

Los principios DEI en el desarrollo de inteligencia artificial para la Administración pública

María E. Cortés-Cediel

Universidad Complutense de Madrid, ICCA

Adrián Rubio Pintado

Universidad Autónoma de Madrid

Iván Cantador

Universidad Autónoma de Madrid

Abstract

El despliegue de la inteligencia artificial (IA) en la Administración pública ha generado efectos adversos en la ciudadanía. Los sesgos algorítmicos de los sistemas de IA provocan distribuciones inequitativas de recursos y discriminación hacia ciertos colectivos vulnerables, mediante clasificaciones orientadas a mayorías, y barreras de accesibilidad y estándares homogeneizadores de “normalidad” que excluyen a personas con necesidades especiales. Aunque el Reglamento (UE) 2024/1689 (Ley de IA) constituye una regulación global relevante, su integración con principios de diversidad, equidad e inclusión (DEI) es limitada. Este trabajo propone un marco teórico para el desarrollo de sistemas de IA basado en el enfoque DEI, sintetizando la literatura académica e incorporando criterios éticos, de justicia social y heterogeneidad como guía operativa hacia una IA pública más inclusiva.

Palabras clave: *inteligencia artificial; DEI; diversidad; equidad; inclusión*

María Elicia Cortés Cediel: María Elicia Cortés Cediel es Profesora Ayudante Doctora del Departamento de Ciencia Política y de la Administración de la Facultad de Ciencias Políticas y Sociología de la Universidad Complutense de Madrid e investigadora del Instituto Complutense de Ciencia de la Administración (ICCA). Investiga el acceso de los ciudadanos a escenarios de toma de decisiones y a entornos de la e-Administración desde un punto de vista de igualdad, participación, justicia social y reducción de brechas digitales, en distintos proyectos.

Adrián Rubio Pintado: Adrián Rubio es científico de datos y doctorando en Ingeniería Informática en la Universidad Autónoma de Madrid, donde también obtuvo su grado y un máster orientado a la investigación en Ciencia de Datos. Tras desarrollar grandes modelos de lenguaje de código abierto en el *Barcelona Supercomputing Center* (2023-2026), actualmente trabaja en CaixaBank Tech integrando inteligencia artificial de vanguardia en sistemas de ciberseguridad.

Iván Cantador: Iván Cantador es Profesor Titular de Ciencias de la Computación e Inteligencia Artificial en la Universidad Autónoma de Madrid. Sus principales áreas de investigación incluyen la recuperación de información, los sistemas de recomendación y el procesamiento del lenguaje natural. Actualmente, su investigación se centra en el estudio de cómo los principios de diversidad, equidad e inclusión pueden incorporarse al diseño y desarrollo de la inteligencia artificial.

1. Introducción

Los sistemas de inteligencia artificial (IA) implementados en la Administración pública optimizan la gestión pública en ámbitos como la provisión de servicios públicos, la seguridad ciudadana, la atención sanitaria y la transformación digital de las instituciones (Borgesano et al., 2025; Fisher et al., 2023). Los algoritmos subyacentes a estos sistemas permiten abordar un amplio abanico de tareas, que van desde la automatización de expedientes hasta el uso de modelos analíticos para la toma de decisiones asistida en el diseño y la ejecución de políticas públicas (Cotilla Conceição y van der Stappen, 2025; Duke, 2026). En este contexto, surge un nuevo escenario de interacción en el que, a las ya existentes comunicaciones C2G (Ciudadano-a-Gobierno) y G2C (Gobierno-a-Ciudadano), recogidas en Gil et al. (2019), se suman nuevas formas de interacción entre la Administración y los ciudadanos. En particular, surgen las comunicaciones C2AI (Ciudadano-a-sistema de IA de la Administración pública) y G2AI (Servidor público-a-sistema de IA). A través de la comunicación C2AI, entre otras tareas, el ciudadano interactúa directamente con sistemas algorítmicos para realizar consultas complejas sobre información gubernamental y llevar a cabo trámites administrativos sin intermediación humana. En la comunicación G2AI, los modelos de IA asisten a los servidores públicos en una amplia gama de actividades, que pueden incluir el análisis de grandes volúmenes de datos y la toma de decisiones complejas.

