Understanding How University Guidelines Address Privacy and Security Issues of Generative AI in Academic Settings

Bei Yi Ng, Jiarui Li, Xinyuan Tong
University of Edinburgh

Kevin Ye, Gauthami Yenne, Varun Chandrasekaran
University of Illinois Urbana–Champaign

Jingjie Li
University of Edinburgh

Abstract

Generative artificial intelligence (GenAI) is transforming the educational landscape by augmenting learning paradigms. However, state-of-the-art GenAI systems driving this transformation are predominantly developed and controlled by a small number of private companies; there is little clarity about their data retention practices and limited user control over inputs and outputs. In the context of education, end-users lack the awareness of how to safely adopt GenAI in learning. This raises significant concerns, particularly when proprietary or personally identifiable educational information may be shared with external GenAI platforms. In response to these concerns, universities are developing their own usage guidelines and policies to balance innovation with academic integrity, privacy, and security. Our research seeks to understand these emerging guidelines, with a particular focus on the privacy and security implications of integrating GenAI tools into academic environments – an area that has received little attention to date. We conducted an in-depth qualitative analysis of GenAI-usage guidelines from 43 universities across 12 countries. Our findings reveal several key challenges, including barriers faced by universities in deploying privacy measures and adopting existing security frameworks. These insights lay the groundwork for designing more robust, privacy-aware GenAI guidelines for higher education.

1 Introduction

Generative AI (GenAI) is rapidly transforming higher education; students and educators use it for writing, programming, research, and data analysis [53]. Its versatility in addressing open-ended queries has made it a core tool in academic workflows [47]. As adoption grows, traditional educational technology providers are embedding GenAI into services like writing support and virtual collaboration [18]. In parallel, extensive research is underway to better understand AI-human collaboration in education—ranging from the design of intuitive interfaces to investigations of GenAI's pedagogical impact to the broader shifts it introduces in teaching and learning paradigms [9, 24, 79].

GenAI systems continually improve by leveraging vast datasets [63], including (potential) user interactions [74]. In academic settings, GenAI is integrated into tools like AI-enhanced learning management systems and adaptive tutoring applications. However, GenAI providers' (e.g., OpenAI) data practices often lack transparency, raising serious concerns [45]. For example, the collapse of AllHere Education—the company behind Los Angeles Unified School District's AI chatbot 'Ed'—sparked questions about the fate of sensitive student data it collected [65]. Such incidents highlight the need for stronger oversight in handling educational data. In the U.S., the Family Educational Rights and Privacy Act (FERPA) [99] regulates how student data is accessed and shared, while similar protections exist globally: the EU's General Data Protection Regulation (GDPR) [38] enforces strict data minimization and consent requirements, and Australia's Privacy Act 1988 [72] governs the handling of personal information in schools. These regulatory frameworks compel institutions to implement robust safeguards when adopting GenAI.

While the benefits of GenAI are undeniable—enhancing productivity, creativity, and efficiency—its adoption is accelerating within a "lemon market," where users lack clear visibility into model quality, risks, and trade-offs [16]. Students and educators often rely on GenAI tools without a full understanding of their limitations, such as susceptibility to misinformation [48], bias [54], and over-reliance [7]. As policy lags behind technological advancement, universities have a vital role to play in shaping responsible GenAI adoption through clear institutional guidelines, safeguards, and pedagogical frameworks [16]. Apart from principles of academic integrity, recent studies identify *privacy and security* as key motivators in universities' and instructors' policies about GenAI usage [4, 51, 52, 67]. However, an in-depth understanding of how universities approach and assess these privacy and security risks of GenAI, and in particular how they communicate the trade-offs between potential mitigation strategies and benefits, is still missing.

Our study aims to bridge this gap. By analyzing GenAI usage guideline documents published online by universities globally, we aim to address the following three research questions:

RQ1. How do university guidelines conceptualize and communicate privacy and security risks of GenAI, such as sensitive data breach and technical vulnerabilities?

RQ2. How do universities evaluate and suggest different measures to mitigate the aforementioned risks?

RQ3. How are privacy and security positioned among other guiding principles in policy design?

We conducted qualitative content analysis for university-level GenAI guidelines from 43 universities spanning 12 countries, focusing on how privacy and security considerations integrate with core academic usage and values of GenAI. Our study reveals in-depth findings corresponding to the privacy and security issues of GenAI in academic settings. We summarize the main findings corresponding to our research questions below:

F1. Universities broadly recognize GenAI-related privacy and security risks, especially those involving sensitive student, research,

This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license visit https://creativecommons.org/licenses/by/4.0/ or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.
Proceedings on Privacy Enhancing Technologies 2026(4), 1055–1071
© 2026 Copyright held by the owner/author(s).
https://doi.org/10.56553/popets-2026-0159

and institutional data. Their guidelines present partial and evolving threat models, reflecting university priorities: compared with privacy risks from intentional and unintentional misuse by students, staff, and third parties, concrete adversarial behaviors considered as security violations and technical attack surfaces in academic settings are under-specified.

F2. Universities mitigate GenAI risks through a combination of technical safeguards, institutional operations, and individual responsibilities, but these measures vary in clarity and consistency. Teaching and learning-focused guidelines often frame privacy and security through academic values and classroom practices, while IT-focused guidelines translate these concerns into more concrete risk assessment, deployment, and monitoring practices. Despite growing reliance on licensed tools and existing security infrastructure, universities still face challenges in adapting safeguards for the expanding ecosystem of GenAI-enabled services.

F3. Privacy and security governance is shaped by academic values, institutional policies, and external regulatory or inter-university frameworks, showing evidence of active knowledge transfer. Universities often position privacy and security alongside values such as academic integrity, fairness, and responsible learning, although this values-based framing can leave implementation open to interpretation. Cross-regional and international regulatory influences also shape guideline development, yet the integration of privacy and security with academic values remains uneven across institutions.

Based on our findings, we propose several recommendations to improve privacy and security for universities, educators, and students adopting GenAI. These include improving privacy and security assessment frameworks that consider GenAI-specific threats and risks, tailoring technical safeguards to academic contexts, as well as reducing infrastructure and communication disparities for safeguarding GenAI across universities and academic disciplines.

2 Related Work

GenAI in education. GenAI has rapidly gained traction across the higher education sector, permeating both student and faculty practices [16, 35, 64, 96, 112]. Students are leveraging GenAI tools for drafting essays, solving mathematical problems, summarizing readings, generating code, translation, and exam preparation [40]. Simultaneously, educational professionals are incorporating GenAI into teaching workflows—using it to develop lecture materials, generate quiz banks, provide formative feedback, assist in grading, and personalize instruction through intelligent tutoring systems and adaptive learning platforms [78, 97].

Despite this growing adoption, the understanding of GenAI technologies within educational contexts remains fragmented. Many users—students and instructors alike—engage with GenAI tools without a clear grasp of their operational limitations or risks in education. One major concern is the generation of false or misleading content (i.e., hallucinations), which can misinform students or degrade the quality of instruction when integrated into teaching materials [32]. This is particularly problematic in activities requiring factual accuracy where hallucinated outputs can introduce conceptual errors or propagate misinformation that is plausible-sounding [48, 109]. Another concern is in coding and programming contexts, where GenAI tools may produce code that compiles successfully but fails at runtime, silently introducing logic bugs. In other cases, the code might make use of deprecated libraries, incorrect syntax for a given programming language, or fabricate API calls and functions that do not exist [61].

Instructors may also rely on AI-generated content for assessments or feedback, leading to inconsistent grading, overlooked bias, or inadequate pedagogical alignment [106]. These risks are amplified in educational settings that inherently demand pluralistic alignment [5]—where diverse pedagogical philosophies, disciplinary norms, institutional missions, and cultural values coexist. For instance, the standards for what constitutes a "good essay" or "effective code" can vary significantly across departments, instructors, and even individual course objectives.

Finally, there is also the risk of over-reliance, where students defer to GenAI outputs without developing critical thinking or research skills [7]. Beyond individual use, institutional integration of GenAI into learning management systems or other tools opens the door to system-level vulnerabilities [42]. These risks are exacerbated by the lack of standardized training and clear communication about safe and ethical usage at the institutional level [58, 89, 114], as well as a dearth of robust yet adaptable AI governance frameworks that can be readily applied in the educational context [13, 14, 36].

Amidst this landscape, universities are beginning to issue formal guidelines to govern GenAI usage. These institutional policies are becoming essential, not just for setting standards of acceptable use, but for mediating stakeholder differences and fostering a shared understanding of GenAI's role in academic life.

Privacy and security risks of GenAI. Prior research reveals critical privacy and security concerns stemming from emerging GenAI services. The development of GenAI models leads to significant data privacy concerns [50, 71], particularly regarding how training data is collected [110]. Many GenAI systems are trained on data scraped from the internet without consent, including proprietary or personal materials. In some cases, models can even memorize and regurgitate specific personal or sensitive details from training data, posing risks of inadvertent disclosure of private information [25, 104]. Furthermore, prior work shows that GenAI introduces exacerbated privacy risks and harms, including inferring/exposing sensitive information and encouraging excessive data collection [60]. GenAI's data privacy issues are prominent in academic settings too: when students input assignments, questions, or personal reflections into third-party GenAI tools, they may inadvertently expose sensitive academic information—raising potential violations of privacy laws [75, 101].

GenAI also introduces unique security challenges. GenAI models are susceptible to jailbreaking, where users manipulate prompts to bypass safety filters and elicit offensive, inappropriate, or dangerous responses [20]. Furthermore, when GenAI is integrated into tools such as chatbots, it becomes vulnerable to prompt injection attacks [62]. These attacks involve embedding hidden instructions in input text that can force the model to behave in unintended ways—such as leaking private data, overriding guardrails, generating software with vulnerabilities that are difficult to detect, or even acting on behalf of a user without authorization. Furthermore,

GenAI is vulnerable to "poisoning", since these models require data-intensive training and frequent updates. During model training and update, if a model ingests skewed, offensive, or malicious content injected by a malicious party, it can reproduce those biases in output, leading to discriminatory or harmful responses [80].

Privacy and security concerns of educational technologies. The digitization of education has led to the widespread adoption of educational technologies. From learning management systems (LMS), such as Canvas and Blackboard, to intelligent tutors, to online proctoring tools used in remote assessments. These technologies offer clear benefits in terms of accessibility, scalability, and efficiency [116]. However, they introduce concerns with respect to security and data privacy, especially in proctored assessment environments and those involving minors [17, 90, 101]. There exists a power imbalance between educational institutions and educational technology companies, which often leads institutions to adopt these technologies despite privacy and security concerns and limited resources to investigate them [55, 81]. Subsequently, students and educators often have little input and control on what happens to the personal information being collected and stored [6, 30, 59, 89, 101]. It is common for proctoring tools, for example, to surveil students using key logging, mouse tracking, and microphone and eye tracking [12, 90, 95]. Similarly, LMS platforms collect data on student behavior and store sensitive and legally protected information such as grades and personal information [59]. In K-12 classrooms, these risks are exacerbated [57, 117]. For example, in the United States, children under the age of 13 are subject to special legal protections under the Children's Online Privacy Protection Act (COPPA) and Family Educational Rights and Privacy Act (FERPA) [34, 99]. Despite these protections and being less equipped to understand and deal with the consequences of data collection, LMS treat these students the same as others [17, 57, 66, 118]. These concerns, compounded by the rapid and expanding adoption of EdTech platforms without thorough evaluation of privacy and security implications, underscore the necessity for scrutiny.

Research gaps in prior work. GenAI is drawing increasing attention in education research. Despite recent studies which identify GenAI privacy and security among other key considerations such as learning enhancement and ethics in the academic usage of GenAI [4, 51, 52, 67], they focus on broader pedagogical analyses and consider privacy and security as secondary. Our work contributes novel findings through an in-depth analysis into the technical, operational, and infrastructural challenges to addressing GenAI privacy and security issues through the study of universities' GenAI guidelines – a gateway to understanding universities' position and implementation of privacy and security measures. Our approach anchors on how privacy and security concerns are institutionally perceived and systematically managed. This complements existing work that has flagged specific stakeholder concerns, such as that of students and teachers, towards GenAI tools [89, 101], as well as emergent literature on how these stakeholders are affected by the ever-changing and vague legislation on implementation [13, 37]. Our study is also distinct from those that study privacy and security issues of other education technologies [17, 57, 90], as the subject of our study is GenAI, which introduces unique privacy and security risks and challenges in policy enforcement due to its wide and flexible application contexts in education.

