
Special issue

Ana Bizberg

<https://orcid.org/0000-0001-8267-4700>

abizberge@pucp.edu.pe

Pontificia Universidad Católica del Perú

Nadia Sabrina Koziner

<https://orcid.org/0000-0003-2834-4799>

nadiakoziner@gmail.com

Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) y Centro de Investigación "Industrias Culturales, Políticas de comunicación y Espacio Público" (ICEP) de la Universidad Nacional de Quilmes (UNQ)

Recibido

14 de noviembre de 2025

Aprobado

11 de mayo de 2026

© 2026

Communication & Society

ISSN 0214-0039

E ISSN 2386-7876

www.communication-society.com

2026 – Vol. 39 (1)

pp. 464-481

Cómo citar este artículo:

Bizberge, A., & Koziner, N. S. (2026). Gobernanza de Meta: Las políticas de moderación de contenidos y sus impactos en América Latina (2018–2025), *Communication & Society*, 39(1), 464-481. <https://doi.org/10.15581/003.39.1.031>

Gobernanza de Meta. Las políticas de moderación de contenidos y sus impactos en América Latina (2018–2025)

Resumen

El trabajo analiza la estructura de gobernanza de Meta, sus políticas de moderación de contenidos entre 2018 y 2025 y sus implicancias para el debate público en América Latina desde la perspectiva de la Economía Política de la Comunicación y los Derechos Humanos. Para ello, combina análisis documental, entrevistas y revisión de declaraciones de organizaciones y expertos regionales en derechos digitales. Concluye que el sofisticado diseño institucional de moderación presenta déficits estructurales para América Latina. Entre ellos, una escasa incidencia regional en el total de casos estudiados y, particularmente en los de desinformación electoral. Además, advierte el desfasaje entre el rol del Oversight Board y el desarrollo de políticas empresariales. El cambio implementado en 2025 supone mayor automatización y externalización de la responsabilidad de moderación hacia los usuarios, a partir de una concepción maximalista de la libertad de expresión. En América Latina, esto representa un riesgo para los procesos electorales. La estrategia de Meta evidencia la insuficiencia sistémica de la autorregulación empresarial como mecanismo de protección de derechos y la necesidad de marcos regulatorios con enfoque de derechos humanos y mayor transparencia sobre la arquitectura de la plataforma.

Palabras clave

Meta, Oversight Board, notas comunitarias, moderación de contenidos, libertad de expresión, América Latina.

1. Introducción

En noviembre de 2018, Meta anunció la creación del Oversight Board (OB), un organismo independiente destinado a revisar decisiones de moderación con el objetivo de reforzar la legitimidad corporativa y alinearse con los estándares internacionales de derechos humanos. Sin embargo, su impacto resultó limitado y funcionó más como mecanismo de validación frente a crecientes presiones regulatorias que como instancia de control.

Por fuera del OB, en 2025 Meta anunció el fin de los acuerdos con verificadores de datos en Estados Unidos y la implementación del sistema de notas de la comunidad, desplazando la moderación profesional hacia formas de verificación centradas en los usuarios y en una mayor moderación algorítmica. El cambio abrió debates en Latinoamérica, donde las plataformas

cumplen un rol central en la configuración del espacio público digital y persisten tensiones entre la autorregulación empresarial y la protección de derechos digitales.

El objetivo del trabajo es analizar la estructura de gobernanza de Meta para la moderación de contenidos entre 2018 y 2025 y evaluar sus implicancias para el debate político en América Latina. Para concretarlo, se organiza en seis apartados. Luego de la introducción, presenta el marco teórico y la estrategia metodológica. Después analiza la estructura de gobernanza de Meta. Seguidamente, aborda el retroceso implementado en 2025 en la política de moderación de la compañía y sus implicancias regionales. Por último, ofrece conclusiones sobre las limitaciones del modelo de autorregulación empresarial.

2. Marco teórico

La formulación e implementación de políticas de comunicación conforman un proceso dinámico en el que circulan supuestos e ideas en competencia sobre la estructura, la conducta y el rendimiento deseables del sistema. Involucra regulación estatal, co-regulación —la industria aplica lineamientos trazados por el regulador, que supervisa su cumplimiento— y autorregulación —el sector define sus propias reglas— (Puppis & Van Den Bulk, 2019). En Internet, la “gobernanza” desafía el modelo estatal de “comando y control” al reconocer la incidencia de múltiples actores —decisores políticos, plataformas, usuarios, anunciantes, periodistas, academia y sociedad civil— en la definición de los límites de la libertad de expresión, la responsabilidad corporativa y el rol estatal (Gorwa, 2019; Iosifidis, 2011).

En la economía digital, las plataformas influyen mediante intervenciones económicas, infraestructurales y regulatorias (Van Dijk et al., 2018). Su poder se asienta en cuatro estrategias interrelacionadas: infraestructuralización —mercados de datos y reglas privadas—, plataformaización —penetración de infraestructuras y marcos de gobierno que reordenan prácticas e imaginarios—, conglomerización —adquisiciones— y financierización —liderazgo bursátil— (Nieborg & Poell, 2025; Poell et al., 2019).

Distintos autores distinguen la gobernanza *por* y *de* plataformas (Suzor, 2018; Gorwa, 2019). La gobernanza *por* las plataformas, aspecto abordado aquí, remite a la autorregulación: los modos en que las empresas tecnológicas gobiernan sus espacios digitales como sistemas de poder privado sobre el derecho público a la expresión (Marchal et al., 2024). Sus políticas, expresadas en normas comunitarias y términos de servicio, definen contenidos y comportamientos aceptables o prohibidos.

Los términos de servicio funcionan como documentos constitucionales que explican cómo se constituyen y gobiernan los espacios compartidos y presuponen una relación desigual entre las partes. No garantizan protección contra decisiones arbitrarias o cambios unilaterales de normas ni ofrecen vías significativas de reclamo y apelación (Suzor, 2018). Así, las plataformas se vuelven árbitros de asuntos sensibles de la esfera pública e instauran un poder discrecional, con déficit de debido proceso, reglas poco inteligibles y responsabilidad difusa (Pita, 2022).

En esta lógica subyace una mirada maximalista de la libertad de expresión, apoyada en la tradición estadounidense de la Primera Enmienda —que prohíbe la injerencia estatal en las expresiones— y en la Sección 230 de la *Communications Decency Act*, que plantea la inmunidad de las plataformas por contenido de terceros —no son consideradas editoras— y por moderación: la cláusula del “buen samaritano” les permite, de “buena fe”, bloquear, filtrar, eliminar o restringir contenidos objetables, con excepciones como propiedad intelectual, trata de personas, explotación sexual o privacidad. A su vez, la *Digital Millennium Copyright Act* las exime de responsabilidad por infracciones a derechos de autor cometidas por usuarios —protección de “puerto seguro”— si actúan diligentemente para retirar el contenido protegido mediante

“notificación y retirada”. Este enfoque, que permitió a las empresas de Silicon Valley evitar responsabilidades en sentido amplio, muestra límites: la promesa de que mayor expresión conduce a la verdad se vuelve falaz cuando la circulación de desinformación y expresiones discriminatorias desalienta otras voces, produce silenciamientos y amenaza derechos (Becerra & Waisbord, 2021a). Frente a la presión pública, las plataformas adaptan sus políticas, aunque hay quienes señalan que esos cambios responden más a cuestionamientos sobre su legitimidad y autoridad que a la afectación de intereses económicos (Marchal et al., 2024).

En Latinoamérica, el Sistema Interamericano de Derechos Humanos enfatiza la dimensión individual y social de la libertad de expresión. El Estado adquiere centralidad como garante del acceso a la información, la diversidad y el pluralismo mediático (Loreti & Lozano, 2014). No obstante, esas obligaciones encuentran límites en la debilidad institucional, la concentración mediática, la captura de sistemas de medios por elites políticas y económicas, la fragilidad de los medios públicos y los ataques a la prensa (Waisbord, 2013; Linares & Segura, 2022; Mastrini & Becerra, 2006). Esto abrió debates sobre las obligaciones positivas del Estado en Internet (Bizberge & Mastrini, 2024; Chenou, 2021; Becerra & Waisbord, 2021b). Ante la incapacidad de las empresas para autorregularse y proteger el interés público y los derechos humanos, crecen iniciativas de gobiernos, sociedad civil y organismos intergubernamentales para regular plataformas (Gorwa, 2019; Pita, 2022).

