

Datificación social y auditabilidad informacional: aplicación exploratoria al caso brasileño

Social Datafication and Informational Auditability: An Exploratory Application to the Brazilian Case

Fabiano Couto Corrêa da Silva

fabiano.c.silva@ufes.br

<https://orcid.org/0000-0001-5014-8853>

Universidade Federal do Espírito Santo, Brasil

<https://ror.org/05sxf4h28>

Recibido: 16/06/2026 **Revisado:** 24/07/2026 **Aprobado:** 16/08/2026 **Publicado:** 01/09/2026

Cómo citar: Couto Corrêa da Silva, F. (2026). Datificación social y auditabilidad informacional: aplicación exploratoria al caso brasileño. *Universitas XXI*, 45, pp. 95-115.
<https://doi.org/10.17163/uni.n45.2026.04>

Resumen

La datificación social convierte prácticas, cuerpos y territorios en datos procesables y, con ello, organiza qué fenómenos pueden capturarse, clasificarse, preservarse y reutilizarse como evidencia. Este artículo construye y aplica exploratoriamente una matriz crítica de auditabilidad informacional al caso brasileño. La investigación es teórico-práctica, crítica y documental. El corpus se cerró el 4 de agosto de 2026 y reúne 40 fuentes publicadas entre 2000 y 2026: 12 registros documentales centrales —normas brasileñas, instrumentos internacionales, informes y políticas de Meta, TikTok y X, y documentación sobre modelos fundacionales— y 28 trabajos científicos empleados en la construcción conceptual. La selección atendió a la relevancia temática, al carácter público, a la vigencia, a la identificabilidad de la fuente y a su relación directa con la recolección, clasificación, moderación, recomendación, acceso, preservación o reutilización de datos. Las ocho dimensiones de la matriz se operacionalizaron mediante 32 indicadores documentales. Cada indicador se codificó como presente, parcial o no verificable públicamente; la suma produjo valoraciones descriptivas de condición sólida, parcial o frágil, sin pretensión de constituir una escala validada. La aplicación sugiere una condición relativamente sólida en transparencia normativa, parcial en transparencia clasificatoria y contestabilidad, y frágil en trazabilidad, documentación, acceso científico, preservación y justicia cognitiva. Se concluye que la auditabilidad informacional es una condición para una ciencia abierta crítica y para la soberanía informacional, y que Brasil necesita mecanismos públicos de acceso, documentación, preservación y verificación de la evidencia producida por plataformas.

Palabras clave

Datificación, auditabilidad, algoritmos, desigualdad, colonialidad, transparencia, plataformas, Brasil.

Abstract

Social datafication converts practices, bodies, and territories into processable data and thereby organizes which phenomena can be captured, classified, preserved, and reused as evidence. This article develops and exploratorily applies a critical matrix of informational auditability to the Brazilian case. The study is theoretical-practical, critical, and documentary. The corpus was closed on August 4, 2026, and comprises 40 sources published between 2000 and 2026: 12 core documentary records—Brazilian regulations, international instruments, Meta, TikTok, and X reports and policies, and documentation on foundation models—and 28 scholarly works used for conceptual construction. Selection criteria included thematic relevance, public availability, currency, source identifiability, and direct relation to data collection, classification, moderation, recommendation, access, preservation, or reuse. The matrix's eight dimensions were operationalized through 32 documentary indicators. Each indicator was coded as present, partial, or not publicly verifiable; the aggregate supported descriptive assessments of solid, partial, or fragile conditions, without claiming a validated measurement scale. The application suggests a relatively solid condition for normative transparency, partial conditions for classificatory transparency and contestability, and fragile conditions for traceability, documentation, scientific access, preservation, and cognitive justice. The study concludes that informational auditability is a prerequisite for critical open science and informational sovereignty, and that Brazil needs public mechanisms for access, documentation, preservation, and verification of platform-produced evidence.

Keywords

Datafication, auditability, algorithms, inequality, coloniality, transparency, platforms, Brazil.

Introducción

Las ciencias sociales y la comunicación han dedicado una atención creciente a los contenidos nocivos en entornos digitales, a sus dinámicas de propagación y a sus efectos sobre la convivencia democrática. Ese debate suele concentrarse en la superficie visible de los fenómenos —los contenidos que circulan, su moderación y sus consecuencias— y deja en segundo plano la capa que organiza esa visibilidad. Antes de que un contenido sea recomendado, restringido o eliminado, ya ha sido capturado, categorizado y procesado por infraestructuras que deciden qué se registra, cómo se nombra y bajo qué condiciones puede examinarse. Este artículo desplaza el foco hacia esa

condición estructural: la datificación social que organiza la clasificación, la circulación y la auditabilidad de los fenómenos sociales.

La datificación, entendida como la conversión de prácticas, relaciones, cuerpos y territorios en datos susceptibles de almacenamiento, análisis y comercialización (van Dijck, 2014; Gendler, 2024), no es un proceso técnicamente neutro ni socialmente homogéneo. Las plataformas digitales y los sistemas algorítmicos producen continuamente datos sobre poblaciones y comportamientos, a la vez que definen, mediante decisiones privadas y poco visibles, qué puede capturarse, clasificarse, preservarse y reutilizarse como evidencia social y científica (Couldry y Mejias, 2019; Zuboff, 2019). En esa medida, la datificación funciona como un sistema de poder, no solo como una práctica de medición.

La pregunta relevante no es únicamente qué circula, sino quiénes pueden observar, documentar y verificar cómo se produce esa circulación. Esta cuestión adquiere una densidad específica en el Sur Global. Los países latinoamericanos, y Brasil de manera destacada, son intensamente datificados por plataformas e infraestructuras de propiedad extranjera, pero disponen de capacidades limitadas para auditar esos sistemas, acceder a sus datos, contestar sus clasificaciones y producir conocimiento autónomo a partir de ellos (Milan y Tréré, 2019; Ricaurte, 2019; Couldry y Mejias, 2021). La asimetría alcanza tanto la propiedad de los datos como las condiciones de su examen y reproduce dependencias epistémicas descritas por las teorías de la colonialidad y el colonialismo de datos (Kwet, 2019; Mejias y Couldry, 2024).