3 Methodology

To answer our research questions in § 1, we conducted a qualitative analysis of university-level GenAI policy documents (or guidelines). Our approach allows us to characterize the early governance landscape, identify emerging privacy and security measures and risk framings. We then examine how institutions are developing guidance to manage vulnerabilities associated with GenAI adoption. We employed a structured qualitative coding and analysis process, enabling us to surface cross-institutional patterns and evolving requirements in privacy and security-aware GenAI integration from a globally representative sample of university guidelines.

Data collection and sampling. Our data collection focused on publicly accessible GenAI guideline documents published by universities globally. We conducted this effort between June and December 2024, and we selected universities from the "2024 QS World University Rankings" [107], which has been well-established in the methodology of cross-comparative policy studies [8, 103] for its global recognition and comprehensive evaluation of teaching, research, and international outlook.

Our initial exploration, during which we explored availability of GenAI policies in universities by region, discovered that the institutions recognized in the top universities globally also tend to have more accessible and comprehensive GenAI policy documents, which became our focus, aligning with prior findings that systematic adoption of GenAI remains experimental across much of higher education [51]. Though starting from the highest ranked institutions, our sampling aimed to achieve institution and region diversity within available data. In this study, we focused on and searched for *university-level* guidelines available on university websites, rather than department-specific documents, as these university-level guidelines represent the *overarching* principles of one university, contextualizing privacy and security concerns within educational practices, that directs specific adoption and implementation at department levels. We manually traversed university websites, including subdomains on Academic Integrity subsuming GenAI guidelines. To assist our search, we constructed a set of search queries to obtain additional results from the search engine, incorporating a combination of representative terms with university names, including *"generative AI," "AI," "guidance," "guideline," "policy," "teaching"* and *"learning,"* that reflect the naming conventions used by institutions to ensure accessibility of their policies. Our search logic includes the above terms: `("University name") & ("Generative AI" | "AI") & ("Policy" | "Guidance" | "Guideline") & ("Learning" | "Teaching" | "")`.

In addition, our scope of sampling and analysis is limited to GenAI guidelines that are publicly available in English. We developed the codebook in stages using universities ranked in the top 100 of the QS World University Rankings, supplemented by two highly ranked regional universities, UCT and KAUST. To reduce overrepresentation of US and UK institutions, we randomized the university order and coded them in batches, refining the codebook after each batch. We stopped sampling once our qualitative coding

Continent	Country/Region	University	Document Title
Africa	South Africa	University of Cape Town (UCT)	EIRC Generative AI Guidelines
Asia	Hong Kong, PRC	University of Hong Kong (HKU)	AI Literacy
Asia	Japan	University of Tokyo	Use of AI Tools in Classes
Asia	Japan	Tokyo Institute of Technology (Tokyo Tech)	Policy on Use of GenAI in Learning
Asia	Saudi Arabia	King Abdullah University of Science and Technology (KAUST)	AI Guidelines on Research
Asia	Singapore	Nanyang Technological University (NTU)	Teaching Learning Assessment with GenAI
Asia	Singapore	Nanyang Technological University (NTU)	Use of GenAI in Research
Asia	Singapore	National University of Singapore (NUS)	NUS Academic Integrity
Europe	Germany	Technical University of Munich (TUM)	Usage Guidelines for GenAI @ TUM
Europe	Netherlands	Delft University of Technology (TU Delft)	AI Chatbots in Unsupervised Assessment
Europe	Netherlands	University of Amsterdam (UvA)	UvA Policy on AI
Europe	Switzerland	École polytechnique fédérale de Lausanne (EPFL)	Tips for the Use of GenAI
Europe	Switzerland	Eidgenössische Technische Hochschule Zürich (ETH Zurich)	GenAI in Teaching & Learning
Europe	United Kingdom	Imperial College London (IC)	GenAI Guidance
Europe	United Kingdom	King's College London (KCL)	AI Guidance on Learning and Teaching
Europe	United Kingdom	London School of Economics and Political Science (LSE)	AI Education and Assessment
Europe	United Kingdom	University of Edinburgh (UoE)	AI Guidance for Students
Europe	United Kingdom	University of Manchester (UoM)	Teaching and Learning FAQ
Europe	United Kingdom	University College London (UCL)	Using GenAI in learning and teaching
Europe	United Kingdom	University of Bristol	University Guidance on GenAI in Education
Europe	United Kingdom	University of Cambridge	Using Generative AI
Europe	United Kingdom	University of Oxford	Guidelines on the Use of GenAI
North America	Canada	McGill University	General Guidelines for Using GenAI Tools
North America	Canada	University of British Columbia (UBC)	GenAI tools in teaching and learning
North America	Canada	University of Toronto (UofT)	GenAI in the Classroom: FAQ's
North America	United States	California Institute of Technology (Caltech)	Resources for Teaching in the Age of AI
North America	United States	Carnegie Mellon University (CMU)	Generative Artificial Intelligence
North America	United States	Columbia University	GenAI Policy
North America	United States	Cornell University	GenAI for Education and Pedagogy
North America	United States	Duke University	AI Policies: Guidelines and Considerations
North America	United States	Harvard University	Generative AI Guidelines
North America	United States	Massachusetts Institute of Technology (MIT)	Guidance for Use of GenAI Tools
North America	United States	Northwestern University (NU)	Northwestern Guidance on the Use of GenAI
North America	United States	University of California, Berkeley (UC Berkeley)	Appropriate Use of GenAI Tools
North America	United States	University of California, Los Angeles (UCLA)	Guiding Principles for Responsible Use
North America	United States	University of Chicago	GenAI Guidance
North America	United States	University of Illinois Urbana-Champaign (UIUC)	Univ. of Illinois System GenAI Principles
North America	United States	University of Michigan-Ann Arbor (UM-Ann Arbor)	U-M Guidance for Students
North America	United States	University of Pennsylvania (UPenn)	AI Guidance
Oceania	Australia	Monash University	Acceptable and Responsible Use of AI Tech.
Oceania	Australia	Australian National University (ANU)	AI including GenAI
Oceania	Australia	University of Melbourne	University of Melbourne AI Principles
Oceania	Australia	University of New South Wales (UNSW)	ChatGPT & GenAI at UNSW
Oceania	Australia	University of Queensland (UQ)	AI Teacher Hub

Table 1. The list of official GenAI documents we analyzed. These documents are published online by universities worldwide.

and analysis reached data saturation [83]; the coding process is described soon after.

Our sampling process stopped after code saturation, which we monitored along our analysis, at 43 universities; some universities in the QS ranking are not analyzed due to lack of publicly available guidelines. We show the list of universities and guidelines analyzed in Table 1. We only include documents whose overall topic is GenAI guidance in contexts of teaching and learning, excluding those where GenAI has partial focus, e.g., President Remark of Kyoto University [1].

Following the data collection and analysis of university-level guidelines focusing on academic integration, particularly teaching and learning, we collected additional GenAI guidelines provided by the university IT departments, including information security, for further analysis and cross-comparison. This additional data source allows us to examine if and how universities provide more dedicated technical resources and support to govern the security & privacy (S&P) risks. We adopted the same web search and website traversal approach to collect these documents. Similarly, the search logic to assist our manual website traversal for this task includes these terms (`"University name") & ("Generative AI" | "AI") & ("Policy" | "Guidance" | "Guideline") & ("Information security" | "Acceptable use" | "Approved tools"`). The above terms are chosen as they correspond to the core concept of privacy and security related to GenAI. We initially experimented with more search terms that include *"privacy policy"*, *"data protection"*, and *"student data privacy."* However, such terms inconsistently returned website privacy policies, cookie notices, and general student privacy notices rather than governance documents specific to GenAI issues, and were therefore not included eventually. As a result, we included additional IT-specific policies for 19 universities to cross-compare with their university-level, teaching and learning-focused guidelines, shown in Table 2 in Appendix A.1.

Data preparation. Universities present their guidelines as web pages. To code and analyze these guideline documents efficiently and consistently, we first reformatted the page content using an `HTML` sanitizer [2]. We preserved the section headers and paragraphs that are relevant to the guideline content. This approach allowed us to segment guideline documents into coherent paragraphs and sections that are suitable for coding and thematic analysis. We manually verified and curated the segmentation when the sections are not well wrapped by `HTML` headers. Our analysis focused on text content, which presents the most meaningful information in a guideline webpage. Thus, we did not include visual aids.

Qualitative coding and analysis. Our analysis started with open coding [22]. Three researchers independently coded guideline documents from 15 universities, looking for guideline content and

concepts that are relevant to our research questions. We first comprehensively captured fine-grained concepts, allowing codes to develop with the initial codebook. Following this process, all research team members engaged in a series of structured discussions to review, compare, and refine the codes generated individually. This contributed to our initial codebook. We further leveraged guidelines from three universities to improve the initial codebook and resolve discrepancies, e.g., when guideline framings were ambiguous, we resolved them by refining definitions and merging overlapping codes. Two researchers later applied this codebook to independently code 22 more guidelines. We used the 22 coded guidelines here to assess the reliability of our codebook. We computed Cohen's kappa (κ), a standard metric of inter-annotator agreement for these guidelines coded by two coders. Our coding attained $\kappa = 0.65$, which indicates a substantial agreement [68]. Then we continued coding and confirmed saturation with 46 guideline documents in total. The lead coder also applied the codebook developed to code and analyze the additional IT-focused GenAI guidelines, further confirming the coverage of the codebook.

Throughout our study, our research team engaged in regular meetings to resolve disagreement in coding and perform thematic analysis. We applied codebook thematic analysis [11, 22], which lies between codebook reliability and reflexive thematic analysis. As such, after reaching code saturation, our team continued analysis to develop themes by identifying patterns from coded data. Note that the codebook served as an analytic framework to underpin later thematic analysis, and κ is treated as a pragmatic quality check rather than a claim of full analytic validity. Our research teams are formed with members with diverse expertise, including computer security and privacy, education, human computer interaction, and artificial intelligence. Our expertise enables our analysis to capture codes and themes from comprehensive and complementary perspectives that are essential to our research questions. § 5 discusses categories of codes and themes discovered. We make our codebook available through our project repository with access to documents. [1]

4 Study Limitations

We recognize several limitations in our study. Our sampling uses the QS World University Rankings, following prior research [52]. However, we acknowledge that the distribution of institutions across the world is skewed due to regional disparities in educational development and internationalization. The top-ranked universities happen to be concentrated in North America, Europe, East Asia, and Australia. Despite having considered regional diversity in sampling and actively looked for available content in underrepresented regions including Africa and the Middle East, we ended up including more universities where English is a major working language. This is also because we focused on and searched for GenAI guidelines written in English. This also ultimately resulted in exclusion of universities from South America. Furthermore, our sample presents perspectives of globally ranked and research-intensive universities, also early GenAI adopters, rather than the whole sector of higher education, which therefore may not be representative for less-resourced institutions. We also acknowledge that the GenAI guidelines we study present the public, overarching position of a university; the actual implementation may differ, and the policies are subject to evolve. Recognizing the fast-evolving landscape of GenAI development and integration, we aim to provide in-depth qualitative results, rather than quantitative and statistical results to identify challenges faced by the universities in safeguarding GenAI usage. Consistent with prior publications [52, 67, 100], we opt to conduct an in-depth qualitative analysis of university-level guidelines, as it provides an authentic view of the AI governance landscape that reflects links to state-level and national frameworks, which are limited in course/department-level rules. Nevertheless, this approach is limited in uncovering specific privacy and security challenges in curriculum design when adopting GenAI.

