Augmented Measurement Framework for Dynamic Validity and Reciprocal Human-AI Collaboration in Assessment

Taiwo Feyijimi[1], Daniel Oyeniran[2], Olukayode Apata[3], Henry Makinde[4], Hope Adegoke[4], John Ajamobe[3], Justice Dadzie[2],
[1] College of Engineering, University of Georgia,
[2]College of Education, University of Alabama,
[3]College of Education and Human Development, Texas A&M University,
[4]College of Education, University of North Carolina, Greensboro.
Correspondence: taiwo.feyijimi@uga.edu

Abstract

The swift penetration of Generative AI presents unprecedented opportunities and profound challenges for educational measurement, straining traditional validity and authenticity. This study introduces the Augmented Measurement Framework (AMF), a novel proactive conceptual model designed to transcend reactive responses by fostering synergistic human-AI collaboration. Grounded in four core principles, the AMF champions: Reciprocal Co-creation, for continuous human-AI learning and development within the assessment lifecycle; a Continuous Validity Ecosystem, ensuring dynamic measurement quality as AI evolves and contexts shift; Explainable Augmentation, enhancing human judgment through transparent AI insights and ethical deployment; and Pedagogical Resonance, aligning AI tools with sound learning outcomes and authentic assessment practices. This framework offers a paradigm shift for creating valid, fair, and pedagogically sound assessments that empower educators and learners alike. The paper explores its practical applications, policy implications, and charts a critical research agenda for advancing this essential framework.

1 Introduction

The educational landscape is currently navigating a period of transformation, largely catalyzed by the rapid advancements and increasing ubiquity of Artificial Intelligence (AI), particularly Generative Artificial Intelligence (GenAI). Tools such as ChatGPT and other sophisticated AI systems have evolved from simple text-generation utilities into powerful engines capable of producing complex, human-like content across a vast spectrum of domains (Khlaif et al., 2025). This technological surge offers immense potential for innovation in teaching, learning, and, critically, educational assessment. However, it simultaneously presents formidable challenges that shake the foundations of established pedagogical and measurement practices.

The core of the current predicament lies in the capacity of GenAI tools to enable students to produce outputs that may not genuinely reflect their own competencies, understanding, knowledge, and skills (Corbin et al., 2025). This capability directly confronts traditional assessment methods designed to ascertain individual student learning. AI's integration into academic practices is so profound that it blurs the line between acceptable assistance and inappropriate delegation at every stage of the assessment process, from initial ideation through to final editing (Corbin et al., 2025). The current situation has fostered a significant degree of uncertainty among educators and students alike, leading to what some describe as a sense of unease, even trauma, arising from the rapid erosion of established pedagogical norms and the challenge to professional identity, as they attempt to reconcile their educational values with the new realities imposed by AI technology (Corbin et al., 2025).

In response to these disruptions, numerous frameworks have emerged to help educators adapt, aiming to articulate boundaries between acceptable and unacceptable uses of AI in assessment (Corbin et al., 2025). Despite variety in the reasons for assessment redesign, most of the reasons are responsive focusing on developing "AI-resistant" assessments or implementing policies of simple prohibition or permission (Khlaif et al., 2025). However, these often reactive and incremental approaches frequently fall short in fostering a deeply integrated future, either limiting the transformative potential of AI or failing to address the fundamental redefinition of learning and assessment required in an AI-pervaded educational environment. This paper posits that the current juncture demands more than incremental adjustments or reactive measures. It calls for a proactive and theoretically grounded rethinking of the role of AI in educational mea-

Proceedings of the Artificial Intelligence in Measurement and Education Conference (AIME-Con) – Volume 1: Full Papers, pages 290–296

surement. The central purpose advanced here is the need for a new conceptual framework, the Augmented Measurement Framework (AMF). The AMF seeks to move beyond the "assessment crisis" by proposing a model where human intelligence and artificial intelligence work synergistically (Feyijimi et al., 2025) to create assessment processes that are more valid, fair, transparent, and ultimately, more conducive to authentic learning.

