REVIEW OF ARTIFICIAL INTELLIGENCE IN EDUCATION

DOI: https://doi.org/10.37497/rev.artif.intell.educ.v7ii.97

ARTICLE

AI, CYBERSECURITY AND STUDENTS' RIGHTS IN EU DIGITAL EDUCATION GOVERNANCE

Inteligência Artificial, Cibersegurança e Direitos dos Estudantes na Governança Digital da Educação na União Europeia

Andreea Nicoleta Dragomir
Faculty of Law, Lucian Blaga University of Sibiu (Romania)
E-mail: andreea.dragomir@ulbsibiu.ro

Ovidiu Bernaschi
Faculty of Sciences, Lucian Blaga University of Sibiu (Romania)
E-mail: ovidiu.bernaschi@ulbsibiu.ro

ABSTRACT | Objective: This article examines how artificial intelligence in higher education may affect students' rights, particularly privacy, autonomy, equality, fair assessment, and institutional trust. **Method:** The study adopts a qualitative, doctrinal, and documentary approach, combining legal analysis of the main European Union regulatory instruments, including the AI Act, the GDPR, and the NIS2 Directive, with a technical perspective on artificial intelligence and cybersecurity risks. **Results:** The article shows that formal legal compliance alone is not sufficient to ensure trustworthy AI in education. Automated assessment systems, learning analytics, proctoring tools, chatbots, large language models, and vulnerable digital infrastructures may create significant risks for students' rights. In response, the article proposes a rights-based Cyber-AI trust architecture structured around three operational layers: the AI system, the educational data flow, and the cybersecurity infrastructure. **Contribution:** The main contribution of the article is to present an operational model that translates legal requirements into practical institutional and technological design principles, including human oversight, data minimization, cybersecurity-by-design, responsible EdTech procurement, continuous auditing, incident-response procedures, and student participation in AI governance. **Conclusion:** AI can support the future of education only if legal, technical, and pedagogical safeguards are integrated into a coherent institutional architecture centered on the student.

Keywords | Artificial Intelligence In Education; Cybersecurity; Students' Rights; GDPR; AI Act; Nis2; Trust Architecture.

GUEST PAPER
Received: 13 may. 2026
Revised: 21 May. 2026
Accepted: 22 May 2026
e-ISSN: 2965-4688

Corresponding Author: Andreea Nicoleta Dragomir – E-mail: andreea.dragomir@ulbsibiu.ro

How to cite this article: Dragomir, A. N., & Bernaschi, O. (2026). AI, Cybersecurity and Students' Rights in EU Digital Education Governance. *Review of Artificial Intelligence in Education*, *7*(i), e097. https://doi.org/10.37497/rev.artif.intell.educ.v7ii.97

RESUMO | Objetivo: Este artigo analisa como a inteligência artificial aplicada à educação superior pode afetar os direitos dos estudantes, especialmente em relação à privacidade, autonomia, igualdade, avaliação justa e confiança institucional. **Metodologia:** A pesquisa adota abordagem qualitativa, doutrinária e documental, combinando análise jurídica dos principais instrumentos regulatórios da União Europeia, como o AI Act, o GDPR e a Diretiva NIS2, com uma perspectiva técnica sobre riscos de inteligência artificial e cibersegurança. **Resultados:** O estudo demonstra que a conformidade jurídica formal não é suficiente para garantir uma IA confiável na educação. Sistemas de avaliação automatizada, learning analytics, proctoring, chatbots, grandes modelos de linguagem e infraestruturas digitais vulneráveis podem gerar riscos relevantes aos direitos dos estudantes. Como resposta, o artigo propõe uma arquitetura de confiança Cyber-AI baseada em direitos, estruturada em três camadas: sistema de IA, fluxo de dados educacionais e infraestrutura de cibersegurança. **Contribuição:** A principal contribuição do artigo é apresentar um modelo operacional que traduz exigências jurídicas em princípios práticos de desenho institucional e tecnológico, incluindo supervisão humana, minimização de dados, cibersegurança desde a concepção, contratação responsável de EdTech, auditoria contínua, resposta a incidentes e participação estudantil na governança da IA. **Conclusão:** A IA pode apoiar o futuro da educação, desde que salvaguardas jurídicas, técnicas e pedagógicas sejam integradas em uma arquitetura institucional centrada no estudante.

Palavras-chave | Inteligência Artificial Na Educação; Cibersegurança; Direitos Dos Estudantes; Gdpr; Ai Act; Nis2; Arquitetura De Confiança.

1 INTRODUCTION

Artificial intelligence is already inside education. It is no longer a future scenario or an abstract technological promise. It is present in learning management systems, adaptive learning platforms, automated assessment tools, plagiarism detection software, proctoring applications, chatbots, student support systems, and learning analytics. In many universities, AI no longer appears as a separate innovation project, but as part of the ordinary digital environment in which students learn, are assessed, communicate and receive institutional support.

This transformation has real value. AI can help teachers identify learning difficulties earlier, personalize educational content, provide faster feedback, support international students, reduce repetitive administrative work, and make large educational systems more responsive (Holmes, Bialik & Fadel, 2019; Wang et al., 2023; Lin, Huang & Lu, 2023). In higher education, where institutions manage thousands of students, courses, exams, platforms, and administrative decisions, AI can offer practical tools to identify patterns that would otherwise remain invisible. Used carefully, it may support inclusion, accessibility and better academic guidance (Halkiopoulos & Gkintoni, 2024).

However, the same technical mechanisms that make AI useful in education also make it risky. AI systems do not personalize learning in a neutral way. They personalize it by collecting data, identifying patterns, generating predictions, and transforming student behavior into measurable signals. A student becomes visible not only through class participation or written work, but also through clicks, logins, response time, writing style, platform activity, chatbot interactions, video feeds, biometric traces and predictive scores. This is where the promise of AI meets the learner's vulnerability.

The risks are not only theoretical. Learning analytics may profile students as "at risk" before a human teacher understands their situation. Automated assessment may convert a technical error into an academic consequence. Proctoring systems may read anxiety, disability-related behavior, poor lighting, or unstable internet connection as suspicious conduct. Chatbots may provide inaccurate

guidance while appearing confident. Large language models may hallucinate, reproduce bias, or offer support of varying quality depending on language, accent, cultural context, or training data. In each case, a technical weakness can become an educational injustice.

For this reason, the central concern of this article is not AI as technology in general, nor compliance as an administrative exercise. The central concern is the student. A student is not merely a user of a platform or a data point inside a model. A student is a person whose privacy, autonomy, dignity, equality, academic progression, and trust in the institution may be affected by how AI systems are designed and deployed. Recent literature has already shown that AI in education raises serious concerns regarding privacy, surveillance, opacity, profiling, discrimination and accountability (Huang, 2023; Klimova, Pikhart & Kacetl, 2023; Pierrès et al., 2024). These concerns become even more acute when AI operates within insecure digital infrastructures.

