Towards an Operational Responsible AI Framework for Learning Analytics in Higher Education

ANONYMOUS AUTHOR(S)

Universities are increasingly adopting data-driven strategies to enhance student success, with AI applications like Learning Analytics (LA) and Predictive Learning Analytics (PLA) playing a key role in identifying at-risk students, personalising learning, supporting teachers, and guiding educational decision-making. However, concerns are rising about potential harms these systems may pose, such as algorithmic biases leading to unequal support for minority students. While many have explored the need for Responsible AI in LA, existing works often lack practical guidance for how institutions can operationalise these principles. In this paper, we propose a novel Responsible AI framework tailored specifically to LA in Higher Education (HE). We started by mapping 11 established Responsible AI frameworks, including those by leading tech companies, to the context of LA in HE. This led to the identification of seven key principles such as transparency, fairness, and accountability. We then conducted a systematic review of the literature to understand how these principles have been applied in practice. Drawing from these findings, we present a novel framework that offers practical guidance to HE institutions and is designed to evolve with community input, ensuring its relevance as LA systems continue to develop.

CCS Concepts: • **Applied computing → Learning management systems**.

Additional Key Words and Phrases: Responsible Artificial Intelligence, Learning Analytics, Higher Education

ACM Reference Format:
Anonymous Author(s). 2024. Towards an Operational Responsible AI Framework for Learning Analytics in Higher Education. In *Proceedings of The International Conference on Learning Analytics & Knowledge (LAK' 25).* ACM, New York, NY, USA, 16 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Learning Analytics (LA)[1] are becoming increasingly central to higher education institutions worldwide. LA systems utilise data to identify at-risk students, support student development, provide personalised and timely feedback, support self-reflection, and enhance the quality of learning and teaching [47, 49].[2]

However, the adoption of AI-powered systems within Higher Education (HE) brings with it a range of ethical concerns. Issues like algorithmic bias, lack of transparency, and potential misuse can have serious implications. For example, systems that automatically identify students at risk may suffer from algorithmic biases and disproportionally under-detect students from certain minority groups, leading to those students not receiving equivalent support to the majority group [7]. Other concerns include the need for these automated systems to be able to explain their decisions or the safe and transparent usage of student data. Addressing these issues is crucial to ensure that AI technologies are deployed in a manner that is fair, equitable, and responsible.

[1]When we refer to Learning Analytics in this paper we consider also Predictive Learning Analytics and any other Learning Analytics approaches supported by Artificial Intelligence
[2]LA definition https://www.solaresearch.org/about/what-is-learning-analytics/

In response to the need of addressing the ethical concerns of AI deployment, tech companies and other organisations have developed Responsible AI frameworks to guide the design and development of AI. These frameworks provide guiding principles such as fairness, transparency, accountability, and data privacy to ensure that AI systems are built in a way that minimises harm and maximises societal benefit. However, while Responsible AI frameworks are essential for guiding the design, development and deployment of AI technologies, these frameworks are frequently designed following high-level concepts and principles that can be applied to any AI application, without considering the specificities of the technology or the environment in which the AI system will be deployed. Similarly, numerous works have emerged from the LA community emphasising the importance of ensuring that LA systems adhere to ethical principles [8, 11, 18, 44]. However, these principles are rarely discussed in practical terms, leaving Higher Education Institutions without clear guidance on how to operationalise them effectively [6, 59].

To address this gap, this paper introduces a novel Responsible AI framework specifically designed for Learning Analytics (LA) in HE. To develop this framework, we first analysed eleven established Responsible AI frameworks, including those from leading technology companies like Google and Microsoft, and mapped them to the context of LA in HE. From this analysis, we identified seven common Responsible AI principles including: fairness and bias, transparency, privacy, accountability, explainability, safety and security. We then conducted a systematic literature review to explore how existing studies addressed these principles in practice within the LA domain. By synthesising the solutions, challenges, and lessons learned from these studies, we propose a new framework aimed at guiding HE institutions in the responsible implementation of LA systems. Our study is motivated by the following questions:

- RQ1: To what extent do existing Responsible AI frameworks address the specific needs and challenges of Learning Analytics in Higher Education?
- RQ2: Which Responsible AI principles from existing frameworks are applicable to the context of Learning Analytics in Higher Education?
- RQ3: How have previous Learning Analytics studies incorporated or addressed Responsible AI principles in practice?

Building on the answers to our research questions, we propose the development of a tailored Responsible AI framework for LA in HE. This framework is designed to offer practical, actionable guidance to HE institutions, addressing the ethical, legal, and social complexities inherent in LA systems. We envision this framework as a dynamic resource that evolves with the community, incorporating real-world examples of LA implementations and continuously adapting based on shared challenges and lessons learned. We believe this framework will serve as a valuable asset to the LA community, providing a much-needed tool for operationalising Responsible AI principles in practical ways.

2 Motivation

Higher Education Institutions (HEIs) face several ethical challenges when integrating Artificial Intelligence (AI) into their operations, teaching, research, and administrative functions. These challenges stem from the complexity of AI technologies, the sensitivity of academic and student data, and the societal implications of widespread AI use in education. The spectrum of technologies used is also broad [12], from Generative AI applications that help to generate new curricula, to AI that can monitor attendance, to LA solutions that could identify students at-risk.

LA systems in particular introduce numerous ethical challenges, especially given their growing use in HE to enhance student success and institutional efficiency. These challenges often stem from the use of vast amounts of student data and algorithmic predictions that can impact decision-making in educational settings. Below is a brief discussion of some of the key ethical issues associated with the use of LA in HE. For a broader overview of the problem, the reader is directed to the following literature [1, 18, 35, 37, 45, 51].

