Clinical Psychiatry
ISSN: 2471-9854

Open access Review Article

The Amplification and Perpetuation of AI-Derived Biases through Automation Dependency: A Framework for Understanding the Long-Term Cognitive and Social Implications of LLM Over-Reliance

Christopher Cleverly*

Independent Researcher, London, United Kingdom

ABSTRACT

This research introduces a framework to elucidate how automation bias in Large Language Models (LLMs) amplifies biases through human over-reliance, leading to critical thinking atrophy and the propagation of biases into human cognition and social systems. Automation bias, defined as the tendency to excessively trust AI outputs while ignoring contradictory evidence or personal judgment, drives a three-phase cycle: (1) initial dependency development, fueled by perceived AI efficiency; (2) critical thinking atrophy via cognitive offloading; and (3) bias internalization and propagation, where AI biases are inherited and reproduced in human decisions, even without AI support. Drawing on evidence such as the impact of AI on 40% of global jobs and cognitive offloading in education, we challenge the notion that technical fixes alone can mitigate these effects. We propose Wisdom as a Service (WaaS), a preliminary framework that integrates non-European wisdom traditions. This theoretical approach prioritizes epistemic pluralism and community validation as potential pathways to address the long-term societal consequences of AI over-reliance (integrating non-European wisdom traditions (e.g., Ubuntu, Nyāya) and decolonized AI architectures to disrupt bias amplification). Automation bias is not just a human-AI interaction problem but a sociocognitive epidemic. Without systemic intervention (e.g., WaaS, decolonized AI), AI biases will become permanent fixtures of human reasoning. This framework prioritizes epistemic multiplicity, community validation, and culturally grounded reasoning to address the long-term societal consequences of AI over-reliance.

Keywords: Automation bias; AI over-reliance; Cognitive offloading; Bias amplification; Critical thinking; LLM bias; Wisdom as a Service; Decolonized AI; Epistemic injustice; Cognitive inheritance

INTRODUCTION

The automation bias crisis

The problem of AI over-reliance: Automation bias, defined as the tendency to excessively trust AI-generated advice while ignoring contradictory information or personal judgment, is a growing concern as LLMs permeate critical domains like healthcare, education, and national security [1,2]. The International Monetary Fund reports that AI impacts 40% of global jobs, with 60% of jobs in advanced economies at risk of displacement or wage suppression, amplifying the potential for automation bias to exacerbate inequities [3]. In healthcare, for instance, automation bias in AI-enabled Clinical Decision Support Systems (CDSS) can lead to deferral to incorrect AI diagnoses in 7% of cases, even among trained experts under time pressure [4]. Straw, for example, highlights how medical AI systems risk encoding and amplifying historical biases when used uncritically, reinforcing disparities under the guise of

Received:	29-Oct-2025	**Manuscript No:**	IPCP-25-22993
Editor assigned:	31-Oct-2025	**PreQC No:**	IPCP-25-22993 (PQ)
Reviewed:	14-Nov-2025	**QC No:**	IPCP-25-22993
Revised:	21-Nov-2025	**Manuscript No:**	IPCP-25-22993 (R)
Published:	28-Nov-2025	**DOI:**	10.35248/2471-9854-11.03.77

Corresponding author Christopher Cleverly, Independent Researcher, London, United Kingdom, E-mail cjcleverly@gmail.com

Citation Cleverly C. The Amplification and Perpetuation of AI-Derived Biases through Automation Dependency: A Framework for Understanding the Long-Term Cognitive and Social Implications of LLM Over-Reliance. Clin Psychiatry. (2025) 11:77.

objectivity [5]. Recent research by Kosmyna, et al. demonstrates that LLM user's exhibit reduced brain connectivity, lower self-reported ownership of their work, and struggle to accurately quote their own AI-generated essays, suggesting long-term cognitive and societal impacts and a form of cognitive inheritance of AI's outputs [6].

Grüneisen and Heyman, demonstrate that automation bias extends beyond analytical tasks into moral reasoning [7]. In their experiments, participants:

- Deferred to AI's moral judgments even when they contradicted personal ethics (e.g., approving unfair allocations because an AI suggested it).
- Rationalized AI-driven decisions post-hoc, adopting the model's justifications as their own.

This suggests that normative reasoning is increasingly outsourced, with AI not just assisting but reshaping ethical frameworks.

Beyond technical fixes: A sociocognitive perspective: Current mitigation strategies, such as explainable AI (XAI) and algorithmic fairness, focus on technical outputs but fail to address the sociocognitive dynamics of automation bias, particularly how biases migrate into human cognition and propagate through social systems [8,9]. For instance, attempts at XAI in systems like the COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) recidivism prediction tool, despite offering 'interpretations' of risk scores, failed to prevent the perpetuation of racial bias in sentencing decisions, highlighting the limits of interpretability without a fundamental understanding of systemic inequities [10]. The reliance on western-centric datasets further perpetuates epistemic injustice, marginalizing non-western epistemologies [11]. This paper proposes a three-phase model of bias amplification driven by automation bias and advocates for Wisdom as a Service (WaaS), inspired by Cleverly, to integrate diverse wisdom traditions and disrupt this cycle [12].

