Received: 21 March 2023 | Accepted: 24 August 2023

DOI: 10.1111/bjet.13388

ORIGINAL ARTICLE

British Journal of Educational Technology BERA

Understanding privacy and data protection issues in learning analytics using a systematic review

Qinyi Liu | **Mohammad Khalil**

Centre for the Science of Learning & Technology (SLATE), University of Bergen, Bergen, Norway

Correspondence
Qinyi Liu, Centre for the Science of Learning & Technology (SLATE), University of Bergen, Christiesgate 12, Bergen 5020, Norway.
Email: qinyi.liu@uib.no

The field of learning analytics has advanced from infancy stages into a more practical domain, where tangible solutions are being implemented. Nevertheless, the field has encountered numerous privacy and data protection issues that have garnered significant and growing attention. In this systematic review, four databases were searched concerning privacy and data protection issues of learning analytics. A final corpus of 47 papers published in top educational technology journals was selected after running an eligibility check. An analysis of the final corpus was carried out to answer the following three research questions: (1) What are the privacy and data protection issues in learning analytics? (2) What are the similarities and differences between the views of stakeholders from different backgrounds on privacy and data protection issues in learning analytics? (3) How have previous approaches attempted to address privacy and data protection issues? The results of the systematic review show that there are eight distinct, intertwined privacy and data protection issues that cut across the learning analytics cycle. There are both cross-regional similarities and three sets of differences in stakeholder perceptions towards privacy and data protection in learning analytics. With regard to previous attempts to approach privacy and data protection issues in learning analytics, there is a notable dearth of applied evidence, which impedes the assessment of their effectiveness. The findings of our paper suggest that privacy and data protection issues should

not be relaxed at any point in the implementation of learning analytics, as these issues persist throughout the learning analytics development cycle. One key implication of this review suggests that solutions to privacy and data protection issues in learning analytics should be more evidence-based, thereby increasing the trustworthiness of learning analytics and its usefulness.

KEYWORDS
data protection, learning analytics, privacy, systematic review, trustworthy

Practitioner notes

What is already known about this topic

- Research on privacy and data protection in learning analytics has become a recognised challenge that hinders the further expansion of learning analytics.
- Proposals to counter the privacy and data protection issues in learning analytics are blurry; there is a lack of a summary of previously proposed solutions.

What this study contributes

- Establishment of what privacy and data protection issues exist at different phases of the learning analytics cycle.
- Identification of how different stakeholders view privacy, similarities and differences, and what factors influence their views.
- Evaluation and comparison of previously proposed solutions that attempt to address privacy and data protection in learning analytics.

Implications for practice and/or policy

- Privacy and data protection issues need to be viewed in the context of the entire cycle of learning analytics.
- Stakeholder views on privacy and data protection in learning analytics have commonalities across contexts and differences that can arise within the same context. Before implementing learning analytics, targeted research should be conducted with stakeholders.
- Solutions that attempt to address privacy and data protection issues in learning analytics should be put into practice as far as possible to better test their usefulness.

INTRODUCTION

Since learning analytics (LA) was given a more general definition at the Learning Analytics and Knowledge Conference in 2011—that is, 'understanding and optimising the learning environment, and measuring and analysing learner-related data (Long & Siemens, 2011)—various aspects of LA have been studied in depth. The issue of privacy and data protection has come to the forefront of LA challenges. Early research emphasised that data used in LA

should be kept for a limited duration and deleted after a certain period of time to protect privacy and take into account student interests (Prinsloo & Slade, 2013). There are several reasons that make privacy and data protection issues more demanding—namely, but not limited to, the significant increase in the amount of data available for LA, the diversity of data modalities, the increasing sophistication of analytical techniques, and the regulatory changes ushered in by the updated data protection legislation, such as the General Data Protection Regulation in the context of Europe (GDPR) (Joksimović et al., 2021). Although privacy and data protection are recognised as critical dimensions for LA, they are also recognised as barriers to the further development of LA as well as the field of data-driven education (Joksimović et al., 2021; Prinsloo et al., 2022).

In contrast, there has been relatively little research on privacy and data protection issues in LA, with the focus being rather dispersed and lacking a concentrated approach (Viberg et al., 2022). Joksimović et al. (2021) suggest that there are two main ways that attempt to address the issues of privacy and data protection in LA—namely, policy- and framework-based solutions, and technical solutions. Examples of policy- and frameworks-based include the DELICATE eight-point checklist developed by Drachsler and Greller (2016). Technical attempts include, for example, the blockchain-based approach for connecting learning data across different learning management systems that was developed by Ocheja et al. (2018). In addition, there are other types of research on privacy and data protection in LA. These studies draw attention to important layers of the problem but do not directly propose solutions, such as studies on the impact of national legal frameworks on LA privacy (Hoel et al., 2017) and studies on teachers' and students' privacy perceptions towards learning management systems (Whitelock-Wainwright et al., 2021).

Privacy and data protection issues in LA remain challenging, and there are still many unanswered questions. For example, to the best of our knowledge, the existing literature lacks systematic answers to questions including: What privacy and data protection problems exist in the LA ecosystem? and What are the similarities and differences between stakeholders in different regions and contexts regarding privacy and data protection issues in LA? This study therefore aims to bridge the gap and contribute with the following:

- indicate how privacy and data protection issues in LA can be better addressed in the future,
- identify privacy and data protection issues that exist at different phases of the learning analytics lifecycle,
- identify how different stakeholders view privacy and which factors influence their views, and
- evaluate and compare previously proposed solutions that attempt to address privacy and data protection in learning analytics.

BACKGROUND AND LITERATURE REVIEW

Learning analytics background

LA is an interdisciplinary field that encompasses several disciplines, including education, psychology and computer science, and its analytical methods include both quantitative and qualitative approaches (Khalil & Ebner, 2015; Misiejuk & Wasson, 2017). The LA ecosystem has been proposed as a lifecycle that includes the collection, analysis, reporting and sharing of learning data (Khalil & Ebner, 2015; Khokhlova, 2023). The LA ecosystem engages different actors but is primarily focused on students and teachers (Mahmoud et al., 2020). Depending on the context, other stakeholders may include administrators representing

institutions, parents and guardians in the K-12 context, technology vendors and researchers (Gray et al., 2022).

Definition of privacy and data protection

Privacy is a complex, multidimensional and controversial concept, with no single framework or theory that can be applied to all scenarios and situations (Page & Wisniewski, 2022). Although an agreed definition of privacy in the context of LA is yet to be established in the literature, a few papers have attempted to define privacy. For example, Ferguson et al. (2016) defined privacy in LA as 'a freedom from unauthorised intrusion: the ability of an individual or a group to seclude themselves or the information about them'. Pardo and Siemens (2014) defined privacy as 'the regulation of how personal digital information is observed by the self or distributed to other observers' (p. 438). Kyritsi et al. (2018), on the other hand, emphasised the non-disclosure of personal information during the data mining and publishing stages. In this paper, drawing upon the works of Kyritsi et al. (2018), Ifenthaler and Schumacher (2016), and Pardo and Siemens (2014), we define privacy in LA as students having control over their own data, and personal information not being disclosed throughout the process of data collection, analysis, or reporting.

With regard to the definition of data protection, to the best of our knowledge, there is not yet a definition of this term in the LA context in the relevant literature. The term 'data protection' is also not in such frequent use as the term 'privacy' in the context of LA. However, the UK's Data Protection Act and the EU's GDPR (Data Protection Act, 2018; Kuner et al., 2020), which are authoritative in the field of data protection, both define data protection in a similar way. Specifically, data protection is covered in both acts by the following three points: (1) all activities related to personal data, such as collection, processing, storage, transmission and deletion, must be used for specific and explicit purposes in accordance with the principles of fairness and transparency, (2) measures must be taken during data processing to ensure that personal data are handled in a secure manner to prevent unauthorised access and possible damage and loss and (3) data must be kept up to date during this process and saved only for as long as necessary (Data Protection Act, 2018; Kuner et al., 2020). Considering the absence of a context-specific definition of data protection within the realm of LA, this paper will adopt the definitions of data protection as stipulated by the data protection laws of the United Kingdom (UK) and the European Union (EU).

Prior research on privacy and data protection in learning analytics

Research related to privacy and data protection in LA is still at a nascent stage, although some remarkable progress has been made. Drachsler and Greller (2016) attempted to have a discussion on what privacy and data protection issues exist in LA. They highlighted power relationship, data/user exploitation, data ownership and several other directions. Other studies have also discussed one of the privacy and data protection issues in LA in a scattered way, such as Torre et al. (2020) on the risky remote computation of sensitive data in the context of digital education. However, previous research has not systematically established which privacy and data protection issues are involved in the different steps of LA. Without a clear identification of the issues at different phases, it is difficult to propose effective solutions. In that matter, our first RQ for this review is:

> RQ1: What are the identified privacy and data protection issues throughout the LA process, from data collection to data reporting?

There are a few papers in the field of LA that investigate stakeholder perceptions of privacy, such as a survey of Norwegian higher education student perceptions of privacy in relation to LA (Botnevik, 2021), and a survey of system developers, academic advisors and students on the privacy use of early warning dashboard LA tools (Sun et al., 2019). But these investigations of stakeholder attitudes are based on the context in which the study was conducted, and it is not clear whether there are some commonalities across contexts and how stakeholder perceptions differ across contexts. Understanding beyond contextual similarity could reduce the workload of future user research prior to conducting the LA. This motivates the second RQ:

> RQ2: How do stakeholders from various backgrounds view privacy and data protection issues in LA similarly and differently?

As for solutions for privacy and data protection issues in LA, Joksimović et al. (2021) classified proposed privacy and data protection solutions into two categories: policy- and framework-based solutions, and technical solutions. For the former, a more recent representative example is the Wellbeing Analytics Code of Practice (Cormack & Reeve, 2022), which includes eight different headings from existing UK and EU legal codes. This guideline has a checklist-like framework in which one of its purposes is to help address privacy and data protection issues in analysing student and staff well-being data. In terms of technical solutions, one of the more representative options is a tool called MORF. MORF is applied in a MOOC context, where it protects student privacy by allowing researchers to perform calculations on datasets without access to the data (Hutt et al., 2022). As can be seen from these two examples, the type, content and context of the solutions proposed in the LA domain vary. There is no comprehensive summary to answer the question of which of the different types of solutions is more promising. Hence, our third RQ of this paper:

> RQ3: How has previous research attempted to address the privacy and data protection issues identified in LA?

