

Systematic Literature Review on Students' Perceptions of Ethical AI: Fairness, and Regulatory Understanding

S. Noorjannah Ibrahim^{1*}, Norsaremah Salleh², Ani Liza Asnawi¹, Nor Fadhillah Md Azmin³,
and S. Fatimah Ibrahim⁴

¹Dept. of Electrical & Computer Engineering, Kulliyah of Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia

²Dept. of Computer Science, Kulliyah of Information and Computer Technology, International Islamic University Malaysia, Kuala Lumpur, Malaysia

³Dept. of Chemical Engineering & Sustainability, Kulliyah of Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia

⁴School of Chemical and Process Engineering, University of Leeds, United Kingdom

ABSTRACT

The increasing demand of Artificial Intelligence (AI) in academia has intensified debate over ethical implications and establishment of regulations. While extensive research focuses on AI technicalities such as algorithm development, algorithmic accountability, and regulatory frameworks, insufficient attention has been given to how students perceive these ethical dimensions. This study presents a systematic literature review (SLR) on students' perceptions of ethical AI, focusing on awareness, fairness, and knowledge of regulatory systems. This SLR adheres to the PRISMA 2020 guidelines and consolidates findings from four academic databases: Scopus, ScienceDirect, IEEE Xplore, and ERIC. The findings reveal that while students generally acknowledge the importance of using AI tools ethically, their understandings are influenced by the Perceived of Usefulness (PoU), Perceived Of Ease of Use (PoEU) and Social and Cultural Perceptions. Students are concerned about the fairness of evaluations, as well as the potential of bias and discrimination, when AI techniques are employed without appropriate regulations or policies. This review highlights the importance of transparency in AI systems, the mitigation of prejudice, the assurance of justice, the reform of the educational paradigm, and the establishment of a comprehensive ethical AI framework and policy. The paper signifies the need to revolutionize teaching and learning while empowering learners with robust ethical foundations, which is critical in developing responsible AI practitioners.

Keywords: AI awareness, AI fairness, AI regulatory, ethical AI, students' perceptions

1. INTRODUCTION

The integration of Artificial Intelligence (AI) into education has brought significant changes to how teaching and learning are conducted, managed, and assessed. From AI-powered chatbots to adaptive learning platforms, these technologies aim to improve educational efficiency and personalized learning experiences. As AI becomes more common in the daily operations of higher education institutions (HEI), it holds great potential to transform how students interact with knowledge and how educators support learning. However, this transformation is not without ethical concerns.

Ethics refers to moral principles, values, and norms that guide the behaviours or actions of an individual or group of individuals to determine what is right or wrong [1]. In practical contexts, ethics serves as the foundation for evaluating actions, systems, and policies within the boundary of virtues and values, e.g., justice, autonomy, responsibility, and fairness. Undeniably, ethics is a complex and complicated concept that weighs multiple factors outlining specific moral

*noorjannah@iium.edu.my

obligations. Within professional domains such as engineering, medicine, and education, ethical frameworks are used to promote accountability, equity, and respect for human dignity [2].

In the context of AI, ethics has established two interrelated but distinct fields of inquiry: *Ethics of AI* and *Ethical AI*. The *Ethics of AI* addresses common questions about how AI should be developed and used [3-5]. The concept elaborates issues related to moral expectations from AI systems e.g., the moral status of autonomous systems, the societal consequences of automation, and the impact of algorithmic decision-making. It is mostly philosophical and future-anticipated, focusing on the values that could guide the growth of AI. On the other hand, *Ethical AI* refers to the practical implementation of ethics principles in the design, deployment, and governance of AI systems. This includes efforts to build transparent, accountable, and fair algorithms; ensure data privacy and protection; reduce discrimination and bias; and promote explainability [3-4]. *Ethical AI* outlines the translation process of ethical theory into policy practices that can be used to make AI systems less harmful and more trustworthy [5], [6]. It addresses how moral norms can protect human interests in the use of AI systems.

The integration of AI technologies in educational settings amplifies the significance of ethical problems. AI has influenced students' educational experiences and outcomes through intelligent tutoring systems, learning analytics, AI-assisted grading, and behavioural monitoring [4], [7-8]. While AI presents considerable advantages, it is consistently associated with issues such as algorithmic bias, privacy and data violations, and vague decision-making processes [4-5], [7-8]. Consequently, the development of AI systems from an ethics point of view, requires meticulous examination, especially from the perspective of those most impacted by the technologies i.e., the students themselves. Incorporating the understanding of *Ethical AI* into education demands not only technical expertise for developing AI systems but also critical awareness among students, who will become future engineers, policymakers, and informed citizens. Despite the increasing emphasis on ethics in the deployment of the AI systems, there is a lack of thorough research study investigating students' perceptions regarding these matters.

This SLR synthesizes empirical studies on students' perceptions of *Ethical AI*, concentrating on : awareness, fairness, and regulatory comprehension. This research seeks to expose existing trends, identify knowledge gaps, and provide direction for future education, technology, and legislative initiatives to embed ethical AI literacy into academic environments. The focus is on capturing insights into how students understand, experience, or evaluate ethical issues associated with AI systems. Fairness involves their judgments about AI bias, discrimination, and equitable outcomes; and regulatory understanding relates to their knowledge of legal frameworks and governance mechanisms that control AI practices. Based on the synthesis of findings, this review aims to **propose directions for future research, education policy, and curriculum development**. The recommendations are intended to support relevant stakeholders, including educators, curriculum designers, and policymakers, in creating educational environments that foster *Ethical AI*, whilst producing *ethically informed, critically engaged, and socially responsible AI users and developers*. To achieve these objectives, this review is structured around the following research questions (RQs):

RQ1: Fairness

How do students perceive fairness in AI-based systems used in education?

