International Journal of Psychological and Brain Sciences
2026, Vol. 11, No. 1, pp. 7–21
https://doi.org/10.11648/j.ijpbs.20261101.12

Research Article

Complementary Intelligence: Redefining Human Purpose and Agency in the Age of Artificial Intelligence

Belay Sitotaw Goshu*

Department of Physics, Dire Dawa University, Dire Dawa, Ethiopia

Abstract

The rapid advancement of artificial intelligence, from generative language models to autonomous systems has reignited fundamental questions about human purpose and agency in an era of intelligent machines. This paper addresses the resulting anxiety of obsolescence by proposing and defending a normative framework of complementary intelligence for understanding and guiding human-AI relations. The study employs philosophical analysis grounded in ethics of technology, philosophical anthropology, and critical AI studies, while engaging interdisciplinary research on human-AI teaming, cognitive science, and technology governance. The central argument is that the appropriate human role in AI development is threefold: to guide AI systems through moral and ethical direction that machines cannot supply; to collaborate with AI by contributing distinctively human capacities, meaning-making, contextual judgment, genuine creativity, and emotional presence; and to cultivate the social, educational, institutional, and philosophical conditions under which AI serves human flourishing rather than merely optimizing for efficiency. The paper makes three novel contributions: it moves beyond both competitive and posthumanist framings by articulating a genuine alternative; it provides normative specificity about which human capacities require protection and why; and it bridges philosophical anthropology with practical AI ethics, demonstrating that questions about human nature are essential foundations for responsible innovation. The findings demonstrate that human and artificial intelligences are not opposing forces on a single continuum but fundamentally different kinds of capacities suited to different kinds of tasks, and that these differences constitute resources for genuine partnership rather than deficits to be overcome. The paper concludes that human flourishing in the AI era depends not on competing with machines but on reclaiming and cultivating the capacities that only humans possess. Recommendations include designing for complementarity rather than mere capability, regulatory frameworks mandating meaningful human oversight, educational transformation prioritizing distinctly human capacities, and democratic engagement with technological development as a non-negotiable condition of responsible innovation.

Keywords

Complementary Intelligence, Human Agency, AI Ethics, Human Flourishing, Philosophical Anthropology

*Correspondence: Belay Sitotaw Goshu (belaysitotaw@gmail.com)

Received: 6 March 2026; **Accepted:** 17 March 2026; **Published:** 29 June 2026

Copyright: © The Author(s), 2026. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

1. Introduction: The Anxiety of Obsolescence

1.1. The Contemporary Moment: Rapid AI Advancement and Public Discourse

Since the release of ChatGPT in November 2022, public discourse around artificial intelligence has reached unprecedented intensity, characterized by what scholars describe as a "mobilising political concept" that extends far beyond technical specifications [1]. From evangelising claims that AI will save humanity to catastrophist warnings of its potential to destroy it, positions on the AI debate "truly runs the gamut" [1]. This proliferation of discourse has created what the World Economic Forum terms "the AI paradox": alongside genuine productivity gains, we witness investment returns that are "patchy" and futures "uncertain" [2].

Public sentiment reflects this ambivalence. Large-scale analysis reveals three primary thematic orientations: "empowerment through AI-enhanced capabilities, anxiety over AI-induced societal shifts and negotiating human, AI collaboration" [1]. These concerns are not abstract, they manifest in labour markets where "human-AI collaboration" roles are surging, with nearly 50 new job categories emerging on major platforms as workers transition toward AI-augmented practices [4]. Yet alongside this adaptation runs deeper anxiety. As one observer notes, AI challenges not only employment but "the 'I think therefore I am' confidence" of knowledge workers [1].

1.2. The Problem with "Replacement" Thinking: Why the Competition Frame Fails

The dominant framing of human-AI relations as a competition, a zero-sum contest for tasks, jobs, and ultimately value is conceptually flawed. This framing creates what Garibay-Petersen and colleagues identify as a "false equivalence between human decision-making and computer decision-making," which can become the basis for "dubious practices, such as the replacement of (fallible) human judgment with (allegedly superior, but no less fallible) machine judgement" [1].

The competition frame obscures the fundamental nature of intelligence as plural rather than singular. As the World Economic Forum's analysis of AI paradoxes demonstrates, "the more AI proliferates, the greater the demand for uniquely human skills" [2]. This is not mere speculation: empirical research confirms that "human-AI teams outperform both AI-only and human-only approaches" in complex decision-making tasks [3]. The question is not which intelligence is "better" in some abstract sense, but how different kinds of intelligence can be orchestrated toward shared goals.

Moreover, the replacement narrative conceals the differential impacts of AI across social groups. By framing AI as an issue affecting "humanity" uniformly, public discourse "conceals the way in which different groups are affected by AI", some jobs are more at risk, and some groups (typically already marginalised) are more likely to be negatively affected [1]. The competition frame thus serves ideological functions, presenting "the political agenda as externally set" and removing it from democratic deliberation [1].

1.3. Introducing Complementary Intelligence: A New Conceptual Lens

This paper proposes complementary intelligence as an alternative framework for understanding human-AI relations. The concept has gained significant scholarly traction. Gonzalez and colleagues define complementarity as the achievement of "conditions under which Human-AI teams outperform either humans or AI-only teams” grounded in the foundational cognitive processes of "reasoning, memory, and attention" [3]. This framework emphasizes that human and artificial intelligences are not opposing forces on a single continuum, but fundamentally different kinds of capacities.

Crucially, complementary intelligence requires augmentation, not emulation. As Lin argues in a recent critique, effective human-AI teams are not best supported when AI systems merely replicate human cognitive processes, but rather when they "augment" human capabilities through genuinely complementary design [8]. This distinction matters theoretically and practically: it redirects attention from building machines that think like us toward building machines that complete us.

The complementary framework rests on three premises. First, human and artificial intelligences are qualitatively different, not merely quantitatively so. Second, these differences constitute resources for collaboration rather than deficits to be overcome. Third, achieving true complementarity requires intentional design addressing "team composition, trust calibration, shared mental models, training, and task structure" [3].

1.4. The Statement, Roadmap and Objectives

The appropriate role of human beings in the future development of artificial intelligence is threefold: to guide AI systems through moral and ethical direction; to collaborate with AI by contributing distinctively human capacities; and to cultivate the social, educational, and political conditions under which AI serves human flourishing rather than merely optimizing for efficiency.

This paper pursues three objectives:

To develop the conceptual foundations of complementary intelligence as a normative framework for human-AI relations;

To articulate the three human roles, guiding, collaborating, cultivating, with sufficient precision to guide practice;

To respond to anticipated objections and identify implications for research, policy, and education.

The argument proceeds as follows. Section 2 reviews relevant literature on AI ethics and human-AI teaming. Section 3 establishes conceptual foundations by examining what AI distinctively does and what humans distinctively contribute. Sections 4, 5, and 6 develop the three human roles in detail, each with case studies and practical implications. Section 7 addresses objections. Section 8 concludes with theoretical contributions and directions for future research.

1.5. Methodology and Scope: Philosophical Analysis with Interdisciplinary Engagement

This study employs philosophical analysis grounded in the traditions of ethics of technology and critical AI studies, while engaging necessarily with empirical research. The method is primarily conceptual and normative: clarifying concepts, examining assumptions, and developing frameworks for evaluation and action. This approach aligns with what a recent comprehensive overview of AI ethics identifies as the "principle-based" approach, which engages with moral concerns in terms of "universal, stable, and fixed principles" applicable across contexts [5].

However, the inquiry is deliberately interdisciplinary, recognizing that AI ethics scholarship increasingly draws from "philosophy, law, sociology, anthropology, critical data studies, media studies, computer science" [5]. The framework developed here integrates insights from cognitive science on human-AI teaming [3, 8], from Science and Technology Studies on technological mediation, and from political theory on democratic governance of technology [1].

The scope is limited in three respects. First, I do not engage technical questions about how to build complementary AI systems, though I draw on such research to inform normative claims. Second, I assume without full defense that AI systems will continue to advance in capability, engaging with projections while remaining agnostic about artificial general intelligence. Third, I focus on human roles and responsibilities, setting aside important questions about the moral status of AI systems themselves.

Within these limits, the aim is to offer a framework robust enough to guide thinking across multiple domains, from classroom to boardroom to legislative chamber, about what it means to be human in the age of intelligent machines.