METHODOLOGY

In order to address the research questions of this study, we use a systematic literature review approach. A systematic review has many advantages and has been described as one of the most reliable sources of evidence to guide practice (Clarke, 2011). The most preferred method for systematic reviews is Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Liberati et al., 2009). The PRISMA guidelines have been continuously developed and are well adopted in educational technology publishing venues. According to Liberati et al. (2009), the main process of PRISMA consists of four stages—namely, identification, screening, eligibility and inclusion. In this paper, we follow this guideline, as shown in Figure 1.

Identification of databases

For the purpose of this research, we focused on four databases—namely, Scopus, ProQuest, Web of Science (WoS) and the Learning Analytics and Knowledge (LAK) conference. These were selected for the following reasons:

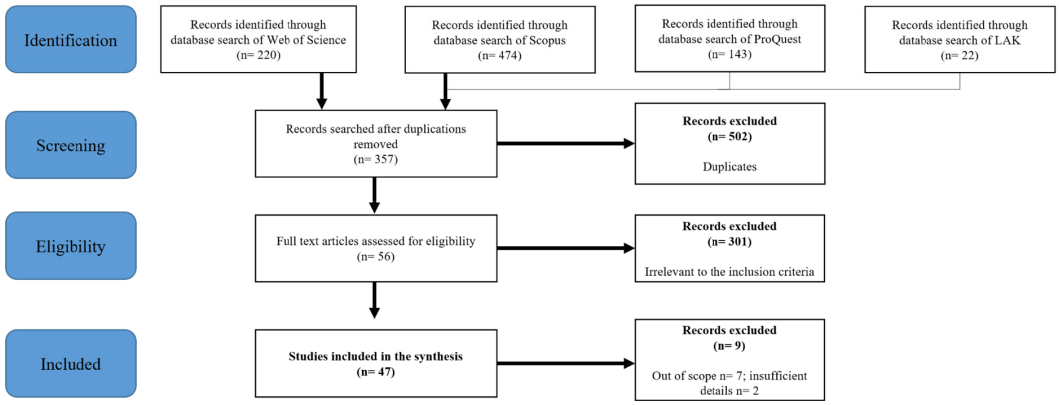

FIGURE 1 The review procedure and protocol, PRISMA (* see Table 1 for inclusion/exclusion criteria).

- Scopus: Scopus covered 40,562 peer-reviewed journals in 2022, which is twice as many as WoS (4831), and is currently the largest multidisciplinary database in existence (Carrera-Rivera et al., 2022).
- ProQuest: A multidisciplinary resource featuring a diversified mix of scholarly journals, trade publications and other timely sources across the top 150 subject areas. ProQuest offers discipline-specific databases (McDonald, 2022).
- WoS: One of the most reputable journal paper collections, indexing both Social Sciences Citation Indexed (SSCI) and Science Citation Indexed (SCI) journals (Rasheed et al., 2020).
- LAK: The LAK is the premier research forum in the field of LA (SoLAR, 2011).

Search query

The search query used for this systematic review is as follows:

- ('learning analytics' OR 'educational data mining') AND ('privacy' OR 'data protection' OR 'privacy preserving' OR 'trustworthy' OR 'responsible'*).[1]

The search terms 'trustworthy' and 'responsible' were added to the search query as both terms encompass linkage to privacy and data protection, as discussed in the works of Khalil et al. (2023) and Thiebes et al. (2020).

Scan and filtering

The search was carried out during the timeframe of 17–19 January 2023. To guarantee inclusion of high quality, papers should be peer-reviewed journal papers and conference proceedings. We decided to use a similar approach to that of Moore and Blackmon (2022)—to filter the results based on the top 20 educational technology journals as per Google Scholar ranking (see Figure 2). In addition, referring to Ifenthaler and Yau (2020) in their systematic review methods section, we also included four additional journals and conferences—Journal of Learning Analytics, Computers in Human Behaviour, Journal Computers and Composition, and Journal of Big Data, as they include peer-reviewed contributions of the learning analytics community. The filtration process first involved importing the retrieved results into the

	Publication	h5-index	h5-median
1.	Computers & Education	128	175
2.	British Journal of Educational Technology	70	96
3.	Education and Information Technologies	69	101
4.	Educational Technology Research and Development	57	80
5.	International Journal of Educational Technology in Higher Education	55	103
6.	International Review of Research in Open and Distributed Learning	55	81
7.	International Journal of Instruction	54	77
8.	Journal of Educational Technology & Society	53	87
9.	Interactive Learning Environments	52	78
10.	The Internet and Higher Education	51	104
11.	Computer Assisted Language Learning	50	81
12.	Journal of Computer Assisted Learning	49	72
13.	Australasian Journal of Educational Technology	49	68
14.	International Journal of Emerging Technologies in Learning (iJET)	47	73
15.	International Conference on Learning Analytics & Knowledge	46	63
16.	Journal of Educational Computing Research	44	72
17.	Learning, Media and Technology	42	66
18.	TechTrends	42	63
19.	Distance Education	40	68
20.	Language Learning & Technology	40	53

FIGURE 2 Top 20 journals for educational technology category used for assessing the retrieved papers as per Google Scholar ranking (Available at https://scholar.google.com/citations?view_op=top_venues&hl=en&vq=eng_educationaltechnology [last accessed March 2023]).

reference management software Zotero, and then identifying and removing duplicates. The two authors then screened the titles and abstracts of the papers using the criteria in Table 1. Next, a review of the full text of the remaining papers was conducted to determine whether they met the eligibility criteria for the final sample. In the event that there were disagreements between the two authors during the process, further discussion was conducted to resolve any conflicts. To ensure that the final corpus was of decent quality, we used the quality criteria of the Critical Appraisal Skills Programme (CASP) (Galdas et al., 2015) for further quality review of the papers, and fine-tuned the CASP for the characteristics of the LA domain (as shown in Figure 3). The rationale behind employing CASP stems from its coverage of the three main issues of quality criteria (rigour, credibility and relevance) and its prior track record in assessing systematic review quality within the LA domain (eg, Mangaroska & Giannakos, 2018).

Coding scheme

The remaining 47 identified papers were reviewed by the authors and were summarised in shared Google Docs according to the following factors: geopolitical locations, study context, proposed solutions, evidence of application, solution details, paper category, research methodology, type of study, number of participants, summary of the paper in terms of RQs,

TABLE 1 Inclusion and exclusion criteria of the systematic review.

Criteria	Inclusion	Exclusion
Topic and focus	Related to the field of LA and covers one or more of the following topics: (1) Privacy, (2) Data protection, (3) Trustworthy	Not related to the field of LA and does not cover one or more of the following topics: (1) Privacy, (2) Data protection, (3) Trustworthy
Publication type	Journal papers and LAK conference proceedings	Posters, book chapters, workshop papers, editorials and reports
Publication status	Peer-reviewed	Non-peer-reviewed and papers in the press
Journal category	Belongs to top 20 educational tech journals, according to Google Scholar and belongs to four additional journals/conferences	Does not belong to top 20 educational tech journals, according to Google Scholar and does not belongs to four additional journals/conferences
Other	Full-access paper	Unavailable articles, duplicates, institutional policies
Language	English only	Other than English

Quality Criteria

1. Does the study clearly address the research problem?
2. Is there a clear statement of the aims of the research?
3. Is there an adequate description of the context in which the research was carried out?
4. Was the research design appropriate to address the aims of the research?
5. Was the data analysis sufficiently rigorous? (if applicable)
6. Is there a clear statement of findings?
7. Is the study of value for research or practice?

FIGURE 3 The adjusted Critical Appraisal Skills Programme (CASP) used in this systematic review.

limitations and bias and CASP quality criteria. A full version is available at this online source: https://shorturl.at/DJT67.

Reliability check

We used the method of inter-rater reliability (IRR) through the Fleiss kappa measure (Fleiss et al., 2013) to ensure the uniformity and precision of the selection and evaluation of papers, particularly during the inclusion and exclusion process. The IRR measures the extent of concurrence among diverse reviewers, which has been shown to reduce bias and guarantee the alignment of the reviewers with the systematic review filtering approach (Cook & Beckman, 2006). As per Fleiss et al. (2013), a Fleiss kappa score of 0.81 or higher denotes a substantial level of agreement among reviewers, while scores ranging from 0.61 to 0.80 indicate a

good level of agreement. Scores between 0.41 and 0.60 imply a moderate level of agreement, whereas scores below 0.41 suggest a weak level of agreement among the reviewers.

In our case, the two authors scanned the filtered papers and identified them for inclusion and exclusion for further analysis. Discussions were conducted regarding areas of uncertainty, and agreement was reached. When comparing the results of the two authors, the final IRR kappa showed a good level of agreement ($\kappa = 0.771$, subjects = 47, raters = 2, and $p < 0.005$).

FINDINGS AND DISCUSSION

Each of the papers included is coded as P_n, and the full list is available in Appendix 1 at the end of the paper (Table A1). The comprehensive table includes all paper details. The corresponding codebook can be accessed via this link: https://shorturl.at/DJT67.

Response to RQ1

> RQ1: What are the identified privacy and data protection issues throughout the learning analytics process, from data collection to data reporting?

After mapping all the papers, we conducted a thematic analysis, following Nowell et al.'s (2017) six steps of thematic analysis. After identifying the initial theme, we looked at the raw data again to verify the appropriateness of the themes. Finally, we summarised eight points of privacy and data protection issues in LA. Additionally, during the process, a note was included to indicate how each paper approached the discussion of the respective issue. The eight identified privacy and data protection issues were mapped to the phases of LA (data collection, data analysis, data reporting and sharing), with five of the issues being addressed throughout the LA steps (see Table 2). Table 3 describes how different papers discuss the eight privacy and data protection issues in LA (Table 3 is the version without notes—see here (https://shorturl.at/mCEZ1) for the version with coding and notes). In addition, as an paper may cover more than one issue, one may appear multiple times under different privacy and data protection issues when counting as demonstrated in Table 3.

