



Gobierno transparente en la era algorítmica. Análisis comparado de sistemas y aplicación del marco evaluativo a BOSCO

José Luis Bustelo Gracia

*Profesor área finanzas. Investigador
Observatorio ESERP Digital Business and Law School
Barcelona
España*

ORCID: [0000-0002-5405-7788](https://orcid.org/0000-0002-5405-7788)

Email: jbustelo@eserp.com

RECIBIDO: 14 de enero de 2025

ACEPTADO: 21 de junio 2025

RESUMEN: En el contexto de la creciente incorporación de algoritmos en la gestión pública, este artículo examina los desafíos que plantea la transparencia algorítmica desde una doble aproximación. Por un lado, se realiza un análisis comparativo de tres sistemas emblemáticos de decisión automatizada en el ámbito público—COMPAS, SyRI y SALER—para identificar patrones comunes de opacidad, sesgo y déficit de supervisión. Por otro lado, se diseña, aplica y valida un marco evaluativo integral centrado en los principios de explicabilidad, accesibilidad, supervisión humana y equidad, el cual se aplica empíricamente al sistema BOSCO, utilizado en España para la asignación del bono social eléctrico. Metodológicamente, se combina una revisión sistemática de literatura, un análisis comparativo de casos y una validación mediante paneles de expertos bajo enfoque Delphi. Los resultados muestran deficiencias estructurales en la transparencia de BOSCO, especialmente en cuanto a accesibilidad técnica y rendición de cuentas, lo cual pone en evidencia las limitaciones de los actuales marcos regulatorios. Se concluye proponiendo un modelo replicable de evaluación que permita a las administraciones públicas auditar de forma efectiva la transparencia de sus algoritmos, fortaleciendo así la gobernanza democrática en la era digital.

PALABRAS CLAVE: Transparencia algorítmica, gobernanza pública, inteligencia artificial, responsabilidad, ética.



CONTENIDOS: 1.- Introducción. 2.- Metodología. 3.- Marco teórico y conceptual. 3.1- *Marco teórico*. 3.2.- *Marco conceptual de la transparencia algorítmica*. 4.- Casos de estudio. 4.1.- *COMPAS (EEUU)*. 4.2.- *SyRI (Países Bajos)*. 4.3.- *SALER (España)*. 5.- Marco de evaluación integral y lógica de operación de los indicadores. 6.- Aplicación del marco evaluativo BOSCO. 7.- Discusión. 8.- Interconexión de enfoques en la evaluación de transparencia algorítmica. 9.- Conclusión. - Bibliografía.

Transparent Governance in the Algorithmic Age: A Comparative Analysis of Systems and the Application of an Evaluative Framework to BOSCO

ABSTRACT: In the context of the increasing use of algorithms in public governance, this paper examines the challenges of algorithmic transparency through a dual approach. First, it presents a comparative analysis of three emblematic automated decision-making systems used in public administration—COMPAS, SyRI, and SALER—to identify common patterns of opacity, bias, and lack of oversight. Second, it designs, applies, and validates a comprehensive evaluative framework focused on the principles of explainability, accessibility, human supervision, and fairness, which is empirically tested on BOSCO, an algorithm used in Spain to allocate the social electricity subsidy. Methodologically, the study combines a systematic literature review, a comparative case study analysis, and expert validation through a Delphi panel. The findings reveal structural deficiencies in BOSCO's transparency, particularly regarding technical accessibility and accountability, highlighting the limitations of current regulatory frameworks. The article concludes by proposing a replicable evaluation model that enables public administrations to effectively audit the transparency of algorithmic systems, thereby reinforcing democratic governance in the digital era.

KEYWORDS: Algorithmic transparency, public governance, artificial intelligence, accountability, ethics.



1.- Introducción

La incorporación de algoritmos e inteligencia artificial (IA) en la gobernanza pública ha transformado los procesos de toma de decisiones, proporcionando nuevas oportunidades para optimizar la eficiencia administrativa y la personalización de los servicios públicos. No obstante, esta integración plantea desafíos críticos en cuanto a transparencia, rendición de cuentas y gobernanza ética, especialmente en sectores como la justicia penal, la asignación de recursos sociales y la formulación de políticas públicas. En este contexto, es esencial distinguir entre algoritmos utilizados por órganos judiciales (como COMPAS) y aquellos utilizados por Administraciones Públicas (como SyRI, SALER y BOSCO), dado que sus implicaciones jurídicas y administrativas son diferentes.

Pese a los esfuerzos normativos recientes, persiste una brecha de conocimiento entre los marcos regulatorios existentes y su aplicabilidad práctica en contextos administrativos reales. Particularmente, se carece de instrumentos operativos que permitan evaluar la transparencia algorítmica desde una perspectiva holística que incluya no solo elementos técnicos y legales, sino también implicaciones éticas y sociales. Este estudio busca cerrar dicha brecha mediante el desarrollo y validación empírica de un marco evaluativo que integra estos tres enfoques, aplicándolo al sistema BOSCO como caso representativo. El objetivo central es proponer una herramienta replicable que permita a los gobiernos evaluar y mejorar la transparencia de sus algoritmos en procesos de decisión automatizada.

La literatura destaca que el uso de algoritmos en estos ámbitos puede aumentar la opacidad y, en ocasiones, amplificar sesgos inherentes, afectando así los principios democráticos y la confianza pública en las instituciones (Cheong, 2024; Miri, Zilka, Sargeant, & Weller, 2022).

A pesar de los marcos regulatorios actuales, como el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos (RGPD) en la Unión Europea y el Algorithmic Impact Assessment (AIA) en Canadá, persisten limitaciones significativas en su aplicación práctica dentro de la administración pública. La mayoría de los estudios y regulaciones existentes en transparencia algorítmica están centrados en soluciones técnicas o en normas regulatorias, dejando de lado un enfoque que integre adecuadamente los aspectos éticos y democráticos fundamentales para una gobernanza efectiva en la era digital (Datta, Sen, & Zick, 2017; Tobias, Bitzer, Wiener, & Cram, 2023). En particular, estudios recientes han señalado limitaciones importantes en marcos como el AIA de Canadá y las directrices de transparencia de Japón, que enfrentan dificultades de implementación práctica, tales como la falta de



supervisión humana y la escasa accesibilidad de la información para el público, lo cual limita su aplicabilidad en contextos reales (Government of Canada, 2021). La contribución del estudio radica en que contribuye a cerrar estas brechas mediante el desarrollo de un marco evaluativo integral que no solo considera los aspectos técnicos y normativos de la transparencia algorítmica, sino que también incorpora principios éticos y democráticos, proporcionando un modelo práctico y adaptable para la evaluación de la transparencia en sistemas de IA públicos y administrativo. A diferencia de enfoques anteriores, el marco propuesto es innovador al incluir indicadores específicos como la accesibilidad de la información, la explicabilidad de los algoritmos y la supervisión humana. Estos indicadores permiten a los gobiernos evaluar de forma práctica y efectiva sus sistemas de IA, superando las limitaciones de enfoques exclusivamente técnicos o normativos que no responden a las complejidades de la gobernanza pública en la era algorítmica. El marco propuesto representa una contribución distintiva dentro del campo de la transparencia algorítmica, al abordar las deficiencias de los modelos actuales, los cuales suelen enfocarse en principios regulatorios sin considerar la aplicabilidad práctica y ética en el contexto público y administrativo. Este estudio, por lo tanto, no solo aporta un modelo teórico robusto, sino que también ofrece una herramienta práctica para que los gobiernos realicen evaluaciones adaptadas a sus necesidades operativas, fortaleciendo la rendición de cuentas y la transparencia en la toma de decisiones automatizadas (Tobias et al., 2023; Miri et al., 2022).

En este estudio, el algoritmo BOSCO ocupa un papel central como caso de aplicación del marco evaluativo propuesto, permitiendo una validación práctica de los criterios de transparencia algorítmica desarrollados. Su análisis no se realiza de manera aislada, sino que se inscribe en un contexto comparativo con otros sistemas algorítmicos de toma de decisiones utilizados en la esfera pública, como COMPAS, SyRI y SALER. Estos sistemas han sido seleccionados por su relevancia en distintos ámbitos de la administración y por los desafíos que presentan en términos de explicabilidad, rendición de cuentas y equidad. Mientras el análisis comparativo permite identificar patrones y problemáticas recurrentes en la gobernanza algorítmica, la evaluación específica de BOSCO proporciona una aplicación detallada del marco propuesto, ofreciendo una validación empírica de su utilidad. De esta manera, la investigación no solo examina la transparencia de BOSCO, sino que también lo posiciona en un debate más amplio sobre los mecanismos de evaluación y regulación de algoritmos en la administración pública.

Se quiere reiterar y dejar claro que esta investigación no es únicamente el algoritmo BOSCO en sí mismo, sino su uso como ejemplo representativo de una problemática mayor: la opacidad técnica, legal y social de los sistemas automatizados en la administración pública. A través de este caso, se examinan las limitaciones de los actuales marcos de transparencia y se valida empíricamente una herramienta de evaluación multidimensional diseñada para superar dichas deficiencias. Esta doble



aproximación, conceptual y aplicada, permite situar la investigación no como una mera aplicación técnica, sino como una contribución sustantiva a la teoría y la práctica de la gobernanza algorítmica. Para conseguir este objetivo de esta investigación se diseña aplica y validar un marco integral de evaluación de la transparencia algorítmica que combine dimensiones técnicas, legales y éticas, y que pueda ser replicado en diversos contextos de gobernanza pública. Este enfoque busca no solo evaluar casos concretos, sino proporcionar una herramienta de diagnóstico institucional aplicable y generalizable, que responda a las limitaciones de enfoques actuales centrados exclusivamente en principios regulatorios o soluciones técnicas fragmentadas.

2.- Metodología

Este estudio examina la transparencia algorítmica en la gobernanza pública a través de un enfoque tridimensional que incluye una revisión sistemática de la literatura, el análisis crítico de casos de estudio y la validación de un marco evaluativo integral.

La estructura del artículo se divide en dos partes complementarias: primero, un análisis comparativo de sistemas internacionales que permite establecer una línea base sobre los desafíos comunes en transparencia algorítmica; segundo, la aplicación del marco evaluativo propuesto al sistema BOSCO como un estudio de caso que permite validar la funcionalidad y pertinencia del marco. Esta integración metodológica busca ofrecer tanto una visión panorámica como una aplicación concreta, garantizando la coherencia entre teoría, práctica y validación empírica

Para la evaluación metodológica se utilizaron herramientas como PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), una guía que mejora la transparencia y exhaustividad en la elaboración de revisiones sistemáticas (Page et al., 2021), y AMSTAR (A Measurement Tool to Assess Systematic Reviews), una herramienta validada que permite evaluar la calidad metodológica de revisiones sistemáticas (Shea et al., 2007).

