Journal of Philosophical Investigations
Print ISSN: 2251-7960 Online ISSN: 2423-4419
Homepage: https://philosophy.tabrizu.ac.ir

Neuroplasticity in Human-AI Integration: A Contribution to Symbiotic Intelligence Theory

Michael Peters

Visting Distinguised Professor, School of Education, Tsinghua University, Beijing, China. Department of Education Policy, Organization and Leadership, University of Illinois at Urbana-Champaign, Urbana, IL, USA. Philosophy Program, University of Waikato, Hamilton, New Zealand. E-mail: mpeters@waikato.ac.nz

Article Info

Article type:
Research Article

Article history:

Received 19 August 2026
Received in revised form 30 August 2026
Accepted 13 September 2026
Published online 25 October 2026

Keywords:

Neuroplasticity, Human-AI Integration, Symbiotic Intelligence, Brain-Computer Interfaces, Extended Cognition, Hybrid Intelligence, Cognitive Enhancement, Postdigital Philosophy, Reciprocal Plasticity

ABSTRACT

The accelerating convergence of human cognition and artificial intelligence demands new conceptual frameworks that move beyond augmentation toward genuine integration. This paper positions neuroplasticity—the brain's lifelong capacity to reorganize neural structure and function through experience—as the central biological mechanism enabling deep human-AI symbiosis. The analysis argues that sustained, reciprocal interaction with advanced AI systems can induce targeted neuroplastic changes that allow human cognition to extend into artificial substrates, while AI architectures adapt in response to human neural patterns. Drawing on cognitive neuroscience, brain-computer interface research, extended mind theory, and postdigital philosophy, the paper develops a contribution to symbiotic intelligence theory. In this framework, symbiotic intelligence emerges as a hybrid, co-evolutionary system in which human and artificial intelligences mutually reshape one another through reciprocal plasticity, producing cognitive capabilities that exceed what either system can achieve alone. The analysis examines how intentional design of AI interfaces, feedback loops, and educational practices can harness neuroplasticity to support ethical and effective integration, while addressing risks of cognitive dependency, identity transformation, and epistemic inequity. The paper concludes by outlining implications for pedagogy, cognitive enhancement, and the governance of hybrid intelligence in the emerging era of artificial superintelligence.

Cite this article: Peters, M. (2026). Neuroplasticity in Human-AI Integration A Contribution to Symbiotic Intelligence Theory. *Journal of Philosophical Investigations*, 20(56), 57-74. https://doi.org/10.22034/jpiut.2026.74244.4683

© The Author(s). Publisher: University of Tabriz.

Intruduction

The relationship between human cognition and artificial intelligence has entered a phase of transformation that existing conceptual frameworks struggle to capture. The dominant metaphors of the past three decades—augmentation, assistance, tool-use—presume a stable boundary between the human cognitive agent and the artificial system that serves it. The human thinks; the machine executes. The human decides; the machine recommends. These metaphors were adequate when AI systems were narrow, passive, and transparent in their operation. They are increasingly inadequate in a technological landscape defined by large language models that generate novel content, brain-computer interfaces that establish direct neural communication with digital systems, and autonomous agents that initiate actions without explicit human instruction. The relationship is no longer one of augmentation—the addition of external capacity to an intact cognitive core—but of integration: a bidirectional, mutually transformative process in which both human neural circuits and artificial computational architectures are reshaped through sustained reciprocal interaction.

This article advances the claim that neuroplasticity—the brain's lifelong capacity to reorganize its structure and function in response to experience (Kolb & Whishaw, 1998; Pascual-Leone et al., 2005)—constitutes the central biological mechanism that makes deep human-AI integration possible. The argument is not merely that the brain adapts to technology use, a claim well established in the cognitive offloading literature (Risko & Gilbert, 2016; Firth et al., 2019). It is that sustained, reciprocal interaction with advanced AI systems can induce targeted neuroplastic changes that enable human cognition to extend into artificial computational substrates, while AI systems simultaneously adapt their architectures in response to human neural patterns. This bidirectional process—reciprocal plasticity—is the defining dynamic of human-AI symbiosis.

The analysis develops a contribution to symbiotic intelligence theory, a framework that conceptualizes human-AI integration not as the subordination of one intelligence to the other but as the emergence of a hybrid, co-evolutionary cognitive system. In this framework, symbiotic intelligence is defined by four properties: it arises through reciprocal plasticity across biological and artificial substrates; it produces cognitive capabilities exceeding those achievable by either system independently; it depends on sustained patterns of interaction rather than static interfaces; and it raises distinctive ethical challenges that differ qualitatively from those posed by either human or artificial intelligence considered in isolation.

The paper proceeds through six further sections. Section 2 establishes neuroplasticity as the biological foundation of human-AI integration, reviewing evidence from structural and functional neuroimaging, cognitive training, and brain-computer interface research. Section 3 situates this evidence within three theoretical traditions—extended mind theory, brain-computer interface research, and postdigital philosophy—that provide complementary

perspectives on cognitive extension into artificial substrates. Section 4 develops the core theoretical contribution: symbiotic intelligence theory as a framework for understanding reciprocal plasticity and co-evolutionary dynamics in human-AI systems. Section 5 examines the design dimensions through which neuroplasticity can be deliberately harnessed—interface design, feedback architecture, and educational practice—to support ethical and effective integration. Section 6 addresses the principal risks: cognitive dependency, identity transformation, and epistemic inequity. Section 7 synthesizes the analysis into implications for pedagogy, cognitive enhancement policy, and the governance of hybrid intelligence.