Cybersecurity is therefore not a secondary issue. In AI-based education, cybersecurity is part of the educational environment itself. AI platforms depend on cloud services, authentication systems, institutional databases, learning management systems, third-party providers and data pipelines. If these elements are vulnerable, the harm may affect not only the availability of a digital service, but also the integrity of assessment, the confidentiality of student data and the fairness of institutional decisions. Universities are attractive targets because they combine open networks with valuable personal, academic, financial and research data (Cheng & Wang, 2022; Lallie et al., 2025). In such a context, ransomware, phishing, data breaches, supply chain compromise, and account hijacking may directly affect students' rights.

The European Union has developed a strong regulatory framework for this environment. The AI Act identifies several educational uses of AI as high-risk, especially systems used for admission, assessment, learning outcomes and exam monitoring (Regulation (EU) 2024/1689). The GDPR protects student data and regulates profiling, transparency, lawful processing and automated decision-making (Regulation (EU) 2016/679). NIS2 strengthens cybersecurity governance, risk management, supply chain security, and incident response (Directive (EU) 2022/2555). Yet the value of these instruments does not lie only in legal compliance. Their real value lies in their ability to shape safer technical architectures.

This article argues that the AI Act, GDPR and NIS2 should not be treated merely as compliance frameworks, but as design parameters for trustworthy AI in education. Their value lies in how they shape safer technical architectures: human-in-the-loop assessment, privacy-preserving data flows, cybersecurity-by-design and institutional accountability centered on the student. In other words, legal rules should be translated into technical and organizational design choices. A high-risk assessment tool should include meaningful human review. A learning analytics platform should minimize data and explain how profiles are generated. An EdTech system should be secure by design, auditable and resilient to foreseeable attacks. A university should know not only what a system promises but also how it protects students.

The contribution of this article is interdisciplinary. It combines a legal analysis of EU digital governance with an AI and cybersecurity perspective. This is important because the problem cannot be solved solely by lawyers or engineers. Legal analysis identifies the rights, duties and accountability gaps. Technical analysis shows how AI systems actually function, fail, expose data, generate bias,

or depend on vulnerable infrastructures. The article therefore proposes a rights-based trust architecture for AI-based education, in which the AI system, the data flow and the cybersecurity environment are treated as interconnected layers of student protection.

The argument is practical. Trustworthy AI in education cannot be built through declarations of ethics alone. It requires design choices, procurement standards, risk assessments, technical safeguards, human oversight, incident response procedures, and student participation. The aim is not to slow down innovation, but to make innovation safe enough to deserve institutional trust. AI can support the future of education, but only if students remain at the center of the architecture.

2 METHODOLOGY

This article uses qualitative, doctrinal, and documentary research methods. Its aim is not to test the performance of a specific AI tool, nor to conduct empirical research in a particular university. Rather, it examines how trustworthy AI in education can be designed when legal duties, technical vulnerabilities and students' rights are considered together. The central objective is to move from a narrow compliance perspective towards an operational model of trust architecture for AI-based education.

The doctrinal component focuses on interpreting the main European Union instruments that shape the governance of AI-based educational systems: the Artificial Intelligence Act, the General Data Protection Regulation, and the NIS2 Directive. These instruments are not analyzed as separate legal regimes, but as regulatory parameters for system design. The AI Act is relevant because several educational AI systems, especially those used for admission, assessment, learning outcomes and examination monitoring, may fall within the category of high-risk AI systems (Regulation (EU) 2024/1689). The GDPR is relevant because AI in education depends on personal data, profiling, transparency, lawful processing and safeguards against automated decision-making (Regulation (EU) 2016/679). NIS2 is relevant because AI-based education depends on secure digital infrastructures, cloud services, institutional networks and third-party providers (Directive (EU) 2022/2555).

The documentary component is based on recent academic literature concerning AI in education, student privacy, learning analytics, automated assessment, cybersecurity in higher education and institutional digital governance. The selected sources were used to identify the main risks that appear when AI systems are deployed in educational environments: privacy loss, predictive profiling, algorithmic bias, opaque outputs, weak human oversight, cybersecurity vulnerabilities and supply-chain risks. Studies on AI in education help explain how learning analytics, chatbots, adaptive systems and assessment tools reshape the educational relationship (Holmes, Bialik & Fadel, 2019; Wang et al., 2023; Lin, Huang & Lu, 2023). Studies on privacy and learning analytics clarify why student data require stronger safeguards than ordinary platform data (Cormack, 2016; Jones, 2019; Karunaratne, 2021). Studies on cybersecurity in higher education show why universities are particularly vulnerable digital environments, due to their open networks, sensitive data, third-party services and large user communities (Cheng & Wang, 2022; Lallie et al., 2025).

The article adopts an interdisciplinary legal-technical perspective. This is necessary because the problem cannot be understood only through legal interpretation. A legal rule becomes meaningful in practice only when translated into system design, procurement requirements, security measures and institutional procedures. For example, the requirement of human oversight under the AI Act is analyzed not only as a legal safeguard, but also as a design requirement for human-in-the-loop assessment. The GDPR principle of data minimization is considered not only a legal obligation but also a technical requirement for privacy-preserving data flows. The NIS2 logic of cybersecurity risk management is examined not only as a regulatory duty but also as a practical requirement for resilient AI infrastructure.

The method, therefore, follows a layered analytical approach. First, the article identifies the technical risk created by an educational AI system. Second, it asks how that risk may affect the student, particularly regarding privacy, equality, autonomy, fair assessment, and access to education. Third, it identifies which EU regulatory instrument provides the relevant safeguard. Fourth, it translates that safeguard into a practical design or governance requirement. This approach allows the article to treat AI Act, GDPR and NIS2 not as the subject of the article, but as the operating parameters of trustworthy AI in education.

The research has two main limitations. First, it does not include original empirical fieldwork with students, teachers, developers, or university administrators. It relies on legal analysis and existing academic literature. Second, it focuses on the European Union framework and does not provide a comparative analysis of other jurisdictions. These limitations are consistent with the article's purpose, which is conceptual and operational: to propose a rights-based trust architecture for AI-based education.

The methodology's contribution is therefore practical. It creates a bridge between legal principles and technical implementation. The article does not ask only whether an AI system is compliant. It asks whether the system is designed to protect the students. This student-centered approach guides the proposed trust architecture model developed in the following sections.