One key ethical issue in LA systems is bias. For instance, a predictive model might unfairly classify students from certain demographics as at-risk or not, leading to unequal treatment and opportunities. If LA decisions guide resource allocation or interventions, some students may receive more support, while others are overlooked. LA systems often inherit bias from the training data [38], reinforcing social inequalities. Bias can also be introduced during data processing, resulting in different levels of support for students based on factors like race, gender, disability, or socioeconomic status [7]. It is also important to consider that the predictions generated by LA tools can bias the behaviour of students or instructors (e.g, by deciding not to continue a course if the prediction is negative in the case of students, or to grade a student unfairly by over-relying on the output of an algorithmic system) [57]. This shows that biases in LA systems are both technical and social, highlighting the need for Responsible AI principles that address both dimensions.

Even when the LA systems are not biased, they are never 100% right, they make mistakes. LA systems often fail to consider the full context of a student's life, such as personal struggles, cultural differences, or family obligations, which can affect academic performance but are not easily captured by quantitative data [25]. Those inaccurate predictions can lead to students either receiving unnecessary interventions or being deprived of necessary support. It is also necessary to reskill staff to the use of LA systems, ensuring these systems are properly understood and used as intended.

Many LA systems function as "black-boxes" [37], meaning their decision-making processes are not easily interpretable. This may lead to mistrust among students and staff. If students don't understand how LA decisions that affected them are made they may feel mistreated. Similarly, if staff does not understand the reasoning behind LA decisions, they may decide not to use them, which could lead to disadvantaging their students.

It is also unclear where the failure lies when the LA system makes an error, and that error leads to harm.[3] Ensuring accountability in LA systems is difficult as they involve various stakeholders, including data scientists, administrators, faculty, third-party vendors, and students. With so many parties involved, it can be difficult to determine who is ultimately accountable for decisions made based on LA insights. For the same reason, when biases and predictive errors occur it is challenging to hold a specific party accountable.

LA systems are based on the analysis of vast amounts of data on students, and staff, including academic records, attendance, or behavioural data from learning management systems. Students may not be fully aware of which data is being collected, for which purposes, and whether it is being shared, as there may be inadequate consent mechanisms. Students may also not have the option to opt out of being monitored by LA systems, which raises ethical questions about autonomy and the right to control one's own data. It is also often unclear who owns the data — the institution, the student, or even third-party providers of analytics platforms. This lack of clarity can lead to disputes about how data can be used. Also, HE institutions are often targets of cyber attacks[4], increasing the risk of sensitive information from students and staff being exposed or misused.

The use of LA systems by different HE stakeholders can also constitute data misuse. It is often unclear within HE institutions which roles should have access to which type of data, and when. Following the principle of data minimisation, staff members should have access to the minimum amount of data required to do their jobs. However, this requires role adaptations of the different LA systems, which are not always possible, specially if the LA solution is acquired from a third-party vendor.

[3]https://www.wired.com/story/alevel-exam-algorithm/
[4]https://www.gov.uk/government/statistics/cyber-security-breaches-survey-2023/cyber-security-breaches-survey-2023-education-institutions-annex

Concerns have been raised about the psychological impact of LA on both students and staff [46]. LA systems can create a sense of constant surveillance, where individuals feel closely monitored. This awareness can affect how staff support students, as teachers might adjust their methods to align with data-driven insights, potentially limiting their teaching style. Similarly, students may limit their creativity and exploration if LA systems restrict their learning options through algorithmic recommendations. This can undermine students' freedom to explore diverse knowledge paths.

Institutions may see LA as a cost-cutting solution. For example, using LA to streamline the effectiveness of academics may lead to job losses or reduce the number of hours within contracts. How and where institutions invest those savings from the use of AI-driven applications also constitutes an ethical dilemma. It is also important to acknowledge the digital divide. As more LA solutions are put in place benefiting both, students and staff, students without access to technology may be at a disadvantage. Similarly, institutions in wealthier countries or regions may have more resources to implement LA solutions effectively, leading to a gap between global education systems.

In addition, many regions in the world are still in the process of developing comprehensive legal frameworks to govern AI [2]. HE institutions therefore need to stay up to date with local, national and international regulations. Also, since different regions may have varying legal standards it makes it difficult for institutions with international students to apply consistent ethical standards. These challenges show the importance of creating a practical framework that could guide HE institutions towards the design, development, deployment and use of their LA solutions.

3 Analysing Existing Responsible AI Frameworks

To address our first two research questions: (i) *RQ1: To what extent do existing Responsible AI frameworks address the specific needs and challenges of Learning Analytics in Higher Education?* and (ii) *RQ2: Which Responsible AI principles from existing frameworks are applicable to the context of Learning Analytics in Higher Education?*, we followed a comparative analysis of eleven well-established Responsible AI frameworks.

Our methodology began by clearly defining our research objectives and scope. To ensure a comprehensive and diverse set of frameworks, we conducted an extensive search using multiple queries through Google's search engine. Key search terms included 'Responsible AI framework,' 'ethical AI adoption,' 'ethical AI in education,' and 'responsible AI in education.' For each query, we reviewed the top 30 results to capture a broad spectrum of frameworks across industry, government, and the education sector. Additionally, we integrated findings from recent literature reviews of Responsible AI frameworks [6, 55], ensuring the inclusion of widely recognised frameworks from leading technology companies like Microsoft, Amazon, and Google.

Given that many Responsible AI frameworks are not published as traditional academic papers, but rather proposed by industry, government, or third-sector organisations, we opted not to limit our search to scholarly databases. This approach allowed us to capture the most relevant and practical frameworks beyond academic literature.

For inclusion in our analysis, we applied the following eligibility criteria:

- Documents must be written in English.
- Frameworks must address the ethical use of data or software development practices, considering the ethical, legal, and social challenges related to AI design, development, and adoption.
- Documents describing policies, guidelines, or codes of practice, rather than full frameworks, were included if: (i) they specifically targeted the education sector, or (ii) they were produced by leading technology companies.

