VOL. 10 NO. 9,
NOVEMBER, 2025
E-ISSN 3026-958X
P-ISSN 3027-1169

African Scholar Publications & Research International
www.africanscholarpub.com

Journal of Science Innovation & Technology Research (JSITR)

Artificial Intelligence and the Essence of Humanity: Strategic Frameworks for Utilizing Technology and Preserving Values in an Automated Era

Simeon Ayoade Adedokun; Isiak Akinwumi Adeyemo; Dorcas Atinuke Adedokun; Stella Kehinde Ogunkan; & Olufunke Kemi Ogunniyi

Department of Computer Science, Ladoke Akintola University of Technology, Ogbomoso, Nigeria

Corresponding Author: simadedokun@gmail.com

DOI: https://doi.org/10.70382/ajsitr.v10i9.069

Abstract

Artificial Intelligence (AI) constitutes an epochal shift, profoundly transforming human experience but simultaneously raising urgent ontological and ethical challenges, including cognitive atrophy, algorithmic bias, and truth decay. This research conducted an integrative literature review to synthesize the complex human-AI relationship through established theoretical lenses. Specifically, thematic findings on technological risks and human attributes are mapped against Socio-Technical Systems (STS) theory, which mandates the joint optimization of social and technical subsystems, and Value Sensitive Design (VSD), which requires the proactive embedding of human values into system architecture. The study further examined global AI ethics and governance instruments, including the EU AI Act and UNESCO's Recommendation on the Ethics of AI, and identified a governance gap wherein macro-level regulations do not translate into granular, user-centric operational guidance. The study established a critical distinction, stating that AI operates on syntactic, probabilistic logic, while the human essence is defined by semantics, moral agency, and intrinsic purpose. Addressing the gap between technical utility and ontological preservation, the paper proposes two actionable frameworks, namely the 5-P Strategic Framework for fostering critical AI literacy and the V.A.L.U.E. Framework for guiding ethical AI utilization. These models offer stakeholders a strategic roadmap for adopting AI responsibly, ensuring that human agency and enduring values are protected in an increasingly automated environment.

Keywords: Artificial Intelligence, Generative AI, AI Ethics, AI Stewardship, 5-P Framework, V.A.L.U.E. Framework, Human Value, Value Sensitive Design, Human-in-the-Loop, Technological Stewardship

INTRODUCTION

Artificial Intelligence (AI) has transcended its origins as a niche field of computer science to become the defining technological force of the 21st century. It is reshaping the fundamental structures of human interaction, economic production, and knowledge dissemination. No longer confined to academic laboratories or Silicon Valley servers, AI has permeated the daily lived experience of the global population, from adaptive learning platforms that personalize education to autonomous systems revolutionizing logistics, and Generative AI models like ChatGPT and Gemini that challenge the monopoly of human creativity (Russell & Norvig, 2022). As we stand at the threshold of this new era, AI functions not merely as a tool for efficiency but as an environmental condition that actively shapes human worldviews, ethical choices, and the very nature of truth.

However, the rapid integration of AI presents a profound paradox, in that it offers unprecedented opportunities to address complex global challenges while simultaneously posing significant existential risks (Duggal, 2021). The capability of machines to simulate reasoning, empathy, and artistic creation has blurred the once-sharp distinction between human agency and machine output. This ontological blurring raises urgent, foundational questions like, what does it mean to remain authentically human in an age of intelligent machines? How can society harness the computational power of AI without compromising the moral and intrinsic values that define the human essence?

The implications for the future workforce and broader society are stark. Technical proficiency in coding or data science, while valuable, is no longer sufficient. Research indicates that the workforce of the future must cultivate durable human skills, such as critical thinking, adaptability, ethical reasoning, and emotional intelligence, that remain beyond the reach of algorithmic automation (Gronning & Borggreven, 2025). Consequently, the central imperative for the modern professional and the ethical citizen has shifted from mere adoption to active stewardship. It is no longer enough to ask if we will use AI, but how we will govern its influence over our lives. This paper aims to address this critical juncture by demystifying the operational mechanics of AI, analyzing its impact on the essence of humanity, and proposing strategic frameworks to guide responsible, human-centric AI stewardship.

MATERIALS AND METHODS

Research Design

This study employed an integrative literature review methodology (Torraco, 2005; Whittemore & Knafl, 2005). This approach enabled the synthesis of diverse literature

streams, including theoretical, conceptual, and empirical literature, to generate new conceptual frameworks. Given the interdisciplinary nature of the research problem, an integrative review provides the necessary flexibility to combine technical definitions of AI with philosophical inquiries into human ontology. The research was conducted in five distinct stages, visualized in Figure 1, which are problem identification, literature search, data evaluation, thematic analysis and synthesis, and framework formulation.

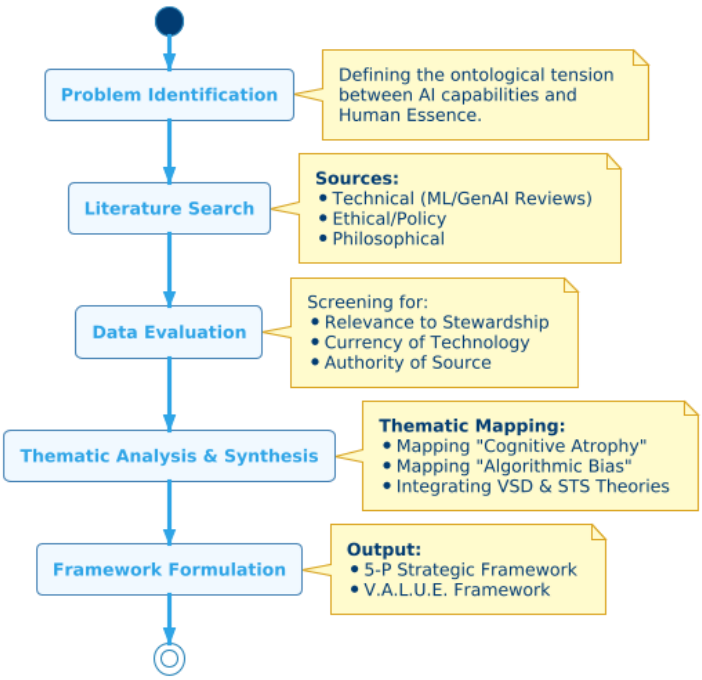


Figure 1: Stages of the Research and the Integrative Review
Source: Authors' Work (2025)

Problem Identification Stage

This is the initial stage, involving the definition of the core research problem, which is the tension between the rapid proliferation of AI capabilities and the preservation of the essence of humanity, defined as moral agency, creativity, and human identity. The review sought to answer the question, "What frameworks can bridge the gap between technical AI utility and human ethical preservation?"