Tras Cambridge Analytica, las Relatorías de libertad de expresión de ONU, OEA y UNESCO advirtieron la necesidad de alinear las políticas de plataformas con estándares de derechos humanos y demandar no solo autorregulación sino intervención de los poderes públicos (Bertoni, 2021; Loreti, 2021).

Luego de un proceso de diálogo multiactor, UNESCO (2023) y la Organización de Estados Americanos (2024) publicaron directrices para la gobernanza de plataformas y la moderación privada de contenidos, con recomendaciones para Estados, plataformas y otros actores. Ambos documentos destacan la adecuación de políticas y prácticas de moderación a estándares de derechos humanos y asignan un lugar especial a los procesos electorales. También piden evaluaciones y mitigación de riesgos sobre la libre circulación de información; transparencia y responsabilidad; identificación de contenido publicitario; transparencia sobre financiamiento y monetización; revisión de algoritmos de recomendación; diálogo entre partes interesadas y alfabetización digital como herramienta de empoderamiento de usuarios.

En América Latina los debates regulatorios son incipientes. La exigencia de transparencia y rendición de cuentas —excepto en Brasil— está prácticamente ausente en las iniciativas (Pita, 2022). Muchas presentan problemas para definir conductas y tienden a criminalizar expresiones con sanciones penales (Bizberge, 2021; CELE, 2018). Además, las plataformas ejercen un fuerte *lobby* para obturar los debates sobre intervención estatal (Peñarredonda et al., 2025).

3. Estrategia metodológica

La investigación adoptó una estrategia de triangulación metodológica con un diseño flexible (Denzin & Lincoln, 2005; Flick, 2014).

Primero, se realizó un análisis documental de fuentes primarias y secundarias: documentos oficiales de Meta; materiales de organizaciones de la sociedad civil latinoamericanas; trabajos académicos; informes y notas periodísticas de especialistas en políticas digitales y moderación de contenidos; y reportes de organismos internacionales. Los textos fueron considerados por su contenido, origen, propósito, función comunicativa y rol como marcos de argumentación y disputa en torno a la moderación de contenidos (Flick, 2014).

En segundo lugar, durante el primer semestre de 2025 se realizaron entrevistas semiestructuradas en profundidad a informantes clave, para incorporar la perspectiva de actores que participan directamente en el debate sobre gobernanza de plataformas digitales. El objetivo fue complementar el análisis documental con visiones de primera mano sobre los efectos concretos de las políticas de Meta en América Latina que los documentos públicos no registran con igual profundidad. El diseño semiestructurado permitió mantener un núcleo común de preguntas sobre el rol del OB en la región, el cambio de políticas de 2025, las tensiones entre autorregulación y regulación estatal y las perspectivas regulatorias en América Latina, junto con preguntas específicas adaptadas al perfil de cada informante. Se buscó equilibrar comparabilidad y exploración de matices interpretativos (Edwards & Holland, 2013; Brinkmann, 2014).

La selección respondió a un muestreo por propósito con criterios de relevancia temática, trayectoria y diversidad institucional. Se buscaron informantes cuya experiencia, posición institucional o conocimiento especializado iluminara las preguntas de investigación (Patton, 2002). Se incluyeron especialistas de organizaciones de la sociedad civil, organismos internacionales y el ámbito académico, diversos en términos regionales:

Joan Barata — Académico y consultor en regulación de plataformas (España);

Carolina Botero — Especialista en derechos digitales y tecnología; ex directora Ejecutiva de la Fundación Karisma (Colombia);

Carolina Fernández — Asociada en Democracia Digital, PNUD Perú;

Gustavo Gómez — Director Ejecutivo, OBSERVACOM (Uruguay/regional);

Javier Pallero — Experto en políticas públicas e incidencia digital (Argentina);

Olivia Sohr — Directora de Impacto y Nuevas Iniciativas, Chequeado (Argentina).

El análisis de las entrevistas siguió un procedimiento de codificación temática (Braun & Clarke, 2006) e identificó cinco patrones transversales: (1) valoración del OB como organismo que amplió su mandato incorporando estándares internacionales de derechos humanos, pero sin mecanismos para obligar a Meta a incorporar esos criterios en sus políticas generales; (2) brecha entre la formalidad contractual con organizaciones de verificación en Latinoamérica y el retroceso efectivo desde 2025; (3) cambio de 2025 como estrategia de reposicionamiento político-regulatorio ante el nuevo clima político en Estados Unidos; (4) valoración de modelos regulatorios híbridos que privilegian la transparencia, la rendición de cuentas y las garantías procedimentales; y (5) amenaza de represalias comerciales de la administración Trump como factor inhibitorio para el avance regulatorio regional. Las tres primeras se triangulan con el análisis documental; la cuarta y la quinta emergen de las entrevistas. Las citas incorporadas operan como evidencia de estos patrones.

4. Gobernanza de Meta

El sistema de gobernanza interna de Meta involucra el conjunto de políticas de las plataformas — sus términos de servicio, normas comunitarias y políticas de publicidad— y dos entidades: el OB y los Foros de la Comunidad (*Meta Community Forums*).

4.1. Las políticas de Meta

Las políticas definen qué está permitido y qué no en sus plataformas y guían las estrategias y mecanismos de moderación de contenidos (Abramovich et al., 2021). Estos van desde la jerarquización a través de algoritmos de recomendación hasta la eliminación de cuentas (Article 19, 2023).

Interesa aquí focalizar en tres conjuntos de políticas por sus efectos sobre la circulación de discursos políticos, en especial, durante períodos electorales. Estos momentos son laboratorios naturales para evaluar cómo dichas reglas condicionan la visibilidad, el alcance y el tratamiento de contenidos informativos o desinformativos.

4.1.1. Normas comunitarias

Las normas comunitarias rigen en Facebook, Instagram, Messenger y Threads. Articulan el ethos de las plataformas y ordenan expectativas sobre lo deseable, tolerable o sancionable en materia de libertad de expresión, seguridad y dignidad. Se publican y actualizan en el Transparency Center, que también explica el proceso para introducir cambios de política.

Como advierte Gillespie, no operan sólo en tanto reglas operativas sino también como “actuación discursiva” (2018, p. 47). Hacia las personas usuarias, prometen equilibrar libertad de expresión y protección frente a abusos; hacia los anunciantes, presentan un entorno “seguro” y comercialmente atractivo; y hacia autoridades y legisladores, exhiben diligencia y autocontrol.

El esquema reúne 27 reglas sobre violencia e incitación; personas y organizaciones peligrosas; coordinación de daños y promoción de delitos; bienes y servicios restringidos; fraude y engaño; suicidio y autolesión; explotación sexual, abuso y desnudos de menores; explotación sexual de adultos; bullying y acoso; explotación de personas; infracciones a la privacidad; conductas de odio; contenido gráfico y violento; desnudos y actividad sexual de adultos; servicios sexuales; integridad de las cuentas; identidad auténtica; spam; ciberseguridad; comportamiento no auténtico; información errónea; cuentas conmemorativas; propiedad intelectual; solicitudes de usuarios; y medidas adicionales de protección para menores. En paralelo, hay políticas específicas que profundizan estos ejes —por ejemplo, conducta de odio, bienes/servicios restringidos, integridad de cuentas—, donde se detallan definiciones, ejemplos y excepciones regulatorias.

La Tabla N° 1 sistematiza las siete reglas vigentes para Latinoamérica y consideradas más relevantes para la circulación de discursos políticos, con foco en el proceso electoral.