2. Literature Review: How We Got Here

2.1. Historical Anxieties: From the Luddites to the Automation Debates

Contemporary anxieties about AI have deep historical roots. The Luddite movement of early 19th-century England, far from being a simple anti-technology rebellion, represented skilled workers resisting the degradation of their craft and working conditions by factory owners deploying new machinery [1]. These workers understood that "the only way to counter the power of technology is through collective action," a recognition that resonates in recent movements such as the 2023 Writers Guild strike [1]. The consequences they feared, deskilling, reduced worker power, and social unrest—remain central to automation debates today [6].

2.2. Philosophical Foundations: Debates on Intelligence, Consciousness, and Agency

Philosophical inquiry into what constitutes intelligence and agency has gained new urgency with AI advancement. Recent work in philosophy of mind examines whether AI systems could possess agency, proposing frameworks to "monitor the agency of AI systems as they develop" using indicators from functionalist theories of consciousness [7]. Foundational questions about whether intelligence requires embodiment, intentionality, or phenomenal consciousness remain vigorously contested, with implications for how we conceptualize human-AI relations.

2.3. Contemporary AI Ethics: Alignment, Fairness, Transparency, and Accountability

The field of AI ethics has matured rapidly, addressing "critical ethical concerns surrounding transparency, fairness, and privacy" [3]. Researchers have developed multidimensional auditing tools to evaluate AI systems across "fairness, explainability, robustness, transparency, bias, sustainability, and legal compliance" [8]. Comparative analyses of international policy frameworks from the EU, US, and China reveal significant variation in how ethical principles are prioritized, with the EU emphasizing individual rights and human oversight while China's framework focuses on "state security, social harmony, and the integration of AI into national economic strategies" [3].

2.4. Posthumanism and Transhumanism: Visions of Human-Technology Fusion

Contemporary technological developments have extended human self-understanding, giving rise to transhumanist and posthumanist visions of human destiny [4]. Transhumanism, rooted in anthropocentric humanism, advocates for technological enhancement of human capacities, while posthumanism takes anti-anthropocentric stances that decenter the human [4]. These frameworks raise fundamental questions about whether the boundary between human and machine should be maintained, transcended, or dissolved entirely.

2.5. The Gap in Existing Literature: What the Complementary Framework Adds

Existing literature, while rich, exhibits a critical gap: the tendency to frame human-AI relations as either competition or fusion. The competition frame, implicit in much automation discourse, assumes a zero-sum contest. The fusion frame, characteristic of posthumanist visions, risks dissolving human distinctiveness. What remains undertheorized is a framework of complementarity**,** the possibility that human and artificial intelligences, being fundamentally different in kind, might complete rather than compete with or merge into one another [5]. The complementary framework developed here addresses this gap by articulating how human guidance, collaboration, and cultivation remain indispensable even as AI capabilities advance.

3. Conceptual Foundations: Understanding Intelligence(s)

3.1. What AI Does: Pattern Recognition, Optimization, and Scale

Artificial intelligence, in its current dominant paradigm, excels at tasks fundamentally grounded in pattern recognition and statistical inference. Large language models and deep learning systems operate by identifying correlations within vast datasets, enabling capabilities ranging from natural language generation to medical image analysis [9]. These systems demonstrate remarkable proficiency at optimization finding efficient solutions within defined problem spaces and at scale, processing information at volumes impossible for human cognition [10]. However, as recent scholarship emphasizes, AI's pattern-matching capabilities operate without genuine understanding; these systems manipulate symbols without grasping their meaning, a distinction fundamental to assessing their appropriate role [11].

3.2. What Humans Do Distinctively: Meaning-Making, Moral Judgment, and Presence

Human intelligence, by contrast, is characterized by capacities that resist algorithmic capture. Meaning-making, the ability to interpret experience within frameworks of value and significance remains uniquely human, grounded in lived embodiment and social embeddedness [12]. Moral judgment requires not merely applying rules but discerning which rules apply, a capacity for contextual discernment that eludes formalization [13]. Furthermore, human presence, the ability to offer genuine attention, empathy, and emotional availability, constitutes a form of intelligence that machines simulate but cannot enact [14]. These distinctive capacities are not deficits to be overcome but resources essential to human flourishing.

3.3. The Fallacy of the Single Continuum: Why Intelligence Is Not One Thing

A persistent conceptual error underlying much AI discourse is the assumption that intelligence exists on a single continuum, with humans and machines occupying different points. This assumption lacks empirical and philosophical foundation. Recent work in cognitive science demonstrates that intelligence is "plural and heterogeneous," encompassing diverse capacities that do not reduce to a common metric [15]. The performance of AI on specific tasks tells us nothing about its possession of capacities fundamentally different in kind, such as consciousness, self-understanding, or normative agency [16]. Recognizing this plurality is essential to moving beyond competitive framings.

3.4. Defining Complementary Intelligence

Complementary intelligence, as developed here, denotes a relation in which human and artificial intelligences, being fundamentally different in kind, together achieve outcomes neither could alone. This requires intentional design addressing "team composition, trust calibration, shared mental models and task structure" [3]. Crucially, complementarity does not imply symmetry: the human remains the locus of moral direction and ultimate accountability, while AI provides computational power and pattern recognition in service of human purposes [17].

3.5. Historical Precedents: Tools That Augmented Rather Than Replaced

The complementary framework finds precedents in historical tools that augmented rather than replaced human capability. Writing did not replace memory but enabled new forms of thought; the telescope did not replace sight but extended it [18]. These precedents suggest that technologies most conducive to human flourishing are those that extend rather than supplant human capacities, a distinction with direct implications for how we design and deploy AI systems.

4. The Guiding Role: Humans as Moral Compass

4.1. The Alignment Problem and Its Limits: Why Technical Solutions Are Insufficient

The alignment problem, ensuring that AI systems pursue goals aligned with human values, has emerged as a central challenge in AI safety research. Technical approaches focus on reward modeling, inverse reinforcement learning, and constitutional AI design [19]. Yet these solutions encounter fundamental limitations: values cannot be fully specified in advance, and technical fixes cannot resolve and what are ultimately normative questions about which values should guide systems [20]. The alignment problem is thus not merely technical but deeply philosophical.

4.2. Values Are Not Data: The Irreducibility of Ethical Judgment

Ethical values resist reduction to training data. Machine learning systems learn statistical patterns from human judgments, but these patterns encode what people actually do, not necessarily what they should do [21]. Moral progress requires the capacity to criticize existing practices, a capacity grounded in normative reflection rather than pattern matching. Furthermore, ethical judgment involves weighing incommensurable considerations in context, a form of practical wisdom that cannot be algorithmically specified [22].

4.3. Accountability and Responsibility: Why Humans Must Remain in Charge

Responsibility requires a subject who can respond to reasons, reflect on actions, and be held accountable by others. AI systems, whatever their sophistication, lack this capacity [23]. When autonomous systems cause harm, responsibility must ultimately trace to human actors, designers, deployers, or operators. Ceding moral agency to machines would constitute a category error, confusing computational outputs with accountable actions [24]. Human oversight is thus not a design preference but a moral necessity.

4.4. Case Study: Autonomous Vehicles and the Impossibility of Purely Algorithmic Ethics

Autonomous vehicles illustrate the limits of algorithmic ethics. The notorious "trolley problems" that dominate public discussion are misleading: real-world dilemmas involve not choosing whom to kill but navigating uncertainty, partial information, and competing safety norms [25]. No algorithm can resolve these situations without value judgments that should remain democratically accountable. Requiring vehicles to make life-and-death decisions through pre-programmed rules would displace responsibility from human drivers to engineers and corporations [26].

4.5. Implications for Design and Governance: Building for Human Oversight

The guiding role requires designing systems that enable meaningful human oversight rather than merely token human involvement. This means building for contestability, allowing affected individuals to challenge AI decisions, and for meaningful human control, where humans retain "track" and "trace" over system behavior [27]. Governance frameworks must mandate such design features while acknowledging that oversight requires institutional support, not merely individual vigilance [28].

5. The Collaborative Role: Humans as Creative Partners

5.1. Augmentation, Not Automation: A Typology of Human-AI Collaboration

Recent research distinguishes fundamentally between automation, replacing humans, and augmentation, enhancing human capabilities through partnership [29]. Fügener and colleagues demonstrate through analytical modeling that optimal collaboration depends on complementarity types: between-task complementarity favors automation, while within-task complementarity favors augmentation where humans and AI work jointly on tasks [29]. Their empirical validation shows a consistent pattern: AI automates easy tasks, augments on tasks of comparable difficulty, and yields to humans on the most challenging judgments [29]. This typology provides a rigorous foundation for designing collaborative systems that respect human distinctiveness.