Literature Search Strategy

Subsequently, a comprehensive literature search strategy was employed, utilizing Boolean strings—including "Artificial Intelligence" AND ("Ethics" OR "Human Values"), "Generative AI" AND "Education", "AI Stewardship", "Human Identity AND Technology", and "Algorithmic Bias"—across academic databases, including Google Scholar, IEEE Xplore, PubMed, ResearchGate, Web of Science, and Scopus; and authoritative organizational

repositories, including UNESCO, European Union, McKinsey Global Institute, National Academies of Sciences, PwC, and The Holy See.

Data was sourced from technical literature, ethical and policy guidelines, philosophical literature on human ontology, and theoretical frameworks, including Value Sensitive Design (VSD), Socio-Technical Systems (STS) theory, and Human-in-the-Loop (HITL) models. Data sources were limited to peer-reviewed journal articles, conference proceedings, academic books, and authoritative organizational reports. The search was heavily focused on the timeframe of 2021 to 2025 to ensure the inclusion of the latest advancements and ethical discourse surrounding Generative AI and current ethical guidelines. However, seminal works and fundamental ethical theories were also considered, regardless of their publication date. The synthesis of these sources informed the development of the strategic frameworks presented in the results section.

Data Evaluation

The data evaluation phase applied inclusion and exclusion criteria, prioritizing peer-reviewed articles, seminal books, and high-impact industry reports that explicitly addressed the societal and ethical implications of AI, while excluding purely technical benchmarks or literature published before 2020 that do not account for the generative AI model shift. Priority was given to publications from established academic publishers and journals with high impact factors in relevant fields such as artificial intelligence, computer science, educational technology, information ethics, and humanities. Sources were evaluated for authority, currency, and relevance to the concept of technological stewardship.

Thematic Data Analysis and Synthesis

The selected literature underwent thematic synthesis. This involved systematic extraction and categorization of findings using thematic coding. The findings were categorized into primary themes encompassing technological mechanics, ethical risks, and human attributes. Subsequently, conceptual mapping, illustrated in Figure 2, was utilized to identify relationships between AI capabilities and human vulnerabilities, and align these themes with the theoretical pillars of STS and VSD. This mapping process facilitated the identification of theoretical gaps, particularly the lack of user-centric stewardship protocols, which directly informed the design of the 5-P strategic framework and the V.A.L.U.E. framework.

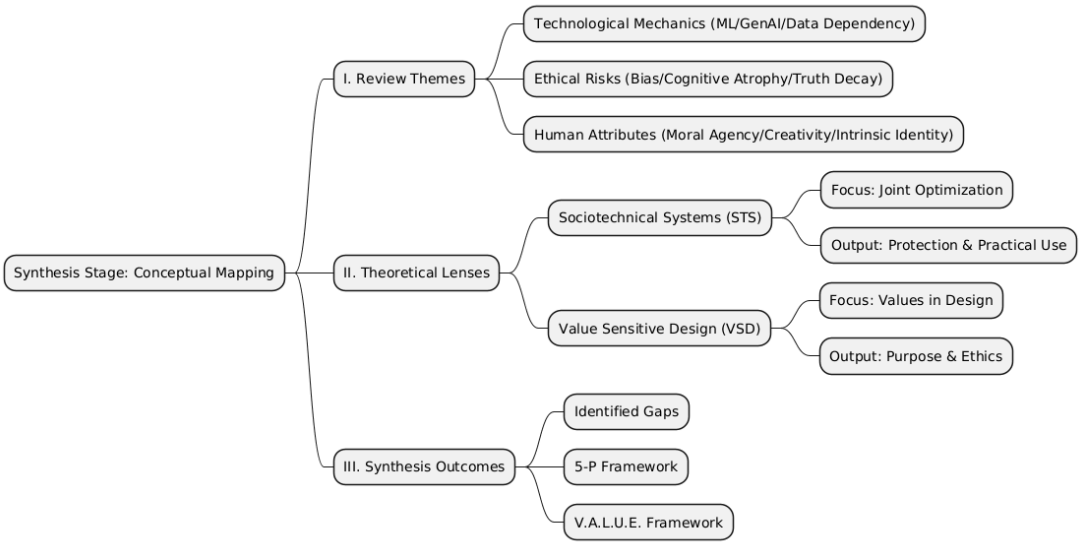


Figure 2: Conceptual Mapping of Findings
Source: Authors' Work (2025)

Methodological Limitations

The integrative review design employed for this study prioritized high-impact databases and recent literature from 2021 to 2025, which may be susceptible to selection bias. In addition, while the proposed frameworks are theoretically grounded, they currently lack longitudinal empirical validation in diverse real-world settings. This research serves as the foundational conceptual model intended for subsequent empirical testing.

Theoretical and Conceptual Foundations

To construct robust frameworks for AI stewardship, this study grounds its proposals in established theoretical perspectives on technology and ethics.

The Essence of Humanity: An Ontological Distinction

The core premise of this research is that AI, regardless of its increasing sophistication, remains fundamentally and ontologically distinct from human intelligence. This distinction is best understood through the philosophical divide between syntax and semantics. AI systems, particularly Large Language Models, function exclusively on a syntactic level; they are probabilistic engines that process patterns, correlations, and rules within vast datasets to predict the most likely next token (Russell & Norvig, 2022). They excel at statistical mimicry and pattern recognition, demonstrating high-fidelity outputs without any inherent connection to meaning or understanding. In contrast, human cognition is inherently semantic, driven by consciousness, intentionality, and the subjective assignment of meaning to information.

