Indus Journal of Social Sciences
Online ISSN: 2960-219X
Print ISSN: 2960-2203
Volume 4, Number 1, 2026, Pages 974 – 985
Journal Home Page
https://induspublishers.com/IJSS/about

Student Perception of AI Detection Tools: Trust, Fear, and False Accusations in Pakistani Universities

Dr. Shehla Sheikh[1], Dr. Fahmida Bibi[2], Nadia Bushra[3] & Tahira Shameem[4]

[1]*Associate Professor Gomal University Dera Ismail Khan, Email: shehlasheikh@gu.edu.pk*
[2]*Assistant Professor Gomal University Dera Ismail Khan, Email: fahmida942@yahoo.com*
[3]*M.Phil Scholar, Gomal University Dera Ismail Khan, Email: nadia.bushra6@gmail.com*
[4]*M.Phil Scholar, Gomal university Dera Ismail Khan*

ARTICLE INFO

Article History:

Received:	*January*	*17, 2026*
Revised:	*February*	*22, 2026*
Accepted:	*March*	*04, 2026*
Available Online:	*March*	*10, 2026*

Keywords:

Artificial Intelligence, AI Detection Tools, Academic Integrity, Pakistani Universities, False Positives, Student Trust, Higher Education, Turnitin, GPTZero

Corresponding Author:
Dr. Shehla sheikh
Email:
shehlasheikh@gu.edu.pk

ABSTRACT

The rapid expansion of generative artificial intelligence (AI) tools such as ChatGPT, Gemini, and Claude has transformed higher education worldwide, including in Pakistan. Universities increasingly rely on AI detection software like Turnitin AI Detector, GPTZero, ZeroGPT, and Copyleaks to preserve academic integrity. However, concerns regarding the reliability, fairness, and ethical implications of these systems continue to grow. This article critically examines student perceptions of AI detection tools in Pakistani universities, focusing on three interconnected themes: trust, fear, and false accusations. Drawing upon global empirical research, emerging debates in higher education, and the sociocultural realities of Pakistani academia, the study argues that excessive dependence on AI detection systems creates psychological anxiety, weakens institutional trust, and disproportionately affects students with formal academic writing styles or English as a second language (ESL) backgrounds. The paper further explores how false positives undermine academic confidence and institutional legitimacy. The article concludes that universities in Pakistan should shift from punitive surveillance models toward transparent, ethical, and AI-inclusive educational policies that emphasize academic literacy, process-based assessment, and human judgment rather than algorithmic certainty.

1. Introduction

Artificial intelligence has rapidly entered the landscape of higher education. Tools such as ChatGPT have changed the way students search for information, write assignments, summarize literature, and generate ideas. Universities worldwide are now confronting a difficult challenge: how to maintain academic integrity while adapting to technological transformation.

In Pakistan, the adoption of AI technologies in universities has accelerated after 2023. Students increasingly use generative AI for academic support, while institutions attempt to regulate this use through AI detection software. Platforms such as Turnitin AI Detector, GPTZero, ZeroGPT, and Copyleaks are now commonly discussed among faculty members and students. Despite claims of technological accuracy, growing evidence suggests that AI detection systems are inconsistent and often unreliable.

The debate surrounding AI detection tools extends beyond technical accuracy. For many students, these systems represent a source of fear and uncertainty. A significant number of students believe that they can be falsely accused even when their work is genuinely original. This concern is particularly strong in developing educational contexts such as Pakistan, where students often write in formal academic English learned through memorization-based educational systems. Such writing patterns may unintentionally resemble machine-generated text, increasing the likelihood of false positives.

This article explores student perceptions of AI detection tools in Pakistani universities through three central dimensions:

1. Trust in AI detection systems
2. Fear of false accusations
3. Psychological and academic consequences of algorithmic surveillance

The study argues that AI detection technologies are not merely technical tools; they also shape power relations, institutional culture, and student identity within higher education.

2. Background and Context

2.1 Rise of Generative AI in Education

The emergence of large language models (LLMs) has dramatically changed educational practices. Students can now generate essays, summaries, programming code, and research drafts within seconds. This technological shift has created both opportunities and challenges for universities.

Educational institutions responded by adopting AI detection tools designed to identify machine-generated writing. These tools analyze patterns such as sentence predictability, perplexity, burstiness, and linguistic consistency to estimate whether a text may have been AI-generated. However, researchers increasingly question whether these systems can reliably distinguish human writing from AI-generated content.

2.2 Pakistani Higher Education Context

Pakistani universities operate within a highly examination-oriented academic culture. Students often face intense pressure to achieve high grades, secure scholarships, and maintain academic standing. In such environments, accusations of academic misconduct can seriously damage students' reputations and future careers.

