ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.01 (2025)

THE ROLE OF AI DETECTION TOOLS IN UPHOLDING ACADEMIC INTEGRITY: AN EVALUATION OF THEIR EFFECTIVENESS

Dr. Shahid Rafiq[1], Qurat-ul-Ain[2], Dr. Ayesha Afzal[3]
[1]Assistant Professor, Department of Social & Behavioral Sciences, Emerson University Multan, Pakistan, Email: shahid.rafiq@eum.edu.pk
[2]Assistant Professor, Department of Social & Behavioral Sciences, Emerson University Multan, Pakistan, Email: quratulain.husnain@eum.edu.pk
[3]Assistant Professor, University of Management and Technology, Lahore Pakistan
Email: ayeshaafzal@umt.edu.pk

Corresponding author: shahid.rafiq@eum.edu.pk

Abstract
The increasing use of artificial intelligence (AI) in academic writing has raised concerns about academic integrity. AI tools like ChatGPT enable students to generate essays and research papers with ease, prompting universities to adopt AI detection tools such as Turnitin AI Detection, GPTZero, and ZeroGPT. However, the effectiveness and ethical implications of these tools remain debated. This study investigates the accuracy, limitations, and ethical concerns of AI detection in academic settings. Data were collected through semi-structured interviews and focus group discussions with educators, academic integrity officers, and postgraduate students. Thematic analysis revealed three key themes: (1) effectiveness of AI detection tools, including false positives, AI evasion techniques, and limitations in detection; (2) ethical concerns, such as algorithmic bias, student academic rights, and privacy risks; and (3) the shift from punitive detection methods to AI literacy education, emphasizing the need for policy development and AI ethics integration. The findings suggest that AI detection alone is insufficient due to inconsistencies and biases. A holistic approach is needed, combining enhanced detection tools, transparent policies, and AI literacy programs to promote responsible AI use. This study contributes to the ongoing discourse on AI and academic integrity, advocating for a balanced, ethical, and educational approach to AI-assisted writing in academia.

***Keywords*:** *Academic integrity, AI detection tools, ethical concerns, AI literacy, higher education, plagiarism detection*

Introduction
The rapid advancements in artificial intelligence (AI) have significantly impacted the education sector, introducing both opportunities and challenges. AI-powered tools such as ChatGPT, Jasper, and Copy.ai have revolutionized the way students approach writing, research, and content creation. However, these technologies have also raised concerns regarding academic integrity, as they enable students to generate essays, reports, and research papers with minimal effort, leading to potential plagiarism, contract cheating, and misrepresentation of authorship. To combat these risks, AI detectors have been developed to identify AI-generated content and ensure that academic work remains authentic and ethically sound. The primary question that emerges, however, is: How effective are these AI detection tools in maintaining academic integrity?

AI detectors such as Turnitin's AI Writing Detection, GPTZero, ZeroGPT, and OpenAI's AI classifier analyze textual content for patterns indicative of AI-generated writing. These tools employ natural language processing (NLP) and machine learning models to assess writing structures, predictability, and coherence to distinguish between human and AI-generated text (Malik & Amjad, 2024). Several studies have explored the effectiveness of these detection tools, with mixed findings. Some researchers argue that AI detectors successfully flag AI-generated text with high accuracy, providing educators with a reliable mechanism to curb academic dishonesty (Karnalim, 2024). Others, however, highlight critical

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.01 (2025)

shortcomings, including false positives, false negatives, and biases in detection models, which can lead to wrongful accusations and erode trust in the assessment process (Biondi-Zoccai et al., 2025).

One of the primary challenges in AI detection lies in the evolution of generative AI models. As AI-generated text becomes more sophisticated and human-like, detection tools struggle to maintain their effectiveness. Some AI writing models, including GPT-4, Claude, and Gemini, have been designed to mimic human writing styles and even incorporate factual references to evade detection. Research suggests that AI writing tools are advancing at a pace that often outstrips the capabilities of AI detectors, making it difficult for institutions to keep up with the latest developments (Giri, 2025). Furthermore, adversarial attacks—where students intentionally modify AI-generated content to bypass detection further reduce the effectiveness of these tools (Lancaster et al., 2024).

Beyond technical limitations, ethical and legal concerns regarding AI detection tools have also been raised. Some researchers argue that over-reliance on AI detectors can lead to privacy violations and false accusations, as students may be penalized based on probabilistic predictions rather than concrete evidence (Reinhardt & Guartuche, 2024). Additionally, educators have expressed concerns about the potential biases in AI detection algorithms, as certain linguistic patterns or non-native English writing styles may be misclassified as AI-generated (Deans et al., 2024). This raises questions about fairness, accountability, and the ethical implications of using automated systems to determine academic integrity violations.

Despite these concerns, AI detectors remain a crucial component in academic integrity frameworks, provided they are used as complementary tools rather than absolute decision-makers. Experts recommend a holistic approach to academic integrity, where AI detection tools are integrated alongside traditional plagiarism detection methods, critical thinking assessments, and educational interventions to foster ethical writing practices among students (Susilo et al., 2024). Additionally, ongoing research and innovation in AI detection are needed to enhance accuracy, reduce biases, and adapt to the ever-changing landscape of AI-generated content (Eslit, 2024).

This paper aims to provide a comprehensive analysis of AI detectors in the context of academic integrity. It explores their effectiveness, limitations, ethical concerns, and future directions based on the latest empirical research. By critically examining the role of AI detection tools, this paper seeks to contribute to the ongoing discourse on maintaining integrity in an era of advanced AI technologies.

Background of the Study

Artificial Intelligence (AI) has become deeply integrated into various sectors, including education, where it serves as both an enabler and a disruptor of academic integrity. The rise of generative AI models, such as OpenAI's GPT series, Google's Gemini, and Anthropic's Claude, has introduced unprecedented challenges to the traditional methods of assessing student work. These tools allow users to generate essays, reports, and even research papers in seconds, raising concerns about originality, authorship, and ethical academic practices. As a response to this evolving landscape, AI detection tools have been developed to identify AI-generated content and mitigate academic dishonesty. However, their effectiveness, reliability, and ethical implications remain subjects of ongoing debate.