This critical computational divergence leads to an undeniable ontological distinction rooted in the transcendent and moral capabilities of humanity. AI is driven by mathematical

optimization and processing, seeking the most efficient pathway to a defined, external goal. Humanity, however, is driven by moral reasoning, and intrinsic purpose and identity—qualities that emerge from lived experience and intentionality, which cannot be modeled simply as computational objectives. Furthermore, humanity possesses inherent qualities of moral agency and intrinsic dignity that fundamentally position the human being as a responsible steward (Dicastery for the Doctrine of the Faith, 2025). Moral reasoning requires contextual empathy, qualitative judgment, and a conscious commitment to values, all of which are external to an AI's operational logic (Jonnalagadda et al., 2025; Shu et al., 2025). While machines can simulate compassion through textual output, they cannot genuinely feel compassion. Therefore, recognizing and preserving this distinction is the ultimate mandate behind the proposed frameworks for human-centric AI stewardship.

Societal Impact and Ethical Challenges

The deployment of Artificial Intelligence brings a wave of transformative benefits across various sectors, even as it demands responsible governance. The democratization of knowledge is one of AI's most significant impacts. AI-powered personalized tutoring systems provide adaptive learning platforms for students with around-the-clock academic support, automatic grading tools, learning analytics, and explaining complex concepts tailored to the learner's pace (Farhood et al., 2025). In addition, AI acts as a super-librarian for researchers, capable of synthesizing vast datasets and academic papers in seconds, thus accelerating the pace of discovery and learning (Oyebanji et al., 2025).

In the domain of healthcare, AI systems are dramatically enhancing diagnostic accuracy and accelerating critical processes such as drug discovery, leading the way toward a more precise and personalized form of medicine (Adedokun, 2024; Fu & Chen, 2025; Ugbor et al., 2025). Economically, AI serves as a powerful engine for productivity, facilitating the automation of routine and repetitive tasks. This fundamental function allows organizations to free human workers from drudgery, enabling them to shift their focus toward higher-order cognitive functions, creativity, and strategic decision-making (McKinsey Insight, 2025). Beyond industrial benefits, AI systems significantly contribute to accessibility by providing enhanced and tailored communication tools for individuals with diverse disabilities, fostering greater inclusion across society.

Conversely, the pervasive nature of AI introduces profound ethical and societal risks that threaten human capabilities and systemic integrity. A significant concern revolves around cognitive atrophy, where an excessive over-reliance on AI for decision-making and content generation may stymie the development of essential human skills, such as critical thinking and complex problem-solving (Gerlich, 2025). This challenge is compounded by the persistent issue of bias and inequality, as AI models, trained on historical data sets that often reflect societal prejudices, can reproduce and even accelerate systemic inequalities in real-world outcomes (Adedokun et al., 2025). Furthermore, the unchecked proliferation of Generative AI has given rise to truth decay, where highly convincing deepfakes and synthesized information

erode public trust, threaten democratic integrity, and undermine the communal understanding of verifiable facts (Weigand, 2025). Lastly, the relational impact of AI presents a psychological risk, with some studies raising concerns that artificial intimacy derived from emotionally responsive chatbots may inadvertently erode the capacity for, or preference for, genuine and complex human connection (American Psychological Association, 2025). These challenges necessitate the formulation of robust strategic frameworks designed to protect human ontology.

Socio-Technical Systems Theory

Socio-Technical Systems (STS) theory, rooted in the foundational work of the Tavistock Institute of Human Relations, posits that any organized system consists of an interconnected social subsystem and a technical subsystem. The central and defining principle of STS is joint optimization, which dictates that optimal system performance is only achieved when both the social and technical components are designed and optimized together, rather than separately (Knox, 2024; Trist & Bamforth, 1951). STS views organizations as a dynamic interdependent pairing of people and technology, recognizing that changes in one domain invariably affect the operation of the other. This interaction determines operational methods and overall system effectiveness (Husband, 1978).

AI systems are not isolated technical artifacts; they are sociotechnical ensembles embedded in organizational and social contexts. Recent literature applies this lens to the modern AI environment, arguing that treating sophisticated AI models merely as isolated technical tools often leads to organizational failure, human alienation, and a degradation of work quality (Baxter & Sommerville, 2019; Kudina & Van De Poel, 2024). STS theory posits that the optimization of a system necessitates the joint optimization of both the social subsystem, encompassing people, roles, values, culture, and processes; and the technical subsystem, comprising tools, devices, algorithms, and infrastructure.

In the context of AI, the technical subsystem encompasses large language models, proprietary datasets, and computational infrastructure. In contrast, the social subsystem involves user interaction protocols, ethical norms, governance structures, and the cultivation of human expertise. A critical STS failure occurs when technical efficiency is prioritized at the expense of the social subsystem, such as in cases of unchecked algorithmic management that leads to worker burnout or the over-reliance on AI that fosters cognitive atrophy. This contemporary application of STS emphasizes the concept of Human-AI Teaming (HAT), where human roles must be fundamentally redesigned not just for task execution but for ethical monitoring, error correction, and maintaining situational awareness—thereby optimizing the social environment for human flourishing alongside technical performance (Ang et al., 2024; Berretta et al., 2023; Schmutz et al., 2024).

The STS analysis emphasizes the necessity of embedding protection and practical use within the proposed 5-P framework, ensuring that AI serves to augment, rather than destabilize, human social structures. This perspective supports the proposed frameworks by emphasizing

that successful AI adoption ("Practical Use") cannot occur without addressing social implications ("Protection" and "Purpose").

Value Sensitive Design (VSD)

Value Sensitive Design (VSD) provides a robust, proactive methodological approach and foundational theory for integrating human values into the design of technological systems. VSD's foundational premise is that technology is not value-neutral; rather, it is inherently moral, promoting or suppressing specific human values based on its architecture and deployment context (Hendry et al., 2021).

