Grenze International Journal of Engineering and Technology, June Issue

AI Literacy for the Future: Responsible usage, Critical Thinking and Ethical Awareness

Bharathi S[1], Binayak Dutta[2] and Somnath Sinha[3]
[1-3]Christ University, Bangalore, India
Email: bharathi.s@bcah.christuniversity.in, binayak.dutta@christuniversity.in, Somnath.sinha@christuniversity.in

***Abstract*— This paper examines the importance of AI literacy beyond technical skills, emphasizing ethical awareness and critical thinking. As AI tools become readily available, users are quick to adopt them without recognizing associated risks such as data bias, privacy issues, and simply too much reliance. This paper explores what grounds AI literacy should contain: responsible usage, interrogating AI outcomes, and knowing the implications of AI in the real world. It considers the inclusion of abstraction of machine learning basics in higher education, and importantly, how much use might help build critical thinking skills and ethical thinking skills in users. This paper contends that it is important to encourage that users be aware of generative AI and the implications without being overly reliant on it; because, without awareness, users may trust AI blindly, ethical thinking as part of the training associated with knowledge of AI, providing ways that machine learning can serve as the foundation needed to engage with AI in a more informed and ethical manner, while advocating for methods of teaching with AI that treats this technology as tool for growth instead of convenience.**

***Index Terms*— AI, ML, Face perception, Facial emotion recognition, Software evaluation, Image processing, MER, Mental health (MH), Guidance counselling.**

I. INTRODUCTION

Artificial Intelligence (AI) has rapidly evolved from a futuristic dream to potentially one of the largest transformational forces of our era. Initially envisioned as science fiction, AI is now integral to everyday life and revolutionizing various sectors, such as education, healthcare, finance, and media. This revolution has introduced new efficiencies and functionality, in addition to altering the nature of human interaction with machines and information. Machine learning (ML) is at the heart of this revolution, a type of AI that allows a machine or model to learn automatically how to process data, recognize patterns, and make predictions. This ability of AI systems to constantly learn and improve from large sets of data and experience is where the power and disruptive potential of such technologies lie. Whereas the potential for good and the potential for advancing human growth, creativity, and problem solving through these innovations are stimulating; so too are many of the moral and ethical dilemmas that cannot be overlooked.

One of the most urgent concerns is the development of algorithmic bias where AI algorithms unwittingly exacerbate or even increase social injustices. E.g., hiring algorithms are biased against certain demographics, lending algorithms unfairly curtail access to capital, and recommendation algorithms influence consumer behavior in ways that reinforce stereotype.

Grenze ID: 01.GIJET.12.2.541

These challenges point out that AI is not merely an impartial tool but one reflecting human values, assumptions, and limitations. Kong [1] contends that these problems need to be addressed by developing ethical awareness early on, if not in primary school, where children start constructing ethical frameworks informing how they understand fairness, responsibility, and justice. Incorporating AI ethics training early can enable future generations to interact responsibly with intelligent systems instead of passively swallowing their outputs. In the same vein, Gabriel [2] emphasizes the importance of generative AI tools as potential access points for AI literacy and stimulating ethical thinking. Rather than just consuming information produced by AI, learners should be invited to question, investigate, and assess what they have encountered. Instead, this exposition will help them cultivate a certain mindset where what constitutes an ethical choice will shape their interactions with technology. Thus, the development of ethical awareness is not just an exercise in regulatory compliance. Ethical awareness is a key part of AI that is sustainable.

With this contextual underlying this paper discusses ethical awareness and critical reasoning as vital and interlinked aspects of responsible practice of AI and machine learning. It considers how ethics can be brought in to the AI lifecycle that comes in more robust ways, such that it is not limited to the beginning of design and planning phases to the implementation, deployment and ultimate use in practice. Using contemporary literature and case studies [3], [4] it discusses the gap between ethics and practice in AI systems. It also discusses emerging possibilities around "Ethics-as-a-Service" which could help organize support to users and developers for ethical reasoning.