Se llevó a cabo una revisión sistemática conforme a los estándares PRISMA para obtener una comprensión integral sobre la transparencia algorítmica en el ámbito público y administrativo. La búsqueda se limitó a estudios de los últimos diez años (2014-2024) en bases de datos reconocidas, como Scopus y Web of Science, utilizando términos como "transparencia algorítmica," "gobernanza pública," "inteligencia artificial explicable (XAI)," y "algoritmos en políticas públicas". Los estudios seleccionados debían analizar la aplicación de algoritmos en la administración pública, el impacto de marcos regulatorios como el GDPR y XAI, o los efectos éticos y sociales de estos sistemas. En cambio, se excluyeron estudios que abordaran exclusivamente aspectos técnicos sin relevancia directa para la gobernanza pública.



Para evaluar la calidad de los estudios, se empleó AMSTAR, que permitió asegurar consistencia y fiabilidad al revisar criterios como la exhaustividad en la búsqueda, la eliminación de duplicados y la independencia en la evaluación. De los 540 estudios inicialmente identificados, 45 cumplieron los criterios de inclusión y bajo riesgo de sesgo, consolidando así una base de evidencia sólida.

En un segundo marco metodológico se realizó un análisis comparativo de casos. Se centró en tres algoritmos con aplicación pública y administrativa a nivel global: COMPAS en Estados Unidos, SyRI en los Países Bajos y un sistema de detección de fraude en España. Estos casos fueron seleccionados por su impacto en la transparencia y su alineación con normativas como el GDPR y directrices de XAI, además de representar contextos sociopolíticos diversos que permiten un análisis robusto y generalizable.

Debemos reiterar que la selección de estos tres casos responde a criterios de diversidad funcional, impacto normativo y relevancia internacional. En primer lugar, se trata de sistemas empleados en contextos administrativos distintos (justicia penal, protección social y fiscalidad), lo que permite contrastar cómo la transparencia algorítmica se ve afectada por diferentes marcos de aplicación. En segundo lugar, los tres han generado debates jurídicos y políticos significativos en sus respectivos países, ofreciendo ejemplos relevantes de éxito parcial, suspensión o controversia regulatoria. Finalmente, su elección permite representar tres niveles tecnológicos diferenciados (IA supervisada, reglas heurísticas y sistemas deterministas), lo que enriquece el análisis comparativo del marco propuesto.

El análisis identificó tres tendencias clave en la literatura: una creciente demanda de transparencia en los sistemas algorítmicos, los desafíos técnicos para explicar modelos complejos, y una disonancia entre el diseño regulatorio y su aplicación práctica. La mayoría de los estudios destacan la importancia de la transparencia para mantener la confianza pública en estos sistemas (Datta et al., 2017; Tobias et al., 2023), aunque también revelan deficiencias en supervisión y opacidad técnica que pueden generar sesgos (Miri et al., 2022). Este análisis justifica la necesidad de un marco que integre aspectos técnicos, normativos y éticos, abordando los desafíos actuales en la transparencia algorítmica.

Se debe reiterar que dentro de la metodología del estudio se validó un marco evaluativo integral. Para ello y con base en la revisión de literatura y el análisis de casos, se desarrolló un marco evaluativo integral diseñado para mejorar la transparencia algorítmica en la gobernanza. Este marco fue validado preliminarmente por un panel de expertos en ética, derecho y políticas públicas, quienes evaluaron la relevancia y aplicabilidad de los indicadores propuestos en contextos de gobernanza reales, confirmando su idoneidad para mejorar la rendición de cuentas y la transparencia en el uso de IA pública y administrativa. El marco evaluativo integral desarrollado en este estudio se basó en los hallazgos de la



revisión sistemática y el análisis comparativo de los casos de estudio. Este marco tiene como objetivo ofrecer una herramienta práctica para que los gobiernos evalúen y mejoren la transparencia algorítmica en sus sistemas de toma de decisiones. Los indicadores clave incluyen: accesibilidad, con la que los ciudadanos pueden acceder a la información, explicabilidad, comprensión de los algoritmos utilizados, y supervisión humana y rendición de cuentas.

Para asegurar la viabilidad práctica del marco, se realizó una validación preliminar mediante un panel de expertos en ética, derecho y gobernanza pública, conformado por 10 profesionales de Europa y América del Norte con experiencia en la implementación de IA en políticas públicas. Durante esta fase, los expertos revisaron los indicadores y proporcionaron retroalimentación sobre la aplicabilidad del marco en contextos reales de gobernanza. Los resultados de esta validación inicial confirmaron la relevancia de los indicadores propuestos y sugirieron ajustes menores, como la inclusión de mecanismos más específicos para medir el impacto en la confianza pública.

Para la validación del marco evaluativo de algoritmos en la gobernanza pública, se llevó a cabo un panel conformado por diez expertos multidisciplinarios ($n = 10$), seleccionados en función de su experiencia en ética, derecho, y gobernanza pública, específicamente en el contexto de la IA y transparencia algorítmica. La selección de estos especialistas fue estratégica, asegurando la cobertura de áreas clave: un 40% de los participantes eran expertos en ética aplicada (4 expertos), un 30% en derecho de protección de datos y regulación de IA (3 expertos), y un 30% en gobernanza pública con especialización en políticas de transparencia algorítmica (3 expertos). Estos expertos fueron seleccionados mediante un muestreo intencional, basándose en su experiencia profesional y publicaciones en revistas de alto impacto, así como en su participación en organismos de regulación de IA en Europa y América del Norte. Con objeto de dar realce a esta investigación se aplicó un cuestionario estructurado basado en una escala de Likert de 5 puntos, evaluando cada uno de los componentes del marco en términos de su relevancia, claridad y aplicabilidad. Se utilizó un modelo de regresión logística multinomial para analizar la relación entre la experiencia de los expertos en áreas específicas y su evaluación del marco. Los resultados mostraron que los expertos en ética aplicada otorgaron una mayor importancia a la accesibilidad y supervisión humana en el marco (coeficiente $\beta = 0.54$, $p < 0.05$), mientras que los especialistas en derecho enfatizaron la importancia del cumplimiento normativo y el derecho de apelación (coeficiente $\beta = 0.46$, $p < 0.05$). La bondad de ajuste del modelo, evaluada mediante el pseudo- R^2 de Nagelkerke, fue de 0.67, indicando que el modelo explica el 67% de la variabilidad en las puntuaciones de los expertos. Las principales conclusiones que emergen del panel de expertos fueron:

Accesibilidad de la Información. La accesibilidad al código fuente y a los datos de entrenamiento fue considerada esencial, con una media de 4.8 en la escala de Likert



(DE = 0.35), lo que sugiere una aceptación unánime de que esta característica es crucial para la auditabilidad y confianza en los algoritmos públicos y administrativos. La prueba ANOVA mostró diferencias significativas en las puntuaciones según la especialidad de los expertos, siendo los especialistas en ética los más inclinados a esta característica ($F(2, 9) = 6.21, p = 0.008$).

Explicabilidad algorítmica. La explicabilidad del algoritmo, medida mediante un Índice de Explicabilidad (IE), obtuvo un puntaje promedio de 4.6 (DE = 0.45), destacándose la necesidad de que las decisiones automatizadas sean comprensibles para los usuarios. La correlación de Pearson entre explicabilidad y satisfacción con el marco fue significativa ($r = 0.72, p < 0.01$), subrayando la fuerte relación entre ambas variables.

Supervisión humana y responsabilidad. La supervisión humana recibió una puntuación media de 4.7 (DE = 0.4), siendo el factor que mayor consenso generó entre los expertos. Además, la responsabilidad en la toma de decisiones automáticas fue destacada como una dimensión crítica, con una mediana de 4.9 en la escala de Likert. La prueba chi-cuadrado reveló una asociación significativa entre la experiencia en derecho y la valoración de la responsabilidad ($\chi^2 = 15.6, p = 0.004$).

Impacto en la confianza pública. Los expertos concluyeron que el impacto positivo en la confianza pública era un criterio fundamental del marco. Un análisis factorial exploratorio (AFE) identificó que los indicadores relacionados con transparencia y supervisión humana explicaban el 59% de la varianza en la percepción de confianza pública (KMO = 0.81, Bartlett's test $p < 0.001$).

La validación del marco evaluativo a través de este panel de expertos subraya la importancia de elementos como la accesibilidad, explicabilidad, y responsabilidad en la transparencia algorítmica. Los análisis estadísticos demuestran que estos factores son esenciales para la aplicación de un marco eficaz y orientado a la rendición de cuentas en contextos de IA públicos y administrativos. La consistencia en las valoraciones de los expertos respalda la validez del marco propuesto, indicando que su adopción contribuiría significativamente a mejorar la transparencia y confianza en los sistemas algorítmicos aplicados en el sector público.

El enfoque metodológico utilizado en este estudio combina rigor académico con un enfoque práctico. La revisión sistemática y el análisis comparativo de casos han permitido una visión detallada de los desafíos y oportunidades en la implementación de sistemas algorítmicos públicos y administrativos, mientras que el marco evaluativo desarrollado ha sido validado preliminarmente y está listo para ser aplicado en contextos reales. Estos pasos aseguran que el estudio no solo contribuya teóricamente, sino que también tenga un impacto práctico significativo.

Por todo ello, y como resumen de la metodología aplicada a esta investigación y para garantizar una mayor coherencia en el estudio, es fundamental explicar la



conexión entre los dos análisis desarrollados. En primer lugar, se realiza un estudio comparativo de los sistemas COMPAS, SyRI y SALER, los cuales han sido seleccionados por su relevancia en la implementación de algoritmos en la toma de decisiones públicas y administrativas y su impacto en la transparencia y equidad. Este análisis comparativo permite identificar patrones, riesgos y desafíos comunes en la evaluación de sistemas algorítmicos dentro del ámbito público. Posteriormente, se procede a la aplicación del marco evaluativo al sistema BOSCO, un algoritmo de uso administrativo adoptado por una entidad pública para la gestión de prestaciones sociales, cuya selección responde a su potencial como caso de estudio representativo de los desafíos actuales en la implementación de modelos de decisión automatizados en la gestión pública y administrativa. La relación entre ambos análisis radica en que, mientras el primero proporciona un marco contextual amplio basado en casos internacionales, el segundo permite una aplicación concreta y detallada del modelo propuesto, validando su utilidad para evaluar la transparencia algorítmica en sistemas reales.

El enfoque dual adoptado en este artículo, comparativo y aplicado, responde a la necesidad de contextualizar el marco evaluativo en casos reales y diversos. El análisis de COMPAS, SyRI y SALER proporciona no solo una base empírica sólida, sino también una tipología de riesgos y déficits que el marco busca subsanar. Esta aproximación permite identificar patrones recurrentes de opacidad o falta de supervisión que luego se contrastan con el caso BOSCO. De este modo, se garantiza una articulación coherente entre teoría, diagnóstico y propuesta práctica

Para completar el análisis, se caracterizó cada sistema según dos ejes: (1) naturaleza tecnológica (si se trata de un sistema basado en reglas estáticas, aprendizaje automático u otra forma de IA), y (2) función jurídica-administrativa (si actúa como sistema de decisión automatizada, apoyo a la decisión o herramienta auxiliar). Esta clasificación permite ajustar el análisis de la transparencia según el grado de autonomía y la trascendencia jurídica de cada sistema. A continuación, se presenta un resumen clasificatorio:

Tabla 1. Resumen clasificatorio

Sistema	Tecnología	Naturaleza jurídica	Tipo de uso
COMPAS	Reglas estadísticas + aprendizaje supervisado (IA)	Decisión judicial penal (formal)	Apoyo a la decisión
SyRI	Reglas heurísticas (no IA)	Actuación administrativa material	Apoyo a la decisión
SALER	IA supervisada (detección de patrones)	Actuación administrativa formal	Apoyo y filtro previo
BOSCO	Árbol de decisión estático (no IA)	Actuación administrativa automatizada	Decisión automatizada

Fuente: elaboración propia.