Tabla 1. Normas comunitarias consideradas centrales para la circulación de contenidos políticos

Norma	Descripción	Última actualización
Incitación a la violencia	Remueve llamados o amenazas creíbles de violencia y contenido que facilita daños fuera de línea; prioriza la prevención de daño físico.	28 de septiembre. 2024
Conducta de odio	Prohíbe ataques directos a personas por características protegidas (raza, religión, orientación sexual), incluyendo deshumanización e insultos.	7 de enero. 2025
Contenido gráfico y violento	Restringe material violento o gráfico (muertes, lesiones graves) y aplica pantallas de advertencia en contextos noticiosos.	1 de marzo. 2024
Spam	Prohíbe contenido diseñado para engañar, confundir o inflar artificialmente el alcance (repetición masiva, enlaces engañosos, engagement-bait).	26 de junio. 2024
Información errónea	Distingue entre desinformación que elimina por riesgo de daño inminente y la que reduce/etiqueta mediante verificaciones externas; incluye penalidades por reincidencia y acuerdos con fact-checkers.	8 de abril. 2025
Integridad de las cuentas	Acciones sobre cuentas que infrinjan reiteradamente las reglas o dañan a la comunidad; armoniza criterios globales y opciones de revisión.	28 de marzo. 2024
Representación de la verdadera identidad	Exige representar la identidad auténtica y restringe suplantaciones/no autenticidad, con excepciones para entidades (como empresas).	31 de julio. 2024

Fuente: Elaboración propia en base al Transparency Center de Meta. Fecha de consulta: 24 de octubre de 2025.

El alcance de las Normas comunitarias depende de cómo se traducen en la práctica —eliminar, reducir, etiquetar, despriorizar— a través de interpretaciones cambiantes y sistemas automatizados. Esa plasticidad, articulada con el resto de las políticas de la empresa y con sus arquitectura institucional, condiciona la visibilidad y el tratamiento de los contenidos. En contextos electorales de América Latina, lo decisivo no es sólo qué dicen las reglas sino cómo, cuándo y para quién se aplican.

4.1.2. Términos de servicio

Los Términos de servicio son un contrato que el usuario está obligado a aceptar para interactuar con las tecnologías de Meta. Además de las condiciones generales, Facebook, Instagram y Threads tienen pautas específicas. Las de Facebook abarcan Marketplace y Messenger y la presentan como una red para la comunidad y el debate. Enfatizan la identidad verificada, normas internas de moderación y el uso de datos con fines publicitarios. El usuario cede una licencia amplia sobre sus contenidos.

Instagram tiene condiciones de servicio propias. Se centra en la expresión visual, permite identidades más flexibles y declara el uso de algoritmos para personalizar la experiencia. La cesión de derechos es similar a la de Facebook, pero enfocada al marketing. Threads incorpora condiciones específicas: la más importante es que forma parte de Fediverse, un ecosistema de redes descentralizadas e interoperables. Si bien Meta no controla los servidores externos, puede compartir contenidos con ellos. Ello amplía su alcance, pero dificulta la trazabilidad de los datos y diluye responsabilidades.

Las organizaciones sociales latinoamericanas se han expresado públicamente en relación con la necesidad de transparencia de los términos de servicio, en sintonía con las propuestas de UNESCO (2023) y Organización de Estados Americanos (2024). El principal reclamo versa sobre la necesidad de documentos claros, accesibles y alineados con estándares internacionales de derechos humanos, especialmente, con la libertad de expresión y el acceso a la información (Pita, 2022). Los contenidos y conductas prohibidas; la explicitación de criterios y mecanismos de curación y moderación; la transparencia algorítmica respecto del ordenamiento, priorización, reducción de alcance o direccionamiento; la delimitación entre automatización y revisión humana; y la claridad sobre qué datos personales se recopilan, almacenan y usan para el direccionamiento de contenidos resultan clave para la transparencia y rendición de cuentas.

4.1.3. Políticas publicitarias

Las Normas de publicidad de Meta definen contenidos y prácticas permitidas y prohibidas en sus tecnologías, remiten al cumplimiento de leyes locales y Normas comunitarias cuando corresponde y habilitan remoción o rechazo de anuncios, desactivación de campañas y otras medidas ante infracciones. Su implementación combina revisión automatizada y humana antes de publicar y durante la circulación.

Meta clasifica el contenido publicitario como inaceptable —por ilegalidad o por contrariar sus normas— o cuestionable. El primero incluye explotación sexual de menores; promoción de actividades delictivas o de personas/grupos peligrosos; prácticas discriminatorias; conductas de odio y explotación; disuasión del uso de vacunas; desinformación y material destinado a “engañar o confundir” —fraude, estafa, prácticas comerciales engañosas y tácticas para eludir la revisión— (Meta, 2025d). También restringe contenidos que infrinjan propiedad intelectual o afecten la salud y la seguridad. Estas categorías se complementan con reglas por producto/servicio y obligaciones legales jurisdiccionales.

El contenido cuestionable comprende publicidad que, sin ser necesariamente ilegal, pueda generar experiencias negativas o contradecir las Normas comunitarias: desnudos y actividad

sexual de adultos; explotación sexual de adultos; oferta/solicitud de servicios sexuales y lenguaje sexual explícito; *bullying* y acoso —con protección reforzada para menores—; groserías; infracciones de privacidad y alusiones a atributos personales —raza, etnia, religión, edad, orientación o prácticas sexuales, identidad de género, discapacidad, enfermedades físicas o mentales, vulnerabilidad financiera, condición de votante, afiliación sindical, antecedentes penales, nombres propios—; contenido violento, gráfico, impactante o sensacionalista; referencias que promuevan o normalicen suicidio, autolesiones o trastornos alimentarios, incluidas piezas ficticias o en vivo, y burlas a víctimas o sobrevivientes.

La publicidad sobre asuntos sociales, elecciones o política (SIEP, por su sigla en inglés) tiene consideraciones especiales. Incluye anuncios vinculados con partidos, figuras políticas, candidatos o comités de acción política; elecciones, referendos o votaciones; y temas sociales: “asuntos sensibles muy debatidos y que pueden influir en los resultados de un proceso electoral, estar relacionados con una legislación vigente o tener como resultado un proyecto de ley” (Meta, 2025e). Meta afirma promover responsabilidad y autenticidad, exige mayor transparencia a candidatos, representantes electos y designados y requiere autorización, excepto para editores de noticias identificados por la empresa¹. También se reserva el derecho de restringir contenidos y contempla particularidades nacionales o regionales. El caso más saliente es la Unión Europea: desde la entrada en vigor del Reglamento (UE) 2024/900 sobre Transparencia y Segmentación de la Publicidad Política (TTPA) en octubre de 2025, Meta anunció que dejará de aceptar anuncios políticos, electorales o sobre asuntos sociales por complejidades operativas y riesgos de cumplimiento. La decisión —compatible con la continuidad del discurso político orgánico— reconfigura el ecosistema publicitario electoral europeo (Chan, 2025; Meta, 2025a).

Meta cuenta con una Biblioteca de Anuncios (*Ad Library*), presentada como base de datos completa y con función de búsqueda (Meta, 2025b). Los avisos, activos o no, se almacenan siete años y, desde mayo de 2025, comenzó a retirar progresivamente los “vencidos”. Para anuncios SIEP, existe un régimen reforzado: definición y alcance; identificación del anunciante; *disclaimer* obligatorio sobre el responsable del pago y registro con metadatos ampliados. Desde 2024–2025, exige divulgaciones explícitas cuando se usen medios digitales creados o alterados con Inteligencia Artificial para reducir riesgos de manipulación en campaña. Las modificaciones se documentan en el Centro de Transparencia con sello temporal.

