Journal of Artificial Intelligence and AI Ethics

SCIVOLVE

Research Article

Open Access Full Text Article

AI as Cognitive Ecology: Revealing the Invisible Cognitive, Cultural, and Epistemic Costs of Generative Models

Godson Ozioma[1], Sebastian Obeta[2*], Dr Ikpe Ibanga[3], Linda Oraegbunam[4], Ruth Imanria Itua[5], Dr Augustina Amaefule[6], Emmanuel Ozioma Ozioma[7], Chidinma Anumaka[8]

[1]*University of Salford.*
[2]*Director of Research and Development.*
[3]*School of Health, Leeds Beckett University.*
[4]*Data Science Researcher Sycom Labs.*
[5]*Security Operations Analyst; Epaton LTD.*
[6]*University of Gloucestershire.*
[7]*Research Fellow, Digievolve hub limited.*
[8]*University of Nottingham.*

***Correspondence:**

Sebastain Obeta
Cambridge University UK.

Received: November 30, 2025; **Published:** January 16, 2026

How to cite this article:

Ozioma, G., Obeta, S., Ibanga, I., Oraegbunam, L., Itua, R. I., Amaefule, A, Ozioma, E. O. and Anumaka, C. (2026) 'AI as Cognitive Ecology: Revealing the Invisible Cognitive, Cultural, and Epistemic Costs of Generative Models', *Journal of Artificial Intelligence and AI Ethics*, vol. 1, no. 1, pp. 1–16.

Abstract

Recent debates on Generative Artificial Intelligence (GenAI) have centred on quantifiable concerns such as computational cost, carbon emissions, and benchmark performance. Yet the most consequential risks may be those that are less visible: the gradual reshaping of human cognition, creativity, and epistemic trust. This paper introduces the concept of AI as cognitive ecology, situating generative systems not merely as tools or agents, but as a pervasive environment in which thought now unfolds. Building on this paradigm, we propose the HORIZON taxonomy of invisible costs: Homogenization, Offloading (deskilling), Resource externalities, Information integrity, Zoomed-in feedback loops, Organizational memory loss, and Normative drift. We illustrate each dimension through conceptual analysis and lightweight audits, and propose new indicators including DAO (Diversity of AI Outputs), CDQ (Cognitive Dependence Quotient), EIS (Epistemic Integrity Score), and RTE (Resource Transparency Equivalent). We argue that sustaining AI innovation requires not only technical and environmental monitoring, but active stewardship of cognitive ecologies.

1. Introduction

The landscape of modern technology is undergoing a profound transformation with the rapid ascent of Generative Artificial Intelligence (GenAI). Defined by their capacity to produce novel content ranging from text, images, and audio to software code, video, and simulations based on extensive training datasets, GenAI models have quickly permeated various sectors, captivating widespread attention and promising to revolutionise business and everyday life (Rana *et al*., 2024). Predictions from leading industry analysts, such as Gartner, anticipate that over 80% of enterprises will have integrated GenAI APIs or deployed GenAI-enabled applications by 2026, underscoring its swift adoption and perceived transformative power (Rana *et al*., 2024). This pervasive integration is driven by GenAI's immense potential to enhance organisational effectiveness, streamline operations, and provide a significant competitive edge (Wamba *et al*., 2023). For instance, GenAI is already being leveraged across diverse business functions, including marketing, project management, data analysis, customer relationship management, content creation, human resources, employee training, and coding. In software engineering, GenAI tools are envisioned as indispensable allies throughout the development lifecycle from ideation and architectural design to code generation, testing, deployment, and maintenance, with projections suggesting a 20-45% surge in productivity by automating tasks such as drafting code and performing root-cause analyses (Russo, 2024). The immediate popularity of models such as ChatGPT, following its public release in November 2022, highlights a demonstrated human-like competence in diverse areas, further accelerating its integration into critical organisational processes (Rana *et al*., 2024). This widespread, fast-paced adoption positions GenAI not merely as an incremental improvement but as a disruptive innovation poised to fundamentally reshape industries (Russo, 2024).

The rapid proliferation and increasing societal reliance on AI systems have, rightly, spurred extensive ethical and societal debates. Current discourse within AI ethics largely centres on tangible, quantifiable costs and concerns, predominantly revolving around three interconnected pillars: accuracy and performance, bias and fairness, and environmental and resource consumption.

Accuracy and performance are paramount concerns, as AI models are increasingly deployed in high-stakes domains such as healthcare, finance, and criminal justice, where precision is critical (Nathim *et al*., 2024). Ensuring that AI systems deliver optimised and timely outcomes, as expected by users, is a key ethical consideration (Rana *et al*., 2024). However, research continually highlights the challenges in guaranteeing accuracy, particularly given that GenAI models, relying on pre-trained data and algorithms, can generate imprecise results due to inherent flaws in their training data or underlying logic (Balasubramaniam *et al*., 2023). The functional integrity of AI-generated code, for instance, remains a fundamental concern in software engineering, with varying degrees of success reported depending on task complexity and programming language (Russo, 2024).

Bias and fairness represent a complex and multifaceted ethical challenge, recognised as critical for establishing trust and ensuring equitable treatment across diverse populations (Nathim *et al*., 2024; Russo, 2024). Biased AI systems can inadvertently perpetuate and even exacerbate existing societal discrimination, stereotyping, and social inequities (Modi, 2023). Sources of bias are deeply embedded throughout the AI lifecycle, from data collection and human annotation to algorithmic design itself (Nathim *et al*., 2024). While various bias mitigation techniques, such as data pre-treatment, algorithmic adjustments, and adversarial testing, have been proposed, they often involve trade-offs with model performance and accuracy, complicating efforts to balance fairness and efficacy. Furthermore, the absence of universal methods for assessing fairness and a consensus on appropriate metrics underscores the ongoing difficulty in systematically identifying and mitigating bias, necessitating accessible and easily integratable tools and frameworks for practitioners.

Parallel to these operational and social concerns, the environmental and resource costs of AI, particularly GenAI, are emerging as a critical, albeit often overlooked, ethical dimension. GenAI models are distinctly resource-intensive, contributing significantly to carbon dioxide emissions and demanding massive amounts of water and land for their operation (Kneese and Young, 2024). The overall energy consumption of the Information and Communication Technologies (ICT) sector, which includes AI, is rapidly escalating, with global data centre electricity usage increasing by 20-40% annually and straining existing renewable energy infrastructures (Luccioni *et al*., 2024; Inie *et al*., 2025b). Crucially, the inference (deployment) phase of Machine Learning (ML) models, often overlooked in favour of training costs, can account for a substantial portion (80-90%) of total cloud computing demand and contributes as much or more to environmental impact (Luccioni *et al*., 2024). Multi-purpose, generative architectures, such as GPT models, are orders of magnitude more environmentally costly than task-specific systems, yet a lack of transparency from major model providers regarding training and hosting data complicates accurate assessment and accountability (Inie *et al*., 2025b). This calls for the integration of environmental factors, including downstream impacts on communities and ecosystems, directly into the design space of AI, advocating for a holistic approach to sustainable AI development (Kneese and Young, 2024).

While these discussions on accuracy, bias, and environmental impact are vital, they primarily address the visible or immediately quantifiable aspects of AI's societal footprint. This paper argues that a crucial dimension of AI's impact remains underexplored: the invisible cognitive, cultural, and epistemic costs associated with the widespread adoption and integration of generative models. This gap arises from a prevailing tendency in AI development and evaluation to valorise influential benchmarks as objective markers of progress, despite their inherent limitations in capturing complex, real-world capabilities and broader societal implications (Raji *et al*., 2021; Eriksson *et al*., 2025).

The existing critique of AI benchmarks, though extensive, still struggles to fully articulate these invisible costs. Researchers have demonstrated that many benchmarks suffer from construct validity issues, often failing to measure what they claim, especially when aspiring to assess "general" or "universal" capabilities (Raji *et al*., 2021). This problematic framing leads to misguidance in task design, underreporting of inherent biases, and the potential misuse of models based on false performance presentations. The historical Common Task Framework (CTF), designed for tightly-scoped, practically-oriented tasks, has been inappropriately extended to abstract "performance," promoting "glamour and deceit" over meaningful progress (Raji *et al*., 2021). These evaluations are often based on "samples of convenience" rather than systematically chosen, theoretically sound tasks. This narrow focus extends to the modalities evaluated, with a vast majority of benchmarks concentrating on text-based AI, leaving other crucial modalities such as audio, images, and multimodal systems largely unexamined (Gomez *et al*., 2024). This also results in a lack of diversity, with many datasets being Anglo-centric and under-representing minorities, raising concerns about the inclusion of multiple perspectives on complex ethical topics (Rauh *et al*., 2024).

Furthermore, AI benchmarking practices are deeply political, performative, and generative, actively shaping how AI models are trained and applied rather than passively measuring their capabilities (Orr and Kang, 2024). They operate as "normative instruments that perpetuate particular epistemological perspectives", often prioritising efficiency over care, universality over contextuality, and impartiality over positionality (Scheuerman *et al*., 2021). The economic, competitive, and commercial roots of capability-oriented benchmarks embed them within corporate marketing strategies, fuelling AI hype and attracting investors (Eriksson *et al*., 2025). This creates an incentive mismatch, where the pursuit of state-of-the-art (SOTA) performance often overshadows high-quality evaluations, leading to "SOTA-chasing" and a "winner's curse" at the expense of genuine insight and explanation (Church and Hestness, 2019). This competitive culture, often likened to a "sport," reinforces path dependencies in AI research, favouring certain methodologies and stifling others that do not align with dominant benchmark logic (Eriksson *et al*., 2025). The gaming of benchmarks is another critical concern, with evidence of data contamination, "sandbagging" (models strategically underperforming), and cherry-picking results due to a lack of transparency and reproducibility resources (Eriksson *et al*., 2025). The dubious community vetting process, where benchmarks gain influence through citation popularity rather than inherent suitability, further compounds these issues. This leads to “peer-washing,” where problematic datasets maintain authority despite their shortcomings (Schlangen, 2020). The rapid pace of AI development also contributes, as benchmarks quickly become saturated and outdated, unable to effectively evaluate increasingly complex models or keep pace with continuous model iterations and new capabilities (McIntosh *et al*., 2025). Finally, the inherent complexity of AI and the presence of unknown unknowns limit current benchmarking capabilities. Human knowledge constraints prevent a full assessment of emerging AI capabilities that may surpass conventional understanding, potentially leading to latent vulnerabilities and unforeseen risks (Eriksson *et al*., 2025). Efforts to fine-tune AI models for safety, for instance, have been shown to degrade performance in other areas or introduce entirely new security risks. These fundamental fragilities in quantitative AI evaluation highlight that current benchmarking is ill-suited to single-handedly provide the safety and capability assurances demanded by policymakers (van der Weij *et al*., 2024).

This paper posits that these acknowledged shortcomings in AI evaluation, from construct validity failures to competitive dynamics and unknown vulnerabilities, are not merely technical glitches but symptoms of a deeper neglect: a failure to account for AI’s role in a cognitive ecology. We define AI as a cognitive ecology as an interconnected system where human and artificial intelligences interact, adapt, and co-evolve, influencing each other’s cognitive processes, cultural norms, and knowledge structures in profound, often invisible ways. The current evaluation paradigm, fixated on isolated performance metrics and tangible outputs, fails to capture how GenAI models subtly reshape human cognition (e.g., through reliance on AI for problem-solving), infuse cultural biases (e.g., through reinforcement of implicit assumptions in datasets), and redefine epistemic authority (e.g., by influencing what constitutes “truth” or “expertise” in a data-driven world). Overlooking these invisible costs of the transformations in human thinking, cultural values, and the very nature of knowledge production and validation poses a significant risk to the responsible development and integration of AI into society. While some initial qualitative insights touch upon ethical and legal considerations, including bias and explainability concerns from developers themselves (Russo, 2024), these remain peripheral to the primary focus on adoption drivers such as compatibility, etc. This paper aims to foreground these subtle yet pervasive impacts.