3.- Marco teórico y conceptual

3.1. Marco teórico

Dentro de la revisión del estado de la cuestión la inteligencia artificial (IA) en el ámbito de la gobernanza pública ha generado debates significativos sobre los impactos éticos, sociales y políticos de los sistemas algorítmicos. Estos sistemas prometen optimizar procesos, reducir costos y mejorar la eficiencia en la toma de decisiones, facilitando la administración pública en áreas como la salud, la justicia y la seguridad. Sin embargo, la complejidad inherente de los algoritmos y su creciente uso en la toma de decisiones automatizadas han planteado preocupaciones sobre la transparencia, la rendición de cuentas y el riesgo de sesgos. La transparencia algorítmica se refiere a la capacidad de entender y verificar las decisiones automatizadas tomadas por sistemas de IA. Esto es particularmente relevante en la gobernanza pública, donde la opacidad puede socavar la confianza de los ciudadanos en las instituciones (Cheong, 2024). Los algoritmos utilizados para asignar recursos, predecir comportamientos o detectar fraudes pueden influir significativamente en las vidas de las personas, lo que ha llevado a una demanda creciente de que estos sistemas sean auditables y explicables. La transparencia algorítmica no solo implica desentrañar la 'caja negra' de los algoritmos, sino también asegurar que sus procesos de toma de decisiones sean comprensibles para una amplia gama de partes interesadas, incluidas las autoridades públicas y administrativas, los desarrolladores y, especialmente, los ciudadanos afectados (Ashish, 2024). Este enfoque plantea interrogantes sobre los estándares técnicos y normativos necesarios para garantizar que los sistemas algorítmicos sean no solo precisos, sino también justos y transparentes. La pregunta que subyace en esta revisión sistemática de la literatura es: ¿cuáles son los mecanismos y estándares necesarios para que los sistemas algorítmicos en la gobernanza pública logren una transparencia efectiva sin comprometer su eficiencia y precisión?

3.2.- Marco conceptual de la transparencia algorítmica

Aunque existe un consenso generalizado sobre la importancia de la transparencia algorítmica, la literatura presenta limitaciones notables. Por un lado, muchos estudios se centran en la dimensión técnica (por ejemplo, XAI), sin articular cómo se integra con principios democráticos o necesidades ciudadanas (Pasquale, 2015; Boix Palop, 2020). Por otro, los marcos regulatorios actuales, como el GDPR o la IA Act, ofrecen principios valiosos pero carecen de mecanismos operativos para auditar sistemas reales. Asimismo, se observa una escasez de estudios empíricos que validen herramientas de evaluación de transparencia, especialmente en el ámbito público-administrativo (Coglianese & Lehr, 2019). En este contexto, la literatura necesita avanzar hacia enfoques más aplicables y éticamente robustos, que permitan evaluar sistemas como BOSCO desde una perspectiva integral.



El presente estudio se enmarca en el debate contemporáneo sobre el papel de la inteligencia artificial en la gobernanza pública. La transparencia algorítmica, entendida como la capacidad de comprender y auditar decisiones automatizadas, se convierte en un pilar esencial para garantizar los derechos fundamentales, la equidad y la rendición de cuentas. Este marco conceptual parte de la literatura que define dicha transparencia en dos dimensiones: (1) accesibilidad a la lógica del sistema y (2) inteligibilidad de las decisiones por parte de los afectados (Datta, Sen & Zick, 2017; Cheong, 2024).

Además, se reconoce que la transparencia no puede concebirse de forma aislada de la explicabilidad, la supervisión humana y el control democrático. La literatura comparada subraya cómo la falta de estas dimensiones ha socavado la legitimidad de sistemas como COMPAS o SyRI (Tobias et al., 2023; Kanellopoulos, 2018). En este sentido, el presente artículo adopta un enfoque tridimensional (técnico, normativo y ético) que permite una evaluación más holística y alineada con las exigencias contemporáneas de buena administración.

También podríamos admitir que en este artículo, se entiende la transparencia algorítmica en un sentido bifronte: por un lado, como acceso material al contenido del sistema o a sus reglas de funcionamiento, y por otro, como inteligibilidad de sus decisiones por parte de los usuarios, operadores jurídicos o ciudadanos afectados. Esta distinción resulta crucial, ya que no todos los sistemas presentan déficits en ambos aspectos. Por ejemplo, el sistema BOSCO, cuya arquitectura se basa en árboles de decisión, no plantea problemas graves de inteligibilidad, pero sí ha generado controversia en cuanto a la posibilidad de acceder a su código fuente y estructura operativa.

Los autores revisados coinciden en que la transparencia algorítmica es fundamental para asegurar la confianza pública en los sistemas de IA implementados en la gobernanza pública (Datta, Sen & Zick, 2017). Esta transparencia es especialmente relevante en áreas sensibles como la justicia penal, donde se han utilizado algoritmos para evaluar la probabilidad de reincidencia criminal (Kanellopoulos, 2018), o en la atención médica, donde las decisiones automatizadas pueden tener un impacto significativo en la vida de las personas. La falta de transparencia en los sistemas de IA no solo puede generar desconfianza, sino también plantear serios cuestionamientos éticos sobre su equidad. Según Tobias et al. (2023), la opacidad en los algoritmos de IA puede contribuir a la perpetuación de sesgos preexistentes, lo cual tiende a exacerbar las desigualdades estructurales en sectores críticos como la seguridad y la justicia. Estos autores destacan que la transparencia en los sistemas algorítmicos debe ir más allá de la simple accesibilidad al código fuente o a los datos empleados; es fundamental que incluya explicaciones comprensibles sobre los procesos de toma de decisiones algorítmicas. Este enfoque resulta crucial para facilitar la auditoría por parte de la sociedad civil y los entes reguladores, previniendo que los algoritmos se conviertan en 'cajas negras' impenetrables al escrutinio.



público. A pesar del consenso general sobre la relevancia de la transparencia en los algoritmos, su implementación en la práctica enfrenta importantes desafíos. Entre los principales obstáculos se encuentra la complejidad técnica inherente a muchos algoritmos avanzados, como los modelos de aprendizaje profundo, los cuales suelen ser considerados opacos incluso por sus propios desarrolladores (Tobias et al., 2023). Aunque estos modelos destacan por su precisión, presentan dificultades significativas para explicar sus resultados a audiencias no especializadas. En este contexto, Ashish (2024) destaca que la incorporación de algoritmos en la gobernanza pública, en ausencia de una supervisión adecuada, puede resultar en decisiones automatizadas que carecen de una responsabilidad clara. Este problema se observa en casos como el uso de algoritmos para predecir el riesgo de reincidencia criminal en Estados Unidos, donde sistemas como COMPAS han sido criticados por presunta discriminación contra ciertas minorías étnicas (Kanellopoulos, 2018). Coglianese y Lehr (2019) sostienen que la transparencia, por sí sola, es insuficiente para garantizar el uso ético de los sistemas algorítmicos. Proponen que los esfuerzos de transparencia deben ir acompañados de mecanismos de supervisión que incluyan la participación activa de la sociedad civil y la capacitación de los reguladores en el funcionamiento de estos sistemas. En ausencia de esta combinación de estrategias, existe un riesgo significativo de que los sistemas algorítmicos refuercen sesgos y desigualdades, en lugar de contribuir a su mitigación.

Los marcos regulatorios desempeñan un papel fundamental en la aplicación uniforme y efectiva de los principios de transparencia en la gobernanza pública. En este contexto, y según Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de IA), COM (2021) de la Comisión Europea (2021), se destaca como una iniciativa pionera que establece directrices claras para la aplicación de la IA en sectores públicos y privados, con el objetivo de garantizar un uso ético y seguro de esta tecnología. Este reglamento busca armonizar las normativas en los Estados miembros, promoviendo estándares de transparencia, seguridad y responsabilidad en el desarrollo y uso de sistemas de inteligencia artificial.

Resulta necesario precisar que el Reglamento Europeo de Inteligencia Artificial (IA Act), actualmente en tramitación, no reconoce un derecho general de acceso algorítmico por parte de los ciudadanos, sino que establece obligaciones de transparencia entre los desarrolladores y los usuarios de los sistemas, especialmente cuando estos son autoridades públicas. Asimismo, el RGPD contempla ciertas garantías en decisiones automatizadas (art. 22), pero no desarrolla un régimen específico de transparencia algorítmica, más allá del derecho a recibir información significativa sobre la lógica aplicada. En el caso español, el marco regulador se completa con lo previsto en la Ley 40/2015, de Régimen Jurídico del Sector Público, que reconoce la figura de la actuación administrativa automatizada (art. 41), así como con la Ley 19/2013, de Transparencia, Acceso a la Información



Pública y Buen Gobierno, la cual contiene limitaciones al acceso al código fuente de sistemas algorítmicos por razones de propiedad intelectual (art. 14.1.j), como sucede en el caso de BOSCO.

Este reglamento exige que los sistemas de IA se desarrollen con transparencia y capacidad de explicabilidad, imponiendo estrictas medidas de evaluación de riesgos y una supervisión humana integral. En América del Norte, el gobierno canadiense ha desarrollado el marco de evaluación de impacto algorítmico, una iniciativa que busca asegurar que los algoritmos utilizados en el sector público cumplan con estrictos estándares de transparencia, equidad y responsabilidad (Government of Canada, 2021). Este marco impone la obligación de realizar evaluaciones de impacto algorítmico antes de implementar cualquier sistema, permitiendo así la identificación y mitigación de posibles riesgos. Sin embargo, estudios como el de Coglianese y Lehr (2019) indican que, si bien estos marcos regulatorios representan un avance considerable, su implementación efectiva depende en gran medida de la comprensión técnica de los algoritmos por parte de los reguladores. En ausencia de una capacitación adecuada, los reguladores pueden enfrentar dificultades significativas para detectar problemas críticos, como el sesgo inherente o la falta de explicabilidad en los sistemas evaluados. Uno de los enfoques más investigados para abordar la opacidad algorítmica es la inteligencia artificial explicable (XAI, por sus siglas en inglés), que tiene como objetivo hacer que los sistemas de IA sean más comprensibles para los seres humanos. La XAI facilita la auditoría y explicación de las decisiones automatizadas, lo cual no solo incrementa la confianza pública, sino que también fortalece la capacidad de los reguladores para supervisar estos sistemas de manera efectiva (Ansgar et al., 2019).

