Preserving Student Privacy While Leveraging Generative AI in Higher Education

Tariq A. Alshugran
Independent Researcher
Connecticut, USA

Lina H. Kloub
School of Computing
University of Connecticut
Storrs, CT, USA

***Abstract*—The integration of Generative Artificial Intelligence (GenAI) in higher education offers significant opportunities for personalized learning and the development of dynamic educational materials. However, the use of GenAI often involves processing sensitive student data, raising concerns about privacy and regulatory compliance. This paper examines these challenges, highlighting key risks such as data breaches and unauthorized data sharing. A comprehensive solution is proposed involving privacy-preserving technologies and robust data governance frameworks. By integrating anonymization techniques and hybrid AI models, institutions can balance local data processing with cloud-based capabilities, ensuring compliance and accountability. The findings underscore the necessity of strong institutional policies to protect student privacy and foster trust in AI-driven educational innovations.**

***Index Terms*—Generative AI, Privacy, Education, Language Models, Data Governance, Student Data Protection**

I. Introduction

The adoption of Generative Artificial Intelligence (GenAI) has revolutionized various sectors, including education. AI-driven tools such as ChatGPT, DALL·E, and GitHub Copilot are now widely used for personalized learning, automated tutoring, and content generation [7]. These advancements enable educators to provide real-time feedback, tailor course materials to individual student needs, and enhance accessibility. However, as GenAI systems become more integrated into higher education, concerns about data privacy, security, and regulatory compliance are becoming increasingly critical [3].

Higher education institutions process vast amounts of sensitive student data, including academic records, behavioral analytics, and personal identifiers [8]. While this data can be leveraged to improve learning outcomes through GenAI, it also introduces significant risks. Unauthorized access, data breaches, and vulnerabilities in AI models pose threats to student privacy [3]. Research has identified privacy risks in GenAI, categorizing them into user-level risks (e.g., re-identification of anonymized data), regulatory risks (e.g., non-compliance with Family Educational Rights and Privacy Act (FERPA) and General Data Protection Regulation (GDPR)), and technological risks (e.g., adversarial attacks and model inversion techniques) [3], [5]. These risks underscore the necessity of robust data governance frameworks to ensure compliance and accountability in AI-driven educational environments.

Regulatory bodies worldwide are increasingly scrutinizing the use of GenAI in educational settings. The temporary ban on ChatGPT by the Italian Data Protection Authority in 2023 raised concerns about the lack of transparency in AI data collection practices [3]. Institutions must navigate a complex regulatory landscape, ensuring that GenAI applications align with existing legal frameworks such as the GDPR in Europe and the FERPA in the United States [6]. Despite some universities implementing AI policies, there remains a lack of standardized governance models for privacy-preserving GenAI integration in higher education [12].

Despite the potential benefits of Large Language Models (LLMs) like ChatGPT, Google's Gemini, and Microsoft's Copilot, as well as local Small Language Models (SLMs) and open-source models, many educational institutions remain hesitant to fully adopt these technologies [2]. Challenges include privacy risks, lack of transparency in AI processing, and concerns about sharing student data with third-party vendors. Moreover, while LLMs provide advanced capabilities, their implementation can lead to significant compliance and ethical dilemmas. Thus, there is a pressing need for a solution that not only leverages the strengths of these models but also addresses the inherent risks associated with their use.

To address these challenges, this paper proposes a novel solution that utilizes an SLM to manage sensitive student data within higher education institutions. By processing data locally, the SLM can scrub personal identifiers and create prompts that maintain student privacy before interacting with a generic LLM like ChatGPT. For instance, when a counselor aims to develop a student improvement plan, they input a prompt into the system, which accesses the student's data, extracts relevant information while anonymizing it, and formulates a query for the LLM. Once the response is generated, the system reintegrates necessary private information and returns a tailored response to the counselor.

This paper explores the intersection of GenAI and student privacy, identifying key challenges and presenting a comprehensive framework for privacy-preserving AI integration in higher education. By implementing privacy-preserving technologies such as the proposed SLM and robust governance policies, institutions can mitigate risks while benefiting from AI-driven education. The findings contribute to ongoing discussions about balancing technological innovation with privacy protection, ensuring that AI-driven education remains

both effective and secure.

While AI privacy policies are evolving, a clear, standardized framework for integrating privacy-preserving AI models into education remains lacking. To bridge this gap, we propose a hybrid AI approach that safeguards student data while enabling AI-driven learning in compliance with privacy regulations.

II. Background and Related Work

Generative AI (GenAI) has demonstrated significant promise in enhancing educational methodologies, improving personalized learning, and automating administrative tasks [13]. Research highlights that AI-driven adaptive learning systems can increase student engagement and performance, allowing course content to be tailored based on individual learning patterns [7]. However, the use of AI also introduces challenges regarding data privacy and security [3]. Unauthorized access, potential misuse of student data, and compliance with existing privacy regulations have emerged as critical concerns [8].