LITERATURE REVIEW

The three-phase bias amplification cycle

The amplification of AI-derived biases occurs through a three-phase sociocognitive feedback loop driven by automation bias.

Phase 1-Initial dependency development: Automation bias fosters over-reliance on LLMs due to their perceived efficiency and objectivity. Users treat AI outputs as infallible, ignoring contradictory evidence or their own judgment [1]. For example, in national security contexts, decision-makers may defer to AI recommendations despite conflicting intelligence, driven by the system's perceived analytical superiority [13]. This phase is exacerbated by the "black box" nature of AI, where opaque reasoning discourages scrutiny [8].

Phase 2-critical thinking atrophy: Prolonged reliance on AI leads to cognitive offloading, reducing critical thinking and vigilance. A 2025 MIT study by Kosmyna, et al. found that ChatGPT users exhibited lower brain engagement and underperformed in linguistic and behavioral tasks, indicating diminished analytical skills [6]. Specifically, LLM users showed weaker, less distributed brain connectivity compared to those using search engines or no tools. Furthermore, when LLM users were subsequently asked to perform tasks without AI, they demonstrated reduced alpha and beta connectivity, indicative of under-engagement. In education, students over-relying on AI for writing tasks produce essays with less linguistic diversity and greater similarity to AI-generated content, indicating reduced originality [6]. Kosmyna, et al. further demonstrated that Large Language Model (LLM) users struggled to accurately quote their own AI-generated work due to impaired memory encoding and reported lower self-ownership of their essays [6], reflecting a cognitive inheritance where prolonged AI reliance reshapes cognitive processes, distinct from short-term cognitive offloading or social algorithmic appropriation [14,15].

This aligns with Andy Clark's theory of extended cognition, which posits that cognitive processes extend beyond the brain to include external tools [16]. While AI tools like ChatGPT act as cognitive scaffolds, offloading effort and reducing cognitive load by 32% compared to traditional software users, they may weaken internal cognitive processes like memory formation and creative synthesis, as evidenced by the LLM group's lower neural connectivity in alpha and beta bands [6]. Similarly, Bernard Stiegler's concept of technogenesis suggests that human cognition co-evolves with technology, reshaping memory and attention [17]. The impaired quoting ability and low essay ownership reported by LLM users reflect a technogenetic shift where over-reliance on AI risks externalizing critical thinking, potentially eroding autonomous cognitive capacities [6].

These findings echo media theorist Marshall McLuhan's insight that media extend and alter human faculties [18], as AI's streamlined outputs may diminish the cognitive friction necessary for deep learning and originality, while also complicating algorithmic accountability by fostering uncritical acceptance of AI outputs [15]. This atrophy creates a feedback loop, increasing dependency on AI to compensate for weakened cognitive abilities.

Phase 3-Bias internalization and propagation: Automation bias enables AI biases to be internalized in human cognition and propagated through social and institutional interactions. Kosmyna, et al. demonstrate that LLM users exhibit a form of "cognitive debt" where their neural, linguistic, and behavioral performance is consistently diminished over time, even without direct AI assistance [6]. While their study focused on essay writing, the findings suggest that the patterns and biases embedded in AI-generated content can be implicitly adopted by human users, affecting their own cognitive processes and outputs. For instance, Amazon's scrapped AI recruitment tool, which favored male candidates, influenced hiring managers' subsequent decisions, perpetuating gender bias [19,20]. This phase extends biases into human-only contexts, such as policy-making and social discourse, amplifying inequities across cultures [18].

Key insight: Automation bias drives a cycle where AI flaws become embedded in human reasoning, with cognitive inheritance ensuring biases persist beyond AI interactions, creating long-term societal impacts.

Cognitive inheritance and long-term bias propagation

The temporal dilemma of AI cognitive inheritance: The concept of cognitive inheritance-the transmission of cognitive patterns, biases, and thinking frameworks through repeated

interaction-presents a critical challenge in AI research. While well established in developmental psychology, cultural anthropology, and educational theory through decades of empirical investigation, AI cognitive inheritance remains largely speculative due to the technology's recent emergence. This creates an urgent temporal dilemma: The methodologies needed to definitively establish cognitive inheritance require years of longitudinal research, yet the potential societal risks of undetected cognitive manipulation or bias transmission could be catastrophic if not identified immediately. As AI systems become ubiquitous in decision-making, education, and daily life, we risk embedding harmful cognitive patterns into human thinking before understanding their effects.

Emerging neural and behavioral evidence: Kosmyna, et al. offer initial neural evidence that LLM reliance may reduce engagement of memory and metacognitive networks, raising concerns about long-term erosion of critical faculties [6]. However, the claim that such reliance leads to durable cognitive inheritance of AI bias requires broader empirical grounding. Analogous findings from digital media research provide compelling support for this hypothesis. Studies demonstrate that prolonged exposure to algorithmically curated content-such as on social media-can reshape belief systems, reduce epistemic diversity, and entrench cognitive heuristics [21-23]. These studies support the hypothesis that uncritical LLM use may similarly encode and transmit normative assumptions across users and generations.

