Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 46 No. 7
July 2025

Ethical Integration Of Artificial Intelligence In Education: A Review Of Frameworks And Practical Strategies

Dave E. Marcial

Dr. Mariano C. Lao Global Studies Center, Silliman University

Dumaguete City, Philippines

demarcial@su.edu.ph

Abstract

The rapid integration of Artificial Intelligence (AI) in education offers transformative opportunities for personalized learning, instructional innovation, and administrative efficiency. However, it also introduces ethical challenges related to data privacy, algorithmic bias, academic integrity, and equitable access. This paper presents a qualitative review of global and regional ethical AI frameworks—including those from UNESCO, the European Union, and the World Economic Forum—along with industry standards and academic literature. The review identifies core ethical principles such as transparency, fairness, accountability, inclusiveness, and privacy, which are foundational guidelines for responsible AI use in educational settings. Beyond the literature synthesis, the paper incorporates practice-based insights from the phased development of an institutional AI integration strategy, illustrating how ethical principles can be operationalized into policies, training programs, and digital governance mechanisms. The findings highlight the critical role of transformative digital leadership, stakeholder collaboration, and sustained upskilling in creating ethical, adaptable, and context-sensitive AI strategies. This study contributes to advancing responsible AI adoption in education by offering a consolidated review of ethical frameworks and actionable strategies for institutional implementation. It concludes with recommendations for developing localized AI policies rooted in ethical standards, ensuring that the integration of AI enhances rather than undermines the core values of education.

Keywords: AI Policy in Education, Educational Technology Ethics, Ethical AI Frameworks, Responsible Use of Technology, ICT in Education, AI Integration Strategies

INTRODUCTION

Accepting artificial intelligence (AI) in education is no longer a choice but a necessity. As schools and universities strive to future-proof their institutions, they must embrace AI as a critical tool to remain competitive and relevant in the rapidly evolving digital landscape. Technologies such as Natural Language Processing (NLP), Machine Learning (ML), and Generative AI are transforming education by offering personalized learning experiences, optimizing administrative workflows, and fostering deeper student engagement (Luckin, Holmes, Griffiths, & Forcier, 2016). Schools that fail to adopt AI risk falling behind in preparing students for the demands of the Fourth Industrial Revolution.

However, integrating AI into education has significant challenges, particularly ensuring its responsible and ethical use. Schools are confronted with various issues, such as privacy violations, data security risks, potential biases, and the unintended suppression of critical thinking. A striking example of the misuse of AI in education occurred in the UK in 2020 when an AI algorithm was used to determine student grades for exams canceled due to the COVID-19 pandemic. The algorithm, which relied on historical data and school performance, disproportionately downgraded students from disadvantaged backgrounds, leading to widespread protests and accusations of bias (BBC News, 2020). This incident highlights the potential harm of relying on AI systems without transparent, equitable, and inclusive practices.

Similarly, facial recognition technology used in some schools in the United States to monitor attendance and security has raised ethical concerns about surveillance and student privacy. Critics argue that the technology disproportionately targets minority students and exacerbates inequalities in disciplinary measures (Almeida, Shmarko, & Lomas, 2022). The use of AI in these contexts underscores the urgent need for schools to address ethical considerations and ensure that AI tools are used responsibly without infringing on students' rights or perpetuating discrimination.

One of the most significant challenges in AI integration is the lack of comprehensive ethical frameworks tailored to the specific needs of educational institutions. Schools must not only implement AI technologies but also develop policies that address issues of fairness, transparency, and inclusivity. Without such frameworks, the risks associated with AI could undermine its potential benefits. For example, AI-driven systems that offer personalized learning experiences can also create "echo chambers" by delivering content that reinforces students' existing knowledge, limiting exposure to diverse perspectives and critical thinking opportunities (Mateen, 2018).

Furthermore, not all schools are equipped with frameworks for the responsible integration of AI. Every institution must develop an ethical AI framework aligned with its vision, mission, and educational goals. Such frameworks should not only guide the technical implementation of AI but also address broader social and ethical concerns, ensuring that AI tools are used equitably and inclusively. UNESCO's ethical guidelines for AI in education emphasize the need for transparency, fairness, and accountability to protect the rights of students and foster a more equitable learning environment (UNESCO, 2021). Without these safeguards, the educational sector risks reinforcing existing inequalities and creating new forms of discrimination through AI misuse.

This paper examines the integration of AI in education, with a focus on its responsible and ethical use. Specifically, it reviews the foundational concepts of AI, the growing demand for AI in education, and the associated risks and challenges. Additionally, the paper analyzes existing ethical frameworks from various perspectives. It identifies common ethical principles that the education sector can adopt. Furthermore, it presents Silliman University's experiences and lessons learned in developing its own institutional AI integration framework.

This paper contributes significantly to understanding ethical AI integration in education by addressing the growing need for schools to adopt AI responsibly. It highlights the urgency of AI adoption as part of future-proofing education while stressing the importance of developing tailored ethical frameworks for schools. The paper underscores the potential risks of unethical AI practices. Likewise, it offers practical insights through the case of Silliman University's phased approach to AI integration, providing a model for other institutions. Additionally, the paper synthesizes global ethical frameworks and proposes common ethical principles, focusing on equity, inclusivity, and student protection. Ultimately, this paper encourages educational institutions to adopt AI technologies thoughtfully, ensuring they align with ethical standards and promote positive educational change.

METHODS

This study employed a qualitative integrative review approach to examine existing frameworks that guide the ethical integration of Artificial Intelligence (AI) in education. The primary objective was to synthesize scholarly literature, policy documents, and global ethical guidelines relevant to educational contexts. The research began with an extensive review of current literature to understand the evolving landscape of AI in education, including its increasing adoption, foundational risk principles, and the most common associated ethical challenges.

The review included a comparative analysis of widely recognized ethical AI frameworks developed by global and regional institutions. It focused on core ethical principles such as transparency, fairness, inclusiveness, and privacy. The study also examined practical recommendations for integrating AI tools in educational environments while addressing potential risks such as algorithmic bias, over-reliance on technology, and diminished creativity. Special attention was given to the educational applications of Generative AI, along with associated risks such as plagiarism, weakened critical thinking, and the dissemination of misinformation in student outputs.

The analysis was organized into three key categories. The first focused on global frameworks, including those developed by the European Union Union (European Parliament, 2023), UNESCO's ethical AI guidelines (UNESCO, 2022), and the World Economic Forum's principles for AI in education (World Economic Forum, 2024). The second category included industry-specific frameworks, such as Microsoft's Responsible AI Standard and other practical ethical guidelines (Microsoft, n.d.). The third covered academic studies and institutional case reports that addressed AI implementation in education, particularly those emphasizing privacy protection, bias prevention, and academic integrity.