One of the primary drivers of this research is the increasing difficulty in distinguishing AI-generated content from human writing. Early plagiarism detection tools, such as Turnitin, were designed to detect text similarity by comparing student submissions to existing databases of academic papers and online sources. However, these tools were not equipped to identify content that is entirely original but generated by AI. AI detection tools

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.01 (2025)

such as GPTZero, Turnitin AI Detection, and ZeroGPT attempt to address this issue by analyzing text coherence, perplexity, and burstiness features indicative of AI-generated writing (Malik & Amjad, 2024). Despite these advancements, recent studies indicate that AI-generated text can be easily modified to evade detection, rendering these tools less effective over time (Giri, 2025). This raises concerns about the arms race between AI content generators and AI detectors, with each side evolving to outpace the other (Karnalim, 2024).

Additionally, the accuracy of AI detectors is inconsistent across different contexts. False positives cases where human-written text is incorrectly flagged as AI-generated—pose a serious risk to students' academic records and reputations. Conversely, false negatives—instances where AI-generated text is undetected—enable students to bypass detection systems, undermining academic integrity. Studies have reported varying degrees of reliability for different AI detection tools, with some failing to detect advanced AI-generated text altogether (Deans et al., 2024). This inconsistency raises concerns about the fairness, transparency, and trustworthiness of AI detection methods.

Beyond technical limitations, ethical and legal concerns surrounding AI detection tools have emerged as significant issues. Academic institutions that integrate AI detection software must consider privacy implications, particularly when student submissions are stored and analyzed by third-party software providers (Lancaster et al., 2024). Moreover, the lack of regulation in AI detection tools means that students may be penalized based on probabilistic assessments rather than definitive proof of misconduct. These ethical dilemmas highlight the need for a balanced approach that promotes academic integrity without infringing on students' rights (Reinhardt & Guartuche, 2024).

Furthermore, the role of educators and institutions in maintaining academic integrity must extend beyond detection tools. While AI detectors serve as valuable tools in identifying AI-generated content, they should not be the sole method of ensuring academic honesty. Instead, educators should adopt a multifaceted strategy, incorporating AI literacy, ethical writing practices, and critical thinking skills into curricula (Susilo et al., 2024). Research has shown that a culture of integrity, supported by proactive education rather than punitive measures, is more effective in deterring academic misconduct in the long run (Eslit, 2024).

Rationale Of The Study

Given the growing reliance on AI in education and the emerging challenges associated with AI-generated content, there is an urgent need for a comprehensive examination of AI detectors and their role in maintaining academic integrity. This study seeks to:

1. Evaluate the effectiveness of AI detection tools by analyzing their accuracy, strengths, and limitations.
2. Identify common challenges faced by educators and institutions when implementing AI detection software.
3. Examine the ethical and legal concerns associated with AI-generated content and AI detectors.
4. Propose alternative strategies to enhance academic integrity, including the role of pedagogy, policy-making, and AI literacy.

By addressing these research objectives, this study aims to contribute to the ongoing discourse on AI and academic integrity, providing data-driven insights and practical recommendations for educators, institutions, and policymakers. With AI continuously evolving, it is essential to adopt a forward-thinking approach that ensures academic integrity while embracing technological advancements in education.

Significance of the Research

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.01 (2025)

The increasing use of artificial intelligence (AI) in education has transformed how students learn and submit academic work, but it also raises concerns about plagiarism, authorship misrepresentation, and critical thinking. AI detection tools have been introduced to address these issues, but their effectiveness, reliability, and ethical implications remain under scrutiny. This research aims to evaluate the accuracy of AI detection tools, addressing challenges such as bias, privacy risks, and false accusations. It contributes to the development of fair AI integrity policies, providing insights for institutions to create balanced approaches to AI use. The study also emphasizes the importance of AI literacy in academia, promoting ethical usage and critical thinking skills. By offering practical recommendations for policy development, AI detection, and education, this research supports the integration of AI ethics in curricula and contributes to ongoing discussions on AI's role in education, ensuring academic integrity is preserved in the digital age.

Literature Review

The emergence of AI-generated content has created new challenges in academic integrity, prompting institutions to adopt AI detection tools to mitigate unethical practices. This literature review explores the effectiveness, limitations, ethical concerns, and evolving strategies surrounding AI detection tools in higher education.

Effectiveness of AI Detection Tools in Maintaining Academic Integrity

AI detection tools such as Turnitin AI Detection, GPTZero, and ZeroGPT have been widely adopted to identify AI-generated content. These tools utilize natural language processing (NLP) and machine learning (ML) models to analyze text coherence, sentence predictability, and linguistic style variations (Balalle & Pannilage, 2025). Research suggests that these tools perform reasonably well in detecting early AI-generated texts, but their reliability declines as AI models evolve to produce more human-like writing (Ahmed, 2025).

Recent studies have compared the effectiveness of different AI detectors. Malik & Amjad (2024) found that while Turnitin AI Detection identified 72% of AI-generated texts, GPTZero's detection rate was only 58%, suggesting variability in effectiveness across different platforms. Additionally, Ashqar et al. (2025) examined the use of explainable AI (XAI) models to improve detection accuracy and found that integrating XAI into AI detection systems enhanced transparency and interpretability.

However, AI detectors still face significant challenges in distinguishing partially AI-generated content, where students modify AI-generated text to evade detection (Najjar et al., 2025). The arms race between generative AI and detection tools continues to intensify, raising concerns about the long-term effectiveness of current AI detectors (Giray, 2025).