2 Conceptual Foundation

As educational institutions grapple with the integration of GenAI, various conceptual models have been proposed to guide practice and policy. One such model, the "Against, Avoid, Adopt, and Explore" framework, offers a pragmatic lens for faculty members to consider their responses to AI in assessment, reflecting differing levels of engagement and concern (Khlaif et al., 2025). Another significant contribution is the Human-Centric AI-First (HCAIF) framework, developed through extensive research and experimentation (Verhoeven & Hor, 2025). The HCAIF model is built upon five pillars: Preparation, Personalized Learning, Classroom Engagement, Summative Assessment, and Personalized Monitoring, and is underpinned by the crucial factors of Attribution (students showing how AI was used) and Reflection (students analyzing their AI use).

The challenges accompanying these adaptations are substantial. A primary concern is the difficulty in distinguishing between authentically student-produced work and AI-assisted or AI-generated submissions, a problem that strikes at the heart of academic integrity (Khlaif et al., 2025). Furthermore, the time and resource demands for redesigning assessments to be "AI-resistant" or to meaningfully incorporate AI are significant, often straining already burdened educators (Khlaif et al., 2025). Equity and accessibility concerns also loom large; disparities in student access to AI tools and digital infrastructure can create unfair assessment landscapes (Khlaif et al., 2025).

Compounding these issues are resistance to change from both faculty and students accustomed to traditional methods, and a pervasive lack of clear institutional guidelines and adequate professional development for educators (Khlaif et al., 2025). To understand how educational communities navigate these changes, particularly in defining acceptable AI use, concepts from sociology, such as boundary work and social boundary theory, offer valuable analytical tools (Corbin et al., 2025). These theories help explain how groups construct, maintain, and negotiate symbolic boundaries between legitimate and illegitimate practices, especially during periods of technological disruption (Corbin et al., 2025). To bridge this gap and cultivate the sophisticated human-AI partnership required, the concept of 'Augmented Intelligence' offers a powerful paradigm for envisioning the future of AI in education.

In an augmented intelligence model, the role of the teacher undergoes a significant transformation. Teachers shift from being the primary purveyors of knowledge to becoming facilitators, mentors, and guides who help students navigate personalized learning paths supported by AI (Chiu & Rospigliosi, 2025; Feyijimi et al., 2025). This shift allows educators to focus on higher-order skills such as critical thinking, ethical reasoning, creativity, and emotional intelligence (Feyijimi et al., 2025).

The integration of AI into educational measurement carries a profound ethical responsibility. As AI systems become more powerful and pervasive, ensuring their ethical development and deployment is not merely an adjunct consideration but a foundational requirement. There are numerous ethical concerns in AI driven assessment such as algorithmic bias (Feyijimi et al., 2025; Walden University, 2025), extensive data collection required (Bhadwal, 2024), academic integrity (Feyijimi et al., 2025; Khlaif et al., 2025) among others. A proactive approach to mitigating the ethical concerns goes beyond reducing harm but designing AI systems that embody and promote pedagogical values and equitable learning opportunities through the very act of measurement (Wynants et al., 2025). This proactive stance, which forms an intrinsic element of the proposed Augmented Measurement Framework, ensures that ethical considerations are not merely reactive safeguards but are foundational to the design and deployment of AI in assessment, fostering human flourishing and equity from inception. For AI to be ethically deployed and effectively integrated into educational measurement, its decision-making processes cannot remain opaque. Explainable AI (XAI) addresses this by referring to AI systems that can reveal how their decisions and recommendations are made, moving away from the "black box" paradigm where AI reasoning is hidden (European Commission, 2024; Feyijimi et al.,

2025). The Defense Advanced Research Projects Agency (DARPA) defines XAI systems as those capable of explaining their reasoning, highlighting their strengths and limitations, and forecasting their future behavior (Gunasekara & Saarela, 2025). XAI is thus not merely a technical feature but a fundamental enabler of ethical AI, fostering trust, enabling informed human oversight, and supporting the pedagogical goals of fairness and understanding in AI-augmented educational measurement.

3 Dynamic Validity in AI-Augmented Assessment

Traditional conceptions of validity in educational measurement, while foundational, face significant challenges when applied to the rapidly evolving landscape of AI-augmented assessment. Classical validity theory, culminating in Messick's unified framework, encompasses various forms of evidence (e.g., content-related, construct-related, criterion-related) to support the interpretation and use of assessment scores for specific purposes (Aghazadeh et al., 2015; Messick, 1995). AI-driven assessments, particularly those employing machine learning, introduce several complexities that traditional validation approaches struggle to accommodate such as adaptivity, continuous learning, opacity and complexity of data.

