.5 %
Voto nulo	39	10.2 %
<b>Total</b>	<b>383</b>	<b>100 %</b>

*Nota: n = 383.*

En otra tentativa de votación estudiantil sobre la propaganda, muchos se abstienen en el tema, por lo que ninguna de las dos opciones principales permite observar una tendencia clara, lo que deja un vacío en los deseos de la comunidad estudiantil sobre el tema.

Finalmente, los estudiantes tienen una variedad extensa de opiniones sobre los lugares donde este tipo de propaganda debería tener prohibido su uso, aunque destacan las aulas y las paredes, seguidas de cerca por las aceras, espacios verdes y lugares de descanso como bancas y mesas, mientras que quedan rezagados el suelo y el alumbrado.

Tabla 11: Espacios donde se prohibiría la propaganda electoral

Espacio	Frecuencia (n)	% del total de resp.
Aulas	78	17.8 %
Paredes internas y externas de edificios	69	15.8 %
Aceras principales de la universidad	57	13.0 %
Espacios verdes y árboles	55	12.6 %
Bancas y mesas	53	12.1 %
Mástiles para banderas	41	9.4 %
Ventanas	38	8.7 %
Alumbrado	29	6.6 %
Suelo	17	3.9 %
<b>Total de respuestas</b>	<b>437</b>	<b>100 %</b>

*Nota: Selección múltiple; porcentajes sobre el total de respuestas emitidas.*

Esto refleja una tendencia que une las tres principales respuestas: los estudiantes repudian más lo que ven en el día a día, dejando más de lado lo que está fuera de su rango visual.

Gao et al. (2024), en su revisión general sobre la contaminación visual, comentan cómo la urbanización y la expansión de las ciudades o lugares en desarrollo incrementan de manera exponencial la contaminación visual. En ese proceso, el individuo comienza a experimentar un estado constante de abrumación, llegando a difuminar su entorno con tal de mantenerse en el día a día. Esto puede verse reflejado en el caso estudiado, al dejar de lado el ambiente verde de las zonas universitarias, las zonas de clases y especialización de estudio, difuminadas en sus funciones, así como la desconexión en el tránsito universitario.

#### 4. Conclusiones

La propaganda electoral universitaria durante las elecciones de 2025 refleja la constante irritación estudiantil frente a la propagación masiva, desorganizada y desregulada, con excepción de lagunas básicas y poco claras estipuladas en el artículo 120 del Reglamento Electoral (UAGRM, 2024), donde solo se indica el cuidado de los bienes,

el ornato y la infraestructura general.

Estos factores han causado una apatía relativamente mayoritaria de parte de la comunidad hacia este intento de comunicación política, lo que en futuros estudios puede profundizarse en relación con su alcance en el proceso y elección de autoridades.

El presente estudio establece una base para futuras investigaciones en relación con el bienestar estudiantil, la regulación y el cuidado de los bienes, la infraestructura, el ornato y el ambiente universitario, así como con el ejercicio responsable de la democracia y la construcción de un pensamiento crítico profundo sobre el medio ambiente, la elección de autoridades y el papel de la Corte Electoral Universitaria.

Manteniéndose en un perfil académico, está abierto a modificaciones con aportes tanto de autoridades pertinentes como de personas interesadas en la materia.

## Referencias

- Mera-Benavides, D. A. (2017). Diagnóstico ambiental de la percepción de la contaminación visual por parte de la población universitaria de la Facultad de Ingeniería Civil y de la Facultad de Ciencias Naturales, Exactas y de la Educación de la Universidad del Cauca. *Revista Luna Azul*, (44), 211-230. <https://doi.org/10.17151/luaz.2017.44.13>.
- Mejía C., J. D., & Ortiz G., M. I. (2019). El panfleto como antítesis de la comunicación política y su impacto en la Universidad de Antioquia. *Diálogos De Derecho Y Política*, (23), 119-125. Recuperado a partir de <https://revistas.udea.edu.co/index.php/derypol/article/view/339695>
- Andrews Kwakwa Peprah, Albert Agbesi Wornyo (2023) propaganda as a means of persuasion by student leaders: a study of graduate students' elections in two universities in ghana
- Comprehensive Review of Noise Pollution Sources, Health Impacts, and Acoustic Environments Affecting College and University Students (H. H. M. Ali, A. H. Farhan, & A. S. Jawad , Trans.). (2023). *Mesopotamian Journal of Civil Engineering*, 2023, 86-97. <https://doi.org/10.58496/MJCE/2023/011>
- Reglamento Electoral Universidad Autónoma "Gabriel René Moreno" 2024 UAGRM (2024) <https://es.scribd.com/document/853261885/Reglamento-Corte-Electoral-Uagrm>
- Estatuto autonómico Universitario UAGRM (2024)
- Vera Leyva, R. (2022). Evolución de la propaganda política en los procesos electorales del siglo XXI. *Ciencia Latina Revista Científica Multidisciplinar*, 6(4), 5809-5825. [https://doi.org/10.37811/cl\\_rcm.v6i4.3050](https://doi.org/10.37811/cl_rcm.v6i4.3050)
- OMS (2011) <https://www.who.int/europe/news-room/fact-sheets/item/noise>
- Gao, H., Bakar, S. A., Maulan, S., Yusof, M. J. M., Mundher, R., Guo, Y., & Chen, B. (2024). A Systematic Literature Review and Analysis of Visual Pollution. *Land*, 13(7), 994. <https://doi.org/10.3390/land13070994>

