



AI, Metaverse and Sustainability in the new organizational design



South
American
Publishing

AI, metaverse and sustainability in the new organizational design

Jesús Alejandro Vázquez Meza
Víctor Manuel Martínez García
Yennifer Díaz Romero
Manuel Iván Tostado Ramírez
Rogelio Estrada Lizárraga
José de Jesús Reyes Machain

Authors

Jesús Alejandro Vázquez Meza

Universidad Autónoma de Sinaloa, Unidad Académica
Preparatoria Rubén Jaramillo, Vicerrectoría, Sinaloa,
México
<https://orcid.org/0000-0001-9655-1576>
avazquezmeza@uas.edu.mx

Víctor Manuel Martínez García

Universidad Autónoma de Sinaloa, Facultad de Arquitectura
y Diseño, Vicerrectoría, Sinaloa, México
<https://orcid.org/0009-0008-2647-9386>
drvictormartinez@uas.edu.mx

Yennifer Díaz Romero

Universidad Autónoma de Sinaloa, Facultad de Ciencias de
la Educación, Vicerrectoría, Sinaloa, México
<https://orcid.org/0009-0009-7232-9749>
yenniferdr07@ms.uas.edu.mx

Manuel Iván Tostado Ramírez

Universidad Autónoma de Sinaloa, Facultad de Informática,
Vicerrectoría, Sinaloa, México
<https://orcid.org/0009-0008-4109-2438>
itostado@uas.edu.mx

Rogelio Estrada Lizárraga

Universidad Autónoma de Sinaloa, Facultad de Informática,
Vicerrectoría, Sinaloa, México
<https://orcid.org/0000-0001-9500-5960>
restrada@uas.edu.mx

José de Jesús Reyes Machain

Universidad Autónoma de Sinaloa, Facultad de Arquitectura
y Diseño, Vicerrectoría, Sinaloa, México
<https://orcid.org/0009-0000-6273-071X>
Jreyes_machain@uas.edu.mx

Copyright Page

© 2026. The authors. This is an open access book, distributed under the terms of a Creative Commons Attribution 4.0 license (<https://creativecommons.org/licenses/by/4.0/>) that allows use, distribution, and reproduction in any medium provided that the original work is properly cited.

This South American Publishing imprint is published by South American Publishing.

For more information, see South American Publishing's Open Access Policy: <https://southam.pub/policies.php>

ISBN (Spanish edition, eBook): 978-9915-9928-8-4

ISBN (Spanish edition, Softcover): 978-9915-9928-9-1

This title is available in both print (softcover) and digital (open access PDF) formats.

DOI: 10.62486/978-9915-9928-8-4

Cataloging Data

Cataloging-in-Publication Data (CIP):

South American Publishing

AI, metaverse and sustainability in the new organizational design / edited by Jesús Alejandro Vázquez Meza, Víctor Manuel Martínez García, Yennifer Díaz Romero, Manuel Iván Tostado Ramírez, Rogelio Estrada Lizárraga, and José de Jesús Reyes Machain. Montevideo: South American Publishing, 2026.

ISBN (Spanish edition, eBook): 978-9915-9928-8-4

ISBN (Spanish edition, Softcover): 978-9915-9928-9-1

THEMA classification codes:

UYQE

BISAC classification:

BUS103000 - BUSINESS & ECONOMICS / Organizational Development

BUS041000 - BUSINESS & ECONOMICS / Management

COM004000 - COMPUTERS / Artificial Intelligence / General

COM057000 - COMPUTERS / Virtual & Augmented Reality

Editorial Notice and Acknowledgments

The publication of this book is part of South American Publishing's mission to promote open, ethical, and rigorous scientific communication across all areas of knowledge.

All books published by South American Publishing undergo a double-blind peer-review process and a technical evaluation in accordance with the publisher's editorial policies, aligned with COPE guidelines.

South American Publishing acknowledges the valuable collaboration of authors, reviewers, designers, and production teams who made this publication possible.

Preface

Today, many organizations continue to operate according to structures designed for an environment that has already changed. The rapid advancement of generative artificial intelligence has made the gap between these structures and the current environment more apparent, as it is transforming how work is organized, who performs it, and which skills ultimately distinguish those who advance from those who fall behind. This premise forms the foundation of AI, Metaverse and Sustainability in the New Organizational Design, a volume in the Management Collection intended for academic and professional readers who are experiencing this transformation in the classroom, in talent management, or through organizational decision-making.

The central idea of the book is the distinction between implementing a technology and transforming an organization. Artificial intelligence and immersive environments enter organizations as tools, but the chapters demonstrate that their impact depends largely on how the organization as a whole is redesigned and the pace at which its members adopt each innovation. From this perspective, integrating generative artificial intelligence into an organization is a comprehensive organizational design initiative rather than merely a technology project.

To develop this argument, the book brings classical frameworks of organizational design and technology adoption into dialogue with recent evidence from both international settings and the Mexican university context, where research has documented that university students and recent graduates have already incorporated artificial intelligence into their everyday intellectual work.

Another defining feature of the book is the caution with which it examines the evidence. Alongside the opportunities created by human-AI collaboration, hybrid work, and immersive environments, the chapters explore the risks associated with this transformation, including technological dependence, algorithmic bias, human vulnerabilities in cybersecurity, and the absence of clear governance mechanisms. They also acknowledge that technologies such as the metaverse are still, in many respects, under development.

Sustainability, as used in the title, refers to an organization's ability to sustain its transformation over time. Throughout the chapters, this capacity is linked to continuous investment in human capabilities and lifelong learning, the well-being of those working alongside artificial intelligence systems, the development of a responsible digital culture, and, above all, trust and legitimacy. As the final chapter argues, no artificial intelligence system, regardless of its technical sophistication, can remain sustainable in the long term without the trust of those who use it or are affected by it.

Rather than offering a manual of ready-made answers, the book seeks to provide a space for critical analysis for those who study organizations and those responsible for leading them through this transformation. Its pages raise questions that organizations can no longer afford to postpone: Which tasks will disappear, and which will be transformed? How can new talent be developed when entry-level positions are declining? What does it mean to belong to a team when geography no longer defines its boundaries? And who makes the decisions when technology itself becomes a decision-maker? The intention is for each reader to find in these questions a starting point for reflecting on their own organizational reality.

Introduction

The arrival of generative artificial intelligence in organizations is often presented as the adoption of yet another tool. However, the evidence examined in this book shows that its effects extend considerably further, reshaping the organization of work and the skills it requires. Employers surveyed in the Future of Jobs Report 2025 anticipate that approximately 39% of core job skills will change by 2030 (World Economic Forum, 2025). A survey by McKinsey & Company (2025) found that 88% of organizations already use artificial intelligence in at least one business function, while only 6% achieve what the consulting firm defines as “high performance.” The gap between these figures illustrates that moving from technology adoption to organizational transformation is one of the central challenges currently facing organizational design.

This book examines how artificial intelligence and immersive environments, including the metaverse, are transforming organizational design. It explores their effects on decision-making, human-system interaction, talent management, workspaces, and organizational learning, as well as the ethical, security, and governance conditions required to sustain this transformation over time.

From a theoretical perspective, the book draws on two complementary approaches. The first, rooted in organizational design, is Galbraith’s Star Model (2014), which proposes that organizational performance depends on the alignment of strategy, structure, processes, rewards, and people. From this perspective, no organization can adequately respond to the disruption caused by artificial intelligence by adjusting only its technology or organizational structure. The second approach is Rogers’ (2003) diffusion of innovations theory, which explains adoption as a social process that unfolds in successive waves, with innovators leading the way and late adopters requiring more time and evidence. Today, however, the duration of these waves has shortened dramatically. Davis’ (1989) Technology Acceptance Model also appears throughout the book, providing a framework for understanding why individuals adopt or reject a technology based on its perceived usefulness and ease of use.

The chapters bring together international reports from organizations such as the World Economic Forum, Deloitte, McKinsey & Company, Stanford University, IBM, and LinkedIn, alongside academic research on human-AI collaboration, the metaverse, ethics, and cybersecurity. These sources are complemented by studies conducted within the Mexican university context, including research undertaken by Cerezo Arreola (2025) at the Autonomous University of Sinaloa involving university students and recent graduates. This combination seeks to examine at a more immediate level the transformations described by global data, since it is in settings such as classrooms and entry-level employment that these changes become tangible.

The book is organized into nine chapters. The first, *The Organization in a New Era of Transformation*, establishes the overarching conceptual framework. It integrates Galbraith’s (2014) model and Rogers’ (2003) theory with global evidence on the labor market and the circumstances of a generation entering the workforce at a time when artificial intelligence has become the norm. The chapter addresses technological acceleration, the transition from hierarchical structures to organizational ecosystems, and transformation as a cultural phenomenon. It concludes that no investment in technological infrastructure can replace the slower, fundamentally human process of developing an organizational culture capable of using technology responsibly.

The second chapter, *Artificial Intelligence: From Technological Tool to Organizational Capability*, begins by distinguishing between possessing data and knowing how to make decisions based on them. It then examines process automation, the differences between traditional and generative artificial intelligence, and the distinction between automating a task and augmenting the capabilities of the person performing it. Building on these foundations, the chapter introduces AI as a cognitive copilot and the concept of the

augmented organization, in which the deliberate redesign of workflows carries greater significance than the number of technological tools deployed. It concludes by arguing that artificial intelligence is moving beyond being a tool that organizations acquire and is becoming a system of capabilities that must be developed, maintained, and actively governed.

The third chapter, *The Human-AI Organization: New Forms of Collaboration*, draws on the work of Frey and Osborne (2017) and Acemoglu and Restrepo (2018) to distinguish between tasks that disappear and those that are transformed. It also examines emerging forms of productivity, the concept of the augmented worker, and the risks of technological dependence, including the cognitive offloading documented by Gerlich (2025). Drawing on the jagged frontier described by Mollick (2024) and the experimental evidence presented by Dell'Acqua et al. (2023), the chapter compares human and artificial intelligence. It further explores shared decision-making, human oversight, emerging organizational roles, and the risk of eroding creativity, critical thinking, and professional judgment. The chapter concludes by examining the transition from workers who use AI to workers who collaborate with it.

The fourth chapter, *AI, Talent, and the Hybrid Organizational Workspace*, examines artificial intelligence as a new component of human capital, drawing on the concept of hybrid human capital proposed by Cerezo Arreola (2025). It considers the implications for professional competencies, recruitment, continuous learning, university-industry relationships, and emerging professional profiles. The second part addresses the dissolution of traditional workplace boundaries, the distinctions between remote, hybrid, and distributed work, and the integration of physical, digital, and immersive spaces. It concludes by examining what happens to employees' sense of belonging when physical proximity is no longer a defining feature of organizational life, highlighting the need to deliberately develop new shared rituals.

The fifth chapter, *The Metaverse as a New Space for Work and Collaboration*, begins by establishing a rigorous definition of the concept, drawing on Ball (2022), Lee et al. (2022), and Feng et al. (2024). It examines immersive and persistent environments, the role of avatars in presence and representation, virtual collaboration, and the possibility of conducting an entire project within a shared digital space. The chapter concludes with a balanced assessment of the opportunities and limitations of these environments, acknowledging that the technology remains under development in many respects and that its promise of greater presence and reduced nonverbal communication demands does not consistently translate into better experiences in everyday practice.

The sixth chapter, *Learning, Innovation, and Organizational Experience in Immersive Environments*, examines immersive training, simulation, competency development, collaborative innovation, and employee and customer experiences. Its analysis draws on recent systematic reviews and the case of Accenture's Nth Floor platform, as documented by Kohli (2025). The final section, like that of the fourth chapter, addresses a challenge that technology alone cannot resolve: building communities in environments that do not depend on physical proximity. This task requires deliberate attention to belonging, trust, and shared rituals.

With the seventh chapter, *Artificial Intelligence Ethics and Organizational Responsibility*, the book turns to the conditions necessary for responsible organizational transformation. The discussion moves from transparency and accountability to employee privacy and well-being, addressing fairness and non-discrimination, algorithmic bias, and human oversight. One of its principal references is UNESCO's Recommendation on the Ethics of Artificial Intelligence (2021). The chapter concludes by arguing that, ultimately, the ethics of artificial intelligence comes down to a concrete question: whether the people working within an organization are better or worse off as a consequence of the decisions made regarding the use of technology.

The eighth chapter, *Cybersecurity, Trust, and Digital Culture*, moves beyond understanding security as an exclusively technical concern and approaches it as an organizational challenge, placing the human factor

at the center of the analysis, drawing on authors such as Da Veiga and Eloff (2010) and Mitnick and Simon (2002). It explores cybersecurity culture, social engineering, whose practices are being reshaped by generative artificial intelligence, and data protection. The final section examines the security challenges specific to hybrid and distributed organizations, where protection can no longer depend solely on perimeter-based defenses but increasingly relies on the behavior of every connected individual.

The ninth and final chapter, Governance of Transformation: Who Decides When Technology Makes Decisions?, begins by asking who governs the algorithms that are increasingly influencing decisions previously made by humans. In addressing this question, the chapter examines algorithmic governance, accountability for automated decisions, regulation, auditing, data protection, ethics committees, and employee participation. Its regulatory references include Regulation (EU) 2024/1689 and the OECD Recommendation. The chapter concludes by focusing on trust and legitimacy, where the different elements discussed throughout the chapter converge.

Following this examination, the book proposes understanding organizational transformation as an ongoing process that must be developed and sustained over time, rather than as an event completed once a technology has been adopted. For organizations, the challenge lies in undertaking comprehensive redesign, investing consistently in human capabilities, and developing the culture, security practices, and governance mechanisms necessary to maintain the trust of their members. The ultimate objective is to ensure that adopting artificial intelligence and immersive environments does not create risks greater than the problems these technologies are intended to solve.

Abstract

Generative artificial intelligence and immersive environments, including the metaverse, are transforming organizational design. The adoption of these technologies goes beyond the introduction of additional tools and constitutes a comprehensive organizational design process that reshapes how work is organized, the skills it requires, and the ethical, security, and governance conditions necessary to sustain transformation over time. This perspective draws on two complementary theoretical frameworks: Galbraith's Star Model, which links organizational performance to the alignment of strategy, structure, processes, rewards, and people, and Rogers' diffusion of innovations theory, which explains technological adoption as a social process that unfolds in successive waves. Through nine chapters, the analysis explores the effects of these technologies on decision-making, human-AI collaboration, talent management, workspaces, and organizational learning, while examining associated risks, including technological dependence, algorithmic bias, human vulnerabilities in cybersecurity, and the absence of clear governance mechanisms. Organizational transformation is therefore understood as an ongoing process that must be developed and sustained over time through comprehensive organizational redesign, continuous investment in human capabilities, and the establishment of a responsible organizational culture, robust security practices, and effective governance mechanisms capable of preserving the trust of organizational members.

Keywords: artificial intelligence; metaverse; organizational design; digital transformation; governance; cybersecurity

Index

Chapter 01. The organization in a new era of transformation | La organización ante una nueva era de transformación 1

Chapter 02. Inteligencia artificial: de herramienta tecnológica a capacidad organizacional Artificial intelligence: from technological tool to organizational capacity 10

Chapter 03. The human-AI organization: new forms of collaboration | La organización humano-IA: nuevas formas de colaboración 17

Chapter 04. AI, talent, and hybrid organizational space | IA, talento y espacio organizacional híbrido 25

Chapter 05. The metaverse as a new workspace and collaboration | El metaverso como nuevo espacio de trabajo y colaboración 33

Chapter 06. Learning, innovation and organizational experience in immersive environments | Aprendizaje, innovación y experiencia organizacional en entornos inmersivos 38

Chapter 07. Artificial intelligence ethics and organizational responsibility | Ética de la inteligencia artificial y responsabilidad organizacional..... 43

Chapter 08. Cybersecurity, trust and digital culture | Ciberseguridad, confianza y cultura digital.. 49

Chapter 09. Transformation governance: who decides when the technology decides | Gobernanza de la transformación: quién decide cuando decide la tecnología 54

References | Referencias 60

Chapter 1 / Capítulo 1



AI, metaverse and sustainability in the new
organizational design
ISBN: 978-9915-9928-8-4
DOI: 10.62486/978-9915-9928-8-4.Ch01
Pages: 1-9

© 2026 The authors. This is an open access
book chapter distributed under the terms of
the Creative Commons Attribution (CC BY)
4.0 License.

The organization in a new era of transformation

La organización ante una nueva era de transformación

Jesús Alejandro Vázquez Meza¹  

Víctor Manuel Martínez García²  

Yennifer Díaz Romero³  

Manuel Iván Tostado Ramírez⁴  

Rogelio Estrada Lizárraga⁴  

José de Jesús Reyes Machain²  

1 Universidad Autónoma de Sinaloa, Unidad Académica Preparatoria Rubén Jaramillo, Vicerrectoría, Sinaloa, México

2 Universidad Autónoma de Sinaloa, Facultad de Arquitectura y Diseño, Vicerrectoría, Sinaloa, México

3 Universidad Autónoma de Sinaloa, Facultad de Ciencias de la Educación, Vicerrectoría, Sinaloa, México

4 Universidad Autónoma de Sinaloa, Facultad de Informática, Vicerrectoría, Sinaloa, México

5 Universidad Autónoma de Sinaloa, Facultad de Informática, Vicerrectoría, Sinaloa, México

ABSTRACT

The chapter examines why adopting artificial intelligence requires organizational redesign rather than merely purchasing technological tools. It connects Galbraith's Star Model, which emphasizes alignment among strategy, structure, processes, rewards, and people, with Rogers's diffusion of innovations, which explains different rates of adoption. International reports on changing employment and skills are considered alongside studies of university students and new professionals. The discussion distinguishes technical familiarity from critical thinking and institutional culture. It also addresses accelerating technological change, pressures on organizational rules, and the shift from rigid hierarchies toward collaborative networks. Rapid adoption can create difficulties when responsible practices and supervision do not develop at the same pace. The main argument is that adaptation depends on coordinated changes in decisions, incentives, learning, and everyday behavior. Organizational transformation is therefore presented as a sustained cultural and structural process. The chapter explores implications for professional development, oversight of AI systems, and the allocation of responsibility, while distinguishing the tendencies described in the reviewed sources from the decisions that individual organizations must make.

Keywords: Organizational design; Digital transformation; Artificial intelligence; Organizational culture; Technology adoption

RESUMEN

El capítulo analiza por qué la incorporación de inteligencia artificial exige revisar el diseño organizacional y no solamente adquirir herramientas tecnológicas. Articula el Modelo de la Estrella de Galbraith, centrado en la coherencia entre estrategia, estructura, procesos, recompensas y personas, con la difusión de innovaciones de Rogers, que explica los distintos ritmos de adopción. Examina informes internacionales sobre transformaciones del empleo y las competencias, junto con estudios sobre universitarios y nuevos profesionales. A partir de esa discusión, distingue entre conocimiento técnico, pensamiento crítico y cultura institucional. Revisa la aceleración de las tecnologías digitales, las tensiones que crea para la actualización normativa y la transición desde jerarquías rígidas hacia redes de colaboración. También aborda el riesgo de que una adopción rápida no vaya acompañada de prácticas responsables. El argumento central es que la capacidad de adaptación depende de cambios coordinados en las decisiones, los incentivos, el aprendizaje y el comportamiento cotidiano. La transformación se presenta como un proceso organizacional y cultural sostenido, con implicaciones para la formación de personas, la supervisión del uso de sistemas y la distribución de responsabilidades. El análisis diferencia las tendencias descritas por las fuentes de las decisiones específicas que deberá adoptar cada institución. La revisión contrapone adopción individual e integración institucional y explica por qué sus ritmos pueden diferir. De este modo, destaca la necesidad de establecer prioridades, acompañamiento y criterios compartidos para observar si los cambios tecnológicos se traducen en prácticas coherentes.

Palabras clave: Diseño organizacional; Transformación digital; Inteligencia artificial; Cultura organizacional; Adopción tecnológica

• DISEÑO ORGANIZACIONAL Y ADOPCIÓN TECNOLÓGICA ANTE LA IRRUPCIÓN DE LA INTELIGENCIA ARTIFICIAL

En la actualidad, buena parte de las organizaciones siguen trabajando con un diseño que fue pensado para un entorno que ya cambió, y es poco común que se detengan a medir qué tanto se han alejado de él, esto hasta que el rezago ya no se puede pasar por alto. Cuando se habla de la inteligencia artificial (IA) generativa, su irrupción no representa un ajuste más en las herramientas de trabajo; con ella cambia de raíz la manera en que se organiza el trabajo y quién lo realiza, así como las habilidades que terminan por separar a quien prospera de quien se queda atrás.

Para tratar de entender de mejor forma este cambio, se retoman dos perspectivas que se complementan. Una de ellas tiene que ver con la forma en que una organización debe modificar su diseño para no quedar obsoleta. En la otra se analiza cómo una innovación se difunde dentro de un sistema social hasta volverse norma. Si bien, el enfoque de cada uno de los dos tienen orientaciones diferentes, es claro que ambos cuentan con evidencia reciente y su destino es un cambio estructural que requiere atención sostenida, es decir, que una respuesta no se puede resolver de una sola vez.

Con respecto a la primera de ellas, Jay Galbraith dedicó buena parte de su carrera a insistir en que ninguna organización cambia de verdad si solo modifica su organigrama, idea que sigue vigente décadas después. Su modelo, llamado Star Model o Modelo de la Estrella, establece que el desempeño depende de la alineación de cinco elementos que se refuerzan uno a otro, los cuales son la estrategia, la estructura, los procesos, las recompensas y las personas (Galbraith, 2014); en dicho modelo, la estrategia es la que define hacia dónde se dirige la organización, en la estructura se determina dónde reside el poder de decisión y los procesos regulan la manera en que circula la información, mientras que los sistemas de recompensa orientan la motivación, siendo las políticas relacionadas con las personas las que construyen las capacidades necesarias para que todo lo anterior funcione.

Lo que mantiene vigente este modelo cuando se habla de IA es la frecuencia en que los cinco elementos se muevan de manera conjunta. Es muy probable que una organización incorpore la inteligencia artificial en sus procesos y que, aun así, se quede a medio camino, esto cuando no hay una modificación real a su estructura, sus incentivos o el perfil de quienes la integran; cabe destacar que Galbraith ya había identificado ese mismo error décadas atrás, en otras oleadas de transformación, en donde se rediseñaba el organigrama sin tocar lo demás y después se preguntaba por qué nadie había cambiado de comportamiento. Con lo anterior se puede ver que el incorporar IA generativa a una organización es un proyecto de diseño organizacional completo, más que un proyecto de tecnología.

• LA DIFUSIÓN DE LA INNOVACIÓN COMO PROCESO SOCIAL

El enfoque de la difusión de innovaciones contradice una idea muy común, la de que adoptar una tecnología en una organización es un evento único, es decir, un antes y un después. Desde esta perspectiva, la adopción se entiende más bien como un proceso gradual, mismo que avanza con ritmos distintos según el grupo del que se trate y que presenta fricciones que se pueden prever una vez que se conocen. Es así como el saber en qué punto de la curva se encuentra cada equipo permite trabajar de manera diferenciada con cada uno de ellos, ya sea acompañando a los innovadores sin sofocarlos, aprovechando a los adoptantes tempranos como agentes de cambio interno o anticipando, sin frustración, la resistencia de quienes necesitan más tiempo y más pruebas.

1.3 LA MAGNITUD DEL CAMBIO: LO QUE DICE LA EVIDENCIA GLOBAL

Los marcos anteriores no se quedan en el plano teórico. Existen datos recientes que confirman la escala real del fenómeno, mismos que a continuación se presentan para tratar de dimensionarlo de mejor forma. El primer hallazgo se encuentra el *Future of Jobs Report 2025*, publicado por el Foro Económico Mundial con información recabada de más de 1000 empleadores que representan a 14 millones de trabajadores en 55 economías. Se trata de uno de los diagnósticos más completos con los que se cuenta hoy en día sobre el rumbo del mercado laboral (World Economic Forum, 2025). De acuerdo con lo encontrado, se prevé que hacia 2030 se creen alrededor de 170 millones de empleos nuevos, mientras que 92 millones de puestos

actuales desaparecerán, lo que representa un crecimiento neto cercano a los 78 millones de empleos a nivel global (World Economic Forum, 2025).

Ahora bien, detrás de esa cifra existe una disrupción digna de un análisis, esto porque los empleadores anticipan que cerca del 39 % de las competencias laborales clave cambiará antes de 2030 (World Economic Forum, 2025). Dentro de las habilidades tecnológicas, el informe da a conocer que la inteligencia artificial y el análisis masivo de datos como las de mayor proyección de crecimiento, incluso por encima de la ciberseguridad. Por su parte, habilidades humanas como el pensamiento creativo, la resiliencia y la curiosidad se mantienen como un complemento indispensable, es decir, no se les ve como un vestigio del pasado (World Economic Forum, 2025). Con lo anterior se puede ver que resulta demasiado simple la idea de que la disrupción tecnológica está eliminando lo humano del trabajo, cuando lo que ocurre es que lo está redefiniendo.

Por ello no sorprende que el 85 % de los empleadores encuestados declare que planea invertir en capacitación y actualización de competencias en los próximos años (World Economic Forum, 2025). Este dato confirma a nivel global lo que Galbraith (2014) anticipaba desde el diseño organizacional, en el sentido de que sin una inversión pensada en el personal ningún cambio estructural o tecnológico puede dar los frutos necesarios.

En referencia a quienes tienen la misión de tomar las decisiones dentro de las organizaciones, Deloitte (2025) elaboró su informe *Global Human Capital Trends* a partir de encuestas aplicadas a casi 10 000 líderes de talento y negocio en 93 países, en donde se encontró que apenas el 19 % de los ejecutivos considera que los modelos tradicionales de organización del trabajo siguen siendo los más adecuados para generar valor, mientras que el 85 % coincide en que sus organizaciones necesitan formas más ágiles de organizarse para no quedarse atrás. Se trata de un patrón muy similar al que documenta el Foro Económico Mundial (World Economic Forum, 2025), solo que visto desde quienes deciden el rediseño organizacional: casi todos coinciden en que hace falta cambiar, pero muy pocos consideran que la estructura actual de su organización ya está lista para hacerlo.

1.4 LA GENERACIÓN QUE ENTRA AL MERCADO LABORAL CON LA IA COMO NORMA

Cuando los datos globales se observan de cerca, adquieren otro sentido. En el contexto local, la investigación de Cerezo Arreola (2025), realizada en la Universidad Autónoma de Sinaloa, analiza cómo los universitarios y recién egresados viven directamente la transición hacia una economía atravesada por la inteligencia artificial generativa. El punto de partida de dicho estudio es la coincidencia con lo planteado por Rogers (2003): esta generación ha ido más allá de probar la IA y la ha incorporado como un componente cotidiano de su trabajo intelectual (Cerezo Arreola, 2025), es decir, pasó de usarla como herramienta complementaria a tenerla como un hábito instalado.

