Mining for Knowledge, Not Trouble: GDPR's Impact on Educational Data Mining

Aytaj Ismayilzada
University of Jyväskylä
Faculty of Information Technology
P.O. Box 35, FI-40014
Jyväskylä, Finland
ayisisma@jyu.fi

Ayaz Karimov
University of Jyväskylä
Faculty of Information Technology
P.O. Box 35, FI-40014
Jyväskylä, Finland
ayaz.karimov@jyu.fi

Mirka Saarela
University of Jyväskylä
Faculty of Information Technology
P.O. Box 35, FI-40014
Jyväskylä, Finland
mirka.saarela@jyu.fi

ABSTRACT

Educational Data Mining (EDM) enables institutions to analyze student data to improve learning outcomes, but this potential must be balanced with data privacy regulations such as the European Union's General Data Protection Regulation (GDPR), which primarily applies to EU residents but influences global data practices. This paper examines how GDPR shapes EDM practices by analyzing key principles, such as lawfulness, purpose limitation, and data minimization, and the challenges they create in educational contexts. We highlight common tensions between regulatory compliance and innovation, including difficulties with consent management, data retention, and the trade-offs of anonymization techniques. Drawing on selected examples from higher education institutions and digital learning platforms, we show strategies adopted to align with GDPR requirements. Our findings suggest that while institutions are developing adaptive governance models, issues such as consent timing, reidentification risks, and limited resources continue to hinder effective EDM. We conclude by proposing future directions for privacy-preserving analytics, dynamic consent frameworks, and collaborative policymaking to support responsible data use in education.

Keywords

educational data mining, general data protection regulation, learning analytics, student data privacy, educational technology

1. INTRODUCTION

Student data is everywhere: schools track attendance, test scores, and online learning activities, while educational platforms record clicks, study patterns, and progress [1]. All of this information contains valuable information that can improve teaching, personalize learning, and help students succeed [2]. This is the power of EDM, which uses data to make education smarter and more effective.

Aytaj Ismayilzada, Mirka Saarela, and Ayaz Karimov. Mining for Knowledge, Not Trouble: GDPR's Impact on Educational Data Mining. In Caitlin Mills, Giora Alexandron, Davide Taibi, Giosuè Lo Bosco, and Luc Paquette (eds.) Proceedings of the 18th International Conference on Educational Data Mining, Palermo, Italy, July, 2025, pp. 608–612. International Educational Data Mining Society (2025).

https://doi.org/10.5281/zenodo.15870232

However, student data are not only numerical, they represent personal and often sensitive information. Who controls these data? How is it protected? Can students decide what happens to their personal information? These questions have become increasingly urgent as data collection in education expands [3]. In response, the European Union introduced the GDPR in 2018, which granted individuals, including students, greater control over their personal data and established strict rules on how organizations collect, store, and share them [4]. Although GDPR is a European regulation, its influence extends globally, shaping how educational institutions worldwide manage student data. For EDM, GDPR presents both challenges and opportunities: it demands stronger security, transparent data policies, and respect for students' rights, such as access and deletion, while also encouraging more ethical and privacy-conscious data practices [5].

This paper focuses primarily on how educational institutions navigate the operational, technical, and ethical challenges of implementing GDPR-compliant practices within data-driven environments. Although the discussion touches on implications for EDM research, the primary emphasis is on institutional compliance and governance strategies. We adopt a qualitative research approach that combines a review of the literature with selected examples from universities, schools, and digital learning platforms. These examples are drawn from documented efforts to adapt to GDPR and reflect a range of institutional responses in diverse educational contexts.

This work contributes to the EDM community by clarifying how GDPR affects the governance, processing, and use of student data. By analyzing core GDPR principles, identifying compliance challenges and presenting representative examples of institutional adaptation, this research supports educators, researchers, and policymakers seeking to align data-driven innovation with robust privacy protections.

2. FUNDAMENTAL GDPR PRINCIPLES IN EDUCATIONAL DATA MINING

The introduction of GDPR has forced EDM practitioners to re-assess how student data is collected, stored, and analyzed. Although intended to improve privacy and accountability, GDPR's legal restrictions often conflict with the data-intensive methods used in EDM. Institutions must bal-

ance innovation with compliance while interpreting complex and evolving regulatory requirements. This section provides an overview of the key GDPR principles most relevant to EDM. This section outlines the key principles of GDPR that shape data practices in educational settings and explains their conceptual relevance to EDM. Specific implementation challenges, such as consent management, data retention, and institutional compliance, are examined in detail in Section 3.

