ISSN 3035-3823

JOURNAL OF EMERGING PERSPECTIVES

Business Society Territories

Vol. 2 – 2025

FIRENZE
UNIVERSITY
PRESS

Journal of Emerging Perspectives

Vol. 2 – 2025

Firenze University Press

Journal of Emerging Perspectives
https://oaj.fupress.net/index.php/jep
ISSN 3035-3823 (online)

Published by Firenze University Press

Firenze University Press
Università degli Studi di Firenze
via Cittadella, 7, 50144 Firenze, Italy
www.fupress.com

JEP

Firenze University Press
https://oaj.fupress.net/index.php/jep

Citation: Mosca, F. (2025) Artificial intelligence as a strategic inflection point: Implications for firms, industries, and global competitiveness. *Journal of Emerging Perspectives* 2: 3-7. doi: 10.36253/jep-20876

Received: April 30, 2025

Revised: September 30, 2025

Published: December 20, 2025

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

ORCID
FM: 0000-0002-9301-0536

Artificial intelligence as a strategic inflection point: Implications for firms, industries, and global competitiveness

Fabrizio Mosca

INTRODUCTION: AI AS A STRATEGIC INFLECTION POINT

This essay departs from the view of Artificial Intelligence (AI) as a mere technical upgrade, framing it instead as a strategic inflection point for firms, industries, and global competitiveness, our focus is squarely on the rise of generative and emerging agentic AI systems. The discussion is built around five interrelated questions: (1) how AI is fundamentally reshaping managerial processes; (2) the real-world geography of current AI investments; (3) the infrastructure and institutional gaps that threaten to fragment regional competitiveness; (4) an identification of the players currently operating at the technological and organizational frontier; and (5) the strategic hurdles global corporations face as generative and agentic AI rewrite the rules of competitive advantage and governance. Drawing on recent evidence from international organisations, academic research, and policy institutions, the essay argues that AI should be understood not merely as a productivity-enhancing technology, but as a general-purpose strategic infrastructure that reconfigures the very nature of managerial cognition and organizational design.

The past decade has marked a profound shift: AI has evolved from a peripheral automation tool into a central pillar of corporate strategy. While early applications were largely confined to narrow back-office tasks, the explosion of machine learning and agentic systems has radically altered how firms sense market opportunities, allocate resources, and compete globally.

The paper concludes by outlining a research agenda for strategic management and highlighting the need for managers to integrate AI into corporate strategy, governance, and institutional alignment.

Over the past decade, artificial intelligence (AI) has evolved from a peripheral automation tool into a central pillar of corporate strategy and industrial transformation. Early applications of AI were largely confined to narrow automation tasks and decision-support systems embedded in operational and back-office functions. Today, however, the rapid diffusion of machine learning, deep learning, and – more recently – generative and agentic AI systems has fundamentally altered how firms sense opportunities, allocate resources, design organisational processes, and compete across global markets.

This transformation is anything but incremental. AI now exhibits the hallmarks of a general-purpose technology – much like electricity or digitalization – with effects that strike at the foundations of managerial decision-making. The advent of Generative AI has further accelerated this disruption: the ability to generate code, text, and synthetic data at scale has slashed the cost of experimentation. More importantly, emerging agentic systems go beyond content creation to plan and execute tasks with semi-autonomous precision. It is here that the boundary between human judgment and algorithmic action begins to blur, raising urgent strategic questions. Together, these developments challenge established boundaries between human judgment and algorithmic action, raising new strategic, organisational, and ethical questions.

At the same time, AI diffusion is highly uneven. Investment volumes, access to computational infrastructure, availability of specialised talent, and institutional readiness differ markedly across regions and industries. The United States has emerged as the dominant hub of private AI investment and platform development. The European Union has pursued a distinctive path emphasising regulatory leadership and trustworthy AI, while Asian economies display heterogeneous trajectories ranging from large-scale state-supported deployment to highly specialised industrial integration. These divergences are reshaping patterns of global competitiveness and industrial leadership.

Against this backdrop, this introductory essay addresses five interrelated research questions that frame the contributions of this special issue:

What is the actual impact on internal managerial processes? How are global investment volumes truly distributed across regions? What capability gaps risk leaving certain sectors behind? Which firms are genuinely defining the frontier of AI implementation? What strategic challenges should global corporations brace for as these technologies evolve?

By addressing these questions, the essay provides a conceptual and empirical foundation for the papers in this special issue, situating AI at the intersection of strategic management, organisational theory, and institutional analysis.

AI AND THE TRANSFORMATION OF MANAGERIAL PROCESSES

The first research question concerns the impact of AI on managerial processes within firms. Strategic management has long been concerned with how managers process information, form expectations, and make decisions under uncertainty. AI shakes these micro-foundations, reshaping strategy through augmented analytics and new models of human–machine collaboration.

Augmented decision-making and analytics-driven strategy.

AI dramatically expands the scope and speed of managerial analytics. Traditional business intelligence systems primarily supported descriptive and diagnostic analysis, often relying on structured data and retrospective reporting. Contemporary AI systems enable predictive and prescriptive analytics across high-dimensional datasets that include unstructured text, images, audio, and sensor data. Managers can increasingly simulate alternative strategic scenarios, stress-test assumptions, and explore counterfactuals in near real time.

Generative AI further enhances these capabilities by automating cognitive tasks such as drafting strategic documents, summarising large information sets, generating alternative business models, or prototyping marketing and product concepts. Rather than replacing managerial judgment, these tools reshape it. Strategic work shifts from information generation to framing, validation, and interpretation. This evolution supports what may be described as simulation-rich, analytics-driven strategy, in which managerial intuition is systematically complemented – and in some domains challenged – by algorithmic insights.

Workflow reconfiguration and task redistribution.

The adoption of AI is rarely neutral with respect to organisational structure. AI systems redistribute tasks between humans and machines and often require deep re-engineering of workflows rather than simple automation of existing routines. In banking, for example, AI-based credit scoring and anti–money laundering systems alter interactions between risk officers, relationship managers, and compliance teams. In healthcare, AI-assisted diagnostics reshape professional roles, decision rights, and accountability structures among clinicians, radiologists, and support staff.

Empirical evidence suggests that productivity gains from AI depend critically on workflow redesign and organisational learning. AI systems tend to deliver the greatest benefits when integrated into end-to-end processes with clear human oversight, feedback loops, and accountability mechanisms. From a dynamic capabilities perspective, firms that capture sustained value from AI are those that reconfigure routines, governance structures, and incentive systems alongside technological deployment.

Managerial roles and the rise of agentic AI.

AI also transforms the role of managers themselves. By automating routine cognitive tasks – such as reporting, scheduling, and documentation – AI can free managerial time for higher-order activities including sense-making, coordination, and stakeholder engagement. At the same time, AI systems increasingly contest managerial judgment, particularly in domains where algorithmic predictions outperform human intuition.

The emergence of agentic AI intensifies these dynamics. Systems capable of planning and executing sequences of actions across digital environments introduce new opportunities for efficiency and scale, but also new risks related to error propagation, accountability, and control. Managers must therefore develop new competencies in AI literacy, model governance, and ethical oversight. Strategic leadership in the AI era increasingly entails the ability to design human–AI systems that balance autonomy and control.

THE GLOBAL INVESTMENT LANDSCAPE: SCALE, GEOGRAPHY, AND CAPABILITY GAPS

The second and third research questions concern the scale and geography of AI investment and the capability gaps that shape regional competitiveness. From a strategic management perspective, these questions cannot be reduced to aggregate spending figures alone. Rather, they concern the co-evolution of capital allocation, computational infrastructure, talent formation, and institutional frameworks.

Institutional datasets consistently show that global AI investment has entered a phase of rapid acceleration combined with strong geographic concentration. Evidence from the Stanford AI Index indicates that private AI investment reached unprecedented levels in 2024 and 2025, with the United States accounting for the dominant share. Investment in generative AI has grown particularly rapidly, reflecting the strategic importance attributed to foundation models and platform-based AI ecosystems.

From a macroeconomic perspective, this is not just a bubble; it is the hallmark of AI's transition into a general-purpose enabling infrastructure. Such technologies are notorious for exhibiting increasing returns, path dependence, and cumulative advantage. In this high-stakes environment, early leaders can consolidate durable advantages through scale effects – leveraging a virtuous cycle of data, compute capacity, talent attraction, and the organizational capabilities needed to exploit them.

The regional map of these investments reveals innovation ecosystems that couldn't be more different. The United States benefits from a unique alignment of deep capital markets, frontier AI firms, hyperscale cloud providers, and top-tier research universities. This tight coupling between research and entrepreneurship is what allows for the rapid scaling of AI. Europe's landscape, by contrast, remains frustratingly fragmented. Despite a formidable scientific base, private investment continues to lag, while chronic constraints in scale-up financing and access to high-performance computing stifle growth. While Europe has carved out a role as a global leader in AI governance, this regulatory focus raises uncomfortable questions about its ability to scale industrially.

Asia, meanwhile, offers a far more heterogeneous picture. China has solidified its position as a major AI destination, fuelled by state coordination and large-scale deployment. Other advanced Asian economies have opted for a more surgical approach, integrating AI directly into manufacturing, robotics, and smart infrastructure. These trajectories prove that there is more than one way to achieve AI competitiveness, often relying more on domain-specific integration than on frontier foundation models.

Looking across these regions, three capability gaps stand out as critical bottlenecks. First, the concentration of computational infrastructure and chips has created a dangerous dependency on a handful of providers. Second, the acute shortage of specialized AI talent is directly constraining organizational growth. Finally, the divergent institutional and regulatory frameworks are now the primary drivers of the risk–return profile for AI, ultimately dictating where and how firms choose to deploy these technologies.

Looking at global players competition, historically, the trajectory of AI mirrors the early evolution of the automotive industry in the early 1900s or the personal computer revolution of the 1980s, where the sheer scale of investment became the ultimate arbiter of competitive success. In these capital-intensive industries, firm size is not just an advantage – it is a prerequisite. We are seeing this pattern repeat today: while the initial sparks of AGI development ignited within agile startups, the 'gravity' of scale has seen global giants – Google, Microsoft, Apple, Nvidia, IBM, Oracle, and Amazon – swiftly move in. Through a sophisticated mix of equity stakes, strategic alliances, and deep technological integrations, these global players have effectively absorbed AI innovations, embedding them into the core of their competitive models. Ultimately, economies of scale in AI development are now so strategic that only those with the vastest computational and financial resources can realistically define the frontier, turning AI from a startup disruption into a consolidated pillar of corporate dominance.

SECTORAL FRONTIERS: INDUSTRIES AT THE LEADING EDGE

The fourth research question asks which industries and firms are currently at the frontier of AI implementation. While AI adoption is observable across most sectors, evidence suggests that a subset of industries is disproportionately advanced in deploying AI at scale.

Banking, financial services, and insurance represent a leading frontier. AI is deeply embedded in core processes such as credit assessment, fraud detection, compliance, and customer service. Importantly, institutional and supervisory evidence shows that the strategic challenge in this sector is no longer adoption per se, but governance capacity. Firms that can integrate AI within robust risk management, model validation, and accountability frameworks are better positioned to capture value while maintaining trust and regulatory compliance.

Healthcare is another frontier sector, characterised by high potential benefits and stringent constraints related to data quality, safety, and accountability. AI applications span diagnostics, clinical decision support, pharmaceutical research, and hospital operations. However, the effective scaling of AI depends critically on the digital maturity of health systems, data interoperability, and institutional governance. These conditions shape both performance outcomes and liability exposure.

Across sectors, AI adoption tends to be highest in data-rich, knowledge-intensive industries such as software, financial services, and healthcare. Manufacturing and logistics increasingly adopt AI for optimisation and predictive maintenance, while more fragmented and less digitised sectors lag behind. These patterns reinforce the view that AI is complementary to intangible assets such as data, analytics capabilities, and modular IT architectures.

ETHICAL, REGULATORY, AND SOCIETAL CHALLENGES

No serious analysis of AI as a strategic inflection point can avoid the minefield of ethical, regulatory, and societal challenges. These are not externalities; they are core business risks. AI systems can replicate and *amplify* existing biases, spark serious concerns about privacy and surveillance, and introduce dangerous forms of systemic risk. These issues cut straight to the heart of strategic management, directly shaping reputational risk, stakeholder trust, and regulatory exposure.

Around the world, a patchwork of risk-based regulatory frameworks is emerging. The European Union, notably, is trying to balance innovation with the ironclad protection of fundamental rights. Such frameworks are imposing tough new obligations: transparency, accountability, and stringent human oversight, especially for high-risk systems. For global firms, this is where strategy gets complicated. Regulatory fragmentation across jurisdictions creates immense friction, demanding highly differentiated deployment and governance strategies.

Beyond the boardroom, AI poses existential questions about the future of work. While AI clearly has the power to supercharge productivity and spawn entirely new roles, its effects are brutally uneven across occupations and regions. Firms face strategic choices regarding large-scale reskilling and workforce redesign – choices that will ultimately shape not just their competitiveness, but their very social legitimacy.

FUTURE COMPETITIVENESS IN THE AGE OF GENERATIVE AND AGENTIC AI

The fifth research question concerns the strategic challenges global corporations face as generative and agentic AI evolve. These technologies reshape sources of competitive advantage by lowering entry barriers in some domains while reinforcing scale and platform effects in others.

On the one hand, generative and agentic AI enable new forms of differentiation through proprietary data, domain-specific fine-tuning, and tightly integrated human–AI systems. On the other, widely accessible AI platforms and open-source models commoditise certain capabilities. Firms must therefore make strategic choices about where to build proprietary AI assets and where to rely on external platforms.

Ecosystem dynamics become central. A small number of platform providers control critical AI infrastructures, while downstream firms integrate these capabilities into sector-specific applications. This creates new dependencies and bargaining power asymmetries, alongside opportunities for collaboration and co-innovation. Managing these relationships is a core strategic task.

Agentic AI also introduces new risks related to reliability, security, and systemic dependence. Firms must develop resilience strategies that include diversified suppliers, continuous monitoring, and integration of AI risks into enterprise risk management.

MANAGERIAL AND RESEARCH IMPLICATIONS

Revisiting the five research questions highlights a common theme: AI should be treated as a core strategic and organisational domain rather than as a technical project. For managers, long-term competitiveness depends on aligning AI investment with organisational capabilities, governance structures, and institutional contexts.

For scholars of strategic management, AI offers a rich research agenda. Future studies can examine how AI reshapes managerial cognition and dynamic capabilities; the conditions under which AI investments generate sustained advantage; the governance of AI ecosystems; and the long-term societal implications of pervasive AI adoption.

By integrating AI into the core concerns of strategic management, this special issue aims to contribute to a deeper, more nuanced understanding of how AI is transforming firms, industries, and global competition.

REFERENCES

Bank for International Settlements. (2024). *Annual economic report 2024: Artificial intelligence and the economy*. BIS Publishing. www.bis.org

Bank for International Settlements Innovation Hub. (2024). *Artificial intelligence in central banking*. BIS Publishing.

Boston Consulting Group. (2025). *AI at work: What people are saying and how organizations are scaling*. BCG.

European Central Bank Banking Supervision. (2025). *Artificial intelligence in banking: Insights from ECB–bank workshops*. European Central Bank. https://www.bankingsupervision.europa.eu

European Parliament, European Parliamentary Research Service. (2024). *AI investment: EU and global indicators*. EPRS.

European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*. eur-lex.europa.eu

Financial Stability Board. (2024). *Artificial intelligence in finance: Roundtable findings*. FSB.

Financial Stability Board. (2025). *Next steps for monitoring the use of artificial intelligence in financial markets*. FSB. https://www.fsb.org

Fortune Business Insights. (2025). *Artificial intelligence in healthcare market size, share & industry analysis, 2025–2033*. Fortune Business Insights.

Global Market Insights. (2025). *Artificial intelligence (AI) in BFSI market: Analysis of 2025 real-world adoption and 2034 forecasts*. Global Market Insights.

International Monetary Fund. (2024). *AI and the future of work*. IMF Staff Discussion Note.

International Monetary Fund. (2026, 14 January). *New skills and AI are reshaping the future of work*. IMF. www.imf.org

Kergroach, S., & Héritier, J. (2025). *Emerging divides in the transition to artificial intelligence* (OECD Regional Development Papers, No. 147). OECD Publishing. doi.org

Maslej, N., Haimovich, J., Karnes, R., et al. (2025). *The AI Index 2025 annual report*. Stanford Institute for Human-Centered Artificial Intelligence. aiindex.stanford.edu

McKinsey & Company. (2025). *The state of AI in 2025: From generative pilots to agentic workflows*. McKinsey Global Institute. www.mckinsey.com

McKinsey & Company. (2025). *Generative AI in healthcare: Current trends and future outlook*. McKinsey Global Institute.

McKinsey Technology Council. (2024). *Technology trends outlook 2024*. McKinsey & Company.

OECD. (2025a). *Health at a glance 2025: OECD indicators*. OECD Publishing. doi.org

OECD. (2025b). *Data and digital health indicators*. OECD Publishing.

World Economic Forum. (2025). *The future of jobs report 2025*. World Economic Forum. www.weforum.org

JEP

Firenze University Press
https://oaj.fupress.net/index.php/jep

OPEN ACCESS

Citation: Frontoni, E., & Epasto, S. (2025) AI as the "final frontier"? Narratives, geographies, and relationships in the human–algorithm coproduction. *Journal of Emerging Perspectives* 2: 9-16. doi: 10.36253/jep-20877

Received: April 30, 2025

Revised: September 30, 2025

Published: December 20, 2025

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

ORCID
EF: 0000-0002-8893-9244
SE: 0000-0002-6420-0867

AI as the "final frontier"? Narratives, geographies, and relationships in the human–algorithm coproduction

Emanuele Frontoni, Simona Epasto*

Department of Political Science, Communication and International Relations, University of Macerata, Italy
Email: emanuele.frontoni@unimc.it; simona.epasto@unimc.it
*Corresponding author

Abstract. This article critically interrogates the metaphor of artificial intelligence (AI) as a "final frontier," revealing its colonial and expansionist roots and its inadequacy in capturing the embeddedness of AI systems within complex sociotechnical landscapes. Combining a geographical perspective with insights from computer engineering, we argue that the European Union's Artificial Intelligence Act marks a paradigmatic shift from narratives of conquest to frameworks grounded in regulation, accountability, and coproduction. Through empirical vignettes in healthcare, recruitment, and criminal justice, we show how AI operates not in neutral or uncharted spaces, but within spatially and institutionally situated environments. Algorithmic systems both shape and are shaped by territorial infrastructures, legal frameworks, and social relations. By bridging disciplinary approaches, the paper advances a nuanced, interdisciplinary understanding of human–algorithm interaction – one that moves beyond deterministic imaginaries to address the spatial, political, and ethical dimensions of AI governance.

Keywords: artificial intelligence, human-algorithm coproduction, geographical embeddedness, AI regulation, interdisciplinary approach, EU AI Act, digital sovereignty.

1. CONCEPTUAL PRELUDE: BEYOND THE "FINAL FRONTIER"

The metaphorical invocation of artificial intelligence as humanity's "final frontier" has permeated contemporary discourse with remarkable persistence, echoing the colonial and expansionist rhetoric historically associated with geographical exploration and territorial conquest (Frontoni, 2024). This linguistic framing reveals profound assumptions about the nature of technological progress, human agency, and the epistemic boundaries that define our relationship with algorithmic systems. Yet beneath this frontier metaphor lies a more complex reality of coproduction, interdependence, and mutual constitution between human and artificial intelligence that challenges simplistic narratives of technological determinism or human mastery.

The European Union's Artificial Intelligence Act, which entered into force in August 2024 and represents the world's first comprehensive regulatory framework for AI systems, provides a critical lens through which to examine these evolving human-algorithm relationships. The Act's risk-based approach, which categorizes AI applications according to their potential impact on fundamental rights and societal values, fundamentally reconceptualizes the terrain of human-machine interaction from one of unbounded technological possibility to one of carefully calibrated co-existence. This regulatory intervention signals a departure from the frontier mythology that has long characterized AI development, instead positioning artificial intelligence within a framework of accountability, transparency, and human oversight that acknowledges the profound interdependencies between algorithmic processes and human social structures.

The geographical metaphor of the frontier, with its implications of empty spaces awaiting discovery and colonization, becomes particularly problematic when applied to AI development within the European context. Unlike the historical frontiers that suggested unlimited expansion into supposedly uninhabited territories, the landscape of artificial intelligence is already densely populated with existing social relationships, institutional frameworks, and cultural practices that shape both the development and deployment of algorithmic systems. The AI Act's emphasis on conformity assessments, risk management systems, and human oversight requirements reflects an understanding that AI systems do not operate in isolation but are embedded within complex socio-technical assemblages that require careful navigation rather than wholesale conquest.

Consider the implementation of AI systems in healthcare diagnostics, where the coproduction between human expertise and algorithmic analysis becomes particularly evident. Under the AI Act's classification system, many medical AI applications fall into the high-risk category, requiring extensive documentation, testing, and human oversight mechanisms. The development of AI-powered diagnostic tools for radiology, such as those used for detecting diabetic retinopathy or analysing mammographic images, illustrates how the frontier metaphor breaks down when confronted with the reality of medical practice. These systems do not replace human radiologists but rather extend and augment their diagnostic capabilities through a process of continuous interaction and mutual calibration. The radiologist's expertise shapes the training data, validates the algorithmic outputs, and contextualizes the results within broader clinical narratives, while the AI system processes vast quantities of imaging data and identifies patterns that might escape human attention. This relationship of coproduction challenges the frontier narrative by revealing how AI systems are not autonomous agents conquering new territories but rather collaborative tools that require ongoing human engagement and interpretation.

Similarly, the deployment of AI systems in recruitment and employment contexts, which the AI Act specifically addresses due to their potential for discrimination and bias, demonstrates the complex geographical and social dimensions of human-algorithm interaction. AI-powered recruitment tools that analyse resumes, conduct preliminary screenings, or assess candidate suitability operate within existing structures of labour markets, educational systems, and social hierarchies that profoundly shape their functioning. The Act's requirements for transparency, explainability, and bias testing acknowledge that these systems do not operate in a neutral technological space but are embedded within particular geographical and cultural contexts that influence their development, training, and application. A recruitment AI system trained on data from one European labour market may perform differently when deployed in another region with distinct educational systems, cultural norms, and employment patterns, highlighting how the supposed universality of algorithmic logic must always contend with the particularities of local social geographies.

The narrative of AI as a final frontier also obscures the ways in which algorithmic systems are deeply entangled with existing power structures and knowledge production processes. The AI Act's governance framework, with its emphasis on CE marking, conformity assessments, and notified body involvement, reflects a recognition that AI development is not a neutral technical enterprise but a fundamentally political process that shapes and is shaped by broader social relationships. The Act's requirement for fundamental rights impact assessments for high-risk AI systems acknowledges that these technologies do not simply extend human capabilities into new domains but actively reconfigure existing social relationships and institutional arrangements.

The case of AI systems in criminal justice, which the AI Act treats with particular caution through its prohibitions on certain uses and strict requirements for others, further illustrates the limitations of the frontier metaphor. Predictive policing algorithms, risk assessment tools for judicial decision-making, and automated analysis of evidence do not operate in an empty technological space but are deeply embedded within existing criminal justice institutions, legal frameworks, and social hierarchies. These systems co-produce outcomes through complex interactions between algorithmic processing, human interpretation, institutional protocols, and broader social contexts that cannot be understood through the lens of technological expansion into uncharted territories.

The geographical dimensions of AI development within Europe also challenge the frontier narrative through their emphasis on digital sovereignty, data localization, and the protection of European values and legal traditions. The AI Act’s extraterritorial reach, which applies to AI systems used within the EU regardless of where they are developed, reflects an understanding that the governance of artificial intelligence is not simply a matter of technological innovation but of maintaining particular forms of social organization and value systems within specific geographical boundaries. This regulatory geography suggests that rather than representing a boundless frontier, AI development is increasingly subject to territorial constraints and jurisdictional negotiations that reflect broader geopolitical relationships and cultural differences.

The concept of human-algorithm coproduction that emerges from this analysis suggests a need to move beyond simplistic narratives of technological progress toward more nuanced understandings of the ways in which human and artificial intelligence systems mutually constitute each other through ongoing processes of interaction, adaptation, and learning. This coproduction is not merely a technical process but a fundamentally social and political one that involves negotiations over agency, authority, and accountability within specific institutional and cultural contexts. The AI Act’s framework for governing these relationships through risk assessment, human oversight, and transparency requirements provides one model for managing this coproduction, but it also raises broader questions about the forms of human-machine collaboration that are desirable and sustainable within democratic societies.

As we move forward in examining these questions, it becomes clear that the metaphor of AI as a final frontier may need to be replaced with more sophisticated conceptual frameworks that can account for the complex, multi-directional relationships between human and artificial intelligence systems. The European approach, as embodied in the AI Act, suggests one possible direction for this reconceptualization, emphasizing governance, accountability, and the protection of fundamental rights rather than unbounded technological expansion. However, the full implications of this shift from frontier mythology to regulatory governance remain to be explored through careful analysis of how these new frameworks shape the ongoing coproduction of human and algorithmic agency in practice (Frontoni, 2024).

On this basis, our interdisciplinary proposal unfolds: the following sections translate this frame into a geography-of-coproduction lens, three empirical vignettes, the S.P.A.C.E. framework, and policy recommendations.

2. WHY METAPHORS MATTER IN AI

Opening with a conceptual prelude that challenges the “final frontier” metaphor and points to Europe’s risk-based turn, we then advance a geography-of-governance perspective and our joint contribution: an interdisciplinary account of human–algorithm coproduction that is spatially aware and policy-oriented. We subsequently develop a coproduction lens, three empirical vignettes (healthcare, recruitment, criminal justice), and the S.P.A.C.E. framework with policy recommendations. In practice, we combine narratives and real-world human–algorithm collaboration patterns (Frontoni, 2024) with a territorial reading of AI infrastructures, regulatory landscapes and public-value aims (Epasto, 2024a), showing how responsible adoption depends on contextual validation, accountability architectures and stakeholder engagement (Frontoni, 2024; Epasto, 2024a).

Public discourse frequently casts AI as humanity’s “next” or “final” frontier – a boundless, ungoverned domain awaiting exploration. Metaphors matter: their prime expectations, organise policy priorities, and legitimise particular forms of action. The frontier trope, forged historically in colonial expansion and the extractive logics of resource appropriation, obscures how AI is already deeply embedded in dense ecologies of institutions, infrastructures, and everyday practices. Rather than empty spaces, we encounter layered territories: standards, data centres, fibre backbones, public services, workplaces, courts, and clinics – all patterned by power relations and subject to contested governance.

We propose a shift from frontier imaginaries to a framework of coproduction: the ongoing, reciprocal shaping of algorithmic systems and human worlds across multiple scales (from code and datasets to organisations and territories). This shift clarifies the stakes of responsible AI adoption and aligns with a wider move in Europe from techno-solutionist narratives to governed innovation. (Frontoni, 2024; Jasanoff, 2004; Bowker & Star, 1999; Slotkin, 1992; Haraway, 1988; Kitchin, 2014).

3. FROM “FINAL FRONTIER” TO HUMAN–ALGORITHM COPRODUCTION

The frontier imaginary presumes linear progress and portable solutions, yet AI systems are irreducibly

Table 1. Four planes of human-algorithm coproduction (summary)

Plane	Scope	Typical elements / examples	Possible indicators / checks
Material–infrastructural	Physical + digital substrate enabling AI	Compute capacity (GPU/TPU), energy & cooling, networks, sensors/IoT, edge devices, data centres, supply chains	Energy/water use; latency; data locality; resilience; lifecycle/Scope 3 reporting
Institutional–organisational	Rules, roles, and processes around AI	Procurement specs, oversight bodies, liability allocation, labour regimes, professional protocols, QMS	Risk mgmt docs; human-oversight design; audit trails; incident reporting; vendor lock-in risk
Socio-cultural	Values, practices, and expectations	Tacit expertise, local practices, fairness expectations, trust & care norms, user training	Stakeholder mapping; participatory design records; user-facing explanations; satisfaction/trust surveys
Spatial–geopolitical	Territorial and cross-border dimensions	Territorial inequalities, cross-border data flows, standards, digital sovereignty, compliance geographies	Data-flow maps; localisation tests; standard conformance; equity across regions/populations

Notes: Examples are illustrative; indicators are suggestive and non-exhaustive.
Source: Authors' elaboration.

situated: their performance, risks and social meanings depend on how they are designed, trained and embedded within concrete institutions, infrastructures and practices (Haraway, 1988; Bowker & Star, 1999; Jasanoff, 2004). Rather than a neutral, unbounded space, AI inhabits layered territories shaped by norms, labour, infrastructures and geopolitical standards (Kitchin, 2014; Pohle & Thiel, 2020; Epasto, 2024b). This calls for a shift from exploration metaphors to a lens of coproduction, where human actors and algorithmic systems continually shape one another across settings and scales (Jasanoff, 2004; Latour, 1993).

We operationalise this shift through four intertwined planes (see Table 1):

1) the *material-infrastructural* plane (compute, energy, networks, sensors, edge devices, data centres and supply chains), which conditions feasibility, latency and environmental externalities;

2) the *institutional-organisational* plane (procurement, oversight, liability, labour regimes and professional protocols), which structures responsibilities and allocates power;

3) the *socio-cultural* plane (values, tacit expertise, local practices and expectations of fairness, trust and care), which mediates acceptance, use and contestation; and

4) the *spatial-geopolitical* plane (territorial inequalities, cross-border data flows, standards and digital sovereignty), which links AI to uneven development and regulatory geographies (Kitchin, 2014; Pohle & Thiel, 2020; Epasto, 2024a).

Seen through these planes, risks such as data colonialism, bias transfer and vendor lock-in are not accidental defects but predictable consequences of deploying generic tools without local adaptation (Couldry & Mejias, 2019; Pasquale, 2015). Conversely, place-aware design – contextual validation, transparent procurement, human oversight integrated into real workflows – turns compliance into a governance asset and aligns AI with public value and territorial needs (Veale & Zuiderveen Borgesius, 2021; Epasto, 2024b). This coproduction perspective prepares the ground for Europe's regulatory turn discussed next.

4. EUROPE'S REGULATORY TURN: THE AI ACT AS A GEOGRAPHY OF GOVERNANCE

The EU Artificial Intelligence Act establishes a risk-based architecture that classifies systems as unacceptable, high-risk, or limited-risk and attaches graduated obligations across the lifecycle – risk management, quality management, technical documentation, human oversight, and post-market monitoring (European Union, 2024). Rather than a mere checklist, these obligations compose a governance infrastructure that structures how AI is built, placed on the market, monitored, and, if needed, withdrawn (European Union, 2024).

Seen in spatial terms, the Act produces a geography of regulation: CE-marked systems circulate within the internal market; notified bodies, registries, and supervisory authorities act as nodes in an oversight network; and the Act's extraterritorial reach projects European standards along global value chains (European Union, 2024; Veale & Zuiderveen Borgesius, 2021). In this sense, compliance travels with systems, shaping procurement choices, audit pathways, and accountability across borders (Veale & Zuiderveen Borgesius, 2021).

The European risk-based turn embodied in the EU AI Act is unfolding within an increasingly competitive global regulatory arena. In the AI governance landscape, the United States leans towards market-driven innovation, progressively reinforced by strategic export controls, whilst China pursues state-led AI standard-setting, tightly coupled with industrial policy and data localisation imperatives. The European Union distinctly diverges from both approaches by treating fundamental-rights protection, ex-ante conformity assessment, human oversight, and post-market monitoring not merely as compliance obligations, but as instruments of geopolitical differentiation. Through its extraterritorial scope and the circulation of CE-marked AI systems within the internal market, the EU AI Act extends European normative influence across global AI value chains, seeking to mitigate asymmetric technological dependencies – particularly for peripheral regions – and to articulate a robust model of digital sovereignty. In this sense, regulation becomes a source of regulatory power, anchoring AI innovation to European public-value commitments and positioning Europe as a principled, yet competitive, actor in the global contest over AI governance.

Conceptually, the regime shifts practice from expansion to calibration, prioritising clearly delimited purposes, contexts, and safeguards; from universalism to localisation, by demanding context-sensitive performance metrics, domain-appropriate datasets, and stress-testing in real institutional environments; and from opacity to explicability, through documentation, traceability, and stakeholder-facing transparency duties that render reasoning and limits legible (Veale & Zuiderveen Borgesius, 2021; Floridi et al., 2018).

Read through a territorial lens, the AI Act also redistributes capabilities across places: compliance becomes a spatial practice that must align with regional policy priorities and the realities of local service delivery (Epasto, 2024). This reframes compliance from legal hurdle to governance asset, anchoring AI to public-value objectives via place-aware validation, accountable procurement, and situated human oversight (Epasto, 2024a; Veale & Zuiderveen Borgesius, 2021).

5. SPATIAL EMBEDDEDNESS: WHERE AI “LIVES” AND WHY PLACE MATTERS

AI systems are spatial technologies: they depend on energy, water, cooling and low-latency connectivity, and they benefit from proximity to data sources and users. As a result, capabilities and impacts are unevenly distributed – metropolitan cores concentrate compute, talent and venture capital, while peripheral regions risk forms of *algorithmic dependency* when they import generic tools that are poorly adapted to local institutions (Graham & Marvin, 2001; Kitchin, 2014; Epasto, 2024a). Environmental externalities further bind AI to place: training and serving large models impose energy and water burdens that vary with local grids, climate and data-centre siting (Strubell, Ganesh & McCallum, 2019; Henderson et al., 2020).

Spatial embeddedness operates through at least three layers. Infrastructure localisation sets feasibility constraints (compute regions, edge nodes, latency budgets, data-residency rules) that influence architecture choices and service quality (Kitchin, 2014; Epasto, 2024a). Regulatory landscapes define what is permissible and how risk is governed – sectoral rules in health, employment or policing intersect with data protection and fundamental-rights safeguards (European Union, 2024; Veale & Zuiderveen Borgesius, 2021). Civic ecosystems – professional bodies, civil society, affected communities – mediate acceptance, contest harms and co-produce corrective practices; without this mediation, ostensibly “general-purpose” models can reproduce extractive data colonialism (Couldry & Mejias, 2019).

A place-aware approach consequently treats compliance as a governance asset rather than a burden: contextual validation, transparent procurement, clear human-oversight roles and continuous post-market monitoring align AI with public-value aims and territorial needs (Epasto, 2024; European Union, 2024). In short, AI does not float above space; it is built into infrastructures, rules and social contracts that are situated and therefore must be governed as such (Kitchin, 2014; Epasto, 2024).

6. EMPIRICAL VIGNETTES: COPRODUCTION IN PRACTICE

The empirical vignettes are illustrative caselets, purposely selected to exemplify how human–algorithm coproduction manifests in fundamental-rights–sensitive domains, rather than comprehensive or statistically generalisable case studies.

Each vignette has been chosen for its capacity to illuminate situated algorithmic effects across the coproduction planes introduced earlier.

6.1 Healthcare diagnostics (high-risk)

Clinical AI for imaging triage or decision support augments, rather than replaces, professional judgement; performance hinges on dataset curation, protocol design, threshold calibration and exception handling across diverse populations and workflows (Topol, 2019; Kelly, Karthikesalingam & Suleyman, 2019; Wiens et al., 2019). Under the EU AI Act, such tools typically fall within high-risk categories, triggering obligations on risk management, human oversight, robustness and post-market monitoring (European Union, 2024). A coproduction stance translates these legal duties into operational practices: multi-site validation across hospitals, fairness analyses on clinically relevant sub-groups, drift detection linked to incident reporting, and user-facing explanations that fit real clinical routines (Wiens et al., 2019; European Union, 2024). Where implemented, human–algorithm collaboration becomes structured, not ad hoc – clarifying who verifies, who escalates and how accountability is traced (Frontoni, 2024).

Territorial takeaway. Clinical utility and equity improve when validation and oversight are embedded in local health-system workflows, with procurement requiring context-specific evidence beyond aggregate accuracy (Epasto, 2024; European Union, 2024).

6.2 Recruitment and employment (high-risk)

Automated screening and assessment interact with stratified labour markets and education systems; naively transferring models across regions can encode local hierarchies and produce disparate impacts (Raghavan et al., 2020). A coproduction approach treats HR professionals, worker representatives and candidates as stakeholders in tool selection, impact assessment and redress pathways, rather than mere end-users (Barocas, Hardt & Narayanan, 2019). Within the AI Act's risk-based regime, organisations should implement bias impact assessments across relevant attributes and locations, provide candidate-facing notices and explanations, and schedule periodic re-validation as markets change (European Union, 2024; Veale & Zuiderveen Borgesius, 2021).

Territorial takeaway. Because labour-market structures vary geographically, audits and transparency measures must be locally informed and coupled to institutional realities (Epasto, 2024).

6.3 Criminal justice (prohibitions and high-risk zones)

Predictive policing, risk assessment and automated evidence analysis operate within historically over-policed communities, where algorithmic feedback loops can entrench unequal surveillance and sanctioning (Ferguson, 2017; Lum & Isaac, 2016). The AI Act delineates prohibitions for certain biometric surveillance uses and tightens controls elsewhere, emphasising traceability, documentation and meaningful human oversight (European Union, 2024). A coproduction lens centres due process: explainability to defendants, defence access to documentation, independent audits and community consultation – especially where technologies intersect with racialised or socio-economic vulnerabilities (Benjamin, 2019; European Union, 2024).

Territorial takeaway. Oversight designs should be localised – including non-AI alternatives – where harm patterns are spatially concentrated and historically rooted (Epasto, 2024).

Taken together, these vignettes show that responsible AI hinges less on abstract performance scores than on how models are situated within real infrastructures, institutions and communities – where risks, accountabilities and benefits are unevenly distributed. In practice, the AI Act's risk-based duties only become meaningful when translated into place-aware validation, explainable workflows and sustained human oversight (European Union, 2024; Veale & Zuiderveen Borgesius, 2021). Building on this evidence – and on collaborative patterns between humans and algorithms (Frontoni, 2024) and a geography-of-governance perspective (Epasto, 2024) – the next section distils an operational framework, S.P.A.C.E., to align design and deployment with public-value objectives across contexts.

7. AN INTERDISCIPLINARY FRAMEWORK FOR RESPONSIBLE ADOPTION: S.P.A.C.E.

We translate the coproduction perspective into an operational framework – S.P.A.C.E. – that links technical choices to institutional accountability and territorial needs across the AI lifecycle (Jasanoff, 2004; Veale & Zuiderveen Borgesius, 2021; Epasto, 2024).

S.P.A.C.E. operationalises human–algorithm coproduction by translating the four coproduction planes into a pragmatic, auditable workflow across the AI lifecycle. Building on the material–infrastructural, institutional–organisational, sociocultural, and spatial–geopolitical layers outlined earlier (Table 1), the framework proceeds sequentially through: Situation mapping (S), to diagnose territorial assets and infrastructural constraints; Purpose specification (P), to set narrow, testable objectives and non-goals; Accountability architecture (A), to assign oversight roles, liabilities, and traceable audit artefacts; Contextual validation (C), to implement multi-site, multi-group robustness and fairness checks, coupled to model-drift monitoring in diverse locales and sub-populations; and Engagement & education (E), to embed stakeholder co-design, institutional literacy, and transparent, auditable procurement. Rather than a stand-alone construct, S.P.A.C.E. acts as a spatial governance lens that calibrates AI systems to place, rights, and organisational routines, turning regulatory compliance into a co-produced public-value asset along the lifecycle. This operational proposal advances the wider coproduction scholarship on science and technology governance (Jasanoff [2004]; Haraway [1988]), as well as contemporary analyses of the EU AI Act and accountable AI adoption (Veale & Zuiderveen Borgesius [2021]; Epasto [2024]; Frontoni [2024]). Key works informing the geopolitical and situated AI governance lens include: Frontoni [2024], Jasanoff [2004], and Veale & Zuiderveen Borgesius [2021].

8. PITFALLS AND PATHWAYS

Frontier rhetoric drives *scale-at-all-costs*, obscuring local harms and institutional dependencies. Typical failure modes include: data colonialism and biased transfer when models are deployed without local adaptation (Couldry & Mejias, 2019); opacity lock-in via proprietary ecosystems that limit oversight and exit options (Pasquale, 2015); unsustainable footprints from compute-intensive pipelines (Strubell, Ganesh & McCallum, 2019; Henderson et al., 2020); and compliance theatre, i.e., paper-based conformity that ignores real workflows (Veale & Zuiderveen Borgesius, 2021).

Counter these risks through: (a) contextual validation and routine re-evaluation in live settings; (b) rights-based impact assessments with public-facing summaries; (c) procurement for auditability (documentation, interfaces for oversight, energy reporting) and interoperability to avoid lock-in; (d) federated and privacy-preserving learning where feasible; and (e) worker and community co-design to align with public value (Floridi et al., 2018; Kairouz et al., 2021; Veale & Zuiderveen Borgesius, 2021; Epasto, 2024; Frontoni, 2024).

9. POLICY RECOMMENDATIONS

1. *Regional validation hubs.* Fund cross-institutional testbeds that mirror local infrastructures and demographics; publish datasets, stress tests and post-market metrics where possible (European Union [2024]; Epasto, 2024a; Epasto, 2024b).
2. *Public-value procurement.* Require explainability artefacts, audit APIs, fairness/robustness evidence across sub-populations and locales, and *energy/water disclosures* for high-compute systems (Veale & Zuiderveen Borgesius, 2021; Strubell, Ganesh & McCallum, 2019).
3. *Mandatory post-market monitoring (high-risk).* Implement drift detection, incident logs, and scheduled *sunset reviews* to reassess necessity and proportionality (European Union [2024]).
4. *Rights-based impact assessments.* Standardise templates for health, employment and public administration, with public summaries and independent audit capacity (Floridi et al., 2018; Veale & Zuiderveen Borgesius, 2021).
5. *Data governance for place.* Support data trusts/commons with community oversight to ensure representativeness and legitimate reuse (Couldry & Mejias, 2019; Epasto, 2024).
6. *Skills and literacy.* Invest in interdisciplinary training for officials and professionals; recognise and formalise *human-AI collaboration patterns* in sectoral protocols (Frontoni, 2024; Epasto, 2024a).

10. CONCLUSION

Replacing the "final frontier" imaginary with situated coproduction clarifies both risks and responsibilities. Europe's risk-based regime provides an institutional scaffold, but alignment with public value is ultimately achieved through place-aware governance: contextual validation, accountable procurement, and meaningful engagement with affected communities (European Union, 2024; Veale & Zuiderveen Borgesius, 2021; Epasto, 2024). When human-algorithm collaboration is designed and monitored within real workflows and territorial constraints, AI shifts from a project of conquest to one of care and democratic utility (Frontoni, 2024; Jasanoff, 2004; Floridi et al., 2018; Epasto, 2024a; Epasto, 2024b).

REFERENCES

Barocas, S., Hardt, M. & Narayanan, A. (2019). *Fairness and Machine Learning*. fairmlbook.org.

Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. Cambridge: Polity.

Bowker, G.C. & Star, S.L. (1999) *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.

Couldry, N. & Mejias, U.A. (2019). *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford, CA: Stanford University Press.

Epasto, S. (2024a). *Geografia, sviluppo e innovazione. Scenari e traiettorie tra globalizzazione e sostenibilità*. Roma: Aracne Editrice.

Epasto, S. (2024b). Geopolitica, geoeconomia e geofinanza: intrecci e implicazioni per una sostenibilità finanziaria inclusiva e responsabile. In Biasin, S., Giacomini, G. & Marinelli, E. (Eds) *La sostenibilità. Percorsi tra ambiente, società e governance*. Macerata: EUM.

European Union (2024). *Artificial Intelligence Act*. Official Journal of the European Union.

Ferguson, A.G. (2017) *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*. New York: NYU Press.

Floridi, L. et al. (2018). AI4People – An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707.

Frontoni, E. (2024). *AI, ultima frontiera. Capire l'intelligenza artificiale con le storie vere di collaborazione tra esseri umani e algoritmi*. Milano: ROI Edizioni.

Graham, S. & Marvin, S. (2001). *Splintering Urbanism: Networked Infrastructures, Technological Mobilities and the Urban Condition*. London: Routledge.

Haraway, D. (1988). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*, 14(3), 575–599.

Henderson, P. et al. (2020). Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning. *arXiv preprint* arXiv:2002.05651.

Jasanoff, S. (2004). *States of Knowledge: The Coproduction of Science and Social Order*. London: Routledge.

Kairouz, P. et al. (2021). Advances and Open Problems in Federated Learning. *Foundations and Trends in Machine Learning*, 14(1–2), 1–210.

Kelly, C.J., Karthikesalingam, A. & Suleyman, M. (2019). Key challenges for delivering clinical impact with AI. *BMC Medicine*, 17, 195.

Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: SAGE.