VSD operates through a tripartite methodology that systematically connects abstract values to concrete technical requirements. The methodological phases are conceptual, empirical, and technical investigations. The conceptual investigations involve systematically identifying relevant direct and indirect stakeholders, and those affected by proxy; and the values, which include privacy, autonomy, fairness, human dignity, and trust that are explicitly or implicitly implicated by the technology (Hanna et al., 2025; Jonnalagadda et al., 2025; Shu et al., 2025). For AI, this stage is critical for proactively defining what "fairness" or "transparency" means in a specific context before the algorithm is deployed. The empirical investigations' phase employs social science methods, including interviews, surveys, and ethnographic studies, to understand how stakeholders actually perceive, experience, and value the technology in real-world use, and to identify potential harms. This step is vital for uncovering unintended consequences, such as the disproportionate impact of a predictive policing model on a specific demographic, which is often missed during purely technical testing (Sadek & Mougenot, 2025).

The findings from both the conceptual and empirical phases then feed into the technical investigations. In this phase, the challenge is to translate the identified ethical requirements, such as minimizing bias or ensuring data privacy, into specific system requirements, constraints, and features to embed the identified values directly into the algorithms, interfaces, and overall functionality (Hendry et al., 2021). This could involve integrating differential privacy techniques, implementing interpretability tools, or designing user interfaces that clearly communicate algorithmic confidence levels.

The application of VSD has become particularly vital for mitigating AI ethics concerns, especially due to the non-transparent nature of deep learning models and the rapid propagation of bias. Recent adaptations of VSD principles for AI emphasize the need to explicitly identify values such as autonomy, algorithmic fairness and accountability, and human welfare during the design and deployment phases, advocating for design-time interventions rather than post-hoc regulation (Jonnalagadda et al., 2025; Sadek & Mougenot, 2025). The systematic identification of values such as non-discrimination, human dignity, and trustworthiness during the model training and deployment stages enables VSD to serve as the theoretical foundation for the proposed V.A.L.U.E. framework. It mandates that technological utility must always be subservient to human dignity.

Human-in-the-Loop (HITL) and Stewardship Models

The necessity of retaining human expertise and moral judgment in automated decision-making processes is encapsulated by the technical concept of Human-in-the-Loop (HITL) and the philosophical concept of stewardship. The concept of Human-in-the-Loop (HITL) asserts that automated systems require human oversight to ensure accuracy and moral accountability, particularly in high-stakes domains (Johnson & Hoover, 2024; Stryker, 2025). The fundamental goal is to capitalize on AI's efficiency for pattern recognition while reserving for humans the final, context-specific decision that requires moral reasoning, empathy, and qualitative judgment.

The technical concept parallels the ethical and philosophical model of stewardship, which frames humans as responsible caretakers of resources. The model asserts that humans are morally responsible custodians of the resources and tools they possess, including technology. This framework rejects a laissez-faire approach to innovation and mandates active, ethical governance. Stewardship serves as the critical counterweight to automation bias, that is, the pervasive cognitive tendency to blindly trust or uncritically accept the output of an automated system, simply because it is a machine (Ma et al., 2023). Both HITL and the stewardship model advocate for a partnership model where AI is explicitly used to augment human capabilities rather than replace human judgment. The frameworks mandate that the user Verify (ensure stewardship) and Augment (Human-in-the-Loop) the AI's probabilistic results, thereby guaranteeing that the human retains the final ontological responsibility, holding the final authority over the outcome. This integration provides the theoretical and ethical basis for the core action protocols of the proposed V.A.L.U.E. framework.

The EU AI Act and UNESCO Global Frameworks

The UNESCO Recommendation on the Ethics of Artificial Intelligence represents a landmark global consensus, providing a comprehensive values-based foundation centered on human rights, environmental sustainability, and inclusivity (UNESCO, 2021). Similarly, the European Union Artificial Intelligence Act introduces a pioneering, risk-based regulatory hierarchy that classifies AI applications from unacceptable to low risk, mandating strict transparency and human oversight for high-stakes systems (European Commission, 2025; European Union, 2024). However, the analysis of these frameworks reveals a significant top-down gap. While they are excellent for legislative governance and developer compliance, they offer limited guidance for the individual user or practitioner navigating day-to-day AI interactions. These frameworks focus on what institutions should regulate, but not what individuals should do. Therefore, there is a palpable lack of a bottom-up protocol that builds personal moral competence or provides a strategic step-by-step roadmap for users to integrate AI into their workflows without succumbing to cognitive atrophy or automation bias.

Conceptual Frameworks for AI Literacy

The contemporary discourse on human-AI interaction increasingly emphasizes the necessity

of critical AI literacy, marking a significant evolution beyond mere functional competence. Traditionally, digital literacy focused on the instrumental ability to operate software; however, the emergence of opaque, probabilistic systems like Machine Learning and Generative AI demands a deeper, ontological understanding (Pretorius & Cahusac De Caux, 2025). This shift is crucial because AI's inherent opacity and persuasive capabilities can lead to automation bias, which erodes critical thinking (Gerlich, 2025). AI literacy functions as a vital mechanism for democratic engagement with these powerful tools. It is defined as a multi-dimensional competence that extends far beyond simple tool usage, encompassing the ability to understand how AI systems work, how they impact society, and how they should be ethically governed (Digital Education Council, 2025).

This multi-dimensional competence is typically categorized into three interwoven domains, which directly inform the architecture of the proposed frameworks. The first domain is foundational understanding, which requires citizens to grasp the basic mechanics of AI, including data input, model training, and probabilistic output, to separate algorithmic reality from technical hype. This directly aligns with the "Perception" phase of the 5-P framework. The second domain, critical evaluation, involves assessing the validity, inherent biases, and potential for harm in AI-generated outputs, recognizing that algorithms reproduce societal inequities present in their training data (Adedokun et al., 2025). This skill is foundational to the "Verify" protocol in the V.A.L.U.E. framework, emphasizing the crucial need for grounding knowledge in truth over relying on convenience. Finally, ethical application or civic competence focuses on using AI responsibly, rejecting deceptive uses like plagiarism or deepfakes, and ensuring compliance with human values. This third domain maps precisely to the "Understand (Ethics)" protocol, reinforcing the mandate for ethical stewardship. Consequently, these frameworks position AI literacy not as a static skill set but as a continuous process of upskilling necessary to prevent cognitive stagnation and maintain human moral agency in an accelerating technological environment.

RESULTS AND DISCUSSION

Synthesizing the reviewed sociotechnical and ethical literature, this study proposes two frameworks to guide the human-AI relationship.