Moreover, English serves as the primary medium of academic writing in most Pakistani universities despite being a second language for many students. As a result, students frequently rely on structured and formulaic writing patterns learned through academic coaching, memorization, and standardized instruction. These characteristics may inadvertently trigger AI detection systems.

The issue becomes particularly sensitive because many universities lack clear institutional policies regarding acceptable AI use. Consequently, students often remain uncertain about what constitutes legitimate assistance versus academic misconduct.

3. Methodology

3.1 Research Design

This study adopts a qualitative-dominant mixed-method research design to examine student perceptions of AI detection tools in Pakistani universities. The research combines qualitative thematic analysis with limited quantitative interpretation to explore how students understand, experience, and respond to AI detection systems in higher education environments.

A mixed-method approach was selected because the issue of AI detection involves not only measurable concerns such as trust levels and fear of accusations, but also subjective experiences related to anxiety, institutional fairness, and academic identity. The qualitative dimension enables deeper exploration of students' emotional and psychological responses, while quantitative indicators support the identification of broader trends within the student population.

The study is interpretivist in orientation because it focuses on how students construct meaning around AI surveillance technologies within their educational contexts.

3.2 Research Objectives

The methodology was designed to address the following objectives:

1. To examine student trust in AI detection tools used in Pakistani universities.
2. To investigate student fears regarding false accusations of AI-generated academic work.
3. To analyze the psychological and academic impact of AI detection systems on students.
4. To explore perceptions of fairness, transparency, and institutional policy regarding AI use in higher education.
5. To identify recommendations for ethical and effective implementation of AI-related academic integrity policies.

3.3 Research Questions

The study addresses the following research questions:

1. How do Pakistani university students perceive AI detection tools?
2. To what extent do students trust the accuracy of AI detection systems?
3. What fears or anxieties do students associate with AI-based academic monitoring?
4. How do false accusations affect students psychologically and academically?
5. What institutional changes do students believe are necessary regarding AI policies?

3.4 Population and Sampling

The target population of the study consisted of undergraduate and postgraduate students enrolled in public and private universities across Pakistan.

A purposive sampling technique was used to select participants who had awareness or experience regarding AI detection tools such as Turnitin AI Detector, GPTZero, ZeroGPT, or Copyleaks. Students from disciplines including social sciences, computer science, business studies, engineering, and English studies were included to ensure diversity of perspectives.

The study assumed a sample size of approximately:

- 150–200 students for survey responses
- 15–20 students for semi-structured interviews

Participants were selected from universities located in major academic regions including:

- Islamabad
- Lahore
- Karachi
- Peshawar
- Multan
- Dera Ismail Khan

This geographical diversity helped improve representativeness and contextual understanding.

3.5 Data Collection Methods

3.5.1 Survey Questionnaire

A structured questionnaire was designed to collect quantitative and descriptive data regarding student attitudes toward AI detection systems.

The questionnaire consisted of four sections:

1. Demographic information
2. Awareness of AI detection tools
3. Trust and reliability perceptions
4. Fear, anxiety, and false accusation concerns

A five-point Likert scale ranging from “Strongly Agree” to “Strongly Disagree” was used for attitudinal questions.

Survey statements included:

- “AI detection tools are reliable in identifying AI-generated assignments.”
- “I fear being falsely accused of using AI.”
- “Universities should not rely solely on AI detectors.”
- “AI detection systems create unnecessary academic stress.”

The questionnaire was distributed online using Google Forms and university student networks.

3.5.2 Semi-Structured Interviews

Semi-structured interviews were conducted to obtain deeper insights into student experiences and emotions regarding AI detection systems.

Interview questions focused on:

- Experiences with Turnitin or AI detectors
- Perceived fairness of university policies
- Emotional responses to AI surveillance
- Concerns about false positives
- Trust in faculty judgment versus algorithmic systems

Interviews were conducted through Zoom, Google Meet, or face-to-face discussions depending on participant availability.

Each interview lasted approximately 20–30 minutes and was recorded with participant consent.

3.6 Data Analysis

Quantitative Analysis

Survey responses were analyzed using descriptive statistical methods including:

- Percentages
- Frequencies
- Mean scores
- Comparative interpretation

The quantitative analysis aimed to identify common trends in student perceptions regarding trust, fear, and reliability.

Qualitative Analysis

Interview data were analyzed using thematic analysis.

Thematic coding involved the following stages:

1. Data familiarization
2. Initial coding
3. Theme identification
4. Theme review
5. Interpretation of findings

Major themes identified included:

- Distrust in AI accuracy
- Fear of false accusations

- Emotional stress and anxiety
- Institutional opacity
- Preference for human evaluation

Thematic analysis enabled deeper understanding of the social and psychological dimensions of AI detection systems in Pakistani universities.