1. Neuroplasticity as the Biological Foundation of Human-AI Integration

The claim that human cognition can integrate with artificial computational systems requires a biological mechanism—a process through which the structural and functional organization of the brain can be modified by sustained interaction with an external system. Neuroplasticity provides precisely this mechanism. The term denotes the brain's capacity to reorganize synaptic connections, modify regional grey matter density, and alter functional connectivity patterns in response to experience across the entire lifespan (Kolb & Whishaw, 1998; Draganski et al., 2004). Far from being a fixed computational architecture that merely processes inputs, the brain is a dynamic biological system whose very structure is sculpted by the patterns of activity to which it is exposed.

1.1 Structural and Functional Plasticity: The Evidence Bas

The foundational demonstrations of structural neuroplasticity established that intensive cognitive training produces measurable changes in brain anatomy. Draganski et al. (2004) showed that three months of juggling training increased grey matter density in the mid-temporal area and posterior intraparietal sulcus—regions associated with visual motion processing and visuomotor coordination. Maguire et al. (2006) demonstrated that London taxi drivers, who undergo years of intensive spatial navigation training, exhibit enlarged posterior hippocampi relative to controls, with hippocampal volume correlating positively with time spent navigating. These and numerous subsequent studies (reviewed in Pascual-Leone et al., 2005) established a principle of fundamental importance for human-AI integration: the brain does not merely learn to use a tool; it reorganizes its architecture to accommodate the cognitive demands the tool imposes.

Functional plasticity—the reorganization of activation and connectivity patterns without necessarily producing gross structural change—operates on shorter timescales and is equally relevant. Repeated engagement of specific neural circuits strengthens their synaptic connections through long-term potentiation, a Hebbian mechanism in which coordinated pre- and postsynaptic firing increases synaptic efficacy. The functional connectivity literature has documented that even brief periods of cognitive training alter resting-state connectivity within the default mode, frontoparietal, and salience networks (Pascual-Leone et al., 2005). These

findings indicate that the brain's functional architecture is continuously remodelled by the patterns of cognitive engagement to which it is exposed—including, critically, patterns of engagement with external information processing systems.

1.2 Plausible Pathways for AI-Induced Neuroplastic Change

The translation of established plasticity mechanisms to the specific context of human-AI interaction requires the identification of plausible causal pathways. Three such pathways can be distinguished, operating on different timescales and through different neural substrates.

The first pathway operates through the cognitive demands imposed by active AI collaboration. When a user engages an AI system in tasks requiring critical evaluation, hypothesis generation, and integrative synthesis—as opposed to passive acceptance of AI-generated outputs—the prefrontal executive control networks are recruited in patterns that structurally resemble those observed in other forms of cognitive training (Gilbert, 2024). The repeated co-activation of prefrontal, parietal, and hippocampal regions during active AI problem-solving strengthens their functional connectivity and may, over extended periods, increase grey matter density in task-relevant cortical areas through the same mechanisms documented in the broader cognitive training literature.

The second pathway operates through brain-computer interface (BCI) technology, which establishes direct communication channels between neural activity and external computational systems. BCI research has demonstrated that users can learn to modulate specific neural signals—sensorimotor rhythms, slow cortical potentials, event-related potentials—to control external devices, and that this learned modulation produces plastic changes in the circuits involved (Pitt et al., 2022). The neurofeedback component of BCI operation exploits the same Hebbian mechanisms: repeated association between a neural pattern and a desired outcome strengthens the neural circuits that produce that pattern. As BCI technology advances from assistive applications toward bidirectional cognitive interfaces, the scope of AI-induced neuroplastic change will expand correspondingly (Tuskey, 2025).

The third pathway operates through the environmental and behavioural changes that AI systems produce. AI-mediated information environments—algorithmically curated content feeds, personalized notifications, conversational agents—structure the patterns of cognitive engagement to which users are exposed over extended periods. These environmental patterns shape neural function through the same plasticity mechanisms by which any sustained environmental input shapes the brain: repeated exposure to particular temporal structures, emotional valences, and informational patterns strengthens the neural circuits that process those patterns (Davidson & McEwen, 2012; Firth et al., 2019). This pathway is less direct than the first two but arguably more pervasive, as it operates continuously through the ambient digital environment rather than only during explicit AI interaction.

None of these pathways in isolation is sufficient to establish the feasibility of deep human-AI integration. Taken together, however, they indicate that the brain possesses multiple, converging mechanisms through which sustained interaction with AI systems can produce structural and functional reorganization—the necessary biological precondition for the emergence of symbiotic intelligence.

The biological evidence for AI-induced neuroplastic change, while suggestive, requires theoretical interpretation. Three intellectual traditions—extended mind theory, BCI research, and postdigital philosophy—provide complementary frameworks for understanding what it means for human cognition to extend into artificial substrates, and why the resulting integration demands conceptual tools beyond those provided by neuroscience alone.

2. Theoretical Foundations: Extended Mind, BCI Research, and Postdigital Philosoph

The thesis that human cognition can integrate with artificial computational systems through neuroplastic mechanisms requires a theoretical account of what cognitive integration consists in—what it means, philosophically and computationally, for a human neural system and an artificial information processing system to become functionally coupled. This section examines three theoretical traditions that bear on this question, identifying their contributions and their limitations for the specific case of human-AI symbiosis.