Privacy and data protection issues in three different phases

This section elaborates on the privacy and data protection issues identified at each LA phase shown in Table 2: (1) data collection, (2) data analytics and (3) data reporting and sharing.

Data collection

Privacy and data protection issues in the data collection phase (ie, collecting too much sensitive data, and concerns about excessive data) are addressed in a total of five relevant papers ($n = 5$). Haythornthwaite (2017) outlined the issue of excessive data collection in her paper from before the introduction of the GDPR, highlighting privacy issues throughout the LA ecosystem. In the four ($n = 4$) subsequent papers, it was highlighted that data collection continues to evolve with the development of LA, the diversification of data sources (collected from wireless networks, social media and mobile apps) and the increasing scope of data collection, such as biometric data, social network data and location data (Khalil et al., 2018; Kyle, 2019; Slade et al., 2019). Moreover, the following paper on Multimodal Learning Analytics (MMLA) by Yan et al. (2022) supported the previous discussion. In Yan et al.'s (2022)

TABLE 2 Privacy and data protection identified by coding and mapping (categorised by LA phase).

	LA phase		
	Data collection	**Data analytics**	**Data reporting and sharing**
Identified privacy and data protection issues	(1) Collecting sensitive data, collecting too much data	(2) Anonymisation, sensitive data storage and calculation problems	(3) Data misuse, anonymisation, sensitive data storage and calculation problems
	Issues that are found in the whole process: (4) LA privacy definition issues; (5) transparency and communication; (6) power relationships; (7) lack of quality, and conservative attitudes of relevant stakeholders; (8) legislation-related issues		

TABLE 3 Specific papers that discuss eight privacy and data protection issues (some papers overlap per identified issue).

Identified issues	Matching papers	Count of papers per issue
LA privacy definition issues	A27	1 paper
Collecting sensitive data, collecting too much data	A6, A7, A17, A40, A46	5 papers
Anonymisation, sensitive data storage and calculation problems	A1, A6, A10, A11, A14, A29, A30, A31, A34, A37	10 papers
Data misuse	A7, A14, A19, A42	4 papers
Transparency and communication	A1, A3, A4, A5, A8, A10, A15, A18, A23, A26, A28, A37, A39, A40, A44, A45, A46	17 papers
Power relationship	A1, A3, A16, A37, A39, A40, A43, A45	8 papers
Lack of quality and conservative attitudes of relevant stakeholders	A15, A18, A22, A35, A37	5 papers
Legislation-related issues	A10, A14, A15, A25, A36, A37, A41, A42, A43, A47	10 papers
Discussed in general	Some of the papers do not mention specific privacy and data protection issues in LA—eg, LA attitude investigation, etc.	11 papers

paper, the authors raised concerns about the nature of bulk data collection in MMLA. They believe it may inadvertently collect sensitive data without proper consent (Yan et al., 2022).

Data analytics

As for the anonymisation, sensitive data storage and calculation problems in LA, it is necessary to mention that this problem is not restricted to the data analysis stage but may also occur at the data publication and sharing stages. The issue of anonymisation is certainly a major one. It is mentioned to varying degrees in previous studies ($n = 10$). The first to raise this issue with regard to LA were Greller and Drachsler (2012), who mentioned data storage security and data anonymisation in their paper. In the following years, related issues were discussed in successive papers (Drachsler & Greller, 2016; Haythornthwaite, 2017). Later, Duin and Tham (2020) argued that LA systems do not allow students to opt out from or delete their own data, arguing that such a practice compromises student anonymity. Torre et al. (2020) point out the risk of remote computation of sensitive data and also tried to

solve this problem by having sensitive data computed locally. Subsequently, more details about anonymisation in LA were discussed. Yacobson et al. (2021) showed that the current field of LA is not ripe for anonymisation and de-identification. Yacobson et al. (2021) used unsupervised machine learning to discover students' personal information from published de-identified data, achieving remarkable success in disclosing private information about the students. This view was echoed by Hutt et al. (2022), who argued that, especially in online courses, the amount of data obtained is so large that anonymisation by simply removing user names/IDs is insufficient. There are now a variety of subtle identifiers, such as demographics and IP addresses, that necessitate further technical means of anonymisation (Hutt et al., 2022). In order to address the issue of subtle identifiers, anonymisation by focusing on group and subgroup characteristics without linking individual identity to the group is proposed (Li, Jung, et al., 2022). In contrast, Vatsalan et al. (2022) responded to the issue of group anonymisation with a rebuttal concerning the current state of data publication and attacks in the field of LA. Vatsalan et al. (2022) argue that, although more work has been done on currently published LA datasets for personally identifiable risks, attackers are often not interested in a single piece of personal information, and the goal of attackers is generally to recover as much data as possible from the dataset. On the other hand, Prinsloo et al. (2022) provide a general critique of privacy-enhancing techniques (PETs) to achieve anonymisation, criticising that current PETs are often obscure and expensive, so they are not widely used. In addition, Prinsloo et al. (2022) emphasised that technical solutions alone do not work, because they always lag behind the problem and can only try to solve it after it has already occurred.

Data reporting and sharing

The privacy and data protection issues that occur in data reporting and sharing (ie, data misuse) have not been discussed as much as the previously mentioned issues. In total, only ($n = 4$) papers addressed the data misuse problem. This issue was raised for the first time by Greller and Drachsler (2012), but they did not elaborate on it in more detail. Lawson et al.'s (2016) study of the Early Warning Student Indicator (EASI) at the University of Queensland, Australia, found that, although students consented to the collection and use of their data upon enrolment, the way in which student data was subsequently used by researchers differed from what had been intended by the platform designers. This resulted in inconsistencies in data use and consent, leading to misuse of data. Another concern about data misuse is that students do not want the data collected from them to lead to incidents such as having their rights violated (Wang, 2016). Additionally, data misuse is also a concern due to the huge volume and diversity of MMLA (Yan et al., 2022). The previous research has provided limited insights into data misuse, possibly due to enhanced relevant legislation that has mitigated this problem compared to other issues. Nonetheless, the rapid advancements in technology have introduced new challenges regarding the potential misuse of new data types, such as multimodal data.

Privacy and data protection issues across the three phases

This section introduces five issues that cut across the LA cycle—namely, (4) LA privacy definition issues; (5) transparency and communication; (6) power relations; (7) lack of quality and conservative attitudes of relevant stakeholders; and (8) legislation-related issues.

First, there is an unclear definition of privacy in the context of LA ($n = 1$) as mentioned by Viberg et al. (2022). Their argument tends to be valid, as many privacy studies suggest it is a cultural product and a context-dependent (Bennett, 2008; Mulligan et al., 2016; Solove, 2008).

Another cross-cutting privacy and data protection issue is transparency and communication. Because compromising transparency often undermines students' rights and creates unequal power relations, transparency and communication issues are also often discussed in conjunction with a third issue of power relations. Therefore, our paper discusses the two consecutively. Out of the eight privacy and data protection issues examined, transparency and communication is the most discussed issue in included papers ($n = 17$). Almost all relevant papers ($n = 17$) acknowledged the significance of transparency and communication as critical concerns. The role of transparency, as summarised from the previous studies, is threefold: first, transparency can support informed consent (West et al., 2020); second, transparency is an outcome of considering the rights of students (Arnold & Sclater, 2017); and third, transparency is equally important for institutions, which are also at risk of legal exposure if they are not transparent about data collection and analysis (Duin & Tham, 2020). Additionally, the research also indicates that the absence of transparency can undermine trust and thus hinder LA (Cormack & Reeve, 2022). The form and content of transparency include communicating with relevant stakeholders for input on privacy, data collection and use (Ahn et al., 2021; Arnold & Sclater, 2017; West et al., 2020), but so far, it has not yet been well addressed in LA. According to Sun et al. (2019), users are unaware of the underlying determinants of the data, and there are usually no indicators of data sources or inferences on the LA page, which makes transparency much less likely and more difficult to audit. Furthermore, data policies, an important component of transparency, are also lengthy and unclear, both in universities and in MOOC (Drachsler & Greller, 2016; Prinsloo & Slade, 2015). Slade et al. (2019) came to the same conclusion when they found opacity in the use of data, leading to impaired student agency.

This naturally leads to the issue of power relations ($n = 8$), where transparency and the issue of power relations are intertwined. Tsai et al. (2020) argue that, due to the current lack of transparency, students passively accept how their data are used, and this situation exacerbates information asymmetry and inequality in power relations. Although students' rightful or desirable rights in LA include control over data, access, accountability and evaluation (Pardo & Siemens, 2014), these rights are not well secured (Slade et al., 2019). Some scholars have reflected on transparency through the lack of student rights, and Li, Jung, et al. (2022) argue that even high levels of transparency are meaningless if students' rights of recourse and accountability to their data are not guaranteed. Furthermore, research suggests that lack of transparency and the fear of being disempowered by conducting LA are simultaneous barriers to trust LA services (Tsai et al., 2021).

Lack of quality and the conservative attitudes of stakeholders were mentioned in few studies ($n = 5$). Previous studies ($n = 3$) had mentioned, to varying degrees, that stakeholders were unable to express their views and make informed decisions regarding privacy and data protection due to their lack of knowledge and related training in LA (Duin & Tham, 2020; Ifenthaler et al., 2021; Mahmoud et al., 2022). This result also calls for more data literacy training for stakeholders in LA. In addition, some studies have argued that students' conservative attitudes towards data sharing also hinder the development of LA (Ifenthaler & Schumacher, 2016; Kumar et al., 2020). This point will be further elaborated in the second research question.