Limitations of AI Detection Tools

Despite their widespread use, AI detectors suffer from false positives, false negatives, and biases that undermine their reliability. Studies indicate that false positive rates—cases where human-written text is mistakenly flagged as AI-generated—range between 15–30% (Hanafi et al., 2025). This can result in wrongful accusations against students, leading to ethical and legal implications.

Additionally, some AI-generated texts can successfully bypass detection. Biondi-Zoccai et al. (2025) argue that prompt engineering techniques—where students modify AI-generated content using specific prompts make it more challenging for detectors to identify AI-assisted writing. Furthermore, the effectiveness of AI detectors varies based on disciplinary context and linguistic style. Research by Shah (2024) suggests that non-native English speakers' writing is more likely to be flagged as AI-generated due to stylistic differences, raising concerns about algorithmic bias.

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.01 (2025)

Another critical limitation is the lack of standardization in AI detection tools. Mondal (2025) highlights that each tool relies on proprietary AI models, making it difficult for educators to compare detection accuracy across platforms. The absence of peer-reviewed benchmarks further complicates the validation of AI detection technologies.

Ethical and Legal Concerns in AI Detection

The use of AI detection tools raises ethical dilemmas related to privacy, bias, and student rights. AI detectors often require cloud-based data storage, raising concerns about student data privacy (Lancaster et al., 2024). Some universities integrate these tools without fully disclosing how student submissions are processed, stored, or shared with third-party companies (Eshet, 2025).

Additionally, algorithmic bias remains a concern. AI detectors trained on English-language datasets may produce inaccurate results for non-English texts, leading to discriminatory practices in AI-assisted academic integrity enforcement (Giri, 2025). Scholars argue that AI detection tools should be supplemented with human oversight to ensure fair and unbiased decision-making (Oates & Johnson, 2025).

Legal challenges also emerge when AI detection tools misidentify students as engaging in academic dishonesty. Some students have successfully appealed AI-based plagiarism accusations, arguing that AI detectors do not provide concrete evidence but rather probabilistic assessments (Hanafi et al., 2025). This has prompted calls for ethical guidelines and regulatory frameworks to govern the use of AI detection tools in academia (Alghazo et al., 2025).

Future Directions: Improving AI Detection and Academic Integrity Policies

Given the limitations of current AI detection tools, scholars propose alternative approaches to maintaining academic integrity. Shepherd (2025) suggests that AI literacy programs should be integrated into university curricula, teaching students how to use AI tools ethically rather than relying solely on detection software. Other researchers advocate for hybrid detection models that combine traditional plagiarism detection with AI-assisted monitoring. Mondal (2025) proposes a multi-layered approach, integrating AI detection with behavioral analytics—tracking student writing habits over time to identify inconsistencies. Additionally, Lancaster et al. (2024) argue that universities should develop AI usage policies that clearly distinguish between permissible AI assistance and academic misconduct. These policies should emphasize critical thinking, citation ethics, and responsible AI usage rather than punitive measures alone.

The literature underscores both the potential and limitations of AI detection tools in maintaining academic integrity. While these tools serve as valuable aids in identifying AI-generated content, their effectiveness is limited by false positives, biases, and the evolving capabilities of AI writing models. Ethical and legal concerns further complicate their implementation, highlighting the need for human oversight, AI literacy education, and standardized policies. Future research should focus on developing more transparent, fair, and adaptive AI detection methodologies to ensure academic integrity in the digital age.

Theoretical Framework

This research draws on three key theoretical perspectives: Technological Determinism Theory, Deontological Ethics and Academic Integrity Theory, and the Adaptive Learning & AI Literacy Framework. Technological Determinism Theory (McLuhan, 1964) suggests that technological advancements, like AI-generated content, drive societal changes, prompting institutions to adopt AI detection tools to combat academic misconduct. However, this theory does not address the ethical concerns related to these tools, which are explored through Deontological Ethics and Academic Integrity Theory. Deontological ethics (Kant, 1785)

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.01 (2025)

stresses adherence to moral principles, while Whitley's (2001) Academic Integrity Theory emphasizes honesty and responsibility in scholarship. AI detection tools align with these values but also raise ethical issues such as false positives, algorithmic biases, and privacy concerns. To address the responsible use of AI, Siemens' (2013) Adaptive Learning Theory and Selwyn's (2020) AI Literacy Framework advocate for AI literacy programs. These programs teach students how to ethically engage with AI, fostering original thought and proper citation practices. This educational approach emphasizes empowerment over punishment, encouraging responsible AI usage and aligning academic integrity with evolving technological landscapes.

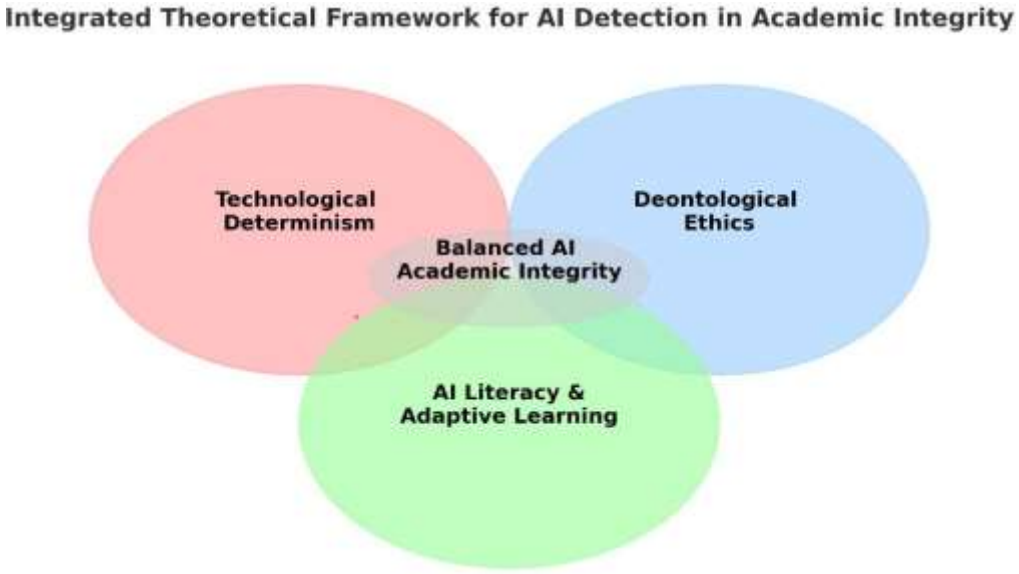

Figure 1: Theoretical Framework

This research proposes an integrated model where Technological Determinism, Deontological Ethics, and AI Literacy work together to shape a balanced approach to AI-generated content in education.