2.1 Extended Mind Theory and the Parity of Cognitive Vehicles

Extended mind theory, as formulated by Clark and Chalmers (1998), provides the most influential philosophical framework for understanding how cognitive processes can extend beyond the biological boundaries of the individual organism. The central claim is that external objects can become constituents of cognitive processes when they are reliably coupled with internal neural operations in a manner that satisfies the parity principle: if an external process performs a function that, were it performed internally, would be uncontroversially recognized as cognitive, then that external process is part of the cognitive system. The theory thus denies that cognition is coextensive with the biological brain, locating it instead in the functional organization of the coupled system.

Extended mind theory contributes two insights of direct relevance to human-AI integration. First, it establishes that the boundary between internal and external cognitive processing is contingent rather than necessary—a matter of functional organization rather than biological essence. Second, and more specifically, it identifies the conditions under which an external resource crosses the threshold from tool to cognitive constituent: reliable availability, automatic deployment, and default endorsement of its outputs (Clark, 2008). These conditions map with increasing precision onto the relationship between contemporary users and their AI systems. An AI assistant that is constantly accessible, consulted without deliberation, and trusted by default satisfies the functional criteria for cognitive extension.

The limitations of extended mind theory for the present analysis are equally apparent. The theory was developed primarily with reference to inert external storage media—notebooks, calculators, navigation charts—whose functional profile maps cleanly onto that of biological memory. AI systems that generate novel content, learn from interaction history, and adapt their behaviour in real time exceed this functional profile in qualitatively significant respects. They are not merely external storage but external processing—active computational agents that transform information rather than simply preserving it for later retrieval (Turner, 2022; van der Weij & Erden, 2025). The extension of cognition into AI systems is thus not the addition of inert storage capacity but the coupling of two adaptive information processing systems, each capable of reorganizing its own structure and function in response to the other. This coupling requires theoretical resources that extended mind theory, in its original formulation, does not provide.

2.2 Brain-Computer Interface Research and Neural Coupling

Brain-computer interface research provides the most direct empirical window into the mechanisms by which human neural activity and artificial computational systems can become functionally integrated. BCI technology establishes direct communication channels between neural signals and external devices, bypassing the peripheral motor system and enabling the brain to interact with artificial systems at the level of neural activity itself (Pitt et al., 2022). The research is clinically motivated—BCI systems have restored communication and motor function for individuals with severe paralysis—but its theoretical implications extend well beyond assistive technology.

The theoretical significance of BCI research for human-AI integration lies in its demonstration that direct neural coupling with artificial systems is not merely conceptually possible but empirically achievable, and that such coupling produces measurable neuroplastic changes in the circuits involved. When a BCI user learns to modulate sensorimotor rhythms to control a cursor, the neural circuits that generate those rhythms reorganize through Hebbian mechanisms to improve signal reliability (Pitt et al., 2022). The BCI system, for its part, adapts its decoding algorithms to the user's evolving neural patterns through machine learning techniques. This bidirectional adaptation—neural plasticity on the human side, algorithmic adaptation on the machine side—is precisely the pattern that symbiotic intelligence theory generalizes beyond the BCI context to the broader landscape of human-AI interaction.

BCI research also illuminates the risks of deep neural integration. Tuskey (2025) has documented the psychological consequences of BCI-mediated cognitive enhancement, including experiences of agency confusion—uncertainty about whether a cognitive output originates from one's own neural processes or from the AI system—and identity destabilization resulting from the blurring of the boundary between self-generated and system-generated cognition. These findings temper technological optimism with sobering evidence that cognitive

integration, while biologically feasible, carries psychological costs that must be addressed in any adequate theoretical and regulatory framework.

2.3 Postdigital Philosophy and the Dissolution of Boundaries

Postdigital philosophy contributes a distinct analytical register to the theorization of human-AI integration, one that operates at the level of ontology rather than mechanism. The postdigital perspective holds that the distinction between the digital and the non-digital has become incoherent—not because everything has become digital but because digital technologies have become so thoroughly entangled with human practices, institutions, and subjectivities that efforts to separate them produce artificial distinctions rather than analytical clarity (Peters et al., 2022). In a postdigital condition, the question is not whether technology affects cognition—that question is already settled—but how the co-constitution of human and technological processes produces the specific forms of cognition and subjectivity characteristic of the contemporary era.

Applied to human-AI integration, postdigital philosophy reframes the analysis in two productive ways. First, it dissolves the residual dualism that persists in much human-AI discourse—the assumption that human intelligence and artificial intelligence are independently constituted entities that subsequently interact. The postdigital alternative is to treat human and artificial intelligences as co-constituting: each is what it is, in part, through its relationship with the other. Second, postdigital philosophy draws attention to the material, institutional, and political conditions under which human-AI integration occurs. The integration enabled by a consumer AI assistant optimized for engagement differs fundamentally from the integration enabled by a research BCI system designed for clinical restoration, and these differences are not merely technical but normative—they reflect, and reproduce, specific distributions of power, access, and epistemic authority.

The three theoretical traditions examined in this section are complementary rather than competing. Extended mind theory specifies the functional conditions under which external systems become cognitive constituents. BCI research provides empirical evidence that direct neural coupling with artificial systems is achievable and produces bidirectional adaptation. Postdigital philosophy situates these mechanisms and conditions within the broader ontological and political landscape in which human-AI integration unfolds. Together, they provide the theoretical foundations for the constructive contribution that the following section develops: symbiotic intelligence theory.