Organizaciones latinoamericanas reportaron dificultades en la trazabilidad de anuncios SIEP. En las últimas elecciones presidenciales de Ecuador (2025), señalaron que la Biblioteca ofreció cierta transparencia, pero no permitió identificar administradores ni montos de gasto con el detalle necesario; además, gastos relevantes de campaña no fueron registrados como gasto electoral y anuncios pagados por terceros escaparon al registro. Buena parte de la desinformación circulante surgió de cuentas que publicitaron en la plataforma (OBSERVACOM, 2025). Durante la primera vuelta presidencial de Bolivia en 2025, periodistas, verificadores y organismos internacionales coincidieron en que la Biblioteca fue la única fuente transparente sobre publicidad política. Esto permitió documentar el financiamiento de desinformación, aunque no derivó en la eliminación de esos anuncios (Martínez Elebi, 2025).

Las políticas publicitarias completan el marco regulatorio de la comunicación paga en Meta. Sus impactos dependen de la implementación concreta y de la calidad de la transparencia ofrecida

¹ Se incluye una lista de tópicos específicos para Australia, Brasil, Canadá, Chile, Unión Europea, India, México, Myanmar, Nueva Zelanda, Filipinas, Singapur, Taiwán, Ucrania y Reino Unido, Estados Unidos y “Otros países y territorios” que se solicita consultar antes de pautar una publicidad.

por la *Ad Library*. Los casos recientes en América Latina muestran su utilidad para documentar financiamientos y la persistencia de zonas opacas de trazabilidad.

4.2. El diseño institucional

El diseño institucional de Meta comprende dos organismos: el OB y los Foros de la Comunidad. Con ellos la compañía busca dar legitimidad a sus decisiones a través de la creación de estructuras consultivas “independientes” donde participan múltiples partes interesadas.

4.2.1. El Oversight Board

El Consejo Asesor nació como producto de una crisis de reputación de la empresa, ante la creciente presión pública y la percepción de que sus decisiones eran inconsistentes y arbitrarias y, por lo tanto, ilegítimas (Douek, 2019, 2024). Diversos episodios —como la injerencia rusa en las elecciones presidenciales de 2016 en Estados Unidos y el caso de Cambridge Analytica (Klonick, 2020)— evidenciaron la necesidad de mecanismos de control que superaran la discrecionalidad corporativa (Gorwa, 2019).

La creación del Consejo fue anunciada por Zuckerberg en noviembre de 2018 con el propósito de construir y proyectar legitimidad para el sistema de moderación de contenidos, de modo que sus decisiones sean percibidas como “justificadas, apropiadas o merecedoras de apoyo” (Douek, 2019, p. 46). Para ello, apuntó a aumentar la transparencia emitiendo explicaciones públicas y auditables, a la vez que trasladó casos controvertidos a una instancia externa. La arquitectura institucional imita a un tribunal independiente que busca producir razonamientos públicos, consistentes y anclados en principios comprensibles y aceptables socialmente. También procura erigirse como contrapeso a la censura reforzando la protección de la libertad de expresión. Finalmente, abre canales de participación con la promesa de morigerar el “déficit democrático” de las plataformas e incorporar voces no corporativas en la definición de sus reglas y el control de su aplicación (Douek, 2024).

Más allá de las intenciones declaradas, el hecho de que el OB hubiera sido creado por una decisión unilateral de la empresa, su mandato definido y su nulo poder coercitivo generaban dudas acerca de su efectividad. Fue calificado como una “locura”, “disfuncional por diseño”, “sin dientes”, una farsa y una fachada (Douek, 2024, p. 378).

Desde la perspectiva de analistas y organizaciones de América Latina, el OB tuvo aciertos al incorporar el marco de derechos humanos para el análisis de casos y limitaciones para incidir en las políticas de la empresa (Observacom, 2019; Derechos Digitales, 2019; Consorcio Al Sur, 2019). Los testimonios recabados valoran al OB como un organismo que, al evaluar casos a partir de estándares internacionales de derechos humanos, amplió su propio mandato y produjo un cuerpo de análisis relevante, pero señalan que carece de mecanismos capaces de obligar a Meta a incorporar esos criterios en sus políticas:

La participación del OB se ha limitado a casos muy específicos en la región [...] Si bien la decisión fue acatada por META, las recomendaciones para prevenir estos casos no han sido de aplicación vinculante, por lo que no han generado un impacto real en la moderación de la conversación digital (Carolina Fernández, comunicación personal, 2 de julio de 2025).

[El OB] se despegó de las normas internas de Facebook y empezó a comparar los casos con los estándares internacionales de derechos humanos. Eso fue un gran avance [...] [No obstante] es un organismo elegido y pagado por la propia Facebook, con muchas limitaciones: no puede estudiar masivamente apelaciones y algunas decisiones no son vinculantes (Gustavo Gómez, comunicación personal, 2 de mayo de 2025).

... su efecto termina siendo más relevante para el mundo académico que para el funcionamiento real de las plataformas. Además, su análisis está fuertemente centrado en la

libertad de expresión, dejando menos desarrollados los abordajes sobre otros derechos como la privacidad (Joan Barata, comunicación personal, 2 de mayo de 2025).

El Consejo fue concebido para tener un máximo de cuarenta integrantes, aunque no superó los veintiséis. Sus miembros exhiben trayectorias de alto perfil —ex jueces, un Premio Nobel de la Paz, académicos, dirigentes políticos y periodistas— y su financiación busca blindar la independencia institucional: Meta aportó inicialmente 130 millones de dólares y luego otros 150 millones a un fideicomiso irrevocable destinado a sostener al organismo (Oversight Board, 2022a).

Su jurisdicción es acotada. No revisa la mayor parte de las operaciones de Meta —cuentas, sistemas publicitarios, grupos y eventos, ni el diseño de algoritmos de amplificación— y se limita a decisiones sobre piezas específicas de contenido. Los casos llegan por apelaciones de usuarios no resueltas dentro de Meta o por derivaciones de la empresa. De más de un millón de presentaciones anuales, el Consejo selecciona una pequeña fracción para una revisión exhaustiva. Admitido un caso, abre un período de comentarios públicos de académicos, ONGs, relatores de ONU y otras organizaciones. El análisis verifica la consistencia de la decisión de Meta con sus políticas, tomando el Derecho Internacional de los Derechos Humanos, aunque se ha criticado la superficialidad de su aplicación (Doueek, 2024).

Los resultados adoptan dos formatos. Las decisiones sobre casos individuales son vinculantes: obligan a Meta a eliminar o restaurar el contenido y a extender el criterio a “contenidos idénticos (...) en contextos paralelos” (Oversight Board, 2022b). En 2024, implementó las decisiones sumarias, resoluciones rápidas y vinculantes para asuntos de interés público o con impactos graves en derechos humanos.

Por otra parte, emite recomendaciones de política más amplias que no son obligatorias, pero la empresa debe responder públicamente qué medidas tomará. Mediante decisiones y recomendaciones, el Consejo impulsó reformas en los procedimientos internos de Meta, aun dentro de estrechos márgenes (Doueek, 2024).

4.2.1.1. El Oversight Board en América Latina

De acuerdo con sus informes anuales, la actuación del OB en Latinoamérica y el Caribe ha representado una proporción constante pero minoritaria en la cantidad de casos entre 2021 y 2025 (Oversight Board, 2022b, 2023, 2024a, 2025a).

Entre octubre de 2020 y diciembre de 2021, el 14,1% de las apelaciones de usuarios presentadas ante el Consejo provinieron de la región, mientras que durante los tres años siguientes bajó a 12%, como ilustra la Tabla N° 2.

Tabla 2. Cantidad de apelaciones recibidas por el OB (2021-2024) y proporción de América Latina y el Caribe

Período	Total de apelaciones recibidas (usuarios)	Proporción de América Latina y el Caribe	Estimación numérica (aprox.)
Oct 2020 – Dic 2021	1.152.181	14,1%	162.458
2022	1.290.942	12%	154.913
2023	398.597	12%	47.832
2024	558.235	12%	67.008

Fuente: elaboración propia en base a los informes anuales del OB (2022, 2023, 2024, 2025).