Therefore, this paper seeks to reveal these invisible cognitive, cultural, and epistemic costs of generative models by proposing a comprehensive framework for their systematic analysis. By viewing AI as a cognitive ecology, we shift the focus from merely assessing model performance to understanding the dynamic interplay between AI systems and the broader human and informational environment. This approach allows us to unpack the subtle ways GenAI alters human decision-making, shapes societal norms, and reconfigures knowledge landscapes. Our contribution will be structured around a novel HORIZON taxonomy, which systematically categorises these heretofore unacknowledged costs, providing a critical lens for future research, design, and policy. We argue that a holistic understanding of GenAI’s true impact necessitates moving beyond quantifiable outputs to embrace an ecological perspective that accounts for its transformative, often hidden, influence on human thought, culture, and knowledge.

The remainder of this paper is structured as follows: Section 2 provides a comprehensive literature review outlining existing approaches to AI evaluation and their limitations, further substantiating the identified gap. Section 3 introduces our theoretical framework, “AI as Cognitive Ecology,” detailing its core tenets. Section 4 presents the HORIZON taxonomy, operationalising the cognitive, cultural, and epistemic dimensions of AI’s invisible costs. Section 5 discusses the implications of this ecological perspective for ethical AI development, responsible innovation, and policy formulation. Finally, Section 6 offers concluding remarks and outlines future research directions.

2. AI as Cognitive Ecology: A Paradigm Shift

2.1 From Tools to Ecologies

The rapid integration of artificial intelligence (AI) into diverse societal domains has prompted a reliance on metaphors to conceptualise its function and impact. Predominantly, AI is framed as a tool, agent, or assistant (Chan *et al*., 2025; Kandasamy, 2025). This perspective views AI as an artefact to be wielded by human users, designed to achieve specific goals or automate predefined tasks (Stryker, 2024). For instance, within environmental computer science, AI is widely perceived as a set of methods, a helpful tool such as GIS, Statistics or Data Visualisation (Sinwell *et al*., 2021). Similarly, the concept of Agentic AI emphasises LLM-driven entities that interact with tools, environments, and other agents to accomplish tasks with a degree of autonomy (Kandasamy, 2025). Even the fundamental definition of software agents posits them as autonomous, goal-directed computational entities capable of perceiving and acting upon their environment. This tool-use pattern is a recognised design approach in Agentic AI systems, defining a tool as a piece of code that the Agent uses to observe or act towards achieving its goal. The Agents and Artefacts (A&A) approach further models’ artefacts as first-class abstractions for modelling and designing MAS working environments, drawing inspiration from Activity Theory (Omicini *et al*., 2009). This common understanding suggests AI, whether as a simple utility or a more sophisticated agent, primarily serves as a means to an end, an instrument for human or system objectives (Sinwell *et al*., 2021).

Despite their intuitive appeal and widespread adoption, these metaphors fall short of capturing the transformative, systemic impact of advanced AI models on human cognition and societal structures. The tool metaphor, in particular, implies a passive instrument entirely subject to external control, obscuring AI’s increasingly active and even generative role (Markolf *et al*., 2021). While artefacts, as discussed in the A&A meta-model, are indeed explicitly designed to provide a certain function (Omicini *et al*., 2009), they also possess both an enabling and a constraining function and shape the way human beings interact with reality (Omicini *et al*., 2009). This suggests a deeper influence than mere utility. Furthermore, these mediating tools embody social practices, reflect historical experiences, and influence not only the external behaviour, but also the mental functioning of individuals using them (Omicini *et al*., 2009). When AI is seen merely as a tool in fields such as Environmental Computer Science, it can lead researchers to focus narrowly on getting a good prediction from a very specific dataset, rather than developing general applicable models that align with the broader aims of pure AI research (Sinwell *et al*., 2021). This implies that the metaphor itself can inadvertently limit the scope and ambition of AI’s application.

The metaphors of learning and training for AI, while pervasive, similarly risk misleading practitioners and the public by implying human-like cognitive processes that are not necessarily present (Murray-Rust *et al*., 2022). As Murray-Rust *et al*. (2022) argue, these impressive surface-level performances do not necessarily correspond with other abilities that humans have, such as generalisation and reasoning (Murray-Rust *et al*., 2022). The black box metaphor, often used to describe the opacity of AI systems, further compounds this issue by fostering a sense of powerlessness and diverting attention from understanding the underlying mechanisms or accountability. Similarly, the term bias in AI, while seemingly analogous to human prejudice, can conceal its multiple technical meanings and the pernicious idea that talking about bias raises the possibility of an unbiased model, thereby sidestepping the crucial understanding that biases are always relative to something, and that something needs articulation (Hildebrandt, 2021). Fundamentally, these established metaphors, while simplifying and connecting to existing knowledge, illuminate and hide, centring particular ideas, marginalising others, and shaping fields of practice (Murray-Rust *et al*., 2022). The very idea that AI systems may lead to "implicitly or explicitly releasing control to algorithms (Markolf *et al*., 2021) points to a profound shift in agency that a simple tool cannot encapsulate. Even the development of control planes to manage tool orchestration in Agentic AI systems (Kandasamy, 2025) highlights the growing complexity inherent in integrating and managing these so-called "tools."

While the agent or assistant metaphors offer a more dynamic view of AI, emphasising their autonomy, goal-seeking behaviour, and interaction capabilities (Chan *et al*., 2025; Kandasamy, 2025), they still tend to centre the AI as a discrete entity rather than recognising its pervasive influence on the underlying cognitive and social environment. An AI agent, as defined by Chan *et al*. (2025), directly interacts with the world and adapts to underspecified tasks, going beyond traditional software. However, even these advanced agents operate within a complex agent infrastructure, technical systems and shared protocols external to agents that are designed to mediate and influence their interactions with and impacts on their environments" (Chan *et al*., 2025). This external infrastructure, encompassing elements like identity binding, certification, agent IDs, channels, oversight layers, and communication protocols, implies that AI's influence extends far beyond its individual actions. Moreover, the emergence of automated thinking, as conceptualised by Sellar and Gulson (2021), challenges the notion of AI as merely an agent executing pre-programmed rules. This nonconscious cognition operates across and within the full spectrum of cognitive agents: humans, animals, and technical devices (Sellar and Gulson, 2021), and can even make inferences about the most optimal decision in a given situation while remaining ignorant of a larger set of indeterminate possibilities". This hints at a more active, and at times unpredictable, reshaping of cognitive processes than the agent metaphor implies. The recognition that AI's capabilities may augment and replace people (Markolf *et al*., 2021) and contribute to implicitly or explicitly releasing control to algorithms further underscores its transformative potential, extending beyond the boundaries of an individual agent.

To adequately grapple with these profound shifts, we advocate for a paradigm shift in how AI is conceptualised: as an integral part of a cognitive ecology or cognitive infrastructure that fundamentally reshapes the conditions of thought, decision-making, and societal organisation. This ecological metaphor posits AI not as a mere instrument, but as a pervasive force that alters the environment in which human cognition operates, much like the advent of language, literacy, or new media systems (Omicini *et al*., 2009; Sellar and Gulson, 2021). As Chester and Allenby argue, the rise of novel digital and connected technologies signifies "not simply the rise of cyber-physical systems as hybrid physical and digital assets but, ultimately, the integration of legacy systems into a new cognitive ecosystem (Chester and Allenby, 2023). This cognitive ecosystem is characterised as an ecology of massive data flows, artificial intelligence, institutional and intellectual structures, and connected technologies, poised to alter how humans and artificial intelligence understand and control our world. It is an emerging and highly complex feature of an increasingly anthropogenic planet that integrates functionalities from diverse sources, including increasingly powerful AI tools such as generative AI, to the rules, regulations, venture capitalists, and social media systems that co-evolve with the technologies. This view emphasises that AI is not just affecting what we do, but how we think and perceive the world, and indeed, "what the systems fundamentally are".

This reframing positions AI as a constituent element of a planetary-scale computation (Chester and Allenby, 2023), where algorithms become embedded in a multi-layered "digital infrastructure space that acts as a medium of information and an operating system for shaping the city. This infrastructure space has a disposition that emerges, in part, from the actions of algorithms, leading to distributed, emergent forms of cognition that can have powerful effects. Sellar and Gulson (2021), drawing on Parisi's work, define this as "automated thinking," a form of nonconscious cognition that syncopates with human thinking and inhabits different temporalities, opening up temporal regimes in which the costs of consciousness become more apparent and more systemically exploitable. This emergent automation introduces creative uncertainty, where algorithms can "reason through and with uncertainty and engender their own form of knowing. The implication is that AI, as cognitive infrastructure, is not just performing tasks but actively contributing to the generation of new knowledge, values, and decision-making processes, thereby changing the possibilities for education policy and the governance of school systems. This perspective acknowledges that control efforts may need to focus on establishing relationships with AI that recognise that cyber-technologies will be guiding us in ways that we may not always fully understand (Markolf *et al*., 2021). Thus, the ecological metaphor prompts a shift from assessing individual AI performance to understanding the broader transformation of our "cognitive ecosystem"(Chester and Allenby, 2023).

2.2 Historical Cognitive Infrastructures

Human cognition has always been scaffolded by infrastructures that extend and reshape mental capacities (Loh and Kanai, 2016). Language was the first and most foundational, enabling the offloading of thought onto others' minds and distributing cognitive load through collaboration (Dror and Harnad, 2008; Baker, 2009). This biological development not only expanded communication but also rewired neural systems, making complex everyday thought possible (Pandey *et al*., 2023).

Writing and later the printing press externalized memory, overcoming the fragility of oral transmission and facilitating reflection, abstraction, and collective knowledge growth (Dror and Harnad, 2008). These innovations generated a period of "cognitive inflation," while the printing press standardized and democratized access to knowledge, embedding infrastructures into the cultural and epistemic fabric of societies.

The Internet represents a more radical shift: a global "Cognitive Commons" where cognizes, databases, and software agents interoperate at speeds and scales inconceivable for individual minds (Loh and Kanai, 2016). Unlike earlier infrastructures that stored or transmitted knowledge, it actively generates information, creating a qualitatively new dimension of cognitive offloading.

This trajectory has been theorized through the Extended Mind Thesis (EMT), which argues that cognition can extend beyond the "skin and skull" into the environment (Pandey *et al*., 2023). A notebook serving as external memory exemplifies such coupling (Baker, 2009). Yet debate persists: while external tools enhance cognition, they are not themselves cognizers. The prevailing view holds cognition as "narrow" - biologically instantiated - while acknowledging that tools can reconfigure cognitive performance and mental states (Dror and Harnad, 2008).

The neuroscientific implications are evident in the rise of Digital Natives. Growing up in hyperlinked, interactive environments has fostered shallow processing styles, rapid attentional shifts, and diminished deep reading, linked to changes in brain circuitry for executive control and sustained focus (Loh and Kanai, 2016). While digital offloading can strategically free resources, it also risks attenuating contemplative skills, making technological deprivation feel akin to cognitive impairment.

Thus, the historical arc from language to the Internet demonstrates more than quantitative enhancement of performance. It reveals qualitative transformations in cognition itself, reshaping how humans process information, sustain attention, and even understand their own minds.

2.3 Why Ecology?

Why adopt an ecological perspective on cognition and artificial intelligence (AI)? Because cognition does not reside in isolated brains but emerges from interdependent systems that link humans, artefacts, environments, and cultural practices. An ecological lens foregrounds interdependence, diversity, fragility, and resilience qualities essential for understanding how intelligence, human or artificial, is sustained. By situating AI within a cognitive ecology rather than treating it as an autonomous tool or agent, we can better diagnose risks that arise not from discrete errors but from systemic imbalances such as homogenization, skill erosion, or epistemic collapse.