II. Literature Review

Choung et al. [5] explored trust and ethics in AI while also providing a conceptual consideration of the risks of bias, opacity, and misuse of AI, when systems are deployed with no oversight but they did not provide or evaluate an actual model. Based on broader concepts of awareness, Abbo et al. [6] developed a conceptual framework for awareness in embodied AI, outlining how AI agents can" sense" and interpret their environments which could later inform how AI tools communicate awareness to learners. Christopher and Khawar's [7] unpublished manuscript discussed new pedagogical ways educators could facilitate critical literacy in an AI-pervaded society, although again this work was only theoretical. Singh et al. [8], however, goes further, outlining an experimental study of using metacognitive prompts in generative AI searches and showed how structured reflective prompts could lead users to critically engage with the answers generated by AI instead of simply accepting them. All together, these works have created a foundation for the study of ethical awareness and reflective practice, but most of it remains theoretical, and only Singh et al., presented a study of an intervention.

Musali and Vermeulen [9] associated a programmatic use of technological pedagogies, whereby AI tools scaffold student reasoning and inquiry, caste AI as a tool shaping deeper engagement in learning environments. Walker et al. [10] build off of this by examining metacognition- or thinking about your own thinking- as a means to consider how it could push safer, more responsible AI use, and integrate reflective practice in AI-aided decision-making processes. Rotaru and Amariei [11] examined ethical risks from AI research such as institutional bias and lack of human-centered evaluation metrics but their work remained primarily conceptual in analysis without practical implications. Zhao et al. [12] provided an empirical study which focused on generative AI use to foster critical thinking and provided evidence of reflectivity through generative use rather than surface rote learning. De la Puente et al. [13] also noted using ChatGPT in an undergraduate debate exercise pushed better argumentation and evaluative reasoning. Regardless, these studies demonstrate the capacities of experience in AI to develop critical thinking and reflectivity on learning, and while some studies highlight pedagogical practice, most concentrate on student outcomes and wholly lack the account to understand how the AI instructional systems themselves could be made transparent, explainable, or ethically aligned.

Wang et al. [14] developed the AI Ethical Reflection Scale (AIERS), a validated measure of how university students' reason about ethical issues in AI, which is one of the few concrete measures developed for this examination. Varghese and Sharma [15] provided a conceptual framework analyzing the impacts of metacognitive awareness on the human–AI interaction with an emphasis on reflection when reasoning about AI outputs of analysis rather than blind acceptance. Moving to the technical context Bakar et al. [16] undertook a systematic review of XAI modeling for predicting diseases including explanatory analysis methods such as SHAP and LIME which require AI reasoning for providing E in healthcare. Salayma et al. [17] and Kim et al. [18] undertook a review of the evaluation lens for XAI in cardiology and additional contexts and both predominantly remain as surveys not accounting for unique challenges in the educational context (e.g., equity for diverse learners or interpretations for teachers). Lecue [19] outlined an explainability framework utilizing knowledge graphs which provides a process for identifying reasoning for actions made by AI, yet to our

knowledge remains untested in classrooms or within learning contexts. At the closest proximity to educational purpose, Maity and Deroy [20] proposed a Human-Centered XAI framework focused on improving the transparency of reasoning in learning tools to enhance trust- although again it remains a theoretical or conceptual framework and has neither been empirically verified nor tested in a real classroom