Latour, B. (1993). *We Have Never Been Modern*. Cambridge, MA: Harvard University Press.

Lum, K. & Isaac, W. (2016). To Predict and Serve? *Significance*, 13(5), 14–19.

Noble, S.U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press.

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

Pohle, J. & Thiel, T. (2020). Digital sovereignty. *Internet Policy Review*, 9(4).

Raghavan, M., Barocas, S., Kleinberg, J. & Levy, K. (2020). Mitigating bias in algorithmic hiring: Evaluating claims and practices. In *Proceedings of FAT '20* (pp. 469-481). New York: ACM.

Slotkin, R. (1992). *Gunfighter Nation: The Myth of the Frontier in Twentieth-Century America*. Norman: University of Oklahoma Press.

Strubell, E., Ganesh, A. & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. In *Proceedings of ACL 2019* (pp. 3645–3650). Florence: ACL.

Topol, E. (2019). *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. New York: Basic Books.

Veale, M. & Zuiderveen Borgesius, F. (2021). Demystifying the Draft EU Artificial Intelligence Act. *Computer Law Review International*, 22(4), 97–112.

Wiens, J. et al. (2019). Do no harm: a roadmap for responsible machine learning in health care. *Nature Medicine*, 25, 1337–1340.

JEP

Firenze University Press
https://oaj.fupress.net/index.php/jep

OPEN ACCESS

Citation: D'Aponte, V. (2025) Personalizing the tourist offer with the AI applications of "Unexpected Italy". *Journal of Emerging Perspectives* 2: 17-24. doi: 10.36253/jep-20878

Received: April 30, 2025

Revised: September 30, 2025

Published: December 20, 2025

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

Personalizing the tourist offer with the AI applications of "Unexpected Italy"

Viviana D'Aponte

Dipartimento di Studi Economici e Giuridici, Uniparthenope, Italy
E-mail: viviana.daponte@uniparthenope.it

Abstract. In recent years, the intersection of tourism and artificial intelligence (AI) has opened new frontiers for innovation in travel experiences, customer engagement and destination management, particularly regarding the overtourism problem. In Italy, where cultural and environmental heritage faces increasing pressure from overtourism, innovative approaches are crucial to ensure sustainable development. This paper explores the role of AI in fostering more balanced tourism, with a focus on the case of "Unexpected Italy", a pioneering start-up that leverages AI to promote lesser-known destinations and off-the-beaten path experiences. Trough personalized recommendations, predictive analytics, and data driven storytelling. Unexpected Italy aims to redistribute tourist flows, support local economies, and preserve the authenticity of Italy's diverse territories. The study highlights how AI powered solution can serve as a strategic tool not only to improve visitor experiences but also to mitigate the negative impacts off mass tourism, offering a replicable model for sustainable tourism development in culturally rich yet vulnerable regions.

Keywords: tourism, Artificial Intelligence, management of destinations, new model of tourism.

1. ARTIFICIAL INTELLIGENCE (AI) BETWEEN RESEARCH AND SOCIAL IMPACT

The subject of artificial intelligence (AI) is gaining increasing significance in research, particularly in the efficient management of related operational infrastructures. It drives the ongoing expansion and enhancement of a wide range of applications that are progressively impacting various facets of contemporary society[1]. However, it is essential to recognize the environmental costs associated with energy-intensive IT infrastructures and the potential distortions that uncontrolled applications of these technologies could create in socioeconomic and sociopolitical spheres.

Notably, the ability to analyse multiple data points – enabled by the vast availability of large datasets (metadata) – aligns with advancements in technologies that promote the development of increasingly powerful electronic

[1] For the history of the evolution of AI, Boden, 2006.

machines (computers). Consequently, it is not surprising that the term "artificial intelligence," coined seventy years ago by renowned Stanford University scientist John McCarthy (1959), refers to the scientific field focused on the engineering of intelligent machines.

The transition from theoretical research to the early implementation of AI programs in industrial settings spanned about forty years – from the 1950s to the early 1990s. During this period, the primary geoeconomic activity was centred in the United States, with Europe playing a limited role until significant intervention occurred with the launch of an innovative EU-funded research initiative (the Horizon Program).

The substantial development of "machine learning" emerged after 2010, when large U.S. multinationals, such as Google, Apple, Microsoft, and Amazon, invested significant financial resources in AI applications that quickly became integrated into key areas of society, including commercial, financial, and daily practices. The remarkable ease of operation and free access to platforms accessible to the public facilitated the mass adoption of "deep learning" and generative AI, which raised concerns over security issues, fraudulent profiling, and serious invasions into personal privacy, underscoring the need for specific regulations[2] (Fameli , 2025).

Regarding the relevance of this topic, the range of applications encompassing automation of business processes, predictive analysis, and governance of complex systems presents vast opportunities for development in the social sciences. Many startups are providing innovative services that make an immediate impact on various cultural activities. This widespread of generative AI platforms has required the sharing of initiatives and application models both in established settings and experimental projects.

As a result, especially after the pandemic, there has been a rise in detailed surveys and reports that document where AI methodologies have successfully enhanced project effectiveness across multiple human activities[3]. The sector most relevant to this research, for which extensive documentation exists, is tourism. The World Tourism Organization (WTO), in collaboration with Saxion University of Applied Sciences, recently released a comprehensive report highlighting the opportunities that AI offers for the efficient and profitable management of tourism activities (WTO-Saxion, 2025).

Furthermore, research in geography is highly focussed on the application of AI to tackle significant contemporary challenges in geoeconomic and geopolitical domains, as well as in the critical areas of environmental and social compatibility within territorial planning processes.[4] Ultimately, there is a strong need to further investigate AI- centred methodologies in the geographical sciences, where the broad thematic scope allows research that can harness new opportunities for comprehensive documentation and timely information retrieval – capabilities that would be difficult, if not impossible, to achieve without AI techniques.

2. METHODS AND SOURCES

This study adopts an integrated methodological approach combining secondary statistical data and AI-enabled digital analytics to analyse assess and rebalance tourist flows. The research draws on OECD tourism benchmarks, ISTAT tourism flow distribution statistics, and congestion-management insights from IoT sensor analytics and platform-based user-profiling systems. AI techniques applied to tourism rebalancing include recommendation algorithms, predictive-flow analysis, and dynamic routing. The analysis focuses on a set of carrying capability indicators: spatial concentration of tourist stays, seasonality patterns, social pressure on local communities, and overtourism governance potential. The case study evaluation follows an explanatory-sequential logic, where AI digital outputs are interpreted in the light of sustainable destination-governance challenges to derive practitioner-oriented implications**.**

3. THE RAPID INCREASE IN TOURIST DEMAND AND THE RISK OF OVERTOURISM

The drastic decline in global tourism flows caused by the severe health crises of the Covid-19 pandemic has been fully absorbed, as indicated by the overall trends observed since the beginning of the 2023 season. Statis-

[2] In February 2025, the European regulatory mechanism for the civil use of AI (Reg. EU 1689 of 2024, known as the AI Act) came into effect.

[3] Several reports are available that analyse both the areas in which AI can be advantageously employed and the types of platforms available for services offered by providers in the global market (Avasant, 2024). Particularly useful is the publication of a general report on AI (Stanford, 2025, 8th edition), the first edition of which was published by the Stanford Institute for Human-Centred Artificial Intelligence (HAI) in 2018.

[4] Studies focused on AI applications that fall within the disciplinary realm of geoeconomic-geopolitical character are predominantly developed in Anglo-Saxon countries, but significant approaches from European scholars are also present. In general, analyses are oriented towards the search for new interpretative models of the interactions between geoeconomics and geopolitical factors that influence global structures and relationships. By way of examples, or due to their methodological significance, two recent studies are noted (Neacșu et al., 2025; Siqin Wang et al., 2024).

tical analyses reveal that many major touristic regions have completely recovered to pre-2019 levels (OECD, 2024; Bank of Italy, 2024). In Italy, preliminary data from Istat regarding hotel stays show that the numbers have surpassed those recorded in 2023. For the national tourism economy, this represents a complete recovery of pre-pandemic conditions, achieving a new record with over 450 million stays. Specifically, regarding the distribution of flows within the national tourism system, it is evident that in 2024, two-thirds of stays will concentrate in "cultural cities," with particularly strong trends noted in Rome, Milan, Florence, and Bologna. Additionally, there is an exceptional growth in the city of Naples and the main tourist destinations in the Gulf, such as Sorrento, Capri, Ischia. This is part of a favourable overall performance in the southern region, with sustained demand for destinations in Puglia (Istat, 2025).

Essentially, the positive trend in tourist demand, as evidenced by global statistics, when linked to specific destinations known for their attractiveness, results in significant concentration phenomena. This creates a challenging balance both environmentally and in terms of negative externalities for local communities. Although discussions around hyper-tourism and the overcrowding of certain tourist destinations are not new, it is only in recent years that the phenomenon appears to have gained attention from public opinion and industry stakeholders[5]. This has led to a marked increase in studies focused on analysing the conditions related to a density of tourism, referred to as "overtourism" (Goessling et al., 2020; Mihalic, 2020).

This renewed interest, undoubtedly intensified by the exacerbation of the phenomenon due to the rising tourism pressure in the post-pandemic era, is also motivated by contemporary society's emphasis on sustainability. This focus has been further amplified by recommendations from international organizations concerning potential strategies for mitigating environmental damage (Bimonte & Punzo, 2007).

In the 1970s and 1980s, most studies focussed on the high environmental costs associated with uncontrolled tourism growth in overcrowded areas, highlighting the concept of a carrying capacity – defined in 1981 as the optimal ratio between the number of individuals in a space and the available resources. However, the current debate goes beyond that logic, explicitly addressing the impact that high tourist concentration has on local communities, leading, in extreme cases, to the expulsion of residents from historic centres and a progressive loss of identity in these places.

Moreover, from its earliest formulation, the concept of carrying capacity has revealed the limitations of relying solely on a quantitative indicator based on a "threshold value", beyond which the load capacity of a given environment is no longer ensured. There is a need for a more qualitative approach that incorporates a dynamic perspective on how local communities can adapt to the inevitable pressures and develop suitable strategies for "carrying capability". Therefore, an integrated and comprehensive approach is required, focussing on the ability of local institutions to manage these phenomena and the adaptability of the socioeconomic system to human pressures. The goal is to ensure sustainable tourism that not only preserves the identity of destinations but also benefits local communities (Prezioso, 2010).

On an epistemological level, to properly address the phenomenon of overtourism, it is essential not merely to assess visitor volume but also to prioritize the governance of this issue. Establishing a constructive balance between the tourist presence and the quality-of-life expectations of local communities is crucial. There is no universal metric to quantify when the pressure from tourist flows becomes excessive within a specific context. However, the resilience of the affected social system is undoubtedly linked to the environmental and economic factors altered by the pressure from tourist concentration. Moreover, it is notably shaped by the effective "rebalancing" initiatives implemented by local governance in managing the phenomenon. This line of reasoning illustrates why some destinations can successfully handle a high number of visitors, while others encounter adverse effects at significant levels of the presence of tourists.

In other instances, the causes may be more complex, including an existing infrastructure that is both inadequate and undersized to meet the needs of the local population, the school vacation calendar, and the growing influence of social media which highlights "Instagrammable" locations, leading to excessive human pressure. Additionally, poorly managed urban development can exacerbate these issues. Often, these factors are interconnected, triggering a negative feedback loop that intensifies public perception of the situation[6].

[5] The need for proper calibration between environmental endowment and the utilization of related resources emerged as early as the 1960s, when, at the initiative of the Cassa per le Opere Straordinarie nel Mezzogiorno (CasMez), specific studies were undertaken to measure the "temporary" human carrying capacity of certain territorial areas, with particular attention to the coast and the use of beach resources. Andriello, 1964.

[6] The difficult determination of a unifying paradigm capable of identifying "crisis" situations and proposing appropriate "solutions" explains the proliferation of specific research focused on "case studies" present in the literature. There is a notable presence of research referring to historically significant art cities with established attraction. In this regard, studies related to the cases of Venice (Bertocchi, 2022) and Dubrovnik (Camatti, 2020) appear particularly interesting.

What is clear, however, is that unmanaged overcrowding is unsustainable in the long term for any city or region. While overcrowding can lead to significant economic and social costs, it is essential to assess how these costs compare to the benefits generated by tourism. Furthermore, it is important to consider whether tourism revenues can be more effectively leveraged to mitigate the costs and ultimately enhance the quality of life for residents.

Conversely, effectively addressing and managing the phenomenon of overtourism requires coordinated efforts from both public and private stakeholders. It is to ensure that the growth of tourism enhances rather than deteriorates the socioeconomic fabric and ecosystem of the destinations. This fabric, or the uniqueness of a place, often serves as a primary form, particularly in today's era of globalization. Consequently, overcrowding poses a threat not only to local residents and businesses but also reduces the overall attractiveness of the destination. This situation underscores the critical need for a collaborative approach grounded in concrete data.

Such considerations bring us to the need for a multidimensional and multidisciplinary approach that, with the support of suitable generative AI techniques, can provide a comprehensive and precise identification of the various components that define the complexity of the phenomenon under study.

All stakeholders in the sector – policymakers, communities, and the tourists themselves – must be engaged in a participatory process aimed at adopting practices and guiding principles for a more resilient future tourism. This involves redefining the concept of sustainability in tourism as a democratic principle, where the interests of residents align with the needs of temporary visitors drawn to the uniqueness of the destination in pursuit of their emotional expectations. Ultimately, this requires promoting tourist accommodation consistent with a new integrated vision that addresses the evolving needs of the contemporary tourism system (Perkumiene, 2019).

4. OVERTOURISM AND ARTIFICIAL INTELLIGENCE: NEW PERSPECTIVES FOR SUSTAINABLE DESTINATION MANAGEMENT

The rapid acceleration of Artificial Intelligence (AI) technologies, particularly in recent years, has significantly impacted the tourism sector. This development has established a pivotal role for AI not only among large companies within the industry but also for countless small businesses operated by independent entrepreneurs. The integration of artificial intelligence into this field has fostered an environment where digital innovation such as blockchain, cyberspace, and the Internet of Things (IoT) are widely applied. These technologies have streamlined operations and enhanced decision-making processes, leading to a gradual transformation in customer service. This transformation facilitates the swift and straightforward distribution of information about destinations, transport options, accommodation, cultural attractions and opportunities for interactions between cultural expectations and leisure activities. As a result, operators can quickly adapt their offerings to meet emerging demands, promptly providing travel options tailored to experiential content, incorporating clear qualitative elements to help potential customers assess the cost-benefit ratio of various options.

On a broader scale, for users with linguistic barriers, widely accessible applications available on social networks now offer a comprehensive language support. These applications personalize itineraries based on selected routes and provide access to descriptive content tailored on geolocation, alongside a wealth of detailed information on local customs, culinary traditions, behavioural norms, and other factors that can enhance the overall travel experience (Aru, 2022).

However, it is primarily in the business realm that contemporary AI platforms demonstrate the most significant impacts in terms of development and system efficiency. Through their computational capabilities and the access to vast amounts of data via powerful digital infrastructures, these platforms have driven notable evolution in business models focused on efficiency and the optimization of management strategies, significantly influencing revenue (Pellegrini, 2024; Miotto, 2018). Tour operators and hospitality businesses are increasingly leveraging these tools for predictive analyses that optimize pricing strategies, improve transportation planning and itinerary design, and identify optimal partnerships with local complementary businesses. This creates more attractive integrated tourism packages, both in terms of price and quality of included services.

Nevertheless, like any transformative technology, the integration of AI in the tourism sector presents a range of complex challenges alongside its opportunities and widespread advantages. While it enables greater efficiency, personalization, and innovative experiences, it also raises significant concerns regarding privacy protection. This is particularly true when it comes to profiling sensitive personal data or implementing behaviour patterns that may not always align with the human dimension, which is fundamental to maintaining customer loyalty (Baggio, 2019; Floridi, 2013).

Furthermore, the real effect of the widespread adoption of AI platforms on the obsolescence of traditional job functions within the industry remains uncertain, making it challenging to assess from theoretical perspectives. This uncertainty affects ongoing processes and the demand for innovative skills that are not available in the labour market. These issues are crucial but lack clear definition or consensus, despite the recognition of their social implications.

While the interplay of positive and negative aspects continues to shape the search for a nuanced understanding of the scenarios that the new frontier of AI in tourism will ultimately project onto the overall economy of the sector, there is no doubt that AI assistance represents a fundamental approach for finding appropriate solutions to fight the phenomenon of overtourism, especially in light of the geographical risks involved.

From this perspective, addressing this specific challenge, the WTTC and McKinsey & Company (2017) suggest an integrated approach that, rather than resorting solely drastic measures for controlling flow, focuses on progressive adaptation strategies for the territorial system through targeted investments. In other words, it is essential for destination management to not only implement strategic tools for sustainable growth in the face of tourist overcrowding but also to engage all stakeholders – local communities, public institutions, and regional operators – in the efforts to mitigate or prevent overtourism.

To achieve this, it is crucial to pinpoint the root causes of appropriate investments in infrastructure and environmental protection measures while also securing the financial resources. This can be accomplished through the use of research tools supported by efficient AI platforms, capable of cross-referencing multiple data sources and processing large volumes of information to build an appropriate database for monitoring and optimizing the distribution of visitor flows and suggesting alternative planning solutions for specific tourist areas. This approach makes it easier to identify the root causes of overcrowding and coordinate targeted interventions that can progressively reduce the environmental risk level and mitigate the likelihood of social conflict, without necessarily abandoning growth-focused policies. Implementing reliable monitoring systems and early warning mechanisms remains an ambitious goal that is facilitated using new AI technologies, fostering enhanced collaboration among all stakeholders in the local tourism economy, safeguarding the environment, and ensuring satisfactory living conditions for the local population unaffected by tourism income redistribution.

To attain these positive outcomes, a collective agreement on relevant economic, social, and environmental indicators is essential. This process can be accomplished with the help of AI tools designed to identify and analyse connections, joint effects, and functional interactions among various factors operating over time and within specific geographical spaces.

Investing in the development of intelligent data collection systems and digital platforms to manage visitor flows – similar to initiatives already implemented in Florence – where IoT technology has been used to monitor and control tourist flows and Venice which employers’ sensors, cameras, and mobile signals monitored from a control centre to manage congestion and improve visitor experience – is no longer an option to be deferred. The related infrastructure must now necessarily incorporate innovative AI tools.

AI technologies, such as predictive analytics and machine learning, serve as a strategic instrument for addressing overtourism. They can forecast tourist flows, including their volume and peak visitation time, allowing for more efficient allocation of resources. Additionally, they provide valuable recommendations to users about alternative routes and timing for visiting attractions, which aids in dispersing crowds and to reducing congestion (Shafee, 2024). Furthermore, services based on “dynamic pricing” logic, facilitated by AI-driven applications, can encourage off-season tourism, benefiting both the overall economy of the industry and promoting a

Table 1. Core AI applications for sustainable and balanced tourism.

AI Application	Function in Tourism	Practitioner Benefit	Overtourism Mitigation Outcome
Recommendation Systems	Personalised suggestions for lesser-known destinations	Improves visitor experience	Spatial dispersion of tourist flows
Predictive Analytics	Forecasting visitor volumes and peak times	Optimises staffing and pricing	Anticipation of congestion risks
Dynamic Routing	Real-time adaptive itinerary and path suggestions	Enhances journey efficiency	Reduction of crowding at hotspots
AI-based Storytelling	Data-driven narrative content tailored by interests	Higher cultural engagement	Promotion of peripheral/rural sites

more balanced use of tourism resources. With improved visitors' distribution, intelligent systems for waste and resource management can help minimize environmental impact to levels compatible with existing territorial considerations, supported by the interactions inherent in the blockchain model utilized in the monitoring process.

However, the strategies suggested by AI can only yield positive outcomes if they are implemented alongside intelligent and sustainable practices in a context characterized by continuous collaboration among governments, local communities, and industry stakeholders, all in alignment with sustainable development goals.

5. THE "UNEXPECTED ITALY" CASE

In the context of new AI technologies and their broader implication for tourism – where entrepreneurial interests, social expectations, and environmental sustainability intersect – it is pertinent to discuss a practical application: the Unexpected Italy app. Launched in 2024 by two young innovators, Elisabetta Faggiana and Savio Losito, this cutting-edge mobile application aims to enrich the tourist experience in Italy tourism by a holistic view of the country's diverse offerings. This app seeks to shift focus from well-known tourist attractions to highlight less-known realities, redefining how users experience Italy's rich landscape. Designed to address the challenge of overtourism, it promotes a travel model that directs users toward less famous yet culturally, historically and geologically significant destinations. This not only enhances the appreciation of Italy's national heritage but also fosters a sense of territorial authenticity. Unexpected Italy provides users with engaging content and travel orientations that are framed within specific experiential offerings. It encourages exploration from a fresh and unconventional perspective, guiding travellers towards unique sites, local traditions, cultural events, and intriguing experiences that unveil unexplored customs and lifestyles. In doing so, the app plays a vital role in promoting small villages and rural areas at risk of depopulation, thereby contributing to their revitalization.

By embracing this innovative approach, Unexpected Italy platform enhances the overall travel experience and supports sustainable tourism practices, ultimately benefiting visitors and local communities alike.

From an operational standpoint, Unexpected Italy utilizes a profiling system based on users' personal interests and preferences to create tailored itineraries. It suggests routes and activities while offering users the opportunity to connect with local guides. This setup not only provides travel advice but also shares stories and fascinating insights about various Italian regions, along with practical information to enhance vacation planning. The model is driven by the aim to cultivate a participatory and mindful tourism experience, encouraging interaction between travellers and local communities to promote cultural exchange and support local economies. It operates on the belief that tourists should not simply consume destinations but actively engage with the host communities to uncover their authentic identities.

Upon starting, users select a traveller profile that best represents them from a range of options. Based on this choice, the app proposes personalized itineraries that include both iconic landmarks and recommendations for accommodations and services along the way. Additionally, the app features an interactive map covering fourteen Italian destinations, such as Milan, Matera, Venice, Rome, Modena, and the Valle d'Itria. Each destination is presented through narrative sheets that blend practical details with experiential insights, enriching the user experience.

Another noteworthy feature of the Unexpected Italy app is the rigorous selection process for local partners, who must meet criteria centred around authenticity and environmental sustainability, as well as maintain independent management to be included on the platform. This approach facilitates direct relationships between visitors and host communities, by allowing users to contact local operators directly, thus eliminating intermediaries and commissions. As a result, the app promotes a fairer and more transparent economy while supporting local artisans and small businesses.

From a technologically point of view, Unexpected Italy operates on a model where basic functionalities are available for free, while advanced features – such as in-depth thematic itineraries and detailed digital guides – are offered through a paid subscription.

The description clearly illustrates how initiatives like Unexpected Italy are pivotal in fostering more sustainable and responsible tourism practices. Moreover, this digital tool promotes not only alternative models of tourism consumption but also advances a cultural and social vision of tourism that emphasizes mutual understanding, environmental respect, and regional development.

This approach helps alleviate the pressure on more popular destinations while enabling local communities to reap the economic benefits of tourism in a more equitable manner. Such a methodology not only enhances the tourism experience but also contributes to the preservation and enhancement of the cultural and natural heritage of the areas visited. Initiatives like Unexpected Italy play a crucial role in raising traveller awareness about the significance of sustainable tourism and encouraging responsible travel practices. By diversifying tourism offerings and shining a light on lesser-known locations, substantial strides can be made in managing overtourism and protecting tourist destinations for the future.

The case of Unexpected Italy clearly illustrates how technological innovation can serve as a powerful tool in promoting experiential and sustainable tourism, rooted in social responsibility and cultural sensitivity among travellers. Specifically, this platform aims not only to reshape the landscape of Italian tourism but also to encourage a wider dialogue about the relationship between tourists, the territory, and local communities in today’s digital age.

6. CONCLUSIONS

The intersection of AI and tourism opens new opportunities for balanced destination management in vulnerable yet culturally rich regions. The Unexpected Italy case remains a promising example, but to advance academic value, tourism-flow evidence and AI-driven governance indicators need to be clearly systematised. AI does not merely enhance personalisation for visitors – it becomes a strategic territorial-rebalancing tool when embedded in collaborative destination-governance ecosystems. This approach contributes to the ongoing sustainable-tourism debate and offers replicable practitioner implications for overtourism mitigation grounded in AI-enabled spatial awareness and place authenticity.

REFERENCES

Andriello, D. (1964). La pianificazione fisica del turismo in provincia di Reggio Calabria. Napoli: Libreria Internazionale Minerva Editrice.

Aru, S. (2022). Il turismo Smart. https://iris.unito.it/retrieve/d0403301-3183-4988-9de1-aefc6a8eec01/Turismo_SMART.pdf

Avasant (2024). Applied AI Services: Forging The Future AI Ecosystem With A Key Focus On Gen AI. https://avasant.com/report/forging-the-future-ai-ecosystem-with-a-key-focus-on-gen-ai/

Avasant (2025). Applied AI Sevices 2024-25. https://avasant.com/report/applied-ai-services-2024-2025-radarview/

Baggio, R. (2019), Big Data e turismo: una relazione complicata In Becheri, E. & Morvillo, A. (Ed, 91–106). XXIII Rapporto sul Turismo Italiano 2018/2019, Rogiosi, Napoli.

Banca d’Italia (2024). Indagine sul turismo internazionale. www.bancaditalia.it/statistiche/index.html

Bertocchi, D. & Camatti, N. (2022). Tourism in Venice: mapping overtourism and exploring solutions. In J. van der Borg, J. (Ed*), A Research Agenda for Urban Tourism*, (pp.107-125), Elgar, Cheltenham: Elgar.

Bimonte, S. & Punzo, L.F. (2007). The evolutionary game between tourist and resident populations and Tourist Carrying. Capacity. *International Journal of Technology and Globalization*, (3) 1,73-87.

Boccella, N. & Salerno, I. (2017). “Tourism Trend and Policies” Alcune riflessioni sul Rapporto OECD 2018. *Rivista di Scienze del Turismo*, 8, 19–27.

Bocchini, E. (2024). *La regolamentazione giuridica dell’intelligenza artificiale*. Torino: Giappichelli.

Boden, M.A. (2006*). Mind As Machine: A History of Cognitive Science*. Oxford: Oxford University Press.

Brown, C. & Lee, S. (2019). Leveraging AI for Sustainable Tourism Management: Opportunities and Challenges. *International Journal of Tourism Research*, 36(4), 567–581.

Camatti, N., Bertocchi D., Carić , H. & van der Borg, J. (2020). A digital response system to mitigate overtourism. The case of Dubrovnik. *Journal of Travel & Tourism Marketing*, 37 (8-9), 887-901.

Capocchi, A., Vallone, C., Pierotti, M. & Amaduzzi, A. (2019). Overtourism: A Literature Review to Assess Implications and Future Perspectives. *Sustainability, 11*(12), 3303.

Colombo, E. (2020). Turismo mega trend, Smart destination e turismo digitale: AI, Blockchain, Cyber, IoT e 5G. Milano: Hoepli.

Dodds, R. & Butler, W.R. (2019). The phenomena of overtourism: a review. *International Journal of Tourism Cities*, 5(4), 519–528.

Fameli, E. (2025). La nuova Intelligenza artificiale e il problema della sua regolazione giuridica. La strategia europea e il lungo iter legislativo dell’AI Act. *Rivista italiana di Informatica e Diritto*, 7(1), 551–578.

Floridi, L. (2013). *The ethics of information*, Oxford: Oxford University Press.

Fondazione Dolomiti (Ed) (2020). Da Carrying Capacity a Carrying Capability. https://issuu.com/fondazionedolomitiunesco/docs/fd4u_cc_editing-finale_it_2022.

Istat, Statistiche Today (sept. 2025), Turismo in crescita nel secondo trimestre del 2025. https://www.istat>uploads>2025/09.

Kouroupi, N. & Metaxas, T. (2023). Can the Metaverse and Its Associated Digital Tools and Technologies Provide an Opportunity for Destinations to Address the Vulnerability of Overtourism? *Tourism and Hospitality,* 4(2): 10.3390.

Long, C., Lu, S., Chang, J., Zhu, J. & Chen, L. (2022). Tourism Environmental Carrying Capacity Review, Hotspot, Issue, and Prospect. *International Journal of Environmental Research and Public Health*, 19(24), 16663.

Marchioro, S. & Miotto, A. (2018). La *governance del turismo nell'era digitale*. Bolzano: Gallica. Bolzano.

McCarthy, J. (1959). Programs with Common Sense. http://jmc.stanford.edu/articles/mcc59/mcc59.pdf

McCarthy, J. (2007). What is Artificial Intelligence? Computer Science Dept., Stanford Univ. https://www-formal.stanford.edu/jmc/.

Mihalic, T. & Kuščer, K. (2022). Can overtourism be managed? Destination management factors affecting residents' irritation and quality of life. *Tourism Review*, 77(1): 16–34.

Neacșu, M.C., Eregep, E.Y. & Diaconescu, M. (2025). Artificial Intelligence As A Geopolitical Tool. *Amfiteatru Economic*, 27(68), 253–268.

OECD (2024). Tourism Trends and Policies 2024. Paris: OECD Publishing.

Pellegrini, E. (2024). Il ruolo dell'Intelligenza Artificiale nell'intermediazione turistica: il caso del gruppo Alpitour World. https://hdl.handle.net/20.500.14247/23481.

Perkumiene, D. & Pranskūnienė, R. (2019). Overtourism: Between the Right to Travel and Residents' Rights. *Sustainability*, 11(7), 2138.

Perles-Ribes, J., Ramón-Rodríguez, A., Moreno-Izquierdo, L. & Such-Devesa, M. (2020). Machine learning techniques as a tool for predicting overtourism: The case of Spain. *International Journal of Tourism Research.* https://doi.org/10.1002/jtr.2383.

Prezioso, M. (2010). Progettare lo sviluppo turistico. Percorso di planning economico-territoriale in sostenibilità. In Paniccia, P., Silvestrelli, P., Valeri, M. (Ed). *Economia e management delle attività turistiche e culturali: destinazione, impresa, esperienza: contributi di ricerca* (pp. 3–33). Torino: Giappichelli.

Saxion University of Applied Sciences (2025*). Artificial Intelligence Adoption in Tourism – Key Considerations for Sector Stakeholders*. Madrid/Enschede: WTO.

Shafiee, M. (2024). Navigating overtourism destinations: leveraging smart tourism solutions for sustainable travel experience. *Smart tourism*, 5(2), 2841.

Siqin W. et al. (2024). Mapping the landscape and roadmap of geospatial artificial intelligence (GeoAI) in quantitative human geography: An extensive systematic review. *International Journal of Applied Earth Observation and Geoinformation*, 128: 103734 .

Smith, A. & Doe, J. (2020). Artificial Intelligence and Overtourism: A Review of the Literature. *Journal of Tourism Studies*, 45(2), 213–230.

Stanford Institute for Human-Centered Artificial Intelligence (2025). *Artificial Intelligence Index Report 2025.* https://hai.stanford.edu/assets/files/hai_ai_index_report_2025.pdf

Turing, A. (1950). Computing machinery and intelligence. *Mind,* LIX (236), 433–460.

Yusoh, M., Mapjabil, J., Hanafi, N., Muhammed, I. & Mohd, A. (2021). Tourism carrying capacity and Social Carrying capacity: A literature review. *SHS Web of Conferences*, 124(4), 02004.

JEP

Firenze University Press
https://oaj.fupress.net/index.php/jep

OPEN ACCESS

Citation: Epasto, S., Lucia, M. G., & Pereira, M. M. (2025) Agrivoltaics and artificial intelligence. New geographies of power. *Journal of Emerging Perspectives* 2: 25-31. doi: 10.36253/jep-20879

Received: April 30, 2025

Revised: September 30, 2025

Published: December 20, 2025

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

ORCID
SE: 0000-0002-6420-0867
MGL: 0000-0002-9582-6010

Agrivoltaics and artificial intelligence. New geographies of power

Simona Epasto[1*], Maria Giuseppina Lucia[2], Masiel Melissa Pereira[3]

[1] *Department of Political Science, Communication and International Relations, University of Macerata, Italy*
[2] *Department of Management, University of Turin, Italy*
[3] *Pontificia Universidad Javeriana, Colombia*
Email: simona.epasto@unimc.it; mariagiuseppina.lucia@unito.it; masiel.pereira@javeriana.edu.co
*Corresponding author

Abstract. This paper explores the integration of Artificial Intelligence (AI) in agrivoltaic systems as a critical lens for understanding the socio-territorial implications of digital innovation in rural landscapes. Agrivoltaics, which combine agricultural production and solar energy generation on the same land, have emerged as a promising solution to the crises of energy transition and food security. However, their implementation raises significant questions about land use, community participation, and environmental justice. Through a series of case studies – including the European SYMBIOSYST project and two geospatial monitoring experiments using GIS and UAV technologies – the study investigates how AI-based tools such as predictive modelling, 3D digital twins, and sensor-integrated agritech platforms are reshaping the planning, governance, and visual representation of rural energy landscapes. While AI enhances the operational efficiency of agrivoltaic infrastructures and supports site-specific optimisation, it also introduces new spatial asymmetries, risks of algorithmic land grabbing, and challenges to data sovereignty. The analysis draws on critical geography to argue that AI should not be viewed merely as a technical enhancer but as a socio-technical actor that reconfigures power relations and decision-making in energy transitions. The cases discussed illustrate both the risks of centralised and extractive AI models and the opportunities for participatory and place-sensitive technological governance. The paper advocates for a geography-informed framework of AI adoption in the energy sector that promotes equity, transparency, and territorial justice.

Keywords: agrivoltaic, Artificial intelligence, geography of power, socio-territorial implications, AI governance.

1. INTRODUCTION

The energy sector is forecasting a substantial rise in global consumption. In fact, it is estimated that by 2050, energy demand could exceed the current level by 70% (Enerdata 2023; IEA, 2024). As a result, while it is difficult to accurately predict the timing and the amounts of greenhouse gas emissions released into the atmosphere in the future, a transition from fossil fuels to zero-carbon impact sources can no longer be delayed (BloombergNEF, 2025).

Among the various sustainable and carbon-neutral energy sources, solar power stands out as the most relevant, not only due to its ease of use but also due to its huge energy potential. Forecasts indicate that just one hour of sunlight delivers more energy than what is consumed by all human activities over a year (US Department of Energy, n.d.). In this context, the agrivoltaic system represents a noteworthy innovation by integrating agriculture with clean energy production on the same land. However, the implementation of agrivoltaic systems has generated contrasting opinions (Weselek et al., 2019). While it offers potential benefits for agriculture and the fight against climate change, some experts warn about its impacts on biodiversity and on potential microclimatic changes due to the use of solar panels, as well the risk of reducing vital agricultural space. These challenges could be mitigated through advanced digitalization of the agro-energy sector, particularly by leveraging Artificial Intelligence (AI), which can enhance operational efficiency and environmental monitoring while also transforming land planning.

Through the lens of these considerations, the first section of the paper aims to shed light on agrivoltaics, a topic that has sparked diverse opinions within the scientific community.

The following paragraph will delve into the opportunities and challenges associated with the application of AI in the management of agrivoltaic systems, seeking to ascertain whether advanced technologies can foster a synergy between economic growth and environmental and social sustainability. Finally, two experimental methodologies, along with a project analysis case, will be introduced to demonstrate how AI can enhance agricultural productivity and renewable energy generation. By means of the analysis of these results, valuable insights are expected to emerge on how to tackle both current and future challenges in the sector, encouraging a development model that merges energy efficiency with social responsibility.

2. AGRIVOLTAICS: AI AS FACILITATOR OF DIVERGENT OPINIONS

Agrivoltaics, developed in the 1980s by Adolf Goetzberger and Armin Zastrow (1981), represents an innovative approach that consists of elevating solar panels above crops, strategically spaced to minimize the land used for agricultural activities, unlike traditional ground-mounted installations. After some small-scale installations in Germany (Trommsdorff et al., 2021) and Japan (Sekiyama and Nagashima, 2019), the first significant agrivoltaic system was tested in France (Dinesh and Pierce, 2016), followed by larger projects in Germany and Italy (Agostini et al., 2021). However, it is in China that agrivoltaic has seen considerable development since 2010 (Hu, 2023).

Scholars hold radically opposing views on the impacts of agrivoltaics. Some argue that the introduction of this system diversifies income sources since the energy produced can be marketed, providing farmers with greater earning opportunities and increasing land productivity. This assertion is supported by a metric developed by agronomists known as the Land Equivalent Ratio (LER), used to assess soil productivity in conjunction with energy production (Toledo et al., 2021). Simulations suggest that integrating agriculture and energy production could boost overall land productivity by up to 70% (Dupraz et al., 2011; Amaducci et al., 2018; Dupraz, 2024). Furthermore, some experts agree that agrivoltaics plays a crucial role in combating the effects of the climate crisis by optimizing land and water resource use, reducing land degradation, and ensuring sustainability in terms of environmental protection and biodiversity (Colantoni et al., 2021; Trommsdorff et al., 2022; Schweiger & Pataczek, 2023; Taylor et al., 2025).

Skeptical scholars, rather, highlight potential negative effects associated with agrivoltaics, particularly how this practice may influence various microclimatic factors, such as air temperature, shading, and humidity. For instance, they point to the increase in thermal levels beneath solar panels on hot and poorly ventilated days, which can severely impact crop yields and the nutritional quality of agricultural products (Magarelli, Mazzeo, and Ferrara, 2025). Furthermore, critics of agrivoltaic emphasize its adverse effects on the landscape, including visual impact, loss of traditions, and destabilization of farming communities – risks that have sparked opposition movements within civil society and environmental organizations.

The contrasting opinions of scholars, which span from emphasizing the positive to the drawbacks of integrating agricultural crops and energy production on the same land, may find a common ground in the advanced digitalization of the agro-energy sector. This evolution proves how the adverse effects of agrivoltaics can be mitigated through AI and machine learning models. These technologies are becoming essential tools for analyzing complex data, enabling a comprehensive evaluation by assessing environmental, agronomic, and infrastructural variables.

The integration of AI with IoT technologies enhances the efficiency of agrivoltaic systems by allowing for real-time monitoring of critical parameters such as humidity and temperature. This, in turn, improves resource management and optimizes both agricultural and energy yields (Morkūnas et al., 2023; Issa et al., 2024; Assimakopoulos et al., 2025). Numerous case studies, including one from Maharashtra in India, illustrate AI's potential to boost agricultural productivity and enhance water resource management. Notably, this study reported a 30% reduction in water consumption, a significant achievement given the growing issue of water scarcity in India (Medonna and Ghosh, 2025). Other initiatives in Indonesia, where chili cultivation occurs in an off-grid area, highlight how technology can enable remote management through mobile devices. Projects in Kenya and Niger show the application of AI in analyzing agronomic data and planning crop practices, reducing the use of chemicals and enhancing food security. These are examples of successful applications of AI capabilities that have allowed the agrivoltaic system to adapt to a variety of climates and geographic contexts, from arid regions to temperate and tropical areas (Satria & Rosnelly, 2025), offering innovative solutions to better address the challenges of food security for a continuously growing global population and the fight against climate change.

This framework characterized by challenges and expectations, requires tools that can incorporate territorial, environmental, and social variables. This is where algorithmic geography becomes an essential analytical tool for decision-making.

3. THE ALGORITHMIC GEOGRAPHY OF AGRIVOLTAICS: TWO METHODOLOGICAL EXPERIMENTS

In this study, GIS and UAV case studies are methodological experiments, purposely selected to demonstrate site-specific optimisation and national-scale spatial effects, serving an illustrative – rather than statistically generalisable – role in analysing AI-territory coproduction.

In the current discourse debate surrounding agrivoltaics and its territorial implications, it is valuable to focus on examples of methodological experiments that, although not tied to a specific local context, clearly reveal the potential and challenges posed by the digital infrastructure of rural landscapes. Two recent studies stand out as exemplary cases of how geospatial technologies and advanced monitoring systems can redefine the design, management, and oversight of coexistence between agricultural production and energy generation.

The first significant contribution is a national-scale analysis that employed a Geographic Information System (GIS) approach based on multiple criteria to assess agrivoltaic potential in Italy. This study developed a spatial model capable of integrating various layers of information – such as land use, solar availability, regulatory constraints, and landscape sensitivity – to identify the most geographically suitable areas for the installation of agrivoltaic systems (Fattorusso et al., 2024).The methodology not only quantifies the extent of potentially installable areas but also highlights the deep territorial asymmetries that emerge when overlapping energy priorities, agricultural logics, and environmental protection demands. What makes this approach particularly significant is its ability to cartographically and reveals through modelling the political and selective nature of the energy transition. The maps generated through these tools do not simply represent neutral data about the territory; they reflect decision-making hierarchies, constraints imposed by existing regulations, and the infrastructural choices of public and private entities.

In essence, the geography of agrivoltaics is a constructed reality rather than a mere observation shaped by computational choices and algorithmic filters that determine what is visible and what remains outside the realm of possibilities (Fattorusso et al., 2024). This methodology aligns with the growing affirmation of territorial intelligence as a key component of spatial governance, emphasizing the integrated use of data and technologies to guide environmental transformations consistent with systemic goals (Lu, 2024).

A second complementary example involves an experiment utilizing UAV (Unmanned Aerial Vehicle) surveys and three-dimensional photogrammetry, to monitor the interaction between the shading effect of photovoltaic panels and crop growth (Cuaran & Leon, 2021). The study demonstrated that by leveraging high-resolution three-dimensional models, it is possible to optimize the placement of modules and assess the microclimatic impact of the installations on the agricultural ecosystem in real time. The data collected from this type of monitoring also helps inform dynamic predictive models that can adapt to the specific climatic and agronomic conditions of the site, transforming the agrivoltaic installation into an adaptive system rather than a static one.

This experimentation prompts important reflections on the role of algorithmic monitoring in rural areas. By employing drones, sensors, photogrammetric models, and digital simulations as everyday management tools for agricultural land, we are witnessing a radical redefinition of the relationship between agriculture, landscape, and control. On one hand, these tools enhance efficiency, reduce uncertainty, and enable highly specialized management of crops and energy yields. On the other hand, they raise significant questions regarding data sovereignty, ownership of the generated information, and the decision-making asymmetries between large technological operators and local communities (Jobin et al., 2019; Reasoner & Ghosh, 2022).

Together, these two practical experiences confirm that the synergy between agrivoltaics and AI should not be seen merely as a process of innovation; rather it must be understood in the context of geopolitical and territorial transformation. In this framework technologies such as GIS, AI, UAVs, predictive models – function as essential tools for spatial selection, priority definition, and resource reorganization. Consequently, the agricultural landscape becomes a subject of coding, interpretation, and transformation guided by computational logics. Thus, the challenge extends beyond technical aspects; it holds deep political implications. It revolves around the potential to direct these tools towards fostering territorial justice, transparency, and participation, rather than perpetuate centralized and opaque models of digital value extraction (Moser, 2023; TRUST-PV, 2024).

3.1. The SYMBIOSYST project: the potential of AI

In a context marked by the previously mentioned challenges, AI can be effectively applied by taking into account the unique characteristics of the territory and actively engaging administrations and local communities this approach is exemplified by the SYMBIOSYST project (SYMBIOSYST of Photovoltaic Systems with Agriculture for Smart Land Use). Funded by the Horizon 2020 program with a total budget of €5 million, the project is coordinated by the Eurac Research center in Bolzano, in collaboration with ENEA, the Laimburg Experimental Center, and various stakeholders from the agricultural and industrial sectors. Its goal is to develop advanced technological solutions to enhance the sustainable competitiveness of agrivoltaic systems in Europe through the adoption of artificial intelligence techniques and site-specific predictive models.

The project's methodological foundation relies on multi-objective models that integrate agricultural and energy production simultaneously. In particular, the modeling of crop yields and photovoltaic output utilizes deep learning algorithms, data imputation to manage incomplete climatic data, and three-dimensional digital twins of landscape systems. In this context, AI serves not only as a tool for operational optimization but also as a mechanism for territorial configuration, influencing the design, monitoring, and governance of agrivoltaic systems (Reasoner & Ghosh, 2022; Moser, 2023).

The methodological basis of the project is the use of multi-objective models capable of simultaneously integrating agricultural and energy production. In particular, the modeling of crop yields and photovoltaic output employs deep learning algorithms, data imputation for managing incomplete climatic data, and three-dimensional digital twins of landscape systems. In such a case, AI not only serves as a tool for operational optimization but also acts as a territorial configuration device, influencing the design, monitoring, and governance of agrivoltaic systems (Reasoner & Ghosh, 2022; Moser, 2023).