Table 1: Theoretical Model

Stage	Technological Determinism	Deontological Ethics	Adaptive Learning & AI Literacy
Challenge	AI-generated content challenges traditional assessments.	AI detection tools raise ethical concerns (false positives, biases).	Lack of AI literacy leads to misuse of AI tools.
Institutional Response	Universities adopt AI detection tools to regulate AI-generated writing.	Policies enforce fairness, transparency, and academic integrity.	AI literacy programs teach responsible AI usage.
Outcome	AI tools evolve alongside detection systems.	Academic integrity is upheld without unethical enforcement.	Students develop AI literacy, reducing reliance on detection tools.

This model underscores that while AI detection tools are valuable, they should not be the sole mechanism for maintaining academic integrity. Instead, a multi-pronged approach integrating technology, ethics, and education is necessary to effectively manage AI's impact on higher education. Technological Determinism explains the increasing reliance on AI detection, Deontological Ethics highlights the need for fair and ethical enforcement, and Adaptive

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.01 (2025)

Learning & AI Literacy offer proactive solutions to equip students with the skills necessary to use AI responsibly. By integrating these perspectives, this research emphasizes that AI detection should be complemented by ethical policies and AI literacy education to foster a balanced academic integrity framework in the AI era.

Research Methodology

This study employed a qualitative research design to explore the effectiveness, limitations, and ethical implications of AI detection tools in maintaining academic integrity. A phenomenological approach was adopted to capture the experiences, perceptions, and challenges faced by educators and students in using AI detection tools within academic institutions.

Participants and Sampling

The study used purposive sampling to select 35 participants, including university professors, academic integrity officers, and postgraduate students from diverse academic disciplines. Participants were chosen based on their experience with AI detection tools such as Turnitin AI Detection, GPTZero, and ZeroGPT. To ensure a diverse representation, participants were recruited from two prominent private-sector universities in Lahore, both of which had purposefully adopted AI detection technologies as part of their academic policies.

Data Collection Methods

Data for this study was collected through semi-structured interviews to gain in-depth insights into participants' experiences and perspectives. Individual interviews were conducted with professors and academic integrity officers to explore their perceptions of AI detection tools, challenges in implementation, and ethical concerns. Each interview lasted 30–45 minutes and was conducted in person, depending. With participant consent, interviews were recorded and transcribed verbatim for analysis.

Two FGDs were conducted with postgraduate students to examine their experiences with AI detection tools and how these tools impacted their writing practices and awareness of academic integrity. Each FGD included 5–7 participants, lasted approximately 60 minutes, and was moderated using guiding questions to encourage open yet focused dialogue on key research themes.

Data Analysis

The study employed thematic analysis following Braun and Clarke's (2006) six-step framework. This included data familiarization, initial coding, theme identification, reviewing and refining themes, defining themes, and reporting findings. The analysis identified recurring patterns in the perceptions and experiences of educators and students, focusing on the effectiveness, limitations, and ethical concerns of AI detection tools in academic integrity. The findings were categorized into three main themes, each with subthemes supported by direct participant quotations, providing a structured and comprehensive understanding of the data.

Ethical Considerations

The study adhered to ethical research guidelines to ensure the protection and confidentiality of participants. Prior to participation, all individuals received an informed consent form detailing the purpose of the study, data confidentiality measures, and voluntary participation rights. Pseudonyms were assigned to participants to ensure anonymity. Additionally, ethical approval was obtained from the respective university Institutional Review Board (IRB) before commencing data collection.

Trustworthiness and Rigor

To ensure the credibility and reliability of the findings, the study employed several validation strategies. Member checking was used, allowing participants to review interview transcripts

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.01 (2025)

and confirm the accuracy of their statements. Triangulation was also applied, where data from interviews and focus group discussions (FGDs) were cross-validated with institutional policies on AI detection to ensure robustness. Additionally, the researcher engaged in reflexivity by maintaining a reflective journal to document personal biases and enhance the objectivity of data interpretation. This qualitative methodology, incorporating semi-structured interviews, FGDs, and thematic analysis, provided rich, contextual insights into the experiences of educators and students regarding AI detection tools. The study effectively captured the real-world challenges and ethical dilemmas surrounding AI-assisted academic integrity, offering valuable perspectives for institutions and policymakers.

Table 1: Thematic Analysis

Codes	Subthemes	Main Themes
"AI flagged my original work as plagiarism"	False Positives & Misclassifications	**Effectiveness of AI Detection Tools**
"Students can bypass AI detection easily"	AI Evasion Techniques	
"Detection tools struggle with paraphrased AI content"	Limitations in Identifying AI-Modified Text	
"AI detectors aren't foolproof"	Reliability and Accuracy Issues	
"I was accused unfairly because of AI detection"	Student Academic Rights	**Ethical Concerns of AI Detection**
"The system is biased against non-native English speakers"	Algorithmic Bias & Fairness Issues	
"Data privacy is a big concern when using AI detectors"	Privacy & Confidentiality Concerns	
"We need to educate students on ethical AI use"	Need for AI Literacy Programs	**Shifting from Detection to Education**
"AI is a tool, not a threat student must learn how to use it ethically"	Integrating AI Ethics in Curriculum	
"Academic integrity policies should adapt to AI"	Institutional Policy Development	

Table 1 categorizes key findings into main themes and subthemes, offering a comprehensive view of the concerns surrounding AI detection tools. The "Effectiveness of AI Detection Tools" highlights issues like false positives and AI evasion techniques, pointing to limitations in the reliability and accuracy of these systems. Ethical concerns, such as bias against non-native English speakers and privacy issues, were also prominent, along with calls for more AI literacy programs. Participants emphasized the importance of adapting academic integrity policies and integrating AI ethics into the curriculum, shifting the focus from merely detecting AI-generated content to educating students on responsible AI use, as shown in figure 2 below.