III. Research Gap

The existing literature on AI ethics and awareness shows an obvious gap. Some, including Choung et al. [5], Christopher and Khawar [7] and Zhao et al. [12] emphasized the need for ethical awareness, literacy and behavior in learners and educators. Others, such as Bakar et al. [16], Lecue [19] and Maity and Deroy [20] emphasized more on the technical aspects of AI and, explicitly frameworks and tools for explainability and transparency, than aspects of educational practices. While these scholarly works are all contributive to their own scholarly landscape, the literature is still fragmented and not interconnecting causal layers of ethical, cognitive and technical education. We see that the educational literature is evolving likely still blotchy, instead of being bound together, evolved into a well-organized, integrated and technical structure of research.
Furthermore, the ethical instruments and frameworks that have been proposed, for example the AIERS ethical reflection scale [14] and the metacognitive prompts designed by Singh et al. [8), are generally focused on measurement or reflective thought, and while important, tend to fall short of showing how ethical awareness and critical reflection can be systematically integrated within functional AI systems. This means that there is currently no model that is empirically driven that embodies ethical awareness, critical thought, and technical explainability, which could benefit education and professional learning, as it would signify the design of a classroom-ready system that maintains transparency, reflection and responsibility from conception to deployment. The absence of this sort of modelling represents a clear research gap, which only underscores the necessity for interdisciplinary approaches to be developed, which marry ethical imperatives to responsible AI practice.

IV. Methodology

The main issue explored in this study is the inadequate consideration of ethical awareness and critical thinking in the domain of AI literacy. Although previous research contributes positive data regarding AI adoption, technical features, and applications, there is a gap in systematically determining how ethical considerations and critical thinking are threaded through the body of AI-related research. By following an empirical study design using survey data as well as a structured analysis of literature features, this research will contribute to bridging the gap. In forming the solution to the problem, the hybrid of quantitative survey data and systematic literary analysis schema was an appropriate choice for the study. This combination of the two approaches accomplished two things; participants' cognizance and awareness of AI ethics was captured and then the researchers were able to critically examine the evaluation of existing research over a number of years. Original research was conducted to identify relevant literature, using a snowball sampling approach to develop a sample size, to enable synthesis and categorization of literature features, and explore patterns of research developments.

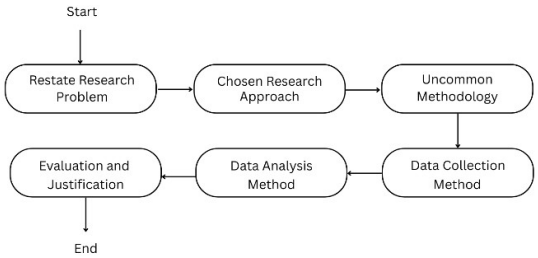

Fig. 1. Workflow for Determining the Methodology

A unique aspect of the study methodology is the automated feature extraction process. In contrast to literature reviews which rely solely on qualitative reading of source documents and summarizing relevant content, this study implemented Python-based scripts to extract features from more than 10 research papers published from 2020 until 2025. Data collection was conducted in two phases: a survey that was administered via Google Forms to evaluate participants critical thinking skills and ethical understanding and the systematic data gathering of literature. Twenty papers were reviewed in detail for gap identification, while more than 100 papers contributed

to the feature extraction and mapping for trends across related studies. This computational approach turned unstructured text into structured numerical datasets which allowed for objective and reproducible comparison across studies.

To conduct the data analysis, the extracted features were merged into a single dataset features into one integrated Excel file to conduct year by year comparison, and generated visualization to expose patterns and sustained features, and identify features that were absent from a given period, which also represents where novelty was found in this study. I justified my methodological approach by incorporating breadth as well as depth to address the research problem.

This study combined perspectives of survey responses and a review of larger scale computational analysis to show ethical awareness is demonstrated in practice, while it also aggregated seldom used features identified in one or two other papers into a consolidated and innovative framework.

V. Implementation

To carry-out the research, a parallel method of data collection incorporating a survey administered instrument and automated literary feature extraction was used to include both participant perspectives and computational analysis over a large number of publications.