Desde la teoría de la difusión de innovaciones, el mismo estudio describe a los jóvenes universitarios como adoptantes tempranos casi por naturaleza, dado que su predisposición al riesgo y su conectividad social los colocan a la vanguardia de la curva de adopción (Cerezo Arreola, 2025). Sin embargo, adoptar desde temprano también tiene repercusiones. La investigación demuestra un efecto dual que cualquier organización tendría que tomar en cuenta: por un lado, se observa un aumento real en la eficiencia y en la capacidad de procesar información compleja; por otro, se corre el riesgo de que se debiliten habilidades críticas como la redacción original, la síntesis profunda y el pensamiento analítico cuando el esfuerzo cognitivo se delega de forma sistemática a la máquina (Cerezo Arreola, 2025).

Este segundo riesgo también es confirmado por Gerlich (2025) con evidencia empírica independiente. En su estudio con 666 participantes de distintas edades se encontró que los más jóvenes muestran, al mismo tiempo, mayor dependencia de las herramientas de IA y puntuaciones más bajas en pruebas de pensamiento crítico que los participantes de mayor edad; es el mismo patrón que documenta Cerezo Arreola (2025) en el contexto universitario mexicano, aunque medido con instrumentos distintos.

Cabe destacar que se trata de la generación que hoy entra al mercado laboral y que las organizaciones tendrán que formar y supervisar durante las próximas décadas. Sus integrantes poseen una fluidez técnica notable en el uso de las herramientas tecnológicas, pero también presentan un riesgo documentado de haber delegado, en el camino, parte de las habilidades analíticas que antes se desarrollaban al no tener esa herramienta a la mano.

Lo anterior se relaciona de manera directa con la evidencia laboral que reporta el estudio de Cerezo Arreola (2025), de acuerdo con la cual dominar la inteligencia artificial dejó de ser una

ventaja competitiva y pasó a ser un requisito para entrar al mercado laboral de nivel inicial. Al momento de hablar de un candidato competitivo, el simple uso de la herramienta no basta; lo que lo distingue es la capacidad de supervisarla, refinarla y aplicarla a problemas complejos, combinándola con juicio ético, creatividad y comunicación, es decir, con aquello que la máquina no logra replicar (Cerezo Arreola, 2025). Con base en ello, la investigación deduce que el nuevo capital humano ya no es puramente técnico ni puramente humano, sino híbrido.

1.5 INTEGRANDO LAS PIEZAS: LO QUE ESTO SIGNIFICA PARA EL DISEÑO ORGANIZACIONAL

Una vez descritos estos cuatro elementos de evidencia, se puede ver que, en conjunto, transmiten el mismo mensaje desde ángulos distintos, tal como se resume a continuación:

a) Partiendo del modelo de Galbraith (2014) se puede entender que ninguna organización absorbe la disrupción de la inteligencia artificial ajustando solo su tecnología o su estructura, esto porque necesita repensar al mismo tiempo su estrategia, sus procesos, sus incentivos y sus políticas de talento.

b) La difusión de innovaciones de Rogers (2003) permite ver que ese cambio no ocurre de golpe, pues avanza en oleadas, con innovadores que abren camino y rezagados que necesitan más tiempo.

c) Con las cifras del Foro Económico Mundial (World Economic Forum, 2025) se confirma que la transformación no es preocupación de unas cuantas empresas visionarias, puesto que un gran número de organizaciones ya viven este proceso, y que la creación y el desplazamiento de empleo que se proyectan exigen una respuesta estratégica y no reactiva.

d) Por último, todo lo anterior apunta a la generación que hoy entra al mundo profesional, tal como lo documenta Cerezo Arreola (2025): personas que ya usan la IA de manera regular, que están en un mercado que exige su dominio y cuyo valor diferencial dependerá de que logren combinar esa competencia técnica con aquello que ninguna máquina puede sustituir.

Con lo anterior se puede deducir que la organización ante esta nueva era de transformación, más que ante una simple decisión entre adoptar o resistir la inteligencia artificial, se encuentra ante una tarea más exigente, la cual consiste en rediseñarse de manera integral, es decir, en su estructura, sus procesos, sus incentivos y sus personas, y al mismo tiempo acompañar un proceso de adopción que, como toda difusión de innovaciones, tomará tiempo, generará resistencias legítimas y exigirá una inversión sostenida en las capacidades humanas que seguirán marcando la diferencia.

La evidencia empírica más reciente confirma que dicho rediseño integral es una condición central del éxito organizacional, más que una opción entre varias. Tal es el caso de la encuesta global *State of AI 2025* de McKinsey, aplicada a casi 2000 participantes en 105 países, en donde se encontró que el 88 % de las organizaciones ya usa inteligencia artificial en al menos una función del negocio, lo que representa un salto considerable con respecto al año anterior (McKinsey & Company, 2025). No obstante, apenas el 6 % logra lo que la propia consultora denomina “alto desempeño”, entendido como un impacto superior al 5 % en utilidades atribuible directamente a la IA (McKinsey & Company, 2025).

Cabe destacar que la tecnología que utilizan unas y otras organizaciones con frecuencia es prácticamente la misma, por lo que la diferencia entre ese 6 % y el resto tiene que ver con otros aspectos. Las organizaciones de alto desempeño son 3.6 veces más propensas a emprender un cambio transformacional, y más de la mitad de ellas rediseña sus flujos de trabajo desde cero, frente a apenas una quinta parte del resto (McKinsey & Company, 2025). En cifras, se trata del mismo argumento de Galbraith (2014), ya que sin un rediseño estructural profundo la tecnología por sí sola no produce ventaja competitiva, y lo que se obtiene es lo que el propio reporte denomina “purgatorio del piloto”, en donde las organizaciones experimentan indefinidamente sin escalar nunca hacia un cambio real.

1.6 LA ACELERACIÓN TECNOLÓGICA

Para caer en cuenta qué tan distinto es el ritmo actual de cambio, basta con hacer una comparación del tiempo que tardaron las tecnologías anteriores en llegar a 100 millones de usuarios con el que requirió la más reciente. Mientras que Instagram necesitó dos años y medio y TikTok nueve meses, ChatGPT alcanzó esa misma cifra en apenas dos meses desde su

lanzamiento, ritmo que analistas del banco UBS describieron como algo nunca antes visto en dos décadas de seguimiento del sector de internet (Hu, 2023). Dicho dato, que podría tomarse como simplemente un número, muestra con claridad que la velocidad misma de la disrupción tecnológica cambió de naturaleza.

Este fenómeno pone a prueba el marco con el que se ha explicado la adopción de innovaciones. Rogers (2003) describió la difusión como un proceso que avanza por oleadas (innovadores, adoptantes tempranos, mayoría temprana, mayoría tardía y rezagados) y que recorrer un sistema social completo toma tiempo; si bien ese proceso sigue ocurriendo, lo que se comprimió drásticamente es la duración de cada oleada, ya que aquello que antes tomaba una generación entera, como fue el caso del teléfono o de la televisión antes de volverse casi universales, hoy ocurre en el lapso de unos cuantos ciclos escolares. Las categorías de Rogers siguen siendo útiles para entender quién se mueve primero y quién se resiste, aunque la escala de tiempo con la que se mide dicho proceso ya no es la misma.

Esta aceleración tampoco es exclusiva del consumo digital, puesto que atraviesa directamente el mundo del trabajo, tal como lo muestran los datos del mercado laboral. El *Future of Jobs Report 2025* identifica el cambio tecnológico como una de las fuerzas de mayor peso que enfrentan las organizaciones de aquí a 2030, además de estimar que cerca del 39 % de las competencias laborales clave cambiará en ese periodo (World Economic Forum, 2025). Dicha cifra describe una situación poco común en la historia económica reciente, ya que una generación entera de profesionales tendrá que renovar buena parte de sus habilidades a la mitad de su carrera y no al final de ella.

El *AI Index Report* de Stanford (2026) confirma, por su parte, que esta aceleración sigue intensificándose. Dicho informe evidencia que la adopción organizacional de inteligencia artificial llegó al 88 % a nivel global y que cuatro de cada cinco estudiantes universitarios ya utiliza IA generativa de forma cotidiana, cifras que hace apenas tres años se habrían visto como una proyección optimista de largo plazo y que hoy describen el presente.

Así mismo, en el informe se reporta un dato técnico que ilustra la velocidad del avance, ya que en una prueba de referencia exigente para evaluar la capacidad de programación, el desempeño de los sistemas líderes pasó de resolver correctamente alrededor del 60 % de las tareas a casi el 100 % en apenas un año. Son pocas las tecnologías que en la historia reciente han mejorado su capacidad técnica central a ese ritmo y en un periodo tan corto.

En el caso particular de México, el contexto universitario muestra el mismo patrón, aunque a una escala más cercana. Cerezo Arreola (2025) documenta cómo la inteligencia artificial generativa pasó, en apenas un lustro, de ser una herramienta de apoyo institucional casi anecdótica a convertirse en un componente cotidiano del trabajo intelectual de estudiantes y egresados. El propio estudio ubica ese salto entre 2020 y 2025, es decir, en cinco años, periodo en el que una tecnología pasó de la periferia al centro de la vida académica de una generación completa (Cerezo Arreola, 2025). Es el mismo patrón de compresión temporal que se observa en las cifras de adopción de ChatGPT, solo que en este caso se mide en hábitos de estudio y no en número de usuarios.

Esta velocidad trae consigo una consecuencia a la que pocas veces se le da la importancia debida, y que se relaciona con la brecha entre la rapidez con la que se adopta una tecnología y la rapidez con la que se desarrolla la cultura necesaria para usarla con responsabilidad. Ruiz Sevilla (2026), en su investigación sobre cultura de ciberseguridad entre estudiantes de informática, lo plantea de manera directa al señalar que la apertura tecnológica de las instituciones educativas avanzó a un ritmo considerablemente más rápido que el desarrollo de una cultura de seguridad digital entre quienes la usan todos los días.

De acuerdo con dicha investigación, los estudiantes más avanzados académicamente identifican con precisión un intento de *phishing* y aun así desactivan de manera deliberada sus propios mecanismos de protección cuando la seguridad les resulta incómoda para su trabajo técnico (Ruiz Sevilla, 2026). Es así como se puede ver que el conocimiento técnico avanzó con rapidez, mientras que los hábitos de protección lo hicieron a un ritmo mucho más lento.

El mismo desfase se repite a escala organizacional. Ríos Navarrete (2025) menciona que los cambios tecnológicos constantes exigen una colaboración estrecha entre gobiernos, fuerza laboral y desarrolladores para regular a tiempo el uso de la inteligencia artificial, esto debido a que la velocidad de adopción superó a la de la regulación. De igual forma, cuando una empresa delega tareas relevantes a un sistema entrenado con grandes volúmenes de datos sin haber definido antes los marcos éticos y legales que deben regirlas, en lugar de gestionar la aceleración, se ve arrastrada por ella (Ríos Navarrete, 2025).

Con base en lo revisado en este apartado, el ritmo de adopción en el consumo masivo,

las proyecciones del mercado laboral global, el caso universitario mexicano y la experiencia de ciberseguridad en un campus local, aunque son cuatro ángulos distintos, coinciden en que la aceleración tecnológica ya no da tiempo para que las estructuras, las normas y los hábitos maduren al mismo ritmo que la herramienta que los provoca. Esta es la razón por lo que las organizaciones que hoy viven esta nueva era de transformación deben decidir qué tecnología adoptar y, sobre todo, cómo construir a la misma velocidad la cultura, la regulación y las capacidades humanas necesarias para que esa adopción no termine siendo un riesgo mayor que el problema que buscaba resolver.

1.7 DE ESTRUCTURAS JERÁRQUICAS A ECOSISTEMAS ORGANIZACIONALES

Durante buena parte del siglo pasado, cuando se hablaba de una organización se pensaba casi de manera automática en un organigrama, es decir, en una pirámide con niveles claros de autoridad en donde la información subía y las decisiones bajaban. Dicho diseño funcionó razonablemente bien mientras el entorno fue estable, con cambios predecibles y decisiones que podían esperar a recorrer la cadena de mando. Sin embargo, ese entorno ya no existe, y las organizaciones que insisten en operar como pirámides rígidas dentro de un ecosistema tecnológico interconectado empiezan a mostrar limitaciones evidentes.

Galbraith (2014) ya lo advertía desde el diseño organizacional clásico: la estructura es solo uno de los cinco elementos que determinan el comportamiento de una organización, y ninguno de ellos, ni siquiera la estructura, puede sostenerse aislado. En su modelo, la estrategia, la estructura, los procesos, las recompensas y las personas deben moverse como un sistema interdependiente, por lo que no pueden ajustarse uno por uno como si fueran piezas separadas; cabe destacar que esta idea, formulada originalmente para explicar por qué fracasaban las reorganizaciones tradicionales, hoy resulta todavía más reveladora, puesto que describe un ecosistema interno en donde cada elemento afecta a los demás, en lugar de una cadena de mando lineal en la que basta con rediseñar el nivel superior para que el resto se ajuste por sí solo.

El paso de la jerarquía al ecosistema tiene, además, una base empírica concreta en la manera en que operan hoy las instituciones digitalizadas, por lo que no se trata solo de una metáfora conveniente. Ruiz Sevilla (2026) describe con precisión el fenómeno al analizar cómo las instituciones de educación superior se transformaron en lo que él mismo denomina “ecosistemas altamente digitalizados”, en donde miles de dispositivos personales de estudiantes, bajo esquemas de conexión abierta conocidos como *Bring Your Own Device*, se conectan simultáneamente a redes institucionales, comparten información y acceden a plataformas en la nube sin que exista ya un centro único de control capaz de regular cada interacción. En este modelo de red distribuida, el control centralizado cede terreno frente a la conectividad masiva, lo que se acerca a la definición misma de un ecosistema, es decir, múltiples nodos interdependientes sin una autoridad única que dicte cada movimiento.

Con base en la teoría de Rogers (2003) se puede explicar por qué este tránsito resulta casi inevitable, ya que en ella la adopción tecnológica se describe como un proceso que se propaga a través de redes interpersonales, líderes de opinión y conexiones horizontales entre pares, más que como una decisión que desciende desde la alta dirección. Esto quiere decir que, incluso dentro de una organización jerárquica, la difusión de una tecnología ya se comportaba desde hace décadas más como un fenómeno de red que como una orden vertical. Lo que cambió no fue tanto el mecanismo de difusión como la estructura formal de la organización, que por fin empieza a parecerse a la manera en que la información siempre se movió en la práctica.

En el ámbito educativo se tiene el caso de “Mi Diario Digital de Aprendizaje”, documentado por Robles Ornelas y Páez López (2026), el cual ofrece un ejemplo aplicado de esa misma lógica. El proyecto surgió a partir de que quienes estaban más cerca del problema lo identificaron directamente, sin que mediara una directriz institucional descendente, esto siguiendo una metodología de Design Thinking que parte de observar necesidades reales antes de imponer soluciones prediseñadas (Robles Ornelas & Páez López, 2026). Aunque se trata de un caso a pequeña escala, muestra cómo la innovación contemporánea tiende a surgir de los nodos periféricos de un sistema, no exclusivamente de su nivel superior, lo cual se ajusta con mayor naturalidad a la lógica de un ecosistema que a la de una pirámide de autoridad.

En la misma dirección apuntan los datos del mercado laboral global. De acuerdo con el *Future of Jobs Report 2025*, las organizaciones que buscan sobrevivir a la próxima década apuestan por la capacitación continua, puesto que el 85 % de los empleadores encuestados planea invertir en ella (World Economic Forum, 2025). Lo anterior puede interpretarse como una

señal de que las estructuras rígidas y los roles fijos ya no logran absorber la velocidad del cambio. Una pirámide jerárquica clásica define roles fijos y rutas de ascenso predecibles; en cambio, un ecosistema organizacional exige que las personas se muevan lateralmente, que las competencias se actualicen constantemente y que la autoridad para decidir se distribuya más cerca del lugar en donde ocurre el problema.

Con respecto al ámbito empresarial, Cusumano et al. (2019) documentan cómo esta lógica de ecosistema dejó de ser una tendencia incipiente y se convirtió en el modelo dominante entre las empresas más valiosas del mundo. De acuerdo con su análisis de estrategia de plataformas, dichas empresas generan valor económico gracias a que conectan a individuos y a otras organizaciones de maneras que una empresa tradicional, estructurada como pirámide cerrada, difícilmente podría replicar. Los mismos autores advierten, no obstante, que dirigir una organización así requiere de otro tipo de liderazgo, en donde quien administra un ecosistema, además de competir dentro de la economía convencional, aprende a cultivar y proteger la red de relaciones que rodea a su organización, tarea para la que ninguna cadena de mando jerárquica tradicional está diseñada de manera adecuada.

Cabe aclarar que el pensar la organización como un ecosistema y no como una jerarquía no implica abandonar la estructura ni la autoridad formal, puesto que Galbraith (2014) es claro en que la estructura sigue siendo indispensable. Lo que implica es reconocer que hoy dicha estructura convive, quiera o no, con redes de colaboración horizontal, con nodos de innovación distribuidos y con una velocidad de adaptación que ninguna cadena de mando vertical puede sostener por sí sola. La organización ante esta nueva era de transformación, además de preguntarse quién manda a quién, tiene que preguntarse cómo diseñar un sistema en donde la información, la innovación y la responsabilidad puedan circular en todas direcciones sin perder coherencia.

1.8 LA TRANSFORMACIÓN COMO FENÓMENO CULTURAL, NO ÚNICAMENTE TECNOLÓGICO

Existe una tentación comprensible, sobre todo entre quienes lideran la adopción tecnológica dentro de una organización, la cual consiste en tratar la transformación digital como un problema de infraestructura, es decir, como un asunto de comprar el *software* correcto, entrenar el modelo adecuado y actualizar el *hardware* necesario. Dicha tentación se entiende, ya que los problemas técnicos tienen soluciones técnicas relativamente claras. Sin embargo, esa lógica, aplicada de manera aislada, ha mostrado sus límites en repetidas ocasiones. La parte más difícil de cualquier transformación tecnológica ha sido siempre cambiar los hábitos, las creencias y las prácticas cotidianas de quienes usan la herramienta, más que instalarla.

Dentro del material revisado, una de las evidencias más claras de este punto es la que ofrece Ruiz Sevilla (2026), quien en su investigación sobre cultura de ciberseguridad retoma la definición clásica de Da Veiga y Eloff (como se citó en Ruiz Sevilla, 2026), en la cual la cultura de seguridad de la información se entiende como la suma de percepciones, actitudes, valores y comportamientos que una comunidad demuestra al interactuar con la tecnología, más que como un cúmulo de conocimientos técnicos; es decir, como una asimilación psicológica de la seguridad en tanto responsabilidad cotidiana, y no como un manual que se memoriza una sola vez.

Los hallazgos del estudio confirman esa distinción de manera consistente. En él se pudo constatar que los estudiantes de semestres avanzados dominan casi a la perfección la teoría de la ingeniería social y, aun así, desactivan voluntariamente sus antivirus o reutilizan contraseñas, esto porque el conocimiento técnico nunca se tradujo en una cultura de autoprotección genuina (Ruiz Sevilla, 2026). Por lo tanto, el autor concluye que la educación en esta materia no puede limitarse a transmitir teoría computacional y que tiene que modificar conductas de forma sistémica (Ruiz Sevilla, 2026).

El patrón de una tecnología bien comprendida frente a una cultura todavía inmadura también se presenta con claridad en el terreno de la inteligencia artificial generativa. Ríos Navarrete (2025) documenta que implementar IA en departamentos de recursos humanos puede generar discriminación algorítmica en los procesos de selección, lo cual ocurre porque las organizaciones la adoptan sin haber construido antes una cultura interna de responsabilidad, transparencia y supervisión ética, más que por un defecto inherente de la tecnología.

El mismo estudio recupera la crítica del sociólogo Zygmunt Bauman (como se citó en Ríos Navarrete, 2025), quien sostiene que las estructuras sociales contemporáneas perdieron solidez y viven en un estado de cambio permanente, al que denomina modernidad líquida. A partir de dicha crítica, Ríos Navarrete (2025) plantea que la inteligencia artificial, al redefinir las interacciones sociales, el trabajo y la vigilancia dentro de las organizaciones, vuelve especialmente urgente

preguntarse cómo se preserva la responsabilidad moral en medio de esa fluidez. De ahí que la ética no pueda tratarse como un anexo técnico de la inteligencia artificial, ya que es una de las preguntas culturales centrales que su llegada obliga a responder.

Una tensión muy parecida es la que documenta Cerezo Arreola (2025) entre universitarios y recién egresados. Ante un vacío normativo institucional sobre el uso legítimo de la inteligencia artificial, los propios estudiantes terminan por desarrollar sus propias normas informales de uso, adoptando la tecnología por eficiencia académica y, además, como una estrategia cultural de supervivencia profesional frente a un mercado laboral que ya exige su dominio (Cerezo Arreola, 2025). Con esto se puede ver que cuando las instituciones no logran construir a tiempo una cultura compartida en torno a una tecnología, dicha cultura no desaparece, ya que se forma de manera espontánea y descentralizada entre quienes la usan a diario, con todos los riesgos éticos que eso conlleva.

El caso de “Mi Diario Digital de Aprendizaje” ilustra la misma idea, aunque desde una perspectiva más esperanzadora. En él, Robles Ornelas y Páez López (2026) no se limitaron a diseñar una herramienta digital de registro académico, pues su propuesta implicó, sobre todo, un cambio cultural en la manera en que los estudiantes se relacionan con su propio aprendizaje, pasando de la memorización pasiva a la reflexión consciente y la metacognición. Los resultados del proyecto respaldan dicho matiz, puesto que la inmensa mayoría de los estudiantes participantes consideró útil la plataforma y reportó, además, haberse vuelto más consciente de cómo aprende (Robles Ornelas & Páez López, 2026). Cabe destacar que la tecnología, una página web sencilla y gratuita, fue solo el medio, mientras que el cambio real se dio en la cultura de reflexión que dicha herramienta ayudó a instalar.

Este mismo principio se reconoce incluso en el marco de Galbraith (2014), que fue pensado originalmente para el diseño estructural de las organizaciones, al incluir a las personas como uno de sus cinco elementos indispensables junto con la estrategia, la estructura, los procesos y las recompensas. Ninguna organización logra un cambio real de comportamiento si solo modifica su tecnología o su organigrama sin invertir, al mismo tiempo, en las creencias, capacidades y valores de quienes tienen que sostener ese cambio todos los días.

Con base en lo expuesto, las evidencias revisadas comparten una advertencia que las organizaciones ya no pueden seguir postergando, la cual tiene que ver con que ninguna cantidad de inversión en infraestructura tecnológica sustituye el trabajo, mucho más lento y mucho más humano, de construir una cultura organizacional capaz de usar esa tecnología con responsabilidad y conciencia colectiva. Es por ello que la transformación que hoy atraviesan las organizaciones no termina el día en que se instala la última actualización de *software*, sino hasta el día en que las personas que las integran cambian de verdad la forma de pensar, como toman decisiones y una familiarización real entre sí. Es ahí en donde radica, en buena medida, la diferencia entre instalar una tecnología y transformar una organización.

Chapter 2 / Capítulo 2



AI, metaverse and sustainability in the new
organizational design
ISBN: 978-9915-9928-8-4
DOI: 10.62486/978-9915-9928-8-4.Ch02
Pages: 10-16

© 2026 The authors. This is an open access
book chapter distributed under the terms of
the Creative Commons Attribution (CC BY)
4.0 License.

Inteligencia artificial: de herramienta tecnológica a capacidad organizacional Artificial intelligence: from technological tool to organizational capacity

Jesús Alejandro Vázquez Meza¹ 

Víctor Manuel Martínez García² 

Yennifer Díaz Romero³ 

Manuel Iván Tostado Ramírez⁴ 

Rogelio Estrada Lizárraga⁴ 

José de Jesús Reyes Machain² 

1 Universidad Autónoma de Sinaloa, Unidad Académica Preparatoria Rubén Jaramillo, Vicerrectoría, Sinaloa, México

2 Universidad Autónoma de Sinaloa, Facultad de Arquitectura y Diseño, Vicerrectoría, Sinaloa, México

3 Universidad Autónoma de Sinaloa, Facultad de Ciencias de la Educación, Vicerrectoría, Sinaloa, México

4 Universidad Autónoma de Sinaloa, Facultad de Informática, Vicerrectoría, Sinaloa, México

5 Universidad Autónoma de Sinaloa, Facultad de Informática, Vicerrectoría, Sinaloa, México

ABSTRACT

This chapter explores the transition from artificial intelligence as an isolated tool to artificial intelligence as an organizational capability. It begins by distinguishing access to data from the knowledge needed to make and review decisions. Frameworks for technology adoption and trust help explain why access to advanced systems does not by itself produce useful outcomes. The text differentiates traditional and generative AI and contrasts task automation with the augmentation of human capabilities. It examines cognitive copilots, their potential contributions, and the need to oversee their outputs. An augmented organization is described as one in which tools, processes, people, and decisions operate together rather than as disconnected elements. The discussion also addresses the limitations of accumulating individual AI users without redesigning workflows or developing shared skills. Attention is given to decision quality, team coordination, and the institutional conditions required for responsible use. The chapter argues that organizational capability does not reside solely in software: it depends on how technology is embedded in daily practices, how appropriate controls are organized, and how the consequences of its use are evaluated.

Keywords: Artificial intelligence; Organizational capability; Automation; Augmentation; Decision-making

RESUMEN

El capítulo examina el tránsito desde la inteligencia artificial entendida como herramienta aislada hacia su integración como capacidad organizacional. Comienza por diferenciar los datos disponibles del conocimiento que una institución puede usar para decidir y revisar sus operaciones. Recupera marcos sobre adopción tecnológica y confianza para explicar por qué el acceso a sistemas avanzados no asegura resultados útiles. Distingue la inteligencia artificial tradicional de la generativa y compara la automatización de tareas con la aumentación de capacidades humanas. En este recorrido, analiza la función del copiloto cognitivo, sus posibles aportaciones y la necesidad de supervisar sus respuestas. El texto presenta la organización aumentada como una configuración donde herramientas, procesos, personas y decisiones se relacionan de forma continua. También explica las limitaciones de acumular usuarios individuales sin rediseñar flujos de trabajo ni desarrollar competencias compartidas. La discusión se centra en la calidad de las decisiones, la coordinación de equipos y las condiciones institucionales que permiten usar los sistemas con responsabilidad. La conclusión conceptual es que la capacidad no reside únicamente en el software, sino en el modo en que la organización lo incorpora a sus prácticas, estructura sus controles y evalúa sus efectos. La exposición insiste en que los resultados deben interpretarse según la tarea, el contexto y las condiciones de uso. Por ello, la formación de usuarios y la definición de controles aparecen como componentes de la misma capacidad organizacional.

Palabras clave: Inteligencia artificial; Capacidad organizacional; Automatización; Aumentación; Toma de decisiones

2.1 DATOS, CONOCIMIENTO Y CAPACIDAD DE DECISIÓN

Durante mucho tiempo se consideró que una organización estaba “informada” cuando tenía acceso a suficientes datos. Sin embargo, esa noción dejó de ser válida, esto porque el volumen de información disponible dejó de ser el problema central y el desafío consiste ahora en transformar ese volumen en conocimiento que de verdad mejore la toma de decisiones. Es en esa distinción entre tener datos y saber decidir con ellos en donde la inteligencia artificial (IA) encontró uno de sus lugares de mayor impacto dentro de las organizaciones.