This systematic process ensures a robust and diverse dataset, allowing for a thorough analysis of Responsible AI principles relevant to Learning Analytics in Higher Education. As detailed in Table 1 we identified eleven relevant

Table 1. Responsible AI initiatives: frameworks, guidelines, policies. AI = Artificial Intelligence, ML = Machine Learning, LA = Learning Analytics, PLA = Predictive Learning Analytics.

#	Organisation name	Document name	Focus	Doc. Type	Year	Number of Principles	Context
1	NIST	AI Risk Management Framework (NIST) [54]	AI	Framework	2023	7	Non-sector-specific AI risk-oriented (design, development, release, and use)
2	Microsoft	Microsoft Responsible AI Standard [36]	AI	Framework	2022	6	AI in industry Product development
3	The Institute for Ethical AI in Education - IEAIE	The Ethical Framework for AI in Education [19]	AI	Framework	2021	9	UK AI in Education
4	Alan Turing Institute	Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector [30]	AI	Guideline	2019	4	UK AI in the public sector
5	The Open University	Data Ethics Policy - OU (2023) [60]	AI, ML, Data	Policy	2023	4	UK. The Open University Data management
6	IBM	AI ethics at IBM [27]	AI	Principles description	2024	5	AI in industry Product development
7	Google	Google's AI Principles [21]	AI	Principles description	2023	7	AI in industry Product development
8	Amazon	Building AI responsibly at AWS [5]	AI	Principles description	NE	8	AI in industry Product development
9	JISC	Code of practice for learning analytics [28]	LA	Code of practice	2023	7	UK Learning Analytics in HEI
10	ICDE - International Council for Open and Distance Education	Global Guidelines: Ethics in Learning Analytics [52]	LA, PLA	Guideline	2019	10	Global Learning Analytics in Education
11	University of Edinburgh	Learning Analytics Principles and Purposes [39]	LA	Policy	2017	7	UK. University of Edinburgh Learning Analytics in HEI

initiatives launched by organisations in different domains. The table details: (i) the originating organisation, (ii) the document's name and URL, (iii) the primary focus of the document (Artificial Intelligence, Machine Learning, Data, Learning Analytics, or Predictive Learning Analytics), (iv) the document type (framework, policy, principles, code of practice, or guidance), (v) the year of release, (vi) the number of Responsible AI principles discussed, and (vii) the context (domain, country, sector) in which the principles are applied. Although marginally relevant, we excluded the SHEILA Framework [55], as it did not fully meet our eligibility criteria, being primarily focused on strategic planning and policy processes for Learning Analytics.

We then conducted a *SWOT (Strengths, Weaknesses, Opportunities, and Threats)* analysis of the eleven selected documents. The Strengths criterion evaluated how well each document addressed the unique needs and challenges of Learning Analytics in Higher Education. Weaknesses identified areas where the documents could be improved to better meet these needs. Opportunities considered how the insights and lessons from each document could contribute to developing a tailored Responsible AI framework for LA in HE. Finally, Threats examined any potential challenges or competition highlighted by the documents that could impact the development of such a framework.

Our analysis revealed that most frameworks adopt a holistic view of AI systems, often focusing on machine learning algorithms and large-scale data science without specific consideration of LA in HE. While some documents, such as those by [28, 39, 52, 60], provide relevant high-level principles and guidelines, **they often lack concrete, actionable steps, tools, or measurable practices that HE institutions could implement in their LA efforts**.

However, we identified significant opportunities within these frameworks. Many of the principles outlined, if refined and tailored specifically to the LA context in HE, could serve as a solid foundation for the development of a robust and applied Responsible AI framework. We then proceeded to analyse the Responsible AI principles mentioned across these documents, extracting their commonalities and unique aspects. This allowed us to identify key principles that could be adapted to meet the needs of LA in HE and inform the creation of our framework.

Table 2 presents the common Responsible AI principles identified across the 11 analysed documents. Each row corresponds to one of the analysed documents and highlights the Responsible AI principles discussed within it. It is important to note that while different documents may use slightly varied terminology, they often refer to the same underlying principles. In the top row, we list the common Responsible AI principles identified across all documents, along with the standardised names we have selected for each principle in this study. Subsequent rows display the individual documents, indicating which of the common principles they cover and how those principles are named in the respective document. Note that certain documents may group two or more principles under a common one (e.g., Transparency and accountability [52] or Privacy and Security [36]). In those cases, the principle appears twice in the table under the different individual ones. The final column, "Unclassified," includes principles that are either unique to a specific policy, guideline, or framework, or that appear only in a small subset of the analysed documents.

Building on the various definitions of common principles found in the 11 analysed documents, we propose here a set of definitions specifically tailored to the context of Learning Analytics in the Higher Education sector:

- **Fairness and Bias**: Ensuring that LA systems are free from biases that could disadvantage certain student groups, such as minorities or underrepresented communities. Fairness in this context means that predictive models do not disproportionately label or classify students based on sensitive attributes like gender, race, or socio-economic background, that any interventions derived from analytics are equitably distributed among all students, and that the outputs of LA systems and are not biasing students and staff in their decisions.
- **Transparency**: Providing clear and accessible information to all stakeholders (students, educators, administrators) about how LA systems operate, including data collection, algorithms used, and the decision-making processes behind predictive analytics. This principle emphasises the importance of openness in the system's design, implementation, and outcomes, ensuring that users understand how predictions and classifications are made and which and how their data is being utilised.
- **Accountability**: Establishing clear lines of responsibility for the design, deployment, and outcomes of LA systems. This includes holding institutions, technology providers, and stakeholders accountable for ensuring ethical practices, addressing unintended consequences, and mitigating harms caused by LA-driven decisions. In the context of Higher Education, accountability particularly ensures that institutions take responsibility for the accuracy and fairness of predictive analytics outcomes and their impact on students.
- **Privacy**: Protecting the personal data of students and staff, ensuring that LA systems comply with legal and ethical standards around data privacy, such as GDPR.[5] This includes collecting only necessary data, securely storing it, and ensuring that students and staff have control over how their data is used. Privacy also involves limiting the sharing of personal data to authorised individuals or systems and ensuring that predictive models do not intrude on the personal lives of students or staff.
- **Security**: Safeguarding LA systems from data breaches, hacking, and unauthorised access to sensitive student and staff data. This principle focuses on implementing robust technical safeguards to protect the integrity, confidentiality, and availability of data and ensuring that predictive models are secure from manipulation or misuse. Security is particularly important in HE, where large volumes of sensitive information are processed.
- **Explainability**: Ensuring that the predictions and decisions made by LA systems can be understood by non-expert users, such as HE staff and students. Explainability in this context involves providing clear, understandable explanations for how specific predictions and decisions were reached and offering insights into the variables that contributed to those outcomes.