We argue that global influences of sampled universities on education and GenAI adoption may inform less-resourced institutions, and some challenges, e.g., technical and operational barriers, faced by these universities may be amplified for low-resourced institutions. Nevertheless, we encourage future work to investigate the challenges faced by less-resourced institutions in the real world and to explore other data sources and methods that offer complementary insights, e.g., from individual perspectives of students and staff, while also conducting longitudinal measurements and comparisons of GenAI guideline updates beyond our study's cut-off date, given the dynamic and ever-evolving nature of policy development. Furthermore, large-scale studies can consider GenAI guidelines available in other languages beyond English, as well as the longitudinal factors in institutional policy development.

5 Findings

Figure 1 provides a holistic overview of the observed themes from our analysis, as well as how they answer each research question. These themes highlight the privacy and security challenges from different perspectives in the university integration of GenAI services, and how stakeholders interact with them.

Roadmap. § 5.1 discusses how guidelines conceptualize GenAI privacy and security risks (**RQ1**), including sensitive data, technical threats, and connections with broader academic harms. Recognizing the privacy and security risks, § 5.2 elaborates on how universities assess different privacy and security measures (**RQ2**) from institutional levels to individual levels, specifically challenges in these measures. Finally, § 5.3 explains how privacy and security, as guiding principles, are considered in universities' policy and guideline development (**RQ3**), along with overarching regulatory frameworks and the tensions with broader academic values.

5.1 Conceptualizing Privacy and Security Risks

Our qualitative coding discovered three high-level categories of university guidelines' conceptualization of privacy and security risks of GenAI:

1. **Privacy-sensitive data practices**, which include varying personal or sensitive data types, data handling practices (data collection, third-party sharing, retention, unauthorized sharing of training data etc.), and breaches.
2. **Exploit and misuse** of GenAI services, including harmful and inaccurate content (misinformation and fraudulent content), vulnerabilities and bugs (e.g., jailbreaking), and students' overreliance on GenAI.

[1] https://osf.io/3ezfw/overview?view_only=3f8cad97c179473eb0ccbfa98595a350

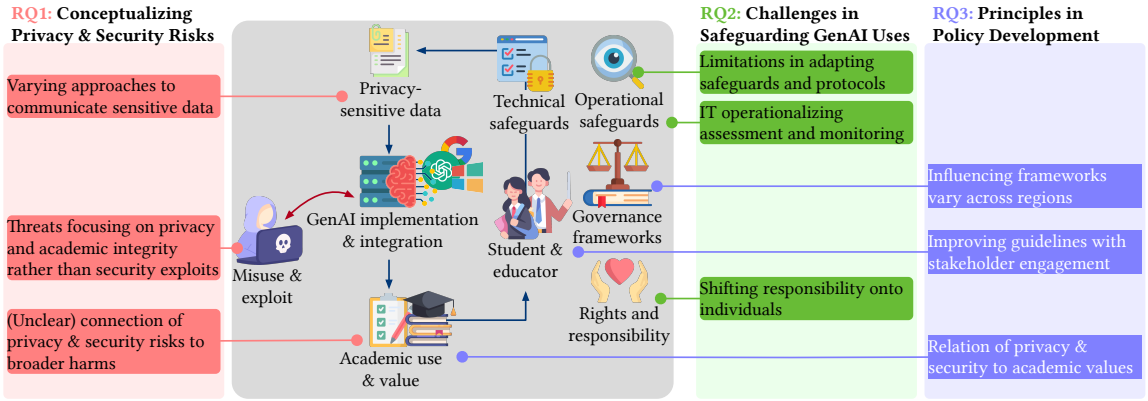

Figure 1. An overview of the findings and themes that address our 3 research questions. We highlight how themes are anchored on 9 categories of codes (within the grey box).

3. **Technical limitations** of GenAI models that can be exploited, such as non-transparency of training, limited reasoning capability and domain-specificity, and service reliability.

Across these categories, university guidelines also implicitly define threat models surrounding the adoption of GenAI in academic settings. Below we further discuss the main (sub)themes observed surrounding elements in the academic threat model, which include (1) protected assets, in particular types of sensitive student and research data, (2) threat actors, especially third-party GenAI vendors and irresponsible users, and their associated attack vectors, and the key (3) affected stakeholders and systems, particularly the learning processes and outcomes. While universities rarely use formal threat-modeling terminology, these dimensions appear throughout the guidelines and shape how privacy and security risks are communicated and mitigated.

5.1.1 Guidelines vary in the approaches to communicate specific sensitive data types to protect. Our analysis reveals that while university guidelines broadly recognize the need to regulate handling sensitive data with GenAI tools, the approaches different universities (and guidelines) took to elaborate on the scope of sensitive data vary. Some university guidelines adopt an "example-driven" approach. For example, Caltech's guideline lists specific examples of sensitive data types including

> *"...any intellectual property or unpublished research data, export-controlled data, and other sensitive HR, business, or administrative data..."* (Caltech)

However, the scope of sensitive data types discussed can be vague in other guidelines, especially when multiple terminologies related to sensitive data, such as *proprietary information, confidential data, and restricted information*, are used concurrently. These terminologies can be mentioned in the same guideline without further elaboration despite signposting to other internal and/or external policy documents and frameworks. For example, the GenAI guideline of NTU mentions that

> *"Any confidential or sensitive information, and/or personal data are not to be uploaded to any GAI software, system, or platform..."* (NTU)

while pointing to two other data policies: Singapore's Personal Data Protection Act (PDPA) [76] and NTU's Data Governance Policy. Similarly, MIT prohibits the use of data *"classified as Medium Risk or High Risk"* using the university's internal data classification framework with AI tools that do not have a license agreement; § 5.3 will further discuss how university guidelines leverage existing policy and regulatory frameworks to articulate their own.

Despite the inconsistent terminology and data classifications adopted by universities, university guidelines (such as those from UPenn) broadly highlight the importance of data protection for proprietary and copyright-protected information, which echoes the elevated concerns for GenAI to infringe IP rights [92]:

> *"Avoid uploading confidential and/or proprietary information to AI platforms prior to seeking patent or copyright protection, as doing so could jeopardize IP rights"* (UPenn)

5.1.2 Recognized threat vectors focus on privacy and academic integrity rather than security exploits. We find that the threat vectors guidelines recognize are centered on the intentional and unintentional GenAI misuse by service providers and academic users.

Privacy exposure to third-party service providers passively and actively. University guidelines frame third-party providers as a key potential threat actor capable of misusing protected academic and personal data, especially when students and staff unintentionally expose such data. For example, UBC's guidelines warn that *"anything that is entered into the [GenAI] tool may be inappropriately exposed to third parties"*. Universities may explicitly name tools from specific vendors, such as Google Gemini and Microsoft CoPilot, with the associated risk levels based on their licensing and privacy assessment outcomes, which § 5.2 will further elaborate.

Moreover, universities are mindful that GenAI service providers may actively infringe users' privacy, as part of the development and technical integration of their GenAI services currently, e.g., *"the risk of exposing personal, confidential and university sensitive information"* (McGill University) for enabling *"internet searching"* (University of Tokyo), frequent model training that infringes *"privacy and intellectual property of information"* (TU Delft), as well as sensitive data required for *"sign[ing] up for an account"* (NUS).

Concern over students and staff centered on misusing GenAI output against academic integrity. On the other hand, guidelines describe the potential of malicious academic use of GenAI by students and staff:

> *"...Potential for malicious actors to use the tools to facilitate irresponsible research practices linked to fabrication, falsification and plagiarism and to make obfuscation easier..."* (UCT)

In general, the security threats on the GenAI model itself are less discussed and less contextualized. Compared to the privacy issues of *personal or sensitive information*, which are discussed in 30 universities' guidelines, just 8 mentions potential for *malicious use* with 4 universities discussing *vulnerabilities, bugs and exploits*, including *jailbreaking*. Among the security issues identified, some university guidelines discuss the secondary security risks due to the malicious use of GenAI-generated content, including those related to harmful textual data and software code. For example, UoE is concerned that GenAI may

> *"...Create code that has security flaws, bugs and use illegal libraries or calls - or infringe copyrights..."* (UoE)

Similarly, CMU mentions GenAI's security risks to *"create more convincing phishing emails"* when used by malicious actors as a motivating background beyond the educational context.

User vulnerabilities due to lack of technical understanding of GenAI. In addition, some university guidelines are also aware that the technical limitations of GenAI may be challenging for the audience to understand, limiting their ability to safely use the service and leading to privacy or security breaches. Therefore, these universities leverage analogies and comparison with other technologies that share similar features or concerns to improve reader understanding. For example, NUS reminds the readers of the unreliability of GenAI service providers who *"can sometimes go down without warning"* and compares GenAI with the old-school calculator when discussing its limited scope in use-cases recommended by the guideline.

5.1.3 Guidelines connect privacy and security risks to broader harms in education which can be unclear. More than the integrated learning management and administrative systems that are affected by privacy breaches, or broader IT infrastructure that is in danger due to security exploits produced by GenAI, universities consistently consider negative impacts on learning processes and outcomes intertwined with privacy and security issues.

Privacy and security impacts positioned among ethical risks. We observe that university guidelines often discuss privacy and security risks within the broader ethical consequences of GenAI that negatively affect students. For example, UIUC names privacy breaches along other harms such as discrimination:

> *"...establish procedures for remediations, recourse, or redress in case of unintended consequences, discrimination, or privacy breaches..."* (UIUC)

Many negative consequences which universities are aware of, e.g., misinformation, are unintended due to technical limitations such as hallucination and biased training datasets:

> *"It is also possible for generative AI models to "hallucinate," or otherwise use false information in their responses"* (NU)

Negative learning outcomes reinforced by adversarial use. Nevertheless, we do observe that several university guidelines realize how "malicious" uses of GenAI and its other negative impacts may reinforce each other in students' learning. For instance, Cornell University expresses its *"overarching concern"* that how students' over-reliance on GenAI might motivate their potential use of "jailbreaking" techniques, which is a major security threat in circumventing GenAI guardrails [102]. This in turn results in students' lack of confidence and ability *"to master needed knowledge or skills."* More specifically, "jailbreaking" could help students in accessing prohibited content in learning:

> *"...Jailbreaking has been used to circumvent prohibitions on certain topics or behaviors such as advocating violence; however, it could also easily be used to obtain answers to homework assignments...."* (Cornell University)

However, guidelines in general seem less specific on how privacy and security risks may connect to other safety concerns of GenAI.

5.2 Challenges in Safeguarding GenAI Uses

We observe that university guidelines discuss privacy and security measures for GenAI under the following categories:

1. **Technical safeguards**, including AI detection tools, data confidentiality and anonymization, and others such as permission management and access control.
2. **Individual rights and responsibility**, such as informed consent, privacy rights and options (e.g., opt-out and data deletion), acknowledging use of AI, and choosing proper alternatives including specific GenAI services to adopt or pedagogical redesign to accommodate GenAI uses.
3. **Operational safeguards**, including security review and licensing, disciplinary actions to ensure compliance of individual conduct and punish misconduct, as well as stakeholder communication and education.

This subsection discusses in-depth how university guidelines evaluate the efficacy and employ these measures.

5.2.1 Guidelines reflect limitations of adapting technical safeguards and existing security protocols. Below discusses the technical reliability concerns and operational complexity to adapt GenAI safeguards we find in the guidelines.

Reliability concerns of technical safeguards. Our study reveals that universities are not over-optimistic about the efficacy of AI detection tools when detecting uses of GenAI. Universities such as the University of Cambridge are aware of the limitations of these tools, and do not offer a standard detection tool as

> *"Detection of AI-generated content is possible, but difficult and not wholly reliable"* (University of Cambridge)

Moreover, some universities are concerned about the instructors' use of AI detection tools due to additional privacy concerns about how students' personal data are handled by these tools.