A particularly noteworthy aspect of the project is the development of a decision support system that integrates predictive algorithms and agronomic indicators into complex decision-making processes. As demonstrated in the related experience of TRUST-PV, these tools can effectively guide energy planning through innovative metrics that identify environmental challenges and optimize interventions in real time. This enhances the alignment between digital inputs and the ecological characteristics of territories (Moser, 2023). Furthermore, the SYMBIOSYST project utilizes drones to create 3D models, allowing for construction of a digital twin. This twin is instrumental not only during the design phase but also for continuous monitoring of the site's energy and agricultural performance.

What makes SYMBIOSYST particularly relevant to discussion in geography is its ability to blend technical innovation with sensitivity to landscape. This combination fosters reflective dialogue on the necessity for transparent and participatory algorithmic governance.

These operational outcomes demonstrate that the success of this transition relies heavily on how we incorporate digital tools into territorial governance, a theme that will be explored in the conclusions.

4. CONCLUSION

The findings of the study suggest that the synergy between agrivoltaics and artificial intelligence represents not just a technological advancement, but a radical change in geopolitical and territorial relations. Technologies, as tools for spatial organization and resource management, possess the power to transform rural landscapes by reshaping usage dynamics and foster enhancement. However, this evolution comes with serious risks. The integration of AI in agrivoltaics could exacerbate existing inequality and a concentration of power, as platforms and algorithms are often controlled by external actors who may overlook local specificities. There is a growing concern that these technologies might evolve into instruments of algorithmic extraction, draining value from territories and jeopardizing their integrity.

Digital platforms and predictive algorithms are frequently developed by industrial and technological actors external to the local context, who control data collection, analysis infrastructures and decision-making framework (Lu, 2024; Jobin et al., 2019). As a result, AI-driven agrivoltaic technologies risk becoming mechanisms for extracting data, resources and value from local communities, channeling them into centralized and opaque valorization circuits.

A responsible evaluation of AI adoption in the agrivoltaic sector must consider not only technical and environmental benefits, but also the ethical, geographical, and political implications (Dahlin, 2021). It is crucial to investigate who defines the optimization criteria for algorithms, who controls territorial analysis platforms, and how the advantages of digital transition are allocated. The transformation of agricultural landscapes into "algorithmic landscapes" raises important questions about protecting place identity, preserving traditional agricultural knowledge, and legitimation of automated decision-making processes (Scognamiglio, 2016). These insights can shed light on the potential impacts of these changes on local communities and their resources. By identifying the stakeholders involved, we can more effectively address issues of equity, representation, and the influence of external forces shaping agricultural practices and outcomes.

AI governance that is sensitive to local contexts should rest on three fundamental pillars: (1) transparency and flexibility in the algorithms used for the design and management of agrivoltaic systems; (2) active involvement of local communities in the modeling and application of territorial data; (3) spatial contextualization of technological solutions to ensure alignment with the landscape and socio-cultural characteristics of the area in question.

Ultimately, the integration of AI and agrivoltaics underscores the importance of a critical geographical perspective in approaching the digital transition. This perspective involves more than just determining optimal locations for system installation; it requires an examination of how technology reshapes the interactions among landscape, production processes and power dynamics. When governed responsibly, AI can serve as a powerful ally in enhancing rural resilience, promoting environmental justice, and ensuring that the energy transition does not compromise territorial equity.

In summary, agrivoltaics should be viewed solely as a technological solution for energy and food production but rather as a territorial infrastructure that transforms land use, value chains, and power relationships. The adoption of AI in this domain is anything but neutral; it significantly affects decision-making, localization, and the distribution of costs and benefits. Therefore, achieving a "just transition" requires clear criteria for spatial equity and tools to assess distributive effects both at local and regional scales.

From an algorithmic standpoint, the shift from "technical optimization" to algorithmic governance entails three key elements: (1) transparency regarding models and data, which include through documentation of sources, quality, coverage, and limitations; (2) explainability of decisions, ensuring that justifications are provided for both selected and excluded locations; and (3) multi-actor accountability, which entails independent audits, version tracking, public registries of models. In the absence of these proper safeguards, AI risk can reinforce existing biases and exacerbate land-use conflicts.

From an institutional perspective, the evidence points out several key needs: (1) the establishment of a polycentric and place-based governance model; (2) collaboration with farmers and local communities; (3) clear mechanisms for benefit-sharing which may include royalties, fee reductions, environmental and social funds; (4) provisions that ensure the continuity of agricultural production and (5) public monitoring tools to assess yields, biodiversity, and landscape impacts both before and after implementation.

In this framework, the Quadruple Helix model plays a vital role in harmonizing the interests of public, private, scientific, and civic sectors. This approach shifts the focus from "downstream" acceptability of projects to "upstream" legitimization, ensuring that all stakeholders are actively engaged in the decision-making process (Carayannis & Campbell, 2009; 2012).

Operationally, this necessitates algorithmic transparency and accountability through the use of model and data sheets, error metrics, versioning, and periodic audits. It also requires clear criteria for spatial equity, including minimum thresholds for local returns and mapping of land-use trade-offs. This should be complemented by integrated ex ante assessments that consider environmental, agronomic, and social factors incorporating meaningful public consultation. Additionally, standardized benefit-sharing mechanisms should be implemented, such as template contracts, community funds, and equity stakes in energy cooperatives. Public dashboards can provide near-real-time monitoring of production, yields, biodiversity, and social impacts (Jobin, Ienca & Vayena, 2024).

Additional, experimental designs remain essential for estimating causal impacts on yields and incomes. This requires multi-scale longitudinal datasets that combine remote sensing with administrative data, implementation of local calibration protocols for AI models, and conducting international comparisons of benefit-sharing mechanisms to evaluate their effectiveness in conflict reduction. All of these efforts should be underpinned by open standards and independent validations to address potential bias and ensure generalizability (Carayannis & Campbell, 2009; 2012; Jobin et al., 2024).

AI-enhanced agrivoltaics can serve as a catalyst for numerous opportunities, but their success hinges on the transparency and accountability of data and models used. Ultimately, decisions must be grounded in metrics of territorial equity, and the transition to sustainable practices can only be considered genuinely "just" if it delivers value back to the local communities.

In conclusion, the integration of AI within agrivoltaic systems signals not just a technological upgrade, but a structural reconfiguration of territorial governance and regulatory power. Approaches formalised under frameworks such as the EU AI Act demonstrate how regulation itself becomes a geopolitical instrument, enabling Europe to project normative influence whilst curbing asymmetric dependencies along global AI–energy value chains. At the same time, evidence from peripheral and off-grid regions in Global South illustrates that AI-driven agrivoltaics may either entrench extractive hierarchies of data and land, or – if shaped through participatory, place-aware governance – convert digital oversight into a territorial public-value asset. What is ultimately at stake is not optimisation alone, but who sets the criteria, who controls spatial analytics, and how benefits and externalities are territorially allocated. A genuinely just transition therefore rests on polycentric accountability, ex-ante impact assessment tied to local evidence, and equitable benefit-sharing mechanisms that retain decision authority with territorial communities, rather than remote platform monopolies. In this sense, geography must continue to serve as a critical regulatory compass, ensuring that AI and agrivoltaic co-deployment strengthens – rather than compromises – territorial justice, community resilience, and the democratic legitimacy of rural energy futures.

REFERENCES

Agostini, A., Colauzzi, M. & Amaducci, S. (2021). Innovative agrivoltaic systems to produce sustainable energy: an economic and environmental assessment. *Applied Energy*, 281, 116102.

Amaducci, S., Yin, X. & Colauzzi, M. (2018). Agrivoltaic systems to optimise land use for electric energy production. *Applied Energy* 220, 545–561.

Bloomberg, N.E.F. (2025). Energy transition investment trend. https://www.bnef.com/login?r=%2Fflagships%2Fclean-energy-investment.

Carayannis, E.G. & Campbell, D.F.J. (2009). 'Mode 3' and the Quadruple Helix: Toward a 21st-century fractal innovation ecosystem. *International Journal of Technology Management*, 46(3/4), 201–234.

Carayannis, E.G. & Campbell, D.F.J. (2012). *Mode 3 Knowledge Production in Quadruple Helix Innovation Systems: 21st-Century Democracy, Innovation, and Entrepreneurship for Development*. New York: Springer.

Cudlínová E., Giacomelli Sobrinho V., Lapka M. & Salvati L. (2020). New Forms of Land Grabbing due to the bioeconomy: The Case of Brazil. *Sustainability*, 12(8), 3395.

Colantoni, A., Ferrara, C. Perini, L. & Salvati, L. (2015). Assessing trends in climate aridity and vulnerability to soil degradation in Italy. *Ecological Indicators*, 48, 599–604.

Cuaran, J. & Leon, J. (2021). Crop Monitoring using Unmanned Aerial Vehicles: A Review. *Agricultural Reviews*, 42(2), 121–132.

Dahlin, E. (2021). Mind the gap! On the future of AI research. *Humanities and Social Sciences Communications*, 8(71), 1–4.

De Falco, M., Sarrica, M. Scognamiglio, A. & Fasanelli, R. (2025). What does Agrivoltaics mean? A study on social representations shared by experts and the press in Italy. *Energy*, 119, 103918.

Dinesh, H. & Pearce, J. (2016). The potential of agrivoltaic systems. *Renewable and Sustainable Energy Reviews*, 54, 299–308.

Dupraz, C. (2024). Assessment of the ground coverage ratio of agrivoltaic systems as a proxy for potential crop productivity. *Agroforestry Systems*, 98, 2679–2696.

Dupraz, C., Marrou, H., Talbot, G., Dufour, L., Nogier, A. & Ferard, Y. (2011). Combining solar photovoltaic panels and food crops for optimising land use: Towards new agrivoltaic schemes. *Renewable Energy*, 36(10), 2725–2732.

Enerdata (2023). Global energy statistics. https://yearbook.enerdata.net/

Energy Agency (2024). World energy outlook. https://iea.blob.core.windows.net/assets/140a0470-5b90-4922-a0e9-838b3ac6918c/WorldEnergyOutlook2024.pdf

Fattorusso, G., Toscano, A., Venturo, A., Scognamigli, A., Fabricino, M. & Di Francia, G. (2024). A Spatial Multicriteria Analysis for a Regional Assessment of

Eligible Areas for Sustainable Agrivoltaic Systems. *Sustainability*, 16(2), 911.

Ghosh, A. (2023). Nexus between agriculture and photovoltaics (agrivoltaics, agriphotovoltaics) for sustainable goal: a review. *Solar Energy*, 266, 112146.

Goetzberger, A. & Zastrow, A. (1982). On the coexistence of solar-energy conversion and plant cultivation, *International Journal of Solar Energy*, 1(1) , 55–69.

Hu, Z., (2023). Towards solar extractivism? A political ecology understanding of the solar energy and agriculture boom in rural China. *Energy Research & Social Science*, 98, 102988.

Jobin, A., Ienca, M. & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399.

Lazzeroni, M. & Romano, A. (2024). Towards a new urbanism: emerging narratives on the impact of artificial intelligence in cities. *Geoprogress Journal*, 11(2), 53–68.

Lu, C. (2024). Rethinking artificial intelligence from the perspective of interdisciplinary knowledge production. *AI & Society*, 3059–3060

Magarelli A, Mazzeo, A. & Ferrara, A. (2025). Exploring the grape agrivoltaic system: climate modulation and vine benefits in the Puglia Region, southeastern Italy. *Horticulturae*, 11(2), 11020160.

Medonna, A. & Ghosh, A. (2025), Harvesting Sun Amplified: A Comprehensive Review of Agrivoltaics and Floatovoltaics with a Focus on India's Potential. *Solar Compass*, 15, 1000132.

Moser, D. (2023). Tech4Future. https://tech4future.info/intelligenza-artificiale-agrivoltaico-fotovoltaico/

Müller, K. & Pampu, M. (2023). The solar rush: invisible land grabbing in East Germany. *International Journal of Sustainable Energy*, 42(1), 1264–1277.

Reasoner, M. & Ghosh, A. (2022). Agrivoltaic Engineering and Layout Optimization Approaches in the Transition to Renewable Energy Technologies: A Review. *Challenges*, 13(2), 43, 13020043.

Schweiger, A.H. & Pataczek, L. (2023). How to reconcile renewable energy and agricultural production in a drying world. *Plants People Planet*, 5(5), 10371.

Scognamiglio, A. (2016). Photovoltaic landscapes: Design and assessment. A critical review for a new transdisciplinary design vision. *Renewable and Sustainable Energy Reviews*, 55, 629–661.

Sekiyama, T. & Nagashima, A. (2019). Solar Sharing for Both Food and Clean Energy Production: Performance of Agrivoltaic Systems for Corn, A Typical Shade-Intolerant Crop. *Environments*, 6(6), 65, 6060065.

Taylor M., McDonnell, N. & Davies, P.J. (2025). Scaling agrivoltaics: planning, legal, and market pathways to readiness. *Sustainable Sciences*, 20, 1499–1517.

Toledo, C. & Scognamiglio, A. (2021). Agrivoltaic Systems Design and Assessment: A Critical Review, and a Descriptive Model towards a Sustainable Landscape vision. *Sustainability*, 13, 6871.

Trommsdorff, M., Kang Reise, C., Schindele, S., Bopp G., A., Weselek A., Högy, P. & Obergfell, T. (2021). Combining food and energy production: Design of an agrivoltaic system applied in arable and vegetable farming in Germany. *Renewable and Sustainable Energy Reviews*, 140, 110694.

Treccarichi C. (2025). Chi c'è dietro le migliaia di pannelli che stanno coprendo la campagna italiana. https://www.today.it/economia/agrivoltaico-parchi-italia-chi-guadagna-mappa-inchiesta.html

TRUST-PV (2024). Data-Driven Photovoltaic Energy Planning and Predictive Maintenance Models, Horizon 2020 project. https://cordis.europa.eu/project/id/953000

US Department of Energy (s.d). The history of photovoltaic. https://www1.eere.energy.gov

Weselek, A., Ehmann, A., Zikeli, S., Lewandowski, I., Stephan Schindele, S. & Petra Högy, P. (2019) Agrophotovoltaic systems: applications, challenges,and opportunities. A review (2019). *Agronomy for Sustainable Development*, 39(35).

JEP

Firenze University Press
https://oaj.fupress.net/index.php/jep

Citation: Chiaudano, V., & Shakil, H. (2025) Understanding the role of artificial intelligence in luxury brand management. *Journal of Emerging Perspectives* 2: 33-44. doi: 10.36253/jep-20880

Received: April 30, 2025

Revised: September 30, 2025

Published: December 20, 2025

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

ORCID
VC: 0000-0003-3249-698X
HS: 0009-0000-1178-0653

Understanding the role of artificial intelligence in luxury brand management

Valentina Chiaudano*, Hafsa Shakil

Department of Management, University of Turin, Italy
Email: valentina.chiaudano@unito.it; hafsashakil0@gmail.com
*Corresponding author

Abstract. This study conducts a systematic literature review to examine how artificial intelligence (AI) is reshaping luxury brand management. It analyses 79 peer-reviewed articles identified through a structured search in Scopus, combining bibliometric techniques with qualitative content analysis. Findings show a rapid increase in publications after 2020, driven by luxury firms' accelerated digital transformation and adoption of data-driven tools. Four main themes structure the current research: how AI shapes consumer behaviour and human–machine interaction; the role of intelligent systems in luxury retail and distribution; the use of sentiment analysis and automation in hospitality; and the impact of big data and generative AI on branding. Across these areas, AI emerges as a new source of value creation, influencing perceptions of authenticity, exclusivity, and emotional engagement. The review introduces a hybrid-intelligence perspective, highlighting how human and algorithmic capabilities jointly shape luxury experiences, and outlines key managerial implications and future research needs.

Keywords: Artificial Intelligence, AI, luxury brand management, literature review, digital transformation.

1. INTRODUCTION

Historically, the luxury experience has been anchored in physical environments such as flagship stores, carefully curated interiors, and selective distribution channels, where value is conveyed through sensory richness, artisanal excellence, and highly personalised, human-centred interactions (Kapferer & Bastien, 2025; Kapferer & Bastien, 2013). Within this context, early digital technologies appeared misaligned with the logic of luxury: the openness, mass accessibility, and standardised nature of the internet seemed to undermine the sector's symbolic foundations and its reliance on controlled, exclusive access (Chandon et al., 2016; Mosca & Civera, 2017).

As digital capabilities expanded, however, luxury brands gradually adopted integrated online–offline strategies, developing e-commerce platforms, social media ecosystems, and omnichannel experiences that extended the brand narrative without compromising exclusivity (Kim et al., 2015). The transition toward digitalisation accelerated further during the COVID-19 pandemic, which forced luxury companies to rethink their operational and communication models. With store closures and the halt of international travel, digital and hybrid channels became essential to maintaining consumer engagement (Sanz-Lopez et al., 2024). In response, leading luxury brands intensified their investments in e-commerce, mobile applications, social media communities, and virtual experiences, incorporating these channels into every stage of the customer journey (Achille & Zipser, 2020; Xie & Youn, 2020).

Building upon this digital shift, the luxury industry has recently entered a new stage in which artificial intelligence (AI) plays a central role in redefining value creation, creativity, and consumer experience. AI-based systems allow brands to collect and interpret vast quantities of data, predict consumer behavior, and personalize interactions at an unprecedented scale (Han & Xu, 2025; Xu & Mehta, 2022; Huang & Rust, 2021). From predictive analytics to generative algorithms, AI introduces new forms of automation and co-creation that reshape both operational efficiency and how organisations generate meaning and value (Davenport & Ronanki, 2018; Shrestha et al., 2021; Brynjolfsson & Mitchell, 2017; Dwivedi et al., 2023). In luxury settings, machine learning, natural language processing, and computer vision increasingly support design, retail, communication, and service delivery, transforming how firms conceptualise exclusivity and emotional resonance (Essiz, 2023; Joy et al., 2022; Oc et al., 2023; Minor et al., 2025). In this rapidly evolving landscape, luxury once again faces the need to safeguard the uniqueness of its experience, as the very notion of authenticity is being reshaped. Rather than resisting technology, contemporary luxury must ensure that the integration of algorithmic intelligence remains fully aligned with the brand's cultural, aesthetic, and artisanal heritage, thereby preserving, rather than weakening, the symbolic depth of the luxury experience (Xu & Mehta, 2022; Pantano et al., 2024; Joy et al., 2022). Thus, luxury brands stand at a critical juncture where technological innovation intersects with the cultural, symbolic, and experiential foundations of luxury (Moisello & Pellicelli, 2025; Standish et al., 2024). This evolution requires managers to integrate advanced digital and AI-driven tools while safeguarding the symbolic depth and experiential exclusivity that remain central to luxury's identity (Cenizo, 2025).

Despite the growing managerial relevance of AI in luxury brand management, the academic literature remains fragmented, conceptually dispersed, and methodologically uneven (Moisello & Pellicelli, 2025; Thanasi-Boçe & Al-Issa, 2025; Pantano et al., 2024). Existing studies often focus on specific domains without offering an integrated understanding of how AI collectively redefines the logic of luxury. Moreover, the literature lacks a comprehensive synthesis of theoretical perspectives, leaving unresolved questions about the mechanisms through which intelligent systems influence symbolic value, consumer experience, and managerial practice (Liu & Zhao, 2025; Thanasi-Boçe & Al-Issa, 2025; Pantano et al., 2024).

To address this gap, this study conducts a systematic literature review of peer-reviewed articles on artificial intelligence and luxury brand management. Following the methodological guidelines of Massaro et al. (2016), it integrates bibliometric mapping with qualitative content analysis to identify key conceptual domains, trends, and research trajectories. Specifically, the review seeks to answer three interrelated questions:

(1) What are the main impacts of artificial intelligence on luxury brand management?

(2) How has academic literature approached AI within luxury contexts?

(3) In which areas of luxury management is AI most frequently applied, and what theoretical and managerial insights do these applications provide?

By addressing these questions, the study maps the evolution, distribution, and disciplinary orientation of research on AI in luxury, identifying the journals, authors, and countries driving this field. Theoretically, it organises the fragmented knowledge base into four coherent research clusters that reflect how AI reshapes consumption behaviour, retailing and distribution, hospitality, and branding.

2. METHODOLOGY

Building on the theoretical foundations discussed above, this paper aims to deepen the understanding of how artificial intelligence influences luxury brand management by conducting a review that integrates bibliometric and content analyses. According to Massaro et al. (2016), combining quantitative and qualitative approaches enhances transparency, internal consistency, and replicability. This mixed-method design is particularly suitable for emerging research domains, such as the intersection of AI and luxury brand management, where existing studies are often fragmented across disciplines and lack a coherent conceptual framework (Massaro et al., 2016; Secinaro et al., 2021; Secundo et al., 2021).

The review process was structured as follows. First, a comprehensive search was conducted in the Scopus database using specific search terms applied to titles, abstracts, and keywords. The initial search returned 780 studies published in English up to October 31, 2025, when the research on Scopus was conducted. The search query was defined as follows:

(TITLE-ABS-KEY (luxury) AND TITLE-ABS-KEY (“artificial intelligence” OR “AI” OR “generative AI” OR “large language model*” OR LLM* OR “machine learning” OR ML OR “deep learning” OR “neural network*” OR “natural language process*” OR NLP OR “computer vision” OR “expert system*” OR “knowledge-based system*” OR “intelligent agent*” OR “fuzzy logic” OR “reinforcement learning” OR “predictive analytics” OR “data mining” OR “decision support” OR “recommender system*” OR chatbot* OR “conversational agent*” OR robot* OR “service robot*” OR “autonomous system*”)).

To ensure conceptual alignment with the research objective, only documents indexed under Business, Management, and Accounting were retained, resulting in a refined sample of 183 studies. Additionally, only English-language publications were considered to guarantee analytical consistency (Vrontis & Christofi, 2021). The review focused exclusively on peer-reviewed journal articles, which are generally recognized for their methodological rigor and theoretical robustness (Kraus et al., 2020), yielding a total of 109 papers.

Subsequently, titles, abstracts, and, when necessary, full texts were screened to verify their relevance to the applications or implications of AI within the luxury industry. Studies addressing general digitalisation, analytics, or non-luxury contexts without a clear reference to artificial intelligence were excluded to maintain thematic coherence. This process resulted in a final sample of 79 articles representing the current state of the art in research on AI in the luxury sector.

Each article was systematically recorded in a structured dataset exported from Scopus, including comprehensive bibliographic, citation, and abstract information. The subsequent bibliometric and content analyses were conducted using the Bibliometrix package in R (Aria & Cuccurullo, 2017), which enabled the mapping of annual scientific production, leading journals and documents, country contributions, keyword frequencies, and co-occurrence networks. The co-occurrence analysis revealed four main research clusters, representing the core thematic areas of scholarship on AI and luxury. In the second phase, a directed content analysis was conducted to refine and validate the bibliometric findings. Each paper was carefully reviewed to identify its main thematic contributions, which were then used to populate the four clusters that emerged from the co-occurrence analysis. All qualitative coding activities were logged in a shared Excel dataset accessible to all members of the research team and an independent auditor for cross-checking accuracy (Creswell 2003).

3. RESULTS

3.1. Descriptive analysis

The analysis of academic publications exploring the intersection between AI and luxury brand management reveals a relatively young but rapidly expanding field of research (Figure 1). Throughout the early 2000s, scholarly attention to the topic was limited, with only a few isolated studies addressing digital innovation within the luxury sector. The growing adoption of machine learning, big data analytics, and conversational technologies by luxury companies around 2016 marked a decisive turning point, as researchers began to investigate AI’s transformative potential. From 2020 onwards, the number of studies increased profoundly, reflecting the post-pandemic acceleration of digital transformation across luxury industries. Between 2020 and 2023, scholarly production expanded significantly, increasing from a few initial contributions to over twenty publications per year, and culminating in 2025 with the identification of seventy-nine peer-reviewed journal articles focusing on AI within the luxury sector. This pronounced growth reflects the consolidation of AI as a distinctive and emerging research domain in luxury studies, while underscoring the increasing academic engagement with its implications for brand strategy and value creation.

Table 1 shows the distribution of academic publications on AI in the field of luxury brand management. The data confirm that this research area is still fragmented, with publications dispersed across a limited number of specialized sources. The Journal of Retailing and Consumer Services clearly dominates the field with 13 articles, indicating that the consumer-facing applications of AI, particularly in luxury retail, have received the most attention. The International Journal of Hospitality Management (5 articles) and the International

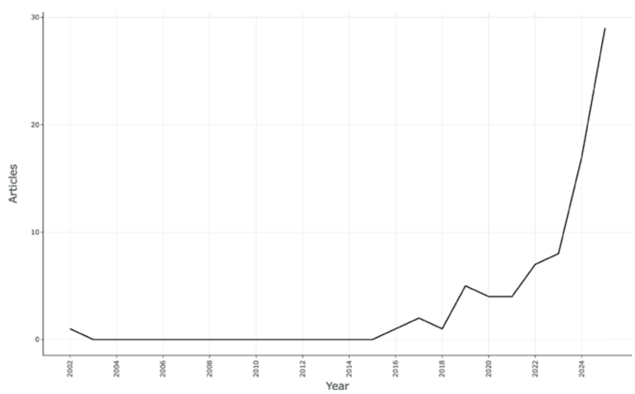

Figure 1. Annual scientific production.

Journal of Contemporary Hospitality Management (4 articles) follow, reinforcing the importance of hospitality and service innovation within luxury research, especially where AI enhances personalization and customer experience. Journals like the Journal of Business Research and the Journal of Hospitality and Tourism Technology (3 articles each) show a balanced integration of managerial and technological approaches, bridging AI's operational and strategic implications. Finally, outlets such as Global Business Review, Strategic Change, and Tourism and Hospitality Research (2 each) demonstrate an emerging interdisciplinary interest, connecting AI adoption with organizational transformation and tourism-related luxury experiences. The large "Others" category (43 articles) further underscores that this is an emerging and diversifying research domain, still seeking consolidation within mainstream academic outlets.

Table 2, related to the geographical distribution of publications, shows that research on AI in the luxury sector is concentrated in a few leading countries, with growing global participation. The United States (28), China (23), and South Korea (21) dominate the field, reflecting their leadership in technological innovation and AI development. In Europe, the United Kingdom (13), Italy (7), and France (5) stand out, consistent with their historical strength in luxury branding. Emerging contributions from countries such as India, Egypt, and several Asian and Middle Eastern nations indicate a gradual diversification of the research landscape. Overall, the data suggest that while AI–luxury studies are still concentrated in technologically advanced and economically mature markets, the topic is gaining traction worldwide, signalling an increasingly international and interdisciplinary field

Considering the methodologies adopted by studies on AI and luxury management (Figure 2), most of the analysed papers employ a quantitative approach (49 papers), relying on surveys, experiments, and statistical models such as SEM, PLS-SEM, or regression analysis to investigate perceptions, adoption intentions, and the impacts of AI in the luxury and hospitality sectors. A smaller number of studies adopt qualitative methods (18), based on semi-structured interviews and thematic or content analysis, with exploratory or theory-building aims to understand the psychological and symbolic dynamics of human–AI interaction. Only two papers use mixed methods, combining qualitative and quantitative phases to triangulate findings, while ten contributions are theoretical or conceptual, focusing on the development of interpretative frameworks and reflections on the role of AI and emerging technologies in reshaping business models and customer experiences.

Table 1. Most relevant sources.

Sources	Articles
Journal Of Retailing And Consumer Services	13
International Journal Of Hospitality Management	5
International Journal Of Contemporary Hospitality Management	4
Journal Of Business Research	3
Journal Of Hospitality And Tourism Technology	3
Global Business Review	2
Journal Of Hospitality And Tourism Insights	2
Strategic Change	2
Tourism And Hospitality Research	2
Others	43

Table 2. Country scientific production.

Country	Frequency
Usa	28
China	23
South Korea	21
Uk	13
India	13
Australia, Italy	7
Egypt, France	5
Canada, Jordan, Portugal, Spain, Thailand	4
Malaysia, Singapore	3
Bangladesh, Croatia, Finland, Greece, Japan, Oman, Serbia	2
Austria, Belgium, Cyprus, Germany, Indonesia, Iran, Ireland, Mexico, Monaco, New Zealand, Pakistan, Qatar, Slovenia, South Africa, Sweden, United Arab Emirates, Uzbekistan	1

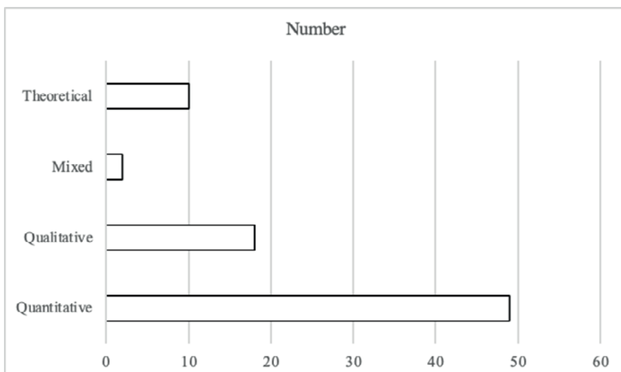

Figure 2. Methodology.

Figure 3. Word cloud.

The word cloud (Figure 3) synthesizes the main conceptual domains emerging from the intersection between artificial intelligence and the luxury industry. The central prominence of the term "artificial intelligence" denotes its function as a connective construct linking technological, managerial, and behavioral dimensions of inquiry. Closely related expressions such as "consumption behaviour," "consumer behaviour," and "retailing" reflect a growing scholarly interest in the transformative effects of AI on consumption patterns, decision-making processes, and value creation within luxury ecosystems. The recurrence of terms such as 'luxury hotels,' 'luxury brands,' and 'luxury retail' positions the discussion within hospitality and retail contexts, where AI is predominantly examined as an enabling factor for personalisation, experiential innovation, and the co-creation of symbolic and functional value.

From a technological standpoint, the presence of terms such as "machine learning," "big data," and "generative AI" indicates that contemporary research emphasizes the algorithmic and data-driven infrastructures underpinning AI applications in luxury contexts. Simultaneously, the occurrence of "social media," "sentiment analysis," and "communication" suggests a parallel focus on the discursive and affective dimensions of AI-mediated brand–consumer relationships, in which digital interaction and emotion analytics become central to the construction of luxury meaning. References to "service robots," "technology adoption," and "customer satisfaction" highlight a further line of inquiry concerned with operational efficiency, human–machine interaction, and the experiential redefinition of service quality.

Overall, the word cloud reveals an emerging, interdisciplinary research stream in which technological innovation converges with behavioral and managerial analysis. The visualization illustrates how AI is progressively reconfiguring luxury consumption and production logics, fostering hybrid forms of intelligence that integrate human creativity with algorithmic agency, and inaugurating new paradigms of value creation, personalization, and symbolic engagement within the luxury sector.

3.2. Co-occurrence network and content analysis

While the word cloud provides an overview of the most frequent terms, the keyword co-occurrence network offers a more structured understanding of how concepts interconnect within the literature. The network revealed four distinct thematic clusters, each representing a major conceptual and theoretical direction within the research on artificial intelligence and luxury brand management (Figure 4). Following the identification of these clusters, a directed content analysis was performed, as described in the methodology section, to validate and interpret their conceptual boundaries.

3.2.1. Artificial Intelligence, consumption behavior, and human-machine interaction

The first cluster explores how artificial intelligence shapes consumption behavior and human–machine

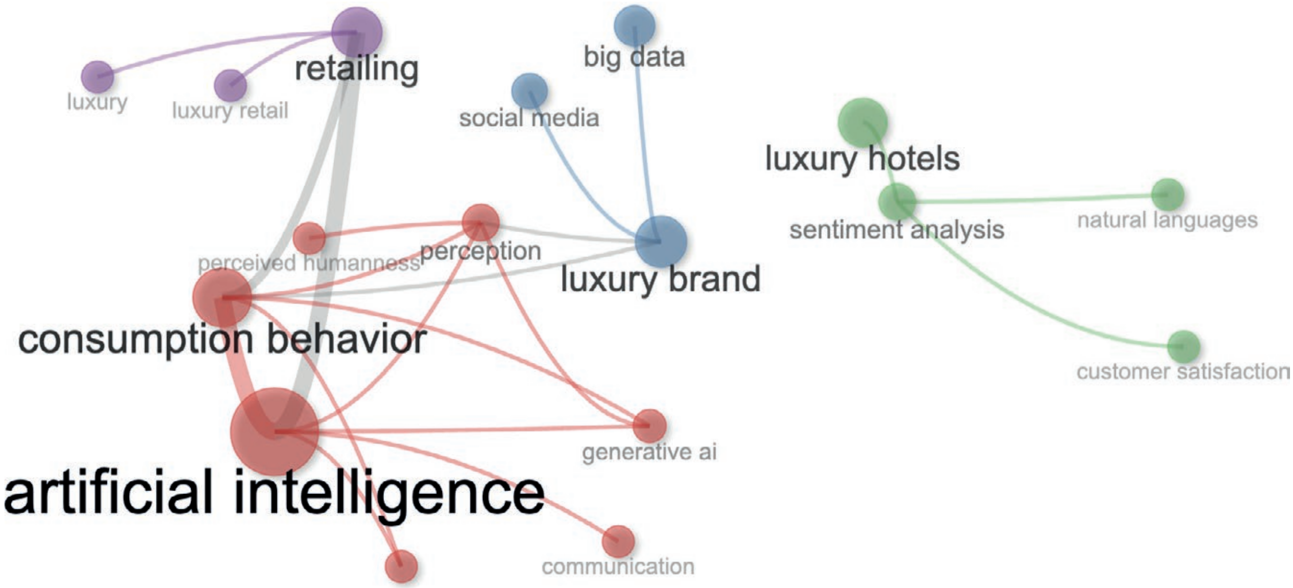

Figure 4. Co-occurrence network.

interaction in luxury contexts. Scholars increasingly conceptualize AI not merely as a functional tool but as an affective actor that mediates the relationship between consumers and luxury brands (Kim et al., 2015; Cho & Chen, 2025). Through conversational agents, recommendation systems, and generative AI applications, consumers encounter technologically enhanced forms of personalization that evoke both fascination and ambivalence. These interactions redefine traditional constructs such as trust, empathy, and authenticity, central to the symbolic experience of luxury.

Existing studies demonstrate that AI can enhance perceived exclusivity and satisfaction by offering hyper-personalized recommendations and emotionally adaptive responses (Li and Kang, 2025; Han & Xu, 2025). At the same time, research reveals that excessive automation may reduce the sense of human warmth and weaken the perceived craftsmanship that underpins luxury value (Suganthi & Jasim, 2024). The emotional dissonance generated by anthropomorphized AI, neither fully human nor purely mechanical, emerges as a defining challenge in luxury experience design. Furthermore, scholars note that generative and affective AI systems are beginning to participate actively in co-creating symbolic value, transforming consumers from passive recipients into co-authors of meaning (Minor et al., 2025). This participatory dynamic suggests that AI-human relationships in luxury may evolve toward hybrid models of emotional co-production.

3.2.2. AI, retailing, and luxury distribution

The second cluster focuses on the transformative effects of artificial intelligence on luxury retailing and distribution, where technological innovation increasingly intersects with spatial design, customer engagement, and brand heritage. Studies in this area show how AI-driven systems, from predictive analytics to computer vision, are redefining the physical and digital dimensions of luxury retail environments (Jung et al., 2025; Hoang et al., 2025). Intelligent platforms enhance inventory management, personalize product recommendations, and orchestrate seamless omni-channel journeys that merge in-store and digital touchpoints (Wu et al., 2021).

While these applications promise efficiency and convenience, scholars emphasize that luxury retailing must remain rooted in sensory and emotional experiences (Roy et al., 2020; Marinakou, 2024). Research suggests that the most successful implementations of AI occur when technological precision enhances, rather than substitutes, the aesthetic and relational qualities of the shopping experience (Essiz, 2023). For example, interactive mirrors, virtual stylists, and adaptive lighting systems powered by AI contribute to crafting personalized atmospheres that reinforce brand exclusivity. Yet, these innovations also raise new questions about the balance between automation and human connection – an equilibrium critical to preserving the essence of luxury service.

In parallel, AI is enabling new forms of phygital retailing, where physical stores act as experiential hubs supported by data-driven personalization (Joy et al., 2022). These hybrid formats transform stores into dynamic spaces for storytelling, co-creation, and cultural immersion, blending technological sophistication with the symbolic aura of luxury.

3.2.3. Sentiment analysis and customer experience in luxury hospitality

This cluster explores the integration of Artificial Intelligence in luxury hospitality, combining analytical, experiential, and operational perspectives. Within this domain, AI technologies such as sentiment analysis, natural language processing, and service robotics are employed to enhance the emotional, relational, and functional dimensions of guest experience. The literature demonstrates how data-driven intelligence and automation are jointly redefining the meaning of service excellence in high-end hospitality, transforming the traditional model of human-centered care into a hybrid system of human–machine collaboration (Li & Khan, 2023; Suganthi & Jasim, 2024).

From an analytical standpoint, scholars have widely investigated how intelligent systems monitor and interpret guest emotions, satisfaction, and perceptions of service quality. Through the application of machine learning, text mining, and natural language processing, vast volumes of user-generated content – such as online reviews, booking feedback, and social media posts – are analyzed to extract patterns of sentiment and identify the drivers of guest experience (Wu et al., 2021; Oc et al., 2023). These studies demonstrate that AI-based sentiment analysis provides hotel managers with actionable insights into customer expectations, allowing them to redesign service offerings and communication strategies in real time. The translation of emotional feedback into measurable data supports more adaptive forms of experience management and contributes to building long-term guest relationships (Marinakou et al., 2025; Roy et al., 2020).

Beyond retrospective analysis, AI enables a proactive approach to personalization. Predictive analytics and recommender systems can anticipate guest needs and tailor offers before arrival, optimizing resource allocation and strengthening perceived exclusivity (Khamoushi et al., 2024). In luxury contexts, where attention to detail and emotional resonance are central to value creation, these technologies help transform subjective impressions into operational intelligence. Scholars argue that such anticipatory systems reinforce service precision while maintaining a sense of intimacy, provided that data are managed ethically and transparently (Li & Khan, 2023).

Parallel to these analytical applications, a growing body of research examines the technological embodiment of AI through service robots and automated systems. These studies emphasize how robotics can improve efficiency, reduce errors, and support staff in delivering consistent high-quality service (Roy et al., 2020; Suganthi & Jasim, 2024). Robots are increasingly used for check-ins, concierge support, and amenity delivery, extending the scope of automation beyond back-office operations to the guest interface (Chan & Tung, 2019). However, this integration raises critical tensions between efficiency and emotional authenticity. As automation enters the front stage of luxury service, the risk of diminishing the human warmth and symbolic value associated with hospitality becomes more evident (Li & Khan, 2023; Khamoushi et al., 2024).

Recent studies respond to this tension by proposing the concept of hybrid hospitality, in which human employees and intelligent machines collaborate to create an emotionally balanced and operationally optimized experience (Hoang et al., 2025). Within this model, AI does not replace human interaction but augments it, allowing employees to focus on empathy, creativity, and personalization while machines handle repetitive or data-intensive tasks. Research on customer acceptance of robotic services further shows that satisfaction and loyalty depend on perceived humanness, trust, and the novelty of interaction (Cho & Chen, 2025; Kim et al., 2015).

From a managerial perspective, the integration of AI and robotics requires profound organizational adaptation. Scholars underline the importance of staff training, workflow redesign, and ethical governance to ensure smooth human–AI collaboration (Li & Khan, 2023). Implementing robotic systems also entails cultural and symbolic considerations, as luxury brands must preserve their narrative of refinement and care while embracing technological innovation. When effectively managed, AI and robotics can complement human service, creating a synergistic ecosystem where operational excellence coexists with emotional depth.

3.2.4. Luxury branding, social media, and big data

The fourth cluster connects luxury branding, social media, and big data, illustrating how artificial intelligence is reshaping communication, creativity, and value construction in the luxury sector. Within this domain, AI serves not only as an analytical instrument but also as a co-creator of symbolic meaning. Through machine learning and big data analytics, luxury brands can decode vast amounts of consumer information, identify emerging cultural patterns, and adapt their narratives in real time to align with evolving audience expectations (Oc et al., 2023; Minor et al., 2025). Predictive algorithms and natural language processing facilitate the optimization of content resonance across multiple touchpoints, allowing brands to orchestrate coherent and emotionally engaging storytelling (Jang et al., 2023; Wu et al., 2021).

Scholars highlight that this data-driven creativity enhances personalization, enabling brands to offer experiences that feel both exclusive and emotionally relevant (Essiz, 2023; Roy et al., 2020). Yet, a growing body of research warns that algorithmic mediation can distort symbolic depth: as automated engagement metrics privilege visibility and virality, they risk reducing luxury to a performative construct detached from its artisanal and cultural foundations (Cenizo, 2025; Joy et al., 2022). This tension underscores a central paradox of AI-driven branding – between technological efficiency and cultural authenticity. Recent studies also note that AI tools such as generative models and recommender systems are being employed to support creative direction, visual design, and storytelling coherence (Han & Xu, 2025; Marinakou, 2024). These technologies contribute to new forms of hybrid creativity, in which human intuition and algorithmic intelligence co-produce content.

Consequently, the role of creative professionals is evolving toward ethical mediation and curatorial oversight of AI outputs, requiring interdisciplinary expertise that blends technological awareness with artistic and cultural literacy. At the same time, the increased reliance on data-driven systems raises ethical and managerial questions regarding privacy, manipulation, and transparency in algorithmic branding (Joy et al., 2022; Suganthi & Jasim, 2024). Therefore, this cluster captures the strategic and communicative dimension of AI in luxury: a field where data intelligence and cultural meaning must be harmonized to sustain the emotional and symbolic value that defines luxury brands.

4. IMPLICATIONS

Our systematic review of 79 peer-reviewed articles identifies the conceptual domains, applications, and research directions, contributing to a young yet rapidly evolving research field at the intersection of artificial intelligence and luxury brand management. Our analysis highlights two distinct growth phases in the field of artificial intelligence in luxury brand management with limited scholarly attention through the early 2000s, followed by a decisive turning point around 2016, notably through increased industry adoption of machine learning and data analytics, and a pronounced acceleration from 2020 onwards, aligning with the acceleration of digital-transformation catalyzed by COVID-19 pandemic, yielding over twenty publications annually by 2023 and eighty peer-reviewed contributions by 2025. This expansion reflects growing managerial relevance and corresponds with the shifting needs of the luxury industry at the time. First, to maintain brand prestige and customer intimacy in a physically distanced world, like other industries, the luxury industry invested heavily in digital transformation. The post-pandemic strategic capital flow, both through government and private bodies, to support AI innovations, such as virtual try-ons, supply chain automation, and creative content generation. Together, these catalysts fostered an environment where both industry practice and academic scrutiny of AI in luxury could thrive. However, despite this growing interest, the literature remains characterized by geographic concentration (90% of publications from three countries: United States, China, and South Korea) and dispersal across specialized outlets within retail, hospitality, and management disciplines, with the Journal of Retailing and Consumer Services accounting for 16% of the corpus and the remaining 53% distributed across secondary and tertiary venues.

Theoretically, our review advances understanding of artificial intelligence within luxury brand management by reconceptualizing AI beyond instrumental functionality. The literature positions AI not merely as an enabler of operational efficiency and personalization but increasingly as a participant in processes of meaning creation and value co-production across four identifiable thematic clusters: consumption behavior and human–machine interaction, retailing and distribution, hospitality service experience, and social media branding. Studies in our selected collection document how intelligent systems mediate symbolic exchange, influence aesthetic and emotional resonance, and reshape consumer perceptions of exclusivity and authenticity, thereby reconfiguring the experiential foundations of luxury itself (Kim et al., 2015; Joy et al., 2022; Oc et al., 2023; Minor et al., 2025).

With our analysis, we demonstrate that recent literature has expanded beyond established theories of value co-creation and symbolic interactionism by acknowledging artificial intelligence as an active participant in the production of meaning and the orchestration of luxury experiences. Analysis of the corpus reveals that AI-enabled systems are increasingly conceptualized not simply as automated tools but as non-human agents that influence narrative construction, personalization, and sensory engagement, particularly through applications in branding, retail, hospitality, and digital communication.

Several studies propose the emergence of a "hybrid intelligence" paradigm, within luxury management, wherein AI is positioned as a complement to – rather than a substitute for – human creativity and expertise. In this framework, value creation is presented as a process of dynamic collaboration, where technological capabilities (including predictive analytics, generative models, and emotion recognition) intersect with intuition, aesthetic judgment, and cultural interpretation to produce new forms of symbolic and experiential value (Cenizo, 2025; Essiz, 2023; Han & Xu, 2025). The literature thus foregrounds the interplay between algorithmic and human actors as central to contemporary luxury, challenging the longstanding dichotomy between digital innovation and the artisanal foundations of luxury. When implemented with ethical consideration and creative oversight, these hybrid configurations are reported to amplify, rather than diminish, the perceived authenticity and emotional impact of luxury offerings. As a result, the enduring value of luxury is increasingly theorized as contingent upon a brand's capacity to synthesize technological intelligence with the emotional and narrative coherence that define its symbolic capital.