Russell y Norvig (2020), autores del texto de referencia más utilizado en la enseñanza universitaria de la inteligencia artificial, definen el campo como el diseño de agentes racionales, es decir, sistemas capaces de percibir su entorno, procesar esa información y actuar de manera que maximicen la probabilidad de alcanzar un objetivo determinado. Si bien se trata de una definición técnica, su implicación organizacional es clara, ya que un sistema de IA no es solo una calculadora más rápida; se trata de un mecanismo diseñado para tomar, o apoyar, decisiones bajo incertidumbre, mismas que antes recaían únicamente en el juicio humano.

Ahora bien, el que una organización tenga acceso a estas capacidades no significa que las use, y mucho menos que confíe en ellas. Una forma de comprender de la mejor forma este punto es recurrir al modelo de aceptación tecnológica de Davis (1989), mismo que, después de tres décadas, sigue siendo el marco más citado para explicar por qué las personas adoptan o rechazan una herramienta digital. En dicho modelo se identifican dos variables que determinan esa decisión: la utilidad percibida, es decir, si la persona cree que la herramienta mejorará realmente su desempeño, y la facilidad de uso percibida, entendida como el esfuerzo que anticipa tener que invertir para usarla (Davis, 1989); al momento de aplicar lo anterior a los sistemas de datos e inteligencia artificial, se puede ver que ningún sistema, por sofisticado que sea, mejora la capacidad de decisión de una empresa si sus colaboradores no perciben que vale la pena usarlo, o si la curva de aprendizaje les resulta demasiado empinada.

En el terreno académico, el estudio de Cerezo Arreola (2025) ofrece evidencia de este mismo fenómeno, ya que retoma de manera explícita el modelo de Davis (1989) para explicar por qué los universitarios adoptan con tanta rapidez las herramientas de inteligencia artificial generativa. De acuerdo con dicho estudio, la utilidad percibida, que en este caso consiste en ahorrar tiempo y estructurar con rapidez información compleja, resulta lo suficientemente alta como para superar cualquier resistencia inicial (Cerezo Arreola, 2025). Con ello se tiene una generación que ya no distingue del todo entre “buscar información” y “consultarle a un sistema de IA”, esto porque en la práctica la utilidad percibida de lo segundo se volvió más alta que la de los métodos tradicionales.

Entre la definición técnica de Russell y Norvig (2020) y el modelo de aceptación de Davis (1989) existe un punto en común que toda organización tendría que tomar en cuenta, en el sentido de que los datos y los algoritmos no mejoran por sí mismos la capacidad de decisión de una empresa. Dicha mejora solo ocurre cuando las personas confían en el sistema, lo entienden lo suficiente para usarlo con criterio y lo perciben como una extensión útil de su propio juicio, más que como una caja negra que sustituye su pensamiento.

La urgencia de construir esa confianza se entiende mejor al observar el ritmo al que mejora la capacidad técnica de estos sistemas. El *AI Index Report* de Stanford (2026) documenta que en SWE-bench Verified, una de las pruebas de referencia más exigentes para evaluar la capacidad de programación de un sistema de IA, el desempeño de los modelos líderes pasó de aproximadamente el 60 % a casi el 100 % en un solo año. Se trata de un salto técnico que las organizaciones no pueden pasar por alto. Sin embargo, dicho salto no garantiza por sí solo que esa capacidad se traduzca en mejores decisiones, ya que la brecha entre lo que un sistema puede hacer técnicamente y lo que una organización logra aprovechar de forma responsable sigue dependiendo de la percepción de las personas, tal como se desprende del modelo de Davis (1989), así como de su confianza, y no solo de la potencia computacional.

2.2 AUTOMATIZACIÓN DE PROCESOS

La automatización no es algo nuevo, ya que las fábricas llevan más de un siglo automatizando tareas físicas repetitivas. Lo que cambió es el tipo de tareas que hoy se pueden automatizar, mismas que ya no se limitan a procesos mecánicos y predecibles e incluyen tareas cognitivas como redactar, resumir, clasificar información, generar código e incluso argumentar. Es ese salto de la automatización física a la cognitiva el que distingue a la ola actual de transformación organizacional de las anteriores.

Russell y Norvig (2020) sitúan este fenómeno dentro del campo más amplio del aprendizaje automático, en donde los sistemas, en lugar de seguir instrucciones fijas programadas paso a paso, identifican patrones a partir de grandes volúmenes de datos y ajustan su comportamiento en consecuencia. Dicha capacidad de aprender, y no solo de ejecutar, es la que permite que hoy un sistema automatice tareas que hace una década se consideraban exclusivamente humanas, tal como es el caso de redactar un informe coherente o sintetizar un documento extenso.

En el caso del trabajo de nivel inicial, Cerezo Arreola (2025) documenta cómo esta automatización cognitiva ya lo está reconfigurando, esto porque los empleos de entrada, que tradicionalmente involucraban el procesamiento rutinario de información, se volvieron altamente susceptibles a la automatización mediante inteligencia artificial generativa. Esto es aún una tarea pendiente en diversas organizaciones que todavía no terminan de asimilar la implicación estratégica de este hallazgo, la cual consiste en que los puestos de inicio de la empresa, que históricamente servían como espacio de formación progresiva dentro de la misma, son los que más rápido se están automatizando, lo que obliga a repensar cómo se desarrolla el talento nuevo cuando esa puerta de entrada es muy reducida.

A escala global, Deloitte (2025) confirma esta tendencia. Según su informe *Global Human Capital Trends* identifica la contracción de los roles de nivel inicial como una de las tensiones principales que hoy enfrentan las organizaciones, misma que genera lo que el propio estudio llama una “brecha de experiencia” cada vez más amplia entre lo que el mercado laboral exige y lo que los recién egresados han tenido oportunidad de practicar en un entorno en el mundo laboral real. Cabe destacar que el automatizar procesos de nivel inicial sin diseñar, al mismo tiempo, nuevas vías de formación que reemplacen la función pedagógica que cumplían esos puestos deja a las organizaciones con un problema de mediano plazo que pocas veces se contempla en el análisis de costos de la automatización. Dicho problema tiene que ver con saber de dónde saldrá, dentro de 5 o 10 años, el talento con experiencia suficiente para ocupar los puestos de mayor responsabilidad.

Por otra parte, la velocidad con la que una organización adopta la automatización de procesos depende tanto de la capacidad técnica del sistema como del ritmo con el que las personas dentro de dicha organización la aceptan. En este punto resulta útil la teoría de Rogers (2003), ya que permite entender que la automatización, como cualquier innovación, no se implementa de un día para otro en toda la organización. Esta avanza primero entre quienes están dispuestos a experimentar y solo después se extiende hacia quienes necesitan evidencia más sólida antes de confiar en que un proceso automatizado no comprometerá la calidad de su trabajo. Tratar la automatización como un interruptor que se enciende de golpe, sin tomar en cuenta esa dinámica, es una de las razones más comunes por las que estos proyectos fracasan o generan una resistencia interna innecesaria.

Con lo anterior se puede ver que automatizar procesos no se limita a una decisión de eficiencia operativa, ya que redefine qué tareas siguen siendo humanas, cómo se forma el talento nuevo y cuánto están dispuestas las personas a confiar en sistemas que, cada vez con más frecuencia, piensan por sí mismos en lugar de solo ejecutar instrucciones.

2.3 IA TRADICIONAL E IA GENERATIVA

Hasta hace relativamente poco, cuando una empresa hablaba de “inteligencia artificial” se refería casi siempre a sistemas de reglas fijas o a modelos predictivos entrenados para una tarea muy concreta, tal como detectar fraude, recomendar un producto o clasificar un correo como *spam*. Russell y Norvig (2020) ubican dichos sistemas dentro de lo que llaman inteligencia artificial débil, es decir, programas diseñados para resolver problemas acotados, sin pretensión alguna de generalidad, los cuales funcionan bien, pero solo dentro del terreno estrecho para el que fueron entrenados.

Con la inteligencia artificial generativa esa situación cambió. Si bien en sentido técnico sigue siendo IA débil, su rango de aplicación se volvió mucho más amplio, al grado de que un mismo modelo puede redactar, programar, resumir, traducir y razonar sobre problemas que nunca antes había visto en esa combinación exacta. Dicha versatilidad es la que atrajo tanta atención académica en tan poco tiempo. Tal como ocurre en el caso del análisis bibliométrico de Gómez Cano y Pérez Gamboa (2025) sobre publicaciones indexadas en Scopus, en donde se documenta un crecimiento exponencial en la producción científica sobre adopción de IA generativa en educación superior, con una colaboración internacional liderada por distintos países de Norteamérica, Europa y Asia y un interés creciente en la personalización del aprendizaje.

Este crecimiento refleja algo que las organizaciones apenas están asimilando, esto es, que la diferencia entre la IA tradicional y la generativa tiene que ver con la escala de su impacto y no solamente con aspectos técnicos. Mientras que la IA tradicional automatiza tareas específicas dentro de un proceso ya definido, la IA generativa se inserta en el propio proceso de pensamiento, es decir, en tareas como redactar un informe, estructurar una idea o generar una primera versión de casi cualquier cosa. Esto explica por qué su adopción se sintió tan distinta a la de cualquier tecnología anterior.

De acuerdo con el *AI Index Report* de Stanford (2026), esa diferencia de escala ya se refleja en cifras de adopción notablemente altas para una tecnología de introducción tan reciente, ya que la adopción organizacional de inteligencia artificial alcanzó el 88 % a nivel global y cuatro de cada cinco estudiantes universitarios ya utiliza la IA generativa de forma regular en su vida académica. El mismo informe documenta que, en 2025, la industria produjo más del 90 % de los modelos de frontera más notables del mundo, concentración que muestra que el ritmo de mejora de estos sistemas depende hoy, principalmente, de laboratorios corporativos que compiten a un ritmo muy acelerado, más que de la investigación académica.

Para las organizaciones, lo anterior sugiere que la diferencia entre la inteligencia artificial tradicional y la generativa puede incidir en los próximos años en la posición competitiva de quienes la adoptan de forma sistemática frente a quienes lo hacen de manera más limitada, por lo que no conviene tratarla como una curiosidad técnica pasajera.

2.4 AUTOMATIZACIÓN VERSUS AUMENTACIÓN

Cuando se habla de inteligencia artificial suele pasarse de largo una distinción importante, la cual consiste en que no es lo mismo automatizar una tarea que aumentar la capacidad de quien la realiza. En la automatización, la máquina hace todo el trabajo de principio a fin y sin intervención humana, mientras que en la aumentación la máquina fortalece lo que una persona ya puede hacer, sin sustituirla.

En el contexto universitario, Díaz Ancoco et al. (2024) documentan esta segunda lógica al señalar que la inteligencia artificial cumple la función de proporcionar conocimientos inmediatos a los estudiantes con el propósito específico de potenciar habilidades y capacidades ya existentes, sin reemplazarlas. Cabe destacar este matiz, puesto que en dicho estudio la IA no sustituye el aprendizaje del estudiante; le da acceso más rápido a información que antes le habría tomado más tiempo encontrar, mientras que es el propio estudiante quien procesa, cuestiona y aplica esa información.

Este matiz se relaciona con lo que Russell y Norvig (2020) llaman agentes racionales, es decir, sistemas diseñados para maximizar un objetivo definido externamente por un humano. En ellos, el éxito lo sigue definiendo la persona, es decir, la máquina no decide qué es exitoso. Cuando una organización pierde de vista esta distinción y automatiza sin criterio tareas que en realidad requerían aumentación, es decir, juicio, contexto y sensibilidad ética, suele terminar con resultados peores que si no hubiera adoptado la tecnología en absoluto.

En este sentido, Brynjolfsson (2022) ofrece el que quizás sea uno de los argumentos con mayor influencia de los últimos años para entender por qué esta distinción, además de técnica, es en buena parte política y económica. En su ensayo sobre lo que llama la “trampa de Turing”, sostiene que existe un estímulo sistemático entre tecnólogos, ejecutivos y responsables de políticas públicas para construir inteligencia artificial que imite y reemplace el trabajo humano, en lugar de construir inteligencia artificial que lo amplíe. Dicha preferencia tiene consecuencias, ya que cuando las máquinas se convierten en mejores sustitutos del trabajo humano, los trabajadores pierden poder de negociación económico y político y se vuelven cada vez más dependientes de quienes controlan la tecnología (Brynjolfsson, 2022).

Por el contrario, cuando la inteligencia artificial se orienta a aumentar las capacidades humanas en lugar de imitarlas, las personas conservan la capacidad de exigir una parte del valor que esa tecnología genera; además, la aumentación tiende a crear productos y servicios enteramente nuevos, lo cual genera más valor total que la simple imitación de tareas ya existentes. Es así como automatizar sin criterio alguno, además de ser un error de diseño técnico, se convierte en una decisión, casi siempre implícita, sobre quién se queda con el valor que genera la inteligencia artificial.

2.5 IA COMO COPILOTO COGNITIVO

El término “copiloto” se volvió popular en el mundo corporativo porque describe algo que las metáforas anteriores no lograban capturar. En este caso no es que la IA vuele en el avión sola, pero tampoco es un simple instrumento pasivo en el tablero, ya que está ahí de manera activa sugiriendo rutas, anticipando errores y revisando cálculos, mientras que la decisión final y la responsabilidad sigue siendo del personal humano.

El rol de copiloto cognitivo implica un tipo de colaboración distinto al que las organizaciones estaban acostumbradas a diseñar. En la automatización tradicional se delega una tarea completa a un sistema y se espera el resultado, mientras que en este caso se tiene un intercambio continuo, en donde la persona plantea, el sistema responde, la persona ajusta y el sistema vuelve a responder. Díaz Ancco et al. (2024) describen esta dinámica de acompañamiento inmediato como el valor central que los estudiantes universitarios encuentran en estas herramientas, debido a que lo que reciben es un punto de partida que ellos mismos deben refinar, más que una respuesta final y cerrada.

Una de las evidencias experimentales más claras sobre qué tan bien, y qué tan mal, puede funcionar un copiloto cognitivo según el tipo de tarea es la que ofrecen Dell'Acqua et al. (2023) en su estudio de campo con consultores de Boston Consulting Group. En las tareas en donde el sistema tenía una ventaja real de capacidad, los consultores que lo usaron como copiloto completaron más subtareas, trabajaron considerablemente más rápido y produjeron resultados de mejor calidad que quienes trabajaron sin él. Sin embargo, en una tarea diseñada específicamente para exigir el tipo de juicio contextual que un copiloto no domina bien, la precisión de quienes usaron el sistema cayó de forma marcada frente a quienes no lo usaron. Con base en lo anterior, se puede deducir que un copiloto cognitivo no mejora el desempeño de manera uniforme en cualquier tarea, sino solo cuando la persona sabe distinguir en qué terreno aporta una ventaja real y en cuál solo aporta la ilusión de ayuda.

2.6 LA ORGANIZACIÓN AUMENTADA

Cuando suficientes personas dentro de una organización empiezan a trabajar con la IA como copiloto cognitivo, se presenta un efecto colectivo que supera la suma de sus usos individuales, en donde la organización misma comienza a operar con una capacidad ampliada, esto al procesar más información, explorar más alternativas e iterar más rápido sobre sus propias ideas. A dicho fenómeno conviene llamarlo organización aumentada, por analogía directa con la aumentación individual descrita anteriormente.

En la práctica, McKinsey & Company (2025) documenta qué distingue a una organización que logró esta aumentación colectiva de otra que solo acumuló usuarios individuales de IA, ya que entre las organizaciones de alto desempeño de su encuesta el 55 % reporta haber rediseñado de forma fundamental sus flujos de trabajo al incorporar la inteligencia artificial, frente a apenas una quinta parte del resto de las organizaciones encuestadas. Con esto se puede ver que la diferencia tiene que ver con si el uso de la IA se tradujo en un rediseño deliberado de la manera en que fluye el trabajo colectivo, más que con la cantidad de IA que usa cada organización. Dicho rediseño es la marca distintiva de lo que en este capítulo se llama organización aumentada, frente a la simple acumulación dispersa de herramientas individuales.

Así mismo, Gómez Cano y Pérez Gamboa (2025) aportan evidencia de que esta transición ya está en marcha a escala institucional, ya que en su análisis bibliométrico identifican a las ciencias de la computación, las ciencias de la educación y las ciencias sociales como las áreas disciplinarias más activas en investigación sobre adopción de IA generativa. Lo anterior sugiere que la aumentación organizacional no se limita a las empresas tecnológicas y que atraviesa sectores completos, incluidas instituciones académicas que tradicionalmente cambian con lentitud.

2.7 DE LA HERRAMIENTA AL SISTEMA DE CAPACIDADES

La distinción entre IA tradicional y generativa, la que existe entre automatización y aumentación, el rol de copiloto cognitivo y la organización aumentada convergen en una idea central con la que se cierra este capítulo: la inteligencia artificial dejó de ser una herramienta que una organización adquiere y empezó a convertirse en un sistema de capacidades que debe construir, mantener y gobernar de manera activa.

Cabe destacar que dicha diferencia tiene consecuencias prácticas, puesto que una

herramienta se compra, se instala y se usa, mientras que un sistema de capacidades requiere procesos, gobernanza, criterios de evaluación continuos y, sobre todo, la disposición a revisarlo constantemente conforme cambian tanto la tecnología como los riesgos que trae consigo.

Autio et al. (2024) lo subrayan en el perfil complementario del AI RMF dedicado específicamente a la inteligencia artificial generativa, mismo que el NIST publicó en 2024 como respuesta directa a la orden ejecutiva estadounidense sobre IA segura y confiable. En dicho documento se establece que los riesgos de estos sistemas se modifican con cada actualización del modelo, con cada nuevo caso de uso y con cada dato adicional que se le entrega. Por lo tanto, gestionar la IA generativa a escala organizacional requiere de un sistema vivo, más que de una decisión de compra que se toma una sola vez.

Desde la propia definición técnica del campo, Russell y Norvig (2020) permiten cerrar el argumento, ya que un agente racional bien diseñado sigue dependiendo, en última instancia, de que alguien defina correctamente sus objetivos y evalúe si los está cumpliendo de forma segura. Dicha responsabilidad no desaparece cuando la organización pasa de usar una herramienta aislada a operar un sistema completo de capacidades de IA; por el contrario, tiende a volverse más exigente, esto porque entran en juego más piezas en movimiento y más decisiones distribuidas por gestionar. Reconocer lo anterior a tiempo puede traducirse en una ventaja más duradera para la organización, mientras que tratar la inteligencia artificial como una simple compra de *software* tiende a mostrar, con el paso del tiempo, los límites de esa visión.

McKinsey & Company (2025) aporta datos que respaldan este último punto. Casi todas las organizaciones que la consultora clasifica como de alto desempeño en IA comparten un rasgo que puede resumirse en que, más que comprar una herramienta, construyeron un sistema. Sus líderes reportan de forma consistente un involucramiento activo y visible en la iniciativa de IA, misma que no delegan por completo al área de tecnología, además de que dichas organizaciones son considerablemente más propensas a implementar reglas claras de supervisión humana, gobernanza centralizada y responsabilidad ejecutiva sobre los resultados (McKinsey & Company, 2025). Otro hallazgo relevante es que estas mismas organizaciones reportan más incidentes relacionados con IA que el resto, lo cual no se debe a que gestionen peor el riesgo, sino a que despliegan la tecnología a mayor escala y, por ello, invierten más en detectar y mitigar esos riesgos de forma proactiva. Con base en estos datos se puede afirmar que es el sistema de capacidades, más que la herramienta aislada, lo que distingue a las organizaciones que logran obtener un valor sostenido de la inteligencia artificial.

Chapter 3 / Capítulo 3

AI, metaverse and sustainability in the new
organizational design
ISBN: 978-9915-9928-8-4
DOI: 10.62486/978-9915-9928-8-4.Ch03
Pages: 17-24



© 2026 The authors. This is an open access
book chapter distributed under the terms of
the Creative Commons Attribution (CC BY)
4.0 License.

The human-AI organization: new forms of collaboration

La organización humano-IA: nuevas formas de colaboración

Jesús Alejandro Vázquez Meza¹ 

Víctor Manuel Martínez García² 

Yennifer Díaz Romero³ 

Manuel Iván Tostado Ramírez⁴ 

Rogelio Estrada Lizárraga⁴ 

José de Jesús Reyes Machain² 

1 Universidad Autónoma de Sinaloa, Unidad Académica Preparatoria Rubén Jaramillo, Vicerrectoría, Sinaloa, México

2 Universidad Autónoma de Sinaloa, Facultad de Arquitectura y Diseño, Vicerrectoría, Sinaloa, México

3 Universidad Autónoma de Sinaloa, Facultad de Ciencias de la Educación, Vicerrectoría, Sinaloa, México

4 Universidad Autónoma de Sinaloa, Facultad de Informática, Vicerrectoría, Sinaloa, México

5 Universidad Autónoma de Sinaloa, Facultad de Informática, Vicerrectoría, Sinaloa, México

ABSTRACT

This chapter addresses collaboration between people and artificial intelligence systems through changes in work. It distinguishes tasks that may be automated, tasks whose content changes, and activities that may emerge when processes are reorganized. Contributions from labor economics are connected to discussions of productivity, augmented workers, and shared decisions. The review considers potential support for routine activities as well as risks related to technological dependence and excessive delegation of judgment. It distinguishes the speed of AI processing from human capacities for contextual interpretation, creativity, and accountability. Particular attention is paid to human oversight and the definition of new responsibilities within teams using these tools. The chapter also considers verification of system outputs and the need for mechanisms to correct errors or challenge automated recommendations. Collaboration is not treated as the automatic replacement of professionals by machines, but as a relationship that changes tasks, requires skills, and redistributes decisions. Critical thinking and professional judgment are presented as essential conditions for integrating artificial intelligence into work while preserving meaningful human deliberation.

Keywords: Human-AI collaboration; Work transformation; Human oversight; Critical thinking; Productivity

RESUMEN

El capítulo aborda la colaboración entre personas y sistemas de inteligencia artificial a partir de las transformaciones del trabajo. Distingue las tareas susceptibles de automatización de aquellas que cambian de contenido y de las nuevas actividades que pueden surgir al reorganizar procesos. Examina aportes de la economía laboral y los conecta con discusiones sobre productividad, trabajadores aumentados y decisiones compartidas. La revisión considera tanto las oportunidades de apoyo en actividades rutinarias como los riesgos asociados con la dependencia tecnológica y la delegación excesiva del juicio. Diferencia la rapidez de procesamiento de la inteligencia artificial de las capacidades humanas de contextualización, creatividad y responsabilidad. El capítulo concede especial importancia a la supervisión humana y a la definición de nuevas funciones dentro de equipos que utilizan estas herramientas. También estudia la necesidad de verificar los resultados y mantener mecanismos para corregir errores o cuestionar recomendaciones automáticas. La colaboración no se interpreta como sustitución automática de profesionales por sistemas, sino como una relación que modifica tareas, exige competencias y redistribuye decisiones. En conjunto, la discusión presenta el pensamiento crítico y el juicio profesional como condiciones para integrar la inteligencia artificial en el trabajo sin perder la capacidad humana de deliberación. Se resalta además que las recomendaciones de un sistema requieren interpretación contextual. La distribución de tareas debe acompañarse de mecanismos explícitos de evaluación, aprendizaje y rectificación para que la colaboración sea examinable dentro de la práctica profesional.

Palabras clave: Colaboración humano-IA; Transformación del trabajo; Supervisión humana; Pensamiento crítico; Productividad

3.1 ¿QUÉ TAREAS DESAPARECEN?

Uno de los estudios más citados, y también de los más discutidos, en la historia reciente de la economía laboral es el de Frey y Osborne (2017), dos investigadores de Oxford. Los autores revisaron 702 ocupaciones de Estados Unidos con ayuda de clasificadores de aprendizaje automático, esto con la finalidad de estimar la probabilidad de automatización de cada una. La cifra a la que llegaron aparece desde entonces en casi cualquier conversación sobre el futuro del trabajo, y es que cerca del 47 % del empleo total de ese país se encontraba en riesgo alto de computarización. Cabe destacar que con ello se referían a empleos que en algún momento resultaban técnicamente automatizables, lo cual es distinto a afirmar que fueran a desaparecer de un momento a otro.

Si bien la cifra llama la atención, el valor de esta metodología radica también en el criterio que la sostiene. De acuerdo con Frey y Osborne (2017), lo que hace que una ocupación esté más o menos expuesta es el tipo de tareas que la componen, independientemente del nombre del puesto; es así como los trabajos que consisten sobre todo en tareas repetitivas y predecibles, como la transcripción de datos, ciertos procesos administrativos o el trabajo en una línea de ensamblaje, resultan mucho más susceptibles que aquellos en donde se requiere percepción compleja, creatividad o inteligencia social. Los autores también documentaron algo que con los años resultaría profético, esto al encontrar una relación negativa muy fuerte entre el salario, el nivel educativo de una ocupación y su probabilidad de automatización, es decir, cuanto más rutinaria y menos calificada es una tarea, más expuesta se encuentra.

Casi una década después, y con la llegada de la inteligencia artificial generativa, esa relación se ha vuelto más compleja de lo que anticipaba el estudio original, punto que se retomará más adelante en este capítulo. Aun así, la lógica de fondo sigue siendo válida, la cual consiste en que las ocupaciones completas no desaparecen de la noche a la mañana; lo que desaparece primero son las tareas específicas dentro de ellas que resultan rutinarias, predecibles y fáciles de codificar.

En su informe *Global Human Capital Trends*, elaborado a partir de una encuesta a casi 10 000 líderes de recursos humanos y de negocio en 93 países, Deloitte (2025) documenta que esta desaparición de tareas dejó de ser una proyección a futuro y forma parte del presente. Un ejemplo de ello son los puestos de nivel de entrada, que tradicionalmente concentraban las tareas más rutinarias de cualquier organización y que hoy se están reduciendo, situación que genera lo que el propio informe llama una “brecha de experiencia” cada vez más amplia entre lo que exige el mercado laboral y lo que los recién egresados han tenido oportunidad de practicar. Se trata de una confirmación reciente, y a gran escala, del patrón que Frey y Osborne (2017) describieron hace casi una década, en el que la automatización ataca primero ciertas tareas, en lugar de ocupaciones enteras, y en este caso aquellas que servían de entrenamiento para los puestos siguientes.

3.2 ¿QUÉ TAREAS SE TRANSFORMAN?

Si el trabajo de Frey y Osborne (2017) se enfoca en lo que se pierde, el de Acemoglu y Restrepo (2018) se ocupa de lo que se gana, algo igual de importante y con frecuencia olvidado, además de explicar por qué la automatización no es un juego de suma cero. Para ello, los autores plantean dos fuerzas que actúan al mismo tiempo, aunque empujan en sentidos contrarios, siendo la primera el efecto desplazamiento, que se da cuando las máquinas sustituyen a las personas en tareas que antes realizaban, con lo que disminuye la demanda de trabajo humano en esas tareas específicas. Frente a este se encuentra el efecto productividad y reinstalación, mismo que se presenta cuando la automatización abarata costos y libera recursos que terminan por financiar la creación de tareas completamente nuevas, es decir, tareas que antes no existían porque no eran económicamente viables y que vuelven a requerir trabajo humano.