[5]https://gdpr-info.eu/

Table 2. Responsible AI common principles identified across analysed documents.

Organisation's name	Common principles							
	Fairness and bias	**Transparency**	**Accountability**	**Privacy**	**Security**	**Explainability**	**Safety**	**Unclassified**
Alan Turing Institute	Fairness	Transparency	Accountability	-	-	-	-	Sustainability
The Open University	Fairness	Transparency	Accountability	-	-	Explainability	-	-
ICDE	Inclusion	- Communications - Transparency	Institutional responsibility and obligation to act	- Consent - Data ownership and control	- Accessibility of data - Validity and reliability of data	-	-	- Cultural values - Student agency and responsibility
Google	- Avoid creating or reinforcing unfair bias - Be socially beneficial	-	Be accountable to people	Incorporate privacy design principles	-	-	- Be built and tested for safety - Be made available for uses that accord with these principles	- Uphold high standards of scientific excellence
IBM	Fairness	Transparency	-	Privacy	Robustness	Explainability	-	-
IEAIE	Equity	*- Informed Participation* *- Transparency and Accountability*	*- Transparency and Accountability*	Privacy	-	-	- Achieving Educational Goals	- Administration and Workload - Autonomy - Ethical Design - Forms of Assessment
JISC	Minimising adverse impacts	- Transparency, legal basis and consent - Access	Responsibility	- Privacy - Stewardship of data	- Validity	-	-	-
University of Edinburgh	- Beneficial to students - Fairness and bias	Transparent	- Be accountable to people - Governance	Privacy	-	Explainability	-	-
Microsoft	- Fairness - Inclusiveness	Transparency	Accountability	*Privacy and Security*	*- Privacy and Security*	-	Reliability and Safety	-
Amazon	Fairness	Transparency	- Governance	*Privacy and security*	*- Privacy and security* *- Veracity and robustness*	- Explainability - Controllability	Safety	-
NIST	Fair with harmful bias managed	*- Accountable and transparent*	*Accountable and transparent*	Privacy-enhanced	- Secure and resilient - Valid and reliable	Explainable and interpretable	- Safe	-

- **Safety**: Ensuring that LA systems are designed to minimise harm to staff and students, whether psychological, emotional, or academic. Safety in this context involves evaluating the potential risks of using LA, such as flawed predictions, over-reliance on predictions or biases derived from the human perceptions of those predictions, and ensuring that interventions based on analytics are supportive rather than punitive. It also means ensuring that these systems do not create undue stress or pressure on students and staff.

Unclassified principles such as Sustainability, Cultural values, or Student agency and Responsibility are not considered in our proposed framework, because they were either less explicitly defined in the context of Learning Analytics or do not directly align with the immediate operational needs and challenges identified in our analysis of existing frameworks. In the following section, our goal has been to review existing works on LA (tools, applications, use cases) and extract valuable lessons learned, including best practices, challenges, opportunities, in relation to the seven identified principles.

4 Analysing LA works with respect to Responsible AI principles

We address in this section the third research question *(RQ3): How have previous Learning Analytics studies incorporated Responsible AI principles in practice?* To answer this, we conducted a systematic literature review of relevant studies.

4.1 A Systematic Literature Review

We initiated our systematic literature review by identifying key terms derived from our research questions and the ethical principles discussed (see Table 2). We compiled a list of synonyms to create a comprehensive search string using Boolean operators (AND, OR). The structured search query was formulated as follows: {domain of interest} + {area of implementation} + {principles} + {focus}. The resulting search string was defined as *('learning analytics' OR 'predictive learning analytics') AND ('higher education') AND (fairness OR transparency OR privacy OR accountability OR safety OR explainability OR ethics OR 'responsible AI') AND (framework OR guideline OR policy OR 'code of practice' OR principles OR 'best practice' OR implications OR 'lessons learn').*

This search string was applied across three digital libraries—ERIC https://eric.ed.gov/, SCOPUS https://www.scopus.com, and ACM https://dl.acm.org —selected for their relevance to our study. Searches were conducted on titles, abstracts,

Table 3. Inclusion and exclusion criteria

Inclusion criteria	
1	Studies that describe ethical concerns in the adoption LA or PLA by HEI
2	Studies that describe ethical challenges or address specific ethical principles for adoption of LA
3	Studies that discuss, suggest or have implemented controls, guidelines or policies for ethical adoption of LA.
Exclusion criteria	
1	Studies written in a language other than English
2	Conference abstracts and editorials
3	Studies that do not meet any of the inclusion criteria
4	Studies that focus on other education organisations other than HEI

and keywords. We obtained: ERIC (54), Scopus (70) and ACM (110) results from each library. Before selection, we removed duplicate results. Subsequently, we established robust inclusion and exclusion criteria based on our research questions (see Table 3). We focused on studies that explore the responsible adoption of LA. This included papers detailing lessons learned from LA implementations, identifying challenges faced by Higher Education Institutions (HEIs), and proposing strategies to address these issues, including policy and guideline development.