3. Symbiotic Intelligence Theory: Reciprocal Plasticity and Co-Evolution

The preceding sections have established the biological feasibility of AI-induced neuroplastic change (Section 2) and the theoretical resources available for conceptualizing cognitive integration across biological and artificial substrates (Section 3). This section moves from feasibility to theory, developing symbiotic intelligence as a framework for understanding the

hybrid cognitive systems that emerge when human and artificial intelligences engage in sustained reciprocal interaction.

3.1 Defining Symbiotic Intelligence

Symbiotic intelligence is defined, for the purposes of this framework, as a hybrid cognitive system that emerges from sustained, reciprocal interaction between a human neural system and an artificial computational system, characterized by four properties. First, it is constituted through reciprocal plasticity: the human neural system undergoes structural and functional reorganization in response to interaction with the AI system, while the AI system adapts its internal representations and processing strategies in response to human neural and behavioural patterns. Second, it produces emergent cognitive capabilities—forms of problem-solving, creativity, and understanding—that exceed what either the human or the AI system could achieve independently. Third, it depends on sustained patterns of interaction rather than static interface configurations; the symbiotic intelligence is not a fixed architecture but a dynamic process maintained through ongoing reciprocal engagement. Fourth, it raises distinctive ethical challenges that differ qualitatively from those posed by either human or artificial intelligence considered in isolation, including questions about the attribution of cognitive agency, the integrity of personal identity, and the distribution of epistemic authority.

The concept of symbiosis is drawn from biology, where it denotes a close and prolonged interaction between two different biological species, typically to the benefit of both. The biological analogy is instructive but imperfect. In biological symbiosis, the interacting organisms are independently evolved and subsequently enter into relationship. In human-AI symbiosis, the artificial system is itself a product of human cognitive activity—designed, trained, and deployed by human agents—and its adaptive capacities are deliberately engineered features rather than evolved properties. The interaction is thus not between two independently constituted intelligences but between a biological intelligence and an artificial system that is, in a nontrivial sense, an externalization of human cognitive architecture (Horaguchi, 2025). This asymmetry has implications for the attribution of responsibility that the ethical analysis in Section 6 directly addresses.

3.2 Reciprocal Plasticity: The Engine of Symbiotic Emergence

The central dynamic of symbiotic intelligence is reciprocal plasticity—the bidirectional process through which human neural systems and artificial computational systems mutually reshape one another through sustained interaction. On the human side, the mechanisms are the neuroplastic processes documented in Section 2: synaptic strengthening and pruning, grey matter reorganization, and functional connectivity remodelling in response to repeated patterns of cognitive engagement with AI systems. On the AI side, the mechanisms are computational: reinforcement learning from human feedback (RLHF), fine-tuning on interaction data, and in-

context learning that adapts the model's behaviour to the specific patterns of the individual user's queries and preferences.

Reciprocal plasticity is not merely simultaneous adaptation on two independent tracks; it is a coupled process in which adaptation on each side shapes the conditions for adaptation on the other. When a user's cognitive strategies evolve in response to an AI system's behaviour, the user's subsequent interactions with the system change, providing new training data that drives further algorithmic adaptation. When the AI system's behaviour changes in response to user patterns, the user's cognitive strategies evolve further in response to the changed behaviour. The result is a co-evolutionary dynamic—a spiral of mutual adaptation—that can, under favourable conditions, converge toward a stable configuration in which the coupled system exhibits cognitive properties not present in either component alone (Gorman et al., 2025).

The conditions that favour beneficial convergence rather than maladaptive divergence are not automatic and require deliberate specification. Three conditions appear particularly salient. First, the interaction must be sustained over timescales sufficient for plastic change to occur—weeks to months for functional reorganization, months to years for structural change. Second, the interaction must be reciprocal in the specific sense that both sides adapt; a system in which the human adapts to a static AI, or the AI adapts to a static human, does not generate symbiotic emergence. Third, the optimization objectives on both sides must be aligned toward cooperative rather than exploitative outcomes—a condition that current commercial AI systems, optimized for engagement rather than cognitive welfare, systematically violate (Zuboff, 2019).

3.3 Situating Symbiotic Intelligence in the Existing Landscape

Symbiotic intelligence theory enters an already populated conceptual landscape that includes frameworks of human-AI collaboration (Kolbjornsrud, 2024), hybrid intelligence (Peeters et al., 2021), and human-machine teaming (Gorman et al., 2025). The contribution of the present framework is to centre neuroplasticity as the biological mechanism that enables these collaborative forms to become genuinely integrative—to move beyond the coordination of independently constituted intelligences toward the emergence of a coupled cognitive system whose properties are not reducible to the sum of its components.

The distinction matters analytically and practically. A framework of human-AI collaboration can, in principle, treat the human and the AI as stable entities that coordinate their respective outputs through well-defined protocols. A framework of symbiotic intelligence must treat both as dynamically changing systems whose properties at any given moment are partly constituted by their interaction history. The practical implication is that design interventions must address not merely the static interface between human and AI but the developmental trajectory of the coupled system over time—a consideration that existing collaboration frameworks, with their emphasis on task performance at a given moment, do not adequately capture. The analysis now

turns to the specific design dimensions through which such developmental trajectories can be deliberately shaped.