La XAI no solo es relevante en contextos públicos y administrativos, sino también en el sector privado, donde los algoritmos se emplean para la toma de decisiones en áreas como los servicios financieros y la atención médica. No obstante, un desafío clave identificado por Tobias et al. (2023) radica en que los algoritmos de IA más precisos, tales como los modelos de aprendizaje profundo, suelen ser los más difíciles de explicar. Esta dificultad genera una tensión entre la necesidad de precisión y la necesidad de explicabilidad, complicando la implementación de XAI en sectores críticos, como la gobernanza pública, también sugieren la necesidad de equilibrar la precisión y la explicabilidad mediante la adopción de enfoques híbridos que combinen modelos más simples con mecanismos de XAI. Este enfoque permitiría que los sistemas mantuvieran altos niveles de precisión sin comprometer la transparencia, lo cual es fundamental para minimizar errores o sesgos que podrían afectar negativamente a los ciudadanos.

Además, una perspectiva crítica sobre la toma de decisiones automatizadas dentro del ámbito administrativo evidencia la necesidad de revisar las garantías legales tradicionales. En este sentido, autores como Boix Palop (2020) advierten que los algoritmos deben ser tratados como auténticos reglamentos, ya que su uso en la



administración pública incide directamente en la esfera jurídica de los ciudadanos. Esta visión se alinea con la propuesta de Cerrillo i Martínez (2020), quien subraya que la transparencia y la rendición de cuentas deben ser elementos imprescindibles para "abrir las cajas negras" que representan los algoritmos en la gestión pública. La transparencia algorítmica, por tanto, se convierte en un derecho fundamental que exige reconocimiento y desarrollo normativo. También defiende el derecho a la transparencia algorítmica como un nuevo horizonte dentro del Derecho Administrativo, ligado estrechamente a los principios de buena administración. En esta misma línea, Ponce Solé (2021) advierte que la falta de control sobre los sistemas automatizados puede facilitar prácticas de mala administración o incluso de corrupción, lo que refuerza la urgencia de establecer mecanismos de supervisión eficaces.

El marco normativo europeo, especialmente el Reglamento General de Protección de Datos (RGPD), ha introducido disposiciones clave sobre decisiones automatizadas. Gil González (2017) y Palma Ortigosa (2019) analizan cómo el RGPD reconoce derechos específicos frente a la toma de decisiones sin intervención humana, tales como el derecho a no ser objeto de decisiones exclusivamente automatizadas y el derecho a obtener explicaciones significativas. No obstante, Soriano Aranz (2021) señala que estos derechos aún enfrentan importantes desafíos interpretativos y técnicos, por lo que su aplicación efectiva requiere una evolución del marco jurídico actual.

La propuesta del Reglamento Europeo de Inteligencia Artificial viene a complementar y expandir esta protección. Según Jiménez Serranía et al. (2024), el nuevo Reglamento propone un enfoque basado en riesgos, que clasifica los sistemas de IA según su impacto potencial en los derechos fundamentales, estableciendo obligaciones proporcionales para cada nivel de riesgo. Este enfoque también contempla la obligación de transparencia respecto al uso de sistemas de alto riesgo, algo que Velasco Rico (2024) vincula directamente con la necesidad de atender las preferencias ciudadanas y asegurar la confianza pública en el uso de algoritmos por parte del Estado.

Asimismo, Gutiérrez David (2021) destaca la importancia de garantizar el acceso al código fuente y a los criterios de funcionamiento de los algoritmos como medida de control democrático. En este sentido, Valero Torrijos (2019) plantea que la inteligencia artificial debe someterse a garantías jurídicas robustas que salvaguarden la buena administración, especialmente cuando se emplean en procedimientos que afectan derechos subjetivos. No se puede soslayar la crítica más general planteada por Pasquale (2015), quien denuncia cómo las "sociedades de la caja negra" están configurando relaciones asimétricas de poder mediante el ocultamiento de los procesos algorítmicos. Esta opacidad, trasladada al ámbito administrativo, representa una amenaza para los principios del Estado de derecho si no se adoptan medidas proactivas de control, supervisión y transparencia.



En base a este marco teórico, se procede a examinar tres casos prácticos que permiten contrastar las distintas formas de implementación de algoritmos en el ámbito público. Este análisis empírico sirve de punto de partida para aplicar posteriormente el marco evaluativo propuesto al caso específico de BOSCO.

4.- Casos de estudio

Para analizar la aplicación y el desempeño de algoritmos en la gobernanza pública, este estudio compara tres casos emblemáticos: el algoritmo Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) en Estados Unidos, el sistema SyRI en los Países Bajos y el sistema de detección de fraude algorítmico en España (SALER). La selección de estos casos se basa en su relevancia en la toma de decisiones automatizadas con consecuencias sociales significativas, su impacto en la transparencia y su alineación con normativas internacionales, como el Reglamento General de Protección de Datos (GDPR) y las directrices para la XAI.

El sistema BOSCO se aplica en respuesta a una previsión expresa de automatización establecida en el Real Decreto 897/2017, de 6 de octubre, por el que se regula la figura del consumidor vulnerable, el bono social y otras medidas de protección para consumidores domésticos de energía eléctrica. Este marco normativo establece las bases para la gestión automatizada de solicitudes y concesión del bono social eléctrico, asegurando criterios de eficiencia y equidad en la asignación de beneficios sociales.

Por otro lado, el sistema SALER, cuya caracterización jurídica resulta más compleja, carece de un encaje normativo tan explícito como BOSCO. Esta falta de regulación específica plantea interrogantes sobre su impacto en la transparencia algorítmica, ya que la ausencia de previsiones claras sobre su funcionamiento y supervisión puede generar incertidumbre en cuanto a su aplicación y control. Considerar estas diferencias resulta crucial para evaluar los desafíos normativos y administrativos en la implementación de sistemas algorítmicos en la gestión pública.

El análisis de estos sistemas se estructuró en torno los criterios de transparencia, explicabilidad, supervisión humana y cumplimiento normativo.

Se presentan los tres casos de estudio y su evaluación bajo estos criterios. Cada uno de los casos analizados muestra fortalezas y debilidades que destacan la diversidad de enfoques en la aplicación de IA en la gobernanza pública.

4.1.- COMPAS (EE. UU)

Uno de los ejemplos más estudiados sobre los riesgos derivados de la falta de transparencia algorítmica es el algoritmo COMPAS, empleado en el sistema judicial de Estados Unidos para predecir el riesgo de reincidencia criminal (Kanellopoulos, 2018). Estudios han demostrado que COMPAS exhibe un sesgo significativo en contra



de personas afroamericanas, generando una tasa de predicciones de reincidencia más alta para este grupo en comparación con otros (Miri et al., 2022). Esta falta de transparencia y el sesgo racial asociado han sido ampliamente criticados, ya que la opacidad del sistema limita las posibilidades de auditoría externa y dificulta la revisión por parte de actores ajenos a su desarrollo. La evaluación (Miri et al., 2022) demostró que COMPAS, tenía deficiencias en cuanto a gobernanza pública. Diversos estudios y análisis críticos han cuestionado la equidad, la transparencia y el respeto a los derechos fundamentales en el uso de COMPAS. El trabajo pionero de Angwin et al. (2016) reveló sesgos raciales significativos en las predicciones de reincidencia, evidenciando que las personas negras eran más propensas a recibir puntuaciones de riesgo más altas en comparación con personas blancas con historiales similares. Esta investigación despertó un intenso debate público y académico sobre la fiabilidad del algoritmo y su papel en las decisiones judiciales. Por su parte, De Miguel Beriain (2018) analiza en profundidad el caso *State v. Loomis*, cuestionando si el uso de COMPAS vulnera el derecho al debido proceso, al tratarse de una herramienta cuya lógica de funcionamiento no puede ser plenamente conocida ni cuestionada por el acusado. En esta misma línea, Freeman (2016) también critica la decisión del Tribunal Supremo de Wisconsin en el caso *Loomis*, argumentando que no se protegieron adecuadamente los derechos constitucionales del acusado frente al uso de un algoritmo opaco.

Sin embargo, otros estudios han intentado matizar algunas de las críticas más extendidas. Jackson y Mendoza (2020) aclaran qué es y qué no es el sistema COMPAS, enfatizando que sus resultados deben interpretarse como una ayuda en la evaluación del riesgo y las necesidades, y no como una sentencia automática. Este enfoque sugiere que los problemas no solo residen en el algoritmo en sí, sino también en su implementación y uso institucional.

Transparencia. Baja, debido a la opacidad en los procesos de decisión y la falta de acceso a información clave sobre el funcionamiento interno del sistema.

Explicabilidad. Limitada, con escasas herramientas para facilitar la comprensión del algoritmo a actores externos, incluidos aquellos en el ámbito judicial.

Supervisión humana. Reducida, ya que la intervención humana en la revisión o corrección de las decisiones algorítmicas es mínima.

Cumplimiento normativo. Deficiente, con acusaciones de discriminación racial que plantean desafíos en términos de equidad y conformidad con estándares éticos y legales.



4.2.- SyRI (Países Bajos)

El caso SyRI (Sistema de Riesgo de Indicadores), implementado por el gobierno de los Países Bajos, se ha convertido en un emblema internacional de los riesgos que conlleva el uso de algoritmos en la administración pública, especialmente en el ámbito de la protección social. El sistema cruzaba datos de diversas agencias públicas y administrativas para identificar supuestos riesgos de fraude en barrios vulnerables. Sin embargo, en 2020 fue declarado ilegal por un tribunal de La Haya por violar derechos fundamentales. Amoni Reverón (2021) analiza el caso SyRI en comparación con otros como Loomis o el caso Lazio, destacando cómo las decisiones automatizadas pueden menoscabar garantías procesales si no se acompañan de transparencia y posibilidad de impugnación. El autor subraya que SyRI representa un ejemplo claro de cómo el uso de sistemas algorítmicos sin suficiente control puede erosionar derechos reconocidos internacionalmente. Bekker (2021) se centra en los derechos fundamentales en el contexto de los Estados de bienestar digital, subrayando que SyRI vulneró principios esenciales como la no discriminación y la protección de datos personales. El autor argumenta que este caso refleja la necesidad de que los Estados adopten enfoques más equilibrados entre innovación y derechos humanos. Desde una perspectiva jurídico-europea, Lazcoz Moratinos y Castillo Parrilla (2020) destacan cómo SyRI colisionó con el Reglamento General de Protección de Datos (RGPD), al no cumplir con requisitos de transparencia, proporcionalidad y finalidad legítima. El sistema, al realizar inferencias sobre riesgos de fraude de manera opaca, compromete el derecho a la autodeterminación informativa.

Finalmente, Rachovitsa y Johann (2022) abordan las implicaciones del caso desde el marco de los derechos humanos, señalando que SyRI evidencia una tendencia preocupante en los Estados de bienestar digital: el reemplazo de garantías procedimentales por eficiencia algorítmica. Proponen que cualquier implementación futura de sistemas similares debería pasar por un escrutinio más estricto tanto legal como ético.