The study selection followed a three-stage process: (a) reviewing the titles and abstracts, (b) reading the introductions and conclusions, and (c) evaluating the full text. At each stage, documents were categorised into three groups: 'important,' 'unsure,' and 'not relevant.' Papers classified as 'unsure' were reviewed collectively by all authors to reach a final decision. Additionally, we employed a snowballing technique to manually include certain papers identified through key citations.

A total of 234 articles were initially identified through database searches. After removing duplicates (N=37), 197 articles proceeded to the three-stage selection process. Ultimately, 167 studies were excluded, and 30 were selected. An additional 15 studies were manually added, resulting in a final total of 45 studies for this literature review.

4.2 Findings

The publication years of the selected papers range from 2013 to 2024. The selected studies include journal articles, book chapters, and conference papers. We first classified the selected works according to the seven principles identified in Section 3. The classification was done in two steps: (a) reading the titles and abstracts to identify the potential Responsible AI principle(s) addressed by each study, and (b) reviewing the full text, with a focus on the methodology, results, and discussion sections, to determine the primary and secondary principles covered. As shown in Table 4, 40% of the analysed studies primarily focus on Privacy. The second most commonly addressed principle is Transparency. We also created a 'Various principles' category for works focusing on more than two principles. In the following section, we review the selected works, discussing how they have applied Responsible AI principles in practice, the challenges encountered, and the lessons learned.

Table 4. Distribution of Studies Across Responsible AI Principles

Focus	Privacy	Transparency	Fairness/Bias	Accountability	Safety	Security	Explainability	Various
Primary	18	9	7	3	1	4	1	2
Secondary	5	6	2	0	2	0	0	6

4.2.1 Accountability. The accountability principle mandates that institutions take responsibility for decisions generated by predictive analytics systems. All stakeholders—such as HEI directors, managers, and data scientists—must understand their roles throughout the lifecycle of Learning Analytics (LA) systems. For instance, [40] highlight the importance of accountability in the design phase, particularly concerning data management. Similarly, [3] emphasise that unclear governance policies can undermine trust in LA systems. [61] provide operational criteria for accountability, such as creating clear documentation of roles for developers and users of LA dashboards. Compliance with GDPR also plays a critical role, in defining key responsibilities among data controllers, processors, and subjects. The literature reveals two dimensions of accountability: forward-looking responsibility, which focuses on identifying stakeholders and their roles, and backwards-looking responsibility, which involves acknowledging the outcomes of LA systems. Overall, addressing accountability is a significant challenge for HEIs. The accountability principle also overlaps with principles such as Safety, Security and Transparency. In their work, [43] suggest that implementing audit measures (mechanisms to inspect what, how, and why predictions and automated decisions were made) could not only improve algorithmic explainability but also support who (developers, admin staff, users) can be held accountable. In summary, the reviewed papers did not provide clear guidelines regarding the allocation of responsibilities among roles or the specific actions to be taken (e.g., escalation procedures) if harm arises from the use of LA systems.

4.2.2 Safety. The Alan Turing Institute's guidelines for safe AI systems [30] highlight accuracy, reliability, and robustness as essential technical characteristics necessary to ensure AI functions safely and avoids harmful outcomes. In the context of LA, components such as data, decision algorithms, and applications (e.g., dashboards, alert systems) must be designed, deployed, and monitored to minimise errors, ensure consistent behaviour aligned with LA goals, and produce trustworthy predictions. To enhance safety it is crucial to provide end users—students, academic staff, and administrators—with clear documentation outlining the responsible use of LA systems, including guidance on interpreting data and engaging with at-risk students [26]. Developers should also receive clear conceptual frameworks and usage guidelines for LA systems [54]. Reliability in decision algorithms is critical; thus, establishing measurable goals for accuracy and expected model performance is vital. Specific considerations regarding acceptable error rates and performance metrics should be implemented. It is important to acknowledge that various factors—such as the choice of machine learning algorithms, missing data, and data noise—can influence predictions. Therefore, setting checkpoints for training and testing data is recommended [36]. Additionally, ensuring data accuracy is paramount; research by [41] underscores the need for HEI policies that guarantee access to up-to-date student data to prevent unreliable predictions. Despite these recommendations, there seems to still be a big gap in the literature on methods and actions that HEIs could put into practice to ensure that LA systems minimise harm to staff and students, whether psychological, emotional or academic.

4.2.3 Security. The security principle includes the implementation of technical, administrative and physical controls to mitigate risks and prevent information assets from being accidentally or deliberately compromised. Key aspects[6] include ensuring confidentiality by controlling access to sensitive data about students and staff, maintaining integrity by preventing unauthorised data alteration or deletion, and guaranteeing availability for authorised users to access LA systems promptly [29]. Literature around Security indicates a dual focus on privacy and data security concerns [10, 17, 53]; and both principles are strongly interlinked. For instance, implementing strong security measures (like data anonymisation) is crucial for protecting personal information, thus supporting privacy. Similarly, adhering to data privacy regulations comprises technical measures and policies to restrict unauthorized access or disclosure of personal information. Given that educational institutions collect extensive socio-demographic and progress data from students,

[6]GDPR Article 32:1b: Security measures must *"ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services"*

any breach of this information could have detrimental effects on individuals and institutions alike. Therefore, the acquisition, processing, storage, and disposal of personal and sensitive data must adhere to strict legal and regulatory compliance standards [29]. In this context, [50] presents a compilation of ethical data governance considerations, which encompass data security aspects such as data process (public, sensitive, personal, high-risk), data storage (local vs. remote), and data audit plans. In the same line, the work by [16] compiles reference questions for managers and decision-makers to consider when implementing LA data security. Other researchers, however, who have looked at security in LA, have used the General Data Protection Regulation (GDPR) as the main guidance to protect personal data; for example, [10] and [53] propose data security aligned with GDPR. While much of the existing literature emphasises legal frameworks and identifies data security issues, there remains a gap in specific operational and technical controls, which can hinder HEIs from effectively delivering the security principle within their LA solutions. Security controls should be tailored to align with organisational objectives and technological capabilities and evolve in response to advancements in technology, the latter being a great challenge for HEIs.