RQ2: Regulatory Understanding

What is students' understanding of AI-related rules, policies, and regulations?

This paper is structured as follows: Section 2 outlines the methodology used for study selection and data extraction. Section 3 presents the findings of the SLR, including the distribution of studies across different domains, the primary ethical issues and legal considerations, the research methods employed, and the solutions and recommendations proposed by authors. Section 4 provides a discussion on the findings and concludes with key insights and recommendations for future research and policy efforts.

2. BACKGROUND

2.1 Ethical AI in Education

While AI has immense potential for innovation and efficiency, it also has complex ethical challenges related to bias, privacy, accountability, and transparency [9-11]. These concerns have led to growing calls for the development and enforcement of *Ethical AI* principles that safeguard human dignity, promote fairness, and ensure the responsible use of technology [6], [12]. Globally, various institutions and governments have introduced frameworks to address these challenges. The Institute of Electrical and Electronics Engineers (IEEE) developed the framework for AI ethics guidelines, which advocate for the creation of AI systems that are transparent, accountable, and aligned with human rights and societal well-being [13]. The European Commission's Ethics Guidelines for Trustworthy AI (2019) outlines seven key pillars—human agency, privacy, transparency, diversity, accountability, societal well-being, and robustness [14]. Meanwhile, UNESCO's Recommendation on the Ethics of Artificial Intelligence emphasizes inclusivity, sustainability, and respect for human autonomy in the deployment of AI technologies [15].

In Malaysia, *Ethical AI* governance is still an emerging area. The National Artificial Intelligence Roadmap 2021–2025, published by the Ministry of Science, Technology and Innovation (MOSTI), promotes responsible AI development by emphasizing digital literacy, public trust, and international collaboration [16]. The roadmap also recognizes the need to develop talent with strong ethical foundations. However, the integration of these values into educational contexts remains promising. Although agencies such as the Malaysian Digital Economy Corporation (MDEC) advocate AI ethics awareness, systematic implementation across schools and universities is still lacking [17]. In educational settings, AI is being rapidly adopted in the form of learning analytics, intelligent tutoring systems, and generative tools such as ChatGPT. These tools are reshaping the ways in which students learn, interact, and are assessed. Consequently, there is a pressing need to understand how students perceive the ethical dimensions of these technologies.

Some students may demonstrate basic awareness of AI ethics and risks, while others remain unaware of potential biases and systemic injustices embedded in algorithmic processes. There is also a lack of clarity on whether students understand current AI regulations or their rights under data protection and fairness laws. Without this understanding, students are unable to critically assess or challenge unethical AI applications. The lack of standardized frameworks and policies to integrate ethical AI in education escalates these problems. Recent research indicates that students hold mixed perceptions of *Ethical AI*. Some appreciate its efficiency, personalization, and convenience in learning [18-19], while others raise concerns over opaque decision-making, algorithmic bias, and misuse of data [1], [20]. Moreover, studies reveal that many students are unaware of the ethical frameworks and governance mechanisms that guide AI usage, highlighting a critical gap in digital ethics education. Scholars have emphasized the need to embed AI ethics into the curriculum. Floridi et. al. in [21] argue that digital literacy must encompass ethical competence. Jobin et. al. in [22] conducted a global mapping of *Ethical AI* principles, revealing significant variation and calling for context-sensitive approaches. Meanwhile, Binns [9] proposed that fairness in AI cannot be universalized but should be contextualized to user values and expectations.

Despite the increasing integration of AI in educational settings, limited understanding of how students perceive the ethical implications of AI technologies remains. While research has extensively explored the technical capabilities of AI, the ethical and human-centered aspects particularly from the perspective of students, are often overlooked. This situation is concerning, as students are the primary stakeholders impacted by AI-driven systems. Furthermore, research in this area is fragmented across disciplines, with limited studies conducted in developing countries or among diverse student populations such as Malaysia. This fragmentation prevents the formation of a coherent global understanding of students' ethical orientations toward AI. In a world where AI is being rapidly globalized, these disparities threaten to reinforce digital inequality and ethical ignorance.

This SLR aims to understand student perceptions regarding components specifically associated with AI ethics: awareness, fairness, and regulatory understanding. It presents a thorough overview of AI ethics integration in education, spanning from 2015 to early 2025. Rather than conceptual or abstract research, this study concentrates on student perceptions of Ethical AI in relation to real-world education initiatives, including teaching and learning methods, content, assessment, and management systems. These elements represent the complex education activities that occur in practice.

3. METHODOLOGY

3.1 Demographic Data

This SLR was conducted following the **PRISMA 2020** guidelines, which provide a structured and transparent approach for identifying, screening, appraising, and synthesizing relevant literature. The PRISMA framework is a preferred reporting system for Systematic Reviews and Meta Analysis, guiding researchers in conducting rigorous and repeatable literature review process [23]. The framework allows systematic selection of literature and review of how students perceive ethical aspects of AI, with particular emphasis on the dimensions of fairness and **regulatory understanding**. The review protocol is to ensure that only empirical studies that directly addressed student perspectives were included, thereby enhancing the relevance and focus of the findings following the inclusion and exclusion criteria in Table 1.