To address these limitations, this paper proposes the concept of Dynamic Validity. Dynamic Validity is conceptualized as an ongoing, context-aware, and adaptive evaluation of an AI assessment system's inferences, decisions, and actions. It is not a one-time certification but a continuous process that monitors and seeks to ensure the system's ability to maintain measurement quality – including accuracy, fairness, relevance, and utility – as the AI model itself evolves, student populations change, instructional contexts shift, and our understanding of the assessed constructs develops.

Dynamic Validity is a multifaceted construct, encompassing several interconnected dimensions which include adaptive validity, continuous validity and ongoing assessment, dynamic fairness and algorithmic equity, explainability's role in dynamic validation and evolving consequential validity. These dimensions collectively underscore the necessity of adaptive mechanisms to re-evaluate validity as AI models evolve, the continuous monitoring of measurement quality for sustained reliability, the proactive pursuit of algorithmic equity to mitigate bias, the integration of explainability for transparent validation, and the iterative assessment of the broader educational and societal impacts of AI systems.

The concept of Dynamic Validity resonates with approaches in other fields that deal with rapidly evolving evidence bases such as Living Systematic Reviews (LSRs) in healthcare which ensures continuously updated review of new research findings and provide current and relevant evidence to support decision-making (Lansky & Wethington, 2020). Analogously, AI-augmented assessment systems require "Living Validity Arguments."

This implies establishing processes for periodic re-validation, transparent reporting of performance metrics, and mechanisms for incorporating new research findings about the AI's effectiveness and impact. Such an approach acknowledges that validity is not a fixed state to be achieved, but a continuous commitment to ensuring the responsible and effective use of AI in educational measurement. Indeed, Dynamic Validity is not merely a theoretical concept but a foundational pillar, intrinsically woven into the fabric of the Augmented Measurement Framework, ensuring that AI's evolution in assessment remains aligned with core educational values and robust measurement standards through constant vigilance.

4 Augmented Measurement Framework (AMF)

The AMF is proposed as a conceptual model (Figure 1) designed to guide the development, implementation, and ongoing evaluation of AI in educational assessment. It aims to foster a synergistic relationship between humans and AI, moving beyond simplistic views of AI as either a threat to be mitigated or a panacea for all assessment challenges (Feyijimi et al., 2025). The AMF is structured around four core, interconnected principles that collectively promote an assessment ecosystem that is valid, fair, transparent, pedagogically sound, and empowering for both educators and learners. This particular combination and emphasis on their synergistic interaction distinguishes the AMF, moving beyond singular focus areas to address the holistic challenges and opportunities of AI in educational measurement by envisioning a truly collaborative future. These principles are not discrete entities but rather deeply interdependent, forming a dynamic ecosystem where Reciprocal

Co-creation informs the ethical development of assessment tools, whose measurement quality is then continuously assured through the Continuous Validity Ecosystem, while Explainable Augmentation fosters trust and provides transparent insights, all serving to optimize Pedagogical Resonance with sound educational practices and desired learning outcomes.

The reciprocal co-creation principle of the AMF posits that educators and AI systems should engage in a continuous, bi-directional process of learning and development within the assessment lifecycle. It moves beyond the traditional paradigm where educators are passive end-users of pre-packaged AI tools or where AI functions merely as a static instrument. Mechanisms for enacting Reciprocal Co-creation include collaborative design from inception, AI-assisted, educator-refined content generation, nuanced feedback loops for AI improvement and interactive machine learning.

The continuous validity ecosystem principle asserts that validity in AI-augmented assessment is not a static property achieved at a single point in time, but an ongoing process embedded within an ecosystem of human oversight and AI adaptation. Mechanisms for establishing a Continuous Validity Ecosystem include systematic human review and oversight, AI-assisted prioritization of human review, ongoing performance monitoring and adaptation, living validity argument documentation and avoiding AI-only training cycles.

The explainable augmentation principle underscores that the primary role of AI in educational measurement is to augment and enhance – not replace – the professional judgment of educators. The principle can be achieved with the following process integration of XAI techniques, actionable insights for educators, supporting pedagogical diagnosis, and building trust through transparency. The final principle of pedagogical resonance proposed that AI assessment tools and processes must be subservient to, and supportive of, sound pedagogical principles and desired learning outcomes. Mechanisms for ensuring Pedagogical Resonance include alignment with learning theories, formative and actionable feedback, educator customization and control, emphasis on authentic assessment and holistic evaluation impact.