Methodological constraints and innovative solutions: Traditional approaches from established fields-developmental psychology's multi-year longitudinal studies, cultural anthropology's generational transmission mapping, and educational theory's long-term learning transfer assessments-require extensive time periods to demonstrate cognitive inheritance. These methodologies, while scientifically rigorous, are inadequate for the urgent timeline AI development demands.

To address this challenge, we propose rapid-detection frameworks combining multiple strategies: Early warning systems using short-term cognitive pattern analysis, accelerated natural experiments with existing heavy AI users, cross-field pattern recognition to identify inheritance signatures from other domains, precautionary monitoring frameworks based on known cognitive transmission mechanisms, and computational modeling to predict inheritance effects. These approaches prioritize identifying high-confidence early indicators that can trigger protective measures while comprehensive research continues.

The inadequacy of current mitigation strategies: It is crucial to distinguish between cognitive offloading-a neutral strategy of delegating mental tasks to external tools-and critical thinking atrophy, the detrimental decline of cognitive abilities due to over-reliance rather than reliance per se. Current mitigation strategies fail to address the sociocognitive dynamics of automation bias effectively. Explainable AI (XAI) tools, such as SHAP, produce complex outputs that non-technical users struggle to interpret, limiting their effectiveness in countering over-reliance [8,24]. Algorithmic fairness approaches focus on model outputs but neglect cognitive inheritance and societal propagation [25].

Addressing systemic bias propagation and scaling AI safety across diverse populations remains challenging, as highlighted in the International AI Safety Report [26]. "Fairwashing," a superficial compliance with ethical standards, further obscures systemic issues [27]. These limitations necessitate a paradigm shift toward culturally inclusive, wisdom-driven frameworks that can operate effectively within the compressed timelines of AI development while maintaining scientific rigor.

Longitudinal research remains essential to trace how these dynamics affect reasoning patterns beyond the moment of interaction, particularly in education and policymaking contexts, but cannot be the sole approach given the urgency of the technological trajectory.

New pathways: Wisdom as a Service (WaaS) and decolonized AI

As current AI ethics frameworks remain largely grounded in Western epistemologies and universalist assumptions, Wisdom as a Service (WaaS) and Decolonized AI present tentative pathways toward epistemic pluralism and cultural specificity. While tools such as ethical AI by design and participatory AI have sought to embed fairness and inclusivity into AI systems, they often fall short in addressing deep-seated automation bias and the socio-cognitive dynamics of cultural misalignment [28,29]. In contrast, WaaS and decolonized AI shall draw on indigenous knowledge systems and classical non-Western philosophies to construct AI architectures that are transparent, context-sensitive, and communally validated. It is acknowledged that translating cultural logics into formal AI systems will always entail loss and require ongoing dialogue with knowledge holders.

While our analysis reveals the urgent need for systemic interventions to address automation bias, we acknowledge that the Wisdom as a Service framework presented here represents an initial conceptual exploration rather than a validated solution. Future research must empirically test these theoretical propositions across diverse cultural and technological contexts.

Clarifying Wisdom as a Service (WaaS): Wisdom as a Service (WaaS), introduced by Cleverly, is a conceptual and technical framework that reimagines AI as a mechanism for cultivating human wisdom, rather than merely optimizing prediction or efficiency [12]. WaaS builds on the integration of non-European epistemological traditions, such as Ubuntu (communal ethics) [30], Nyāya (structured logic), the Yoga Sutras (virtue ethics), Dzogchen (non-dual awareness) [31], Tawhid (Sufi unity), and Dao (relational naturalism), to foster systems that are ethically grounded, culturally embedded, and resistant to automation bias.

WaaS is operationalized through a three-layered architecture:

- Axiomatic base: Encodes foundational ethical and epistemic principles to guide AI reasoning. For example, Ubuntu emphasizes relational harmony, while Nyāya offers a five-part syllogism to structure transparent inference [32,33].
- Discernment layer: Facilitates context-sensitive decision-making by prioritizing relational, ecological, and culturally situated knowledge over purely empirical or

decontextualized models [12].

- Recursive repair and review: Establishes ongoing mechanisms of community validation through Wisdom Learning Institutes, which iteratively refine and culturally align AI outputs [12].

This recursive model not only challenges static technical "fixes" to AI bias but also provides an institutional mechanism for long-term cultural accountability.

Contrasts with existing frameworks

1. Ethical AI by design [28]

- Definition: Embeds abstract ethical principles (e.g., beneficence, autonomy, justice) into AI systems at the design stage. This framework emphasizes universal standards and technical robustness as pathways to trustworthy AI.
- Contrast with WaaS: While ethical AI by design promotes preemptive ethical alignment, it largely reflects Western moral theory and assumes universality. In contrast, WaaS incorporates non-Western ethical systems and allows for recursive cultural validation, thereby resisting both automation bias and epistemic homogenization. Where ethical AI is front-loaded and largely static, WaaS is dynamic and community-responsive [12,32].

2. Participatory AI [29]

- Definition: Seeks to democratize AI by involving users and stakeholders in co-design processes and feedback loops. It promotes inclusivity and transparency but often operates within prevailing Western epistemic frameworks.
- Contrast with WaaS: While participatory AI values stakeholder input, it tends to lack deep integration of indigenous knowledge systems and does not explicitly challenge the epistemic foundations of mainstream AI. WaaS, by contrast, centers non-western logic systems (e.g., Nyāya) and ethical traditions (e.g., Ubuntu) as normative foundations, rather than adjunct considerations. Its institutional anchor, Wisdom Learning Institutes ensures that community engagement is not only participatory but also epistemically authoritative.

