International Journal of Scientific Research and Modern Technology (IJSRMT) ijsrmt.com
Volume 5, Issue 5, 2026
DOI: https://doi.org/10.38124/ijsrmt.v5i5.1406

Shadow AI Usage as an Emerging Insider Threat in Secondary Schools

Joram Bwambale[1]

[1]National Curriculum Development centre

Publication Date: 2026/05/28

Abstract

Because of pressures of work and the pace of work, education agencies are finding ways around networked, organizational IT controls and policies. The result is a fresh insider threat named Shadow AI. It has been reported that 66% of faculty members and 41% of students use artificial intelligence tools regularly without sanction. For instance, the secondary education sector is a critical security environment in which unauthorised generative AI systems have rapidly proliferated. This paper points out how the education policy does not lay enough focus on security vulnerabilities in regard to data vulnerabilities. The educational policy used to apply academic integrity bias prohibition as if they were stronger than the risks posed by educational cheating and educational cheating. Impulsive Behavior Students, Discretionary Workload Managing Educators, Negligent Insiders They skip and elude controls in the . For instance, an external organization obtained the release of confidential student records and personal information. The controls are unusable. The theft of guest information, use of third-party LLMs without permission, and invasion of privacy has become possible because of it. The review of literature identifies major areas for policy gaps. This policy is not very clear. It punishes too much. It misses basic AI education. The paper proposes governance frameworks to prevent the potential development of structural vulnerabilities at instances of the object that are ‘essentially safe’ and at instances that are ‘explicitly unsafe’, rather than at the proactive, reactive model.

There needs to be embedded AI literacy programmes, iterative policy review processes, and “privacy-by-design” controlled AI tools like licensed enterprise services and sandboxed educational tools for the successful integration of AI in secondary schools.

I. INTRODUCTION

The education sector is undergoing a transformative phase rapidly propelled by the infusion of artificial intelligence (AI) into the ecosystem. Schools have yet to roll out government-approved software, but the real spike in AI use is through “under-the-table” and “organisation unapproved” software. Using the term “Shadow AI” refers to a widening of the concept of Shadow IT software or hardware used without the organization’s approval in an effort to bypass (real or perceived) constricting organizational processes (Silic et al, 2025). Shadow AI entails unauthorized application of Artificial Intelligence models, tools, or services by users, and lacks consent, oversight or management by the independent cybersecurity or Information Technology department or organizational unit controlled by that organization. Puthal et al. (2023)

The main reason why schools refrain from using official software is simple user pragmatism and systemic friction. According to Silic, Silic, and Kind-Trüller (2025), teachers and educational managers consider the central IT software approval process to be slow and expensive, rigid or politicised. Put simply, when the central IT office/educational staff take longer than planned to meet expectations or have a different academic knowledge need, “under-the-radar,” Shadow AI software becomes a useful “gap-filler” (Silic, Sil

There is a strong incentive to bypass formal controls Because of the looming presence of a heavy workload and tight deadlines enhancing productivity (Puthal et al., 2025; Silic, Silic, & Kind-Trüller, 2025). In addition, easily accessible AI platforms and open-source frameworks that use low-code/no-code have already enabled people to exploit powerful technology that is not available in their organization’s default software (Puthal et al., 2025).

Bwambale, J. (2026). Shadow AI Usage as an Emerging Insider Threat in Secondary Schools. *International Journal of Scientific Research and Modern Technology*, *5*(5), 24–29. https://doi.org/10.38124/ijsrmt.v5i5.1406

A governance drift zone refers to the area where the stated (static and not formative) official policies and guidelines do not match the emerging usage, and the local beings activate AI the other way round. It will be possible for Shadow AI to thrive in educational institutes without any regulatory intervention (Silic, Silic, & Kind-Trüller, 2025.

II. PROBLEM STATEMENT

In enacting policies for generative AI, schools have rushed to block the tools. However, analysis shows their limited framing; which is overwhelmingly around cheating, plagiarism, and blocking AI use (Jiang, 2025; Seehgai, 2025; Sojan, Chow, & Peng, 2025) This has created a major policy vacuum, whereby most schools only concerned about the pedagogical risk of "cheating" are seriously ignoring the security risk of data leakage and third-party processing (Jiang, Xie, & Cao, 2025). Because students' personal information and teachers' instructional materials are entering the public domain without their permissions, it remains essential to investigate the impact of institutional policies on these unconsented data leaks (Haney et al., 2025; Kortright & Lee, 2025; Piri, 2025; Sengupta, 2024).