Table 1 Inclusion and exclusion criteria of the SLR

Inclusion Criteria	Exclusion Criteria
Empirical studies involving students' perspectives on AI ethics	Articles not involving students' perspectives on AI ethics
Ethical issues in AI	General AI without ethics-oriented issues
Studies focus on awareness, fairness, or regulation	Other issues not related to the three themes
Empirical studies, primary research	Commentary, Opinion or editorials
Published in peer-reviewed sources	Grey literature with no academic validation
Higher education	Studies other than higher education

A comprehensive search was conducted across four major academic databases, **Scopus, Science Direct, IEEE Xplore**, and **ERIC**, to gather literature from various fields, including education, AI ethics, and technology adoption. The search strategy utilized a combination of targeted keywords and Boolean operators, such as: *(ethical AI OR AI ethics OR Ethics OR AI) AND (students OR learners) AND (perceptions OR awareness OR understanding) AND (fairness OR bias OR justice) AND (regulations OR laws OR policy)*. Filters were applied to limit the results to English-language

publications from 2015 to 2025, and only peer-reviewed journal articles and conference proceedings were considered.

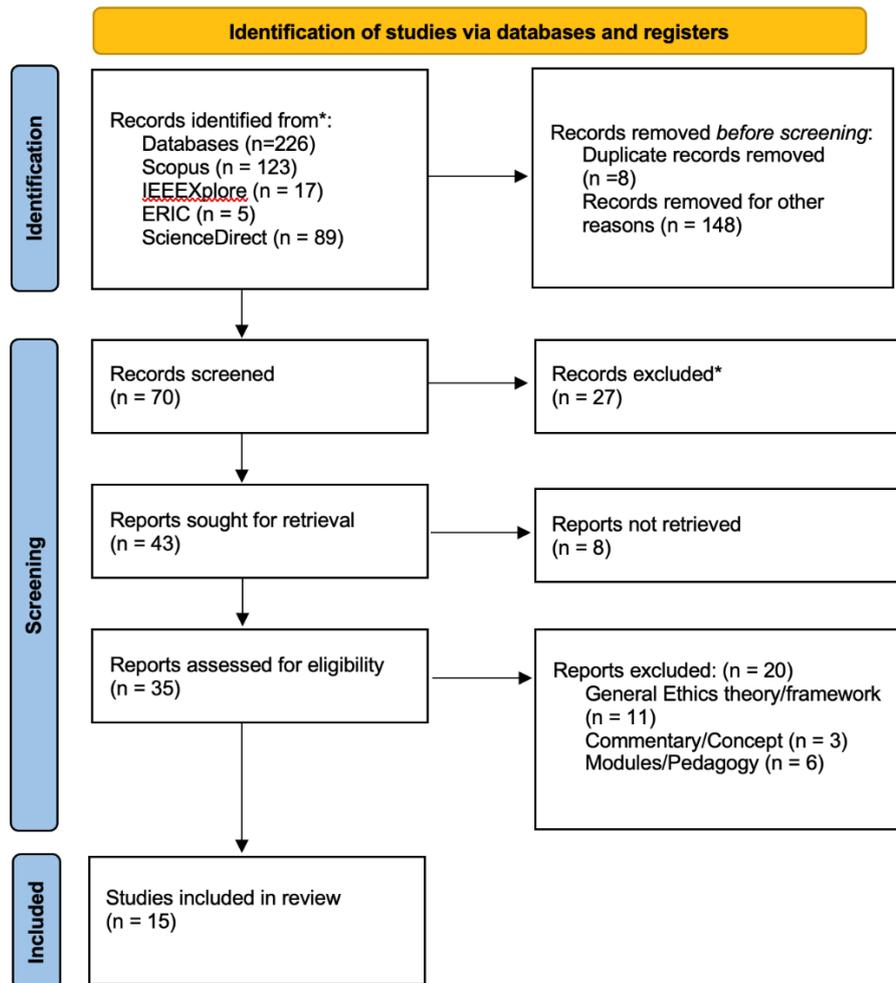


Figure 1. Flow diagram of identification, screening and inclusion process of literature using PRISMA 2020 protocol [12].

The inclusion criteria were formulated to select studies that focused specifically on students or learners and addressed at least one of the core themes: ethical awareness of AI, perceived fairness in AI systems, or understanding of AI-related regulations. Only empirical studies employing qualitative, quantitative, or mixed-methods approaches were considered and studies were excluded if they presented general ethical frameworks unrelated to learners, commentary or concept papers, or focused on module development or pedagogical practices without analyzing student perspectives.

An initial search yielded **234 articles**, of which **226 unique articles related to ‘AI Ethics and students’** remained after the removal of duplicates. Following a title and abstract screening, 148 articles were excluded for not meeting the relevance criteria. The first screening stage involved 70 articles and 27 articles were excluded. In the second screening stage, 43 articles are retrieved from the databases, and only 35 are available for full-text review, resulting in the exclusion of 11 general ethics theory or framework papers, 3 concept or commentary papers and 6 papers centered on pedagogical module design. In total, **15 empirical studies** were included in the final synthesis. Fig. 1 illustrates whole process of literature collection. These 15 studies provide the data foundation for the thematic analysis presented in subsequent sections.