Cabe destacar que Acemoglu y Restrepo (2018) insisten en un matiz que suele perderse en el debate público, el cual consiste en que el efecto reinstalación no es automático ni está garantizado. Para que ocurra, la economía tiene que crear en verdad esas nuevas tareas intensivas en trabajo, y la velocidad de la automatización no debe superar la capacidad del sistema para generarlas. Cuando la automatización avanza más rápido de lo que el mercado logra reinventar nuevas formas de trabajo humano, los autores hablan de “automatización excesiva”, un desequilibrio que destruye empleo neto y que, además, desperdicia el potencial de productividad que la tecnología debería estar generando.

El efecto reinstalación del que hablan Acemoglu y Restrepo (2018) también se puede ver en los datos del *AI Index Report* de Stanford (2026), en donde se documenta que la adopción

organizacional de IA llegó al 88 % a nivel global. Se trata de un salto considerable que, de haber seguido solamente la lógica del desplazamiento, tendría que haber provocado una contracción generalizada del empleo, lo cual no ha ocurrido; por el contrario, el mismo índice registra en paralelo el crecimiento de roles nuevos ligados a la gobernanza, la supervisión y la implementación de estos sistemas. Estos roles corresponden al tipo de tarea reinstalada que anticipa la teoría, con la particularidad de que solo aparecen cuando la organización invierte deliberadamente en crearlos.

3.3 AUTOMATIZACIÓN DE ACTIVIDADES RUTINARIAS

Con la evidencia más reciente, y con una precisión que Frey y Osborne no podían tener en 2017, se observa que el riesgo de automatización sigue concentrado en las tareas rutinarias. Sin embargo, el umbral de lo que cuenta como “rutinario” se ha movido de manera considerable desde entonces.

Para tratar de entender de mejor forma por qué ese umbral se ha movido tan rápido, conviene revisar un dato técnico que aporta el *AI Index Report* de Stanford (2026), en donde se documenta que en SWE-bench Verified, una prueba de referencia exigente para evaluar la capacidad de programación, el desempeño de los sistemas líderes pasó de alrededor del 60 % a casi el 100 % en un solo año. Es así como tareas de programación que hasta hace poco se consideraban claramente no rutinarias, por exigir un juicio técnico considerable, empezaron a comportarse casi como rutinarias desde la perspectiva de la automatización. Con lo anterior se puede ver que la frontera entre lo automatizable y lo que no lo es cambia de lugar con cada mejora en la capacidad de estos sistemas, por lo que las organizaciones que la definieron hace apenas un par de años probablemente ya la tienen desactualizada.

3.4 NUEVAS FORMAS DE PRODUCTIVIDAD

El automatizar tareas rutinarias libera tiempo, por lo que la pregunta que más le importa a una organización es qué hacer con ese tiempo liberado.

McKinsey & Company (2025) documenta en dónde ya se está materializando esta productividad, y en dónde todavía no. Su encuesta encontró que las funciones de ingeniería de software y de tecnología de la información reportan reducciones de costos de entre 10 y 20 % atribuibles directamente a la inteligencia artificial, mientras que en mercadotecnia y desarrollo de producto se observan incrementos de ingresos superiores al 10 %. Si bien se trata de ganancias considerables, estas se encuentran concentradas, puesto que el propio informe advierte que la mayoría de las organizaciones no ha logrado que sus proyectos de IA escalen más allá de la fase experimental, esto debido a que no ha podido traducir esas ganancias funcionales aisladas en una reorganización más amplia del trabajo.

Esto obliga a replantear la pregunta que muchas organizaciones se hacen hoy, ya que en la práctica la discusión sobre si conviene o no adoptar IA ya quedó resuelta. Lo que ahora corresponde preguntarse es qué tan rápido, y con qué grado de intención, se está reorganizando el trabajo alrededor de las nuevas formas de productividad que la tecnología hace posibles.

3.5 EL TRABAJADOR AUMENTADO

En referencia a cómo se ve, en la práctica, el trabajador que sobrevive y prospera en medio de esta transición, el *Future of Jobs Report 2025* identifica a la inteligencia artificial y al análisis de datos como las competencias tecnológicas con mayor proyección de crecimiento en importancia (World Economic Forum, 2025). Cabe destacar que el mismo informe sitúa a las habilidades humanas como un complemento indispensable, lejos de considerarlas un vestigio en vías de extinción.

Por su parte, el reporte de LinkedIn Learning (2025) documenta que lo que separa a un trabajador que solo convive con la IA de otro que en verdad se ve aumentado por ella es la inversión activa en su propio desarrollo. Dicho reporte encontró que el 68 % de los profesionales encuestados coincide en que aprender de forma continua le ayuda a adaptarse al cambio, 4 puntos más que el año anterior, y que las organizaciones que priorizan el desarrollo de carrera de su gente tienen significativamente más probabilidades de posicionarse a la vanguardia en la adopción de inteligencia artificial generativa. Con esto se puede ver que el trabajador aumentado

se construye a través de una inversión deliberada en desarrollo continuo, tanto individual como organizacional, más que surgir espontáneamente por el simple contacto con la tecnología.

3.6 EL RIESGO DE LA DEPENDENCIA TECNOLÓGICA

Lo revisado hasta aquí permite tener una visión moderadamente optimista, siempre y cuando la automatización se gestione de forma adecuada. Sin embargo, conviene detenerse en una advertencia que aparece de manera constante en la evidencia revisada, la cual tiene que ver con el costo de la dependencia tecnológica excesiva, mismo que no siempre se hace visible de inmediato.

Ese costo ya lo anticipaban Acemoglu y Restrepo (2018) con su concepto de automatización excesiva, el cual describe lo que ocurre cuando una organización automatiza más rápido de lo que su propia capacidad de reinventar el trabajo puede sostener. En ese caso, el resultado incluye una posible pérdida de empleo y, además, una ineficiencia estructural que frena el crecimiento de la productividad que la automatización debería haber generado. En cierto sentido, esta es la paradoja central del capítulo, pues la tecnología que promete productividad puede terminar por restarla si se implementa sin criterio.

El otro lado de esta advertencia lo aportan Frey y Osborne (2017) con su hallazgo sobre la relación entre el salario, la educación y el riesgo de automatización. A partir de este hallazgo se puede ver que las organizaciones que automatizan sin invertir al mismo tiempo en la formación de su fuerza laboral corren el riesgo de ampliar la brecha que la evidencia identifica como la más peligrosa, la cual consiste en concentrar cada vez más valor en quienes ya cuentan con las competencias para trabajar junto a la tecnología y en dejar cada vez más atrás a quienes no las tienen. Visto así, el gestionar bien la automatización es tanto una decisión técnica como una decisión sobre qué tipo de organización, y qué tipo de mercado laboral, se quiere construir con el tiempo que la tecnología libera.

Existe también un riesgo de dependencia que opera en un nivel más íntimo que el del mercado laboral, el de las capacidades cognitivas de quienes usan estas herramientas todos los días. En un estudio con 666 participantes de distintas edades y niveles educativos, Gerlich (2025) encontró una correlación negativa significativa entre el uso frecuente de herramientas de inteligencia artificial y el desempeño en pruebas de pensamiento crítico. Esta relación se encuentra mediada por lo que el propio estudio llama “descarga cognitiva”, es decir, la tendencia a delegar en la herramienta procesos mentales que antes se realizaban de forma autónoma. Uno de los hallazgos más preocupantes de la investigación es que los participantes más jóvenes mostraron, al mismo tiempo, mayor dependencia de la IA y puntuaciones más bajas en pensamiento crítico que los de mayor edad (Gerlich, 2025), lo cual cobra relevancia si se toma en cuenta que se trata del segmento que hoy está entrando al mercado laboral y al que las organizaciones están formando para los próximos 20 o 30 años de su vida profesional.

Este hallazgo no significa que la inteligencia artificial sea perjudicial por naturaleza para el pensamiento humano, ya que el propio Gerlich (2025) sugiere que el impacto depende, en buena medida, de la forma en que se usa la herramienta, más que de la herramienta en sí. Lo que sí confirma es que la dependencia tecnológica ha dejado de ser únicamente un riesgo abstracto o futurista para convertirse en un fenómeno que ya se puede medir, por lo que las organizaciones que despliegan IA generativa a gran escala entre su fuerza laboral, sobre todo entre los perfiles más jóvenes, deberían tomarla tan en serio como cualquier otro riesgo operativo.

3.7 INTELIGENCIA HUMANA E INTELIGENCIA ARTIFICIAL

Entre las metáforas que circulan sobre la inteligencia artificial, una de las más útiles es la que propone Ethan Mollick, profesor de Wharton, quien plantea pensar en la IA como un alien. Con ello se refiere a algo que por naturaleza no es hostil ni benevolente, pero que resulta distinto a la inteligencia humana, ya que sus fortalezas y debilidades no siguen el mismo patrón que las de las personas (Mollick, 2024). Es así como la IA puede resultar sorprendentemente capaz en tareas que a las personas les exigen un esfuerzo considerable, como lo son sintetizar grandes volúmenes de texto o generar variaciones creativas con rapidez, y, de forma igual de sorprendente, mostrarse torpe en tareas que a cualquier persona le resultarían triviales.

A esa asimetría Mollick (2024) la llama frontera irregular, esto porque el límite entre lo que la IA hace bien y lo que hace mal no obedece a una lógica intuitiva y solo se descubre al

experimentar directamente con la herramienta en tareas reales. Dado que ningún manual anticipa con precisión dónde cae esa frontera en cada tarea específica, lo que queda es ponerla a prueba, de ahí que ninguna organización pueda diseñar desde un escritorio una estrategia completa de colaboración humano-IA sin que experimenten directamente quienes van a trabajar con la IA todos los días.

Por lo demás, la frontera irregular tiene su origen en un hallazgo experimental, más que en una metáfora. En un estudio de campo realizado junto con la consultora Boston Consulting Group, Dell'Acqua et al. (2023) asignaron aleatoriamente a 758 consultores para trabajar con o sin acceso a GPT-4 en tareas realistas de consultoría, y sus resultados mostraron de manera muy concreta la existencia de dicha frontera. En las tareas que caían dentro de las capacidades del sistema, quienes usaron IA completaron un 12 % más de sub tareas, trabajaron alrededor de un 25 % más rápido y produjeron resultados que los evaluadores humanos calificaron consistentemente como superiores (Dell'Acqua et al., 2023). Sin embargo, en una tarea diseñada a propósito para quedar fuera de esa frontera, la cual consistía en un caso de estrategia de marca en el que se tenían que reconciliar datos numéricos con pistas sutiles contenidas en notas de entrevistas, el patrón se invirtió por completo. Los consultores sin acceso a IA acertaron el 84 % de las veces, mientras que quienes usaron el sistema vieron caer su precisión de forma marcada (Dell'Acqua et al., 2023). En esa tarea específica, la IA terminó por empeorar el desempeño de quienes confiaron en ella sin saber que se encontraban del lado equivocado de la frontera.

3.8 DECISIONES COMPARTIDAS

El que la inteligencia humana y la artificial sean distintas no garantiza, por sí solo, que al combinarlas se obtengan mejores resultados que al usar cualquiera de las dos por separado. La investigación reciente sugiere que ese resultado ideal, al que Hemmer et al. (2024) llaman desempeño complementario de equipo, es más difícil de lograr de lo que se podría pensar. En su trabajo, en donde se formaliza el concepto de complementariedad en la colaboración humano-IA, los autores encontraron que dicho desempeño, en el que el equipo humano-máquina supera lo que cualquiera de los dos lograría por su cuenta, rara vez se observa en la práctica.

Con respecto a las fuentes de las que depende la complementariedad, Hemmer et al. (2024) identifican dos. La primera es la asimetría de información, que se da cuando la persona y el sistema tienen acceso a información distinta y complementaria sobre el problema, mientras que la segunda, la asimetría de capacidad, aparece cuando cada uno resuelve mejor distintos tipos de subproblemas dentro de una misma tarea. Si estas asimetrías no están presentes, o si no se diseñan deliberadamente dentro del proceso de decisión, la colaboración termina pareciéndose más a una suma de dos opiniones redundantes que a un trabajo en equipo; es por ello que diseñar decisiones compartidas efectivas implica algo más que “poner a un humano y a una IA en la misma sala”, ya que se requiere identificar explícitamente qué sabe cada uno que el otro no sabe, para después estructurar el proceso alrededor de esa diferencia.

El experimento de Dell'Acqua et al. (2023) sirve para ilustrar de forma concreta por qué esta asimetría es tan importante. En las tareas en las que la IA ayudó, es decir, dentro de la frontera, existía una asimetría de capacidad, puesto que el sistema podía generar y organizar grandes volúmenes de contenido más rápido de lo que cualquier consultor lograría por sí solo, mientras que la persona aportaba el criterio sobre qué contenido era relevante para el cliente en cuestión. Por el contrario, en la tarea en la que la IA empeoró el desempeño, fuera de la frontera, esa asimetría de capacidad no existía, porque el problema exigía integrar señales sutiles y propias del contexto que el sistema no podía procesar mejor que un humano con experiencia. Aun así, los consultores confiaron en el sistema sin antes identificar en qué consistía la ventaja comparativa de cada parte. De esta manera, una decisión compartida mal diseñada deja de ser neutral, ya que puede llegar a ser peor que decidir sin ayuda alguna.

3.9 SUPERVISIÓN HUMANA

Mollick (2024) resume de una manera memorable una de sus reglas de co-inteligencia, la cual consiste en mantener siempre al humano en el ciclo, lo que debe entenderse como una condición estructural de cualquier colaboración que funcione bien, más que como una formalidad burocrática. Su argumento es sencillo y a la vez sólido. Dado que la frontera irregular de la IA es impredecible, la única manera confiable de detectar cuándo el sistema se equivoca es que haya

una persona revisando activamente sus resultados, en lugar de limitarse a aprobarlos de forma automática.

En la misma línea, McKinsey & Company (2025) encontró que esta supervisión, lejos de ser un lujo reservado a organizaciones particularmente cautelosas, es una de las prácticas que distinguen a quienes obtienen valor de la inteligencia artificial. Entre las organizaciones de alto desempeño que se identifican en su encuesta, contar con reglas claras de supervisión humana en el ciclo de decisión es sustancialmente más común que en el resto, al igual que los mecanismos de gobernanza centralizada y la responsabilidad ejecutiva explícita sobre los resultados del sistema. El hallazgo pone en duda una suposición extendida, según la cual la supervisión humana hace más lenta la adopción de IA y reduce la ventaja competitiva, ya que la evidencia apunta en sentido contrario, al mostrar que la supervisión disciplinada es uno de los factores que separan a las organizaciones que capturan valor de aquellas que solo experimentan sin rumbo.

3.10 NUEVOS ROLES Y RESPONSABILIDADES

Conforme la colaboración humano-IA se vuelve más profunda, las organizaciones empiezan a enfrentar preguntas para las que sus estructuras tradicionales no fueron diseñadas. Al respecto, Castillo Martínez et al. (2024) documentan cómo la adopción de inteligencia artificial generativa complica directamente las cuestiones de autenticidad y de propiedad intelectual, sobre todo en organizaciones que ya operan bajo modelos ágiles, planos y basados en proyectos, lo cual corresponde al tipo de estructura hacia la que se están moviendo muchas empresas, como se discutió en capítulos anteriores.

Entre esas preguntas están la de quién responde por un contenido cuando una parte sustancial fue generada por un sistema de IA, o la de cómo se atribuye la autoría intelectual dentro de un equipo en el que algunos miembros usan estas herramientas extensamente y otros no. Castillo Martínez et al. (2024) plantean estas cuestiones sin ofrecer respuestas universales prefabricadas. Cada organización tiene que construir sus propios criterios, los cuales cobran urgencia conforme la estructura organizacional se vuelve menos jerárquica y más distribuida. Los roles nuevos que empiezan a aparecer en las organizaciones más avanzadas en este terreno, entre ellos los responsables de gobernanza de IA, los auditores de contenido generado y los curadores de calidad de las salidas de los modelos, constituyen respuestas institucionales a esa misma pregunta que sigue sin resolverse.

De acuerdo con McKinsey & Company (2025), la aparición de estos roles ha dejado de ser una previsión especulativa para convertirse en una consecuencia medible de la adopción de IA agéntica dentro de las organizaciones. Su encuesta encontró que, entre las organizaciones que ya escalan agentes de inteligencia artificial en al menos una función, es conveniente la creación de nuevas responsabilidades específicas, vinculadas a la supervisión, la validación y la corrección de las decisiones que estos sistemas toman de forma semiautónoma. Con esto se puede ver que la colaboración humano-IA, además de transformar la manera en que se realiza el trabajo existente, genera categorías de trabajo completamente nuevas, las cuales no tenían equivalente antes de que estos sistemas empezaran a actuar con cierto grado de autonomía.

3.11 CREATIVIDAD, PENSAMIENTO CRÍTICO Y JUICIO PROFESIONAL

Uno de los temores más comunes frente a la inteligencia artificial generativa es que termine por erosionar las capacidades humanas que más valor aportan, como lo son la creatividad genuina, el pensamiento crítico y el juicio profesional que se forja con la experiencia. Castillo Martínez et al. (2024) documentan que dicho temor no carece del todo de fundamento, esto debido a que la facilidad con la que la IA generativa produce contenido plausible hace más difícil, en la práctica, distinguir entre un trabajo creativo y un contenido que simplemente fue generado. Esto tiene implicaciones directas en la manera en que se evalúa y se reconoce el trabajo profesional.

Frente a este temor, Mollick (2024) ofrece un contrapeso útil, aunque sin negarlo del todo. Una de sus reglas de co-inteligencia consiste en tratar a la IA como una persona a la que se le asigna un rol y una personalidad específicos, lo que en la práctica obliga a quien la usa a pensar con mayor claridad qué tipo de tarea le está delegando y por qué. El definir con precisión el rol de la IA, en lugar de solo pedirle “que escriba algo”, termina por convertirse, de manera paradójica, en un ejercicio de pensamiento crítico en sí mismo, puesto que exige que la persona tenga claro lo que quiere lograr antes de delegar cualquier parte del proceso. Si bien el riesgo de erosión

existe, este no se da de forma automática, ya que depende de si la organización cultiva un uso deliberado de la herramienta o si permite un uso pasivo y sin criterio.

El estudio de Gerlich (2025), que se revisó en el apartado 3.6, aporta evidencia empírica que confirma ese riesgo cuando dicho uso deliberado no se cultiva. En él, la correlación negativa entre el uso frecuente de estas herramientas y el desempeño en pensamiento crítico aparece mediada por la descarga cognitiva, es decir, por delegar en la herramienta procesos mentales que antes se hacían de forma autónoma. Este hallazgo se relaciona directamente con la advertencia de Castillo Martínez et al. (2024) sobre la erosión de la creatividad genuina. Cuando delegar se convierte en el hábito por defecto, en lugar de ser una decisión deliberada que se toma caso por caso, el juicio profesional que distingue a un colaborador con experiencia puede empezar a atrofiarse, riesgo que ambos estudios, desde ángulos distintos, permiten anticipar.

3.12 DEL TRABAJADOR QUE UTILIZA IA AL TRABAJADOR QUE COLABORA CON IA

Entre usar una herramienta y colaborar con ella existe una diferencia sutil, pero decisiva. En el primer caso el flujo va en un solo sentido, ya que la persona pide, el sistema entrega y la persona toma o descarta el resultado, mientras que colaborar implica un proceso iterativo y en ambos sentidos, en donde tanto la persona como el sistema ajustan su comportamiento en función de lo que aporta la otra parte.

A partir del trabajo de Hemmer et al. (2024) se puede plantear que esta transición, de usuario a colaborador, es lo que determina si una organización logra que el trabajo conjunto entre personas e IA rinda más que cada parte por separado, o si solo añade una herramienta más a su caja sin cambiar de fondo la manera en que trabaja. El hallazgo central de estos autores, que la complementariedad rara vez ocurre de manera espontánea, puede leerse como una señal de que la mayoría de las organizaciones se encuentra todavía en la etapa de “usar IA” y aún no llega a la de “colaborar con IA”.

Con datos de adopción a escala global, McKinsey & Company (2025) coincide con este diagnóstico, puesto que en su encuesta se encontró que, si bien el 88 % de las organizaciones ya usa inteligencia artificial en al menos una función, apenas un tercio ha logrado escalar ese uso más allá de proyectos piloto aislados; además, solo el 6 % alcanza lo que la propia consultora clasifica como alto desempeño, grupo que se caracteriza, entre otras cosas, por haber rediseñado sus flujos de trabajo alrededor de la colaboración con IA en lugar de solo añadir la herramienta al proceso existente sin modificarlo. Se trata de la misma brecha, pero medida ahora en cifras organizacionales concretas, en donde la mayoría de las empresas sigue en la etapa de “usar IA” que se desprende del trabajo de Hemmer et al. (2024) y solo una minoría ha logrado dar el paso hacia la colaboración. Para dar ese paso se requiere lo que Mollick (2024) plantea en varias de sus reglas, como lo son invitar activamente a la IA al proceso de trabajo desde el principio, sin dejarla como último recurso, mantener una supervisión humana constante y asumir, de entrada, que la versión actual de la herramienta es la peor que se usará en el futuro. Esto último obliga a la organización a mantenerse en un proceso permanente de aprendizaje y ajuste, en lugar de tratar la adopción de IA como un proyecto que en algún momento se da por terminado.

Chapter 4 / Capítulo 4

AI, metaverse and sustainability in the new
organizational design
ISBN: 978-9915-9928-8-4
DOI: 10.62486/978-9915-9928-8-4.Ch04
Pages: 25-32



© 2026 The authors. This is an open access
book chapter distributed under the terms of
the Creative Commons Attribution (CC BY)
4.0 License.

AI, talent, and hybrid organizational space

IA, talento y espacio organizacional híbrido

Jesús Alejandro Vázquez Meza¹ 

Víctor Manuel Martínez García² 

Yennifer Díaz Romero³ 

Manuel Iván Tostado Ramírez⁴ 

Rogelio Estrada Lizárraga⁴ 

José de Jesús Reyes Machain² 

1 Universidad Autónoma de Sinaloa, Unidad Académica Preparatoria Rubén Jaramillo, Vicerrectoría, Sinaloa, México

2 Universidad Autónoma de Sinaloa, Facultad de Arquitectura y Diseño, Vicerrectoría, Sinaloa, México

3 Universidad Autónoma de Sinaloa, Facultad de Ciencias de la Educación, Vicerrectoría, Sinaloa, México

4 Universidad Autónoma de Sinaloa, Facultad de Informática, Vicerrectoría, Sinaloa, México

5 Universidad Autónoma de Sinaloa, Facultad de Informática, Vicerrectoría, Sinaloa, México

ABSTRACT

This chapter examines the role of artificial intelligence in talent management and the configuration of hybrid workplaces. It begins with the idea that professional skills increasingly include interaction with intelligent tools, while technical proficiency does not replace human judgment. The discussion reviews skills-based recruitment, continuous learning, the university-to-work transition, and evolving professional profiles. It then considers the declining centrality of a single physical workplace and the spread of remote, hybrid, and distributed arrangements. Opportunities to combine physical, digital, and immersive settings for team coordination are explored. At the same time, geographical distance and technological mediation create challenges for belonging, communication, and institutional culture. Changes in talent development are connected to the need to redesign learning and collaborative processes. The chapter does not treat one spatial model as a universal solution; instead, it considers the organizational conditions in which different arrangements can function. Its central proposition is that choices about technology, skills, and the workplace are interrelated elements of organizational change rather than separate management decisions.

Keywords: Talent; Digital skills; Hybrid work; Lifelong learning; Distributed organizations

RESUMEN

El capítulo analiza cómo la inteligencia artificial interviene en la gestión del talento y en la configuración de espacios laborales híbridos. Parte de la idea de que las competencias profesionales incorporan nuevas formas de interacción con herramientas inteligentes, sin que las habilidades técnicas sustituyan el criterio humano. Revisa la selección basada en capacidades, la formación continua, las relaciones entre universidad y empresa y los cambios en los perfiles profesionales. El texto examina además la pérdida de centralidad del lugar físico único como sede de la organización y la expansión de modalidades remotas, híbridas y distribuidas. Considera las posibilidades de combinar espacios presenciales, digitales e inmersivos para la coordinación de equipos. En paralelo, reconoce que la distancia geográfica y la mediación tecnológica plantean desafíos para la pertenencia, la comunicación y la cultura institucional. La discusión conecta la transformación del talento con la necesidad de rediseñar procesos de aprendizaje y colaboración. No presenta un único modelo espacial como solución universal, sino que explora sus condiciones de funcionamiento. Su planteamiento es que las decisiones sobre tecnologías, competencias y espacios de trabajo deben entenderse como partes relacionadas de un mismo cambio organizacional. El texto vincula estas dimensiones con la experiencia de pertenecer a una organización y con las formas en que se sostiene el aprendizaje entre equipos. Así, considera conjuntamente la infraestructura, las competencias y las relaciones de trabajo.

Palabras clave: Talento; Competencias digitales; Trabajo híbrido; Formación continua; Organización distribuida

4.1 LA IA COMO NUEVO COMPONENTE DEL CAPITAL HUMANO

Durante mucho tiempo, el capital humano de una organización se medía a partir de variables relativamente estables, tal es el caso de los años de escolaridad, la experiencia acumulada y las certificaciones obtenidas. Con el avance e implementación de la inteligencia artificial generativa se agregó una variable nueva a esa ecuación, misma que llegó con una velocidad que ninguna de las anteriores tuvo. Cerezo Arreola (2025) documenta este fenómeno al identificar lo que llama capital humano híbrido, el cual consiste en la combinación entre el dominio técnico de las herramientas de IA y las habilidades humanas que ningún algoritmo puede replicar, como el juicio ético, la creatividad no algorítmica y la comunicación compleja.

Desde la gestión organizacional, Canossa-Montes de Oca y Peraza-Villarreal (2023) llegan a una conclusión compatible con la anterior. Su investigación, de carácter documental y cualitativo, analiza la función que cumple la inteligencia artificial dentro de la gestión del talento humano de las organizaciones, además de cerrar con propuestas concretas para su implementación en las áreas de recursos humanos. De ambos estudios se desprende que la IA dejó de ser una herramienta externa al capital humano de una organización para convertirse en uno de sus componentes, es decir, en algo que las personas deben dominar de la misma manera en que antes tenían que dominar una hoja de cálculo o un idioma extranjero.

A escala global, Deloitte (2025) aporta datos que respaldan esta transformación. En su informe *Global Human Capital Trends*, elaborado a partir de una encuesta a casi 10 000 líderes de talento y de negocio en 93 países, se encontró que más del 70 % de gerentes y colaboradores se muestra más propenso a permanecer en una organización que los ayude a crecer y prosperar en un mundo moldeado por la inteligencia artificial. Sin embargo, el mismo informe advierte que más de la mitad de los trabajadores y líderes encuestados expresa preocupación por la creciente ambigüedad entre el trabajo que realiza una persona y el que realiza la tecnología (Deloitte, 2025). Con esto se puede ver que el capital humano híbrido que describe Cerezo Arreola (2025) ha dejado de ser un concepto opcional para convertirse en la respuesta necesaria a una pregunta que la fuerza laboral ya se está haciendo.