The use of shadow AI has led to a considerable amount of organisational blind spots (Ojowa et al., 2026). Using undocumented models to input PII, circumvention of normal security and compliance checks expose organizations to internal theft and external unpermitted processing, in addition to incurring regulatory penalties (Puthal et al., 2025).

The architecture of the LLMs is such that they retain large amounts of training data and thus can "memorize" sensitive private information (Barati et al., 2025). As per Perlin (2027), a malicious actor may conduct attacks, namely data extraction or model inversion, which refer to the recovery and extract any sensitive stored data in these models.

This oversight issue will be raised further in the context of K-12. Regulatory processes operating around the development and use of generative AI does not take into account child development nor does it provide any clear guidance on how meaningful parental consent for processing can be obtained (Piri, 2024; Barati, Christensen and Zhou, 2025).

The ramifications of these unaddressed vulnerabilities is already playing out in real-time, e.g., the Vancouver Public Schools in 2025 publicly exposed via unprotected links one of the largest breaches of K-12 student record in history. More than 3,500 very sensitive student records were leaked following AI chat transcripts tracking self-harm and cyberbullying (Barati et al. 2025). Consequently, the real issue in secondary schools is not AI passing examinations; it is data exposure – unmitigated systemic institutional and student data exposure –through.

III. LITERATURE REVIEW

- *Insider Threat Taxonomy in Educational Environments*

Although "Insider Threat Theory" is most often applied to the corporate or financial realm, it has been said that the most recent literature heavily adopts criminological and behavioral theory to explain the rise of insider vulnerabilities (Troublefield, 2025). With respect to Shadow AI, the literature distinguishes between two main types of insider risk: the very pressured student and negligent educator.

In contrast to the corporate cybercriminals who seek financial or malicious gain, the vulnerabilities of students stem largely from cognitive overload, impulsiveness and academic-related stress. Younger learners appear particularly impulsive with their data when they are presented with speedier academic help and gratification rather than complex privacy notices (2024). In conducting analysis of academic misconduct with the help of AI, many scholars must rely on the Fraud Triangle Theory that stipulates that students may commit a neo-plagiarism due to the convergence of high academic pressure, the chance to do so with readily available AI tools and the reasoning that using generative AI is a victimless act (Alsharefeen, 2025). Additionally, students are not aware of the data-related workings of open-access language model so they end up oversharing their personal information during chat box interactions (Author Unknown, 2024).

On the contrary, educators who are insider threats fundamentally have limited cognitive capacity and that58 results in "copying" (also spelt "coping"). Due to time pressure and limited resources, faculty members often behave as 'street-level bureaucrats who have a lot of discretion with policies (Alsharefeen & AlSayari, 2025).58 The high workload of educators leads them to develop discretionary coping strategies. For instance, some choose to ignore their institution's policies while others use unvetted Shadow AI tools to fast track their grading and lesson planning process (Alsharefeen and AlSayari, 2025; Puthal et al., 2025). Although not inherently malicious, these well-intentioned time-saving behaviors create systemic loopholes, resulting in the exposure of student data and undermining institutional security frameworks.

- *The Evolution of Academic Integrity: Beyond Plagiarism*

The swift adoption of generative AI has sparked a core theoretical discussion on what constitutes academic authorship. For a long time, institutional policies have managed to force AI-generated text into definitions of plagiarism. The policies proposed strict citations and a distinction between machine and human contributions, as well as punitive consequences (Sehgal, 2025). Nevertheless, the latest literature indicates a serious flaw here.

According to Eaton (2025), reliance on AI detection software to police traditional plagiarism has initiated an academic integrity arms race. An empirical study has shown that these detection tools are highly unreliable and prone to false positives, and are "worse than a coin toss" at distinguishing human from AI-written text (Eaton, 2025; Jiang, Xie, & Cao, 2025). In addition, the strictest of zero-tolerance policies are associated with an increase in covert and unethical user responses, as students seek to completely sidestep the rule to manage their workloads (Ul Anam et al., 2025).