Extensive research has examined privacy risks in GenAI adoption, highlighting concerns at user, regulatory, and technological levels. Golda et al. [3] provide a comprehensive survey on privacy and security concerns in generative AI, categorizing risks into user-level, regulatory, and technological domains. Ismail et al. [5] further examine data privacy challenges and discuss encryption and anonymization techniques that could mitigate risks. Wang et al. [12] analyze university policies on AI integration, noting that despite efforts to establish guidelines, inconsistencies persist across institutions. Additionally, Olohunfunmi and Khairuddin [10] explore the ethical implications of AI-generated content, highlighting the risks of misinformation and biased outputs in educational environments.

While FERPA and GDPR impose stringent data protection requirements, their applicability to GenAI remains ambiguous, creating compliance challenges for educational institutions [6]. Many institutions struggle with compliance, particularly when relying on cloud-based AI tools for student data processing.

Several universities, including Columbia [1], Harvard [4], NIU [9], and Vanderbilt [11], have taken different approaches to managing GenAI integration in education while addressing privacy concerns. For example, some institutions have opted for **institutionally managed AI tools** rather than relying on public LLMs, reducing third-party risks. Other universities have implemented **AI disclosure policies**, requiring students and faculty to declare when AI-generated content is used in coursework, fostering transparency while maintaining data accountability.

Recent initiatives highlight attempts to create **privacy-preserving AI frameworks** tailored to education. For instance, projects focusing on **federated learning** aim to decentralize AI model training, keeping sensitive student data on local devices rather than sending it to external servers [5]. Additionally, open-source AI initiatives are gaining traction as institutions seek customizable and **self-hosted AI models** to mitigate vendor lock-in and proprietary data concerns.

A comparative analysis of university AI policies in higher education suggests that while AI-driven education is advancing rapidly, there remains a gap in **standardized privacy governance**. Universities differ in their approaches to GenAI regulation, from **strict AI bans in assessments** to **AI-integrated curricula with clear ethical guidelines**. Table I summarizes different institutional policies regarding AI in education and privacy protection.

tabularx

TABLE I
Comparison of AI Policies in Higher Education

University	**AI Usage Policy**	**Privacy Protection Strategy**
Columbia University [1]	AI Governance Policy	Ethical Use and Institutional Oversight
Harvard University [4]	AI Syllabus Guidelines	Instructor-Led AI Integration Framework
Northern Illinois University [9]	Prohibition of AI in Assignments	Academic Integrity Policy Enforcement
Vanderbilt University [11]	Instructor Discretion on AI Usage	Course-Specific AI Adaptation

The growing reliance on AI in education underscores the importance of **proactive regulatory frameworks** that balance innovation and privacy. Without clear policies, institutions risk exposing student data to **unregulated AI vendors**. The next sections of this paper propose a structured methodology for ensuring privacy compliance while leveraging AI to enhance education.

III. Current Approaches and Their Limitations

Large Language Models (LLMs): LLMs such as ChatGPT and Gemini offer powerful AI-driven solutions for text generation, automated tutoring, and educational support. However, they primarily rely on cloud-based infrastructure, leading to privacy risks related to data storage and potential unauthorized access.

Small Language Models (SLMs): SLMs process data locally, reducing reliance on external servers and mitigating data privacy concerns. However, these models may lack the extensive training and diverse knowledge base that LLMs provide, which could limit their effectiveness in complex educational tasks.

Open-Source Models: Open-source AI models provide institutions with greater transparency and customization opportunities. Universities can deploy these models within controlled environments, ensuring compliance with regulatory policies. However, maintaining and securing open-source implementations requires significant technical expertise and resources, which may pose adoption challenges. Table II provides a comparative analysis of these approaches, highlighting their advantages and limitations in the context of AI-driven education.

TABLE II
COMPARISON OF AI MODELS FOR EDUCATIONAL APPLICATIONS

AI Model	Advantages	Limitations
Large Language Models (LLMs)	• High accuracy in text generation • Can process complex queries • Continuously updated by providers	• Requires cloud-based processing, raising privacy concerns • High computational cost • Black-box nature limits explainability
Small Language Models (SLMs)	• Local processing enhances data privacy • Lower latency • Customizable for institutional needs	• Less powerful than LLMs • Requires on-premise computational resources • Limited external knowledge
Open-Source AI Models	• Full transparency • Customizable and modifiable • No dependency on third-party vendors	• Requires significant technical expertise for deployment • Security risks if not properly maintained • May not match LLM performance

IV. METHODOLOGY

A. Overview

This section outlines the proposed privacy-preserving framework for leveraging GenAI in higher education. The methodology integrates privacy-preserving techniques with a hybrid AI model that processes sensitive student data locally before engaging with external AI models. This approach ensures compliance with data protection regulations while maximizing the benefits of AI-driven educational tools.