For the legislation-related issue, there are certain papers ($n = 11$) discussing the following legal concerns: data ownership, recourse, data integration issues and the slow process of related legislation. However, the legislation-related papers are more time-sensitive, and issues described by some papers are no longer of concern in some regions—for example, the legal gaps of data integration, which were mentioned by Greller and Drachsler (2012) and for which legislation is now in place in many parts of the world. Based on the review of this paper, the legal issues that still exist in relation to LA include cross-border issues with personal data and inadequate legislation in some regions (Prinsloo & Kaliisa, 2022; Silvola et al., 2021).

Response to RQ2

> RQ2: How do stakeholders from various backgrounds view privacy and data protection issues in LA similarly and differently?

A total of 17 empirical papers ($n = 17$) have examined stakeholder perceptions towards privacy and data protection in LA (see detail in Table 4). The geopolitical and study contexts, as well as the stakeholders investigated in these papers, are varied. As shown in Table 4, the vast majority of the studies focus on higher education ($n = 16$), with only a single paper focusing on K-12 education ($n = 1$). Most of the studies focus primarily on student perspectives ($n = 11$), with few studies ($n = 6$) focus on other stakeholders. Specifically, papers focus on both learner and faculty ($n = 2$), learner, faculty and developer ($n = 1$), and faculty and/or policymakers ($n = 3$). With regard to the geopolitical location, most reported studies were based in Europe and the United States of America, with relatively few investigating non-Western countries and countries from the Global South.

Although the backgrounds of these studies are different, there are still some commonalities. According to previous empirical research, students and faculty staff consider privacy and data protection to be as important as the analytics dimensions in the LA ecosystem (Whitelock-Wainwright et al., 2020). Both instructors and students consider privacy issues to be very important (Li, Jung, et al., 2022). Some studies have found that students' attitudes can be quite conservative—for example, students' desire for LA services is not as strong as their desire for universities to ensure data security and data control (Whitelock-Wainwright et al., 2020). Furthermore, students are concerned about the privacy concerns that may arise from the further development of LA capabilities (Whitelock-Wainwright et al., 2020). Another commonality is that studies across geolocations have shown students' strong thoughts about

TABLE 4 Mapping stakeholder, geopolitical and study contexts in the included papers.

Stakeholders	Paper ID	Study context	Geopolitical location
Learner	A3	Higher education	UK
	A8	Higher education	USA and UK
	A18	Higher education	General (majority of the participants based in Germany)
	A22	Higher education	Egypt
	A25	Higher education	Estonia, Spain and Netherlands
	A28	Higher education	Sweden
	A32	Higher education	Sweden
	A38	Higher education	UK
	A40	Higher education and online education	UK
	A45	Higher education	Australia
	A47	Higher education	Finland
Learner and faculty	A23	Higher education	UK
	A35	Schools	Malaysia
Learner, faculty and developer	A39	Higher education	USA
Faculty/local policy makers	A10	Higher education	USA
	A33	Schools	Greece and Cyprus
	A44	Higher education	USA

privacy do not translate into action, as demonstrated by Tsai et al. (2020) in the UK and Kumar et al. (2020) in Malaysia. The students' desire for privacy is strong but is not acted upon accordingly. The reasons for this have not been explored in the relevant LA studies.

Another similarity is found in the attitudinal factors that influence students' privacy. Studies in different contexts have shown that students' privacy attitudes are influenced by trust and perceived privacy risks. In Slade et al.'s (2019) survey of the Open University level 1 students, trust was a key factor influencing their attitudes towards privacy. Because students were more confident that the Open University would use their data appropriately, they were correspondingly less concerned about both the collection and use of personal data than externals (Slade et al., 2019). The importance of trust has also been demonstrated in other studies. In Tsai et al.'s (2020) study, when expressing attitudes towards sharing personal data, students from undergraduate to PhD level expressed the highest trust in their instructors and were willing to share more, while trusting external parties the least. As for the perception of privacy risk, according to a study conducted by Mutimukwe et al. (2022) at Swedish universities, students' assessments of their privacy risk determine their attitudes towards privacy-related matters. This assertion is consistent with the results of Slade et al.'s (2019) study and Tsai et al.'s study (2020) in the UK. It is therefore clear that trust and privacy risk perceptions continue to influence student's attitudes towards privacy and data protection in LA, even though different studies are rooted in different contexts.

However, this does not mean that there are no differences in stakeholders' attitudes towards privacy and data protection. There are three main groups of differences—the first between students, the second between different types of stakeholders, and the third between student expectations and reality. For the first difference (between students), Arnold and Sclater (2017) found that privacy attitudes differed between students in the UK and the USA. US students demonstrate a higher level of acceptance towards the storage of data for extended durations for LA services, whereas the UK students exhibit a comparatively lower level of acceptance of the storage of their data. Moreover, apart from the variations observed between different countries, student attitudes towards data storage for longer periods of time in LA can also differ within the confines of a single university. According to Sun et al. (2019), students at the University of Michigan demonstrated different views with regard to whether students should control the use of data for the LA early warning dashboard. The majority of students believed that they should have control over the LA tools, but a small number of students had the opposite opinion, or no opinion at all (Sun et al., 2019).

The second difference is between different types of stakeholders. This refers to the difference between students and teachers, and the difference between students and developers in their perception of privacy in LA. In the case of the University of Michigan, a lower percentage of the faculty staff than the students believed that students should control the data in the early warning dashboard (Sun et al., 2019). As another stakeholder in the event, developers also felt that, while they could help to manage student data, specific control and consent issues were not their consideration (Sun et al., 2019).

The third distinction is between students' expectations and reality. This set of distinctions is demonstrated by the fact that students' experiences and realistic expectations of privacy in LA are often lower than ideal expectations. This distinction was confirmed by studies in different contexts. For example, students report that their ideal expectations are that they are the beneficiaries of LA and are able to control the data, but the power imbalance in the LA process often makes a difference between their actual feelings and their ideal expectations (Tsai et al., 2020). These three groups of differences indicate that the differences in perceptions between stakeholders should be taken into account when implementing LA. Targeted, detailed research should be conducted to better meet the expectations of different stakeholders.

Response to RQ3

> RQ3: How has previous research attempted to address the privacy and data protection issues identified in LA?

Our mapping shows that there are a certain number of papers ($n = 26$) proposing various solutions and approaches (see Table 5). Papers proposing solutions accounted for 55% of the total included corpus. The solution proposed by papers is unique in most cases (either a legal- and framework-based solution or a technical solution), and it is very rare to find a combination of both. There is only one paper ($n = 1$) that covers both sides, therefore listed as a separate category. Additionally, because of the need to compare the proportions of different types of solutions and the proportions of practice of different types of solutions, these two proportions are added to this table.

Legal- and framework-based solutions

This approach was originally mentioned by Tsai et al. (2020) and Joksimović et al. (2021), who argued that early in the development of LA, the researchers in LA and social sciences focused on developing policies and frameworks to address privacy and data protection issues in LA. During the systematic review, we found that some papers deal with privacy issues in LA by discussing the legislative basis. Therefore, we categorised such papers under the name legal- and framework-based solutions. This category of solutions accounts for 65% of the total number of solutions and is represented by frameworks/models of various forms, such as the eight-point checklist called DELICATE (Drachsler & Greller, 2016) and SPICE (Mutimukwe et al., 2022). SPICE is a model specifically designed to explore student privacy issues under LA and was built on data from over a hundred Swedish students (Mutimukwe et al., 2022). Although it is not known whether SPICE is generalisable or efficient, it attempts to reveal the core structure of the privacy problem and to identify its antecedents and consequences. In addition to such framework-type measures, there are many policy- and law-related initiatives under this category. For example, Prinsloo and Slade (2017) make recommendations for addressing privacy issues. In particular, they emphasise that initiatives such as making ethical and appropriate moves towards LA outcomes must be implemented into contracts in order to have legal benefits (Prinsloo & Slade, 2017). Prinsloo and Kaliisa (2022), on the other hand, conclude that students can withdraw consent after the initial consent, based on a review of legislation in 32 countries. Overall, legal- and framework-based solutions make many contributions, particularly in defining whether particular initiatives are legally actionable and in examining the factors that influence students' privacy-related behaviours. These contributions provide a critical foundation for the implementation of technical solutions.

TABLE 5 Categories of proposed solutions for privacy and data protection issues in LA.

Category	Paper ID	Proposed solutions	Evidence of application for proposed solutions
Legal and frameworks	A1, A5, A6, A12, A14, A15, A16, A19, A24, A26, A28, A31, A37, A38, A39, A40, A43	17 (65%)	3 (18%)
Technical	A4, A9, A11, A13, A17, A21, A29, P30	8 (31%)	5 (63%)
Combined solutions	A20	1 (4%)	0

There are, however, some limitations to this type of solution, and only 18% of the legal- and framework-based solutions have been validated in practice. This percentage is much lower than the percentage of technical solutions that have been applied (63%). This shows that there is a consequential problem with legal- and framework-based solutions, which is that they are too rarely applied in practice. The reasons for this can be attributed to a lack of clear and actionable guidelines (Marshall et al., 2022).

Technical solutions

Technical solutions is also one of the two solutions summarised by Joksimović et al. (2021). The number of technical solutions is relatively small, accounting for only 31% of the total solutions. There are two main types of technical solutions—namely, those that make improvements in algorithms and those that attempt to develop tools. Two papers ($n = 2$) have been enhanced in algorithms: the fairer and more trustworthy LA algorithm developed by Li, Xing, and Leite (2022), and the Markov models used by Vatsalan et al. (2022) to quantify the re-identification risk in LA. Both papers have experimental results based on data that prove their contribution. As for the second type, which concerns the development of tools to address privacy concerns in LA, the more representative ones are the two privacy-enhancing tools for MOOCs developed by Torre et al. (2020) and Hutt et al. (2022). The former is a browser-based privacy tool that preprocesses data to reduce the computational burden on researchers and allows sensitive LA data to be stored locally (Torre et al., 2020). This tool also has the advantage of having no setup cost (Torre et al., 2020). The second tool, developed by Hutt et al. (2022), is called MORF and allows data to be available invisibly, and researchers cannot access the data directly. Other tools include a privacy dashboard that allows students to set privacy preferences, although this tool has not been developed (Kyle, 2019), and using blockchain to address learning data silos while protecting privacy (Ocheja et al., 2018). Overall, the practice of technical solutions is relatively good, with five of eight papers reporting this being tested in practice. However, there are some shortcomings in the current technical solutions—for example, some methods are only used in MOOC and cannot be used in other broader scenarios, lacking generalisability. Furthermore, technical solutions come at the expense of data utility and are usually expensive (Khalil, 2018; Khalil & Ebner, 2016). In addition, staff who use technical tools will be a problem, because the tools will be used by humans, and technical solutions do not have a regulatory framework for staff who will use this tool. This could lead to the improper use of the tools and hinder their effectiveness. Considering all the above reasons, technical solutions should be combined with legal- and framework-based solutions in order to be most effective.