- Survey Instrument Data Collection and Pre-Processing. To evaluate participants' understanding of AI biases, think about ethical decisions, and ways they engaged with AI in a critical way, we conducted a structured survey within Google Forms. The survey was intended spend considerable time on measuring participants' understanding of AI ethics in both breadth and depth by understanding their everyday usage of AI-based tools. We posed some general questions to assess their baseline awareness, such as whether they noticed AI systems were biased, how often they thought about ethics when using AI, and how often they thought about the consequences of AI outputs.
- In addition to the survey responses, I identified and systematically collected more than 100 AI-related journal articles and conference papers that had been published in the years 2020 to 2025. The data presented in these publications were identified in order to present a balanced sample of work from the research community in educational and learning research that captured recent developments in AI research. The works selected for inclusion captured a variety of research projects that were informed by and discussed ethical awareness, explainability, literacy, and critically examining AI. Each paper was transformed into machine-readable text in order to computationally analyze them using textual mining. Alongside the main body of text, the metadata publication year, type of text, and subject focus of the article was captured and logged into a structured dataset.
- Developed text mining scripts in Python to analyze the corpus. These scripts revolved around the following preprocessing steps: tokenization, stop-word removal, and lemmatization using NLTK (Natural Language Toolkit). The keywords and thematic terms that were identified across ethical awareness, critical thinking, explainability, and pedagogical integration categories. The appearances that were included in our corpus were then mapped against these four categories, resulting in structured datasets from the previously unstructured text.
- In evaluating the survey data and literature features, both quantitative survey data and literary analysis data were analyzed and visualized. In the literature data, year-by-year trends of thematic emphasis were indicated on bar charts and various types of visualizations. In the survey data, descriptive statistics and visualizations conveyed participant's ethically aware and critical thinking awareness. The analyses of each data type together provided two insights: user cognition in the wild (probably in daily life) and documented priorities in research.
- The results were converged to triangulate findings. Whereas the analysis of the literature showed that Explainability was emphasized and integrated ethical-critical frameworks were relatively absent, the survey showed participants were often limited in the reflective awareness in using AI outputs. The combined finding suggested that this dual gap is addressed in the present papers to emphasize the need to embed ethical awareness and critical thinking into AI rather than using it as a simple tool of convenience.

VI. Results and Discussions

A. Survey Findings

High Adoption, Low Reflection: Although 80 percent of respondents use ChatGPT and more than 63 percent report using AI frequently, only 17 percent regularly cross-check output.

Trust vs. Awareness Gap: Although 85 percent of respondents reported having a medium to high level of trust in AI, only 26.7 percent felt comfortable critically evaluating AI outputs.
Ethics vs. Practice: 70 percent agreed that ethics should be ingrained in AI, but fewer than half had received formal training on responsible AI usage.

TABLE I. SURVEY FINDINGS

Findings	Percentage (%)
High Adoption	80
Frequent AI Use	63
Cross-checking Outputs	17
Trust in AI	85
Confidence in Evaluation	26.7
Agree Ethics in AI	70
Formal Instruction Received	< 50
Primary Risks: Overreliance & Data Privacy	-

B. Literature Feature Extraction

Focus on Explainability (2020–2021): Early studies emphasized XAI frameworks and model transparency, with limited practical adoption.
Emergence of ethics (2022-2023): Ethics considerations gained attention, mostly conceptual, with few actionable frameworks.
Critical Thinking and Pedagogy (2024–2025): Recent literature explored AI literacy and pedagogical strategies, although practical implementation remains limited.
Trend Summary: The focus shifted from explainability to ethics discourse and partial integration of ethics, explainability and critical thinking.
Key Issues Across Literature: Bias, fairness, overreliance, and data privacy were recurrent concerns. As seen in Fig. 2, the first five parameters identified for influencing AI literacy coming from an expansive sheaf of 120 academic references are presented. The results indicate that Ethical Awareness carries the most weight (1.5), foreshadowing the level of agreement in the literature pertaining to the issue of fairness, transparency, and accountability of AI. Close behind were exploring Critical Thinking, Creativity Problem Solving, and Societal Impact Awareness (1.4). The results demonstrate the importance of critical reasoning, building innovative solutions, and knowing the social aspects of AI impact. Finally, Policy Governance Understanding, which had less weight at approximately 1.0, remains important as the literature reflects a dimension of regulatory frameworks and governance mechanisms to guide responsible AI adoptions. These findings, overall, demonstrate that as identified by prior literature before, AI literacy is a multidimensional construct and relates beyond technical skills to ethical, cognitive, social, and governance skills.