4.2 COMPETENCIAS HÍBRIDAS

El concepto de capital humano híbrido que propone Cerezo Arreola (2025) merece un desarrollo más detallado, esto porque describe el tipo de perfil que hoy resulta más competitivo en el mercado laboral. Dicho perfil no consiste en dominar la inteligencia artificial en sustitución de las competencias tradicionales, ni en ignorarla para proteger esas competencias. Se trata de una combinación deliberada en la que la herramienta se usa para lo que hace bien, como procesar volumen, generar borradores y acelerar tareas repetitivas, mientras la persona aporta lo que la herramienta no puede aportar por sí sola.

Esta lógica también se puede ver a escala global en el *Future of Jobs Report 2025*, en donde los empleadores encuestados identifican a la inteligencia artificial y al análisis de datos como las competencias tecnológicas con mayor proyección de crecimiento, si bien mantienen al pensamiento creativo, la resiliencia y la curiosidad como habilidades humanas igual de indispensables (World Economic Forum, 2025). Ese patrón, en el que la competencia técnica se suma a la humana en lugar de reemplazarla, es la definición operativa de lo que en este capítulo se llama competencias híbridas. Cabe destacar que coincide con otro dato del mismo informe, según el cual se espera que el 39 % de las competencias laborales clave cambie antes de 2030.

4.3 RECLUTAMIENTO BASADO EN CAPACIDADES

Si el capital humano relevante ya no se mide solo por credenciales tradicionales, el proceso de reclutamiento tiene que cambiar en consecuencia. Con respecto a este cambio, Canossa-Montes de Oca y Peraza-Villarreal (2023) plantean propuestas concretas, las cuales consisten en incorporar la evaluación de competencias en IA dentro de los propios procesos de selección de personal, en lugar de asumir que se trata de una habilidad que se puede enseñar sin ningún costo después de que la persona ha sido contratada.

Los datos de Deloitte (2025) permiten ver que el tránsito hacia un reclutamiento basado en capacidades ya es una urgencia reconocida por la propia alta dirección, más que una recomendación aspiracional. Apenas el 19 % de los ejecutivos encuestados considera que los modelos tradicionales de organización del trabajo, anclados en puestos fijos y credenciales

estáticas, siguen siendo los más adecuados para generar valor, mientras que el 85 % coincide en que sus organizaciones necesitan formas más ágiles de organizar el trabajo para adaptarse con la rapidez que exige el mercado actual. Dicho contraste, de 19 frente a 85, deja ver la distancia entre la manera en que la mayoría de las organizaciones sigue organizando el trabajo, y con ello reclutando, y la forma en que sus propios líderes reconocen que deberían hacerlo.

4.4 FORMACIÓN CONTINUA

Ningún cambio de esta magnitud puede sostenerse sin una inversión sistemática en formación. De acuerdo con el *Future of Jobs Report 2025*, el 85 % de los empleadores encuestados planea invertir en programas de capacitación y actualización de competencias para su fuerza laboral en los próximos años (World Economic Forum, 2025). Con esta cifra se puede ver que la formación continua dejó de ser un beneficio adicional para convertirse en una necesidad estructural de cualquier estrategia de talento seria.

En la misma línea se encuentra el *Reporte de Aprendizaje en el Lugar de Trabajo 2025* de LinkedIn Learning, construido a partir de encuestas a más de 900 profesionales de aprendizaje y desarrollo y de datos agregados de 1000 millones de usuarios de la plataforma. Dicho informe muestra que esta urgencia ya se refleja en la práctica cotidiana de las diversas áreas, ya que el 91 % de estos profesionales coincide en que la formación continua es hoy más importante que nunca para el éxito profesional (LinkedIn Learning, 2025). El mismo informe documenta un hallazgo revelador sobre quién se beneficia de la disrupción de la IA. Las organizaciones que el estudio llama “campeonas del desarrollo de carrera”, es decir, aquellas que invierten de forma sistemática en el crecimiento de su gente, tienen 42 % más probabilidades de posicionarse como pioneras en la adopción de inteligencia artificial generativa que aquellas que no lo hacen (LinkedIn Learning, 2025). Dicho de otra forma, la formación continua, además de ser una respuesta a la disrupción tecnológica, es en sí misma una de las variables que influyen en qué tan bien logra aprovecharla una organización.

Ricra Ruiz et al. (2025), en una revisión sistemática centrada en las implicaciones éticas de la inteligencia artificial generativa en la educación superior, encuentran un patrón que las organizaciones deberían tomar como un dato relevante. De acuerdo con su revisión, la falta de marcos regulatorios claros y de formación ética explícita en torno al uso de estas herramientas es una de las categorías de riesgo más recurrentes en la literatura académica reciente. Su análisis, en el que se revisaron más de 40 referencias publicadas entre 2023 y 2025, concluye que integrar la ética y la alfabetización en IA generativa dentro de los programas formativos es una necesidad, lejos de ser un complemento opcional, esto para evitar que la formación continua se limite a enseñar el manejo técnico de la herramienta sin desarrollar el criterio necesario para usarla de forma responsable.

4.5 UNIVERSIDAD, EMPRESA Y TRANSICIÓN AL TRABAJO

El puente entre la formación universitaria y el mercado laboral siempre ha sido un punto de fricción; sin embargo, la inteligencia artificial generativa lo está sometiendo a una presión particular. Diaz Ancco et al. (2024) documentan que las instituciones de educación superior enfrentan el reto de integrar la IA en sus procesos formativos justo cuando el mercado laboral ya exige que los egresados dominen estas herramientas desde el primer día de trabajo, situación que deja a las universidades compitiendo contra un reloj que corre más rápido que sus propios ciclos de actualización curricular.

Desde la perspectiva institucional, Ricra Ruiz et al. (2025) documentan un patrón similar. Entre los hallazgos más consistentes de su revisión sistemática se encuentra la existencia de desafíos institucionales significativos para integrar la inteligencia artificial generativa de forma coherente en los planes de estudio, junto con una percepción todavía dispar dentro de la propia comunidad universitaria sobre la manera en que debería hacerse. Como resultado, y tal como lo documenta también Cerezo Arreola (2025) en el contexto mexicano, buena parte del aprendizaje real sobre estas herramientas ocurre fuera del aula, de manera informal, entre estudiantes que ya las usan de forma cotidiana sin que exista todavía una estructura curricular que oriente ese uso.

Frente a este panorama de desajuste institucional, el caso de “Mi Diario Digital de Aprendizaje”, documentado por Robles Ornelas y Páez López (2026), ofrece un contraste. Dicho proyecto nació de la identificación directa de una necesidad por parte de quienes estaban más

cerca del problema, sin esperar a que existiera una política curricular formal sobre reflexión académica y metacognición, y se implementó como una herramienta funcional antes de que cualquier estructura institucional la respaldara de manera oficial. Con este caso se puede ver que el puente entre universidad y empresa no siempre tiene que construirse desde arriba, a la espera de que las instituciones se pongan al día; en muchos casos se construye desde abajo, a través de soluciones concretas que después la institución termina por adoptar o formalizar.

4.6 EL NUEVO PERFIL PROFESIONAL

Lo desarrollado hasta aquí converge en un mismo retrato. El perfil profesional que hoy resulta más competitivo ya no se define únicamente por su especialidad técnica; lo que lo define es la manera en que combina esa especialidad con el dominio crítico de la inteligencia artificial. Cerezo Arreola (2025) lo resume de forma clara al señalar que lo que distingue a un profesional competitivo es la capacidad de supervisar la IA, refinarla y aplicarla a problemas complejos, combinándola con juicio ético, creatividad y comunicación, más que el simple uso de la herramienta.

Una lectura similar se desprende del *Future of Jobs Report 2025*, ahora a escala de todo el mercado laboral global, puesto que las organizaciones que buscan sobrevivir a la transformación tecnológica de aquí a 2030 ya no están buscando personas que sepan hacer una sola cosa muy bien; lo que buscan son personas capaces de combinar competencias técnicas y humanas de forma adaptable, en un contexto en el que ambas seguirán cambiando (World Economic Forum, 2025). Ese es el nuevo perfil profesional que atraviesa esta primera parte del capítulo. No corresponde al trabajador reemplazado por la máquina ni al que solo aprendió a usarla, ya que se trata de quien construyó de manera deliberada un capital humano híbrido, capaz de sostenerse en el tiempo mientras la tecnología que lo rodea sigue cambiando.

4.7 LA DESAPARICIÓN DE LAS FRONTERAS TRADICIONALES DEL LUGAR DE TRABAJO

Durante generaciones, “ir a trabajar” significó desplazarse físicamente hacia un lugar específico, en donde la oficina, además de ser un espacio funcional, era en un sentido casi literal la organización misma. Ese vínculo entre organización y lugar físico, que parecía tan natural que no llegaba a cuestionarse, se ha ido disolviendo de manera acelerada. La pandemia que aceleró el trabajo remoto no es la única explicación de este proceso, puesto que fue apenas el catalizador que expuso una tendencia que venía gestándose desde antes.

Para entender por qué esta disolución tiene un carácter más estructural que coyuntural, Lee et al. (2022) ofrecen un marco útil, esto al describir el metaverso como un espacio mixto físico-virtual en donde múltiples usuarios pueden interactuar de manera concurrente con un entorno generado por computadora y entre sí, además de considerarlo el siguiente hito importante del ciberespacio actual. Si bien dicha definición fue pensada originalmente para el consumo masivo y el entretenimiento, describe bien lo que ya empieza a ocurrir dentro de las organizaciones, en donde el espacio de trabajo deja de tener una ubicación única y se convierte en una superposición de capas físicas y digitales que coexisten.

Ball (2022) llega a una conclusión compatible, aunque desde un ángulo más orientado a la infraestructura que a la interacción social, ya que define lo que él llama el Metaverso como una red masivamente escalada e interoperable de mundos virtuales tridimensionales renderizados en tiempo real, que pueden ser experimentados de manera sincrónica y persistente por un número prácticamente ilimitado de usuarios, cada uno con su propio sentido de presencia y con continuidad de datos, es decir, de identidad, historial, objetos y comunicaciones. Trasladada al terreno organizacional, esta definición trae consigo una consecuencia de fondo, la cual consiste en que un colaborador ya no necesita estar en un lugar determinado para que su presencia, su historial de trabajo y sus interacciones con el equipo mantengan continuidad; es así como la frontera que antes definía “estar en la oficina” o “estar fuera de ella” deja de ser la variable relevante.

4.8 TRABAJO REMOTO, HÍBRIDO Y DISTRIBUIDO

Estos tres términos suelen usarse de forma intercambiable, aunque no son equivalentes. El trabajo remoto describe la ausencia de un espacio físico compartido, mientras que el híbrido

combina momentos presenciales con momentos remotos, por lo general bajo algún esquema definido por la organización. Por su parte, el trabajo distribuido, el más radical de los tres, no asume que exista una oficina central de referencia en absoluto; en su lugar, la organización opera desde su propio diseño como una red de nodos dispersos geográficamente.

Las preferencias de las personas por estos esquemas están bien documentadas y se repiten en distintas mediciones. Tal es el caso de una encuesta de Gallup, citada en el reporte de estadísticas de trabajo híbrido de Zoom (2025), la cual encontró que seis de cada diez empleados en puestos en los que el trabajo remoto es técnicamente posible prefieren un esquema híbrido. Una tercera parte prefiere trabajar completamente a distancia y menos del 10 % elegiría regresar a un esquema totalmente presencial si tuviera la opción. El mismo informe documenta que, desde 2021, quienes trabajan de forma presencial reportan de manera sistemática menor satisfacción con sus interacciones laborales que quienes trabajan en esquemas híbridos o remotos, hallazgo que contradice la idea de que la presencia física garantiza, por sí sola, mejores relaciones de equipo.

En su revisión sistemática sobre el metaverso en contextos educativos, Feng et al. (2024) documentan un patrón que también se puede aplicar al terreno organizacional, ya que identificaron 37 estudios desde 2008 que exploran cómo los entornos virtuales tridimensionales y los avatares, usados originalmente en plataformas como Second Life, fueron efectivos para sostener distintas metodologías de trabajo colaborativo y mejorar la experiencia de quienes participaban en ellos a distancia. Un aspecto relevante de este hallazgo es su cronología, puesto que la posibilidad técnica de sostener colaboración distribuida con sentido de presencia compartida lleva más de década y media de desarrollo. Lo que cambió recientemente fue, sobre todo, la disposición organizacional a adoptar esa tecnología como norma y ya no como excepción.

Este tránsito de lo remoto a lo distribuido tiene una implicación que muchas organizaciones aún no terminan de asimilar, la cual consiste en que, cuando el trabajo distribuido se convierte en el diseño por defecto en lugar de una adaptación temporal, cambia también la manera en que se definen los equipos, los horarios y hasta la cultura organizacional.

4.9 ESPACIOS FÍSICOS, DIGITALES E INMERSIVOS

Resulta tentador ordenar estos tres tipos de espacio como si fueran una escala evolutiva, primero lo físico, después lo digital y al final lo inmersivo; sin embargo, la evidencia apunta en otra dirección, ya que las organizaciones más efectivas aprenden a combinarlos según lo que exige cada actividad, en lugar de sustituir un tipo de espacio por otro. Hay interacciones para las que el espacio físico sigue siendo insustituible, como la confianza que se construye al compartir un café o la lectura del lenguaje corporal en una negociación delicada, mientras que la coordinación cotidiana la resuelve con una eficiencia razonable el espacio digital tradicional, conformado por videollamadas, documentos compartidos y mensajería; por su parte, el espacio inmersivo, es decir, lo que Ball (2022) describe como mundos tridimensionales renderizados en tiempo real, abre una tercera posibilidad que ninguno de los otros dos cubre bien, la cual tiene que ver con la sensación de presencia compartida en un entorno común sin necesidad de coincidir físicamente.

Para que este tercer tipo de espacio funcione a escala organizacional, Lee et al. (2022) identifican tres habilitadores tecnológicos críticos, los cuales son las redes, los sistemas y los usuarios. Las redes deben soportar la latencia y el ancho de banda que exige la renderización en tiempo real y los sistemas tienen que resolver la interoperabilidad entre plataformas distintas, mientras que los usuarios, el elemento más difícil de estandarizar, deben desarrollar la alfabetización necesaria para desenvolverse con naturalidad en un entorno que todavía les resulta poco familiar. Sin estos tres elementos alineados, el espacio inmersivo se queda en una promesa tecnológica que no llega a convertirse en una herramienta organizacional real.

El trabajo de Bailenson (2021) ayuda a explicar por qué el espacio digital tradicional, a pesar de su eficiencia, no logra sustituir del todo ni al espacio físico ni al inmersivo. Su investigación sobre lo que llama sobrecarga no verbal documenta que la videollamada convencional exige un esfuerzo cognitivo considerable para producir e interpretar señales no verbales de forma artificial, tal es el caso de mantener contacto visual con una cámara en lugar de con una persona, permanecer dentro del encuadre y verse a uno mismo constantemente en pantalla. Dicho esfuerzo, sostenido durante horas de reuniones consecutivas, explica buena parte de la fatiga que muchas organizaciones ya identifican como un costo oculto del trabajo remoto intensivo. Cabe destacar que ni el espacio físico ni el inmersivo comparten exactamente ese mismo problema, aunque por razones distintas, puesto que en el primero las señales no verbales ocurren de forma

natural y el segundo, al menos en teoría, permite representarlas a través de un avatar sin la carga adicional de la autovigilancia constante.

4.10 LA ORGANIZACIÓN SIN FRONTERAS GEOGRÁFICAS

Cuando el espacio de trabajo deja de estar anclado a un lugar, la geografía deja de ser, por primera vez en la historia moderna del trabajo, una variable determinante en decisiones que antes dependían por completo de ella, como lo son dónde contratar talento, dónde ubicar un equipo o qué zona horaria define el horario laboral “normal” de la organización.

Esta transición tiene implicaciones que rebasan lo logístico, ya que reconfigura el propio concepto de organización. Ball (2022) advierte que el metaverso, y por extensión la infraestructura organizacional que se construye sobre lógicas similares, presenta una oportunidad para usuarios y plataformas, pero también para nuevas reglas, estándares y cuerpos de gobernanza que hasta ahora no existían, esto debido a que los marcos legales y culturales tradicionales se diseñaron pensando en organizaciones ancladas a un territorio específico. En este sentido, una empresa que opera sin fronteras geográficas fijas tiene que resolver preguntas que antes ni siquiera se planteaba, entre ellas bajo qué legislación laboral opera cada colaborador, cómo se construye una cultura compartida entre personas que nunca coinciden en un mismo huso horario y cómo se garantiza la equidad entre quienes tienen acceso a una infraestructura digital robusta y quienes no.

Los datos de Deloitte (2025) permiten ver una de las fricciones que pueden acompañar a esta reconfiguración. En su informe *Global Human Capital Trends* se encontró que más de la mitad de los trabajadores y líderes encuestados en 93 países expresa preocupación por la creciente dificultad para distinguir con claridad qué trabajo realiza una persona y qué trabajo queda mediado por sistemas digitales y de inteligencia artificial, ambigüedad que puede acentuarse cuando la organización deja de operar desde un espacio físico compartido. Dicha ambigüedad difícilmente se resuelve solo con una mejor tecnología de coordinación, por lo que la organización tiene que redefinir de forma explícita qué significa pertenecer a un equipo cuando la geografía deja de ser el elemento que, históricamente, definía sus límites.

4.11 NUEVAS FORMAS DE PERTENENCIA ORGANIZACIONAL

Uno de los cambios más profundos, y de los menos visibles en las conversaciones sobre transformación digital, tiene que ver con lo que ocurre con el sentido de pertenencia cuando desaparece el espacio físico compartido. Durante mucho tiempo, buena parte de la identidad organizacional se construía ahí mismo, ya fuera en los pasillos, en las conversaciones informales junto al café o en la simple repetición diaria de compartir un mismo edificio con las mismas personas.

Feng et al. (2024) sugieren que la claridad conceptual sobre la manera en que el metaverso puede sostener experiencias de comunidad seguirá evolucionando junto con el desarrollo tecnológico. Los autores también señalan de forma explícita que la preparación de quienes lideran estos entornos, docentes en su estudio y líderes organizacionales en el contexto que aquí se analiza, necesita mucha más atención de la que ha recibido hasta ahora. Se trata de una advertencia relevante, ya que el uso de la tecnología para sostener presencia compartida a distancia ya existe y sigue avanzando; sin embargo, la tecnología por sí sola no resuelve la construcción de un sentido de pertenencia real de dentro de ese entorno.

El caso de “Mi Diario Digital de Aprendizaje”, documentado por Robles Ornelas y Páez López (2026), ofrece una pista sobre la manera en que podría construirse ese sentido de pertenencia cuando el espacio físico compartido ya no está garantizado. En lugar de intentar replicar la experiencia de un salón de clases a distancia, el proyecto creó un ritual digital propio, el cual consistía en el registro diario y reflexivo del propio aprendizaje; según lo reportan los estudiantes participantes, dicho ritual generó un sentido de continuidad y autoconciencia que la simple asistencia presencial no necesariamente producía. Se trata de un ejemplo, a pequeña escala, de un principio que probablemente aplica también al contexto organizacional, el cual consiste en que la pertenencia en un entorno distribuido no se recupera imitando lo que antes ocurría en la oficina física; lo que hace falta es diseñar de forma deliberada nuevos rituales compartidos, propios del entorno digital, que cumplan la misma función psicológica de una manera distinta, lo cual requiere una intención organizacional explícita y sostenida en el tiempo sobre la forma

en que se cultivan la confianza, la identidad compartida y el sentido de comunidad cuando ya no se puede dar por sentado que las personas coincidan físicamente. Es probable que las organizaciones que atiendan esta distinción de manera temprana enfrenten menos dificultades que aquellas que se limiten a trasladar sus prácticas presenciales a una pantalla sin rediseñar los mecanismos que sostienen la pertenencia.

Chapter 5 / Capítulo 5

AI, metaverse and sustainability in the new
organizational design
ISBN: 978-9915-9928-8-4
DOI: 10.62486/978-9915-9928-8-4.Ch05
Pages: 33-37



© 2026 The authors. This is an open access
book chapter distributed under the terms of
the Creative Commons Attribution (CC BY)
4.0 License.

The metaverse as a new workspace and collaboration

El metaverso como nuevo espacio de trabajo y colaboración

Jesús Alejandro Vázquez Meza¹  

Víctor Manuel Martínez García²  

Yennifer Díaz Romero³  

Manuel Iván Tostado Ramírez⁴  

Rogelio Estrada Lizárraga⁴  

José de Jesús Reyes Machain²  

1 Universidad Autónoma de Sinaloa, Unidad Académica Preparatoria Rubén Jaramillo, Vicerrectoría, Sinaloa, México

2 Universidad Autónoma de Sinaloa, Facultad de Arquitectura y Diseño, Vicerrectoría, Sinaloa, México

3 Universidad Autónoma de Sinaloa, Facultad de Ciencias de la Educación, Vicerrectoría, Sinaloa, México

4 Universidad Autónoma de Sinaloa, Facultad de Informática, Vicerrectoría, Sinaloa, México

5 Universidad Autónoma de Sinaloa, Facultad de Informática, Vicerrectoría, Sinaloa, México

ABSTRACT

This chapter examines the concept of the metaverse and its potential applications to organizational work and collaboration. It first distinguishes immersive and persistent environments from other forms of digital communication by considering their technical and social characteristics. The roles of avatars, presence, and personal representation in virtual spaces are discussed. On this basis, the chapter explores meetings, projects, and activities involving distributed teams through immersive platforms. It considers opportunities for remote interaction and coordination among participants who do not share a physical location. The review also emphasizes the technical, economic, and human limitations of these experiences and avoids equating a novel interface with proven organizational improvement. Accessibility, adoption, usefulness, and participation are relevant to assessing the value of virtual environments. The metaverse is presented as a collection of possibilities to be evaluated against specific organizational needs, not as the inevitable replacement of face-to-face relationships. The discussion concludes that applications should be connected to clear collaborative purposes and examined in light of the practical conditions under which they operate.

Keywords: Metaverse; Immersive environments; Avatars; Virtual collaboration; Distributed work

RESUMEN

El capítulo estudia el concepto de metaverso y sus posibles aplicaciones en el trabajo y la colaboración organizacional. Comienza por distinguir los entornos inmersivos y persistentes de otras formas de comunicación digital, atendiendo a las características técnicas y sociales que justifican esa diferenciación. Analiza la función de los avatares, la sensación de presencia y las formas de representación personal en espacios virtuales. A partir de esas nociones, examina reuniones, proyectos y actividades de equipos distribuidos que pueden realizarse mediante plataformas inmersivas. El texto considera las oportunidades para interactuar a distancia y coordinar participantes que no comparten un mismo lugar físico. También subraya los límites técnicos, económicos y humanos de estas experiencias y evita equiparar la novedad de una interfaz con una mejora organizacional demostrada. La discusión incluye aspectos de accesibilidad, adopción, utilidad y condiciones de participación, en la medida en que determinan el valor de los entornos virtuales. El metaverso se presenta como un conjunto de posibilidades que deben evaluarse según necesidades concretas, no como reemplazo inevitable de toda relación presencial. La conclusión destaca la importancia de vincular cada aplicación con objetivos de colaboración verificables. De esa forma, el capítulo diferencia la presencia virtual de la cooperación efectiva. Plantea considerar los recursos necesarios, el grado de participación y la calidad de las interacciones antes de atribuir beneficios a una experiencia inmersiva.

Palabras clave: Metaverso; Entornos inmersivos; Avatares; Colaboración virtual; Trabajo distribuido

5.1 ¿QUÉ ENTENDEMOS POR METAVERSO?

Pocas palabras han generado tanta confusión conceptual en tan poco tiempo como “metaverso”. Hay quienes lo consideran sinónimo de videojuegos, haciendo alusión a un mundo virtual, llevado a un nivel de sofisticación mayor, mientras que otros lo ven como una promesa de marketing tecnológico sin sustancia real, e incluso hay quienes lo describen como la infraestructura llamada a suceder a internet tal como se conoce en la actualidad. Por ello, antes de discutir su potencial organizacional, conviene fijar una definición con cierto rigor.

La definición más citada y trabajada del término es, probablemente, la de Ball (2022), quien describe al metaverso como una red masivamente escalada e interoperable de mundos virtuales tridimensionales renderizados en tiempo real, que pueden experimentarse de manera sincrónica y persistente por un número prácticamente ilimitado de usuarios, cada uno con su propio sentido de presencia, y con continuidad de datos como identidad, historial, pertenencias, comunicaciones y pagos. En dicha definición cada palabra importa, esto porque “masivamente escalada” deja fuera cualquier entorno virtual limitado a un puñado de usuarios, del mismo modo que “interoperable” excluye a las plataformas cerradas que no se conectan entre sí y “persistente” descarta cualquier experiencia que se reinicie o desaparezca cuando el usuario se desconecta. Desde la perspectiva organizacional, el elemento más importante es quizás la “continuidad de datos”, ya que significa que lo que ocurre dentro del entorno tiene consecuencias reales y acumulativas, lejos de ser una simulación desechable.

Como contrapunto a esta definición ambiciosa, Feng et al. (2024) recuerdan en su revisión sistemática sobre metaverso y educación que el término, lejos de ser una novedad reciente, lleva más de década y media como objeto de investigación empírica, puesto que en su análisis se identificaron estudios desde 2008 que exploraban entornos virtuales tridimensionales con fines educativos y colaborativos, mucho antes de que la palabra “metaverso” se popularizara como categoría de marketing tecnológico en 2021 y 2022. Este dato es importante porque desmonta una percepción errónea extendida, la cual consiste en creer que la tecnología detrás del metaverso corporativo actual nació de la nada hace un par de años, cuando en realidad se construye sobre una base de investigación y experimentación considerablemente más madura de lo que sugiere la cobertura mediática reciente.

Por su parte, Lee et al. (2022) complementan esta definición con un enfoque más técnico, esto al describir el metaverso como un espacio mixto físico-virtual en donde el ciberespacio actual da su siguiente salto evolutivo. Sin embargo, más allá de la definición en sí, su aporte más útil es la insistencia en que el metaverso no constituye una tecnología única, puesto que lo describen como la convergencia de varias, las cuales son las redes capaces de sostener interacción en tiempo real, los sistemas de renderización e interoperabilidad y los usuarios con la alfabetización necesaria para participar en él con naturalidad. Sin esa convergencia, lo que existe es apenas una versión primitiva y fragmentada de lo que el término promete.

5.2 ENTORNOS INMERSIVOS Y PERSISTENTES

Si hay una palabra clave en este capítulo es persistencia, por lo que conviene detenerse en ella. Un entorno inmersivo tradicional, como una videollamada con avatares, deja de existir en el momento en que todos se desconectan, mientras que un entorno persistente continúa existiendo, cambiando y acumulando historia incluso cuando nadie lo está observando en ese momento.

Dicha distinción tiene una implicación organizacional concreta. Una sala de reuniones virtual que solo es inmersiva, sin ser persistente, funciona como una videollamada con más adornos visuales, útil pero conceptualmente idéntica a lo que ya existía; en cambio, si esa misma sala es persistente, puede conservar pizarras de trabajo, prototipos tridimensionales y configuraciones específicas del equipo tal como quedaron la última vez, disponibles para retomarse en cualquier momento sin necesidad de reconstruir el contexto desde cero. Ball (2022) enfatiza que esta continuidad de identidad, historial y objetos es lo que distingue a un verdadero entorno de metaverso de una simple sala de videoconferencia con gráficos tridimensionales. En términos prácticos, esa combinación de identidad, historial y objetos persistentes es lo que separa a un entorno que se visita de vez en cuando de otro que se habita como espacio de trabajo.