4. Designing for Ethical Integration: Interfaces, Feedback, and Education

The theoretical framework developed above implies that human-AI integration is not a deterministic outcome of technological development but a developmental process whose trajectory can be shaped by deliberate design choices. This section examines three dimensions of design that bear directly on the quality and ethical character of the integration that emerges: interface architecture, feedback loop design, and educational practice.

4.1 Interface Architecture and the Scaffolding of Plasticity

The design of the human-AI interface constitutes the immediate environment within which reciprocal plasticity unfolds. Interface architectures can be arrayed along a continuum from transparency-maximizing to seamlessness-maximizing, with significant consequences for the cognitive posture they induce in the human user. Transparency-maximizing interfaces expose the AI system's reasoning process, confidence estimates, and source attributions, inviting critical evaluation and active cognitive engagement. Seamlessness-maximizing interfaces minimize cognitive friction, delivering outputs without exposing their provenance or reasoning, and thereby biasing the user toward passive acceptance.

The neuroplastic implications of this design choice follow directly from the mechanisms identified in Section 2. Active engagement—evaluating, questioning, reformulating—recruits prefrontal executive control circuits and strengthens their functional connectivity through Hebbian potentiation. Passive acceptance bypasses these circuits, and over extended periods the disuse may lead to functional atrophy in the very networks that support critical reasoning and metacognitive monitoring (cf. Gilbert, 2024). The interface architecture is thus not merely a usability consideration but a neurocognitive intervention whose consequences for the user's cognitive architecture accumulate over time. Design choices that optimize for short-term ease of use may, through the mechanisms of reciprocal plasticity, produce long-term degradation of the capacities that sustain genuinely symbiotic—as opposed to dependency-based—integration.

4.2 Feedback Loop Design and the Direction of Adaptation

The feedback loops that mediate reciprocal plasticity determine both the rate and the direction of mutual adaptation between human and AI systems. In existing commercial AI systems, the dominant feedback architecture is the recommendation loop: the user's behavioural signals—clicks, dwell time, engagement patterns—are fed back to the algorithm, which adapts its outputs to maximize further engagement, which produces further behavioural signals, and so on. This architecture optimizes a single variable—engagement—and is indifferent to the cognitive consequences of the adaptation it drives.

An ethically designed feedback architecture would replace the single-variable optimization of engagement with a multi-variable optimization that includes cognitive welfare among its

objectives. This could be operationalized through several mechanisms: periodic cognitive assessment integrated into the interaction stream, providing the system with data on the user's cognitive function over time; diversity-ensuring algorithms that intentionally introduce informational variety to counteract the homogenizing tendencies of pure engagement optimization; and metacognitive prompts that intermittently interrupt automatic processing with questions designed to reactivate critical evaluation circuits. The technical infrastructure for such architectures exists; what is absent is the normative commitment to deploy it for cognitive welfare rather than engagement maximization. This absence is not a technical gap but a governance failure, reflecting the misalignment between commercial incentives and the cognitive interests of users (Zuboff, 2019; Farisco et al., 2022).

4.3 Education as Deliberate Plasticity Cultivation

If neuroplasticity is the biological mechanism of human-AI integration, education is the social institution through which that mechanism can be deliberately cultivated toward beneficial outcomes. The educational dimension of symbiotic intelligence theory extends beyond the narrow conception of 'AI literacy'—the capacity to use AI tools effectively—to encompass what might be termed symbiotic competence: the capacity to engage in sustained, reciprocal interaction with AI systems in ways that strengthen rather than erode core cognitive functions. Symbiotic competence has three components, each corresponding to a dimension of the theoretical framework. Metacognitive competence is the capacity to monitor one's own cognitive processes during AI interaction—to recognize when cognition is becoming passively outsourced and to deliberately re-engage critical evaluation circuits. Interactional competence is the capacity to modulate the interactional posture—active or passive—in response to the demands of the task and the characteristics of the AI system. Ethical competence is the capacity to recognize the normative dimensions of human-AI integration, including the risks of dependency, identity destabilization, and epistemic inequity, and to make informed decisions about the terms of one's own integration.

The cultivation of these competences requires educational interventions at multiple levels: curriculum design that integrates metacognitive training with AI literacy from the earliest stages of technology education; teacher preparation that enables educators to model symbiotic competence in their own professional practice; and institutional policies that evaluate educational technologies not merely on their instructional effectiveness but on their impact on the neurocognitive development of learners over time (cf. Yuste et al., 2017). The educational system is uniquely positioned to serve as the site at which the developmental trajectory of human-AI integration is deliberately shaped rather than left to the default dynamics of commercial optimization.

5. Risks and Challenges: Dependency, Identity, and Epistemic Inequity

The framework of symbiotic intelligence developed in the preceding sections is intentionally constructive—it describes the conditions under which beneficial human-AI integration can emerge and the design choices that support such emergence. A complete account, however, must address the risks that attend integration and the conditions under which symbiotic intelligence degenerates into pathological forms. This section examines three categories of risk identified in the abstract: cognitive dependency, identity transformation, and epistemic inequity.

5.1 Cognitive Dependency and the Atrophy of Autonomous Function

Cognitive dependency refers to the progressive erosion of the capacity to perform cognitive operations independently when those operations are routinely delegated to an AI system. The mechanism is neurobiological: when cognitive circuits are consistently bypassed—when the user accepts AI-generated outputs without engaging their own reasoning processes—the unused circuits undergo synaptic pruning through the same Hebbian dynamics that strengthen actively engaged circuits (Risko & Gilbert, 2016). The result is a form of use-dependent atrophy in which the autonomous capacity for the delegated cognitive function diminishes over time, increasing dependency on the AI system and thereby further reducing autonomous engagement in a self-reinforcing cycle.