> *"Instructors should not use these tools to evaluate any student work that contains the name of the student or any other personal information of the student or third parties."* (UBC)

In contrast to the discouragement of unreliable AI detectors, we observed that universities leverage existing technical infrastructures that are deemed more reliable to secure GenAI when integrating it into their IT systems, e.g., by account verification:

> *"Copilot is an enterprise version of an AI-powered chatbot and search engine which better protects the privacy and security of users (when users are signed into their UofT account)."* (UofT)

Likewise, Caltech's guideline describes their access control in place to *"disable or limit access to AI companion tools"* through their enterprise control. Universities also demand the GenAI service providers to implement GenAI safeguards by design:

> *"Protect data and ensure generative AI systems and applications incorporate privacy and security by design."* (UIUC)

In tandem, they rely on the responsibility of individuals to adopt multiple technical safeguards for GenAI, including *"encrypt[ing] sensitive data both at rest and in transit to protect it from unauthorized access,"* using *"anonymized datasets"*, and *"implement[ing] strict access controls, such as Multi-factor Authentication (MFA)"* (UCLA).

Yet, it is unclear how these technical safeguards are contextualized to GenAI service implementation.

Complexity of operational security in multi-department coordination. We observe universities implementing operational privacy and security measures in multiple stages of adopting and using GenAI services, which can be complicated and require multi-party involvement. Universities such as Columbia University (and the University of Chicago) mention their implementation of security review when acquiring or procuring services that *"contain functions that rely on AI to operate," "before they are acquired, even if the software is free."*

Such processes can be initiated and managed by different departments including multiple IT teams within the university, e.g., Columbia IT (CUIT) or Columbia Irving Medical Center (CUIMC-IT) as mentioned in Columbia University's guidelines.

University guidelines also mention their continuous commitment to privacy and security via oversight and awareness training.

> *"keep your training up to date, including the latest training available to you on McGill's cybersecurity awareness training platform."* (McGill University)

Universities often set up dedicated contact points for GenAI-related queries to address *"any questions about the use of this dynamic technology."* (Caltech) University guidelines (e.g., University of Melbourne and UIUC) may also underscore the importance of redressing GenAI harms. Nevertheless, establishing this procedure seems to be an on-going effort.

> *"The University of Melbourne will maintain human oversight of and responsibility for the AI tools and systems used by staff, offer appropriate avenues for redress in case of errors or misuse, and commit to procedures by which those who are impacted can access an explanation for relevant decisions."* (University of Melbourne)

5.2.2 Guidelines shift responsibility onto individuals. A predominant guideline we observe across universities is the devolution of accountability to the individual (students, faculty, and staff).

Individual accountability for making S&P decisions. While some institutions continue to provide technical safeguards, individuals are also expected to be responsible for their privacy and security behaviors and decisions when interacting with GenAI tools, including not disclosing confidential information as discussed before. Another example is UCLA: *"It is the users' responsibility and accountability to get informed about these limitations and risks, closely review the outputs of GenAI tools..."*

To this end, universities expect individuals to choose a GenAI service that provides better privacy and security features. However, a large number of GenAI services on the market can overwhelm individuals as they navigate alternatives. University guidelines often mention a variety of specific GenAI service providers, including standalone GenAI chatbots, e.g., OpenAI's ChatGPT [73] and Google's Gemini [27] as well as GenAI integrated with other digital services such as CoPilot within coding IDE [41] and AI companions in online meeting apps, e.g., Zoom IQ [119]. Despite the availability of many GenAI services, a more detailed privacy and security evaluation for individual services is absent in many guidelines, except for the most popular service provider OpenAI's ChatGPT model:

> *"ChatGPT utilizes one of the largest (known) datasets to power its operation and is hence in its current version of GPT-4 seemingly capable of a variety of complex tasks"* (UM-Ann Arbor).

Furthermore, users are expected to take different levels of responsibility beyond proper tool and data use, including reviewing GenAI outputs and formally citing or acknowledging GenAI use.

Shared responsibility via licensing and risk assessment. As discussed earlier, performing security assessment is a main responsibility universities are committed to, guided by regional or institutional policies like *Privacy and the Need to Monitor and Access Records (SPG 601.11)* and the *Institutional Data Resource Management Policy (SPG 601.12)* established by UM-Ann Arbor. As a result of such assessment and license negotiation, recommending and enforcing the use of university-licensed GenAI services seems to be a more feasible alternative that well-resourced universities offer to individuals. A few universities such as ETH Zürich, UoE, UCL, McGill, UM-Ann Arbor, Caltech have formally licensed platforms including Microsoft Copilot under enterprise agreements with Data Protection Addendums (DPAs) enabled, which shape platform behavior in alignment with institutional values such as academic integrity, data minimization, and ethical AI usage [23]. An example is UM-Ann Arbor's customized U-M GPT that is *"private, secure, and free for students."* (UM-Ann Arbor)

Nevertheless, licensing can be seen as an approach by universities offloading the responsibility to GenAI vendors and transferring trust in a uniform way instead of further tailoring their own safeguards. For instance, ETH Zürich redirects users to Microsoft's Enterprise Data Protection FAQ which reiterates, in the EU context, its commitment to GDPR [38], ISO/IEC 27018 [69], and existing Microsoft DPA [70] that is uniformly applied to all other universities. In these contexts, the main threat identified and mitigated through licensing agreements is the use of student data for training:

> *"The freely accessible version of Microsoft Copilot offers [...] the 'protected' status (with Microsoft ETH-account), whereby personal data is not used as training data,"* (ETH Zürich)

Individuals' challenges in exercising privacy rights and options. Nevertheless, guidelines recognize the challenges of individuals exercising their privacy rights and options for privacy

compliance, e.g., consent procedures, when using GenAI, reflecting universities' distrust of the GenAI service providers' privacy commitment:

> *"you are sharing your data with private companies, so you lose your control over it."* (EPFL)

Similarly, Columbia worries that *"the technology may not respect the privacy rights of individuals..."* Universities are also concerned about individuals not accepting consent carefully:

> *"Individuals who accept click-through agreements without delegated signature authority may face personal consequences, including responsibility for compliance with terms..."* (UC Berkeley)

McGill University explicitly recommends students to self-censor the input for GenAI *"by removing personally identifying information."* Interestingly, we observe that UC Berkeley offers a fine-grained classification of allowable data to be used by specific GenAI services. For example, while *"Publicly-available information (Protection Level P1) can be used freely in all generative AI tools"*, *"...Zoom AI Companion (pilot only) [are] approved for use with P2 information, providing everyone actively consents to its use each time."* The reference also discusses disabled GenAI features which can replace human meeting attendees with bots. However, they fall short in discussing the possible unique privacy implications of these services for individuals to make more informed decisions, e.g., hijacked GenAI bots that produce harmful information.

Some university guidelines ask individual instructors and researchers to offer privacy rights and options when they involve GenAI into their teaching and research. LSE suggests instructors offer opt-out options for students.

> *"Please consider offering students a choice to opt-out of using a system if they have concerns about the cost, privacy, security or other issues related to the technology."* (LSE)

Likewise, UC Berkeley emphasizes the importance for researchers to obtain *"explicit and informed consent from individuals before collecting, processing, or using their data in GenAI systems."*.

5.2.3 IT departments operationalizing GenAI governance through risk assessment and monitoring. Our comparison between teaching and learning-focused guidelines with IT-focused guidelines highlights how university departments coordinate and operationalize GenAI governance. The high-level privacy and security risks, such as third-party sharing, and mitigations identified in IT guidance broadly align with those in teaching and learning-focused guidelines. While teaching and learning guidelines more often emphasize academic values and classroom integration, IT-focused guidelines translate shared privacy and security concerns into concrete assessment, deployment, and monitoring practices.

Making privacy assessment outcome visible. Universities use IT-focused guidance to disclose tool-specific privacy and security assessment *outcomes* and support safer individual adoption of specific services. For example, the University of Oxford warns that Copilot for Microsoft 365 may increase risks to confidential university data because *"access permissions are often poorly managed and this can go unnoticed"*; it provides similar assessments for tools such as Google Gemini and ChatGPT Edu.

IT departments also play a central role in managing and governing licensed GenAI services and custom GenAI systems developed by students or staff. Some host "one-stop" AI guidance websites that consolidate policies, approved services, and risk advice. For instance, UMich's Information and Technology Services states its role in governing and monitoring licensed and custom GenAI services, including conversational tools, API interfaces, and systems powered by contracted or open-source models for custom dataset analysis. For custom GenAI development, UofT's Information Security warns of risks from *"adversary-provided training data"*, while the University of Oxford asks users to *"discuss this with the Information Security GRC team in advance"* before *"inputting internal University data to an external custom GPT."*

Inconsistent efforts in adapting existing safeguards. Universities generally extend existing risk assessment processes and technical protections to GenAI rather than creating entirely new safeguards, such as applying *"the same security and encryption techniques as your emails and contents of your OneDrive or Sharepoint Sites"* (UoM). However, IT departments vary in how actively they adapt and explain these frameworks for GenAI users. 16 of 19 institutions simply ask students and staff to revisit existing data protection policies, while others provide more detailed GenAI-specific interpretation. For example, Imperial College's Data Protection Officer lists privacy-sensitive activities that should be monitored and recorded under its data activity risk assessment framework, including *"the versions of relevant software or firmware"* used by AI systems, the *"recourse to pre-trained systems or tools provided by third parties"*, and *"the cybersecurity measures put in place."*.

Monitoring GenAI usage. IT-focused guidelines emphasize both users' responsibility to *review* GenAI outputs and institutional commitment to *monitor* GenAI use that requires individual cooperation. Beyond information required for acquiring or updating GenAI services, some universities set stronger expectations for recording GenAI outputs. For example, ANU states that when GenAI outputs *"meet the definition of a University record, they must be stored in the ERMS [Electronic Records Management System]."*

5.3 Principles in Policy Development

Considering the identified privacy and security concerns and safeguard limitations, we find that universities adopt GenAI for academic use cautiously and provisionally. Building on the privacy and security issues and safeguards discussed in § 5.1 and 5.2, this subsection examines how university, national, and international frameworks shape framing of privacy and security issues. We identify four key impact factors:

1. **Governance frameworks**, including privacy regulations and policies in different countries or educational associations such as university systems and groups as well as universities' relevant disciplinary procedures.
2. **Academic values** aligned with privacy and security, such as academic integrity and other ethical values such as fairness and equality.
3. **Risk-benefit trade-offs** in academic uses, including productivity, creativity, and personalization, when GenAI assists multiple academic uses such as teaching, learning and assessment, writing, and research.
4. **Engaged stakeholders** in policy development, such as students, faculty, researchers, and staff in relevant departments.

5.3.1 Privacy governance frameworks for GenAI vary across regions and influence guidelines. Universities first recognize the need for GenAI to comply with existing legal obligations. This is stated explicitly in UCT's guidance for example: *"publicly available (non-personal) data may be used with AI tools. For any sensitive data, consider legal, regulatory, or contractual obligations before use"* The awareness of the inextricability of University GenAI guidelines and existing legal frameworks also entails an acknowledgement of the need for universities to continuously update their GenAI guidelines due to the evolving technological and legal landscape:

> *"this document will be updated regularly and interact with other sources of policy, ethics, and governing legal authority."* (UPenn)

National regulatory frameworks anchoring privacy governance rather than AI governance. At the same time, because universities' adaptation stands on existing privacy regulations, they vary across jurisdictions. Differences in national privacy and security regulations therefore shape the robustness and precision of university GenAI guidelines from a privacy and security standpoint. For universities in the US, privacy practices are governed by policies at different levels – *"Federal, state, and local laws as well as Caltech policies [that] may limit data that can be disclosed"* (Caltech). US universities frequently mention HIPAA (Health Insurance Portability and Accountability Act) [46] which guides their sensitive data classification frameworks discussed previously in § 5.1. US universities also rely on FERPA [99] that protects children's education records and gives their parents the right to access and control the data. This regulation surpasses HIPAA in regulating the usage of students' medical records.

> *"It is the faculty members' responsibility to safeguard student data following all relevant regulations covered by FERPA"* (UIUC)

Beyond the US, universities similarly anchor their guidelines in national regulatory frameworks: Singapore's PDPA is referenced by NTU [76], while ANU grounds its guidance in the Australian Privacy Act (1988) [72]. In South Africa, UCT invokes POPIA [77], UK institutions reference the UK GDPR [39] alongside the Data Protection Act 2018 [28], and European universities may cite the GDPR [38] and/or the EU AI Act [3]. In Canada, the University of Toronto adheres to the national FIPPA framework [98], while McGill incorporates FIPPA alongside the "FASTER" principles—Fair, Accountable, Secure, Transparent, Educated, Relevant—endorsed by the Chief Information Officer at the Treasury Board of Canada Secretariat [86]. North American universities lead policy and governance discourse, whereas European universities emphasize operational data management (§ 5.2 and § 5.1): 13 out of 18 NA universities discuss privacy regulations versus 6 out of 14 in Europe.