En el informe anual 2023² ante la Comisión de Bolsa y Valores, la región se incluye bajo la categoría “Rest of World”—que agrupa también a África y Medio Oriente— y concentra el 32,4 % de los usuarios diarios activos de Facebook a nivel mundial (683 de 2.110 millones) (Meta, 2023). Teniendo en cuenta que combina regiones heterogéneas, la contribución específica de América Latina en cantidad de usuarios queda diluida, lo que sugiere una posible subrepresentación. Ello puede deberse a barreras de acceso a información sobre el mecanismo, así como a sesgos de selección y priorización temática. El propio Consejo informó acciones de acercamiento con actores de estas regiones para ampliar su alcance (Oversight Board, 2022b).

Como se observa en la Tabla N° 3, la proporción de los casos atendidos por el OB entre 2021 y 2025³ fue variable. La actuación del OB aporta evidencia de correcciones a la aplicación de reglas, a la vez que plantea el desafío de alinear la accesibilidad y la representación de América Latina y el Caribe en las apelaciones con su peso entre los usuarios y traducirlo en mejoras en la arquitectura de moderación.

Tabla 3. Decisiones del OB en América Latina (2021-2025)

Año	Cantidad total de decisiones	Casos América Latina y El Caribe	Contenido de la decisión del OB
2020 /2021	20	4 (20%)	Revocó la eliminación en Instagram de una imagen para concientizar sobre cáncer de mama por infringir la norma de desnudos y actividad sexual de adultos, dado que contempla excepciones educativas o de concientización (Brasil, 2020-004-IG-UA). Confirmó mantener en Facebook una publicación de un consejo médico estatal de Brasil que afirmaba que los confinamientos por COVID-19 son ineficaces y que la OMS los había condenado. Evaluó medidas menos intrusivas, como pantallas de advertencia o etiquetado (Brasil, 2021-008-FB-FBR). Revocó la eliminación de una publicación de Instagram sobre ayahuasca y recomendó modificar las normas para permitir el debate positivo sobre usos tradicionales y religiosos de drogas sin fines medicinales cuando existan evidencias históricas (Brasil, 2021-013-IG-UA). Revocó la eliminación de un video en Facebook donde manifestantes criticaban al presidente Iván Duque con insultos homofóbicos por lenguaje de odio. Aplicó la “concesión de interés periodístico” (Colombia, 2021-010-FB-UA).
2022	12	1 (8,33%)	Revocó la eliminación de una caricatura similar al escudo oficial de la Policía Nacional de Colombia que criticaba la brutalidad policial. El contenido había sido eliminado automáticamente por coincidencia con imágenes del Servicio de coincidencias de Meta (Colombia, 2022-004-FB-UA).

2 En estos informes la compañía rinde cuentas sobre su desempeño del año fiscal. Históricamente, incluía usuarios diarios anuales de Facebook y mensuales diarios. Hasta 2023, ofrecía un desglose por regiones que luego dejó de reportar (Meta, 2023).

3 Se incluyen los casos publicados hasta octubre de 2025 en el sitio de OB.

Año	Cantidad total de decisiones	Casos América Latina y El Caribe	Contenido de la decisión del OB
2023	53	3 (5,66%)	Revocó la eliminación de un video donde personas ingresaban a una comisaría y proferían amenazas por violencia e incitación. Valoró el registro documental, la libertad de expresión y la rendición de cuentas, y propuso alternativas de mitigación (Haití, 2023-21-FB-MR). Revocó la eliminación de un video que llamaba a manifestarse contra el gobierno. Subrayó la protección del discurso político y la protesta, y consideró descontextualizada su clasificación como “Lenguaje que incita al odio” (Cuba, 2023-014-IG-UA). Revocó la decisión de conservar un video donde un general instaba a “ir al Congreso Nacional y al Supremo Tribunal”. Evaluó riesgo de daño y potencial incitación a acciones perjudiciales, concluyendo que correspondía aplicar violencia e incitación / organización de actos dañinos (Brasil, 2023-001-FB-UA).
2024	65	4 (6,15%)	Emitió dos decisiones expeditas sobre videos en Instagram y Facebook tras las elecciones presidenciales de 2024. Respaldó conservar el primero en Instagram y anuló la remoción del segundo en Facebook, dado que permitirlos era coherente con las políticas, valores y responsabilidades en derechos humanos (Venezuela, IG-BLFI4MP4 y FB-SV81R3HF). Revocó la decisión de conservar una publicación con un dibujo animado que insinuaba una violación inminente. Ponderó el alto riesgo de normalización de la violencia sexual y revictimización (México, 2024-052-FB-UA). Revocó la eliminación de una imagen con una cita de Yesenia Zamudio. Destacó el interés público y el carácter de protesta frente a la violencia de género, señalando que no constituía llamado a la violencia ni ataque a características protegidas (México, 2024-045-FB-UA).
2025	En curso	5	Revocó la eliminación de una publicación de Facebook donde una persona mayor mostraba un cartel: “Cambio chocolate Bienestar por insulina”. Ponderó el interés público y el daño potencial de la remoción. Consideró que denunciaba una situación de acceso a la salud y no facilitaba transacciones prohibidas de “artículos regulados” (Decisión expedita, México, 2025-057-FB-UA). Revocó la decisión de conservar una publicación de Facebook con un video manipulado por IA que presentaba a un futbolista promoviendo juegos de azar. Enfatizó que inducía a engaño y podía facilitar prácticas comerciales prohibidas, agravadas por el carácter fotorrealista de la deepfake (Brasil, 2025-018-FB-UA). Revocó la decisión de conservar contenido contra una defensora de derechos humanos. Destacó el riesgo de daño fuera de línea, la dimensión de género y el efecto amedrentador sobre la participación cívica (Perú, 2025-012-FB-UA). Revocó la decisión de conservar un video en Facebook que mostraba a facciones disidentes de las FARC. Concluyó que debía ajustarse a las políticas sobre “personas y organizaciones peligrosas”, priorizando la prevención de daños (Decisión expedita, Colombia, 2025-026-FB-UA). Revocó la eliminación de una publicación de Instagram con un poema vinculado a protestas políticas sobre la comunidad LGBT. Señaló que no constituía ataque a características protegidas ni llamado a la violencia (Argentina, 2025-043-IG-UA).

Fuente: elaboración propia en base a informes anuales del Oversight Board (2022, 2023, 2024, 2025) y los casos publicados en su sitio

La presencia de América Latina y el Caribe en la agenda del OB ha sido marginal y temáticamente acotada. El corpus revisado privilegia casos de salud, violencia de género, protestas y seguridad pública. Prácticamente no hubo pronunciamientos sobre integridad electoral pese a las controversias sobre el rol de las redes sociales en la circulación de desinformación, como en el caso de Brasil (von Bülow, Avritzer, Santana, 2024).

4.2.2. El Meta Community Forum: deliberación y consulta pública

Meta lanzó en 2022 los Foros Comunitarios, concebidos como espacios de participación deliberativa. Su diseño se realizó en colaboración con el *Deliberative Democracy Lab* de la Universidad de Stanford y con el *Behavioural Insights Team* del Reino Unido. Según la empresa (2025c), se estructuran como espacios temáticos de discusión, en los que se convoca a usuarios y especialistas a debatir sobre emergentes de la industria y de las plataformas.

En los *Community Forums* no se toman decisiones vinculantes sobre casos concretos ni se interpretan normas comunitarias. Meta presentó estas experiencias como evidencia de su disposición a incorporar la voz de las comunidades de usuarios en la definición de políticas, aunque sin comprometerse a implementar sus recomendaciones (Malkin & Alnemr, 2024).

Hasta el momento, se realizaron cuatro foros⁴. Cada uno comenzó con el reparto de materiales educativos sobre un tema y continuó con discusiones en pequeños grupos, asistidas por expertos. Luego, los participantes emitieron su opinión mediante encuestas.