4.2.4 Fairness and bias. The principles of fairness and bias have garnered considerable attention within the Learning Analytics (LA) community, exemplified by dedicated workshops such as FairLAK[7]. Research has underscored the necessity of evaluating LA algorithms for biases and implementing effective mitigation strategies [44]. Some studies focus on fairness metrics to assess biases affecting specific groups, including minority ethnic students [7]. Advancing this work, Deho and colleagues [15] shifted from merely detecting bias to actively mitigating it, conducting a comparative evaluation of selected bias mitigation approaches. Their findings reveal that fairness lacks a universal definition, making the choice of definition a crucial first step in determining appropriate mitigation strategies. Furthermore, both studies indicate that enhancing fairness in LA systems may not need a compromise on predictive performance. [44] recommend the de-weighting or removal of sensitive attributes (and potential proxies, such as socio-economic status) from the training data of LA algorithms. However, Deho et al. [13] clarify that the inclusion or exclusion of a protected attribute impacts performance and fairness only if it is correlated with the target label and deemed significant. Importantly, LA models that demonstrate fairness based on historical data may not maintain this fairness when applied to current or future datasets. Deho and colleagues therefore advocate for ensuring robustness against dataset drifts prior to deployment [14]. In a recent survey on biases in education, Li et al. [31] emphasised the importance of considering intersectionality—how multiple sensitive attributes like gender and ethnicity interact—when evaluating algorithmic bias. They cautioned that applying fairness metrics to inappropriate tasks could lead to false conclusions and potentially harmful decisions. On the social aspect of fairness, [57] utilised questionnaires to gather insights from students and staff regarding the implications of bias in decision-making processes. Students expressed concerns about bias perpetuation and the fear of unfair assessments, while staff highlighted apprehensions regarding decisions made about them based on LA, such as managers using LA for performance evaluations. While the existing identification and mitigation methods for bias provide valuable insights for HEIs, there remains a significant gap. The lack of clear guidelines on which definitions of fairness should be adopted based on specific objectives, as well as which bias mitigation methods are most suitable depending on the context (data, algorithm, etc.), leaves many open questions when operationalising this principle.

4.2.5 Transparency. The principle of transparency is crucial in ensuring that all stakeholders involved in a LA system (students, staff, and other relevant parties) are well-informed about its operations. Numerous studies highlight the

[7] https://sites.google.com/view/fairlak

necessity of transparency in LA [8, 34, 59], yet many fall short of providing practical guidance on how to achieve it. Drachsler and Greller [16] note the inherent complexity in data collection and algorithmic processes, emphasising the challenge of conveying this information to non-technical stakeholders, including learners, teachers, and education managers. They advocate for giving data subjects access to their analytics results, empowering them to decide whether to seek pedagogical support or interventions, thereby placing the learner in control. The work also stresses the importance of obtaining clear consent prior to data collection, including the need for straightforward yes/no questions and the option to opt-out without repercussions. Hakami [24] reviewed 37 LAK papers mentioning LA and Transparency. They highlighted the need to ensure stakeholders know how the algorithm works (algorithmic transparency), the need to set policies that reveal what data is collected and how they are used (institutional transparency), and the need to inform learners that they are being tracked (transparency and data). They also highlight that transparency can enhance understanding, sense-making and reflection, technology acceptance and adoption and trust. Tsai et al.[56] further highlight the significance of effective communication in achieving transparency, while Veljanova et al. [62] propose technical design features aimed at operationalising transparency within LA systems. Collectively, these works demonstrate the multifaceted nature of transparency, illustrating how open communication regarding data practices, algorithms, and decision-making processes fosters a more informed environment for all users involved in LA. However, significant gaps remain regarding when and how these communications should be implemented, as well as the most effective strategies to ensure that both students and staff receive, understand, and assimilate this information. [56] suggest practical steps such as organising workshops or meetings with students and incorporating relevant training on digital literacy into academic development programmes. These initiatives aim to raise awareness about the importance of data protection and empower students to take informed actions regarding their data in the context of LA. However, despite these valuable recommendations, the effectiveness of these proposed actions has yet to be evaluated.

4.2.6 Privacy. A wide range of studies have considered the privacy implications of LA. In a concept mapping exercise with experts, privacy as well as transparency were identified as the most important elements of LA policy [48]. A key issue identified in prior studies is how and whether the student has agency over use of their data in the LA system. Recommendations include giving students the option to opt-out [55]. Alternatively, LA could be offered on an opt-in basis [42] in which LA is presented in terms of how it can improve their learning experience [22]. To ensure consent is genuinely informed [40, 55, 63] students need greater awareness of what data is used and how [42] and the expected impact of granting or withdrawing consent [3]. Workshops or meetings with students may be used to ensure students have appropriate knowledge of data literacy and data protection [58]. Consent-seeking procedures should be defined at an early stage [4] to help ensure initial or changing student preferences can be handled in the LA infrastructure [17]. Similarly, staff should be given better guidance on the appropriate use of data and also the consequences of misuse (e.g. loss of confidentiality, negative publicity, legal action) [20]. Such guidance could be informed by a Privacy Impact Assessment, covering impact on individuals, groups and wider society [9]. Privacy should be initially considered at the point of the initial business case as part of the risk analysis for the initiative [9]. Data should be anonymised wherever possible [55, 58], for example when aggregated to inform curriculum improvements [22]. More broadly, data governance guidelines should inform data sharing and ownership [3] and be used to continually assess data access rights for different stakeholders [4], minimising access to student data [58]. Privacy enhancing technologies such as anonymisation, encryption and digital signatures should be considered to improve the security of personal data [42] and any stakeholders should have access to an independent complaints body if they have any grievance over how their data has been accessed and used [9].