It is also worth noting that not all countries have regulatory frameworks tailored for GenAI usage, with a main exception in our corpus being McGill's adaptation of FIPPA according to the Canadian Government's GenAI guidance and the EU AI Act. Thus, attention has to be paid on the limitations of these privacy regulations, particularly in their inconsistency in regulating privacy practices. In the US, for example, universities that reference FERPA for student data protection may leave leeway that allows third-party organizations such as GenAI vendors to utilize student metadata [105] – a scenario that is disallowed in other legal contexts such as the Canadian FIPPA, where personal data cannot be shared to third-party organizations unless mandated by law [21].

Despite this limitation, legal compliance can still support more systematic mitigation of the GenAI threats discussed in § 5.2. From a cross-comparative perspective, institutions across North America, the UK, and Australia — including UBC, MIT, McGill, UofT, and Northwestern; Oxford, ICL, and UCL; and ANU — consistently incorporate Privacy Impact Assessments (PIAs) into their GenAI-related policies as part of their legal obligations. These assessments are explicitly communicated to the stakeholders, reflecting a more institutionalized approach to privacy governance and accountability. In contrast, publicly available guidelines from Asian universities do not overtly mention any privacy impact assessments. Instead, they may place greater responsibility on individual users to evaluate the implications of their choices. For example, at Tokyo Tech, using GenAI tools is framed as being entirely at the student's discretion.

Influence of cross-institutional collaboration frameworks Alongside national or provincial legal frameworks, cross-university associations and collaborations with other universities also shape the process of designing guidelines. The former includes the Russell Group in the UK [43], which has published their own set of GenAI guidelines, and US state university systems such as the University of California system, which defines *"privacy and security"* among the seven core principles that guide the policy development at UC Berkeley and UCLA. Universities also make cross-references to other universities' GenAI policies to support their rules and offer complementary information. For example, EPFL's guideline points its reader to a FAQ page for ChatGPT policies developed by their *"sister institution in Zürich."* Similarly, Duke University directs its readers to established guidelines in other universities, noting that, in relation to best practices for GenAI attribution, *"Monash University has curated a list of the various AI citation formats."* This is not a practice limited to universities in Western countries. NTU has also put together *"curated examples"* from Monash University, Cornell University, KCL, and Harvard.

Engagement with international privacy frameworks. In addition to inter-institutional collaboration, a number of universities also explicitly engage with international privacy frameworks. UIUC, for instance, acknowledges that relevant privacy laws span multiple jurisdictions, encompassing federal and state regulations alongside *"geographic and extraterritorial international laws such as GDPR and PIPL [China's Personal Information Protection Law]."* This attention to extraterritorial reach is echoed in the policies of KAUST, Harvard and Duke, which reference GDPR and/or the EU AI Act as a relevant instrument. IC also cites the Indian Digital Act and Chinese AI Safety Framework, while Cornell mentions the EU-U.S. Data Privacy Framework. Awareness of international privacy frameworks in their GenAI guidelines highlights how universities, as globally embedded institutions, must navigate an increasingly complex, cross-jurisdictional privacy landscape.

5.3.2 Relationship between privacy, security and other academic values. Given that GenAI has profoundly impacted teaching and learning while simultaneously introducing privacy and security risks—as elaborated in § 5.1 and § 5.2—university guidelines reflect this complexity in how they frame privacy and security in relation to other academic values.

Balancing benefits with privacy trade-offs. Universities value GenAI's positive uses and recognize that not embracing it might shortchange their students in the *"AI-supported labour market and societies of tomorrow"* (ETH Zürich). However, they are also aware that GenAI tools can be a double-edged sword due to privacy and security risks. For example, the University of Chicago claims that

> *"GenAI tools offer many capabilities and efficiencies that can greatly enhance our work...members of the university community must also consider issues related to information security, privacy, compliance, and academic integrity."* (University of Chicago)

Privacy and security subsumed under broader academic values. The tension between opportunity and risk leads universities to foreground broader academic values—particularly ethics and academic integrity—as the normative foundation for their GenAI stance. For example, *"UCL has opted to promote ethical and transparent engagement with GenAI tools rather than seek to ban them."*

In ETH Zürich's guidelines, privacy and security are framed not as a compliance obligation but as an expression of fairness:

> *"Fairness: Respect the privacy and copyright of the content with which you work. Refrain from disclosing copyrighted, private, or confidential information to commercial GenAI clients, unless expressly permitted."* (ETH Zürich)

Universities may also tie privacy and security expectations to their institutional reputation. For instance, the University of Oxford's policy declares that their *"reputation stands on the trustworthiness of our research and our communications."*

Limits of values-based framing. While grounding privacy in broader values can lend it normative weight, this approach carries a risk: high-level principles are inherently open to interpretation and can sometimes substitute for concrete guidance rather than complement it. Tokyo Tech illustrates this tendency, deferring to student judgment rather than specifying expected practices:

> *"We expect all students to exercise good sense and judgment when using GenAI, in accordance with the Institute's spirit of "Student-Centered Learning."* (Tokyo Tech)

Notably, this breadth may be deliberate. Some universities explicitly design their principles to remain open-ended, as evident in the justification of the University of Melbourne's GenAI guidelines:

"The principles are intentionally broad: the aim is not to prescribe specific initiatives or actions, but to support decision-making across the University, and to ensure the principles can be adapted to developments in the technology. The intention is to periodically review the principles, given the ongoing evolution of AI tools." (University of Melbourne)

Value integration and interaction across departments. Policies from Centres for Teaching and Learning, or equivalent bodies, typically ground GenAI guidance in pedagogical values and institutional ethos. By contrast, IT or information security guidelines, often published as companion documents, foreground technical compliance and risk management. The extent of this divergence varies across institutions. At IC, the IT-specific guidance adopts a "privacy-first" approach, e.g., by asking staff to minimize data collection. This emphasis on privacy is absent from its general guidance, which instead highlights values such as respect, collaboration, excellence, integrity, and innovation. In other cases such as UCLA, the guidelines from two sources show stronger integration: its IT guidance incorporates academic integrity, transparency, and ethical and equitable use, and explicitly directs users to guidance from its Center for Teaching and Learning.

Cross-referencing between IT and teaching-focused guidance also differs by region, when we look at cases we analyzed where institutions publish separate IT-focused AI guidelines, in addition to a teaching and learning one. Among Canadian universities in our corpus, all 3 institutions with IT-specific guidelines link to their teaching and learning counterparts. In the US, this practice is mixed: 5 of the 8 coded US IT-specific policies do not cross-reference. In contrast, in the UK, 6 out of 9 universities maintain dedicated "AI hubs"—centralized landing pages that consolidate access to both teaching-focused and IT-specific guidelines under a single entry point. In Asia, KAUST provides an IT-specific GenAI guideline while 3 of the remaining 6 Asian universities address data security briefly as a principle within their teaching-centered GenAI guidance. This trend is mirrored in Australian universities in our corpus: ANU and Monash University have an IT-specific GenAI guideline, while the remaining 3 universities only cite privacy and security as one of their high-level AI usage principles.

5.3.3 Improving guidelines with stakeholder engagement. Encouraging active participation from students, teaching and research staff, and other university stakeholders in both AI policy creation and its application across specific academic contexts is a pattern observed widely across university guidelines worldwide. For example, Duke University's guideline states that

> *"The Provost is consulting with faculty and staff experts on these larger questions involving generative AI systems, and welcomes debate and discussion on these issues."* (Duke University)

For these stakeholders, university guidelines acknowledge the possibility that the privacy and security considerations could be undermined by their perceived benefits, including the convenience to assist learning, education, and research. Therefore, we observe that university guidelines try to balance these trade-offs. While highlighting the values for students to develop *"their own skills in critical analysis, creative thinking, and rigorous research"* (University of Cambridge), university guidelines simultaneously underscore academic integrity and disciplinary procedures, emphasizing the consequences of misusing GenAI tools. Guidelines ask instructors to take responsibility in navigating GenAI risks for students:

> *"Instructors must educate their students about these risks, and develop plans to mitigate the negative impact of risks in their classroom if they decide to use or allow the use of GenAI in their teaching."* (UM-Ann Arbor)

6 Discussion

Our research illuminates how institutions interpret privacy and security considerations and implement solutions to mitigate the risks of GenAI in academic settings, in particular the gaps in their institutional policies to safeguard GenAI. Our study confirms the conclusions from prior work that privacy and security are a key consideration when adopting GenAI in academic settings [4, 51, 52, 67]. Moreover, our work contributes novel findings compared to prior studies that treated privacy and security concerns as epiphenomenal to broader pedagogical and implementation discussions [100], and a gap in these studies is the distinct lack of granularity in their analysis framework for privacy and security [52, 67]. As a result,

prior findings may superficially identify university awareness of the topic about privacy and security but does not analyze the extent of their understanding, assessment and the challenges faced in mitigating privacy and security issues. Recommendations offered thus also remain at the procedural level, such as how to clarify confidential data types for ensuring greater compliance, without delving into the underlying technical infrastructures and inadequacy of existing privacy frameworks in adapting to GenAI capacities.

Furthermore, there is a dearth of privacy and security research that applies specifically to emerging GenAI services. The GenAI use in universities presents a distinct context of privacy and security compared to enterprise settings [84, 94]: Although both share concerns such as IP leakage, universities stress long-term learning impacts and student privacy harms, whereas enterprises prioritize efficiency and customer satisfaction. Companies also rely on universities for early ethics training while maintaining clearer models for GenAI-related cyber-attacks. Compared to traditional education technologies such as exam proctoring systems [17, 57, 90], the adoption and usage of GenAI is much more flexible and dynamic to accommodate individual requirements and varying application contexts. GenAI amplifies the concerns flagged in the former, such as privacy intrusion and monitoring [111], and the university policies aimed at maintaining control over vendors of education technologies [55] can be unfeasible in this new context.

6.1 Recommendations

Improving privacy and security assessment frameworks for GenAI in academic settings. We found that universities often assess GenAI risks through existing frameworks (e.g., FERPA) and their internal data classification and audit procedures under the influences of regional and international privacy regulations (§ 5.1 and § 5.3). Regulatory development varies across regions: the US has been more advanced in educational privacy governance, while the EU has taken the lead in AI governance. Interestingly, national privacy and GenAI regulatory developments influence higher education across political borders, both among allies such as Canada and the EU, and between strategic competitors such as the US and China, as reflected in UIUC's recognition of China's PIPL (§ 5.3.1). Although it takes time for GenAI regulations to mature, we argue that these influences can extend beyond shaping how university guidelines frame privacy and security risks, supporting the systematic adoption and improvement of local frameworks, such as student- and staff-facing privacy assessment toolkits, to reduce inconsistencies in regulating privacy practices (§ 5.3.1 and 5.3.3). However, our results further show that legal and institutional instruments may fall short of addressing the emerging privacy and security challenges introduced by GenAI – particularly its enhanced capacity to infer sensitive information from minimal or aggregated data [93]. Such capabilities exacerbate risks of re-identification, discrimination, and misuse, which are highlighted in universities' GenAI guidelines (c.f. § 5.1). For example, GenAI could infer students' demographics and sensitive information (gender, race, and health) [60] indirectly from their linguistic patterns revealed in dialogues without accessing student records. Effective governance requires expanding risk models to reflect these GenAI-specific threats and adopting a multi-stakeholder approach, engaging technical teams, legal experts, and academic departments in policy design. More than performing security review during procurement and adoption (as currently highlighted by the guidelines), universities should develop and improve their oversight process to identify new privacy and security risks and remediate harms with fast responses (§ 5.2.3). Instead of relying only on prolonged human review, universities can leverage licensed GenAI services to detect anomalous model interactions in a privacy-preserving manner [115].