2.1 Lawfulness, Fairness, and Transparency

The principles of lawfulness, fairness, and transparency are fundamental to the GDPR framework. They require that personal data be collected and processed on a legal basis, that purposes and methods are clearly communicated, and that individuals are treated equitably throughout the data lifecycle [6]. In educational contexts, these principles require institutions to provide clear justifications for data use and ensure that students, parents, or guardians understand how data will be collected, stored and analyzed [7]. This is especially relevant in EDM, where continuous data collection occurs through learning management systems, assessment tools, and behavioral tracking technologies. Ensuring transparency in such environments is complex, particularly when data-driven systems influence learning outcomes, feedback, or academic decisions.

The principle of fairness also highlights the importance of avoiding bias and discrimination in data use. For EDM, this means ensuring that machine learning models do not unintentionally disadvantage specific student groups based on gender, socioeconomic status, or learning style. Although these principles offer ethical safeguards, their implementation in EDM settings requires careful design of consent mechanisms and algorithmic transparency, which are explored in more detail in Section 3.

2.2 Purpose Limitation and Data Minimization

GDPR's principle of purpose limitation requires that personal data be collected for a specific, explicit purpose and not reused for other objectives without additional consent [8]. This presents challenges in EDM, where data is often repurposed for new analyses or refined predictive models over time. Each new application may trigger renewed consent requirements, which complicates long-term data use.

Closely related is the principle of data minimization, which requires collecting only the data necessary for the intended task [9]. In EDM, this raises concerns about the balance of legal compliance with analytic utility, as more comprehensive datasets often improve predictive model performance. Institutions must navigate this tension while also ensuring that data remain proportional to its stated purpose. These considerations affect both system design and policy frameworks, which are further explored in Section 3.

2.3 Data Integrity and Security

GDPR requires that personal data be accurate, up-to-date, securely stored and retained only as long as necessary [10]. These requirements reflect the principles of data integrity, security, and storage limitation. Institutions are expected to implement safeguards to prevent unauthorized access and correct inaccurate records promptly. They must also establish clear retention policies to ensure that data is not held longer than needed.

In the context of EDM, these principles are especially relevant given the dependence on longitudinal data for trend analysis, predictive modeling, and personalized learning [11]. Ensuring that student records remain current and accurate is essential for analytic reliability, while limiting data storage duration may constrain long-term model development. Similarly, the expectation of robust data security presents added complexity for institutions that manage large volumes of sensitive information on multiple digital platforms.

3. CHALLENGES AND RESPONSES

This section presents the key challenges the GDPR introduces to EDM, followed by real-world case examples illustrating how institutions have responded.

3.1 Challenges of GDPR in Educational Data Mining

Although GDPR strengthens data protection, its implementation presents substantial obstacles for EDM [9]. The regulation imposes strict controls on how data are collected, stored, and processed, controls that can conflict with the needs of large-scale iterative data analysis. Institutions must therefore strike a balance between regulatory obligations and research-driven innovation [1].

3.1.1 Data Collection and Consent

One of the most complex challenges is to obtain explicit and informed consent before processing student data [12, 13]. EDM often relies on passively collected data from learning platforms and virtual classrooms. However, GDPR requires active, documented consent, which is logistically difficult at scale, especially in online or hybrid environments. Further complications arise because GDPR allows students to withdraw consent at any time [13]. This creates substantial risks to data integrity, as consent withdrawal can force the deletion of records mid-analysis, disrupting longitudinal studies, and invalidating models. In addition, privacy notices are often lengthy or technical, leading to uninformed or superficial consent. This raises ethical questions about whether consent is truly voluntary or simply a procedural formality [14].