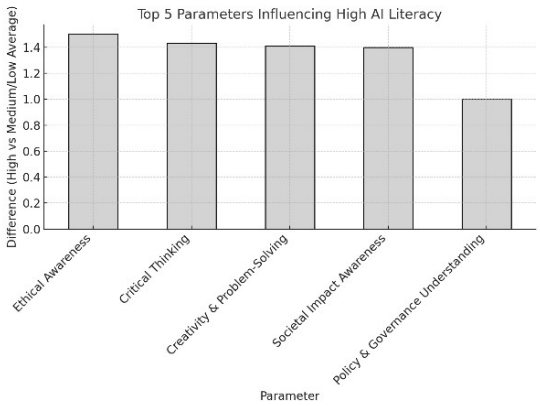

Fig. 2. Top Five Parameters Influencing High AI Literacy, Derived from A Literature Review Of 120 References

C. Convergent Findings

Triangulation of survey data and literature analysis highlights a dual gap:
Literature vs. Practice: Although academic research emphasizes explainability, it often lacks ethical integration and pedagogical application.

User Awareness vs. Engagement: Users recognize the importance of ethics, but frequently lack reflective skills to critically engage with AI outputs. These findings underscore the need for integrated AI literacy frameworks that integrate ethical awareness, critical thinking, and explainability, promoting responsible and reflective adoption of AI rather than blind dependence. Fig 3 illustrates the proportional contribution of the top five AI literacy parameters for each year based on a cumulative literature review of 120 references from 2020 to 2025. As shown in the data, Ethical Awareness consistently occupies the top position through the timeline for all five AI literacies, while Critical Thinking and Creativity Problem-Solving each emerged to play similar roles, while contributing to a greater proportion overall, although all five parameters appeared to peak in proportions in 2022. Societal Impact Awareness and Policy Governance Understanding appeared to hold lower contributions in proportions, but both of these parameters appeared to increase overall, indicating an increasing recognition of the importance of the regulatory, governance, and/or societal aspects of AI adoption. Thus, this trend over time shows, within the academic literature reviewed, that the ethical and cognitive aspects have consistently held more of priority overall, while governance related aspects have only recently emerged, indicating a growing recognition of AI literacy as a multi-faceted construct.

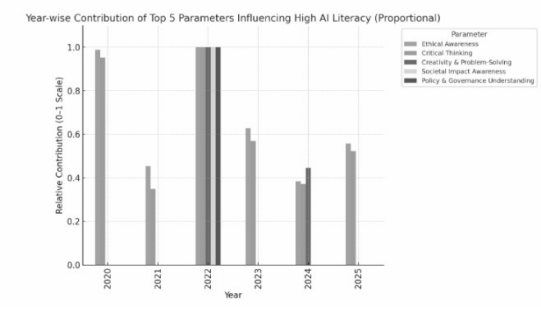

Fig. 3. Year-Wise Contribution of Top Five Parameters Influencing High AI Literacy (2020–2025)

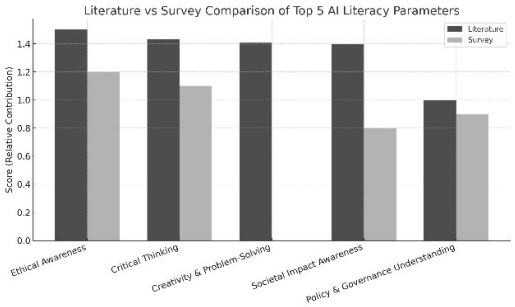

Fig. 4. Comparison Of Top Five AI Literacy Parameters: Literature Review vs. Survey Results