4. RESULTS

4.1 Key Findings

The qualitative and quantitative methodologies are frequently used to investigate perceptions. In this SLR, 40% of the studies used qualitative research method, while 27% of the studies used quantitative research method. The remaining studies employ mixed-method (hybrid) to investigate students' perceptions of ethical AI. These studies include viewpoints from various backgrounds; technical, non-technical, and others; to give comprehensive insights on how distinct student groups see ethical issues related to AI technologies. We classified students with backgrounds in engineering [20], [24], and computer science [25] as part of the technical group, while those with backgrounds in language [26], psychology [27], management [28] and marketing [29], were categorized into the non-technical group. Students with biology [30], mixed [12], [20] or unspecified backgrounds [32-35] were designated as 'others' as listed in Table 2.

Table 2 Group of Students Included in Reference Studies

Method	Number of References	Percentage (%)
Technical	3	20
Non-Technical	4	27
Others (mixed-group or not specified)	8	53

We divide students' perceptions on ethical AI into positive and negative perceptions. Issues related to students' awareness on this matter are listed in Table 3.

4.1.1 Positive Perceptions on AI

Students exhibit a positive attitude towards integrating AI in their studies, primarily driven by their perception of its usefulness. This is consistently identified as a significant factor influencing students' intention to use AI. Students believe AI tools can make them more productive and offer more effective and efficient ways of learning [28], [29], [36-37]. A survey of 173 learners of non-English languages revealed that a significant majority (74%) opted against utilizing ChatGPT for language acquisition. But certain students claimed that AI tools such as ChatGPT could be more advantageous in STEAM disciplines due to the unique characteristics of language acquisition [26]. Students widely recognize the potential of AI tools to improve learning outcomes and efficiency. A significant portion relies on AI to minimize time spent on less engaging activities (58%), to prioritize important topics (67%) and to translate texts not written in their native language (55%) [34].

Students agree that AI tools enhance efficiency and save time in assignments [29], [34], optimize research, enhance drafts, and help in study management [37]. They value AI for its ability to deliver immediate access to information and explanations on complex topics [29]. A key perceived benefit is AI's capacity to deliver personalized learning experiences, individualized feedback, and customized contents [29], [33-34], [37]. They also recognize its value in developing professional documents like resumes and refining communication skills, thereby enhancing their personal brand and job search self-efficacy [29], [37]. They believe that AI knowledge is relevant for future careers [38]. Another advantage of AI is enhanced accessibility to knowledge, as AI-driven services and solutions, such as speech-to-text software, are regarded as advantageous for students with impairments [37].

Table 3 Students Awareness on AI Technologies in Education

Positive Perceptions on AI	Reference
Perceived of Usefulness (POU)	[24], [26], [28], [29], [34], [36], [37]
Perceived of Ease of Use (PEOU)	[28], [29], [36]
Social and Cultural Contexts	[24], [27], [28], [34], [36]
Negative Perceptions on AI	Reference
AI limitations and Risks	[20], [24], [26], [29], [31], [33], [34], [35], [37]
Lack of Transparency and Guidelines	[24], [31], [33], [34]

Students describe AI tools as easy to use (user-friendly) and accessible, providing a variety of options that make learning more engaging [36]. AI tools enhance self-efficacy and promote joy in learning. Students are more inclined to use new technology if they find it enjoyable. AI tools boost students' confidence in their capacity to use and leverage AI effectively [29], [36]. Therefore, students are more inclined to implement AI if they perceive it as advantageous over traditional methods, which reduces resistance and uncertainty during teaching and learning [28].

The influence of peers, educators, and university environment plays a pivotal role in shaping students' attitudes and intentions towards AI adoption [28]. When individuals within a university; lecturers, influencers or administrators; use AI, students tend to perceive it as enhancing their image and boosting their intention to use AI technology [27], [36]. Institutional endorsements, such as university or government approval, also lend credibility to AI tools and enhance trust [34]. Student attitudes are also significantly influenced by instructor guidance and perceptions [28]. Perceptions of technology, including AI, can vary significantly across different cultures and regions [28]. For instance, German students might approach AI tools with more caution due to cultural emphasis on data privacy, while Chinese students might be more willing to adopt new technologies quickly, focusing on practical benefits and efficiency [34]. This highlights the need for context-specific and culturally sensitive approaches to AI integration.

4.1.2 Negative Perceptions on AI

Despite the perceived benefits, students have significant concerns about the ethical implications and potential risks of AI, which can lead to resistance or **'adopt with caution'** approach to AI usage in learning. Students mistakenly assume that AI tool acquires information from the internet in real time, but in actual, it draws information from static training data. The comprehension of the process by which these outputs are generated can be inadequate, even though students generally comprehend that AI generates a variety of media types (text, calculations, images, and codes) [24].

The primary ethical concern is academic integrity or plagiarism [24], [29], [33-35], [37], [39]. Students worry about the high possibility of cheating or the ease of generating text that resembles human writing [26], [33], [37]. Students often recognize the potential for abuse, despite their desire for AI's benefits [29]. Additionally, a substantial number of students are of the opinion that AI tools are not considered as original authors, as they primarily compile or rephrase preexisting information [24]. The mixed perception of AI, which is influenced by contradictory information and policies, is evident in the fact that students frequently report feeling more apprehensive than excited about it [24]. Students are concerned that AI systems can perpetuate and amplify biases present in their training data, which can potentially lead to unfair treatment or inaccurate information [33-34], [37], [40]. If students have personally encountered discrimination, their perceptions of individual and societal damages from AI applications, such as when using any Academic Performance Prediction (APP) tools [1], are intensified.

