Computing&AIConnect
2025, Vol. 2, Cite as: doi:10.x/journal.x.x.x

SCIFINITI PUBLISHING

OPEN ACCESS

Research Article

Rewiring Education for a Super-Smart Society: Cognitive Integrity, AI Ethics, and the Future of Knowledge

Constantine Andoniou
College of Arts and Sciences, Abu Dhabi University, Khalifa City, Abu Dhabi, United Arab Emirates
constantine.andoniou@adu.ac.ae
ORCID ID: [0000-0001-8805-3569]

Abstract: Education systems need to establish methods for protecting cognitive integrity and epistemic trust as artificial intelligence increasingly controls knowledge delivery, educational processes and institutional decision processes. Society 5.0 presents a human-centered digital civilization that uses AI to enhance educational foundations instead of substituting them. The proposed *Cognitive-AI Interaction Framework (CAIF)* serves as a model to study how artificial intelligence transforms educational knowledge production and interpretation processes. The framework uses socio-technical systems (STS) theory together with *Ubuntu*, *Kaitiakitanga*, *Confucian* and *Buddhist* thoughts, *Buen Vivir* and *Human Rights* principles to identify three epistemic modalities: (a) *Epistemic Anchoring*, (b) *Cognitive Mediation Systems*, and (c) *Synthetic Knowledge Environments*. These modalities demonstrate how AI affects trust mechanisms and personalization processes and meaning construction in learning systems. Education protection in Society 5.0 demands both inclusive technology implementation and epistemic governance to maintain ethical design, interpretive transparency and human cognitive autonomy. The paper recommends strategic outputs to help implement ethical AI systems in curriculum development and policy creation and pedagogical design.

Keywords: Society 5.0; artificial intelligence in education; Cognitive-AI Interaction Framework (CAIF); epistemic integrity; ethical AI governance; synthetic knowledge; global AI ethics

1 Introduction

Education finds itself at a decisive point because our world is witnessing rapid technological progress along with increasing uncertainty about knowledge. The Japanese government first presented Society 5.0 as a worldwide movement which now drives human-centered super-smart development beyond the information society of Society 4.0. The evolution of society embeds AI with cyber-physical systems and advanced data analytics throughout social and institutional operations. Education functions as a vital component to this transition because it produces knowledge while shaping values through its AI tool applications and learning epistemic protection mechanisms.

The transition from previous digital reforms that focused on automation and efficiency has given way to Society 5.0 which emphasizes technologies that support human well-being through ethical design and inclusive knowledge systems [1] [2]. The connection between educational processes emerges from the fusion of technological resources with the process of teaching and learning alongside institutional authority in knowledge verification and credentialing.

The *Cognitive-AI Interaction Framework (CAIF)* serves as a conceptual framework which evaluates multiple levels of AI effects on educational systems. The framework shows how AI transforms knowledge systems in educational spaces through three distinct modalities: (a) *Epistemic Anchoring,* which describes how AI enhances knowledge artifact credibility together with institutional trust and traceability; (b) *Cognitive Mediation Systems,* which include AI tools that personalize educational content while providing students with learning assistance and behavioral guidance; and (c) *Synthetic Knowledge Environments,* which consist of generative AI content together with immersive simulations and synthetic instructional agents. The framework (discussed in more detail in Section 2.3) establishes a systematic approach to assess epistemic effects of AI through the examination of knowledge preservation and transformation and their effects on educational approaches and assessment methods.

The framework draws its foundation from the *Socio-Technical Systems (STS)* theory which demonstrates that technological advancement depends on institutional development [3]. The *CAIF* framework also draws support from multiple worldwide ethical traditions which provide unique perspectives to develop educational and artificial technologies that respect cultural values and moral principles. The paper establishes *CAIF* as an analytic method which examines AI implementation in Society 5.0 education through an epistemological framework. The framework defends cognitive integrity by stating that AI implementation in education must protect interpretability, provenance and reflective agency to achieve sustainability. To this end strategic recommendations are provided through a declaration and policy brief series to support educational institutions and policy decision makers. The main aim is to integrate AI in a critical manner to preserve education as a domain where knowledge sovereignty and ethical learning and collective human purpose thrive. To operationalize this aim, the study adopts a conceptual methodology grounded in framework development and global epistemic traditions.

1.1 Methods

This conceptual study adopts a theory synthesis and integrative framework development approach. The Cognitive-AI Interaction Framework (CAIF) was developed by integrating socio-technical systems (STS) theory with a range of ethical and philosophical traditions from various global contexts. The process involved a structured review of interdisciplinary literature, policy frameworks, and ethical codes, including documents from UNESCO, the European Commission, and national and indigenous knowledge traditions. This methodology supports the construction of a normative and analytic framework for evaluating AI's impact on education.