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.01 (2025)

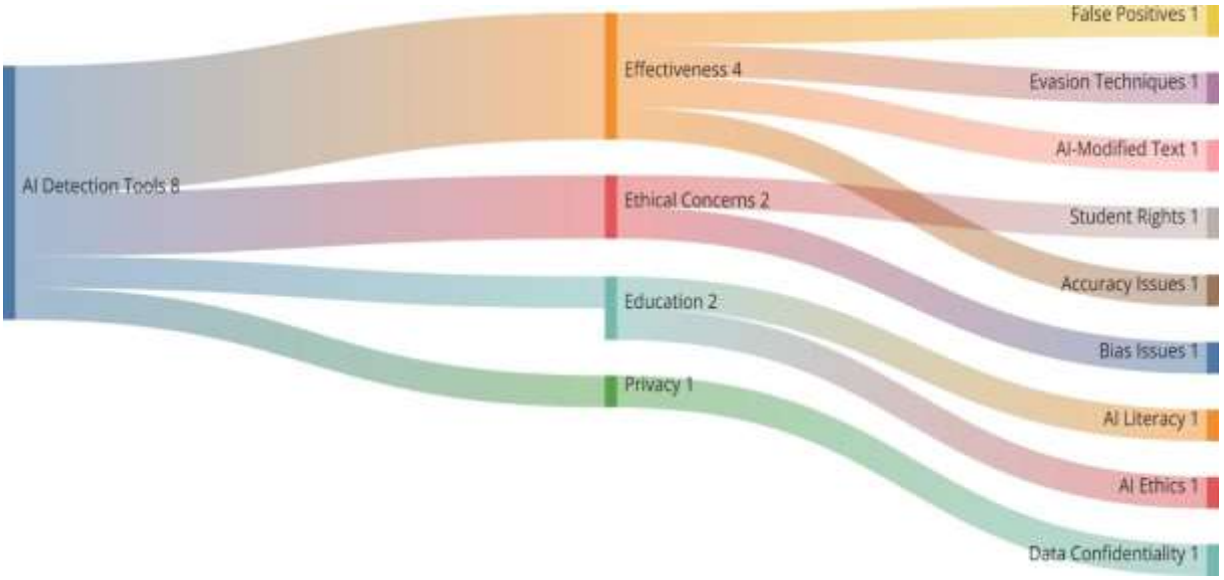

Figure 2: Main findings of Analysis

Effectiveness of AI Detection Tools

AI detection tools are widely used in academic institutions to identify AI-generated content in student submissions. However, participants highlighted several issues regarding their reliability, including false positives, AI evasion techniques, and accuracy limitations.

False Positives & Misclassifications

This theme underscores the significant issue of false positives in AI detection tools, where human-written content is incorrectly flagged as AI-generated. Participants expressed frustration with the distress this causes for students, who are often forced to prove the originality of their work despite having written it themselves. Educators also highlighted the added burden of manually reviewing flagged content, which increases their workload and raises concerns about the reliability of AI detection systems. These false positives highlight the need for more accurate and nuanced detection tools to ensure fairness and prevent unnecessary stress for students.

> *"I had a case where a student's paper was flagged for AI-generated content, but after reviewing it manually, it was clearly their own work. These false positives are a real issue." (Professor, 1)*
>
> *"I wrote my essay myself, and yet the AI detector flagged it. It was frustrating trying to prove my originality." (Postgraduate Student, 2)*

AI Evasion Techniques

This theme highlights the effectiveness of students who intentionally modify AI-generated content to bypass detection systems. Participants noted that savvy students use paraphrasing tools or manually edit AI-generated text, making it difficult for AI detection tools to identify the content as artificial. This raises concerns about the limitations of current detection technologies, as they struggle to keep up with the pace of advancements in AI writing tools. Educators and integrity officers emphasized that detection systems must evolve as rapidly as AI writing tools to remain effective in maintaining academic integrity.

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.01 (2025)

> *"Savvy students can rephrase AI-generated content using paraphrasing tools, and the AI detectors fail to catch it. So, how effective are these systems really?" (Academic Integrity Officer, 3)*

> *"Detection tools need to evolve as fast as AI writing tools; otherwise, they become useless." (Professor, 4)*

Limitations in Identifying AI-Modified Text

This theme highlights the growing challenge of distinguishing AI-generated text from human writing, as AI tools become more sophisticated. Participants noted that when students personalize or edit AI-generated content, it becomes nearly indistinguishable from authentic work, making it difficult for detection tools to identify potential misuse. This presents a significant challenge for academic integrity, as traditional detection methods struggle to keep up with the evolving capabilities of AI, requiring a reassessment of how academic institutions approach AI-assisted content.

> *"AI-generated content is no longer robotic. When students edit and personalize it, it becomes nearly impossible to detect." (Lecturer, 7)*

2. Ethical Concerns of AI Detection

The use of AI detection tools has raised ethical issues related to fairness, bias, and privacy. Many participants believed that over-reliance on these tools without human oversight could lead to wrongful accusations and unfair penalties.