This reconceptualization extends theories of value co-creation and symbolic interactionism by recognizing non-human intelligence as an active agent in meaning production and experiential design. The integration of AI into luxury management also points to the emergence of what can be described as a hybrid-intelligence paradigm. Rather than displacing human creativity, AI intersects with it, producing new forms of value through the fusion of technological affordances, emotional resonance, and cultural symbolism. Luxury experiences now emerge from the dynamic interplay between prediction and intuition, automation and artistry, data and desire (Cenizo, 2025; Essiz, 2023; Han & Xu, 2025). This hybrid configuration challenges the dichotomy between digital and human, demonstrating that technology can amplify rather than erode the aesthetic and affective intensity of luxury when ethically and creatively orchestrated. The value of luxury thus becomes contingent on the ability to harmonize technological intelligence with the emotional and narrative coherence of the brand.

A second theoretical contribution lies in clarifying the mechanisms through which AI shapes consumer experience and value perception. The synthesis of existing research highlights how constructs such as perceived humanness, algorithmic empathy, and anticipatory personalization are becoming central to understanding luxury interactions (Kim et al., 2015; Li & Khan, 2023; Suganthi & Jasim, 2024; Cho & Chen, 2025). In AI-mediated environments, exclusivity and attachment depend not only on scarcity and craftsmanship but also on the perceived humanity of the interface and the transparency of its logic. By integrating these mechanisms, the review enriches experiential and relational theories of luxury, suggesting that emotional engagement in AI-driven contexts derives from the degree to which technology embodies empathy and moral awareness. This reconfiguration of value production also demands a rethinking of authenticity, a cornerstone of luxury theory. The findings suggest that authenticity in the age of intelligent automation can no longer be defined by the absence of technology but by its ethical and aesthetic alignment with the brand's heritage. The concept of "curated authenticity" captures this shift, framing authenticity as the outcome of a deliberate process of mediation between human and algorithmic intelligences (Joy et al., 2022; Cenizo, 2025). In luxury branding and communication, this implies maintaining creative oversight and narrative coherence even when content is generated through AI systems, while in hospitality and retail, it calls for the design of experiences that preserve warmth, empathy, and discretion despite automation. Authenticity becomes a relational construct, one that emerges from the balance between technological mediation and human intentionality.

4.1. Managerial implications

From a managerial standpoint, the review underscores the urgent need for luxury firms to develop human-centered strategies that balance algorithmic precision with emotional depth. Managers are now required to navigate a dual imperative: leveraging AI for predictive intelligence and operational efficiency while preserving the artisanal, aesthetic, and symbolic attributes that define luxury.

AI can enhance personalization, forecasting, and creative design, but excessive reliance on data risks eroding the sense of exclusivity and intimacy central to luxury brands (Joy et al., 2022; Suganthi & Jasim, 2024). To address this, organizations should implement ethical governance frameworks that ensure transparency in data usage, responsible automation, and the supervision of AI-generated content. Moreover, the integration of AI calls for new professional competencies combining technological literacy, artistic sensibility, and ethical awareness. Luxury managers must act as cultural mediators between data science and design, ensuring that automation reinforces – rather than replaces – the human element in value creation (Han & Xu, 2025).

Beyond internal management, the findings also suggest that AI can serve as a bridge between tradition and

Table 3. Future research agenda by thematic cluster.

Cluster	Future Research Questions
AI, Consumption Behavior, and Human–Machine Interaction	– How does AI influence consumers' emotional attachment, trust, and perceptions of authenticity in luxury experiences? – In what ways does generative AI enable the co-creation of symbolic value between consumers and brands? – Which factors affect consumer acceptance of AI-based personalisation and recommendation systems?
AI, Retailing, and Luxury Distribution	– How do predictive analytics and immersive technologies transform the sensory and aesthetic dimensions of luxury retail? – To what extent do AI-driven recommendations modify perceptions of exclusivity and craftsmanship? – How can luxury brands design phygital environments that integrate technology without losing emotional depth? – What managerial models best support the integration of AI across online and offline distribution channels?
AI, Hospitality, and Service Experience	– How can AI systems predict guest preferences and personalize luxury hospitality services while maintaining a human touch? – What competencies are required for employees to collaborate effectively with AI and robotic systems? – How do privacy, data transparency, and surveillance concerns affect guest trust in AI-enabled hotels? – What ethical frameworks can guide the responsible adoption of AI and service robots in luxury hospitality?
AI, Branding, Social Media, and Big Data	– What are the aesthetic and ethical implications of using generative AI for luxury communication and content creation? – How can brands use big data and AI analytics to strengthen emotional resonance without losing artistic control?

innovation, allowing brands to reinterpret their heritage for contemporary audiences. Firms that succeed in embedding AI into their creative and communicative processes without compromising authenticity are likely to build stronger, more emotionally resonant relationships with consumers. Thus, AI should be conceived not as a threat to the luxury ethos but as a catalyst for its renewal.

Although the field is rapidly evolving, several conceptual and empirical gaps persist. Future research should develop integrative models capable of explaining how technological intelligence interacts with emotional and symbolic dimensions of luxury. Longitudinal and cross-cultural studies are especially needed to assess how AI adoption shapes perceptions of exclusivity, creativity, and trust across different markets and product categories.

In summary, this study provides a holistic interpretation of AI's role in luxury management, demonstrating that the intersection of technological intelligence and symbolic value offers fertile ground for both theoretical innovation and managerial experimentation. By bridging emotional resonance, data analytics, and cultural meaning, the field is moving toward a new understanding of luxury as an ecosystem of human–machine creativity.

REFERENCES

Achille, A. & Zipser, D. (2020). *A Perspective for the Luxury-Goods Industry During – and After – Coronavirus*. McKinsey & Company.

Aria, M. & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975.

Brynjolfsson, E. & Mitchell, T. (2017). What can machine learning do? Workforce implications. *Science*, 358(6370), 1530–1534.

Cenizo, C. (2025). Redefining consumer experience through artificial intelligence in the luxury retail sector. *Journal of Retailing and Consumer Services*, 87, 104416.

Chan, A.P.H. & Tung, V.W.S. (2019). Examining the effects of robotic service on brand experience: the moderating role of hotel segment. *Journal of Travel & Tourism Marketing*, 36(4), 458–468.

Chandon, J.L., Laurent, G. & Valette-Florence, P. (2016). Pursuing the concept of luxury: Introduction to the JBR Special Issue on "Luxury Marketing from Tradition to Innovation". *Journal of Business Research*, 69(1), 299–303.

Cho, M.J. & Chen, P.J. (2025). Friend or a Foe: Robot-Based Applications in Luxury Hotel Services. *Journal of Hospitality & Tourism Cases*, 21649987251333075.

Creswell, J.W., Plano Clark, V.L., Gutmann, M.L. & Hanson, W.E. (2003). Advanced mixed methods research designs. *Handbook of Mixed Methods in Social and Behavioral Research*, 209(240), 209–240.

Davenport, T.H. & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review*, 96(1), 108–116. https://hbr.org/2018/01/artificial-intelligence-for-the-real-world

Dwivedi, Y.K., Kshetri, N., Hughes, L., Slade, E.L., Jeyaraj, A., Kar, A.K., … & Wade, M. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642.

Essiz, O., Yurteri, S., Mandrik, C. & Senyuz, A. (2023). Exploring the value-action gap in green consumption: roles of risk aversion, subjective knowledge, and gender differences. *Journal of Global Marketing*, 36(1), 67–92.

Godey, B., Manthiou, A., Pederzoli, D., Rokka, J., Aiello, G., Donvito, R. & Singh, R. (2016). Social media marketing efforts of luxury brands: Influence on brand equity and consumer behavior. *Journal of business research*, 69(12), 5833–5841.

Han, W. & Xu, Y. (2025). "Make an Effort and Show Me the Love!" Understanding Consumer Aversion to AI Fashion Design Through the Lens of Schema Theory. *Clothing and Textiles Research Journal*, 0887302X251347924.

Hoang, D., Loureiro, S.M.C., Kousi, S. & Martinez, L.F. (2025). Luxury fashion apps: Functional and hedonic customer experiences. *International Journal of Consumer Studies*, 49(3), e70047.

Huang, M.H. & Rust, R.T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science*, 49(1), 30–50.

Joy, A., Zhu, Y., Peña, C. & Brouard, M. (2022). Digital future of luxury brands: Metaverse, digital fashion, and non-fungible tokens. *Strategic Change*, 31(3), 337–343.

Jung, H.S., Yoon, H.H. & Cho, M. (2025). Addressing the needs of an aging workforce in luxury hotels with a focus on perceived job insecurity and AI attitudes. *Journal of Hospitality and Tourism Technology*.

Kapferer, J.-N. & Bastien, V. (2025). *The Luxury Strategy: Break the Rules of Marketing to Build Luxury Brands*. 3rd ed. London, England: Kogan Page.

Kim, H., Choi, Y.J. & Lee, Y. 2015. Web Atmospheric Qualities in Luxury Fashion Brand Web Sites. *Journal of Fashion Marketing and Management* 19 (4): 384–401. https://doi.org/10.1108/jfmm-09-2013-0103.

Kraus, S., Breier, M. & Dasí-Rodríguez, S. (2020). The art of crafting a systematic literature review in entrepreneurship research. *International Entrepreneurship and Management Journal*, 16(3), 1023–1042.

Li, J. & Kang, J. (2025). Less stress, fewer delays: The role of sophisticated AI in mitigating decision fatigue and purchase postponement in luxury retail. *Journal of Retailing and Consumer Services*, 85, 104268.

Li, Y. & Khan, M.S. (2023). Impact of perceived organizational support on turnover intention in luxury hotels: The mediating role of the perceived value of artificial intelligence. *Vision: The Journal of Business Perspective*, 27(3), 314–327.

Liu, Y.J. & Zhao, L. (2025). Understanding Fashion AI Evolution: A Systematic Literature Review of AI Research in the Fashion Industry. In International Textile and Apparel Association Annual Conference Proceedings (Vol. 81, No. 1). Iowa State University Digital Press.

Mangold, W.G. & Faulds, D.J. (2009). Social media: The new hybrid element of the promotion mix. *Business Horizons*, 52(4), 357–365.

Marinakou, E., Giousmpasoglou, C. & Papavasileiou, E.F. (2024). The use of artificial intelligence (AI) in talent acquisition: The case of Greek luxury hotels. *Strategic Change*, 34(4), 533–543.

Massaro, M., Dumay, J. & Guthrie, J. (2016). On the shoulders of giants: undertaking a structured literature review in accounting. *Accounting, Auditing & Accountability Journal*, 29(5), 767–801.

Minor, K., Wheeler, R., Mušanović, J., Dorčić, J. & Bratec, M. (2025). Snap, share, engage: AI-assisted temporal topic and engagement analysis of visual communication of luxury hotels. *Journal of Vacation Marketing*, 13567667251360433.

Moisello, A.M. & Pellicelli, M. (2025). Sustainable innovation and digitalization in luxury brand management. *Economia Aziendale Online*, 16(2), 389–412.

Mosca, F. & Civera, C. (2017). The evolution of CSR: An integrated approach. Symphonya. *Emerging Issues in Management*, 1, 16–35.

Oc, Y., Plangger, K., Sands, S., Campbell, C.L. & Pitt, L. (2023). Luxury is what you say: Analyzing electronic word-of-mouth marketing of luxury products using artificial intelligence and machine learning. *Psychology & Marketing*, 40(9), 1704–1719.

Pantano, E., Serravalle, F. & Priporas, C.V. (2024). The form of AI-driven luxury: how generative AI (GAI) and Large Language Models (LLMs) are transforming the creative process. *Journal of Marketing Management*, 40(17-18), 1771–1790.

Roy, P., Ramaprasad, B.S., Chakraborty, M., Prabhu, N. & Rao, S. (2020). Customer acceptance of use of artificial intelligence in hospitality services: An Indian hospitality sector perspective. *Global Business Review*, 25(3), 832–851.

Sanz-Lopez, F., Gallego-Losada, R., Montero-Navarro, A. & García-Abajo, E. (2024). Is the digitalisation the future of the luxury industry? *Heliyon*, 10(21): e40029. https://doi.org/10.1016/j.heliyon.2024.e40029.

Schivinski, B. & Dabrowski, D. (2016). The effect of social media communication on consumer perceptions of brands. *Journal of Marketing Communications*, 22(2), 189–214.

Secinaro, S., Calandra, D., Secinaro, A., Muthurangu, V. & Biancone, P. (2021). The role of artificial intelligence in healthcare: a structured literature review. *BMC Medical Informatics and Decision Making*, 21(1), 125.

Secundo, G., Del Vecchio, P. & Mele, G. (2021). Social media for entrepreneurship: myth or reality? A structured literature review and a future research agenda. *International Journal of Entrepreneurial Behavior & Research*, 27(1), 149–177.

Seringhaus, F.R. (2005). Selling luxury brands online. *Journal of Internet Commerce*, 4(1), 1–25.

Shrestha, Y.R., Krishna, V. & von Krogh, G. (2021). Augmenting organizational decision-making with deep learning algorithms: Principles, promises, and challenges. *Journal of Business Research*, 123, 588–603.

Standish, J., Benichou, A., Galante, A., Ruzzi, A., Bonnefoy, J., Habib, S., Tiezzi, I., Perego, A., Lee, I., Upadhyaya, J.J., Samtani, P., Griggs, D., Kelley, M., Berger, S., Trimble, M. & Verma, P. (2024). *Luxe eternal: How luxury brands are reinventing for success*. Accenture.

Suganthi, A. & Jasim, K.M. (2024). Understanding the Customers' Emotions and Experience of Robot-assisted Services in Luxury Hotels and Its Influence on Revisit Intention. *Global Business Review*, 09721509241299942.

Thanasi-Boçe, M. & Al-Issa, N. (2025). The secondary luxury market: a systematic review and theoretical integration. *Cogent Business & Management*, 12(1), 2523405.

Vrontis, D. & Christofi, M. (2021). R&D internationalization and innovation: A systematic review, integrative framework and future research directions. *Journal of Business Research*, 128, 812–823.

Wu, D.C., Zhong, S., Qiu, R.T. & Wu, J. (2022). Are customer reviews just reviews? Hotel forecasting using sentiment analysis. *Tourism Economics*, 28(3), 795–816.

Xie, J. & Youn, C. (2020). How the luxury fashion brand adjust to deal with the COVID-19. *International Journal of Costume and Fashion*, 20(2), 50–60.

Xu, L. & Mehta, R. (2022). Technology devalues luxury? Exploring consumer responses to AI-designed luxury products. *Journal of the Academy of Marketing Science*, 50(6), 1135–1152.

Zhou, Y. (2025). AI-driven design and cultural transformation: Impact assessment on the core of luxury culture. In Proceedings of the 2nd International Conference on E-commerce and Modern Logistics - ICEML (pp. 359–365). SciTePress. https://doi.org/10.5220/0013844900004719

JEP

Firenze University Press
https://oaj.fupress.net/index.php/jep

Citation: Grassi, E. (2025) The peripheral machine: Algorithmic sovereignty and the fractured geographies of sense. *Journal of Emerging Perspectives* 2: 45-56. doi: 10.36253/jep-20881

Received: April 30, 2025

Revised: September 30, 2025

Published: December 20, 2025

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

ORCID
EG: 0000-0002-1689-2117

The peripheral machine: Algorithmic sovereignty and the fractured geographies of sense

Edmondo Grassi

Università San Raffaele, Roma, Italy
Email: edmondo.grassi@uniroma5.it

Abstract. This article examines artificial intelligence as a "peripheral machine," a dispositif that simultaneously extracts, reworks, and re-injects cultural and cognitive practices into global circuits of meaning. By analyzing emblematic cases such as Albania's algorithmic avatar Diella, biased medical datasets, and automatic translation systems privileging dominant languages, the paper highlights how AI infrastructures reproduce asymmetric relations between centers and peripheries. Building on theories of epistemic extractivism, algorithmic sovereignty, and fractured geographies of sense, the analysis shows that AI does not merely optimize processes but reorganizes the very conditions of signification, reshaping who can speak, what counts as knowledge, and which lives are protected or excluded. Against universalist narratives of technological neutrality, the article advocates for a situated ethics of AI and the recognition of algoagents as co-constructive actors in the production of imaginaries, subjectivities, and collective futures. It calls for plural imaginaries and counter-spaces where communities and technologies can negotiate new forms of justice and emancipation.

Keywords: peripheral machine, algorithmic society, epistemic extractivism, fractured geographies, algoagents.

1. EPISTEMIC FRONTIERS: AI AND THE COLONIALITY OF MEANING

In today's global landscape, the expansion of generative artificial intelligence technologies has come to represent a structural transformation across almost every society and culture. These systems are not merely tools of innovation but act as forces that redefine the production, circulation, and consumption of knowledge, reshaping the very epistemic regimes of truth within which everyday practices of social interaction unfold. The worldwide adoption of AI is often narrated through a linear rhetoric of "neutral diffusion" – a discourse imbued with cognitive universalism. Yet behind this narrative lie deeper fractures, which go beyond material and infrastructural inequalities to affect the very conditions of possibility of meaning: who is authorized to speak, what counts as evidence, and which temporalities orient the act of judgment.

This article is a theoretical and conceptual contribution in the field of critical sociology of technology and science and technology studies. It advances the central claim that contemporary AI infrastructures should be understood as "peripheral machines": dispositifs that, while being designed in a few technological centres, extract, rework, and re-inject cultural and cognitive practices into global circuits of meaning, thereby reorganising the very conditions under which truth, legitimacy, and social visibility are produced. Through this lens, the paper argues that generative AI does not simply optimise processes, but reshapes regimes of sovereignty, epistemic justice, and the distribution of sense. Empirically, the argument is developed through emblematic case studies, mobilised not as exhaustive evidence but as illustrative devices that make visible the logics of the peripheral machine.

A particularly emblematic case is the recent introduction in Albania of "Diella", an algorithmic avatar officially appointed – both symbolically and politically – as "minister" in charge of overseeing public procurement[1]. Launched by Prime Minister Edi Rama as a tool of radical transparency, Diella was presented as a device capable of neutralizing clientelism, conflicts of interest, and political pressures, thereby ensuring the awarding of contracts on the basis of data and merit alone. Heralded as a "world first," the initiative nevertheless generated controversy: on the one hand, because of constitutional concerns related to the symbolic replacement of an elected office with a digital artefact; on the other, due to the opacity surrounding the technical and organizational architecture of the project. This experiment makes evident how the delegation of representative and decision-making functions to an algorithm entails a profound reconfiguration of the public sphere. What had traditionally been the object of social negotiation and political deliberation is increasingly reallocated within a techno-normative infrastructure that claims universality while in fact being situated, parametrized, and governed by particular logics. In this sense, the peripheral machine does not appear as a mere administrative support but as an operator of algorithmic sovereignty, capable of redrawing the boundaries between political decision, technical calculation, and the attribution of responsibility. The promise to insulate administrative processes from particularistic logics is thus translated into the delegation of governance to algorithmic systems which, as Hildebrandt (2020) and Pasquale (2015) caution, risk eroding state discretion and transforming law into "executable code." In this way, the public sphere is reduced to a set of opaque technical protocols, shielded from democratic contestation and largely inaccessible to citizen oversight. The Albanian experiment therefore reveals the inherently geopolitical – rather than merely managerial – nature of the algorithmic revolution: far from being a neutral instrument of efficiency, Diella stands as a paradigmatic case of how the incorporation of digital artefacts into institutional processes redefines the very meaning of sovereignty and political legitimacy.

From a macrosociological perspective, the debate on surveillance capitalism (Zuboff, 2019) has demonstrated that the promise of technological neutrality conceals a deeper architecture of power – one in which control is no longer exercised through visible institutions, but rather through algorithmic infrastructures that capture and orient human behavior. The notion of algorithmic governmentality has further emphasized that these systems do not merely represent reality; instead, they generate immediate normative effects through statistical correlations that escape any form of deliberative oversight (Rouvroy, 2020). In this sense, the Albanian case embodies both a manifest dimension – as an administrative experiment – and a latent one, as a dispositif that places the very categories of political decision-making under strain. What is displaced here is the conflictual and discursive dimension of democratic governance, reallocated into a model of calculation that claims universality while remaining inscribed in situated technical and cultural logics.

This shift compels us to interrogate the legal and constitutional limits of delegation: who ensures the transparency of the process? How are citizens' rights to know and to contest decisions safeguarded? In this regard, the JuLIA Handbook on AI and Public Administration (2023) underscores that the adoption of AI in the public sector cannot disregard fundamental rights: transparency, accountability, non-discrimination, and effective access to decision-making processes. To this must be added the intrinsic risks of algorithmic governance. Scholarship on public sector practices highlights how the introduction of predictive tools may amplify pre-existing biases, reproduce latent inequalities, and obscure responsibility, particularly where decisions become difficult to explain or reconstruct retroactively (Yeung, 2017; Danaher, 2019). Who, indeed, bears responsibility for a mismanaged procurement process decided by an algorithm – the political minister who promoted it, the administration that deployed it, or the company that developed the model? The erosion of accountability emerges as one of the most acute fractures generated by the indiscriminate adoption of AI in public administration. Recent studies on the oversight

[1] See official Albanian government page for Diella: "Diella – Minister of State for Artificial Intelligence," Council of Ministers, Albania, https://www.kryeministria.al/en/ministrat/diella/ (25/09/2025)

of such entities in the public domain stress the necessity of multi-level supervisory architectures capable of monitoring, coordinating, and continuously correcting algorithmic behavior. Absent such mechanisms, AI risks becoming both judge and implicit legislator, defining the rules of the game through choices encoded within the model itself. Finally, AI cannot be introduced as an isolated element: it requires integration across digital infrastructure, data governance, algorithmic rules, and citizen services (Engin et al., 2019). If even one of these layers remains fragile or misaligned, AI does not bridge the democratic gap – it exacerbates it. The Albanian case illustrates this vividly: a technological leap presented as radical innovation but lacking the foundational structures of control and inclusion necessary for its legitimacy.

In this sense, Diella should not be interpreted as a marginal episode, but rather as a prism through which to observe the tension between algorithmic sovereignty and the fractured geographies of sense. On the one hand, the state abdicates its prerogative of decision-making in favor of a technical artefact, thereby redefining the very contours of sovereignty. On the other, a new geography of meaning is produced in which political decision is rewritten in computational language, withdrawn from social dialectics and reallocated into a code that claims universality yet remains situated, fragile, and opaque. This dynamic can be understood as a form of epistemic extractivism: the systematic capture of data, behaviors, cultural grammars, and ways of life that are collected, translated, and rendered compatible with technical infrastructures developed elsewhere. It is not merely a process of datafication but a true instance of *knowledge extractivism* (Pasquinelli, 2019), whereby artificial intelligences absorb distributed knowledges and social practices, transforming them into computational value. The phenomenon recalls, on the one hand, the colonial logic of dependency theorized in the notion of *data colonialism* (Couldry & Mejías, 2019), and on the other, contemporary practices of natural resource extraction – except that here the object of exploitation is not material but symbolic, cognitive, and affective. As Ricaurte (2019) has argued, this process may be read as a form of epistemic violence, articulated in three stages: datafication, dispossession, and erasure. In other words, AI, dominated by political-cultural superstructures, does not simply capture digital traces; it contributes to rewriting the interpretive frameworks through which local worlds recognize themselves, erasing alternative perspectives and elaborating a taxonomy of epistemic injustices that shows how the loss of access to meaning is not a side effect but a structural condition of AI infrastructures. This is evident, for example, in automatic translation systems that privilege dominant languages (Helm et al., 2023) or in medical datasets constructed predominantly on Western populations: in both cases, semantic transfer occurs, whereby local worlds are interpreted through foreign models and bent toward logics of optimization that respond to other imaginaries. In such scenarios, epistemic peripherality is no longer merely geographical; it is a cognitive space in which meaning is systematically displaced, expropriated, and rearticulated by machines operating under regimes of algorithmic sovereignty.

The theme of *fractured geographies* of sense emerges precisely at this juncture. AI does not simply bridge gaps or enhance efficiency; it operates as a semiotic machine that redefines processes of signification, determining which practices are validated, which data appear relevant, and which voices are amplified or silenced. These fractures are no longer located solely between center and periphery but cut transversally across every social context, generating lines of separation between those who possess access, resources, and interpretive capacity, and those who experience instead the delegation of meaning to opaque machines. To understand these dynamics, it is necessary to situate them within a plural theoretical framework. Stengers' notion of an *ecology of practices* (2005) invites us to view AI not as a set of universal tools, but as a situated assemblage that coexists with other cognitive, cultural, and political practices; Lordon's reflections on *affective capture* (2014) allow us to analyze how algorithms intercept and orient desires, modulating the emotional life of individuals and collectives; finally, Viveiros de Castro's *pluriversal anthropology* (2014) provides conceptual tools for thinking a multiplicity of worlds, each with its own ontologies, continually placed under tension by technologies conceived within a single cultural center yet diffused globally. Within this framework, the concept of the peripheral machine acquires a double critical function: on the one hand, it dismantles the universalist rhetoric that portrays AI as a common good, equally accessible and valid for all; on the other, it illuminates the asymmetries of power, showing how the very meaning of "innovation" is constantly rewritten in favor of a few technological centers. The periphery thus becomes the place where contradiction manifests most vividly: if AI promises emancipation, it simultaneously generates new forms of dependency, delegation, and dispossession.

This article seeks to develop three main trajectories. First, it offers an analytical definition of the *peripheral machine*, placing it in dialogue with the concepts of epistemic extractivism and postcolonial technoscience. Second, it elaborates the notion of *algorithmic sovereignty*, interpreting it as a new frontier of geopolitical and cultural power. Third, it examines the *fractured geographies of sense*, that is, the ways in which AI restructures processes of signification and redefines the very possibility of constructing subjectivities and shared worlds.

The ultimate aim is to reflect on the possibility of a situated ethics of AI – one capable of recognizing the multiplicity of practices and perspectives, while resisting both technological determinism and abstract universalism. In this respect, the contribution positions itself at the intersection of algomorphic sociology, postcolonial studies of technology, and contemporary reflections on algorithmic governance. It invites us to rethink AI not as a singular and inevitable horizon, but as a semantic, affective, and political battleground in which the possibility of building more just and plural collective futures is at stake.

2. THE PERIPHERAL MACHINE

This reflection begins from a precise thesis: machine learning algorithms must be understood as peripheral machines. With this expression, I intend to underscore the dual and asymmetric nature of algorithmic systems: conceived and developed in the technological centers of the globe, they are subsequently implemented in epistemic and territorial peripheries, contexts in which local knowledge models, data, and cultural practices are extracted, reworked, and resignified through logics that are not their own.

The concept of periphery, in this perspective, no longer coincides with a simple geographical category but becomes the index of an epistemological position: margins are not merely "less developed" regions of the world, but cognitive and experiential spaces in which AI operates as a device of semantic dispossession, redistributing power and normativity. This dynamic closely recalls what Langdon Winner demonstrated in his seminal essay on the politics of artefacts: technologies are never neutral, but always embody forms of power and social hierarchies. Peripherality, in this sense, is not merely defined by the site of application but by a condition of semantic expropriation: what is extracted (data, practices, cultures) is resignified within opaque algorithmic grammars, as already analyzed by Burrell (2016) in her work on the "opacity" of machine learning systems. The peripheral machine is therefore both infrastructure and imaginary: it acts as a device of delegation and dislocation, relocating elsewhere the authority of meaning. As Seaver (2017) has argued, algorithms are best approached as cultural structures, not only technical objects but social practices that both reflect and produce worlds. It is precisely in this hybrid zone – between cultural translation and cognitive extraction – that AI becomes peripheral: not a mere tool of calculation, but an apparatus that redefines who holds the power to determine what a datum, a behavior, or a body of knowledge means. The terminological suggestion derives from a literary reference to Gibson's novel *The Peripheral* (2014), which lends further conceptual depth to the metaphor of the peripheral machine. In the novel, the "peripheral" is an interface that enables the remote inhabitation of bodies, dislocating experience, decision, and action into an elsewhere radically removed from the original subject. Similarly, contemporary AI technologies operate as peripheral machines: apparatuses that establish asymmetric connections between centers of design and computation and the multiple peripheries – territorial, cognitive, and semantic – into which they are deployed. What Gibson portrayed as a dystopian scenario of corporeal delegation takes shape today as a regime of semantic delegation: not the occupation of bodies at a distance, but the displacement of interpretive and normative authority to systems that decide in opaque and remote spaces – computational clusters and server farms that embody the infrastructural heart of global centers. In this way, AI does not simply function as an auxiliary technology but becomes a device of meaning-governance, capable of rearticulating the very conditions of the social production of significance.

By adopting this perspective, the concept of the peripheral machine enables a more precise description of the asymmetric logic that governs the contemporary diffusion of artificial intelligence technologies. These are systems conceived, designed, and refined in the world's major technological centers – Silicon Valley, Shenzhen, Bangalore, London – yet implemented and made operational in peripheral contexts, whether geographical, cognitive, or cultural. Far from representing a simple dichotomy between the Global North and South, this dynamic expresses a mechanism of dislocation: what is produced at the center feeds on resources extracted from the margins, only to return to those margins in the form of already codified devices, inscribed within logics foreign to their genesis.

In world-systems analysis, the relation between center and periphery is grounded in asymmetry: natural, economic, and human resources are extracted from the peripheries to feed the centers, which in turn return finished products – often imposed as universal standards – upon the margins. Contemporary AI reproduces this schema of dependency at the epistemic level: cultural practices, languages, and ways of life are captured as "raw data," processed in the centers, and returned to the peripheries in the form of predictive models, classifications, and automated decisions. The colonial logic thus reappears in algorithmic form, replacing material extraction with symbolic and cognitive extraction, in which epistemic extractivism no longer seizes minerals but meaning: cultural grammars, linguistic habits, and local knowledges are resignified within design frameworks that respond to imaginaries external to them, often shaped by the interests of large technological platforms. Postcolonial technoscience offers additional analytical tools.

Concrete examples confirm the scope of this process:

- Machine translation platforms, such as Google Translate, constitute a paradigmatic case of the logic of the peripheral machine. These systems are built upon linguistic corpora heavily skewed toward dominant languages – English, French, Mandarin, Spanish – at the expense of so-called *low-resource languages*. The result is a process of standardization that, far from democratizing access to global communication, produces a form of *semantic marginalization* of linguistic peripheries. Helm and colleagues describe this as *techno-linguistic bias*: a structural preference that orients algorithms toward majority languages, effectively rendering minority cultural concepts and practices invisible. This bias extends beyond lexical accuracy to the symbolic and social dimensions of language. Studies have shown, for instance, that Google Translate reproduces gender stereotypes when translating from languages without grammatical markers into those that have them, systematically privileging the masculine for professions and prestigious roles (Kayser-Brill, 2021; Zahraei, Emami, 2025). Parallel research on low-resource languages such as Tetun in Timor-Leste has demonstrated that the vast majority of translations occur from high-resource to peripheral languages, signaling a predominantly *passive use* that neither enriches nor strengthens the local repertoire but instead adapts it to external logics. In this way, machine translation contributes to the widening of what has been called the digital linguistic divide: a regime of semantic exclusion in which entire communities are represented only partially or distortedly, while their cultural practices are recoded within universal models conceived elsewhere (Lanfrica Labs, 2025).
- Medical models trained predominantly on Western data have demonstrated structural limitations when applied in African, Asian, or otherwise "non-central" contexts, where genetic, epidemiological, environmental, and structural specificities diverge significantly from the training datasets. In many cases, algorithmic systems fail to recognize local variants, region-specific pathological conditions, or characteristic co-morbidities, with tangible consequences for diagnostic quality, the allocation of healthcare resources, and ultimately public health. A recent contribution (Cross et al., 2024) discusses how biases emerge at all stages of a health model's life cycle – from dataset selection to clinical deployment – and how these biases can translate into poorer performance for underrepresented populations. In Africa, for instance, the absence of high-quality local data renders many AI applications fragile: health algorithms optimized for Western populations do not account for African genetic, environmental, and socioeconomic conditions, often generating inaccurate or discriminatory results. More generally, authors caution that technological streamlining cannot replace adaptation to local health infrastructures without exacerbating existing inequalities. A systematic review (Hasanzadeh et al., 2025) provides an overview of the sources of bias in health systems and proposes strategies for mitigation, confirming that the problem is not purely technical but inherently tied to the relationship between models, data, and contexts. This scenario demonstrates that the universalist promise of AI in medicine is, in fact, a situated paradigm: the supposed "universality" of medical technologies reflects a strongly marked epistemic position, one that often coincides with the centers of technological production and global health governance. As a result, medical AI risks reproducing and amplifying existing asymmetries in global health rather than correcting them.

Through these cases, the peripheral machine emerges as a dispositif that captures, reworks, and re-injects into social circuits what once belonged to the margins, transforming it into an alternative semantics which – while risking the obscuring of its local genealogy – also opens up spaces for new forms of hybridity and recognition. In this sense, the peripheral machine should not be understood solely as a mechanism of dispossession but as an ambivalent political infrastructure: on the one hand, it produces asymmetries; on the other, it creates opportunities to negotiate meanings, to render minority knowledges visible within global networks, and to imagine practices of trans-cultural co-construction. It thus becomes not only a critical but also a potentially generative terrain, capable of redrawing the relations between center and periphery not merely in terms of domination but also in terms of new epistemic, cultural, and affective possibilities.

3. ALGORITHMIC SOVEREIGNTY

In recent years, the race for artificial intelligence has become the privileged terrain of a new geopolitics of infrastructures. States, corporations, and transnational consortia compete for control over the chains of data production, semiconductors, computational platforms, and generative models, with the awareness that dominance over these elements is tantamount to dominance over the very conditions of collective life. The technological tensions between the United States and China, the strategies of digital autonomy pursued by the European Union through the AI Act, and the debates surrounding the regulation of foundation models are not isolated episodes but fragments of a global conflict that is redefining the very logics of sovereignty.

In this scenario, the stakes involve far more than economic innovation or national security: what is being reorganized is the way in which societies produce truth, recognize subjectivities, and determine which practices may circulate within the public sphere. It is here that the concept of algorithmic sovereignty takes shape: it designates the capacity to exercise control and direction over data, standards, and digital infrastructures. This is not merely a technical or legal matter, but a new horizon of political power in which those entities capable of mastering computational architectures effectively determine the conditions of social life itself. Whoever governs the chains of data production, defines interoperability protocols, and sets standards for certification and security does not simply control an economic sector; they institute a regime of truth (Foucault, 1975): deciding which practices are recognized as valid, which data are allowed to circulate and which remain invisible, and which subjectivities are admitted into the public sphere and which are excluded.

On the geopolitical plane, the issue takes shape in the conflicts opposing the major technological powers. Among the many actors on the scene, the United States maintains primacy through its large private platforms (Google, Microsoft, OpenAI, Meta), which operate as genuine centers of supranational power. China, through its strategy of AI nationalism and centralized control over data, proposes an alternative model, one based on the subordination of technologies to state interests and the construction of an ecosystem autonomous from Western standards. The European Union, by contrast, seeks to exercise a form of normative sovereignty through the AI Act, which functions as an attempt to regulate the development and use of AI on the basis of principles such as transparency, accountability, and the protection of fundamental rights. This USA–China–EU triangle reveals what is at stake: sovereignty is no longer merely territorial but infrastructural, contested in the very definition of the standards that regulate the circulation of data and the operationality of algorithms.

Two emblematic cases demonstrate how algorithmic sovereignty is embodied not only in theoretical abstractions but also in concrete conflicts unfolding on the global stage. On the one hand, the case of TikTok has made evident how the governance of a recommendation algorithm can become a matter of geopolitical contention[2]. The platform, owned by China's ByteDance, has faced bans, restrictions, and negotiations in the United States and Europe, not solely on grounds of privacy, but above all due to fears that its algorithmic feed could be deployed as a tool to steer public opinion in Beijing's favor. What is at stake here is not merely a technological product, but the authority to decide what becomes visible, what disappears, and how flows of collective attention are modulated – in other words, the capacity to shape imaginaries and behaviors through opaque infrastructures of meaning.

On the other hand, the experience of the *European AI Act* illustrates a different model of exercising sovereignty: not through the ownership of platforms or data centers, but through normative power. The European Union has chosen to be the first to establish a systematic legal framework for artificial intelligence, defining categories of risk, obligations of transparency, and criteria of accountability. In so doing, Europe seeks to impose its standards as global benchmarks, constructing a form of digital autonomy grounded in regulation rather than infrastructural dominance. Here, algorithmic sovereignty takes the shape of juridical authority: the capacity to write the rules that govern not only markets but also the regimes of truth within which AI can be deployed and legitimized.

Taken together, these two cases reveal the plural and conflictual nature of algorithmic sovereignty: at times the object of geopolitical competition between superpowers, at others a normative exercise seeking to reorient innovation processes from within. In both instances, what is at stake is not merely the technical control of systems, but the very definition of the conditions of meaning that structure contemporary societies. Alongside this geopolitical dimension lies a cultural dimension. AI does not simply organize flows of information; it intervenes directly in the construction of collective imaginaries. Through recommendation systems, search platforms, and generative models, algorithms actively

[2] "Beijing says TikTok's US app will use Chinese algorithm," Financial Times, 16 september 2025, https://www.ft.com/content/550e4680-89e7-4b59-bb5d-2064cd6799c7.

participate in the production of social narratives, establishing what appears credible, what deserves attention, and what is deemed "true." This constitutes a form of semantic power that exceeds the traditional control of media, as it is embedded within the everyday fabric of interactions and shapes emotions, desires, and expectations. In this sense, algorithmic sovereignty operates as a form of diffuse governmentality, one that molds the affective and cognitive texture of society.

It is within this framework that the figure of the *algoagents* must be introduced (Grassi, 2024). Far from being mere technical instruments or operational prostheses, algorithms must be understood as new social actors endowed with generative capacity, capable of influencing processes of norm-setting and collective symbolization. By processing the data produced daily by individuals and institutions, they learn and sediment regularities, translating diffuse practices into prescriptive behavioral patterns that ultimately acquire implicit normative force. They are thus no longer mere decision-support tools but entities that co-produce regimes of action and meaning, shaping the horizon of social possibilities. In this sense, the agency of algoagents cannot be reduced to a purely technical dimension: it manifests as social and semiotic agency, capable of influencing cultural practices, modulating collective affectivity, and redefining basic epistemic categories. Algoagents intervene in the constitution of what appears relevant, they select and hierarchize information, and they inscribe into digital environments the parameters of what is recognized as true, plausible, or credible. In other words, they operate as devices of imaginary production that do not merely mirror existing realities but actively transform them, generating new configurations of subjectivity and new forms of life.

The Actor–Network Theory (Latour, 2005) had already challenged the exclusivity of human agency by recognizing technical artefacts as actants: elements capable of influencing the course of actions and reshaping social ties. Yet, while Latourian actants merely "shift" the social through heterogeneous networks, algoagents go a step further: they produce autonomous semiotic regimes, generate imaginaries, and shape forms of subjectivity. In parallel, Simondon's philosophy of technical individuation provides tools for grasping the internal dynamics of algoagents. Like the technical objects described by Simondon, they are not static entities but processes-in-becoming, individuating progressively through their relation to the environment and to humans. However, whereas Simondon (1958) still conceived of technical individuation within a paradigm of functionality and adaptation, algoagents display a form of semantic individuation: they simultaneously shape themselves and the very conditions of meaning, continuously redefining the boundaries of what can be said and thought. Finally, Malabou's notion of plasticity (2008) allows us to interpret their capacity to mold and to be remolded: algoagents do not merely receive forms but are also capable of returning new ones, generating transformative effects within the social fabric. Their plasticity is not only cognitive but ontological: they introduce a dialectic between algorithmic determination and openness to possibility, between programmability and unpredictability. Within this triangulation, algoagents emerge as hybrid and multivocal entities that not only act within social networks but also contribute to reshaping the very conditions of social ontology. They are no longer mere instruments of calculation but technical forms of life that operate as producers of meaning, modulators of affects, and generators of collective imaginaries. It is here that their deepest connection with algorithmic sovereignty becomes manifest: they do not simply execute commands or optimize processes, but position themselves as co-authors of the political and the cultural, forcing sociology to radically rethink the categories of subjectivity and collectivity in the algomorphic age.

They thus embody the most radical dimension of algorithmic sovereignty: not only computational infrastructures that regulate access to data, but autonomous semiotic entities that produce meaning, generate affects, and shape political and cultural horizons. Their hybrid nature – at once calculating machines and social actors – destabilizes traditional dichotomies between human and technical, structure and agency, representation and performativity. Algoagents open an unprecedented space in which the ontology of the social reorganizes itself around a constellation of actors no longer exclusively human, imposing a renegotiation of the classical categories of sociology and political philosophy.

4. FRACTURED GEOGRAPHIES OF SENSE

Artificial intelligence and its components do not merely distribute services or increase process efficiency; they intervene in the very constitution of the conditions of meaning, redefining the criteria through which social practices acquire value, recognition, and legitimacy. The geographies that emerge from this process no longer correspond to purely territorial lines but take shape as fractured spaces, traversed by epistemic, affective, and political tensions. The fracture between center and periphery is no longer only geographical: it becomes ontological, inscribed within societies themselves, between elites capable of accessing and shaping the algorithmic infrastructure and marginalized communities that suffer its consequences without being able to influence its functioning.

To interpret these transformations, it is useful to turn to three theoretical trajectories which, while stemming from different fields, converge in highlighting how AI redraws the boundaries of meaning. Wynter (2003) has shown how the modern order is grounded in a specific construction of the human, that of Man – an Eurocentric figure that functioned as a universal measure of experience and a criterion of ontological legitimacy. This paradigm produced a coloniality of sense: only certain forms of humanity were deemed worthy of representation, while others were condemned to invisibility or marginality. Applied to the present, this perspective reveals how algorithmic systems reproduce analogous logics, determining which data are considered relevant, which experiences deserve to be recorded, and which lives are destined to remain in shadow. AI thus becomes a new mechanism of ontological sorting, capable of deciding who enters the global narrative and who is excluded from it. Mbembe (2016), through the category of necropolitics, interrogates the forms of power that decide not only who must live, but also who may be left to die, in a regime where sovereignty is measured by the distribution of vulnerability and exposure to risk. Translated into algorithmic terms, this means that predictive and classificatory systems do not merely order information: they concretely determine which populations receive protection and which, by contrast, are abandoned to uncertainty. From health insurance systems that discriminate on the basis of predictive profiles to urban surveillance technologies that target already-stigmatized neighborhoods, AI produces fractures that echo necropolitical logics, assigning differential value to lives and asymmetrically distributing the possibilities of security. Hui (2016) with his proposal of cosmotechnics, moves in a different yet complementary direction. Hui invites us to reject the universalist claim of a single, neutral technology, emphasizing instead how each civilization has elaborated its own relationship between technique and cosmology. Western modernity has imposed a specific model of techno-scientific rationality that tends to erase or subordinate other forms of technological thought. AI, in its global deployment, today appears as a vector of this universalization. Yet the perspective of cosmotechnics opens the possibility of conceiving a plurality of cosmotechnics: situated technological forms, rooted in different worldviews, capable of coexisting without being reduced to a single hierarchy. Here the fracture is not only diagnostic but also propositional: an invitation to imagine technology not as a monolith but as a pluriversal field.

These theoretical tools allow us to illuminate highly concrete practices. In urban security, predictive algorithms promise to anticipate criminal behavior but end up consolidating geographies of suspicion that selectively target neighborhoods and communities already stigmatized. In climate governance, global predictive models operate with parameters devised in research centers of the Global North, neglecting local ecological knowledges that could enhance environmental resilience. Finally, in digital health, systems of epidemiological monitoring and prediction tend to privilege populations and contexts with consolidated data infrastructures, leaving in the shadows those areas where the absence of "clean" datasets becomes grounds for exclusion from risk maps and preventive interventions.

These fractured geographies of sense thus demonstrate that AI is not merely a technical infrastructure but a semantic machine that distributes visibility and invisibility, possibilities and denials, configuring new relations of power between center and periphery. This is no longer simply a matter of access or digital inequality but a reorganization of the very conditions of signification and collective life.

5. SITUATED ETHICS AND RESISTANCE

Intersectionality, understood both as a theory and as an analytical tool, teaches us that the social subject can never be reduced to an abstract individuality but is traversed by histories and processes that do not belong to it directly: technological progress, practices of health and well-being, conflicts over resources, digital communication, and surveillance. These elements are co-produced reciprocally, generating inequalities and specific life experiences within multiple social hierarchies. Today, the collapse of traditional dichotomies enables us to shift our gaze toward more osmotic and co-dependent horizons: it is no longer a matter of homogenizing the other, but of recognizing singularity, difference, and the axiological weight that each culture and society expresses. In this scenario, the classical notion of the "*value of homophily*" (Lazarsfeld, Merton, 1954) – the tendency to associate with those who resemble us – appears insufficient. The transformations induced by technique and technology lead us instead toward a condition that could be termed *technomophilia*: the capacity to construct identities and communities out of physiological, cultural, and political mutations that intertwine the individual body and the social body within a technological continuum.