El diseño de los sistemas de IA, sin embargo, puede generar barreras de accesibilidad en estas comunicaciones, potenciando una doble discriminación hacia colectivos minoritarios y vulnerables. Específicamente, a los potenciales efectos discriminatorios derivados de la toma de decisiones automatizada mediante algoritmos se suma la falta de accesibilidad atribuible a los propios sistemas. La IA puede reproducir estándares rígidos de «normalidad» al basarse en datos procedentes de colectivos dominantes para establecer patrones y modelos universales de conducta, apariencia y capacidad, lo que puede conducir a la discriminación de quienes se desvían de dichos patrones y modelos (Zhang et al., 2024). Esta cuestión resulta especialmente relevante en el ámbito de la Administración pública, donde el principio de universalidad implica garantizar el acceso a los servicios públicos de todos los ciudadanos. Al implementar herramientas digitales con sesgos de diseño, la discriminación puede adquirir un carácter sistémico, de modo que el Estado deja de ser inclusivo y pasa a actuar como un filtro automatizado que profundiza las desigualdades al excluir a colectivos minoritarios y vulnerables.

En este contexto, la integración de enfoques como DEI (Diversidad, Equidad e Inclusión) resulta fundamental para garantizar una accesibilidad tecnológica universal. Sin embargo, hasta donde alcanza nuestro conocimiento, esta perspectiva apenas ha sido explorada tanto en la literatura como en la práctica. Por ello, este trabajo tiene como objetivo elaborar un marco teórico que integre y desarrolle los principios de DEI en el diseño, la implementación y la gobernanza de las herramientas de IA implementadas en la Administración pública.

2. Conclusiones y trabajo futuro

Este trabajo ha tenido como objetivo la elaboración de un marco teórico integrador del enfoque DEI para el diseño, la implementación y la gobernanza de sistemas de IA en la Administración pública. Para ello, se ha adoptado un enfoque integrador que comprende no solo las cuestiones que deben abordar los algoritmos subyacentes a estos sistemas para evitar la discriminación de distintos sectores de la población, sino también una serie de prácticas que deben considerarse para garantizar una comunicación C2AI y G2AI inclusiva entre la ciudadanía y las instituciones, con el objeto de garantizar el principio de universalidad de la Administración pública. En particular, se han identificado como relevantes aspectos como el despliegue de herramientas multimodales, los modelos de gobernanza inclusivos, las estrategias de alfabetización digital y los mecanismos de acompañamiento y asistencia por parte de servidores públicos (humanos).

El trabajo presentado es de carácter teórico y, por tanto, debe ser complementado con estudios empíricos orientados al desarrollo y la evaluación de sistemas de IA para la Administración pública que integren el enfoque DEI, materializando las cuestiones revisadas. En este sentido, se considera fundamental la participación de actores pertenecientes a diversos colectivos en todas las fases del ciclo de vida de los sistemas, desde su concepción y diseño hasta su validación y seguimiento una vez desplegados.