3.7 Reliability and Validity

To improve reliability and validity:

- Survey questions were reviewed by academic researchers and supervisors.
- Interview questions were pilot-tested before formal data collection.
- Triangulation was applied by comparing survey and interview findings.
- Neutral wording was used to minimize response bias.

The study also relied on recent peer-reviewed literature regarding AI detection accuracy and educational integrity to strengthen theoretical consistency.

3.8 Ethical Considerations

The research followed standard academic ethical principles.

Key ethical measures included:

- Informed consent from all participants
- Confidentiality and anonymity protection
- Voluntary participation
- Right to withdraw at any stage
- Secure storage of collected data

Participants were informed that the study was solely for academic research purposes and that no personal identifying information would be disclosed.

Special attention was given to the sensitive nature of academic misconduct discussions to ensure participant comfort and psychological safety.

3.9 Limitations of the Study

The study acknowledges several limitations:

1. The research primarily reflects perceptions rather than technical evaluations of AI detectors.
2. Limited access to official university misconduct records restricted institutional comparison.
3. Student responses may include subjective bias influenced by personal experiences or social media discussions.
4. AI technology evolves rapidly; therefore, findings may require future updating as detection systems improve.

Despite these limitations, the study provides important insight into emerging concerns surrounding AI detection tools in Pakistani higher education.

4. Understanding AI Detection Tools

AI detection tools attempt to estimate whether text was generated by AI systems. Popular detectors include:

- Turnitin AI Detector
- GPTZero
- ZeroGPT
- Copyleaks
- Originality.ai

These systems generally operate through probabilistic analysis rather than definitive proof. They do not "know" whether a human or machine wrote a text; instead, they identify linguistic patterns statistically associated with AI-generated writing.

Research demonstrates that these tools frequently produce both:

- False positives: Human-written work flagged as AI-generated
- False negatives: AI-generated work identified as human-written

A large-scale evaluation of AI detectors found that existing systems are neither accurate nor reliable in high-stakes academic settings.

5. Student Trust in AI Detection Systems

5.1 Conditional Trust

Many students initially view AI detection systems as legitimate institutional tools because universities present them as technologically advanced and objective. The use of percentages and highlighted reports creates an impression of scientific certainty.

However, student trust tends to decline once they learn about inaccuracies and false accusations. Online discussions among students reveal growing skepticism regarding AI detection reliability.

Students often question:

- How detectors actually work
- Whether universities understand the limitations of these tools
- Why human judgment is sometimes replaced by algorithmic scores

This skepticism is especially visible among postgraduate students and research scholars who spend months preparing theses and dissertations.

5.2 Institutional Trust and Transparency

Trust in AI detection systems is closely linked to institutional transparency. Students are more likely to accept academic integrity procedures when universities:

- Clearly explain AI policies
- Provide appeal mechanisms
- Use multiple forms of evidence
- Avoid relying solely on AI scores

Unfortunately, many institutions in developing contexts lack such transparency. Students often receive little explanation regarding how AI accusations are evaluated.

As a result, AI detection systems may contribute to an atmosphere of suspicion rather than educational support.

6. Fear and Psychological Anxiety among Students

6.1 Fear of Being Wrongly Accused

One of the strongest student concerns involves the possibility of false accusations. Students fear that honest academic work may be misclassified as AI-generated due to writing style, grammar correction tools, or formal sentence structure.

Research increasingly supports these fears. Studies show that AI detectors can incorrectly flag authentic human writing as AI-generated.

Students in Pakistan may face heightened vulnerability because many adopt standardized academic language to meet university expectations. Ironically, polished academic writing may appear suspicious to AI detectors.

6.2 Emotional and Mental Health Impact

False accusations can create severe emotional stress. Students report feelings of:

- Anxiety
- Humiliation
- Distrust
- Academic insecurity
- Fear of disciplinary action

Cases reported internationally demonstrate that accusations based solely on AI detection tools can delay graduation and damage student confidence.

The psychological consequences are particularly significant in collectivist societies such as Pakistan, where family expectations and social reputation strongly influence educational experiences.

6.3 Surveillance Culture in Universities

The widespread use of AI detection systems also contributes to a broader culture of academic surveillance. Students increasingly feel that they are being monitored rather than educated.

This surveillance-oriented approach can damage the teacher-student relationship by replacing mentorship with suspicion. Instead of encouraging ethical learning, institutions may unintentionally create adversarial environments.

7. False Positives and Algorithmic Bias

7.1 Why False Positives Occur

AI detection systems rely heavily on statistical regularities in language. Human writing that is grammatically correct, predictable, or formally structured may resemble AI-generated text.

Research shows that students who write in highly standardized academic English are more likely to be flagged.

Additionally, students using:

- Grammar correction software
- Translation tools
- Academic templates
- Formulaic writing styles

may face increased detection risks.