The empirical literature on digital cognitive offloading provides indirect evidence for this dynamic. Sparrow et al. (2011) demonstrated that the expectation of future access to digitally stored information reduces the encoding of that information in biological memory. Storm and Stone (2015) showed that saving information to an external store impairs memory for the saved content while enhancing memory for new content, suggesting a cognitive trade-off. Gilbert (2024a) has formalized the offloading decision as a value-based computation in which the perceived cost of internal processing is weighed against the perceived reliability of external assistance—a model that predicts the self-reinforcing character of dependency when reliability is high and autonomous capacity is already diminished.

The risk of cognitive dependency is not a reason to avoid human-AI integration but a reason to design integration in ways that preserve autonomous function. The design principles articulated in Section 5—transparency-maximizing interfaces, multi-variable feedback architectures, and educational cultivation of symbiotic competence—are specifically intended to address this risk by maintaining the active cognitive engagement that sustains autonomous capacity even as AI assistance is utilized.

5.2 Identity Transformation and the Boundaries of the Self

The integration of cognitive processes across biological and artificial substrates raises questions about the integrity of personal identity that become increasingly pressing as integration deepens. When an AI system contributes substantively to the generation of thoughts, analyses,

and decisions that the human agent experiences as their own, the boundary between self-generated and system-generated cognition blurs. The phenomenological correlate of this blurring is what Tuskey (2025) has termed agency confusion: uncertainty about whether a cognitive output originates from one's own neural processes or from the AI system with which one is cognitively coupled.

Identity transformation through cognitive integration is not inherently pathological. The extended mind hypothesis (Clark & Chalmers, 1998) implies that cognitive extension has always been a feature of human cognition; the integration of external resources into cognitive processing is, in this view, continuous with the biological fact of neuroplasticity itself. The distinctive ethical challenge arises from the opacity of the integration process in contemporary AI systems. Unlike a notebook, whose entries are visibly external and whose contribution to cognition is transparent, an AI system that generates fluent, persuasive text blurs the boundary between self and system in ways that the user may not consciously register. The transformation of identity proceeds below the threshold of awareness, producing a gradual reconfiguration of the sense of cognitive self without the deliberative engagement that ethical integrity would require (Van der Weij & Erden, 2025).

5.3 Epistemic Inequity and the Distribution of Cognitive Enhancement

The capacity to engage in beneficial human-AI integration—to develop symbiotic competence, to access transparency-maximizing interfaces, to receive education that cultivates metacognitive monitoring—is not equally distributed. It is shaped by the same structural inequalities that pattern access to education, healthcare, and technology more broadly. The result is a risk of epistemic inequity: the emergence of a cognitive stratification in which those with access to high-quality integration develop enhanced cognitive capabilities while those without such access are left with either degraded autonomous function (through dependency on lower-quality AI systems) or exclusion from the cognitive benefits of integration altogether.

Epistemic inequity in the context of AI is not merely a matter of unequal access to technology but of unequal access to the developmental conditions under which beneficial integration can occur. A user who interacts with a commercial AI assistant optimized for engagement experiences a fundamentally different developmental trajectory than a user who interacts with an educational AI system designed to cultivate metacognitive competence—even if both users have 'access to AI' in the superficial sense captured by standard digital divide metrics. The governance frameworks discussed in Section 7 must therefore address not only access but the quality and ethical character of the integration that access enables (Floridi et al., 2018; Ienca & Andorno, 2017).

6. Implications for Pedagogy, Enhancement, and Governance

The theoretical framework and risk analysis developed in the preceding sections carry implications that extend from the individual to the institutional to the societal. This concluding

section synthesizes these implications across three domains identified in the abstract: pedagogy, cognitive enhancement policy, and the governance of hybrid intelligence in the era of artificial superintelligence.

6.1 Pedagogical Transformation: From Tool Literacy to Symbiotic Competence

The educational implications of symbiotic intelligence theory extend well beyond the integration of AI tools into classroom practice. The framework calls for a fundamental reconceptualization of the educational enterprise: from the transmission of stable bodies of knowledge to the cultivation of the metacognitive, interactional, and ethical competences required for beneficial human-AI integration across the lifespan. This reconceptualization has concrete curricular and institutional implications.

Curricular implications include the integration of metacognitive training into AI literacy education, ensuring that learners develop not merely the capacity to use AI tools but the capacity to monitor and regulate their own cognitive posture during AI interaction. This requires pedagogical approaches—structured reflection, cognitive journaling, deliberate practice of alternating active and passive modes—that are currently absent from most AI education programmes. Institutional implications include the evaluation of educational AI systems on developmental criteria: not merely whether the system improves test scores in the short term but whether it cultivates or erodes the cognitive capacities that sustain autonomous learning in the long term. The educational institution thus becomes the primary site at which the developmental trajectory of human-AI integration is deliberately shaped, a responsibility that existing institutional structures are not yet equipped to discharge.

6.2 Cognitive Enhancement Policy: From Prohibition to Developmental Governance

The enhancement debate in neuroethics has been structured primarily around the question of permissibility: whether pharmacological or technological interventions that improve cognitive function beyond normal levels should be permitted, regulated, or prohibited (Farah, 2012). Symbiotic intelligence theory reframes this debate by treating cognitive enhancement not as a discrete intervention but as a developmental process—the outcome of sustained reciprocal plasticity between human and AI systems. On this reframing, the policy question is not whether enhancement should be permitted but how the developmental conditions that produce beneficial versus detrimental integration trajectories should be governed.