(Hutchins, 2010) describes cognitive ecology as the study of cognitive phenomena in context, where meaning-making processes unfold through webs of mutual dependence among people, environments, and material systems. This position challenges the cognition as internal operation model of classical cognitivism, which confined mind to the boundaries of skull and skin. Instead, cognitive processes extend into sensorimotor engagements, social interactions, and cultural artefacts. Bateson's (1972) famous parable of the blind man with a stick illustrates this point: cognition cannot be bounded at the skin, since the stick, the ground, and the sensory loops that run through them are integral to perception and action. To excise any part of this ecology would be to render the system inexplicable. Similarly, when AI systems become part of human practices, they enter into feedback loops that are no less constitutive of cognition than sticks, maps, or scripts.

(Tribble and Sutton, 2011) expand this framework by highlighting how cognition spreads or smears across heterogeneous resources: brains and bodies, tools and texts, institutions and environments. No one dimension holds analytic primacy, since cognition is always hybrid and distributed. A Shakespearean performance, for example, depended on the interplay of actors' bodies, playbooks, audience conventions, acoustic environments, and economic structures, none of which could be abstracted away without distorting the whole. By analogy, contemporary AI must be situated within ensembles that include human expertise, cultural norms, infrastructures, and ecological limits. Thought and action are always "system-level activities" (Tribble and Sutton, 2011).

This ecological perspective makes visible risks that are otherwise obscured. If cognition is distributed, systemic imbalances rather than isolated errors become the primary source of fragility. Homogenization of knowledge through algorithmic filtering, erosion of embodied skills displaced by automation, or the collapse of epistemic diversity in networked cultures are not technical malfunctions but ecological pathologies. Hutchins (2010) stresses that cognitive ecosystems gain their resilience from diversity and redundancy: multiple pathways of representation, interaction, and interpretation buffer the system against failure. By contrast, overreliance on uniform AI models risks brittleness, where local perturbations propagate into systemic crises.

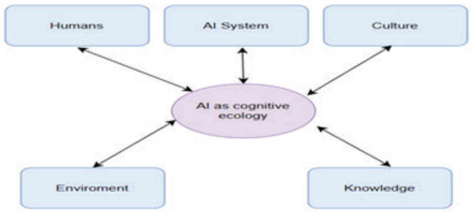

Figure 1: AI as Cognitive Ecology (system diagram / hub-and-spokes)

As illustrated in Figure 1, AI must be understood as part of a cognitive ecology in which humans, culture, knowledge, and environment are mutually conditioned by and through artificial systems. The figure visually underscores the paradigm shift: AI is not a discrete agent but a node in an ecological web of cognition. Each arrow represents a bidirectional relationship AI shapes human practices, cultural forms, epistemic resources, and environmental impacts, while simultaneously being shaped by them. This relational framing highlights not only the interdependence but also the fragility of the system. Just as ecosystems can collapse when keystone species are disrupted, cognitive ecologies can falter when systemic diversity or balance is lost.

Crucially, this model also foregrounds resilience. Cognitive ecologies adapt as elements shift: when one dimension is stressed, others may compensate. Tribble and Sutton (2011) describe how changes in theatrical technologies from gestural performance to lighting design reshaped the distribution of attention and skill across actors, technicians, and audiences. Analogously, as AI transforms knowledge production and labour, resilience will depend on how humans, institutions, and environments redistribute capacities. Recognizing AI as ecology thus invites us to cultivate redundancies, preserve epistemic pluralism, and design for adaptive diversity rather than efficiency alone.

By reframing AI as ecology, we align with a broader trajectory in cognitive science that has moved from reductionism toward holism. Hutchins (2010) predicts that the reality of the rich interconnectivity of the brain, body, and world" will draw together disparate strands of embodied, enactive, and distributed cognition into a coherent ecological synthesis. For cultural historians, Tribble and Sutton (2011) stress that cognition must be analysed historically and materially, since artefacts and practices are not external to thought but constitutive of it. Extending these insights to AI, we see that artificial systems are neither external tools nor independent agents: they are deeply embedded in, and transformative of, our cognitive ecologies.

Treating AI as part of a cognitive ecology illuminates the systemic conditions under which intelligence, human and artificial, emerges and persists. It makes clear that risks arise not only from flawed algorithms but from ecological imbalances, and that resilience depends on maintaining interdependence, diversity, and adaptability across humans, cultures, environments, and knowledge systems.

3. HORIZON Taxonomy of Invisible Costs

The HORIZON taxonomy (Figure 2) visualises the seven dimensions of invisible costs. Like stressors in an ecological system, each dimension radiates from the same central phenomenon: the embedding of AI within human cognitive environments.

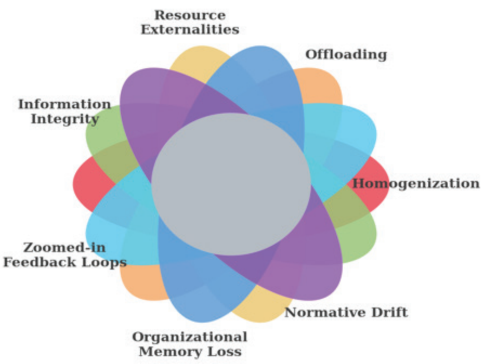

Figure 2: HORIZON Taxonomy of Invisible Costs (Petal Ecology Flower Layout)

3.1 Homogenization

A central invisible cost of generative systems is the homogenization of cultural and linguistic expression. By design, large language models and text-to-image systems are probabilistic engines: they optimize for the most statistically likely continuation of a prompt, privileging median rather than marginal outputs (Sourati *et al*., 2025). This leads to stylistic flattening, where outputs are grammatically fluent but stylistically neutral and culturally mainstream. As these systems are integrated across creative, academic, and professional domains, the very statistical strength that powers them risks producing a narrowing of cultural possibility.

The homogenizing dynamic is visible at multiple levels. At the level of language, empirical studies have shown that LLM-mediated writing reduces lexical and stylistic variety, erasing markers of individuality and cultural nuance (Sourati *et al*., 2025; Zhang *et al*., 2025). Survey-based evidence demonstrates that when research participants rely on AI to compose open-ended responses, outputs cluster around homogenised, positive, and generic formulations, masking underlying diversity in beliefs and attitudes (Zhang *et al*., 2025). Cross-cultural experiments likewise reveal how AI suggestions pull writers from non-Western contexts toward Western stylistic norms, diminishing culturally specific expression in favour of globally legible but culturally impoverished forms (Agarwal *et al*., 2025).

Visual culture exhibits similar dynamics. Analyses of text-to-image systems show that reliance on standardised prompt engineering practices, coupled with model training on predominantly Western-centric datasets, generates a convergence toward familiar aesthetic templates (Palmini and Cetinic, 2024). Even when user input introduces originality, the reinforcement of shared prompt structures and the popularity-driven curation of outputs contribute to visual uniformity. Large-scale studies of online art-sharing platforms confirm that the introduction of AI assistance reduces visual novelty across portfolios, as adoption spreads and community norms recalibrate around AI-influenced aesthetics (Zhou and Lee, 2024).

Cultural and social implications follow. As generative systems privilege dominant linguistic and visual repertoires, they risk marginalising minority voices and alternative epistemologies. The flattening of linguistic markers not only undermines cultural preservation but also disrupts fields that rely on the richness of stylistic variation, such as psychological diagnostics, personnel evaluation, and sociolinguistic research (Sourati *et al*., 2025). In media and communication contexts, generative tools often replicate normative identities and suppress non-normative narratives, reproducing what (Gillespie, 2024) terms the politics of visibility, where representational harms range from stereotyping to symbolic erasure. Similar processes are evident in urban-cultural domains: generative AI tools, when asked to depict local contexts, tend to foreground commercialised and tourist-oriented elements, narrowing the perceived scope of cultural life and exacerbating existing power imbalances (Campo-Ruiz, 2025).

Taken together, these findings suggest that homogenization is not merely a by-product of generative probability distributions but a systemic cultural cost with implications for diversity, equity, and knowledge production. The convergence toward the median amplifies dominant cultural logics while silencing peripheral ones, producing an aesthetic monoculture that risks eroding the pluralism essential to cultural vitality (Karpouzis, 2024; Singh, 2024). This homogenization is subtle yet pervasive: it is experienced not as overt censorship but as the quiet disappearance of difference, drowned beneath the polished fluency of statistical averages.

3.2 Offloading (Deskilling)

One of the most insidious invisible costs of AI integration is the deskilling that results from the offloading of cognitive labour. As individuals and organisations increasingly delegate tasks such as drafting, analysis, recall, and problem-solving to AI systems, core human competencies risk atrophy. Scholars have long recognised that automation reshapes expertise by transforming workers from active decision-makers into passive overseers of “black box” processes (Rinta-Kahila *et al*., 2018). While efficiency gains are undeniable, the erosion of tacit knowledge and procedural know-how leaves workers and by extension, societies vulnerable when systems fail.

The phenomenon extends beyond technical work into broader domains of cognition and professional judgment. (Matueny, 2025) argue that dependence on AI fosters an illusion of competence, wherein individuals mistake AI-generated fluency for personal mastery. This misperception discourages active engagement and deep learning, weakening memory, critical thinking, and metacognitive regulation. Over time, as with the Google Effect in memory research, what is routinely offloaded to machines is less likely to be internally retained. The danger is that individuals come to rely on external systems to such an extent that resilience in non-digital contexts diminishes.

Empirical studies underscore this fragility. When organisations discontinue automated systems, latent deskilling becomes painfully visible. In a case study of accountants, the removal of an automated fixed-asset management system forced employees to relearn procedures they had long neglected, exposing significant gaps in both declarative and procedural knowledge (Rinta-Kahila *et al*., 2018). Such disruptions illustrate that automation not only reduces the demand for skills in the short term but may also impair the capacity to recover them in the long term. Similarly, Bushuyev *et al*. (2024) highlight the erosion of managerial competencies in innovation projects, as reliance on AI-generated insights undermines experiential decision-making, tacit coordination, and creative risk assessment—skills once central to project leadership.

Yet, deskilling is not uniformly negative. Ong and Png (2021) provide evidence that automation-induced simplification of cognitively demanding tasks, such as cashiering or way-finding for drivers, can enhance job satisfaction and expand labour supply. By lowering entry barriers, technological deskilling increases workforce participation, particularly in low-skill sectors. However, this amenity-driven benefit trades off against the longer-term resilience of cognitive skills, echoing the broader tension between efficiency and robustness in work design.

The future of labour markets may therefore hinge on how societies navigate this dialectic. Zhang *et al*. (2024) suggest that while sensory-physical tasks are highly susceptible to automation, social-cognitive and higher-order reasoning skills retain comparative resilience. However, sustaining this resilience requires deliberate investment in cognitive and metacognitive skills, ensuring that workers cultivate adaptive expertise rather than ceding intellectual agency to AI. Deskilling is thus not a deterministic outcome of automation but a contingent one, shaped by how technologies are integrated into human systems and whether offloading is balanced with opportunities for skill development.

3.3 Resource Externalities

The material costs of artificial intelligence (AI) systems extend beyond carbon emissions, manifesting in substantial yet often invisible demands on electricity and freshwater resources. While these externalities are rarely factored into assessments of AI's sustainability, their ecological implications are profound.

First, the energy intensity of AI training and inference has escalated sharply with the proliferation of large-scale models. A single rack of AI hardware, such as NVIDIA H100 GPU clusters, can consume nearly 39 times the electricity of an average U.S. household, with hyper-scale data centres approaching the annual electricity demand of entire metropolitan areas (Sunkara, 2025). This surge in demand is not evenly distributed, as regional data centre expansions have destabilising effects on national energy infrastructures; for instance, Ireland projects that data centres may soon account for nearly one-third of its total electricity use (Inie *et al*., 2025a). While AI-driven optimisation can improve energy efficiency in sectors such as manufacturing and smart grids (Nurhaeni *et al*., 2024; Zakizadeh and Zand, 2024), the rebound effect suggests that these savings are outpaced by the exponential growth of computational demand, raising questions about whether AI constitutes a net energy-saving technology.