2 Theoretical Foundations

2.1 The Evolution Toward Society 5.0

The development of societies exists as a sequence of stages which describe the arrangement of knowledge together with labor and value. Society 1.0 operated as a hunter-gatherer system with oral knowledge transmission followed by Society 2.0 which introduced agrarian communities and spiritual institutions, followed by Society 3.0 which emerged from industrialization and standardized schooling before Society 4.0 brought global connectivity and digital information systems (Figure 1). The proposed super-smart society of Society 5.0 merges intelligent infrastructure with human-centered values [1], [2].

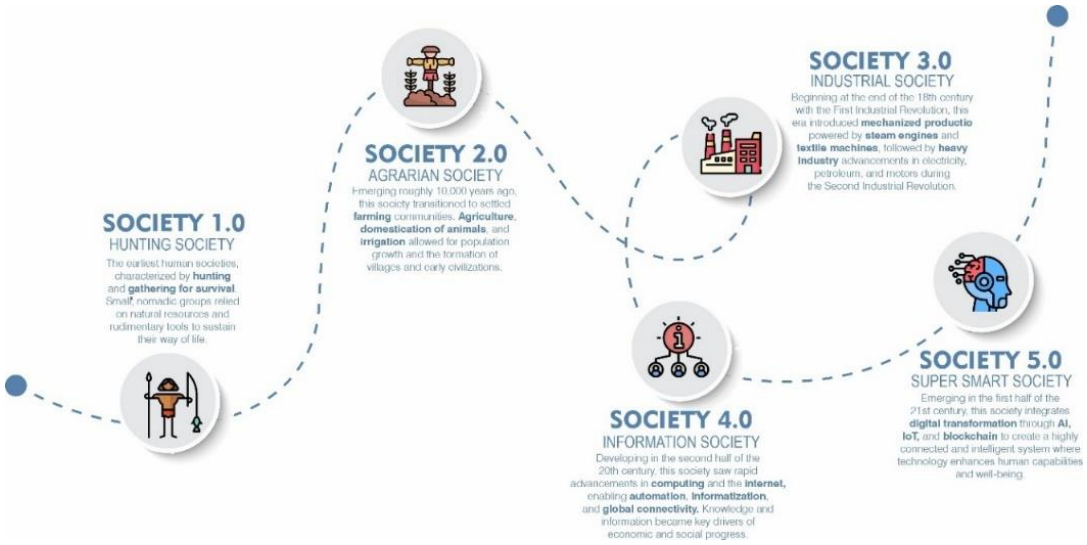

Figure 1: Major societal stages of the evolution of human civilization.

The evolution from Society 4.0 to Society 5.0 represents an epistemological transformation which surpasses technological advancement. Knowledge undergoes transformation regarding its origin and verification processes as well as interpretation methods. This development has profound effects on educational systems. The educational system of Society 5.0 needs to transform its knowledge transmission methods into programs that develop learners who can

use AI environments while making decisions and practicing ethical reasoning. Society 5.0 differs from other visions because it advocates *human-AI harmony* without embracing either technological utopianism or automated dystopia. The realization of this vision depends heavily on education which needs to establish governance and curricular frameworks that emphasize interpretability and pluralism alongside cognitive sovereignty.

2.2 Key Characteristics of Society 5.0

The philosophical foundation of Society 5.0 combines technological optimism with social responsibility [10]. Five core characteristics define this model and guide its application across societal domains [1], [2]: (a) *Human-Centric Innovation*: Society 5.0 prioritizes human needs over market efficiency. AI tools are designed for empathetic communication and inclusive hiring, supporting well-being and social equity; (b) *Cyber-Physical Integration*: The fusion of cyberspace with the physical world enables AI-powered systems such as collaborative robots and wearable IoT devices that enhance workplace safety and preserve human oversight; (c) *Sustainability and Inclusivity*: Equitable access to digital infrastructure is a central goal. AI and IoT technologies support marginalized communities in urban agriculture and remote work through accessible, inclusive platforms; (d) *Data-Driven Decision-Making*: Big data analytics guide targeted interventions in education and workforce development, helping institutions respond to future employment challenges; and (e) *Decentralized Knowledge and Collaboration*: Platform-based models and blockchain technologies empower individuals through credential verification, transparent compensation, and ethical, worker-owned ecosystems (Figure 2). Together, these characteristics articulate a vision where intelligent systems reinforce human dignity, autonomy, and shared progress.

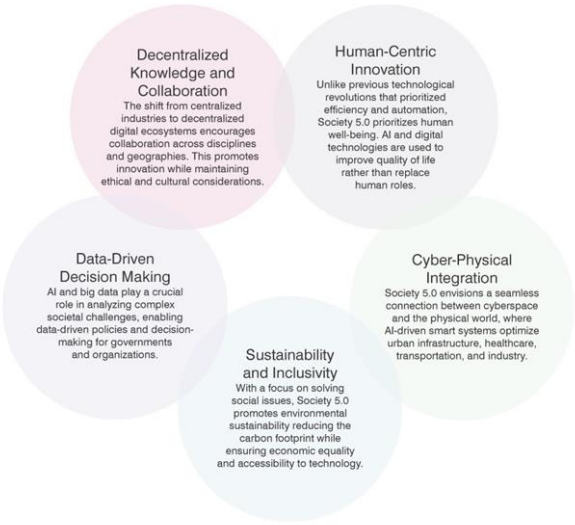

Figure 2: Key characteristics of the Society 5.0 model.