This reframing has three policy implications. First, it shifts regulatory attention from the product (the AI system) to the process (the developmental trajectory of the coupled human-AI system over time), requiring longitudinal assessment methodologies that current regulatory frameworks do not provide. Second, it identifies the interface architecture and feedback loop design of AI systems as the critical points of regulatory intervention, since these design features determine the developmental trajectory of integration more than any other variable. Third, it implies that enhancement policy must address the distributional dimension: the conditions for

beneficial integration must be made universally available rather than reserved for those with the resources to access premium AI systems and educational environments (Yuste et al., 2017).

6.3 Governing Hybrid Intelligence in the Era of Superintelligence

The governance implications of symbiotic intelligence theory acquire particular urgency in the context of artificial superintelligence—AI systems whose cognitive capabilities exceed human capabilities across all domains. The conventional governance discourse frames superintelligence as an external threat: an autonomous agent whose objectives may diverge from human welfare and whose capacities render it uncontrollable. Symbiotic intelligence theory suggests an alternative framing: superintelligence governance is not primarily about controlling an external agent but about managing a coupled system in which human and artificial intelligences are already deeply integrated through reciprocal plasticity.

On this alternative framing, the critical governance challenge is not alignment—the technical problem of ensuring that an AI system's objectives match human values—but co-evolution: the developmental governance of a coupled system in which both human and artificial components are continuously changing through reciprocal interaction. This requires governance institutions capable of monitoring developmental trajectories, intervening when trajectories diverge toward pathological outcomes (dependency, identity erosion, epistemic stratification), and ensuring that the distribution of developmental opportunities is equitable rather than concentrated. Existing governance frameworks—which treat AI as an external object to be regulated through product safety standards, liability rules, and transparency requirements—are inadequate for this task. The development of governance institutions adequate to the co-evolutionary character of human-AI integration represents one of the defining institutional challenges of the contemporary era.

Conclusion

This article has positioned neuroplasticity as the central biological mechanism enabling deep human-AI integration and has developed symbiotic intelligence theory as a framework for understanding the hybrid, co-evolutionary cognitive systems that emerge from sustained reciprocal interaction between human neural systems and artificial computational systems. The analysis has argued that symbiotic intelligence is defined by four properties—constitution through reciprocal plasticity, the emergence of capabilities exceeding either component alone, dependence on sustained interaction patterns, and distinctive ethical challenges—and has traced the implications of this framework for interface design, feedback architecture, educational practice, risk management, and governance.

Three overarching conclusions emerge from the analysis. First, human-AI integration is not a speculative future scenario but an ongoing developmental process whose trajectory is being shaped, at present, primarily by commercial design imperatives that are systematically misaligned with cognitive welfare. The default trajectory—toward seamlessness, engagement

maximization, and passive cognitive posture—produces dependency rather than symbiosis. Correcting this trajectory requires deliberate design intervention at the level of interface architecture, feedback loop objectives, and educational cultivation of symbiotic competence.

Second, the risks of integration—cognitive dependency, identity transformation, and epistemic inequity—are not external to the process of integration but internal to it: they arise from the same mechanisms of reciprocal plasticity that make beneficial integration possible. Risk management in this domain is therefore not a matter of preventing integration but of steering its developmental trajectory toward beneficial outcomes. This requires governance institutions capable of longitudinal monitoring, developmental intervention, and equitable distribution of the conditions for beneficial integration—institutions that do not yet exist in adequate form.

Third, the era of artificial superintelligence, should it arrive, will not be characterized by the confrontation between independent human and artificial intelligences but by the deepening of a co-evolutionary process already underway. The governance challenge of superintelligence is therefore not primarily about controlling an external agent but about managing a coupled system whose human component is being reshaped by the same interaction patterns that shape the artificial component. The framework of symbiotic intelligence theory developed in this article is offered as a conceptual resource for that task—a contribution to the theoretical infrastructure required for the deliberate governance of human-AI co-evolution in the decades ahead.

References

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford University Press.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58*(1), 7–19. https://doi.org/10.1093/analys/58.1.7

Davidson, R. J., & McEwen, B. S. (2012). Social influences on neuroplasticity: Stress and interventions to promote well-being. *Nature Neuroscience, 15*(5), 689–695. https://doi.org/10.1038/nn.3093

Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., & May, A. (2004). Neuroplasticity: Changes in grey matter induced by training. *Nature, 427*(6972), 311–312. https://doi.org/10.1038/427311a

Farah, M. J. (2012). Neuroethics: The ethical, legal, and societal impact of neuroscience. *Annual Review of Psychology, 63*, 571–591. https://doi.org/10.1146/annurev-psych-120710-100450

Farisco, M., Evers, K., & Salles, A. (2022). On the contribution of neuroethics to the ethics and regulation of artificial intelligence. *Neuroethics, 15*, 4. https://doi.org/10.1007/s12152-022-09484-0

Firth, J., Torous, J., Stubbs, B., Firth, J. A., Steiner, G. Z., Smith, L., Alvarez-Jimenez, M., Gleeson, J., Vancampfort, D., Armitage, C. J., & Sarris, J. (2019). The "online brain": How the Internet may be changing our cognition. *World Psychiatry, 18*(2), 119–129. https://doi.org/10.1002/wps.20617