Fig. 4 shows the comparison of top five parameters of AI literacy by 120 literature references with the survey results. The literature review showed Ethical Awareness, Critical Thinking, and Creativity Problem-Solving as the most potent indicators of AI literacy factors, noting a representative score of more than 1.4 for each factor on the relative measures, while Societal Impact Awareness and Policy Governance Understanding were noted as slightly lower but still impactful parameters. However, in the survey, relative scores were noted lower in all parameters, with larger gaps between the relative scores recorded in Societal Impact Awareness and Critical Thinking. This suggests that while there is a agreement with scholars, they recognized these dimensions rather robustly, in practice, they felt these dimensions were lower. This view of the evidence situations literatures and practitioner use of these various literacies in terms of Ethical Dimensions and Cognitive Literacy competing while indicating lagging Societal Impact Awareness and Policy Governance Understanding literacies, acknowledgement to support a reason for range for implementation consideration.

VII. CONCLUSION

This article highlights an important gap in existing AI literacy conversations. Explainability and ethics are both important factors in user experiences and research areas, however, these ideas are undermined by treating these

subjects in isolation instead of understanding them as a part of a framework. This contributes to a disjointed conceptualization of AI literacy, in practice. Survey data supports this gap; while many users indicate a clear interest in ethics, their use of AI may appear rather superficial. Users have a tendency to accept outputs as they were presented, comfortable enough with the ease and authority of the system to avoid probing things like biases, limits, and assumptions. Much of the research on explainability continues to treat it as an isolated technical design feature relevant to the use of AI without relating to the broader ethical, cognitive, and educational context. This absence relates back to an underdeveloped idea of AI literacy that does not help people become critical, responsible citizens regarding AI decisions and interaction.

The collective nature of these observations shows that integrated models of AI literacy are required to prioritize ethical reasoning, critical appraisal, and technical explainability, rather than treating them separately. Such models would help learners and practitioners to move beyond understanding how AI systems work, to thinking about their social effects, and ethical consequences.

For example, professional preparation program training interventions could incorporate aspects of AI ethics, evaluative reasoning, and media literacy to enable reflective engagement with AI-generated outputs. Interweaving these aspects would mean that users could move from being passive consumers of the algorithmic outputs to active, responsible dwellers in AI-augmented settings, which is a value to offer. This has applicability in practice: it enables practitioners to more effectively navigate processes of decision-making, increases accountability, and fosters responsible practice in utilizing AI in personal and professional practice. Future research should focus on empirically developing, assessing, and using these types of frameworks in educational contexts, as a way to create overlap between ethical thinking, transparency, and reflective thinking
about practice. In this way, it is possible to shift AI literacy from a technical skills-based approach to a lived competency grounded in a practice of utilizing technology for human good and societal purpose.

REFERENCES

[1] S. C. Kong, ”Developing and validating an artificial intelligence ethical awareness intervention in primary education,” in Proc. 2025 Int. Conf. AI in Education (ICAIE), Location, Month 2025, pp. xx–yy.

[2] S. Gabriel, ”Generative AI in writing workshops: A path to AI literacy,” in Proc. 2024 Int. Conf. AI Research (ICAIR), Reading, UK, Dec. 2024, pp. 126–132, doi: 10.34190/icair.4.1.3022.

[3] M. Ryan, J. Antoniou, L. Brooks, T. Jiya, K. Macnish, and B. Stahl, ”Research and practice of AI ethics: A case study approach juxtaposing academic discourse with organisational reality,” Science and Engineering Ethics, vol. 27, no. 2, Art. 16, Mar. 2021, doi: 10.1007/ s11948-021-00293-x.

[4] J. Morley, A. Elhalal, F. Garcia, L. Kinsey, J. M¨okander, and L. Floridi, ”Ethics as a service: A pragmatic operationalisation of AI ethics,” Minds and Machines, vol. 31, no. 2, pp. 239–256, June 2021, doi: 10.1007/s11023-021-09563-w.

[5] H. Choung, P. David, and A. Ross, ”Trust and ethics in AI,” AI & Society, vol. 37, no. 3, pp. 821–835, May 2022, doi: 10.1007/s00146-022-01473-4.