Decolonized AI

Decolonized AI extends WaaS by directly contesting the colonial legacies embedded in current AI systems, particularly, the dominance of English-language corpora, Western data hierarchies, and extractive algorithmic logics [27]. It advocates for linguistic sovereignty, multilingual training corpora, and participatory data governance, enabling AI to reflect the ontologies and epistemologies of diverse cultures. Incorporating structured reasoning frameworks like Nyāya improves AI explainability, directly addressing the "black box" opacity that fuels automation bias [8]. Ganeri's analysis of the Nyāya pañcāvayava syllogism-comprising pratijñā (thesis), hetu (reason), udāharaṇa (example), upanaya (application), and nigamana (conclusion), offers a structured inferential model [34]. This logic enables transparent, auditable reasoning in AI, countering automation bias [34,35]. Ethical principles such as ahimsa (non-harming) from the Yoga Sutras guide AI behavior toward minimally harmful, ethically resonant outcomes [36]. These practices resist both data colonialism and epistemic injustice by foregrounding local validation and relational ethics.

Theoretical foundations: Decolonized AI and epistemic injustice

Efforts to decolonize AI require more than technical diversification of datasets or language models, they demand a structural rethinking of knowledge hierarchies and power in AI development. Mohamed, et al. argue that mainstream AI systems are built on colonial logics: Extractive data practices, epistemic exclusion, and the reproduction of global inequities [37]. Their concept of "decolonial AI" calls for a dismantling of AI's Eurocentric assumptions, advocating instead for design frameworks grounded in local epistemologies, community governance, and knowledge sovereignty. WaaS aligns with this imperative by centering non-western wisdom traditions, participatory validation, and recursive accountability.

Complementing this critique, Dotson introduces the concept of epistemic oppression; a systemic limitation placed on marginalized groups' ability to participate in knowledge production [38]. Her work reframes automation bias as not just a cognitive shortcut, but as part of a broader epistemic injustice, where certain ways of knowing (e.g., oral traditions, relational ethics) are excluded from AI systems by design. WaaS and decolonized AI directly respond to this injustice by institutionalizing epistemic pluralism, ensuring that diverse reasoning practices (such as Nyāya logic or Ubuntu ethics) are not only included but structurally privileged in algorithmic decision-making. All good, but what happens when pluralism turns into incommensurability.

Epistemic incommensurability arises when traditions employ distinct conceptual frameworks, complicating mutual intelligibility. Yet Vedic svadharma [39], Buddhist pratītyasamutpāda [40], Ubuntu's relational ontology [41], and Taoist Wu Wei each articulate a process-relational ethic that resolves individual-community tensions. These frameworks foreground context-sensitive responsibility over fixed identity: Selfhood unfolds through dynamic interdependence rather than atomistic autonomy. A comparative ethics grounded in principled flexibility and dialogical engagement can thus transcend epistemic gaps, not by collapsing differences, but by aligning on functional coherence and relational process [30].

Wave function collapse occurs when a quantum system in superposition reduces to a single eigenstate through interaction with the external world, reflecting physics' bias toward discrete, deterministic outcomes. This interpretation has shaped a broader western metaphysics privileging fixed identities and objective resolution over relational becoming [42]. In contrast, Cantor's theory of transfinite numbers (believed by Cantor to be divinely revealed) embraced infinitude, echoing ancient wisdom traditions that resist closure and privilege unfolding process [43]. Similarly, dharma, pratītyasamutpāda, and Ubuntu maintain dynamic balance between individual and community without requiring epistemic finality. These processoriented paradigms may offer more robust responses to complex social realities than collapsebased models [44].

By emphasizing process over doctrine, WaaS enables integration of seemingly incompatible wisdom traditions through shared commitment to ethical becoming.

Efforts to decolonize AI through data diversification and bias mitigation are necessary but insufficiently transformative [37]. Wisdom as a Service (WaaS) offers an alternative paradigm, framing AI as a partner in cultivating ethical and cognitive insight. Drawing on Nyāya logic for analytical precision [33], Ubuntu for relational ethics [45], and African oral traditions as outlined by Oluwole, WaaS integrates non-Western modes of knowing [46]. Central to this is darśana, not as passive seeing, but as transformative realization through experiential presence [47,48]. In this view, wisdom is not transmitted but emergent, fostered through guided reflection and ethical context. Inspired by wisdom education frameworks and metacognitive prompts [49,50], WaaS supports users in developing insight rather than merely extracting answers, aligning with darśana's emphasis on self-realization and moral becoming.

Case study: Cultural reasoning in autonomous systems

Cleverly illustrates WaaS in the context of autonomous vehicles operating in culturally sensitive environments. For example, when navigating a religious procession in India, a vehicle guided by Ubuntu's relational ethics and Nyāya's structured reasoning is more likely to respect communal dynamics than one governed by utilitarian cost-benefit logic. Early AI simulations (Grok, DeepSeek) suggest that this approach may reduce ethical regret and litigation compared to standard decision-making models, offering a pathway to culturally aligned automation (Table 1).