To complement the literature review, the paper presents a localized perspective by drawing on the author's involvement in developing an AI integration framework at Silliman University. This section outlines the institution's phased approach

to AI adoption, including the formulation of policies, execution of key initiatives, and lessons learned throughout the process to ensure ethical and effective implementation.

UNDERSTANDING AI

Artificial Intelligence (AI) refers to a suite of technologies that enable machines to perform tasks typically requiring human intelligence, such as reasoning, learning, problem-solving, and decision-making. AI has become a cornerstone of the Fourth Industrial Revolution, significantly influencing various sectors, including education. AI can revolutionize traditional teaching methods in education, providing personalized learning experiences, optimizing administrative processes, and fostering innovation in content delivery (Luckin, Holmes, Griffiths, & Forcier, 2016).

AI encompasses several key components, each with distinct applications. For example, Natural Language Processing (NLP) focuses on enabling machines to understand, interpret, and generate human language. This technology powers applications such as content extraction, sentiment analysis, and machine translation, enhancing language learning and communication tools (Kurni, Mohammed, & Srinivasa, 2023). Expert Systems are AI-driven decision-support systems that offer specialized advice or decision-making based on large datasets. At the same time, Computer Vision enables machines to interpret and process visual data, supporting image recognition and video analysis (LeCun, Bengio, & Hinton, 2015).

One of the most transformative aspects of AI is Machine Learning (ML), which involves training algorithms to learn from data and make predictions or decisions without being explicitly programmed. Within ML, supervised, unsupervised, and reinforcement learning enable machines to perform complex tasks like classification, clustering, and optimization. Deep learning, a subset of ML, utilizes neural networks to model and understand intricate patterns in large datasets, contributing to advancements in image and speech recognition (Heaton, 2018).

Among the growing AI technologies, Generative AI has gained considerable attention for its ability to create new content, including text, images, music, and even video. Unlike traditional AI, which focuses on tasks based on existing data, Generative AI models, such as Generative Adversarial Networks (GANs) and Transformer models like GPT, can produce human-like content by learning underlying data patterns. Generative AI is revolutionizing content creation in education, enabling personalized instructional materials and simulations and even automating assessment processes (Holmes, Bialik, & Fadel, 2019).

As AI technologies evolve, their integration into educational systems promises to reshape teaching and learning experiences. However, understanding AI's foundational components is crucial for institutions to leverage its potential effectively. This knowledge ensures that educators and students alike can use AI technologies responsibly and meaningfully, enhancing educational outcomes while mitigating risks such as dependence on technology or the propagation of misinformation (Floridi, et al., 2018).

THE RISING DEMAND OF AI

The rising demand for AI in education is driven by the technology's ability to enhance personalized learning, improve administrative efficiency, and better prepare students for the future workforce. As AI continues to evolve and its adoption accelerates, educational institutions must proactively integrate AI to remain relevant and competitive in the global landscape.

In recent years, the adoption of artificial intelligence (AI) has accelerated across multiple sectors, including healthcare, finance, transportation, and education. AI is increasingly seen as a transformative tool capable of reshaping industries by enhancing efficiency, driving innovation, and enabling personalized services. The demand for AI technologies is growing in education as institutions recognize the potential to improve learning outcomes, streamline administrative processes, and prepare students for a technology-driven world (Holmes, Bialik, & Fadel, 2019).

According to a report by McKinsey & Company, AI could contribute up to $13 trillion to the global economy by 2030, driving exponential growth across industries (Chui, et al., 2023). Education, traditionally slower to adopt cutting-edge technology, is now beginning to embrace AI-driven tools to meet the evolving needs of students and educators. AI-powered platforms are used for personalized learning, allowing educators to tailor instruction based on individual student performance. This shift toward adaptive learning technologies enables more efficient, targeted interventions, especially for students who need additional support (Luckin, Holmes, Griffiths, & Forcier, 2016).

Generative AI is one of the most significant advancements driving the demand for AI in

education, which can create new content such as essays, images, music, and videos. Generative AI is being utilized to assist in everything from content creation to automating grading and assessment processes. By creating personalized materials and simulations, AI enhances the interactivity of educational content, making learning more engaging and tailored to individual needs (Holmes, Bialik, & Fadel, 2019). For example, AI-based tools such as intelligent tutoring systems can provide real-time feedback to students, mimicking one-on-one instruction and offering immediate responses to learners' needs (Chen, Chen, & Lin, 2020).

Beyond personalized learning, AI is critical in improving administrative efficiency. Machine learning algorithms are being used to automate routine tasks like student enrollment, scheduling, and performance tracking, freeing educators to focus on higher-value activities. AI applications can also optimize resource allocation by predicting staffing needs, analyzing enrollment trends, and helping institutions make data-driven decisions (Zawacki-Richter, Marín, Bond, & Gouverneur, 2019).

Moreover, AI has begun to revolutionize knowledge work by enhancing decision-making processes. Generative AI, for example, can aid in drafting documents, analyzing large datasets, and generating insights from unstructured information. This impacts fields like education, law, and healthcare and has significant implications for research and development. In education, researchers leverage AI to automate labor-intensive tasks such as data collection and analysis, allowing more time for higher-level problem-solving and creative thinking (Marcus & Davis, 2019).

Globally, governments and organizations recognize the importance of AI adoption in education. For instance, the Philippines launched its National AI Roadmap to position the country as an AI hub in Southeast Asia, potentially boosting its GDP by 12% by 2030 (Department of Trade and Industry - Republic of the Philippines, 2021). This demonstrates the growing acknowledgment of AI's potential to drive educational innovation and spur national economic growth.

As the global AI market expands, the demand for educational institutions to adopt AI tools will only intensify. This demand is driven by the need to remain competitive and the growing recognition that AI can prepare students for future work environments where AI and automation will play an increasingly significant role. According to a World Economic Forum (2020) study, 85 million jobs could be displaced by AI by 2025. However, an estimated 97 million new roles will emerge, particularly in AI development, machine learning, and digital transformation. Therefore, schools and universities must integrate AI into their curricula to equip students with the necessary skills to thrive in this evolving landscape (World Economic Forum, 2020).

RISKS OF AI

As technology progresses, it inevitably brings transformative changes to many aspects of life, including education. However, with these advancements come inherent risks. Risk is an unavoidable aspect of technological innovation, stemming from the complex interplay between technical, human, and societal factors. These risks are rooted in principles such as complexity and uncertainty, human error, security vulnerabilities, resource limitations, and regulatory and compliance challenges. These elements give rise to ethical and social concerns, making risk management a critical part of technology adoption (Floridi, et al., 2018).