Frente a este escenario, el artículo propone una matriz crítica de auditabilidad informacional. La auditabilidad informacional designa la capacidad social, científica e institucional de rastrear, documentar, interpretar, verificar y reutilizar registros sobre los procesos mediante los cuales los datos son recogidos, clasificados, moderados, recomendados, almacenados, circulados y reutilizados. La matriz no se limita a evaluar la transparencia declarada por las plataformas: examina si existen condiciones documentales efectivas de reproducibilidad, preservación, contestación y control público.

La pregunta de investigación se reformula en términos compatibles con el alcance documental del estudio: ¿qué condiciones de auditabilidad informacional pueden identificarse documentalmente en las infraestructuras de datificación que operan en Brasil, qué dimensiones aparecen más frágiles y cómo se relacionan esas condiciones con la justicia cognitiva, la transparencia pública y la reutilización científica de datos? El objetivo general consiste

en construir y aplicar exploratoriamente una matriz crítica de auditabilidad informacional al caso brasileño, con el propósito de identificar fortalezas, limitaciones y necesidades de infraestructura pública de evidencia.

El texto se organiza en seis secciones. Tras esta introducción, la segunda aborda el estado de la cuestión sobre datificación, poder algorítmico, colonialidad de datos, justicia cognitiva y auditabilidad. La tercera expone el diseño metodológico, el corpus, los criterios de selección y el protocolo de codificación. La cuarta presenta la matriz y su aplicación exploratoria al caso brasileño. La quinta examina su utilidad para una ciencia abierta crítica y para la soberanía informacional. Finalmente, la sexta sintetiza las principales contribuciones del estudio, sus límites y las líneas de validación futura.

Estado de la cuestión

Datificación social, poder algorítmico y desigualdad

La noción de datificación describe el proceso por el cual aspectos de la vida que antes no se registraban de forma sistemática pasan a representarse como datos cuantificables, agregables y analizables (van Dijck, 2014). Lo que se cuantifica no constituye un reflejo transparente de lo real, sino el resultado de decisiones sobre qué medir, con qué categorías y con qué finalidad. Gendler (2024) traza una genealogía de la datificación social y muestra cómo las tecnologías digitales codificaron rasgos individuales, cuantificaron interacciones y modelaron representaciones sociales, en un proceso que la inteligencia artificial intensifica. La creencia de que los datos ofrecen un acceso objetivo a la realidad —el dataísmo— oculta el carácter construido y situado de todo dato.

La datificación se materializa en plataformas. La sociedad de plataformas organiza la interacción social en arquitecturas privadas que registran cada gesto y lo convierten en insumo para la predicción y la influencia del comportamiento (van Dijck *et al.*, 2018). Zuboff (2019) denominó capitalismo de vigilancia al modelo que extrae datos de la experiencia humana para producir predicciones comercializables, y Couldry y Mejias (2019) mostraron que esa extracción instaure relaciones de datos que naturalizan la conversión de la vida cotidiana en flujo de información apropiable. Los datos no son un recurso encontrado, sino un recurso producido mediante infraestructuras que concentran poder.

El poder algorítmico es la cara operativa de ese orden. Los algoritmos no se limitan a procesar datos: intervienen en la realidad que ordenan, jerarquizan visibilidades y producen efectos sociales. Mittelstadt *et al.* (2016) sistematizaron cuestiones éticas como opacidad, equidad y trazabilidad. La clasificación algorítmica distribuye oportunidades y riesgos, define quién es sospechoso o prioritario y quién resulta invisible, y puede reproducir desigualdades preexistentes. Eubanks (2018) mostró cómo sistemas automatizados de elegibilidad penalizan a poblaciones pobres; Noble (2018) documentó representaciones racistas y sexistas en motores de búsqueda; y Benjamin (2019) conceptualizó el modo en que la tecnología codifica discriminación bajo apariencia de neutralidad.

Estas asimetrías operan en varios planos: visibilidad, porque las plataformas deciden qué se muestra y qué se silencia; la clasificación, porque las categorías que ordenan a los sujetos rara vez son públicas o discutibles; acceso, porque quien produce los datos no es necesariamente quien puede examinarlos; y extracción y control, porque el valor generado se concentra en quienes poseen las infraestructuras. Pasquale (2015) describió una sociedad de cajas negras, mientras Crawford (2021) situó la inteligencia artificial dentro de cadenas materiales de extracción de recursos, trabajo y datos que reproducen jerarquías globales.

La perspectiva feminista de los datos añade una atención sostenida a quién queda fuera de la cuenta. D'Ignazio y Klein (2020) proponen examinar quién recoge los datos, sobre quién y con qué propósito, y recuerdan que aquello que no se mide tiende a volverse invisible para las políticas públicas. La datificación produce desigualdad tanto por lo que clasifica como por lo que omite. Reconocerlo obliga a desplazar la atención desde el contenido de los datos hacia las condiciones de su producción, examen y reutilización, que es el terreno de la auditabilidad.

Colonialidad de datos y justicia cognitiva en el Sur Global

La crítica de la datificación adquiere un giro decolonial cuando se examina su distribución geopolítica. Couldry y Mejias (2019) propusieron la categoría de colonialismo de datos para nombrar la apropiación de la vida social mediante la extracción de datos, en continuidad histórica con el colonialismo territorial. Mejias y Couldry (2024) sostienen que este nuevo colonialismo se superpone a efectos persistentes del antiguo y concentra el valor extraído

en centros corporativos del Norte global. La metáfora colonial designa una estructura de apropiación, dependencia y gobierno a distancia.