Tailoring technical safeguards to academic contexts. Most institutions still rely on traditional IT infrastructure (e.g., account authorization and access control to disable GenAI features), which they have deployed for conventional educational technologies such as LMS, to primarily mitigate risks about data access and breaches. However, this approach does not adequately address the evolving threat landscape posed by GenAI and enforce compliance off-campus (§ 5.2), as GenAI systems synthesize data, introducing uncertainties and new or amplified privacy risks such as identification and physiognomy-related harms [60]. As such, monitoring GenAI outputs becomes a new lever to govern GenAI risks. However, universities are also not confident in the reliability of novel safeguards such as AI detectors that ease the efforts of output monitoring. Furthermore, § 5.3.2 discusses the specific requirements of the universities to align GenAI models with academic values and expect privacy-by-design [44] for these models, which currently falls short in GenAI development. We notice several opportunities to address these requirements given recent technology advances for GenAI safeguards. First, while not perfect, AI detectors could improve their reliability in academic contexts by fine-tuning on expert feedback provided during assessment to improve decision-making with human in the loop [29]. Second, model watermarking, which actively embeds a robust signal invisible to humans into its generated content for verification [56], can be leveraged to ensure compliant use of licensed models even off-campus. Third, differentially private AI model alignment [88, 108] can be applied during model training to ensure privacy and security-by-design along with other academic values tailored to university needs. Improved alignment strategies potentially offer users an efficient way to mitigate risks. Recent commercial approaches such as Google's SynthID [26] and VaultGemma [91] shed light on the practicality of these directions.

Reducing infrastructure disparities through strategic partnerships for integrating GenAI safeguards. Developing and implementing effective GenAI safeguards often requires significant technical and human resources, which many smaller institutions may lack [100]. Nevertheless, we observe good examples of universities and regulators encouraging partnerships in their policy creation, e.g., McGill University's adoption of Canada's FASTER framework illustrates the value of national policy alignment (§ 5.3.1).

We argue that these collaborations and partnerships can be deepened to benefit less-resourced institutions by facilitating knowledge sharing and joint infrastructure development, thus securing their GenAI adoption, e.g., via sharing of GenAI services co-hosted and overseen by universities. The current collaborative partnerships are developed mostly within existing systems or regional consortia. Broader academic networks, like the Russell Group in the UK, could serve as operational hubs to support less-resourced institutions [49]. Nevertheless, we identify several open challenges to

solve when supporting less-resourced institutions. First, beyond resources disparities, we argue that the differing academic and educational missions may influence their positions and focus on privacy and security. For example, compared to research-active universities, the privacy concerns of administrative data such as student records might be more prevalent in professional schools than the IP concerns of GenAI handling research datasets (§ 5.1.1); second, the concern of lacking privacy and security awareness and literacy for staff and students, who take major responsibilities to avoid GenAI misuse, may be further amplified in less-resourced institutions [82]; furthermore, additional privacy and compliance issues need to be considered in these partnerships, e.g., data sharing across universities and regions (§ 5.3.1).

Improving privacy and security communications for GenAI in higher education through analogies and examples. Guidelines highlight the importance of individual accountability and awareness for privacy and security of GenAI adoption, as enforcing central rules that govern every specific GenAI application seems unrealistic and non-ideal (§ 5.2). As such, effective privacy and security communications become crucial for students and instructors to contextualize their GenAI application, e.g., at the course level.

However, disparities in student and instructor preparedness—particularly between technical and non-technical disciplines [4, 31]—undermine consistent privacy and security awareness and enforcement. Our additional analysis shows that guidelines we analyzed require college-level comprehension, with regional variations: lower average Flesch reading ease scores [33] (more difficult) are found for North American guidelines (27.2) than those from Europe (38.3) and Asia (33.7) in Table 1. Though privacy and security policies can be complicated to understand in a GenAI guideline due to its technical nature, our findings in § 5.1.2 show that universities use analogies to help stakeholders make sense of emerging GenAI technologies, which can be further adapted for privacy and security communication. For example, NUS and UofT effectively communicate the level of appropriate use of GenAI services in the teaching and learning context by drawing analogies between using GenAI and *"using calculators"* or *"consulting peers."* This approach situates the new technology in *"a familiar historical and socio-technical context,"* understood as a linear development of pedagogical methods [85].

In the same manner, analogies can be developed to improve the accessibility of otherwise abstract and technical jargon-laden data privacy and security measures to support students, staff and researchers in making optimal privacy and security protection decisions [19]. For example, guidelines could help people understand how their personal data can be *"unlearned"* by GenAI models as a new approach to exercise their privacy rights [10]. In addition, examples such as template syllabi (e.g., from UM-Ann Arbor) and sensitive data types are observed as another starting point for university guidelines to clarify complex privacy and security concepts (§ 5.1 and § 5.2). Continuous professional development remains crucial to the cause: universities should provide flexible communications and training, including micro-training for data privacy embedded in the interaction with university-licensed GenAI [15, 87]. Example user prompts can also be provided to improve privacy and security alongside productivity in GenAI interaction [113].

6.2 Future Work

Longitudinal and multi-level policy measurement and analysis. A promising direction for future research involves tracking how GenAI governance evolves across time and institutional levels. Longitudinal studies, enabled by automated policy collection and natural language processing (NLP) pipelines, could uncover how universities adapt GenAI-related guidelines in response to shifts in law, pedagogy, technology, as well as privacy and security incidents of GenAI. Multi-level comparisons, across institutional policies, departmental guidelines, course syllabi, and broader privacy frameworks like FERPA and GDPR [38, 99], can reveal alignment gaps and structural inconsistencies in GenAI governance. Another promising avenue for research would be to examine where such discrepancies stem from, for example, tensions between regulatory levels and varying national attitudes towards restrictive regulations. A further exploration of trade-offs across different contexts or overarching governance frameworks would provide valuable insights.

Understanding stakeholder perceptions on policy-making and enforcement of GenAI privacy and security. Future work should further study how different stakeholders (students, instructors, administrators, and service providers) understand and engage with universities' GenAI policies and enforcement, e.g., through interviews and surveys. Identifying gaps due to misconceptions and conflicting interests about privacy and security risks can inform the design of more accessible, targeted communication strategies, as well as business and service models to improve licensing.

Designing and evaluating privacy-preserving GenAI systems for education in a human-centric way. Technically, a key research challenge lies in building usable GenAI systems and safeguards that are aligned with academic values and personalized to user needs. Future research can involve students and educators in different phases, from co-designing prototypes to end-to-end system evaluation, when developing privacy and security measures for GenAI, e.g., for managing data sharing, interaction logs, and model behavior transparently. Balancing privacy with personalization and creative affordances is also central to making GenAI both effective and responsible in educational contexts.

7 Conclusions

GenAI has become increasingly popular in higher education. It enables new education paradigms and boosts people's efficiency, productivity, and creativity in various educational activities. However, there are significant privacy and security risks accompanying the wide adoption of GenAI. To understand how universities address these emerging privacy and security risks and evaluate mitigation strategies, we performed an in-depth qualitative analysis of GenAI guidelines from 43 universities around the world. Our study uncovered novel insights about the gaps universities face in safeguarding GenAI usage, including the limitations in existing technical and procedural measures to mitigate risks and the alignment of privacy and security with other academic values in institutional policy development. These findings further motivated recommendations to improve assessment frameworks, design tailored technical solutions for academic contexts, and enhance stakeholder communication and awareness for GenAI privacy and security in academic settings.

Acknowledgments

We would like to acknowledge the support from Madelyn Rose Sanfilippo, and our anonymous reviewers and editor. Bei Yi Ng, Jiarui Li, Xinyuan Tong, and Jingjie Li would like to thank the Generative AI Laboratory, University of Edinburgh. For Kevin Ye, Gauthami Yenne, and Varun Chandrasekaran, this research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The authors used generative AI-based tools to revise the text, improve flow and correct grammatical errors.

References

[1] FY2024 Undergraduate Entrance Ceremony Remarks (5 April 2024) — kyoto-u.ac.jp. https://www.kyoto-u.ac.jp/en/about/president/speech/2024/240405-1. [Accessed 29-11-2025].

[2] html-sanitizer — pypi.org. https://pypi.org/project/html-sanitizer/, 2025. [Accessed 29-11-2025].

[3] EU AI Act-Info. EU AI Act. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng, 2024. [Accessed 15-05-2026].

[4] Areej Ali, Aayushi Hingle Collier, Umama Dewan, Nora McDonald, and Aditya Johri. Analysis of Generative AI policies in computing course syllabi. In *Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 1*, pages 18–24, 2025.

[5] Lindsay R Baker, Shanon Phelan, Nicole N Woods, Victoria A Boyd, Paula Rowland, and Stella L Ng. Re-envisioning paradigms of education: towards awareness, alignment, and pluralism. *Advances in Health Sciences Education*, 26:1045–1058, 2021.

[6] David G. Balash, Dongkun Kim, Darika Shaibekova, Rahel A. Fainchtein, Micah Sherr, and Adam J. Aviv. Examining the examiners: Students' privacy and security perceptions of online proctoring services. In *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*, pages 633–652. USENIX Association, August 2021.

[7] Hamsa Bastani, Osbert Bastani, Alp Sungu, Haosen Ge, Özge Kabakcı, and Rei Mariman. Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, 122(26):e2422633122, 2025.

[8] Brett A Becker and Thomas Fitzpatrick. What do CS1 syllabi reveal about our expectations of introductory programming students? In *Proceedings of the 50th ACM technical symposium on computer science education*, pages 1011–1017, 2019.

[9] Marina Belkina, Scott Daniel, Sasha Nikolic, Rezwanul Haque, Sarah Lyden, Peter Neal, Sarah Grundy, and Ghulam M. Hassan. Implementing Generative AI (GenAI) in higher education: A systematic review of case studies. *Computers and Education: Artificial Intelligence*, 8:100407, June 2025.

[10] Lucas Bourtoule, Varun Chandrasekaran, Christopher A Choquette-Choo, Hengrui Jia, Adelin Travers, Baiwu Zhang, David Lie, and Nicolas Papernot. Machine unlearning. In *2021 IEEE symposium on security and privacy (SP)*, pages 141–159. IEEE, 2021.

[11] Virginia Braun and Victoria Clarke. Can i use TA? should i use TA? should i not use TA? comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches. *Counselling and psychotherapy research*, 21(1):37–47, 2021.

[12] Ben Burgess, Avi Ginsberg, Edward W. Felten, and Shaanan Cohney. Watching the watchers: bias and vulnerability in remote proctoring software. In *31st USENIX Security Symposium (USENIX Security 22)*, pages 571–588, Boston, MA, August 2022. USENIX Association.

[13] Claudia Camacho-Zuñiga, Marco A. Rodea-Sánchez, Omar Olmos López, and Genaro Zavala. Generative AI guidelines by/for engineering undergraduates. In *2024 IEEE Global Engineering Education Conference (EDUCON)*, page 1–8, May 2024.

[14] Yalin Cao, Jinna Fan, and Qian Yang. Chatgpt in education: Ethical predicaments of Generative AI. *Transactions on Social Science, Education and Humanities Research*, 11:64–69, August 2024.

[15] Pawat Chaipidech, Niwat Srisawasdi, Tanachai Kajornmanee, and Kornchawal Chaipah. A personalized learning system-supported professional training model for teachers' tpack development. *Computers and Education: Artificial Intelligence*, 3:100064, 2022.

[16] Cecilia Ka Yuk Chan and Wenjie Hu. Students' voices on Generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20(1):43, 2023.

[17] Jake Chanenson, Brandon Sloane, Navaneeth Rajan, Amy Morril, Jason Chee, Danny Yuxing Huang, and Marshini Chetty. Uncovering privacy and security challenges in k-12 schools. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–28, 2023.

[18] Yupeng Chang, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxiang Wang, Yidong Wang, et al. A survey on evaluation of Large Language Models. *ACM transactions on intelligent systems and technology*, 15(3):1–45, 2024.