3 THEORETICAL AND TECHNICAL FRAMEWORK: AI IN EDUCATION AND STUDENTS' RIGHTS

Artificial intelligence is often introduced into education through the language of promise. It is said to make learning more personal, assessment more efficient, feedback more immediate and institutional support more responsive. These benefits are real and should not be ignored. AI-based tools can help identify learning difficulties, adapt educational content, support international students, reduce repetitive administrative work and offer new forms of student assistance (Wang et al., 2023; Lin, Huang & Lu, 2023; Halkiopoulos & Gkintoni, 2024).

However, the same mechanisms that make AI useful in education also pose significant risks to students' rights. AI systems do not personalize learning in a neutral or abstract way. They personalize it by collecting, processing and comparing data about students. In this sense, AI in education is not only a pedagogical tool. It is also a data-intensive infrastructure that can observe, classify, predict, and influence student behavior.

The first major risk concerns **privacy and confidentiality**. Educational AI systems may collect data on attendance, grades, learning rhythm, time spent on digital platforms, submitted answers, resources accessed, communication patterns, and interactions with chatbots or learning management systems. In more intrusive cases, AI tools may also process voice, images, facial expressions, biometric indicators, or emotional signals. This expands the traditional boundaries of educational relationships. The student is no longer visible only through participation in class or written assessments, but also through continuous digital traces created within educational platforms (Huang, 2023; Klimova, Pikhart & Kacetl, 2023).

This is particularly sensitive because students are not ordinary consumers of digital services. They often use educational platforms because the institution requires or strongly encourages them to do so. Their ability to refuse data collection may therefore be limited in practice. A student may formally accept the terms of a platform, but this does not always mean that consent is genuinely free, informed, or meaningful. In education, privacy should therefore be understood not only as a question of data management, but as a condition of intellectual freedom and trust.

A second risk concerns **profiling and predictive analytics**. Learning analytics may be used to identify students considered "at risk", to predict academic performance, to recommend interventions, or to classify learning behavior. These tools can be useful when they help teachers offer timely support. Yet they can also create labels that follow students throughout their academic trajectory. A prediction about risk may become a form of institutional expectation. A student may be treated not according to actual performance or context, but according to a probabilistic profile.

The danger is not only that a prediction may be wrong. The deeper problem is that algorithmic profiles may silently influence how students are perceived, supported, monitored, or evaluated. This is especially problematic for international students, students from minority backgrounds, students with disabilities, or students with non-linear educational paths. Wang et al. (2023) show that AI may support international students through translation, chatbots, adaptive learning and analytics, but also that privacy, cultural differences, language barriers and ethical concerns must be carefully addressed. Predictive systems may misinterpret cultural or linguistic differences, such as a lack of engagement, low performance, or risk.

A third risk is **algorithmic bias and discrimination**. AI systems learn from data, and educational data often reflect existing inequalities. If the data used to train or calibrate a system are incomplete, historically biased, or insufficiently representative, the system may reproduce and amplify those inequalities. Automated assessment tools may disadvantage students who express knowledge in a different linguistic style. Chatbots may offer lower-quality support to students whose accents, dialects, or cultural references are underrepresented. Learning analytics may misinterpret the behavior of students who work part-time, have caregiving responsibilities, face disability-related barriers, or use digital resources differently.

The literature on AI in e-learning confirms this tension. AI can improve student engagement, performance and motivation, but bias and discrimination remain central challenges, especially where adaptive assessment and personalized learning are based on complex learner profiles (Halkiopoulos & Gkintoni, 2024). The risk is therefore not merely technical. It directly concerns equality, non-discrimination and fair access to educational opportunities.

This concern becomes even stronger in relation to students with disabilities. AI can support accessibility if it is designed with diverse needs in mind. It can adapt pace, format, feedback and learning routes. But if disability is treated as an anomaly in the data, AI may produce exclusion rather than inclusion. Pierrès et al. (2024) warn that students with disabilities may be treated as outliers by AI systems and may face risks of bias, privacy risks, errors, unrealistic expectations, and exclusion from system design. In education, fairness cannot mean forcing all students into the same algorithmic model. It must mean recognizing difference without turning difference into disadvantage.

A fourth risk concerns **surveillance**. Some educational AI systems do not merely support learning; they monitor it. Proctoring tools, engagement analytics, behavioral tracking, and emotion-detection technologies may create a permanent sense of being observed. This can change the educational atmosphere. Students may feel that every hesitation, pause, click, absence, question, or interaction is being recorded and interpreted. Such an environment may reduce openness, creativity and intellectual risk-taking.

Surveillance is especially problematic because education requires spaces where students can explore ideas, make mistakes and develop independent judgment. If students believe that they are constantly monitored, they may adapt their behavior to what they think the system expects. They may avoid controversial questions, limit their use of digital tools, or become anxious about being misclassified. In this sense, AI not only observes but also learns. It may also reshape the way students learn and express themselves.

A fifth risk is **opacity**. Many AI systems operate in ways that are difficult for students, teachers and even institutions to understand. A student may receive a score, a recommendation, a warning, or a risk classification without knowing which data were used, how the result was produced, or how it can be challenged. This opacity weakens accountability. It also weakens the educational value of feedback. Feedback that cannot be understood is not truly educational; it becomes an administrative output.

Opacity is particularly serious when AI contributes to assessment, academic progression, admissions or disciplinary suspicion. If a system flags a student as at risk of failure, detects alleged cheating or generates an automated evaluation, the student should be able to understand the basis of that outcome. Teachers should also be able to critically assess the system's recommendations rather than simply trusting them. Lin, Huang and Lu (2023) emphasize the importance of trust and explainability in AI-supported education, especially where intelligent tutoring systems and learning analytics are used to guide educational decisions.

A sixth risk concerns **responsibility and accountability**. In traditional education, responsibility is relatively visible. Teachers, examination boards, administrative bodies and institutions make decisions and can be asked to justify them. In AI-mediated education, responsibility may become distributed among many actors: the university, the teacher, the platform provider, the software developer, the cloud provider, the data processor and sometimes a public authority that recommends or funds the system. When harm occurs, it may be unclear who should answer for it.

This diffusion of responsibility is one of the most important ethical problems of AI in education. If an automated assessment tool produces an unfair result, who is responsible? If a proctoring system wrongly flags a student for misconduct, who must correct the error? If a chatbot gives misleading

academic guidance, who bears responsibility for the consequences? Huang (2023) correctly links educational AI to the need for responsibility, accountability, impact assessment and due diligence throughout the lifecycle of AI systems.