Para el Sur Global, la cuestión decisiva es su posición subordinada en la economía política de los datos. Milan y Treré (2019) advirtieron contra el universalismo de datos y reclamaron una desoccidentalización de los estudios críticos de datos como reparación de una injusticia cognitiva. Ricaurte (2019) propuso un modelo para analizar la colonialidad del poder a través de los datos. Estas contribuciones reactivan una tradición latinoamericana que incluye la colonialidad del poder de Quijano (2000) y que hoy abarca la datificación de plataformas y la representación algorítmica de cuerpos racializados (Robledo Maturana, 2023; Valente & Grohmann, 2024).

La dependencia no es solo económica, sino infraestructural. Nubes de cómputo, modelos de lenguaje, sistemas de identificación y plataformas que organizan la vida social en la región pertenecen mayoritariamente a corporaciones extranjeras (Kwet, 2019; Birhane, 2020). Kwet (2019) describió el colonialismo digital como una forma de imperio basada en el control de infraestructura, datos y estándares; Birhane (2020) mostró cómo la colonización algorítmica de África puede presentarse bajo el lenguaje del desarrollo; y Salami (2024) subrayó la necesidad de marcos de gobernanza propios. En América Latina, Ricaurte *et al.* (2024) analizaron cómo imaginarios socio-técnicos externos pueden incorporarse de manera acrítica y erosionar capacidades locales de gobernanza.

El concepto de justicia cognitiva ofrece una clave normativa. Las epistemologías del Sur sostienen que la injusticia social contiene una injusticia epistémica: la desvalorización sistemática de formas de conocer ajenas al canon dominante (Santos, 2014). Aplicada a la datificación, esta perspectiva exige que las poblaciones datificadas no sean tratadas solo como fuentes de datos, sino como sujetos capaces de definir qué se mide, cómo se interpreta y para qué se usa. Los principios CARE para la gobernanza de datos indígenas sitúan el beneficio colectivo, la autoridad para controlar, la responsabilidad y la ética como condiciones de un tratamiento justo, en complemento de los principios FAIR de gestión técnica (Carroll *et al.*, 2020; Wilkinson *et al.*, 2016).

La relación con la ciencia abierta es directa y problemática. La ciencia abierta promueve acceso, reutilización y reproducibilidad, pero una parte sustancial de los datos que describen la vida social está bajo control privado y resulta inaccesible para la investigación pública. En el Sur Global esta restricción se agrava porque universidades e instituciones públicas dispo-

nen de menor poder de negociación y de infraestructuras propias limitadas para preservar y reutilizar datos a gran escala (Valente y Grohmann, 2024; Grohmann, 2025). La soberanía informacional —la capacidad de gobernar las condiciones de producción, circulación y uso de los datos— aparece así como horizonte normativo y material.

De la transparencia a la auditabilidad informacional

La transparencia se ha convertido en la respuesta dominante frente a la opacidad de los sistemas algorítmicos. Informes, políticas de uso, descripciones técnicas y portales de datos prometen abrir las cajas negras a la inspección pública. Ananny y Crawford (2018) identificaron los límites de ese ideal: ver el interior de un sistema no equivale a comprenderlo ni a poder gobernarlo. La transparencia puede ofrecer información parcial o no verificable y desplazar la responsabilidad hacia quien observa sin dotarlo de capacidad de intervención. En el Sur Global, la información publicada por las plataformas suele presentarse en agregados globales o regionales, sin microdatos ni garantías de continuidad.

Conviene distinguir entre disponibilidad y verificabilidad. La transparencia alude a que cierta información sea visible o publicada. La auditabilidad alude a que esa información permita reconstruir, examinar y contrastar de manera independiente los procesos que describe. Un sistema puede ser nominalmente transparente y, a la vez, no auditable cuando la información disponible no permite reproducir sus resultados, identificar cambios ni contestar decisiones.

Las herramientas habituales de transparencia resultan insuficientes ante este criterio. Los informes corporativos presentan cifras agregadas sobre contenidos removidos o solicitudes atendidas, pero rara vez permiten verificar cómo se clasificaron casos concretos. Las interfaces de programación condicionan el acceso de la investigación a términos definidos unilateralmente por las empresas. Como referencia comparativa, el Reglamento de Servicios Digitales de la Unión Europea prevé acceso de investigadores acreditados a datos de grandes plataformas (European Union, 2022), y el Reglamento Delegado (UE) 2025/2050 desarrolló los procedimientos técnicos y jurídicos para ese acceso (European Commission, 2025). No existe un mecanismo equivalente de alcance general en Brasil.

La literatura sobre documentación de sistemas ha avanzado instrumentos próximos a la auditabilidad. Las fichas de conjuntos de datos documentan

procedencia, composición y usos previstos para que terceros evalúen pertinencia y límites (Gebru *et al.*, 2021). El Índice de Transparencia de Modelos Fundacionales de 2025 evaluó 13 empresas mediante indicadores sobre datos, cómputo, riesgos, uso e impactos y encontró una media de 40 puntos sobre 100, inferior a los 58 puntos de 2024, con opacidad persistente sobre datos de entrenamiento, cómputo y efectos posteriores al despliegue (Wan *et al.*, 2025). La documentación es, por tanto, un componente indispensable, pero no garantiza por sí sola verificación independiente ni control público.

La auditabilidad informacional articula tres exigencias inseparables: reproducibilidad, para rehacer y comprobar resultados; preservación, para evitar que la evidencia social se degrade o desaparezca; y control social, para que sociedad civil, universidades e instituciones públicas puedan examinar y contestar sistemas. Estas exigencias se descomponen en dimensiones observables y orientan el diseño metodológico presentado a continuación.

Metodología

Diseño y alcance

El estudio adopta un diseño teórico-práctico, crítico y documental. Es teórico porque su producto principal es una matriz conceptual que articula categorías de los estudios críticos de datos, la comunicación y la ciencia de la información. Es práctico porque la matriz se aplica al caso brasileño para examinar su utilidad analítica. Es crítico porque interroga las relaciones de poder que constituyen las infraestructuras de datificación. Es documental porque su base empírica son fuentes públicas y no datos de recolección primaria. El trabajo no incluye entrevistas, encuestas, experimentos ni análisis estadístico de bases de plataformas, y no formula inferencias poblacionales.