While some studies note a relatively lower emphasis on privacy compared to academic integrity, students still feel uneasy about their personal information being used for training AI, even with consent [29]. Students with higher negative perceptions are less likely to voluntarily provide their data and demand more control over what data is used. The nature of AI systems is complex and very technical and can affect trust and accountability [20]. Students generally prefer understanding the reasons behind AI outputs, which means clarity and transparency significantly impacted students' trust in AI [20], [34]. Interestingly, some studies suggest students can perceive AI evaluations as fairer due to explicit assessment criteria and well defined processes, even if the underlying algorithms are complex [20]. This highlights the importance of effective communication about how AI tools function to make students understand how AI works [33].

Another significant concern is the potential for AI to impede creative and independent thinking [24], [29], [31], [33-34], [37]. Students express concern that overreliance on AI tools could weaken critical thinking skills [34], [37]. They worry about developing tendencies for procrastination and memory loss [31]. Students' perceptions of AI's directly influence their interactions and judgments of AI outputs as they may be less inclined to use their own critical judgment [31]. Despite students' awareness of AI benefits, students suffer a higher level of nervousness than excitement about AI, often due to mixed messages and contradictory policies. Students also express anxiety about AI's potential to replace human jobs [34] and they may be unsure about AI's use in the future, especially when it personally affects them.

The absence of consistent guidelines is a major barrier, creating confusion and undermining students' trusts [34]. Existing AI policies in education are often too broad, fragmented and outdated, particularly struggling to keep pace with rapid AI tool advancements [33]. The wide range of instructor attitudes and policies, from strict bans to permissive use, leaves students confused and anxious about being penalized due to unclear rules [24], [30]. There are limited examples of education based on AI ethical principles. This lack of understanding hinders critical and responsible engagement with AI tools.

4.1.3 RQ1: How do students perceive fairness in AI-based systems used in education?

A. Fairness in Evaluation

Fairness in AI involves ensuring equitable treatment, mitigating biases, providing equal access, and fostering just outcomes in AI systems and their application. The perceptions of fairness and bias are critical factors that influence AI adoption in teaching and learning. Students exhibit a unique perception of fairness when it comes to AI, particularly in evaluation contexts. Interestingly, some studies indicate that college students perceive AI algorithms as fairer evaluators than human educators [20]. While this perception seems controversial, it is largely attributed to the information transparency that AI system provides. For instance, AI algorithms can assess candidates based on objective criteria like academic performance and test scores, unaffected by subjective influence [20]. Moreover, the assessment processes can enhance transparency, providing students with a clear understanding of evaluative criteria and assessment methodologies. This level of clarity fosters a stronger sense of trust and fairness in the evaluation system.

However, this perception of fairness and bias lessens when explanations for evaluation outcomes are provided. In such cases, the difference in perceived fairness between AI and educators becomes insignificant [20]. This highlights that while AI's objective process is valued, human-provided explanations can equally build trust. Students perceive fairness when the AI judgments align with their own, while students' feelings of unfairness arise when the opposite is true. Despite this perceived fairness in evaluation, students show significant concerns about other aspects of fairness and the potential for bias within AI systems.

B. Bias and Discrimination

The ethical landscape of AI-created materials is multifaceted, underlining the need for bias identification in trained models. A frequent concern is that AI systems can amplify biases present in their training data, potentially leading to unfair treatment or inaccurate information, particularly for underrepresented or underprivileged students [24], [26], [33-34], [37]. Students worry about the biased content that results from AI training on human-created data. For example, AI systems in grading or admissions processes trained on historical data may unintentionally disadvantage certain student groups [20], [33]. Some institutional guidance on diversity, equity and inclusion (DEI) policies specifically emphasizes data bias because of training on publicly available data that can perpetuate discrimination or lead to inaccurate information [27]. What is considered fair can vary among individuals and groups, adding complexity to addressing fairness in AI systems. Fairness concerns include equal access to technology and resources, as disparities can widen the educational gap [27], [33], [35], [37]. Disparities in access (e.g., due to financial hurdles for paid subscriptions, lack of internet infrastructure, or disabilities) can widen existing educational inequalities. Hence, students advocate for systems that reflect the diversity of the student population and are accessible to all [35]. Students reported relatively low emphasis on privacy and ethical concerns compared to academic integrity issues.

4.1.4 RQ3: What are students' understanding of AI-related rules, policies, and regulations?

Regulations encompass the policies, guidelines, and frameworks to govern the ethical, responsible, and effective use of AI in educational settings. In general, students welcome and advocate for clear institutional policies, guidance or regulations on AI use in education [24], [26], [30], [34]. The current landscape of fragmented or underdeveloped AI policies leads to significant confusion and anxiety among students [24], [30], [33]. The ambiguities, particularly around acceptable use of AI tools and academic integrity, can affect trust, which negatively impacts learning outcomes [30].

A. Clear and Consistent on AI Tool Rules

Clear, detailed, and research-based approaches to generative AI tools are essential to prevent educators from implementing policies that are not supportive of learning. Students explicitly call for definitive guidance from universities and educators on how to use AI responsibly and ethically, especially concerning graded assignments [20], [24], [26], [29], [30], [33-34], [37]. They want clear boundaries to avoid cheating due to inconsistent instructor approaches [30]. While acknowledging the risks, most students support the permitted use of AI, provided it is regulated and integrated into the curriculum [24], [26], [30].