La variedad de países representados fue cambiante. El foro sobre acoso en Metaverso (2022) contó con la participación de cerca de 6.500 usuarios de 32 países. Por América Latina, asistieron usuarios de Brasil, México, Colombia y Chile (Stanford Deliberative Democracy Lab, 2023). La cantidad de personas se redujo a prácticamente un cuarto en 2023 y la región contó sólo con representantes de Brasil. En 2024 participaron representantes de cinco países de Asia y África (Meta, 2025c).

A partir del análisis de la estructura y funcionamiento de los foros y del OB se observa la baja participación de Latinoamérica en la gobernanza de Meta. Más significativo es que ese diseño opera como un mecanismo de evasión de responsabilidades: Meta busca legitimidad a través de procesos consultivos con baja incidencia en la elaboración e implementación de sus políticas.

5. Cambio de ciclo: nuevas políticas, viejas prácticas opacas

Desde 2018, Meta enfrenta demandas judiciales e investigaciones por prácticas anticompetitivas en Europa, Estados Unidos y, más recientemente, en América Latina —en Brasil, por abuso de posición dominante en IA, mensajería y autopreferencia en las búsquedas—. La empresa no ha podido eludir sanciones legales por sus efectos sobre el mercado ni sociales por su diseño adictivo, como demostró el fallo del Tribunal de Massachusetts (Observacom, 2026), pero logra evitar la intervención estatal a través de regulaciones que limiten su poder. El retroceso en sus políticas de moderación de contenidos se inscribe en ese patrón y puede leerse como una estrategia adaptativa en el contexto político de desregulación desde la llegada de Trump al gobierno de Estados Unidos.

A comienzos de 2025, Meta anunció decisiones que marcaron un retroceso en sus compromisos de derechos humanos y reconfiguraron su arquitectura de moderación: dismanteló instancias profesionales de verificación y revisión *ex post*; automatizó y

4 Se realizaron dos foros en 2022 —sobre desinformación climática y sobre acoso e intimidación en el Metaverso—; uno en 2023 y otro en 2024, ambos dedicados a la inteligencia artificial generativa. Al momento de elaboración de este trabajo, Meta anunciaba la realización de un nuevo foro sobre agentes de inteligencia artificial.

externalizó “comunitariamente” la moderación; y reorientó su discurso hacia la “libertad de expresión” y la “reducción de errores” (Kaplan, 2025). Como señala la especialista Carolina Botero, fue “un baldazo de agua fría al trabajo que venía haciendo la empresa en considerar los derechos humanos en su proceso de moderación”: la aproximación previa podía ser objeto de críticas, pero el compromiso con ese marco se mantenía; con el anuncio de enero, desaparece (Botero, comunicación personal, 7 de mayo de 2025).

Meta había incorporado el programa de *fact-checking* en 2016, a partir de acuerdos con organizaciones de cada país certificadas por la Red Internacional de Verificación de Datos (IFCN, por su sigla en inglés). El proceso de verificación consistía en la remoción de piezas sospechosas de información errónea hasta su revisión y, en caso de confirmarse, quedaba etiquetada y se limitaba su alcance.

Hay trabajos empíricos que dan cuenta de la actividad de *fact-checkers* (Dafonte-Gómez et al., 2022). Mientras algunos identifican limitaciones para frenar la diseminación de desinformación (Wasike, 2023), otros encuentran que incrementa la precisión de las creencias de las personas al reducir sus percepciones erróneas sobre la información (Bachmann & Valenzuela, 2023).

El sistema de Notas de la comunidad no implica sanciones ni reducciones algorítmicas automáticas, sino que depende de contribuciones de usuarios. Organizaciones y analistas advierten que puede aumentar la exposición a desinformación y discursos de odio, trasladar la carga de la prueba a las personas usuarias y reducir los márgenes de control democrático al erosionar la transparencia y la rendición de cuentas, especialmente en ciclos electorales. Aunque reconocen los límites del *fact-checking*, los actores entrevistados leen el cambio de 2025 como estrategia de reposicionamiento político-regulatorio y advierten sobre la brecha entre la continuidad formal de los acuerdos y la inacción operativa en América Latina. Botero agrega que la comparación con modelos exitosos de moderación comunitaria —como Wikipedia o Reddit— resulta engañosa, porque estas operan bajo una lógica de negocio ajena a la economía de la atención y a la monetización de datos de interacción. Sus elementos centrales —incentivos, responsabilidades y reglas de moderación comunitaria— difícilmente pueden transferirse a plataformas cuyo diseño estructural favorece la viralización sobre la deliberación (Comunicación personal, 7 de mayo de 2025).

Se suma así otra preocupación: la vulnerabilidad de las políticas de moderación frente a los cambios de ciclo políticos y los riesgos que ello implica.

Aunque el modelo de *fact-checking* tenía limitaciones, como su baja capacidad de respuesta en tiempo real frente al volumen de contenidos, estas se compensaban con herramientas automatizadas que facilitaban la detección de publicaciones desinformativas. Con este cambio (...) la responsabilidad de la verificación recae en usuarios no especializados que pueden ser vulnerables a la polarización, las brechas digitales y campañas de desinformación. Este modelo podría amplificar narrativas falsas, afectando la calidad del debate público y reduciendo la efectividad de las medidas contra la desinformación en procesos electorales, conflictos sociales o emergencias sanitarias (...) Meta busca proteger su legitimidad y reducir riesgos reputacionales, regulatorios o de pérdida de usuarios (...) sin asumir un rol explícito de moderador editorial (Carolina Fernández, comunicación personal, 2 de julio de 2025).

Meta es una empresa muy sensible al entorno político (...) El caso de Cambridge Analytica fue el disparador de la creación del OB (...). Durante la pandemia del COVID-19 también modificaron su estrategia de moderación y transparencia. Ahora, con el regreso de Trump, Meta gira nuevamente...Esto no implica obediencia directa, pero sí una lectura adaptativa del clima político. Meta intenta alinear su política editorial con las corrientes dominantes

para evitar sanciones o conflictos legales en sus principales mercados (Javier Pallero, comunicación personal, 2 de mayo de 2025).

Creo que el problema de fondo con las notas de la comunidad es que un usuario hace una nota y después esa nota tiene que ser considerada útil por x número de personas que hayan tenido opiniones diferentes en el pasado. El problema es que con cualquier tema controversial no necesariamente se van a poner de acuerdo y entonces esa nota nunca va a aparecer (Olivia Sohr, comunicación personal, 22 de mayo de 2025).

Hasta ahora, Meta no ha publicado una evaluación sobre la fase de prueba de las medidas en Estados Unidos. Los primeros reportes van en sintonía con lo planteado por Sohr sobre la dificultad de alcanzar “consensos” para que las notas cobren visibilidad: el periodista Geoffrey Fowler (*The Washington Post*) redactó 65 notas en cuatro meses para desmentir desinformación de distinto signo, de las cuales se publicaron sólo tres, lo que ilustra demoras y escasa eficacia. (Fowler, 2025).

Aún cuando Meta afirma que se trata de una prueba circunscrita a los Estados Unidos, el alcance transnacional de la plataforma anticipa efectos extraterritoriales. Los casos de Ecuador y Bolivia exhiben una disonancia entre la continuidad formal de convenios con *fact-checkers* y la ausencia de reacciones: no hubo etiquetado ni moderación como en ciclos previos y aumentó la circulación de desinformación electoral (Martínez Elebi, 2025; OBSERVACOM, 2025). En Perú, las autoridades electorales requirieron compromisos explícitos a Meta, X y TikTok en materia de desinformación, fiscalización del gasto en anuncios y verificación de cuentas, señalando la necesidad de garantías procedimentales para las elecciones generales de 2026 (PNUD Perú, 2025).

La transición de Meta hacia un modelo basado en “notas de la comunidad” evidencia en América Latina una desconexión entre la continuidad formal de los convenios de *fact-checking* y su inacción operativa. Además, el sistema exige consensos entre usuarios con opiniones divergentes para ganar visibilidad. En contextos de alta polarización política, sistemas mediáticos concentrados, ausencia de medios auténticamente públicos, cooptación institucional, desigualdades profundas y brechas digitales, el cambio de reglas de Meta aumenta los riesgos para las garantías democráticas.