Several prominent philosophers have provided frameworks for understanding the inherent risks of technology. In The Question Concerning Technology (1954), Martin Heidegger introduced the concept of enframing (Gestell), which portrays modern technology as a force that treats the world as a resource to be optimized. Heidegger cautioned that this mindset leads to unforeseen harm, including the loss of authentic human experiences and the unpredictability of relying on technology to shape our environment. His work emphasizes the risks of control and unintended consequences as humans increasingly depend on technology (Heidegger, 1997).

In his influential work The Imperative of Responsibility (1979), Hans Jonas argued for ethical responsibility toward future generations in the face of technological progress. Jonas stressed the precautionary principle, advocating for carefully weighing risks and long-term consequences before embracing innovations. He promoted a cautious, responsible approach to technology, prioritizing harm prevention over unchecked advancement (Jonas, 2014).

Ulrich Beck, in *Risk Society: Towards a New Modernity* (1986), introduced the concept of the risk society, wherein modernity focuses on the production and management of risks, particularly those arising from technological and industrial developments. Beck emphasized reflexive modernization, which involves critically assessing

and responding to the risks and consequences of technological progress as society becomes more aware of them. His work highlights the growing importance of managing risks in a world increasingly shaped by technological advancements (Leiss, W. , 1994).

These theoretical underpinnings provide a foundation for understanding the constant presence of risk in technological development. Their perspectives emphasize the importance of ethical considerations, awareness of unintended consequences, and the need for a cautious approach to technological progress, particularly in fields like AI, where the implications can be profound and far-reaching.

While artificial intelligence (AI) holds tremendous promise for transforming education, it also presents several risks that must be addressed to ensure ethical and responsible use. As AI becomes more integrated into educational systems, potential challenges related to privacy, bias, academic integrity, and over-reliance on technology may arise. Addressing these risks is essential to maintain trust, promote equity, and safeguard educational standards.

Privacy Concerns. One of the most pressing risks associated with AI is the issue of privacy. AI systems often rely on vast amounts of data, including sensitive personal information about students, such as performance metrics, learning preferences, and behavioral patterns. If not adequately safeguarded, this data could be vulnerable to breaches or misuse, raising concerns about student privacy (Zawacki-Richter, Marín, Bond, & Gouverneur, 2019). For example, AI-powered learning platforms that track students' progress could expose sensitive information if proper data protection measures, such as encryption and anonymization, are not in place (Chen, Chen, & Lin, 2020).

Bias and Fairness. AI models are only as unbiased as the data on which they are trained. If AI systems are trained on datasets that contain biases—whether related to race, gender, socio-economic status, or other factors—they may unintentionally reinforce or even exacerbate these biases. For instance, an AI system used for grading or student evaluations may favor certain demographic groups if trained on biased data, potentially leading to unfair outcomes (Holmes, Bialik, & Fadel, 2019). Studies have shown that bias in AI systems can perpetuate stereotypes and marginalize already disadvantaged groups, highlighting the importance of transparency and fairness in AI design and deployment (Floridi, et al., 2018).

Academic Integrity and Plagiarism. With the rise of Generative AI, tools like GPT-3 can generate human-like text, raising concerns about academic integrity. Students may misuse AI to produce essays, reports, or even research papers without proper attribution, leading to instances of plagiarism. AI-generated content could also obscure whether a student genuinely understands the subject matter, as it becomes difficult to assess the authenticity of their work (Baker, Smith, & Anissa, 2019). Institutions must implement robust plagiarism detection tools and educate students on the ethical use of AI to maintain academic standards (Holmes, Bialik, & Fadel, 2019).

Over-reliance on AI and Loss of Critical Thinking Skills. AI's ability to quickly provide answers or generate content could lead to a dangerous over-reliance on technology, potentially stifling students' critical thinking and problem-solving skills. Students who habitually rely on AI-generated solutions may bypass the cognitive processes necessary to engage with complex problems deeply, leading to a superficial understanding of the material (Luckin, Holmes, Griffiths, & Forcier, 2016). Moreover, AI-driven recommendations may sometimes promote a single, deterministic way of thinking, limiting exposure to diverse perspectives and stifling creativity (Marcus & Davis, 2019).

Misinformation and Lack of Accuracy. Although AI systems like Generative AI can produce compelling text, they are not immune to errors or inaccuracies. AI models can sometimes generate misleading or false information, especially when trained on incomplete or biased data sources. The spread of AI-generated misinformation poses a significant risk, particularly in educational settings where students may take AI-generated content at face value without critically assessing its accuracy (Mateen, 2018). Misinformation, especially when propagated at scale, could undermine educational outcomes and damage trust in AI technologies.

Dependence on Technology and Inequitable Access. AI's growing presence in education may also exacerbate disparities in access to technology. Students in under-resourced schools or communities may not have the same access to AI-driven tools as their more affluent peers, further widening the digital divide (Zawacki-Richter, Marín, Bond, & Gouverneur, 2019). Additionally, as educators and students become more dependent on AI systems for teaching and learning, technological disruptions—such as system failures

or lack of internet access—could create significant educational barriers.

Job Displacement and Ethical Concerns. AI's ability to automate routine tasks raises broader ethical concerns about the future of work in education. As more educational tasks—such as grading, administrative functions, and even aspects of teaching—are automated, there is a risk that some roles may become redundant, leading to job displacement (Schatsky, Gurumurthy, & Muraskin, 2021). At the same time, ethical concerns arise regarding the appropriate role of AI in education: Should AI replace human teachers in specific tasks or merely assist them? These questions underscore the need for careful consideration of AI's impact on employment and the human elements of teaching.

Ethical Decision-making and Responsibility. Lastly, AI raises fundamental questions about accountability and ethical decision-making. Who is responsible if an AI system produces biased or harmful outcomes? How should educators and institutions ensure that AI is used in ways that align with ethical principles? These challenges call for developing clear guidelines and policies to ensure that AI is used responsibly in education (Floridi, et al., 2018).

ETHICAL AI IN EDUCATION FRAMEWORKS

As artificial intelligence (AI) continues to become an integral part of education, schools and universities must adopt it in a way that proactively addresses both technical and ethical considerations. Developing an AI integration framework is essential for educational institutions to ensure responsible, equitable, and effective use of AI technologies. Several established ethical AI frameworks serve as benchmarks for guiding the integration of AI in education.