Búsqueda, composición del corpus y criterios de selección

La búsqueda, la revisión y la actualización basada en documentos concluyeron el 4 de agosto de 2026. Se consultaron portales normativos oficiales de Brasil y de la Unión Europea, documentos de UNESCO, páginas institucionales y centros de transparencia de plataformas, registros DOI y páginas

editoriales de artículos y libros. Las combinaciones de búsqueda se formularon en español, portugués e inglés con los términos datificación/datafication, auditabilidad/auditability, transparencia algorítmica/algorithmic transparency, acceso a datos de plataformas/platform data access, colonialismo de datos/data colonialism, justicia cognitiva/cognitive justice, ciencia abierta/open science, Brasil/Brazil y PL 2338/2023.

El corpus final reúne 40 fuentes publicadas entre 2000 y 2026. Doce registros documentales centrales sustentan la aplicación: tres marcos brasileños; cinco instrumentos internacionales, de ciencia abierta o de gobernanza de datos; tres documentos corporativos de Meta, TikTok y X; y el Índice de Transparencia de Modelos Fundacionales de 2025. Las otras 28 obras científicas fundamentan la construcción conceptual y la discusión. La Tabla 1 explicita la composición, los periodos y la función de cada bloque.

Tabla 1

Composición del corpus documental y función analítica

Bloque del corpus	Registros considerados y periodo	Cantidad	Función analítica
Marcos normativos brasileños	Marco Civil de Internet (2014); LGPD (2018); ficha de tramitación del PL 2.338/2023, actualizada en 2026	3	Examinar reglas, derechos, obligaciones de gobernanza y mecanismos de contestación.
Gobernanza internacional y ciencia abierta	Principios FAIR (2016); principios CARE (2020); Recomendación de UNESCO sobre Ciencia Abierta (2021); Reglamento de Servicios Digitales (2022); Reglamento Delegado (UE) 2025/2050	5	Relacionar apertura, documentación, preservación, reutilización, ética y acceso de investigadores.
Documentos de plataformas	Meta, Community Standards Enforcement Report, primer trimestre de 2025; TikTok, Community Guidelines Enforcement Report, julio-septiembre de 2024; X, Global Transparency Report, enero-junio de 2024. Consulta: 4 de agosto de 2026	3	Identificar información publicada sobre reglas, moderación, clasificación, acceso, trazabilidad y recursos.
Transparencia de modelos fundacionales	The 2025 Foundation Model Transparency Index (publicado en 2025)	1	Contrastar la documentación y la transparencia de datos, cómputo, riesgos, usos e impactos.
Literatura científica	Artículos y libros sobre datificación, plataformas, colonialismo de datos, justicia cognitiva, documentación, auditabilidad y ciencia abierta (2000-2025)	28	Derivar dimensiones e indicadores y sustentar la interpretación crítica.
Total	Fuentes incluidas en la lista de referencias	40	Corpus teórico-documental del estudio.

Nota. Las fechas de consulta de los documentos corporativos corresponden al cierre del corpus.

Se incluyeron fuentes con autoría institucional o personal identificable, fecha o versión verificable, acceso público y relación directa con al menos uno de los procesos analizados: recolección, clasificación, moderación, recomendación, almacenamiento, acceso, preservación, reutilización o contestación de datos. Se priorizaron documentos vigentes y literatura reciente, sin excluir referencias fundacionales indispensables. Se excluyeron duplicados, piezas de opinión sin sustento documental, materiales sin texto completo accesible, versiones sustituidas por documentos posteriores y textos exclusivamente técnicos sin implicaciones para gobernanza o auditabilidad.

Construcción, codificación y control de consistencia

La matriz se construyó mediante síntesis conceptual deductiva. A partir de la literatura sobre transparencia, rendición de cuentas algorítmica, documentación de datos y modelos, gobernanza de datos científicos, justicia cognitiva y colonialismo de datos, se identificaron propiedades vinculadas con la posibilidad de examinar y reutilizar evidencias. Las propiedades convergentes se agruparon en ocho dimensiones. Cada dimensión se operacionalizó con cuatro indicadores documentales, para un total de 32 indicadores.

La aplicación al caso brasileño se realizó mediante una ficha de codificación organizada por dimensiones y fuentes. A cada indicador se le asignó 1 punto cuando existía información pública directa y suficientemente específica; 0,5 cuando esta era parcial, indirecta o dependía de condiciones no documentadas; y 0 cuando no fue posible verificar información pública pertinente. La suma de los cuatro indicadores permitió establecer una valoración descriptiva: sólida (3-4 puntos), parcial (1,5-2,5 puntos) y frágil (0-1 punto). Esta puntuación tuvo como único propósito sistematizar la interpretación documental; no constituye una escala psicométrica, un ranking de plataformas ni una medida generalizable.

La consistencia del análisis se aseguró mediante dos revisiones independientes del corpus y una segunda evaluación diferida, contrastada con una matriz de respaldo documental y las referencias precisas de cada fuente. Las discrepancias se resolvieron revisando los criterios operativos y privilegiando la interpretación más conservadora cuando la información resultaba ambigua. El estudio no contó con codificadores independientes ni con validación por un panel de expertos, aspecto que se reconoce como una limitación y, al mismo tiempo, como una línea de fortalecimiento para futuras investigaciones.

Durante la revisión posterior al dictamen se utilizó ChatGPT (OpenAI, versión disponible en agosto de 2026) como herramienta complementaria para apoyar la revisión lingüística y estructural, la sistematización de observaciones y la comprobación de consistencia interna. No se utilizó para generar, alterar o inventar datos, resultados y referencias. El autor revisó cada propuesta y contrastó la información normativa, documental y bibliográfica con las fuentes citadas.