2.3 Socio-Technical Systems (STS) Theory

The *Socio-Technical Systems (STS)* theory establishes a fundamental framework for analyzing how artificial intelligence transforms educational systems (Figure 3). According to *STS* technology exists as part of cultural, institutional and human systems, and develops through its evolution [4]. The educational application of AI functions beyond being a tool because it gradually becomes essential for transforming knowledge structures and teaching practices and student learning engagement. Orlikowski [3] demonstrated that technology interacts with social systems through a continuous feedback process which transforms both the technology and the social structures.

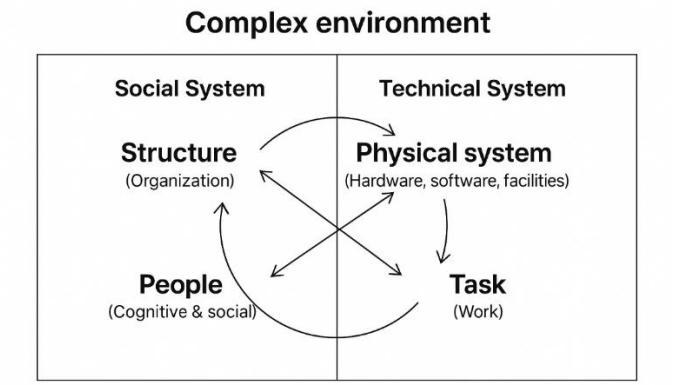

Figure 3: The impact of new technology on complex sociotechnical systems (adapted from van Engelen) [5].

The introduction of AI in classrooms starts with supporting functions such as assessment automation and content delivery yet produces profound changes in teaching authority and student learning independence. The complete understanding of these changes needs more than technical knowledge because it requires awareness of cognitive, ethical and cultural factors that interact with technology. *STS* provides an effective method to counter technological determinism through its support for human-AI collaborative design processes. Educational technology development should occur through collaborative efforts between humans and AI systems while maintaining focus on local values and specific learning environments. The co-development process becomes essential for AI systems because they determine how learners understand concepts such as knowledge and evidence and truth. The *STS* framework enables ethical and context-sensitive implementation of AI in education.

2.4 The Cognitive-AI Interaction Framework (CAIF)

The *Cognitive-AI Interaction Framework (CAIF)* bases its development on *STS* theory to describe three distinct ways AI interacts with knowledge and learning systems:

Epistemic Anchoring: Educational knowledge verification and institutional trust receive support from AI tools, Blockchain-based credentialing, semantic citation validation and plagiarism detection systems serve as examples [6]. The systems ensure content authenticity but need to stop students from depending too heavily on machine verification systems [7].

Cognitive Mediation Systems: Adaptive learning environments, AI tutors, and recommendation engines that personalize instructional pathways [8]. Although these educational tools enhance student outcomes, they simultaneously minimize intellectual diversity and generate algorithmic dependencies. Educators must ensure interpretive agency through complete transparency together with human supervision and diverse pedagogical approaches.

Synthetic Knowledge Environments: AI-created materials combined with computer-generated characters and virtual educational simulations create interactive learning scenarios. The educational tools open up new teaching methods but they make it harder to distinguish between authentic and fabricated content. Students tend to accept artificial representations without question due to the lack of epistemic labeling and critical reflection protocols [9].

The modalities operate as both a diagnostic and design framework which helps educational professionals and institutions understand the implications of AI on knowledge development and educational ethics. The *CAIF* framework provides an analytical framework to predict and direct educational evolution in AI-driven societies but it does not specify a single solution. The analysis of technological form in relation to epistemic function enables researchers to differentiate between AI systems that assist learning from those that distort meaning. These theoretical foundations provide the basis for the following results, which summarize the CAIF framework's core modalities and their implications for AI-integrated education.

3 Results

3.1 Global Ethical Anchors for Human-Centric AI in Education

The increasing integration of AI systems into educational structures requires stronger ethical frameworks that respect cultural diversity. Society 5.0 began as a Japanese philosophical concept with *ikigai* and *wa* values [1] yet its worldwide adoption requires an ethical framework that extends beyond Japan. The governance of AI in education requires multiple global perspectives which value human dignity together with collective knowledge autonomy. This section briefly analyzes five regional ethical systems from Africa to Oceania and Asia and Americas and Europe to determine their impact on CAIF modality design and deployment.