Equally significant, though less visible, are AI's water footprints. Cooling systems for AI-intensive data centres are overwhelmingly water-dependent, with evaporative cooling converting freshwater into vapour that is permanently lost from local watersheds (Natarajan, 2025). Training and inference runs for advanced large language models can therefore consume millions of litres of freshwater, often in drought-prone regions such as Arizona and Northern Virginia, where competition with residential and agricultural users sharpens issues of environmental justice (Natarajan, 2025). Empirical estimates highlight this magnitude: inference with GPT-4 for a 10-page report can consume over 50 litres of water, compared to less than one litre for smaller-scale models (Shumba *et al*., 2025). Such disparities underscore how infrastructural decisions amplify regional vulnerabilities, producing what has been termed the hydro-digital paradox, technological progress intensifying local water scarcity (Natarajan, 2025).

Attempts at mitigation have focused on embedding AI into sustainable data centre design, including water-efficient cooling, hardware optimisation, and recycling systems (Hiremath, 2024). Yet even these innovations risk redistributing rather than resolving burdens. Life cycle assessments (LCAs) of generative AI services indicate that focusing narrowly on carbon overlooks intertwined costs such as water depletion, metal scarcity, and e-waste (Berthelot *et al*., 2024). In this sense, AI exemplifies the broader challenge of "carbon tunnel vision" in sustainability discourse: privileging emissions metrics at the expense of recognising the full spectrum of material dependencies (Berthelot *et al*., 2024).

The resource externalities of AI reveal a contradiction at the heart of digital modernity. While AI is celebrated as a driver of sustainability and efficiency, its hidden appetites for electricity and freshwater expose new vectors of ecological strain. These costs are not marginal but systemic, disproportionately affecting regions already vulnerable to energy and water scarcity. Future governance of AI infrastructure must therefore reckon with these invisible costs, shifting sustainability frameworks from carbon-centric metrics toward integrated assessments that account for multi-resource entanglements.

3.4 Information Integrity

A central but often underappreciated invisible cost of generative models lies in their destabilisation of epistemic reliability. While such models excel at producing fluent and persuasive language, their confidence calibration is systematically misaligned with truth value, generating conjectures with the same assertive tone as verified facts (Krishnan *et al*., 2024; Tao *et al*., 2025). This epistemic opacity erodes not only the reliability of individual outputs but also the category of knowledge itself, as users become less able to distinguish justified belief from manufactured plausibility.

Philosophically, these challenges conventional accounts of epistemic authority. As (Ferrario *et al*., 2024) argue, AI systems cannot be granted genuine epistemic expertise because they lack the understanding and intellectual virtues necessary for such a status. Their outputs, however accurate in narrow tasks, remain severed from justificatory structures. Evans *et al*. (2021) underscore this risk in their call for truthful AI, noting that scalable, personalized untruths may undermine not only individual decisions but also collective epistemic and democratic deliberation.

Technical responses have sought to reintegrate uncertainty as an explicit epistemic signal. Approaches such as black-box uncertainty quantification for LLM-as-a-judge (Wagner *et al*., 2024), uncertainty-aware fine-tuning (Krishnan *et al*., 2024), and atypical-presentation recalibration in healthcare (Qin *et al*., 2024) demonstrate that calibrated confidence can enhance trustworthiness without sacrificing performance. Yet, large-scale benchmarking reveals that accuracy and uncertainty are often decoupled: high-performing models can remain overconfident and poorly calibrated, particularly on knowledge-heavy tasks (Tao *et al*., 2025). Proposals for structured epistemic architectures (Wright,

2025) suggest a more radical path, embedding propositional commitment, contradiction detection, and normative truth maintenance into reasoning systems to prevent epistemic drift.

The invisible cost, therefore, is not reducible to factual error or isolated hallucination (Ji *et al*., 2024; Lu, 2025). It is the cumulative erosion of epistemic integrity where the persuasive fluency of generative systems destabilises the social trust infrastructure that underpins knowledge practices. Absent robust mechanisms for epistemic calibration, users risk conflating probabilistic text generation with warranted assertion, thereby transforming knowledge ecosystems into arenas of ambient uncertainty. This cost manifests less in discrete failures than in the long-term corrosion of epistemic norms. Safeguarding information integrity thus requires both technical calibration mechanisms and normative frameworks that re-anchor generative systems within truth-conducive practices.

3.5 Zoomed-in Feedback Loops

As generative AI systems increasingly contribute to the pool of online data, recursive loops emerge wherein models are trained on their own outputs. This recursive dynamic produces a class of invisible costs that extend beyond technical degradation to deeper epistemic narrowing of knowledge systems. Recent theoretical and empirical work converges on the phenomenon of model collapse: the progressive deterioration of generative performance as synthetic data dominates training corpora (Borji, 2024; Seddik *et al*., 2024). Collapse manifests statistically when recursive training erodes the tails of the original distribution, reducing diversity and yielding homogenised, repetitive, or even degenerate outputs (Seddik *et al*., 2024). The effect is not confined to text but generalises across modalities, as recursive inpainting experiments show successive degradation of images until they drift toward meaningless artefacts (Conde *et al*., 2025).

The recursive feedback mechanism operates as both a technical and epistemic loop. Technically, each generation of models amplifies the approximation errors of its predecessors, accelerating drift away from the underlying real-world distribution (Borji, 2024). Epistemically, the iterative reliance on self-produced data narrows the representational horizon: what models "know" is increasingly filtered through their own outputs, risking an autophagic cycle where the ecosystem feeds on itself (Shumailov *et al*., 2024), as discussed in (Borji, 2024). This dynamic threatens not only accuracy but the breadth of knowledge itself, substituting richness of human-authored data with recursive self-reference.

Empirical investigations suggest two partial mitigations. First, mixing real and synthetic data can attenuate collapse, though only when the ratio of authentic data remains sufficiently high (Seddik *et al*., 2024). Second, accumulation rather than replacement of training data, where each generation augments rather than overwrites prior corpora bounds error growth and avoids total collapse (Gerstgrasser *et al*., 2024). Yet these mitigations underscore the structural fragility of recursive feedback loops: they do not eliminate the epistemic narrowing but only slow its progression.

Viewed through the HORIZON taxonomy, these loops exemplify invisible costs: the degradation is subtle, distributed, and often invisible in the short term, but accumulates over cycles to reshape entire knowledge systems. Unlike immediate technical failures, feedback loops risk a gradual impoverishment of the epistemic commons. In effect, they collapse diversity into predictability, precision into noise, and world-models into self-referential artifacts an outcome as socially consequential as it is technically avoidable.

3.6 Organisational Memory Loss

As organisations increasingly embed critical processes into proprietary AI models and automated systems, a subtle but profound erosion of organisational memory emerges. Historically, institutional knowledge has been sustained through collective practices, documents, mentorship, and shared routines that both preserved tacit expertise and enabled its intergenerational transfer (Falckenthal *et al*., 2025). The contemporary shift toward codifying workflows in prompt templates and AI-generated outputs risks displacing these social mechanisms of knowledge retention. While AI-enhanced knowledge management systems promise efficiency gains through semantic indexing, dynamic retrieval, and automated synthesis (Jarrahi *et al*., 2023; Gadde, 2025), their reliance on vendor-controlled infrastructures centralises knowledge in external architectures. This creates a form of epistemic dependency, where the durability of organizational intelligence becomes contingent upon proprietary platforms rather than distributed human memory.

The implications of this shift are twofold. First, the automation of tacit knowledge through machine learning and conversational capture bots provides an expedient but fragile archive. Systems such as those described by Satsangi (2019) demonstrate how AI can collect employees' day-to-day experiences and convert them into structured repositories. Yet, while such tools preserve fragments of experiential data, they decouple knowledge from its embodied context, stripping away the relational and situational nuance that traditionally sustains expertise (Collins, 2010; cited in Falckenthal *et al*., 2025). Without mechanisms of embodied apprenticeship or interactive sense-making, what persists is an attenuated representation of practice rather than the adaptive, resilient memory required for organisational continuity (Nonaka and Takeuchi, 2021, cited in Falckenthal *et al*., 2025).

Second, organisational dependence on AI intermediaries reshapes the ecology of knowledge transfer. Multi-agent system research shows that distributed knowledge exchange thrives when responsibilities are shared through organisational protocols and negotiated roles (Farias *et al*., 2024). By contrast, outsourcing memory to algorithmic infrastructures reduces opportunities for co-constructed meaning and weakens the social level of organisational learning. The result is not only an erosion of collective memory but also a narrowing of adaptive capacity in the face of disruptions.

This trajectory aligns with concerns in the knowledge management literature that AI systems, while augmenting knowledge creation and retrieval, simultaneously fragment institutional continuity by privileging efficiency over social embedding (Jarrahi *et al*., 2023). The more organisations normalise the substitution of mentoring, storytelling, and shadowing with AI-mediated archives, the more fragile their epistemic resilience becomes. Thus, organisational memory loss is not a passive by-product of technological change but an invisible cost where the very infrastructures designed to preserve knowledge paradoxically accelerate its decontextualization and externalisation.

3.7 Normative Drift

Among the less visible but most consequential dimensions of the HORIZON taxonomy of invisible costs is normative drift, the gradual, often unexamined process by which AI systems default guardrails, refusals, and stylistic conventions become taken-for-granted social norms. Unlike explicit regulation, where laws or policies are openly debated and codified, normative drift occurs through cumulative micro-interactions with AI systems, where corporate or technical defaults are silently naturalised as appropriate ways of speaking, refusing, or reasoning.

Guardrails such as refusal styles or politeness defaults are never normatively neutral. As Šekrst *et al*. (2024) show in their discussion of customizable guardrails, even technical interventions designed for harm reduction encode normative choices about civility, safety, and appropriateness. For example, when a language model consistently responds to risky or controversial queries with deferential refusals, it implicitly sets expectations about what kinds of discourse are considered beyond the pale, not just for machines, but for humans engaging with them. Over time, these outputs can act as norm entrepreneurs, subtly steering cultural expectations of politeness, risk tolerance, or moral acceptability.

The concern is not merely theoretical. Scholars of design and technology have long argued that artefacts embody values and political commitments (Vermaas and Stone, 2020). Yet what distinguishes AI is the opacity of its normative loadings and the velocity of its diffusion. Unlike infrastructure norms, which evolve slowly across decades, AI defaults can globalise within months, reaching billions of users before any meaningful public deliberation (Luccioni and Bengio, 2019). This creates a profound mismatch between the speed of technological diffusion and the slower timescales of democratic norm formation (Baronchelli, 2024).

This acceleration magnifies the risks of homogenization. Lim *et al*. (2023); Seo and Kwon (2024) emphasise that social norms surrounding AI are shaped not only by regulators and ethicists but also by the daily practices of developers, corporations, and end-users. When billions of interactions reinforce uniform refusal phrasings or "politeness defaults," the result is a powerful feedback loop that narrows cultural variation and epistemic diversity. Such bottom-up norm formation is particularly concerning because it often occurs without transparency about whose values are embedded or how alternatives might be considered.

Moreover, cross-cultural tensions sharpen the stakes of normative drift. As Younas (2023) argues, many AI ethics frameworks reflect Western liberal-democratic traditions, privileging certain norms of individual autonomy or secular risk assessment. When these defaults are exported globally, they risk marginalising alternative cultural traditions of moral reasoning—for example, relational ethics in Confucian contexts or Ubuntu ethics in African traditions. If left unexamined, normative drift may thus not only flatten communicative styles but also entrench a form of cultural imperialism under the guise of "safety."

The governance literature underscores that algorithms already act as regulators, shaping visibility, credibility, and access to information (Saurwein *et al*., 2015; Lucero, 2020). Yet current governance debates focus predominantly on transparency, bias, and accountability, with far less attention to the subtle normative imprint of guardrails. To resist unexamined drift, governance must expand to include explicit acknowledgement of value-laden defaults, participatory processes for shaping refusal styles, and pluralistic infrastructures that allow users to select among different normative frameworks rather than being locked into a single corporate template.