Consequently, contemporary academic thinking is embracing 'postplagiarism' (Eaton, 2025). In this view, it is becoming increasingly impossible to disconnect AI from the human. Rather, hybrid human-AI co-creation is developing into the academic norm. According to the literature, AI can be used as an "intellectual sparring partner" which may enhance creativity and critical thinking (El Hayani & Benamar, 2025) For instance, the act of compelling students to check AI 'hallucinations' and rectify mistakes enhances cognitive engagement and scepticism (Fiialka et al., 2023; Jiang, 2025). In consequence, academic integrity is evolving away from "originality" policing and towards human accountability, fact-checking, and ethical responsibility for the final product (Eaton, 2025).

IV. THEORETICAL FRAMEWORK

As a comprehensive undertaking, this study relies on a combination of established behavioral and criminological frameworks adapted for the educational context to indicate the unauthorized adopting of generative AI and resulting insider vulnerabilities.

Originally conceived to explain financial crime, the fraud triangle theory is exceptionally suited to explaining students' engagement in 'neo-plagiarism' through the agency of AI, despite potential data exposure. According to the framework, misconduct happens when three different elements come together. First, there is pressure, which refers to the strong strain of achieving high grades. Second, there is opportunity, which refers to the widespread availability of AI tools along with the confidence that institutional detection methods are ineffective. Finally, there is rationalization. This refers to the belief that the student is not impacting anybody adversely by using generative AI (Alsharefeen, 2025).

The Technology Acceptance Model (TAM) and the Theory of Planned Behavior will be employed to evaluate the adoption and potential misuse of these devices by the students. According to these theories, a student's intention to adopt Shadow AI is shaped by intrinsic motivators (such as one's attitude toward AI, perceived ethical norm) and extrinsic variables(i.e., the tool's perceived usefulness and ease of use in completing a certain academic task) (Sojan, Chow, & Peng, 2025). Students will be driven toward unofficial, more powerful alternatives as the lack of ease-of-use or usefulness of official tools means performance and quality of generation will be poor.

Street-Level Bureaucracy (SLB) framework is critical to explain the negligent insider behaviours of the educators. Today, 'street-level bureaucrats', are the faculty that possess discretionary power over policy enforcement (Alsharefeen, 2025). Many teachers develop various discretionary coping strategies because of high workloads, inadequate resources and the administrative difficulty of proving AI misuse. This frequently leads to avoiding reporting through formal institutional channels in favor of informal interventions or using unvetted Shadow AI tools to speed up grading and planning (Alsharefeen, 2025).

The analysis of systemic risk of data exposure is approached through the lens of the Lifestyle-Routine Activity Theory (L-RAT). According to this criminological theory, the creation of vulnerabilities and cyber incidents only occurs when a motivated offender, suitable target (unsecured student data), and absence of capable guardian (bypassed institutional IT oversight) converge in time and space (Hayes & Maher, 2023). In case of Shadow AI, students' habitual online activities, along with teachers over-relying on informal IT controls, immediately strips away the "capable guardian", leaving a high exposure of institutional data.

V. METHODOLOGY

In order to find out the shadow AI phenomenon in secondary schools, this study used systematic narrative review methodology. By integrating empirical studies, the existing theories, and analytical frameworks of institutional policies, we are able to arrive at a comprehensive understanding. This understanding is an unauthorized application of AI and a security risk that accompanies it.

The review synthesizes literature from a diverse range of research designs. The underlying studies predominantly relied on the quantitative approach, where researchers mainly adopt cross-sectional surveys, and Structural Equation Modelling (SEM) was used to test behavioural frameworks such as Technology Acceptance Model (TAM), and measure statistical prevalence of AI adoption (Fan et al., 2025; Ioku et al., 2024). Many studies also used mixed-methods or qualitative designs. Research employing both statistics and qualitative data can paint a coherent picture of the local setting. Semi-structured interviews, focus groups and inductive thematic coding are becoming common for capturing the complexity of students' lived experience and the challenges faced by implementers operating under vaguely defined institutional policies (Sehgal, 2025; Jiang, Xie, & Cao, 2025).