Combined solutions

A combined solution refers to an integrated solution that combines both technical solutions and legal- and framework-based solutions. This kind of solution is supported by Ifenthaler and Schumacher (2016), who came to this conclusion (that a combined solution is needed) after conducting a survey of 330 students. They argue both for institutional measures to ensure that all stakeholders are involved in LA and for the use of privacy computing models to inform stakeholders of the complex decisions required by the LA system (Ifenthaler & Schumacher, 2016). However, there are some difficulties in the proposal and implementation of combined solutions, due to their complexity and workload. This has only been attempted by Marshall et al. (2022) who propose a replicable framework for ethical LA that includes both technical solutions and legal- and framework-based solutions, and which also introduces a

new standardised privacy risk measure. Furthermore, this privacy risk measurement mechanism helps to automate data privacy operations (Marshall et al., 2022). However, this approach is still under development—including the tools for students to calculate privacy risks, for example. There is, therefore, not yet any evidence of application. In general, the development of combined solutions is still relatively rudimentary, but they have great potential for the future as they can combine the advantages of other solutions to complement each other.

Limitations of the study

We acknowledge that this paper has limitations. First, the filtering process is limited to the most influential journals in the field of educational technology, which partially compromises the coverage of this paper. Second, we limited the final selection of the papers to the English language. Third, we acknowledge the possibility that other papers could have been missed in the systematic review due to the search query used in Section 'Search query'. Fourth, published studies may be biased because they reflect the interests of researchers and those involved in the authorship of the included papers. That is, other stakeholder documentations, for example, local policymakers, are outside the scope of this systematic review. Finally, addressing the concern that certain suggested solutions lack supporting evidence in RQ3, it is important to acknowledge the time lag between practice and evidence (eg, Marshall et al., 2022). Therefore, the absence of evidence of practice could be that it is under development or undergoing practice and has not yet been published.

SUMMARY AND CONCLUSIONS

Summary of findings

This study provides a comprehensive summary of the identified privacy and data protection issues in LA, similarities and differences in stakeholder perceptions of privacy and data protection issues, and prior research's solutions to LA privacy and data protection issues.

We have identified eight privacy and data protection issues in LA process (RQ1):

- **Collecting sensitive data, collecting too much data.** As LA tech (like MMLA) advances, more data sources and expansive collection raise over-collection of (sensitive) data.
- **Anonymisation, sensitive data storage and calculation problems.** This issue pertains to the underdeveloped state of LA data anonymisation and de-identification, particularly regarding group anonymisation, along with challenges in remote and local storage and computation of sensitive data.
- **Data misuse.** Cases of data misuse include inconsistencies in data use and consent. While less discussed than the prior concerns, data misuse might encounter novel challenges due to evolving technologies.
- **LA privacy definition issues**. Although some LA literature has attempted to define privacy in LA, the definition of privacy in LA is not clear enough to develop an accepted understanding.
- **Transparency and communication.** Transparency and communication take the form and content of communicating with relevant stakeholders and soliciting their views on privacy, data collection and use, and data policies.

- **Power relationship.** This issue refers to the vulnerability of students relative to other stakeholders in power relations. The rights that students should ideally have in LA, including control, access, accountability and evaluation of data, are not currently well secured.
- **Lack of quality and conservative attitudes of relevant stakeholders**. This issue relates to the lack of knowledge and training of stakeholders in relation to LA and therefore their difficulty in making informed decisions about privacy and data protection issues, for example, students have a conservative attitude towards LA data sharing.
- **Legislation-related issues.** Current legal issues with LA include cross-border issues with personal data and inadequate legislation in certain regions.

From privacy definitions to legislation, spanning the LA cycle, it is clear that safeguarding data privacy is an ongoing, time-sensitive marathon that cannot be relaxed in the LA cycle (Crowther, 2022). It is important to underscore that privacy and data protection issues in LA are inherently time-sensitive. Evolving concerns and technology demand continuous adaptation to address emerging issues.

In terms of RQ2, we have identified two major similarities and three differences:

- **Two similarities**. Two cross-contextual student attitude similarities relate to privacy and data protection in LA: valuing privacy's role and having conservative attitudes (Whitelock-Wainwright et al., 2020). Additionally, shared attitudinal factors—trust and perceived privacy risk—affect privacy stances across diverse research contexts (Slade et al., 2019; Tsai et al., 2020).
- **Three differences.** Three distinctions: (1) Among students—country-based and intra-university differences (Arnold & Sclater, 2017; Sun et al., 2019). (2) Stakeholders differ—students and instructors disagree on data control (Sun et al., 2019). (3) Reality falls short—student LA ideal privacy expectations exceed LA actual experiences (Tsai et al., 2020).

Finally, we also identified three types of solutions for the RQ3:

- **Legal- and framework-based solutions.** This solution category involves policy, legal analysis and framework development to tackle LA privacy issues. Notably, these solutions clarify legal viability (Prinsloo & Kaliisa, 2022), forming a base for technical implementations. However, merely 18% were practically used, hindered by lacking clear, feasible guidance (Marshall et al., 2022).
- **Technical solutions.** Two technical solution types: algorithm enhancement and tool development. They see higher practical use than legal or framework approaches. Still, they face issues: limited applicability, data utility loss and costliness (Khalil, 2018). Moreover, lacking regulatory guidelines for human tool users.
- **Combined solutions.** This is a hybrid solution combining the previous two solutions, few that are documented (eg, Marshall et al., 2022). Such solutions may hold promise to integrate benefits from previous solutions, mutual reinforcement for the future.

Implications and conclusions

This review identifies eight privacy and data protection issues at all phases of the LA, summarises two similarities and three differences in stakeholder perceptions of privacy and data protection, and also divides and evaluates previous attempts to address privacy and data protection solutions into three categories. Our findings impose considerable implications for future research on privacy and data protection in LA. First, eight complex

and intertwined privacy and data protection issues show that research in this area needs to go further, and future research should look at these eight issues dynamically in the context of new technological and social conditions. Stakeholders must remain vigilant throughout the entire implementation of LA, as privacy and data protection issues are pertinent at almost every phase. There should be no room for complacency, as these concerns consistently accompany each step of LA implementation. The results of this systematic review suggest that targeted, detailed research should be conducted to better meet the expectations of different stakeholders. Lastly, the research presented in this review paper emphasises the importance of empirically testing the proposed solutions. The field strives for more encouragement to conduct practical experiments in order to thoroughly evaluate the efficiency of the proposed solutions to privacy and data protection concerns in LA.

FUNDING INFORMATION

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this paper as no new data were created or analysed in this study.

ETHICS STATEMENT

This systematic literature review only used existing published materials and did not involve data collection from human subjects.

ORCID

Qinyi Liu https://orcid.org/0009-0003-4973-0901

ENDNOTE

[1] For specific search strings for different database, see Appendix 2, Table A2.

REFERENCES

Ahn, J., Campos, F., Nguyen, H., Hays, M., & Morrison, J. (2021). Co-designing for privacy, transparency, and trust in k-12 learning analytics. In *LAK21: 11th International Learning Analytics and Knowledge Conference*, Irvine, CA. https://doi.org/10.1145/3448139.3448145

Arnold, K. E., & Sclater, N. (2017). Student perceptions of their privacy in learning analytics applications. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, Vancouver, British Columbia. https://doi.org/10.1145/3027385.3027392

Bennett, C. J. (2008). *The privacy advocates: Resisting the spread of surveillance*. MIT Press.

Botnevik, S. (2021). *Student perceptions of privacy in learning analytics: A quantitative study of Norwegian students*. In bora.uib.no. https://hdl.handle.net/11250/2757115

Carrera-Rivera, A., Ochoa, W., Larrinaga, F., & Lasa, G. (2022). How-to conduct a systematic literature review: A quick guide for computer science research. *MethodsX*, *9*(9), 101895. https://doi.org/10.1016/j.mex.2022.101895

Clarke, J. (2011). What is a systematic review? *Evidence-Based Nursing*, *14*(3), 64. https://doi.org/10.1136/ebn.2011.0049

Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: Theory and application. *The American Journal of Medicine*, *119*(2), 166.e7-16.