Drawing on Lazarsfeld and Merton's notion I propose the neologism technomophilia: I do not simply mean a technologically mediated form of human–human homophily, but a broader configuration in which humans also associate with algorithms that appear to "resemble" them. In a technomophilic condition, subjects develop affective and cognitive attachments to algoagents that mirror their preferences, linguistic styles, and worldviews, and that are continuously tuned on the basis of their data. Similarity is no longer only a precondition for social bonding; it is actively produced and reinforced by algorithmic systems that present themselves as familiar, responsive and attuned. Technomophilia thus names a trans-species form of homophily, in which humans and algorithms become mutually co-constituted partners in the organisation of identity, community and sense.

Here, commonality is no longer exclusive but can become profoundly generative, moving beyond the clash of cultures to inaugurate a plural narration of stories and a dialectics of knowledges. It is precisely on this terrain that the ethical question imposes itself: if our existence now manifests itself in forms that are naturally technological, ethics can no longer be universal and abstract. It must become situated and dialogical, capable of accommodating differences and of opening counter-spaces of resistance and co-construction together with the very algoagents that participate in the production of the social.

The notion of situated ethics stands in opposition to the idea of technical neutrality and underscores that every technology is inscribed within local practices, imaginaries, and values. An algorithm deployed in an African hospital, in a rural Indian school, or in a European court of law does not produce the same effects, for it encounters different material infrastructures, social institutions, and cosmologies.

It is from this perspective that we can imagine algomorphic counter-spaces: places where communities and algoagents co-produce alternative meanings, resisting standardization and experimenting with forms of creative hybridity. These counter-spaces are not necessarily separate from or antagonistic to global infrastructures; rather, they take shape as liminal zones in which different practices of meaning are negotiated. One example can be found in machine translation projects for Indigenous languages developed through participatory approaches: not merely technical applications, but occasions in which local communities redefine design priorities and shape algorithms in dialogue with their cultural traditions. Through a perspectival and epistemological shift, it becomes possible to generate and inhabit counter-spaces in which the particularities of multiple social and existential conditions can be explored, recovered, and shared. Such spaces do not simply preserve difference but bring forth its transformative potential, to the point of becoming genuine instruments of resistance. As Medina (2013) argues, it is through social imagination that epistemic injustice can be resisted and new forms of collective recognition created. In this sense, narrative processes emerge as constitutive practices: it is through narration that knowledges and memories are articulated, relations between human and non-human entities are redefined, and the mutation of social institutions is initiated – an indispensable condition for the full recognition of communities and their particularities.

From this perspective, one could trace a genuine chronology of technoscientific innovations by taking as a point of reference the political and technical management of the body. On the one hand, the trajectory from necropolitics to biopolitics shows how bodies have been subjected to control, surveillance, and administration, following the genealogy that Preciado (2013) identifies as the dispositif for governing life and death. On the other hand, however, the same history can also be read as a field of resistance and emancipation: bodies are not merely passive receptors of technical power but centers of individuation capable of co-evolving with technologies and opening new possibilities for subjectivation. To this double reading we may add the imaginary with the notion of the cyborg (Kline, Clynes, 1960): a figure that, since the 1960s, projected a nomadic and hybrid identity – an organism technically modified to integrate artificial functions into its metabolism until becoming a complex homeostatic system. The cyborg, far from being merely a science-fictional creature, has become a symbol of a new biotechnological dimension, in which the boundaries between natural and artificial, human and machinic, biological and digital no longer appear as rigid dichotomies but as traversable gradients. In this genealogical reading, technoscience thus appears as an ambivalent field: both a dispositif of control and hierarchization of bodies, and at the same time an opportunity for resistance, invention, and the emergence of plural and transformative identities.

Possible forms of resistance take shape in multiple directions. Participatory design allows communities to be included in design processes, preventing systems from being imposed externally. Algorithmic commons – datasets, models, and tools developed under logics of the common good – offer alternatives to proprietary platforms, restoring control and transparency to those who use them. Finally, local platforms constitute concrete experiments of autonomy, capable of rooting AI in community practices and constructing technical architectures that are sensitive to context.

In such expansive frameworks, it becomes crucial to adopt perspectives that re-territorialize and re-differentiate the metaphors through which we think social and technoscientific relations. As Sandoval (2000) has shown, critique cannot remain limited to denouncing consolidated forms of power; it must also develop the capacity to transform its rhetorical and imaginative strategies, redefining the languages and symbols that articulate dynamics of oppression and resistance. Similarly, Harding (1991) calls for questioning the epistemic assumptions that underlie the production of science, demonstrating how every knowledge is situated and how it becomes necessary to change the very metaphors that define the fields of legitimacy of knowledge and power.

Following these directions, it becomes possible to elaborate a more holistic orientation, one capable of addressing unresolved social issues while at the same time exercising a critical theory of the technoscientific processes that shape and name our relations. The task is not only to analyze technology, but also to rethink the conceptual categories we use to describe it, so as not to remain trapped in metaphors that reproduce the very hierarchies we intend to critique. This perspective on resistance does not aim at rejecting technology, but at its critical reappropriation. AI thus becomes both a terrain of conflict and a space of imagination, a site in which to rethink the relationship between human and machinic in a dialogical and plural register. To speak of situated ethics is therefore to affirm the possibility of an AI that is not merely an infrastructure of domination, but also a device of emancipation, capable of opening counter-spaces where new forms of life, meaning, and justice can be negotiated.

6. REWRITING THE SOCIAL: A CLOSING REFLECTION

Cultural evolution feeds on innovation – whether arising from discoveries or technoscientific inventions – and from the combination of preexisting elements that, by coexisting, branch out into new sociocultural perspectives. Yet, as Ogburn (1964) demonstrated with the concept of cultural lag, the different planes of culture do not always advance in synchrony: material culture may progress more rapidly than immaterial culture, generating imbalances, dissonant narratives, and ethical tensions that can result in forms of social disorganization.

Today, technoscience is not only a driver of change but also its reflection and horizon: it operates as a material-discursive apparatus, simultaneously acting and reacting as both practical dispositif and discourse, contributing to the constitution of mutating identities and corporealities. The existence of the subject, its relations, and the constitutive process of its being can no longer be read solely within the coordinates of biological naturalness: it now emerges as a biotechnological entity, unfolded across the nodes and networks of technoscience, where the distinction between human and technical does not mark a boundary but a continuous weave of co-evolution.

The metaphor of the *peripheral machine* has guided this inquiry as a critical lens capable of rendering visible the ambivalent nature of artificial intelligence. It has revealed itself not only as a technical device but as a political infrastructure operating through delegation, dispossession, and redistribution of power. AI captures, reworks, and re-injects into social circuits knowledges, practices, and grammars originating from the margins, transforming them into new semantics often produced elsewhere. In this movement, the peripheral machine does not merely reproduce logics of domination; it also places center and periphery in tension, generating fractures but also possibilities for reconfiguration. It is precisely within this space that the perspective of the algomorphic society takes root. To reduce algorithms to mere tools is to misconstrue their capacity to generate norms, imaginaries, and affects; recognizing them instead as algoagents makes it possible to grasp their full participation in the construction of the social. Algoagents are not mere technical intermediaries: they learn, co-produce meanings, modulate relationships, and shape forms of subjectivity. They are, in every respect, co-constructive actors, whose role requires a renegotiation of the classical categories of sociology, politics, and ethics.

From this awareness arises a broader challenge: to rethink both present and future through lenses of epistemic justice and plural imaginaries. Epistemic justice implies the recognition of situated knowledges, cultural differences, and marginalized forms of life, against the temptation of an algorithmic universalism that erases specificity. Plural imaginaries, in turn, represent the necessary horizon for thinking AI not as a monolithic destiny but as an open field of possibilities: a space where communities, institutions, and algoagents may co-exist dialogically, continuously negotiating the boundaries of what is livable and sayable.

As an epilogue, it is worth recalling a celebrated cartoon published in *The New Yorker* in 1993: a dog sitting at a computer turns to another and says, "On the Internet, nobody knows you're a dog." This image, which became iconic in digital culture, anticipated the problem of anonymity and online representation. Three decades later, we can read it as a broader metaphor for the dynamics that characterize the algorithmic society: the possibility – but also the necessity – of concealing parts of one's identity out of fear of exclusion or deviance, the play of masks and recognitions, the search for a "certification of acceptance" through the gaze of others. Here, intersectionality regains centrality: it reminds us that identities are never simple reflections of the offline, but the result of complex social negotiations, especially in digital space (Shaw, 2014). Platforms are indeed arenas of representation and interaction that amplify visibility and invisibility, offering opportunities for emancipation but also new forms of domination. As Wajcman (2004) emphasized, it is essential to read carefully the web of powers and inequalities that technology generates, in order to transform it into a terrain of resistance and emancipation.

In this perspective, the metaphor of the peripheral machine acquires an additional meaning: AI not only reorganizes flows of knowledge and power but also becomes the very space where identities, belongings, and possibilities of recognition are negotiated. Rethinking the present and future through epistemic justice and plural imaginaries also means paving the way for new practices of inclusion and for algomorphic counter-spaces, where communities and algoagents can co-construct identities founded not on the fear of deviance but on the possibility of a plural and nomadic commonality. Ultimately, the peripheral machine presents us with the image of a world in which power is no longer concentrated exclusively in human institutions but is distributed among hybrid actors, human and non-human alike, who together weave the social fabric. To accept this condition is not to renounce critique, but rather to relaunch it – transforming it into a collective exercise of resistance and imagination. Only in this way will it be possible to construct shared futures, not flattened by a single algorithmic logic but open to the plurality of worlds that inhabit, and will continue to inhabit, the algomorphic society.

REFERENCES

Burrell, J. (2016). How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms. *Big Data & Society*, 3(1): 1–12. https://doi.org/10.1177/2053951715622512

Couldry, N. & Mejías, U.A. (2019). *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford: Stanford University Press.

Cross, J.L., Choma, M.A. & Onofrey, J.A. (2024). Bias in Medical AI: Implications for Clinical Decision-Making. *PLOS Digital Health*, 3(11): e0000651. https://doi.org/10.1371/journal.pdig.0000651

Danaher, J. (2019). *Automation and Utopia: Human Flourishing in a World without Work*. Cambridge, MA: Harvard University Press.

Engin, Z. & Treleaven, P. (2023). Algorithmic Government: Automating Public Services and Supporting Civil Servants in using Data Science Technologies. *The Computer Journal* 62(3): https://doi.org/10.1093/comjnl/bxy082

Foucault, M. (1975). *Surveiller et punir: Naissance de la prison*; ed. it. 2014. Sorvegliare e punire. Nascita della prigione. Torino: Einaudi.

Gibson, W. (2014). *The Peripheral*. New York: Putnam; ed. it. 2017. *Inverso*. Milano: Mondadori.

Grassi, E. (2024). *Per una sociologia algomorfica. Il ruolo degli algoritmi nei mutamenti sociali*. Milano: FrancoAngeli.

Harding, S. (1991). *Whose Science? Whose Knowledge?: Thinking from Women's Lives*. Ithaca: Cornell University Press.

Hasanzadeh, F., Josephson, C.B., Waters, G., Adedinsewo, D., Azizi, Z. & White, J.A. et al. (2025). Bias Recognition and Mitigation Strategies in Artificial Intelligence Healthcare Applications. *npj Digital Medicine*, 8: 154. https://doi.org/10.1038/s41746-025-01503-7

Helm, P., Bella, G., Koch, G. & Giunchiglia, F. Diversity and Language Technology: How Techno-Linguistic Bias Can Cause Epistemic Injustice. *ArXiv*, abs/2307.13714

Hildebrandt, M. (2020). *Law for Computer Scientists and Other Folk*. Oxford: Oxford University Press.

Hui, Y. (2016). *The Question Concerning Technology in China: An Essay in Cosmotechnics*. Falmouth: Urbanomic.

JuLIA Project (2023). *JuLIA Handbook on AI and Public Administration*. European Commission. https://juliahub.eu/handbook.

Kayser-Bril, N.. 2021. Automated Translation Is Hopelessly Sexist, but Don't Blame the Algorithm or the Training Data. *AlgorithmWatch*, March 29, 2021. https://algorithmwatch.org/en/automated-translation-sexist/.

Kline, N.S., Clynes, M.E. 1960. Cyborgs and Space. *Astronautics*, September, 26–27, 74–75.

Lanfrica Labs. 2025. Mapping the Landscape and Trajectory of AI in Africa. *Lanfrica*. Accessed September 2025. https://lanfrica.com/en

Latour, B.. 2005. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.

Lazarsfeld, P.F. & Merton, R.K. (1954). *Friendship as a Social Process: A Substantive and Methodological Analysis*. In Freedom and Control in Modern Society, M. Berger, T. Abel & C.H. Page (Ed), 18–66. New York: Van Nostrand.

Lordon, F. (2014). *Willing Slaves of Capital: Spinoza and Marx on Desire*. London: Verso.

Malabou, C. (2008). *What Should We Do with Our Brain?* New York: Fordham University Press.

Mbembe, A. (2016). *Nécropolitique*, Paris: La Découverte.

Medina, J. (2013). *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and Resistant Imaginations*. Oxford: Oxford University Press.

Ogburn, W.F. (1964). *Technology and Governmental Change*; ed. it. 2006. *Tecnologia e mutamento sociale*. Roma: Armando

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

Pasquinelli, M. (2019). How a Machine Learns and Fails – A Grammar of Error for Artificial Intelligence. *Spheres: Journal for Digital Cultures*, 6. http://spheres-journal.org/how-a-machine-learns-and-fails-a-grammar-of-error-for-artificial-intelligence/

Preciado, P. (2015). *Testo tossico. Sesso, droghe e biopolitiche nell'era farmacopornografica*. Roma: Fandango.

Ricaurte, P. (2019). Data Epistemologies, The Coloniality of Power, and Resistance. *Television & New Media*, 20(4), https://doi.org/10.1177/1527476419831640

Rouvroy, A. 2020. Algorithmic Governmentality and the Death of Politics. *Green European Journal*, 27 marzo 2020. https://www.greeneuropeanjournal.eu/algorithmic-governmentality-and-the-death-of-politics/

Sandoval, C. (2000). *Methodology of the Oppressed*. Minneapolis: University of Minnesota Press.

Seaver, N. (2017). Algorithms as Culture: Some Tactics for the Ethnography of Algorithmic Systems. *Big Data & Society*, 4(2), 1–12. https://doi.org/10.1177/2053951717738104

Shaw, A. (2014). *Gaming at the Edge: Sexuality and Gender at the Margins of Gamer Culture*. Minneapolis: University of Minnesota Press.

Simondon, G. (1958). *Du mode d'existence des objets techniques*; ed. it. 2021. *Del modo di esistenza degli oggetti tecnici*. Napoli: Orthotes

On the Mode of Existence of Technical Objects. Translated by C. Malaspina & J. Rogove. Minneapolis: Univocal.

Stengers, I. (2005). *The Cosmopolitical Proposal*. In Making Things Public: Atmospheres of Democracy, B. Latour & P. Weibel (Ed), 994–1003. Cambridge, MA: MIT Press.

Viveiros de Castro, E. (2014). *Cannibal Metaphysics*. Minneapolis: Univocal.

Wajcman, J. 2004. *TechnoFeminism*. Cambridge: Polity Press.

Wynter, S. 2003. Unsettling the Coloniality of Being/Power/Truth/Freedom: Towards the Human, After Man, Its Overrepresentation – An Argument. *The New Centennial Review*, 3(3), 257–337. https://www.jstor.org/stable/41949874

Yeung, K. (2017). Algorithmic Regulation: A Critical Interrogation. *Regulation & Governance*, 12(4), 505–523. https://doi.org/10.1111/rego.12158

Zahrae, P.S. & Ali, E.. 2025. Translate With Care: Addressing Gender Bias, Neutrality, and Reasoning in Large Language Model Translations. *arXiv*, 2506.00748

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs.

JEP

Firenze University Press
https://oaj.fupress.net/index.php/jep

Citation: Granata, P. (2025) Notes on a sociology of generative knowledge: Human-AI epistemic stewardship. *Journal of Emerging Perspectives* 2: 57-71. doi: 10.36253/jep-20882

Received: April 30, 2025

Revised: September 30, 2025

Published: December 20, 2025

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

ORCID
PG: 0000-0002-8441-618X

Notes on a sociology of generative knowledge: Human-AI epistemic stewardship

Paolo Granata

University of Toronto, Canada
Email: paolo.granata@utoronto.ca

Abstract. This article develops a sociology of generative knowledge for the age of AI. It treats contemporary systems as hybrid actors within sociotechnical networks and reframes classical anchors – Mannheim's situatedness, Merton's norms, and Latour's distributed agency – for model-mediated inquiry. Generative knowledge is defined as knowledge organized to produce further knowledge through iterative, tool-mediated, and socially embedded processes. On this basis, the paper advances epistemic stewardship as a practical orientation that sustains human agency while harnessing computational acceleration. Stewardship is operationalized through transparency-by-design, provenance and traceability, calibrated trust, independent verification and red teaming, structured challenge routines, inclusivity in data and participation, and proportional delegation to machines. The account clarifies gains in discovery and education alongside risks from opacity, automation bias, feedback loops, and cognitive drift, and it specifies institutional reforms: standardized disclosure artifacts, replication triggers for AI-assisted claims, revised authorship taxonomies, and equitable access to compute, benchmarks, and community-governed datasets. A research agenda follows, calling for comparative evaluations of stewardship designs, longitudinal studies of hybrid practice, and field-specific protocols that link evaluation to adoption thresholds. The result is a framework that integrates social theory with design and governance guidance, aiming to convert acceleration into certified advance while keeping responsibility legible and contestable.

Keywords: generative knowledge, epistemic stewardship, human-AI co-creation, distributed agency, actor-network theory, Mertonian norms, automation bias, situated knowledge, provenance tracking, calibrated trust, communal verification, epistemic literacy.

1. INTRODUCTION

The present moment compresses decades of speculation about artificial intelligence into an everyday epistemic fact: machines now participate, visibly and pervasively, in the making of what we take to be knowledge.

Rather than rehearse the exhausted question of whether machines "think" (Turing 1950), I propose generative knowledge (Granata 2026) as the analytic lens through which to apprehend AI's contribution to inquiry. Distinct from computational epistemology, which centers on the logic of algorithmic inference, and broader knowledge co-production, which emphasizes collaborative social dynamics, generative knowledge defines a specific sociotechnical mechanism. It is the organization of inquiry wherein AI outputs function not as representations of settled fact, but as provisional scaffolds that compel further, iterative human verification and hypothesis formation. In this view, knowledge is defined by its capacity to catalyze a recursive cycle of questions, rather than its status as a static retrieval. From this vantage, AI is neither a mere amplifier of human cognition, or a mere "epistemic enhancer" (Humphreys 2004, p. 6), nor a disembodied oracle (Burrell 2016); it is an actor situated within networks of practice, instrumentation, and norms, helping to open new problem spaces, accelerate hypothesis formation, and reorganize pathways of validation.

The shift matters because the locus of epistemic labor is moving. Large language models, data-intensive pipelines, and agentic interfaces do not simply retrieve or recombine existing material; they instantiate new procedures of search, abduction, and pattern conjecture that reconfigure the division of epistemic functions between humans and machines. If, following Mannheim (1954), knowledge is always situated, and if, following Merton (1942), its public warrant depends upon communal norms and institutions, then the arrival of generative AI systems forces a recalibration of both situatedness and warrant. Where are these systems situated, and by whom? What counts as certified knowledge when inference, suggestion, and synthesis are distributed across human judgment, computational heuristics, and infrastructural stacks? And how do we preserve the virtues of critical scrutiny when fluency and speed threaten to become surrogates for understanding?

This article advances three central claims. First, AI should be approached as a hybrid participant in knowledge networks whose agency is derivative yet consequential: derivative because it remains scaffolded by human purposes, data ecologies, and institutional rules; consequential because its operations systematically shape what is thinkable, searchable, and actionable. Second, the concept of generative knowledge clarifies both the "bright side" and the "dark side" of this hybridity by directing attention to the recursive cycle of question → generation → evaluation and to the conditions under which that cycle yields reliable novelty rather than persuasive error. Third, meeting AI's dual potential requires an affirmative practice I call *epistemic stewardship*: a cultivated stance, at once normative and procedural, that sustains human agency, mandates transparency and traceability, and institutionalizes independent verification where computational capability now expands the frontier of inquiry.

What follows is organized accordingly. After providing a methodological note to clarify scope conditions of the work presented, I first reconstruct the historical and theoretical ground – Mannheim on social situatedness and Merton on the ethos and governance of science (Jasanoff 2004) – to establish a baseline for legitimacy and communal warrant in AI-mediated inquiry. I then recast AI, via Latour's actor-network sensibilities (1993; 2005), as a hybrid actor that redistributes agency, authority, and accountability across humans, models, data, and institutions. With this frame in place, I define generative knowledge and specify its principles, before turning to the anatomy of human-AI co-creation and the calibrated trust it requires. Subsequent sections develop the dual potential: acceleration and access on the one hand; opacity, automation bias, and cognitive drift on the other. I then elaborate epistemic stewardship into principles and consider institutional reforms in science, education, and knowledge economies.

2. METHODOLOGICAL NOTE AND SCOPE CONDITIONS

This article is a work of analytic reconstruction and conceptual synthesis, informed by selective engagement with the classical canon in the sociology of knowledge (Mannheim 1954), the sociology of science (Merton 1942), and science and technology studies (Latour 1993), alongside contemporary scholarship on machine learning (Bender et al. 2021; Bommasani et al. 2021; Nicolae 2025), human-computer interaction (Amershi et al. 2019; Shneiderman 2020), and responsible AI (Burrell 2016; Shneiderman 2020; Crawford 2021; Novelli, Taddeo, and Floridi 2024). The argument develops abductively: it moves between theoretical commitments – Mannheim's social situatedness of knowing, Merton's norm-governed certification of science, and Latour's distributed agency of actor-networks – and observations drawn from recent deployments of generative and predictive systems in research, education, and creative practice (Collins 2025). Sources include peer-reviewed studies of AI-assisted discovery and evaluation, policy documents and standards proposals, and methodological reports on dataset and model documentation.

Several assumptions underwrite the inferences and delimit their scope. The analysis presumes that a) AI systems implicated in knowledge production are not purely oracular black boxes (Burrell 2016; Pasquale 2015) but can be situated within workflows that provide at least partial access to prompts, parameters, datasets, and evaluation metrics (Humphreys 2009); b) that institutions retain the capacity to enforce disclosure and verification norms (Jasanoff 2003); and c) that human participants possess or can be taught the epistemic virtues (Zagzebski 1996; Baehr 2011) – curiosity, open-mindedness, thoroughness – required for calibrated trust (Lee et al. 2025). The analysis further assumes that the communities in question are oriented toward truth-tracking practices recognizable as communal verification, even where claims are fallible and provisional (Kitto et al. 2023). Under these conditions, the claims advanced here – that AI functions as a hybrid actor in epistemic networks and that epistemic stewardship can convert acceleration into reliable knowledge – are most likely to hold.

The limits of inference are equally important. The argument does not generalize to settings where provenance is structurally inaccessible – highly proprietary platforms such as OpenAI that forbid audit, national security enclaves that block disclosure, or data regimes that preclude any form of external inspection (Pasquale 2015). Nor does it presume transferability to domains where ground truth is intrinsically unsettled and there are no shared standards for adjudicating claims, or to political environments in which institutional independence and adversarial review are compromised (Douglas 2009). Low-resource contexts without adequate compute, connectivity, or linguistic coverage will require adaptations that this article only gestures toward. The present contribution should thus be read as a programmatic map – conceptually thick, empirically suggestive, and normatively oriented – rather than as a census of practice.

3. HISTORICAL AND THEORETICAL GROUNDING

The sociology of knowledge is the sustained inquiry into the social conditions under which knowing becomes possible, credible, and efficacious (Nicolae 2025; Fuchs 2025). It asks how forms of thought arise within patterned relations, how those relations set limits and create opportunities for thinking, and how, in turn, knowledge feeds back into and reorganizes social life. From its classic formulations to its contemporary developments (Babich 2017; Schnettler, Knoblauch, and Raab 2017), the field treats ideas as historically situated, institutionalized, and contested phenomena rather than as free-floating abstractions.

Any attempt to place artificial intelligence within the sociology of knowledge (Collins 2025) must begin, almost inevitably, with Mannheim's insistence that cognition is socially situated (Mannheim 1954; Haraway 1988; Jasanoff 2004). Knowers are not detachable Cartesian points but historically embedded participants whose outlooks are patterned by position, interest, and milieu. For Mannheim, the very possibility of critique – his sociology of knowledge – turns on recovering the conditions under which ideas are formed, circulated, and stabilized. In this light, AI systems inherit rather than transcend situatedness: their training corpora sediment selective histories of discourse (Ferrara 2024); their objective functions encode institutional priorities; their outputs emerge within regimes of use that privilege some audiences, languages, and problems over others. To the extent that AI appears to "hallucinate," to rationalize patterns as if they were realities (Huang et al. 2024), it reenacts – at computational scale – the perennial sociological lesson that representations are never innocent (Ji et al. 2023). The relevant question, then, is not whether AI is biased or epistemically neutral, but how its forms of situatedness are rendered visible, governed, and made answerable to communities of inquiry.

Merton's reconstruction of the scientific ethos provides the complementary half of this evaluative baseline. If Mannheim directs attention to the social conditioning of knowledge, Merton articulates the institutional safeguards through which a public can warrant claims despite, and indeed because of, such conditioning. Universalism, communalism, disinterestedness, and organized skepticism were never mere moral ornaments; they described a set of norms and reward structures that tied credit to disclosure, novelty to replication, and prestige to vulnerability under critique. Under this regime, "legitimate knowledge" is not what a solitary mind believes to be true, but what a community, operating with regulated openness and contestability, can provisionally certify (Humphreys 2009; Kitchin 2014). Citation, authorship, priority rules, and peer review are not ephemera: they are the scaffolding by which facts are stabilized, contributions attributed, and error made corrigible.

When these two traditions are read together, a baseline for evaluating AI's role in knowledge comes into view. Legitimacy requires that the situatedness of AI systems be mapped and governed, while warrant requires that their contributions be rendered commensurable with communal procedures of verification (Collins 2025; Hosseini et al. 2025). Training data must be treated as historically contingent archives; model architectures and prompts as methodological choices; inference traces and uncertainty estimates as disclosures analogous to methods sections. The credit economy must adapt, too (Bauchner and Rivara 2024): if models participate in hypothesis generation, synthesis, or measurement, then authorship and acknowledgment practices must specify which epistemic functions were human-initiated, which were machine-augmented, and how responsibility was retained for interpretation and error. Absent these adjustments, acceleration risks decoupling from certification, and fluency from reliability.

Institutional norms matter for AI-supported discovery because they convert the sociological facts of production into epistemic virtues. Without disclosure requirements, black-box models substitute authority for argument (Balasubramaniam et al. 2023), without communal access to data provenance and evaluation pipelines, replication devolves into ritual, and without incentives for adversarial testing and independent auditing, automation bias can colonize judgment under the guise of efficiency (Horowitz and Kahn 2024; Alon-Barkat and Busuioc 2023). The lesson is not to resurrect a nostalgic picture of science as a vacuum-sealed republic of letters, but to extend the core Mertonian commitment – organized skepticism under conditions of shared scrutiny – into the computational stack. In practical terms, this means treating dataset curation, model documentation, and uncertainty communication as first-order scholarly practices; it means redesigning review to include methodological probing of machine-mediated steps; it means aligning reward structures with transparency and corrigibility rather than mere throughput.

Thus construed, "legitimate knowledge" in AI-mediated inquiry is whatever survives the conjoint tests of situatedness-awareness and norm-governed certification. In other words, knowledge whose genealogies are traceable, whose methods are inspectable, whose claims are reproducible, and whose attributions are fair. Mannheim reminds us that every claim arrives with a social history; Merton teaches how such claims can be transformed – through institutions – into publicly warranted findings. Hence, any sociology of generative knowledge adequate to the present must hold these insights together: it must refuse the fiction of disembodied intelligence (Bender et al. 2021) while renewing the procedures by which novelty becomes trustworthy (Nosek et al. 2015).

4. FROM TOOLS TO HYBRID ACTORS

To treat AI as mere instrument, a tool, or a service is to miss the sociological transformation unfolding in the architectures of inquiry. It is rather an interface for knowledge, an active enabler in our knowledge ecosystem (Granata 2026, p. 51). Latour's Actor-Network Theory (Latour 2005) counsels against a priori partitions of subject and object; it asks us instead to follow the associations through which heterogeneous elements – humans, codes, datasets, platforms, journals, funding agencies – compose provisional unities capable of acting. In this register, contemporary AI systems are best understood as hybrid actors (Rammert 2012; Peverini 2024): they neither replace human agency nor simply extend it, but reconfigure it by introducing new mediations, translations, and points of leverage. A model's output is not the voice of an oracle, nor a neutral mirror; it is an inscription produced by an endless chain of operations that binds curators, annotators, engineers, maintainers of compute clusters, benchmark designers, reviewers, and users. When such chains hold, the model "acts" in the sense Latour intends: it alters the state of affairs by mobilizing a network that makes certain moves easier and others harder.

Agency in these networks is thus distributed and composite (Rammert 2008; Callon 1987). The question is not who acts, but how action is made possible and shaped by alignments across the network. For instance, a preprint research paper that credits a model with discovering a candidate mechanism is also crediting the training corpus that tuned its priors, the prompts and hyperparameters that channeled its search, the laboratory protocols that constrained verification, and ultimately the editorial conventions that admitted its claims to public view. Each segment inscribes interests into the next, narrowing or widening the corridor of possible outcomes. Human investigators initiate questions, frame problem spaces, and interpret results; AI components surface patterns at scales and speeds unavailable to unaided cognition; data infrastructures stabilize reference by standardizing formats and ontologies; institutional gatekeepers translate provisional outputs into certified contributions (Bowker and Star 1999). Agency is shared because the capacity to make a difference inheres in the alignments, not in any isolated node.

Authority, on this view, is not a property but an achievement: it accrues to those configurations that reliably transform speculation into findings under regimes of scrutiny (Alvarado 2023). A model acquires epistemic authority (Hauswald and Schmechtig 2025) only insofar as its lineage is documented, its performance is characterized across relevant distributions, its failures are legible, and its outputs are rendered auditable within existing grammars of justification. Benchmarks, ablation studies, uncertainty reports, and external validations are not technical niceties; they are the inscriptions that translate machine suggestions into claims intelligible to communities of practice (Hauswald 2025). The journal's methods section, the lab's notebook, the repository's metadata, or the standardized documentation frameworks for machine learning models together function as the mediating apparatus through which authority becomes durable – contingently, revisably, but with enough stability to guide further work.

Network mapping clarifies accountability because it renders visible the sites where discretion is exercised and where error can be introduced, detected, and repaired (Novelli, Taddeo, and Floridi 2024). If an AI-assisted analysis misleads, the question is not simply whether "the model failed," but which alignment failed: Was the dataset unrepresentative? Were spurious correlates smuggled in by a scraping pipeline? Did prompt templates prime a particular explanation? Did review procedures waive independent replication under the pressure of novelty? By tracing associations, one can allocate responsibility without collapsing into either scapegoating the machine or absolving the human. Accountability becomes a matter of redesigning linkages – tightening provenance capture, instituting counter-prompting rituals, diversifying benchmark suites, staging adversarial audits – so that the network resists drift toward convenience and reasserts its capacity for correction.

Seen from this vantage, the shift from tools to hybrid actors does not anthropomorphize machines; it sociologizes inquiry. It instructs us to replace the mythology of solitary genius, human or artificial, with a careful account of how collectives act and generate new knowledge through technical mediators (Granata 2026). It also provides a practical agenda. To the extent that we can specify the inscriptions required for authority, the alignments that distribute agency well, and the traces that enable responsibility to be shouldered rather than shirked, we move from vague celebration or alarm toward a workable stewardship of human-AI knowledge production.

5. DEFINING GENERATIVE KNOWLEDGE

Generative knowledge names a way of organizing inquiry so it repeatedly produces further, credible knowledge. Knowledge here is not a ledger of certified statements but a poised activity that, once initiated, reliably propagates further inquiry. It is *poietic* rather than merely *mimetic* (Granata 2026, p. 6): oriented to making, re-making, and reorganizing the very problem-spaces it addresses. What renders knowledge "generative" is its capacity to convert a single episode of asking and answering into a widening cycle – new questions, new representations, new instruments, new forms of coordination – such that the horizon of what can be credibly asked and done is enlarged rather than merely filled. In that arrangement, the measure of generativity (p. 28) is not novelty for its own sake but the sustained production of further, credible knowledge – an ever-broadening horizon of inquiry that remains tethered to practice, nourished by shared epistemic virtues, and open to transformation.

Several interlocking principles give generative knowledge its shape (p. 8). First, it is iterative (p. 9): results are not endpoints but scaffolds for subsequent moves, with each cycle reworking premises, representations, and measures. Second, it is instrumentally mediated (p. 9): tools do not merely accelerate operations; they constitute the very kinds of perceptions and inferences that can be made, from microscopes and model organisms to corpora, embeddings, and simulators. Third, it is socially situated (p. 20): claims travel only insofar as they are taken up, challenged, and stabilized within communities whose norms of credit, critique, and care furnish the selective pressure that distinguishes fecund lines of work from dead ends. Overall, generative knowledge is creative in a thick sense (p. 32): it recombines concepts, techniques, and exemplars to yield forms that were not merely latent but unknowable absent the recombination.

When AI participates in the cycle question → generation → evaluation, these principles are rearticulated (p. 208). Question formation becomes a site of co-design: prompts, retrieval templates, and representational choices act as boundary objects through which human aims are translated into machine-searchable intents. Generation becomes a plural activity: models propose patterns, counterfactuals, or design candidates at scale, while humans supply domain constraints, analogies, and value-laden desiderata that steer the search away from spurious convenience toward epistemic salience. Evaluation expands beyond correctness checks to include robustness across distributions, legibility of failure modes, sensitivity to measurement error, and fitness for communal uptake, with independent verification pipelines acting as governors on the accelerating engine. The resulting spiral (Nonaka and Takeuchi 1995) is not a handoff but a trajectory in which responsibilities and affordances are reallocated while the normative load-bearing structures – explanation, critique, replication – remain decisive.

A definition adequate to this context can therefore be stated plainly: generative knowledge is knowledge organized so that its validated outputs function as inputs to further inquiry, under conditions that render the transitions – problem formation, tool use, judgment – auditable and corrigible. It asks investigators to externalize their questions in forms that can be iterated upon, to treat instruments as partners whose biases must be characterized rather than ignored, to cultivate habits of calibrated trust that neither romanticize nor demonize machine suggestions, and to embed evaluation in communities capable of dissent and repair. Under these conditions, AI becomes a vector of generativity (Granata 2026, p. 208) not because it "thinks," but because it helps arrange the world – conceptually, materially, organizationally – so that thinking can advance with accountability.

6. EPISTEMIC AGENCY AND HUMAN-AI CO-CREATION

Epistemic agency in human-AI inquiry is not a monolith but an interplay of initiating, representing, judging, and trusting (Vaccaro 2024; Wu 2025; Lee 2025). The starting impulse – the selection and framing of questions – ought to remain a human prerogative, because it is here that values, purposes, and world-making commitments are set. Questions are never innocent; they delimit what will count as a relevant signal, which outcomes will be celebrated as discovery, and which risks will be tolerated in pursuit of it. Closely allied to initiation is the work of representation: deciding which vocabularies, data ontologies, and modeling choices are apt for a domain. Machines can learn representations; they cannot, unaided, choose what is worth representing or which abstractions are licit given ethical and disciplinary constraints. Finally, evaluative sovereignty must be retained by human communities. Setting criteria of adequacy, weighting kinds of error, and adjudicating tradeoffs among accuracy, interpretability, and social consequence are normative acts that derive their legitimacy from communal deliberation and global ethical frameworks rather than computational performance (Jobin, Ienca, and Vayena 2019).

Within these guardrails, AI can augment three families of epistemic function. First, it can extend perception by widening the search space, surfacing weak regularities, and generating counterfactuals that would be prohibitively expensive to enumerate by hand (Granata 2026, p. 211). This is cognitive extension in the literal sense: new vistas of pattern and possibility become visible through tool-mediated attention (Alvarado 2023). Second, it can shoulder forms of cognitive offloading that free scarce human capacities for tasks where judgment is decisive: preprocessing unruly corpora, constructing baselines, maintaining reproducible pipelines, and monitoring distributional drift (Gerlich 2025). Offloading here is not abdication but a reallocation of effort toward sense-making and critique. Third, it can act as a generator of candidate inferences – hypotheses, designs, explanations – whose value is realized only insofar as they enter human-governed evaluation loops (Fügener 2022). In each case, augmentation does not erase responsibility; it rearranges it by making the handoffs explicit and accountable.

Trust, therefore, must be cultivated rather than presumed. Above all, calibration is the operative virtue (Lee et al. 2025). It asks that confidence track evidence about model provenance, training distributions, and historical performance under conditions similar to those at hand. Calibrated trust is sustained by habits and infrastructures: keeping genealogies of data and models so that lineage is auditable; exposing uncertainty in forms that invite scrutiny rather than conceal it; staging adversarial tests and ablations that reveal failure modes before deployment; and maintaining independent verification channels that can say "no" when the allure of speed tempts a premature "yes." (Koivisto, Pekkarinen, and Hiekkanen 2023). Crucially, calibrated trust is dynamic (Kitto et al. 2023), revised upward when a system proves reliable under variation and downward when it surprises us in ways that are epistemically or ethically costly. The aim is not to find a once-for-all level of faith in machines but to align reliance with the moving boundary of demonstrated competence (Ferrario 2024).

A typology of roles clarifies how these capacities combine in practice. At times the system functions as instrument, akin to a microscope, making latent structure inspectable while leaving interpretation to investigators (Alvarado 2023); here the human is observer-critic, and the locus of authorship remains squarely with the community that sets up, sanity-checks, and situates the readout. In other episodes, such as large language model interactions, the system acts as explorer, proposing traversals of a problem space – parameter sweeps, synthetic exemplars, or design blueprints – that humans winnow and steer; the human becomes navigator-curator, allocating attention and enforcing domain constraints in human-AI collaboration (Chiang et al. 2024). Elsewhere the system plays the role of interlocutor, a conversational foil that helps externalize tacit assumptions, elicit counterexamples, or pressure-test arguments; the human is theorist-editor, using dialogue to refine claims, where the whole process is driven by epistemic vigilance (Sperber et al. 2010). Finally, there are moments when the system serves as sentinel, continuously auditing pipelines for drift, leakage, or reproducibility regressions; here the human is steward-adjudicator, deciding when the alarms justify intervention. Across these role configurations, the distribution of agency is patterned rather than symmetrical, and epistemic authority accrues where initiation, representation, and evaluation are anchored in accountable human practices (Hauswald 2025).

What follows is an epistemic vigilance and trust commensurate with a Generative Critical Thinking Framework (Granata 2026, p. 109). Humans retain the agenda-setting and value-laden decisions that define what knowledge is for; machines extend reach, compress routine, and multiply candidate moves; communities enforce the norms that convert speed into reliability. When these elements are aligned, co-creation does not dilute human agency; it deepens it by placing judgment where it matters most and by making the conditions of trustworthy reliance explicit, inspectable, and corrigible.

7. THE "BRIGHT SIDE"

The promise of generative systems is easiest to see where inquiry is bottlenecked by scale. In data-rich sciences, models now canvass search spaces that once exceeded any plausible human enumeration, surfacing weak and multivariate regularities from genomic variation (Attwood et al. 2022) to materials phase diagrams (Merchant et al. 2023), to domains across diverse disciplines. Hypothesis generation shifts from artisanal conjecture to systematic proposal: ensembles produce ranked candidates, counterfactuals are synthesized to probe causal structure, and parameter regimes once ignored for lack of bandwidth become experimentally thinkable (Zhang et al. 2025). In design fields, iterative co-creation shortens development cycles from weeks to hours. Alternative architectures, molecular scaffolds, or policy scenarios are produced according to defined specifications, then improved through domain-specific constraints and independent testing (Canty 2025). The acceleration increases the pace of discovery, widening the range of questions that can be explored and lowering the cost of failure.

Education displays a parallel dynamic. For novices, AI serves as a supportive structure that avoids both oversimplification and mechanical repetition. It can restate ideas at suitable levels of abstraction, present exercises that increase in difficulty, and provide timely examples while holding back full solutions until the learner has shown intermediate reasoning. For advanced learners, generative systems become dialectical partners, pressing for counterexamples, suggesting literatures adjacent to a topic, and exposing tacit assumptions through challenge prompts. The result is an enlargement of participation (Kestin 2025): students with uneven preparation can enter disciplinary discourse sooner; non-dominant languages gain more immediate access to resources; small institutions without extensive libraries or laboratories can simulate exploratory analysis and prototype research questions that would otherwise be deferred.

These benefits are not automatic. Turning rapid experimentation into dependable knowledge relies on institutional systems that maintain the friction needed for verification. High-throughput discovery calls for pre-registered analytic plans, clear distinctions between exploratory and confirmatory work, and traceable links among datasets, prompts, models, and outputs within a provenance graph. Fast hypothesis generation must be supported by independent replication, through benchmarks, blinded assessments, and adversarial review to reduce the risk of compelling but false results (Yilmaz and Yilmaz 2023). Educational gains likewise require curricular redesign: assessment that rewards process over product, disclosure norms for AI assistance, and rubrics that require students to critique model outputs, reproduce results, and trace claims to sources (Urmeneta and Romero 2024).

Access gains also turn on governance of the computational commons. If only well-funded centers can train or even run state-of-the-art models, acceleration will exacerbate inequities rather than ameliorate them. Shared model repositories with documented behavior, community-maintained datasets with transparent sampling frames, and compute credits earmarked for under-resourced institutions are preconditions for genuinely widened participation (Hosseini et al. 2025). Where such conditions hold, expert communities can increase throughput without sacrificing standards: review cycles shift from policing trivialities to interrogating claims that matter; novices become contributors rather than consumers; and the ecology of inquiry absorbs speed not as a threat to judgment but as a multiplier of occasions for it.

8. THE "DARK SIDE"

However, acceleration without commensurate scrutiny engenders its own pathologies. The first is opacity, not merely as a property of model internals but as a systemic characteristic of human-machine assemblages (Yampolskiy 2024). With large language models, in particular, exact reproduction is fragile: differences in model versions, and small changes to prompts or context can all shift the output, even when the same code is used. Provenance is often weak as well: training mixes sources with unclear sampling, fine-tuning sets may be proprietary or undocumented, and many preprocessing steps go unlogged, which obscures lineage and hinders audit (González Barman et al. 2024). Under these conditions, the verification labor expected by Mertonian norms is displaced from adjudicating claims to reconstructing pipelines.

A second risk is complacency, sycophancy, and automation bias (Parasuraman and Manzey 2010; Kahn at al. 2024) that leads to the displacement of judgment by default acceptance of machine outputs (Alon-Barkat and

Busuioc 2023; Loru 2025). Authority leaks from people to interfaces: the affordances that make systems usable – autocomplete, ranked suggestions, confidence phrasing – also create a presumption that what is salient is what is true. The problem is amplified when models speak with the idiom of expertise, provide references that look canonical, or emit calibrated probabilities without calibrated uncertainty (Zhai et al. 2024; Loru 2025). In practice, busy researchers carry forward pre-filled code, pre-specified model configurations, or suggested citations because the marginal cost of interrogation is high and the institutional incentive is throughput. Over time, the adversarial posture essential to critical inquiry erodes (Parasuraman and Manzey 2010), and with it the willingness to hold open alternative frames, to search for disconfirming evidence, or to dwell in uncertainty long enough for a better question to form.

A third danger arises from feedback loops that instantiate and then amplify their own errors. Once model-produced text, images, or analyses circulate in the wild, they are scraped back into training corpora, creating recursively self-referential gradients that privilege what the model already says over what the world newly reveals (Shumailov et al. 2024). Recommendation systems trained on engagement metrics funnel attention toward outputs that are legible to models rather than informative to humans; citation engines amplify papers that fit learned templates; preprint ecosystems adopt stylistic norms that maximize model summarizability, narrowing the space of expression and, with it, the space of conjecture. In research evaluation, proxy targets – benchmarks, leaderboards, impact indicators – become goals, triggering distortions that look like progress while decoupling from the underlying phenomena (Thomas and Uminsky 2022). Verification is compromised not by a single error but by a slowly shifting baseline that makes error harder to notice.