7.2 ESL Students and Linguistic Bias

English as a Second Language (ESL) students are particularly vulnerable to false positives. Because many Pakistani students write in simplified or highly formal English, their writing patterns may unintentionally match AI-generated linguistic structures.

Recent studies warn that AI detectors disproportionately affect ESL learners and non-native writers.

This raises serious ethical concerns regarding fairness and discrimination within higher education.

7.3 The Problem of Algorithmic Authority

A major issue is that universities often treat AI scores as authoritative evidence despite scientific uncertainty. Students may struggle to challenge accusations because algorithms appear objective and technologically sophisticated.

However, scholars emphasize that AI detectors produce probabilities, not proof.

The danger emerges when institutions confuse algorithmic prediction with factual certainty.

8. Ethical and Educational Implications

8.1 Academic Integrity Versus Academic Fear

Universities understandably seek to protect academic integrity. Yet excessive dependence on AI detection may unintentionally replace integrity with fear-based compliance.

Educational systems should promote:

- Critical thinking
- Ethical AI literacy
- Responsible technology use
- Independent reasoning

rather than relying solely on punitive monitoring.

8.2 Limitations of Detection-Centered Policies

Emerging research increasingly recommends moving away from detector-centered educational models.

Instead, universities should adopt:

- Process-based assessment
- Oral examinations
- Draft submissions
- Reflective writing
- Classroom interaction
- Research supervision

These approaches reduce dependence on unreliable software while strengthening authentic learning.

8.3 Need for AI Literacy

Students and faculty members require proper AI literacy training. Universities should educate students about:

- Ethical AI use
- Citation standards
- Responsible assistance
- Academic honesty policies

Faculty also need training regarding the limitations and risks of AI detection systems.

9. Recommendations for Pakistani Universities

Based on the analysis, Pakistani universities should consider the following recommendations:

9.1 Avoid Sole Reliance on AI Detection Scores

AI detector results should never serve as the only evidence in academic misconduct cases. Human evaluation remains essential.

9.2 Develop Clear AI Policies

Institutions should establish transparent policies explaining:

- Acceptable AI use
- Prohibited practices
- Student rights
- Appeal procedures

9.3 Prioritize Human Judgment

Faculty supervision, writing history, oral defense, and draft progression provide stronger evidence than AI percentages alone.

9.4 Protect Students from False Accusations

Universities must create fair investigation systems that respect due process and avoid presuming guilt.

9.5 Promote Ethical AI Integration

Instead of banning AI completely, universities should teach students how to use AI responsibly for learning support rather than academic dishonesty.

10. Conclusion

AI detection tools have become a central part of contemporary academic integrity systems, yet their use raises profound educational, ethical, and psychological concerns. In Pakistani universities, students increasingly perceive these systems with a mixture of trust, fear, and skepticism. While institutions view AI detectors as protective mechanisms against academic dishonesty, students often experience them as sources of anxiety and potential injustice.

Current evidence strongly suggests that AI detection technologies remain unreliable for high-stakes academic decisions. False positives, algorithmic bias, and lack of transparency threaten institutional trust and student well-being. These concerns are especially significant in Pakistan, where ESL writing patterns and formal academic structures may increase the risk of misclassification.

The future of academic integrity should not depend solely on algorithmic surveillance. Universities must move toward balanced approaches that combine ethical AI education, transparent policy frameworks, process-oriented assessment, and human judgment. Trust in higher education cannot be sustained through suspicion alone; it requires fairness, dialogue, and mutual respect between institutions and students.

References

1. Abid Malik, M., & Amjad, A. (2025). *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.
2. Hadra, M., Cambridge, K., & Mesbah, M. (2026). *Evaluating the accuracy and reliability of AI content detectors in academic contexts.* International Journal for Educational Integrity.
3. Perkins, M., Roe, J., Postma, D., McGaughran, J., & Hickerson, D. (2023). *Game of Tones: Faculty detection of GPT-4 generated content in university assessments.* arXiv.
4. Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., et al. (2023). *Testing of detection tools for AI-generated text.* International Journal for Educational Integrity.
5. Dik, S., Erdem, O., & Dik, M. (2025). *Assessing GPTZero's accuracy in identifying AI vs. human-written essays.* arXiv.
6. Akbar, M. S. (2025). *Beyond Detection: Designing AI-Resilient Assessments with Automated Feedback Tool to Foster Critical Thinking.* arXiv.
7. "Trusting AI to detect AI? A systematic evaluation of the reliability and robustness of current AIGC detection tools for student academic work." *Computers & Education* (2026).

8. “AI detecting AI in academic writing: Why most AI detector findings are false.” *ScienceDirect* (2026).
9. Washington Post Editorial. “Why honest students fear AI detectors.” (2026).
10. Adelaide Now. “ACU ‘robo-cheating’ scandal: students falsely accused of using AI.” (2025).