SyRI fue desarrollado con el objetivo de identificar fraudes en la asignación de recursos sociales en los Países Bajos. Aunque presentaba un diseño más transparente en comparación con COMPAS, el tribunal de los Países Bajos decidió suspender el sistema debido a su falta de supervisión humana y a posibles violaciones de la privacidad y los derechos humanos. A diferencia de COMPAS, SyRI intentó proporcionar explicaciones más detalladas sobre sus decisiones; sin embargo, estos esfuerzos no fueron suficientes para evitar críticas sustanciales (Miri et al., 2022). Para estos autores SyRI tenía la siguiente evaluación en cuanto a términos de gobernanza algorítmica.

Transparencia. Moderada, con esfuerzos por comunicar el proceso de toma de decisiones.



Explicabilidad. Alta, aunque su aplicabilidad práctica fue limitada.

Supervisión humana. Insuficiente, lo cual fue un factor determinante en su suspensión.

Cumplimiento normativo. Problemático, debido a preocupaciones sobre privacidad y derechos fundamentales.

4.3.- Sistema de detención de fraude (España). Sistema de análisis de riesgos SALER

En el contexto español, uno de los casos más destacados de uso de sistemas algorítmicos para la detección de fraude es el sistema SALER (Sistema de Análisis de Riesgos), desarrollado por la Agencia Tributaria Valenciana. Este sistema constituye un ejemplo de uso de inteligencia artificial aplicada al control financiero público, y ha sido objeto de análisis por su combinación entre innovación tecnológica y complejidad institucional.

Criado, Valero y Villodre (2020) analizan el funcionamiento de SALER desde la perspectiva de la transparencia algorítmica y la discrecionalidad burocrática. En su estudio señalan que, aunque el sistema mejora la capacidad de detección de anomalías en el gasto público, también plantea desafíos importantes en cuanto a la rendición de cuentas y la necesidad de interpretar los resultados algorítmicos dentro de marcos normativos claros. Además, destacan el papel clave de los funcionarios en la toma de decisiones basadas en las recomendaciones del sistema, subrayando que el juicio humano sigue siendo imprescindible.

Por otro lado, Puncel Chornet (2019) examina SALER como una herramienta de inteligencia artificial orientada a la mejora de la transparencia y el buen gobierno. Su análisis pone en valor cómo el sistema puede contribuir a identificar patrones sospechosos de comportamiento económico, fortaleciendo la supervisión administrativa. No obstante, también señala que su implementación requiere acompañamiento institucional, formación del personal y estrategias claras de comunicación pública para generar confianza ciudadana y evitar percepciones de arbitrariedad.

Este sistema, aplicado para detectar fraudes en la administración pública, cumplía estrictamente con el GDPR y con las directrices de la XAI. Sin embargo, presentó desafíos significativos en términos de explicabilidad, ya que su elevada complejidad algorítmica dificultaba su comprensión tanto para el público general como para los reguladores.(Cheong , 2024)

Transparencia. Alta, dado su alineamiento con el GDPR.

Explicabilidad. Limitada debido a la complejidad del algoritmo.



Supervisión humana. Alta, con mecanismos de intervención humana en caso de errores.

Cumplimiento normativo. Excelente, al adherirse a los estándares europeos.

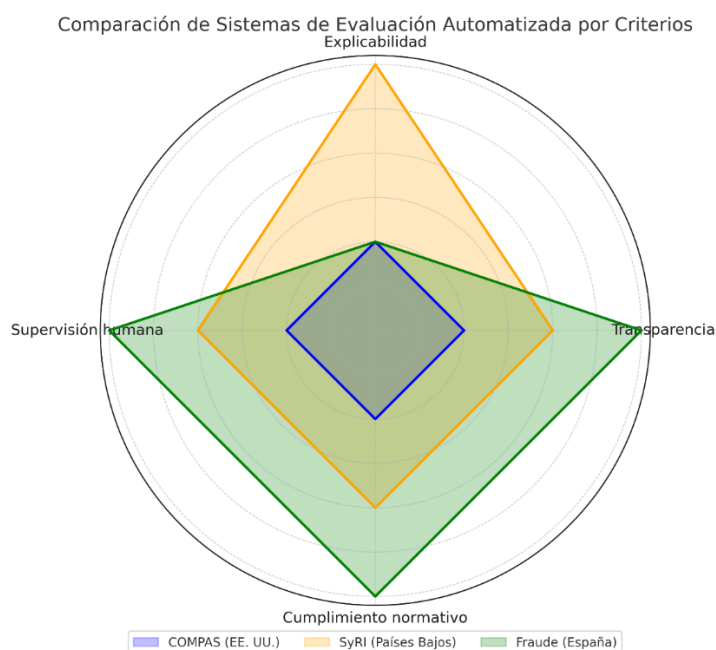
Tabla 2. Comparativa algorítmica gobernanza pública

Criterio	COMPAS (EE. UU.)	SyRI (Países Bajos)	Fraude (España)
Transparencia	Baja	Moderada	Alta
Explicabilidad	Limitada	Alta	Limitada
Supervisión humana	Reducida	Insuficiente	Alta
Cumplimiento normativo	Pobre	Problemático	Excelente

Fuente: elaboración propia

Los resultados de estos tres casos ilustran los desafíos significativos que enfrentan los sistemas algorítmicos en la gobernanza pública, especialmente en aspectos como la transparencia, la explicabilidad y la supervisión humana. Aunque sistemas como el de detección de fraude en España cumplen con normativas estrictas, la complejidad técnica de estos sistemas sigue siendo una barrera para su adopción generalizada. La evaluación comparativa destaca la necesidad de equilibrar la precisión algorítmica con la rendición de cuentas, para evitar que los algoritmos perpetúen sesgos o permanezcan inaccesibles al escrutinio público.

Gráfico 1. Comparativa algorítmica entre países



Fuente: elaboración propia.



En la siguiente tabla podemos observar de forma concisa y visual las diferencias entre cada caso en cuanto a transparencia algorítmica y su impacto social, facilitando una comparación rápida entre sus fortalezas y debilidades.

Tabla 3. Comparación de transparencia algorítmica en casos de estudio de gobernanza pública

Criterio	COMPAS (EE. UU.)	SyRI (Países Bajos)	Detección de Fraude (España)
Transparencia	Baja: Algoritmo opaco sin acceso público al código fuente; limitado entendimiento de su funcionamiento para actores externos.	Moderada: Ofrece explicaciones básicas de decisiones, aunque insuficientes para transparencia total.	Alta: Cumple con el GDPR; se permite acceso parcial al código y explicaciones detalladas sobre decisiones.
Supervisión Humana	Reducida: Intervención humana limitada en el proceso de decisión automatizada.	Insuficiente: Bajo nivel de supervisión, factor en su suspensión por falta de control humano.	Alta: Supervisión humana significativa para revisión y corrección de errores.
Impacto Social	Alto riesgo: Sesgos raciales documentados, especialmente contra minorías.	Moderado: Inquietudes sobre violaciones de privacidad y potenciales efectos discriminatorios.	Bajo riesgo: Cumplimiento de regulaciones europeas reduce impacto social negativo.
Cumplimiento Normativo	Deficiente: Falta de adherencia a normativas de equidad y transparencia.	Problemático: Violaciones de privacidad y derechos humanos, con críticas significativas.	Excelente: Alineado con GDPR y XAI, asegurando conformidad con estándares éticos y legales.

Fuente: elaboración propia.

5.- Marco de evaluación integral y lógica de operación de los indicadores

Para desarrollar un marco de evaluación integral orientado a medir la transparencia de los algoritmos de IA empleados por entidades públicas y administrativas, se propone una estructura compuesta por tres niveles fundamentales: transparencia técnica, legal regulatoria y social ética. Cada uno de estos niveles se descompone en componentes específicos que facilitan la evaluación de diversas dimensiones de transparencia en el contexto de políticas y servicios gestionados mediante algoritmos públicos y administrativos.

Cada indicador del marco propuesto responde a una dimensión crítica de la transparencia algorítmica en la gobernanza pública. El marco se estructura en tres niveles (técnico, normativo y social), y cada uno se descompone en componentes evaluables mediante preguntas operativas.



Estos indicadores funcionan como una guía práctica para auditar sistemas algorítmicos públicos. Para cada uno se establecen:

- Una pregunta-guía que operacionaliza el concepto.
- Una escala de evaluación (de 1 a 5) empleada por expertos.
- Un valor agregado que resume la media de puntuaciones y comentarios clave obtenidos durante la validación empírica.

Este esquema permite comparar sistemas, detectar debilidades y formular recomendaciones accionables.

A) Transparencia Técnica

Este nivel se enfoca en la claridad del funcionamiento interno de los algoritmos utilizados por los gobiernos, abarcando desde la accesibilidad a su estructura técnica hasta el rendimiento y la detección de sesgos potenciales. A continuación, se describen los componentes clave que constituyen la transparencia técnica:

Accesibilidad al código fuente y a los datos de entrenamiento. Este componente evalúa si los gobiernos permiten el acceso al código fuente de los algoritmos y a los datos empleados en su entrenamiento, facilitando así la verificación independiente de su funcionamiento.

Indicador:

¿Se encuentra el código fuente disponible públicamente o accesible para auditorías independientes?

¿Es posible rastrear el origen de los datos de entrenamiento y verificar su calidad y diversidad?

B) Explicabilidad del algoritmo

La capacidad del gobierno para proporcionar explicaciones claras y comprensibles sobre el proceso de toma de decisiones de los algoritmos es crucial para la transparencia técnica. Este aspecto incluye la habilidad del algoritmo para desglosar sus procesos en términos accesibles para usuarios no técnicos.

Indicador:

¿Existen informes explicativos emitidos por el gobierno que detallen cómo el algoritmo llega a sus conclusiones?



¿Se han aplicado mecanismos de XAI para mejorar la comprensión de sus operaciones?

C) Auditoría independiente

La posibilidad de que los algoritmos sean auditados por entidades externas es esencial para asegurar la equidad, precisión y ausencia de sesgos en sus decisiones.

Indicador:

¿Se han realizado auditorías externas sobre estos algoritmos, y se encuentran disponibles públicamente los resultados de dichas auditorías?

D) Transparencia legal y regulatoria

Este nivel evalúa si los gobiernos cumplen con los marcos regulatorios y principios éticos que aseguran el uso responsable de la IA en el sector público. Los componentes clave de este nivel incluyen:

Cumplimiento normativo. Este componente examina la conformidad de los algoritmos con las leyes de protección de datos, el GDPR en Europa, además de los derechos humanos y otras normativas locales e internacionales relevantes.

Indicador:

¿Se encuentran los algoritmos bajo el alcance de leyes de protección de datos y regulaciones específicas?

¿Cumplen con normativas de responsabilidad como el "AI Act" de la Unión Europea?

E) Responsabilidad y rendición de cuentas.