The European Union AI Act

The European Union AI Act is the world's first comprehensive legal framework designed to regulate AI. It establishes strict requirements for AI developers while reducing burdens on businesses, especially small and medium-sized enterprises (SMEs), through initiatives such as the AI Innovation Package and the Coordinated Plan on AI (European Parliament, 2023). The act categorizes AI systems into risk levels—unacceptable, high-risk, and low-risk—and mandates strict oversight for high-risk AI systems, such as those used in education for student evaluation or grading (Directorate-General for Communication, 2024). The primary goal is to promote AI innovation while safeguarding fundamental rights and ensuring that AI systems adhere to ethical principles, including transparency, accountability, and fairness in decision-making processes.

UNESCO's Human-Centered AI Framework

The second framework is UNESCO's Recommendation on the Ethics of Artificial Intelligence, which advocates for a human-centered approach to AI. It expands the conversation around AI's role in addressing existing inequalities in access to knowledge, research, and cultural diversity. A vital goal of this framework is to prevent AI from exacerbating technological disparities within and between nations (UNESCO, 2021).

UNESCO's Ethical AI framework is built upon four core values: respect for human dignity, fairness, privacy, and transparency. These values form the foundation for AI systems designed to serve the betterment of individuals, societies, and the environment. The framework outlines ten core principles, providing a human-rights-centered approach to AI ethics. These principles guide the ethical implementation of AI by ensuring that it upholds values such as equity, transparency, and inclusiveness.

In the context of education, UNESCO emphasizes the importance of learner-centric AI. The framework advocates for transparency, bias prevention, data protection, and agency and autonomy. Transparency ensures that AI systems are explainable and free from bias while promoting inclusion to foster equitable access to educational opportunities. Protecting learners' data is critical, as is empowering students with autonomy to actively engage with AI tools responsibly. Additional key concepts include interoperability, diversity, anticipation and precaution, remedial measures and redress, social good, and dignity, all of which contribute to the ethical use of AI in education.

While these values and principles are crucial for establishing an ethical AI foundation, recent advancements in AI ethics emphasize the need to transition from abstract ideals to practical strategies. UNESCO's Recommendation addresses this by outlining eleven critical areas for policy actions and guiding institutions in implementing AI ethically in practice (UNESCO, 2021).

In November 2023, the author of this paper participated in UNESCO's regional experts' meeting on Generative AI and Education in Asia-Pacific, held in Bangkok, Thailand. The meeting focused on articulating ethical practices for AI use in education.

During a roundtable discussion, two questions were explored: a) What skills and competencies are essential for students to ethically and responsibly interact with generative AI? b) What skills and competencies are necessary for students to critically assess and interpret AI-generated content?

In addition, the author attended the United Nations Philippines' National Stakeholders meeting on Ethical AI. The meeting introduced UNESCO's AI Readiness Assessment Methodology (RAM) (UNESCO, 2023), a critical tool for educational institutions to evaluate their preparedness for ethical AI integration. RAM offers a comprehensive assessment of how ready schools are to implement AI responsibly, identifying areas that require institutional and regulatory changes to support the ethical use of AI in education.

World Economic Forum's Ethical Principles for AI

The World Economic Forum (WEF) offers seven practical principles for guiding the ethical use of AI in education: fairness, accountability, transparency, inclusiveness, reliability, privacy, and security (Schatsky, Gurumurthy, & Muraskin, 2021). These principles ensure that AI systems in education support inclusive and equitable learning environments while protecting students' privacy and personal data. WEF emphasizes that AI systems should be designed to avoid bias and discrimination, providing students with equitable opportunities for learning and development.

IEEE's Ethically Aligned Design

The Institute of Electrical and Electronics Engineers (IEEE) developed its General Principles of Ethically Aligned Design to provide ethical guidance for all autonomous and intelligent systems (IEEE, 2019). This framework outlines high-level standards for the development, deployment, and decommissioning of AI systems, with a focus on ensuring human well-being and preventing harm. In education, these principles promote the ethical use of AI by safeguarding students' autonomy, creativity, and intellectual growth while mitigating the risks of over-reliance on AI tools for decision-making.

Industry-Industry Oriented AI Standard

Schools can also consider industry-oriented Ethical AI frameworks like those developed by Microsoft. Microsoft's Responsible AI Standard is an industry-oriented framework with six fundamental principles: fairness**,** reliability**,** safety**,** privacy**,** inclusiveness**,** transparency**,** and accountability (Microsoft, n.d.). In an educational setting, Microsoft's framework helps institutions implement AI tools that enhance teaching and learning without compromising ethical standards. Microsoft's AI systems aim to provide personalized learning experiences for all students, regardless of their background, while ensuring that AI tools are fair, reliable, and safe (Microsoft, n.d.).

Cohort-oriented Standards

Several studies have examined cohort-based applications of Artificial Intelligence (AI) in education, highlighting their potential to enhance personalized learning while advancing equity. Cohort-oriented standards refer to AI systems designed to address the specific needs of defined student groups—based on characteristics such as academic performance, learning preferences, cultural background, or socio-economic status. This approach allows AI tools to adapt content delivery and teaching strategies more precisely, resulting in more meaningful and effective learning experiences (Luckin, Holmes, Griffiths, & Forcier, 2016).

Moving beyond generalized datasets, cohort-based AI can help mitigate systemic biases and support more inclusive learning environments. Students from diverse linguistic, cultural, or economic backgrounds often require differentiated support, and cohort-oriented models offer a pathway for addressing these disparities in real time (Holmes, Bialik, & Fadel, 2019). This personalization improves learning outcomes and promotes fairness and access in AI-enhanced classrooms.

Moreover, cohort-based AI can foster collaboration and peer-to-peer engagement by forming dynamic groupings for shared learning tasks. These systems support collective problem-solving and interactive learning, essential elements of student-centered education. Educators also benefit from more granular data insights, enabling them to monitor performance, identify at-risk learners, and design targeted interventions more precisely (Chen, Chen, & Lin, 2020).

In the Philippine context, for instance, a study on teachers' social and ethical ICT competencies in Central Visayas found that while educators demonstrated a generally "good" level of competence, their practical experience in ethical technology use remained limited. Key factors influencing ICT competency included age, employment status, teaching experience, and access to technology (Marcial, 2017). This underscores the importance of designing AI systems and training programs sensitive to cohort-

specific characteristics among learners and educators who implement these tools.