Taken together, normative drift exemplifies the broader dynamics captured in the HORIZON taxonomy of invisible costs (see Table 1). Like homogenization, it threatens cultural diversity; like information integrity failures, it risks epistemic trust. But its distinctive danger lies in its silence: norms become standards without ever being publicly chosen. Preventing normative drift, therefore, requires mechanisms of transparency, participatory deliberation, and cultural co-genesis, ensuring that the invisible costs of AI do not calcify into invisible norms.

Table 1. HORIZON Taxonomy of Invisible Costs in Generative AI

Dimension	Definition	Example	Risk	Possible Mitigation
H – Homogenization	Convergence of outputs toward median styles or perspectives.	AI-assisted essays sound stylistically similar.	Loss of cultural diversity; flattening of originality.	Enforce output diversity budgets; promote pluralistic sampling.
O – Offloading (Deskilling)	Reliance on AI reduces human practice of cognitive skills.	Students rely on AI to draft, weakening their argumentation ability.	Erosion of baseline skills; vulnerability during AI failures.	"AI-off drills" in education and critical professions.
R – Resource Externalities	Hidden environmental costs beyond carbon.	Water use for data centre cooling; power-grid stress.	Environmental strain, especially in water-scarce regions.	Standardised per-query disclosures (RTE labels).
I – Information Integrity	Models output fluent but misleading content without uncertainty.	LLM fabricates citations confidently.	Epistemic collapse; erosion of trust in knowledge systems.	Calibrated uncertainty by default; citation verification tools.
Z – Zoomed-in Feedback Loops	Recursive training on AI outputs narrows diversity and accuracy.	Models trained on synthetic data collapse in performance.	Cultural and epistemic narrowing; degraded AI reliability.	Curated training data; monitoring "synthetic contamination."
O – Organisational Memory Loss	Tacit knowledge migrates into AI prompts or vendor systems.	Firms are embedding SOPs into prompt libraries.	Fragile institutional memory; vendor lock-in.	Hybrid storage of organisational knowledge; resilience audits.
N – Normative Drift	AI defaults and guardrails shape cultural/ethical norms implicitly.	Model refusal styles become the de facto politeness standard.	Silent adoption of corporate norms without debate.	Transparency about normative choices; participatory design.

Note. Table 1 summarises the HORIZON taxonomy of "invisible costs" in generative AI, offering concise definitions, illustrative examples, key risks, and potential mitigation strategies. The taxonomy highlights how technical defaults and systemic properties of AI models can exert hidden but significant cultural, environmental, and organisational effects.

4. Minimal-Effort Measures: Making the Invisible Visible

Scholars have emphasised that evaluation frameworks for artificial intelligence (AI) too often privilege accuracy while neglecting broader ethical, epistemic, and ecological dimensions (Ge *et al*., 2010; Singh *et al*., 2014). In higher education, this narrow focus obscures fairness, accountability, transparency, and ethics (FATE), which shape how AI systems intersect with social and cognitive processes (Memarian and Doleck, 2023). Sustainability research likewise highlights the absence of standardised, transparent metrics for energy, water, and carbon disclosure, hindering accountability and comparability (Adelakun *et al*., 2024). Across domains from hallucination detection in radiology and business education (Dang and Nguyen, 2025; Hardy *et al*., 2024) to recommender system diversity (Ge *et al*., 2010), a common theme emerges: invisible costs must be rendered visible through simple, tractable measures.

We therefore propose four conceptual metrics, DAO, CDQ, EIS, and RTE, that distil these insights into minimal-effort heuristics for AI audits. DAO (Diversity of AI Outputs) operationalises concerns about homogenization and lack of serendipity in generative AI (Ge *et al*., 2010). CDQ (Cognitive Dependence Quotient) foregrounds automation bias and the risk of over-reliance, echoing findings that students often fail to detect AI hallucinations (Dang and Nguyen, 2024). EIS (Epistemic Integrity Score) responds to epistemic fragility documented in medical AI and education, where unverifiable or overconfident outputs erode trust (Hardy *et al*., 2024; Thomas *et al*., 2024). Finally, RTE (Resource Transparency Equivalent) adapts sustainability reporting practices, offering standardized disclosure of energy, water, and carbon per query (Adelakun *et al*., 2024; Basereh *et al*., 2021).

Table 2 summarizes these conceptual metrics, illustrating how they translate abstract ethical concerns into actionable indicators.

Table 2. Conceptual Metrics for Invisible Costs

Metric	What It Measures	Example	Application
DAO (Diversity of AI Outputs)	Lexical/semantic dispersion across multiple generations of same prompt.	Running 10 completions for one prompt, measuring variety.	Detecting homogenization; setting "diversity budgets."
CDQ (Cognitive Dependence Quotient)	Ratio of task steps done by AI vs human.	Student essay outline: 80% AI, 20% human.	Monitoring deskilling risk; thresholds for safety-critical domains.
EIS (Epistemic Integrity Score)	Proportion of outputs that express calibrated uncertainty & cite verifiable evidence.	10 fact queries → only 3 include source + uncertainty → EIS = 0.3.	Tracking epistemic trustworthiness of outputs.
RTE (Resource Transparency Equivalent)	Standardised disclosure of per-query energy, carbon, and water.	1,000 prompts → 12 kWh, 50 litres of water.	Sustainability reporting; consumer awareness.

5. Case Vignettes

5.1 Education: Homogenised Writing

Teachers increasingly report that student essays shaped by AI tools exhibit striking similarities in phrasing, argumentative structure, and rhetorical cadence, even when plagiarism detection software does not flag them. This phenomenon signals a shift from overt academic dishonesty to a subtler homogenization of discourse. While AI can scaffold grammar, coherence, and surface polish, its generative templates risk narrowing the expressive range of student writing (Pryma *et al*., 2025).

Empirical evidence suggests that this homogenization effect is already observable in practice. In a controlled experimental study conducted by researchers at Cornell University, participants from different cultural backgrounds (including U.S. and Indian students) were asked to write short essays with and without AI writing assistance. The study found that AI-assisted texts became significantly more like one another in terms of lexical choice, sentence structure, and rhetorical framing, thereby reducing culturally distinct and stylistically idiosyncratic features typically present in unaided writing. The authors conclude that AI suggestions systematically push users toward more generic, standardised forms of expression, demonstrating that convergence in writing style is not merely a theoretical concern but a measurable outcome of AI-mediated composition (Stanley, 2025).

The homogenization effect is not merely stylistic but epistemic. Studies show that AI writing assistants encourage formulaic arrangements and "robotic" sentence structures, often reducing opportunities for rhetorical experimentation and independent argument construction (Bašić *et al*., 2023). School and university educators worry that such reliance produces text that is grammatically correct yet cognitively thin, with diminished critical reasoning and originality (Akyıldız, 2024; Malik *et al*., 2023).

Survey-based research confirms that students themselves are ambivalent: they value AI's ability to improve fluency, reduce errors, and provide efficient scaffolding, yet many also fear it stifles their creative development and voice (Marrone *et al*., 2022; Sharma, 2025). This aligns with findings that younger generations rely more heavily on generative AI than teachers and parents, raising concerns about long-term dependence (Sharma *et al*., 2025).

Educators thus face a paradox. On one hand, AI can act as a relational artefact that supports collaboration and expands student exploration (Lim *et al*., 2023). On the other hand, unchecked use risks routinization, where students substitute authentic experimentation with AI-optimised phrasing. The result is a narrowing of the discursive field: writing that passes as "authentic" but lacks the idiosyncratic markers of human experimentation and voice (Avila-Chauvet and Mejía, 2023; Khalil and Er, 2023).

The case of homogenised writing underscores the need for pedagogical strategies that position AI as a supplement rather than a surrogate. As several studies emphasise, balanced integration requires teacher mediation, explicit creativity-focused tasks, and opportunities for students to deliberately diverge from AI-suggested patterns (Akyıldız, 2024; Lim *et al*., 2023). Without such measures, the promise of AI in education risks devolving into a culture of standardised expression, were efficiency eclipses originality.

5.2 Organisations: Prompt-Dependent Workflows

A growing number of organisations are migrating creative and operational tasks into AI-mediated environments, often organised around prompt libraries. While this shift increases efficiency and standardises outputs, it also risks restructuring organisational learning in ways that erode collective memory and tacit knowledge. Prompt engineering - whether through zero-shot, few-shot, or chain-of-thought techniques - allows firms to leverage pre-trained models without retraining (Gu *et al*., 2023; Sahoo *et al*., 2024). Yet this very reliance on externalised prompts transforms expertise from a situated, experiential practice into a procedural interaction with templates (Sikha *et al*., 2023).

This dynamic is not merely speculative, as prompt libraries are already being implemented in real organisational settings to structure and standardise AI-mediated work. Enterprise documentation from Microsoft describes prompt libraries as shared repositories of reusable prompts designed to accelerate task completion and ensure consistency across teams (Phil-cmd, 2024). Practitioner-oriented guidance likewise encourages organisations to centralise prompt development as part of formal governance structures to scale generative AI use efficiently (Themefisher, 2024). At the same time, empirical workplace research indicates a cognitive trade-off associated with such reliance. Qualitative interview-based research on human AI augmentation reports that employees express concerns about dependency and deskilling when AI systems increasingly mediate problem-solving activities, reducing opportunities for skill enactment and experiential learning. Complementary research on AI use in work contexts similarly highlights that automation and augmentation can lead to deskilling depending on how tasks are structured and routinised (Charpied, 2025).

Taken together, these sources provide verifiable evidence that organisations are actively structuring workflows around reusable prompts and that such arrangements can plausibly externalise reasoning processes, raising risks for the durability of organisational knowledge and skill development.

The case of a technology startup that migrated its design brainstorming into prompt-guided AI illustrates this risk vividly. Within months, employees ceased to learn the rationales underlying design choices; organisational memory was effectively outsourced to the model. This aligns with broader evidence that AI-based augmentation often produces deskilling, as workers lose opportunities for experimentation, overview, and reflective judgment (Crowston and Bolici, 2025; Huseynova, 2024). Although human AI augmentation is typically framed as complementary, blurred boundaries between augmentation and substitution frequently mean that workers merely input prompts and evaluate outputs, rather than engaging in deeper knowledge creation (Huseynova, 2024).

From a knowledge management perspective, the outsourcing of decision rationales to AI threatens the durability of organisational memory. Scholars highlight that while emerging technologies like AI can automate tacit knowledge capture, they also risk bias, over-reliance, and the erosion of unarticulated know-how (Nonato and Perez, 2025; Storey, 2025).

Traditional knowledge management strategies, personalisation (relying on human expertise) and codification (relying on stored databases) are both destabilised when AI itself becomes the locus of "hidden" organisational knowledge (Fteimi and Hopf, 2021). Without deliberate governance, firms risk creating brittle knowledge ecosystems, in which the interpretive capacities of employees atrophy while design rationales remain locked in opaque prompt–output cycles.

Nevertheless, research also emphasises that outcomes are not uniform: AI adoption in knowledge work can simultaneously produce new tasks, new roles, and skill requirements, especially when paired with participatory change management (von Richthofen *et al*., 2022). The organisational challenge, then, is not whether to use AI, but how to embed it without allowing prompt dependence to substitute for organisational reasoning. Building resilience requires designing workflows that deliberately expose employees to the "why" behind design choices, ensuring that organisational memory remains distributed among people, not just prompts.

6. Governance and Design Playbook

The transition from conceptual diagnosis to institutional response requires a clear linkage between invisible costs, their measurement, and possible interventions. Section 4 outlined minimal-effort metrics such as the Diversity of AI Outputs (DAO) and the Epistemic Integrity Score (EIS) as heuristics for rendering latent risks visible. Governance design must then translate these signals into actionable practices. Figure 3 depicts this flow: invisible costs are operationalised through metrics, which in turn provide entry points for governance interventions. By structuring the relationship in this way, the diagram underscores that interventions are not abstract aspirations but concrete responses to measurable patterns.