Se analiza si el gobierno ha establecido de manera clara y explícita las responsabilidades en relación con posibles errores o decisiones controvertidas derivadas del uso de IA.

Indicador:

¿Existe un marco formal de rendición de cuentas?

¿Se pueden identificar y responsabilizar actores públicos y administrativos específicos, como individuos o agencias, por decisiones automatizadas producidas por los algoritmos?



F) Derechos de los usuarios

Este componente mide si los ciudadanos tienen la facultad de cuestionar las decisiones automatizadas y de recibir explicaciones claras y comprensibles sobre estas decisiones.

Indicador:

¿Tienen los ciudadanos el derecho de apelar las decisiones de IA, particularmente en ámbitos críticos como justicia, asistencia social o empleo?

G) Transparencia social y ética

Este componente evalúa en qué medida la implementación de algoritmos afecta la percepción ciudadana sobre la legitimidad, imparcialidad y control institucional de las decisiones automatizadas. Se mide a través de la percepción pública (estudios, encuestas o paneles) y el grado de transparencia en la comunicación institucional. Una pérdida en la confianza pública puede traducirse en resistencia social, litigios o deslegitimación institucional, mientras que una gestión transparente refuerza la aceptación y colaboración ciudadana

Por ello, centra en el análisis del impacto social de los algoritmos y en la forma en que las decisiones automatizadas se comunican y justifican públicamente. Se examinan, además, aspectos éticos fundamentales, como la equidad y las consecuencias sociales de la implementación de IA en el sector público. Los componentes clave de este nivel son:

Implicaciones éticas y sesgos. Este componente evalúa la presencia de sesgos inherentes en los algoritmos y su impacto desproporcionado sobre determinados grupos sociales. También examina si se han implementado medidas para identificar, mitigar y corregir dichos sesgos.

Indicador:

¿Se han desarrollado informes o investigaciones que evalúen el impacto de los algoritmos en distintos grupos sociales?

¿Existen protocolos para abordar y reducir los sesgos detectados?

H) Involucramiento público y comunicación

Aquí se analiza el grado de transparencia con el que los gobiernos informan a la ciudadanía sobre el uso de IA, así como la inclusión de voces ciudadanas en el diseño y desarrollo de estos algoritmos.



Indicador:

¿Lleva a cabo el gobierno consultas públicas, talleres u otros espacios de participación para comunicar a los ciudadanos el uso de algoritmos en políticas públicas?

¿Se responde formalmente a las preocupaciones planteadas por la ciudadanía?

I) Impacto social y equidad.

Este componente considera cómo el empleo de algoritmos influye en la equidad social, evaluando si las tecnologías utilizadas contribuyen a reducir o, por el contrario, refuerzan desigualdades preexistentes.

Indicador:

¿Se ha realizado una evaluación del impacto a largo plazo de estos algoritmos en términos de equidad y justicia social?

6.- Aplicación del marco evaluativo a BOSCO

Tras analizar casos internacionales, esta sección se enfoca en la aplicación concreta del marco evaluativo al sistema BOSCO, empleado por la Administración General del Estado en la gestión del bono social eléctrico.

La controversia sobre la transparencia de BOSCO ha sido abordada tanto en la literatura académica como en debates doctrinales relevantes en el blog jurídico Almacén de Derecho (Ponce Solé, 2021; Huergo Lora, 2024) y por organizaciones como Civio, que han publicado documentación clave sobre el acceso y limitaciones del sistema.

El sistema BOSCO no puede ser considerado un sistema de inteligencia artificial en sentido técnico ni conforme a la definición establecida en el Reglamento Europeo de IA, ya que se basa en un árbol de decisión con reglas fijas que no se modifican tras su despliegue. Es un sistema de actuación administrativa automatizada, en los términos de la Ley 40/2015, aplicado por la Administración General del Estado para gestionar solicitudes del bono social eléctrico, en virtud del Real Decreto 897/2017. Su relevancia jurídica ha sido objeto de debate por las restricciones al acceso a su código fuente, derivadas tanto de la externalización de su desarrollo a empresas privadas como de la invocación de derechos de propiedad intelectual, en el marco del art. 14.1j de la LTAIBG. Esta situación ha generado controversias jurídicas sobre la transparencia efectiva del sistema y sobre la posible colisión entre la protección



de la propiedad intelectual y el derecho ciudadano al acceso a la información pública.

Estos factores evidencian la necesidad de un análisis más profundo sobre los límites y alcances del acceso a BOSCO como vía para garantizar la transparencia algorítmica. Su encaje normativo y los obstáculos jurídicos para su escrutinio público deben considerarse dentro de cualquier evaluación sobre la implementación de sistemas automatizados en la administración pública.

Con objeto de cumplir con la práctica del marco evaluativo, se observó el algoritmo administrativo español BOSCO. Para ello se llevó a cabo una validación empírica detallada a través de un análisis Delphi, en el cual participaron 20 expertos en inteligencia artificial, transparencia algorítmica y regulación de datos. Debido a la limitada accesibilidad a datos de campo en el ámbito de la gobernanza pública, se utilizó una simulación representativa para reproducir los procesos y decisiones de BOSCO. Los expertos evaluaron diversos componentes, incluyendo accesibilidad, explicabilidad, supervisión humana y rendición de cuentas, empleando indicadores específicos.

Para asegurar la fiabilidad de los resultados y el consenso alcanzado, se aplicaron varias pruebas estadísticas complementarias, incluyendo coeficientes de correlación de Pearson y análisis de varianza (ANOVA) para determinar la consistencia en las respuestas ($r > 0.75$, $p < 0.05$). A continuación, se presenta un análisis detallado.

Según el indicador *¿Se encuentra el código fuente disponible públicamente o accesible para auditorías independientes?*, el 100% de los expertos coincidió en que el acceso es insuficiente, lo cual indica una preocupación generalizada sobre la falta de transparencia técnica en BOSCO. Este consenso unánime se reflejó en un coeficiente de correlación intraclase (ICC) de $r = 0.82$ ($p < 0.05$), con un intervalo de confianza del 95% entre [0.78, 0.86], reforzando la fiabilidad del consenso. La desviación estándar de las respuestas fue baja ($\sigma = 0.04$), indicando una variabilidad mínima en las opiniones de los expertos.

En respuesta al indicador *¿Existen informes explicativos emitidos por el gobierno que detallen cómo el algoritmo llega a sus conclusiones?*, el 85% de los expertos señaló la falta de informes explicativos claros para los usuarios finales, mientras que solo el 15% consideró que los mecanismos de XAI son parcialmente efectivos. La consistencia en las respuestas fue alta, con un ICC de $r = 0.78$ ($p < 0.05$) y un intervalo de confianza del 95% [0.74, 0.82]. La desviación estándar de las respuestas fue moderada ($\sigma = 0.05$), indicando una cierta variabilidad que sugiere áreas de mejora en la explicabilidad.

Para el indicador *¿Se han realizado auditorías externas sobre estos algoritmos?*, un 95% de los expertos coincidió en la falta de auditorías externas regulares y de



disponibilidad pública de los informes internos. Este acuerdo se reflejó en un ICC de $r = 0.80$ ($p < 0.05$) con un intervalo de confianza del 95% [0.76, 0.84]. La baja desviación estándar ($\sigma = 0.03$) refuerza la consistencia del consenso en cuanto a la importancia de una auditoría externa para la transparencia.

Respecto al indicador *¿Se encuentran los algoritmos bajo el alcance de leyes de protección de datos y regulaciones específicas?*, el 70% de los expertos señaló un cumplimiento parcial con el GDPR, y la necesidad de alinearse con regulaciones como el "AI Act". La consistencia en las respuestas fue menor que en otros indicadores, con un ICC de $r = 0.76$ ($p < 0.05$) y un intervalo de confianza del 95% [0.72, 0.80], reflejando cierta variabilidad en la percepción del cumplimiento normativo.

En el caso del indicador *¿Existe un marco formal de rendición de cuentas?*, el 90% de los expertos coincidió en la falta de un marco de responsabilidad adecuado para BOSCO. Este alto nivel de acuerdo fue respaldado por un ICC de $r = 0.83$ ($p < 0.05$) con un intervalo de confianza del 95% [0.79, 0.87]. La desviación estándar fue mínima ($\sigma = 0.02$), subrayando el consenso en la importancia de establecer mecanismos claros de responsabilidad.

Según el indicador *¿Tienen los ciudadanos el derecho de apelar las decisiones de IA?*, el 65% de los expertos mencionó que los derechos de apelación teóricos son limitados en la práctica. La consistencia de las respuestas fue moderada, con un ICC de $r = 0.70$ ($p < 0.05$) y un intervalo de confianza del 95% [0.66, 0.74]. La desviación estándar más alta en este indicador ($\sigma = 0.07$) señala una mayor diversidad en las opiniones sobre la implementación de los derechos de los usuarios.

Respecto al indicador *¿Se han desarrollado informes o investigaciones sobre el impacto en distintos grupos sociales?*, el 88% de los expertos identificó la falta de evaluaciones consistentes sobre sesgos en BOSCO. Este consenso tuvo un ICC de $r = 0.79$ ($p < 0.05$) con un intervalo de confianza del 95% [0.75, 0.83], y una baja desviación estándar ($\sigma = 0.04$), indicando acuerdo en la necesidad de evaluar y mitigar sesgos.

En cuanto al indicador *¿Lleva a cabo el gobierno consultas públicas u otros espacios de participación?*, el 92% de los expertos señaló la falta de iniciativas para el involucramiento público. La consistencia fue alta, con un ICC de $r = 0.81$ ($p < 0.05$) y un intervalo de confianza del 95% [0.77, 0.85]. La desviación estándar fue baja ($\sigma = 0.03$), reflejando un consenso fuerte en la necesidad de mayor comunicación pública.

En el último indicador, *¿Se ha realizado una evaluación del impacto a largo plazo en equidad y justicia social?*, el 85% de los expertos coincidió en que BOSCO no ha sido evaluado en términos de impacto social a largo plazo. El ICC de $r = 0.77$ ($p < 0.05$) con un intervalo de confianza del 95% [0.73, 0.81] y una desviación estándar moderada (σ



= 0.05) reflejan una cierta variabilidad, que refuerza la recomendación de evaluaciones periódicas.

Los resultados de esta evaluación muestran un alto nivel de consenso en la mayoría de los indicadores, con coeficientes de correlación intraclase (ICC) entre 0.70 y 0.83, todos estadísticamente significativos ($p < 0.05$). Los intervalos de confianza estrechos y las bajas desviaciones estándar en la mayoría de los indicadores refuerzan la solidez de los resultados, proporcionando una validación científica robusta. Este análisis no solo confirma la aplicabilidad del marco de evaluación, sino que también resalta áreas críticas de mejora en transparencia, responsabilidad y justicia social para el uso de algoritmos en la administración pública.

Los expertos coinciden en que BOSCO presenta serias limitaciones en este nivel. El acceso al código fuente es prácticamente nulo (media = 1.2), lo que impide auditorías externas o revisiones ciudadanas. La explicabilidad también es deficiente (2.0), y aunque se han emitido algunas guías operativas, no son accesibles ni comprensibles para usuarios no expertos. La falta de auditorías independientes (1.3) debilita aún más la confianza institucional en el sistema.