3.1.2 Right to be Forgotten and Data Retention

The GDPR's right to erasure (the "right to be forgotten") directly clashes with EDM's reliance on historical data [8, 9]. Predictive analytics, longitudinal tracking, and personalized learning models are all dependent on long-term data continuity. When students request deletion, institutions must erase records that may be integral to existing datasets. This not only disrupts research continuity, but also degrades model accuracy by introducing gaps [8]. Some institutions have responded by anonymizing data to retain analytical value while complying with the principle of storage limitation. However, anonymization removes the ability to track individual learning progress over time, limiting its usefulness for personalized feedback and adaptive learning strategies.

3.1.3 Anonymization and Pseudonymization Trade-offs

GDPR encourages anonymization and pseudonymization as privacy-preserving techniques [12, 15]. However, these approaches introduce technical and ethical challenges in the context of EDM. Anonymized data permanently remove personal identifiers, which supports compliance but prevents linking records over time [16]. This limits the ability to provide personalized feedback or detect learning patterns between semesters. Pseudonymization, while more flexible, introduces its own burdens: institutions must implement strict access controls, document their processes, and still face the risk of reidentification, especially when combined with external datasets [17].

3.1.4 Compliance Costs and Institutional Burden

GDPR compliance requires significant institutional resources [18, 13]. Universities must invest in secure data infrastructure, employ Data Protection Officers (DPOs), and develop procedures for ongoing compliance, such as conducting data audits and updating privacy policies. These financial and administrative burdens are particularly high for smaller institutions. Even when willing to adopt EDM tools, many are discouraged by the fear of noncompliance and the complexity of managing consent withdrawals, breach notifications, and cross-border data sharing. Paradoxically, a regulation designed to protect student rights can inhibit the adoption of beneficial learning analytics due to institutional risk aversion and uncertainty.

3.2 Institutional Responses to GDPR: Case Examples from Educational Settings

To better understand how different institutions are addressing GDPR compliance, this subsection examines three types of educational settings: higher education institutions, schools and e-learning platforms, and further education colleges. These cases demonstrate how institutions with varying capacities and educational roles have responded to GDPR challenges in data governance. These examples show diverse and practical efforts by educational institutions to align their data practices with core GDPR principles such as data minimization, transparency, and lawful processing, each tailored to their unique operational contexts.

3.2.1 Higher Education Institutions: Developing Comprehensive Data Protection Frameworks

Universities manage large volumes of student data, including academic records, research data, and interactions within learning management systems. Given the complexity of their data processing activities, GDPR compliance has required universities to adopt comprehensive data protection frameworks that integrate data governance policies with enhanced security measures [6, 15]. Many institutions have established centralized data protection units, ensuring that all departments comply with GDPR requirements. These units oversee data audits, consent management protocols, and anonymization techniques to enable compliance while minimizing disruptions to research activities [15]. To address GDPR's requirements for storage limitation and transparency, several European universities have developed automated data retention systems that periodically delete or archive data in accordance with institutional policy [15].

Despite these measures, research data management remains a persistent challenge. Universities conducting long-term educational research must comply with the right to be forgotten while maintaining data integrity [8]. Some institutions have addressed this by pseudonymizing research data, replacing personal identifiers with coded references, allowing data to remain useful without directly exposing personal information. However, pseudonymization still presents privacy risks under GDPR's accountability principle, as reidentification remains possible in specific research contexts.

3.2.2 Schools and E-learning Platforms: Managing GDPR in Digital Learning Environments

Rapid expansion of e-learning platforms, particularly after the COVID-19 pandemic, increased the volume of student data processing in schools. Primary and secondary schools now face significant challenges in ensuring secure data collection and storage, especially when relying on third-party EdTech providers [13, 15].

To comply with GDPR, schools have implemented strict data protection agreements with educational technology providers, which ensure that external providers follow privacy regulations [13]. Some platforms have introduced userfriendly consent management systems, which allowed parents and students to control their data settings and request deletion of non-essential data [15]. A notable example involves a group of European schools that introduced privacy dashboards, where students and parents can view what data is being collected and opt out of non-essential tracking [13]. These dashboards improve transparency and empower families to make informed decisions about data sharing, aligning with GDPR's lawfulness and fairness principles. Their adoption also reflects an institutional shift toward user-centered and proactive privacy practices.