Ubuntu (Sub-Saharan Africa): *Ubuntu* represents a key relational philosophy in sub-Saharan Africa which states "*I am because we are*". This approach emphasizes the importance of interdependence along with community and shared knowledge production responsibilities [10]. Through education *Ubuntu* challenges learning personalization that creates individual learning isolation and breaks down collaborative knowledge systems. According to Higgs [11] *Ubuntu*-based ethics reject individualistic Western approaches by promoting communal values, collective intelligence, and dialogical participation in education. *Ubuntu* matches the principles of *cognitive mediation systems* within the *CAIF* framework by supporting the development of AI tools that promote peer learning and social knowledge scaffolding and co-engagement activities.

Kaitiakitanga (Oceania-Māori Worldview*):* In Māori thought, *Kaitiakitanga* refers to stewardship and guardianship, particularly over cultural and environmental resources [12]. Applied to AI in education, this concept insists that knowledge must be preserved with reverence, especially when it concerns endangered languages, indigenous epistemologies, or oral traditions. *Kaitiakitanga* underscores the value of *epistemic anchoring***,** ensuring that AI systems do not extract or commodify knowledge but rather act as stewards of cultural continuity [13]. AI-driven archival systems, for example, can be designed to preserve community ownership over data and embed indigenous consent protocols into content generation.

Confucian and Buddhist Ethics (East and Southeast Asia): Philosophical systems in East and Southeast Asia emphasize ethical self-cultivation, interdependence, and social harmony. *Confucian* ethics**,** particularly the value of *li* (ritual propriety), prioritize structured moral development and relational responsibility in learning [14]. *Buddhist* thought adds a contemplative dimension, promoting mindfulness, non-attachment, and moral discernment [15]. These frameworks challenge the techno-solutionism often seen in AI-based education and argue for the preservation of pedagogical intentionality. In *CAIF* terms, this maps onto *cognitive mediation systems***,** where AI systems must support - not manipulate - learner reflection, autonomy, and moral reasoning.

Buen Vivir (Latin America): *Buen Vivir* (or Sumak Kawsay), rooted in Andean cosmologies, emphasizes holistic well-being, reciprocity, and harmony with the natural world [16]. It critiques linear, industrial notions of development and advocates for culturally situated and ecologically balanced knowledge systems. As applied to AI in education, *Buen Vivir* problematizes *Synthetic Knowledge Environments* that are disconnected from learners' lived realities or that erase indigenous perspectives. AI-generated simulations and content must be grounded in contextual relevance and moral depth. As Walsh [17] explains, digital technologies in Latin America must be reoriented to serve epistemic justice and community-based validation.

Human Rights Tradition (Europe and Global Institutions): The *Human Rights* tradition, emerged from Enlightenment philosophy and post-war global institutions, emphasizes equality and freedom of expression and access to information. The UNESCO Recommendation on the Ethics of Artificial Intelligence [18] demonstrates this tradition by establishing obligations for transparency, non-discrimination and inclusive access. These principles apply across all *CAIF* modalities - particularly the principles function to regulate algorithmic bias and protect learner data while ensuring explainability.

To clarify their relevance across the framework, Figure 4 summarizes the core principles of each ethical tradition and how they align with the modalities of the *CAIF:*

Ethical Tradition	Core Principle	Relevance to Education	CAIF Alignment
Ubuntu	Relational humanity, interdependence	Promotes collective responsibility and dignity in learning	Epistemic Anchoring
Kaitiakitanga	Guardianship and sustainability	Encourages stewardship of knowledge and ethical tech use	Cognitive Mediation Systems
Buen Vivir	Harmony with nature and community	Frames education as holistic, context-rooted, and ethical	Synthetic Knowledge Environments
Confucian Ethics	Respect, self-cultivation, social harmony	Aligns learning with moral development and communal obligation	Epistemic Anchoring / Mediation Systems
Indigenous Knowledge Systems	Embodied knowledge, reciprocity, spiritual grounding	Values context, oral tradition, and narrative legitimacy	All three

Figure 4: Ethical Anchors Aligned with the Cognitive-AI Interaction Framework (CAIF).

The application of AI systems requires compliance with legal standards as well as epistemically just principles to allow learners to develop their own independent understanding. The ethical anchors operate as a system to develop a multicultural epistemic framework which leads the governance and development of AI in education. The philosophical framework of *CAIF* modalities receives its foundation from these principles which ensure AI tools meet the ethical and cognitive and cultural requirements of learners across diverse cultural backgrounds. Educational systems can transition from technological advancement to moral development. The integration of values into policy and pedagogy and platform design leads to the development of both smart and wise learners.

3.2 Applying the Cognitive-AI Interaction Framework (CAIF) to Educational Systems

The *Cognitive-AI Interaction Framework (CAIF)* provides a framework to study how artificial intelligence transforms education by changing the verification process of knowledge, and the experience of knowledge and trust in knowledge. This section applies the three *CAIF* modalities (*Epistemic Anchoring, Cognitive Mediation Systems, Synthetic Knowledge Environments*) to educational practices, with attention to curriculum, pedagogy, assessment, and governance. While Section 2.4 outlined the foundational principles and theoretical architecture of *CAIF*, the following section applies the framework to concrete educational processes including curriculum, pedagogy, and assessment.