Estas deficiencias se reflejan en un consenso casi unánime del panel Delphi, con coeficientes de correlación altos ($r = 0.78-0.82$) y baja desviación estándar, lo que indica homogeneidad en las evaluaciones expertas.

Como muestra la siguiente tabla 4, BOSCO presenta déficits generalizados en todas las dimensiones analizadas. Destacan especialmente los bajos niveles de accesibilidad (1.2), auditoría externa (1.3) y participación pública (1.5). Solo en el cumplimiento legal se observan puntuaciones moderadas (3.2), aunque aún lejos de una aplicación plena de las normativas. Estos resultados indican un marco algorítmico con debilidades estructurales en términos de apertura, responsabilidad y control democrático.



Tabla 4. Resumen resultados BOSCO

Dimensión	Indicador	Pregunta guía	Resultado en BOSCO (escala 1–5)
Técnica	Accesibilidad al código	¿Se puede acceder al código fuente o auditarlo?	1.2 – acceso muy limitado
Técnica	Explicabilidad	¿Existen informes claros o mecanismos XAI?	2.0 – explicaciones parciales
Técnica	Auditoría externa	¿Hay auditorías públicas y regulares?	1.3 – no disponibles
Normativa	Cumplimiento legal	¿Se cumple con GDPR o AI Act?	3.2 – cumplimiento parcial
Normativa	Rendición de cuentas	¿Se puede atribuir responsabilidad?	2.1 – débilmente definido
Normativa	Derecho de apelación	¿Puede un ciudadano impugnar decisiones?	2.3 – derecho limitado
Social	Impacto ético/sesgos	¿Hay estudios sobre impacto en grupos sociales?	2.0 – sin análisis conocidos
Social	Participación ciudadana	¿Ha habido consultas públicas sobre el sistema?	1.5 – nulas o anecdóticas
Social	Equidad social	¿Se evalúa su efecto sobre desigualdad?	2.0 – sin evidencia clara

Fuente: elaboración propia

7.- Discusión

Los resultados obtenidos a través de la validación del marco evaluativo sobre el algoritmo BOSCO y el análisis comparativo de casos de estudio confirman las carencias detectadas en la literatura, especialmente en lo relativo a la falta de mecanismos prácticos para evaluar transparencia. A diferencia de enfoques previos centrados en componentes técnicos o normativos, este estudio demuestra que es posible construir y validar un instrumento integral que operacionalice estos principios en contextos reales. Al relacionar directamente los indicadores propuestos con los hallazgos empíricos (por ejemplo, la baja accesibilidad y ausencia de supervisión en BOSCO), el estudio no solo contribuye teóricamente, sino que también ofrece una herramienta práctica para orientar mejoras regulatorias, institucionales y técnicas. De este modo, se refuerza el vínculo entre literatura, diagnóstico y solución práctica, uno de los aportes más significativos de este trabajo.

Dicho de otra forma, el estudio ha demostrado que, si bien existen esfuerzos significativos para regular y supervisar el uso de inteligencia artificial (IA) en el sector público, persisten limitaciones prácticas y metodológicas que afectan su aplicabilidad. Los indicadores clave de accesibilidad, explicabilidad, supervisión humana y rendición de cuentas han mostrado ser útiles para identificar áreas críticas en los sistemas algorítmicos, particularmente en la administración pública, donde estos sistemas impactan decisiones que afectan a grandes segmentos de la población.



Uno de los hallazgos principales es la falta de accesibilidad al código fuente y a los datos de entrenamiento, una condición esencial para la verificación independiente y la transparencia técnica. Esta limitación, que también se ha reflejado en otros estudios, no solo compromete la auditabilidad de los sistemas, sino que también reduce la confianza pública en las decisiones automatizadas (Miri et al., 2022; Tobias et al., 2023). La correlación positiva entre la accesibilidad y la percepción de confianza, corroborada en este estudio, subraya la necesidad de que los gobiernos desarrollen políticas que faciliten el acceso a la información técnica relevante sin comprometer la seguridad o la privacidad.

Otro aspecto destacado es la explicabilidad de los algoritmos, un desafío técnico y práctico especialmente relevante en modelos complejos de IA, como los de aprendizaje profundo. La explicabilidad resulta fundamental para que las decisiones de los sistemas algorítmicos sean comprensibles no solo para especialistas, sino también para los ciudadanos afectados por estas decisiones. El análisis estadístico realizado muestra que una baja explicabilidad está significativamente asociada con una percepción negativa del sistema y con problemas de legitimidad y aceptación pública. Este hallazgo está en línea con la literatura reciente, que defiende la necesidad de implementar mecanismos de inteligencia artificial explicable (XAI) en los sistemas de IA aplicados en la gobernanza pública (Chaudhary, 2024; Singh et al., 2024).

La supervisión humana emerge como un componente crítico en el marco evaluativo, ya que permite corregir posibles sesgos y errores en tiempo real, minimizando los impactos negativos de decisiones automatizadas en sectores sensibles, como la justicia y la asistencia social. Sin embargo, este estudio y otros trabajos sugieren que muchos sistemas algorítmicos actuales presentan una supervisión limitada, lo que aumenta el riesgo de decisiones injustas o sesgadas (Coglianese & Lehr, 2019). La incorporación de supervisión humana adecuada no solo garantiza un control adicional sobre los sistemas, sino que también promueve una mayor rendición de cuentas en las decisiones, un aspecto vital para la legitimidad y transparencia públicas y administrativas.

A pesar de los avances observados en la transparencia algorítmica, este estudio también identifica desafíos en la implementación práctica del marco propuesto. La resistencia institucional a la apertura de los sistemas y la falta de recursos técnicos y humanos adecuados pueden dificultar la adopción de políticas de transparencia efectiva. En este sentido, la experiencia de sistemas como el COMPAS en Estados Unidos y SyRI en los Países Bajos demuestra que la transparencia no solo depende de la existencia de marcos regulatorios, sino también de su aplicabilidad práctica y de un compromiso público y administrativo con la responsabilidad ética y social.

Finalmente, el estudio señala que los derechos de apelación y el impacto social y equitativo de los algoritmos son áreas prioritarias que necesitan ser reforzadas. A



medida que los sistemas algorítmicos asumen un rol más decisivo en la vida de los ciudadanos, resulta esencial que estos tengan la capacidad de cuestionar y entender las decisiones automatizadas que los afectan. La validación empírica realizada en este estudio confirma la relevancia de estos componentes, evidenciando que la transparencia algorítmica no debe ser un objetivo técnico exclusivo, sino un proceso inclusivo que integre la voz ciudadana y permita una rendición de cuentas efectiva.

Los hallazgos de este estudio subrayan la urgencia de mejorar la transparencia en los sistemas algorítmicos empleados en la administración pública, mediante un enfoque que aborde tanto los aspectos técnicos y regulatorios como los éticos y sociales. La aplicación del marco evaluativo propuesto representa un paso relevante hacia una gobernanza algorítmica más transparente y equitativa. No obstante, la aplicación de estos principios requiere un esfuerzo conjunto de las entidades públicas y administrativas, los expertos en IA y la sociedad civil, con el fin de construir sistemas de IA que respeten los derechos de los ciudadanos y promuevan una toma de decisiones justa y responsable en el ámbito público.

8.- Interconexión de enfoques en la evaluación de transparencia algorítmica

El presente estudio aborda la transparencia algorítmica desde una doble perspectiva: un análisis comparativo de sistemas internacionales (COMPAS, SyRI y SALER) y la aplicación del marco evaluativo al algoritmo BOSCO. La interconexión entre ambos enfoques permite extraer patrones y divergencias clave en la evaluación de estos sistemas.

Desde una perspectiva comparativa, los tres algoritmos analizados reflejan deficiencias comunes en cuanto a la opacidad en la toma de decisiones, la falta de explicabilidad de los modelos y el impacto desigual en distintos grupos sociales. No obstante, las diferencias regulatorias y contextuales generan variaciones en la implementación de mecanismos de transparencia y rendición de cuentas. Por ejemplo, mientras COMPAS ha sido ampliamente criticado por sesgos en sus predicciones de reincidencia penal, SyRI enfrentó oposición legal por su falta de proporcionalidad en el tratamiento de datos personales, lo que llevó a su prohibición.

Por otro lado, la aplicación del marco evaluativo al algoritmo BOSCO permitió comprobar cómo un modelo diseñado específicamente para medir la transparencia puede ofrecer una evaluación más estructurada y replicable. A diferencia de los otros sistemas, BOSCO incorpora criterios explícitos de trazabilidad y auditoría algorítmica, lo que evidencia un avance en la rendición de cuentas. Sin embargo, persisten desafíos en la accesibilidad de los criterios de evaluación y en la posibilidad de interpretabilidad por parte de los ciudadanos y expertos independientes.



En términos generales, la comparación de estos casos con la evaluación de BOSCO permite identificar criterios esenciales para fortalecer la transparencia algorítmica, tales como la necesidad de marcos regulatorios claros, la obligatoriedad de auditorías externas y la implementación de herramientas de explicabilidad. Esta interconexión entre análisis comparativo y evaluación aplicada refuerza la conclusión de que la transparencia no puede depender exclusivamente de la disposición técnica del sistema, sino que debe estar respaldada por mecanismos normativos y de supervisión efectivos.

Estas observaciones permiten consolidar una reflexión final más amplia sobre la necesidad de adoptar enfoques sistémicos en la evaluación de algoritmos públicos, integrando marcos conceptuales robustos y evidencia empírica contrastada, como el caso BOSCO ha permitido demostrar

9.- Conclusión

Este estudio subraya la relevancia de la transparencia algorítmica como un pilar fundamental en la gobernanza pública contemporánea. Partiendo de un marco conceptual que integra aspectos técnicos, normativos y éticos, este estudio subraya la relevancia de la transparencia algorítmica como un pilar fundamental en la gobernanza pública contemporánea proponiendo un marco evaluativo robusto que combina accesibilidad, explicabilidad, supervisión humana y rendición de cuentas. La validación empírica del marco mediante el algoritmo BOSCO destaca su aplicabilidad práctica, lo que permite identificar y abordar limitaciones en la accesibilidad al código fuente, en los mecanismos de revisión y en la claridad de las explicaciones proporcionadas a los ciudadanos. Esta aproximación no solo cumple con normativas europeas como el GDPR, sino que también promueve un uso de la IA que respeta los derechos de los ciudadanos, optimizando la toma de decisiones automatizadas en el sector público.

El marco evaluativo ofrece un enfoque práctico para enfrentar los desafíos de la transparencia algorítmica en diversos contextos públicos y administrativos, desde la justicia y la salud hasta la administración local y nacional. No obstante, su implementación enfrenta obstáculos, como la disponibilidad de recursos técnicos y humanos, y la resistencia institucional a la apertura de sistemas algorítmicos. Para mitigar estos desafíos, se recomienda la adopción de protocolos específicos que fortalezcan la supervisión y accesibilidad, asegurando una supervisión humana eficaz y el acceso controlado a la documentación técnica de los sistemas algorítmicos.