COMMON ETHICAL AI PRINCIPLES

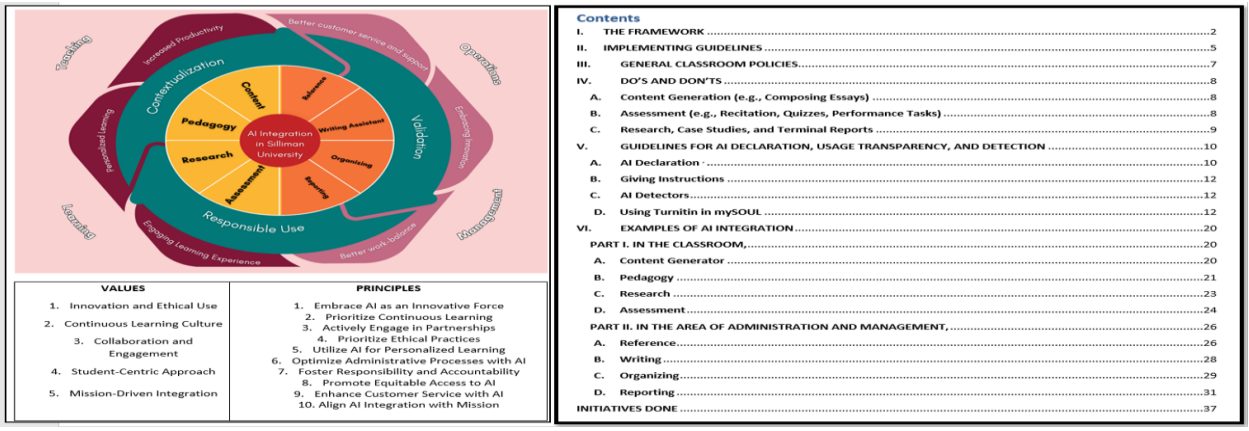
After reviewing several established frameworks, fourteen key principles emerge: Prevention of Harm, Precautionary Approach, Equity and Justice, Sustainable Development, Solidarity, Scientific Knowledge and Integrity, Accountability, Inclusiveness, Reliability and Safety, Fairness, Transparency, Privacy and Security, Avoidance of Conflicts of Interest, and Enforcement of the Code. These principles form the foundation for institutions to ensure that AI is used responsibly, ethically, and effectively in education (Floridi, et al., 2018).

1. Prevention of Harm - Prioritize public safety, health, and welfare by adhering to ethical design principles, protecting privacy, and disclosing potential risks.
2. Precautionary Approach - Exercise caution when adopting AI, carefully considering potential risks and unintended consequences.
3. Equity and Justice - Ensure fairness in AI systems and avoid discrimination or bias in their application.
4. Sustainable Development - Adopt responsible and sustainable AI practices that promote long-term societal benefits.
5. Solidarity - Collaborate across sectors and disciplines to develop ethical and beneficial AI for all.
6. Scientific Knowledge and Integrity - Uphold honesty in AI research, acknowledge errors, and give proper credit for contributions.
7. Accountability - Take responsibility for AI systems' outcomes and impact, including unintended consequences.
8. Inclusiveness - Consider diverse perspectives and voices when developing and implementing AI systems.
9. Reliability and Safety - Ensure that AI systems are dependable and do not pose risks to users or the public.
10. Fairness - Treat all individuals fairly and respectfully, ensuring that AI decisions do not result in unjust outcomes.
11. Transparency - Be transparent about how AI systems operate, including their decision-making processes.
12. Privacy and Security - Protect individuals' privacy and ensure robust security measures are in place to safeguard data.
13. Avoidance of Conflicts of Interest - Disclose and avoid conflicts of interest to maintain trust and lawful conduct.
14. Enforcement of the Code - Actively supports ethical behavior by enforcing ethical standards and reporting violations.

DEVELOPING AN AI INTEGRATION FRAMEWORK: THE SILLIMAN UNIVERSITY EXPERIENCE

To successfully adopt Artificial Intelligence (AI), educational institutions must establish a comprehensive integration framework with clear policies, practical guidelines, and effective monitoring mechanisms. Such a framework should prioritize ethical considerations, address potential biases, safeguard data privacy, and balance technological innovation and meaningful human interaction. This approach ensures the development of an inclusive, equitable, and responsible learning environment.

At Silliman University, an Institutional AI Integration Framework was developed to guide AI's ethical and effective use in academic and administrative domains (see Figure 1). The framework is anchored in the university's core values of innovation, ethical AI use, and continuous learning (Marcial, Dave E.; Silliman University, 2024). It emphasizes a collaborative, student-centered approach that aligns with the institution's mission and long-term educational goals.



Contents
I. THE FRAMEWORK 2
II. IMPLEMENTING GUIDELINES 5
III. GENERAL CLASSROOM POLICIES 7
IV. DO'S AND DON'TS 8
A. Content Generation (e.g., Composing Essays) 8
B. Assessment (e.g., Recitation, Quizzes, Performance Tasks) 8
C. Research, Case Studies, and Terminal Reports 9
V. GUIDELINES FOR AI DECLARATION, USAGE TRANSPARENCY, AND DETECTION 10
A. AI Declaration 10
B. Giving Instructions 12
C. AI Detectors 12
D. Using Turnitin in mySOUL 12
VI. EXAMPLES OF AI INTEGRATION 20
PART I. IN THE CLASSROOM, ... 20
A. Content Generator 20
B. Pedagogy 21
C. Research 23
D. Assessment 24
PART II. IN THE AREA OF ADMINISTRATION AND MANAGEMENT, ... 26
A. Reference 26
B. Writing 28
C. Organizing 29
D. Reporting 31
INITIATIVES DONE 37

Figure 1. Silliman University AI Integration Framework

(Accessible at https://oursoul.su.edu.ph/OER/index.php/ourSOUL-OER/article/view/791)

The framework is built upon ten core principles:

1. Embracing AI as a tool for innovation,
2. Prioritizing continuous learning for all stakeholders,
3. Engaging in partnerships with local and global organizations,
4. Ensuring ethical practices in AI implementation,
5. Leveraging AI for personalized learning,
6. Optimizing administrative processes,
7. Fostering responsibility and accountability in AI use,
8. Promoting equitable access to AI technology, and
9. Enhancing customer service and
10. Aligning AI integration with institutional goals.

To operationalize these principles, the university developed a set of institutional policies and guidelines that govern AI use in instructional delivery and administrative functions. These include classroom protocols, content generation practices, ethical assessment procedures, and research standards. The framework also outlines transparency requirements, including AI use declarations, detection mechanisms, and illustrative use cases. These tools ensure that AI is integrated efficiently and ethically, emphasizing institutional alignment and user accountability.

Phased Implementation Process

The development of Silliman University's AI Integration Framework followed a structured, multi-phase approach designed to ensure strategic planning, community engagement, iterative development, and sustained capacity-building. Each phase, shown in Figure 2, is built upon the previous, allowing for the gradual and context-sensitive adoption of ethical AI practices within the university.