## INSTRUCCIONES PARA AUTORES

*Conocimiento i Política* es una revista física y online, creada el año 2005 por el Centro de Investigación Regional en Ciencia Política y Ciencias Sociales de la Universidad Autónoma Gabriel René Moreno. La revista admite trabajos inéditos en Ciencia Política y Ciencias Sociales, con argumentación clara, métodos empíricos o teóricos bien definidos, datos comprobables y citas precisas. Los manuscritos no deben ser difundidos simultáneamente en otras publicaciones, ni contener duplicaciones, datos inventados, citas no verificables o atribuciones de autoría que no correspondan. La edición no alterará el sentido del texto: las intervenciones editoriales se limitarán a correcciones ortográficas, gramaticales, de puntuación, normalización formal y adecuación al estilo de la revista, salvo autorización expresa de los autores.

La revista publica dos números por año y acepta seis modalidades de manuscritos: artículos de investigación originales, artículos de revisión, artículos de reflexión, notas o comunicaciones de investigación, ensayos bibliográficos y reseñas críticas. Todos los envíos deben mantener un tono académico, conciso y verificable.

### Metadatos editoriales e indexación

Para facilitar la identificación, citación, preservación e indexación de los trabajos, cada envío debe incluir información completa y consistente en el manuscrito y en la plataforma OJS:

- título en español e inglés;
- nombre completo de cada autor, afiliación institucional, ciudad y país;
- correo electrónico de contacto y autor de correspondencia;
- ORCID de los autores, cuando esté disponible;
- tipo de manuscrito: artículo de investigación, artículo de revisión, artículo de reflexión, nota o comunicación, ensayo bibliográfico o reseña;
- resumen y palabras clave en español e inglés, cuando corresponda según la modalidad;
- declaración de financiamiento, conflictos de interés, contribución de autoría y consideraciones éticas;

- disponibilidad de datos, materiales, instrumentos o repositorios, cuando el manuscrito los utilice;
- referencias completas, DOI o URL persistente cuando existan.

La información declarada por los autores debe coincidir con la versión final publicada. La revista incorporará, cuando corresponda, fechas de recepción y aceptación, DOI, licencia de uso y forma de citación.

### **Formato general**

Los manuscritos deben presentarse en hoja tamaño carta (21 × 28 cm), interlineado 1.5, letra Times New Roman de 12 puntos, texto justificado y numeración de páginas consecutiva. Las tablas y figuras deberán insertarse cerca del párrafo que las mencione, numerarse con arábigos y llevar título o descripción breve. Se solicita revisar la coherencia de la estructura, la correspondencia entre citas y referencias, y la aplicación uniforme de la norma APA 7.

**Tablas y figuras.** Cada tabla, figura o gráfica debe mencionarse en el texto por orden de aparición e incluir una nota de fuente. Cuando corresponda, podrá utilizarse la fórmula *Fuente: elaboración propia*; si se emplean datos, imágenes o materiales de terceros, deberá indicarse la fuente específica. No se deben utilizar abreviaturas para referirse a tablas o figuras en el cuerpo del manuscrito. Las ilustraciones, gráficos, mapas y fotografías deben enviarse en JPG o PNG, con resolución superior a 300 dpi. Las tablas deben remitirse en formato editable, preferentemente Excel; si exceden una página, podrán incorporarse como anexos con la numeración correspondiente.

**Idioma.** La revista acepta trabajos en español o inglés. Los manuscritos en español deben incluir título, resumen y palabras clave en inglés; los manuscritos en inglés deben incluir título, resumen y palabras clave en español.

### **Artículos científicos y de reflexión**

Los artículos deben presentar título en español e inglés. Se recomienda que el título sea informativo, preciso y breve, sin abreviaturas innecesarias, fórmulas o expresiones ambiguas. A continuación se incluirán los nombres de los autores, afiliaciones, correo electrónico y datos del autor de correspondencia.

Los artículos científicos y de reflexión deben incluir resumen y abstract de 200 a 300 palabras, sin citas bibliográficas, además de un máximo de 10 palabras clave en español e inglés. Para asegurar consistencia terminológica, se recomienda seleccionar descriptores de tesauros o vocabularios normalizados del área.