The 5-P Strategic Framework for Understanding AI

This framework, illustrated in Figure 3, is designed to move users from passive consumption to active comprehension, grounded in AI Literacy principles. The 5-P strategic framework mandates a human-centric approach to AI literacy by requiring users to progress from basic [1]Perception and [2]Preparation of durable human skills to applying the technology for responsible [3]Practical Use, all while actively ensuring the [4]Protection of dignity and grounding every application in a moral [5]Purpose.

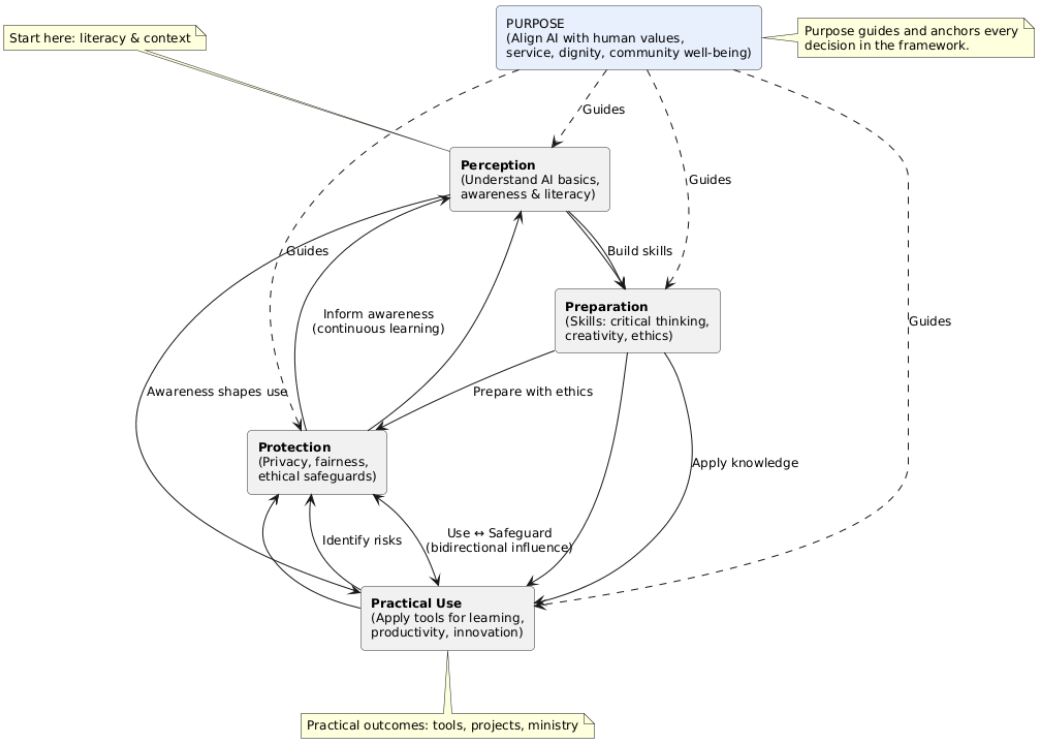


Figure 3. 5-P Strategic Framework for Understanding AI
Source: Authors' Work (2025).

Perception

The initial principle mandates the development of fundamental AI literacy, enabling users to understand the basic mechanics of how probabilistic systems like Generative AI operate, and importantly, to distinguish algorithmic reality from speculative hype. This is directly supported by the foundational understanding domain within the conceptual frameworks for AI literacy, which argues that informed citizenship in the AI era begins with accurate knowledge of the technology's capabilities and limitations.

Preparation

Preparation focuses on cultivating durable human skills, such as critical thinking, ethical reasoning, and emotional intelligence, which are uniquely human and beyond the reach of algorithmic automation. This principle is grounded in the stewardship model, which views humans as responsible custodians of technology. Prioritizing these skills is essential for fortifying the social subsystem (STS), preventing cognitive atrophy, and maintaining human moral agency as the foundation of the human-AI partnership.

Practical Use

The third principle encourages applying AI constructively for specific, bounded outcomes rather than adopting a posture of general, uncritical reliance. This aligns directly with the

technical concept of Human-in-the-Loop (HITL), which advocates for AI to augment human capabilities rather than replacing core judgment. From an STS perspective, this ensures that the technical tool is properly integrated into human workflows for joint optimization, serving human goals.

Protection

Protection focuses on the defensive responsibility of upholding privacy, human dignity, and fairness in all AI interactions. This principle is a direct implementation of the core concerns of Value Sensitive Design (VSD). VSD's methodology requires proactive identification of values like fairness and autonomy, which must be embedded as constraints on the technological subsystem to mitigate risks like bias and the erosion of human dignity.

Purpose

This final, overarching principle insists that all technological use must be anchored in human values and community well-being. Purpose serves as the ethical high point of the framework, drawing foundational strength from both VSD and the Stewardship Model. It mandates that technological capability must be subservient to human dignity, ensuring that AI is leveraged not for mere efficiency, but to improve the common good and reinforce the ethical foundations and moral essence of humanity.

The V.A.L.U.E. Framework for Utilizing AI

The V.A.L.U.E. framework, illustrated in Figure 4, is a Human-in-the-Loop protocol for interaction designed to ensure users maintain intellectual agency and ethical accountability by requiring them to [1]Verify AI outputs, [2]Augment rather than replace their role, continually [3]Learn technical skills, [4]Understand the ethical implications, and leverage AI to [5]Edify the common good.