Referencias

- Ali, A., & Iskandar, Y. H. P. (2025). The role of artificial intelligence in financial inclusion in emerging markets: A conceptual review. *Edelweiss Applied Science and Technology*, 9(5), 1303-1323.
- Atf, Z., & Lewis, P. R. (2025). Towards inclusive explainable artificial intelligence: a thematic analysis and scoping review on tools for persons with disabilities. *Disability and Rehabilitation: Assistive Technology*, 20(8), 2836-2857.
- Borgesano, F., De Maio, A., Laghi, P., & Musmanno, R. (2025). Artificial intelligence and justice: a systematic literature review and future research perspectives on Justice 5.0. *European Journal of Innovation Management*, 28(11), 349-385.
- Buch, D., Upadhyay, S., Lakhani, S., & Srivastava, H. (2026). Artificial Intelligence in Inclusive and Special Education: A Narrative Survey. *International Journal of Special Education*, 41(8s), 273-282.
- Chedrawi, C., Haddad, G., Haddad, G., & ElAli, R. (2026). Can artificial intelligence bridge the diversity, equity, and inclusion gap for a sustainable future? A PRISMA systematic review. *Journal of Innovation & Knowledge*, 15, 100986.
- Cong-Lem, N., Van Bui, C. N., Nguyen, N. P. N., & Huynh, M. Q. (2025). Bridging or breaking? A systematic review of how generative AI shapes equity in foreign language education. *International Journal of Applied Linguistics*.
- Cotilla Conceição, J. M., & van der Stappen, E. (2025). The Impact of AI on Inclusivity in Higher Education: A Rapid Review. *Education Sciences*, 15(9), 1255.

- Dwork, C. (2006). Differential privacy. In *International colloquium on automata, languages, and programming* (pp. 1-12). Springer.
- Duke, B. (2026). Artificial intelligence for inclusive and sustainable international education: A critical conceptual review. *Journal of International Students*, 16(10), 167-198.
- Fisher, E., Flynn, M. A., Pratap, P., & Vietas, J. A. (2023). Occupational safety and health equity impacts of artificial intelligence: a scoping review. *International Journal of Environmental Research and Public Health*, 20(13), 6221.
- Gazi, M. A. I., Shinwary, S. S., Khan, W., Alam, M. N., Hashim, F., Senathirajah, A. R. B. S., & Hossain, A. I. (2026). Ethical implications of AI-driven educational management systems on equity and social justice in sustainability education: A comprehensive review. *Multidisciplinary Reviews*, 9(11), 2026537.
- Ghanem, S., Moraleja, M., Gravesande, D., & Rooney, J. (2025). Integrating health equity in artificial intelligence for public health in Canada: A rapid narrative review. *Frontiers in Public Health*, 13, 1524616.
- Gil, O., Cortés-Cediel, M. E., & Cantador, I. (2019). Citizen participation and the rise of digital media platforms in smart governance and smart cities. *International Journal of E-Planning Research (IJEPR)*, 8(1), 19-34.
- Hassan, S., Ibrahim, S., Bielecki, J., Stanimirovic, A., Mathew, S., Hooey, R., ... & Rac, V. E. (2025). Integrating artificial intelligence in community-based diabetes care programmes: enhancing inclusiveness, diversity, equity and accessibility a realist review protocol. *BMJ Open*, 15(7), e100512.
- Jaime-Vargas, J. A. (2025). Generative AI for enhancing accessibility and inclusion in higher education: A systematic review. *HighTech and Innovation Journal*, 6(2), 723-734.
- Kalim, U., Kanwar, A., Xu, L., Xu, H., Chen, X., Irfan, S., & Rasool, A. (2025). Female gender bias in artificial intelligence applications for education: a systematic review of regional disparities and equity implications. *AI & SOCIETY*, 1-17.
- Iriqat, D., & Ashour, Y. (2026). Digital Exclusion or Zero Hunger? A Sustainability Review of Ethical AI in Fragile Contexts. *Sustainability*, 18(9), 4171.
- Lee, H., Kim, H., & Yan, W. (2025). Promoting inclusive AI and technology in K-12: A review of context, instructional strategies, and learning outcomes. *Computers and Education: Artificial Intelligence*, 100478.
- Li, J., Yan, Y., & Zeng, X. (2025). Exploring Artificial Intelligence in Inclusive Education: A Systematic Review of Empirical Studies. *Applied Sciences* (2076-3417), 15(23), 12624.
- Li, X., Li, B., Li, J., & Cho, S. J. (2025). Technology review of magic school AI: An intelligent way for education inclusivity and teacher workload reduction. *Education Sciences*, 15(8), 963.