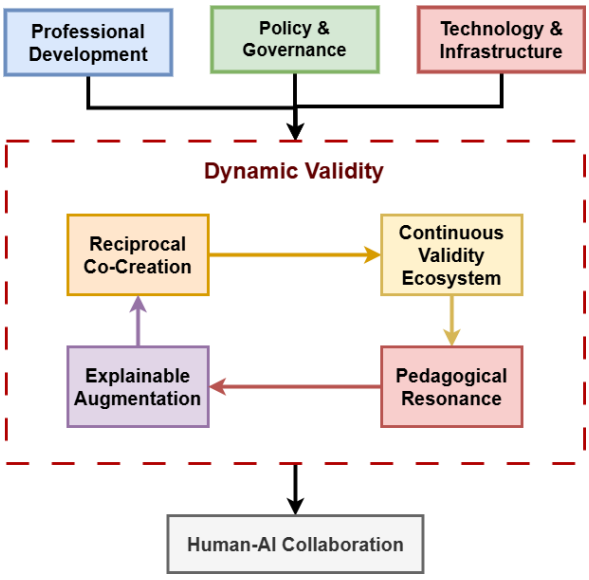

Figure 1: Augmented Measurement Framework. This figure visualizes the interconnectedness of the pillars and components within the AMF. Three enabling pillars: Professional Development, Policy & Governance, and Technology & Infrastructure feed into a Dynamic Validity envelope that continuously surrounds a four-step co-creative cycle consisting of Reciprocal Co-creation, Continuous Validity Ecosystem, Explainable Augmentation, and Pedagogical Resonance. This validity lens ensures that AI-supported assessments are fair, transparent, and aligned with learning goals. The cycle culminates in true human-AI collaboration, where educators and algorithms work together to produce valid, interpretable, and learner-centered results.

5 Operationalizing the Augmented Measurement Framework

The principles of the Augmented Measurement Framework (AMF) – Reciprocal Co-creation, Continuous Validity Ecosystem, Explainable Augmentation, and Pedagogical Resonance – provide a conceptual blueprint. Translating this blueprint into practice requires considering illustrative scenarios, implications for teacher professional development, necessary policy adjustments, and supportive technological enablers. Operationalizing the AMF is not about implementing a fixed set of tools, but fostering an evolving ecosystem where human expertise and AI capabilities synergize to improve educational measurement.

To concretize how the AMF might function, an illustrative scenario will be discussed in this proposal.

A consortium of STEM educators, representing diverse cultural backgrounds, collaborates with AI developers to create more culturally responsive as-

sessment items. The process begins with educators identifying potential biases (e.g., culturally specific contexts, language) in existing STEM assessment banks. The AI, equipped with NLP capabilities and access to diverse textual and cultural knowledge bases, assists in generating alternative assessment items or scenarios that are potentially more inclusive or relevant to different student populations (Martin et al., 2025). Educators then meticulously vet these AI-generated alternatives for cultural appropriateness, linguistic clarity, pedagogical soundness, and alignment with STEM learning objectives (Reciprocal Co-creation, Pedagogical Resonance). They might, for instance, adapt an AI-suggested physics problem to use examples or contexts familiar to students in their specific communities. The AI learns from these educator modifications and preferences, improving its ability to generate culturally sensitive content over time. Furthermore, the AI's Explainable AI (XAI) capabilities provide transparency into why certain items were flagged for bias or how alternative items were generated, offering educators actionable insights into the underlying patterns and reasoning, thereby enhancing trust and facilitating more informed collaborative decisions (Explainable Augmentation). The Continuous Validity Ecosystem involves ongoing monitoring of item performance across different demographic groups to ensure fairness and identify any emergent biases, with educators playing a key role in interpreting these data and guiding further AI refinement (Continuous Validity for fairness).