In response to growing cybersecurity concerns, several secondary schools in western Europe have implemented realtime monitoring tools to detect unauthorized access attempts. These tools send alerts if unusual data requests occur, enabling proactive responses to potential security breaches [15]. However, despite these efforts, many schools still lack dedicated Data Protection Officers, which limits their ability to fully enforce GDPR compliance.

3.2.3 Higher Education Colleges: Agile Implementation of GDPR Compliance Measures

Colleges and vocational institutions, often operating with fewer resources than large universities, have adopted agile, phased approaches to GDPR compliance [18, 21]. Rather than implementing all regulations at once, compliance measures are integrated into existing data systems. An example is a network of technical colleges that introduced tiered access controls, ensuring that staff only access the data necessary for their roles [18]. Instructors can view student performance data without accessing personal contact details, while administrative staff manage enrollment records without viewing academic data. This approach supports GDPR's data minimization principle while enhancing data security [21].

Operating with limited budgets and fewer specialized per-

Table 1: Overview of GDPR's Impact on Educational Data Mining

Aspect	Key principles (GDPR guidelines)	Challenges (implementation issues)
Data collection and consent	Lawfulness, fairness, transparency [6, 7]	Obtaining informed consent is complex in large-scale research. Many students do not fully understand privacy policies, and consent withdrawals disrupt datasets [12, 14, 13].
Data usage and minimization	Purpose limitation, data minimization [12, 6, 8]	Strict minimization reduces model accuracy, affecting learning analytics. Purpose limitations restrict AI-driven updates without repeated consent [19, 9].
Data protection and security	Confidentiality, integrity [5, 15]	High cybersecurity costs and lack of resources make student data vulnerable. Compliance requirements add administrative burdens [15, 20, 13].
Data retention and right to be forgotten	Storage limitation [8, 21]	Deleting student data weakens predictive models and disrupts long-term research. Anonymization limits tracking for personalized learning [9, 17].

sonnel, vocational colleges have taken incremental steps toward compliance, often focusing on data minimization and role-based access control. These institutions demonstrate how GDPR compliance can be pursued progressively, emphasizing that adaptation is feasible even without enterprise-level infrastructure.

In addition, cross-border data transfers present challenges for institutions participating in international programs [21]. To mitigate risks, some colleges have adopted GDPR compliant cloud services, which ensure student data protection when shared internationally [18]. Others restrict access to sensitive data unless covered by formal data-transfer agreements.

4. CONCLUSION

The integration of EDM and GDPR creates a dynamic tension between ethical data governance and innovation in education. GDPR establishes foundational principles, such as lawfulness, fairness, transparency, purpose limitation, and data minimization, that fundamentally shape how institutions handle student data. While these principles enhance individual privacy and ethical accountability, they also introduce constraints that challenge the flexibility and scalability of data-driven educational tools.

This paper examined the operational challenges that arise when implementing GDPR-compliant EDM systems. Explicit consent requirements complicate the collection of large-scale educational data, while the right to eliminate undermines the continuity required for longitudinal learning analytics. Techniques such as anonymization and pseudonymization, although critical to protecting privacy, often diminish the effectiveness of adaptive learning models. In addition, compliance involves considerable financial and administrative commitments, particularly for smaller institutions.

The case examples reviewed in this study show how institutions adapt to these regulatory pressures. Universities have adopted centralized data governance structures, schools have implemented privacy dashboards to enhance transparency and user agency, and vocational colleges have pursued phased and role-based access strategies to meet compliance goals with limited resources. These responses demonstrate that, while there is no universal solution, institutions are developing viable paths to reconcile privacy and data utility.

Looking ahead, three strategic areas deserve continued attention. First, investing in privacy-preserving analytics, such as federated learning and differential privacy, can help institutions analyze student data without compromising individual identities. Second, dynamic consent systems offer a promising way to grant students granular control over data use while maintaining system flexibility. Third, stronger collaboration between regulators and educators is needed to create GDPR interpretations tailored to educational contexts, allowing innovation to proceed within clearly defined ethical boundaries.

GDPR has reshaped the EDM landscape by elevating privacy from a technical safeguard to a central design principle. As institutions adapt, the challenge will not be whether to comply, but how to do so in ways that sustain innovation, protect individual rights, and support the transformative potential of data in education.