B. Data Anonymization and Privacy-Preserving Techniques

To mitigate privacy risks, the framework employs a combination of k-anonymity, differential privacy, and encryption. The anonymization process follows these key steps:

1) **Named Entity Recognition (NER)**: Extract personally identifiable information (PII) from student inputs.
2) **Pseudonymization and Masking**: Replace sensitive identifiers with pseudonyms or generalizations.
3) **K-Anonymity**: Ensure that each anonymized record is indistinguishable from at least k-1 other records.
4) **Differential Privacy**: Introduce Laplace noise to numerical or categorical data, reducing the risk of re-identification.
5) **Encryption**: Apply homomorphic encryption to safeguard sensitive data even when transmitted externally.

The workflow for this anonymization process is depicted in Fig. 1.

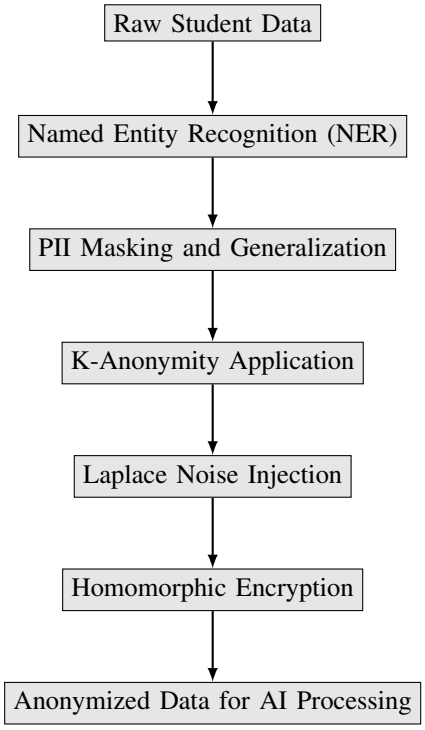

Fig. 1. Flowchart of the data anonymization process.

C. Security Measures for Data Transmission to LLMs

To protect sensitive student data during LLM interactions, we employ the following multi-layered security measures:

- **Zero-Knowledge Proofs (ZKP)**: Ensures that institutions can verify anonymization integrity without revealing the actual data.
- **Federated Learning**: Keeps training data localized while allowing aggregated model improvements.
- **Secure Multi-Party Computation (SMPC)**: Enables encrypted AI interactions without exposing private student data.
- **Access Control Policies**: Role-based access ensures that only authorized personnel can interact with sensitive datasets.

These security measures provide multiple layers of protection, ensuring that sensitive student information does not get exposed to unauthorized third parties.

D. Latency vs. Accuracy Trade-offs in SLM and LLM Interactions

The hybrid AI model balances computational efficiency and accuracy when utilizing both SLMs and LLMs. The trade-offs between the two are summarized in Table III.

TABLE III
LATENCY VS. ACCURACY TRADE-OFFS BETWEEN SLM AND LLM

Model Type	Latency	Accuracy
SLM	Low latency (local processing)	Moderate accuracy due to limited dataset
LLM	Higher latency (cloud-dependent)	High accuracy but potential privacy risks

E. Hybrid AI Model Implementation

The framework employs a **hybrid approach**, where an SLM pre-processes student queries locally and generates anonymized prompts before forwarding them to an LLM. This approach reduces latency while maintaining acceptable accuracy levels.

F. System Workflow and Data Flow Diagram

The workflow involves: 1. Collecting student input data and preprocessing it via the SLM. 2. Anonymizing and structuring the input to remove identifiable elements. 3. Sending anonymized prompts to an external LLM. 4. Reintegrating necessary information into the AI-generated response before presenting it to the user.

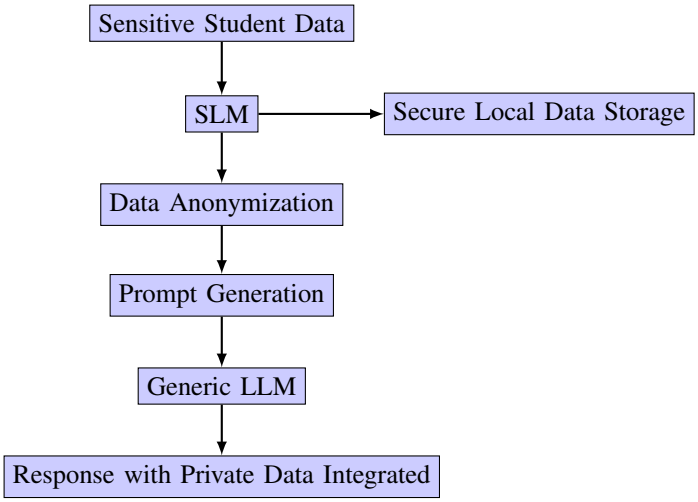

Fig. 2. Data flow diagram for the proposed methodology.