Resultados

Matriz de auditabilidad informacional

La matriz organiza el análisis en ocho dimensiones complementarias. Las cuatro primeras se refieren a la inteligibilidad de los sistemas: reglas, categorías, decisiones y documentación. Las cuatro siguientes se refieren a la apropiabilidad pública de la evidencia: acceso, preservación, contestación e incorporación de perspectivas del Sur Global. Ninguna dimensión es suficiente por sí sola; una infraestructura puede ser fuerte en reglas publicadas y débil en acceso científico, trazabilidad o justicia cognitiva.

Tabla 2
Matriz de auditabilidad informacional

Dimensión	Pregunta orientadora	Indicadores documentales	Riesgos en ausencia	Aporte a la justicia cognitiva
Transparencia normativa	¿Existen reglas públicas y claras sobre recolección, clasificación, recomendación, moderación y reutilización?	Políticas publicadas; términos de uso; marcos legales; normas técnicas declaradas	Arbitrariedad; imposibilidad de exigir cumplimiento; asimetría entre empresa y usuarios	Somete el tratamiento de datos a reglas discutibles colectivamente.
Transparencia clasificatoria	¿Se explicitan las categorías con que se clasifican sujetos, contenidos, riesgos y comportamientos?	Taxonomías; definiciones operativas; criterios de etiquetado; criterios de moderación	Estigmatización o invisibilización sin posibilidad de revisión	Permite cuestionar categorías importadas que no reflejan realidades locales.
Trazabilidad	¿Es posible reconstruir decisiones, flujos, registros, alteraciones e intervenciones?	Registros de actividad; versionado; historiales de decisión; identificadores persistentes	Decisiones irreconstruibles; pérdida de evidencia; responsabilidad difusa	Habilita verificación independiente y atribución de responsabilidad.

Dimensión	Pregunta orientadora	Indicadores documentales	Riesgos en ausencia	Aporte a la justicia cognitiva
Documentación y metadatos	¿Existe información técnica, semántica y contextual suficiente para interpretar los datos?	Fichas de datos/modelos; diccionarios; metadatos estandarizados; procedencia	Datos ininterpretables; reutilización errónea; dependencia del productor	Distribuye capacidad interpretativa más allá de quien produce los datos.
Acceso científico	¿En qué condiciones acceden los investigadores a los datos?	Interfases documentadas; programas de acceso; criterios de muestreo; restricciones explícitas	Acceso revocable; sesgos por acceso parcial; exclusión del Sur Global	Reconoce a universidades públicas como sujetos legítimos de examen.
Preservación y reutilización	¿Pueden preservarse, interoperar y reutilizarse los datos de forma ética y reproducible?	Repositorios sostenibles; interoperabilidad; licencias; planes de preservación	Evidencia efímera; irreproducibilidad; pérdida del registro social	Convierte la evidencia social en bien común duradero.
Contestabilidad	¿Existen mecanismos para impugnar clasificaciones o decisiones automatizadas?	Apelación; revisión humana; recursos administrativos/judiciales; reclamación	Decisiones inapelables; indefensión; concentración de poder	Restituye agencia a las poblaciones datificadas.
Justicia cognitiva	¿La infraestructura incorpora perspectivas, necesidades e instituciones del Sur Global?	Participación local; adecuación cultural; reconocimiento institucional; gobernanza inclusiva	Universalismo de datos; dependencia epistémica; jerarquías coloniales	Sitúa la equidad epistémica como criterio explícito.

La matriz admite un uso descriptivo, orientado a identificar fortalezas y fragilidades, y un uso normativo, destinado a orientar el diseño de infraestructuras con mayores niveles de auditabilidad. En este estudio se utiliza con un propósito cualitativo y exploratorio. La puntuación únicamente sintetiza la presencia documental de los indicadores y debe interpretarse conjuntamente con la evidencia específica correspondiente a cada dimensión.

Aplicación exploratoria al caso brasileño

El caso brasileño combina densidad normativa creciente y dependencia de infraestructuras privadas. El Marco Civil de Internet fija principios de uso de la red (Brasil, 2014), y la LGPD establece derechos sobre el tratamiento de datos personales y una autoridad nacional (Brasil, 2018). El PL 2.338/2023, aprobado por el Senado y aún en análisis por una comisión especial de la Cámara de Diputados en 2026, incorpora un enfoque basado en riesgos, de-

beres de transparencia y mecanismos de gobernanza (Brasil, 2026). Estos instrumentos ofrecen evidencia directa para tres de los cuatro indicadores de transparencia normativa, pero no garantizan por sí mismos acceso a registros técnicos ni auditabilidad de sistemas privados.

En transparencia clasificatoria, Meta, TikTok y X publican reglas de comunidad y categorías generales de infracción. TikTok, por ejemplo, describe detección automatizada, revisión adicional, notificación y apelación; Meta ofrece un panel de aplicación de normas; y X divulga categorías agregadas de cumplimiento. Sin embargo, los documentos no exponen taxonomías operativas completas, umbrales de clasificación, tasas de error por idioma o criterios detallados de recomendación para Brasil. Por ello, la evidencia satisface parcialmente dos indicadores y la dimensión recibe valoración parcial.

La trazabilidad es frágil. Ninguno de los tres documentos corporativos ofrece registros públicos que permitan reconstruir la cadena de una decisión concreta de recomendación, reducción de alcance, etiquetado o remoción. Tampoco se publican historiales de versión suficientemente granulares para vincular un cambio de regla con casos y resultados posteriores. La información agregada permite conocer volúmenes, pero no reconstituir decisiones individualizadas o flujos algorítmicos.

La documentación y los metadatos también son frágiles. Los informes corporativos describen categorías y resultados agregados, pero no proporcionan diccionarios de variables, información de procedencia, criterios de muestreo completos ni fichas reutilizables de los conjuntos de datos que sustentan las cifras. En el campo de los modelos fundacionales, el FMTI de 2025 muestra que la opacidad sobre datos de entrenamiento, cómputo, uso e impactos no es excepcional, sino sistémica (Wan *et al.*, 2025).