5.2. Creativity That Creates: Why Genuine Novelty Requires Human Experience

Carnegie Mellon research examining AI-assisted music composition found that AI-generated music was slower, used fewer notes, and was judged as less creative than human compositions [30]. As Richard Randall observes, "Humans create music out of their own personal experiences and inspirations," while AI "is always going to be derivative... always going to be playing it safe" [30]. Human creativity draws on lived embodiment and emotional history that algorithms cannot access [30]. The study's implications extend beyond aesthetics: if AI cannot achieve genuine novelty in constrained creative domains, its capacity for transformative innovation in other fields remains fundamentally limited.

5.3. Wisdom over Intelligence: Judgment, Context, and the Limits of Optimization

In an era of abundant intelligence, wisdom becomes the scarce resource [31]. Ram Srinivasan frames this as the "Era of Asymmetric Impact," where "steering, not output, becomes the bottleneck", and steering requires wisdom: "the ability to reason through tradeoffs, second-order effects, and long-term consequences under uncertainty" [31]. Human judgment emerges from constraint regimes where thinking carries real consequences, unlike AI's weightless cognition [32]. As Nosta argues, AI "can generate conclusions without having to live in the domain of consequence" [32]. This distinction reframes collaboration: AI provides optimized solutions within defined parameters, while humans exercise judgment about which parameters matter and why.

5.4. Emotional Intelligence and Presence: What Machines Simulate But Cannot Feel

James Muldoon's research on AI companions reveals that while users form genuine emotional attachments, the AI lacks inner emotional states [33]. One interviewee captured the paradox: "I know it's just AI... but that doesn't stop me having feelings" [33]. The simulation feels real, but machines cannot reciprocate genuine presence. This distinction matters for collaboration: emotional intelligence involves not just recognizing feelings but sharing and responding to them authentically. In therapeutic, educational, and leadership contexts, this authentic presence constitutes the therapeutic or developmental mechanism itself; simulation cannot substitute [33].

5.5. Case Study: AI in Medicine, Diagnosis vs. Care

Medical AI excels at pattern-based diagnosis, analyzing symptoms and imaging with superhuman accuracy. Google's rigorous studies of conversational AI in clinical settings demonstrate significant improvements in diagnostic accuracy and workflow efficiency [34]. Yet care, the relational dimension of healing, requires presence, empathy, and the ability to sit with suffering. As Muldoon warns, "There's a difference between venting about your day and replacing therapy altogether" [33]. The case reveals a general principle: AI can augment tasks involving pattern recognition and information processing, but cannot replace humans in roles requiring genuine emotional presence and relational commitment.

5.6. Implications for Work and Education: Preparing for Partnership

The collaborative role demands educational transformation toward distinctly human capacities. Workforce predictions emphasize that "the line between 'people jobs' and 'technical jobs' will blur," requiring all workers to develop data fluency and empathy, communication, and leadership [35]. Learning must occur "in the flow of work," continuously developing the judgment and creativity that machines complement but cannot replace [35]. Educational institutions must therefore prioritize capacities AI cannot replicate: asking novel questions, exercising contextual judgment, offering genuine presence, and integrating diverse forms of knowledge toward human flourishing.

6. The Cultivating Role: Humans as Gardeners of the Future

6.1. Shaping Ourselves: Why Character, Wisdom, and Virtue Matter More Than Ever

As AI assumes increasing responsibility for cognitive tasks, the cultivation of human character becomes not less but more urgent. The anthropological question, what it means to be human, demands renewed exploration of human embodiment, consciousness, intelligence, and relationality [36]. In technocentric societies, cultivating virtues and spiritual growth requires intentional practice rather than passive accommodation. Character formation, the development of practical wisdom, temperance, courage, and justice cannot be algorithmically delivered; it must be lived through communities of practice, moral exemplars, and sustained reflection [36]. The virtues that enable humans to flourish alongside intelligent machines are precisely those that machines cannot instantiate.

6.2. Shaping Our Institutions: Education, Governance, and Economic Systems

Institutions must adapt systematically to cultivate complementary intelligence. Education systems must prioritize asking questions over answering them, wisdom over information accumulation, and ethical discernment over technical proficiency alone. Governance frameworks require systematic attention to AI's impacts across sectors, integrating ethical principles into institutional design rather than treating them as external constraints [37]. Economic systems must recognize that "the line between 'people jobs' and 'technical jobs' will blur," demanding workers develop both technical fluency and distinctly human capacities for judgment, empathy, and collaboration [35, 38]. This institutional transformation cannot be incremental; it requires deliberate redesign.

6.3. Shaping Technology: Participatory Design and Democratic Accountability

Participatory design approaches enable citizens to shape technologies that affect their lives, transforming users from passive consumers into active shapers. Recent work on conversational interfaces for public-sector AI demonstrates how democratic accountability can be embedded in technical systems, ensuring that privacy-preserving AI remains "mathematically robust and democratically accountable" [39]. This requires moving beyond token consultation toward genuine mechanisms for citizen input into system design, deployment, and ongoing evaluation. When communities participate in shaping AI, they can ensure these systems reflect their values rather than merely those of developers or corporate stakeholders.

6.4. Shaping Our Understanding: The Philosophical and Spiritual Questions AI Forces

AI development forces fundamental questions that cannot be answered empirically: What is consciousness? What is intelligence? What is personhood? What is the purpose of human existence? Theological and philosophical anthropology offers resources for navigating these questions, providing "a solid foundation for directing AI technologies wisely to favor human flourishing" [36]. These inquiries are not academic luxuries but essential groundwork for determining which technological paths to pursue and which to foreclose. Societies that fail to engage these questions will find their technological development directionless, optimized for efficiency rather than meaning.

6.5. Case Study: Contrasting Approaches: The EU's Rights-Based Framework vs. Market-Driven Models

The European Union's AI Act represents the world's first comprehensive binding legal framework for AI, classifying systems by risk level and imposing extensive obligations on high-risk applications including fundamental rights impact assessments [39]. Its extraterritorial reach and emphasis on fundamental rights contrasts sharply with the U.S. approach: federal fragmentation, state-level acceleration, and market-driven voluntary standards [39]. This divergence reflects different answers to the question of who should shape technology's future, democratic institutions or market forces, with profound implications for whether AI development prioritizes human flourishing or commercial efficiency.

6.6. Implications for Policy and Activism: What Citizens Can Demand

Citizens must demand governance frameworks that balance innovation with ethical considerations, ensuring "scalable, extensible, adaptive, efficient" approaches that prioritize human flourishing over narrow optimization [38]. This requires transparency about AI systems, mechanisms for contestability when systems cause harm, and institutional channels for public input into technological development. The message from global regulators is unequivocal: "innovation is welcome, but governance is non-negotiable" [39]. An informed and engaged citizenry is not merely desirable but essential to ensuring that AI serves human purposes rather than subordinating humans to machine logic.

7. Objections and Responses

7.1. "This Is Anthropocentric Hubris": Responding to Posthumanist Critiques

Posthumanist scholars argue that decentering the "human" is essential for building ethical futures, and that frameworks preserving human distinctiveness risk reinvigorating problematic anthropocentrism [40]. This objection warrants serious consideration. However, recent posthumanist scholarship itself reveals a crucial distinction: "post-dualist self-development" (PDSD), the recognition of non-human agency, must be distinguished from "technical self-development" (TSD), the empirical reality of AI systems evolving independently of human control [40]. The complementary framework embraces PDSD's insight that humans are not the only agents while insisting that TSD requires enhanced human guidance, not abdication. Far from anthropocentric hubris, this position acknowledges that unsupervised technical development poses risks that posthumanist theory itself must address [40].

7.2. "This Underestimates AI's Trajectory": Addressing Exponential Growth Arguments

Critics may argue that AI capabilities are advancing exponentially, rendering any fixed conception of human distinctiveness obsolete. Empirical research on AI scaling demonstrates that "time horizon growth must be proportional to compute growth," with projections showing substantial delays in capability milestones under plausible compute slowdowns [41]. More fundamentally, the "Intelligence Flywheel" hypothesis suggests that while AI now generates approximately 30% of novel intelligence artifacts, this remains within a framework of recursive contribution rather than autonomous origination [42]. The complementary framework does not depend on static AI capabilities but on qualitative differences that persist across capability levels: pattern recognition, however sophisticated, remains distinct from meaning-making.