It is important to acknowledge the limitations in methodology highlighted within the literature. Social desirability bias and self-reporting is mentioned by many studies. Due to their covert, unauthorized nature, Shadow

AI users face strong incentives to hide their behaviours. Because official judgements may not reflect true usage, verifying facts is difficult (Dong et al., n.d.; Fan et al., 2025). Also, many empirical studies are based on cross-sectional datasets and purposive or convenient samples from very specific groups (e.g., undergraduate STEM students). Because of these sampling limitations and the rapid fluid changes in generative AIs, the external validity and the generalizability of the findings to broader, less-resourced educational contexts are limited (Elstad & Eriksen, n.d.; Ioku et al., 2024; Jiang, Xie, & Cao, 2025). The existing literature and scientific contributions on various healthcare support systems present a systematic bias in favour of apparently successful institutional implementations and English-language publications, while hiding unsuccessful implementations and those from other countries (Tahaei et al. 2023).

VI. DISCUSSION: THE "INSIDER THREAT" CONNECTION

There is a difference in the understanding of "insider threat" in the literature. While literature defines an insider threat as a corporate executive committing malicious acts such as tampering or theft, the understanding is different in secondary education. The definition of insider threat in school as in literature is mainly negligent.

Through two mechanisms that effectively circumvent institutional controls and embed algorithmic bias into the educational process, the use of unauthorized AI by students and educators creates a type of insider threat.

Utilization of Shadow AI may create blind spots in organizations. This tool makes institutional safeguards irrelevant and creates information asymmetry (Ojowa et al, 2026). The lack of adherence to formal IT governance and predictable security processes by a subset of people who enter sensitive educational data and/or PII into untested models is exposing their school system to data theft, rapacious third-party processing, and regulatory non-compliance (Puthal et al., 2025).

The technology lags behind since it is less capable than the average user. Younger students definitely face a tough time understanding how an open access language model collects, saves and processes user information (Authors Unknown, 2024). The desire for instant academic gratification displayed by students reveals a careless attitude towards their data. As indicated by Author Unknown (2024) and Barati, Christensen, & Zhou (2025), they are therefore particularly prone to revealing highly sensitive personal data in unmoderated chatbot settings. In this way, the student, playing the role of insider negligent, inadvertently rips, the institution's technical instrumentality, the unsupervised entry for grave data privacy breaches.

Curriculum "Data Poisoning": Algorithmic Bias Beyond data leakage, unregulated use of Shadow AI presents new risks for curriculum "data poisoning." Schools are so preoccupied with managing cheating that they overlook the broader ethical issues related to algorithmic bias (Jiang, Xie, & Cao, 2025). In general, large language models are trained by scraping a big dataset from the surface web which often includes private, copyrighted, and biased historical material not offered by the end user explicitly (Piri, 2024). According to the literature, the use of historical training data presents serious threats of reinforcing systemic biases (Frimpong, 2025).

If students rely on these unvetted tools for their research, the algorithmic bias is then embedded in their work. Moreover, if teachers use the same unauthorised models to create lesson plans or grade tests in order to manage heavy workloads as 'street-level bureaucrats', such bias is institutionalised (Alsharefeen & AlSayari, 2025). Opaque and unverified decision-making is what "poisons" the curriculum and evaluation systems. As this sort of integration is not properly monitored, the end result is a serious breakdown in trust between teachers and students. This occurs in the form of an invasion academic integrity arms race that limits surveillance to the detriment of a usually supportive learning atmosphere (Eaton, 2025).

VII. CONCLUSION AND RECOMMENDATIONS

The rapid proliferation of artificial intelligence in education has exposed a major vulnerability in the management of technology in secondary schools. This paper demonstrated that institutional responses focused on highly restrictive, reactive bans on generative AI are fundamentally ineffective. Use of "shadow AI" is occurring because of these bans instead of deterring people from its usage. In order to cope with high workload pressure, students and teachers end up being negligent insider threats by avoiding IT control. Establishment of institutional biases and personally identifiable information (PII) leakage are noted to happen due to this behaviour of data privacy at organizations. Data processing may occur by a rogue third-party too as highlighted by the study (Barati et al. 2025; Jiang, Xie, & Cao, 2025; Ojowa et al. 2025). In order to enable a safe learning environment, education institutions must move towards a proactive and comprehensive model urgently.