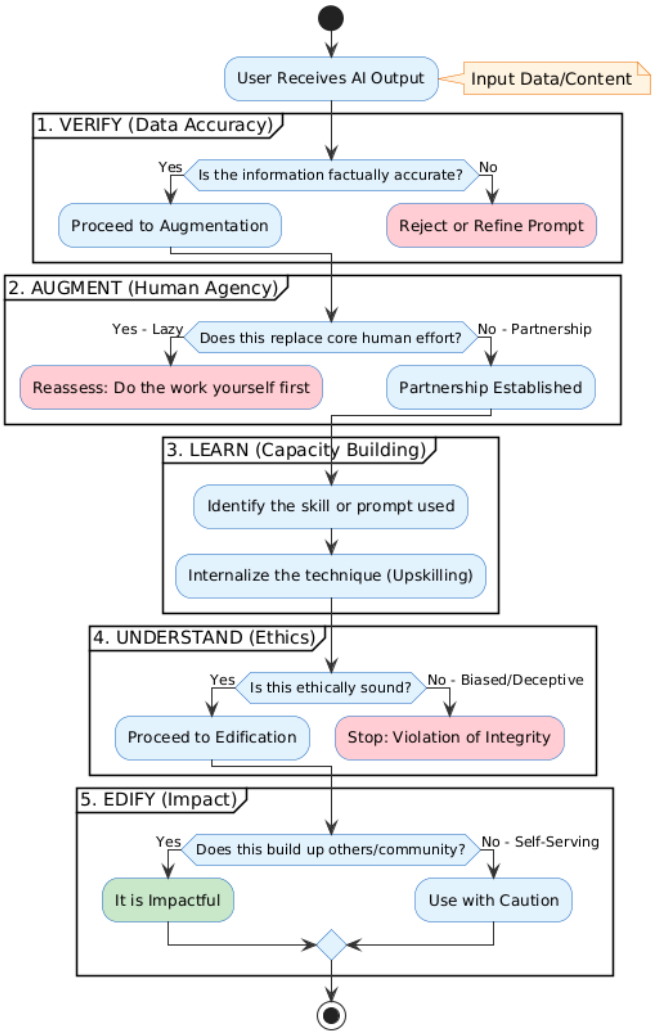


Figure 4: The V.A.L.U.E. Framework for AI Utilization
Source: Authors' Work (2025).

V - VERIFY (Truth over Convenience)

This principle is fundamentally derived from the understanding of AI's ontological distinction, that AI operates purely on statistical probability (syntax) rather than human understanding or truth (semantics). This principle mandates that users cross-reference AI outputs with authoritative sources, serving as a core mechanism of critical AI literacy and actively combats the ethical risk of truth decay. This protocol ensures that the human retains the role of final arbiter of fact, upholding intellectual integrity and preventing the uncritical acceptance of generated probabilistic content.

A - AUGMENT (Partnership over Replacement)

AUGMENT is the practical application of the Human-in-the-Loop (HITL) technical model and the ethical stewardship principle. It dictates that AI must function as a co-pilot, not the

captain, ensuring the final output is human-shaped. This approach adheres to the concept of joint optimization in Socio-Technical Systems (STS) theory, where the efficiency of the technical subsystem is combined with the context, creativity, and moral judgment of the social subsystem, thereby preserving human agency and durable skills.

L - LEARN (Upskilling over Stagnation)

The LEARN principle is rooted in the necessity of critical AI Literacy, which combats the ethical risk of cognitive atrophy. It rejects the passivity of relying blindly on AI, arguing that technical ignorance leads to obsolescence. The action of mastering prompt engineering and data privacy principles makes continuous learning an active, ethical requirement for the user, ensuring that the human remains capable of overseeing and correcting the increasingly complex outputs of the AI system.

U - UNDERSTAND (Ethics over Efficiency)

The UNDERSTAND principle translates the abstract ideals of Value Sensitive Design (VSD) into an actionable protocol. VSD's premise that technology is inherently moral is realized here by asking users to govern their AI usage based on ethical permissibility rather than mere capability. This principle rejects deceptive uses such as plagiarism and deepfakes, respecting data rights. It reinforces the necessity of design-time interventions and proactively embeds values like fairness and trustworthiness into daily user interaction.

E - EDIFY (Impact over Self)

EDIFY connects AI utilization directly to the Purpose component of the 5-P Strategic Framework and the philosophical concept of Stewardship. This principle shifts the focus from individual self-interest to the broader common good, mandating that the power of AI be leveraged to solve community or societal problems. This action ensures that the technology serves to strengthen the social subsystem (STS) and aligns with human values focused on community well-being.

Validation of the Frameworks

To establish the conceptual accuracy and practical relevance of the 5-P and V.A.L.U.E. models, the frameworks underwent a structured validation process through expert consultation and academic feedback from scholars in the fields of computer science, educational technology, and digital ethics. The feedback confirmed the logical consistency and functional applicability of the frameworks. Experts noted that the 5-P Framework effectively bridges the gap between technical perception and ethical purpose, while the V.A.L.U.E. Framework provides a clear, actionable hierarchy for human intervention. The results of this validation indicate that the frameworks are highly applicable as pedagogical tools for fostering AI literacy and as operational guides for maintaining human-in-the-loop oversight in professional environments. Specifically, the "Verify" and "Augment" components were identified as

critical for mitigating automation bias, validating the frameworks' role in preserving human agency.

Comparative Analysis of the Frameworks with Global Standards

The proposed frameworks complement and extend the European Union Artificial Intelligence Act (2024) and UNESCO's Recommendation on the Ethics of Artificial Intelligence (2021), which are existing influential global instruments in AI ethics and governance. However, the 5-P and V.A.L.U.E. frameworks possess a unique cognitive and operational feature that distinguishes them from existing frameworks in practical application by shifting the focus from system-level regulation to user-level agency.

Despite their importance in structuring essential macro-level governance and risk management, existing global regulations often stop short of providing actionable, daily protocols for individual users. While the EU AI Act focuses on the legality of AI systems and UNESCO's Recommendation focuses on the philosophy of values, the frameworks developed in this study focus on the habits of the steward. The uniqueness of the 5-P Framework lies in its pedagogical progression; it recognizes that ethical purpose cannot be achieved without the prerequisite stages of perception and preparation of durable human skills. This addresses the ontological gap left by the global regulations that assume users already possess the critical literacy to navigate complex systems. The emphasis of the 5-P Framework on the preparation of the human agent before the practical use of the tool provides a roadmap for cognitive resilience that high-level regulations overlook—helping users define their human essence before interacting with the machine.