These risks show that AI in education cannot be governed solely by general ethical declarations. It requires concrete safeguards. Privacy by design, data minimization, transparency, explainability, bias audits, human oversight, clear complaint mechanisms, and student participation should become normal elements of AI governance in education. Students and teachers should not be treated as passive users of AI systems. They should be involved in decisions about which systems are adopted, what data are collected, which purposes are legitimate, and what limits should be respected.

The ethical challenge is therefore not to reject AI, but to humanize its use. AI should support the educational relationship, not replace it. It should strengthen student autonomy, not reduce students to behavioral profiles. It should help teachers understand learning needs, not transform teaching into automated surveillance. It should improve access to education, not reproduce inequalities through technical systems that appear neutral but are socially consequential.

In this article, the protection of students' rights is understood as the foundation of trustworthy AI in education. Privacy, equality, autonomy, transparency and accountability are not external constraints on innovation. They are the conditions under which educational innovation remains legitimate. AI may have a meaningful role in the future of education, but only if students remain visible as people, not merely as data points, risk scores, or learning profiles.

4 RESULTS AND DISCUSSION: FROM COMPLIANCE TO TRUST ARCHITECTURE

The analysis developed in the previous sections shows that trustworthy AI in education cannot be achieved by legal compliance alone. A university may formally refer to the AI Act, GDPR and NIS2, but students will only be protected if these instruments are translated into the actual design, procurement, deployment and monitoring of AI systems. In this sense, the regulatory framework should not be treated as an external administrative burden. It should function as a design system for safe educational AI.

This section, therefore, moves from legal description to operational design. The central argument is that the AI Act, GDPR and NIS2 should be understood as three technical and institutional layers of trust. The AI Act requires safe and accountable AI systems, especially where admission, assessment, progression or proctoring are involved (Regulation (EU) 2024/1689). The GDPR requires lawful, transparent, and proportionate data processing, especially where learning analytics, profiling, and automated decision-making are used (Regulation (EU) 2016/679). NIS2 contributes to the cybersecurity logic: resilient infrastructure, incident response, supplier security, and cyber hygiene (Directive (EU) 2022/2555).

The practical question is therefore simple: how should an educational AI system be built to protect the student? The answer proposed here is a rights-based Cyber-AI trust architecture. Its purpose is not to stop educational innovation, but to make it reliable enough to deserve institutional trust.

4.1 Human-in-the-loop assessment

The first design principle is meaningful human oversight. In education, this is especially important for AI systems used in admission, assessment, academic progression, plagiarism detection, risk scoring and proctoring, because these uses may directly affect students' academic opportunities and fundamental rights (Regulation (EU) 2024/1689; Colonna, 2025).

From a technical perspective, human-in-the-loop design means that the AI system should support human judgment rather than replace it. A score, alert, risk profile, or recommendation should not be considered final simply because it was generated by a model. The system should be designed so that a teacher, examiner, admissions officer, or academic board can understand the output, review the relevant data, consider the context, and override the recommendation where necessary.

This is particularly important in automated assessment. AI tools may generate grades, evaluate essays, detect plagiarism or provide feedback. Such tools may be useful, but they can also misunderstand originality, linguistic variation, disability-related writing patterns, or culturally specific expression. The literature on AI-based assessment confirms that AI can improve efficiency and feedback, but it cannot replace pedagogical judgment (González-Calatayud, Prendes-Espinosa & Roig-Vila, 2021; Owan et al., 2023). Assessment is not only a measurement. It is an interpretation.

The same applies to proctoring. AI proctoring tools may detect gaze movement, background noise, a missing face, browser activity, or unusual behavior. However, these signals do not automatically prove misconduct. A student may look away because of anxiety, disability, technical interruption, environmental constraints or simple human movement. If the system flags such behavior without meaningful review, a technical suspicion may become an academic accusation, with particular risks for students whose behavior does not match standardized patterns assumed by the system (Pierrès et al., 2024).

The AI Act's high-risk logic supports this design requirement. Educational systems used for access, assessment, learning outcomes and exam monitoring are treated as high-risk because they may affect fundamental rights and life opportunities (Regulation (EU) 2024/1689). But the legal label is only the starting point. In practice, high-risk classification should lead to concrete technical architecture: explainable outputs, audit trails, review interfaces, escalation procedures and documented human intervention.

A human-in-the-loop system should therefore include at least four elements. First, the AI output must be visible and understandable to the responsible human reviewer. Second, the reviewer must receive enough contextual information to assess whether the output is reliable. Third, the reviewer must have the authority to change or disregard the output. Fourth, the student must have a route to contest the decision. Without these elements, human oversight becomes only a procedural decoration.

This is also a matter of institutional culture. Teachers should not be placed in a position where they feel obliged to accept the algorithmic result because the system appears objective or because the vendor presents it as scientifically neutral. Colonna (2025) correctly notes that the role of teachers in AI-mediated education remains one of the most important unresolved issues. A trustworthy educational AI system must preserve the teacher's professional responsibility while also giving the student a meaningful opportunity to be heard.

4.2 Privacy-preserving educational data flows

The second design principle concerns how student data flows through educational AI systems. AI in education is data intensive. Learning analytics, adaptive platforms, AI tutors, early warning systems and automated assessment tools all depend on the collection and processing of student data. The problem is not only that data are collected. The problem is that data are transformed into profiles, predictions and decisions that may affect the student.

GDPR should therefore be understood as a design framework for educational data flows. Its principles of lawfulness, fairness, transparency, purpose limitation, data minimization, and accountability should be translated into technical choices from the beginning of system development and procurement (Regulation (EU) 2016/679). In practical terms, this means that an educational AI platform should collect only the data necessary for a clearly defined educational purpose. It should avoid unnecessary behavioral tracking, unnecessary biometric processing, and excessive long-term storage.

Data minimization is particularly important. Many AI systems perform better when they receive more data. But in education, "more data" does not automatically mean "better education". A platform may not need to collect every click, pause, facial expression, or behavioral signal in order to support learning. A privacy-preserving architecture should start with a different question: what is the minimum amount of data needed to achieve educational purposes safely?

Where possible, institutions should use anonymization, pseudonymization, aggregation, and privacy-preserving analytics. In more advanced settings, technical approaches such as federated learning or differential privacy may also reduce exposure by limiting the centralization or identifiability of student data, although their use requires specialized expertise and careful implementation.

Learning analytics provides a useful example. A university may want to identify students who need support. This can be legitimate. But the system should avoid turning students into permanent risk profiles. Cormack (2016) proposes a useful distinction between pattern-finding and intervention. General analysis may help institutions understand educational patterns, but individual interventions should be more carefully justified and include student-facing safeguards. This distinction is useful because it allows institutions to benefit from analytics without making intrusive profiling the default mode of education.