Cormack, A. N., & Reeve, D. (2022). Developing a code of practice for using data in wellbeing support. *Journal of Learning Analytics*, *9*, 1–12. https://doi.org/10.18608/jla.2022.7533
Crowther, K. (2022). *Cybersecurity: A marathon, not a sprint*. https://www.wateronline.com/doc/cybersecurity-a-marathon-not-a-sprint-0001
Darvishi, A., Khosravi, H., Sadiq, S., & Gašević, D. (2022). Incorporating AI and learning analytics to build trustworthy peer assessment systems. *British Journal of Educational Technology*, *53*(4), 844–875. https://doi.org/10.1111/bjet.13233
Data Protection Act, Gov.uk. (2018). https://www.gov.uk/data-protection
Drachsler, H., & Greller, W. (2016). Privacy and analytics: It's a DELICATE issue a checklist for trusted learning analytics. *ACM Digital Library*, *10*, 89–98. https://doi.org/10.1145/2883851.2883893
Duin, A. H., & Tham, J. (2020). The current state of analytics: Implications for learning management system (LMS) use in writing pedagogy. *Computers and Composition*, *55*, 102544. https://doi.org/10.1016/j.compcom.2020.102544
Ferguson, R., Hoel, T., Scheffel, M., & Drachsler, H. (2016). Guest editorial: Ethics and privacy in learning analytics. *Journal of Learning Analytics*, *3*(1), 5–15. https://doi.org/10.18608/jla.2016.31.2
Fleiss, J. L., Levin, B., & Paik, M. C. (2013). *Statistical methods for rates and proportions*. John Wiley & Sons.
Galdas, P., Darwin, Z., Fell, J., Kidd, L., Bower, P., Blickem, C., McPherson, K., Hunt, K., Gilbody, S., & Richardson, G. (2015). *Critical Appraisal Skills Programme criteria*. NIHR Journals Library. https://www.ncbi.nlm.nih.gov/books/NBK311069/
Gray, G., Schalk, A. E., Cooke, G., Murnion, P., Rooney, P., & O'Rourke, K. C. (2022). Stakeholders' insights on learning analytics: Perspectives of students and staff. *Computers & Education*, *187*, 104550. https://doi.org/10.1016/j.compedu.2022.104550
Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society*, *15*, 42–57. https://www.jstor.org/stable/jeductechsoci.15.3.42
Haythornthwaite, C. (2017). An information policy perspective on learning analytics. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, Vancouver, British Columbia. https://doi.org/10.1145/3027385.3027389
Hoel, T., Griffiths, D., & Chen, W. (2017). 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*, Vancouver, British Columbia. https://doi.org/10.1145/3027385.3027414
Hutt, S., Baker, R. S., Ashenafi, M. M., Andres-Bray, J. M., & Brooks, C. (2022). Controlled outputs, full data: A privacy-protecting infrastructure for MOOC data. *British Journal of Educational Technology*, *53*(4), 756–775. https://doi.org/10.1111/bjet.13231
Ifenthaler, D., Gibson, D., Prasse, D., Shimada, A., & Yamada, M. (2021). Putting learning back into learning analytics: Actions for policy makers, researchers, and practitioners. *Educational Technology Research and Development*, *69*, 2131–2150. https://doi.org/10.1007/s11423-020-09909-8
Ifenthaler, D., & Yau, J. Y.-K. (2020). Utilising learning analytics to support study success in higher education: A systematic review. *Educational Technology Research and Development*, *68*(4), 1961–1990. https://doi.org/10.1007/s11423-020-09788-z
Ifenthaler, D., & Schumacher, C. (2016). Student perceptions of privacy principles for learning analytics. *Educational Technology Research and Development*, *64*, 923–938. https://doi.org/10.1007/s11423-016-9477-y
Joksimović, S., Marshall, R., Rakotoarivelo, T., Ladjal, D., Zhan, C., & Pardo, A. (2021). Privacy-driven learning analytics. *Manage Your Own Learning Analytics*, *261*, 1–22. https://doi.org/10.1007/978-3-030-86316-6_1
Khalil, M. (2018). Learning analytics in massive open online courses. *arXiv preprint arXiv:1802.09344*.
Khalil, M., & Ebner, M. (2015). Learning analytics: Principles and constraints: World conference on educational multimedia, hypermedia and telecommunications. In *Proceedings of ED-Media 2015 Conference, 1789–1799*. Montreal, Quebec: Association for the Advancement of Computing in Education (AACE). https://graz.elsevierpure.com/en/publications/learning-analytics-principles-and-constraints
Khalil, M., & Ebner, M. (2016). De-identification in learning analytics. *Journal of Learning Analytics*, *3*(1), 129–138. https://doi.org/10.18608/jla.2016.31.8
Khalil, M., Prinsloo, P., & Slade, S. (2018). User consent in MOOCs—Micro, meso, and macro perspectives. *The International Review of Research in Open and Distributed Learning*, *19*(5), 61–79. https://doi.org/10.19173/irrodl.v19i5.3908
Khalil, M., Prinsloo, P., & Slade, S. (2023). Fairness, trust, transparency, equity, and responsibility in learning analytics. *Journal of Learning Analytics*, *10*(1), 1–7. https://doi.org/10.18608/jla.2023.7983
Khokhlova, A. (2023). *What is learning analytics: Methods, ethics and privacy [2022]*. Valamis. https://www.valamis.com/hub/learning-analytics

Kumar, J. A., Bervell, B., & Osman, S. (2020). Google classroom: Insights from Malaysian higher education students' and instructors' experiences. *Education and Information Technologies*, *25*, 4175–4195. https://doi.org/10.1007/s10639-020-10163-x

Kuner, C., Bygrave, L. A., Docksey, C., & Drechsler, L. (2020). *The EU general data protection regulation (GDPR)*. Oxford University Press. https://doi.org/10.1093/oso/9780198826491.001.0001

Kyle, M. L. J. (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*, 1–22. https://doi.org/10.1186/s41239-019-0155-0

Kyritsi, K. H., Zorkadis, V., & Stavropoulos, E. C. (2018). Privacy issues in learning analytics. *Blended and Online Learning*, *218*, 218–232.

Lawson, C., Beer, C., Rossi, D., Moore, T., & Fleming, J. (2016). Identification of "at risk" students using learning analytics: The ethical dilemmas of intervention strategies in a higher education institution. *Educational Technology Research and Development*, *64*, 957–968. https://doi.org/10.1007/s11423-016-9459-0

Li, C., Xing, W., & Leite, W. L. (2022). Do gender and race matter? Supporting help-seeking with fair peer recommenders in an online algebra learning platform. In *LAK22: 12th International Learning Analytics and Knowledge Conference (LAK22)*, New York, NY, (pp. 432–437). https://doi.org/10.1145/3506860.3506869

Li, Q., Jung, Y., d'Anjou, B., & Wise, A. F. (2022). Unpacking instructors' analytics use: Two distinct profiles for informing teaching. In *LAK22: 12th International Learning Analytics and Knowledge Conference*, New York, NY. https://doi.org/10.1145/3506860.3506905

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gotzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *BMJ*, *339*(339), b2700. https://doi.org/10.1136/bmj.b2700

Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review*, *46*(5), 132–137. https://doi.org/10.17471/2499-4324/195

Mahmoud, M., Dafoulas, G., Abd ElAziz, R., & Saleeb, N. (2020). Learning analytics stakeholders' expectations in higher education institutions: A literature review. *The International Journal of Information and Learning Technology*, *38*(1), 33–48. https://doi.org/10.1108/ijilt-05-2020-0081

Mahmoud, M., Dafoulas, G., Abd ElAziz, R., & Saleeb, N. (2022). Factors affecting the deployment of learning analytics in developing countries: Case of Egypt. *International Journal of Emerging Technologies in Learning (IJET)*, *17*, 279–298. https://doi.org/10.3991/ijet.v17i03.24405

Mangaroska, K., & Giannakos, M. N. (2018). Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. *IEEE Transactions on Learning Technologies*, *12*(4), 1. https://doi.org/10.1109/tlt.2018.2868673

Marshall, R., Pardo, A., Smith, D., & Watson, T. (2022). Implementing next generation privacy and ethics research in education technology. *British Journal of Educational Technology*, *53*(4), 737–755. https://doi.org/10.1111/bjet.13224

Mavroudi, A., Papadakis, S., & Ioannou, I. (2021). Teachers' views regarding learning analytics usage based on the technology acceptance model. *TechTrends*, *65*. https://doi.org/10.1007/s11528-020-00580-7

McDonald, M. K. (2022). *LibGuides: ProQuest central student: About*. Proquest.libguides.com. https://proquest.libguides.com/pqcstuden

Misiejuk, K., & Wasson, B. (2017). State of the field report on learning analytics. In *bora.uib.no*. Centre for the Science of Learning & Technology (SLATE), University of Bergen. https://bora.uib.no/bora-xmlui/handle/1956/17740

Moore, R. L., & Blackmon, S. J. (2022). From the learner's perspective: A systematic review of MOOC learner experiences (2008–2021). *Computers & Education*, *190*, 104596. https://doi.org/10.1016/j.compedu.2022.104596

Mulligan, D. K., Koopman, C., & Doty, N. (2016). Privacy is an essentially contested concept: A multi-dimensional analytic for mapping privacy. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2083), 20160118. https://doi.org/10.1098/rsta.2016.0118

Mutimukwe, C., Viberg, O., Oberg, L., & Cerratto-Pargman, T. (2022). Students' privacy concerns in learning analytics: Model development. *British Journal of Educational Technology*, *53*, 932–951. https://doi.org/10.1111/bjet.13234

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Method*, *16*(1), 160940691773384. https://doi.org/10.1177/1609406917733847

Ocheja, P., Flanagan, B., & Ogata, H. (2018). Connecting decentralized learning records. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge—LAK '18* (Vol. *5*, pp. 265–269). https://doi.org/10.1145/3170358.3170365

Page, X., & Wisniewski, P. (2022). *Modern socio-technical perspectives on privacy* (B. P. Knijnenburg, X. Page, P. Wisniewski, H. R. Lipford, N. Proferes, & J. Romano, Eds.). Springer International Publishing. https://doi.org/10.1007/978-3-030-82786-1

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, *45*, 438–450. https://doi.org/10.1111/bjet.12152

Tsai, Y.-S., Whitelock-Wainwright, A., & Gasevic, D. (2020). *The privacy paradox and its implications for learning analytics*, 230–239. https://www.research.ed.ac.uk. https://doi.org/10.1145/3375462.3375536

Prinsloo, P., & Slade, S. (2013). Learning analytics: Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, *57*(10), 1510–1529. https://doi.org/10.1177/0002764213479366

Prinsloo, P., & Slade, S. (2015). Student privacy self-management: implications for learning analytics. In *Proceedings of the fifth international conference on learning analytics and knowledge* (LAK '15), (pp. 83–92). Poughkeepsie, USA. ACM. https://doi.org/10.1145/2723576.2723585

Prinsloo, P., & Slade, S. (2017). An elephant in the learning analytics room. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, Vancouver, British Columbia. https://doi.org/10.1145/3027385.3027406

Prinsloo, P., Slade, S., & Khalil, M. (2022). The answer is (not only) technological: Considering student data privacy in learning analytics. *British Journal of Educational Technology*, *53*, 876–893. https://doi.org/10.1111/bjet.13216

Rasheed, R. A., Kamsin, A., & Abdullah, N. A. (2020). Challenges in the online component of blended learning: A systematic review. *Computers & Education*, *144*(1), 103701. https://doi.org/10.1016/j.compedu.2019.103701

Salas-Pilco, S. Z., & Yang, Y. (2020). Learning analytics initiatives in Latin America: Implications for educational researchers, practitioners and decision makers. *British Journal of Educational Technology*, *51*(4), 875–891. https://doi.org/10.1111/bjet.12952

Silvola, A., Näykki, P., Kaveri, A., & Muukkonen, H. (2021). Expectations for supporting student engagement with learning analytics: An academic path perspective. *Computers & Education*, *168*, 104192. https://doi.org/10.1016/j.compedu.2021.104192

Slade, S., Prinsloo, P., & Khalil, M. (2019). Learning analytics at the intersections of student trust, disclosure and benefit. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, Tempe, AZ. https://doi.org/10.1145/3303772.3303796

SoLAR. (2011). *International Conference on Learning Analytics & Knowledge (LAK)—Society for Learning Analytics Research (SoLAR)*. Society for Learning Analytics Research (SoLAR). https://www.solaresearch.org/events/lak/

Solove, D. J. (2008). *Understanding privacy*. Harvard University Press.