El acceso científico obtiene valoración frágil. Los programas e interfaces disponibles dependen de decisiones empresariales y pueden imponer elegibilidad, límites, formatos y condiciones de uso revocables. El régimen europeo ofrece una referencia distinta: el DSA y su reglamento delegado establecen un procedimiento jurídico para solicitudes de investigadores acreditados (European Union, 2022; European Commission, 2025). En el corpus no se identificó un mecanismo brasileño general equivalente que obligue a grandes plataformas a ofrecer datos verificables a universidades y centros públicos de investigación.

La preservación y la reutilización presentan una brecha entre la cultura de documentación exigida a la ciencia pública y la evidencia social alojada

en plataformas. Los principios FAIR y la Recomendación de UNESCO promueven interoperabilidad, sostenibilidad y reutilización (UNESCO, 2021; Wilkinson *et al.*, 2016), pero los informes de plataforma no incluyen planes de preservación, licencias de reutilización de microdatos ni garantías de continuidad. La evidencia puede desaparecer cuando cambian productos, interfaces o políticas corporativas.

La contestabilidad es parcial. La LGPD reconoce derechos de los titulares y las plataformas describen canales de apelación o revisión. No obstante, los documentos examinados ofrecen información limitada sobre accesibilidad, tiempos de respuesta, revisión humana, resultados desagregados y recursos efectivos para decisiones de recomendación o visibilidad. La existencia formal de un canal no equivale a capacidad simétrica de contestación.

La justicia cognitiva es frágil porque los documentos contienen escasa evidencia de participación de universidades públicas brasileñas, organizaciones locales o poblaciones afectadas en la definición de categorías, métricas y mecanismos de acceso. La disponibilidad de reglas en portugués mejora la accesibilidad lingüística, pero no demuestra gobernanza local ni autoridad compartida sobre los datos. La valoración se sustenta, por tanto, se apoya en la ausencia documental de mecanismos de participación, adecuación contextual verificable y control institucional regional.

Tabla 3
Mapa de evidencias y valoración descriptiva para el caso brasileño

Dimensión	Fuentes principales	Evidencia localizada / ausencia relevante	Puntuación (0-4)	Valoración
Transparencia normativa	Marco Civil; LGPD; PL 2.338/2023	Reglas, derechos y obligaciones públicas; implementación y acceso técnico no garantizados.	3,0	Sólida
Transparencia clasificatoria	Meta; TikTok; X	Reglas y categorías generales publicadas; faltan taxonomías completas, umbrales y desagregación brasileña.	2,0	Parcial
Trazabilidad	Meta; TikTok; X	No se localizaron logs públicos, historiales de decisión o versionado suficiente para reconstruir casos.	0,5	Frágil
Documentación y metadatos	Meta; TikTok; X; FMTI 2025	Informes agregados; faltan diccionarios, procedencia y documentación reutilizable de conjuntos de datos.	1,0	Frágil
Acceso científico	DSA/Reg. 2025/2050 como contraste; documentos corporativos	Acceso condicionado por plataformas; no se identificó obligación brasileña general equivalente.	0,5	Frágil

Dimensión	Fuentes principales	Evidencia localizada / ausencia relevante	Puntuación (0-4)	Valoración
Preservación y reutilización	UNESCO; FAIR; documentos corporativos	Existen principios para ciencia; faltan licencias, planes de preservación y garantías para datos de plataforma.	1,0	Frágil
Contestabilidad	LGPD; Meta; TikTok; X	Derechos y canales de apelación; información parcial sobre revisión humana, plazos y resultados.	2,0	Parcial
Justicia cognitiva	CARE; literatura latinoamericana; documentos corporativos	Escasa evidencia de participación local, gobernanza compartida o control de instituciones públicas regionales.	0,5	Frágil

Nota. Los valores resumen presencia documental de cuatro indicadores por dimensión: 1 = evidencia directa; 0,5 = evidencia parcial; 0 = no verificable públicamente. No constituyen una medición validada ni un ranking de plataformas.

El mapa permite precisar la conclusión del caso: la aplicación documental no demuestra un nivel general de auditabilidad de todas las plataformas que operan en Brasil, sino que identifica una combinación específica de fortaleza normativa y debilidades documentales e infraestructurales. La principal brecha se sitúa entre la publicación de reglas y cifras agregadas, por un lado, y la capacidad de reconstruir, preservar, verificar y reutilizar evidencia, por otro.

Discusión

Los resultados sugieren ampliar el programa de la ciencia abierta. En su formulación más difundida, la ciencia abierta se ha centrado en abrir artículos y conjuntos de datos producidos por la comunidad científica (UNESCO, 2021; Wilkinson et al., 2016). Esa agenda es necesaria, pero insuficiente si deja fuera las infraestructuras privadas que producen una parte decisiva de los datos sobre la vida social. Una ciencia abierta crítica debe ocuparse no solo de abrir lo que la ciencia produce, sino de auditar las condiciones bajo las cuales se producen, clasifican y circulan los datos generados al habitar plataformas.

La matriz ofrece utilidad tanto para la comparación como para el diseño. Puede aplicarse a una plataforma, un sistema de inteligencia artificial, una política pública o una iniciativa de datos abiertos; permite localizar qué tipo de evidencia falta y distinguir publicación de información de capacidad efectiva de verificación. También evita que la transparencia normativa se tome como sustituto de trazabilidad, acceso o preservación. Su uso exige, sin embargo,

adaptar indicadores y ponderaciones al objeto: en un sistema público de alto riesgo puede ser prioritario aumentar el peso de contestabilidad y trazabilidad; en un repositorio científico, documentación, preservación e interoperabilidad pueden ser más determinantes.