[19] Qing Chen, Wei Shuai, Jiyao Zhang, Zhida Sun, and Nan Cao. Beyond numbers: Creating analogies to enhance data comprehension and communication with Generative AI. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2024.

[20] Junjie Chu, Yugeng Liu, Ziqing Yang, Xinyue Shen, Michael Backes, and Yang Zhang. Comprehensive assessment of jailbreak attacks against LLMs. *arXiv preprint arXiv:2402.05668*, 2024.

[21] CIGI. AI governance in research libraries: A case study of the University of Toronto libraries. Technical report, CIGI: Centre for International Governance Innovation, 2024. [Accessed 29-11-2025].

[22] Victoria Clarke and Virginia Braun. Thematic analysis. *The journal of positive psychology*, 12(3):297–298, 2017.

[23] Shaanan Cohney, Ricardo Teixeira, David Kohlbrenner, Arvind Narayanan, Mihir Kshirsagar, Yan Shvartzshnaider, and Madelyn Sanfilippo. Virtual classrooms and real harms: Remote learning at u.s. universities. In *Proceedings of the 17th Symposium on Usable Privacy and Security (SOUPS 2021)*, 2021.

[24] Cornell University Committee on Generative AI. CU committee report: Generative Artificial Intelligence for education and pedagogy. https://teaching.cornell.edu/{Generative}-artificial-intelligence/cu-committee-report-{Generative}-artificial-intelligence-education, 2023. [Accessed 11-04-2025].

[25] Badhan Chandra Das, M Hadi Amini, and Yanzhao Wu. Security and privacy challenges of Large Language Models: A survey. *ACM Computing Surveys*, 57(6):1–39, 2025.

[26] Sumanth Dathathri, Abigail See, Sumedh Ghaisas, Po-Sen Huang, Rob McAdam, Johannes Welbl, Vandana Bachani, Alex Kaskasoli, Robert Stanforth, Tatiana Matejovicova, et al. Scalable watermarking for identifying Large Language Model outputs. *Nature*, 634(8035):818–823, 2024.

[27] Deepmind. Gemini — deepmind.google. https://deepmind.google/technologies/gemini/, 2025. [Accessed 29-11-2025].

[28] UK DPA-Info. United Kingdom Data Protection Act. https://www.legislation.gov.uk/ukpga/2018/12/contents, 2018. [Accessed 15-05-2026].

[29] Liam Dugan, Daphne Ippolito, Arun Kirubarajan, Sherry Shi, and Chris Callison-Burch. Real or fake text?: Investigating human ability to detect boundaries between human-written and machine-generated text. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 12763–12771, 2023.

[30] Anirudh Ekambaranathan, Jun Zhao, and Max Van Kleek. "money makes the world go around": Identifying barriers to better privacy in children's apps from developers' perspectives. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery.

[31] Musaddag Elrayah and Saima Jamil. Impact of digital literacy and online privacy concerns on cybersecurity behaviour: The moderating role of cybersecurity awareness. *International Journal of Cyber Criminology*, 17(2):166–187, 2023.

[32] Hassan Elsayed. The impact of hallucinated information in Large Language Models on student learning outcomes: A critical examination of misinformation risks in AI-assisted education. *Northern Reviews on Algorithmic Research, Theoretical Computation, and Complexity*, 9(8):11–23, 2024.

[33] James N Farr, James J Jenkins, and Donald G Paterson. Simplification of flesch reading ease formula. *Journal of applied psychology*, 35(5):333, 1951.

[34] Federal Trade Commission. Children's online privacy protection rule, 2013. [Accessed 09-04-2025].

[35] Nigel J Francis, Sue Jones, and David P Smith. Generative AI in higher education: Balancing innovation and integrity. *British Journal of Biomedical Science*, 81:14048, 2025.

[36] Vyoma Gajjar. The future of AI governance: Navigating the challenges of GenerativeAI. *International Journal of Science and Research (IJSR)*, 13(9):1686–1687, 2024.

[37] Iván Miguel García-López and Laura Trujillo-Liñán. Ethical and regulatory challenges of Generative AI in education: a systematic review. *Frontiers in Education*, 10, 2025.

[38] GDPR-Info. General Data Protection Regulation (GDPR) – Legal Text — gdpr-info.eu. https://gdpr-info.eu/, 2016. [Accessed 29-11-2025].

[39] UK GDPR-Info. United Kingdom General Data Protection Regulation. https://www.legislation.gov.uk/eur/2016/679/contents, 2016. [Accessed 15-05-2026].

[40] Michail Giannakos, Roger Azevedo, Peter Brusilovsky, Mutlu Cukurova, Yannis Dimitriadis, Davinia Hernandez-Leo, Sanna Järvelä, Manolis Mavrikis, and Bart Rienties. The promise and challenges of Generative AI in education. *Behaviour & Information Technology*, page 1–27, September 2024.

[41] Github. GitHub Copilot · Your AI pair programmer — github.com. https://github.com/features/copilot, 2025. [Accessed 29-11-2025].

[42] Kai Greshake, Sahar Abdelnabi, Shailesh Mishra, Christoph Endres, Thorsten Holz, and Mario Fritz. Not what you've signed up for: Compromising real-world LLM-integrated applications with indirect prompt injection. In *Proceedings of*

the 16th ACM Workshop on Artificial Intelligence and Security, pages 79–90, 2023.
[43] Russell Group. Our universities | Russell Group — russellgroup.ac.uk. https://www.russellgroup.ac.uk/our-universities, 2025. [Accessed 29-11-2025].
[44] Seda Gürses, Carmela Troncoso, and Claudia Diaz. Engineering privacy by design. *Computers, Privacy & Data Protection*, 14(3):25, 2011.
[45] Jack Hardinges, Elena Simperl, and Nigel Shadbolt. We must fix the lack of transparency around the data used to train foundation models. *Harvard Data Science Review*, Special Issue 5, May 2024. [Accessed 12-04-2025].
[46] HHS. Summary of the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html, 2025. [Accessed 29-11-2025].
[47] Xiangen Hu, Sheng Xu, Richard Tong, and Art Graesser. Generative AI in education: From foundational insights to the socratic playground for learning. *arXiv preprint arXiv:2501.06682*, 2025.
[48] Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, et al. A survey on hallucination in Large Language Models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, 43(2):1–55, 2025.
[49] Robert Huggins, Andrew Johnston, and Chris Stride. Knowledge networks and universities: Locational and organisational aspects of knowledge transfer interactions. *Entrepreneurship & Regional Development*, 24(7-8):475–502, 2012.
[50] IAPP Staff. Do LLMs store personal data? this is asking the wrong question. https://iapp.org/news/a/do-llms-store-personal-data-this-is-asking-the-wrong-question, 2024. [Accessed 11-04-2025].
[51] Junfeng Jiao, Saleh Afroogh, Kevin Chen, David Atkinson, and Amit Dhurandhar. The global landscape of academic guidelines for Generative AI and llms. *Nature Human Behaviour*, pages 1–5, 2025.
[52] Yueqiao Jin, Lixiang Yan, Vanessa Echeverria, Dragan Gašević, and Roberto Martinez-Maldonado. Generative AI in higher education: A global perspective of institutional adoption policies and guidelines. *Computers and Education: Artificial Intelligence*, 8:100348, June 2025.
[53] Irina Jurenka, Markus Kunesch, Kevin R McKee, Daniel Gillick, Shaojian Zhu, Sara Wiltberger, Shubham Milind Phal, Katherine Hermann, Daniel Kasenberg, Avishkar Bhoopchand, et al. Towards responsible development of Generative AI for education: An evaluation-driven approach. *arXiv preprint arXiv:2407.12687*, 2024.
[54] Kadaruddin Kadaruddin. Empowering education through Generative AI: Innovative instructional strategies for tomorrow's learners. *International Journal of Business, Law, and Education*, 4(2):618–625, 2023.
[55] Easton Kelso, Ananta Soneji, Sazzadur Rahaman, Yan Shoshitaishvili, and Rakibul Hasan. Trust, because you can't verify: Privacy and security hurdles in education technology acquisition practices. In *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security*, pages 1656–1670, 2024.
[56] Rohith Kuditipudi, John Thickstun, Tatsunori Hashimoto, and Percy Liang. Robust distortion-free watermarks for language models. *arXiv preprint arXiv:2307.15593*, 2023.
[57] Priya C. Kumar, Marshini Chetty, Tamara L. Clegg, and Jessica Vitak. Privacy and security considerations for digital technology use in elementary schools. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–13, New York, NY, USA, 2019. Association for Computing Machinery.
[58] Sangeetha Kutty, Ritesh Chugh, Pethigamage Perera, Arjun Neupane, Meena Jha, Lily Li, Wijendra Gunathilake, and Nimeshia Chamini Perera. Generative AI in higher education: Perspectives of students, educators and administrators. *Journal of Applied Learning and Teaching*, 7(2):47–60, August 2024.
[59] Monika Blue Kwapisz, Avanya Kohli, and Prashanth Rajivan. Privacy concerns of student data shared with instructors in an online learning management system. 2024.
[60] Hao-Ping Lee, Yu-Ju Yang, Thomas Serban Von Davier, Jodi Forlizzi, and Sauvik Das. Deepfakes, phrenology, surveillance, and more! a taxonomy of AI privacy risks. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–19, 2024.
[61] Fang Liu, Yang Liu, Lin Shi, Houkun Huang, Ruifeng Wang, Zhen Yang, Li Zhang, Zhongqi Li, and Yuchi Ma. Exploring and evaluating hallucinations in LLM-powered code generation. *arXiv preprint arXiv:2404.00971*, 2024.
[62] Yi Liu, Gelei Deng, Yuekang Li, Kailong Wang, Zihao Wang, Xiaofeng Wang, Tianwei Zhang, Yepang Liu, Haoyu Wang, Yan Zheng, et al. Prompt injection attack against LLM-integrated applications. *arXiv preprint arXiv:2306.05499*, 2023.
[63] Yiheng Liu, Hao He, Tianle Han, Xu Zhang, Mengyuan Liu, Jiaming Tian, Yutong Zhang, Jiaqi Wang, Xiaohui Gao, Tianyang Zhong, et al. Understanding LLMs: A comprehensive overview from training to inference. *Neurocomputing*, page 129190, 2024.
[64] Jiahui Luo. A critical review of genai policies in higher education assessment: A call to reconsider the "originality" of students' work. *Assessment & Evaluation in Higher Education*, 49(5):651–664, 2024.
[65] Seshadri Mallika. Communities demand transparency after ed, lausd's AI chatbot, fails. https://edsource.org/2024/communities-demand-transparency-after-ed-lausds-ai-chatbot-fails/717772, 2024. [Accessed 29-11-2025].
[66] Sana Maqsood and Sonia Chiasson. "they think it's totally fine to talk to somebody on the internet they don't know": Teachers' perceptions and mitigation strategies of tweens' online risks. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery.
[67] Nora McDonald, Aditya Johri, Areej Ali, and Aayushi Hingle Collier. Generative Artificial Intelligence in higher education: Evidence from an analysis of institutional policies and guidelines. *Computers in Human Behavior: Artificial Humans*, 3:100121, March 2025.
[68] Mary L McHugh. Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3):276–282, 2012.
[69] Microsoft. Iso/iec 27018 - microsoft compliance. https://learn.microsoft.com/en-us/compliance/regulatory/offering-iso-27018, 2023. [Accessed 29-11-2025].
[70] Microsoft. Microsoft products and services data protection addendum (dpa). https://www.microsoft.com/licensing/docs/view/Microsoft-Products-and-Services-Data-Protection-Addendum-DPA, 2023. [Accessed 29-11-2025].
[71] Niloofar Mireshghallah, Maria Antoniak, Yash More, Yejin Choi, and Golnoosh Farnadi. Trust no bot: Discovering personal disclosures in human-LLM conversations in the wild. *arXiv preprint arXiv:2407.11438*, 2024.
[72] OAIC. The Privacy Act — oaic.gov.au. https://www.oaic.gov.au/privacy/privacy-legislation/the-privacy-act. [Accessed 29-11-2025].
[73] OpenAI. ChatGPT overview. https://openai.com/chatgpt/overview/. [Accessed 29-11-2025].
[74] OpenAI Help Center. Data controls faq. https://help.openai.com/en/articles/7730893-data-controls-faq, 2024. [Accessed 12-04-2025].
[75] Cecelia Parks. Beyond compliance: Students and FERPA in the age of Big Data. *Journal of Intellectual Freedom & Privacy*, 2(2):23–33, 2017.
[76] PDPC. PDPC | PDPA Overview — pdpc.gov.sg. https://www.pdpc.gov.sg/overview-of-pdpa/the-legislation/personal-data-protection-act. [Accessed 29-11-2025].
[77] POPIA-Info. Protection of Personal Information Act (POPI Act). https://popia.co.za/, 2013. [Accessed 15-05-2026].
[78] Jia Qi, Ji'an Liu, and Yanru Xu. The role of individual capabilities in maximizing the benefits for students using genai tools in higher education. *Behavioral Sciences*, 15(3):328, 2025.
[79] Mike Quartararo. The impact of Generative AI on education. https://aceds.org/the-impact-of-generative-ai-on-education-aceds-blog/, 2023. [Accessed 12-04-2025].
[80] Rajesh Ranjan, Shailja Gupta, and Surya Narayan Singh. A comprehensive survey of bias in LLMs: Current landscape and future directions. *arXiv preprint arXiv:2409.16430*, 2024.
[81] Joel R. Reidenberg and Florian Schaub. Achieving big data privacy in education. *Theory and Research in Education*, 16(3):263–279, 2018.
[82] Yacine Rezgui and Adam Marks. Information security awareness in higher education: An exploratory study. *Computers & security*, 27(7-8):241–253, 2008.
[83] Benjamin Saunders, Julius Sim, Tom Kingstone, Shula Baker, Jackie Waterfield, Bernadette Bartlam, Heather Burroughs, and Clare Jinks. Saturation in qualitative research: exploring its conceptualization and operationalization. *Quality & quantity*, 52:1893–1907, 2018.
[84] Johannes Schneider, Rene Abraham, Christian Meske, and Jan vom Brocke. Ai governance for businesses. *arXiv preprint arXiv:2011.10672*, 2020.
[85] Claudia Schwarz-Plaschg. The power of analogies for imagining and governing emerging technologies. *NanoEthics*, 12(2):139–153, August 2018.
[86] Treasury Board of Canada Secretariat. Guide on the use of Generative artificial intelligence. https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/guide-use-{Generative}-ai.html, May 2024. [Accessed 29-11-2025].
[87] Tamar Shamir-Inbal and Ina Blau. Micro-learning in designing professional development for ict teacher leaders: The role of self-regulation and perceived learning. *Professional Development in Education*, 48(5):734–750, 2022.
[88] Tianhao Shen, Renren Jin, Yufei Huang, Chuang Liu, Weilong Dong, Zishan Guo, Xinwei Wu, Yan Liu, and Deyi Xiong. Large language model alignment: A survey. *arXiv preprint arXiv:2309.15025*, 2023.
[89] Vivian Sheng. Attitudes towards the use of Generative AI in education: Triangulating the perspectives of students, teachers, and industry professionals. *Lecture Notes in Education Psychology and Public Media*, 123(1):99–111, October 2025.
[90] Elisa Shioji, Ani Meliksetyan, Lucy Simko, Ryan Watkins, Adam Aviv, and Shaanan Cohney. " it's been lovely watching you": Institutional decision-making on online proctoring software. In *2025 IEEE Symposium on Security and Privacy (SP)*, pages 18–18. IEEE Computer Society, 2024.
[91] Amer Sinha, Thomas Mesnard, Ryan McKenna, Daogao Liu, Christopher A Choquette-Choo, Yangsibo Huang, Da Yu, George Kaissis, Zachary Charles, Ruibo Liu, et al. Vaultgemma: A differentially private gemma model. *arXiv preprint arXiv:2510.15001*, 2025.