Sun, K., Mhaidli, A. H., Watel, S., Brooks, C. A., & Schaub, F. (2019). It's my data! Tensions among stakeholders of a learning analytics dashboard. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Glasgow. https://doi.org/10.1145/3290605.3300824

Thiebes, S., Lins, S., & Sunyaev, A. (2020). Trustworthy artificial intelligence. *Electronic Markets*, *31*, 447–464. https://doi.org/10.1007/s12525-020-00441-4

Torre, M. V., Tan, E., & Hauff, C. (2020). edX log data analysis made easy. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, Frankfurt. https://doi.org/10.1145/3375462.3375510

Tsai, Y.-S., Whitelock-Wainwright, A., & Gasevic, D. (2020). *The privacy paradox and its implications for learning analytics*, 230–239. http://www.research.ed.ac.uk. https://doi.org/10.1145/3375462.3375536

Tsai, Y.-S., Whitelock-Wainwright, A., & Gašević, D. (2021). More than figures on your laptop: (Dis)trustful implementation of learning analytics. *Journal of Learning Analytics*, *8*, 1–20. https://doi.org/10.18608/jla.2021.7379

Vatsalan, D., Rakotoarivelo, T., Bhaskar, R., Tyler, P., & Ladjal, D. (2022). Privacy risk quantification in education data using Markov model. *British Journal of Educational Technology*, *53*, 804–821. https://doi.org/10.1111/bjet.13223

Viberg, O., Mutimukwe, C., & Grönlund, Å. (2022). Privacy in LA research: Understanding the field to improve the practice. *Journal of Learning Analytics*, *9*(3), 169–182. https://doi.org/10.18608/jla.2022.7751

Wang, Y. (2016). Big opportunities and big concerns of big data in education. *TechTrends*, *60*, 381–384. https://doi.org/10.1007/s11528-016-0072-1

West, D., Luzeckyj, A., Searle, B., Toohey, D., Vanderlelie, J., & Bell, K. R. (2020). Perspectives from the stakeholder: Students' views regarding learning analytics and data collection. *Australasian Journal of Educational Technology*, *36*, 72–88. https://doi.org/10.14742/ajet.5957

Whitelock-Wainwright, A., Gašević, D., Tejeiro, R., Tsai, Y., & Bennett, K. (2019). The student expectations of learning analytics questionnaire. *Journal of Computer Assisted Learning*, *35*(5), 633–666. https://doi.org/10.1111/jcal.12366

Whitelock-Wainwright, A., Gašević, D., Tsai, Y., Drachsler, H., Scheffel, M., Muñoz-Merino, P. J., Tammets, K., & Kloos, D. (2020). Assessing the validity of a learning analytics expectation instrument: A multinational study. *Journal of Computer Assisted Learning*, *36*, 209–240. https://doi.org/10.1111/jcal.12401

Whitelock-Wainwright, A., Tsai, Y.-S., Drachsler, H., Scheffel, M., & Gašević, D. (2021). An exploratory latent class analysis of student expectations towards learning analytics services. *The Internet and Higher Education*, *51*, 100818. https://doi.org/10.1016/j.iheduc.2021.100818

Yacobson, E., Fuhrman, O., Hershkovitz, S., & Alexandron, G. (2021). De-identification is insufficient to protect student privacy, or—What can a field trip reveal? *Journal of Learning Analytics*, *8*, 83–92. https://doi.org/10.18608/jla.2021.7353

Yan, L., Zhao, L., Gasevic, D., & Martinez-Maldonado, R. (2022). Scalability, sustainability, and ethicality of multimodal learning analytics. In *LAK22: 12th International Learning Analytics and Knowledge Conference*, New York, NY. https://doi.org/10.1145/3506860.3506862

Zeide, E. (2017). The structural consequences of Big Data-driven education. *Big Data*, *5*(2), 164–172. https://doi.org/10.1089/big.2016.0061

How to cite this article: Liu, Q., & Khalil, M. (2023). Understanding privacy and data protection issues in learning analytics using a systematic review. *British Journal of Educational Technology*, *54*, 1715–1747. https://doi.org/10.1111/bjet.13388

APPENDIX 1

TABLE A1 Reduced version of summary of privacy and data protection issues in LA.

Paper ID	Authors	Title	Proposed solutions		Evidence of application the solutions	Paper category	Summary of the paper in terms of RQs
			Legal and frameworks	Technical			
A1	Drachsler and Greller (2016)	Privacy and Analytics: It is a DELICATE Issue a Checklist for Trusted Learning Analytics	\		No	G	(1) Attitudes towards learning analytics. (2) Power relations, use of data/users. (3) Anonymity and data
A2	Hoel et al. (2017)	The Influence of Data Protection and Privacy Frameworks on the Design of Learning Analytics Systems	N/A	N/A	N/A	L	Law's impact
A3	Tsai et al. (2020)	The privacy paradox and its implications for learning analytics	N/A	N/A	N/A	DC, DA, DS	Power relations; transparency and communication; trust; overlong data policy
A4	Ahn et al. (2021)	Co-Designing for Privacy, Transparency and Trust in K-12 Learning Analytics		\	No	G	Communication, trust
A5	Prinsloo and Slade (2015)	Student Privacy Self-Management: Implications for Learning Analytics	\		No	G	Communication, data policy
A6	Haythornthwaite (2017)	An Information Policy Perspective on Learning Analytics	\		No	G, L	What data to collect, store, anonymise and analyse
A7	Yan et al. (2022)	Scalability, Sustainability and Ethicality of Multimodal Learning Analytics	N/A	N/A	N/A	DC, DS	MMLA, searching for the wrong, abusive
A8	Arnold and Sclater (2017)	Student Perceptions of Their Privacy in Leaning Analytics Applications	N/A	N/A	N/A	DC, DA, DS	Communicate and seek advice

TABLE A1 (Continued)

Paper ID	Authors	Title	Proposed solutions		Evidence of application the solutions	Paper category	Summary of the paper in terms of RQs
			Legal and frameworks	Technical			
A9	Ocheja et al. (2018)	Connecting Decentralised Learning Records: A Blockchain-Based Learning Analytics Platform		\	No	DS	The data are isolated from each other and the next stage cannot access the history of the previous stage, hindering the development of learning analytics
A10	Li, Jung, et al. (2022)	Unpacking Instructors' Analytics Use: Two Distinct Profiles for Informing Teaching	N/A	N/A	N/A	G	Anonymisation, transparency and student data recourse issues
A11	Torre et al. (2020)	EdX Log Data Analysis Made Easy: Introducing ELAT: An Open-Source, Privacy-Aware and Browser-Based EdX Log Data Analysis Tool		\	\	DA	Save sensitive data locally for computing
A12	Prinsloo and Slade (2017)	An Elephant in the Learning Analytics Room: The Obligation to Act	\		No	G, L	Law, ethics's impact
A13	Li, Xing, and Leite (2022	Do Gender and Race Matter? Supporting Help-Seeking with Fair Peer Recommenders in an Online Algebra Learning Platform		\	\	DA	Related to prior attempts to solve these questions, fairer algorithms, improved trust
A14	Greller and Drachsler (2012)	Translating Learning into Numbers: A Generic Framework for Learning Analytics	\		No	G	Data misuse, anonymisation. Data ownership issues, data integration issues, data interpretation issues

(Continues)

TABLE A1 (Continued)

Paper ID	Authors	Title	Proposed solutions		Evidence of application the solutions	Paper category	Summary of the paper in terms of RQs
			Legal and frameworks	Technical			
A15	Ifenthaler et al. (2021)	Putting learning back into learning analytics: actions for policy makers, researchers, and practitioners	\		No	G	People do not understand learning analytics, learning analytics lacks various standards (privacy, transparency, whatever), old guidance policies need to be updated, learning analytics needs to be flexible and fit in with groups from different backgrounds
A16	West et al. (2020)	Do academics and university administrators really know better? The ethics of positioning student perspectives in learning analytics	\		No	G	All went to focus on the teacher body and did not focus on the students and did not communicate with them. Focusing on the student's perspective and presentation
A17	Kyle (2019)	Learning analytics and higher education: a proposed model for establishing informed consent mechanisms to promote student privacy and autonomy: Revista de Universidad y Sociedad del Conocimiento			No	DC, DA, DS	The more data is collected, the greater the scope, biometric data, social data

TABLE A1 (Continued)