A shared viewpoint among students from different fields is a preference for regulated use and institutional embracement of AI technologies over absolute bans of AI usage. They view AI as a valuable supplementary learning resource rather than a tool for directly completing assignments [24], [30], [34], [37]. Students also suggest that policies should require them to disclose their use of AI tools in assignments and for educators to disclose detection tools for assessment [33-34]. Clear consequences for AI misuse and boundaries for AI involvement in academic work must be defined. However, less than half of institutions discuss privacy implications and ethics [34]. Concerns about academic dishonesty and plagiarism are significant, as AI systems can fabricate facts and sources. Hence, it requires the reevaluation of current academic integrity policies.

B. AI Literacy

It is crucial to establish clear guidelines on ethical AI use, including comprehensive training for both educators and students on appropriate use, and clarity on when and how these tools should be employed. Students recognize the critical importance of AI literacy training to understand AI's capabilities, limitations, ethical considerations, and practical applications [20], [25], [31], [33]. Students believe that the responsibility for ensuring ethical AI use should be shared among students, educators, and university administration [24], [30]. Furthermore, they advocate for active involvement in policy discussions and the co-creation of guidelines [30]. This approach can foster mutual respect, reciprocity, and shared responsibility, which is seen as a vital aspect for effective AI integration [30]. Students largely reject total bans on AI as they recognize AI benefits in their future careers [24], [26], [30], [37]. They suggest that instead of prohibitions, educational institutions should adapt evaluation methods and focus on fostering higher-order thinking skills that are resilient to AI assistance [27], [30], [37].

5. DISCUSSION

5.1 Recommendations

From the findings, students' perceptions on ethics are more positive and trusting when the stakeholders created a more positive and trusting environment for AI usage. Several key recommendations are as follows:

5.1.1 Increase Transparency of AI Systems

When AI tools are used in evaluations, explanations for evaluation outcomes significantly impact perceived fairness [20]. Higher education institutions should ensure that AI tools used in teaching and learning are understandable by students with clear descriptions of their functions, potential biases, and impacts [33-34]. Clarity and explainability are crucial for trust, with over 65% of students emphasizing their importance [34]. Policies should require students to disclose their use of AI tools in assignments and assessments to enhance transparency and accountability [33-34]. Similarly, institutions and educators should be transparent about their own use of AI tools and how they are monitored [30].

Institutions must proactively address bias by requiring greater transparency in AI model training and the inclusion of diverse datasets. This is to minimize algorithmic discrimination [25], [37]. Regular audits should be conducted to maintain ethical and fair decision-making [25], [33]. Institutions need to guarantee AI tools are accessible to all students and that resources are distributed fairly [33], [35]. Providing opt-out options for students who are concerned about AI tool costs or their privacy is also important. AI-generated content should reflect the diversity of the student population and avoid perpetuating biases [35].

5.1.2 Transform Education Paradigms

Formal education on AI and its ethical aspects should be embedded into academic curricula across disciplines, not just STEM fields [30], [33], [41]. This includes teaching students about AI terminology, capabilities, limitations, and the practical application of AI tools, alongside digital ethics and responsible AI practices [29], [34], [41]. Structured orientation and ongoing professional development programmes for both students and educators should be offered by institutions to support AI adaptation [33-34], [37], [41]. For students, training should cover effective prompt engineering, research using AI, content generation, and understanding AI's role in the learning process [29], [33]. Students should be trained on iterative workflows when using AI tools, emphasizing on refinement technique to finish the specific task given [34]. This involves

starting with broad prompts and iteratively refining them for more tailored results and encouraging follow-up questions for diverse perspectives. This approach helps build trust and deeper comprehension of subjects. Meanwhile, for educators, training should focus on both the technical aspects of AI tools and their pedagogical implications, ensuring they can integrate AI effectively while maintaining control over its application and understanding how students might perceive and use it [4], [30], [33], [37].

Design educational activities that specifically encourage critical AI engagement rather than passive acceptance of AI-generated content [31]. This includes exercises that require students to critically assess AI outputs, identify biases, reconstruct arguments, and develop independent reasoning skills [31], [37]. Offer personalised learning experiences through AI tools that adapt to individual learning styles and paces. This can include AI-driven simulations, brainstorming assistants, and tools for generating and refining ideas. AI should be positioned as a cognitive amplifier that enhances learning, creativity, and problem-solving, rather than a shortcut or a replacement for human intellect [29], [31-34], [37].

Higher education institutions should also ensure equitable access to AI technologies for all students, regardless of socioeconomic status or disability [27], [33], [37]. Instead of total bans of AI tools, institutions should adapt changes in evaluation, rethink assessment methods with methods that consider AI use [20], [28], [34]. Focus should be put on fostering higher-order thinking skills that are resilient to AI assistance [27], [37]. This includes process-based evaluations like oral exams, debates, and discussions, and scaffolding assignments into smaller steps. Policies should require students to disclose their use of AI tools in assignments. This promotes transparency and allows for a better assessment of genuine learning outcomes. It is crucial to define consequences for AI misuse and set clear boundaries on its involvement in academic work.