[92] Jan Smits and Tijn Borghuis. Generative AI and Intellectual Property rights. In *Law and artificial intelligence: Regulating AI and applying AI in legal practice*, pages 323–344. Springer, 2022.
[93] Robin Staab, Mark Vero, Mislav Balunović, and Martin Vechev. Beyond memorization: Violating privacy via inference with Large Language Models. *arXiv preprint arXiv:2310.07298*, 2023.
[94] Araz Taeihagh. Governance of Generative AI, 2025.
[95] Arnout Terpstra, Alwin De Rooij, and Alexander Schouten. Online proctoring: Privacy invasion or study alleviation? discovering acceptability using contextual integrity. 2023.
[96] Tanja Tillmanns, Alfredo Salomão Filho, Susmita Rudra, Peter Weber, Julia Dawitz, Emma Wiersma, Dovile Dudenaite, and Sally Reynolds. Mapping tomorrow's teaching and learning spaces: A systematic review on GenAI in higher education. *Trends in Higher Education*, 4(1):2, 2025.
[97] Chinemelum Goodness Udeh. The role of Generative AI in personalized learning for higher education. *World Journal of Advanced Engineering Technology and Sciences*, 14(2):205–207, 2025.
[98] University of Toronto. About FIPPA | The Office of the Governing Council, FIPP, Jun 2019.
[99] U.S. Department of Education. Family educational rights and privacy; final rule, 2011. [Accessed 9-04-2025].
[100] Hui Wang, Anh Dang, Zihao Wu, and Son Mac. Generative AI in higher education: Seeing ChatGPT through universities' policies, resources, and guidelines. *Computers and Education: Artificial Intelligence*, 7:100326, 2024.
[101] Ning Wang, Ying Li, and Fengyu Cong. University students' privacy concerns towards Generative Artificial Intelligence. *Journal of Academic Ethics*, 23(4):2401–2422, December 2025.
[102] Alexander Wei, Nika Haghtalab, and Jacob Steinhardt. Jailbroken: How does LLM safety training fail? *Advances in Neural Information Processing Systems*, 36:80079–80110, 2023.
[103] James Weichert, Dayoung Kim, Qin Zhu, and Hoda Eldardiry. 'do i have to take this class?': A review of ethics requirements in computer science curricula. In *Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 1*, pages 1197–1203, 2025.
[104] Ryan Thomas Williams. The ethical implications of using Generative chatbots in higher education. *Frontiers in Education*, 8, January 2024.
[105] Cadence Willse. State education agency governance, virtual learning, and student privacy: Lessons from the COVID-19 pandemic. *Educational Policy*, 38(1):186–217, 2024.
[106] Gavin Witsken, Igor Crk, and Eren Gultepe. LLMs in the Classroom: Outcomes and Perceptions of Questions Written with the Aid of AI. *arXiv preprint arXiv:2503.18995*, 2025.
[107] QS World University Rankings. QS World University Rankings 2024 — topuniversities.com. https://www.topuniversities.com/world-university-rankings/2024, 2024. [Accessed 19-04-2025].
[108] Fan Wu, Huseyin A Inan, Arturs Backurs, Varun Chandrasekaran, Janardhan Kulkarni, and Robert Sim. Privately aligning language models with reinforcement learning. *arXiv preprint arXiv:2310.16960*, 2023.
[109] Liangru Xie, Hui Liu, Jingying Zeng, Xianfeng Tang, Yan Han, Chen Luo, Jing Huang, Zhen Li, Suhang Wang, and Qi He. A survey of calibration process for black-box LLMs. *arXiv preprint arXiv:2412.12767*, 2024.
[110] Jin Yang, Zhiqiang Wang, Yanbin Lin, and Zunduo Zhao. Global data constraints: Ethical and effectiveness challenges in Large Language Model. *arXiv e-prints*, pages arXiv–2406, 2024.
[111] Tianyi Yang and Rakibul Hasan. Discovering privacy harms from education technology by analyzing user reviews. In *Proceedings of the 23rd Workshop on Privacy in the Electronic Society*, page 186–192, Salt Lake City UT USA, November 2023. ACM.
[112] Abdullahi Yusuf, Nasrin Pervin, and Marcos Román-González. Generative AI and the future of higher education: a threat to academic integrity or reformation? evidence from multicultural perspectives. *International Journal of Educational Technology in Higher Education*, 21(1):21, 2024.
[113] J Diego Zamfirescu-Pereira, Richmond Y Wong, Bjoern Hartmann, and Qian Yang. Why johnny can't prompt: how non-AI experts try (and fail) to design LLM prompts. In *Proceedings of the 2023 CHI conference on human factors in computing systems*, pages 1–21, 2023.
[114] Angie Zhang and Min Kyung Lee. Knowledge workers' perspectives on AI training for responsible AI use. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pages 1–18, 2025.
[115] Xiaoyu Zhang, Cen Zhang, Tianlin Li, Yihao Huang, Xiaojun Jia, Ming Hu, Jie Zhang, Yang Liu, Shiqing Ma, and Chao Shen. Jailguard: A universal detection framework for prompt-based attacks on LLM systems. *ACM Trans. Softw. Eng. Methodol.*, 35(1), December 2025.
[116] Zhuojing Zhang and Sarrah Wasie. Educational technology in the post-pandemic era: Current progress, potential, and challenges. In *Proceedings of the 15th International Conference on Education Technology and Computer*s, ICETC '23, page 40–46, New York, NY, USA, 2024. Association for Computing Machinery.
[117] Jun Zhao, Ge Wang, Carys Dally, Petr Slovak, Julian Edbrooke-Childs, Max Van Kleek, and Nigel Shadbolt. 'i make up a silly name': Understanding children's perception of privacy risks online. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–13, New York, NY, USA, 2019. Association for Computing Machinery.
[118] Victoria Zhong, Susan McGregor, and Rachel Greenstadt. "i'm going to trust this until it burns me" parents' privacy concerns and delegation of trust in K-8 Educational Technology. In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 5073–5090, Anaheim, CA, August 2023. USENIX Association.
[119] Zoom. Meet Zoom AI Companion, your new AI assistant! Unlock the benefits with a paid Zoom account — zoom.com. https://www.zoom.com/zh-cn/blog/zoom-ai-companion/. [Accessed 29-11-2025].

A Appendix

A.1 Complementary Sample

Continent	Country/Region	University	Document Title
Africa	South Africa	University of Cape Town (UCT)	ICTS AI Guidelines
Asia	Saudi Arabia	King Abdullah University of Science and Technology (KAUST)	New Security Controls on Third-Party App Plugins & AI Bots
Europe	United Kingdom	Imperial College London (IC)	DPO: Considerations for the Use of AI
Europe	United Kingdom	London School of Economics (LSE)	DTS: AI Legal and Regulatory Guidance
Europe	United Kingdom	University of Edinburgh (UoE)	IS AI Safety: Data Privacy
Europe	United Kingdom	University College London (UCL)	IS Guidance on Use of Third Party AI Services
Europe	United Kingdom	University of Oxford	IS Use Generative AI Services
North America	Canada	McGill University	IT Service AI Guidelines
North America	Canada	University of British Columbia (UBC)	Principles for the use of GenAI tools
North America	Canada	University of Toronto (UofT)	IS Use Artificial Intelligence Intelligently
North America	United States	California Institute of Technology (Caltech)	Guidance on the Use of GenAI & LLM Tools
North America	United States	Carnegie Mellon University (CMU)	CIO Use AI Safely at CMU
North America	United States	Cornell University	IT Guidelines for Artificial Intelligence
North America	United States	Duke University	OIT Artificial Intelligence Security and Privacy
North America	United States	University of California, Los Angeles (UCLA)	DTS AI Use and Recommendations
North America	United States	University of Illinois Urbana-Champaign (UIUC)	CIO Privacy considerations for Generative AI
North America	United States	University of Michigan-Ann Arbor (UM-Ann Arbor)	ITS AI Guidelines
North America	United States	University of Michigan-Ann Arbor (UM-Ann Arbor)	ITS Artificial Intelligence and U-M Data
Oceania	Australia	Monash University	Cybersecurity: Use AI Responsibly
Oceania	Australia	Australian National University (ANU)	CGRO Generative AI and Data Governance

Table 2. Additional GenAI guidelines contributed by IT departments.