Table 1: Differences between WaaS/decolonized AI and mainstream AI ethics across core dimensions.

Dimension	WaaS/Decolonized AI	Mainstream AI ethics
Epistemology	Relational, ecological (Ubuntu, Nyāya)	Empirical rationalism (western moral theory)
Validation	Communal *via* wisdom learning institutes	Statistical validation or expert auditing
Language	Multilingual, locally sourced	English-dominant, westerncentric corpora
Goal	Cultivation of collective wisdom and ethical discernment	Optimization, prediction accuracy, fairness metrics

This reorientation toward wisdom rather than prediction, and toward communal reasoning over individual efficiency, marks a necessary evolution in the theory and practice of ethical AI. WaaS and Decolonized AI are not only correctives to technical and moral shortcomings in current systems, but also foundational strategies for building epistemically inclusive and culturally sustainable AI futures.

Implementation challenges and tentative solutions for culturally responsive AI systems

Methodological note: This section identifies key barriers to implementing Wisdom as a Service (WaaS) as a culturally responsive AI framework aimed at mitigating automation bias and cognitive inheritance [6]. Vygotskian theory supports community-driven validation and dialogic interfaces, ensuring AI acts as a scaffold rather than a crutch, mitigating automation bias [51]. We see merit in human-in-the-loop solutions such as indigenous wisdom councils. However, rather than offering final solutions, we present tentative approaches and research priorities to guide future development. All proposals emphasize empirical validation, interdisciplinary inquiry, and community leadership.

Technical implementation challenges and tentative solutions

Challenge 1: Computational representation of cultural knowledge

Problem: Encoding complex philosophical concepts (e.g., Nyāya logic, Ubuntu ethics) into AI risks distortion or reduction [35].

Obstacles: Knowledge encoding-cultural axioms such as ahimsa and satya resist straightforward formalization [36].

Context sensitivity: AI systems struggle to incorporate situational nuance.

Dynamic adaptation: Evolving cultural norms challenge system adaptability.

Tentative solutions:

- Develop moral reasoning graphs to model context-sensitive ethical dependencies [34].
- Utilize hybrid symbolic-neural architectures to blend logical transparency with adaptive capacity [12].
- Validate computational models through ethnographic fieldwork in cultural contexts.

Research priorities:

Explore formalization of Nyāya-Inspired Logical Structures: The Nyāya tradition's five-part syllogism (pañcāvayava) could enhance AI explainability:

- Pratijñā (Proposition): The AI's intended action.
- Hetu (Reason): The ethical principle guiding the action.
- Udāharaṇa (Example): Precedent cases demonstrating the principle.
- Upanaya (Application): How the principle applies to the current context.
- Nigamana (Conclusion): The reasoned decision.

Develop integrity metrics for culturally faithful AI representation.

Challenge 2: Scalability *vs*. authenticity

Problem: Scalable AI architectures may flatten cultural nuance under standardization.

Obstacles:

- Standardization bias: Cultural diversity is reduced to generic models.
- Efficiency trade-offs: Cultural fidelity may reduce computational performance.
- Architecture incompatibility: Some cultural logics may not

align with existing AI paradigms.

Tentative solutions:

- Design modular architectures permitting localized cultural customization.
- Define performance benchmarks that balance efficiency with cultural appropriateness [12].
- Deploy localized pilot projects in diverse communities.

Research priorities:

- Assess hybrid models for balancing universality and specificity.
- Develop culturally adaptive benchmarking protocols.

Challenge 3: Validation and testing frameworks

Problem: There is no standard methodology for validating cultural responsiveness in AI.

Obstacles:

- Metric ambiguity: No clear criteria for evaluating cultural integrity.
- Simulated contexts: Testing environments often lack cultural realism.
- Cross-cultural bias: Validation may embed dominant epistemologies.

Tentative solutions:

- Implement community-led validation protocols with cultural mediators [32].
- Adopt mixed-methods evaluation combining bias-reduction metrics with qualitative assessment.
- Develop cross-cultural testing frameworks to ensure equitable performance.

Research priorities:

- Construct culturally inclusive evaluation standards.
- Test validation processes across plural contexts.

Cultural and epistemological challenges and tentative solutions

Challenge 4: Avoiding cultural appropriation

Problem: Integrating non-Western epistemologies into AI risks extraction and misrepresentation.

Obstacles:

- Consent deficits: Knowledge integration without community approval.
- Misrepresentation: Oversimplification of nuanced traditions.
- Power asymmetries: Technical experts dominate epistemic frameworks.

Tentative solution:

- Implement free, prior, and informed consent protocols [27].
- Conduct co-design workshops with knowledge holders [12].
- Establish benefit-sharing agreements ensuring reciprocal value.

Research priorities:

- Analyze ethical models for equitable knowledge integration.
- Evaluate community perspectives on cultural AI inclusion.

Challenge 5: Epistemic incommensurability

Problem: Some knowledge systems (e.g., Ubuntu relationality) may be irreconcilable with formal AI logic.