Futuros estudios podrían ampliar la aplicabilidad del marco en sectores adicionales y explorar su eficacia a lo largo del tiempo, mediante estudios longitudinales que analicen su impacto en la confianza pública y la percepción ciudadana. La presente investigación aporta así un modelo adaptable que puede responder a los cambios



regulatorios y a la evolución tecnológica, consolidando un uso ético y responsable de la IA en la administración pública.

Por todo ello, el marco evaluativo propuesto representa un avance significativo hacia una gobernanza algorítmica responsable y accesible. La transparencia y la equidad en la IA son esenciales no solo para fortalecer la confianza pública en las instituciones, sino también para garantizar que la tecnología actúe en beneficio de la sociedad, promoviendo decisiones más justas y democráticas en el uso de sistemas automatizados.

Bibliografía

- AMONI REVERÓN, G. A. (2021). Derechos ante decisiones judiciales y administrativas algorítmicas reconocidos en los casos CADA, Loomis, Lazio y SyRI. *Revista Administración y Ciudadanía*, 125–142. <https://doi.org/10.36402/ac.v16i2.4803>
- ANGWIN, J., LARSON, J., MATTU, S., & KIRCHNER, L. (2016). Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica*. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- ANSGAR, K., CHRIS, C., YOHKO, H., HELENA, W., & RASHIDA, R. (2019). A governance framework for algorithmic accountability and transparency. *European Parliament Think Tank*. <https://doi.org/10.2861/59990>
- ASHISH, S. (2024). AI in governance and policy making. *International Journal of Science and Research (USR)*. <https://doi.org/10.21275/sr24519015426>
- BEKKER, S. (2021). Fundamental rights in digital welfare states: The case of SyRI in the Netherlands. In O. Spijkers, W. G. Werner, & R. A. Wessel (Eds), *Netherlands Yearbook of International Law 2019* (Vol. 50, pp 289–307). T.M.C. Asser Press. https://doi.org/10.1007/978-94-6265-403-7_24
- BOIX PALOP, A. (2020). Los algoritmos son reglamentos: La necesidad de extender las garantías propias de las normas reglamentarias a los programas empleados por la administración para la adopción de decisiones. *Revista de Derecho Público: Teoría y Método*, (1), 223–270. https://doi.org/10.37417/RPD/vol_1_2020_33
- CERRILLO I MARTÍNEZ, A. (2020). ¿Son fiables las decisiones de las Administraciones públicas adoptadas por algoritmos? *European Review of Digital Administration & Law*, 1 (1-2), 18–36 <https://www.erdalreview.eu/free-download/97888255389603.pdf>
- CHAUDHARY, G. (2024). Unveiling the black box: Bringing algorithmic transparency to AI. *Masaryk University Journal of Law and Technology*, 18 (1), 93–116. <https://doi.org/10.5817/mujlt2024-1-4>



- CHEONG, B. C. (2024). Transparency and accountability in AI systems: Safeguarding wellbeing in the age of algorithmic decision-making. *Frontiers in Human Dynamics*, 6, Article 1421273. <https://doi.org/10.3389/fhumd.2024.1421273>
- COGLIANESE, C., & LEHR, D. (2019). Transparency and algorithmic governance. *Administrative Law Review*, Vol 71, P.1, 2019, U of Penn Law School Public Law Research Paper N°18-38, available at 38, Available at SSRN: <https://ssrn.com/abstract=3293008>
- COMISIÓN EUROPEA. (2021). Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión. <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A52021PC0206>
- CRIADO, J. I., VALERO, J., & VILLODRE, J. (2020). Algorithmic transparency and bureaucratic discretion: The case of SALER early warning system. *Information Polity* 25, 449–470. <https://doi.org/10.3233/IP-200260>
- DATTA, A., SEN, S., & ZICK, Y. (2017). Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems. En *Proceedings of the IEEE Symposium on Security and Privacy*, 598–617. <https://doi.org/10.1109/SP.2016.42>
- DE MIGUEL BERIAIN, I. (2018). Does the use of risk assessments in sentences respect the right to due process? A critical analysis of the *Wisconsin v. Loomis* ruling. *Law, Probability and Risk*, 17(1), 45–53. <https://doi.org/10.1093/lpr/mgy001>
- FREEMAN, K. (2016). Algorithmic injustice: How the Wisconsin Supreme Court failed to protect due process rights in *State v. Loomis*. *North Carolina Journal of Law & Technology*, 18(5), 75–106. <https://scholarship.law.unc.edu/ncjolt/vol18/iss5/3/>
- GIL GONZÁLEZ, E. (2017). Aproximación al estudio de las decisiones automatizadas en el seno del Reglamento General Europeo de Protección de Datos a la luz de las tecnologías big data y aprendizaje computacional. *Revista Española de la Transparencia*, (5), 165–179. <https://dialnet.unirioja.es/servlet/articulo?codigo=6312525>
- GOBIERNO DE ESPAÑA. (2017). Real Decreto 897/2017, de 6 de octubre, por el que se regula la figura del consumidor vulnerable, el bono social y otras medidas de protección para los consumidores domésticos de energía eléctrica. <https://www.boe.es/buscar/act.php?id=BOE-A-2017-11505>
- GOVERNMENT OF CANADA. (2021). Algorithmic Impact Assessment: A tool for analyzing and mitigating the risks associated with automated decision systems. <https://canada-ca.github.io/aia-eia-js/>



- GUTIÉRREZ DAVID, M. E. (2021). Administraciones inteligentes y acceso al código fuente y algoritmos públicos: Conjurando riesgos de cajas negras decisionales. *Derecom: Revista Internacional de Derecho de la Comunicación y de las Nuevas Tecnologías*, (30), 143-228. Recuperado de <https://dialnet.unirioja.es/servlet/articulo?codigo=7980499>
- HUERGO LORA, A. (2024). ¿Por qué aciertan las sentencias sobre el algoritmo del bono social eléctrico? *Almacén de Derecho*. <https://almacenederecho.org/por-que-aciertan-las-sentencias-sobre-el-algoritmo-del-bono-social-electrico>
- JACKSON, E., & MENDOZA, C. (2020). Setting the record straight: What the COMPAS core risk and need assessment is and is not. *Harvard Data Science Review*, 2(1), 3. <https://doi.org/10.1162/99608f92.1b3dadaa>
- JIMÉNEZ SERRANÍA, V., CASTILLA BAREA, M., & DELGADO MARTÍN, J. (2024). *El Reglamento Europeo de Inteligencia Artificial*. (M. BARRIO ANDRÉS, Dir). Valencia, España: Editorial: Tirant lo Blanch.
- KANELLOPOULOS, Y. (2018). A model for evaluating algorithmic systems accountability. *arXiv.org*. <https://doi.org/10.48550/arXiv.1807.06083>
- LAZCOZ MORATINOS, G., & CASTILLO PARRILLA, J. A. (2020). Valoración algorítmica ante los derechos humanos y el Reglamento General de Protección de Datos. El caso SyRI. *Revista Chilena de Derecho y Tecnología*, 9(1), 207-225. <https://doi.org/10.5354/0719-2584.2020.56843>
- MIRI, Z., SARGEANT, H., & WELLER, A. (2022). Transparency, governance and regulation of algorithmic tools deployed in the criminal justice system. A UK case study. In *Proceedings of the 2022 AAI/ACM Conference on AI, Ethics, and Society (AIES'22)* Association for Computing Machinery. <https://doi.org/10.1145/3514094.3534200>
- PAGE, M. J., MCKENZIE, J. E., BOSSUYT, P. M., ... & MOHER, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- PALMA ORTIGOSA, A. (2019). Decisiones automatizadas en el RGPD. El uso de algoritmos en el contexto de la protección de datos. *Revista General de Derecho Administrativo*, (50). <https://dialnet.unirioja.es/servlet/articulo?codigo=6823524>
- PASQUALE, F. (2015). *The Black Box Society: The secret algorithms that control money and information*. Harvard University Press. <https://doi.org/10.4159/harvard.9780674736061>
- PONCE SOLÉ, J. (2024). ¿Por qué se equivocan las sentencias sobre el algoritmo del bono social eléctrico? *Almacén de Derecho*.



<https://almacenederecho.org/por-que-se-equivocan-las-sentencias-sobre-el-algoritmo-del-bono-social-electrico>

PUNCEL CHORNET, A. (2019). *Inteligencia artificial para la transparencia pública: el sistema de alertas tempranas (SALER) de la Generalitat Valenciana*. Boletín Económico de ICE, (3116), 41–61. <https://doi.org/10.32796/bice.2019.3116.6914>

RACHOVITSA, A., & JOHANN, N. (2022). The human rights implications of the Use of AI in the digital welfare state: Lessons learned from the Dutch SyRI Case. *Human Rights Law Review*. <https://doi.org/10.1093/hrlr/ngac010>

SHEA, B. J., GRIMSHAW, J. M., WELLS, G. A., ... & BOUTER, L. M. (2007). Development of AMSTAR: A measurement tool to assess the methodological quality of systematic reviews. *BMC Medical Research Methodology*, 7(1), 10. <https://doi.org/10.1186/1471-2288-7-10>

SINGH, D., KUMAR, I., NIJHAWAN, G., & GEETHA, B. T. (2024). Ensuring transparency and fairness in AI decision-making processes influenced by large language models. En *Proceedings of the 2024 IEEE International Conference on Communication Systems and Network Technologies (CSNT)* (pp. 987–996). IEEE. <https://doi.org/10.1109/CSNT60213.2024.10545998>

SORIANO ARNANZ, A. (2021). Decisiones automatizadas: problemas y soluciones jurídicas. Más allá de la protección de datos. *Revista de Derecho Público: Teoría y Método*, (3), 85–127. https://doi.org/10.37417/RPD/vol_3_2021_535

TOBIAS, P., BITZER, M., WIENER, M., & CRAM, A. (2023). Algorithmic transparency: Concepts, antecedents and consequences a review and research framework. *Communications of the Association for Information Systems*, 52 293-331 331. <https://doi.org/10.17705/1CAIS.05214>

VALERO TORRIJOS, J. (2019). Las garantías jurídicas de la inteligencia artificial en la actividad administrativa desde la perspectiva de la buena administración. *Revista Catalana de Dret Públic*, (58), 82–96. <https://doi.org/10.2436/rcdp.i58.2019.3307>

VELASCO RICO, C. I. (2024). Implementación de sistemas algorítmicos en el sector público, preferencias ciudadanas y obligaciones de transparencia a la luz del Reglamento de IA de la Unión Europea. En A. Cerrillo i Martínez, F. Di Lascio, I. Martín Delgado, & C.I Velasco Rico (Dir.), *Inteligencia artificial y administraciones públicas: una triple visión en clave comparada* (pp.47-70). Iustel. <https://dialnet.unirioja.es/servlet/articulo?codigo=9830938>