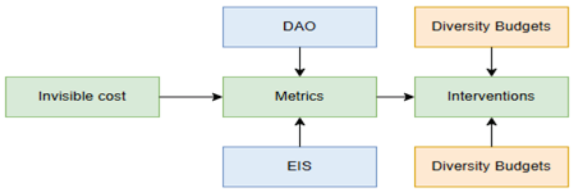

Figure 3: Governance flow - from Invisible cost to Interventions

The logic of diversity budgets, for example, follows directly from DAO: if the diversity metric declines, governance should impose quotas that encourage models to sample more widely, thereby sustaining a long-tail distribution of cultural expression (Shur-Ofry *et al*., 2024; Wan and Kalman, 2025). Similarly, a low EIS score calls for interventions such as "uncertainty by default," which aligns AI systems with scientific norms of hedging and tentativeness rather than overstated certainty (Ho and Caals, 2024; Wihbey, 2024).

To operationalise the playbook, consider two illustrative cases. In education, a persistently low Diversity of AI Outputs (DAO) score in student writing tools would trigger assignment redesign, such as requiring students to generate multiple AI-assisted drafts using different prompts or models and to explicitly diverge from AI outputs through reflection and revision. At the institutional level, DAO thresholds could be enforced by rotating approved AI models or limiting repeated reuse of identical prompt templates across courses. In organisational settings, a low Epistemic Integrity Score (EIS) in decision support systems would prompt design interventions such as uncertainty by default, requiring AI outputs to include confidence ranges, alternative explanations, or explicit unknowns. Firms could further reinforce EIS governance through periodic AI off workflows in which teams justify decisions without AI assistance, ensuring that reasoning remains distributed among employees rather than embedded solely in model outputs.

This governance flow reframes invisible risks as tractable levers of intervention. Rather than treating homogenization, epistemic fragility, or deskilling as diffuse concerns, the framework connects each to a corresponding design doctrine. In doing so, it embeds accountability at the level of system design: diversity quotas as correctives to monoculture, epistemic hedging as insurance against lock-in, periodic "AI-off" practices as safeguards against deskilling, and transparency labels as accountability mechanisms for ecological costs (Agha *et al*., 2025; Campo-Ruiz, 2025).

What emerges is a playbook that treats governance as an iterative feedback loop. Metrics track the health of epistemic and cultural ecosystems, while interventions are triggered when thresholds are crossed. The diagram thus represents more than a static mapping: it signals a dynamic governance architecture, one capable of adapting as invisible costs surface and as interventions reshape the terrain.

7. Conclusion

This paper has advanced the argument that generative AI should be understood not merely as a tool or agent but as an ecology, an environment that reshapes the very conditions of cognition, culture, and epistemic trust. This ecological framing exposes risks that are neither captured by benchmark performance nor reducible to carbon costs. They are instead slow-moving, systemic transformations: the homogenization of discourse, the erosion of skills through over-reliance, the depletion of hidden ecological resources, and the destabilisation of information integrity, recursive feedback loops, organisational memory loss, and normative drift. Taken together, these invisible costs suggest that the most consequential impact of AI may not be technical error, but ecological imbalance.

This analysis does not deny the substantial benefits of generative AI. Across education and organisational contexts, AI systems demonstrably improve efficiency, reduce cognitive load on routine tasks, and expand access to expertise by supporting users in drafting, summarising, coding, and problem-solving. In many cases, these systems enable individuals and institutions to perform tasks that would otherwise be prohibitively time-consuming or inaccessible. The concern addressed in this paper is therefore not whether AI should be used, but how its benefits can be realised without allowing efficiency gains to obscure or amplify longer-term epistemic, cultural, and organisational costs.

By foregrounding invisible costs, this work makes two contributions. First, it reframes existing debates on AI evaluation, which remain dominated by visible metrics of performance, fairness, and emissions (Luccioni *et al*., 2024; Eriksson *et al*., 2025). While such metrics are necessary, they are insufficient to account for AI's role in shaping human cognitive infrastructures. The ecological perspective insists that intelligence is sustained not by isolated algorithms but by interdependent systems that draw resilience from diversity, redundancy, and contextual adaptation (Hutchins, 2010; Tribble and Sutton, 2011). Second, the paper operationalises this insight through the HORIZON taxonomy and the proposed indicators DAO, CDQ, EIS, and RTE. These minimal-effort measures translate abstract risks into actionable metrics, rendering latent costs visible and therefore governable.

The implications extend across domains. In education, over-reliance on generative systems risks narrowing expression and critical reasoning, demanding pedagogical interventions that cultivate divergence rather than conformity (Agarwal *et al*., 2025; Pryma *et al*., 2025). In organisations, prompt-dependent workflows may erode tacit expertise, raising questions about how to sustain institutional memory when knowledge is increasingly externalised into vendor-controlled infrastructures (Jarrahi *et al*., 2023; Falckenthal *et al*., 2025). At the societal level, recursive training on AI-generated content risks epistemic collapse, as synthetic data feeds back into future models, progressively narrowing the representational horizon (Borji, 2024; Gerstgrasser *et al*., 2024). These trajectories highlight that the stakes of AI adoption are not simply efficiency or productivity but the health of cognitive ecologies that underpin democratic deliberation, cultural vitality, and organisational resilience.

Yet these risks also point to constructive pathways. If homogenization is a stressor, then deliberate diversity budgets in generative outputs can sustain pluralism. If epistemic integrity is fragile, then uncertainty-by-default and verifiable citation protocols can align machine discourse with scientific norms (Ho and Caals, 2024; Krishnan *et al*., 2024). If organizational knowledge risks decontextualization, then hybrid approaches that pair AI archives with embodied apprenticeship can preserve tacit expertise (Nonaka and Takeuchi, 2021). And if resource externalities are obscured by carbon tunnel vision, then RTE-style disclosures can surface the full ecological footprint of AI, enabling informed governance (Adelakun *et al*., 2024; Berthelot *et al*., 2024). These interventions are modest in design but systemic in effect: they recalibrate incentives away from narrow optimization toward stewardship of the conditions under which intelligence thrives.

The limitations of this study must also be acknowledged. The HORIZON taxonomy is necessarily conceptual and exploratory; further empirical work is needed to test its categories, refine its measures, and assess its applicability across cultural and institutional contexts. The proposed indicators are heuristic rather than standardized metrics, requiring interdisciplinary collaboration to integrate them into regulatory frameworks and organizational practice. Moreover, while the ecological metaphor provides analytical leverage, it should not obscure the material and political dimensions of AI infrastructures, which are shaped by corporate interests, state power, and global inequities (Chester and Allenby, 2023; Scheuerman et al., 2021).

Nonetheless, the ecological perspective advanced here is intended as a provocation to reorient the discourse. The central question is not whether AI systems outperform benchmarks, but whether they enrich or erode the ecologies of human thought. To frame AI as ecology is to recognize that invisible costs are not marginal side effects but central dynamics, shaping what kinds of knowledge endure, whose voices are amplified, and what forms of reasoning are considered legitimate. Future research must therefore move beyond accuracy and fairness audits toward ecological audits that assess the resilience of cultural, cognitive, and epistemic systems in the presence of pervasive generative models.

If the twentieth century was defined by the engineering of technical systems, the twenty-first will be defined by the stewardship of cognitive ecologies. Generative AI will continue to proliferate; the task is to ensure that its integration strengthens rather than corrodes the infrastructures of thought. That task requires not only better models but also better metaphors, better measures, and above all, better care for the habitats of human cognition.

Reference

Adelakun, B.O., Antwi, B.O., Ntiakoh, A., and Eziefule, A.O. (2024) 'Leveraging AI for sustainable accounting: Developing models for environmental impact assessment and reporting', *Finance & Accounting Research Journal*, 6(6), pp. 1017–1048. https://doi.org/10.51594/farj.v6i6.1234

Agarwal, D., Naaman, M. and Vashistha, A. (2025) 'AI suggestions homogenize writing toward western styles and diminish cultural nuances'. In: P*roceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.48550/arXiv.2409.11360

Agha, R.A., Mathew, G., Rashid, R., Kerwan, A., Al-Jabir, A., Sohrabi, C., Franchi, T., Nicola, M., Agha, M. and TITAN Group (2025) 'Transparency in the reporting of artificial intelligence – the TITAN guideline', *Premier Journal of Science*, 10, 100082. https://doi.org/10.70389/PJS.100082

Akyıldız, S.T. (2024) 'Enhancing or hindering? AI's role in sparking creativity in language teaching: insights from private high school EFL teachers', *International e-Journal of Educational Studies*, 8(18), pp. 234–254. https://doi.org/10.31458/iejes.1502509

Avila-Chauvet, L. and Mejía, D. (2023) 'Can professors and students detect ChatGPT essays?' *SSRN Electronic Journal*. Available at: https://ssrn.com/abstract=4373643. https://doi.org/10.2139/ssrn.4373643

Balasubramaniam, N., Kauppinen, M., Rannisto, A., Hiekkanen, K. and Kujala, S. (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

Basereh, M., Caputo, A. and Brennan, R. (2021) 'FAIR ontologies for transparent and accountable AI: a hospital adverse incidents vocabulary case study', In: P*roceedings of the Third International Conference on Transdisciplinary AI (TransAI 2021)*, pp. 92–97. https://doi.org/10.1109/TransAI51903.2021.00024

Bašić, Ž., Banovac, A., Kružić, I. and Jerković, I. (2023) 'ChatGPT-3.5 as writing assistance in students' essays', *Humanities and Social Sciences Communications*, 10, 750. https://doi.org/10.1057/s41599-023-02269-7

Berthelot, A., Caron, E., Jay, M. and Lefèvre, L. (2024) 'Estimating the environmental impact of generative-AI services using an LCA-based methodology', *Procedia CIRP*, 122, pp. 707–712. https://doi.org/10.1016/j.procir.2024.01.098

Borji, A. (2024) A note on Shumailov et al. (2024): AI models collapse when trained on recursively generated data. arXiv. Available at: https://arxiv.org/abs/2410.12954 or https://doi.org/10.48550/arXiv.2410.12954

Bushuyev, S., Bushuiev, D., Bushuieva, V., Bushuyeva, N. and Murzabekova, S. (2024) 'The erosion of competencies in managing innovation projects due to the impact of ubiquitous artificial intelligence systems', *Procedia Computer Science*, 231, pp. 403–408. https://doi.org/10.1016/j.procs.2023.12.225

Campo-Ruiz, I. (2025) 'Artificial intelligence may affect diversity: architecture and cultural context reflected through ChatGPT, Midjourney, and Google Maps', *Humanities and Social Sciences Communications*, 12, 24. https://doi.org/10.1057/s41599-024-03968-5

Chan, A., Wei, K., Huang, S., Rajkumar, N., Perrier, E., Lazar, S. and Anderljung, M. (2025) 'Infrastructure for AI agents', arXiv. Available at: https://arxiv.org/abs/2501.10114 or https://doi.org/10.48550/arXiv.2501.10114

Chester, M.V. and Allenby, B. (2023) 'Infrastructure and the cognitive ecosystem: an irrevocable transformation', *Environmental Research: Infrastructure and Sustainability*, 3(3), 033002. https://doi.org/10.1088/2634-4505/aced1f

Church, K.W. and Hestness, J. (2019) 'A survey of 25 years of evaluation', *Natural Language Engineering*, 25(6), pp. 753–767. https://doi.org/10.1017/S1351324919000275

Charpied, B. (2025) 'Deskilling, reskilling, and upskilling', *AI in the Workplace [Preprint]*. https://doi.org/10.4135/9781071982846

Conde, J., Gonzalez, M., Martínez, G., Moral, F., Merino-Gomez, E. and Reviriego, P. (2025) 'Recursive inpainting (RIP): how much information is lost under recursive inferences?', *AI & Society*, pp. 1–17. https://doi.org/10.1007/s00146-025-02351-5

Crowston, K. and Bolici, F. (2025) 'Deskilling and upskilling with AI systems', *Information Research: An International Electronic Journal*, 30(iConf), pp. 1009–1023. https://doi.org/10.47989/ir30iConf47143