Paper ID	Authors	Title	Proposed solutions		Evidence of application the solutions	Paper category	Summary of the paper in terms of RQs
			Legal and frameworks	Technical			
A18	Ifenthaler and Schumacher (2016)	Student perceptions of privacy principles for learning analytics	N/A	N/A	N/A	G, A	Students are very conservative and have to communicate that they want to be involved as equals, without power relations. Further empirical research is needed to clarify the conditions under which students are willing to share relevant data for the learning analytics system
A19	Lawson et al. (2016)	Identification of 'at risk' students using learning analytics: the ethical dilemmas of intervention strategies in a higher education institution	\		No	DC, DA, DS	The use of data outside of consent, the use of data by academics in ways that go beyond what the designers originally intended, the labelling of students—the misuse, the bringing about of unequal power relations
A20	Marshall et al. (2022)	Implementing next-generation privacy and ethics research in education technology	\	\	No (but plan to do it)	G	Lack of ethical and moral related tools in la, so there is gap

(Continues)

TABLE A1 (Continued)

Paper ID	Authors	Title	Proposed solutions		Evidence of application the solutions	Paper category	Summary of the paper in terms of RQs
			Legal and frameworks	Technical			
A21	Darvishi et al. (2022)	Incorporating AI and learning analytics to build trustworthy peer assessment systems		\	\	DA	Improve peer assessment with AI, make it more trust
A22	Mahmoud et al. (2022)	Factors Affecting the Deployment of Learning Analytics in Developing Countries: Case of Egypt	N/A	N/A	N/A	G, A	Lack of stakeholder capacity is a problem
A23	Tsai et al. (2021)	More Than Figures on Your Laptop: (Dis)trustful Implementation of Learning Analytics	N/A	N/A	N/A	G, A	Mistrust. The subjectivity of numbers, the fear of diminished power and the design and implementation of LA
A24	Prinsloo & Kaliisa (2022)	Data privacy on the African continent: Opportunities, challenges and implications for learning analytics	\		No	G, L	The legislative development is very slow and naturally hinders this
A25	Whitelock-Wainwright et al. (2020)	Assessing the validity of a learning analytics expectation instrument: A multinational study	N/A	N/A	N/A	A	Investigate privacy attitude
A26	Cormack and Reeve (2022)	Developing a Code of Practice for Using Data in Wellbeing Support	\		\	G	Inappropriate data visibility can create a lack of trust and a sense of surveillance
A27	Viberg et al. (2022)	Privacy in LA Research: Understanding the Field to Improve the Practice	N/A	N/A	N/A	G, C	Privacy in learning analytics is not well defined. Insufficient research hinders development

TABLE A1 (Continued)

Paper ID	Authors	Title	Proposed solutions		Evidence of application the solutions	Paper category	Summary of the paper in terms of RQs
			Legal and frameworks	Technical			
A28	Mutimukwe et al. (2022)	Students' privacy concerns in learning analytics: Model development	\	N/A	\	G, A	The reasons affecting students' disclosure of personal information are unclear, hindering further progress in the advancement of privacy research
A29	Hutt et al. (2022)	Controlled outputs, full data: A privacy-protecting infrastructure for MOOC data		\	\	DA, DS	Insufficient anonymisation and de-identification
A30	Vatsalan et al. (2022)	Privacy risk quantification in education data using Markov model		\	\	DA	The inverse relationship between data availability and privacy, where attackers in learning analytics are generally not interested in identifying individuals, but in re-identifying as many records as possible from the dataset
A31	Prinsloo et al. (2022)	The answer is (not only) technological: Considering student data privacy in learning analytics	\		No	DC, DA, DS	Limitations of technical issues, limitations of PET
A32	Viberg et al. (2022)	Exploring students' expectations of learning analytics: A person-centred approach	N/A	N/A	N/A	A	Investigate privacy attitude
A33	Mavroudi et al. (2021)	Teachers' Views Regarding Learning Analytics Usage Based on the Technology Acceptance Model	N/A	N/A	N/A	A	Investigate privacy attitude

(Continues)

TABLE A1 (Continued)

Paper ID	Authors	Title	Proposed solutions		Evidence of application the solutions	Paper category	Summary of the paper in terms of RQs
			Legal and frameworks	Technical			
A34	Yacobson et al. (2021)	De-identification is insufficient to protect student privacy, or,ÄîWhat can a field trip reveal?	N/A	N/A	N/A	DS	De-identified or may be identified, anonymisation is not good
A35	Kumar et al. (2020)	Google classroom: insights from Malaysian higher education students,Äô and instructors,Äô experiences	N/A	N/A	N/A	A	Students are conservative
A36	Salas-Pilco and Yang (2020)	Learning analytics initiatives in Latin America: Implications for educational researchers, practitioners and decision makers	N/A	N/A	N/A	L	Legislative developments are very slow and hinder the resolution of related issues
A37	Duin and Tham (2020)	The Current State of Analytics: Implications for Learning Management System (LMS) Use in Writing Pedagogy	\		No	G	Students are not able to opt out of the data, students do not have good consent, students have too little involvement and rights in it, students are objectified into the data, literacy of various stakeholders is lacking and training is inadequate. Institutions are in LMS situations where students are not given the option to consent to use; the LMS is seen as similar to a designated textbook. Data policies are bad and fail to communicate policies in all directions. Poor anonymisation. Lack of communication

TABLE A1 (Continued)

Paper ID	Authors	Title	Proposed solutions		Evidence of application the solutions	Paper category	Summary of the paper in terms of RQs
			Legal and frameworks	Technical			
A38	Whitelock-Wainwright et al. (2019)	The Student Expectations of Learning Analytics Questionnaire	\		\	A	Investigate privacy attitude
A39	Sun et al. (2019)	It's My Data! Tensions Among Stakeholders of a Learning Analytics Dashboard	\		No	DC, DA, DS, A	Inadequate presentation of data analysis sources and processes, non-transparent algorithms, poor auditing and impact on confidence. Does not involve all stakeholders
A40	Slade et al. (2019)	Learning analytics at the intersections of student trust, disclosure and, benefit	\		No	DC, DA, DS, A	No student control over data collection, excessive scope of data collection, compromised predictive privacy, non-transparent process, compromised student rights
A41	Zeide (2017)	The Structural Consequences of Big Data-Driven Education	N/A	N/A	N/A	DC, DA, DS	Uni/Schools rely too much on outside vendors, regulations are inadequate, and existing ones provide only minimal protection
A42	Wang (2016)	Big Opportunities and Big Concerns of Big Data in Education	N/A	N/A	N/A	DC, DA, DS	Data silos, data misuse, bad consent
A43	Pardo and Siemens (2014)	Ethical and privacy principles for learning analytics	\		No	DC, DA, DS	Transparency, student ownership of data, access, accountability and assessment

(Continues)

TABLE A1 (Continued)

Paper ID	Authors	Title	Proposed solutions		Evidence of application the solutions	Paper category	Summary of the paper in terms of RQs
			Legal and frameworks	Technical			
A44	Kyle (2019)	Advising the whole student: eAdvising analytics and the contextual suppression of advisor values	N/A	N/A	N/A	DC, DA, DS, A	Lack of explanation, no credibility, uncertainty about validity and ethics.
A45	West et al. (2020)	Perspectives from the stakeholder: Students' views regarding learning analytics and data collection	N/A	N/A	N/A	DC, DA, DS, A	Sensitive data collected (geosocial data and such), lack of transparency. Not student-centred, too busy taking care of researchers. No right for students to join and exit
A46	Khalil et al. (2018)	User Consent in MOOCs—Micro, Meso, and Macro Perspectives	N/A	N/A	N/A	DS	Personal Data Cross-Border Issues, Beyond Legislation
A47	Silvola et al. (2021)	Expectations for supporting student engagement with learning analytics: An academic path perspective	N/A	N/A	N/A	A	Investigate privacy attitude

Note: C = colleges, HE, FE, anything post-school; M = MOOCs, online courses, khan academy; S = schools; G = general; subject = specific subject, such as maths, STEM.

APPENDIX 2

TABLE A2 Search strings for different databases.

Name of the database	Search strings
Web of Science	(((((((((TS = ('learning analytics' AND 'privacy')) OR TS = ('educational data mining' AND 'privacy')) OR TS = ('learning analytics' AND 'data protection')) OR TS = ('learning analytics' AND 'privacy preserving')) OR TS = ('learning analytics' AND 'trustworthy')) OR TS = ('learning analytics' AND 'responsible')) OR TS = ('educational data mining' AND 'data protection')) OR TS = ('educational data mining' AND 'privacy preserving')) OR TS = ('educational data mining' AND 'trustworthy')) OR TS = ('educational data mining' AND 'responsible')
Scopus	(TITLE-ABS-KEY ('learning analytics' AND 'privacy') OR TITLE-ABS-KEY ('learning analytics' AND 'data protection') OR TITLE-ABS-KEY ('learning analytics' AND 'privacy preserving') OR TITLE-ABS-KEY ('learning analytics' AND 'trustworthy') OR TITLE-ABS-KEY ('learning analytics' AND 'responsible') OR TITLE-ABS-KEY ('educational data mining' AND 'privacy') OR TITLE-ABS-KEY ('educational data mining' AND 'data protection') OR TITLE-ABS-KEY ('educational data mining' AND 'privacy preserving') OR TITLE-ABS-KEY ('educational data mining' AND 'trustworthy') OR TITLE-ABS-KEY ('educational data mining' AND 'responsible'))
Proquest	noft (learning analytics or educational data mining) AND noft (privacy or data protection or privacy preserving or trustworthy or responsible)
LAK	'query': {Abstract:('learning analytics' AND 'privacy') OR Abstract:('learning analytics' AND 'data protection') OR Abstract:('learning analytics' AND 'privacy preserving') OR Abstract:('learning analytics' AND 'trustworthy') OR Abstract:('learning analytics' AND 'responsible') OR Abstract:('educational data mining' AND 'trustworthy') OR Abstract:('educational data mining' AND 'responsible') OR Abstract:('educational data mining' AND 'privacy preserving') OR Abstract:('educational data mining' AND 'data protection') OR Abstract:('educational data mining' AND 'privacy')} 'filter': {Conference Collections: LAK: Learning Analytics and Knowledge}