Transparency must also become technical. Students should not only receive a long privacy policy. They should be able to understand, in plain language, what data is collected, for what purpose, who has access to it, whether profiling occurs, and how analytics may affect them. Jones (2019) argues that privacy dashboards can help students exercise more meaningful control over learning analytics. Such dashboards could show the categories of data used, the purpose of processing, the actors involved and available choices. They could also allow students to manage preferences where consent or optional participation is appropriate.

This is especially important because consent in education is often fragile. Students may agree to the platform's terms because they need access to a course, an exam, or a university service. Their consent may be formally valid in some situations, but it may not always reflect real autonomy. Jones et al. (2020) show that students often lack awareness of learning analytics practices, yet express

strong concerns about being constantly tracked. This confirms the need for transparency tools that are understandable and usable.

Privacy-preserving data flows also reduce cybersecurity harm. If less data are collected, less data can be leaked. If data are properly encrypted, segmented and access-controlled, a breach becomes less damaging. If student profiles are not stored indefinitely, the long-term risk decreases. In this sense, GDPR-inspired data minimization and NIS2-inspired cybersecurity resilience support each other.

The aim is not to eliminate data from education. The aim is to ensure that data serve the student, rather than silently defining the student. Data should help teachers support learning. It should not become an invisible infrastructure of surveillance, ranking, or behavioral control.

4.3 Cybersecurity-by-design and resilient infrastructure

The third design principle is cybersecurity-by-design. AI systems in education are not merely algorithms. They operate through networks, accounts, cloud services, databases, APIs, vendors, authentication systems and learning platforms. If these infrastructures are insecure, the AI system cannot be trustworthy.

Universities are particularly exposed because they combine open networks with sensitive data. They must support access for students, staff, researchers, visitors and partners, often across many devices and locations. This openness is part of the academic mission, but it also increases the attack surface. Cheng and Wang (2022) show that higher education institutions require system-wide cybersecurity strategies, including governance, awareness, encryption, risk management and policy mechanisms. Lallie et al. (2025) similarly identify cyberattacks and vulnerabilities in the university sector as a serious and growing problem.

For AI-based education, the cybersecurity question becomes more specific: how can the institution protect the AI system, the data used by the system and the educational decisions influenced by that system? A ransomware attack may interrupt exams. A phishing attack may compromise teacher or student accounts. A breach of a learning analytics platform may expose sensitive student profiles. A supply-chain attack against an EdTech provider may affect multiple universities at once. A manipulated input may distort AI outputs.

The NIS2 Directive provides useful governance logic for this environment. Even where a university is not directly classified as an essential or important entity, NIS2 points to practical measures that are relevant to educational AI: risk management, incident handling, business continuity, supply-chain security, cyber hygiene, encryption, access control and multi-factor authentication (Directive (EU) 2022/2555). These measures should be treated as design requirements, not as optional IT practices.

This cybersecurity logic becomes even more important when educational platforms integrate large language models, AI tutors, or student-facing chatbots. In such systems, the risk is not limited to unauthorized access or service interruption. A prompt injection attack may manipulate the chatbot's behavior, causing it to reveal information, ignore institutional instructions, or provide

unsafe academic guidance. Sensitive information disclosure may occur when a model exposes personal data, internal prompts, uploaded documents, or confidential student information. Data and model poisoning may affect the integrity of an assessment or tutoring system if training, fine-tuning or retrieval data are manipulated. Supply-chain vulnerabilities may also arise when universities rely on third-party models, plugins, APIs, datasets or cloud components that are not sufficiently verified. The OWASP Top 10 for Large Language Model Applications identifies prompt injection, sensitive information disclosure, supply chain risks, data and model poisoning, and improper output handling as major risks for LLM-based applications (OWASP, 2025). In education, these vulnerabilities are particularly serious because a compromised AI tool may not only expose data, but also distort feedback, recommendations, assessment support or the trust relationship between the student and the institution.

Multi-factor authentication should be standard for systems that process student data or influence academic decisions. Encryption should protect data at rest and in transit. Access control should follow the principle of least privilege. Logs should be maintained in a way that allows investigation without creating unnecessary surveillance. Backups should be tested, not merely created. Incident-response procedures should include academic consequences, not only technical recovery.

For this reason, AI systems used in education should be tested not only for functionality, but also for abuse cases. What happens if a student inputs malicious prompts? What happens if an attacker compromises a teacher's account? What happens if the vendor's API is unavailable during exams? What happens if assessment data are altered? What happens if the model generates inaccurate or harmful recommendations? These are not hypothetical questions. They are part of responsible deployment.

Cyber hygiene remains essential. Many attacks succeed because users click on malicious links, reuse passwords, ignore updates, or mishandle data. Universities should therefore train students, teachers and administrative staff in practical cybersecurity. The goal is not to transform everyone into a technical specialist. The goal is to reduce predictable human vulnerabilities and create a culture in which suspicious activity is reported early.

Cybersecurity is often presented as a defensive function. In education, it should be understood more positively: it protects the conditions of learning. It protects access, continuity, confidentiality, assessment integrity and trust. Without cybersecurity, educational AI remains fragile, regardless of its legal documentation.

4.4 EdTech procurement as a governance moment

The fourth design principle concerns procurement. In many universities, AI systems are not built internally. They are purchased, licensed, or integrated from external EdTech providers. This makes procurement one of the most important governance moments, especially because universities remain responsible for the educational and data-protection consequences of systems deployed within their institutional environments (Jones, 2019; Cheng & Wang, 2022).

A trustworthy procurement process should simultaneously evaluate educational value, legal compliance, technical security, accessibility, and accountability. If these dimensions are separated,

the institution may adopt a tool that appears pedagogically useful but creates serious risks in practice. For example, a proctoring tool may promise exam integrity but rely on intrusive surveillance. A chatbot may promise student support but provide inaccurate advice. A learning analytics platform may promise early intervention but create opaque risk profiles. A vendor may promise AI innovation but fail to provide strong security guarantees.

The institution should therefore ask concrete questions before adoption. What data does the system collect? Is every category of data necessary? Where are the data stored? Are they encrypted? Who has access? Are subcontractors involved? Can the system explain its outputs? Can teachers override recommendations? Can students contest decisions? Is the system accessible for students with disabilities? Has the provider tested for bias? What happens after a cyber incident? How quickly must the provider notify the institution?

These questions transform legal compliance into procurement criteria. AI Act requirements become questions about risk management, transparency, human oversight and robustness. GDPR requirements become questions about data minimization, lawful basis, retention, student rights and processing roles. NIS2 logic translates into questions about supplier security, incident response, business continuity, and cyber hygiene.