5.3 AVATARES, PRESENCIA Y REPRESENTACIÓN

En cierto sentido, el avatar es el elemento más visible y, al mismo tiempo, el menos comprendido de la colaboración en entornos inmersivos. Además de ser una representación gráfica de la persona, funciona como el vehículo a través del cual esa persona proyecta su presencia dentro de un entorno compartido; es por ello que, para efectos de colaboración, lo que importa es la presencia, no la imagen.

Lee et al. (2022) identifican a los usuarios como uno de los tres habilitadores tecnológicos críticos del metaverso, lo cual cobra sentido si se considera que la experiencia de presencia depende en gran medida de la manera en que cada persona se representa y se percibe dentro del entorno. Esto plantea preguntas organizacionales que antes no existían, tal es el caso de qué tan fiel debe ser un avatar a la apariencia real de la persona para que la colaboración se perciba como auténtica, o si la dinámica de poder dentro de un equipo cambia cuando todos los avatares cuentan con una presencia visual más equitativa que en una videollamada tradicional, en donde la calidad de la cámara o la iluminación de cada quien varía considerablemente. Si bien estas preguntas no tienen todavía respuestas consolidadas, cualquier organización que empiece a experimentar con espacios inmersivos de trabajo tendrá que empezar a formularse las.

Bailenson (2021), director del Laboratorio de Interacción Humana Virtual de Stanford, ofrece una explicación teórica influyente sobre el cansancio de las videollamadas, misma que ayuda a entender por qué representar la presencia a través de un avatar puede resultar, de manera paradójica, menos agotador. Su argumento sobre lo que llama “sobrecarga no verbal” identifica varios factores específicos detrás del cansancio que produce videoconferenciar durante horas, como lo son el contacto visual prolongado y a corta distancia, la carga cognitiva de procesar señales no verbales de varias personas a la vez y lo que describe como el efecto de “espejo todo el día”, es decir, verse a uno mismo constantemente en pantalla, algo que genera ansiedad de autoevaluación (Bailenson, 2021).

5.4 COLABORACIÓN VIRTUAL

La colaboración virtual no nació con el metaverso, puesto que lleva décadas desarrollándose a través del correo electrónico, las videollamadas y los documentos compartidos; sin embargo, los entornos inmersivos añaden una dimensión que esas herramientas no lograban replicar bien, la cual tiene que ver con la sensación de compartir un espacio físico común, con todo lo que eso implica en términos de atención sostenida, lenguaje corporal simulado y sentido de copresencia.

Lin et al. (2022) sitúan esta capacidad dentro de un marco más amplio, esto al describir al metaverso como la convergencia de múltiples tecnologías, entre ellas big data, interacción, inteligencia artificial, diseño de juegos, computación en internet, internet de las cosas y blockchain, mismas que en conjunto generan un nuevo tipo de espacio social capaz de sostener colaboración compleja. Si bien su análisis se centra en el ámbito educativo, resulta aplicable al organizacional, ya que la colaboración virtual profunda depende de que varias piezas técnicas distintas trabajen en conjunto de forma coherente, más que de una sola tecnología estelar. Es así como una organización puede tener acceso a hardware de realidad virtual de última generación y, aun así, no lograr una colaboración efectiva si la infraestructura de red, los sistemas de gestión de datos o la alfabetización digital de su equipo no están igualmente desarrollados.

5.5 REUNIONES, PROYECTOS Y EQUIPOS DISTRIBUIDOS

Dentro de una organización, el caso de uso más inmediato y tangible de estos entornos es, probablemente, la reunión de trabajo. Sin embargo, limitar la conversación a las reuniones subestima el potencial de estos espacios para sostener proyectos completos, más allá de encuentros puntuales.

Un proyecto que se desarrolla dentro de un entorno persistente puede mantener, de forma continua y accesible para todo el equipo distribuido, prototipos tridimensionales en evolución, líneas de tiempo visuales y espacios de trabajo específicos por fase del proyecto, todo disponible en el lugar en donde se dejó y sin necesidad de reconstruir el contexto en cada sesión. Con base en el trabajo de Lin et al. (2022) se puede plantear, además, que la combinación de inteligencia artificial y entornos inmersivos permite que estos espacios se adapten de manera dinámica al tipo de tarea que el equipo realiza en cada momento, algo que ninguna sala de reuniones física, ni siquiera la más sofisticada, puede ofrecer con la misma flexibilidad. Con lo anterior se puede

ver que el cambio más significativo tiene que ver con que un proyecto completo pueda vivir, de forma persistente y accesible, dentro de un espacio compartido que trasciende la geografía de quienes lo integran, más que con hacer que una reunión “se vea mejor”.

5.6 OPORTUNIDADES Y LÍMITES ORGANIZACIONALES

Para cerrar este capítulo conviene reconocer que estos entornos, por prometedores que sean, todavía enfrentan límites reales. Al respecto, Lee et al. (2022) identifican de forma explícita una serie de desafíos técnicos y de ecosistema que todavía distan de estar resueltos, los cuales van desde las limitaciones de la infraestructura de red hasta la falta de estándares de interoperabilidad entre plataformas. Por ello, ninguna organización debería adoptar estos entornos asumiendo que la tecnología ya maduró por completo, puesto que en muchos sentidos sigue en construcción.

La propia investigación sobre fatiga en videoconferencias aporta un matiz importante a esta cautela. Estudios recientes, como el de Frontzkowski et al. (2025) que ponen a prueba el supuesto de que la realidad virtual reduce el cansancio de las reuniones virtuales encuentran resultados menos concluyentes de lo esperado, puesto que la tecnología actual de realidad virtual no necesariamente resuelve por completo la fatiga de las videoconferencias, además de que factores como avatares poco naturales o una expresividad facial limitada pueden generar formas distintas de agotamiento, aunque no necesariamente menores. Con esto se puede ver que la promesa teórica de un entorno inmersivo, es decir, mayor presencia y menor sobrecarga no verbal, todavía no siempre se traduce en una mejor experiencia en la práctica cotidiana.

Dicho esto, las oportunidades son igual de reales. Ball (2022) plantea que el metaverso representa una oportunidad para usuarios y desarrolladores, pero también para que surjan nuevas reglas, estándares y estructuras de gobernanza más confiables que las actuales, lo que él describe, con un optimismo deliberado, como una posible “carrera hacia la confianza”. Para una organización, esto significa que adoptar espacios de colaboración inmersiva es una decisión tecnológica y, al mismo tiempo, una apuesta a que este tipo de infraestructura madurará con el tiempo lo suficiente como para sostener, además de reuniones más inmersivas, formas de trabajo colectivo que hoy apenas se empiezan a imaginar. En ese sentido, la decisión que enfrenta cada organización incluye, además de saber si esta tecnología estará lista en algún momento, qué tan pronto conviene empezar a familiarizarse con su uso.

Chapter 6 / Capítulo 6

AI, metaverse and sustainability in the new
organizational design
ISBN: 978-9915-9928-8-4
DOI: 10.62486/978-9915-9928-8-4.Ch06
Pages: 38-42



© 2026 The authors. This is an open access
book chapter distributed under the terms of
the Creative Commons Attribution (CC BY)
4.0 License.

Learning, innovation and organizational experience in immersive environments

Aprendizaje, innovación y experiencia organizacional en entornos inmersivos

Jesús Alejandro Vázquez Meza¹ 

Víctor Manuel Martínez García² 

Yennifer Díaz Romero³ 

Manuel Iván Tostado Ramírez⁴ 

Rogelio Estrada Lizárraga⁴ 

José de Jesús Reyes Machain² 

1 Universidad Autónoma de Sinaloa, Unidad Académica Preparatoria Rubén Jaramillo, Vicerrectoría, Sinaloa, México

2 Universidad Autónoma de Sinaloa, Facultad de Arquitectura y Diseño, Vicerrectoría, Sinaloa, México

3 Universidad Autónoma de Sinaloa, Facultad de Ciencias de la Educación, Vicerrectoría, Sinaloa, México

4 Universidad Autónoma de Sinaloa, Facultad de Informática, Vicerrectoría, Sinaloa, México

5 Universidad Autónoma de Sinaloa, Facultad de Informática, Vicerrectoría, Sinaloa, México

ABSTRACT

This chapter examines immersive environments for learning, innovation, and the experiences of people within organizations. Immersive training and simulation are considered as ways to represent working situations and support experimentation. The discussion connects these formats with skills development, practice, and exchange among participants. It also explores collaborative innovation and the experiences of employees and customers. Organizational examples described in the source text are considered, while specific initiatives are distinguished from general evidence of effectiveness. A recurring argument is that the value of a platform depends on the educational or collaborative quality of the activities it supports, not merely on the feeling of immersion. The chapter also considers the difficulty of building community through technologically mediated interaction and the potential effects of communication overload. Training, experimentation, and belonging are therefore connected to choices about design and management. The discussion suggests that immersive environments need clear purposes, appropriate conditions for participation, and a critical assessment of their benefits and limitations before they become permanent organizational practices.

Keywords: Immersive learning; Simulation; Collaborative innovation; Employee experience; Organizational community

RESUMEN

El capítulo examina la utilización de entornos inmersivos para el aprendizaje, la innovación y la experiencia de las personas dentro de las organizaciones. Aborda la capacitación inmersiva y la simulación como recursos que pueden representar situaciones de trabajo y facilitar la experimentación. Discute cómo estos formatos se vinculan con el desarrollo de competencias, la práctica de habilidades y el intercambio entre participantes. También considera su relación con procesos de innovación colaborativa y con experiencias de empleados y clientes. La revisión incorpora ejemplos organizacionales discutidos en el propio texto, pero distingue las iniciativas descritas de una prueba general de efectividad. Un eje transversal es que el valor de una plataforma depende de la calidad pedagógica o colaborativa de las actividades, no solo de la sensación de inmersión. Asimismo, el capítulo analiza las dificultades para construir comunidad mediante interacciones mediadas tecnológicamente y los posibles efectos de la sobrecarga comunicativa. La argumentación conecta formación, experimentación y pertenencia con decisiones sobre diseño y gestión. Su aporte principal consiste en mostrar que los entornos inmersivos requieren propósitos claros, condiciones de acceso y evaluación crítica de sus beneficios y límites antes de integrarse como prácticas permanentes. El capítulo relaciona el diseño de estas actividades con las necesidades de quienes participan y con las condiciones materiales de acceso. Por ello, insiste en observar qué experiencias favorecen el aprendizaje, la coordinación y la construcción de vínculos.

Palabras clave: Aprendizaje inmersivo; Simulación; Innovación colaborativa; Experiencia del empleado; Comunidad organizacional

6.1 CAPACITACIÓN INMERSIVA

El enseñar un procedimiento complejo a través de su descripción en un manual y el enseñarlo dejando que alguien lo practique en un entorno simulado son experiencias de aprendizaje notablemente distintas, y casi cualquier persona que haya vivido ambas coincidiría en cuál de las dos retiene mejor. Esa intuición, que llevaba décadas confirmándose en simuladores de vuelo y en entrenamientos militares, es la que ahora empieza a extenderse hacia la capacitación organizacional convencional por medio de entornos inmersivos accesibles.

De Felice et al. (2023) presentan una revisión sistemática sobre metaverso y educación, misma que se construyó siguiendo el protocolo PRISMA a partir de 84 documentos indexados en Scopus. En ella se encontró que el campo dominante de aplicación, por un margen considerable, son las ciencias de la computación, seguidas de la ingeniería y las ciencias sociales, aunque los autores anotan de forma explícita que el potencial de la tecnología se extiende también a la medicina y a otros campos en donde la práctica situada resulta decisiva. Cabe destacar que su revisión se enfoca de manera deliberada en el vínculo entre el metaverso y el futuro del trabajo, además de documentar que la capacitación inmersiva es objeto de una producción académica interdisciplinaria considerable, con una presencia notable de investigación proveniente de distintos países, lo que aleja a la capacitación inmersiva de ser un nicho experimental.

Shi y Park (2024) aportan un matiz importante a esta imagen entusiasta. Su revisión sistemática, centrada en el uso del metaverso en la educación formal, abarca 34 estudios de entre 2009 y 2023 y encuentra que, si bien la investigación se concentra con claridad en la educación superior, pocos estudios adoptan diseños experimentales robustos. Se trata de una advertencia útil para cualquier organización que esté considerando invertir en capacitación inmersiva, ya que en muchos casos la evidencia de que funciona mejor que los métodos tradicionales todavía no está tan consolidada como sugiere el entusiasmo que rodea a la tecnología.

6.2 SIMULACIÓN Y EXPERIMENTACIÓN

Para efectos de formación, la ventaja más evidente de un entorno inmersivo es la posibilidad de simular situaciones que serían costosas, peligrosas o hasta imposibles de replicar en el mundo físico, desde una negociación de alto riesgo hasta una falla de equipo industrial, esto sin las consecuencias reales que tendría cometer un error en el contexto verdadero.

En su revisión sistemática sobre marcos de aprendizaje inmersivo, Fernandes et al. (2023) identifican que la literatura existente distingue dos tipos de marcos, los cuales son los modelos teóricos que explican la relación entre inmersión y aprendizaje, y las guías de desarrollo de software orientadas a construir estas experiencias. Su análisis muestra algo importante sobre el estado del campo, ya que, a pesar de que las tecnologías inmersivas llevan años en uso educativo, existen relativamente pocos trabajos que ofrezcan marcos sólidos para diseñar y evaluar con rigor estos entornos de simulación, lo que deja a muchas organizaciones con la tarea de diseñar sus propios criterios de éxito sin un consenso metodológico claro al cual recurrir.

Lo anterior no invalida el valor de la simulación como herramienta de aprendizaje organizacional, puesto que la evidencia acumulada sobre su efectividad en dominios específicos, tal es el caso del entrenamiento técnico o de seguridad, es sólida. Lo que sí sugiere esta falta de marcos es que las organizaciones deben ser cuidadosas al diseñar sus propios programas de simulación inmersiva, en lugar de asumir que cualquier experiencia tridimensional produce de manera automática mejores resultados de aprendizaje que un método más convencional.

6.3 DESARROLLO DE COMPETENCIAS

La decisión de una persona de usar, o de rechazar, un entorno inmersivo para desarrollar una competencia nueva depende de factores identificables que ya se han estudiado en la literatura sobre adopción tecnológica. Al-kfairy et al. (2024), en su revisión sistemática de 35 estudios empíricos sobre la adopción del metaverso en contextos educativos, aplican marcos teóricos ya conocidos, como lo son el Modelo de Aceptación Tecnológica, la Teoría Unificada de Aceptación y Uso de la Tecnología y la Teoría del Comportamiento Planificado, esto para responder a esa misma pregunta.

Sus hallazgos identifican factores concretos que influyen en la adopción en los niveles individual, tecnológico y ambiental, los cuales son la expectativa de esfuerzo, la intención conductual, la autoeficacia, el disfrute y la sensación de inmersión (Al-kfairy et al., 2024). Este

último punto merece una atención especial desde la perspectiva del desarrollo de competencias organizacionales, esto porque el disfrute y la inmersión, lejos de ser variables accesorias, se asocian con que una persona se comprometa con el proceso de aprendizaje, en lugar de solo tolerarlo. Una organización que diseñe capacitación inmersiva sin prestar atención deliberada a estos factores, es decir, que trate el entorno virtual como un simple contenedor técnico para transmitir información, corre el riesgo de replicar, con un mayor costo tecnológico, los mismos problemas de desconexión que ya tenía con la capacitación tradicional.

El caso de Accenture, documentado por Kohli (2025) y que se discute con mayor detalle más adelante en este capítulo, aporta una confirmación práctica de este mismo principio a gran escala. La firma no se limitó a instalar la tecnología de realidad virtual y esperar resultados; por el contrario, diseñó de forma deliberada la experiencia de incorporación alrededor de factores de disfrute e inmersión, lo que se tradujo en una mejora del 33 % en la retención de contenido y en una calificación de satisfacción de 4.7 sobre 5 entre 150 000 nuevas contrataciones. Este resultado coincide con lo que se desprende de Al-kfairy et al. (2024) cuando una organización toma en serio los factores de adopción individual, en lugar de asumir que la tecnología por sí sola produce un compromiso genuino con el aprendizaje.

6.4 INNOVACIÓN COLABORATIVA

Además de servir para transmitir conocimiento ya existente, los entornos inmersivos pueden convertirse en espacios en donde equipos completos generan ideas nuevas de forma colaborativa, esto al aprovechar la sensación de copresencia que permiten incluso entre personas geográficamente distantes.

Lin et al. (2022), cuyo trabajo también se revisó en el capítulo anterior, subrayan que la convergencia entre inteligencia artificial, análisis de datos masivos y entornos inmersivos abre posibilidades de innovación colaborativa que ninguna de esas tecnologías lograría por separado. A partir de ello se puede pensar en sistemas capaces de sugerir conexiones entre ideas dispersas del equipo, de visualizar prototipos en tiempo real mientras se discuten y de mantener persistente todo el proceso creativo para que ningún avance se pierda entre una sesión y otra. Esta capacidad de sostener un proceso de innovación de forma continua y visualmente compartida, en lugar de fragmentarlo entre correos, documentos y reuniones aisladas, es probablemente uno de los usos organizacionales menos reconocidos de estos entornos.

6.5 EXPERIENCIA DEL EMPLEADO Y DEL CLIENTE

Hasta ahora la conversación se ha centrado en el aprendizaje interno; sin embargo, los entornos inmersivos también están reconfigurando la manera en que las organizaciones diseñan la experiencia de quienes las integran y de quienes consumen sus productos o servicios.

De Felice et al. (2023) documentan que el impacto potencial del metaverso se extiende más allá del ámbito estrictamente educativo, alcanzando terrenos como la medicina, los materiales y otros sectores en los que la experiencia situada, es decir, vivir algo en lugar de solo leerlo, resulta decisiva para el aprendizaje y la satisfacción. Trasladado al terreno de la experiencia del empleado, esto sugiere que los procesos de incorporación, de desarrollo de carrera y hasta de bienestar laboral pueden beneficiarse de experiencias inmersivas diseñadas con la misma atención al detalle que hoy se dedica al diseño de productos digitales para clientes. En sentido inverso, la experiencia del cliente también empieza a beneficiarse de espacios en donde puede experimentar un producto o servicio de forma inmersiva antes de decidirse, algo que ninguna fotografía o video convencional logra replicar del todo.

El caso de Accenture ofrece, probablemente, la evidencia empresarial más completa disponible hasta ahora sobre lo que esto significa en la práctica. Kohli (2025) documenta la manera en que la firma desplegó su plataforma inmersiva Nth Floor a través de una fuerza laboral de 700 000 empleados, así como el uso de dicha plataforma para incorporar a 150 000 nuevas contrataciones mediante 60 000 dispositivos de realidad virtual, en donde entre los resultados documentados se encuentran una mejora del 33 % en la retención de lo aprendido durante el proceso de incorporación y una calificación de satisfacción de los empleados de 4.7 sobre 5 (Kohli, 2025). Se trata de un caso que confirma, a una escala considerablemente mayor que la de la mayoría de los estudios experimentales revisados en este capítulo, que la experiencia del empleado en entornos inmersivos puede sostenerse de forma exitosa incluso en despliegues

masivos, siempre que la organización invierta en la planeación estratégica y en la gestión del cambio con la misma seriedad que dedica a la tecnología misma.

6.6 NUEVAS FORMAS DE CONSTRUIR COMUNIDAD

Al igual que el anterior, este capítulo cierra con un tema que la tecnología por sí sola no resuelve, el cual tiene que ver con la construcción de una comunidad genuina en un entorno que no depende de la coincidencia física.

Shi y Park (2024) ofrecen una pista importante al señalar que la exploración de las experiencias estudiantiles dentro de estos entornos sigue siendo un área de investigación incipiente. Esto sugiere que entender cómo las personas construyen vínculos genuinos dentro de espacios inmersivos, más allá de completar tareas en ellos, es todavía un terreno poco explorado, tanto en la educación como, por extensión, en el ámbito organizacional.

Por su parte, Fernandes et al. (2023) añaden que los elementos que componen un buen marco de aprendizaje inmersivo abarcan aspectos tecnológicos, pedagógicos y psicológicos, lo que refuerza la idea de que construir comunidad dentro de estos entornos es un problema que no se resuelve únicamente con mejor tecnología.

Bailenson (2021), cuyo trabajo sobre sobrecarga no verbal ya se discutió en el capítulo anterior, aporta un último elemento relevante para esta discusión. Su investigación sugiere que buena parte del agotamiento que las personas experimentan en la videoconferencia convencional proviene de señales no verbales mal calibradas, es decir, del contacto visual forzado y de la ausencia de movimiento natural. A partir de ello se puede pensar que esas mismas señales también generan desconexión, puesto que dificultan que se forme un sentido genuino de conexión social, incluso cuando la comunicación técnica funciona sin problemas. Si estos entornos inmersivos logran resolver aunque sea en parte, de la pesadez no verbal, es probable que, además de mejorar la eficiencia de las reuniones, contribuyan a sostener algo parecido al sentido de comunidad que antes solo parecía posible al compartir un espacio físico; sin embargo, esto requiere aplicar, a un espacio que puede compartirse sin que las personas coincidan físicamente en el mismo lugar, el mismo tipo de diseño intencional que cualquier comunidad humana ha necesitado siempre, es decir, atención a la pertenencia, a la confianza y a los rituales compartidos.

Chapter 7 / Capítulo 7

AI, metaverse and sustainability in the new
organizational design
ISBN: 978-9915-9928-8-4
DOI: 10.62486/978-9915-9928-8-4.Ch07
Pages: 43-48



© 2026 The authors. This is an open access
book chapter distributed under the terms of
the Creative Commons Attribution (CC BY)
4.0 License.

Artificial intelligence ethics and organizational responsibility

Ética de la inteligencia artificial y responsabilidad organizacional

Jesús Alejandro Vázquez Meza¹ 

Víctor Manuel Martínez García² 

Yennifer Díaz Romero³ 

Manuel Iván Tostado Ramírez⁴ 

Rogelio Estrada Lizárraga⁴ 

José de Jesús Reyes Machain² 

1 Universidad Autónoma de Sinaloa, Unidad Académica Preparatoria Rubén Jaramillo, Vicerrectoría, Sinaloa, México

2 Universidad Autónoma de Sinaloa, Facultad de Arquitectura y Diseño, Vicerrectoría, Sinaloa, México

3 Universidad Autónoma de Sinaloa, Facultad de Ciencias de la Educación, Vicerrectoría, Sinaloa, México

4 Universidad Autónoma de Sinaloa, Facultad de Informática, Vicerrectoría, Sinaloa, México

5 Universidad Autónoma de Sinaloa, Facultad de Informática, Vicerrectoría, Sinaloa, México

ABSTRACT

This chapter reviews ethical issues that arise when artificial intelligence influences organizational processes and decisions. The discussion is organized around transparency, accountability, fairness, non-discrimination, and algorithmic bias. It examines the importance of enabling people to understand the scope and limits of AI systems, challenge their outputs, and maintain meaningful oversight. Institutional responsibilities are connected with the privacy of data used or processed by AI applications. Employee well-being is also considered, including possible effects on autonomy, safety, and the experience of work. Ethical principles are not treated as sufficient declarations in themselves; they require procedures, identifiable responsibilities, and controls. The review distinguishes technical performance from the social conditions in which a tool is used. The chapter argues that responsible integration requires clarity about who makes decisions, who reviews them, and who responds to adverse consequences. Data practices and human participation are therefore integral to the ethical assessment of the design and operation of these systems, rather than additional issues to consider only after deployment.

Keywords: AI ethics; Algorithmic bias; Transparency; Privacy; Organizational accountability

RESUMEN

El capítulo revisa los problemas éticos que surgen cuando la inteligencia artificial interviene en procesos y decisiones organizacionales. Ordena la discusión en torno a la transparencia, la responsabilidad, la equidad, la no discriminación y los sesgos algorítmicos. Examina la importancia de que las personas puedan comprender los alcances y límites de los sistemas, cuestionar sus resultados y mantener formas significativas de supervisión. El texto relaciona las obligaciones institucionales con la privacidad de los datos que alimentan o atraviesan las aplicaciones de inteligencia artificial. También incorpora el bienestar de los trabajadores como criterio de evaluación, considerando posibles efectos sobre su autonomía, seguridad y experiencia laboral. Los principios éticos no se presentan como declaraciones suficientes por sí mismas, sino como orientaciones que requieren procedimientos, responsables y controles. La revisión diferencia el desempeño técnico de las condiciones sociales bajo las cuales se utiliza una herramienta. El capítulo sostiene que integrar inteligencia artificial de manera responsable exige identificar quién decide, quién revisa y quién responde ante consecuencias adversas. Asimismo, muestra que el tratamiento de los datos y la participación humana forman parte de la evaluación ética del diseño y la operación de estos sistemas. La exposición examina estas exigencias de manera interdependiente: una práctica transparente puede seguir siendo problemática si reproduce sesgos, y una herramienta precisa puede resultar inadecuada si las personas afectadas no disponen de mecanismos de revisión.

Palabras clave: Ética de la inteligencia artificial; Sesgo algorítmico; Transparencia; Privacidad; Responsabilidad organizacional

7.1 TRANSPARENCIA

Cuando un sistema de inteligencia artificial toma, o influye en, una decisión que afecta a una persona, esa persona tiene derecho a entender, al menos en términos generales, cómo se llegó a esa decisión. Aunque el principio parece de sentido común, en ocasiones resulta difícil de garantizar en la práctica, sobre todo cuando el sistema en cuestión opera con modelos cuya lógica interna ni siquiera sus propios desarrolladores pueden explicar del todo.

En su *Recomendación sobre la Ética de la Inteligencia Artificial*, la UNESCO (2021) establece la transparencia como uno de los principios rectores de cualquier desarrollo o uso responsable de estas tecnologías, junto con la rendición de cuentas, la inclusión y la sostenibilidad. De acuerdo con el documento, la transparencia consiste en garantizar que exista suficiente claridad sobre cómo y cuándo se utiliza la IA, de manera que las personas afectadas puedan comprender sus resultados y, cuando corresponda, cuestionarlos, sin necesidad de exponer aspectos técnicos que probablemente para la mayoría de las personas pasarían desapercibidos.

Barroso Camiade y Pérez Castrejón (2025) sitúan este principio dentro de un debate más amplio sobre los supuestos filosóficos de la propia inteligencia artificial, esto al retomar la crítica clásica de Hubert Dreyfus a las asunciones sobre las que se construyeron los primeros sistemas inteligentes. Su argumento, actualizado al contexto actual, plantea que la opacidad de muchos sistemas de IA, además de ser un problema técnico, es una consecuencia directa de haber construido la tecnología sobre supuestos que sus propios diseñadores nunca sometieron a suficiente escrutinio, de modo que, entendida así, la transparencia deja de ser únicamente una buena práctica de comunicación para convertirse también en una exigencia epistemológica sobre qué tan bien se entiende lo que estos sistemas hacen.