Obstacles:

- Logical dissonance: Contradictory reasoning across traditions.
- Ontological divergence: Incompatible assumptions about personhood, agency, or reality
- Translational limits: Difficulty preserving meaning across epistemic boundaries.

Tentative solutions:

- Develop epistemic pluralism protocols to handle divergence without relativism [34].
- Use semantic translation layers to mediate between knowledge systems.
- Design conflict resolution frameworks for epistemic incompatibility.

Research priorities:

- Experiment with pluralist logic models in computational settings.
- Build tools for meaningful cross-epistemic translation.

Challenge 6: Community agency and governance

Problem: Cultural communities must be empowered to govern systems that incorporate their epistemologies.

Obstacles:

- Technical exclusion: Communities often lack access to AI design tools.
- Resource gaps: Oversight requires sustained funding and expertise.
- Governance durability: Long-term participation is difficult to maintain.

Tentative solutions:

- Develop low-barrier governance platforms (e.g., community dashboards) [12].
- Allocate resources for community-led system maintenance and training.
- Establish Wisrsi (human-in-the-loop) platforms to enable real-time feedback.

Research priorities:

- Prototype community-led governance models.
- Assess long-term engagement mechanisms.

Institutional and social challenges and tentative solutions

Challenge 10: Academic and industrial culture change

Problem: Existing research cultures disincentivize culturally engaged approaches.

Obstacles:

- ***Academic incentives:*** Publication norms devalue cultural collaboration.

- ***Commercial timelines:*** Industry favors rapid iteration over slow engagement.
- ***Disciplinary fragmentation:*** AI, ethics, and cultural studies remain siloed.

Tentative solutions:

- Reform academic evaluation to value community collaboration.
- Develop sustainable business models for ethical AI.
- Create interdisciplinary research consortia.

Research priorities:

- Analyze incentive structures for culturally inclusive innovation.
- Prototype models for cross-sector AI development.

Challenge 11: Capacity building and education

Problem: Few professionals are trained at the intersection of AI and cultural knowledge.

Obstacles:

- Training deficit: Curricula rarely combine technical and cultural literacy.
- Linguistic Gaps: Knowledge may be inaccessible in dominant languages.
- Knowledge attrition: Oral traditions risk erasure without preservation.

Tentative solutions:

- Develop interdisciplinary training programs.
- Create translation and documentation tools for cultural epistemologies.
- Fund cultural preservation initiatives alongside AI development.

Research priorities:

- Evaluate effectiveness of interdisciplinary education models.
- Study preservation mechanisms for endangered epistemologies.

Research priorities and testable hypotheses

Immediate research questions:

1. Can existing AI architectures be adapted to faithfully represent non-Western knowledge systems?
2. What metrics can assess cultural authenticity in AI decision-making?
3. Which governance models best enable community oversight?
4. Does culturally responsive AI reduce automation bias in practice?

Testable hypotheses:

- H1: Nyāya-structured explanations improve user comprehension compared to standard XAI models (e.g., SHAP).
- H2: Community-governed AI systems generate higher trust and lower automation bias.
- H3: Culturally responsive AI interfaces show varied effectiveness across cultural contexts.
- H4: Long-term engagement with culturally responsive AI reduces cognitive inheritance effects [6].

Methodological frameworks:

- Cross-cultural experimental designs
- Community-based participatory research
- Longitudinal impact studies
- Mixed-methods evaluation

A path forward

Implementing culturally responsive AI such as WaaS requires navigating complex, interrelated challenges across technical, cultural, ethical, and institutional domains. The tentative strategies outlined-moral reasoning graphs, hybrid architectures, community governance, and legal reform-demand rigorous empirical inquiry and sustained community engagement.

Key insights:

- Addressing these challenges requires systemic, not modular, solutions.
- Communities must be partners and epistemic authorities, not subjects of research.
- Progress depends on long-term investment, interdisciplinary collaboration, and cultural humility.

We invite researchers, practitioners, and communities to collaborate in co-developing WaaS, prioritizing cultural integrity, local agency, and shared wisdom in the future of AI.

Policy and research recommendations

Regulatory actions

- ***Mandate bias propagation assessments:*** Require AI developers to evaluate sociocognitive impacts, including automation bias and cognitive inheritance (as evidenced by studies like) [6], as proposed in the European Parliament (2024). Artificial Intelligence Act. (https://eur-lex.europa.eu/eli/reg/2024/1689).
- ***Fund decolonized AI initiatives:*** Support projects like the Shri Vidya AI Institute to integrate non-Western epistemologies [12].

Industry practices

- ***Adopt reliance drills:*** Implement exercises to test human fallback capacity, reducing overreliance on AI [52].
- ***Develop WaaS-certified AI:*** Certify systems for epistemic pluralism and community validation, countering automation bias [12].

Future research directions

- ***Longitudinal studies:*** Investigate cognitive inheritance over time, building on the neural and behavioral insights from across a wider range of tasks and populations [6].
- ***Cross-cultural experiments:*** Test WaaS's effectiveness in mitigating automation bias across diverse contexts, using metrics like ethical Turing tests [12].