5.1.3 Establish Robust Ethical AI Policy/Framework

It is crucial for universities to establish explicit, comprehensive policies and guidelines on AI use and define acceptable boundaries for AI assistance in academic work. These guidelines should clearly define what constitutes appropriate use (e.g., drafting, brainstorming) versus academic misconduct (e.g., submitting AI-generated text as one's own graded assignment) [29], [30], [33]. Higher education institutions should actively involve students in the development of AI policies and guidelines, engaging in open dialogue to understand their perspectives, needs, and goals [30]. A "Students as Partners" approach can lead to shared standards and innovative uses that support learning. Educators should also be transparent about their own use of AI and any concerns they have, fostering a reciprocal relationship built on trust. A dedicated multi-discipline task force or AI Governance Office should be established at higher education institutions to draft comprehensive AI policies. This office would create and enforce ethical frameworks, conduct technical audits, manage violation procedures, and provide training [34]. This aligns with the broader institutional responsibility to ensure AI supports teaching and learning ethically.

6. CONCLUSION

Students' perceptions of ethical AI in education are complex, reflecting both enthusiasm for its potential and significant apprehension regarding its responsible integration. Students voice significant concerns about the potential negative impacts of AI on academic integrity in their learning processes. A primary ethical concern is the risk of AI facilitating plagiarism and cheating. Many students feel that using AI tools, especially for graded assignments, constitutes intellectual theft and undermines fairness. They also worry that depending too heavily on AI tools could weaken their deep analytical thinking, problem-solving, and creativity and may lead to superficial learning habits. Hence, students and educators should advocate for disclosure and proper

attribution of AI use in academic work to ensure transparency and accountability. Regarding bias and fairness, students express apprehension that AI systems, if trained on biased data or designed with flawed algorithms, may perpetuate or even amplify existing biases, potentially leading to unequal treatment or disadvantage for certain student groups.

Clearly, an urgent need is identified for comprehensive AI literacy and training programs for both students and educators. Such training should encompass technical skills, the pedagogical implications of AI, and crucially, how to critically evaluate AI outputs, including fact-checking and bias detection. Finally, students and educators generally agree that AI should function as a supportive tool that complements, rather than replaces, traditional teaching methods and essential human interaction. This leads to a call for balanced AI integration that augments, rather than diminishes, the role of educators.

REFERENCES

- [1] Lünich, M., Keller, B., Marcinkowski, F., 2024. Diverging perceptions of artificial intelligence in higher education: A comparison of student and public assessments on risks and damages of academic performance prediction in Germany. *Comput. Educ. Artif. Intell.* 7, 100305.
- [2] Krsmanovic, G., Deek, F. P., 2024. AI in the learning environment: Examination of pedagogical, psychological, and ethical implications, in: *Proc. Int. Conf. AI Res.*, 4(1), 210–217.
- [3] Siau, K., Wang, W., 2020. Artificial intelligence (AI) ethics: Ethics of AI and ethical AI. *J. Database Manag.* 31(2), 74–87.
- [4] Belkina, M., et al., 2025. Implementing generative AI (GenAI) in higher education: A systematic review of case studies. *Comput. Educ. Artif. Intell.* 8, 100407.
- [5] Ajailia, N., Zerai, M., Debbabi, N., Damergi, A., Dchicha, M., 2024. Ethics, AI, and society: Shaping responsible AI engineers with CDIO framework, in: *Proc. Int. CDIO Conf.*, Chalmers University of Technology, 2024, pp. 253–262.
- [6] Arbelaez Ossa, L., Lorenzini, G., Milford, S. R., Shaw, D., Elger, B. S., Rost, M., 2024. Integrating ethics in AI development: A qualitative study. *BMC Med. Ethics* 25(1), 10.
- [7] Da Silva, M., Ferro, M., Mourão, E., Seixas, E. F. R., Viterbo, J., Salgado, L. C. C., 2024. Ethics and AI in higher education: A study on students' perceptions, in: Rocha, Á., Ferrás, C., Hochstetter Diez, J., Diéguez Rebolledo, M. (Eds.), *Information Technology and Systems*, Springer, 2024, pp. 149–158.
- [8] Kajiwara, Y., Kawabata, K., 2024. AI literacy for ethical use of chatbot: Will students accept AI ethics? *Comput. Educ. Artif. Intell.* 6, 100251.
- [9] Binns, R., 2021. Fairness in machine learning: Lessons from political philosophy. *arXiv preprint arXiv:1712.03586*.
- [10] Memarian, B., Doleck, T., 2023. Fairness, accountability, transparency, and ethics (FATE) in artificial intelligence (AI) and higher education: A systematic review. *Comput. Educ. Artif. Intell.* 5, 100152.
- [11] Pham, N., Ngoc, H. P., Nguyen-Duc, A., 2025. Fairness for machine learning software in education: A systematic mapping study. *J. Syst. Softw.* 219, 112244.
- [12] Behera, R. K., Janssen, M., Rana, N. P., Bala, P. K., Chakraborty, D., 2024. Responsible metaverse: Ethical metaverse principles for guiding decision-making and maintaining complex relationships for businesses in 3D virtual spaces. *Decis. Support Syst.* 187, 114337.
- [13] Malhotra, C., Kotwal, V., Dalal, S., 2018. Ethical framework for machine learning, in: *Proc. ITU Kaleidoscope: Machine Learning for a 5G Future*, 2018, pp. 1–8.
- [14] European Commission, 2019. Ethics guidelines for trustworthy AI, Publications Office of the European Union, 2019.
- [15] UNESCO, 2022. Ethics of artificial intelligence, United Nations Educational, Scientific and Cultural Organization.