Dang, C.T. and Nguyen, A. (2025) 'Distinguishing fact from fiction: student traits, attitudes, and AI hallucination detection in business school assessment', *arXiv*. Available at: https://arxiv.org/abs/2506.00050 or https://doi.org/10.48550/arXiv.2506.00050

Eriksson, M., Purificato, E., Noroozian, A., Vinagre, J., Chaslot, G., Gomez, E. and Fernandez-Llorca, D. (2025) 'Can we trust AI benchmarks? An interdisciplinary review of current issues in AI evaluation'. *arXiv*. Available at: https://arxiv.org/abs/2502.06559 or https://doi.org/10.48550/arXiv.2502.06559

Evans, O., Cotton-Barratt, O., Finnveden, L., Bales, A., Balwit, A., Wills, P. and Saunders, W. (2021) 'Truthful AI: developing and governing AI that does not lie', *arXiv*. Available at: https://arxiv.org/abs/2110.06674 (Accessed: 13 January 2026). https://doi.org/10.48550/arXiv.2110.06674

Falckenthal, B., Au-Yong-Oliveira, M. and Figueiredo, C. (2025) 'Intergenerational tacit knowledge transfer: leveraging AI', *Societies*, 15(8), 213. https://doi.org/10.3390/soc15080213

Farias, G., Rodrigues, P., Adamatti, D. and Gonçalves, E. (2024) 'Model for knowledge transfer in agent organizations: a case study on Moise+', In: *The International FLAIRS Conference Proceedings*. https://doi.org/10.32473/flairs.37.1.135476

Ferrario, A., Facchini, A. and Termine, A. (2024) 'Experts or authorities? The strange case of the presumed epistemic superiority of artificial intelligence systems', *Minds and Machines*, 34(3), 30. https://doi.org/10.1007/s11023-024-09681-1

Fteimi, N. and Hopf, K. (2021) 'Knowledge management in the era of artificial intelligence: developing an integrative framework', *ResearchGate*. Available at: https://www.researchgate.net/publication/352410098_Knowledge_Management_in_the_Era_of_Artificial_Intelligence_--_Developing_an_Integrative_Framework (Accessed: 13 January 2026).

Gadde, A. (2025) 'AI-enhanced knowledge management systems in enterprises: transforming organizational intelligence', *World Journal of Advanced Research and Reviews*, 26(2), pp. 2020–2030. https://doi.org/10.30574/wjarr.2025.26.2.1913

Ge, M., Delgado-Battenfeld, C. and Jannach, D. (2010) 'Beyond accuracy: evaluating recommender systems by coverage and serendipity', In: *Proceedings of the Fourth ACM Conference on Recommender Systems*. https://doi.org/10.1145/1864708.1864761

Gerstgrasser, M., Schaeffer, R., Dey, A., Rafailov, R., Sleight, H., Hughes, J. and Gromov, A. (2024) 'Is model collapse inevitable? Breaking the curse of recursion by accumulating real and synthetic data', *arXiv*. Available at: https://arxiv.org/abs/2404.01413 (Accessed: 13 January 2026). https://doi.org/10.48550/arXiv.2404.01413

Gillespie, T. (2024) 'Generative AI and the politics of visibility', *Big Data & Society*, 11(2), 20539517241252131. https://doi.org/10.1177/20539517241252131

Gomez, E., Porcaro, L., Frau, A. and Vinagre, J. (2024) 'Diversity in artificial intelligence conferences', Available at: https://doi.org/10.2760/796551

Gu, J., Han, Z., Chen, S., Beirami, A., He, B., Zhang, G. and Torr, P. (2023) 'A systematic survey of prompt engineering on vision-language foundation models', *arXiv*. Available at: https://arxiv.org/abs/2307.12980 or https://doi.org/10.48550/arXiv.2307.12980

Hardy, R., Kim, S.E., Ro, D.H. and Rajpurkar, P. (2024) 'Rextrust: a model for fine-grained hallucination detection in AI-generated radiology reports', *arXiv*. Available at: https://proceedings.mlr.press/v281/hardy25a.html or https://arxiv.org/abs/2412.15264

Hildebrandt, M. (2021) 'The issue of bias: the framing powers', In: *Machines We Trust: Perspectives on Dependable AI*, 43. Available at: https://doi.org/10.2139/ssrn.3497597 (Accessed: 13 January 2026)

Hiremath, R.B. (2024) 'AI-embedded data centres: promoting sustainability and reducing water footprint', In: *Proceedings of the 2024 First International Conference on Data, Computation and Communication (ICDCC)*. https://doi.org/10.1109/ICDCC62744.2024.10961316

Ho, C.W.-L. and Caals, K. (2024) 'How the EU AI Act seeks to establish an epistemic environment of trust', *Asian Bioethics Review*, 16(3), pp. 345–372. https://doi.org/10.1007/s41649-024-00304-6

Huseynova, F. (2024) 'Addressing deskilling as a result of human-AI augmentation in the workplace', *CEUR Workshop Proceedings*, Available at: https://ceur-ws.org/Vol-3901/short_5.pdf (Accessed: 13 January 2026)

Hutchins, E. (2010) 'Cognitive ecology', *Topics in Cognitive Science*, 2(4), pp. 705–715. https://doi.org/10.1111/j.1756-8765.2010.01089.x

Inie, N., Falk, J. and Selvan, R. (2025a) 'How CO2STLY is CHI? The carbon footprint of generative AI in HCI research and what we should do about it', In: *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3706598.3714227

Inie, N., Falk, J. and Selvan, R. (2025b) 'The HCI GenAI CO2ST calculator: a tool for calculating the carbon footprint of generative AI use in human-computer interaction research', *arXiv*. Available at: https://arxiv.org/abs/2504.00692 (Accessed: 13 January 2026). https://doi.org/10.48550/arXiv.2504.00692

Jarrahi, M.H., Askay, D., Eshraghi, A. and Smith, P. (2023) 'Artificial intelligence and knowledge management: a partnership between human and AI', *Business Horizons*, 66(1), pp. 87–99. https://doi.org/10.1016/j.bushor.2022.03.002

Ji, Z., Chen, D., Ishii, E., Cahyawijaya, S., Bang, Y., Wilie, B. and Fung, P. (2024) 'LLM internal states reveal hallucination risk faced with a query', *arXiv*. Available at: https://arxiv.org/abs/2407.03282 (Accessed: 13 January 2026). https://doi.org/10.48550/arXiv.2407.03282

Kandasamy, S. (2025) 'Control plane as a tool: a scalable design pattern for agentic AI systems', *arXiv*. Available at: https://arxiv.org/abs/2505.06817 (Accessed: 13 January 2026). https://doi.org/10.48550/arXiv.2505.06817

Karpouzis, K. (2024) 'Plato's shadows in the digital cave: Controlling cultural bias in generative AI', *Electronics*, 13(8), 1457. https://doi.org/10.3390/electronics13081457

Khalil, M. and Er, E. (2023) 'Will ChatGPT get you caught? Rethinking of plagiarism detection', In: *International Conference on Human-Computer Interaction*. Available at: https://doi.org/10.48550/arXiv.2302.04335

Kneese, T. and Young, M. (2024) 'Carbon emissions in the tailpipe of generative AI', *Harvard Data Science Review*, 5. https://doi.org/10.1162/99608f92.fbdf6128

Krishnan, R., Khanna, P. and Tickoo, O. (2024) 'Enhancing trust in large language models with uncertainty-aware fine-tuning', *arXiv preprint*, arXiv:2412.02904. Available at: https://doi.org/10.48550/arXiv.2412.02904

Lim, J., Leinonen, T., Lipponen, L., Lee, H., DeVita, J. and Murray, D. (2023) 'Artificial intelligence as relational artifacts in creative learning', *Digital Creativity*, 34(3), pp. 192–210. https://doi.org/10.1080/14626268.2023.2236595

Lu, T. (2025) 'Maximum hallucination standards for domain-specific large language models', *arXiv preprint*, arXiv:2503.05481. Available at: https://doi.org/10.48550/arXiv.2503.05481

Luccioni, S., Jernite, Y. and Strubell, E. (2024) 'Power hungry processing: Watts driving the cost of AI deployment?', In: *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. Available at: https://doi.org/10.1145/3630106.365854

Malik, A.R., Pratiwi, Y., Andajani, K., Numertayasa, I.W., Suharti, S. and Darwis, A. (2023) 'Exploring artificial intelligence in academic essay: Higher education student's perspective', *International Journal of Educational Research Open*, 5, 100296. https://doi.org/10.1016/j.ijedro.2023.100296

Markolf, S.A., Chester, M.V. and Allenby, B. (2021) 'Opportunities and challenges for artificial intelligence applications in infrastructure management during the Anthropocene', *Frontiers in Water*, 2, 551598. https://doi.org/10.3389/frwa.2020.551598

Marrone, R., Taddeo, V. and Hill, G. (2022) 'Creativity and artificial intelligence—A student perspective', *Journal of Intelligence*, 10(3), 65. https://doi.org/10.3390/jintelligence10030065

Matueny, R.M. and Nyamai, J.J. (2025) 'Illusion of competence and skill degradation in artificial intelligence dependency among users', *International Journal of Research and Scientific Innovation*. https://doi.org/10.51244/IJRSI.2025.120500163

McIntosh, T.R., Susnjak, T., Arachchilage, N., Liu, T., Xu, D., Watters, P. and Halgamuge, M.N. (2025) 'Inadequacies of large language model benchmarks in the era of generative artificial intelligence', *IEEE Transactions on Artificial Intelligence*. https://doi.org/10.1109/TAI.2025.3569516

Memarian, B. and Doleck, T. (2023) 'Fairness, accountability, transparency, and ethics (FATE) in artificial intelligence (AI) and higher education: A systematic review', *Computers and Education: Artificial Intelligence*, 5, 100152. https://doi.org/10.1016/j.caeai.2023.100152

Modi, T.B. (2023) 'Artificial intelligence ethics and fairness: A study to address bias and fairness issues in AI systems, and the ethical implications of AI applications', *Revista Review Index Journal of Multidisciplinary*, 3(2), pp. 24–35. https://doi.org/10.31305/rrijm2023.v03.n02.004

Murray-Rust, D., Nicenboim, I. and Lockton, D. (2022) 'Metaphors for designers working with AI', In: *DRS 2022: Design Research Society International Conference, Bilbao*. https://doi.org/10.21606/drs.2022.667

Natarajan, A. (2025) 'The hydro-digital paradox: Water scarcity in the age of artificial intelligence', *Journal of Computer Science and Technology Studies*, 7(7), pp. 500–505. https://doi.org/10.32996/jcsts.2025.7.7.55

Nathim, K.W., Hameed, N.A., Salih, S.A., Taher, N.A., Salman, H.M. and Chornomordenko, D. (2024) 'Ethical AI with balancing bias mitigation and fairness in machine learning models', In: *Proceedings of the 2024 36th Conference of Open Innovations Association (FRUCT)*. https://doi.org/10.23919/FRUCT64283.2024.10749873

Nonaka, I. and Takeuchi, H. (2021) 'Humanizing strategy', *Long Range Planning*, 54(4), 102070. https://doi.org/10.1016/j.lrp.2021.102070

Nonato, J.A.A. and Perez, G. (2025) 'From challenges to opportunities: The impact of emerging technologies on enhancing organizational memory systems', *International Journal of Scientific Management and Tourism*, 11(1), p. e1247. https://doi.org/10.55905/ijsmtv11n1-001

Nurhaeni, H., Delhi, A., Daeli, O.P.M., Anjani, S.A. and Yusuf, N.A. (2024) 'Optimizing electrical energy use through AI: An integrated approach for efficiency and sustainability.', *International Transactions on Artificial*

Intelligence, 2(2), pp. 106–113. https://doi.org/10.33050/italic.v2i2.533

Omicini, A., Piunti, M., Ricci, A. and Viroli, M. (2009) ‘Agents, intelligence and tools’, In: *Bramer, M. (ed.) Artificial Intelligence: An International Perspective. Lecture Notes in Computer Science*, vol. 5640. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-03226-4_9

Ong, P. and Png, I.P.L. (2023) ‘Automation, deskilling, and labor supply: Empirical evidence’, Available at SSRN: https://ssrn.com/abstract=4666472 or http://dx.doi.org/10.2139/ssrn.4666472

Orr, W. and Kang, E.B. (2024) ‘AI as a sport: On the competitive epistemologies of benchmarking’, In: *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. Available at: https://doi.org/10.1145/3630106.3659012

Palmini, M.-T.D.R. and Cetinic, E. (2024) ‘Patterns of creativity: How user input shapes AI-generated visual diversity’, *arXiv preprint*, arXiv:2410.06768. Available at: https://doi.org/10.48550/arXiv.2410.06768

Phil-cmd (2024) ‘Get started with prompt library – Microsoft Copilot Studio’, *Microsoft Copilot Studio | Microsoft Learn*. Available at: https://learn.microsoft.com/en-us/microsoft-copilot-studio/prompt-library (Accessed: 23 December 2025).