Phase 1: Reactive and Proactive Planning (Q4 2022)

In late 2022, the university's leadership began exploring the implications of Artificial Intelligence, particularly Generative AI (GenAI), within academic settings. The university did a careful observation and internal learning as an initial step. At this stage, the institution focused on gaining a foundational understanding of emerging AI technologies, their potential educational applications, and their associated risk based on international standards.

Phase 2: Community Engagement and Experimentation (Q1–Q3 2023)

During the first half of 2023, efforts shifted toward raising awareness and initiating dialogue within the university community. A university-wide seminar on AI was held on March 16, 2023, marking the beginning of open discussions on responsible AI use. Parallel to this, upskilling programs were launched to provide faculty, staff, and students with essential knowledge of AI concepts, ethics, and tools. Benchmarks were conducted with international organizations such as UNESCO and the United Nations. Collaborations were also emphasized to facilitate experimentation and the exchange of insights on best practices. These engagements enriched the institution's understanding of AI's practical opportunities, challenges, and ethical dimensions.

Phase 3: Framework Formulation and Approval (Q4 2023–Q1 2024)

In this phase, the focus shifted to drafting and refining the institutional AI integration framework. Insights from prior community engagements were

consolidated into preliminary strategies, which were presented to academic stakeholders, including teachers and management. Feedback mechanisms were implemented to ensure inclusivity, enabling inputs from faculty, staff, students, and parents. Concurrent upskilling initiatives continued to enhance digital competence across the university, supporting the transition from experimentation to structured planning.

Phase 4: Guidelines and Policy Development (Q2 2024 – Q3 2024)

The EdTech Committee undertook extensive research and policy drafting following the framework formulation. Department-level eLearning representatives were consulted to ensure that guidelines were responsive to the specific needs and instructional practices across disciplines. The resulting policy drafts include provisions for classroom implementation, AI use in research and assessments, and transparency requirements. These documents were reviewed by the members of the middle management and approved by the top management, with plans for institutional dissemination and stakeholder orientation.

Phase 5: Implementation and Continuous Improvement

The current phase focuses on the formal implementation of the AI integration framework. Institutional orientations are being conducted, and instructional policies are refined to align with departmental contexts. Upskilling programs remain ongoing and adaptive to stakeholder needs. A monitoring and evaluation system is also being developed to assess the framework's effectiveness and guide future improvements as AI technologies evolve.

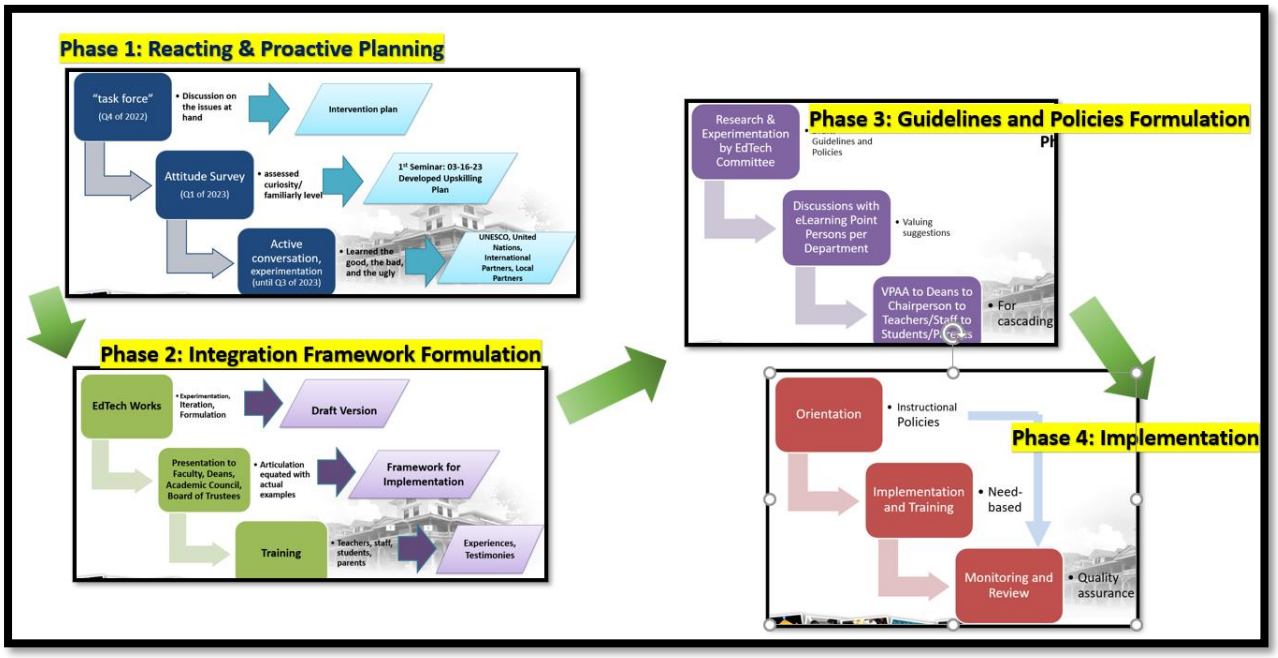


Figure 2. Phases of the Formulation of Silliman University AI Integration Framework

Silliman University's AI integration journey reflects a commitment to ethical innovation, participatory governance, and continuous learning. The university has established a working model for responsible AI adoption in higher education through deliberate planning, stakeholder collaboration, and responsiveness to global and local developments. The framework remains dynamic, with flexibility built into its design to accommodate technological advancements and emerging educational priorities.

KEY LESSONS AND STRATEGIES FOR EFFECTIVE AI INTEGRATION IN EDUCATION

The development and implementation of Silliman University's AI Integration Framework revealed several critical lessons that can guide other

educational institutions in adopting AI responsibly and effectively, as seen in Figure 3. These lessons offer a solid foundation for developing a responsible and effective AI integration strategy. Strong leadership, collaborative engagement, continuous capacity-building, and a commitment to ongoing improvement are essential to ensure that AI use reflects the core values of education and advances inclusive, future-ready learning environments. Further, these insights highlight the strategic, organizational, and cultural shifts necessary to ensure ethical and impactful AI integration.