Los datos del *AI Index Report* de la Universidad de Stanford (2025) muestran que esta exigencia sigue lejos de cumplirse en la práctica. El índice de transparencia entre los principales desarrolladores de modelos de frontera pasó de un 37 % en octubre de 2023 a un 58 % en mayo de 2024, lo cual representa una mejora real, aunque la mayoría de estos sistemas todavía está lejos de un nivel de divulgación que permita una auditoría externa sobre sus datos de entrenamiento, su consumo de cómputo y su impacto posterior al despliegue. El mismo informe menciona, además, un aumento en los incidentes de IA reportados formalmente, los cuales llegaron a 233 en 2024, lo que representa un aumento de más del 56 % con respecto al año anterior (Stanford University, 2025). La brecha entre la velocidad con la que se despliegan estos sistemas y la velocidad con la que se documentan sus fallas es, en sí misma, uno de los indicios más claros de por qué la transparencia sigue siendo un principio aspiracional, más que una práctica consolidada.

7.2 RESPONSABILIDAD

El determinar quién responde cuando un sistema de inteligencia artificial se equivoca es uno de los vacíos legales y organizacionales más persistentes en este terreno. Al respecto, Méndez Reyes (2025) documenta que, conforme la IA se expande hacia *chatbots*, redes neuronales y conducción autónoma, entre muchas otras aplicaciones, trae consigo problemas de sesgo, opacidad algorítmica y preocupaciones sobre la privacidad y la seguridad de los datos, mismos que hacen indispensable establecer marcos claros de responsabilidad organizacional.

Por su parte, la UNESCO (2021) plantea que la responsabilidad debe distribuirse de forma proporcional entre quienes diseñan, despliegan y utilizan estos sistemas, con lo cual se evita tanto el extremo de exonerar por completo a la organización que decide implementarlos como el de atribuir toda la responsabilidad a quien programó el algoritmo original. En la práctica, esto significa que ninguna organización puede refugiarse en el argumento de que “el sistema decidió así” para eludir su propia responsabilidad sobre las consecuencias de haber elegido implementar, sin una supervisión adecuada, una herramienta capaz de afectar a personas reales.

El *AI Index Report* de Stanford (2025) documenta que las organizaciones están empezando a internalizar esta exigencia, aunque de forma todavía desigual, y señala que la proporción de empresas que declara no tener ninguna política de inteligencia artificial responsable cayó de forma notable, de 24 a 11 % en apenas un año. Se trata de una señal alentadora; no obstante, el mismo informe advierte que reconocer los riesgos y mitigarlos activamente son cosas distintas. Tal es el caso de una encuesta de McKinsey citada en el índice, la cual muestra que, aunque la mayoría de las organizaciones identifica correctamente los riesgos centrales de la IA responsable, no todas traducen ese reconocimiento en acciones concretas de mitigación. Dicho de otra forma, saber que se es responsable todavía no equivale a actuar en consecuencia.

Barroso Camiade y Pérez Castrejón (2025) añaden un matiz filosófico importante a esta discusión al retomar la distinción clásica entre ética descriptiva y ética normativa, a partir de la cual señalan que buena parte del debate público sobre la responsabilidad de la IA se queda atrapado en describir lo que distintos grupos consideran correcto o incorrecto, sin avanzar hacia principios normativos claros sobre cómo debería actuar una organización en la práctica. Dicha distinción trasciende lo académico, ya que explica por qué tantas organizaciones logran redactar códigos de ética extensos y bien intencionados que, sin embargo, no se traducen en ningún cambio observable en la manera en que se toman las decisiones. Una responsabilidad genuina exige cruzar ese puente entre la descripción de valores y la normatividad operativa, algo que la mayoría de los marcos éticos corporativos todavía no logra del todo.

7.3 EQUIDAD Y NO DISCRIMINACIÓN

Un sistema de inteligencia artificial entrenado con datos que reflejan sesgos históricos, como los de género, origen étnico o clase socioeconómica, tiende, salvo que haya una intervención deliberada, a reproducir y en ocasiones a amplificar esos mismos sesgos en sus resultados. Esto ocurre sin que el sistema sea “malintencionado” en ningún sentido, puesto que aprendió lo que los datos le enseñaron, incluidas sus injusticias.

Guaña-Moya y Chipuxi-Fajardo (2023) analizan este fenómeno desde la intersección entre la ética y la privacidad de los datos, en donde señalan que garantizar la equidad requiere algo más que buenas intenciones declaradas, esto es, principios operativos concretos, como lo son la transparencia, la responsabilidad y el respeto a los derechos fundamentales de las personas, principios que tienen que traducirse en procesos verificables, más allá de las declaraciones de valores corporativos. La UNESCO (2021) refuerza esta idea de que la inclusión sea uno de sus principios centrales, misma que, lejos de ser un gesto simbólico, se entiende como la exigencia activa de que los sistemas de IA no perpetúen ni profundicen las desigualdades existentes.

El *Future of Jobs Report 2025* aporta a esta discusión una dimensión adicional, poco tratada fuera del ámbito laboral, según la cual la equidad algorítmica importa dentro de un sistema de IA ya desplegado, pero también en quién tiene acceso, desde el principio, a formarse en las competencias que esa IA exige. De acuerdo con el informe, el 85 % de los empleadores planea invertir en capacitación en la capacitación de su personal de forma general, en lo que para la capacitación ligada a la IA es del 77% (World Economic Forum, 2025); sin embargo, si esa misma inversión no se distribuye de forma deliberadamente equitativa entre distintos niveles jerárquicos, géneros y regiones dentro de una organización, corre el riesgo de convertirse en un mecanismo que amplía, en lugar de reducir, las brechas existentes entre quienes ya tenían acceso a oportunidades de desarrollo y quienes no. En ese sentido, la equidad algorítmica empieza antes de que un sistema tome su primera decisión, es decir, en quién tuvo la oportunidad de aprender a trabajar con él.

7.4 SESGOS ALGORÍTMICOS

Conviene distinguir el sesgo algorítmico, como fenómeno técnico específico, de la discusión más amplia sobre la equidad. Este sesgo ocurre cuando un modelo produce resultados sistemáticamente distintos a los esperados, y que por lo general resultan perjudiciales, para determinados grupos de personas; en ocasiones quienes diseñaron o implementaron el sistema no se percataron de esta situación en tiempo y forma.

Méndez Reyes (2025) identifica los sesgos algorítmicos como uno de los problemas centrales que la inteligencia artificial trae consigo en su expansión reciente, junto con la falta de transparencia y las preocupaciones sobre la privacidad. Su análisis sugiere que estos sesgos rara vez son producto de una decisión de diseño explícita, ya que suelen colarse a través de los datos de entrenamiento, de variables *proxy* aparentemente neutrales que en realidad se correlacionan con características protegidas o de la falta de diversidad entre quienes diseñan y prueban el sistema antes de su despliegue. Para detectarlos se requieren auditorías activas y deliberadas, lo cual va más allá de la simple confianza en que “el sistema funcionará bien porque fue diseñado con buena intención”.

La evidencia empírica más reciente muestra que ni siquiera los esfuerzos explícitos por eliminar el sesgo garantizan resultados neutrales. El *AI Index Report* de Stanford (2025) documenta que varios de los modelos de lenguaje más avanzados, incluidos sistemas diseñados

de forma deliberada con medidas para reducir sesgos explícitos, siguen exhibiendo sesgos implícitos considerables. Estos modelos asocian con más frecuencia términos negativos con personas afrodescendientes, vinculan a las mujeres con las humanidades en lugar de con disciplinas científicas y tecnológicas y favorecen a los hombres para roles de liderazgo, con lo cual refuerzan patrones discriminatorios de género y raza en la toma de decisiones. Si bien el hallazgo resulta incómodo, es necesario reconocer que mejorar el desempeño de un sistema en pruebas estandarizadas de sesgo no equivale a haber resuelto el problema de fondo, y que una organización que confía ciegamente en esas pruebas puede seguir desplegando decisiones sesgadas sin saberlo.

7.5 SUPERVISIÓN HUMANA

Ninguno de los principios éticos revisados hasta aquí, como lo son la transparencia, la responsabilidad o la equidad, se sostiene sin un mecanismo concreto que lo haga cumplir en la práctica, y ese mecanismo es, casi invariablemente, la supervisión humana activa.

Autio et al. (2024), en el perfil complementario de este mismo marco dedicado a la inteligencia artificial generativa, subrayan que los riesgos de estos sistemas cambian constantemente conforme evolucionan los modelos, por lo que la supervisión humana tiene que concebirse como un proceso continuo que se ajuste a la misma velocidad que la propia tecnología, en lugar de una revisión puntual al momento de la implementación. Una organización que instale un sistema de IA y considere terminado el trabajo de supervisión ha malentendido, de raíz, en qué consiste la responsabilidad ética sobre estas herramientas.

El estudio experimental de Dell'Acqua et al. (2023) ofrece evidencia concreta de por qué esta supervisión continua, más que un formalismo burocrático, es una necesidad práctica medible. En la tarea diseñada para exigir un juicio contextual complejo, fuera de lo que los autores llaman la frontera de capacidad del sistema, los consultores que confiaron en la IA sin una supervisión crítica vieron caer su precisión de forma marcada frente a quienes trabajaron sin ella. A partir de ello se puede ver que la falla estuvo menos en la tecnología que en la ausencia de un mecanismo que alertara a los consultores sobre cuándo debían aplicar un mayor escrutinio a las sugerencias del sistema. Se trata, en esencia, de la misma advertencia que hacen Tabassi y Autio et al. desde el terreno normativo, en el sentido de que, sin una supervisión humana activa y constante, ni el sistema más sofisticado garantiza decisiones correctas.

7.6 PRIVACIDAD

Para funcionar bien, los sistemas de inteligencia artificial necesitan datos, a menudo muchos, y con frecuencia datos personales sensibles. Esa dependencia estructural pone a la privacidad en tensión directa con la utilidad misma de estas herramientas, por lo que resolver esa tensión de forma responsable es uno de los retos éticos más persistentes del sector.

Guaña-Moya y Chipuxi-Fajardo (2023) concluyen su análisis con un llamado directo, el cual consiste en que garantizar un desarrollo responsable y respetuoso de la inteligencia artificial exige tanto reflexión ética como una regulación legal adecuada, lo que incluye la adopción de principios éticos sólidos, la promoción activa de la transparencia y la protección explícita de la privacidad de los datos de las personas. En la misma línea, la UNESCO (2021) coincide en que ningún desarrollo de IA puede considerarse ético si compromete de manera sistemática el derecho de las personas a controlar qué se sabe de ellas y cómo se usa esa información.

7.7 BIENESTAR DE LOS TRABAJADORES

Este capítulo cierra con un principio que rara vez encabeza las listas de prioridades éticas sobre la IA, pero que debería estar mucho más presente, el cual tiene que ver con el impacto de estos sistemas en la salud mental, la seguridad laboral y el sentido de propósito de las personas que trabajan junto a ellos o bajo su supervisión algorítmica.

Méndez Reyes (2025) sitúa esta preocupación dentro de un marco de teoría del discurso, esto al argumentar que la manera en que las organizaciones hablan y deciden sobre la inteligencia artificial tiene consecuencias directas en la forma en que las personas experimentan su propio trabajo cuando esa tecnología entra en juego. Un sistema de IA que optimiza la productividad sin ninguna consideración sobre el bienestar de quienes trabajan con él puede generar ganancias

de corto plazo mientras erosiona, de forma silenciosa y acumulativa, la confianza, la motivación y la salud de la fuerza laboral. En última instancia, más que un ejercicio abstracto de principios bien redactados, la ética de la inteligencia artificial es la pregunta concreta de si las personas que trabajan dentro de una organización terminan mejor o peor como consecuencia de las decisiones que esa organización tomó sobre la manera de usar esta tecnología.

Los datos de Deloitte (2025) muestran que esta preocupación no se limita a los especialistas en ética y que ya es ampliamente compartida por la propia fuerza laboral, tal como se observa en su informe *Global Human Capital Trends*, en el cual se encontró que más de la mitad de los trabajadores y líderes encuestados en 93 países expresa preocupación por la creciente ambigüedad entre el trabajo que realiza una persona y el que realiza la inteligencia artificial. Si no se atiende de forma deliberada, esa incertidumbre puede convertirse con facilidad en ansiedad laboral sostenida. Es así como las organizaciones que ignoran esta señal, al tratar el bienestar como un asunto separado de su estrategia de IA en lugar de considerarlo una dimensión inseparable de ella, terminan pagando ese costo más tarde en forma de rotación, desmotivación y pérdida de confianza institucional.

Chapter 8 / Capítulo 8

AI, metaverse and sustainability in the new
organizational design
ISBN: 978-9915-9928-8-4
DOI: 10.62486/978-9915-9928-8-4.Ch08
Pages: 49-53



© 2026 The authors. This is an open access
book chapter distributed under the terms of
the Creative Commons Attribution (CC BY)
4.0 License.

Cybersecurity, trust and digital culture

Ciberseguridad, confianza y cultura digital

Jesús Alejandro Vázquez Meza¹  

Víctor Manuel Martínez García²  

Yennifer Díaz Romero³  

Manuel Iván Tostado Ramírez⁴  

Rogelio Estrada Lizárraga⁴  

José de Jesús Reyes Machain²  

1 Universidad Autónoma de Sinaloa, Unidad Académica Preparatoria Rubén Jaramillo, Vicerrectoría, Sinaloa, México

2 Universidad Autónoma de Sinaloa, Facultad de Arquitectura y Diseño, Vicerrectoría, Sinaloa, México

3 Universidad Autónoma de Sinaloa, Facultad de Ciencias de la Educación, Vicerrectoría, Sinaloa, México

4 Universidad Autónoma de Sinaloa, Facultad de Informática, Vicerrectoría, Sinaloa, México

5 Universidad Autónoma de Sinaloa, Facultad de Informática, Vicerrectoría, Sinaloa, México

ABSTRACT

This chapter analyzes cybersecurity as an organization-wide responsibility that cannot be reduced to technical equipment or solutions. It begins with the role of processes, everyday decisions, and attitudes toward information risks. Human factors and social engineering are considered, including attempts to manipulate individuals into unsafe behavior. The discussion distinguishes awareness of security recommendations from consistent safe practices, a distinction linked to cybersecurity culture. Risks concerning data protection and the use of AI tools outside institutional oversight are examined. The chapter also addresses hybrid and distributed workplaces, where devices, providers, and locations multiply. Management frameworks discussed in the manuscript help explain why security practices require continuity, training, and evaluation. The central argument is that digital trust depends on both suitable systems and organizational habits. Security is therefore presented as a shared task involving technical specialists, managers, and users, one that must adapt to changing patterns of work and collaboration. The chapter connects the human and technical dimensions of protection without treating either as a complete solution in isolation.

Keywords: Cybersecurity; Digital culture; Social engineering; Human factors; Data protection

RESUMEN

El capítulo analiza la ciberseguridad como una responsabilidad que atraviesa a toda la organización y no se agota en equipos o soluciones técnicas. Parte del papel que desempeñan los procesos, las decisiones cotidianas y las actitudes frente a los riesgos de información. Examina el factor humano y la ingeniería social, incluidas las formas de manipulación que buscan inducir comportamientos inseguros. Distingue entre conocer recomendaciones de seguridad y practicarlas de manera consistente, una diferencia vinculada con la cultura de ciberseguridad. La revisión considera los riesgos asociados a la protección de datos y a la incorporación de herramientas de inteligencia artificial fuera de los mecanismos de control institucional. También aborda las dificultades que presentan los entornos laborales híbridos y distribuidos, donde los dispositivos, proveedores y ubicaciones se multiplican. Los marcos de gestión mencionados en el capítulo sirven para mostrar por qué las prácticas de seguridad requieren continuidad, formación y evaluación. Su argumento central es que la confianza digital depende tanto de sistemas adecuados como de hábitos organizacionales. La seguridad aparece así como una tarea compartida entre responsables técnicos, directivos y usuarios, que debe adaptarse a nuevas modalidades de trabajo y colaboración. Esta perspectiva conecta las medidas preventivas con la comprensión real de los riesgos y con las prácticas que se repiten en el trabajo cotidiano. Se subraya así la relevancia de la formación y de la responsabilidad organizacional.

Palabras clave: Ciberseguridad; Cultura digital; Ingeniería social; Factor humano; Protección de datos

8.1 LA SEGURIDAD COMO PROBLEMA ORGANIZACIONAL

Durante mucho tiempo, la ciberseguridad se trató como un asunto exclusivamente técnico, centrado en *firewalls*, antivirus y parches de *software*; si bien dicha visión no está equivocada, resulta gravemente incompleta. Da Veiga y Eloff (2010) plantean desde las primeras líneas de su trabajo, en donde establecen que el enfoque de una organización hacia la seguridad de la información debe centrarse en el comportamiento de sus empleados, esto porque el éxito o el fracaso de esa seguridad depende, en última instancia, de lo que esas personas hacen o dejan de hacer todos los días.

Con esta idea se desplaza el centro de gravedad del problema. Además de comprar la herramienta técnica correcta, se vuelve necesario entender la seguridad como una propiedad emergente de toda la organización, es decir, de sus procesos, su cultura, sus incentivos y, sobre todo, de las decisiones cotidianas y aparentemente pequeñas que toma cada persona con acceso a un sistema. Cabe destacar que una organización puede invertir fuertemente en infraestructura de seguridad de última generación y, aun así, quedar expuesta si nadie dentro de ella entiende por qué esas medidas importan o cómo evitar comprometerlas sin darse cuenta.

8.2 FACTOR HUMANO

Si la seguridad es un problema organizacional, además de técnico, el factor humano se convierte casi por definición en la variable más determinante y, al mismo tiempo, en la más difícil de controlar. En su obra ya clásica sobre ingeniería social, Mitnick y Simon (2002) documentan con detalle cómo los atacantes más efectivos rara vez necesitan vulnerar un sistema técnico complejo cuando pueden, con mucha menos dificultad, convencer a una persona de que les entregue de manera voluntaria el acceso que necesitan.

Esta observación, formulada hace más de dos décadas, sigue siendo tan vigente como entonces, y quizás más. Por su parte, Da Veiga y Eloff (2010) señalan que una cultura organizacional consciente de la seguridad reduce el riesgo de mal comportamiento de los empleados y de interacciones dañinas con los activos de información. Se trata, en este sentido, del tipo de vulnerabilidad que ningún *firewall* puede prevenir por sí solo, ya que es un problema de comportamiento humano bajo presión, curiosidad o simple desconocimiento, ajeno a la infraestructura.

El reporte anual de IBM (2025) sobre el costo de las violaciones de datos aporta la evidencia económica más reciente de esta misma idea. En dicho estudio, el error humano fue la causa directa del 26 % de las violaciones de datos analizadas, mientras que el *phishing*, la técnica de manipulación más básica de la ingeniería social, volvió a ser el vector de ataque más común, presente en el 16 % de los incidentes y con un costo promedio de 4.8 millones de dólares por incidente (IBM, 2025). El mismo informe documenta algo inquietante sobre la velocidad con la que esto está empeorando, al señalar que la inteligencia artificial generativa redujo el tiempo necesario para redactar un correo de *phishing* convincente de hasta 16 horas a apenas 5 minutos, lo que multiplicó tanto la escala como la personalización de estos ataques. Con esto se puede ver que el factor humano, lejos de perder relevancia conforme avanza la tecnología de defensa, se está convirtiendo en un blanco cada vez más fácil de explotar con la misma inteligencia artificial que las organizaciones esperaban usar para protegerse.

8.3 CULTURA DE CIBERSEGURIDAD

El concepto de “cultura de ciberseguridad” suele usarse como frase hecha, sin que su contenido quede del todo claro. Da Veiga y Eloff (2010) ofrecen una respuesta precisa, en la que la cultura de seguridad de la información se entiende como la suma de percepciones, actitudes, valores y comportamientos que una comunidad organizacional demuestra al interactuar con la tecnología; con ello se puede ver que dicha cultura no se reduce a un cúmulo de conocimientos técnicos memorizados. Su marco distingue de forma explícita entre saber qué es una buena práctica de seguridad y haberla asimilado como parte natural del propio comportamiento cotidiano.

Esta distinción, que no es meramente académica, explica por qué muchas organizaciones invierten de forma considerable en programas de capacitación técnica sobre seguridad y, aun así, siguen siendo vulnerables, puesto que el haber transmitido información no equivale a haber cambiado el comportamiento. De acuerdo con este marco, construir una cultura genuina de

ciberseguridad requiere actuar al mismo tiempo sobre tres niveles, los cuales son el técnico, el procedimental y el humano-conductual, reconociendo que ninguno de ellos produce por sí solo el resultado que sí produce su combinación sostenida en el tiempo.

El estudio de Ruiz Sevilla (2026) sobre estudiantes de informática en la Facultad de Informática Mazatlán aporta evidencia empírica directa de esta brecha entre conocimiento y comportamiento, en donde se encontró que los estudiantes de semestres avanzados identifican correctamente, en un 86 % de los casos, los vectores de ataque basados en engaño textual o en suplantación de identidad corporativa, lo que revela un dominio conceptual sólido. Sin embargo, ese mismo grupo mostró un patrón de comportamiento que el estudio describe como “exposición técnica deliberada”, el cual consiste en la desactivación consciente de *firewalls* y antivirus, justificada por la necesidad percibida de mayor libertad técnica, y en un exceso de confianza que el propio autor vincula con el efecto Dunning-Kruger, es decir, con la creencia de que la habilidad técnica propia es suficiente para “controlar” un entorno inseguro sin las protecciones estándar (Ruiz Sevilla, 2026). En otras palabras, el conocimiento teórico no elimina la conducta de riesgo y, en este caso específico, incluso la transforma en una forma distinta y más deliberada de exposición.

8.4 INGENIERÍA SOCIAL

Más de 20 años después de que Mitnick y Simon (2002) la documentaran con detalle, la ingeniería social, entendida como el conjunto de técnicas de manipulación psicológica diseñadas para inducir a una persona a comprometer la seguridad de forma voluntaria, aunque involuntariamente perjudicial, sigue siendo uno de los vectores de ataque más efectivos que existen, esto porque ataca la confianza humana en lugar de los sistemas técnicos.

El propio texto de Mitnick y Simon (2002) plantea algo que resulta incómodo, pero que es necesario reconocer, en el sentido de que el elemento humano de la seguridad es, con frecuencia, el más débil de toda la cadena, lo cual no se debe a la incompetencia de las personas, ya que los atacantes explotan de forma deliberada rasgos humanos universales, como el deseo de ser útil, el respeto a la autoridad percibida y la urgencia inducida artificialmente, mismos que ninguna capacitación técnica logra desactivar por completo. Por ello, para combatir la ingeniería social no bastan las reglas memorizadas y se requiere una cultura organizacional en donde cuestionar una solicitud sospechosa, incluso si aparenta venir de una autoridad superior, se perciba como un comportamiento correcto, sin que se interprete como insubordinación.

La inteligencia artificial generativa está reescribiendo, en tiempo real, las reglas de este juego. De acuerdo con IBM (2025), el 16 % de las violaciones de datos analizadas en su estudio ya involucró el uso de IA por parte del propio atacante, principalmente para generar correos de *phishing*, presentes en un 37 % de esos casos, y contenido falsificado mediante *deepfakes*, presente en un 35 %. Lo que antes requería horas de investigación manual sobre una víctima potencial para redactar un mensaje convincente y personalizado hoy puede automatizarse casi por completo. Si bien las técnicas de manipulación que documentaron Mitnick y Simon (2002) dependían del ingenio y la paciencia de un atacante humano, la ingeniería social asistida por IA elimina buena parte de esa fricción, lo que exige que la cultura organizacional de sospecha saludable madure a la misma velocidad que la sofisticación de estos ataques.

8.5 PROTECCIÓN DE DATOS

Con la protección de datos, todos los principios anteriores se convierten en una obligación operativa concreta. En su publicación sobre programas de concientización y capacitación en seguridad, el Instituto Nacional de Estándares y Tecnología (2003) establece que ningún programa de este tipo puede considerarse completo si se limita a explicar amenazas en abstracto, sin conectar esa información con las responsabilidades específicas que cada persona tiene sobre los datos con los que trabaja a diario.

Por su parte, la norma ISO/IEC 27001:2022 aporta el marco de gestión formal para sostener esta protección de manera sistemática. En ella se establecen los requisitos que debe cumplir todo sistema de gestión de seguridad de la información para considerarse robusto y auditable, en lugar de quedarse en un conjunto de buenas intenciones sin verificación. La combinación de ambos enfoques, el educativo y el normativo, resulta indispensable, puesto que sin capacitación las normas se cumplen de forma mecánica y sin una comprensión real, mientras que sin normas la capacitación carece de estructura y de mecanismos de verificación que confirmen si está

funcionando.

Un riesgo particular y cada vez más frecuente que ilustra bien por qué ambos enfoques deben ir juntos es el de la llamada “IA en la sombra”, es decir, el uso de herramientas de inteligencia artificial por parte de colaboradores sin autorización ni supervisión de la organización. IBM (2025) encontró que el 97 % de las violaciones de datos relacionadas con IA ocurrieron en organizaciones que carecían de controles de acceso adecuados sobre estas herramientas, y que las violaciones vinculadas a la IA en la sombra generaron, en promedio, 670 000 dólares adicionales en costos para las organizaciones con un uso más extendido de este tipo de herramientas no autorizadas. El dato revela una paradoja incómoda, la cual consiste en que, mientras más entusiasmo pone una organización en adoptar la inteligencia artificial sin gobernarla, más se expone al tipo de fuga de datos que la protección formal está diseñada para prevenir.

8.6 SEGURIDAD EN ORGANIZACIONES HÍBRIDAS Y DISTRIBUIDAS

Todo lo discutido en este capítulo se vuelve considerablemente más complejo cuando la organización deja de operar desde un perímetro físico único y controlado para hacerlo de forma híbrida y distribuida, que es el mismo tipo de organización que se discutió en capítulos anteriores de este libro. En su informe sobre el panorama global de ciberseguridad, el Foro Económico Mundial (2024) documenta una brecha creciente entre las organizaciones que logran mantenerse resilientes ante estas nuevas condiciones y las que no, a la vez que señala que la escasez de talento especializado en ciberseguridad continúa ampliándose de manera notable en paralelo a esta transición.

Cabe destacar que el mismo informe advierte que solo el 15 % de las organizaciones se muestra optimista respecto a que esta brecha de habilidades mejore significativamente en los próximos dos años (Foro Económico Mundial, 2024).

Los datos de IBM (2025) permiten ver que este entorno disperso encarece de forma directa la respuesta de una organización ante un incidente. Su estudio documenta que las violaciones de datos toman, en promedio, 254 días en detectarse y contenerse por completo, plazo que se alarga considerablemente cuando la infraestructura de la organización está distribuida entre múltiples ubicaciones, proveedores de nube y dispositivos. El caso documentado por Ruiz Sevilla (2026) en el entorno universitario ilustra esta misma dinámica a menor escala, en donde, cuando miles de dispositivos personales de estudiantes se conectan al mismo tiempo a la red institucional bajo un esquema de *Bring Your Own Device*, el perímetro de seguridad tradicional deja de existir como línea de defensa única y la organización queda dependiendo, casi por completo, de que cada persona conectada mantenga por su cuenta buenos hábitos de higiene digital. Esta observación converge con lo discutido a lo largo del capítulo, esto es, con la idea de que en una organización híbrida y distribuida la seguridad deja de resolverse únicamente en el perímetro y pasa a depender, en buena medida, del comportamiento de cada persona conectada.