- [16] Malaysian Science and Technology Information Centre, 2025. Official portal, MASTIC, Malaysia.
- [17] Holstein, K., Vaughan, J. W., Daumé III, H., Dudík, M., Wallach, H., 2019. Improving fairness in machine learning systems: What do industry practitioners need?, in: Proc. CHI Conf. Hum. Factors Comput. Syst., 2019, pp. 1–16.
- [18] Xu, Q., Gu, J., Lu, J., 2024. Leveraging artificial intelligence and large language models for enhanced teaching and learning: A systematic literature review, in: Proc. Int. Conf. Comput. Technol. Dev., IEEE, 2024, pp. 73–77.
- [19] Chai, F., Ma, J., Wang, Y., Zhu, J., Han, T., 2024. Grading by AI makes me feel fairer? How different evaluators affect college students' perception of fairness. *Front. Psychol.* 15.
- [20] Floridi, L., et al., 2018. AI4People — An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds Mach.* 28(4), 689–707.
- [21] Jobin, A., Ienca, M., Vayena, E., 2019. The global landscape of AI ethics guidelines. *Nat. Mach. Intell.* 1(9), 389–399.
- [22] Sarkis-Onofre, R., Catalá-López, F., Aromataris, E., Lockwood, C., 2021. How to properly use the PRISMA statement. *Syst. Rev.* 10(1), 117.
- [23] Mohammed, C., Rocke, S., 2025. Electrical and computer engineering freshmen and generative AI: Awareness, attitudes, and ethics, in: Proc. IEEE Global Eng. Educ. Conf. (EDUCON), 2025, pp. 1–8.
- [24] Shrivastava, P., 2025. Understanding acceptance and resistance toward generative AI technologies: A multi-theoretical framework integrating functional, risk, and sociolegal factors. *Front. Artif. Intell.* 8.
- [25] Hellmich, E. A., Vinal, K., Brandt, Z. M., Chen, S., Sparks, M. M., 2024. ChatGPT in language education: Centering learner voices. *Technol. Lang. Teach. Learn.* 6(3).
- [26] Stone, B. W., 2025. Generative AI in higher education: Uncertain students, ambiguous use cases, and mercenary perspectives. *Teach. Psychol.* 52(3), 347–356.
- [27] Huang, Q., Kumarasinghe, P. J., Jayarathna, G. S., 2024. Navigating the future: Exploring AI adoption in Chinese higher education through the lens of diffusion theory. *Interdiscip. J. Inf. Knowl. Manag.* 19.
- [28] Hazari, S., 2025. Marketing students' perceptions towards ChatGPT: An AI-assisted thematic analysis. *Mark. Educ. Rev.*
- [29] Coelho, R., et al., 2025. A call for clarity: Biology students advocate for guidelines for the use of generative AI in higher education. *J. Sci. Educ. Technol.*
- [30] Oliveira, L., Tavares, C., Strzelecki, A., Silva, M., 2025. Prompting minds: Evaluating how students perceive generative AI's critical thinking dispositions. *Electron. J. E-Learn.* 23(2).
- [31] Alkamli, S., Al-Yahya, M., Alyahya, K., 2024. Ethical and legal considerations of large language models: A systematic review of the literature, in: Proc. Int. Conf. Foundation Large Lang. Models, IEEE, 2024, pp. 576–586.
- [32] Funa, A. A., Gabay, R. A. E., 2025. Policy guidelines and recommendations on AI use in teaching and learning: A meta-synthesis study. *Soc. Sci. Humanit. Open* 11.
- [33] Zhang, Y., Reusch, P., 2025. Trust in and adoption of generative AI in university education: Opportunities, challenges, and implications, in: Proc. IEEE Global Eng. Educ. Conf. (EDUCON), 2025, pp. 1–10.
- [34] Kinney, M., Anastasiadou, M., Naranjo-Zolotov, M., Santos, V., 2024. Expectation management in AI: A framework for understanding stakeholder trust and acceptance of artificial intelligence systems. *Heliyon* 10(7), e28562.
- [35] Duy, N. B. P., Phuong, T. N. M., Chau, V. N. M., Nhi, N. V. H., Khuyen, V. T. M., Giang, N. T. P., 2025. AI-assisted learning: An empirical study on student application behavior. *Multidiscip. Sci. J.* 7(6).
- [36] Mambile, C., Mwogosi, A., 2025. Transforming higher education in Tanzania: Unleashing the true potential of AI as a transformative learning tool. *Technol. Sustain.* 4(1), 51–76.
- [37] McDonald, N., Johri, A., Ali, A., Collier, A. H., 2025. Generative artificial intelligence in higher education: Evidence from an analysis of institutional policies and guidelines. *Comput. Hum. Behav. Artif. Hum.* 3, 100121.

- [38] Alrayes, A., Henari, T. F., Ahmed, D. A., 2024. ChatGPT in education – Understanding the Bahraini academics' perspective. *Electron. J. E-Learn.* 22(2), 112–134.
- [39] Rosdiana, Sunandar, E., Purnama, A., Arribathi, A. H., Yusuf, D. A., Daeli, O.P.M., 2024. Strategies and consequences of AI-enhanced predictive models for early identification of students at risk, in: *Proc. Int. Conf. Creative Commun. Innov. Technol.*, IEEE, 2024, pp. 1–6.
- [40] Chan, M. M., Rosales, M., Hernandez-Rizzardini, R., Amado-Salvatierra, H.R., 2025. Ethical AI in education: A proposed model for responsible integration, in: *Proc. IEEE Global Eng. Educ. Conf. (EDUCON)*, 2025, pp. 1–7.