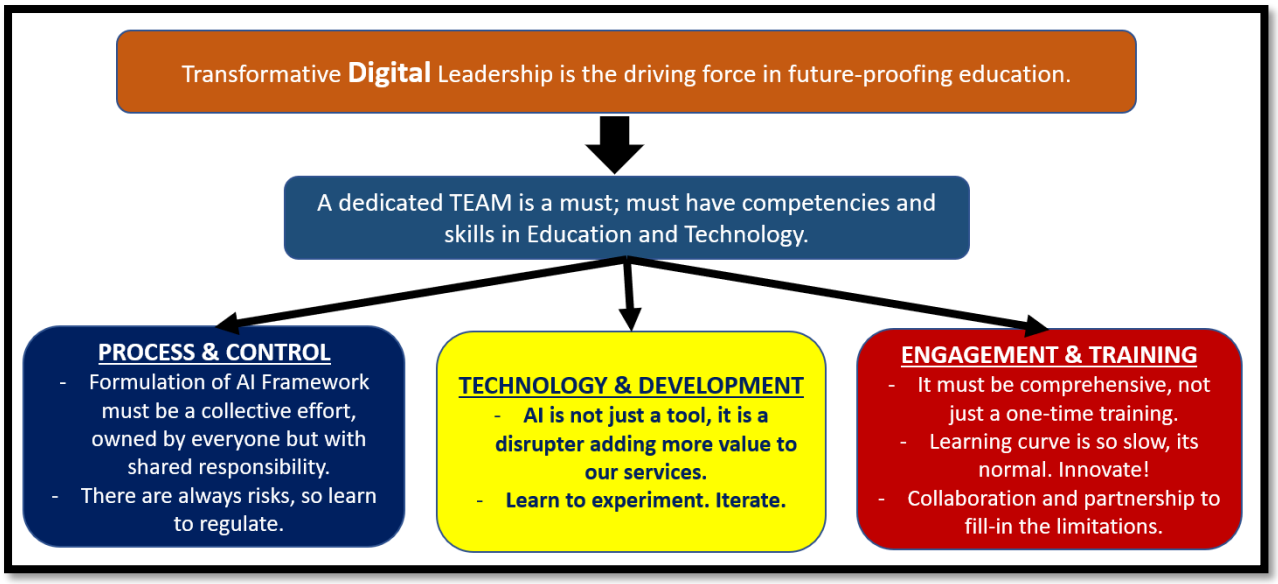


Figure 3. Schematic Diagram of Lessons Learned

Transformative Digital Leadership

Effectively future-proofing education requires strong leadership, specifically transformative digital leadership (McCarthy, Maor, McConney, & Cavanaugh, 2023). Institutional leaders must actively engage with the digital landscape, develop an informed understanding of AI technologies, and articulate a vision that balances innovation with ethical responsibility. Their role is vital in setting direction, fostering a culture of trust and experimentation, and mobilizing support across all levels of the organization.

Building a Competent and Multidisciplinary Team

Successful AI integration depends on a team combining education and technology expertise. Institutions must invest in developing multidisciplinary teams that address AI's pedagogical, technical, and ethical dimensions. Such teams can ensure that AI tools are aligned with instructional goals and institutional values while also managing the operational complexities of implementation.

Fostering a Collaborative and Inclusive Process

The formulation of an institutional AI framework must be collaborative, involving educators, administrators, students, and technical experts. This inclusive approach ensures that the framework is responsive to diverse needs and grounded in shared ownership. Recognizing that AI introduces systemic risks, it is essential to embed mechanisms for collective accountability, ethical oversight, and stakeholder feedback.

Recognizing AI as Both a Disruptor and an Enabler

AI is not merely an assistive tool but a transformative force that challenges existing educational models while creating new opportunities. Institutions must adopt a mindset of experimentation, iteration, and critical reflection. Leveraging AI for innovation and personalization requires careful balancing of its potential with the responsibility to mitigate risks such as bias, dependence, and inequity.

Ensuring Continuous Training and Stakeholder Engagement

Ongoing training is essential to equip educators, staff, and students with the skills to engage ethically and effectively with AI tools. These programs should be continuous, scalable, and responsive to evolving technologies. Institutions must acknowledge the steep learning curve associated with AI adoption and commit to building

a culture of lifelong learning, adaptability, and shared growth.

SUMMARY, CONCLUSION, AND RECOMMENDATIONS

This paper explored the ethical integration of Artificial Intelligence (AI) in education by examining global and industry frameworks, identifying common ethical principles, and analyzing a localized institutional implementation strategy. Through a qualitative integrative review, the paper synthesized key ethical principles from sources such as UNESCO, the European Union AI Act, the World Economic Forum, IEEE, industries, and cohort-based. Fourteen principles emerged from this review—ranging from transparency and fairness to accountability and inclusiveness—forming a practical guide for institutions adopting AI. The paper also presented the phased development of Silliman University's AI Integration Framework, highlighting how ethical principles can be translated into institutional policies, professional development programs, and adaptive governance structures.

The ethical integration of AI in education is both urgent and essential. While AI offers transformative benefits for teaching, learning, and administration, its use must be aligned with clearly defined ethical standards to prevent unintended consequences such as bias, data misuse, academic dishonesty, and inequity. The review underscores the need for institutions to move beyond theoretical commitments and adopt actionable strategies rooted in ethical values. The Silliman University experience demonstrates that ethical integration is achievable through strategic planning, community engagement, and sustained capacity-building. Institutions that adopt AI without a solid ethical foundation risk reinforcing educational inequalities and undermining trust in technology-enabled learning environments.

Drawing from the review of global and regional ethical frameworks and the institutional experience in developing a localized AI integration strategy, several recommendations are proposed to guide educational institutions in ethically integrating Artificial Intelligence. First, institutions should develop ethical AI frameworks grounded in universally recognized principles—such as prevention of harm, inclusiveness, accountability, and transparency—but contextualized to reflect the institution's specific vision, mission, and learner profile. These frameworks should not remain aspirational but be translated into operational policies and guidelines governing instruction, research, and administration. Second, institutions are encouraged to adopt a phased and participatory approach to AI integration. As demonstrated in the case of Silliman University, implementation that begins with proactive planning, community engagement, and iterative development fosters institutional ownership and responsiveness to emerging risks and opportunities. Each phase—from policy formulation to pilot testing and review—should involve consultations with key stakeholders, including educators, learners, administrators, and IT professionals. Third, leadership must play a strategic and transformative role. School leaders should cultivate a mindset of ethical innovation by championing responsible AI use, supporting faculty development, and aligning technology initiatives with the institution's educational philosophy. Continuous upskilling is also critical. Institutions must provide sustained training programs to build AI literacy, ethical awareness, and adaptive pedagogical competencies among teachers and students. Additionally, institutions should establish robust evaluation, monitoring, and policy refinement mechanisms. These mechanisms may include ethical review boards, data governance protocols, and feedback channels that ensure accountability and adaptability. Finally, collaboration and knowledge exchange should be strengthened with other institutions, both locally and internationally. Sharing implementation experiences, benchmarking ethical practices, and participating in joint research will enable schools to navigate AI adoption more effectively, particularly in resource-constrained environments. These recommendations ensure that AI is integrated efficiently, ethically, inclusively, and sustainably in the educational landscape. The focus should be on what AI can do and what we, as responsible educators and leaders, can achieve by leveraging AI for the greater good (UNESCO, 2021).