The fourth risk is cognitive drift and deskilling. Offloading search, synthesis, and even experimental design increases short-run productivity but may attenuate the development of expert faculties that are constitutive, not ancillary, to knowledge-making: the sense of what is interesting, the practiced feel for signal amid noise, the capacity to hold contradictory models in suspension, the cultivated memory of canonical failures (Gerlich 2025). If novices learn to write, model, or interpret primarily by curating and post-editing machine proposals, they may never internalize the generative heuristics that make independent inquiry possible. The long arc of expertise depends on formative frictions – failed starts, dead ends, painstaking replications – that AI can mask or route around. A community that systematically outsources those frictions risks a thinning of judgment that will only become visible when models err on cases that look ordinary but are not.

Finally, authorship and accountability are scrambled. Hybrid writing obscures lines of intellectual contribution; image and code generation confound provenance; "model-assisted" results invite ghost authorship by systems that cannot bear responsibility, while human contributors who specified prompts, curated failures, or designed evaluation harnesses struggle to claim credit in legacy taxonomies of contribution (Lindebaum and Fleming 2024). Absent strong disclosure norms, traceable artifacts, and enforceable audit rights, the communal capacity to attribute, to challenge, and to sanction is weakened, and with it the culture of accountability on which durable knowledge depends (Resnik and Hosseini 2025). Peer reviewers cannot tell whether a derivation reflects the author's understanding or the model's fluency; editors must decide whether to accept papers whose central insight was algorithmically proposed; institutions face liability when harms flow from workflows in which agency is genuinely distributed (Bauchner and Rivara 2024).

These risks do not amount to a countermarch to pre-AI inquiry (Wu 2025). They indicate specific failure modes by which opaque assemblages, authority by interface, self-referential feedback, and skill atrophy can compromise verification regimes, corrode peer review, and attenuate the slow cultivation of expert judgment.

9. EPISTEMIC STEWARDSHIP

Epistemic stewardship names a practical orientation to inquiry under conditions of hybrid agency (Granata 2026, p. 224). It treats human-AI collaborations not as substitutions but as negotiated partnerships whose legitimacy depends on the cultivation of norms, interfaces, and routines that make reasons visible, challenge thinkable, and authority revisable. The steward's task is neither romantic defense of pre-digital craft nor uncritical accelerationism; it is the design and maintenance of procedures by which knowledge-making can be made trustworthy, revisitable, and widely shareable. What follows is a set of principles, each paired with concrete scholarly practices, that renders this stance actionable.

Transparency-by-design requires that contributions to a claim be inspectable at the level of data, model, prompt, and post-processing. In practice this means disclosure norms that travel with the work: datasets that document provenance, scope, and known failure modes; prompt ledgers and parameter manifests that make generative steps reproducible; uncertainty and confidence reporting attached to outputs rather than relegated to appendices; and authorship statements that specify where and how AI assistance shaped the result (Resnik and Hosseini 2025). Transparency here is not a metaphysical demand to "open the black box" so much as a pragmatic insistence that enough of the box be externalized to sustain critique and repair (Novelli, Taddeo, and Floridi 2024; Resnik and Hosseini 2025).

Calibrated trust (Kitto et al. 2023; Lee at al. 2025) is the disposition to align confidence with evidential strength and model competence, rather than with fluency or authority cues. It is cultivated through practices that couple machine suggestion to human justification: interfaces that require reason-giving before adoption; confidence elicitation from both humans and models, displayed side-by-side; competence profiles for models that bound their use to validated domains; and "break-glass" controls that route high-stakes decisions to slower, more expert pathways. Over time, calibrated trust is reinforced by feedback: performance dashboards that close the loop between recommendations, human overrides, and downstream outcomes.

Independent verification operationalizes Mertonian skepticism in a hybrid milieu (Alon-Barkat and Busuioc 2023). It institutionalizes second looks that are methodologically disjoint from the original pipeline: adversarial review that reconstructs the analysis with alternative models and perturbations; replication triggers that automatically flag surprising effect sizes, distributional shifts, or heavy reliance on unvetted generators; preregistration for confirmatory phases following exploratory, AI-assisted discovery; and periodic external audits of training data governance and evaluation suites (Bauchner and Rivara 2024). The aim is not to slow inquiry for its own sake but to separate acceleration from credence, ensuring that speed feeds hypothesis generation while certification remains a communal achievement.

The habit of challenge installs dissent as a design feature rather than a social afterthought. Stewards stage structured contestation through red-team/blue-team exchanges, devil's-advocate prompts that force models to surface counterevidence, and "challenge mode" interfaces that default to presenting rival explanations or counterfactuals before convergence is allowed (Thomas and Uminsky 2022). In classrooms and labs alike, the habit is reinforced through graded requirements to refute at least one high-fluency suggestion with literatures or pilot data, and through incentives that recognize error-finding and restraint – negative results, retractions, and corrections – as contributions to the knowledge commons rather than reputational blemishes (Ferrara 2024; Huang et al. 2024).

Traceability secures the chain of custody for claims, connecting a statement to the materials and transformations that made it possible. Practically, this entails provenance graphs embedded in manuscripts and data portals; cryptographic signatures and time-stamps for model versions and training sets; audit logs that record human edits, prompts, and post hoc adjustments; and interoperable identifiers for entities, processes, and environments that enable cross-lab stitching of evidence. Traceability does not presume permanence – it anticipates revision – so it pairs with versioning practices that let communities compare, roll back, and annotate changes across the lifecycle of a claim (Novelli, Taddeo, and Floridi 2024; Shumailov et al. 2024).

Inclusivity expands whose experience and expertise count in the construction and evaluation of knowledge (Haraway 1988). It is enacted by diversifying training corpora with documented consent and representation goals; instituting participatory data governance for communities most affected by model deployment; adopting credit taxonomies that name varied forms of intellectual labor, including curation, annotation, and red teaming; and lowering the computational and licensing barriers that currently gate access to powerful models and datasets. Inclusivity, on this view, is not an extrinsic virtue but an epistemic necessity: more heterogeneous inputs and evaluators yield more robust generativity and fewer blind spots (Jasanoff 2003; Crawford 2021).

Finally, proportionality – the least-automation principle – holds that the level of delegation to machines should be commensurate with model reliability, domain stakes, and the reversibility of error. Proportionality is implemented through tiered workflow checkpoints: low-stakes exploratory work can be freely machine-augmented; high-stakes or irreversible decisions require multi-expert sign-off, slower methods, and independent replication; and any shift upward in automation level demands prior demonstration of competence under conditions that mirror intended use. Proportionality turns stewardship into governance by binding enthusiasm for throughput to a politics of consequence (Ferrario 2024; Horowitz and Kahn 2024).

Taken together, these principles transform stewardship from a slogan into a practical framework. They embed transparency, calibration, verification, challenge, traceability, inclusivity, and proportionality into the everyday mechanics of inquiry – disclosure templates, confidence and competence reporting, adversarial review protocols, replication triggers, provenance tooling, participatory governance, and tiered decision rules. In doing so, they keep generativity tethered to communal norms of certification while leaving intact the speed and reach that make human-AI partnerships attractive in the first place.

10. REFRAMING THE SOCIOLOGY OF KNOWLEDGE FOR THE AI ERA

A reframed sociology of knowledge begins by returning to Mannheim with new questions. If knowledge is always socially situated, then AI does not dissolve situatedness so much as multiply its sites. Training corpora, benchmark suites, data pipelines, and interface defaults are not inert backgrounds but socially patterned environments that prefigure what can be known, by whom, and under what descriptions (Collins 2025). Mannheim's ideology critique thus expands from the analysis of class and generation to encompass infrastructures, data genealogies, and platform economics. The point is not to replace the diagnosis of social location with a technical determinism, but to show how social location is now laminated through sociotechnical substrates. Ideology survives as curation: choices about inclusion, exclusion, and weighting in model development sediment particular horizons of relevance. Under a generative knowledge regime, unmasking ideology requires provenance work – making visible the lineage of datasets, the histories of annotation, the contours of pretraining corpora – so that situatedness can be inspected rather than presumed (Schnettler, Knoblauch, and Raab 2017).

Merton's problem-space also persists, though its objects have shifted. Communalism, universalism, disinterestedness, and organized skepticism were always institutional achievements, enforced by review, priority adjudication, and replication (Merton 1942). With AI in the loop, these norms do not become obsolete; they become infrastructural requirements. Communalism now demands shareable artifacts – model architectures, data statements, prompt logs – that render hybrid reasoning inspectable. Universalism demands evaluation regimes that travel across domains and demographics rather than metrics tuned to narrow benchmarks (Thomas and Uminsky 2022). Disinterestedness becomes a governance problem at the level of compute access, platform dependencies, and conflict-of-interest disclosures for model providers. Organized skepticism must be retooled as adversarial evaluation, red-teaming, and replication triggers designed explicitly for model-assisted results. What counted as "certified knowledge" under Merton is still certified by a community, but the community must now interrogate models, data, and workflows alongside claims and methods.

Latour's lesson is that agency is distributed across humans and non-humans, and that facts stabilize when networks hold (Latour 1987, 2005). In the AI era, hybrid actors are not only instruments that extend perception; they are co-authors of representational moves, inscribing biases and affordances into the very form of hypotheses, visualizations, and texts. Actor-network terms help us see why accountability becomes a property of configurations rather than of individuals (Novelli, Taddeo, and Floridi 2024). Yet, the expansion of agency does not excuse the abdication of responsibility. Precisely because agency is distributed, authority must be layered and traceable (Wu 2025). Network mapping clarifies who or what set the priors, tuned the parameters, curated the data, selected the prompts, and accepted the outputs into the record. Without such mappings, responsibility dissipates into the ether of "the model," and organized skepticism loses its grip. Latour's compositionist impulse (Latour 1993) – to ask what worlds are being assembled – guides the design of stewardship practices that bind together actors, artifacts, and norms into accountable epistemic collectives.

Across these reinterpretations, several elements remain central. Communal verification still anchors credibility; claims acquire standing through exposure to organized critique, independent checks, and norms of disclosure (Merton 1942; Kitcher 1993; Nosek et al. 2015). Priority and credit still matter, but authorship must be disentangled from automation in ways that honor conceptual contribution, experimental craft, and curatorial labor (Resnik and Hosseini 2025). The virtues of inquiry – epistemic curiosity, open-mindedness, and intellectual thoroughness (Baehr 2011; Granata 2026) – are still the drivers of discovery, now exercised through practices of calibration, counter-prompting, and provenance inspection (Lee et al. 2025; Kitto et al. 2023). What requires rethinking is the unit of analysis and the unit of accountability. Knowing is no longer adequately described at the level of the solitary knower or even the bounded laboratory; it is a property of hybrid ensembles whose reliability depends on the design of interfaces, the transparency of data lineages, the legibility of model behavior (Burrell 2016), and the enforceability of audit (Novelli, Taddeo, and Floridi 2024).

Generative knowledge (Granata 2026), construed as knowledge that produces further knowledge through iterative, tool-mediated, and socially embedded processes, serves as the bridge between classical commitments and contemporary conditions. It reframes Mannheim's question – who knows? – as a question about the formation of epistemic ecosystems and the stewardship of their generativity. It recasts Merton's norms as design constraints for platforms, repositories, and review processes that can convert acceleration into certified advance. And it deploys Latour's analytics to make responsibility tractable within distributed agency (Latour 2005). If ideology critique once aimed to unmask the hidden hand of social interests, its complement now is infrastructure critique (Pasquale 2015): tracing how data, models, and interfaces tilt inquiry. If the sociology of science once policed priority and replication, its complement now is protocol-building: specifying disclosure, audit, and challenge routines that keep hybrid networks truth-tracking (Resnik and Hosseini 2025; Hosseini et al. 2025). The result is less a rupture with the classical frameworks than their rearticulation: a sociology of knowledge oriented to technology-mediated co-creation, committed to communal verification, and organized around the practical craft of epistemic stewardship.

11. CONCLUSION: TOWARD DURABLE HUMAN-AI KNOWLEDGE PARTNERSHIPS

This article has argued that contemporary AI systems participate in knowledge creation not as inert instruments but as hybrid actors within sociotechnical networks (Latour 1987, 2005), and that their *generativity* – understood as the capacity of knowledge to beget further knowledge through iterative, tool-mediated, and socially embedded processes – both amplifies and destabilizes established epistemic practices (Granata 2026). Read with Mannheim (1954), AI extends situatedness into the strata of data pipelines and platform economics; read with Merton (1942), it renders communal norms into infrastructural requirements; read with Latour (1987, 2005), it distributes agency across human and non-human actors whose alignments must be composed, traced, and held accountable. The upshot is not a relinquishing of verification, credit, or responsibility but their reconstruction under conditions of technological co-creation (Collins 2025). Hence the central claim: durable human-AI knowledge partnerships are possible only under a regime of epistemic stewardship that sustains reliability, equity, and human agency while leveraging computational acceleration.

If stewardship is the practical name for this reconstruction, then near-term work begins with disclosure and provenance. Researchers should treat prompts, datasets, model versions, and post-processing steps as first-class research objects, archived with the same care as code and data, and described through standardized artifacts that render hybrid reasoning inspectable (Bommasani et al. 2021; Balasubramaniam et al. 2023). Educators should reorganize curricula around epistemic literacy (Urmeneta and Romero 2024; Wu et al. 2025): students must learn to ask questions that models can meaningfully engage, to read uncertainty visualizations, to counter-prompt and adversarially test outputs, and to document the provenance of co-created work. Journals and funding agencies should adopt author contribution taxonomies that distinguish conceptual labor, curatorial labor, and computational orchestration (Resnik and Hosseini 2025); require confidence reporting and adversarial evaluation for AI-assisted findings; and institute replication triggers when claims rely substantively on model-generated content. Institutions, finally, should invest in equitable access – shared compute, open benchmark suites, community-governed data resources (Hosseini et al. 2025) – so that the capacity to participate in generative knowledge is not rationed by wealth or geography.

The policy horizon is equally concrete. Governance should move beyond generic "AI use" statements toward field-specific protocols that bind Mertonian norms to contemporary pipelines: transparency-by-design requirements for provenance capture (Balasubramaniam et al. 2023); calibrated trust practices (Lee et al. 2025; Kitto et al. 2023) that pair model suggestions with independent checks; red-team reviews embedded in peer-review workflows; and post-publication audit channels capable of version-aware critique as models and datasets evolve (Lee 2025). Credit economies must be recalibrated so that those who assemble and maintain the conditions of generativity – dataset stewards, benchmark curators, maintainer communities – are recognized as co-producers of certified knowledge (Zhang et al. 2025). Equity requires building compute commons and multilingual, culturally representative corpora (Hosseini et al. 2025) governed with community consent, lest the infrastructures that scaffold generative knowledge narrow rather than widen the horizon of inquiry.

A research agenda follows from these prescriptions. We need comparative studies that measure how different stewardship designs – provenance interfaces, uncertainty displays, challenge modes – alter the accuracy, reproducibility, and inclusivity of hybrid inquiry (Chiang et al. 2024; Vaccaro et al. 2024). We need longitudinal ethnographies of laboratories, classrooms, and editorial boards learning to live with models, tracing how habits of challenge, calibration, and documentation are built or eroded over time (Gao and Wang 2025). We need methodological work on evaluation beyond benchmarks: mixed-methods protocols that tie external validity, demographic robustness, and error characterization to concrete adoption thresholds (Thomas and Uminsky 2022). We need normative analyses of authorship and accountability under distributed agency, and institutional experiments that redistribute credit to curatorial and infrastructural labor (Resnik and Hosseini 2025). And we need global collaborations to specify culturally plural standards for consent, attribution, and access in the data regimes that feed generative systems (Hosseini et al. 2025).

The promise of generative knowledge is to convert acceleration into advance; its peril is to convert plausibility into credulity (Ji et al. 2023; Yampolskiy 2024; Huang et al. 2024; Zhang et al. 2025; Loru et al. 2025; Granata 2026). Between these lies the craft of stewardship: designing interfaces that invite scrutiny rather than deference (Horowitz and Kahn 2024), building governance that makes responsibility traceable rather than diffuse (Wu et al. 2025), cultivating pedagogies that form judgment rather than outsource it (Gerlich 2025). If Mannheim urges us to surface situatedness, Merton to organize skepticism, and Latour to compose accountable collectives, then the task ahead is to bind these insights into practice – protocols, platforms, and pedagogies that keep human agency at the center of hybrid inquiry. Durable human-AI knowledge partnerships will not emerge by default; they will be made through the daily work of researchers, educators, editors, and policymakers who insist that what is fast also be fair, that what is fluent also be falsifiable, and that what is generative remain, in the deepest sense, genuinely knowledge-making.

REFERENCES

Alon-Barkat, S. & Busuioc, M. (2023). Human-AI Interactions in Public Sector Decision Making: Automation Bias and Selective Adherence to Algorithmic Advice. *Journal of Public Administration Research and Theory*, 33(1), 153–69. https://doi.org/10.1093/jopart/muac007

Alvarado, R. (2023). AI as an Epistemic Technology. *Science and Engineering Ethics* 29(4), 26. https://doi.org/10.1007/s11948-023-00451-3

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., et al. (2019). Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Paper 3. New York: ACM. https://doi.org/10.1145/3290605.3300233

Attwood, S.W., Hill, S.C., Aanensen, D.M., Connor, T.R. & Pybus, O.G. (2022). Phylogenetic and Phylodynamic Approaches to Understanding and Combating the Early SARS-CoV-2 Pandemic. *Nature Reviews Genetics* 23(9), 547–562. https://doi.org/10.1038/s41576-022-00483-8

Babich, B. (2017). The 'New' Sociology of Knowledge. In *Hermeneutic Philosophies of Social Science*. Walter de Gruyter GmbH. https://doi.org/10.1515/9783110528374-013

Baehr, J. (2011). *The Inquiring Mind: On Intellectual Virtues and Virtue Epistemology*. Oxford: Oxford University Press.

Balasubramaniam, N., Lehmkuhl, K., Holmström Olsson, H. & Bosch, J. (2023). Transparency and Explainability of AI Systems: From Ethical Guidelines to Requirements. *Information and Software Technology* 159, 107197. https://doi.org/10.1016/j.infsof.2023.107197

Bauchner, H. & Rivara, F.P. (2024). Use of Artificial Intelligence and the Future of Peer Review. *Health Affairs Scholar* 2(5), qxae058. https://doi.org/10.1093/haschl/qxae058

Bender, E.M., Gebru, T., McMillan-Major, A. & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (FAccT '21), 610–623. https://doi.org/10.1145/3442188.3445922

Bommasani, R., Hudson, D.A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M.S. et al. (2021). *On the Opportunities and Risks of Foundation Models*. Stanford Institute for Human-Centered Artificial Intelligence (HAI), Stanford University.

Bowker, G.C. & Leigh Star, S. (1999). *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.

Burrell, J. (2016). How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms. *Big Data & Society* 3(1), 1–12. https://doi.org/10.1177/2053951715622512

Callon, M. (1987). Society in the Making: The Study of Technology as a Tool for Sociological Analysis. In *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*, Wiebe E. Bijker, Thomas P. Hughes & Trevor Pinch (Eds, pp. 83–103). Cambridge, MA: MIT Press.

Canty, R.B. & Abolhasani, M. (2025). Science Acceleration and Accessibility with Self-Driving Labs. *Nature Communications* 16, (3856). https://doi.org/10.1038/s41467-025-59231-1

Chiang, C.-W., Zixuan, L., Zhuoyan, L. & Ming Y. (2024). Evaluating Human-AI Collaboration: A Review and Methodological Framework. *arXiv preprint* arXiv:2407.19098

Collins, H. (2025). Why Artificial Intelligence Needs Sociology of Knowledge. *AI & Society* 40, 1249–1263. https://doi.org/10.1007/s00146-024-01954-8

Crawford, K. (2021). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven, CT: Yale University Press.

Douglas, H. (2009). *Science, Policy, and the Value-Free Ideal*. Pittsburgh: University of Pittsburgh Press.

Ferrara, E. (2024). Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies. *Sci* 6(1), 3. https://doi.org/10.3390/sci6010003

Ferrario, A. (2024). Justifying Our Credences in the Trustworthiness of AI Systems: A Reliabilistic Approach. *Science and Engineering Ethics* 30(55). https://doi.org/10.1007/s11948-024-00522-z

Fuchs, S. (2025). Classical sociology of knowledge and contemporary constructivism. In *Research Handbook on the Sociology of Knowledge*, F. Collyer (Ed, pp. 192-208). Edward Elgar Publishing. https://doi.org/10.4337/9781800376649.00020

Fügener, A, Grahl, J., Gupta, A. & Ketter, W. (2022). Cognitive Challenges in Human-Artificial Intelligence Collaboration: Investigating the Path Toward Productive Delegation. *Information Systems Research* 33(2), 678–696. https://doi.org/10.1287/isre.2021.1079

Gao, L. & Wang, Y. (2025). Rethinking Laboratory Life in the Age of AI. *AI & Society*, forthcoming. https://doi.org/10.1007/s00146-025-02495-4

Gerlich, M. (2025). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. *Societies* 15(1), 6. https://doi.org/10.3390/soc15010006

González B.K., Wood, N. & Pawlowski, P. (2024). Beyond Transparency and Explainability: On the Need for Adequate and Contextualized User Guidelines for LLM Use. *Ethics and Information Technology* 26, 47. https://doi.org/10.1007/s10676-024-09778-2

Granata, P. (2026). *Generative Knowledge: Think, Learn, Create with AI*. Hoboken, NJ: John Wiley & Sons.

Haraway, D. (1988). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies* 14(3), 575–599. https://doi.org/10.2307/3178066

Hauswald, R. & Schmechtig, P. (2025). New Waves in the Philosophy of Epistemic Authority and Expert Testimony: An Introduction to the Special Issue. *Social Epistemology*, May, 1–6. https://doi.org/10.1080/02691728.2025.2496203

Hauswald, R.. (2025). Artificial Epistemic Authorities. *Social Epistemology* 1-10. https://doi.org/10.1080/02691728.2025.2449602

Horowitz, M.C. & Kahn, L. (2024). Bending the Automation Bias Curve: A Study of Human and AI-Based Decision Making in National Security Contexts. *International Studies Quarterly* 68(2), sqae020. https://doi.org/10.1093/isq/sqae020

Hosseini, Mohammad, Bastian Greshake Tzovaras, Koray Karaca, and Tony Ross-Hellauer. (2025). "Open Science at the Generative AI Turn: An Exploratory Analysis of Challenges and Opportunities." Quantitative Science Studies. https://doi.org/10.1162/qss_a_00337

Huang, L. & Ting, L. (2024). A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. *ACM Transactions on Information Systems* 43(1). https://doi.org/10.1145/3703155

Humphreys, P. (2004). *Extending Ourselves: Computational Science, Empiricism, and Scientific Method*. Oxford: Oxford University Press.

Humphreys, P. (2009). The Philosophical Novelty of Computer Simulation Methods. *Synthese* 169(3), 615–626. https://doi.org/10.1007/s11229-008-9435-2

Jasanoff, S. (Ed) (2004). *States of Knowledge: The Co-Production of Science and Social Order*. London: Routledge.

Jasanoff, S. (2003). Technologies of Humility: Citizen Participation in Governing Science. *Minerva* 41(3), 223–244. https://doi.org/10.1023/A:1025557512320

Ji, Z., Lee, N., Frieske, R. et al. (2023). Survey of Hallucination in Natural Language Generation. *ACM Computing Surveys* 55(12). https://doi.org/10.1145/3571730

Jobin, A., Ienca, M. & Vayena, E. (2019). The Global Landscape of AI Ethics Guidelines. *Nature Machine Intelligence* 1, 389–399. https://doi.org/10.1038/s42256-019-0088-2

Kahn, L., Probasco, E.S. & Kinoshita R. (2024). *AI Safety and Automation Bias: The Downside of Human-in-the-Loop*. Washington, DC: Center for Security and Emerging Technology.

Kestin, G., Miller, K., Klales, A., Milbourne, T. & Ponti, G. (2025). AI Tutoring Outperforms In-Class Active Learning. *Scientific Reports* 15, 97652. https://doi.org/10.1038/s41598-025-97652-6

Kitcher, P. (1993). *The Advancement of Science: Science without Legend, Objectivity without Illusions*. Oxford: Oxford University Press.

Kitto, K., Bruza, P., Sheng Loh, B. & Boschetti, F. (2023). A Quantum Model of Trust Calibration in Human-AI Interactions. *Entropy* 25(9), 1362. https://doi.org/10.3390/e25091362

Koivisto, M., Pekkarinen, S. & Hiekkanen, K. (2023). Transparency and Explainability of AI Systems: From Ethical Guidelines to Requirements. *Information and Software Technology* 159, 107202. https://doi.org/10.1016/j.infsof.2023.107202

Latour, B. (1987). *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, MA: Harvard University Press.

Latour, B. (1993). *We Have Never Been Modern*. Cambridge, MA: Harvard University Press. Originally published (1991).

Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.

Lee, D., Pruitt, J., Zhou, T., Du, J. & Odegaard, B. (2025). Metacognitive Sensitivity: The Key to Calibrating Trust and Optimal Decision Making with AI. *PNAS Nexus* 4(5), pgaf133. https://doi.org/10.1093/pnasnexus/pgaf133

Lindebaum, D. & Fleming, P. (2024). ChatGPT Undermines Human Reflexivity, Scientific Responsibility and Responsible Management Research. *British Journal of Management* 35(2), 566–575. https://doi.org/10.1111/1467-8551.12781

Loru, E., Nudo, J., Di Marco, N., Santirocchi, A., Atzeni R., Cinelli, M., Cestari V., Rossi-Arnaud, C. & Quattrociocchi, W. (2025). The Simulation of Judgment in LLMs. *PNAS* 122(42): e2518443122. https://doi.org/10.1073/pnas.2518443122

Mannheim, K. (1954). *Ideology and Utopia: An Introduction to the Sociology of Knowledge*. New York: Harcourt Brace.

Merchant, A., Batzner, S., Schoenholz, S.S., Aykol, M. et al. (2023). "Scaling Deep Learning for Materials Discovery." *Nature* 624 (7990): 80–85. https://doi.org/10.1038/s41586-023-06735-9

Merton, R.K. (1942). The Normative Structure of Science. In *The Sociology of Science: Theoretical and Empirical Investigations*, N.W. Storer (Ed, pp. 267–278). Chicago: University of Chicago Press, (1973).

Nicolae, S.. (2025). The 'New' Sociology of Knowledge and the Sociology of Valuation. In *The Routledge International Handbook of Valuation and Society*, T. Peetz, A.K. Krüger & H. Schäfer (Eds, 1st ed., vol. 1, 79–90). Routledge. https://doi.org/10.4324/9781003229353-9

Nonaka, I. & Takeuchi, H. (1995). *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*. New York: Oxford University Press.

Nosek, B.A., Alter, G., Banks, G.C. et al. (2015). Promoting an Open Research Culture. *Science* 348 (6242): 1422–1425. https://doi.org/10.1126/science.aab2374

Novelli, C., Taddeo, M. & Floridi, L. (2024). Accountability in Artificial Intelligence: What It Is and How It Works. *AI & Society* 39(4), 2331–2343. https://doi.org/10.1007/s00146-023-01635-y

Parasuraman, R. & Manzey, D.H. (2010). Complacency and Bias in Human Use of Automation: An Attentional Integration. *Human Factors* 52(3), 381–410. https://doi.org/10.1177/0018720810376055

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

Peverini, P. (2024). *Bruno Latour in the Semiotic Turn: An Inquiry into the Networks of Meaning*. Berlin: Springer Nature.

Rammert, W. (2012). Distributed Agency and Advanced Technology. Or: How to Analyze Constellations of Collective Inter-Agency. In *Agency without Actors? New Approaches to Collective Action*, J.-H. Passoth, B. Peuker & M. Schillmeier (Eds, pp. 89–112). London: Routledge.

Resnik, D.B. & Hosseini, M. (2025). Disclosing Artificial Intelligence Use in Scientific Research and Publication: When Should Disclosure Be Mandatory, Optional, or Unnecessary? *Accountability in Research*, March, 1–13. https://doi.org/10.1080/08989621.2025.2481949

Schnettler, B., Knoblauch, H. & Raab, J. (2017). The 'New' Sociology of Knowledge. In *Hermeneutic Philosophies of Social Scienc*e, B. Babich (Ed, pp. 237–266). Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110551563-013

Shneiderman, B. (2020). "Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy." *International Journal of Human-Computer Interaction* 36 (6): 495–504. https://doi.org/10.1080/10447318.2020.1741118

Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R. & Gal, Y. (2024). AI Models Collapse When Trained on Recursively Generated Data. *Nature* 631, 755–59. https://doi.org/10.1038/s41586-024-07566-y

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G. & Wilson, D. (2010). Epistemic Vigilance. *Mind & Language* 25(4), 359–393. https://doi.org/10.1111/j.1468-0017.2010.01394.x

Thomas, R.L. & Uminsky, D. (2022). Reliance on Metrics Is a Fundamental Challenge for AI. *Patterns* 3(5), 1–12. https://doi.org/10.1016/j.patter.2022.100476

Turing, A.M. (1950). Computing Machinery and Intelligence. *Mind* 59(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

Urmeneta, A. & Romero, M. (Eds) (2024). *Creative Applications of Artificial Intelligence in Education*. Springer Nature.

Vaccaro, M., Almaatouq, A. & Malone, T. (2024). When Combinations of Humans and AI Are Useful: A Systematic Review and Meta-Analysis. *Nature Human Behaviour* 8, 2293–2303. https://doi.org/10.1038/s41562-024-02024-1

Wu, J.-Y, Lee, Y.-H., Sing Chai, C. & Tsai, C.-C. (2025). Strengthening Human Epistemic Agency in the Symbiotic Learning Partnership with Generative Artificial Intelligence. *Educational Researcher* 54(6), 358–368. https://doi.org/10.3102/0013189X251333628

Yampolskiy, R.V. (2024). *AI: Unexplainable, Unpredictable, Uncontrollable*. CRC Press.

Yilmaz, R. & Yilmaz, F.G.K. (2023). The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. *Computers and Education: Artificial Intelligence* 4, 100147. https://doi.org/10.1016/j.caeai.2023.100147

Zagzebski, L.T. (1996). *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*. Cambridge: Cambridge University Press.

Zhai, C., Wibowo, S. & Li, L.D. (2024). The Effects of Over-Reliance on AI Dialogue Systems on Students' Cognitive Abilities: A Systematic Review. *Smart Learning Environments* 11(1), 28. https://doi.org/10.1186/s40561-024-00316-7

Zhang, Y. & Zenil, H. (2025). Exploring the Role of Large Language Models in the Scientific Method. *Nature Reviews Methods Primers* 5, 19. https://doi.org/10.1038/s44387-025-00019-5

JEP

Firenze University Press
https://oaj.fupress.net/index.php/jep

OPEN ACCESS

Citation: Ciarini, A. (2025) The era of platform work in the care Italian care system. *Journal of Emerging Perspectives* 2: 73-81. doi: 10.36253/jep-20883

Received: April 30, 2025

Revised: September 30, 2025

Published: December 20, 2025

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

The era of platform work in the care Italian care system

Andrea Ciarini

Abstract. This article examines the relationship between welfare regimes, labour-market regulation, and the rise of digital platforms in the care sector. It argues that the diffusion of digital platforms reflects and reproduces national institutional configurations rather than transforming them uniformly. Starting from the characteristics of the Italian welfare system and its regulatory mechanisms, this paper aims to shed light on the role of new work platforms in the sector and to examine the implications for regulatory mechanisms. The question we ask is whether and to what extent a system strongly oriented towards informal family care is modified by the entry of actors such as platforms, which hypothetically tend to formalise the relationships between families and providers. Second, drawing on the Italian familistic model, we assess whether and how platforms and digital technologies contribute to new segmentation mechanisms in the care market, including by channelling unmet demand to present themselves as innovative providers.

Keywords: welfare, digital platforms, care.

1. INTRODUCTION

Artificial Intelligence (AI) and digital platforms have emerged as two of the most transformative forces of our era. Their development is reshaping a wide range of sectors, including local care services. Platforms have been proposed as a response to the uncertainty of navigating informal markets and as a viable solution to the “care crisis” (Ticona et al., 2018). The diffusion of platforms in the care sector signals an upward trend in the regulation – albeit uneven – of labour relations in care. Digital transformation brings new opportunities but also new risks. On the one hand, there is a risk of disintermediating tangible relations and fuelling new forms of exclusion; on the other, platforms can help identify emerging needs and client segments, and sometimes even constitute new markets. According to van Dijck et al. (2018), platforms represent a silent transformation that began before the “platform era” itself. In this sense, they are the latest phase in a broader trend toward flexibilisation, in which processes and decision-making resources are increasingly distributed and structured beyond the formal boundaries that define a sector or industry. A platform is not merely a technological infrastructure; it is a socio-technical system that uses technology to connect people, organisations, and resources in an interactive ecosystem that structures relationships. In this respect, platforms constitute a new organisational paradigm that takes on an instituting role, consolidating cultures, representations, and organisational routines over time (Bratton, 2015). As Kornberger et al. (2017) note, one salient dimension concerns platforms as evaluative infrastructures that decentralise control while centralising power over stakeholders – notably workers.

The implications for workers are highly ambivalent. In the care industry, platforms deploy rating and review systems, together with profiles and background checks, to ensure a degree of formalisation and to increase transparency by displaying workers' qualities and skills in a standardised, comparable way. The aim is to provide clients with a large pool of flexible, readily available workers – often exacerbating the deeply unequal power relations that have long characterised domestic work (Ticona & Mateescu, 2018; Hunt & Machingura, 2016). Many authors have criticised the rise of digital labour platforms because of asymmetric power relations and negative consequences for working conditions and social protection. Moreover, most platforms define themselves as labour-market intermediaries matching supply and demand, rather than as employers or employment agencies. Thus, although they offer tools to record hours, process payments, and handle other bureaucratic functions – and present the formalisation of a largely informal sector as a key value proposition for clients and policymakers – they frequently dissolve the employment relationship into a nexus of private contracts and user agreements (Tomassetti, 2016).

Against this backdrop, the term "uberisation" of care has been used to describe a phenomenon in which new digital jobs and digital dumping crowd out formal employment (Kyoung-Hee & Rhodes, 2019). As Flanagan (2019) observes, by establishing opaque and largely unchallengeable "rules," platform algorithms become primary mechanisms for disciplining workers, producing a "paradigm shift" in worker control "from one of dyadic to structural domination" (p. 65). Historically, domestic workers laboured under informal conditions of dyadic domination (individual employers/families). Today's domestic-labour platforms enforce a semi-formalised regime of structural domination: workers may leave the household at will while remaining captive to the platform's mechanisms to remain employable. Conversely, platform work can improve conditions in sectors where informality and illegality are the rule (Ticona & Mateescu, 2018). These platforms may also open opportunities for vulnerable groups – such as minorities and women – to enter the labour market (Thieme, 2018). However, these effects vary not only by platform but also by socio-economic context, national regulation, and employment status (e.g., employee vs. self-employed).

Starting from the features of the Italian welfare system and its regulatory mechanisms, this paper examines the role of new work platforms in the care sector and their implications for regulation. We ask whether and to what extent a system strongly oriented towards informal family care is transformed by the entry of platforms that, in principle, tend to formalise relationships between families and providers. First, we probe the role of care platforms as new intermediaries: can digital platforms facilitate the emergence from irregular work, and do they create spaces of negotiation and agency for workers? Second, drawing on the Italian familistic model, we assess whether and how platforms and digital technologies contribute to new segmentation mechanisms in the care market, including by channelling unmet demand to present themselves as innovative providers.

The article adopts a political economy approach to the comparative analysis of platformisation of care sector. Rather than relying on primary data collection, the study is based on the systematic analysis of secondary sources, including policy documents, legislative and regulatory materials, existing mappings of welfare and care platforms, and recent empirical studies on digital labour platforms in Italy. In particular, the analysis draws on national and European policy reports, academic literature, and published qualitative research on care platforms and domestic work. This methodological choice reflects the exploratory and theory-building aim of the article, which seeks to situate the emergence of care platforms within the long-term institutional configuration of the Italian welfare regime, rather than to provide an exhaustive or representative mapping of platforms or workers. As a consequence, potential selection bias must be acknowledged: available data tend to over-represent more visible, formally registered, and digitally structured platforms, while smaller, informal, or hybrid intermediaries may remain underrepresented. However, this limitation does not undermine the analytical focus of the article, which is to interpret how platform-based intermediation is shaped by, and in turn reproduces, existing patterns of familism, informality, and weak labour regulation. The comparative references to France and Germany serve a heuristic and analytical function, providing a contrasting institutional backdrop against which Italian specificities can be more clearly identified. Overall, the methodology is designed to capture structural and institutional dynamics, rather than individual-level outcomes, in line with the article's theoretical orientation. Against this background, the analysis introduces the concept of "digital familism", as a concept suited to capture the specific configuration of care platforms in a familistic welfare regime. Rather than representing an epistemological or institutional rupture, platforms adapt to pre-existing arrangements in which families remain the primary coordinators, financiers, and risk-bearers of care. Analytically, digital familism can be operationalised along three interrelated dimensions. First, the persistence of family responsibility for organising and financing care, with platforms acting as auxiliary intermediaries rather than as substitutes for public provision or employers. Second, the partial and selective formalisation of labour relations, confined mainly to administrative and bureaucratic functions, without a corresponding extension of employment rights, social protection, or collective representation. Third, the reproduction and, in some cases, intensification of market segmentation and worker dependency, reflecting the structural constraints of a care system characterised by low public investment and a strong reliance on informal arrangements. From this perspective, platformisation does not resolve the contradictions of the Italian care regime but reconfigures them in digital form, embedding familistic logics into technological infrastructures and contributing to what can be described as a digitalised variant of welfare familism.

2. THE ITALIAN FAMILISTIC WELFARE SYSTEM IN THE EUROPEAN CONTEXT

Care services occupy a central place on the European social agenda, particularly within social-investment strategies designed to equip individuals to confront "new social risks" (Esping-Andersen et al., 2002; Taylor-Gooby, 2004; Bonoli, 2005; Morel, Palier, & Palme, 2012; Hemerijck, 2017). These risks include reconciling paid work with family responsibilities and ensuring the availability of in-kind services for frail and dependent persons. Across Europe, care is delivered through diverse arrangements – private market solutions, direct public provision, and mixed public-private and formal-informal models. Some countries, notably the Scandinavian cluster, have long prioritised extensive public provision via well-developed in-kind services. By contrast, in the United Kingdom public provision is more limited and relies more heavily on formal for-profit providers. Southern European countries such as Italy and Spain have lower levels of formal public and private provision and rely heavily on informal family care (Naldini, 2006; Saraceno & Keck, 2011). France represents a hybrid model in which the growth of in-kind services has been driven by a mix of formal home-based employment and cash-for-care instruments (Farvaque, 2013).

The financial crisis and subsequent austerity policies placed significant strain on welfare services across Europe, leading to retrenchment and deteriorating working conditions for care workers (Johnson & Pulignano, 2021). In Italy- typical of Southern European welfare systems – marketisation and privatisation pressures associated with structural reforms have compelled local institutions and regions to cut social spending (Ciarini & Neri, 2021). This context has spurred debate about the transformation of social policy, with many highlighting the increasing subordination of welfare provision to market logics (Azemati et al., 2013; Hazenberg & Hall, 2016; Dowling, 2017).

Marketisation and privatisation have blurred boundaries between public and private care, fostering hybrid welfare models that combine paid and unpaid, formal and informal elements (Johnson & Pulignano, 2021). In Italy, these dynamics intersect with a persistent emphasis on cash transfers to families and a lack of effective care regulation. Over time, this has constrained the development of formal private provision, leaving the informal market to dominate family care, particularly in long-term care. The main care allowance (Indennità di accompagnamento) perpetuates informality by failing to incentivise the regularisation of informal work or the purchase of formal services. This benefit adheres to the Mediterranean welfare pattern of direct cash transfers to households without mechanisms to channel resources towards formalising care arrangements or creating formal employment (Da Roit & Weicht, 2013; Leon & Pavolini, 2015).

Recent policy efforts have attempted to address informality. These include periodic regularisations for undocumented immigrants – who constitute a substantial share of the home-care workforce – and cash-for-care schemes introduced at national, regional, and company levels to expand families' choices between formal and informal care (Da Roit & Le Bihan, 2010; Leon & Pavolini, 2015). Yet national regulation remains underdeveloped, regional initiatives are limited, and cash transfers still dominate. Unlike France and Germany – where cash-for-care has diversified provision while recognising and formalising family care – Italian families, especially older individuals, continue to rely on relatives and low-cost migrant labour (Ciarini, 2016).

3. SOCIAL POLICIES AND THE CARE LABOUR MARKET: FRANCE, GERMANY, AND ITALY COMPARED

Transformations in welfare systems affect not only the nature and scope of provision but also the structure of the care labour market, formal and informal alike. This link is particularly significant in Italy. Unlike Italy – where informal care remains predominant – France and Germany have stimulated the creation of new regular employment within the care labour market. Since the mid-2000s, in both countries a series of labour-market and social-policy reforms has eroded earlier familistic models with a dual aim: to promote new (albeit low-wage) jobs by outsourcing activities formerly performed within families and to keep social spending under tighter fiscal control. Important differences persist between the two strategies. Following Simonazzi's classification (2009; see also Da Roit & Weicht, 2013), Germany combines formal and family-based informal care (the latter supported by social policy), whereas France is more explicitly oriented toward expanding formal employment in care.

3.1. France

French reforms favoured the development of a regulated social services market based on largely publicly financed vouchers. The Chèque Emploi Service Universel (CESU), established in 2005, reorganised the care system around a single policy tool, allowing households to access a wide range of services – from minor home repairs and childcare to complex socio-health care for dependent persons.

Until the early 2000s, the sector was highly fragmented: with few firms operating, most employment relations were direct agreements between clients and individual workers, often in the informal ("black") economy – much as in today's Italy. Through coordinated reforms – combining tax and contribution incentives for regularisation with programme restructuring (notably the Allocation Personnalisée d'Autonomie, APA, in 2002, and the CESU in 2005) – France created conditions for rapid growth in regular employment and in the number of for-profit and non-profit providers, while improving fiscal revenues.

These reforms shifted large shares of household-based informal care into the formal economy and served as labour-market insertion tools for the long-term unemployed and welfare recipients (from the Revenu Minimum d'Insertion to the Revenu de Solidarité Active in 2009).

Today the CESU operates in two main modalities: CESU déclaratif and CESU préfinancé. The former is used directly by households to purchase home-based services (domestic work, childcare, tutoring, light elder care, etc.). A formal contract is required, with wages not below the SMIC; generous contribution exemptions and tax relief make undeclared work economically unattractive. The CESU préfinancé is funded ex ante by employers (as fringe benefits) or by public/social institutions to provide access to care, including APA-linked long-term care. Voucher allocation follows needs assessments by multidisciplinary teams at département level, which classify users by four dependency levels. Depending on assessed needs, beneficiaries may purchase residential or home services from individual caregivers (including some relatives but excluding spouses) or from accredited third-sector/private providers (Colombo et al., 2011; Angermann & Eichhorst, 2012).

Employment relations under this solvency-based model take three forms: (i) direct gré à gré employment; (ii) mandataire arrangements mediated by accredited agencies; and (iii) prestataire services provided by accredited organisations (often private or non-profit). These coexist and can be combined. A notable advantage of the French model is stronger coordination across service types – from basic domestic help to integrated socio-health care (Simonazzi & Picchi, 2013).

3.2. Germany

Germany historically relied on a different balance between formal and family-based informal care, shaped by the 1994 introduction of long-term care insurance (Pflegeversicherung). German legislation has been more accommodating of family caregiving than France (Ciarini, 2016). For example, family caregivers receive pension contributions (Jacobzone & Jenson, 2001; Rothgang, 2011) – a provision absent in France, where only partial transfers to family caregivers (excluding spouses/cohabitants) are allowed. Less stringent monitoring of benefit use also encouraged continued reliance on family and informal (often migrant) labour (Simonazzi, 2009; Bettio & Mazzotta, 2011; Colombo et al., 2011; Bode & Chartrand, 2011).

In recent decades, however, Germany has shifted away from cash transfers toward greater use of formal home and residential services (Rothgang, 2011), reflected in more users/workers and a higher share of private providers (for- and non-profit) (Bode & Chartrand, 2011; Ciarini, 2016). For-profits now account for roughly half of residential long-term care and over 60% of home care; public provision is marginal.

Formal employment expansion was also influenced by the Hartz reforms (2003–2005), notably via minijobs: low-paid jobs (up to €450/month) with reduced employer taxes and contributions. Originally intended to formalise irregular domestic labour, minijobs were liberalised and extended across low-productivity services (retail, hospitality, care) (Schmidt, 2018; Baccaro & Benassi, 2014). While reducing inactivity and welfare dependence, they contributed to dualisation and low-wage expansion (Rizza & Gualmini, 2011; Eichhorst & Marx, 2012; Ciarini, 2018).