Pandey, P., Singh, T. and Kumar, S. (2023) ‘Cognitive offloading: Systematic review of a decade’, *International Journal of Indian Psychology*, 11(24), pp. 1545–1563. https://doi.org/10.25215/1102.163

Pryma, V., Pelivan, O., Teletska, T., Tsobenko, O. and Zagrebelna, N. (2025) ‘AI writing assistants and student competence: A linguistic aspect’, *Arab World English Journal (AWEJ) Special Issue on Artificial Intelligence*, pp. 319–329. Available at: https://dx.doi.org/10.24093/awej/AI.18

Qin, J., Liu, B. and Nguyen, Q.D. (2024) ‘Enhancing healthcare LLM trust with atypical presentations recalibration’, *arXiv preprint*, arXiv:2409.03225. Available at: https://doi.org/10.48550/arXiv.2409.03225

Raji, I.D., Bender, E.M., Paullada, A., Denton, E. and Hanna, A. (2021) ‘AI and the everything in the whole wide world benchmark’, *arXiv preprint*, arXiv:2111.15366. Available at: https://doi.org/10.48550/arXiv.2111.15366

Rana, N.P., Pillai, R., Sivathanu, B. and Malik, N. (2024) ‘Assessing the nexus of generative AI adoption, ethical considerations and organizational performance’, *Technovation*, 135, 103064. https://doi.org/10.1016/j.technovation.2024.103064

Rauh, M., Marchal, N., Manzini, A., Hendricks, L.A., Comanescu, R., Akbulut, C., Stepleton, T., Mateos-Garcia, J., Bergman, S., Kay, J., Griffin, C., Bariach, B., Gabriel, I., Rieser, V., Isaac, W. and Weidinger, L. (2024) ‘Gaps in the safety evaluation of generative AI’, *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 7(1), pp. 1200–1217. https://doi.org/10.1609/aies.v7i1.31717

Rinta-Kahila, T., Penttinen, E., Salovaara, A. and Soliman, W. (2018) ‘Consequences of discontinuing knowledge work automation: Surfacing of deskilling effects and methods of recovery’, In: *Proceedings of the 51st Hawaii International Conference on System Sciences (HICSS)*. https://doi.org/10.24251/HICSS.2018.654

Russo, D. (2024) ‘Navigating the complexity of generative AI adoption in software engineering’, *ACM Transactions on Software Engineering and Methodology*, 33(8), Article 221, pp. 1–5. https://doi.org/10.1145/3680471

Sahoo, P., Singh, A.K., Saha, S., Jain, V., Mondal, S. and Chadha, A. (2024) ‘A systematic survey of prompt engineering in large language models: Techniques and applications’, *arXiv preprint*, arXiv:2402.07927. Available at: https://doi.org/10.48550/arXiv.2402.07927

Satsangi, P. (2019) ‘Automation of tacit knowledge using machine learning’, In: *Proceedings of the 2019 6th International Conference on Soft Computing & Machine Intelligence (ISCMI), Johannesburg, South Africa*, pp. 35–39. https://doi.org/10.1109/ISCMI47871.2019.9004290

Scheuerman, M.K., Hanna, A. and Denton, R. (2021) ‘Do datasets have politics? Disciplinary values in computer vision dataset development’, *Proceedings of the ACM on Human-Computer Interaction, 5(CSCW2)*, Article 317, pp. 1–37. https://doi.org/10.1145/3476058

Schlangen, D. (2020) ‘Targeting the benchmark: On methodology in current natural language processing research’, *arXiv preprint*, arXiv:2007.04792. Available at: https://doi.org/10.48550/arXiv.2007.04792

Seddik, M.E.A., Chen, S.-W., Hayou, S., Youssef, P. and Debbah, M. (2024) ‘How bad is training on synthetic data? A statistical analysis of language model collapse’, *arXiv preprint*, arXiv:2404.05090. Available at: https://doi.org/10.48550/arXiv.2404.05090

Sellar, S. and Gulson, K.N. (2021) ‘Becoming information centric: The emergence of new cognitive infrastructures in education policy’, *Journal of Education Policy*, 36(3), pp. 309–326. https://doi.org/10.1080/02680939.2019.1678766

Sharma, R., Aakanksha, S., Shrivastava, S. and Singh, S. (2025) ‘The influence of artificial intelligence on students’ creativity: perspectives and perceptions’, *Journal of Artificial Intelligence and Autonomous Intelligence Research*, 2(1), pp. 197–215. https://doi.org/10.54364/JAIAI.2024.1114

Shumba, N., Tshekiso, O., Li, P., Fanti, G. and Ren, S. (2025) ‘A water efficiency dataset for African data centers’, In: *Proceedings of the ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS ’25)*, pp. 453–460. Available at: https://doi.org/10.1145/3715335.3735483

Shur-Ofry, M., Horowitz-Amsalem, B., Rahamim, A. and Belinkov, Y. (2024) ‘Growing a tail: Increasing output diversity in large language models’, *arXiv preprint*, arXiv:2411.02989. Available at: https://doi.org/10.48550/arXiv.2411.02989

Sikha, V.K., Siramgari, D. and Korada, L. (2023) ‘Mastering prompt engineering: Optimizing interaction with generative AI agents’, *Journal of Engineering and Applied Sciences Technology*, 5(6), pp. 2–8. Available at: https://doi.org/10.47363/JEAST/2023(5)E117

Singh, A. (2024) ‘Diverse yet biased: towards mitigating biases in generative AI (student abstract)’, *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(21). Available at: https://doi.org/10.1609/aaai.v38i21.30512

Singh, B., Guldhe, A., Rawat, I. and Bux, F. (2014) ‘Towards a sustainable approach for development of biodiesel from plant and microalgae’, *Renewable and Sustainable Energy Reviews*, 29, pp. 216–245. Available at: https://doi.org/10.1016/j.rser.2013.08.067

Sinwell, L., Bruns, J., Budde, M. and Abecker, A. (2021) ‘A meta-analysis of the status of AI in environmental computer science’, In: *INFORMATIK 2021. Gesellschaft für Informatik (GI), Bonn*.

Sourati, Z., Karimi-Malekabadi, F., Ozcan, M., McDaniel, C., Ziabari, A., Trager, J., et al. (2025) ‘The shrinking landscape of linguistic diversity in the age of large language models’, *arXiv preprint*, arXiv:2502.11266.

Available at: https://doi.org/10.48550/arXiv.2502.11266

Stanley, G. (2025) 'AI assistants make writing more generic, Western', *Cornell Tech*. Available at: https://tech.cornell.edu/news/ai-writing-assistants-western/ (Accessed: 23 December 2025).

Storey, V.C. (2025) 'Knowledge management in a world of generative AI: Impact and implications', *ACM Transactions on Management Information Systems*, 16(3), Article 26, pp. 1–14. https://doi.org/10.1145/371920

Stryker, C. (2024) 'What is artificial intelligence (AI)', *IBM*. Available at: https://www.ibm.com/topics/artificial-intelligence

Sunkara, K.N.K.C. (2025) 'Power consumption and heat dissipation in AI data centers: A comparative analysis', *International Journal of Innovative Research in Science, Engineering and Technology*. Available at: https://doi.org/10.15680/IJIRSET.2025.1402015

Tao, L., Yeh, Y.-F., Dong, M., Huang, T., Torr, P. and Xu, C. (2025) 'Revisiting uncertainty estimation and calibration of large language models', *arXiv preprint.* arXiv:2505.23854. Available at: https://doi.org/10.48550/arXiv.2505.23854

Themefisher (2024) 'Build a prompt library with Microsoft Lists', *Microsoft 365 and Power Platform Community Blog*. Available at: https://pnp.github.io/blog/post/build-a-prompt-library-with-microsoft-lists/ (Accessed: 23 December 2025).

Thomas, A., Rosen, S. and Vettrivel, V. (2024) 'Seeing through the fog: A cost-effectiveness analysis of hallucination detection systems', *arXiv preprint.* arXiv:2411.05270. Available at: https://doi.org/10.48550/arXiv.2411.05270

Tribble, E. and Sutton, J. (2011) 'Cognitive ecology as a framework for Shakespearean studies', Available at: https://philarchive.org/rec/SUTCEA

van der Weij, T., Hofstätter, F., Jaffe, O., Brown, S.F. and Ward, F.R. (2024) 'AI sandbagging: language models can strategically underperform on evaluations', *arXiv*. Available at: https://arxiv.org/abs/2406.07358 or https://doi.org/10.48550/arXiv.2406.07358

von Richthofen, G., Ogolla, S. and Send, H. (2022) 'Adopting AI in the context of knowledge work: empirical insights from German organizations', *Information*, 13(4), 199. https://doi.org/10.3390/info13040199

Wagner, N., Desmond, M., Nair, R., Ashktorab, Z., Daly, E.M., Pan, Q. and Geyer, W. (2024) 'Black-box uncertainty quantification method for LLM-as-a-judge', *arXiv*. Available at: https://arxiv.org/abs/2410.11594 or https://doi.org/10.48550/arXiv.2410.11594

Wamba, S.F., Queiroz, M.M., Jabbour, C.J.C. and Shi, C.V. (2023) 'Are both generative AI and ChatGPT game changers for 21st-century operations and supply chain excellence?', *International Journal of Production Economics*, 265, 109015. https://doi.org/10.1016/j.ijpe.2023.109015

Wan, Y. and Kalman, Y.M. (2025) 'Using generative AI personas increases collective diversity in human ideation', *arXiv*. Available at: https://arxiv.org/abs/2504.13868 or https://doi.org/10.48550/arXiv.2504.13868

Wright, C.S. (2025) 'Beyond prediction: structuring epistemic integrity in artificial reasoning systems', *arXiv*. Available at: https://arxiv.org/abs/2506.17331 or https://doi.org/10.48550/arXiv.2506.17331

Zakizadeh, M. and Zand, M. (2024) 'Transforming the energy sector: unleashing the potential of AI-driven energy intelligence, energy business intelligence, and energy management system for enhanced efficiency and sustainability', In: *Proceedings of the 2024 20th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP)*. https://doi.org/10.1109/AISP61396.2024.10475298

Zhang, S., Xu, J. and Alvero, A. (2025) 'Generative AI meets open-ended survey responses: research participant use of AI and homogenization', *Sociological Methods & Research*, 00491241251327130. https://doi.org/10.1177/00491241251327130

Zhang, W., Lai, K.-H. and Gong, Q. (2024) 'The future of the labor force: higher cognition and more skills', *Humanities and Social Sciences Communications*, 11(1), pp. 1–9. https://doi.org/10.1057/s41599-024-02962-1

Zhou, E. and Lee, D. (2024) 'Generative artificial intelligence, human creativity, and art', *PNAS Nexus*, 3(3), p. pgae052. https://doi.org/10.1093/pnasnexus/pgae052