4.2.7 Explainability. The principle of explainability highlights the need for LA systems to offer clear insights into how their predictions and decisions are made. For example, [33] note that traditional machine learning methods, like decision trees, provide higher interpretability compared to modern deep learning models, which often function as "black boxes." As a result, despite their lower accuracy, these interpretable systems may be preferred in scenarios where understanding the decision-making process is critical. This observation is further supported by Gunasekara et al. [23], who reviewed explainability research within Educational Data Mining (EDM) and Learning Analytics, underscoring the importance of clarity in these systems. To address the explainability of more complex models, Li et al. [32] utilise two widely used explainable AI tools: Local Interpretable Model-agnostic Explanations (LIME) and Shapley Additive Explanations (SHAP). These tools, also adopted by well-known predictive analytics platforms [25], help break down complex model outputs into understandable components. Li and colleagues stress the importance of developing robust evaluation metrics to assess the quality of these explanations. Indeed, Gunasekara et al.'s review [23] reveals a critical gap in research regarding the metrics necessary to effectively evaluate the quality and utility of these explanations. Taken together, these works reflect the complex nature of explainability in LA. While existing studies underscore the importance of offering transparent and comprehensible explanations for decision-making processes, a significant gap remains in providing practical guidance on how to implement explainability effectively in educational settings. The challenge lies not only in ensuring that explanations are available but also in communicating them in ways that are meaningful and accessible to a diverse range of users, including both students and educators.

5 Responsible AI Framework for Learning Analytics

In this section, we introduce our Responsible AI (RAI) framework tailored to Learning Analytics (LA) in Higher Education (HE) - see Figure 1. The primary aim of this framework is to provide higher education institutions with actionable guidance on how to incorporate responsible AI principles effectively into their LA initiatives. Recognising that institutions are at various stages of their LA adoption, we have structured the framework to follow the stages of the software development lifecycle: Requirements and Data Collection, Design, Development, Testing, Release, and Monitoring. By aligning the framework with these stages, we address a key limitation of many existing resources, allowing HEIs to engage with the specific stage of development they are currently in. This approach enables a more flexible, actionable pathway for integrating responsible AI principles.

While our ultimate goal is to provide both a list of actions HEIs can take to ensure their LA systems incorporate responsible AI principles and how to implement these actions, our literature review reveals a significant lack of real-world examples of how HEIs have operationalised these principles—if they have done so at all. We acknowledge that this leaves our framework incomplete, particularly in offering specific, practical steps that have already been tested in the field. However, we see this as an opportunity for continued growth. Our ambition is to refine this resource in collaboration with the wider academic and practitioner community, learning from best practices as they emerge.

For the purposes of this paper, a version of our proposed framework is summarised in Figure 1 and accessible via an anonymised URL[8], where we have presented the relevant elements through a PowerPoint presentation (due to the constraints of the double-blind review process). The framework is hosted on a dedicated project website, where each step is linked to available documentation and real-world case studies from HEIs. This evolving resource will allow the community to contribute relevant materials, such as code libraries, consent forms, and other practical examples, fostering a collaborative environment where institutions can learn from one another.

[8]Project repository:https://osf.io/at97f/?view_only=15b10f42abaa466691ffcce8a61226c1

Ultimately, we hope that this framework becomes a dynamic tool for HEIs seeking to responsibly implement LA systems, enabling them to align their practices with responsible AI principles.

6 Discussion and Conclusions

In this paper, we have sought to address a pressing need within Higher Education Institutions (HEIs) for practical guidance on implementing Responsible AI principles within Learning Analytics solutions. Our aim is to establish a comprehensive Responsible AI framework tailored specifically for LA applications in HE. This framework is rooted in existing literature, but we have taken an important step further by examining not only high-level principles but also detailed accounts from studies that have operationalised these principles in various contexts. We have gathered lessons learned and challenges encountered, providing a richer foundation for our proposed framework.

Our framework is organised around the Software Development Lifecycle, encompassing critical stages from requirements gathering and data collection to deployment and monitoring. By doing this, we aim to create a structured approach that allows HEIs to systematically integrate Responsible AI principles into their LA practices. Each stage of the lifecycle includes a list of actionable items based on insights gleaned from the literature, offering guidance on operationalising these principles.

This resource, which is accessible online (URL withheld due to the double-blind review process), is intentionally a work in progress. We envision it as a starting point for collaboration and dialogue between the academic community and practitioners. Discussing our framework at relevant venues is crucial for refining and enhancing this resource, as it enables us to gather feedback and share best practices with stakeholders in the LA community.

We acknowledge that there may be relevant works from other disciplines—such as Computer Science and Sociology—that explore aspects of Responsible AI, albeit not specifically within the HE context. While these works have not been incorporated into this initial literature overview, our goal is to continually expand our understanding and enrich our framework with interdisciplinary insights that can aid HEIs in operationalising Responsible AI principles.

It is important to note that for certain principles, particularly accountability, there remain significant gaps in practical examples. We hope that this paper will stimulate new research directions that can help the community address these challenges. By highlighting these gaps, we aim to foster a collaborative environment where researchers can share insights and develop concrete strategies for accountability in LA systems.

In summary, our proposed Responsible AI framework serves as a vital resource for HEIs looking to implement ethical and responsible practices in their learning analytics efforts. We believe that by engaging with the broader academic and practitioner communities, we can collectively enhance the application of Responsible AI principles, ultimately benefiting both educators and students in the evolving landscape of higher education.

References

[1] Circle U. European University Alliance. 2022. Legal and Ethical Aspects of Using Learning Analytics from a Univeristy Perspective.

[2] Jose M Alvarez, Alejandra Bringas Colmenarejo, Alaa Elobaid, Simone Fabbrizzi, Miriam Fahimi, Antonio Ferrara, Siamak Ghodsi, Carlos Mougan, Ioanna Papageorgiou, Paula Reyero, et al. 2024. Policy advice and best practices on bias and fairness in AI. *Ethics and Information Technology* (2024).