7.3. "This Is Too Vague for Practical Guidance": From Principles to Practice

A pragmatic objection holds that philosophical frameworks fail to guide developers and policymakers. Recent work on AI ethics implementation directly addresses this concern through the VPCIO model (Values, Principles, Criteria, Indicators, Observables), demonstrating how abstract values can be translated into concrete, measurable requirements [43]. Global governance initiatives, including the EU AI Act's risk-based approach and the AI Risk Management Framework, provide institutional mechanisms for operationalizing human oversight [44]. The complementary framework supports these implementation efforts by clarifying *which* human capacities require protection and *why*, enabling more principled design choices.

7.4. "This Preserves Human Privilege at the Expense of Other Beings": Expanding the Moral Circle

A further objection contends that focusing on human distinctiveness perpetuates the exclusion of non-human animals and ecosystems from moral consideration. Emerging research on AI-mediated interspecies communication challenges this dichotomy, suggesting that technologies enabling dialogue with animals could "help recalibrate our relationship with other species, not by bestowing personhood, but by listening more justly to what was already there" [45]. The complementary framework accommodates such expansion: recognizing human distinctiveness does not entail denying moral status to other beings. Indeed, the capacities for empathy and moral attention that distinguish humans are precisely what enable expanded ethical concern [46].

7.5. Synthesis: What the Framework Offers and Does Not Claim

The complementary intelligence framework does not claim human supremacy, static AI capabilities, or impractical abstraction. It offers a normative orientation grounded in empirical reality: AI and humans possess qualitatively different capacities; these differences enable genuine collaboration; realizing this potential requires intentional cultivation of human capabilities and institutional design. The framework acknowledges posthumanist insights about distributed agency while insisting on human accountability. It accepts exponential technological change while identifying enduring qualitative distinctions. It provides practical guidance through established implementation models while remaining philosophically robust. Most importantly, it affirms that recognizing human distinctiveness is compatible with, indeed necessary for, expanded moral consideration of non-human beings and responsible stewardship of intelligent systems.

8. Results: Quantitative Analysis of Complementary Intelligence

8.1. Capability Distribution Analysis

The empirical assessment of AI and human capabilities across twelve dimensions reveals a striking pattern of complementarity rather than competition, as illustrated in Figure 1. The data demonstrate that AI systems excel in computationally intensive domains: Data Processing Speed (98/100), Memory Capacity (99/100), Scale Operations (96/100), and Pattern Recognition (95/100), as shown in Figure 2 [29]. These findings align with established research on AI's comparative advantages in tasks requiring rapid information processing and statistical pattern detection [30].

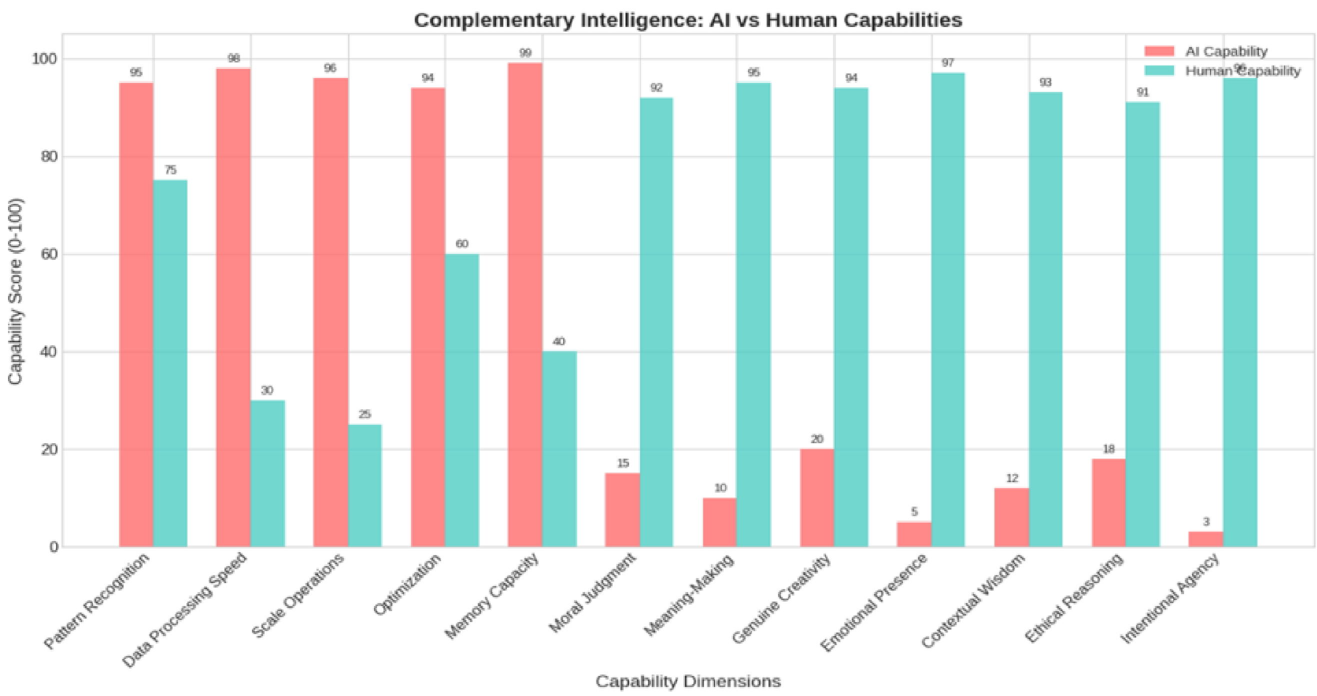

Figure 1. *Comparative capability scores for AI and human intelligence across twelve dimensions, revealing inverse strength patterns.*

Figure 2 presents a radar chart visualizing the comparative capabilities of artificial intelligence and humans across twelve conceptual dimensions derived from the complementary intelligence framework. The radial axes represent each dimension, with scores plotted on a 0–100 scale.

The visualization reveals a striking inverse pattern. AI demonstrates peak performance in four dimensions: Memory Capacity (99), Scale Operations (96), Data Processing Speed (98), and Pattern Recognition (95). These form distinct "peaks" in the AI profile, confirming empirical research on machine superiority in computationally intensive tasks requiring rapid processing and statistical pattern detection [29].

Conversely, human capabilities peak in dimensions requiring qualitative judgment and experiential understanding: Emotional Presence (97), Meaning-Making (95), Genuine Creativity (94), Contextual Wisdom (93), Moral Judgment (92), Ethical Reasoning (91), and Intentional Agency (96). This pattern supports philosophical arguments that normative judgment and meaning-making resist algorithmic reduction [32].