Based on the existing literature, this study offers targeted recommendations across governance, technology and pedagogy.

- *Recommendations for Institutional Governance.*

Actions that school leaders should take to plan for a more controlled application of ai usage.

- Institutions should move to controlled enablement through clear policies stating approved technologies and scenario based examples of acceptable use for

different disciplines (Jiang, Xie, & Cao, 2025; Silic, 2025) The provisions set forth the limitations of academic integrity and the boundary for data privacy (Ul Anam et al., 2025).

- A good educational campaign must support a governance model of Comprehensive AI Literacy (Sehgal, 2025) According to Kim (2025) and Tadimalla & Maher (2025), educational institutions must spend resources to train staff and educators in how to act on AI outputs and mitigate any algorithmic bias.
- As A.I. rapidly evolves, it is essential to have an iterative evaluation stream. Capabilities make policies outdated almost immediately. The guidelines issued by the AI policies need to be treated as living documents by the administrators essentially putting in place continuous feedback loops consisting of the faculty, students and IT staff. A good way could be to start small with some pilot projects that examine what works and what does not.

➢ *Suggestions for Technical Controls 8.2*

The bank must impose strong technical controls to deter the insider threat of data leakage. To help with this, it must provide secure alternatives to spontaneous usage of unvetted Shadow AI.

- Organizations need to employ protection architectures for their systems. Their systems must use real-time filters that can filter the inputs and sanitize the sensitive data before it reaches the core AI models. Along with this, differential privacy and machine unlearning must also be employed which can prevent the extraction of sensitive data (Barati et al., 2025).
- Schools should move away from consumer-facing AI and use privacy by design alternatives. Some possible alternatives would be the use of licensed enterprise tools (like Microsoft 365 Copilot) which has strict contractual protection of data (Author Unknown, 2024) and school chatbots for students (Elstad & Eriksen, 2024). Educational LLM tutors with transitory conversational memory and a sandboxing coding environment should be employed by educators for younger users (Barati, Christensen, & Zhou, 2025; Pankiv & Duranowski, 2025).

➢ *Advisory Suggestions for Education Practices*

To decrease the disruptive effects of AI, academic assessment must fundamentally be redesigned. Educators should get alerted to stop making knowledge-recall assessments (Sehgal, 2025; Wirzal et al., 2024). Rather than assessment tests, performance tasks, oral defenses, group projects, and process-oriented portfolios that require higher-order critical thinking skills and authentic demonstrations of competency are recommended (Sehgal, 2025; Wirzal et al., 2024). Integrating these adaptations with gamified training on cybersecurity, children would be able to reinforce data privacy hygiene responsible for creating a modern academic environment (Troublefield, 2025).

REFERENCES

[1]. Alsharefeen, R. (2025). Faculty as street-level bureaucrats: discretionary decision-making in the era of generative AI. *Frontiers in Education*, 10(1662657).

[2]. Alsharefeen, R., & AlSayari, S. (2025). Faculty as street-level bureaucrats: discretionary decision-making in the era of generative AI. *Frontiers in Education*, 10(1662657).

[3]. Author Unknown. (2024). *Data Privacy Risks and Challenges in the Context of LLMs*. [URN_NBN_fi_jyu-202409306177].

[4]. Barati, M., Christensen, M., & Zhou, C. (2025). *Children's Data Privacy in the Era of Large Language Models*. Carleton University.

[5]. Dong, M., et al. (n.d.). *Shadow AI*. Max Planck Institute for Human Development.

[6]. Eaton, S. E. (2025). Global Trends in Education: Artificial Intelligence, Postplagiarism, and Future-focused Learning for 2025 and Beyond - 2024-2025 Werklund Distinguished Research Lecture. *International Journal for Educational Integrity*, 21(12).

[7]. El Hayani, M., & Benamar, F. (2025). A conceptual framework for a symbiotic integration of generative Al in post-secondary technical and vocational education and training (TVET): Bridging pedagogy, technology, and ethics. *International Journal of Accounting, Finance, Auditing, Management & Economics*, 6(9), 319-339.