- Liu, R., & Glave, C. Z. (2026). Justice, Equity, Diversity, and Inclusion in Artificial Intelligence and Childhood Education: A Critical Systematic Review. *Journal of Research in Childhood Education*, 40(1), 5-23.
- Marko, J. G. O., Neagu, C. D., & Anand, P. B. (2025). Examining inclusivity: the use of AI and diverse populations in health and social care: A systematic review. *BMC Medical Informatics and Decision Making*, 25(1), 57.
- Moral, S. V., & Moreno-Tallón, F. (2025). Inteligencia Artificial y Educación Inclusiva: soluciones tecnológicas para una enseñanza accesible. Revisión Sistemática. *Digital Education Review*, (47), 62-77.
- Mushkani, R., & Koseki, S. (2026). Street review: A participatory AI-based framework for assessing streetscape inclusivity. *Cities*, 170, 106602.
- Nantaburom, S., & Wetcho, S. (2026). The role of artificial intelligence in special and inclusive education: A systematic literature review. *Contemporary Educational Technology*, 18(2), ep652.
- Naoum, R. F., Szakadáti, T., & Balogh, G. (2026). Artificial Intelligence (AI) in human resource management (HRM): a systematic review of its dual impact on diversity, equity, and inclusion (DEI). *Management Review Quarterly*, 1-72.
- NAPA (2005). Sounding the Call to the Public Administration Community: The Social Equity Challenge in the US. *National Academy of Public Administration, Panel on Social Equity, Research Committee, Washington, DC*.
- Ni, W., & Li, J. (2026). Mapping the impact of generative AI in higher education: a scoping review of psychological and equity dimensions. *Frontiers in Psychology*, 17, 1856854.
- Pagliara, S. M., Bonavolonta, G., Pia, M., Falchi, S., Zurru, A. L., Fenu, G., & Mura, A. (2024). The integration of artificial intelligence in inclusive education: A scoping review. *Information*, 15(12), 774.
- Paglieri, L., Costumato, L., & Bonomi Savignon, A. (2025). Does artificial intelligence improve social equity in public services? A systematic literature review on opportunities and pitfalls. *International Journal of Public Sector Management*, 1-23.
- Pulijala, S. (2024). Artificial intelligence in governance: Opportunities, challenges, and ethical implications for public administration. *International Journal for Multidisciplinary Research*, 6(6), 1-10.
- Salas-Pilco, S. Z., Xiao, K., & Oshima, J. (2022). Artificial intelligence and new technologies in inclusive education for minority students: A systematic review. *Sustainability*, 14(20), 13572.
- Siddaway, A. P., Wood, A. M., & Hedges, L. V. (2019). How to do a systematic review: a best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annual Review of Psychology*, 70(1), 747-770.

Suprihatiningrum, J., & Syaripudin, A. (2026). Technology for equity: a systematic review of AI, adaptive tools, and digital platforms in differentiated instruction. *Texto Livre*, 19, e61658.

Tronnier, F., Löbner, S., Azanbayev, A., & Walter, M. L. (2024) From Bias to Balance—A Systematic Literature Review on Gender Bias in AI. *Proceedings of the 28th Pacific-Asia Conference on Information Systems*.

Wang, S., Mei, Y. Q., Yue, J. C., & Xia, H. M. (2026). Does artificial intelligence improve accessibility in cultural and heritage tourism? Evidence from a design-led review and the Inclusive Human–AI Mediation (IHAM) framework. *Information Technology & Tourism*, 28(1), 43.

Xu, Z., Liu, F., Xia, G., Duan, Y., & Yu, L. (2026). A scoping review of inclusive and adaptive human–AI interaction design for neurodivergent users. *Disability and Rehabilitation: Assistive Technology*, 21(4), 943-961.

Zhang, X., Antwi-Afari, M. F., Zhang, Y., & Xing, X. (2024). The impact of artificial intelligence on organizational justice and project performance: A systematic literature and science mapping review. *Buildings*, 14(1), 259.