The successful implementation of the AMF heavily relies on equipping educators with new knowledge, skills, and dispositions. Professional development (PD) must evolve beyond basic operational training for specific AI tools. Key areas for PD include comprehensive AI literacy, data literacy for assessment, ethical AI use and bias mitigation, skills for human-AI collaboration and co-creation and pedagogical adaptability. Effective PD for the AMF is not just about transferring technical skills; it's about fostering a new "assessment mindset." Educators should see themselves as augmented decision-makers, critical evaluators of AI-generated insights, and active agents in shaping the AI assessment ecosystem, rather than passive recipients of AI directives or mere implementers of externally developed tools.

Supportive institutional and systemic policies are crucial for the AMF to take root. Achieving this requires ethical guidelines and data governance, investment in research and AMF aligned development, procurement and development standards, equity and access, and educator agency and time allocation. Policy frameworks supporting the AMF must strike a balance between fostering innovation and establishing robust safeguards.

The vision of the AMF relies on the availability and continuous improvement of certain technological capabilities such as user friendly XAI interfaces, platforms for collaborative designs and validation, robust and secure data infrastructure, interoperability standard and advanced AI models. Operationalizing the AMF is an ambitious but necessary endeavor. It requires a concerted effort from researchers, developers, educators, and policymakers to build the tools, practices, and supportive environments that can realize the promise of a truly augmented and human-centered approach to educational measurement.

6 Navigating Challenges and Charting Future Directions

While the Augmented Measurement Framework (AMF) offers a promising vision for the future of AI in educational assessment, its realization is not without significant challenges. Successfully navigating these hurdles and advancing the field requires a clear understanding of the practical, technical, ethical, and societal complexities involved, alongside a focused research agenda. However, the very design of the AMF, with its emphasis on Reciprocal Co-creation, a Continuous Validity Ecosystem, Explainable Augmentation, and Pedagogical Resonance, inherently provides mechanisms to proactively address many of these challenges, transforming potential obstacles into opportunities for innovation within a human-centered design.

The implementation of AMF-aligned AI assessment systems faces several practical and technical obstacles such as scalability of human oversight, data requirements and quality, computational costs and infrastructure, developing intuitive XAI interfaces and interoperability and integration.

Beyond technical issues, the ethical and societal implications of AI-augmented measurement demand careful and continuous attention such as mitigating algorithmic bias and ensuring fairness, protecting student data privacy and security, academic integrity in the age of GenAI, the digital divide and equitable access and over-reliance on AI

and De-skiling (Feyijimi et al., 2025). Successfully navigating these multifaceted challenges hinges on fostering deep, interdisciplinary collaboration.

The AMF is a conceptual framework that requires empirical validation and refinement. A robust research agenda is needed to explore its principles and operationalization. There is need to develop and validate dynamic validity metrics, structure a longitudinal impact study, and investigating pedagogies for human-AI co-creation in assessment are some of the most crucial research agenda for achieving the goals of AMF. Among these, the immediate imperative lies in developing and empirically validating dynamic validity metrics, as these are fundamental to establishing the trustworthiness and utility of AMF-aligned systems in diverse educational contexts.

7 Conclusion

The integration of Artificial Intelligence into educational measurement stands at a critical juncture. The transformative power of AI, particularly GenAI, offers unprecedented opportunities to personalize learning, provide richer feedback, and gain deeper insights into student understanding. However, this potential is accompanied by significant challenges to traditional assessment paradigms, raising fundamental questions about validity, fairness, authenticity, and the very nature of human learning in an AI-pervaded world. This paper has argued that navigating this complex terrain requires a paradigm shift, a move away from viewing AI as a mere instrument of automation or efficiency, towards conceptualizing it as a collaborative partner in a dynamic and ethically grounded assessment ecosystem.

The Augmented Measurement Framework (AMF) has been proposed as a conceptual guide for this shift. Unlike prior reactive or incremental approaches, the AMF provides a proactive, holistic blueprint for integrating AI, championing a synergistic relationship where the unique strengths of human intelligence are weaved into the fabrics of AI-augmented assessment. Built upon four interconnected principles: Reciprocal Co-creation, Continuous Validity Ecosystem, Explainable Augmentation, and Pedagogical Resonance – the AMF offers a pathway to harness AI's capabilities responsibly. It champions a synergistic relationship where the unique strengths of human intelligence (pedagogical expertise, ethical judgment, contextual understanding) and artificial intelligence (data processing, pattern recognition, adaptive capabilities) are combined to create assessment processes that are more than the sum of their parts.