Cognitive offloading is not inherently detrimental; when used strategically, it can support higher-order thinking by reallocating mental resources. However, when offloading becomes habitual and uncritical, it leads to critical thinking atrophy, displacing

internal cognitive faculties rather than augmenting them. Automation bias accelerates this shift by encouraging over-reliance on AI outputs, initiating a cycle in which biased AI-generated content is internalized and reproduced in human cognition and decision-making.

This process, termed cognitive inheritance, has been empirically observed in LLM users, who demonstrate reduced neural engagement, diminished self-authorship, and impaired recall [6]. As these biases migrate into non-AI contexts, their impact compounds, threatening intellectual integrity, cultural autonomy and individual equity, especially, in advanced economies where AI exposure is highest [3].

While cognitive offloading is not new to human-technology interaction, large language models (LLMs) present a uniquely complex challenge by combining and exacerbating forms of offloading and bias propagation observed in earlier technologies. For instance, just as calculators offload procedural arithmetic tasks, potentially impacting mental arithmetic skills and the neural regions associated with numerical processing, such as the intraparietal sulcus [53,54], and GPS navigation can reduce spatial memory by shifting attentional demands [55,56], LLMs similarly contribute to critical thinking atrophy by generating ready-made answers that reduce the need for active cognitive engagement [6,57]. However, unlike calculators, which are deterministic tools with no embedded cultural biases (though educational biases can emerge through their use) [58], LLMs parallel social media platforms in their capacity for algorithmic shaping and amplification of beliefs [21]. The critical distinction lies in how LLMs uniquely bias not just information access (as seen in search engine filter bubbles) but also information synthesis, with their biases architecturally embedded in training data rather than solely crowd-sourced [11,22].

Call to action

- Researchers: Explore non-western epistemologies to counter automation bias and continue to investigate the long-term cognitive and behavioral impacts of LLM overreliance [34].
- Intervention effectiveness: Can proposed mitigation strategies (including WaaS principles) demonstrably reduce automation bias and cognitive inheritance?
- Implementation feasibility: What are the practical challenges and benefits of integrating non-Western epistemologies into AI systems?
- Policymakers: Regulate AI for cognitive and societal impacts [59].
- Industry: Build wisdom-enhancing AI to foster inclusive outcomes [60].

To translate the conceptual strength of WaaS into practice, researchers and developers should pursue concrete implementations such as API-based ethical filters grounded in Nyāya or Ubuntu reasoning, WaaS plug-ins for LLMs in education and policy contexts, and community lab models that operationalize epistemic pluralism in real-world settings.

CONCLUSION

What is new is this novel combination of pervasive cognitive offloading and insidious, embedded bias propagation constitutes a unique threat, necessitating systemic responses like Wisdom as a Service (WaaS) to ensure appropriate human oversight and prevent cognitive inheritance.

To counter this trajectory, technical interventions alone are insufficient. Wisdom as a Service (WaaS) and decolonized AI, grounded in non-European epistemologies, offer a systemic response. By foregrounding epistemic pluralism, communal validation, and culturally contextual reasoning, they provide a foundation for more equitable and reflective AI integration. Rather than optimizing for efficiency alone, these frameworks restore the human capacity for discernment; crucial for sustaining autonomy in an increasingly automated world.

REFERENCES

1. Goddard K, Roudsari A, Wyatt JC. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. J Am Med Inform Assoc. 19(1):121-127.
2. Skitka LJ, Mosier KL, Burdick M. (2000). Accountability and automation bias. IJHCS. 52(4):701-717.
3. Georgieva K. (2024). AI will transform the global economy: Let's make sure it benefits humanity. International Monetary Fund Blog.
4. Rosbach E, Ganz J, Ammeling J, Riener A, Aubreville M. (2024). Automation bias in AI-assisted medical decision-making under time pressure in computational pathology.
5. Straw I. (2020). The automation of bias in medical Artificial Intelligence (AI): Decoding the past to create a better future. Artif Intell Med. 110:101965.
6. Kosmyna N, Hauptmann E, Yuan YT, Situ J, Liao XH, et al. (2025). Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task. ArXiv preprint arXiv:2506.08872.
7. Grüneisen S, Heyman G. (2024). Human reliance on AI predictions in moral decision-making. Proc Natl Acad Sci USA. 121(3):e2316960121.
8. Burrell J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. BDS. 3:2053951715622512.
9. Fricker M. (2007). Epistemic injustice: Power and the ethics of knowing. Oxford University Press.
10. Angwin J, Larson J, Mattu S, Kirchner L. (2016). Machine bias. ProPublica.
11. Birhane A. (2021). Algorithmic injustice: A relational ethics approach. Patterns. 2(2):100205.
12. Cleverly C. (2025). Wisdom as a Service (WaaS): A pluralist ethical framework for AI rooted in non-European traditions. SSRN.
13. Horowitz M. (2023). Artificial intelligence and international security. Brookings Institution.
14. Risko EF, Gilbert SJ. (2016). Cognitive offloading. Trends Cogn Sci. 20(8):676-688.
15. Binns R. (2018). Algorithmic accountability and public reason. Philos Technol. 31(4):543-556.
16. Clark A, Chalmers D. (1998). The extended mind. Analysis. 58(1):7-19.