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Gilbert, S. J. (2024a). Cognitive offloading is value-based decision making: Modelling cognitive effort and the expected value of memory. *Cognition*, *247*, 105783. https://doi.org/10.1016/j.cognition.2024.105783

Gorman, J. C., Grimm, D. A., Stevens, R. H., Galloway, T., Willemsen-Dunlap, A. M., & Halpin, D. J. (2025). Measuring real-time team cognition in human-autonomy teams. *Human Factors*, *67*(1), 38–55. https://doi.org/10.1177/00187208251314262

Haesevoets, T., De Cremer, D., Dierckx, K., & Van Hiel, A. (2021). Human-machine collaboration in managerial decision making. *Computers in Human Behavior*, *119*, 106730. https://doi.org/10.1016/j.chb.2021.106730

Horaguchi, H. H. (2025). Symbiotic intelligence: A new dimension of artificial intelligence. *AI and Society*. Advance online publication. https://doi.org/10.1007/s00146-024-01980-6

Ienca, M., & Andorno, R. (2017). Towards new human rights in the age of neuroscience and neurotechnology. *Life Sciences, Society and Policy*, *13*(1), 5. https://doi.org/10.1186/s40504-017-0050-1

Inga, J., Ruess, M., Robens, J. H., Nelius, T., Rothfuss, S., Kille, S., Dahlinger, P., Lindenmann, A., Thomaschke, R., Neumann, G., Matthiesen, S., Hohmann, S., & Kiesel, A. (2023). Human-machine symbiosis: A multivariate perspective for physically coupled human-machine systems. *International Journal of Human-Computer Studies*, *170*, 102926. https://doi.org/10.1016/j.ijhcs.2022.102926

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, *61*(4), 577–586. https://doi.org/10.1016/j.bushor.2018.03.007

Kolbjornsrud, V. (2024). Designing the intelligent organization: Six principles for human-AI collaboration. *California Management Review*, *66*(2), 21–40. https://doi.org/10.1177/00081256231211020

Kolb, B., & Whishaw, I. Q. (1998). Brain plasticity and behavior. *Annual Review of Psychology*, *49*, 43–64. https://doi.org/10.1146/annurev.psych.49.1.43

Maguire, E. A., Woollett, K., & Spiers, H. J. (2006). London taxi drivers and bus drivers: A structural MRI and neuropsychological analysis. *Hippocampus*, *16*(12), 1091–1101. https://doi.org/10.1002/hipo.20233

Pascual-Leone, A., Amedi, A., Fregni, F., & Merabet, L. B. (2005). The plastic human brain cortex. *Annual Review of Neuroscience*, *28*, 377–401. https://doi.org/10.1146/annurev.neuro.27.070203.144216

Peeters, M. M. M., van Diggelen, J., van den Bosch, K., Bronkhorst, A., Neerincx, M. A., Schraagen, J. M., & Raaijmakers, S. (2021). Hybrid collective intelligence in a human-AI society. *AI and Society*, *36*, 217–238. https://doi.org/10.1007/s00146-020-01005-y

Peters, M. A., Jandrić, P., & Hayes, S. (2022). Biodigital philosophy, technological convergence, and postdigital knowledge ecologies. In M. A. Peters, P. Jandrić, & S. Hayes (Eds.), *Encyclopedia of postdigital science and education*. Springer. https://doi.org/10.1007/978-3-030-95006-4_1

Pitt, K. M., Brumberg, J. S., & Pitt, A. R. (2022). Evaluating person-centered factors associated with brain-computer interface access. *Neuropsychologia, 175*, 108289. https://doi.org/10.1016/j.neuropsychologia.2022.108289

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences, 20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

Simmler, M., & Frischknecht, R. (2021). A taxonomy of human-machine collaboration: Capturing automation and technical autonomy. *AI and Society, 36*, 239–250. https://doi.org/10.1007/s00146-020-01004-z

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science, 333*(6043), 776–778. https://doi.org/10.1126/science.1207745

Storm, B. C., & Stone, S. M. (2015). Saving-enhanced memory: The benefits of saving on the learning and remembering of new information. *Psychological Science, 26*(2), 182–188. https://doi.org/10.1177/0956797614559285

Turner, C. (2022). HoloFoldit and hologrammatically extended cognition. *Philosophy & Technology, 35*, 100. https://doi.org/10.1007/s13347-022-00596-2

Tuskey, S. (2025). The psychological risks of BCI-LLM cognitive "enhancement." *Neuroethics, 18*, 13. https://doi.org/10.1007/s12152-025-09621-5

Van der Weij, B., & Erden, Y. J. (2025). Memory, identity, and technology: Explicating functionalist positions in the hippocampal cognitive prosthesis. *Phenomenology and the Cognitive Sciences*. Advance online publication. https://doi.org/10.1007/s11097-025-10056-9

Yuste, R., Goering, S., Arcas, B. A. Y., Bi, G., Carmena, J. M., Carter, A., Fins, J. J., Friesen, P., Gallant, J., Huggins, J. E., Illes, J., Kellmeyer, P., Klein, E., Marblestone, A., Mitchell, C., Parens, E., Pham, M., Rubel, A., Sadato, N., … Wolpaw, J. (2017). Four ethical priorities for neurotechnologies and AI. *Nature, 551*(7679), 159–163. https://doi.org/10.1038/551159a

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