Operationalizing the AMF is undoubtedly a complex undertaking, fraught with technical, practical, and ethical challenges. It demands significant investment in research, the development of new technologies and professional competencies, and the establishment of supportive policy environments. However, the pursuit of such a framework is not merely a technical or academic exercise; it is an ethical imperative. The ultimate success of AI in educational measurement will not be judged solely by its technical sophistication or its efficiency gains. Instead, it will be measured by its capacity to foster more humane, equitable, and meaningful learning experiences for all students (Khlaif et al., 2025). AI must be a tool that helps to close achievement gaps, not widen them; that promotes critical thinking and creativity, not rote compliance; and that empowers educators, not diminishes their professional role.

Adopting frameworks like the AMF represents more than a technical upgrade for educational institutions; it signifies a cultural shift. It requires a collective commitment to ongoing learning, critical reflection, interdisciplinary collaboration, and adaptive governance of AI technologies. Realizing this vision demands a concerted, sustained effort from researchers, developers, educators, and policymakers alike, ensuring that the necessary tools, practices, and supportive environments are co-created to serve this profoundly human endeavor. The journey towards this future is ongoing, but by embracing principles of synergy, dynamic validity, transparency, and pedagogical integrity, we can strive to ensure that AI serves to make educational assessment not only more powerful but also more purposeful and profoundly human.

References

Aghazadeh, M. A., Jayaratna, I. S., Hung, A. J., Pan, M. M., Desai, M. M., Gill, I. S., & Goh, A. C. (2015). External validation of global evaluative assessment of robotic skills (GEARS). Surgical Endoscopy, 29, 3261–3266.

Bhadwal, A. (2024, September). Benefits and challenges of implementing AI in student assessment. Mera Tutor. https://www.meratutor.ai/blog/ai-in-student-

assessment/, accessed on June 12, 2025.

Chiu, T. K., & Rospigliosi, P. A. (2025). Encouraging human-AI collaboration in interactive learning environments. Interactive Learning Environments, 33(2), 921–924.

Corbin, T., Dawson, P., Nicola–Richmond, K., & Partridge, H. (2025). 'Where's the line? It's an absurd line': towards a framework for acceptable uses of AI in assessment. *Assessment & Evaluation in Higher Education*, 1–13.

Feyijimi, T. R., Aliu, J. O., Oke, A. E., & Aghimien, D. O. (2025). ChatGPT's Expanding Horizons and Transformative Impact Across Domains: A Critical Review of Capabilities, Challenges, and Future Directions. Computers, 14(9), 366. https://doi.org/10.3390/computers14090366

Khlaif, Z. N., Alkouk, W. A., Salama, N., & Abu Eideh, B. (2025). Redesigning assessments for AI-enhanced learning: A framework for educators in the generative AI era. Education Sciences, 15(2), 174.

Lansky, A., & Wethington, H. R. (2020). Living systematic reviews and other approaches for updating evidence. American Journal of Public Health, 110(11), 1687.

Martin, A. F., Tubaltseva, S., Harrison, A., & Rubin, G. J. (2025). Participatory co-design and evaluation of a novel approach to generative AI-integrated coursework assessment in higher education.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American psychologist, 50(9), 741.

Troussas, C., Krouska, A., & Sgouropoulou, C. (2025). The role of augmented intelligence and pedagogical theories in digital learning. In Human-computer interaction and augmented intelligence: The paradigm of interactive machine learning in educational software (pp. 39–93). Cham: Springer Nature Switzerland.

Verhoeven, B., & Hor, T. (2025, February). Higher education is undergoing a paradigm shift toward more human-centric, adaptive learning models that will equip students for an AI-powered future. Association to Advance Collegiate Schools of Business (AACSB). https://www.aacsb.edu/insights/articles/2025/02/a-framework-for-human-centric-ai-first-teaching, accessed on June 12, 2025.

Walden University. (2025). 5 pros and cons of AI in the education sector. What teachers need to know about artificial intelligence in the classroom. Available at https://www.waldenu.edu/programs/education/resource/five-pros-and-cons-of-ai-in-the-education-sector, accessed on June 12, 2025.

Wynants, S., Childers, G., De La Torre Roman, Y., Budar-Turner, D., & Vasquez, P. (2025). ETHICAL principles AI framework for higher education.