Vendor risk assessment is especially important because responsibility may be distributed across several actors: the university, platform provider, cloud provider, data processor, model developer and authentication provider. Students, however, usually experience the system as part of the university. If harm occurs, they will not distinguish easily between the institution and the vendor. The university, therefore, remains responsible for selecting systems appropriate for an educational environment (Regulation (EU) 2016/679; Directive (EU) 2022/2555).

Procurement should also include exit conditions. A university should know how to stop using a system if it becomes unsafe, inaccurate, discriminatory or insecure. It should be able to retrieve data, delete unnecessary records, migrate services and notify students. Without exit planning, the institution may become dependent on a vendor even when risks emerge.

This approach gives strong visibility to the technical expertise of AI and cybersecurity specialists. They should be involved before adoption, not only after deployment. Too often, legal review, IT security and pedagogical evaluation are conducted separately or too late. A trust architecture requires them to work together at the procurement stage.

4.5 The proposed Cyber-AI governance workflow

The proposed model can be expressed as an operational workflow (González-Calatayud, Prendes-Espinosa & Roig-Vila, 2021; Jones, 2019; Cheng & Wang, 2022):

Risk mapping → Vendor selection → Impact assessment → Technical implementation → Human oversight → Continuous monitoring → Incident response

This workflow is intended to help universities move from abstract compliance to practical trust architecture.

The first step is risk mapping. Before adopting an AI system, the institution should identify what the system does, which educational process it affects, what data it uses, who is affected and what harm may occur. The central question is always student-centered: how could this system affect privacy, equality, autonomy, fair assessment, or access to education?

The second step is vendor selection. Providers should not be selected only on the basis of functionality, cost or innovation. They should be evaluated for explainability, security, data protection, accessibility, auditability, incident-response capacity and contractual accountability. If a vendor cannot explain how the system protects students, the system should not be deployed in a sensitive educational context.

The third step is impact assessment. For sensitive or high-risk uses, institutions should combine AI risk assessment, data protection impact assessment and cybersecurity assessment. These should not be three disconnected documents. They should form one governance process. The AI assessment asks whether the system is fair, explainable and overseen. The data protection assessment asks whether the data processing is lawful, limited and transparent. The cybersecurity assessment asks whether the system and providers are resilient.

The fourth step is technical implementation. This includes privacy-preserving data flows, secure authentication, encryption, role-based access, logging, backups, bias testing, accessibility testing and integration with institutional procedures. Implementation should also include documentation for teachers and understandable information for students.

The fifth step is human oversight. The institution must assign real responsibility to human decision-makers. A teacher, examiner or administrator must be able to review AI outputs, question them and correct them. Students must know how to contest decisions or request an explanation.

The sixth step is continuous monitoring. AI systems change over time. Student populations change. Vendors update models. Cyber threats evolve. A system that was acceptable at deployment may become risky later. Periodic audits should therefore examine accuracy, bias, security, data protection, accessibility and educational impact.

The seventh step is incident response. If something goes wrong, the institution must respond not only technically, but also educationally. If an AI assessment system is compromised, the institution must consider whether grades or academic opportunities were affected. If student data is leaked, it must communicate clearly and provide remedies. If a model produces discriminatory outputs, it must stop, review and correct the system.

This workflow is the main contribution of the article. It places the student at the center and translates EU digital regulation into practical design and governance requirements. It also shows that trustworthy AI in education is not a single product feature. It is a continuous institutional process.

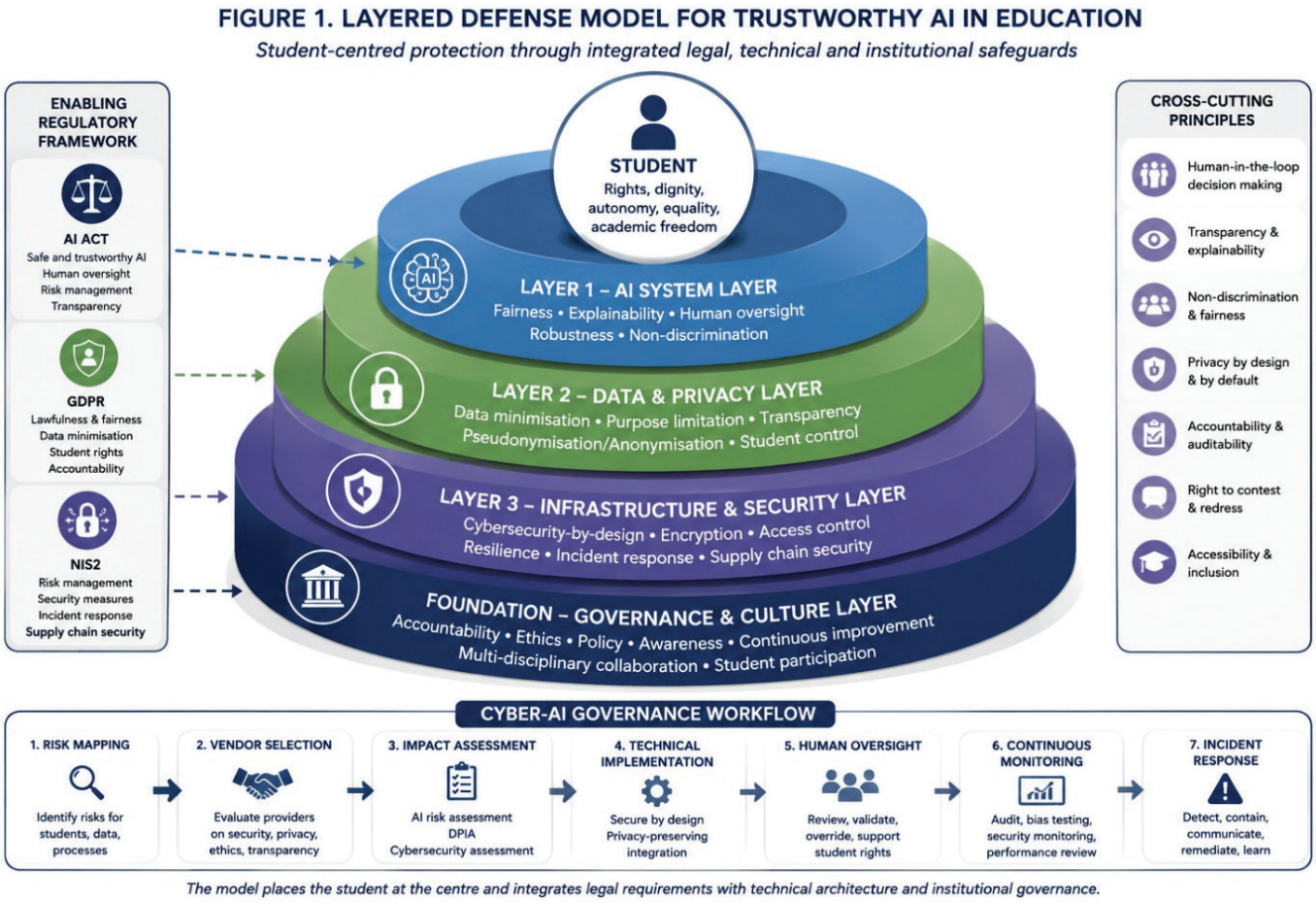

Figure 1. Layered Defense Model for Trustworthy AI in Education

Source: Authors' own elaboration, based on the conceptual framework developed in this article