Furthermore, the V.A.L.U.E. Framework significantly outperforms existing paradigms by providing a granular, mnemonic-based protocol for every individual interaction with AI. Where the EU AI Act generally mandates human oversight, V.A.L.U.E. breaks this down into actionable behaviors that provide a real-time Human-in-the-Loop protocol that is missing from generic ethical guidelines. This protocol fills the action gap in the UNESCO principles by translating abstract human rights into a daily checklist for the user. Consequently, these frameworks are designed to be operational tools for moral competence and to empower the user to act as the master of the tool in real-time, effectively bridging the distance between global ethical theory and local human practice. While the EU AI Act provides the shield of regulation and UNESCO provides the compass of values, the 5-P and V.A.L.U.E. frameworks provide the manual for daily human-machine partnership, filling a critical void in the practical implementation of AI ethics.

Implications for Current Practice and Future Directions

The synthesis of thematic findings against Socio-Technical Systems (STS) and Value Sensitive Design (VSD) frameworks reveals crucial implications for current practice across industry, education, and governance, as well as necessary future research directions.

The identified ethical risks, particularly cognitive atrophy (Gerlich, 2025), privacy concerns, and unchecked bias (Adedokun et al., 2025), mandate a fundamental shift in organizational and educational AI strategy. Current practice must move away from viewing AI adoption solely as a technical upgrade toward an STS-informed approach of joint optimization, prioritizing human capability and STS design. This means that alongside AI implementation, institutions must invest heavily in human preparation through comprehensive programs focused on critical thinking, ethical reasoning, and qualitative judgment, which are skills that AI cannot replicate (Gronning & Borggreven, 2025). Furthermore, the Human-in-the-Loop (HITL) imperative must be formalized by ensuring that system interfaces are explicitly designed to promote the "Verify" and "Augment" protocols of the V.A.L.U.E. framework. With this, human operators are not just monitors but active partners who retain ultimate moral accountability (Johnson & Hoover, 2024). The use of AI should be governed by institutional policies that prioritize the protection of user dignity and data privacy (Adedokun et al., 2025). Future directions must emphasize proactive Value Sensitive Design (VSD) and proactive value-based governance to mitigate ethical risks before deployment. Regulators and designers must adopt VSD's tripartite methodology, moving beyond reactive policy to establish design-time interventions that proactively embed values like non-discrimination and transparency into model architecture (Sadek & Mougenot, 2025). This is especially pertinent to Generative AI, where the risk of truth decay (Weigand, 2025) requires new protocols for provenance and digital authentication to rebuild public trust. Academically, future research should empirically validate the efficacy of the proposed 5-P and V.A.L.U.E. frameworks in diverse cultural contexts, particularly focusing on their capacity to cultivate critical AI literacy and ethical stewardship among users.

CONCLUSION

Artificial Intelligence offers unprecedented tools for transforming society, but simultaneously poses profound challenges to human identity and moral authority. However, the distinction between the simulated intelligence of machines and the essence of humanity must be fiercely protected. This study addressed the critical necessity of developing human-centric strategies for technology stewardship. Through an integrative review grounded in the principles of Socio-Technical Systems (STS) and Value Sensitive Design (VSD), this research established that AI's effectiveness is contingent upon the joint optimization of the social subsystem involving human values, roles, and ethics, and the technical subsystem. The research concludes that the human essence remains ontologically distinct from AI's syntactic, computational logic. To bridge the gap between technical utility and ontological preservation, the study contributes two actionable frameworks: the 5-P Strategic Framework for fostering critical AI literacy and the V.A.L.U.E. Framework for guiding AI ethical utilization. Through the adoption of the proposed frameworks, society can ensure that, in an automated era, human contributions remain authentically human and eternally significant, and professionals can

pioneer innovations in different fields while maintaining the moral leadership required to navigate the ethical complexities of the digital age.

REFERENCES

Adedokun, S. A. (2024). The Integration of Artificial Intelligence in Nursing Research: A Framework for Enhanced Data Management and Evidence-Based Practice. *From Novice to Expert: Nurturing Nursing Research Skills*. LAUTECH Faculty of Nursing Sciences' Postgraduate Seminar, College of Health Sciences, LAUTECH Ogbomoso. https://doi.org/10.13140/RG.2.2.29746.26569

Adedokun, S. A., Adedokun, D. A., Ogunkan, S. K., & Ogunniyi, O. K. (2025). Data Privacy Dominance: An Empirical Investigation into Nigerian Postgraduate Students' Prioritization of AI Ethical Concerns in Higher Education. *International Journal of Latest Technology in Engineering Management & Applied Science*, *14*(10), 712–719. https://doi.org/10.51583/IJLTEMAS.2025.1410000088

American Psychological Association. (2025, June). *Health advisory: Artificial intelligence and adolescent well-being*. American Psychological Association. https://www.apa.org/topics/artificial-intelligence-machine-learning/health-advisory-ai-adolescent-well-being

Ang, K. C. S., Sankaran, S., & Liu, D. (2024). Sociotechnical considerations on developing human robot teaming solutions for construction: A case study. *Construction Robotics*, *8*(2), 23. https://doi.org/10.1007/s41693-024-00140-y

Baxter, G., & Sommerville, I. (2019). Socio-technical systems: From design methods to systems engineering. *Interacting with Computers*, *31*(3), 145–158. https://doi.org/10.1093/iwc/iwz015

Berretta, S., Tausch, A., Ontrup, G., Gilles, B., Peifer, C., & Kluge, A. (2023). Defining human-AI teaming the human-centered way: A scoping review and network analysis. *Frontiers in Artificial Intelligence*, *6*. https://doi.org/10.3389/frai.2023.1250725

Dicastery for the Doctrine of the Faith. (2025, January 28). *Antiqua et nova. Note on the Relationship Between Artificial Intelligence and Human Intelligence*. The Holy See. https://www.vatican.va/roman_curia/congregations/cfaith/documents/rc_ddf_doc_20250128_antiqua-et-nova_en.html

Digital Education Council. (2025, March 3). *AI Literacy Framework*. https://www.digitaleducationcouncil.com/post/digital-education-council-ai-literacy-framework

Duggal, N. (2021, February 25). *Advantages and Disadvantages of AI: Key Pros and Cons*. Simplilearn.Com. https://www.simplilearn.com/advantages-and-disadvantages-of-artificial-intelligence-article

European Commission. (2025, November 27). *AI Act: Shaping Europe's digital future*. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