3.2.1 Safeguarding Knowledge Provenance with Epistemic Anchoring

Artificial Intelligence systems performing *epistemic anchoring* verify the trustworthiness of information along with its origin. The widespread adoption of AI-generated content creates new risks for misinformation, fake academic outputs and incorrect citations. Institutions need to implement semantic citation validators and AI- detection systems and blockchain-based credentialing platforms [6], [7] to support epistemic integrity in their academic programs.

The combination of AI-detection systems with blockchain-based transcript verification allows student references to be checked against academic databases while preventing fraudulent credentials and enabling worldwide transcript transfer. These technologies help build trust regarding the genuine nature of academic products. These systems must complement human judgment. Educators should help students contextualize AI outputs and develop epistemic literacy skills to teach students about source reliability assessment. The excessive use of verification automation could result in *interpretive passivity* because students might follow system outputs without performing proper critical evaluation. The implementation of human-in-the-loop design remains essential for *epistemic anchoring* since machine logic functions to support reflective reasoning while maintaining human control.

3.2.2 Structuring Personalization and Pedagogical Flow in Cognitive Mediation Systems

The term *cognitive mediation systems* describe artificial intelligence platforms which provide learning content management through adaptation or recommendation features. The field of adaptive assessments functions alongside recommendation engines and intelligent tutoring systems. These tools provide the best results by boosting student interest while providing instant feedback and enabling customized learning approaches [12], [19].

Mediation brings multiple potential risks to the table. The implementation of algorithmic personalization in education leads to intellectual narrowing since students repeatedly encounter identical content which strengthens their cognitive biases and restricts their ability to explore new concepts [20]. A learner who interacts only with AI-curated readings on a theory might miss out on different or opposing perspectives. Systems need to adopt pluralistic design principles to incorporate multiple perspectives and contradictory epistemological viewpoints. Education professionals must maintain their authority to alter system recommendations while leading students toward more than algorithmically generated content. *Interpretive dependency* presents an additional risk because students tend to consider AI outputs as absolute facts particularly when responses carry authoritative tones. AI systems require features which explain recommendation rationales to users and students need education about analyzing AI decision-making processes. Teacher training should treat *cognitive mediation* as an educational partnership that requires human involvement instead of automated processes. Instructors need to recognize how learning algorithms modify student educational paths and they must know when to intervene with their own evaluation during key instructional and assessment phases.

3.2.3 Navigating Synthetic Pedagogy in Synthetic Knowledge Environments

Synthetic knowledge environments consist of AI-produced interactive learning environments which include virtual laboratories, historical avatar interfaces and synthetic storytelling features to boost student involvement and practical learning experiences. Through these tools students can now experience reconstructed historical cities and interact with AI avatars of important thinkers [9]. Simulations add depth to educational experiences yet they make it challenging to distinguish between authentic content and simulated material. Learners tend to mistake simulated content for genuine historical facts and academic consensus if the simulated experience lacks proper critical guidance. Humanities subjects face significant challenges because interpretation alongside context serves as their foundation.

The educational application of simulation requires three essential transparency protocols: (a) the simulation must include a distinct label indicating it is a synthetic creation; (b) the system must show both the sources used for creation and the algorithms that guide its design; and (c) students need proper guidance to evaluate the content selection and omission in the representation.

A virtual trial simulation about history should contain marginalized perspectives unless designers make a conscious effort to include them. The improper implementation of AI leads to the continuation of dominant stories while eliminating marginalized voices. Educational use of simulated experiences requires students to participate in interpretive reflection tasks that examine the foundational assumptions of the simulation, evaluate different representation perspectives and assess ethical implications of avatar use. The activities help students understand simulations better and recognize them as discursive instruments instead of passive displays.

The implementation of simulated environments creates new problems related to accessibility and fairness. Virtual learning systems with advanced features need strong infrastructure to operate properly which creates challenges for schools with limited resources. Educational institutions need to develop such tools inclusively through open-source models that provide localization capabilities and accessible versions to avoid worsening digital inequities.

The three *CAIF* modalities provide an organized framework for educational institutions to implement artificial intelligence systems. The method of *epistemic anchoring* guarantees the truth-value of knowledge while *cognitive mediation systems* deliver personalized education with strict oversight needs and *synthetic knowledge environments* need ethical oversight to properly analyze their new experiences. Educators need to approach AI by combining resistance with critical evaluation to confirm that new systems fulfill educational objectives which include understanding and interpretation and meaning-making.

These applied insights now lead to a broader conceptual reflection, situating CAIF within existing literature and outlining future directions for empirical validation and institutional adaptation.