V. DISCUSSION

A. Effectiveness of Privacy-Preserving Models

The proposed SLM-LLM hybrid framework effectively mitigates data exposure risks by ensuring that personal identifiers are removed before AI processing. By leveraging k-anonymity and differential privacy, student records remain protected against adversarial re-identification attacks.

B. Comparison With Existing Approaches

Unlike traditional AI deployments that rely on fully centralized LLMs, our model offers a privacy-first architecture by balancing local processing (SLM) with LLM scalability. Compared to open-source alternatives, this approach maintains higher adaptability while ensuring compliance with AI governance policies.

C. Challenges and Limitations

Despite its advantages, the hybrid model introduces trade-offs in processing speed, accuracy, and institutional costs. Local processing requires computational resources, and integrating federated learning mechanisms across institutions may be technically complex. Future research should explore scalable, cost-effective deployment strategies tailored to diverse educational environments.

VI. FUTURE WORK AND RECOMMENDATIONS

Future research should focus on enhancing the robustness of privacy-preserving AI models and evaluating their real-world impact in educational settings. Expanding the dataset diversity and refining the integration of SLM and LLM architectures can further optimize performance. Additionally, collaboration with policymakers can help establish clearer guidelines for AI-driven educational applications while ensuring compliance with evolving data protection laws.

VII. CONCLUSION

This paper has explored the challenges of integrating GenAI in higher education while maintaining student privacy. A hybrid AI model leveraging SLMs and LLMs was proposed to address data privacy concerns. This methodology protects student information while allowing institutions to use AI-driven tools effectively. Advancing AI-driven education requires continuous refinement of privacy-preserving models while ensuring robust data security and ethical integrity.

REFERENCES

[1] Columbia University. (2024) Generative ai policy. [Online]. Available: https://provost.columbia.edu/content/office-senior-vice-provost/ai-policy
[2] A. Garry, "Creating ieps with genai while ensuring data privacy," *eSchool News*, September 2024. [Online]. Available: https://www.eschoolnews.com/it-leadership/2024/09/25/creating-ieps-with-genai-while-ensuring-data-privacy
[3] A. Golda, K. Mekonen, A. Pandey, A. Singh, V. Hassija, V. Chamola, and B. Sikdar, "Privacy and security concerns in generative ai: A comprehensive survey," *IEEE Access*, vol. 12, pp. 12 345–12 367, 2024.
[4] Harvard University. (2024) Ai guidance for instructors and students. [Online]. Available: https://oue.fas.harvard.edu/ai-guidance
[5] I. A. Ismail, "Protecting privacy in ai-enhanced education: A comprehensive examination of data privacy concerns and solutions in ai-based learning," in *Impacts of Generative AI on the Future of Research and Education*, A. Mutawa, Ed. IGI Global, 2025, pp. 117–142.
[6] I. A. Ismail and J. Aloshi, "Data privacy in AI-driven education: An in-depth exploration into the Synergy of AI and Data Privacy in Education," in *Encyclopedia of Information Science and Technology, Sixth Edition*. IGI Global, 2024.
[7] L. Kloub and A. Gupta, "Chatgpt in computer science education: Exploring benefits, challenges, and ethical considerations," in *Proceedings of the 2024 ASEE North East Section*. Fairfield, Connecticut: ASEE Conferences, April 2024.

[8] P. K. Konakalla and G. Simuni, "Security and privacy concerns in generative ai," *SSRN Electronic Journal*, 2024.
[9] Northern Illinois University. (2024) Class policies for ai tools. [Online]. Available: https://www.niu.edu/citl/resources/guides/class-policies-for-ai-tools.shtml
[10] I. A. Olohunfunmi and A. Z. Khairuddin, "Exploring ethical dilemmas of ai generative tools among higher education students: A systematic review," in *Proceedings of the International Conference on Innovation & Entrepreneurship in Computing, Engineering & Science Education (InvENT 2024)*. Atlantis Press, 2024, pp. 255–275.
[11] Vanderbilt University. (2024) Ai and syllabus policies. [Online]. Available: https://as.vanderbilt.edu/gci-ai/syllabus-ai-policies/
[12] H. Wang, A. Dang, Z. Wu, and S. Mac, "Generative ai in higher education: Seeing chatgpt through universities' policies, resources, and guidelines," *Computers and Education: Artificial Intelligence*, vol. 7, p. 100326, 2024.
[13] J. Zhang, "Generative ai in higher education: Challenges and opportunities for course learning," *Advances in Social Sciences Research Journal*, vol. 12, no. 1, pp. 11–18, 2025.