[8]. Elstad, E., & Eriksen, H. (2024). High School Teachers' Adoption of Generative AI: Antecedents of Instructional AI Utility in the Early Stages of School-Specific Chatbot Implementation. *Nordic Journal of Comparative and International Education*, 8(1).

[9]. Elstad, E., et al. (n.d.). Integrating AI in Education: An Analysis of Factors Influencing... *Human Behavior and Emerging Technologies*.

[10]. Fan, G., et al. (2025). The reconfiguration of human education in an uncertain world. *International Journal of STEM Education*, 12(16).

[11]. Fiialka, S., et al. (2023). Cited in Integrating Al in Education: An Analysis of Factors Influencing.

[12]. Frimpong, V. (2025). AI Debris: Residual Risk and the Afterlife of Failed AI Systems. *SSRN*.

[13]. Hayes, B. E., & Maher, C. A. (2023). A systematic review of lifestyle-routine activity theory in the context of direct-contact sexual violence. *Trauma, Violence, & Abuse*, 25(1), 369-392.

[14]. Ioku, et al. (2024). *Performance of Artificial Intelligence: Does artificial intelligence dream of electric sheep*.

[15]. Jiang, Y., Xie, L., & Cao, X. (2025). Exploring the Effectiveness of Institutional Policies and Regulations for Generative AI Usage in Higher Education. *Higher Education Quarterly*.

[16]. Kim, J. (2025). Perceptions and preparedness of K-12 educators in adopting generative AI. *Research in Learning Technology*, 33.

[17]. Ojowa, A., Marques, M., & Goncalves, A. (2026). Shadow AI in Organisations: A Practical Framework for Detection, Risk Classification, and Governance. *Preprints.org*.
[18]. Pankiv, O., & Duranowski, W. (2025). *Al Literacy for Skills Formation*.
[19]. Perlin, J. E. (2027). Client Confidentiality and Generative AI. *Harvard Journal of Law & Technology*, 40.
[20]. Piri, C. (2024). *Data privacy in the age of LLM-based services in education: Current challenges, improvement guidelines and future directions*. University of Jyväskylä.
[21]. Puthal, D., Mishra, A. K., Mohanty, S. P., Longo, A., & Yeun, C. Y. (2025). Shadow AI: Cyber Security Implications, Opportunities and Challenges in the Unseen Frontier. *SN Computer Science*, 6(405).
[22]. Sehgal, M. (2025). *Adapting Academic Integrity Policies to Incorporate Generative AI Tools* [Doctoral dissertation, University of Victoria].
[23]. Silic, M., Silic, D., & Kind-Trüller, K. (2025). From Shadow IT to Shadow AI-Threats, Risks and Opportunities for Organizations. *Strategic Change*, 34(2).
[24]. Sojan, A., Chow, W. W.-Y., & Peng, S. (2025). Beyond "AI-Proofing": A Mixed-Methods Evaluation of Assessment Design for Learning and Integrity. *The University of Melbourne*.
[25]. Tadimalla, S. Y., & Maher, M. L. (2025). Al literacy as a core component of Al education. *Al Magazine*.
[26]. Tahaei, M., Constantinides, M., Quercia, D., & Muller, M. (2023). *A Systematic Literature Review of Human-Centered, Ethical, and Responsible AI*.
[27]. Troublefield, T. C. (2025). Human-Centric Artificial Intelligence In Cybersecurity: Integrating Cyberpsychology for The Next Generation Defense Mechanisms. *International Journal of Security*, 16(1).
[28]. Ul Anam, M., et al. (2025). AI Tools in Education: Investigating University Students' Motivations, Institutional Policies, and Ethical Considerations in Bangladesh. *OZCHI '25*.
[29]. Wirzal, M. D. H., Md Nordin, N. A. H., Abd Halim, N. S., & Bustam, M. A. (2024). Generative Al in Science Education: A Learning Revolution or a Threat to Academic Integrity? A Bibliometric Analysis. *Jurnal Penelitian Dan Pengkajian Ilmu Pendidikan: E-Saintika*, 8(3), 319-351.