Para el Sur Global, el problema no se reduce a una diferencia técnica de capacidades. Si la datificación concentra el poder de producir y examinar datos en corporaciones del Norte, la ciencia abierta regional permanece dependiente de infraestructuras que no controla. La soberanía informacional exige capacidades propias de documentación, preservación y reutilización, además de marcos jurídicos de acceso. La comparación con el DSA muestra que el acceso científico puede ser construido como derecho regulado y no solo como concesión corporativa, aunque la transferencia de ese modelo a Brasil requeriría adecuación institucional, presupuestaria y jurídica.

El vínculo con la justicia cognitiva es directo. La justicia de datos trata los datos como cuestión de derechos y equidad, no solo de eficiencia (Taylor, 2017). La justicia cognitiva añade quién puede definir categorías, interpretar resultados y participar de la gobernanza. Los principios CARE muestran que la apertura sin autoridad para controlar y sin beneficio colectivo puede reproducir extracción (Carroll *et al.*, 2020). Por ello, una infraestructura auditable debe incorporar participación local, adecuación lingüística y cultural, capacidad pública de investigación y mecanismos de contestación accesibles.

Las valoraciones también pueden admitir explicaciones alternativas. La ausencia de evidencia pública no prueba necesariamente que una práctica interna no exista; puede indicar que la empresa no la documenta externamente, que la información está fragmentada o que el formato de publicación no permite recuperarla. Del mismo modo, algunas restricciones al acceso pueden responder a privacidad, seguridad o secretos comerciales legítimos. La matriz no elimina estos conflictos: los hace explícitos y permite evaluar si existen salvaguardas proporcionadas, supervisión independiente y formas de acceso controlado.

La aplicación tiene límites. Se basa en documentos públicos, no prueba directamente sistemas ni compara decisiones observadas. El corpus corporativo se restringe a Meta, TikTok y X y a periodos distintos; los informes son dinámicos y pueden cambiar después de la fecha de consulta. La codificación fue realizada por un único investigador, sin acuerdo intercodificador ni validación de expertos. Las puntuaciones son heurísticas y sensibles a los indicadores y umbrales elegidos. Por estas razones, los resultados deben interpretarse como diagnóstico exploratorio y transferirse con cautela.

Una agenda de validación futura debería incluir panel Delphi con especialistas de comunicación, ciencia de la información, derecho, estudios críticos de datos y organizaciones de la sociedad civil; prueba piloto con casos y plataformas adicionales; doble codificación independiente; análisis de sensibilidad de pesos y umbrales; y triangulación con entrevistas, solicitudes de acceso, auditorías técnicas y observación de recursos de apelación. Ese programa permitiría convertir la matriz conceptual en un instrumento comparativo robusto sin perder su orientación crítica.

Conclusiones

El estudio preguntó qué condiciones de auditabilidad pueden identificarse documentalmente en infraestructuras de datificación que operan en Brasil y cuáles de sus dimensiones presentan mayores niveles de fragilidad. La matriz y el mapa de evidencias permiten responder de manera acotada: el corpus muestra una base relativamente sólida de transparencia normativa; niveles parciales de transparencia clasificatoria y contestabilidad; y debilidades en materia de trazabilidad, documentación, acceso científico, preservación y justicia cognitiva. Esta conclusión se limita a la evidencia pública localizada y no constituye una evaluación exhaustiva del funcionamiento interno de las plataformas.

La contribución del estudio se articula en tres planos. En el conceptual, la auditabilidad informacional integra disponibilidad, verificabilidad, preservación, contestación y justicia cognitiva. En el metodológico, la matriz operacionaliza ocho dimensiones mediante 32 indicadores y explicita un protocolo replicable de codificación documental. En el empírico-exploratorio, la aplicación al caso brasileño identifica la distancia entre densidad normativa y capacidad pública de reconstruir, preservar y reutilizar evidencia producida por infraestructuras privadas.

Las implicaciones apuntan a mecanismos públicos de acceso a datos de plataforma para investigación acreditada; requisitos de documentación, metadatos y versionado; infraestructuras nacionales o consorciadas de preservación; obligaciones de transparencia con desagregación contextual; y canales de contestación con revisión humana verificable. Estas medidas no eliminan conflictos con privacidad o seguridad, pero pueden someterlos a criterios públicos, proporcionados y auditables.

El carácter documental, el número limitado de plataformas y la ausencia de validación independiente restringen la generalización. La etapa siguiente consiste en validar la matriz con expertos y múltiples codificadores, aplicarla a sistemas concretos y combinar documentos con evidencia técnica y experiencia de usuarios e investigadores. Dentro de esos límites, la auditabilidad informacional se presenta como condición para una ciencia abierta crítica, para la justicia cognitiva y para una soberanía informacional capaz de transformar datos sociales en evidencia pública verificable.

Apoyos y soporte financiero de la investigación

Esta investigación contó con el apoyo del Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), mediante la Bolsa de Productividad en Investigación, proceso N.º 305515/2024-1.

Referencias bibliográficas

- Ananny, M. y Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 20(3), 973-989. <https://doi.org/10.1177/1461444816676645>
- Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. Polity Press.
- Birhane, A. (2020). Algorithmic colonization of Africa. *SCRIPTed*, 17(2), 389-409. <https://doi.org/10.2966/scrip.170220.389>
- Brasil. (2014). *Lei nº 12.965, de 23 de abril de 2014. Marco Civil da Internet. Diário Oficial da União*. <https://bit.ly/4fPceYb>
- Brasil. (2018). *Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Diário Oficial da União*. <https://bit.ly/4pVAg8H>
- Brasil. Câmara dos Deputados. (2026). *Projeto de Lei nº 2.338, de 2023: Ficha de tramitação*. <https://bit.ly/4fPDjKZ>
- Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Matechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J. y Hudson, M. (2020). The CARE Principles for Indigenous Data Governance. *Data Science Journal*, 19, 43. <https://doi.org/10.5334/dsj-2020-043>