Student Academic Rights

This theme highlights concerns about the reliability and transparency of AI detection tools in academic settings. Students and faculty expressed frustration over false accusations of AI-generated plagiarism, emphasizing that the tools are not flawless yet are often treated as definitive evidence. These issues point to the need for academic institutions to exercise caution, ensure transparency, and implement fair verification processes before penalizing students based solely on AI detection results. The participants revealed that;

> *"A student came to me in distress after being falsely accused of using AI. The detection tool isn't perfect, and yet universities treat it as infallible." (Professor, 13)*

> *"Universities need to be cautious before punishing students based solely on AI detection results." (Academic Integrity Officer, 9)*

Algorithmic Bias & Fairness Issues

This theme underscores concerns about linguistic bias in AI detection tools, particularly their disproportionate impact on non-native English speakers. Participants highlighted that the algorithms tend to flag complex phrasing or unconventional sentence structures as indicative of AI-generated content, even when the work is original. This bias raises questions about the fairness of AI detection tools, suggesting that students from diverse linguistic backgrounds may be unfairly targeted by these systems, which could undermine academic integrity. Some participants noted that;

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.01 (2025)

> *"The AI detectors disproportionately flag work by non-native English speakers. It's as if they assume complex phrasing means AI involvement." (Linguistics Professor, 19)*

> *"Students from diverse backgrounds struggle more with AI detection. The algorithms seem biased." (Postgraduate Student, 3)*

Privacy & Confidentiality Concerns

This theme highlights concerns about the privacy of student submissions, particularly when AI detection tools store or analyze their work in external databases. Educators and students expressed discomfort over the lack of explicit consent for the use of their academic work by third-party tools. This raises significant ethical concerns regarding data security and privacy, as students may not be fully aware of or agree to how their work is being handled, potentially violating their rights and undermining trust in these detection systems.

> *"I worry about student work being stored and analyzed by third-party AI detection tools without their explicit consent." (Ethics Researcher, 2)*

Shifting from Detection to Education

This theme emphasizes the importance of shifting the focus from merely policing AI usage to fostering AI literacy and ethical understanding among students. Educators highlighted the need for institutions to integrate AI ethics and literacy into their curricula, ensuring students are equipped to use AI responsibly. Participants stressed that developing a clear institutional policy that encourages ethical AI practices, rather than focusing solely on detection and punishment, is essential for preparing students to navigate AI technologies effectively and responsibly in their academic work.

Need for AI Literacy Programs

This theme emphasizes the importance of shifting the focus from merely policing AI usage to fostering AI literacy and ethical understanding among students. Educators highlighted the need for institutions to integrate AI ethics and literacy into their curricula, ensuring students are equipped to use AI responsibly. Participants stressed that developing a clear institutional policy that encourages ethical AI practices, rather than focusing solely on detection and punishment, is essential for preparing students to navigate AI technologies effectively and responsibly in their academic work.

> *"We need to shift from catching students to teaching them how to use AI responsibly." (Professor, 27)*

> *"Students don't always know where the ethical line is when using AI. More guidance is needed." (Lecturer, 23)*

Integrating AI Ethics in Curriculum

This theme emphasizes the importance of viewing AI as a learning tool rather than a threat to academic integrity. Participants argued that instead of banning AI, universities should focus on integrating ethical AI use into teaching and learning. Some suggested that AI literacy courses be incorporated into university curricula to help students understand how to use AI responsibly and effectively. This approach encourages a more constructive perspective on AI, recognizing its potential to enhance education while addressing ethical concerns.

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.01 (2025)

> *"AI should be treated like a learning tool, not just a threat to academic integrity." (Education Policy Researcher, 31)*

> *"Banning AI is unrealistic. Instead, we should focus on ethical AI integration in teaching." (Professor, 33)*

Institutional Policy Development

This theme highlights the need for academic integrity policies to adapt to the growing presence of AI in education. Participants argued that traditional approaches, designed for plagiarism detection, are no longer sufficient to address the complexities of AI-generated content. They emphasized that academic integrity policies must evolve to keep pace with technological advancements, ensuring they remain effective in maintaining fairness and accountability in an era of widespread AI use. This adaptation would allow institutions to better address the challenges posed by AI in academic settings.

> *"Academic integrity policies need to evolve alongside AI. What worked for plagiarism detection doesn't necessarily work for AI-generated content." (University Administrator, 9)*

The thematic analysis revealed a complex and evolving landscape of AI detection in academia. While AI detection tools serve as valuable instruments for identifying AI-generated content, their accuracy remains inconsistent, leading to false positives, detection loopholes, and evasion tactics by students. Ethical concerns, particularly fairness, privacy, and bias, also emerged as significant challenges.

Many participants emphasized that AI detection alone cannot uphold academic integrity. Instead, universities should transition from a punitive approach to an educational one, promoting AI literacy, ethical writing practices, and adaptive academic integrity policies. By integrating technology, ethics, and education, institutions can create a balanced approach that ensures both academic integrity and responsible AI use in higher education.