5. ACKNOWLEDGMENTS

This work was supported by the Academy of Finland (project no. 356314).

6. REFERENCES

[1] Aytaj Ismayilzada, Ayaz Karimov, and Mirka Saarela. Serious games analytics in vr environments: A two-stage systematic literature review. *Journal of Interactive Learning Research*, 36(1):57–69, 2025.

[2] Paul Prinsloo, Sharon Slade, and Mohammad Khalil. The answer is (not only) technological: Considering student data privacy in learning analytics. volume 53, pages 876–893. Wiley Online Library, 2022.

[3] Joel R Reidenberg and Florian Schaub. Achieving big data privacy in education. *Theory and Research in Education*, 16(3):263–279, 2018.

[4] Maja Gligora Marković, Sandra Debeljak, and Nikola Kadoić. Preparing students for the era of the general data protection regulation (gdpr). *TEM Journal*, 8(1), 2019.

[5] Ayaz Karimov, Mirka Saarela, and Tommi Kärkkäinen. Ethical educational data processing differences of students with special needs in post-soviet countries. In *Proceedings of the 17th International Conference on Educational Data Mining*, pages 898–902, 2024.

[6] Aurimas Šidlauskas and Tadas Limba. General data protection regulation implementation in higher education institutions. In *EDULEARN19 Proceedings*, pages 2040–2047. IATED, 2019.

[7] Polydorou Eleni. Towards a secure and privacy compliant framework for educational data mining. In *International Conference on Research Challenges in Information Science*, pages 534–541. Springer, 2023.

[8] Stephen Hutt, Sanchari Das, and Ryan S Baker. The right to be forgotten and educational data mining: Challenges and paths forward. *International Educational Data Mining Society*, 2023.

[9] Isak Potgieter. Privacy concerns in educational data mining and learning analytics. *The International Review of Information Ethics*, 28, 2020.

[10] Eirini Mougiakou and Maria Virvou. Based on gdpr privacy in uml: Case of e-learning program. In *2017 8th International Conference on Information, Intelligence, Systems & Applications (IISA)*, pages 1–8. IEEE, 2017.

[11] Roland J Petrasch and Richard R Petrasch. Data integration and interoperability: Towards a model-driven and pattern-oriented approach. *Modelling*, 3(1):105–126, 2022.

[12] Ingo Siegert, Vered Silber Varod, Nehoray Carmi, and Pawel Kamocki. Personal data protection and academia: Gdpr issues and multi-modal data collections. *Online Journal of Applied Knowledge Management (OJAKM)*, 8(1):16–31, 2020.

[13] Aikaterini Daoultzoglou. Gdpr and education: an approach for e-learning in greek schools. *Ανοικτη Εκπαιδευση*, 19(1):191–209, 2023.

[14] Tore Hoel, Dai Griffiths, and Weiqin Chen. The influence of data protection and privacy frameworks on the design of learning analytics systems. In *Proceedings of the seventh international learning analytics & knowledge conference*, pages 243–252, 2017.

[15] Evgeniya Nikolova, Mariya Monova-Zheleva, and Yanislav Zhelev. Personal data processing in a digital educational environment. *Mathematics & Informatics*, 65(4), 2022.

[16] Mark Klose, Vasvi Desai, Yang Song, and Edward Gehringer. Edm and privacy: Ethics and legalities of data collection, usage, and storage. *International Educational Data Mining Society*, 2020.

[17] Alexander Askinadze and Stefan Conrad. Respecting data privacy in educational data mining: an approach to the transparent handling of student data and dealing with the resulting missing value problem. In *2018 IEEE 27th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, pages 160–164. IEEE, 2018.

[18] Alex Harding. An agile approach to gdpr implementation within a further education college, 2018.

[19] Thashmee Karunaratne. For learning analytics to be sustainable under gdpr—consequences and way forward. *Sustainability*, 13(20):11524, 2021.

[20] Maria da Conceição Freitas and Miguel Mira da Silva. Gdpr and distance teaching and learning. In *2021 16th Iberian Conference on Information Systems and Technologies (CISTI)*, pages 1–6. IEEE, 2021.

[21] Ágota Albert. Gdpr and educational institutions: Implementing gdpr in 13+1 steps. 2019.