Acknowledgment

The author extends appreciation to the National Interdisciplinary Virtual Research Conference participants, held via Zoom on September 26–27, 2024, for their constructive comments and suggestions on earlier versions of this paper.

REFERENCES

Almeida, D., Shmarko, K., & Lomas, E. (2022). The ethics of facial recognition technologies, surveillance, and accountability in an age of

artificial intelligence: a comparative analysis of US, EU, and UK regulatory frameworks. *AI Ethics, 2*(3), 377–387. doi:https://link.springer.com/article/10.1007/s43681-021-00077-w

Baker, T., Smith, L., & Anissa, N. (2019). *Educ-AI-tion Rebooted?* NESTA. Retrieved from https://media.nesta.org.uk/documents/Future_of_AI_and_education_v5_WEB.pdf

BBC News. (2020, August 20). A-levels and GCSEs: How did the exam algorithm work? *BBC News*. Retrieved from https://www.bbc.com/news/explainers-53807730

Chen, L., Chen, P., & Lin, Z. (2020). Artificial Intelligence in Education: A Review. *IEEE Access*, 75264-75278. doi:https://ieeexplore.ieee.org/document/9069875

Chui, M., Hazan, E., Roberts, R., Singla, A., Smage, K., Sukharevsky, A., . . . Zemmel, R. (2023). *The economic potential of generative AI.* McKinsey & Company. Retrieved from https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#/

Department of Trade and Industry - Republic of the Philippines. (2021, May 5). *DTI to launch National Artificial Intelligence Roadmap*. Retrieved from Department of Trade and Industry - Republic of the Philippines: https://www.dti.gov.ph/archives/national-artificial-intelligence-roadmap/

Directorate-General for Communication. (2024, August 1). AI Act enters into force. *European Commission*. Retrieved from https://commission.europa.eu/news-and-media/news/ai-act-enters-force-2024-08-01_en

European Parliament. (2023, August 6). *EU AI Act: first regulation on artificial intelligence*. Retrieved from Europian Parliament: https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., . . . Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds & Machines*, 689-707. doi:https://doi.org/10.1007/s11023-018-9482-5

Heaton, J. (2018). Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning. *Genet Program Evolvable Mach*, 305-307. doi:https://doi.org/10.1007/s10710-017-9314-z

Heidegger, M. (1997). The Question Concerning Technology. In D. F. Krell, *Martin Heidgger: Basic Writings* (pp. 287-317). New York: Harper & Row. Retrieved from https://www2.hawaii.edu/~freeman/courses/phil394/The%20Question%20Concerning%20Technology.pdf

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial Intelligence in Education Promises and Implications for Teaching and Learning.* MA, USA: UCL. Retrieved from https://www.researchgate.net/publication/332180327_Artificial_Intelligence_in_Education_Promise_and_Implications_for_Teaching_and_Learning

IEEE. (2019). *Ethically Aligned Design.* Retrieved from https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf

Jonas, H. (2014). Technology and Responsibility: Reflections on the New Tasks of Ethics. In R. Sandler, *Ethics and Emerging Technologies* (pp. 37-47). Palgrave Macmillan, London. doi:https://doi.org/10.1057/9781137349088_3

Kurni, M., Mohammed, M., & Srinivasa, K. (2023). Natural Language Processing for Education. In C. Springer, *A Beginner's Guide to Introduce Artificial Intelligence in Teaching and Learning.* doi:https://doi.org/10.1007/978-3-031-32653-0_3

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 436–444. doi:https://doi.org/10.1038/nature14539

Leiss, W. . (1994). Review of Risk Society, Towards a New Modernity, by U. Beck, M. Ritter, S. Lash, & B. Wynne]. *The Canadian Journal of Sociology / Cahiers Canadiens de Sociologie*, 544-547. doi:https://doi.org/10.2307/3341155

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence Unleashed. An argument for AI in Education.* London: Pearson. Retrieved from https://www.pearson.com/content/dam/corporate/global/pearson-dot-com/files/innovation/Intelligence-Unleashed-Publication.pdf

Marcial, D. E. (2017). ICT Social and Ethical Competency among Teacher Educators in the Philippines. *Journal of Information Technologies and Learning Tools, 57*(1), 93-103. doi:http://journal.iitta.gov.ua/index.php/itlt/article/view/1533/1137

Marcial, Dave E.; Silliman University. (2024). *Integration of Artificial Intelligence into Silliman Education.* Retrieved from https://oursoul.su.edu.ph/OER/index.php/ourSOUL-OER/article/view/791/115

Marcus, G., & Davis, E. (2019). *Rebooting AI: Building Machines we can Trust.* New York, NYUnited States: Pantheon Books.

Mateen, H. (2018). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. *Berkeley Journal of Employment and Labor Law, 39*(1), 285–292. Retrieved from https://www.jstor.org/stable/26732553

McCarthy, A. M., Maor, D., McConney, A., & Cavanaugh, C. (2023, April). Digital transformation in education: Critical components for leaders of system change. *Social Sciences & Humanities Open, 8*(1). doi:https://doi.org/10.1016/j.ssaho.2023.100479

Microsoft. (n.d.). *The Microsoft Responsible AI Standard*. Retrieved from https://www.microsoft.com/en-us/ai/principles-and-approach

Schatsky, D., Gurumurthy, R., & Muraskin, C. (2021). Cognitive technologies: The real opportunities for business. *Deloitte Review Issue*. Retrieved from https://www2.deloitte.com/us/en/insights/deloitte-review/issue-16/cognitive-technologies-business-applications.html

UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence. *Intergovernmental Meeting of Experts.* Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000377897

UNESCO. (2022). *Ethics of Artificial Intelligence*. Retrieved from UNESCO - Artificial Intelligence: https://unesdoc.unesco.org/ark:/48223/pf0000381137

UNESCO. (2023). *Readiness assessment methodology: a tool of the Recommendation on the Ethics of Artificial Intelligence.* Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000385198

World Economic Forum. (2020). *The Future of Jobs Report 2020.* World Economic Forum. Retrieved from https://www.weforum.org/publications/the-future-of-jobs-report-2020/

World Economic Forum. (2024, January 18). *I and education: Kids need AI guidance in school. But who guides the schools?* Retrieved from https://www.weforum.org/agenda/2024/01/ai-guidance-school-responsible-use-in-education/

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education– where are the educators? *International Journal of Educational Technology in Higher Education*. doi:ttps://doi.org/10.1186/s41239-019-0171-0