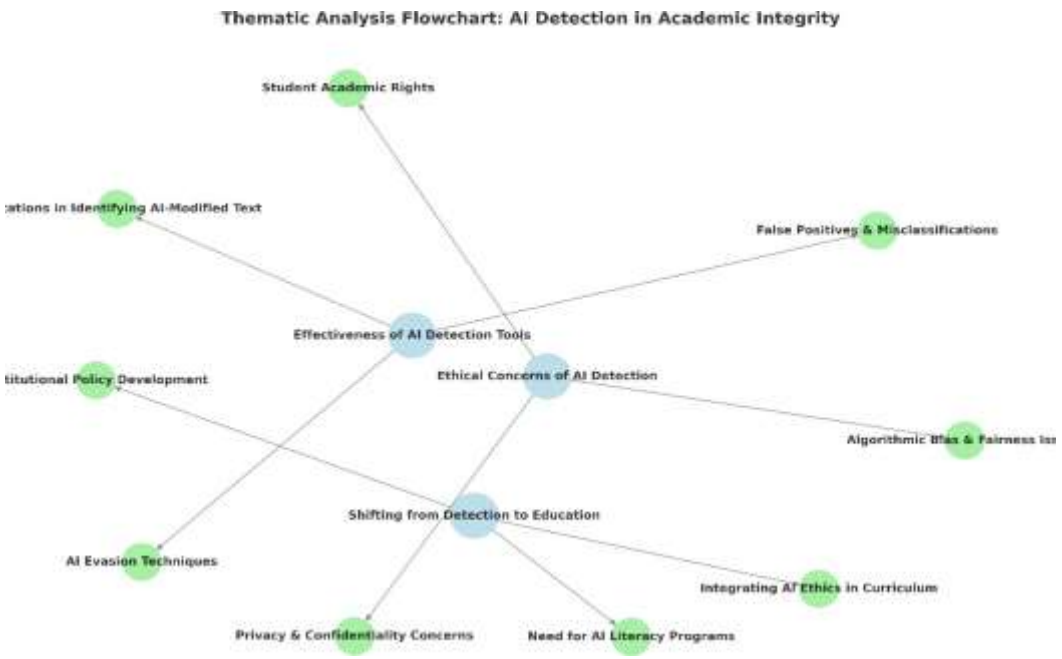

Figure 3: Thematic Analysis Map

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.01 (2025)

Discussion

The findings of this study align with recent research examining the effectiveness, limitations, and ethical considerations of AI detection tools in academic integrity. As artificial intelligence becomes increasingly integrated into education, institutions have turned to AI detection tools such as Turnitin AI Detection, GPTZero, and ZeroGPT to address concerns about AI-assisted plagiarism. However, the reliability of these tools remains contested, particularly as AI-generated text becomes more sophisticated. Research indicates that AI detection tools struggle with false positives and evasion techniques, raising concerns about their accuracy and practicality in real-world academic settings (William, 2024). While detection tools can identify AI-generated text with a reasonable degree of accuracy, they often misclassify human-written work, leading to wrongful academic penalties for students (Nnaka, 2024). This study also found that students have developed workarounds to bypass detection, such as paraphrasing AI-generated content or combining AI-assisted writing with manual edits, making it increasingly difficult for detection algorithms to remain effective (Tripathi & Thakar, 2024).

Beyond the limitations in accuracy, bias in AI detection tools emerged as a significant ethical concern. Recent studies suggest that AI detection models are more likely to flag content written by non-native English speakers, even when their work is original (Zapata-Rivera et al., 2024). This linguistic bias places international students at a higher risk of false accusations, reinforcing concerns about fairness and equity in AI-assisted assessments (Aleynikova & Yarotskaya, 2024). Additionally, AI detection tools lack transparency in their decision-making processes, meaning that students and educators often cannot challenge or verify AI-based classifications (Hanafi et al., 2025). These concerns highlight the need for human oversight in AI detection practices, as over-reliance on automated tools can exacerbate existing academic inequalities (Indolfi et al., 2024).

Another critical issue identified in this study is the lack of clear data privacy policies surrounding AI detection tools. Many AI-based plagiarism detection systems store student submissions in external databases, raising concerns about data ownership, confidentiality, and long-term storage (Raimi et al., 2024). Some institutions have adopted these tools without fully disclosing how student work is processed, which conflicts with privacy regulations such as GDPR and FERPA (Hanafi et al., 2025). Addressing these concerns requires greater transparency and policy development to ensure that AI detection technologies uphold both academic integrity and student rights.

Given these challenges, a growing body of research advocates for moving beyond AI detection toward AI literacy and ethical AI use in education. Rather than solely focusing on punitive measures, universities should integrate AI ethics education into curricula, helping students understand how to use AI responsibly (Zapata-Rivera et al., 2024). This aligns with findings from this study, where educators emphasized the need for proactive AI literacy programs that teach students how to ethically engage with AI tools without violating academic integrity principles (Tripathi & Thakar, 2024). Additionally, institutions should update academic integrity policies to reflect the realities of AI-assisted writing, ensuring that guidelines are clear on acceptable AI usage while promoting fair and transparent assessment methods (William, 2024).

Ultimately, the findings suggest that AI detection alone cannot sustain academic integrity in an era where AI-generated content is becoming more advanced and accessible. Instead, a holistic approach that combines AI detection, ethical education, and institutional policy reform is necessary to maintain fairness, transparency, and trust in academic assessments. Future research should focus on improving AI detection accuracy, reducing

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.01 (2025)

bias, strengthening privacy protections, and developing adaptive learning strategies that prepare students for responsible AI use in education and beyond.

Conclusion

The increasing use of AI-generated content in academic writing has led to the adoption of AI detection tools to uphold academic integrity. However, this study reveals significant challenges regarding their accuracy, fairness, and ethical implications. While tools like Turnitin AI Detection, GPTZero, and ZeroGPT play a crucial role in identifying AI-assisted writing, issues such as false positives, detection loopholes, and algorithmic bias remain prevalent. As AI-generated text becomes more sophisticated, students have found ways to bypass detection, raising concerns about the sustainability of AI-based academic integrity enforcement.

Ethical issues, such as bias against non-native English speakers and privacy risks, also hinder the widespread adoption of these tools. Bias in AI models disproportionately flags work by non-native speakers, potentially leading to unfair penalties. Data privacy concerns regarding how AI detectors store and process student submissions also require greater transparency and regulatory oversight.

This study supports the argument that AI detection should not be the sole method of maintaining academic integrity. Instead, a holistic approach combining AI detection tools with AI literacy programs and updated policies is essential. Future research should focus on improving AI detection accuracy, reducing bias, and developing comprehensive policies to foster a responsible AI-integrated learning environment.

Implications for Policy Development

To address the challenges of AI in academic integrity, universities should update policies to clearly define acceptable AI usage, distinguishing between permissible AI-assisted work and academic dishonesty. This may ensure consistent enforcement and eliminate confusion. Institutions must also develop transparent AI detection practices, including human oversight, appeals processes for false flags, and regular reviews to prevent bias and errors.