La estructura sugerida para artículos de investigación es: introducción, metodología, resultados, discusión, conclusiones, agradecimientos cuando correspondan y literatura citada. Los artículos de revisión deben presentar una estrategia de búsqueda y selección documental claramente descrita, una síntesis crítica de la literatura y conclusiones sustentadas en las fuentes examinadas. Los artículos de reflexión deben sostener una argumentación teórica o crítica claramente organizada, con problema definido, desarrollo analítico, discusión y conclusiones.

### **Notas, ensayos y reseñaciones**

Las notas o comunicaciones presentan avances de investigación, resultados preliminares, hallazgos parciales o novedades metodológicas. Su extensión no debe superar las 2000 palabras y su organización puede incluir título, autores, afiliaciones, palabras clave, cuerpo de la nota, agradecimientos y literatura citada.

Los ensayos bibliográficos constituyen análisis críticos y comparativos de dos o más obras recientes que comparten un eje temático o una problemática teórica. Deben superar la descripción aislada de cada obra y producir un diálogo analítico entre los textos revisados.

Las reseñaciones se orientan a la evaluación crítica de una publicación reciente en Ciencia Política o Ciencias Sociales. Deben exponer objetivos, contenido central, metodología, aportes, limitaciones y pertinencia de la obra reseñada. Se recomienda no superar las 2000 palabras.

### **Citación y referencias**

Las citas y referencias deben ajustarse de manera uniforme a APA 7. Toda referencia incluida en la bibliografía debe estar citada en el texto y toda cita del texto debe aparecer en la lista final. Los nombres de revistas deberán escribirse completos, sin abreviaturas. Siempre que sea posible, se debe incluir DOI o URL estable.

Ejemplos básicos:

- **Artículo científico:** Smith, J. R. (2010). Electoral systems and political stability in Latin America. *Journal of Political Science*, 45(3), 345–367.
- **Libro:** Dahl, R. A., & Lindblom, C. E. (1995). *Politics, economics, and welfare*. Yale University Press.
- **Capítulo de libro:** Norris, P., & Stevenson, M. (2012). Comparative political regimes. En J. Smith & A. Brown (Eds.), *Political systems in a globalized world* (pp. 85–123). Oxford University Press.
- **Tesis:** Melgar, N. (2023). *Acción exterior subnacional en el municipio de Puerto Quijarro* [Tesis de grado, Universidad Autónoma Gabriel René Moreno].
- **Página web:** UNDP. (2023). *Human Development Report*. <http://www.hdr.undp.org>

## Envío y revisión

Los manuscritos deberán enviarse a través de la plataforma OJS, disponible en <https://ojs.uagr.edu.bo/revista-conocimiento-politica>, mediante la opción “Envíos”. El manuscrito debe presentarse anonimizado para evaluación, sin datos que permitan identificar a los autores en el archivo o en sus metadatos. Las referencias a obras propias podrán incorporarse en la versión final, si el trabajo es aceptado.

El envío implica el conocimiento y aceptación de estas instrucciones. Mientras el manuscrito esté en evaluación por *Conocimiento i Política*, los autores no deben someterlo simultáneamente a otra revista. La persona responsable del envío será destinataria de las comunicaciones editoriales y deberá informar a los coautores sobre el proceso.

Los manuscritos recibidos son evaluados inicialmente por la Editora en Jefe y el Consejo de Redacción, que verifican pertinencia temática, originalidad y cumplimiento formal. Superada esta fase, el manuscrito se deriva a especialistas externos para valorar rigor académico, claridad argumentativa, calidad metodológica y aportes al campo.

## Ética, originalidad y transparencia

La revista mantiene una política de tolerancia cero frente al plagio, la fabricación o falsificación de datos, la omisión deliberada de información, la publicación duplicada y los conflictos de autoría. Toda fuente de financiamiento, apoyo institucional, conflicto

de interés, uso de datos sensibles o requerimiento de consentimiento informado debe declararse de forma explícita. Si se utilizaron herramientas de inteligencia artificial para apoyo lingüístico, búsqueda, procesamiento de datos u otras tareas, su uso deberá declararse cuando corresponda, sin atribuirles autoría.

Cualquier incumplimiento de estas normas autoriza a la revista a tomar medidas editoriales, incluidas la suspensión del proceso, la retractación del texto, la notificación a la institución de afiliación del autor responsable y la inhabilitación para futuras colaboraciones.

### **Derechos de autor y licencia**

Los autores conservan plenamente sus derechos de autor al publicar en esta revista, pero conceden a *Conocimiento i Política* el derecho de primera publicación de la obra. El contenido se publica bajo licencia Creative Commons Reconocimiento-No Comercial-Sin Obra Derivada 4.0 Internacional, lo que permite compartir la obra siempre que se reconozca la autoría y se cite la revista como fuente original, sin usos comerciales ni elaboración de obras derivadas.

Los autores pueden establecer acuerdos de licencia no exclusiva para distribuir la versión final publicada, por ejemplo en repositorios institucionales o compilaciones académicas, siempre que indiquen que el texto fue publicado inicialmente en *Conocimiento i Política*.