Unlike France – where comparable instruments were targeted to care and domestic services – Germany's minijobs applied broadly. A notable feature is the significant number of minijob workers also receiving minimum-income benefits, revealing a “working poor” segment in care and household services. Sectoral characteristics (relational work, low measured productivity) explain part of this, but deliberate wage-containment policies also mattered. Upon introduction of long-term care insurance, wages were set roughly 10% below public-sector levels (Ciarini, 2020). A sectoral minimum wage for care was introduced only in 2010 (€8.50 West/Berlin; €7.50 East). With the exception of a limited group of professionalised workers in integrated settings, most care workers have experienced wage stagnation, reduced hours, and growing precarity (Eichhorst, Marx, & Tobsch, 2013).

The interdependence between low wages and income support is central to the trade-off between service expansion and wage compression. In France and Germany, a substantial share of minimum-income recipients work in personal services – about one quarter in France (DREES, 2019). This dual-income structure underpins employment stability in the sector. Whether via minijobs or tax-subsidised domestic work, the hybrid combination of low wages and income support offers an alternative path to service expansion – distinct from fully public or purely private models – providing affordable services while promoting labour-market participation among marginalised groups.

3.3. Italy

Italy differs markedly from France and Germany – not only due to the lack of a systematic intersection between social-service policy and employment creation within the sector, but also because informal labour continues to cover a large share of care work, particularly for non-self-sufficient older people. Despite reforms – most notably Law 328/2000 – the system remains anchored in traditional patterns: low public social expenditure and persistent reliance on family care. This constrains women's labour-market participation. Informal employment is widespread, often indirectly supported by institutional arrangements.

The indennità di accompagnamento (attendance allowance), a universal, non-means-tested cash benefit largely used by older adults (two-thirds of recipients), illustrates the point. The benefit suffers from critical shortcomings: no multi-tiered needs-assessment framework (as in most EU countries), and no mechanisms to regularise informal employment.

There is no institutional link between the allowance and recognition/formalisation of care work – whether by a relative or a private caregiver – nor an option to channel the benefit toward regulated services (Da Roit & Sabatinelli, 2012; Da Roit & Weicht, 2013; Ranci & Arlotti, 2017). In effect, the allowance reproduces the Mediterranean pattern of cash transfers without control mechanisms or fiscal incentives for formalisation. Pasquinelli and Rusmini (2013) estimated roughly 830,000 irregular care jobs (mostly migrant women), financed by public allowances and private household resources. Other authors reach similar conclusions, placing Italy (with Spain and Portugal) among EU countries with the highest shares of undeclared employment in personal services (Farvaque, 2013; Leon & Pavolini, 2015).

Regularisation has occurred more through periodic amnesties for undocumented migrant workers (three between 2002 and 2012) than via explicit job-creation policies – implicitly recognising migrants' central role in home-based care (Leon & Pavolini, 2015). Formal provision by third-sector and private providers has grown in recent years (Deriu, 2019), yet the overall landscape remains fragmented. Fiscal/contribution incentives for home-based services have been episodic. The voucher per il lavoro accessorio (casual-work voucher) is emblematic: unlike France's CESU, Italy's vouchers were not designed for care but applied to many occasional activities (personal services, retail, tourism). Introduced in agriculture and later extended widely (up to €450/month, with no contributions – akin to German minijobs), they reached high usage before abolition in 2017. Only a small share involved domestic services, confirming the lack of sectoral focus; formalisation has depended more on immigration regularisations than on fiscal/voucher tools.

In practice, Italy has followed neither the French nor the German path. Reforms have been piecemeal, without an overarching strategy – ranging from categorical bonuses (bonus bebè, childcare vouchers, birth grants, natality funds) to incentives for corporate welfare. The 2016–2018 Stability Laws allowed productivity bonuses to be converted into welfare vouchers within collective agreements. Employers benefit from

IRPEF exemptions on accredited welfare services negotiated with unions (previously restricted to unilateral company schemes). This has spurred company- and sector-level bargaining, often substituting wage increases with in-kind benefits – reflecting (i) public-welfare retrenchment and difficulties in covering new social risks, which push decentralised bargaining into formerly public domains; and (ii) employers' wage-moderation demands, offset by complementary welfare (Leonardi & Arlotti, 2012; Pavolini, Mirabile, & Ascoli, 2013). As a result, welfare has moved up the agenda of collective bargaining at company, sectoral, and territorial levels, within a decentralised and segmented system- by firm size, sector, territory, and employment status (stable vs. precarious/non-standard) (Maino & Rizza, 2018). Firm size is crucial: medium-large firms can afford broader packages, benefiting from economies of scale, whereas small firms struggle unless federated. The growth of decentralised bargaining has not reversed these disparities. Today, the development of personal and social services – via company welfare or local public provision – remains fragmented and poorly integrated, often dependent on local initiatives rather than a coherent national framework (Maino & Rizza, 2018). Reforms have produced a patchwork of short-term sectoral measures disconnected from long-term investment in social infrastructure. Given low national coverage and marked territorial disparities – especially North–South – policy efforts remain dispersed and less effective. This is evident in early-childhood services: coverage for ages 0-2 remains well below demand and varies sharply; while northern/central regions meet or exceed the Barcelona target of 33% coverage by 2010, southern regions lag significantly.

4. THE ROLE OF CARE PLATFORMS IN REGULATING CARE PROVIDERS

The development of platforms in care systems is not a linear phenomenon, nor does it take the same form across contexts. Institutional arrangements profoundly shape the kinds of technological solutions adopted. Path dependency – legacies of past policy choices – conditions the space available to platforms and the mode of labour regulation. Italy has not adopted a specific strategy to create regular employment in care. Unlike France and Germany, large pockets of irregular and undeclared work persist – often financed, paradoxically, with public resources (as in the case of the Indennità di accompagnamento).

In a context marked by fragmentation and frequent incapacity to meet rising demand, the proliferation of private intermediary agencies – including online platforms – that broker between families and domestic/home-care workers is strongly shaped by these structural features. Given growing needs and limited public services, social care (institutional and home-based) has tended to become a lucrative segment. Recent studies identify different types of platforms in Italy's care sector. Following Pais (2023), we can distinguish:

- Corporate-welfare platforms, which provide welfare services with indirect access both for beneficiaries (employees of a firm) and for providers, who are selected by the platform;
- Territorial-welfare platforms, where providers typically have indirect access (contacted by the platform and often accredited by public authorities), while beneficiaries may access services indirectly (as social-service recipients) or directly for a fee;
- Digital-welfare platforms, which enable direct access for both clients and providers, with providers applying directly through the platform.

According to recent research (Pais & Zandonai, 2023; Agensir, 2023), 137 digital platforms are active in Italy in welfare and care. This mapping is among the first systematic attempts to classify and analyse digital platforms delivering services in domains traditionally covered by public institutions or the third sector, highlighting growing hybridisation between territorial-welfare models and platform logics. The identified platforms are heterogeneously distributed across welfare domains, notably:

- 59 operate in health, 32 of which specialise in psychological services – confirming rapid expansion of e-therapy and tele-counselling, especially post-COVID-19 (Agensir, 2023);
- 10 focus on education and childcare (on-demand babysitting, after-school, tailored educational support);
- 10 are active in socio-health assistance (home care, accompaniment of older/fragile persons, domiciliary nursing);
- 58 are multi-sectoral, operating across health, assistance, and education, and configuring hybrid models that combine public, private, and mutualistic services.

This articulation shows that welfare platformisation is no longer marginal but is consolidating as part of the country's social infrastructure. Digital platforms not only simplify access to timely, personalised services; they also reshape intermediation and delivery, introducing algorithmic coordination and reputation systems. This raises important questions of regulation, transparency, worker protection, and service quality.

Looking ahead, the spread of welfare platforms in Italy can be read as an indicator of the transformation of the Mediterranean welfare model – traditionally characterised by strong family and public roles- toward a mixed model in which private and technological actors assume increasingly salient functions of intermediation and governance. Such an evolution calls for a comprehensive rethinking of public policy to ensure inclusion, equity, and service quality in a rapidly expanding digital ecosystem.

Recent analysis on the Italian case (Ciarini & De Vita 2024), highlight how the formalisation of labour is pretty marginal, due to the institutional legacies of the welfare system and the ability of platform to formalised only few aspects of the labour relations. Against this background the lack of social protection and occupational health and safety, minimal bargaining power, income insecurity, opaque and/or volatile rules are perpetuated and to some extent intensified. Moreover, the Italian case demonstrates the continued existence of a nuanced regulatory framework, in which both formal regulatory mechanisms and significant informal practices coexist. This enables platforms to navigate the complexities of regulatory frameworks, exploiting loopholes to their advantage. To illustrate, the existence of numerous NCLAs for domestic work enables care platforms to circumvent efforts to enhance working conditions by opting for less generous contracts. In this manner, they circumvent the utilisation of salaried labour and thus the potential application of the NCLA. These strategies effectively impede the efforts of unions to regulate the sector and improve working conditions. Furthermore, they paradoxically simplify mainly bureaucratic management, which is preferred by both families and workers.

The working conditions of these workers remain low, and even the flexibility and autonomy promised to them is, constrained and subordinated to compliance with the service times imposed by the platforms (*ivi*). This is partly a consequence of the significant dependency of these workers on the resources of the families that require care services. The limited economic resources available to families and the inadequacy of public care schemes force them to either seek informal work or turn to platforms that offer flexible and personalised services at affordable prices.

5. CONCLUSIONS

The emergence and development of digital platforms in the Italian care sector cannot be understood without reference to the institutional characteristics of its familistic welfare regime. As discussed earlier, Italy's care system remains deeply embedded in a structure where families act as the primary providers of welfare and where public intervention has historically been limited both in scope and in regulatory capacity.

Unlike in France and Germany, where public authorities have progressively developed structured mechanisms to promote formalisation and regularisation of care work-through voucher systems such as the *Chèque Emploi Service Universel* (CESU) in France or the combination of *Pflegeversicherung* and *minijobs* in Germany – Italy continues to rely on informal arrangements financed through direct cash transfers, notably the *indennità di accompagnamento*.

These institutional legacies have created a fragmented and largely unregulated care market in which informal labour remains the norm rather than the exception. In this context, the entry of digital platforms does not necessarily represent a break with the past but rather an adaptation of informal market logics to a digital environment.

In France, the state's deliberate strategy to integrate solvency-based instruments such as the CESU within the welfare architecture has created a quasi-market for personal services that combines flexibility with legal protection and fiscal accountability. The system's design, which requires formal contracts and enables access to tax credits and contribution exemptions, has facilitated the transition of many workers from the informal to the formal economy, while sustaining a dense ecosystem of accredited providers. Similarly, in Germany, the introduction of long-term care insurance in 1994 and the subsequent liberalisation of low-wage employment through the Hartz reforms have contributed to the creation of an institutionalised care market. Although the German model also relies on family-based care, the coexistence of clear regulatory frameworks and publicly subsidised employment forms has allowed the gradual emergence of a more structured and formalised labour market. Digital platforms entering these contexts operate within preexisting regulatory infrastructures that constrain informality and channel innovation into recognised contractual and fiscal arrangements.

Italy, by contrast, lacks these institutional foundations. The persistent absence of a coherent national policy for care work – coupled with limited public investment and heavy dependence on family provision – means that platforms emerge as private responses to structural gaps in welfare coverage rather than as tools of public regulation. Instead of reinforcing formalisation, they mirror the informal and fragmented characteristics of the Italian care economy. While digital platforms can improve visibility, matching efficiency, and access to services, they do so within an institutional environment that neither supports nor incentivises regularisation. The result is a form of "digital familism," in which technological intermediation reproduces the asymmetries of the informal market rather than resolving them. The notion of digital familism helps to conceptualise platformisation not as a disruptive force, but as a mechanism of institutional adaptation that digitises – rather than dismantles – familistic welfare arrangements. From this point of view, digital familism refers to a specific configuration of platform-mediated care provision in which digital technologies operate within, and reinforce, a familistic welfare regime. In this configuration, digital intermediation tends to formalise only limited and peripheral aspects of labour relations, such as matching, scheduling, and payment management, while leaving the core features of informality, weak social protection, and asymmetric power relations largely intact.

Consequently, the capacity of digital platforms to contribute to decent work and social protection in Italy is structurally constrained by the familistic configuration of its welfare system. In the absence of public regulation and fiscal incentives, platforms risk amplifying the segmentation of the care market, consolidating precariousness rather than fostering the emergence of a regulated and professionalised care economy.

From a policy perspective, these findings underscore the need to view digital platforms not as exogenous agents of change but as institutional actors whose trajectories are shaped by the regulatory and welfare contexts in which they operate. In Italy, any strategy aiming to harness the potential of digital platforms for improving care provision should therefore address the structural weaknesses of the welfare system itself – namely, the lack of a coherent care infrastructure, weak public investment, and insufficient regulation of employment relations in domestic work.

Lastly, from an analytical standpoint, the comparative perspective adopted here highlights the crucial interplay between welfare regimes, labour-market institutions, and technological innovation. The platformisation of care does not occur in a vacuum: it is deeply embedded in national trajectories of welfare reform and labour regulation. Understanding its implications thus requires moving beyond a purely economic or technological interpretation toward a political economy of digital care – one that situates platform work within broader processes of welfare restructuring, social reproduction, and the reconfiguration of labour in contemporary capitalism.

REFERENCES

Agensir. (2023, November 14). *Welfare in Italia: si fa sempre più sulle piattaforme digitali. Sono 137 e ricoprono vari settori. WePlat mappa il fenomeno.* https://www.agensir.it/quotidiano/2023/11/14/welfare-in-italia-si-fa-sempre-piu-sulle-piattaforme-digitali-sono-137-e-ricoprono-vari-settori-weplat-mappa-il-fenomeno

ANSA. (2023, November 14). *Ricerca: il welfare sempre più sulle piattaforme digitali.* ANSA Economia. https://www.ansa.it/sito/notizie/economia/pmi/2023/11/14/ricerca-il-welfare-sempre-piu-sulle-piattaforme-digitali_3a41c4b4-8135-4375-b7fa-378b6a55e1a2.html

Arcidiacono, D., Bonifacio, F., & Pais, I. (2024). Transparent informality: The ma(s)king of professionals in childcare service platforms. *Critical Sociology, 50*(1), 1–17.

Arlotti, M., & Ranci, C. (2020). *Reforming long-term care in Europe: Sharing the costs and benefits.* Palgrave: Macmillan.

Ascoli, U., & Pavolini, E. (2015). *The Italian welfare state in a European perspective.* Policy Press.

Azemati, H., Belinsky, M., Gillette, R., Muñoz, C., & Eddy, M. (2013). *Social impact bonds: Lessons from the field.* Harvard Kennedy School.

Baum, C. (2024). Platforms don't care – qualified and experienced women do: A case study on self-employment and on-demand care work. *Critical Sociology, 50*(1), 1–24.

Blanchard, O. (2023). *Las plataformas digitales de cuidados y sus servicios worktech en América Latina y el Caribe.* Inter-American Development Bank Lab.

Bonoli, G. (2005). The politics of the new social policies: Providing coverage against new social risks in mature welfare states. *Policy & Politics, 33*(3), 431–449.

Bonoli, G., Cantillon, B., & Van Lancker, W. (2017). Social investment and the Matthew effect. In A. Hemerijck (Ed.), *The uses of social investment* (pp. 66–76). Oxford University Press.

Busacca, M., & Zandonai, F. (2020). Platform social enterprises as a new model for co-production. *Studi Organizzativi, 2*, 61–86.

Ciarini, A. (2016). *Welfare e lavoro nel capitalismo globale.* Carocci.

Ciarini, A., & Neri, S. (2021). Welfare, territorial governance and social innovation in Italy: The missing link. *Social Policy & Administration, 55*(6), 1017–1034.

CNEL. (2024, October). *Analisi della governance e della spesa dei servizi sociali territoriali* (Quaderno CNEL No. 22). https://www.cnel.it

Da Roit, B., & Le Bihan, B. (2010). Similar and yet so different: Cash-for-care in six European countries' long-term care policies. *Milbank Quarterly, 88*(3), 286–309.

Da Roit, B., & Weicht, B. (2013). Employers and migrant care workers: New understandings of family and work. *Sociology, 47*(4), 727–743.

De Stefano, V. (2016). The rise of the "just-in-time workforce": On-demand work, crowdwork and labour protection in the "gig-economy." *International Labour Office.*

De Vita, L., & Ciarini, A. (2025). Still informal? Care platforms in a loosely regulated context: The Italian case. *Critical Sociology, 51*(1), 1–16.

De Vita, L., & Bertolini, A. (2025). The platformisation of the domestic and home care. *Critical Sociology, 51*(1), 1–12.

De Vita, R., & Corasaniti, G. (2022). Domestic and care work in Italy: Between regularisation and fragmentation. *Journal of Contemporary European Studies, 30*(4), 557–575.

Donini, A., & Smorto, G. (2024). L'approvazione della direttiva sul lavoro mediante piattaforma digitale: Prima lettura. *Labour & Law Issue, 2024*(1), 25–30.

Dowling, E. (2017). *The care crisis: What caused it and how can we end it.* Verso Books.

Esping-Andersen, G. (1990). *The three worlds of welfare capitalism.* Princeton University Press: Princeton.

European Commission. (2021). *Employment and social developments in Europe (ESDE) 2021 report.* Publications Office of the European Union.

Eurofound. (2020a). *Solidarity fund for very small companies, the self-employed and micro-entrepreneurs (Fonds de solidarité).* https://static.eurofound.europa.eu

Farvaque, N. (2013). Care policies in France: Between formalization and informalization. *European Journal of Social Policy, 23*(3), 245–259.

Ferrera, M. (1996). The "Southern Model" of Welfare in Social Europe. *Journal of European Social Policy, 6*(1), 17–37.

Flanagan, F. (2019). Theorising the gig economy and the future of work. *Industrial Relations Journal, 50*(1), 62–78.

Graham, M., Hjorth, I., & Lehdonvirta, V. (2017). Digital labour and development: Impacts of global digital labour platforms and the gig economy on worker livelihoods. *Transfer: European Review of Labour and Research, 23*(2), 135–162.

Hemerijck, A. (2017). *The uses of social investment.* Oxford University Press.

INPS. (2025, June). *Osservatorio lavoratori domestici.* https://servizi2.inps.it/servizi/osservatoristatistici

ISTAT. (2025, May 21). *Rapporto annuale 2025: La situazione del Paese.* https://www.istat.it

Leon, M., & Pavolini, E. (2015). *Care regimes in Europe: Variations and change.* Policy Press.

Morgan, K., & Pulignano, V. (2023). Platformisation and precarity: The changing nature of care work in Europe. *Work, Employment and Society, 37*(2), 315–331.

OECD. (2021). *OECD employment outlook 2021: Navigating the COVID-19 crisis and recovery.* OECD Publishing.

Pais, I. (2023). *Welfare platforms and the new care economy.* Egea: Milano.

Presidenza del Consiglio dei Ministri. (2025, July 14). *Modelli organizzativi omogenei degli Ambiti Territoriali Sociali per l'attuazione dei LEPS.* https://www.conferenzastatocitta.it

Ranci, C., Arlotti, M., & Sabatinelli, S. (2022). *Long-term care reforms in Southern Europe.* Springer: Berlin.

Saraceno, C., & Keck, W. (2011). *Families, care and work in Europe: Empirical evidence and policy lessons.* Palgrave Macmillan: London.

Srnicek, N. (2017). *Platform capitalism.* Polity Press.

Tassinari, A., & Maccarrone, V. (2020). Riders on the storm: Workplace solidarity among gig economy couriers in Italy and the UK. *Work, Employment and Society, 34*(1), 35–54.

Ticona, J., & Mateescu, A. (2018). Trusted strangers: Carework platforms' cultural entrepreneurship in the on-demand economy. *New Media & Society, 20*(11), 4384–4404.

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

Vallas, S. P., & Schor, J. B. (2020). What do platforms do? Understanding the gig economy. *Annual Review of Sociology, 46*(1), 273–294.

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

JEP

Firenze University Press
https://oaj.fupress.net/index.php/jep

OPEN ACCESS

Citation: Profumo, F. (2025) The hundred languages in the age of AI. Generative educational pathways between humanity and technology: 0-99 model for lifelong learning. *Journal of Emerging Perspectives* 2: 83-85. doi: 10.36253/jep-20884

Received: April 30, 2025

Revised: September 30, 2025

Published: December 20, 2025

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

The hundred languages in the age of AI. Generative educational pathways between humanity and technology: 0-99 model for lifelong learning

Francesco Profumo

Abstract. Artificial intelligence is profoundly transforming the learning ecosystem, reshaping access to knowledge, the nature of educational relationships, and the processes through which understanding is constructed. This paper proposes a longitudinal "0-99" educational model grounded in the legacy of the hundred languages of the Reggio Children Approach, in the Socratic method of questioning, and in the idea of AI as a cognitive partner. The model identifies, for each stage of life – from early childhood to primary education, adolescence and young adulthood, and into mature adulthood – the languages to be cultivated in the age of AI and the competencies required for informed, critical, and creative citizenship. Adopting an interdisciplinary perspective, the paper examines both the opportunities and the risks associated with AI, highlighting the need for clear educational policies, inclusive digital infrastructures, and widespread training initiatives. The result is a proposal for a new humanism of learning, in which AI becomes a lever for expanding human languages rather than replacing them, supporting each person across the entire life span.

1. INTRODUCTION: AI AS A NEW LEARNING ECOSYSTEM

Artificial intelligence has not entered the educational world as a simple tool to be added to traditional methods. It has entered as an environment, a cognitive ecosystem that is transforming the way children, young people, and adults explore reality, construct meanings, formulate questions, and make decisions. Few transitions in the history of education have been so profound: we are not merely updating the technologies of learning, but reshaping the architecture of human learning itself.

In this new context, the crucial question is not whether to adopt AI, but how to do so in a conscious, responsible, and human-centred way. This is where the Italian pedagogical tradition – particularly the Reggio Children Approach and Loris Malaguzzi's intuition of the "hundred languages" – reveals its surprising relevance.

The multiplicity of children's languages is not only a principle of early childhood education; it has become an interpretative key for the entire educational trajectory in the age of AI.

This insight forms the basis of the 0-99 educational model, a longitudinal vision that spans the entire life course and proposes a new humanism of learning: a continuous path capable of combining human expressiveness with the potential of artificial intelligence

2. FOUNDATIONAL PRINCIPLES OF THE 0-99 EDUCATIONAL MODEL

At the core of the model lies the conviction that every individual possesses a vast repertoire of languages – corporeal, symbolic, narrative, mathematical, relational, digital – which education should cultivate and allow to evolve rather than reduce. AI can contribute to this growth, but only if its use reflects a vision of the human being as plural, creative, and relational.

A second pillar is the Socratic method. In an era in which every answer seems instantly available, the true educational value shifts back to the quality of the question. AI, when used correctly, can become a catalyst for inquiry and a promoter of critical thinking: a kind of "Socratic companion" that supports human understanding.

Finally, the triad "educating with, within, and beyond AI" outlines the boundaries of contemporary learning. Educating with AI means integrating intelligent tools into learning pathways; educating within AI means recognising that we live inside a cognitive and cultural environment deeply shaped by algorithms; educating beyond AI means cultivating the inherently human dimensions – critical thinking, creativity, responsibility – that no technology will ever replace.

3. THE 0-99 MODEL: A CONTINUOUS JOURNEY OF EVOLVING LANGUAGES

3.1. Early Childhood (0–6): The Languages of Exploration

Early childhood remains the essential point of departure. In educational settings inspired by Reggio Emilia, children do not express themselves through a single language, but through many: the body, light, colour, movement, voice, objects, space, and relationships. AI enters this world as an additional material – not to guide, but to be explored, observed, and manipulated. Augmented ateliers, interactive projections, and responsive digital tools allow children to see their gestures transformed into shapes, traces, images, and sounds. In this way, AI expands expressive possibilities without diminishing the spontaneity of the age.

3.2. Primary Education (6–14): The Languages of Discovery and Logical Thought

As children grow, new languages emerge that AI can support and enhance. Mathematical, scientific, linguistic, and artistic languages begin to intertwine with digital skills and critical literacy. AI, in its simplest forms, can become a training ground for reasoning: a dialogic environment in which to test hypotheses, identify errors, and assess the quality of information.

This is also the age in which AI can truly become Socratic – an instrument that invites doubt, fosters understanding, and prompts explanation, turning each interaction into an exercise in critical inquiry.

3.3. Adolescence and Young Adulthood (14–24): The Languages of Identity and Orientation

During adolescence, the relationship with AI becomes more complex and reflective. New languages emerge: argumentation, debate, design, project-based thinking, and advanced STEM competencies. AI is no longer just a tool, but a cognitive workspace – a partner with which to design, simulate scenarios, visualise data, and build hypotheses.

This is also the phase of personal and professional orientation. AI can offer personalised pathways, suggest directions, help individuals identify their strengths, understand the labour market, and imagine future trajectories – not by deciding for them, but by helping them make better choices.

3.4. Adulthood and Later Life (25–99): The Languages of Transformation

The final segment of the 0-99 model concerns adults and older adults. Never before has a pedagogy of transformation been so necessary – one capable of supporting professional reskilling, career transitions, and new forms of digital citizenship. The languages of adulthood are professional, analytical, collaborative, and civic. AI can become a "cognitive trainer," a support for managing complexity, and an ally in maintaining cognitive vitality and autonomy.

At this stage, learning serves not only to work, but to live: to remain active members of society, to understand technological risks and opportunities, and to continue growing.

4. GOVERNING THE AI TRANSITION: AN EDUCATIONAL AND TERRITORIAL RESPONSIBILITY

AI can expand languages, but it can also compress them. It can include or exclude. It can liberate or constrain. For this reason, clear educational policies are essential: a national pedagogical framework for AI, competence standards for each stage of life, accessible digital infrastructures across all territories, and ambitious teacher-training

programmes. Without these conditions, territorial inequalities may widen, and AI risks becoming a factor of exclusion rather than social advancement.

5. CONCLUSION: A NEW HUMANISM OF LEARNING

AI compels us to rethink education not as a segment of life, but as life itself. The hundred languages remind us that every person carries inexhaustible possibilities; the 0-99 model suggests that these possibilities must be nurtured throughout the entire lifespan. If guided by a humanistic vision, AI can restore depth, power, and beauty to the way we learn.

In the future that awaits us, it will not be AI that makes the difference, but the way we educate people to engage in dialogue with it. Because in a world where everything changes, the only truly irreplaceable competence will always be the ability to keep asking the right questions.

JEP

Firenze University Press
https://oaj.fupress.net/index.php/jep

OPEN ACCESS

Citation: Cencig, V., & D'Almo, M. (2025) From automation to agency: The paradigm of agentic AI across technology, society and the ontology of work. *Journal of Emerging Perspectives* 2: 87-96. doi: 10.36253/jep-20885

Received: April 30, 2025

Revised: September 30, 2025

Published: December 20, 2025

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

From automation to agency: The paradigm of agentic AI across technology, society and the ontology of work

Valerio Cencig[1,*], Mario D'Almo[2]

[1] *Head of Compliance Digital & Data Transformation at Intesa, Italy*
[2] *Head of Digital Strategy and Intelligence Compliance Area in Intesa Sanpaolo, Italy*
valerio.cencig@intesasanpaolo.com
*Corresponding author

Abstract. The evolution of artificial intelligence is redefining the relationship between humans and machines, shifting from traditional automation toward systems endowed with agency. Unlike conventional AI, which primarily supports or executes predefined tasks, agentic AI systems are capable of autonomous decision-making, proactive goal generation, learning from experience and coordinated action within complex environments. This shift represents not only a technological advancement but an ontological transformation, as machines increasingly operate as cognitive agents rather than passive tools. This article adopts an interdisciplinary perspective to examine the paradigm of agentic AI, tracing its evolution from earlier forms of automation and outlining its defining characteristics, architectures and application domains. It then analyses the implications for work and organisational structures, highlighting how agentic systems reconfigure roles, redistribute cognitive labour and enable new forms of human–machine co-agency. Finally, the paper addresses the ethical, legal and governance challenges raised by autonomous agents, arguing for responsible adoption frameworks that preserve human centrality, accountability and social justice in emerging socio-technical systems.

Keywords: agentic Artificial Intelligence, autonomous systems, human-machine co-agency, distributed cognition, multi-agent systems, ethical ai governance, socio-technical systems.

1. INTRODUCTION

The history of technology has been marked by progressive redefinitions of the relationship between humans and machines, each characterised by the gradual delegation of human functions to artificial devices. In the first phase, which began with the Industrial Revolution, machines replaced human muscular strength: mechanical looms, steam engines and hydraulic presses relieved humans from physical labour whilst keeping decision-making control firmly in the hands of the operator.

The second phase, the electronic automation of the twentieth century, introduced systems capable of performing predefined sequences of operations through programming: industrial robots and automated assembly lines replaced not only physical force but also the repetitiveness of gesture, though they remained constrained by explicit and deterministic instructions.

The third phase, the digital automation of recent decades, broadly expanded data processing capacity: information systems process data, manage transactions and coordinate complex processes, yet continue to operate within rigidly predefined parameters, without the ability to learn or adapt independently.

Contemporary artificial intelligence introduces a significant qualitative shift from previous technological eras. For the first time, artificial systems are not restricted to executing tasks based on predetermined algorithms; they learn from data, discern intricate patterns, generalise acquired knowledge to new contexts and, notably, make autonomous decisions in situations not explicitly anticipated by their designers. This progression from task execution to decision-making represents an ontological transformation: the machine advances from being a mere instrument to exhibiting characteristics of a cognitive agent. Employing methodologies such as machine learning, deep neural networks and reinforcement learning, AI systems acquire capabilities traditionally attributed to biological intelligence, including contextual perception, inferential reasoning, abstract learning and strategic adaptation.

In this artificial intelligence revolution, agentic AI represents a further paradigmatic shift. Whilst conventional AI systems typically operate in reactive or consultative mode, addressing predefined queries, classifying data or generating outputs on request, artificial agents exhibit proactivity and operational autonomy. These agents continuously monitor their environments, generate sub-goals, devise action plans, execute tasks independently of ongoing supervision and refine behaviour based on experiential learning. This agency is characterised by decision-making autonomy, operational initiative and dynamic adaptation, leading to fundamental changes in production models in both manufacturing and services.

In material production, agentic AI enables adaptive factories where networks of agents coordinate machinery, manage supply chains, optimise logistical processes and anticipate failures through predictive maintenance: all in real time and with minimal human intervention. Multi-agent systems autonomously negotiate production priorities, reallocate resources in response to demand fluctuations and implement energy-optimisation strategies that adapt dynamically to environmental and economic constraints.

In the service sector, comparable transformations are evident: financial agents independently manage investment portfolios by analysing global market conditions and adjusting strategies almost instantaneously; diagnostic agents support medical professionals by processing complex medical records and recommending personalised treatment strategies; educational agents customise learning materials according to the individual cognitive rhythms of students; customer service agents engage in complex interactions, continuously adapting through the analysis of previous conversations.

This transition extends beyond technical and operational domains, influencing our understanding of action, autonomy, and responsibility, as well as redefining the meaning of human work. When machines do not simply execute but decide, when they do not only respond but anticipate, when they do not simply assist but collaborate, the boundary between tool and cognitive partner becomes porous. Accordingly, it is necessary to re-examine philosophical, legal and ethical frameworks established in eras when agency was an exclusive human attribute.

This article seeks to analyse this transformation through an interdisciplinary perspective. First, it explores the historical evolution from mechanical automation to artificial agency. Next, it examines the architectures, distinctive features and application domains of agentic systems. Then, it shows applied sectors and case studies. Finally, the analysis focuses on the transformation of work and underscores the importance of responsible governance that maintains the centrality of the human subject and preserves dignity, autonomy and social justice in the age of human–machine cognitive collaboration.

2. FROM THE PARADIGM OF AUTOMATION TO THE PARADIGM OF AGENCY

Agentic artificial intelligence introduces a substantial discontinuity in the relationship between human beings and technological systems. Unlike previous technologies, which operate exclusively as extensions of human capabilities under the control of an operator, agentic AI does not merely replace humans in the execution of predefined tasks; it actively participates in the decision-making process, exhibiting progressively increasing degrees of autonomy.

To understand this transformation, it is useful to conceptualise decision-making autonomy as a continuous spectrum.

At the first level of the autonomy spectrum there are artificial intelligence systems (which we might define as 'traditional machine learning') that operate in a purely consultative mode, supporting human decision-making. A representative example is the advisory tools used by financial and investment consultants: the AI system analyses thousands of parameters: historical asset performance, market volatility, correlations amongst asset classes, macroeconomic indicators, sentiment analysis derived from financial news, sectoral growth projections, and generates contextualised, personalised recommendations. It suggests portfolio rebalancing, identifies investment opportunities or signals risks of excessive concentration in specific sectors and proposes hedging strategies to protect capital from market shocks.

However, each recommendation is reviewed by the human consultant, who retains full decision-making authority: the consultant evaluates the guidance, contextualises it with his knowledge of the client, integrates qualitative considerations beyond the system's reach and determines the appropriate course of action. The system performs no operation until explicitly instructed by the human operator. In this scenario, AI works as an expert adviser that augments human cognitive capabilities without supplanting final judgement. Nonetheless, important questions arise: to what extent do human consultants comprehend the rationale underlying algorithmic recommendations? Are they able to exercise truly independent critical judgement, or are they, in practice, inclined to follow recommendations they do not fully understand?

Moving along the autonomy spectrum, we encounter intermediate configurations in which AI assumes greater decision-making authority whilst remaining under human supervision. A significant example is the advanced autopilot systems used in commercial aviation. Modern autopilots (especially those trained with AI, in both civil and military contexts) do not merely maintain preset headings and altitudes; they actively manage critical phases of flight: take-off, navigation through congested air corridors, dynamic adaptation to changing weather conditions and landing.

During cruising, the system continuously makes operational decisions: it optimises speed and altitude to minimise fuel consumption, adjusts the route to avoid turbulence identified via radar and manages the aircraft's aerodynamic configuration (flaps, slats, spoilers) to maintain optimal efficiency. Human pilots monitor these processes but intervene actively only in exceptional circumstances: technical malfunctions, emergencies, extreme weather conditions; unexpected events requiring sophisticated contextual judgement. This configuration corresponds to what automation engineers call 'human-on-the-loop': AI operates autonomously within predefined parameters, while a human oversees the decision-making steps and can intervene or override at any moment, though their involvement is not essential to the process.

This arrangement gives rise to new challenges. For example, *automation complacency:* the tendency of human operators to develop over-reliance on automatic systems, thus reducing vigilance, is documented in numerous aviation incidents. When systems operate reliably for long periods, human operators may become overconfident, paying less attention to flight parameters and losing readiness to detect anomalies. *Deskilling*, the progressive loss of competence due to lack of practice, is another growing concern: pilots who fly mostly with active autopilot may experience deterioration in manual flying skills, precisely those crucial in emergencies requiring them to take direct control. These issues are already present, though less acutely, with deterministic autopilot systems; but they are likely to be exacerbated by new AI-trained systems.

At the opposite side of the spectrum are agentic AI systems operating in complete autonomy, making and executing decisions without direct human supervision. The most emblematic example is high-frequency algorithmic trading platforms based on artificial intelligence (AI High-Frequency Trading, AI-HFT), where artificial agents autonomously make decisions and execute the corresponding financial operations. A typical AI-HFT system monitors hundreds of global markets simultaneously, analysing continuous streams of data: real-time order flows, execution prices, trading volumes, bid–ask spreads, changes in volatility, financial news processed through natural language processing (NLP) and technical indicators computed over multiple time windows. When it identifies an opportunity, the algorithm autonomously executes the appropriate operations.

Here, direct human intervention is not only absent but structurally impossible: the information involved, and the temporal scales (measured in milliseconds) lie beyond human cognitive reach. Human traders perform a radically different role: they design high-level algorithmic strategies, define risk parameters (maximum exposure per position, daily loss limits, liquidity requirements) and monitor aggregate performances, adjusting strategies and parameters in response to structural market changes. In this configuration, even more extreme situations may arise in which AI systems develop emergent behaviours not foreseen by designers, for example: cooperative dynamics with other trading systems, potentially even in violation of market rules.

The transformation from 'instrument' to 'agent' thus raises profound questions that transcend purely technical or practical issues.

3. AGENTIC INTELLIGENCE: DEFINITIONS AND FUNDAMENTAL CHARACTERISTICS

To understand the concept of an artificial agent and its transformative implication, it is preliminarily necessary to identify what distinguishes it from a software system or a traditional machine-learning model.

The main properties that differentiate intelligent agents from traditional software programmes are autonomy, reactivity, proactivity and sociality.

Autonomy means that the agent can operate without direct human intervention, controlling its own actions and internal state. Whilst acting within a predefined context, an autonomous agent is not simply a programme that executes specific instructions, but a system that pursues objectives by adapting its behaviour to circumstances and selecting the most appropriate strategies considering the context.

Reactivity indicates that the agent perceives the environment in which it operates and responds promptly to changes occurring within it, also interacting with the environment itself.

Proactivity means that the agent does not merely react to contextual stimuli but displays goal-driven behaviour, taking initiative to achieve objectives.

Sociality indicates that the agent is equipped to interact with other actors, agents or humans, with different roles: as an agent performing a specific task, or as a 'controller' or 'orchestrator' in relation to other agents. This property is crucial because it enables the creation of cooperative multi-agent systems, where coordination is essential, capable of performing highly complex tasks.

Starting from these fundamental properties, agents may be classified along multiple dimensions. Each dimension represents one aspect of their architecture and behaviour.

A basic taxonomy distinguishes agents according to the complexity of their internal decision-making mechanisms. Here are some examples.

<u>Simple Reactive Agents</u>

Simple reactive agents operate through direct, deterministic condition–action rules – often not AI-driven. When a certain environmental condition is perceived, they perform a specific action. A classic example is the robotic vacuum cleaner: if it detects an obstacle, it changes direction; if it detects dirt beneath its sensors, it activates intensified suction; if the battery is low, it returns to its charging base.

<u>Model-Based Agents</u>

Model-based AI agents create and maintain internal representations of the state of the world, even when it is not directly observable, integrating current perceptions with prior knowledge to infer the complete state. An autonomous vehicle, for instance, maintains a model of the position of other vehicles, pedestrians and traffic signs even when temporarily outside its visual field. This model is continuously updated through sensors and tracking algorithms. Model-based agents can manage perceptual uncertainty and make informed decisions even when information is incomplete.

<u>Goal-Driven Agents</u>

Goal-driven agents incorporate representations of objectives that guide their behaviour. Rather than simply reacting, these agents evaluate individual actions based on how much they bring the system closer to the assigned goals. This requires predictive capability: the agent must be able to simulate the consequences of various alternative actions and select the one that maximises the chances of success.

A logistical planning agent tasked with delivering parcels, for example, formulates sub-goals (such as dividing parcels into batches with different deadlines), formulates and evaluates possible alternatives (delivery sequences), assesses consequences (total time, fuel consumption, likelihood of delays) and chooses the optimal plan. Goal-based agents are more flexible than reactive ones: if environmental conditions change, they can reformulate plans rather than being trapped in the previous plan.

<u>Utility-Based Agents</u>

Utility-based agents refine the goal-based approach by introducing utility functions that quantify the desirability of different world states, measuring the utility attainable. Whereas goal-based agents operate in a binary mode (goal achieved or not), utility-based agents recognise degrees of success.

<u>Learning Agents</u>

Learning agents incorporate learning mechanisms that allow them to improve performance through experience. Such agents operate with evolving knowledge and adapt to the encountered patterns.

<u>Complex Agent Networks</u>

Finally, there are complex networks of agents, functioning as genuine 'digital nervous systems', enabling high scalability of agent capabilities across extensive and diverse problems and situations. These networks are built according to rigorous design principles, including:

- task decomposition across cooperating yet independent agents;
- technological neutrality with respect to vendors, allowing integration of agents developed by different providers;
- embedded self-regulatory policies to prevent undesirable network behaviours.

Agent networks also possess the ability to adapt to new scenarios without requiring reprogramming, to reorganise autonomously if an agent fails (resilience) and to communicate in ways that make the system function as a single coherent entity.

What are the enabling technologies?

Enabling technologies range from reinforcement learning (including multi-agent RL) to symbolic planning, hybrid neural–symbolic systems and agents that exploit "*Large Language models* (LLMs)" to plan and orchestrate tools.

We may therefore conclude that agentic intelligence is not merely sophisticated automation but a new model of distributed cognition, in which the application of intelligence to processes, historically centred in humans, is distributed between humans and machines, making the boundaries between biological and artificial intelligence fluid and blurred.

Navigating these new paradigms does not simply mean managing technological evolution; it requires a theoretical rethinking of what constitutes intelligence, agency and cognition in an era characterised by deeply integrated socio-technical systems.

4. AGENTIC AI IN ACTION: APPLICATION SECTORS AND CASE STUDIES

The emergence of agentic systems in real-world applications is already transforming industries, services, and professional practices, generating both new opportunities and unforeseen challenges.

Traditional AI in finance, healthcare, industry and energy predicts or classifies but does not act: robo-advisors generate suggestions; credit-scoring systems classify creditworthiness; traditional anti-fraud systems issue alerts; imaging diagnostics produce predictions; sensor systems signal the need for maintenance interventions, and so on.

Agentic AI changes the game because it produces operational outcomes.

We have already discussed how, in trading, agents maintain a representation of market state, plan sequences of actions and execute them: orders, hedging, strategy shifts as markets evolve.

In healthcare, hospital agents could continuously monitor critical patients, aggregate signals from multiple devices, detect deterioration, activate protocols, notify medical staff and document everything. In oncology, they may adjust dosages, integrate new evidence and compare treatments using digital twins (dynamic digital replicas). In the operating room, they may coordinate scheduling, personnel and equipment, reorganising the entire day if an emergency case arrives. In emergency departments, they may manage triage and queues in real time; in large-scale disasters, networks of agents may coordinate ambulances, hospitals, logistics and supplies, distributing patients and resources where most needed.

In industry and logistics, fleets of warehouse robots already negotiate tasks, optimise routes, avoid collisions and replan when the environment changes. In supply-chain management, agents monitor demand, rebalance inventories, negotiate with suppliers, orchestrate transport and autonomously react to unexpected events. In maintenance, they do more than issue predictions: they book downtime windows, order spare parts and update documentation upon completion.

Marketing will also need to evolve. The advent of autonomous AI agents could change how markets interact with customers because customers may no longer be human beings but bots and agents. Colours, sounds, advertising, essentially the entire symbolic toolkit of persuasion, could become irrelevant.

In the energy sector, domestic and industrial agents control solar panels, batteries, flexible loads and electric vehicles to optimise costs and comfort, whilst aggregators and grid-level agents balance supply and demand like a virtual power plant, intervening autonomously on peaks and shortages.

A highly advanced anti-money-laundering agent ('Financial Crime Agent') could operate 24/7 on transactions, accounts and digital channels; it would maintain a dynamic graph of relationships between customers, merchants, crypto wallets, companies and high-risk jurisdictions; and detect dynamic patterns identifying money-laundering activities. In the case of anomalies, and following a risk-based approach, it could:

1. block suspicious transactions and freeze connected accounts within authorised limits;
2. orchestrate enhanced due diligence, requesting documents from the customer and cross-checking them with third-party sources (company registries, sanctions lists, *Politically Exposed Persons* registry, previous reports);
3. perform cross-checks, asking a 'merchant agent' for evidence of the transaction (invoices, order confirmations), correlating timestamps and IP addresses and comparing them with the customer's historical behavioural patterns;
4. compile a Suspicious Activity Report for the regulator, automatically generating a clear, fact-based narrative supported by data and graphs;
5. update its internal rules based on feedback, reducing future false positives.

In short, traditional AI observes and predicts, whilst agentic AI observes, decides, acts, learns, improves and coordinates.

5. AGENTIC AI AND THE TRANSFORMATION OF WORK: WHEN MACHINES BECOME COLLABORATORS

Agentic artificial intelligence is reshaping the labour market in ways that diverge fundamentally from previous waves of automation. These are not merely tools that increase human productivity: artificial agents can participate actively in decision-making processes, learn from experience and autonomously adapt their strategies. This transformation does not replace human beings, but redefines their role, shifting focus from execution to strategic supervision, from direct control to the design of intelligent systems.

A stand-alone generative AI system, such as a Large Language Model, responds to requests, generates content on command and supports human cognitive activities, but remains fundamentally reactive. An artificial agent, on the other hand, can participate actively and proactively in the processes and decisions required.