4 Discussion

4.1 Governance and Policy Implications for Society 5.0 Education

Educational governance must establish new methods to handle technical aspects and intellectual consequences of AI systems when integrating them into educational infrastructure. The current ethical governance system based on abstract principles and delayed monitoring no longer meets requirements so we need a governance system that predicts problems and involves stakeholders through *cognitive integrity* principles. AI systems in Society 5.0 educational ecosystems should operate as *collaborative meaning-makers* instead of independent authorities while maintaining responsibility to human teachers, students and their communities. *CAIF* functions as a base model to demonstrate how AI affects knowledge through its three distinct processes, *epistemic anchoring*, *cognitive mediation systems*, and *synthetic knowledge environments*. AI system integration into educational knowledge integrity requires unique governance approaches for each modality.

Transparency Protocols and *Epistemic Labeling:* All educational institutions must establish mandatory transparency standards that apply to all AI systems which include content generation and assessment and recommendation systems. The requirements follow international policy directions such as the European Commission's guidelines for trustworthy AI in education which stress algorithmic transparency and accountability and explainability in educational tools [21]. Educational platforms must label AI-generated content while maintaining records of algorithmic decisions that impact learner outcomes and developing standards for epistemic metadata which display authorship information together with source credibility and interpretive scope. Such practices are essential for maintaining *epistemic anchoring* and preventing the normalization of synthetic content as factually equivalent to human-authored knowledge.

*Human-AI Co-Education Policies: T*he implementation of policies needs to ensure educators maintain their authority to interpret AI systems' outputs. The following rights belong to educational staff members: (a) the educational system must allow human instructors to reject AI-generated curriculum and assessment recommendations; (b) the system should enable educators to stop automated feedback systems that lead to student misdirection or de-motivation; and (c) educators must apply their judgment to place AI-generated insights in proper context. *Cognitive mediation systems* need constant monitoring by educators who must learn both AI system operation and output evaluation methods. The implementation of AI systems should prevent educational inequalities from becoming more pronounced. Public institutions and national organizations need to establish policies which provide AI tools to students regardless of their location or financial status or language abilities. AI systems need training data that includes diverse collections which encompass multiple epistemological perspectives together with different linguistic and cultural knowledge domains. *Synthetic knowledge environments* will maintain dominant narratives unless deliberate steps are implemented to incorporate diverse epistemic content. The verification process and external auditing function needs to be implemented. The establishment of effective AI governance requires verification systems that operate through independent third-party verification structures. Third-party audit mechanisms should assess educational AI systems based on their fairness level as well as their transparency and epistemic reliability. The audits should evaluate system performance data beyond performance metrics by assessing interpretive fidelity and cognitive bias elimination and learner agency safeguards. Blockchain technology allows educational credential verification which protects student achievements through secure transferable and tamper-proof systems.

Policy Literacy and Participatory Design: The process of governance needs to become more democratic. Educational systems require stakeholders including educators and learners and parents and community leaders to collaborate for designing and implementing and assessing AI tools. Educational institutions should integrate AI policy literacy into

training programs for teachers and school leadership programs and curriculum development to establish AI policy literacy. Co-design approaches enhance the relevance of AI tools while building institutional trust and protecting interpretive autonomy. Society 5.0 governance serves a purpose that goes beyond technical control because it enables collective development of knowledge systems. Through the *CAIF* framework institutions can shift from regulatory responses to design-oriented approaches which transform AI into an educational partner rather than an authority seeking control.

4.2 Strategic Outputs for Educational Futures

The successful implementation of the *Cognitive-AI Interaction Framework (CAIF)* in educational institutions demands both practical implementation tools and clear conceptual understanding from educational professionals and policymakers. Strategic outputs help maintain the principles of epistemic integrity and human-centered equitable learning during AI implementation. This section introduces two vital initiatives that include, a *Global Declaration* for value alignment, and a *Policy Briefing Series* for local implementation execution.

4.2.1 Declaration for Human-Centered AI in Education

A worldwide endorsement of a declaration would define essential ethical guidelines for incorporating artificial intelligence in educational settings. The declaration should incorporate ethical principles that resemble the UNESCO Recommendation on the Ethics of Artificial Intelligence [21] and the AI4People guidelines [22]. This declaration should ensure: (a) the protection of cognitive integrity requires maintaining epistemic transparency while allowing learner autonomy to maintain knowledge provenance; (b) the declaration should embrace epistemic diversity through *Ubuntu* along with *Kaitiakitanga* and *Buen Vivir* principles [16], [12], [10]; (c) the development of inclusive governance models needs to include teachers alongside learners and their communities (Figure 5). The declaration serves as a normative framework which enables educational institutions to reform their curriculum and platforms as well as develop policies aligned with AI integration. Appendix A contains a draft declaration that institutions can adapt.