6. Conclusiones

Este trabajo analizó las estructuras de gobernanza de Meta para la moderación de contenidos entre 2018 y 2025 y sus implicancias para la circulación del debate político en América Latina.

Del análisis se desprende que los organismos creados por la empresa representan un avance formal al incluir múltiples actores en la toma de decisiones corporativas. El OB fue central para incorporar el enfoque de derechos humanos. Sin embargo, los casos de América Latina tuvieron una presencia marginal y una notoria ausencia de casos de desinformación en contextos electorales. Esto ilustra las limitaciones del OB para abordar controversias de este tipo en la región.

El trabajo permitió ver el desfase entre esa arquitectura sofisticada y su limitada incidencia en el diseño de políticas de la empresa, cuyo funcionamiento permanece opaco para la sociedad.

El retroceso anunciado en 2025 supuso un cambio en la aproximación de la empresa al debate público. Si desde 2018 había decidido intervenir activamente —aunque con las limitaciones expuestas—, el nuevo enfoque *hands off*, bajo el argumento de permitir la circulación de todas las expresiones, delega más decisiones a los algoritmos y externaliza la responsabilidad en las personas usuarias. Aunque la empresa sostiene que el cambio se

circunscribe por ahora a Estados Unidos, los riesgos regionales frente a procesos electorales se evidencian en déficits de transparencia publicitaria y en el incremento de la desinformación electoral.

Más allá del debate sobre la relevancia de los *fact-checkers*, la pregunta central sigue siendo por el funcionamiento de la arquitectura de las plataformas y por la necesidad de mayor transparencia sobre sus criterios de selección y promoción de contenidos, así como de empoderar a los usuarios para la deliberación en esos espacios digitales. El caso de Meta ilustra la insuficiencia estructural de la autorregulación corporativa como garantía de derechos: un modelo que, cuando los intereses empresariales y políticos lo requieren, puede desmantelarse unilateralmente sin rendición de cuentas. Las propuestas de UNESCO (2023) y Organización de Estados Americanos (2024) ofrecen elementos claves para desarrollar políticas alineadas con estándares internacionales y asumir responsabilidades concretas por el rol de las plataformas en la democracia.

Los especialistas de la región consultados coinciden en que las respuestas regulatorias no deberían orientarse hacia esquemas de autorregulación ilimitada ni hacia formas de intervención estatal directa sobre los contenidos. En cambio, advierten la necesidad de avanzar hacia marcos híbridos con obligaciones de transparencia, rendición de cuentas y garantías procedimentales —mecanismos de apelación, instancias de consulta pública y explicaciones sobre los criterios de decisión algorítmica— sin intervenir sobre los contenidos. No obstante, esta agenda está fuertemente condicionada en términos políticos: la articulación entre las grandes plataformas y la administración de Trump genera, en distintos gobiernos de América Latina, temores por eventuales represalias comerciales o diplomáticas, lo que puede inhibir la acción legislativa aun cuando existe voluntad política para avanzar. A ello se suma el temor a la censura estatal en una región moldeada por dictaduras y regímenes autoritarios. La tensión entre establecer reglas públicas para actores privados con capacidad de ordenar la circulación de información y las presiones geopolíticas que desalientan estos procesos constituye uno de los principales desafíos que el caso de Meta plantea para la investigación académica y la formulación de políticas públicas en América Latina.

References

- Abramovich, V., Guembe, M. J., & Capurro Robles, M. (2021). *El límite democrático de las expresiones de odio: principios constitucionales, modelos regulatorios y políticas públicas*. Teseo.
- Article 19. (2023). *Manual de moderación de contenidos y libertad de expresión*. UNESCO, European Union.
- Bachmann, I., & Valenzuela, S. (2023). Studying the Downstream Effects of Fact-Checking on Social Media: Experiments on Correction Formats, Belief Accuracy, and Media Trust. *Social Media + Society*, 9(2). <https://doi.org/10.1177/20563051231179694>
- Becerra, M., & Waisbord, S. (2021a). The Need to Reconsider the Orthodoxy of Freedom of Speech in Digital Communication. *Revista de Desarrollo Económico*, 60(232), 295–313.
- Becerra, M., & Waisbord, S. R. (2021b). The curious absence of cybernationalism in Latin America: Lessons for the study of digital sovereignty and governance. *Communication and the Public*, 6(1–4), 67–79.
- Behavioural Insights Team (2022). *Deliberative democracy in action*. <https://www.bi.team/blogs/deliberative-democracy-in-action/>
- Bertoni, E. (2021). Responsabilidad de Actores No Estatales por Violaciones a la Libertad de Expresión. OBSERVACOM. <https://301.tv/XkaAs>

- Bizberge, A. (2021). El modelo de negocio de las plataformas, su concentración y las decisiones que adoptan agravan problemas que están afectando a la democracia: Entrevista a Edison Lanza. *Eptic*, 23(3), 109–115. <https://periodicos.ufs.br/epitic/article/view/16925/12385>
- Bizberge, A., & Mastrini, G. (2024). Regulación de intermediarios de Internet. Análisis de los Foros Latinoamericanos de Gobernanza (2018–2021). *Comunicación Y Sociedad*, 1–28. <https://doi.org/10.32870/cys.v2024.8560>
- Braun, V., & Clarke, V. (2008). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Brinkmann, S. (2014). Unstructured and Semi-Structured Interviewing. En Leavy, P. (Ed.). *The Oxford Handbook of Qualitative Research* (2nd ed.). pp. 277–299.
- CELE (2018). La regulación de internet y su impacto en la libertad de expresión en América Latina. <https://301.tv/mxsOM>
- Chan, K. (2025). *Meta will cease political ads in European Union by fall, blaming bloc's new rules*. AP News. <https://301.tv/TdTRe>
- Chenou, J.M. (2021). Varieties of digital capitalism and the role of the state in internet governance. A view from Latin America. In Haggart, B., Tusikov, N., & Scholte, J.A. (Eds.). *Power and Authority in Internet Governance: Return of the State?* Routledge. <https://doi.org/10.4324/9781003008309-13>
- Consorcio Al Sur (2019). “Oversight Board” de Facebook: una perspectiva desde América Latina y el Caribe. <https://301.tv/ojTye>
- Dafonte-Gómez, A., Míguez-González, M.-I., & Ramahí-García, D. (2022). Fact-checkers on social networks: analysis of their presence and content distribution channels. *Communication & Society*, 35(3), 73–89. <https://doi.org/10.15581/003.35.3.73-89>
- Derechos Digitales (2019). Consejo Asesor de Facebook: ¿La bala de plata para los problemas de la moderación de contenidos? *Derechos Digitales*. <https://301.tv/HwHiL>
- Denzin, N. K., & Lincoln, Y. S. (2005). *The Sage Handbook of Qualitative Research*. Sage Publications.
- Douek, E. (2019). Facebook’s “Oversight Board:” Move fast with stable infrastructure and humility. *North Carolina Journal of Law & Technology*, 21(1), 1–78. <https://301.tv/lErFp>
- Douek, E. (2024). The Meta Oversight Board and the empty promise of legitimacy. *Harvard Journal of Law & Technology*, 37(2), 373–445.
- Edwards, R., & Holland, J. (2013). What is Qualitative Interviewing? In “What is?” *Research Methods Series* (1^o, Vol. 7). Bloomsbury Publishing Plc. <https://doi.org/10.5040/9781472545244>
- Flick, U. (2014). *An Introduction to Qualitative Research*. SAGE.
- Fowler, G. A. (2025, August 4). Zuckerberg fired the fact-checkers. We tested their replacement. *The Washington Post*. <https://301.tv/Xjwgx>
- Gillespie, T. (2018). *Custodians of the internet. Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.
- Gorwa, R. (2019). What is platform governance? *Information Communication and Society*, 22(6), 854–871. <https://doi.org/10.1080/1369118X.2019.1573914>
- Iosifidis, P. (2011), *Global Media and Communication Policy*, Basingstoke and New York: Palgrave Macmillan.
- Kaplan, J. (2025). *More Speech and Fewer Mistakes*. <https://301.tv/RlixV>
- Klonick, K. (2020). The facebook oversight board: Creating an independent institution to adjudicate online free expression. *Yale Law Journal*, 129(8), 2418–2499.
- Linares, A., & Segura, M. S. (2022). Instituciones participativas de políticas de comunicación en Latinoamérica. Modelos, barreras, logros y desafíos. *Comunicación Y Medios*, 31(46), pp. 68–81. <https://doi.org/10.5354/0719-1529.2022.67688>
- Loreti, D. (2021). Redes sociales, plataformas digitales y libertad de expresión: obligaciones de los actores no estatales. *OBSERVACOM*. <https://301.tv/cFYcV>