[6] G. A. Abbo, S. Marchesi, K. Ciupi´nska, A. Wykowska, and T. Bel paeme, ”Towards a definition of awareness for embodied AI,” in Proc. 16th Int. Conf. Agents and Artificial Intelligence (ICAART), Rome, Italy, Feb. 24–26, 2024, vol. 3, pp. 1399–1404.

[7] B. Christopher and C. Khawar, ”Fostering critical thinking in the AI era: Innovative educational approaches for a data-driven society,” unpublished manuscript, Apr. 2024.

[8] A. Singh, Z. Guan, and S. Y. Rieh, ”Enhancing critical thinking in generative AI search with metacognitive prompts,” in Proc. 88th Annual Meeting Assoc. for Information Science & Technology (ASIS&T), Washington, DC, USA, Nov. 14–18, 2025.

[9] S. Musali and M. Vermeulen, ”AI-driven learning: Enhancing critical thinking through technology-integrated pedagogy,” unpublished preprint, Sep. 2024, doi: 10.13140/RG.2.2.31693.45285.

[10] P. B. Walker, J. J. Haase, M. L. Mehalick, C. T. Steele, D. W. Russell, and I. N. Davidson, ”Harnessing metacognition for safe and responsible AI,” Technologies, vol. 13, no. 3, art. 107, Mar. 2025, doi: 10.3390/technologies13030107.

[11] T.-S ,. Rotaru and C. Amariei, ”Ethical issues in research with artificial intelligence systems,” in Ethics in Scientific Research– New Perspectives, London, UK: IntechOpen, 2023, ch. X, doi: 10.5772/in techopen.1001451.

[12] G. Zhao, H. Sheng, Y. Wang, X. Cai, and T. Long, ”Generative artificial intelligence amplifies the role of critical thinking skills and reduces reliance on prior knowledge while promoting in-depth learning,” Education Sciences, vol. 15, no. 5, art. 554, May 2025, doi: 10.3390/educsci15050554.

[13] M. de la Puente, J. Torres, A. L. B. Troncoso, Y. Y. Hern´andez Meza, and J. X. Marquez Carrascal, ”Investigating the use of ChatGPT as a tool for enhancing critical thinking and argumentation skills in international relations debates among undergraduate students,” Smart Learning Environments, vol. 11, art. 55, Dec. 2024, doi: 10.1186/s40561-024 00347-0.

[14] Z. Wang, C.-S. Chai, J. Li, and V. W. Y. Lee, ”Assessment of AI ethical reflection: The development and validation of the AI ethical reflection scale (AIERS) for university students,” Int. J. Educ. Technol. High. Educ., vol. 22, art. 19, Mar. 2025, doi: 10.1186/s41239-025-00519-z.
[15] M. A. Varghese and P. Sharma, ”Thinking about thinking: How metacognitive awareness shapes human–AI interaction,” in Artificial Intelligence and ChatGPT (A Transformative Approach), Art. 24, Oct. 2024, doi: 10.25215/9392917538.24.
[16] A. F. S. Bakar, M. I. Muhamad, R. A. A. Rahman, and N. Mustaffa, ”The role of explainable AI in disease prediction: A systematic literature review,” BMC Med. Inform. Decis. Mak., vol. 25, no. 1, p. 94, May 2025.
[17] A. A. Salayma, A. H. M. Nordin, A. A. M. Isa, M. A. Maarof, and W. A. W. Adnan, ”A review of evaluation approaches for explainable AI with applications in cardiology,” Artif. Intell. Rev., vol. 58, pp. 4875–4912, May 2024.
[18] J. Kim, H. Maathuis, and D. Sent, ”Human centered evaluation of explainable AI applications: A systematic review,” Frontiers in Artificial Intelligence, vol. 7, p. 1456486, Nov. 2024.
[19] F. Lecue, ”On the role of knowledge graphs in explainable AI,” presented at PROFILES/SEMEX @ ISWC 2019.
[20] S. Maity and A. Deroy, ”Human centric eXplainable AI in education,” arXiv preprint arXiv:2410.1982.