Chapter 9 / Capítulo 9

AI, metaverse and sustainability in the new
organizational design
ISBN: 978-9915-9928-8-4
DOI: 10.62486/978-9915-9928-8-4.Ch09
Pages: 54-59



© 2026 The authors. This is an open access
book chapter distributed under the terms of
the Creative Commons Attribution (CC BY)
4.0 License.

Transformation governance: who decides when the technology decides

Gobernanza de la transformación: quién decide cuando decide la tecnología

Jesús Alejandro Vázquez Meza¹ 

Víctor Manuel Martínez García² 

Yennifer Díaz Romero³ 

Manuel Iván Tostado Ramírez⁴ 

Rogelio Estrada Lizárraga⁴ 

José de Jesús Reyes Machain² 

1 Universidad Autónoma de Sinaloa, Unidad Académica Preparatoria Rubén Jaramillo, Vicerrectoría, Sinaloa, México

2 Universidad Autónoma de Sinaloa, Facultad de Arquitectura y Diseño, Vicerrectoría, Sinaloa, México

3 Universidad Autónoma de Sinaloa, Facultad de Ciencias de la Educación, Vicerrectoría, Sinaloa, México

4 Universidad Autónoma de Sinaloa, Facultad de Informática, Vicerrectoría, Sinaloa, México

5 Universidad Autónoma de Sinaloa, Facultad de Informática, Vicerrectoría, Sinaloa, México

ABSTRACT

This chapter examines the governance of artificial intelligence when systems systematically influence organizational decisions. It begins by asking who determines their purposes, controls their operation, and remains responsible for their consequences. The discussion reviews responsibility for automated decisions and regulatory frameworks addressed in the manuscript, together with auditing and verifiable documentation. Data protection is treated as a matter of authority, purpose, and consent as well as technical security. Ethical committees and worker participation are considered as ways of introducing diverse perspectives into oversight. The chapter also discusses trust and the legitimacy of algorithmically mediated decisions. It emphasizes that adopting a tool and governing its consequences are distinct but inseparable processes. Governance requires arrangements for assessing risks, explaining decisions, correcting problems, and preserving identifiable human responsibility. Institutional oversight and deliberation are presented as conditions for sustaining technological transformation with legitimacy over time. Rather than assuming that one framework solves every problem, the chapter brings together the requirements that should guide organizational implementation.

Keywords: Algorithmic governance; Automated decisions; AI auditing; Regulation; Worker participation

RESUMEN

El capítulo examina la gobernanza de la inteligencia artificial cuando los sistemas participan de manera sistemática en decisiones organizacionales. Parte de la necesidad de identificar quién define sus fines, controla su funcionamiento y responde por los efectos que generan. Revisa las responsabilidades vinculadas con decisiones automatizadas y los marcos regulatorios discutidos en la obra, así como la función de la auditoría y de la documentación verificable. La protección de datos se aborda como una cuestión de autoridad, propósito y consentimiento, además de seguridad técnica. El texto considera los comités éticos y la participación de los trabajadores como vías para incorporar perspectivas diversas a la supervisión. También analiza la confianza y la legitimidad de las decisiones mediadas por algoritmos. En conjunto, la discusión destaca que adoptar una herramienta y gobernar sus consecuencias son procesos diferentes pero inseparables. La gobernanza requiere procedimientos para evaluar riesgos, explicar decisiones, corregir problemas y mantener responsabilidades humanas identificables. El capítulo presenta la supervisión y la deliberación institucional como condiciones para que la transformación tecnológica pueda sostenerse en el tiempo con legitimidad. No da por resuelta la eficacia de un marco único, sino que organiza las exigencias que deberían orientar su implementación. El recorrido muestra conexiones entre procedimientos institucionales y derechos de las personas afectadas. Insiste en que la revisión de decisiones, la documentación y la posibilidad de participación forman parte de una misma arquitectura de responsabilidad organizacional.

Palabras clave: Gobernanza algorítmica; Decisiones automatizadas; Auditoría de IA; Regulación; Participación laboral

9.1 GOBERNANZA ALGORÍTMICA

Cuando un algoritmo empieza a influir de forma sistemática en decisiones que antes tomaban personas, como a quién contratar, qué precio ofrecer o qué contenido mostrar, surge una pregunta que ninguna organización puede seguir postergando, la de quién gobierna ese algoritmo. El planteamiento de esta pregunta se hace desde un sentido político, sin pensar en el aspecto técnico de quien lo mantiene funcionando y este se refiere mas que nada a quién decide lo que se codifica, sus límites y ante quién rinde cuentas.

La UNESCO (2021) sitúa esta pregunta en el centro de su *Recomendación sobre la Ética de la Inteligencia Artificial*, en donde la supervisión humana significativa, lejos de ser un accesorio opcional, se entiende como una condición necesaria; a partir de ello se puede pensar que, sin dicha supervisión, ningún sistema de IA podría considerarse legítimo dentro de una organización o una sociedad. Entendida de esta manera, la gobernanza algorítmica es el conjunto de estructuras, procesos y personas responsables de que esa supervisión se ejerza en la práctica, más allá de su declaración formal, lo que implica que exista alguien con autoridad efectiva para cuestionar, ajustar o incluso desactivar un sistema cuando sus resultados dejan de ser aceptables.

Si bien la brecha entre esta aspiración y la práctica real se está cerrando con rapidez, todavía es considerable. Al respecto, McKinsey & Company (2025) encontró que solo el 28 % de las organizaciones tiene un director ejecutivo que asume responsabilidad directa sobre la gobernanza de la IA, y que apenas un 17 % reporta que esa responsabilidad recae en otras áreas. En otras palabras, la existencia de un comité no garantiza que haya una autoridad real y visible detrás de él, distinción que separa a la gobernanza algorítmica auténtica de la simple formalidad organizacional.

9.2 RESPONSABILIDAD ANTE DECISIONES AUTOMATIZADAS

La pregunta de quién responde cuando una decisión automatizada causa daño, discutida ya en el capítulo 7 desde la perspectiva ética, adquiere aquí una dimensión más estructural, esto porque, además de determinar la responsabilidad después de que algo salió mal, es necesario que desde el inicio de las operaciones existan mecanismos que permitan asignar la responsabilidad de forma clara.

Loayza Villanueva (2025) analiza este vacío desde la perspectiva del derecho laboral y documenta cómo la incorporación acelerada de la inteligencia artificial a las relaciones de trabajo genera una necesidad urgente de reconfigurar los marcos legales existentes, mismos que en su mayoría fueron diseñados pensando en decisiones tomadas exclusivamente por personas. De acuerdo con su análisis, sin una regulación equitativa y responsable que distribuya con claridad esta responsabilidad, las organizaciones quedan operando en una zona gris en donde la ausencia de reglas explícitas termina beneficiando, casi siempre, a quien tiene más poder dentro de la relación, que por lo general es el empleador que decide implementar el sistema y no la persona afectada por sus decisiones.

Por su parte, Castillo Martínez et al. (2024), analizados con más detalle en el Capítulo 3, documentan una dimensión adicional de este vacío que rara vez se discute junto con la responsabilidad legal y que tiene que ver con la responsabilidad sobre la autenticidad y la propiedad intelectual del contenido que un sistema de IA genera dentro de una organización. Cuando un producto de trabajo combina contribución humana y generación algorítmica en proporciones difíciles de deslindar, el atribuir responsabilidad, tanto sobre errores como sobre créditos y reconocimiento, exige criterios que la mayoría de las organizaciones todavía no ha definido de forma explícita. Se trata de la misma zona gris que documenta Loayza Villanueva (2025), vista ahora desde el ángulo de la autoría, que se suma al del daño, lo que muestra que la responsabilidad ante decisiones automatizadas va más allá de quién responde cuando algo sale mal e incluye, también, quién responde, y quién recibe crédito, cuando todo camina de la mejor forma en los resultados de la implementación de la misma.

9.3 REGULACIÓN

En los últimos años, el panorama regulatorio sobre inteligencia artificial dejó de ser un terreno mayoritariamente voluntario para convertirse en un campo con obligaciones legales concretas y exigibles. Tal es el caso de la Unión Europea (2024), que a través del Reglamento (UE) 2024/1689, conocido como la Ley de Inteligencia Artificial, implementó el primer marco legal

integral del mundo sobre esta tecnología, mismo que se basa en lo niveles de riesgos, en donde, cuanto mayor es el riesgo potencial de un sistema de IA para los derechos y la seguridad de las personas, mayores son las obligaciones que debe cumplir quien lo desarrolla o lo utiliza.

Cabe destacar que la atención al riesgo no es exclusiva de la regulación europea. La OCDE (2019), en su *Recomendación del Consejo sobre Inteligencia Artificial*, adoptada por más de 40 países y actualizada en 2024 para incorporar los desafíos específicos de la IA generativa, estableció principios que hoy sirven de referencia normativa para buena parte de las regulaciones nacionales del mundo, un trabajo en paralelo con la Ley de IA europea. Sus cinco principios centrales, los cuales son el desarrollo sostenible, los valores democráticos, con la equidad y la privacidad y la seguridad en el sentido de la protección, funcionan como un lenguaje común que permite a las organizaciones que operan en múltiples países entender qué se espera de ellas, incluso cuando las leyes locales específicas todavía están en desarrollo.

Por su parte, Loayza Villanueva (2025) documenta cómo este panorama regulatorio internacional se traduce, o no, en marcos nacionales concretos dentro del ámbito laboral, y en su análisis encuentra que, si bien la Ley de IA europea ofrece principios generales sólidos, algunos países todavía carecen de una legislación laboral específica que traduzca esos principios en obligaciones concretas para el empleador que decide usar IA en procesos de contratación, evaluación o supervisión de su personal. Esta brecha entre la regulación general y la regulación laboral específica hace que muchas organizaciones sigan operando bajo un criterio de autorregulación que depende enteramente de su propia voluntad de actuar con responsabilidad, situación que, como advierte la propia autora, tiende a beneficiar de forma desproporcionada a quien ya tiene más poder dentro de la relación laboral.

9.4 AUDITORÍA

Ninguna regulación, por bien diseñada que esté, tiene efecto real sin mecanismos que verifiquen su cumplimiento efectivo. La auditoría de sistemas de inteligencia artificial, la cual consiste en revisar con metodología y de forma periódica si un sistema opera dentro de los límites éticos y legales que se supone debe respetar, se ha convertido en una función organizacional cada vez más indispensable y cada vez menos opcional.

Para los sistemas de IA clasificados como de alto riesgo, la Unión Europea (2024) exige documentación técnica detallada y mecanismos de apelación disponibles para las personas afectadas por sus decisiones, lo cual presupone la existencia de procesos de auditoría capaces de generar y verificar esa documentación. Una organización que despliega un sistema de este tipo sin haber establecido antes cómo lo auditará incumple una formalidad legal y, sobre todo, opera sin ninguna manera confiable de saber si el sistema hace lo que se supone que debe hacer.

El *AI Index Report* de Stanford (2025) muestra que la infraestructura de auditoría todavía va muy por detrás de la velocidad de despliegue de estos sistemas, esto al documentar que evaluar sistemas de IA con criterios de responsabilidad sigue siendo poco común, aunque empiezan a surgir puntos de referencia estandarizados diseñados para cerrar esta brecha, como los que evalúan la factualidad y la seguridad de forma sistemática. Mientras tanto, prácticamente todos los desarrolladores de modelos de frontera reportan resultados detallados sobre pruebas de capacidad técnica, es decir, sobre qué tan bien resuelve problemas un modelo, los principales desarrolladores prueban sus modelos con los mismos puntos de referencia sobre sus riesgos éticos y sociales. Esa asimetría entre cuánto se audita la capacidad y cuánto se audita la responsabilidad es el vacío que las organizaciones tienen que llenar por su cuenta cuando ninguna autoridad externa las obliga todavía a hacerlo.

9.5 PROTECCIÓN DE DATOS

La protección de datos, ya discutida en el capítulo anterior desde la perspectiva de la ciberseguridad, reaparece aquí como una pregunta de gobernanza que, junto con la forma en que se protegen técnicamente los datos, abarca quién decide qué datos se recolectan, con qué propósito y bajo qué mecanismo de consentimiento real de las personas involucradas.

La Unión Europea (2024) refuerza este punto desde el propio texto de la Ley de Inteligencia Artificial, que exige a los proveedores de sistemas de alto riesgo demostrar, con documentación verificable, cómo se recolectaron y gobernaron los datos usados para entrenar y operar sus sistemas. Dicha exigencia conecta de forma directa con los principios de la OCDE (2019) sobre

transparencia y explicabilidad, puesto que una organización no puede afirmar de forma creíble que gobierna bien sus datos si no puede explicar, con suficiente detalle, de dónde provienen y quién tiene acceso a ellos en cada etapa del proceso. Cuando se mira desde la gobernanza, además de hacerlo desde la ciberseguridad técnica, la protección de datos es, en esencia, una cuestión de poder, es decir, de quién decide, quién controla y quién rinde cuentas sobre la información más sensible que una organización posee.

9.6 COMITÉS ÉTICOS

Frente a la complejidad creciente de estas decisiones, cada vez más organizaciones están adoptando una estructura de gobernanza que hasta hace poco parecía reservada casi exclusivamente al ámbito de la investigación médica, como lo es el comité de ética, ahora aplicado a la supervisión de sistemas de inteligencia artificial.

Estos comités, integrados idealmente por personas de distintas áreas de la organización, además de las del departamento técnico que desarrolla o implementa el sistema, cumplen una función que ningún individuo aislado puede replicar bien, la cual consiste en someter las decisiones sobre IA a un proceso deliberativo real antes de que se conviertan en política operativa. La UNESCO (2021) respalda este tipo de mecanismos como parte de una gobernanza de IA responsable.

9.7 PARTICIPACIÓN DE LOS TRABAJADORES

Un vacío recurrente en la gobernanza algorítmica es la ausencia casi total de las personas directamente afectadas por estos sistemas, es decir, de los propios trabajadores, en las decisiones sobre cómo se diseñan, implementan y supervisan. Loayza Villanueva (2025) documenta que esta exclusión, que no es accidental, obedece a marcos legales que todavía no contemplan mecanismos claros de consulta o participación laboral diseñados para el contexto de la inteligencia artificial en el trabajo.

Corregir este vacío es una cuestión de justicia procedimental, aunque no se limita a ella. Las personas que usan un sistema todos los días, o que son evaluadas por él, suelen detectar problemas prácticos, sesgos sutiles o efectos no anticipados mucho antes que cualquier auditoría externa formal. Una organización que excluye de manera sistemática a sus trabajadores de las decisiones sobre gobernanza algorítmica genera desconfianza y, al mismo tiempo, renuncia a una fuente valiosa de información temprana sobre cuándo un sistema empieza a fallar.

El caso documentado por Ruiz Sevilla (2026) en el ámbito universitario ilustra este punto desde un ángulo distinto pero complementario. En su investigación se encontró que los propios estudiantes, quienes en ese contexto equivalen a los “trabajadores” cuyo comportamiento cotidiano determina buena parte de la seguridad institucional, identificaban con precisión los riesgos técnicos de ingeniería social, aunque nadie dentro de la institución había diseñado un mecanismo para canalizar ese conocimiento hacia mejoras concretas en la política de seguridad. Si bien el conocimiento estaba ahí, distribuido entre quienes vivían el problema todos los días, hacía falta un canal formal para que influyera en las decisiones de gobernanza. Se trata del mismo vacío que Loayza Villanueva (2025) documenta en el terreno laboral, en el sentido de que, cuando no hay participación, el conocimiento útil sigue existiendo, solo que no tiene ningún lugar formal adonde llegar.

9.8 CONFIANZA Y LEGITIMIDAD

Todos los elementos discutidos en este capítulo, es decir, la gobernanza algorítmica, la responsabilidad, la regulación, la auditoría, la protección de datos, los comités éticos y la participación laboral, convergen, en última instancia, en una sola pregunta sobre si las personas confían en que la organización está usando estas tecnologías de manera legítima.

Con base en la Recomendación de la UNESCO (2021), se puede plantear que ningún sistema de inteligencia artificial, por técnicamente sofisticado que sea, puede sostenerse a largo plazo si carece de la confianza de quienes lo usan o son afectados por él. Dicha confianza, que no se decreta ni se compra con una campaña de comunicación bien diseñada, se construye de forma acumulativa a través de decisiones consistentes de gobernanza que demuestren, con hechos verificables, más allá de las declaraciones de intención, que la organización prioriza la

responsabilidad sobre la velocidad de implementación.

Los datos más recientes muestran que esta confianza es un recurso frágil y desigualmente distribuido. Tal es el caso del *AI Index Report* de Stanford (2025), en el cual se documentan diferencias considerables en el optimismo público hacia la inteligencia artificial según el país. Existe una amplia mayoría de la población en países como China, Indonesia y Tailandia que percibe que los productos y servicios de IA son de beneficio mas que de perjuicios, ese optimismo no es igual en países como Canadá, Estados Unidos y los Países Bajos, donde apenas cuatro de cada diez personas comparten esa misma percepción. Esta brecha de confianza, que es algo más que un simple dato cultural curioso, significa para cualquier organización que opera en múltiples países que la legitimidad de su uso de la inteligencia artificial no puede asumirse como uniforme y tiene que construirse de manera deliberada y adaptada a cada contexto, dado que esa confianza parte, de entrada, de niveles muy distintos.

En este sentido, la organización que atiende estos elementos tiende a tratar la gobernanza de su transformación tecnológica como la base sobre la que se sostiene, a largo plazo, su legitimidad para seguir operando, en lugar de verla como un obstáculo burocrático que hay que sortear.

References

Bibliografía

BIBLIOGRAPHY / REFERENCIAS

- Acemoglu, D., & Restrepo, P. (2018). Artificial intelligence, automation and work. En A. Agrawal, J. Gans, & A. Goldfarb, *The economics of artificial intelligence: An agenda* (págs. 197-236). University of Chicago Press.
- Al-kfairy, M., Ahmed, S., & Khalil, A. (2024). Factors impacting users' willingness to adopt and utilize the metaverse in education: A systematic review. *Computers in Human Behavior Reports*.
- Autio, C., Dunietz, J., Hall, P., Jain, S., Roberts, K., Schwartz, R., . . . Tabassi, E. (2024). Artificial intelligence risk management framework: Generative artificial intelligence profile. National Institute of Standards and Technology.
- Bailenson, J. (2021). Nonverbal overload: A theoretical argument for the causes of Zoom fatigue. *Technology, Mind, and Behavior*, 2(1).
- Ball, M. (2022). *The metaverse: And how it will revolutionize everything*. Liveright.
- Barroso Camiade, C., & Pérez Castrejón, E. M. (2025). Desafíos éticos y legales en el uso de la inteligencia artificial (IA). *Sintaxis*, 102-118.
- Brynolfsson, E. (2022). The Turing trap: The promise & peril of human-like artificial intelligence. *Daedalus*, 272-287.
- Canossa-Montes de Oca, H. A., & Peraza-Villarreal, N. S. (2023). Gestión del Talento Humano en la Era de la Inteligencia Artificial: Retos y Oportunidades en el Entorno Laboral. 593 *Digital Publisher CEIT*, 9(1), 301-319.
- Castillo Martínez, K., Aguilar Rodríguez, J. A., & Madrigal Rentería, A. S. (2024). Desafíos éticos de la inteligencia artificial generativa en las nuevas formas organizacionales. *Revista Digital de Tecnologías Informáticas y Sistemas (REDTIS)*, 8(1), 46-55.
- Cerezo Arreola, C. (2025). so de IA en universitarios/egresados y el ámbito laboral. Tesina, Universidad Autónoma de Sinaloa, Facultad de Informática de Mazatlán, Mazatlán, Sinaloa.
- Cusumano, M. A., Gawer, A., & Yoffie, D. B. (2019). *The Business of Platforms: Strategy in the Age of Digital Competition, Innovation, and Power*. Harper Business.
- Da Veiga, A., & Eloff, J. (2019). A framework and assessment instrument for information security culture. *Computers & Security*, 196-207.
- Davis, F. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 319-340.
- De Felice, F., Petrillo, A., Lovine, G., Salzano, C., & Baffo, I. (2023). How does the metaverse shape education? A systematic literature review. *Applied Sciences*, 5682.
- Dell'Acqua, F., McFowland III, E., Mollick, E., Lifshitz, H., Kellogg, K., Rajendran, S., . . . Lakhani, K. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality. *Organization Science*, 22.
- Deloitte. (2025). 2025 global human capital trends: Turning tensions into triumphs. Obtenido de Deloitte Insights.: <https://www.deloitte.com/us/en/insights/topics/talent/human-capital-trends/2025.html>
- Diaz Ancco, F., Rodriguez Gonzales, K., & Estrada Chacon, L. H. (2024). El impacto de la inteligencia artificial en la formación de estudiantes de educación superior. *Yachay - Revista Científico Cultural*, 44-61.
- Feng, S., Xu, X., Li, S., Li, Z., & Gibson, D. (2024). Is metaverse a buzzword in education? Insights from a systematic review. *Educational Technology Research and Development*, 72(6), 3349-3390.
- Fernandes, F. A., Rodrigues, C., Teixeira, E., & Werner, C. (2023). Immersive learning frameworks: A systematic literature review. *arXiv*.
- Foro Económico Mundial. (2024). Global cybersecurity outlook 2024. World Economic Forum.
- Frey, C. B., & Osborne, M. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 254-280.
- Frontzkowski, Y., Münstermann, F., Gubernator, P., Bruhn, K., & Felfe, J. (2025). Virtual, but Not (Yet) Natural: The Role of Nonverbal Communication Quality, Cognitive Load, and Fatigue in Virtual Reality Compared to Videoconferencing and Face-to-Face Team Meetings. *Research Square*.
- Galbraith, J. (2014). *Designing Organizations: Strategy, Structure, and Process at the Business Unit and Enterprise Levels*. Jossey-Bass.
- Gelrich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, 15(1), 6.
- Guaña-Moya, J., & Chipuxi-Fajardo, L. (2023). Impacto de la inteligencia artificial en la ética y la privacidad de los datos. *RECIAMUC (Revista Científica de Investigación Multidisciplinaria Arbitrada Arbitraje por Pares)*, 923-930.

- Gómez Cano, C. A., & Pérez Gamboa, A. J. (2025). La adopción de la inteligencia artificial generativa en la Educación Superior: análisis bibliométrico y cienciométrico. *Maestro y Sociedad*, 299-311.
- Hemmer, P., Schemmer, M., Köhl, N., Vössing, M., & Satzger, G. (2024). Complementarity in human-AI collaboration: Concept, sources, and evidence. *arXiv*.
- Hu, K. (2 de Feb de 2023). ChatGPT sets record for fastest-growing user base - analyst note. Obtenido de Reuters: <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>
- IBM. (2025). Cost of a data breach report 2025. IBM Security.
- Instituto Nacional de Estándares y Tecnología (NIST, por sus siglas en inglés). (2003). Building an information technology security awareness and training program. U.S. Department of Commerce.
- Kohli, S. (2025). Corporate metaverse adoption: A case study of Accenture's Nth Floor virtual onboarding and training platform. En M. Gonçalves Rodrigues, M. Carvalho, & J. Monteiro, *Cases on metaverse and consumer experiences* (págs. 271-302). IGI Global.
- Lee, L.-H., Zhou, P., Braud, T., & Hui, P. (2022). What is the metaverse? An immersive cyberspace and open challenges. *arXiv*.
- Lin, H., Wan, S., Gan, W., Chen, J., & Chao, H.-C. (2022). Metaverse in education: Vision, opportunities, and challenges. 2022 IEEE International Conference on Big Data (págs. 2857-2866). IEEE.
- LinkedIn Learning. (2025). 2025 workplace learning report. LinkedIn.
- Loayza Villanueva, S. (2025). Desafíos y oportunidades jurídicas de la inteligencia artificial en el ámbito laboral: hacia una regulación equitativa y responsable. *Laborem*, 139-161.
- McKinsey & Company. (2025). The state of AI in 2025: Agents, innovation, and transformation. Obtenido de McKinsey & Company: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>
- Méndez Reyes, J. (2025). Ética e inteligencia artificial en las organizaciones. *Clío. Revista de Historia, Ciencias Humanas y Pensamiento Crítico*, 698-738.
- Mitnick, K., & Simon, W. (2002). *The art of deception: Controlling the human element of security*. John Wiley & Sons.
- Mollick, E. (2024). Co-intelligence: Living and working with AI. Portfolio.
- OCDE. (2019). Recomendación del Consejo sobre Inteligencia Artificial. Organización para la Cooperación y el Desarrollo Económicos.
- Ricra Ruiz, R. A., Queque Luque, E. F., Vega Lazo, F. H., Martínez Horna, D. J., Ross Audureau, J., & Lara Tapia, L. M. (2025). Implicaciones éticas de la inteligencia artificial generativa en la educación superior: una revisión sistemática. *Revista InveCom*, 6(2), 1-8.
- Rios Navarrete, R. G. (2025). Ética y regulación de la inteligencia artificial en entornos laborales en México. Mazatlán, Sinaloa: Universidad Autónoma de Sinaloa.
- Robles Ornelas, U., & Paez Lopez, M. (2026). *Mi diario digital de aprendizaje: Plataforma web para la reflexión académica y el aprendizaje consciente*. Mazatlán, Sinaloa: Universidad Autónoma de Sinaloa.
- Rogers, E. (2003). *Diffusion of Innovations*. Free Press.
- Ruiz Sevilla, D. A. (2026). Evaluación del nivel de cultura en ciberseguridad y hábitos de higiene digital en los estudiantes de la Facultad de Informática Mazatlán. Mazatlán, Sinaloa: Universidad Autónoma de Sinaloa.
- Russell, S., & Norvig, P. (2022). *Artificial Intelligence A Modern Approach*. Pearson Education.
- Shi, C., & Park, J. Y. (2024). A systematic review of the metaverse in formal education. *Journal of Applied Research in Higher Education*, 17(5), 1850-1868.
- Stanford University. (2025). Stanford University, Institute for Human-Centered Artificial Intelligence. Obtenido de Stanford University, Institute for Human-Centered Artificial Intelligence: <https://hai.stanford.edu/ai-index/2025-ai-index-report>
- Stanford University. (2026). Stanford University, Institute for Human-Centered Artificial Intelligence. Obtenido de Stanford University, Institute for Human-Centered Artificial Intelligence: <https://hai.stanford.edu/ai-index/2026-ai-index-report>
- UNESCO. (2021). Recommendation on the ethics of artificial intelligence. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura.
- Unión Europea. (2024). Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial). Diario Oficial de la Unión Europea.
- World Economic Forum. (2025). The future of jobs report 2025. Obtenido de World Economic

Forum: <https://www.weforum.org/publications/the-future-of-jobs-report-2025/>
Zoom. (2025). 24+ hybrid work statistics for the evolving workplace [2025]. Obtenido de
Zoom Blog: <https://www.zoom.com/en/blog/hybrid-work-statistics>