In addition to detection tools, universities should implement AI literacy programs to teach students ethical AI use. This can include workshops on AI-assisted writing, faculty development programs, and clear guidelines on when AI tools are acceptable.

Lastly, universities should collaborate with AI developers to improve detection systems. By refining algorithms, conducting regular audits, and creating adaptive systems that evolve with AI advancements, universities can ensure academic integrity while embracing responsible AI use.

References

Aleynikova, D. V., & Yarotskaya, L. V. (2024). AI Bias in Academic Integrity Detection: Challenges and Solutions. Tambov University Review. https://elibrary.ru/item.asp?id=60234916

Balalle, H., & Pannilage, S. (2025). Reassessing academic integrity in the age of AI: A systematic literature review on AI and academic integrity. Social Sciences & Humanities Open. DOI: 10.1016/j.ssaho.2025.100026

Biondi-Zoccai, G., Cazzaro, A., et al. (2025). *Artificial Intelligence Tools for Scientific Writing: The Good, The Bad and The Ugly*. Top Italian Scientists. https://journal.topitalianscientists.org/Artificial_Intelligence_Tools_for_Scientific_Writing_The_Good_The_Bad_and_The_Ugly

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. DOI: 10.1191/1478088706qp063oa

Deans, K. R., Jones, J., & Harvey, J. B. (2024). *Artificial Intelligence in Higher Education: A Comparative Study on the Performance and Detectability of AI-Generated Graduate-Level Coursework Utilizing ChatGPT.* IngentaConnect. DOI: 10.1016/j.iheduc.2024.100986

Eslit, E. (2024). *Challenges and Benefits of AI-driven Plagiarism Detection in Higher Education.* Preprints. DOI: 10.20944/preprints202403.0567.v1

Giri, A. (2025). *Revolutionizing Student Evaluation: The Power of AI-Powered Assessment.* IGI Global. DOI: 10.4018/978-1-6684-9058-5.ch013

Hanafi, A. M., Al-Mansi, M. M., & Al-Sharif, O. A. (2025). Generative AI in Academia: A Comprehensive Review of Applications and Implications for the Research Process. October 6 University. https://journals.ekb.eg/article_404708.html

Indolfi, C., Klain, A., Dinardo, G., & Decimo, F. (2024). Artificial Intelligence in the Transition of Academic Integrity: Ethical Considerations and Privacy Concerns. Frontiers in Medicine. https://www.frontiersin.org/journals/medicine/articles/10.3389/fmed.2024.1469161/full

Karnalim, O. (2024). *Identifying AI-Generated Code with Parallel KNN Weight Outlier Detection.* Springer. DOI: 10.1007/978-3-031-71530-3_29

Lancaster, T., Draper, M., Dann, S., & Crockett, R. (2024). *Developing Policies to Address Historic Contract Cheating and Misuse of Generative Artificial Intelligence.* Loughborough University Repository. https://repository.lboro.ac.uk/articles/journal_contribution/28194365

Malik, M. A., & Amjad, A. I. (2024). *AI vs AI: How effective are Turnitin, ZeroGPT, GPTZero, and Writer AI in detecting text generated by ChatGPT, Perplexity, and Gemini?* Journal of Applied Learning and Teaching. DOI: 10.24135/jalt.v7i1.2411

Mondal, H. (2025). Ethical engagement with artificial intelligence in medical education. Advances in Physiology Education. DOI: 10.1152/advan.00188.2024

Najjar, A. A., Ashqar, H. I., & Darwish, O. A. (2025). Detecting AI-Generated Text in Educational Content: Leveraging Machine Learning and Explainable AI for Academic Integrity. arXiv. https://arxiv.org/abs/2501.03203

Nnaka, C. (2024). Preserving Academic Integrity–Challenges and Solutions In The Era Of Artificial Intelligence. ResearchGate. https://www.researchgate.net/publication/384969429

Oates, A., & Johnson, D. (2025). ChatGPT in the Classroom: Evaluating its Role in Fostering Critical Evaluation Skills. International Journal of Artificial Intelligence in Education. DOI: 10.1007/s40593-024-00452-8

Raimi, L., Bamiro, N. B., & Lim, S. A. (2024). Two Facets of AI-Driven Applications for Sustainable Learning and Development: A Systematic Review of Tech-Entrepreneurial Benefits and Threats to Creative Education. Emerald. https://www.emerald.com/insight/content/doi/10.1108/S2043-052320240000023012/full/html

Reinhardt, K. S., & Guartuche Jr., O. (2024). *Curriculum Integration of AI Technology for Student Autonomy and Self-Efficacy.* Texas Association of Teacher Education. https://txate.org/resources/Documents/The%20Forum%20Volume%2016%20Winter%2024.pdf#page=40

Shepherd, C. (2025). Generative AI Misuse Potential in Cyber Security Education: A Case Study of a UK Degree Program. arXiv. https://arxiv.org/abs/2501.12883

Susilo, C. H., et al. (2024). *ChatGPT: The Future Research Assistant or an Academic Fraud?* Asia Pacific Fraud Journal. http://www.apfjournal.or.id/index.php/apf/article/view/347

Tripathi, A., & Thakar, S. V. (2024). Ethical Use of AI for Academic Integrity: Preventing Plagiarism and Cheating. Google Books. https://books.google.com/books?id=xSAaEQAAQBAJ

William, F. K. A. (2024). AI in Academic Writing: Ally or Foe? ResearchGate. https://www.researchgate.net/publication/380399233

Zapata-Rivera, D., Torre, I., & Lee, C. S. (2024). Generative AI in Education: Bias, Privacy, and Ethical Implications. Frontiers in Artificial Intelligence. https://www.frontiersin.org/articles/10.3389/frai.2024.1532896/full