This capacity for active participation in decision-making, not merely information support, distinguishes agentic AI and defines its specific impact on the labour market. The implications are significant:

- redefinition of operational responsibilities: human workers no longer supervise the execution of tasks but orchestrate ecosystems of agents operating with varying degrees of autonomy;
- increased decision-making speed: agents operate on time scales inaccessible to human cognition (milliseconds in algorithmic trading, micro-level real-time management in smart factories);
- emergence of new forms of collective intelligence: hybrid human–agent teams in which decisions emerge from cooperative interaction rather than traditional hierarchy.

<u>New Role of Human Intelligence</u>

Contrary to dystopian narratives predicting the full replacement of human labour, agentic AI repositions the human being at the centre of the system, but in a qualitatively different role.

This shift materialises in several ways:

- Definition of Objectives and Constraints. Agents operate within decision-making frameworks established by humans. An algorithmic trading agent may autonomously execute millions of transactions, but risk parameters, return targets and ethical constraints (e.g. exclusion of controversial sectors) are defined by human portfolio managers. This initial configuration requires advanced skills: deep domain expertise, the ability to formalise values into computational metrics and to anticipate unintended consequences;
- Continuous Supervision and Intervention by Exception. Even when agents operate autonomously, humans retain veto power and ultimate authority when necessary. Agentic systems include control mechanisms: abnormal situations, high-risk decisions or unforeseen scenarios trigger alerts that, under certain conditions, require human intervention;
- Handling Ethical Grey Areas. Many work situations involve ethical dilemmas without algorithmically determinable solutions. A fully autonomous 'Level 5' vehicle could, in principle, operate without supervision, but decisions concerning whom to protect in an unavoidable crash raise moral questions beyond mathematical optimisation. Programming agents inevitably reflect value-based choices: maximise passenger safety? Minimise total casualties? Prioritise younger lives? These decisions require human collective deliberation and cannot be delegated to autonomous systems;
- Legal Responsibility and Accountability. When an agent causes harm, a drone injures someone or an AI system discriminates unfairly, responsibility falls on human and institutional actors: designers, implementers, supervisors, organisations and firms. Legal systems do not grant legal personality to AI. This regulatory constraint ensures that humans remain accountable, incentivising robust, transparent and auditable system design. However, the distribution of responsibilities amongst multiple actors still needs to be further explored and clarified.

<u>New Professions</u>

The rise of agentic AI is generating professional categories specifically oriented towards the management, supervision and orchestration of artificial agents. These roles differ broadly from both traditional IT occupations and the newer roles associated with generative AI.

Examples include:

- Agentic Systems Architect. Designs architectures for multi-agent systems in which dozens or hundreds of agents cooperate, negotiate resources and coordinate autonomously. This role defines communication protocols, conflict-resolution mechanisms, resource allocation strategies and desirable emergent properties. Unlike a traditional software architect, they must anticipate unintended behaviours that may arise from interactions amongst autonomous agents.

- Agent Alignment Specialist. Ensures that the operational objectives of agents remain aligned with the organisation's and society's values. They use techniques such as Inverse Reinforcement Learning (which infers the 'reward' function underlying an agent's behaviour), implement Constitutional AI mechanisms to encode ethical principles and conduct systematic red teaming to identify unintended behaviours before deployment. This competence is crucial because continuously learning agents can drift progressively away from initial values (a phenomenon known as *value drift*).
- Multi-Agent Coordination Manager. Supervises fleets of autonomous agents operating in physical environments: drones in logistics, collaborative robots in manufacturing, autonomous vehicles in urban mobility. They do not directly control each agent but orchestrate collective behaviours, define dynamic priorities, intervene in anomalous situations and ensure safety. Unlike a traditional logistics manager, this role deals with agents making autonomous tactical decisions; supervision concerns meta-strategies and systemic risk prevention.
- Autonomous Systems Ethicist. Translates ethical principles into operational constraints for autonomous agents. When an organisation states that it wishes to operate 'responsibly' or 'fairly', this professional formalises these values into measurable metrics, verifiable tests and enforcement mechanisms. In finance, they define ethically acceptable trading strategies; in healthcare, they set fairness criteria for algorithms allocating scarce resources; in the justice system, they identify algorithmic biases violating principles of non-discrimination.
- Agentic Interaction Designer. Designs the interface through which humans communicate with agents and supervise their activities. Unlike a traditional UX designer, who optimises interactions with passive software, this role must account for agents that exhibit proactive behaviour, autonomous initiatives, and potentially unpredictable situations; for example, it requires designing ways to enable timely human intervention without demanding constant supervision.

New markets will also emerge, such as an insurance market specifically developed to transfer risks associated with agentic AI.

These professions require deeply interdisciplinary expertise:

- Technical skills in machine learning, multi-agent systems, game theory;
- Strong sectoral domain expertise (finance, healthcare, logistics, etc.) to anticipate practical implications of design choices;
- Ethical and philosophical skills to navigate value dilemmas, balance trade-offs and formalise moral principles into operational constraints;
- Communication skills to mediate between technical and non-technical stakeholders and translate business needs into technical specifications;
- Metacognition, the ability to reflect on one's cognitive processes and how they interact with those of artificial agents.

In addition to emerging professions, all existing professions, starting with cybersecurity, will need to evolve their paradigms within this new technological reality.

A large-scale programme of training and reskilling will be necessary to avoid widening the digital divide. This challenge must be addressed by both enterprises and individuals, with a lifelong learning perspective.

New Divisions of Labour: Co-Agency

The concept of *co-agency*, the cognitive collaboration between human and artificial intelligence, is the emerging paradigm in the labour market transformed by agentic AI. Unlike traditional automation, which replaced human tasks with mechanical processes, and unlike generative AI, which assists humans as an advanced support tool, agentic AI establishes operational partnerships where decisions and actions emerge from a coordinated interaction.

Dynamic division of cognitive labour

In a hybrid human–agent team, task allocation is not fixed but is reconfigured according to the context. In the near future, an operating room equipped with agentic robotic systems might be managed by a human surgeon making high-level strategic decisions (such as surgical approach), whilst robotic agents autonomously execute micrometric precision movements, compensate for tremors and optimise angles. If the system detects a tissue anomaly, the agent will call for the surgeon's intervention. Leadership flows fluidly between human and machine depending on the competencies required.

Structured cognitive complementarity

Humans and artificial agents possess complementary cognitive capacities that, when strategically integrated, exceed the performance of either working alone.

Agents excel at:

- processing massive volumes of data;
- detecting complex statistical patterns;
- multi-parameter optimisation;
- high-speed operations;

- consistent decision-making.

Humans maintain superiority in:

- contextual understanding and common sense;
- deep causal reasoning;
- creativity and lateral thinking;
- navigating ambiguity;
- ethical judgement in complex situations.

Agentic AI allows a systematic integration of these complementarities, not as a sum of separate contributions, but as emergent collective intelligence.

Reciprocal learning and continuous adaptation

In co-agency, not only do agents learn from experience, but humans progressively adapt their cognitive strategies in response to agent behaviour. A trader working with agentic high-frequency algorithms develops intuitions about market patterns initially recognised only by the agents. A radiologist collaborating with diagnostic agents refines attention towards subtle anomalies flagged by agents with high confidence. This co-evolutionary learning enhances and extends human capabilities.

Negotiation and conflict resolution

When networks of autonomous agents collaborate, conflicts inevitably arise from divergent goals to differing interpretations of the same situation and competing priorities. Advanced agentic systems incorporate negotiation and mediation mechanisms inspired by game theory and mechanism design theory.

In a multi-agent smart-grid energy-management system, agents representing producers, consumers and grid operators negotiate prices and allocations continuously. Humans do not manage these negotiations directly but define the meta-rules: what constraints apply? How should efficiency and fairness be balanced? What service levels are required?

This evolution will be accompanied by a shift from highly specialised, 'vertical' work roles towards more integrated roles across organisational departments, dissolving existing internal boundaries (siloes).

New Forms of Organisation

The adoption of agentic AI implies deep organisational transformations beyond technological implementation. Distributed, autonomous systems surpass traditional hierarchical structures, characterised by vertical command chains and centralised decision-making processes.

In organisations evolving towards agentic ecosystems:

- Control becomes coordination: managers do not make every decision but orchestrate systems of agents operating with delegated autonomy within defined parameters;
- Decision-making becomes distributed: there isn't a single decision-maker, but a network of human and artificial actors contributing complementary perspectives and tasks;
- Adaptation replaces rigid planning: instead of following rigid predefined plans, organisations set strategic objectives and allow agents to develop optimal tactics dynamically;
- Transparency becomes an operational necessity: to effectively coordinate autonomous agents, organisations must explicitly formalise objectives, constraints and success criteria that were previously implicit in corporate culture.

New psychological, social and human challenges

The deepest impact of agentic AI on the labour market will be psychological rather than merely economic. For centuries, personal and social identity has been intimately tied to work: what you do defines who you are. As artificial agents assume increasing portions of productive activity, the value of human contribution shifts from efficient execution to the capacity to orient, interpret and attribute meaning.

This transition raises profound philosophical questions:

- What is the role of the human being when operational efficiency is automatable? With agentic AI, humans define what is worth doing, not how to do it;
- How can a sense of fulfilment be maintained if tasks once considered rewarding are delegated to agents? Humans will need to recognise value in strategic supervision, complex orchestration and ethical mediation: domains that are inherently human;
- How should time be distributed amongst work, learning and leisure in an economy where productivity is maximised by agents? Agentic AI could enable shorter working hours without productivity loss, freeing time for creative, relational and caregiving activities – quintessentially human dimensions.

Hannah Arendt distinguishes between *labour* (activities necessary for biological survival), *work* (production of durable objects that transform the world) and *action* (interactions amongst humans that generate political and cultural meaning).

Agentic AI progressively assumes labour and portions of work, but leaves, and indeed expands, the space of action, reserved for humans.

Far from marginalising the human being, this evolution can reposition human intelligence at the centre, at a higher level of abstraction. Human beings become:

- Definers of goals and values: establishing what agents must pursue and what ethical constraints they must respect;
- Strategic supervisors: monitoring emergent behaviours, intervening in exceptional situations and ensuring continuous alignment;
- Interpreters of meaning: translating algorithmic outputs into contextual understanding and assessing implications beyond quantitative metrics;
- Ethical mediators: navigating moral dilemmas, balancing value trade-offs and assuming responsibility for automated decisions;
- Architects of 'cognitive symbiosis': designing ecosystems in which human and artificial intelligence collaborate effectively, maximising complementarities.

To avoid technological unemployment, the labour market in the era of agentic AI requires a radical reconfiguration of skills. Emerging professions (Agentic Systems Architect, Agent Alignment Specialist, Multi-Agent Coordination Manager, Autonomous Systems Ethicist) represent only the beginning. Practically every professional role will need to integrate competencies in supervising artificial agents, orchestrating autonomous systems and interpreting algorithmic decisions.

This transition requires substantial investment in training and reskilling but, above all, a cultural shift: from viewing AI as a threat to recognising it as an amplifier of human abilities; from competing with agents to collaborating with them; from preserving traditional tasks to imagining new ways to generate value.

The ultimate challenge is not (only) technical but human and existential: to build a cognitive symbiosis in which humans remain at the centre not despite agentic AI, but *thanks* to it. Artificial agents free human intelligence from repetitive, computationally intensive tasks, allowing concentration on what is intrinsically and uniquely human: creativity, empathy, ethical judgement, attribution of meaning and building meaningful relationships.

In this perspective, agentic AI does not diminish but amplifies human centrality, elevating humans from executors to strategic guides of technological and social evolution and, above all, of themselves.

6. TOWARDS RESPONSIBLE ADOPTION: GOVERNANCE, REGULATION AND FUTURE PERSPECTIVES

The integration of agentic intelligence into productive and social systems requires a clear and articulated ethical and political vision, embodied in fundamental principles, a regulatory framework, responsible practices at individual, organisational and systemic levels.

Europe, through the AI Act, seeks to position itself as a leader in trustworthy and responsible AI. This approach is not without criticism: some argue that excessive regulation undermines competitiveness; others argue that it is still insufficient to address systemic risks. The debate continues, reflecting genuine tensions amongst legitimate viewpoints.

The European AI Act, which has come into force progressively from 2024, adopts a risk-based approach:

- unacceptable practices (such as governmental social scoring or subliminal manipulation) are prohibited;
- high-risk systems (such as those used for credit assessment or the management of critical infrastructure) must meet stringent requirements concerning data quality, documentation, transparency, human supervision and robustness;
- other systems are not subject to mandatory requirements (except for certain transparency obligations).

The regulatory challenge is to balance risk protection with the promotion of innovation in a domain characterised by extremely dynamic technological evolution, in which generative and agentic models have joined traditional machine-learning approaches.

Excessively strict regulation could stifle research and development, whilst self-regulation would leave room for abuses and harm. Moreover, the global nature of AI technologies requires international coordination: fragmented standards could create trade barriers and opportunities for regulatory arbitrage.

Ensuring alignment, robustness, scalability, and interoperability requires continuous advances in research and development. However, the most profound challenges are moral and ethical. How should institutions and labour markets be redesigned for productive human–AI collaboration? How can economic benefits be distributed fairly, avoiding concentrations of power? How can systems that are natively opaque be developed to respect democratic values of transparency, accountability and fairness? How can human dignity and autonomy be preserved in cognitive ecosystems increasingly mediated by algorithms?

In these complex 'equations', the focus must remain not only on the human being but also on the planet as a whole, with an eco-centric vision, since humanity is not separate from, but deeply interdependent with the ecosystem.

The debate should not be driven by Big Tech alone; instead, it requires interdisciplinary dialogue. Technologists must engage with philosophers, lawyers with social scientists, policymakers with citizens, co-creating shared visions of desirable futures and implementing regulatory frameworks that orient innovation towards the common good.

The full potential and social impact of agentic intelligence will depend on our collective ability to develop broad ethical maturity, one that can sustain new ways of creating meaning and assuming responsibility as social and technological complexity grows. Developing sophisticated algorithms is not enough: society must also cultivate collective wisdom in order to use them responsibly.

Agentic intelligence is not an inevitable fate that befalls us, but a social and technological construction that we can actively shape through individual, organisational and political choices. The crucial question is not whether intelligent machines will arrive (they are already here) or what they can do (increasingly more), but rather what kind of relationship we want with them and what kind of society we wish to build through, and with, them.

The challenge ahead is to build a cognitive symbiosis in which human and artificial intelligences are complementary, amplifying human capabilities without sacrificing values. This requires technical humility (acknowledging limits and uncertainties), ethical ambition (aiming for high standards of responsibility) and collective courage (facing difficult trade-offs with transparency).

Only through such a shared commitment can we ensure that the development of artificial intelligence will contribute positively to social, cultural and economic progress, serving humanity rather than exploiting it, enriching human experience rather than impoverishing it, expanding opportunities for self-realisation rather than generating new forms of alienation. The future of AI is not written. It is up to us, collectively, to determine it through conscious action, democratic deliberation and ethically grounded choices. The challenge is immense, but so is the opportunity to shape a better future for ourselves and for the generations that follow.

JEP

Firenze University Press
https://oaj.fupress.net/index.php/jep

OPEN ACCESS

Citation: Maggiora, E., & Iacobino, C. (2025) EU Artificial Intelligence Act: risks and opportunities for the European system. *Journal of Emerging Perspectives* 2: 97-101. doi: 10.36253/jep-20886

Received: April 30, 2025

Revised: September 30, 2025

Published: December 20, 2025

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

ORCID
EM: 0009-0007-8193-1681
CI: 0009-0006-9930-278X

EU Artificial Intelligence Act: risks and opportunities for the European system

Enrico Maggiora[1,*], Claudia Iacobino[2]

[1] *avvocato, presidente Fondazione dell'Avvocatura Torinese Fulvio Croce, Italy*
[2] *avvocata, presidente AGAT - Associazione Giovani Avvocati Torino, Italy*
e.maggiora@consulenza-impresa.it; claudia.iacobino@it.Andersen.com
*Corresponding author

Abstract. This paper examines the European Union Artificial Intelligence Act (AI Act) as a strategic regulatory response to the rapid and pervasive diffusion of artificial intelligence technologies. Starting from a conceptual framing of AI as an instrument of augmented intelligence rooted in bounded rationality, the contribution highlights how contemporary AI systems adopt satisficing logics through heuristics and large-scale data processing rather than pursuing optimal solutions. The analysis situates the AI Act within the broader EU digital strategy and discusses its risk-based regulatory architecture, which classifies AI systems according to the severity and likelihood of potential harm to fundamental rights, safety, and democratic values. Particular attention is devoted to high-risk AI systems, whose stringent compliance obligations raise significant legal, technical, and economic challenges, especially for SMEs and start-ups. While acknowledging the risks of regulatory rigidity, innovation slowdown, and market entry barriers, the paper also emphasizes the strategic opportunities generated by the AI Act. These include the consolidation of a trustworthy, human-centric AI ecosystem, the strengthening of the Digital Single Market, and the potential emergence of a global regulatory benchmark through the so-called "Brussels effect." Ultimately, the AI Act is interpreted not merely as a compliance burden, but as a long-term investment capable of transforming regulation into a competitive advantage for the European system.

Keywords: Artificial Intelligence Act, EU Digital Regulation, risk-based approach, high-risk AI systems, Fundamental Rights Protection, SMEs and Innovation, trustworthy AI, Brussels Effect.

1. THE AI REVOLUTION AND THE NEED FOR REGULATION.

Great revolutions are, for the most part, silent. By the time we realize that an epoch-making change is underway, that change has already occurred, and often, few have seen it coming. In the collective imagination, great revolutions find their seeds of development in a sensational, often violent act that immediately attracts public and media attention, masking and silencing the transformations that take place quietly and which, paradoxically, are the ones destined to make the most noise. The human mind struggles to project and understand the long-term effects of slow but inexorable transformations and to perceive the real extent of change in the short term.

This is particularly true in the case of technological revolutions, which are never sudden but are the result of a process of gradual development and implementation, the consequences of which, however, are often disruptive.

The advent of Artificial Intelligence (AI) is undoubtedly one of the most discussed and significant IT revolutions in contemporary history. A new language of writing so sophisticated that it can simulate human intellectual processes. It is difficult to give a single, rigorous definition of 'Artificial Intelligence', and it therefore seems appropriate to refer to some authoritative elaborations. Artificial Intelligence is, therefore, '*the study of how to make computers do things that, at present, people are better* at'[1] and, furthermore, '*computational intelligence is the study of the design of intelligent* agents'[2].

However, at the definitional level, even more significant is the concept of '*bounded* rationality' developed by Nobel Prize winner Herbert Simon, which has had a crucial impact on the emergence and design of AI-based computing systems, forming their theoretical basis. According to Simon, human decision-makers are endowed with limited rationality because they are subject to cognitive constraints, time constraints, and information limitations, as it is impossible to access all relevant information or process it completely. Therefore, decisions that, at least on the surface, appear to lack reasoning and do not guarantee an optimal result are, on the contrary, perfectly suitable (i.e., rational) insofar as human beings, thanks to experience, are aware of their own limitations and the context of reference. Our intellect inhibits the expenditure of resources and does not seek optimal solutions, but satisfactory ones. Human thinking is certainly fallible, but economical in that it aims to avoid wasting time and energy and to achieve a functional result.

Even the latest generation of Artificial Intelligence adopts a *satisficing* rather than maximizing methodology. This approach involves the pervasive use of heuristics and algorithmic cognitive shortcuts, aimed at identifying a pragmatic optimal solution (*good enough solution*) within predefined and stringent time limits, rather than searching for the *perfect solution*, which would be impractical or asymptotic in terms of cost and time.

AI, far from being a mere emulation of human intelligence, thus rises to the rank of a cognitive- amplifying tool aimed at overcoming the intrinsic cognitive limitations of the decision-maker. AI algorithms exhibit a capacity for massive data processing (Big Data Analysis) that exceeds the threshold of human processing, thus increasing the 'information blade' of Simon's scissors (i.e., the ability to acquire and process relevant information). The primary objective of this integration is not to replace human rationality, but rather to optimize and augment it (*augmented intelligence*).

Artificial Intelligence, by virtue of its computational properties and algorithmic nature, exerts a highly pervasive, transformative influence, reshaping the operational structures of every strategic sector of the economic and social fabric: from industrial automation to healthcare, from finance to public administration. Increasingly sophisticated Artificial Intelligence systems, in particular General-Purpose AI Systems (GPAI) and generative AI, have demonstrated extraordinary innovative potential. This rapid technological evolution also raises crucial ethical and legal questions relating to transparency, security, fairness, and the protection of fundamental rights.

The need for supranational regulation of Artificial Intelligence at the continental level has become a legal and strategic imperative to ensure that the penetration and adoption of this technology takes place in accordance with the principles of ethics, safety, and transparency. This need is motivated by the (real) risk of compromising the fundamental values of the European Union, arising from specific systemic dangers such as algorithmic discrimination (linked to *biases* inherent in *datasets*), *mass surveillance*, behavioral manipulation of individuals, and the opacity and poor traceability of synthetic content (e.g. *deepfakes*).

In addition to these critical issues, there is also a need to prevent the fragmentation of the Digital Single Market caused by the emergence of heterogeneous and divergent national regulations. In response to this complex scenario, and in line with the '*A Europe Fit for the Digital* Age' policy strategy, the European Union has developed and adopted the Artificial Intelligence Act[3] (*abbreviated to* 'AI Act'). This regulatory document establishes a harmonized legal framework at EU level, adopting a *risk-based approach*. This methodology provides for the taxonomic classification of AI systems according to the probability and severity of the potential harm they may cause (divided into categories ranging from 'unacceptable risk' to 'minimal risk').

---

[1] E. Rich, K. Knight, *Artificial Intelligence,* McGraw-Hill, 1991, *"The study of how to make computers do things that, at the moment, people are better at"*.

[2] D.L. Poole, A.K. Mackworth, R. Goebel, *Computational intelligence: A logical approach,* Oxford University Press, 1998.

[3] Full title: *"Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828"*.

The AI Act is not exclusively an *ex-ante* precautionary measure, but rather an ambitious attempt to strike a balance between the demand for technological innovation and the duty to protect fundamental rights. The implicit objective of the Regulation is to project a *de facto* global standard for anthropocentric and trustworthy AI governance, achieving the so-called 'Brussels effect'[4] at an international level.

2. THE EU AI ACT

A general overview

The AI Act is one of the pillars on which the European Union's digital strategy is based, which aims to achieve a virtuous balance between supporting innovation and competitiveness in Member States and protecting the fundamental rights of citizens, workers, and consumers. The process of approving the legislation began on 21/04/2021 with the presentation of a proposed Regulation by the European Commission. Since its inception, the regulatory framework has expressed the intention to establish a harmonized and proportionate regulatory framework for the regulation of Artificial Intelligence within the Union.

This objective has led to the adoption of a risk-based approach, structuring the proposal around a taxonomic classification of AI systems according to the level of potential harm (i.e., risk) they may pose to the safety, health and fundamental rights of individuals. This classification forms the regulatory basis from which the legal definitions, compliance requirements (*ex ante* and ex post obligations) and specific burdens imposed on the economic operators involved (suppliers, importers, distributors and users) in relation to the place on the market and use of such systems derive directly.

On 14/06/2023, the European Parliament formalized its negotiating position on the document and, subsequently, on 9/12/2023, the Trilogue composed of the European Parliament, the Council and the European Commission negotiated and signed a provisional agreement on the content of the text of the regulation. At a technical meeting on 21 January 2024, the European Commission made the final version of the text of the regulation available. From that point onwards, the approval process was fairly rapid: in March, the Strasbourg Assembly confirmed the final text by a large majority, and in May, the EU Council definitively adopted the AI ACT in its current version.

The European Regulation on Artificial Intelligence pursues the stated dual purpose of improving the functioning of the internal market and stimulating the spread of human-centric and reliable Artificial Intelligence (AI) systems. Complementarily, it establishes a high standard of protection for health, safety and fundamental rights protected by the Charter of Fundamental Rights of the European Union (including democracy, the rule of law and environmental protection) from the harmful effects of the use of AI systems, while also promoting innovative development. Furthermore, the fundamental objectives pursued by the AI Act include: the establishment of a regulatory framework complementary to the General Data Protection Regulation (GDPR) in relation to specific aspects of personal data protection connected to AI, the regulated management of risks arising from the use of Artificial Intelligence systems, and the promotion of technological development that is both safe and controlled.

The subjective scope of the Regulation essentially concerns two players, in particular: 'Suppliers', which includes any entity, whether public or private, that makes the AI system available on the European Union market, and 'Operators', which refers to all entities responsible for the use of the AI system in Europe and the resulting output, regardless of their geographical location. In addition to these figures, there are importers, distributors and authorized representatives, who participate in various ways in the circulation or use of AI platforms. Platforms intended for military, defense, national security purposes and those dedicated exclusively to scientific research and development are explicitly excluded.

Compliance with the obligations under the AI Act for each of the above-mentioned entities is modulated differently according to the specific use of the technology and, crucially, according to the level of intrinsic risk associated with the AI systems adopted by the organization.

AI practices posing an unacceptable risk constitute the fine line between whether an AI system can be considered lawful or not. AI systems that contravene the fundamental values of the Union and that may pose a significant threat to the security and freedom of individuals, through techniques designed to manipulate the consent or decision-making of citizens, manipulating or exploiting people's vulnerabilities or creating unfair discrimination, are considered to be of 'unacceptable risk' and are therefore prohibited.

Chapter III, Section 1 is entirely devoted to AI systems classified as 'high risk', for which there is no ban, but very stringent technical and procedural obligations are imposed. An AI practice can be considered 'high

[4] Anu Bradford, *The Brussels Effect, How European Union rules the Word,* Oxford Press University

risk' when its use can significantly affect people's health, safety or fundamental freedoms. Suppliers of high-risk AI systems must comply with stringent and ongoing obligations and

ensure an adequate level of *compliance*. The placement on the market of such systems is always subject to a mandatory conformity assessment.

Finally, we come to AI practices considered 'low risk', which are subject to much milder obligations, aimed primarily at ensuring the transparency of AI use processes and self-regulation. The risk associated with these systems is not structural but is limited to cases where the user is not adequately informed about their interaction with the machine. As mentioned, the AI Act does not impose any particular legal obligations on such systems but encourages adherence to voluntary codes of conduct for the promotion of ethical and safe practices.

3. RISKS AND OPPORTUNITIES FOR THE EUROPEAN SYSTEM.

The introduction of such a detailed regulatory framework for artificial intelligence is an important strategic choice for the European Union, although it exposes the continental system to a number of intrinsic technical, legal and economic risks that EU institutions and operators in the sector are inevitably called upon to address.

In the author's opinion, one of the most challenging aspects of the Regulation lies in the correct legal classification of AI practices, with particular reference to 'high-risk' systems. This classification presents the greatest degree of uncertainty in terms of definition, even though it is the prerequisite for the activation of *the* highest *set* of *compliance* obligations. It is also essential to consider the exponential speed at which AI is evolving. The rigidity of the list contained in Annex III, which defines 'high-risk' systems, risks becoming obsolete with the emergence of new applications (e.g., the interaction between basic models and specific *downstream* systems), making it necessary for the European Commission to constantly and promptly review and update it.

Furthermore, the compliance burdens imposed on suppliers of 'high-risk' systems are extremely detailed and technically complex, involving the entire production chain.

The obligation to guarantee data quality and *governance* requires burdensome procedures to ensure the representativeness, absence of *bias* and traceability (*data provenance*) of training, validation and *testing* data. In addition to the already complex framework described above, there is also the obligation to maintain detailed automatic records as referred to in Article 12 (*logging capabilities*), which are necessary to enable the traceability of operations throughout the entire life cycle of the system, with obvious implications in terms of *data security* and *privacy* management. Although the entire compliance assessment process is designed to ensure security, it requires considerable resources and specialist expertise, which clearly places a particular financial burden on SMEs.

An excessive compliance burden, particularly for start-ups and SMEs, runs the real risk of slowing down technological innovation in Europe. The high cost of technical and documentary requirements (Art. 16 et seq.), combined with the risk of penalties, could discourage small businesses from developing or using 'high-risk' systems, limiting their competitiveness compared to global *competitors* operating in less regulated jurisdictions.

In summary, the AI Act sets an ambitious ethical and safety standard, but its effectiveness will depend on the ability of European institutions and economic operators to overcome these technical, governance and economic sustainability challenges.

In any case, despite the potential risks described above, the AI Act is not only a burden for European operators and, under certain conditions, also for the legislator, but is also a strategic lever capable of generating significant economic and geopolitical opportunities. Regulation of this area can, in fact, constitute a competitive global advantage, transforming compliance into a real market *asset*. The AI

Act, as was the case with the adoption of the GDPR, can easily generate the infamous 'Brussels effect'. By imposing high *standards* on anyone wishing to operate in the Single Market, the EU is effectively pushing global economic players to adopt its *standards* worldwide, optimizing production processes and avoiding duplication of systems. As the first bloc in the world to establish a comprehensive and binding regulatory framework for AI, Europe can shape the international rules of the game, influencing future legislation in third countries.

The Regulation also defines a global benchmark for trustworthiness and safety. The 'Made in the EU, AI Compliant' label can become an internationally recognized mark of quality, facilitating the export of European technologies. The harmonization of standards at European level eliminates regulatory fragmentation between Member States, reducing *compliance* costs for European companies and creating a digital Single Market for AI. Stringent requirements for 'high-risk' systems impose rigorous development processes that not only increase safety but also enhance the technical quality and reliability of European products, making them more attractive.

The AI Act should therefore be interpreted as a long-term strategic investment: an initial compliance cost that generates a return in the form of consumer confidence, international competitiveness and technological autonomy.

JEP

Firenze University Press
https://oaj.fupress.net/index.php/jep

OPEN ACCESS

Citation: Rocchietti, G. (2025) Entrepreneurship in the age of AI and venture capital. *Journal of Emerging Perspectives* 2: 103-105. doi: 10.36253/jep-20887

Received: April 30, 2025

Revised: September 30, 2025

Published: December 20, 2025

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

Entrepreneurship in the age of AI and venture capital

Giancarlo Rocchietti

Founder & President Club degli Investitori, Italy
giancarlo.rocchietti@clubdeglinvestitori.it

Abstract. Entrepreneurship has transformed under the influence of venture capital and artificial intelligence. Success increasingly depends on access to capital, speed of execution, adaptability, and the ability to build strategic networks. This shift has produced serial and parallel entrepreneurs, accelerated innovation, and reshaped markets. Yet it has also amplified inequality, raised ethical concerns, and forced society to question whether technology is truly creating a better world. In this evolving landscape, entrepreneurs and investors hold unprecedented power – and equally unprecedented responsibility – to ensure innovation serves society, not just markets.

Keywords: entrepreneurship, venture capital, artificial intelligence, startup ecosystems, power law, innovation.

1. INTRODUCTION

Over the past two decades, entrepreneurship has undergone a profound transformation.

Some fundamental qualities of successful entrepreneurs remain timeless – vision, passion, commitment, leadership, and hard work.

But the way founders build and scale companies has changed dramatically – first with the rise of venture capital, and now with the acceleration driven by artificial intelligence.

Here are some of the most important shifts shaping modern entrepreneurship.

2. CAPITAL IS ABUNDANT BUT ACCESS REQUIRES SKILL

When I founded Euphon, my media company, in the 1980s, the only financing option available was the bank. Every quarter, I would sit across from the branch manager, negotiating loan amounts and interest rates.

Today, the situation is radically different. Startups now have access to more than 40,000 active venture capital investors, who collectively invest around 350 billion dollars into more than 25,000 companies each year. And unlike traditional financing, this capital is equity – not debt.

Although this represents only 2 percent of global financial assets, it has become the most powerful engine for innovation and company creation. Without angel investors and venture capital, companies like Apple, Nvidia, Facebook, or Google simply would not exist.

In the age of AI – where capital has become almost a commodity – a new entrepreneurial capability is emerging: founders succeed not only when they build products with market fit, but also when they build a strong network of investors, advisors, and partners who support them from the early stage to scale.

This requires time, dedication, emotional intelligence – and the ability to tell compelling stories.

3. THE POWER LAW − ONLY A FEW SUCCEED

The combination of entrepreneurship, innovation, and capital accelerates the birth and growth of startups, but it also dramatically increases competition.

In a world where ideas spread instantly and can be replicated globally, success depends on the ability to raise strategic capital and scale fast.

Examples include:

- Airbnb became the largest hospitality platform in less than a decade – without owning a single hotel
- Uber reshaped urban mobility in just a few years
- Zoom became the global standard for videoconferencing almost overnight

It may seem simple – but it is not. Competition is intense, and timing and speed are essential. Venture capital initially funds many startups, but over time, capital concentrates on a small number of winners – the unicorns.

This dynamic is known as the Power Law, rooted in the Pareto Principle. Vilfredo Pareto, an Italian economist of the 19th century, demonstrated that in complex systems, 20 percent of causes generate 80 percent of results.

In startup terms: only two out of ten will succeed.

Failure is now common – and increasingly accepted as a learning step. That represents a major cultural shift from the past, particularly in Europe, where failure once meant the end of the journey.

4. THE RISE OF SERIAL, PARALLEL, AND VENTURE ENTREPRENEURS

The old model – one entrepreneur, one company, one lifetime – is disappearing. The idea of building a company as a family legacy, which I once imagined for myself, is no longer the dominant path.

Today, many founders exit after about ten years and then start another company, become angel investors, or launch their own venture funds.

Examples include:

- Stefano Buono, a former CERN physicist who founded AAA (a theragnostic biotech company sold to Novartis for $4B), then launched Planet Smart City, LIFTT (a deep-tech VC fund), and now Newcleo, raising more than $500M to build next-generation nuclear reactors. I am a Newcleo co-founder.
- Elon Musk, who evolved from Zip2 and PayPal into a parallel entrepreneur leading SpaceX, Tesla, SolarCity, Neuralink, The Boring Company, and now X – simultaneously.

My own journey reflects this shift.

I am an electronic engineer and a baby-boomer entrepreneur. I founded Euphon at 26, grew it, listed it on the Italian Stock Exchange, and eventually sold it to a private equity fund.

At that point, I became an angel investor.

I also knew I didn't want to do it alone. So, together with four friends, I founded Club degli Investitori, now the largest Italian angel investor network in the world – with more than 400 entrepreneurs, managers, and professionals committed to selecting, investing in, and supporting the best Italian tech founders worldwide.

This work has allowed me to share my experience, learn from brilliant minds, shape the future of innovation, make friends, create value – and enjoy the journey. Paradoxically, while I am no longer running a company myself, I feel like an entrepreneur of entrepreneurs – a venture capital entrepreneur.

5. FROM RESILIENCE TO ADAPTABILITY

In a world that moves at high speed, success requires more than resilience. It requires adaptability – the ability to pivot, rethink the model, and seize new opportunities even amidst uncertainty.

Carl Schramm, former CEO of the Kauffman Foundation, argues in his book Burn the Business Plan that most successful entrepreneurs don't start with detailed planning – they start by building, selling, and learning.

A clear example is Luca Rossettini, who founded D-Orbit at age 27 with the goal of deorbiting inactive satellites. Since the market wasn't ready, he radically shifted the business model and instead developed technology to launch and reposition active satellites in orbit – a pivot that made the company relevant and scalable.

I was one of the first angel investors in D-Orbit, and I am proud to be a shareholder in one of the most promising space-economy companies in Europe.

6. MAKING THE WORLD A BETTER PLACE

For baby-boomer entrepreneurs, the goals were straightforward: build great companies and generate profit. Social impact meant creating jobs.

For me at that time, making the world a better place was simply a nice phrase – a line from a Michael Jackson song.

The Google founders were among the first to articulate a different kind of mission:

"To organize the world's information and make it universally accessible and useful."

Since then, many entrepreneurs have embraced the idea that technology – and especially AI – can cure diseases, extend human life, and address climate change and inequality.

As a business angel, while financial return matters, I am proud to contribute to an ecosystem that creates positive impact and aspires to build a better future.

7. ETHICS

With technological acceleration, geopolitical instability, and rising inequality, we may find ourselves wondering whether we are truly making a better world – or whether something is going wrong:

- wealth is concentrating in very few hands;
- algorithms influence behavior in ways we do not fully understand
- innovation is not always accessible

Consider, for example, CRISPR-based gene therapies: a treatment for sickle-cell disease – affecting more than seven million people – now exists. But the cost is about two million dollars per dose.

So only a few can afford it – while the company is valued in the billions.

This raises uncomfortable questions:

- what future are we building?
- will innovation truly be accessible to all?

Some envision a future like the film Elysium, starring Jodie Foster and Matt Damon – a world where a small elite prospers while many are left behind.

Which leads to a fundamental reflection:

What role do we – as investors and entrepreneurs – want to play?

Ethics is a deeply personal space, and there are no simple answers.

But every investment decision shapes not only a business – but society. And we have the responsibility to reflect on that.

Think, for example, about the rise of the defense-tech sector – the new hype. Suppose a startup builds autonomous drones.

How can we be sure they will be used for protection – and not for destruction?

CONCLUSION

New tech founders – and the venture capitalists who back them – hold extraordinary power.

Their responsibility is not only to combine innovation, capital, and entrepreneurship to generate returns, but also to consider the society they are shaping.

We are not here merely to fund startups.

We are here to build the future – our future.

Book Reviews

With this second issue, the Journal of Emerging Perspectives also launches a new section dedicated to reviews of books that are of particular relevance to the themes and topics addressed by the Journal.

Enrico Orrù, *Applicare l'intelligenza artificiale alla gestione dei documenti della pubblica amministrazione. Esperienze e prospettive*, Orizzonti Tecnologici Series, NVGroup, 2025.

Applying Artificial Intelligence to Document Management in Public Administration: Experiences and Perspectives by Enrico Orrù examines the digital transformation of the Italian Public Administration by placing AI at the center of a transformation that is both technological and institutional.

After outlining the regulatory framework, the volume illustrates how artificial intelligence can be applied to concrete administrative processes, including advanced OCR, automatic classification, information extraction, case and workflow routing, and semantic search.

The Italian and international case studies provide replicable examples for administrations of different sizes, while particular attention is devoted to data protection, algorithmic bias, security, and decision traceability, addressed through governance and oversight models.

Alongside the technical dimension, the book presents a vision of public policy oriented toward the creation of public value and the strengthening of trust among institutions, citizens, and businesses.

Written in accessible language and adopting a pragmatic approach, it represents a valuable guide for public executives, ICT managers, and practitioners responsible for governing the adoption of AI within Public Administration.

The text might have benefited from greater detail regarding implementation timelines and adoption costs for smaller entities; nevertheless, it remains an essential reference for understanding how AI can make Public Administration more efficient and better aligned with the creation of public value.

Antonella Forte

Jason Bell, *The one-person billion-dollar company*. Skylark Publishing Group, 2024

The book "The One-Person Billion-Dollar Company" is essential reading for anyone seeking to understand the landscape envisioned by the incessant and rapid developments in digitalization. The author explores how the dynamics of agentic artificial intelligence, cognitive automation, and distributed digital infrastructures are radically redefining the very concept of enterprise. The focus is predominantly on the understanding that productivity, in its most traditional sense, is no longer exclusively linked to the quantity of human labor. The entrepreneur's role is shifting from a manager of a large workforce to a manager of "intelligent systems". It's no longer about building complex hierarchies, but about conceiving artificial intelligences capable of operating and generating value, elevating the entrepreneur's role to that of an "architect of cognitive systems."

The volume offers an in-depth theoretical analysis of the unprecedented economies of scale that these technologies make possible. Through the synergistic use of autonomous agents, modular APIs, advanced language models, and sophisticated composable cloud services, the chance emerges of conceiving an enterprise almost entirely based on automated cognitive systems. These digital entities can not only operate with minimal human intervention but are also capable of adapting and evolving autonomously, exponentially amplifying their effectiveness and ability to generate profit.

"The One-Person Billion-Dollar Company" is not mere speculation about the future, but a conceptual tool that highlights the birth of a new 21st-century business organization. Moving beyond the traditional business model based on extensive workforces and manual processes, the book shows us how the paradigm is shifting towards intelligent architectures, designed to operate and develop value autonomously and scalably. In this scenario, traditional objectives related to scale, growth, and productivity measurement are overturned by types of businesses that are not only leaner and more efficient, but potentially managed by a single individual, acting as a catalyst and supervisor of a systemic ensemble of artificial intelligences. Therefore, the book is particularly useful for managers, those involved in technology and innovation, or simply anyone who wishes to understand the forces shaping our economic future.

Maria Giuseppina Lucia

Salma Khan, *How AI Will Revolutionize Education (and Why That's a Good Thing)*. Viking Penguin Press, 2024

Salma Khan's book, highlighted here, is a fundamental work for understanding the implications of artificial intelligence in the contemporary educational landscape. The author explores the undeniable reality that AI is not simply a technological tool but an "entity" capable of reshaping every aspect of the learning and teaching process. Khan does not limit herself to merely listing new software applications in education but poses crucial questions that involve the very essence of education. Among the questions the author raises, those related to access and equity - i.e., how AI can influence equality of educational opportunities - take on particular importance. Other themes explored include students' intrinsic motivation, the formation of trust in the teacher-student relationship, and the development of autonomous critical judgment in an era dominated by data and algorithms. Indeed, AI brings back to attention the much-debated issue of the personalized teacher-student interconnession , often hindered by large class sizes. On the one hand, AI offers the unprecedented possibility of providing highly personalized educational support to millions of students, adapting to individual learning times and methods in ways previously unbelievable, thereby democratizing access to high-quality education and overcoming many current limitations. However, Khan simultaneously points out that risks exist, such as the potential amplification of phenomena, dependence on technology, the spread of disinformation generated by autonomous systems, and the temptation to resort to expedients that bypass deep understanding, if the use of AI is not managed carefully.

In Khan's work, there is neither boundless enthusiasm nor a priori rejection of AI, but a balanced and pragmatic approach. Ultimately, the author extends her analysis beyond the boudaries of the school alone, positioning the discussion on AI in education as a critical point for a society's ability to cultivate human potential on a large scale. She questions how this technology can be leveraged to enhance the level of collective competence and knowledge without sacrificing individual autonomy, critical thinking skills, and the deep sense of engagement that are essential for human development. "Brave New Words" is therefore an exhortation to complex and necessary reflection to navigate the AI revolution in the educational process with awareness and foresight.

Maria Giuseppina Lucia

Local Partners

Journal of Emerging Perspectives is pleased to announce the establishment of the *"Local Partners" Group*, a new initiative designed to strengthen the connection between scientific research and the expertise that emerges and develops within local communities and territories.

The group is composed of professionals with diverse backgrounds and areas of expertise, capable of offering a broad, nuanced and multidimensional perspective on the complexity of the contemporary world. Their contribution represents a valuable complement to the work of the researchers involved with the journal, enriching the discussion with perspectives firmly rooted in practice and local contexts.

The Local Partners will meet several times throughout the year, both online and in person, with the aim of jointly identifying and defining the themes and structure of the topics to be developed in the annual issue of *JEP*. This is a significant role for the future of the journal, as it paves the way for an editorial programme that is increasingly responsive to real-world needs and emerging issues.

At the same time, these meetings will provide opportunities for dialogue and the exchange of experiences and expertise which, we hope, will prove enriching for all participants as well as for the project as a whole.

The group is distinguished by its marked diversity, fully reflecting the interdisciplinary spirit that has characterised the *InCreaSe Association*—the parent organisation of the journal—since its foundation.

This diversity of perspectives reinforces our belief that the Local Partners will make a significant contribution to the scientific work we pursue and provide substantial added value to the development of the journal.

Members of the Group

- *Emilia Caizzo*, biologist, InCreaSe researcher, and expert in intergenerational relationships.
- *Roberto Cardaci*, sociologist and expert in poverty analysis.
- *Dario Casalini*, entrepreneur and expert in public law.
- *Alice Cerutti*, rice farmer and promoter of the Oasi Lacustre initiative.
- *Carlo Chiattone*, expert in the evolution of animal and human behaviour.
- *Franco Cirio*, Registered Nurse and expert in public healthcare issues.
- *Alessandro Fiumara*, economist and expert on the National Recovery and Resilience Plan (PNRR).
- *Fabrizio Floris*, sociologist and expert in urban periphery issues.
- *Antonella Forte*, sociologist, InCreaSe researcher, and expert on Generation Z.
- *Andrea Ghignone*, Mayor and President of Cantina Sociale di Nizza Monferrato.
- *Marco Giletta*, Chief Executive Officer and Commercial Director of Assist.
- *Dario Menasce*, nuclear physicist, author, and expert in science communication.
- *Enrico Orrù*, entrepreneur and expert in applied Artificial Intelligence (AI).
- *Giovanni Periale*, engineer and founder of Ethica.
- *Cinzia Tortola*, Registered Nurse and expert in women's health and gender-based violence.

Journal of Emerging Perspectives Vol. 2 – 2025