- Couldry, N. y Mejias, U. A. (2019). Data colonialism: Rethinking Big Data's relation to the contemporary subject. *Television & New Media*, 20(4), 336-349. <https://doi.org/10.1177/1527476418796632>
- Couldry, N., & Mejias, U. A. (2021). The decolonial turn in data and technology research: What is at stake and where is it heading? *Information, Communication & Society*, 26(4), 786–802. <https://doi.org/10.1080/1369118X.2021.1986102>
- Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
- D'Ignazio, C. y Klein, L. F. (2020). *Data feminism*. MIT Press.
- European Commission. (2025). Commission Delegated Regulation (EU) 2025/2050 of 1 July 2025 supplementing Regulation (EU) 2022/2065 by laying down technical conditions and procedures for data access by vetted researchers. Official Journal of the European Union. <https://bit.ly/4wixADK>
- European Union. (2022). Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services (Digital Services Act). Official Journal of the European Union, L 277, 1-102. <https://bit.ly/4wFISDR>
- Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
- Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H. y Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, 64(12), 86-92. <https://doi.org/10.1145/3458723>
- Gendler, M. (2024). Datificación social e inteligencia artificial: ¿hacia un nuevo “salto de escala”? *Resonancias. Revista de Filosofía*, 17, 121-141. <https://doi.org/10.5354/0719-790X.2024.74503>
- Grohmann, R. (2025). Latin American critical data studies [Editorial]. *Big Data & Society*, 12(2). <https://doi.org/10.1177/20539517251330160>
- Kwet, M. (2019). Digital colonialism: US empire and the new imperialism in the Global South. *Race & Class*, 60(4), 3-26. <https://doi.org/10.1177/0306396818823172>
- Mejias, U. A. y Couldry, N. (2024). *Data grab: The new colonialism of Big Tech and how to fight back*. University of Chicago Press.
- Meta. (2025). Community Standards Enforcement Report: First quarter 2025. Meta Transparency Center. <https://bit.ly/4wFIZPN>
- Milan, S. y Treré, E. (2019). Big Data from the South(s): Beyond data universalism. *Television & New Media*, 20(4), 319-335. <https://doi.org/10.1177/1527476419837739>

- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S. y Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2). <https://doi.org/10.1177/2053951716679679>
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.
- Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.
- Quijano, A. (2000). Colonialidad del poder, eurocentrismo y América Latina. En E. Lander (ed.), *La colonialidad del saber: Eurocentrismo y ciencias sociales. Perspectivas latinoamericanas* (pp. 201-246). CLACSO.
- Ricaurte, P. (2019). Data epistemologies, the coloniality of power, and resistance. *Television & New Media*, 20(4), 350-365. <https://doi.org/10.1177/1527476419831640>
- Ricaurte, P., Gómez-Cruz, E. y Siles, I. (2024). Algorithmic governmentality in Latin America: Sociotechnical imaginaries, neocolonial soft power, and authoritarianism. *Big Data & Society*, 11(1). <https://doi.org/10.1177/20539517241229697>
- Robledo Maturana, R. (2023). Los condenados del algoritmo: Inteligencia artificial y cuerpos racializados latinoamericanos. *Pléyade*, 32, 130-159. <https://doi.org/10.4067/S0719-36962023000200130>
- Salami, A. O. (2024). Artificial intelligence, digital colonialism, and the implications for Africa's future development. *Data & Policy*, 6, e67. <https://doi.org/10.1017/dap.2024.75>
- Santos, B. de S. (2014). *Epistemologies of the South: Justice against epistemicide*. Paradigm Publishers.
- Taylor, L. (2017). What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society*, 4(2). <https://doi.org/10.1177/2053951717736335>
- TikTok. (2024). Community Guidelines Enforcement Report: July-September 2024. TikTok Transparency Center. <https://bit.ly/4wcAk5s>
- UNESCO. (2021). UNESCO Recommendation on Open Science. United Nations Educational, Scientific and Cultural Organization. <https://bit.ly/4fEVswe>
- Valente, J. C. L. y Grohmann, R. (2024). Critical data studies with Latin America: Theorizing beyond data colonialism. *Big Data & Society*, 11(1). <https://doi.org/10.1177/20539517241227875>
- van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, 12(2), 197-208. <https://doi.org/10.24908/ss.v12i2.4776>

van Dijck, J., Poell, T. y de Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford University Press.

Wan, A., Klyman, K., Kapoor, S., Maslej, N., Longpre, S., Xiong, B., Liang, P. y Bommasani, R. (2025). *The 2025 Foundation Model Transparency Index (arXiv:2512.10169)*. *arXiv*. <https://doi.org/10.48550/arXiv.2512.10169>

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., *et al.* (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>

X Corp. (2024). Global Transparency Report: H1 2024. X Transparency Center. <https://bit.ly/4wco1Wx>

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.

Declaración de Autoría - Taxonomía CRediT	
Autor	Contribuciones
Fabiano Couto Corrêa da Silva	Conceptualización, metodología, investigación, análisis formal, redacción del borrador original, redacción: revisión y edición, visualización, recursos, validación.

Declaración de Uso de inteligencia artificial
El autor DECLARA que, en la elaboración del artículo titulado Datificación social y auditabilidad informacional: aplicación exploratoria al caso brasileño, sí se utilizaron herramientas de inteligencia artificial, de manera complementaria y no sustitutiva del trabajo intelectual del autor. Se utilizó ChatGPT (OpenAI, versión disponible en agosto de 2026), exclusivamente para apoyar la revisión lingüística y estructural del manuscrito, la sistematización de las observaciones de los evaluadores y la comprobación de consistencia entre secciones, tablas y referencias. La herramienta no se utilizó para generar, alterar o inventar datos, resultados, citas o fuentes. El autor revisó integralmente cada propuesta, contrastó la información normativa y bibliográfica con las fuentes oficiales y citadas, verificó la coherencia de las modificaciones y asume plena responsabilidad por la exactitud, originalidad e integridad del contenido final. La herramienta de inteligencia artificial no figura ni debe figurar como autora.