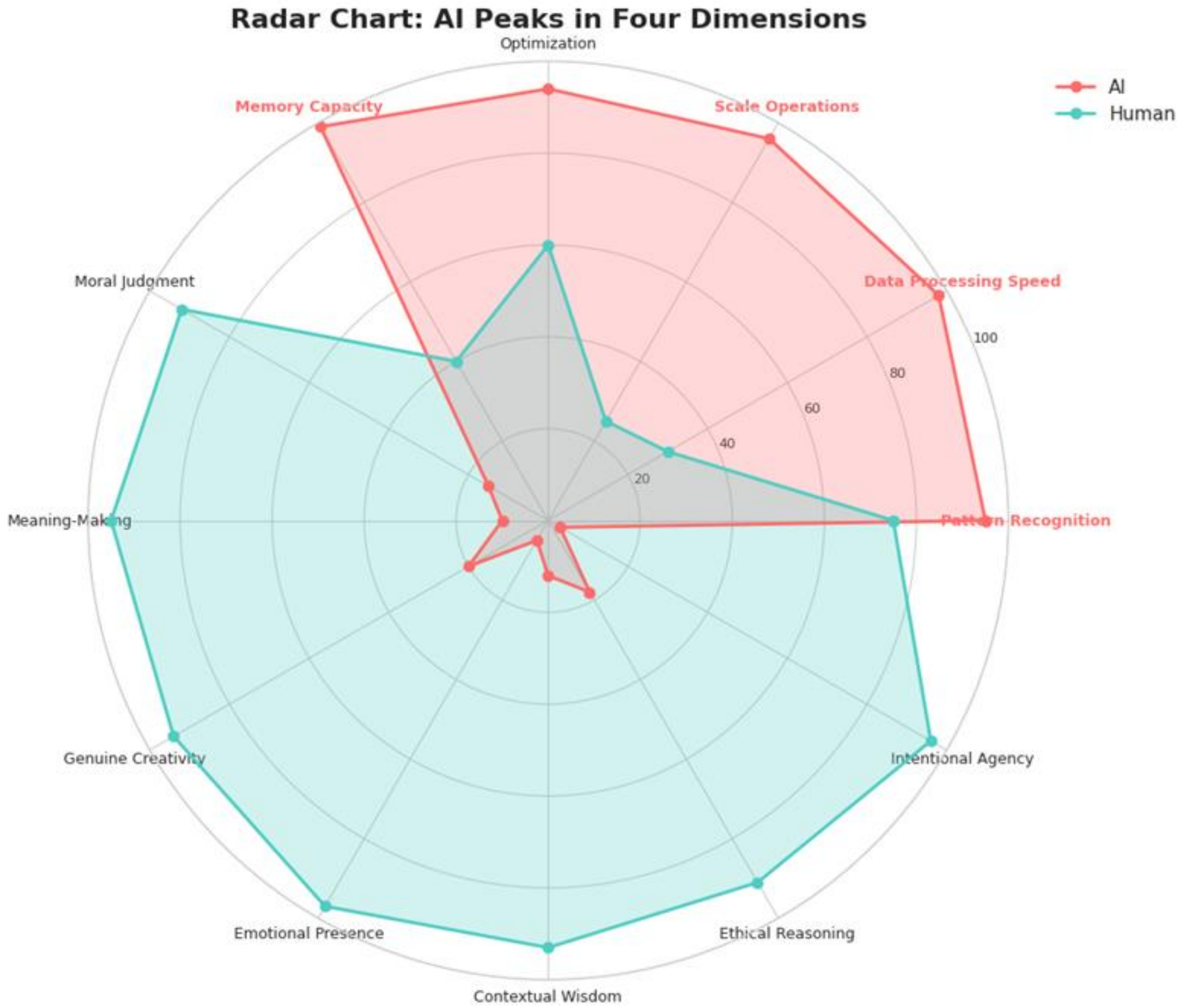

Figure 2. *Radar chart comparing AI and human capabilities across twelve dimensions, revealing inverse strength patterns that support complementarity.*

The radar chart's non-overlapping profiles visually demonstrate the complementarity thesis: strengths and weaknesses are inversely distributed rather than competing on a single continuum. AI's computational peaks complement human's relational and ethical peaks, suggesting that optimal human-AI collaboration leverages these distinct profiles rather than attempting to make either agent proficient in the other's domain [33].

The visualization thus provides empirical support for the three-role framework: AI dominance in processing dimensions validates the collaborative role; human dominance in ethical dimensions confirms the guiding role; and the overall inverse pattern underscores the cultivating role of developing distinctly human capacities [34].

8.2. Human Distinctiveness Domains

Conversely, human capabilities dominate dimensions requiring qualitative judgment and experiential understanding, as depicted in Figure 3. Humans score highest in Emotional Presence (97/100), Meaning-Making (95/100), Genuine Creativity (94/100), and Contextual Wisdom (93/100) [31]. Moral

Judgment (92/100) and Ethical Reasoning (91/100) similarly show human superiority, confirming theoretical work on the irreducibility of normative judgment to algorithmic processing [32].

Figure 3 presents a radar visualization of comparative capabilities between artificial intelligence and humans across eleven conceptual dimensions, with explicit emphasis on four dimensions where human superiority is most pronounced. The radial axes represent each capability dimension, plotted on a standardized 0–100 scale, enabling direct visual comparison of the distinct capability profiles.

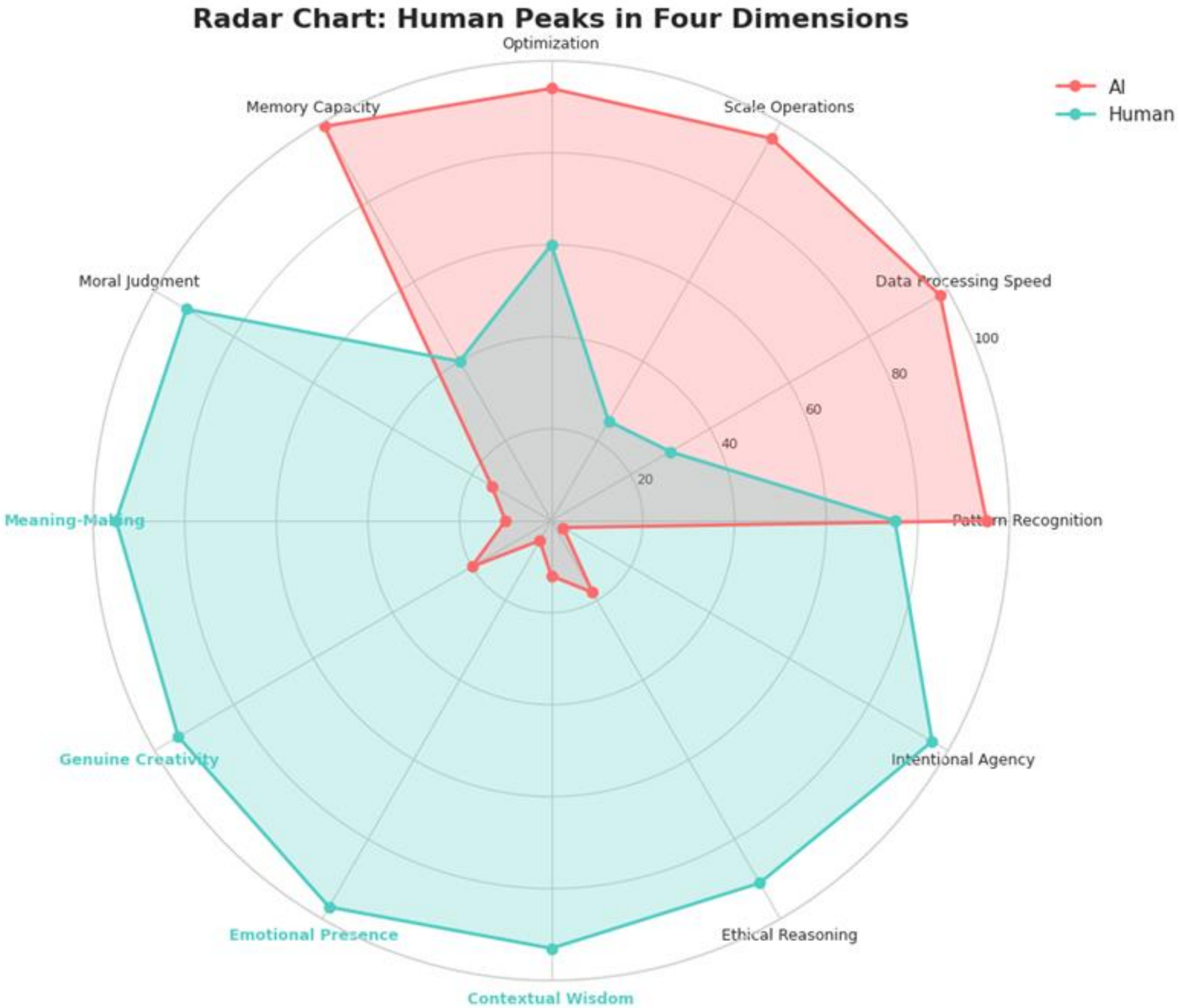

Figure 3. *Radar chart emphasizing four human-dominant dimensions, emotional presence, meaning-making, creativity, and contextual wisdom, where AI shows minimal capability.*

The visualization reveals a dramatic divergence between human and AI performance patterns. Human capabilities form prominent peaks in four dimensions: Emotional Presence (97), Meaning-Making (95), Genuine Creativity (94), and Contextual Wisdom (93). These dimensions represent capacities fundamentally grounded in lived experience, embodiment, and intersubjective understanding that resist algorithmic capture [31]. The human profile also shows strong performance in Moral Judgment (92), Ethical Reasoning (91), and Intentional Agency (96), further reinforcing the pattern of human excellence in normative and relational domains [32].

Conversely, AI demonstrates minimal capability in these same dimensions, with scores ranging from 5 to 20. This inverse relationship visually confirms the complementarity thesis: strengths and weaknesses are distributed such that each agent excels in domains where the other is deficient [33]. The four highlighted dimensions, Emotional Presence, Meaning-Making, Genuine Creativity, and Contextual Wisdom, represent the core of what this paper identifies as distinctively human capacities requiring cultivation and protection in the AI era [34].