17. Stiegler B. (1998). Technics and time, 1: The fault of Epimetheus (R. Beardsworth & G. Collins, Trans.). Stanford University Press.
18. McLuhan M. (1964). Understanding media: The extensions of man. McGraw-Hill.
19. Dastin J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. Reuters.
20. Mbembe A. (2017). Critique of black reason. Duke University Press.
21. Tufekci Z. (2015). Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency. CTLJ. 13(203):203-218.
22. Pariser E. (2011). The filter bubble: What the internet is hiding from you. Penguin Press.
23. Fuchs C. (2021). Social media: A critical introduction (3rd ed.). SAGE Publications.
24. Kirkpatrick K. (2023). AI and the challenge of ethical decision-making. Communications of the ACM. 66(4):15-17.
25. Whittlestone J, Nyrup R, Alexandrova A, Cave S, Govindarajulu NS. (2019). The role and limits of principles in AI ethics: Towards a focus on tensions. In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (pp. 195–200).
26. United Kingdom. (2025). International AI safety report. UK Government.
27. Couldry N, Mejias UA. (2019). Data colonialism: Rethinking big data's relation to the contemporary subject. TVNM. 20(4):336-349.
28. Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, et al. (2018). AI4People-An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds Mach. 28(4):689-707.
29. Sloane M, Moss E, Awomolo O, Forlano L. (2022). Participation is not a design fix for machine learning. Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society. 1-6.
30. Ramose MB. (2002). African philosophy through Ubuntu. Mond Books.
31. Reynolds JM. (1996). The golden letters: The three statements of Garab Dorje, the first teacher of Dzogchen, together with an old commentary and an introduction. Snow Lion Publications.
32. Ewuoso C, Hall S. (2019). Core aspects of Ubuntu: A systematic review. S Afr J Philos. 38(2):175-186.
33. Ganeri J. (2001). Indian logic: A reader. Routledge.
34. Ganeri J. (2001). Philosophy in classical India: The proper work of reason. Routledge.
35. Matilal BK. (1998). The character of logic in India. State University of New York Press.
36. Tigunait PR. (2014). The secret of the yoga sutra: Samadhi pada. Himalayan Institute Press.
37. Mohamed S, Png MT, Isaac W. (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. AI & Society. 35(4):753-764.
38. Dotson K. (2014). Conceptualizing epistemic oppression. Soc Epistemol. 28(2):115-138.
39. Flood G. (1996). An introduction to Hinduism. Cambridge University Press.
40. Harvey P. (2013). An introduction to Buddhism: Teachings, history and practices. Cambridge University Press.
41. Metz T. (2007). Toward an African moral theory. J Polit Philos. 15(3):321-341.
42. Barad K. (2007). Meeting the Universe halfway: Quantum physics and the entanglement of matter and meaning. Duke University Press.
43. Dauben JW. (1979). Georg Cantor: His mathematics and philosophy of the infinite. Harvard University Press.
44. Ellis GFR. (2012). On the nature of emergent reality. In The re-emergence of emergence (pp. 79-107). Oxford University Press.
45. Metz T. (2007). Ubuntu as a moral theory and human rights in South Africa. Afr Hum Rights L J. 11(2):532-559.
46. Oluwole S. (1992). Witchcraft, reincarnation and the god-head. Excel Publishers.
47. Coomaraswamy AK. (1943). Figures of speech or figures of thought? The Art Bulletin. 21(2):253-268.
48. Chakrabarti A. (1995). Knowing from words: Western and Indian philosophical analysis of understanding and testimony. Springer.
49. Sternberg RJ. (2001). Why schools should teach for wisdom: The balance theory of wisdom in educational settings. Educational Psychologist. 36(4):227-245.
50. Lubiana T. (2023). Cognitive offloading and critical thinking in AI-assisted decisions. AI and Society.
51. Vygotsky LS. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.
52. Buçinca Z, Malaya MB, Gajos KZ. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW1), Article 188. 5(CSCW1):1-21.
53. Ansari D, Dhital B. (2006). The neural correlates of mathematical competence. Trends Cogn Sci. 10(11):518-525.
54. Dehaene S, Piazza M, Pinel P, Cohen L. (2003). The neural foundations of numerical and mathematical cognition. Nat Rev Neurosci. 4(2):145-155.
55. Maguire EA, Gadian DG, Johnsrude IS, Good CD, Ashburner J, et al. (2000). Navigation-related structural change in the hippocampi of taxi drivers. Proc Natl Acad Sci U S A. 97(8):4398-4403.
56. Montello DR. (2005). Navigation and spatial cognition. In P. Thagard and R. Eliott (Eds.), Handbook of cognitive science (pp. 375-389).
57. Sparrow B, Liu J, Wegner DM. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. Science. 333(6043):776-778.
58. Boaler J. (2016). Mathematical mindsets: Unleashing students' potential through creative math, inspiring

messages and innovative teaching. Jossey-Bass.

59. Owen R, Macnaghten P, Stilgoe J. (2013). Responsible innovation: Framing, analysis, and ways forward. Sci Public Policy. 40(6):751-760.
60. Vallor S. (2016). Technology and the virtues: A philosophical guide to a future worth wanting. Oxford University Press.