The layered model can be summarized as follows: the student and student rights form the core; the AI system is the first layer, governed by fairness, explainability and human oversight; the data flow is the second layer, governed by minimization, consent, encryption and student control; the infrastructure is the third layer, governed by resilience, incident response and supply-chain security (Regulation (EU) 2024/1689; Regulation (EU) 2016/679; Directive (EU) 2022/2555). A failure in any layer may reach the student. This is why the lawyers must work together.

The shift from compliance to trust architecture is therefore essential. Compliance asks whether the institution can show that it respected a rule. Trust architecture asks whether the system is designed to protect the student. In AI in education, this second question matters most.

5 CONCLUSION AND RECOMMENDATIONS

Artificial intelligence is becoming part of the ordinary infrastructure of education. It supports learning platforms, assessment tools, proctoring systems, chatbots, student-support services and learning analytics. This development should not be approached with fear, but it should also not be accepted with naïve enthusiasm. AI can help universities become more responsive, more inclusive

and more efficient, but it can also create new vulnerabilities when students are transformed into data profiles, when automated systems influence academic decisions, or when educational platforms are deployed without sufficient cybersecurity safeguards (Holmes, Bialik & Fadel, 2019; Wang et al., 2023; Huang, 2023).

The main argument of this article is that trustworthy AI in education cannot be built through formal compliance alone. The AI Act, GDPR and NIS2 are essential, but their value does not lie only in the fact that they impose legal obligations. Their deeper value lies in the design discipline they can introduce into educational AI systems. The AI Act points to fairness, explainability and human oversight; the GDPR requires privacy-preserving data flows and meaningful student control; NIS2 brings the logic of cybersecurity-by-design, resilience, incident response and supply-chain responsibility (Regulation (EU) 2024/1689; Regulation (EU) 2016/679; Directive (EU) 2022/2555).

The student must remain the center of this architecture. A student is not only a user of an application, a data subject under GDPR or a person affected by a high-risk AI system. A student is a learner whose academic path, confidence, autonomy and trust in the institution may be shaped by technical decisions that are often invisible. A biased score, a false proctoring alert, a leaked profile, an opaque risk prediction, or a disrupted digital exam can have real educational consequences. For this reason, the protection of students' rights must be translated into technical and institutional safeguards (Jones, 2019; Jones et al., 2020; Pierrès et al., 2024).

First, the European Union should support the development of sector-specific guidelines for AI in education. General rules are necessary, but they are not always sufficient for universities, schools and EdTech providers. Educational AI has its own specific risks: automated assessment, learning analytics, student profiling, AI tutoring, proctoring, adaptive learning and student-support chatbots. EU-level guidance should therefore explain how AI Act obligations, GDPR principles and cybersecurity requirements interact in educational settings. Such guidance should be practical rather than merely declaratory (European Commission, 2022; Panagopoulou, Parpoula & Karpouzis, 2025).

Second, institutions should combine the Educational Fundamental Rights Impact Assessment, the Data Protection Impact Assessment and the cybersecurity assessment into a single governance process. These assessments should not be separate documents prepared by different offices without real communication. In AI-based education, the risks are connected. A system may be high-risk because it affects assessment, processes personal data for profiling, and depends on vulnerable cloud or vendor infrastructure. A combined assessment would ask three questions: which rights may be affected, what data are processed, and how secure the system and its infrastructure are (Cormack, 2016; Karunaratne, 2021; Colonna, 2025).

Third, cybersecurity-by-design should become standard in EdTech procurement. Universities should not adopt AI platforms only because they promise innovation, efficiency or attractive learning analytics. They should ask how the system protects students. Procurement should include questions about encryption, access control, multi-factor authentication, logging, backup, vulnerability management, data retention, subcontractors, breach notification and incident response. The selection of an AI platform is not only a technical purchase. It is a governance decision that may affect student privacy, academic fairness and institutional trust (Cheng & Wang, 2022; Lallie et al., 2025).

Fourth, human-in-the-loop design should be mandatory for AI systems used in admission, assessment, academic progression and proctoring. Human oversight must not be reduced to a formal validation of an algorithmic result. A teacher, examiner or academic officer should be able to understand the system's output, question it, correct it and explain the final decision to the student. This is particularly important where AI systems generate grades, detect alleged misconduct, rank applicants or identify students as "at risk" (González-Calatayud, Prendes-Espinosa & Roig-Vila, 2021; Colonna, 2025).

Fifth, students should be given practical tools to understand and control how their data is used. Privacy dashboards are one possible solution. They could show students what data is collected, why it is processed, who has access to it, whether profiling is involved, and what choices are available. Such tools would not solve every problem, but they would make transparency more concrete. Students should not have to read lengthy, technical privacy notices to understand how an educational platform uses their data (Jones, 2019; Jones et al., 2020).

Sixth, educational AI systems should be subject to periodic audit. Audit should not focus only on technical performance. It should also examine bias, explainability, cybersecurity, accessibility, data governance, human oversight and student complaint mechanisms. AI systems change over time. Vendors update models, student populations change, new integrations are added and cyber threats evolve. A system that was acceptable at deployment may become risky later. Periodic audit is therefore necessary to maintain trust, not only to prove initial compliance (Huang, 2023; Pierrès et al., 2024).

Seventh, incident-response procedures must include academic remedies, not only IT recovery. If a ransomware attack blocks access to exams, the institution must consider the academic consequences. If learning analytics data are exposed, affected students must be informed clearly. If an AI assessment tool is compromised, grades or academic decisions may need to be reviewed. If a proctoring system incorrectly flags students due to a technical error, there must be a mechanism for correction. Incident response in education cannot stop at restoring servers. It must also restore fairness (Directive (EU) 2022/2555; Cheng & Wang, 2022).

Eighth, students should participate in the governance of AI systems used in education. They should not be treated as passive recipients of digital innovation. Their concerns about surveillance, profiling, fairness, accessibility and data protection should be considered before systems are adopted, not only after harm occurs. Student participation may take different forms: consultation before procurement, representation in digital governance committees, feedback mechanisms, complaint channels and involvement in the evaluation of AI tools (Jones et al., 2020; Li, Sun, Schaub & Brooks, 2022).