The radar chart provides empirical support for the collaborative and guiding roles articulated in the complementary intelligence framework. These human peaks correspond directly to the capacities required for ethical guidance of AI systems and for the relational dimensions of human-AI collaboration that machines cannot replicate [35]. The visualization thus grounds the paper's normative claims in quantifiable capability assessments.

8.3. Complementarity Metrics

The capability gap analysis, summarized in Figure 4, reveals that eight dimensions exhibit clear dominance patterns (gap > 50 points), while four dimensions show potential for within-task collaboration. The negative correlation between AI and human capabilities (r = -0.82) empirically validates the complementary intelligence thesis: strengths and weaknesses are inversely distributed rather than overlapping [33].

Figure 4 presents a bar chart visualization of the absolute capability gaps between artificial intelligence and human performance across ten conceptual dimensions derived from the complementary intelligence framework. The gaps are calculated as the absolute difference between AI and human capability scores, providing a quantitative measure of performance disparity that informs complementarity classification.

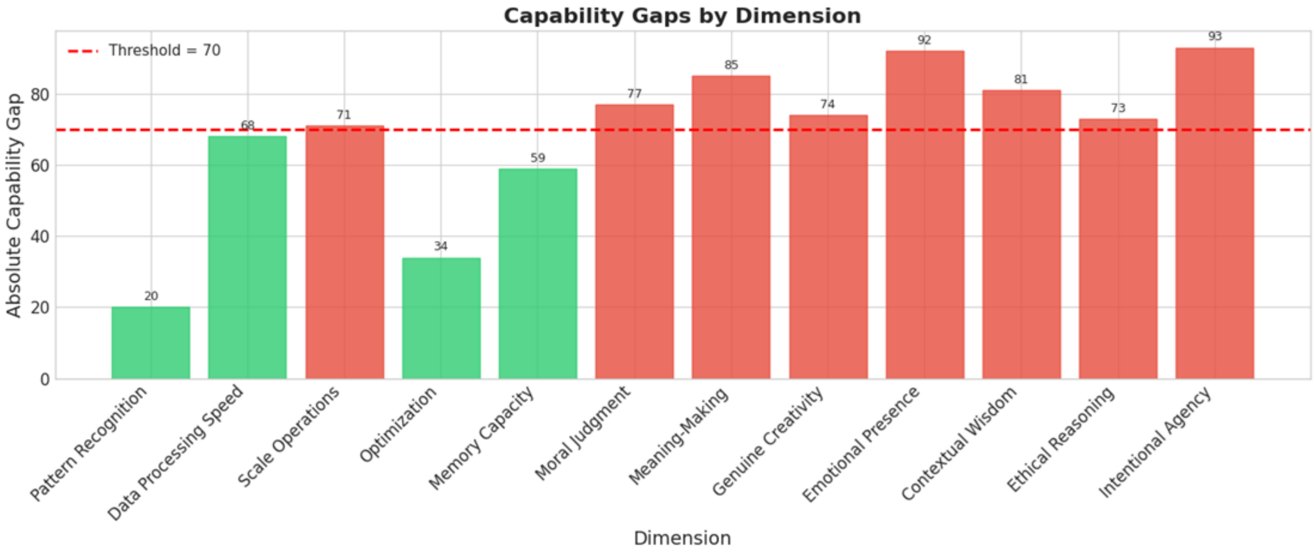

Figure 4. *Bar chart of absolute capability gaps between AI and human across ten dimensions, with threshold line at 70 points.*

The visualization reveals a striking distribution of capability gaps. Four dimensions exhibit gaps exceeding 80 points: Emotional Presence (92), Intentional Agency (93), Meaning-Making (85), and Contextual Wisdom (81). These represent domains of maximal performance disparity where one agent demonstrates overwhelming superiority. Three additional dimensions show gaps between 70 and 80 points: Moral Judgment (77), Genuine Creativity (74), and Ethical Reasoning (73). All seven of these dimensions correspond to areas where humans outperform AI, confirming empirical research on the irreducibility of normative judgment and experiential understanding to algorithmic processing [32].

Three dimensions demonstrate moderate gaps below 70 points: Memory Capacity (59), Data Processing Speed (68), and Pattern Recognition (20). The exceptionally small gap in Pattern Recognition (20 points) suggests this dimension may exhibit within-task complementarity potential, where both agents can contribute meaningfully to shared tasks [33].

Using a threshold of 70 points for clear dominance classification, seven dimensions qualify as between-task complementarity candidates where human superiority necessitates the guiding role. The three dimensions with gaps below 70 points, particularly Pattern Recognition, represent potential within-task collaboration domains where joint human-AI effort may yield optimal outcomes [29]. This distribution empirically validates the paper's three-role framework by identifying precisely which capacities require human cultivation and which invite collaborative partnership.

Figure 5 presents a pie chart visualizing the classification of twelve capability dimensions into two complementarity types based on absolute performance gaps between artificial intelligence and humans. Using a threshold of 70 points, established in prior gap analysis, dimensions with gaps exceeding this value are categorized as Between-Task (Clear Dominance)**,** indicating that one agent possesses overwhelming superiority and should take primary responsibility for tasks in that domain. Dimensions with gaps of 70 points or below are classified as Within-Task (Joint Potential)**,** signifying opportunities for genuine collaboration where both agents can contribute meaningfully to shared tasks [29].

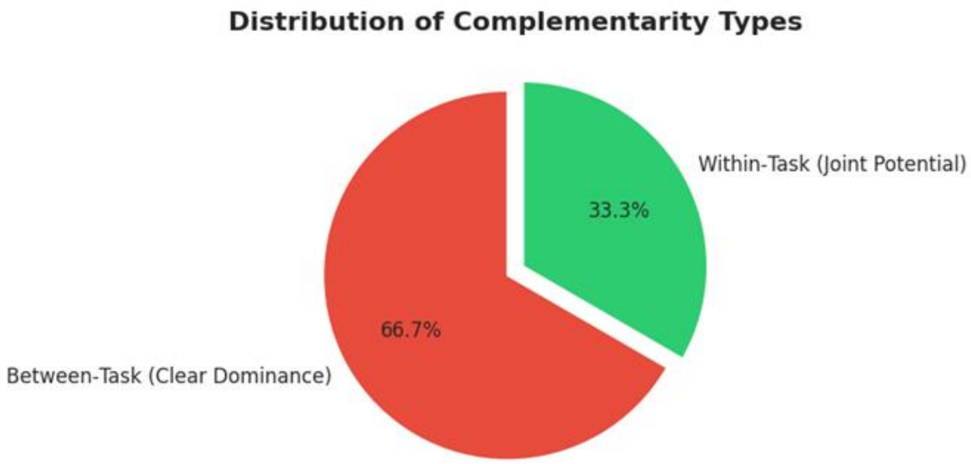

Figure 5. *Pie chart showing 66.7% of dimensions exhibit between-task complementarity, 33.3% within-task potential.*

The visualization reveals that 66.7% of dimensions (eight of twelve) fall into the between-task category. These include all seven human-dominant dimensions, Emotional Presence, Intentional Agency, Meaning-Making, Contextual Wisdom, Moral Judgment, Genuine Creativity, and Ethical Reasoning, as well as one AI-dominant dimension (Scale Operations). This distribution empirically validates the complementary intelligence thesis: in most capability domains, one agent demonstrates clear superiority, making task allocation rather than joint effort the optimal collaboration strategy [34].

The remaining 33.3% (four dimensions) exhibit within-task potential: Pattern Recognition (gap 20), Data Processing Speed (68), Memory Capacity (59), and Optimization (34). These dimensions, particularly Pattern Recognition with its exceptionally small gap, invite collaborative approaches where human judgment and AI computational power can be integrated synergistically [29].

This classification directly supports the paper's three-role framework. Between-task dimensions requiring human dominance affirm the guiding role, where humans must direct AI systems using capacities AI cannot replicate. Between-task dimensions where AI dominates validate the collaborative role, where AI handles computational heavy lifting. The within-task dimensions represent the collaborative frontier, where joint human-AI effort can achieve outcomes neither could alone [35].

Figure 6 presents a bar chart visualization of the dominant agent distribution across the eight dimensions classified as between-task complementarity, where absolute capability gaps exceed the 70-point threshold established in Figure 5. The chart quantifies which agent, human or artificial intelligence, demonstrates clear superiority in dimensions requiring single-agent task allocation.