- Loreti, D. y Lozano, L. (2014). El derecho a comunicar. Los conflictos en torno a la libertad de expresión en las sociedades contemporáneas. Buenos Aires: Siglo XXI.
- Malkin, C., & Alnemr, N. (2024). *Big Tech-driven deliberative projects*. 5, 1–31.
<https://301.tv/jaqEo>.
- Marchal, N., Hoes, E., Klüser, K. J., Hamborg, F., Alizadeh, M., Kubli, M., & Katzenbach, C. (2024). How Negative Media Coverage Impacts Platform Governance: Evidence from Facebook, Twitter, and YouTube. *Political Communication*, 00(00), 1–19.
<https://doi.org/10.1080/10584609.2024.2377992>
- Martínez Elebi, C. (2025). *Elecciones en Bolivia: más desinformación electoral y sin participación activa de las plataformas*. OBSERVACOM. <https://301.tv/OGaWC>
- Mastrini, G., & Becerra, M. (2006), *Periodistas y magnates: estructura y concentración de las industrias culturales en América Latina*, Buenos Aires, Prometeo.
- Meta (2023). *Form 10-K: Annual report pursuant to Section 13 or 15(d) of the Securities Exchange Act of 1934 for the fiscal year ended December 31, 2023*. <https://301.tv/itHYv>
- Meta (2025a). *Ending Political, Electoral and Social Issue Advertising in the EU in Response to Incoming European Regulation*. Meta. Newsroom for Europe. <https://301.tv/ZWmBN>
- Meta (2025b). *Información sobre la biblioteca de anuncios de Meta*. <https://301.tv/kkCjx>
- Meta (2025c). *Community Forums at Meta*. Transparency Center. <https://301.tv/cdigG>
- Meta (2025d). *Introducción a las Normas de publicidad*. <https://301.tv/FcaLM>
- Meta (2025e). *Información sobre temas sociales*. <https://301.tv/IUjOF>
- Nieborg, D. B., & Poell, T. (2025). Analyzing institutional platform power: Evolving relations of dependence in the mobile digital advertising ecosystem. *New Media & Society*, 27(4), 1909–1927. <https://doi.org/10.1177/14614448251314405>
- OBSERVACOM (2019). *Comentarios preliminares de OBSERVACOM sobre el Oversight Board de Facebook*. OBSERVACOM. <https://301.tv/MrOuV>
- OBSERVACOM (2025). *Menos moderación, más desinformación: el impacto de las nuevas políticas de Meta en las elecciones de Ecuador*. OBSERVACOM. <https://301.tv/YxwTK>
- OBSERVACOM (2026). *Tribunal de Massachusetts habilita demanda contra Meta por diseño adictivo en Instagram*. Observacom. <http://bit.ly/4cSUGUB>
- Organización de los Estados Americanos (2024). *Inclusión digital y gobernanza de contenidos en internet*. Comisión Interamericana de Derechos Humanos. Relatoría Especial para la Libertad de Expresión.
https://www.oas.org/es/cidh/expresion/informes/Inclusion_digital_esp.pdf
- Oversight Board (2022a). *Securing ongoing funding for the Oversight Board*. Disponible en <https://www.oversightboard.com/news/1111826643064185-securing-ongoing-funding-for-the-oversight-board/>
- Oversight Board (2022b). *Oversight Board Annual Report 2021*. Oversight Board LLC. Disponible en <https://www.oversightboard.com/wp-content/uploads/2023/11/Annual-Report-English.pdf>.
- Oversight Board (2023). *Informe anual 2022 del Consejo Asesor de Contenidos*. <https://oversightboard.com/attachment/263717339654593/>
- Oversight Board (2024a). *2023 Annual Report. Improving how Meta treats people and communities around the world*. Disponible en <https://www.oversightboard.com/wp-content/uploads/2024/06/Oversight-Board-2023-Annual-Report.pdf>
- Oversight Board (2024b). *Contenido antiolektivistas tras las elecciones en Venezuela*. Disponible en <https://www.oversightboard.com/decision/bun-xgwxj6hs/?lang=es>.
- Oversight Board. (2025a). *2024 Annual Report: Improving how Meta treats people and communities around the world*. Oversight Board. Disponible en <https://www.oversightboard.com/wp-content/uploads/2025/08/2024-Annual-Report.pdf>
- Oversight Board (2025b). *Oversight Board Charter*. Oversight Board. Disponible en <https://www.oversightboard.com/governance/>
- Patton, M. Q. (2002). *Qualitative Research & Evaluation Methods*. Sage Publications.

- Peñarredonda, J.L.; Ronderos, M.T.; Scofield, L.; Rincón, A.; Buitrago, E.; Skoknic, F.; Almeida, M.; Mena, P. Así presiona la “mano invisible” de las gigantes tecnológicas a las autoridades latinoamericanas. Centro Latinoamericano de Investigación Periodística.
- Pita, M. (2022). *Transparencia de la moderación privada de contenidos - una mirada de las propuestas de sociedad civil y legisladores de América Latina*
https://www.observacom.org/wp-content/uploads/2022/04/PapersTransparencia-MarinaPita-ESP-WEB.pdf?utm_source=chatgpt.com
- PNUD Perú (2025). Encuentro internacional “Empresas tecnológicas y procesos electorales” [Video]. Youtube. https://www.youtube.com/watch?v=_ztKx4JA3DE
- Poell, T., Nieborg, D., & van Dijck, J. (2019). Platformisation. *Internet Policy Review*, 8(4).
<https://doi.org/10.14763/2019.4.1425>
- Puppis, M., & Van den Bulk, H. (2019). Introducción: las políticas de medios y la investigación de políticas. En Van den Bulck, H., Puppis, M., Donders, K., & Van Audenhove, L. (eds). *The Palgrave Handbook of Methods for Media Policy Research* (pp. 1–11). Palgrave Macmillan.
- Stanford Deliberative Democracy Lab. (2023). *Meta Community Forum: Results Analysis*.
- Suzor, N. (2018). Digital Constitutionalism: Using the Rule of Law to Evaluate the Legitimacy of Governance by Platforms. *Social Media + Society*, 4(3).
<https://doi.org/10.1177/2056305118787812>
- UNESCO. (2023). *Directrices para la gobernanza de las plataformas digitales*. <https://301.tv/DjrKC>
- van Dijck, J., Poell, T., & de Waal, M. (2018). *The Platform Society*, (New York, 2018; online edn, Oxford Academic, 18 Oct. 2018. <https://doi.org/10.1093/oso/9780190889760.001.0001>
- von Bülow, M., Avritzer, L., & Santana, E. (eds). (2024). *Democracy under attack. Social media and disinformation in Brazilian elections*. Springer. Cham, Switzerland.
- Waisbord, S. (2013). *Vox populista. Medios, periodismo, democracia*. Gedisa: Buenos Aires.
- Wasike, B. (2023) You’ve been fact-checked! Examining the effectiveness of social media fact-checking against the spread of misinformation. *Telematics and Informatics Reports*, Volume 1. <https://doi.org/10.1016/j.teler.2023.100090>