Appendix A

Declaration on Human-Centered AI in Education
(Draft for Institutional Adoption)

We, the undersigned educators, researchers, technologists, policymakers, and learners, affirm that the integration of artificial intelligence (AI) into educational systems must uphold the dignity, autonomy, and cognitive freedom of all learners. In the spirit of **Society 5.0** and guided by the ***Cognitive-AI Interaction Framework (CAIF)***, we declare the following principles to ensure ethical, equitable, and epistemically just AI deployment in education:

1. **Cognitive Integrity**
   Educational AI must preserve the interpretability, provenance, and truth-value of knowledge. AI tools must be designed and governed to support - not replace - human judgment and reflective reasoning.
2. **Transparency and Accountability**
   All AI systems used in learning environments must disclose the source, authorship, and logic of their outputs. Learners and educators have the right to understand, question, and override AI recommendations.
3. **Pluralism and Cultural Coherence**
   AI in education must reflect diverse epistemologies, languages, and worldviews. Design and data practices must avoid cultural erasure, algorithmic bias, or colonial encoding.
4. **Inclusive Access and Equity**
   The benefits of AI must be equitably distributed. Policy must ensure that under-resourced regions, schools, and learners are not excluded from ethical and high-quality educational technologies.
5. **Collaborative Governance**
   AI in education must be co-designed and governed by multiple stakeholders - including educators, students, parents, and local communities. No AI system should be deployed without human-centered ethical review.
6. **Educational Purpose over Technological Efficiency**
   AI should serve educational goals rooted in critical inquiry, moral development, and democratic participation - not solely efficiency, automation, or data extraction.

We call upon all educational institutions, ministries, and technology developers to endorse, adopt, and adapt these principles to guide AI implementation strategies worldwide.

CAIF Alignment with Declaration Principles

Declaration Principle	CAIF Modality Alignment	Practical Focus
Cognitive Integrity	Epistemic Anchoring	Preserve truth-value, authorship, and provenance
Transparency and Accountability	All Modalities	Source disclosure, auditability, override rights
Pluralism and Cultural Coherence	Synthetic Knowledge Environments	Cultural representation, epistemic diversity
Inclusive Access and Equity	Cognitive Mediation Systems	Infrastructure, language inclusion, affordability
Collaborative Governance	All Modalities	Stakeholder inclusion in AI design and review
Educational Purpose over Efficiency	Cognitive Mediation & Simulation	Align tech with pedagogy, ethics, and reflection

Figure 5: Declaration on Human-Centered AI in Education.

4.2.2 AI in Education Policy Briefing Series

A modular *Policy Briefing Series* (Figure 6) is proposed to assist decision-making at both national ministry and educational leadership levels. These briefs transform *CAIF* into governance-friendly tools by analyzing major policy domains and using ethical frameworks to provide recommendations. Each brief exists to assist both pilot programs and national implementations by providing adaptable and scalable solutions. Proposed briefs include:

Brief Title	CAIF Modality	Policy Focus Area	Key Recommendations
Epistemic Integrity and Verification Infrastructure	Epistemic Anchoring	Content traceability, academic records, authorship	- Blockchain-secured credentials - AI-generated content labeling - Human-in-the-loop validation
Regulating Synthetic Knowledge Environments	Synthetic Knowledge Environments	AI avatars, immersive simulations, synthetic content	- Mandatory simulation disclosure - Epistemic training modules - Oversight boards for AI narratives
AI Governance for Equity and Access	All (cross-modal equity approach)	Infrastructure, multilingual tools, open access	- Funding for open-source AI tools - Language and localization support - Shared infrastructure models

Each brief is structured to include:

- A policy rationale grounded in ethical frameworks (e.g. [4], [19]);
- Implementation checklists and example use cases;
- Governance guidance tailored to educational contexts.

These policy tools translate CAIF from a conceptual framework into a **governance-ready instrument** - empowering decision-makers to embed cognitive and cultural integrity into AI adoption across education systems.

Figure 6: AI in Education Policy Briefing Series.

The briefs feature adaptable structures for different regions and systems with templates and checklists and policy metrics included. The transformation of CAIF into an inclusive governance system and educational foresight framework occurs through these briefs.

5 Conclusion

The main issue currently is how to implement AI responsibly instead of deciding whether to integrate it. This paper has demonstrated that effective integration needs to consider both functional aspects and the epistemic and ethical implications of AI. Education within Society 5.0 needs to safeguard both the availability of knowledge and the complete integrity of interpretation and meaning [1], [2].

The *Cognitive-AI Interaction Framework (CAIF)* serves as an evaluation a nd guidance system for assessing AI's effects on learning processes. *CAIF* establishes a connection between AI integration through its three modalities (*Epistemic Anchoring, Cognitive Mediation Systems,* and *Synthetic Knowledge Environments)* to examine knowledge production and experience processes. The framework establishes a link between system design elements and educational objectives and cognitive learning results. The framework bases its principles on worldwide ethical traditions including *Ubuntu* and *Kaitiakitanga* and *Confucian* thought and *Buen Vivir* and *Human Rights* to establish that Society 5.0 education needs to preserve its pluralistic and interpretive and human-centered approach [16], [10], [18]. The traditions support *CAIF*'s recommendation for technology systems that work with established values instead of replacing them.