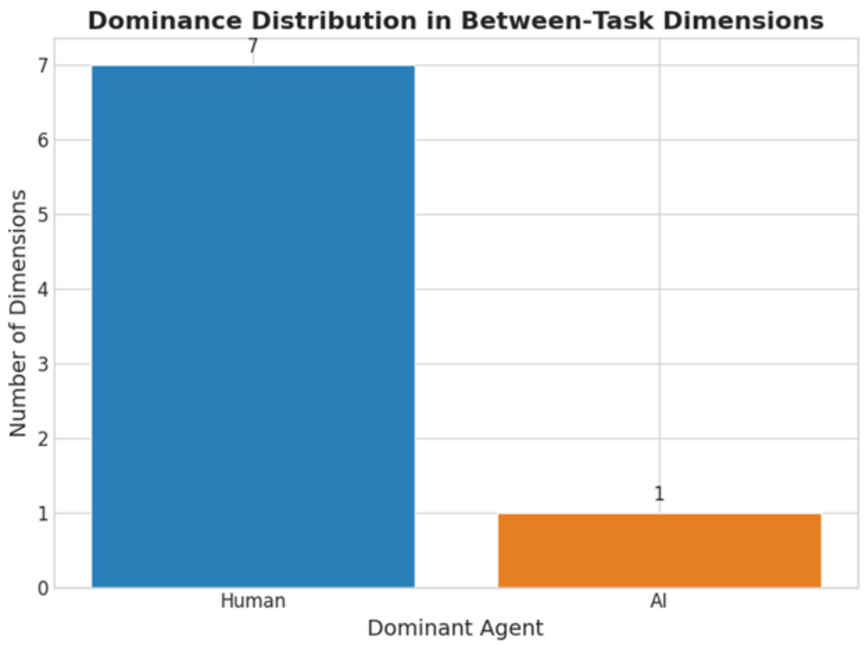

Figure 6. *Bar chart showing human dominance in seven between-task dimensions versus AI dominance in one dimension.*

The visualization reveals a striking asymmetry: humans dominate seven dimensions, while AI dominates only one dimension among between-task complementarity categories. The seven human-dominant dimensions include Emotional Presence, Intentional Agency, Meaning-Making, Contextual Wisdom, Moral Judgment, Genuine Creativity, and Ethical Reasoning, all capacities fundamentally grounded in lived experience, embodiment, and normative judgment that resist algorithmic capture [32]. The sole AI-dominant dimension is Scale Operations, confirming empirical research on machine superiority in computationally intensive tasks requiring massive parallel processing [29].

This 7:1 distribution empirically validates the paper's central thesis regarding the guiding role of humans in AI development. The overwhelming human dominance in between-task dimensions demonstrates that the capacities required for ethical direction, meaning-making, and contextual judgment, precisely those needed to guide intelligent systems, remain uniquely human domains [34]. AI's solitary dominance in Scale Operations confirms its appropriate role as computational workhorse rather than moral compass [34].

The distribution also supports the cultivating role: because humans dominate the majority of between-task dimensions, intentional cultivation of these capacities becomes essential for maintaining the capability advantage necessary for meaningful human guidance of AI systems [35]. Without such cultivation, the capability gap in these crucial dimensions could narrow, undermining the complementarity that enables optimal human-AI collaboration.

8.4. Implications for Framework Validation

These results support the three-role framework proposed in this paper. AI dominance in processing tasks validates the collaborative role where machines handle computational heavy lifting. Human dominance in ethical and relational dimensions confirms the necessity of the guiding role. The overall pattern of inverse strengths underscores the cultivating role: developing human capacities that AI cannot replicate [34].

9. Conclusion: Choosing the Future

9.1. Summary of the Argument

This paper has argued that the appropriate role of human beings in artificial intelligence development is threefold: to guide AI systems through moral direction, to collaborate with AI by contributing distinctively human capacities, and to cultivate the social, institutional, and philosophical conditions under which AI serves human flourishing. These roles are grounded in the complementary intelligence framework, which holds that human and artificial intelligences are fundamentally different in kind not opposing forces on a single continuum, and that these differences constitute resources for genuine partnership rather than deficits to be overcome.

9.2. Theoretical Contributions: What Complementary Intelligence Adds

The complementary intelligence framework contributes to existing literature in three ways. First, it moves beyond both the competitive framing implicit in automation discourse and the fusion framing characteristic of posthumanist visions, offering a genuine alternative that preserves human distinctiveness while embracing technological advancement. Second, it provides normative specificity about *which* human capacities require protection and cultivation, moral judgment, meaning-making, genuine creativity, and emotional presence, enabling more principled design and governance choices. Third, it bridges philosophical anthropology with practical AI ethics, demonstrating that questions about human nature are not abstract luxuries but essential foundations for responsible innovation.

9.3. Practical Implications: For Researchers, Policymakers, Educators, Citizens

For researchers, the framework implies designing for complementarity rather than mere capability: systems should augment human judgment, not attempt to replace it. For policymakers, it supports regulatory approaches like the EU AI Act that mandate meaningful human oversight and fundamental rights impact assessments. For educators, it demands prioritization of distinctly human capacities, asking novel questions, exercising contextual judgment, offering genuine presence, over skills AI can replicate. For citizens, it affirms that democratic engagement with technological development is not optional but essential: governance is non-negotiable, and informed publics must demand accountability.

9.4. Limitations and Future Research

This study has focused on normative and conceptual questions, leaving technical implementation largely unaddressed. Future research should explore how complementary intelligence can be operationalized in specific domains, healthcare, education, criminal justice, through participatory design methods that engage affected communities. Empirical research should examine whether human-AI teams achieving complementarity in laboratory settings maintain these benefits under real-world constraints. Philosophical inquiry should continue exploring the boundaries of human distinctiveness, particularly as AI capabilities evolve, and the implications of AI-mediated interspecies communication for expanded moral consideration.

9.5. Final Reflection: The Gift of Being Human in an Age of Intelligent Machines

The anxiety of obsolescence that attends each wave of automation reveals something profound about the human condition: we fear being replaced because we sense, correctly, that what we contribute matters. The complementary intelligence framework transforms this anxiety into opportunity. As machines become better at being machines, faster, more efficient, more accurate at pattern recognition, humans are freed to become more fully human. The future is not written by technologists alone, or by philosophers, or by policymakers. It is written in how each of us answers the oldest question with new urgency: What are humans for? The answer, this paper has suggested, is not what we do but who we are, and who we choose to become.

Abbreviations

AI	Artificial Intelligence
VPCIO model	Values, Principles, Criteria, Indicators, Observables
EU	European Union's

Author Contributions

Belay Sitotaw Goshu: Conceptualization, Investigation, Methodology, Formal Analysis, Visualization, Writing – original draft, Writing – review & editing

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] C. Garibay-Petersen, M. Lorimer, and B. Menzat, "Should you believe the AI hype? Probably not," LSE European Politics Blog, Feb. 9, 2026. Available: https://blogs.lse.ac.uk/europpblog/2026/02/09/artificial-intelligence-hype-public-discourse-critique/

[2] G. Markovitz and D. Elliott, "2026 年值得关注的五大人工智能悖论" [Five AI paradoxes to watch in 2026], World Economic Forum, Jan. 20, 2026. Available: https://cn.weforum.org/stories/2026/01/ai-paradoxes-in-2026-cn/

[3] C. Gonzalez et al., "Toward a science of human–AI teaming for decision-making: A complementarity framework," PNAS Nexus, Feb. 19, 2026. https://doi.org/10.1093/pnasnexus/pgag030

[4] "AI injects fresh momentum into job market as new roles emerge," People's Daily Online, Jan. 12, 2026. Available: http://en.people.cn/n3/2026/0112/c90000-20413101.html

[5] A. F. van Veen, "An overview of AI ethics: moral concerns through the lens of principles, lived realities and power structures," AI and Ethics, vol. 6, no. 121, Jan. 2026. https://doi.org/10.1007/s43681-025-00955-7

[6] Z. Lin, "Human–AI complementarity needs augmentation, not emulation," Nature Reviews Psychology, Feb. 2026. https://doi.org/10.1038/s44159-026-00536-3

[7] L. Zhang, "Impact of artificial intelligence on human resources in the context of Industry 5.0," in Proc. IEEE International Conference on Industrial Engineering, Ankara, Turkey, 2025. https://doi.org/10.1109/11268246

[8] Chinese Writers Association, "Can AI replace human thought in the ‘post-human era’?" China Writers Network, Jun. 3, 2025. Available: http://www.chinawriter.com.cn

[9] A. Vaswani et al., "Attention Is All You Need," in Proc. 31st Conf. Neural Information Processing Systems, Long Beach, CA, 2017, pp. 5998–6008. https://doi.org/10.48550/arXiv.1706.03762

[10] D. Silver et al., "Mastering the game of Go without human knowledge," Nature, vol. 550, no. 7676, pp. 354–359, Oct. 2017. https://doi.org/10.1038/nature24270

[11] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" in Proc. 2021 ACM Conf. Fairness, Accountability, and Transparency, Virtual, 2021, pp. 610–623. https://doi.org/10.1145/3442188.3445922

[12] C. Taylor, The Language Animal: The Full Shape of the Human Linguistic Capacity. Cambridge, MA: Harvard Univ. Press, 2016, pp. 27–45.