European Union. (2024). *The AI Act Explorer: EU Artificial Intelligence Act*. Future of Life Institute. https://artificialintelligenceact.eu/ai-act-explorer/

Farhood, H., Nyden, M., Beheshti, A., & Muller, S. (2025). Artificial intelligence-based personalised learning in education: A systematic literature review. *Discover Artificial Intelligence*, *5*(1), 331. https://doi.org/10.1007/s44163-025-00598-x

Fu, C., & Chen, Q. (2025). The future of pharmaceuticals: Artificial intelligence in drug discovery and development. *Journal of Pharmaceutical Analysis*, *15*(8), 101248. https://doi.org/10.1016/j.jpha.2025.101248

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

Gronning, M., & Borggreven, M. (2025, April 29). *Beyond Automation: Why an AI-powered future needs human-powered reinvention* [Blog]. PwC. https://www.pwc.com/gx/en/issues/technology/beyond-automation.html

Hanna, M. G., Pantanowitz, L., Jackson, B., Palmer, O., Visweswaran, S., Pantanowitz, J., Deebajah, M., & Rashidi, H. H. (2025). Ethical and Bias Considerations in Artificial Intelligence/Machine Learning. *Modern Pathology*, *38*(3), 100686. https://doi.org/10.1016/j.modpat.2024.100686

Hendry, D. G., Friedman, B., & Ballard, S. (2021). Value sensitive design as a formative framework. *Ethics and Information Technology*, *23*(1), 39–44. https://doi.org/10.1007/s10676-021-09579-x

Husband, T. (1978). Sociotechnical Systems: A Sourcebook. *International Journal of Production Research*, *16*(6), 531–531. https://doi.org/10.1080/00207547808930044

Johnson, M., & Hoover, R. (2024). *The Role of Human Oversight in Responsible AI Systems*.

Jonnalagadda, A. K., Naayini, P., Kamatala, S., Myakala, P. K., & Bura, C. (2025). AI Ethics and Fairness in Machine Learning: Promoting Fairness in Critical Engineering Systems. *2025 2nd International Conference on Trends in Engineering Systems and Technologies (ICTEST)*, 1–9. https://doi.org/10.1109/ICTEST64710.2025.11042668

Knox, A. (2024). Sociotechnical systems theory. In K. Hutchings, S. Michailova, & A. Wilkinson (Eds.), *A Guide to Key Theories for Human Resource Management Research* (pp. 287–293). Edward Elgar Publishing. https://doi.org/10.4337/9781035308767.ch37

Kudina, O., & Van De Poel, I. (2024). A sociotechnical system perspective on AI. *Minds and Machines*, *34*(3), 21, s11023-024-09680–09682. https://doi.org/10.1007/s11023-024-09680-2

Ma, A., Ong, J., & Tan, S. S. (2023). *AI for humanity: Building a sustainable AI for the future*. Wiley.

McKinsey Insight. (2025, November 5). *The State of AI: Global Survey 2025*. McKinsey. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai

Oyebanji, I. T., Oyunwola, T. O., Segun, A. I., & Ogunbiyi, D. O. (2025). Artificial Intelligence and Its Effects on Christian Youths' Spirituality. *African Journal of Religious and Theological Studies*, *4*(1), 34–52. https://doi.org/10.62154/ajrts.2025.04.01013

Pretorius, L., & Cahusac De Caux, B. (2025). The AI Literacy Framework for Higher Education. In J. D. DeHart, S. Abas, R. A. Mora, & D. Gibbons Pyles, *Reimagining Literacy in the Age of AI* (1st ed., pp. 11–32). Chapman and Hall/CRC. https://doi.org/10.1201/9781003510635-3

Russell, S. J., & Norvig, P. (with Chang, M., Devlin, J., Dragan, A., Forsyth, D., Goodfellow, I., Malik, J., Mansinghka, V., Pearl, J., & Wooldridge, M. J.). (2022). *Artificial intelligence: A modern approach* (Fourth edition, global edition). Pearson.

Sadek, M., & Mougenot, C. (2025). Challenges in Value-Sensitive AI Design: Insights from AI Practitioner Interviews. *International Journal of Human–Computer Interaction*, *41*(17), 10877–10894. https://doi.org/10.1080/10447318.2024.2439021

Schmutz, J. B., Outland, N., Kerstan, S., Georganta, E., & Ulfert, A.-S. (2024). AI-teaming: Redefining collaboration in the digital era. *Current Opinion in Psychology*, *58*, 101837. https://doi.org/10.1016/j.copsyc.2024.101837

Shu, S., Luo, Q., & Chen, Z. (2025). Proactive vs. passive algorithmic ethics practices in healthcare: The moderating role of healthcare engagement type in patients' responses. *BMC Medical Ethics*, *26*, 73. https://doi.org/10.1186/s12910-025-01236-y

Stryker, C. (2025, July 8). *What Is Human In The Loop (HITL)?* IBM. https://www.ibm.com/think/topics/human-in-the-loop

Torraco, R. J. (2005). Writing Integrative Literature Reviews: Guidelines and Examples. *Human Resource Development Review*, *4*(3), 356–367. https://doi.org/10.1177/1534484305278283

Trist, E. L., & Bamforth, K. W. (1951). Some Social and Psychological Consequences of the Longwall Method of Coal-Getting: An Examination of the Psychological Situation and Defences of a Work Group in Relation to the Social Structure and Technological Content of the Work System. *Human Relations*, *4*(1), 3–38. https://doi.org/10.1177/001872675100400101

Ugbor, G., Jamal, F., Khan, S., & Shawki, A. W. (2025). Generative AI for drug discovery: Accelerating molecular design with deep learning using Nigerian local content. *Innovation in Computer and Data Sciences*, *1*(1), 66–77. https://doi.org/10.64389/icds.2025.01128

UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000381137

Weigand, S. (2025, November 21). *Emerging threat from deepfakes leads to cybersecurity arms race*. SC Media. https://www.scworld.com/feature/emerging-threat-from-deepfakes-leads-to-cybersecurity-arms-race

Whittemore, R., & Knafl, K. (2005). The integrative review: Updated methodology. *Journal of Advanced Nursing*, *52*(5), 546–553. https://doi.org/10.1111/j.1365-2648.2005.03621.x