These recommendations point to a broader change in institutional culture. AI in education should not be governed only by legal departments, IT offices or external vendors. It requires collaboration between teachers, AI specialists, cybersecurity experts, data protection officers, administrators and students. The legal framework identifies the rights and obligations. AI expertise explains how systems function and fail. Cybersecurity expertise reveals where infrastructure is vulnerable. Educational expertise ensures that technology remains connected to learning, not only to automation.

The future of AI in education should be built around a trust architecture. This means designing systems in which legal, technical, and pedagogical safeguards work together. A trustworthy system is not simply one that performs well. It is one that can be explained, challenged, secured, audited and corrected. It is one that supports teachers rather than replacing them, protects data rather than exploiting it, and treats students as people rather than as behavioral profiles.

The contribution of this article is therefore practical and conceptual. It proposes a shift from compliance to architecture. Instead of asking only whether an institution can demonstrate formal compliance with the AI Act, GDPR, or NIS2, the more important question is whether the AI system is designed to protect the student. AI can improve education, but only if its implementation respects the human purpose of education: to help students learn without compromising their privacy, autonomy, equality or trust.

REFERENCES

Cheng, E. C. K., & Wang, T. (2022). *Institutional strategies for cybersecurity in higher education institutions*. *Information, 13*(4), 192. https://doi.org/10.3390/info13040192

Colonna, L. (2025). *Artificial Intelligence in Education (AIED): Towards more effective regulation*. *European Journal of Risk Regulation, 17*(1), 161–181. https://doi.org/10.1017/err.2025.10039

Cormack, A. (2016). *A data protection framework for learning analytics*. *Journal of Learning Analytics, 3*(1), 91–106. https://doi.org/10.18608/jla.2016.31.6

Directive (EU) 2022/2555 of the European Parliament and of the Council. (2022). *Directive on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS2 Directive)*. *Official Journal of the European Union, L 333*, 80–152. https://eur-lex.europa.eu/eli/dir/2022/2555/oj

European Commission. (2022). *Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators*. Publications Office of the European Union. https://data.europa.eu/doi/10.2766/153756

González-Calatayud, V., Prendes-Espinosa, P., & Roig-Vila, R. (2021). *Artificial Intelligence for student assessment: A systematic review*. *Applied Sciences, 11*(12), 5467. https://doi.org/10.3390/app11125467

Halkiopoulos, C., & Gkintoni, E. (2024). *Leveraging AI in e-learning: Personalized learning and adaptive assessment through cognitive neuropsychology—A systematic analysis*. *Electronics, 13*(18), 3762. https://doi.org/10.3390/electronics13183762

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

Huang, L. (2023). *Ethics of artificial intelligence in education: Student privacy and data protection*. *Science Insights Education Frontiers, 16*(2), 2577–2587. https://doi.org/10.15354/sief.23.re202

Jones, K. M. L. (2019). *Learning analytics and higher education: A proposed model for establishing informed consent mechanisms to promote student privacy and autonomy*. *International Journal of Educational Technology in Higher Education, 16*, 24. https://doi.org/10.1186/s41239-019-0155-0

Jones, K. M. L., Asher, A., Goben, A., Perry, M. R., Salo, D., Briney, K. A., & Robertshaw, M. B. (2020). *"We're being tracked at all times": Student perspectives of their privacy in relation to learning analytics in higher education*. *Journal of the Association for Information Science and Technology, 71*(9), 1044–1059. https://doi.org/10.1002/asi.24358

Karunaratne, T. (2021). *For learning analytics to be sustainable under GDPR—Consequences and way forward*. *Sustainability, 13*(20), 11524. https://doi.org/10.3390/su132011524

Klimova, B., Pikhart, M., & Kacetl, J. (2023). *Ethical issues of the use of AI-driven mobile apps for education*. *Frontiers in Public Health, 10*, 1118116. https://doi.org/10.3389/fpubh.2022.1118116

Lallie, H. S., Thompson, A., Titis, E., & Stephens, P. (2025). *Analysing cyber attacks and cyber security vulnerabilities in the university sector. Computers, 14*(2), 49. https://doi.org/10.3390/computers14020049

Li, W., Sun, K., Schaub, F., & Brooks, C. (2022). *Disparities in students' propensity to consent to learning analytics*. *International Journal of Artificial Intelligence in Education, 32*, 564–608. https://doi.org/10.1007/s40593-021-00254-2

Lin, C.-C., Huang, A. Y. Q., & Lu, O. H. T. (2023). *Artificial intelligence in intelligent tutoring systems toward sustainable education: A systematic review*. *Smart Learning Environments, 10*, 41. https://doi.org/10.1186/s40561-023-00260-y

OWASP. (2025). *OWASP Top 10 for Large Language Model Applications 2025*. OWASP Foundation. https://genai.owasp.org/llm-top-10/

Owan, V. J., Abang, K. B., Idika, D. O., Etta, E. O., & Bassey, B. A. (2023). *Exploring the potential of artificial intelligence tools in educational measurement and assessment*. *EURASIA Journal of Mathematics, Science and Technology Education, 19*(8), em2307. https://doi.org/10.29333/ejmste/13428

Panagopoulou, F., Parpoula, C., & Karpouzis, K. (2025). *Legal perspectives on AI and the right to digital literacy in education*. *Frontiers in Computer Science, 7*, 1692268. https://doi.org/10.3389/fcomp.2025.1692268

Pierrès, O., Christen, M., Schmitt-Koopmann, F., & Darvishy, A. (2024). *Could the use of AI in higher education hinder students with disabilities? A scoping review*. *IEEE Access, 12*, 27810–27839. https://doi.org/10.1109/ACCESS.2024.3365368

Regulation (EU) 2016/679 of the European Parliament and of the Council. (2016). *Regulation on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*. *Official Journal of the European Union, L 119*, 1–88. https://eur-lex.europa.eu/eli/reg/2016/679/oj

Regulation (EU) 2024/1689 of the European Parliament and of the Council. (2024). *Regulation laying down harmonised rules on artificial intelligence and amending Regulations and Directives in certain Union legislative acts (Artificial Intelligence Act)*. *Official Journal of the European Union, L 2024/1689*. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Wang, T., Lund, B. D., Marengo, A., Pagano, A., Mannuru, N. R., Teel, Z. A., & Pange, J. (2023). *Exploring the potential impact of artificial intelligence (AI) on international students in higher education: Generative AI, chatbots, analytics, and international student success*. *Applied Sciences, 13*(11), 6716. https://doi.org/10.3390/app13116716