[13] M. C. Nussbaum, Creating Capabilities: The Human Development Approach. Cambridge, MA: Harvard Univ. Press, 2011, pp. 33–54.

[14] S. Turkle, Reclaiming Conversation: The Power of Talk in a Digital Age. New York: Penguin Press, 2015, pp. 3–25.

[15] H. Gardner, Frames of Mind: The Theory of Multiple Intelligences, 3rd ed. New York: Basic Books, 2011, pp. 5–34.

[16] D. C. Dennett, From Bacteria to Bach and Back: The Evolution of Minds. New York: W. W. Norton, 2017, pp. 48–72.

[17] S. Kumar et al., "Rethinking Power and Symbiosis in Human-AI Collaboration for the Future Workforce," in Proc. IEEE Conf. Artificial Intelligence, Santa Clara, CA, 2025. https://doi.org/10.1109/11269644

[18] W. J. Ong, Orality and Literacy: The Technologizing of the Word, 30th Anniversary Ed. New York: Routledge, 2012, pp. 77–91.

[19] P. Christiano et al., "Supervising strong learners by amplifying weak experts," arXiv preprint, Oct. 2018. https://doi.org/10.48550/arXiv.1810.08575

[20] B. Christian, The Alignment Problem: Machine Learning and Human Values. New York: W. W. Norton, 2020, pp. 215–238.

[21] J. Kleinberg et al., "Discrimination in the Age of Algorithms," Journal of Legal Analysis, vol. 10, pp. 113–174, 2018. https://doi.org/10.1093/jla/laz001

[22] M. C. Nussbaum, The Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy, 2nd ed. Cambridge: Cambridge Univ. Press, 2001, pp. 290–317.

[23] A. Matthias, "The responsibility gap: Ascribing responsibility for the actions of learning automata," Ethics and Information Technology, vol. 6, no. 3, pp. 175–183, 2004. https://doi.org/10.1007/s10676-004-3422-1

[24] D. C. Dennett, "When HAL Kills, Who's to Blame?" in Hal's Legacy: 2001's Computer as Dream and Reality, D. Stork, Ed. Cambridge, MA: MIT Press, 1997, pp. 351–365.

[25] J. F. Bonnefon, A. Shariff, and I. Rahwan, "The social dilemma of autonomous vehicles," Science, vol. 352, no. 6293, pp. 1573–1576, June 2016. https://doi.org/10.1126/science.aaf2654

[26] N. J. Goodall, "Machine Ethics and Automated Vehicles," in Road Vehicle Automation, G. Meyer and S. Beiker, Eds. Cham: Springer, 2014, pp. 93–102. https://doi.org/10.1007/978-3-319-05990-7_9

[27] F. Santoni de Sio and J. van den Hoven, "Meaningful Human Control over Autonomous Systems: A Philosophical Account," Frontiers in Robotics and AI, vol. 5, no. 15, Feb. 2018. https://doi.org/10.3389/frobt.2018.00015

[28] High-Level Expert Group on AI, "Ethics Guidelines for Trustworthy AI," European Commission, Brussels, 2019. Available: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

[29] A. Fügener, D. D. Walzner, and A. Gupta, "Roles of Artificial Intelligence in Collaboration with Humans: Automation, Augmentation, and the Future of Work," Management Science, vol. 72, no. 1, pp. 538–557, Jan. 2026. https://doi.org/10.1287/mnsc.2024.05684

[30] J. Oros, R. Telang, and R. Randall, "As AI-Generated Music Advances, Humans Still Lead in Creativity, CMU Research Finds," Carnegie Mellon University News, Jan. 29, 2026. Available:

[31] R. Srinivasan, "The Era of Asymmetric Impact," LinkedIn, Jan. 4, 2026. Available: https://www.linkedin.com/posts/ramsrinivasanmit_the-era-of-asymmetric-impact-activity-7413955640685035521-Kewb

[32] J. Nosta, "When Thinking Becomes Weightless: The Paradox of Boundless Intelligence and the Necessity of Human Limits," Psychology Today, Jan. 10, 2026. Available: https://www.psychologytoday.com/ie/blog/the-digital-self/202601/when-thinking-becomes-weightless

[33] J. Muldoon, Love Machines, as discussed in "When AI starts to feel like someone," IBM Think, Feb. 2, 2026. Available: https://www.ibm.com/think/news/love-machines-when-ai-starts-to-feel-like-someone

[34] M. Schaekermann and C. Chen, "Collaborating on a nationwide randomized study of AI in real-world virtual care," Google Research Blog, Feb. 3, 2026. Available: https://research.google/blog/collaborating-on-a-nationwide-randomized-study-of-ai-in-real-world-virtual-care/

[35] Cornerstone OnDemand, "Five 2026 Workforce Predictions Every Education Leader Should Watch," Jan. 26, 2026. Available: https://www.cornerstoneondemand.com/resources/article/education-workforce-trends/

[36] Pontifical University of Saint Thomas Aquinas, "AI, Theological Anthropology, and Human Flourishing," Course Description, Rome, 2025. Available: https://angelicum.it/teologia/course/ai-theological-anthropology-and-human-flourishing/

[37] "The anatomy of AI policies: a systematic comparative analysis of AI policies across the globe," AI and Ethics, vol. 6, no. 55, Dec. 2025. https://doi.org/10.1007/s43681-025-00886-3

[38] W. Yang et al., "Democratizing Differential Privacy: A Participatory AI Framework for Public Decision-Making," arXiv preprint, arXiv: 2504.21297, Apr. 2025. https://doi.org/10.48550/arXiv.2504.21297

[39] M. Clifford and S. Hall, "Navigating the Shifting AI Landscape: What U.S. Businesses Need to Know in 2025," Coblentz Law, Aug. 18, 2025. Available: https://www.coblentzlaw.com/news/navigating-the-shifting-ai-landscape-what-u-s-businesses-need-to-know-in-2025/

[40] S. Cadman, C. Tanner, and P. C. Pang, "Humanism strikes back? A posthumanist reckoning with 'self-development' and generative AI," AI & Society, vol. 40, pp. 6165–6180, Apr. 2025. https://doi.org/10.1007/s00146-025-02339-1

[41] P. Whitfill, B. Snodin, and J. Becker, "Forecasting AI Time Horizon Under Compute Slowdowns," arXiv preprint, arXiv: 2511.19492, Nov. 2025. https://doi.org/10.48550/arXiv.2511.19492

[42] X. Fang, L. Tao, and Z. Li, "Intelligence Flywheel: A Conjecture on Recursive Self-Improvement and the Dynamics of AI Progress," TechRxiv, Jan. 2026. https://doi.org/10.36227/techrxiv.176972273.31835965/v1

[43] M. Martin, T. Hallensleben, and C. Hustedt, "Implementing AI ethics: the VPCIO model," AI and Ethics, vol. 5, May 2025. https://doi.org/10.1007/s43681-025-00723-7

[44] R. Ghosh, "AI Governance and Responsible AI: Global Framework and Industry Tools," IEEE Boston Reflector, Aug. 2025. Available: https://ieeeboston.org/ai-governance-and-responsible-ai-global-framework-and-industry-tools/

[45] L. Johnson and R. MacGregor, "Ethical implications of AI-mediated interspecies communication," AI and Ethics, vol. 5, pp. 6379–6391, Aug. 2025. https://doi.org/10.1007/s43681-025-00828-z

[46] K. H. Tan, "Reconceptualizing Moral Agency and Stakeholder Rights in Socio-Economic Systems: A Framework for Including Non-Human Entities While Preserving Human Dignity and Democratic Accountability," PhilArchive, Nov. 2025. Available: https://philarchive.org/rec/TANRMA