The strategic outputs which include policy tools, curricular reforms and literacy programs show how *CAIF* can transition from theoretical concepts to practical applications. The outputs assist educational professionals and policy-makers to safeguard both the content of student learning and their abilities to understand, question and place information in context. Education's future development will result from our intentional design choices rather than AI's intellectual capabilities. Through frameworks like *CAIF* we can establish AI as an augmentation tool which preserves educational domains of ethical inquiry and epistemic freedom.

List of Abbreviations
CAIF – Cognitive-AI Interaction Framework
STS – Socio-Technical Systems
AI – Artificial Intelligence
IoT – Internet of Things

Authors' Contributions
The author confirms that he was solely responsible for the conception, design, analysis, interpretation, drafting, visualization, and final approval of the article.

Availability of Data and Materials
This is a theoretical manuscript. No empirical data were generated or analyzed. All referenced materials are publicly accessible through the cited sources.

Conflicts of Interest
The author declares no conflicts of interest regarding this manuscript.

Funding
No external funding was received for this research.

Acknowledgment
The author acknowledges the institutional support of Abu Dhabi University.

References

[1] H. Deguchi, Y. Hirai, and S. Muto, *Society 5.0: A People-Centric Super-Smart Society*. Springer, 2020.
[2] M. Kovacic, "The epistemic vision of Society 5.0: A framework for digital civilization," *AI Soc.*, vol. 38, no. 1, pp. 45–61, 2023.
[3] W. J. Orlikowski, "The duality of technology: Rethinking the concept of technology in organizations," *Organ. Sci.*, vol. 3, no. 3, pp. 398–427, 1992.
[4] A. Oosthuizen and L. Pretorius, "Systemic change and socio-technical systems thinking in education," *S. Afr. J. Educ.*, vol. 36, no. 3, pp. 1–10, 2016.
[5] Y. van Engelen, "Explaining Socio-Technical Systems," Medium, 2020. [Online]. Available: https://yaelvengelen.medium.com/explaining-socio-technical-systems-94993a68948c. [Accessed: Aug. 21, 2025].
[6] A. Grech and A. F. Camilleri, *Blockchain in Education*. Joint Res. Centre, European Commission, 2017.

[7] L. Floridi and J. Cowls, "A unified framework of five principles for AI in society," *Harvard Data Sci. Rev.*, vol. 1, no. 1, 2019, doi: 10.1162/99608f92.8cd550d1.
[8] K. Holstein, J. Wortman Vaughan, H. Daumé III, M. Dudik, and H. Wallach, "Improving fairness in machine learning systems: What do industry practitioners need?," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2019.
[9] S. M. West and J. R. Allen, "Deepfakes and education: The coming crisis of trust," *Brookings Inst.*, 2021. [Online]. Available: https://www.brookings.edu
[10] T. Metz, "Ubuntu as a moral theory and Human Rights in South Africa," *Afr. Hum. Rights Law J.*, vol. 22, no. 1, pp. 1–17, 2022.
[11] P. Higgs, "Ubuntu and the politics of communalism in South African education," *Educ. Philos. Theory*, vol. 44, no. S2, pp. 37–55, 2012, doi: 10.1111/j.1469-5812.2011.00795.x.
[12] J. Hutchings, J. Tipene, and H. Phillips, *Kaitiakitanga: Māori Perspectives on Conservation*. Huia Publishers, 2012.
[13] L. T. Smith, *Decolonizing Methodologies: Research and Indigenous Peoples*, 3rd ed. Zed Books, 2021.
[14] J. Li, "The Confucian ideal of harmony and its modern interpretations," *Philos. East West*, vol. 63, no. 4, pp. 568–582, 2013.
[15] D. R. Loy, *Money, Sex, War, Karma: Notes for a Buddhist Revolution*. Wisdom Publications, 2008.
[16] E. Gudynas, "Buen Vivir: Today's tomorrow," *Dev.*, vol. 54, no. 4, pp. 441–447, 2011.
[17] C. E. Walsh, "Development as Buen Vivir: Institutional arrangements and (de)colonial entanglements," *Dev.*, vol. 53, no. 1, pp. 15–21, 2010.
[18] UNESCO, *Recommendation on the Ethics of Artificial Intelligence*, 2021. [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000381137
[19] R. Luckin, W. Holmes, M. Griffiths, and L. B. Forcier, *Intelligence Unleashed: An Argument for AI in Education*. Pearson, 2016.
[20] C. R. Sunstein, *#Republic: Divided Democracy in the Age of Social Media*. Princeton Univ. Press, 2017.
[21] European Commission, *Ethical Guidelines for Trustworthy AI: Education Sector Applications*. EU AI Office, 2024.
[22] L. Floridi *et al.*, "AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations," *Minds Mach.*, vol. 28, no. 4, pp. 689–707, 2018.