[3] Asma Shannan Alzahrani, Yi-Shan Tsai, Naif Aljohani, Emma Whitelock-wainwright, and Dragan Gasevic. 2023. Do teaching staff trust stakeholders and tools in learning analytics? A mixed methods study. *Educational technology research and development* (Aug. 2023).

[4] Asma Shannan Alzahrani, Yi-Shan Tsai, Sehrish Iqbal, Pedro Manuel Moreno Marcos, Maren Scheffel, Hendrik Drachsler, Carlos Delgado Kloos, Naif Aljohani, and Dragan Gasevic. 2023. Untangling Connections between Challenges in the Adoption of Learning Analytics in Higher Education. *Education and Information Technologies* (April 2023).

[5] Amazon. [n. d.]. Responsible AI. https://aws.amazon.com/machine-learning/responsible-ai/

[6] Vita Santa Barletta, Danilo Caivano, Domenico Gigante, and Azzurra Ragone. 2023. A rapid review of responsible ai frameworks: How to guide the development of ethical ai. In *Proceedings of the 27th International Conference on Evaluation and Assessment in Software Engineering*.

[7] Vaclav Bayer, Martin Hlosta, and Miriam Fernandez. 2021. Learning analytics and fairness: do existing algorithms serve everyone equally?. In *International Conference on Artificial Intelligence in Education*. Springer.

[8] Teresa Cerratto Pargman and Cormac McGrath. 2021. Mapping the Ethics of Learning Analytics in Higher Education: A Systematic Literature Review of Empirical Research. *Journal of Learning Analytics* 2 (Sept. 2021). https://doi.org/10.18608/jla.2021.1

[9] Roger Clarke. 2018. Guidelines for the responsible application of data analytics. *Computer Law & Security Review* (June 2018).

[10] Andrew Nicholas Cormack. 2016. A Data Protection Framework for Learning Analytics. (April 2016). https://doi.org/10.18608/jla.2016.31.6

[11] Linda Corrin, Gregor Kennedy, Sarah French, Simon Buckingham Shum, Kirsty Kitto, Abelardo Pardo, Deborah West, Negin Mirriahi, and Cassandra Colvin. 2019. The Ethics of Learning Analytics in Australian Higher Education. A Discussion Paper. https://resolver.ebscohost.com/Redirect/PRL?EPPackageLocationID=2418090.1356208.27427588&epcustomerid=s2947694

[12] Helen Crompton and Diane Burke. 2023. Artificial intelligence in higher education: the state of the field. *International Journal of Educational Technology in Higher Education* (April 2023). https://doi.org/10.1186/s41239-023-00392-8

[13] Oscar Blessed Deho, Srecko Joksimovic, Jiuyong Li, Chen Zhan, Jixue Liu, and Lin Liu. 2022. Should learning analytics models include sensitive attributes? Explaining the why. *IEEE Transactions on Learning Technologies* 16, 4 (2022), 560–572.

[14] Oscar Blessed Deho, Lin Liu, Jiuyong Li, Jixue Liu, Chen Zhan, and Srecko Joksimovic. 2024. When the past!= the future: Assessing the Impact of Dataset Drift on the Fairness of Learning Analytics Models. *IEEE Transactions on Learning Technologies* (2024).

[15] Oscar Blessed Deho, Chen Zhan, Jiuyong Li, Jixue Liu, Lin Liu, and Thuc Duy Le. 2022. How do the existing fairness metrics and unfairness mitigation algorithms contribute to ethical learning analytics? *British Journal of Educational Technology* (2022).

[16] Hendrik Drachsler and Wolfgang Greller. 2016. Privacy and Analytics – it's a DELICATE Issue A Checklist for Trusted Learning Analytics. Edinburgh.

[17] Polydorou Eleni. 2023. *Towards a Secure and Privacy Compliant Framework for Educational Data Mining*. Springer Nature Switzerland.

[18] Rebecca Ferguson. 2019. Ethical Challenges for Learning Analytics. *Journal of Learning Analytics* (2019).

[19] The Institute for Ethical Al in Education. 2021. *The Ethical Framework for AI in Education*. https://www.buckingham.ac.uk/research/research-in-applied-computing/the-institute-for-ethical-ai-in-education/

[20] Mary Francis, Mejai Bola Mike Avoseh, Karen Card, Lisa Newland, and Kevin Streff. 2023. Student Privacy and Learning Analytics: Investigating the Application of Privacy within a Student Success Information System in Higher Education. *Journal of Learning Analytics* (2023).

[21] Google. 2023. AI Principles Progress Update 2023. https://ai.google/static/documents/ai-principles-2023-progress-update.pdf

[22] Nicole Gosch, David Andrews, Carla Barreiros, Philipp Leitner, Elisabeth Staudegger, Martin Ebner, and Stefanie Lindstaedt. 2021. Learning Analytics as a Service for Empowered Learners: From Data Subjects to Controllers. In *LAK21: 11th International Learning Analytics and Knowledge Conference*. https://doi.org/10.1145/3448139.3448186

[23] Sachini Gunasekara and Mirka Saarela. 2024. Explainability in Educational Data Mining and Learning Analytics: An Umbrella Review. In *International conference on educational data mining*. International Educational Data Mining Society.

[24] Eyad Hakami and Davinia Hernández Leo. 2020. How are learning analytics considering the societal values of fairness, accountability, transparency and human well-being?: A literature review. *Martínez-Monés A, Álvarez A, Caeiro-Rodríguez M, Dimitriadis Y, editors. LASI-SPAIN 2020: Learning Analytics Summer Institute Spain 2020: Learning Analytics. Time for Adoption?; 2020 Jun 15-16; Valladolid, Spain. Aachen: CEUR; 2020.* (2020).

[25] Martin Hlosta, Christothea Herodotou, Tina Papathoma, Anna Gillespie, and Per Bergamin. 2022. Predictive learning analytics in online education: A deeper understanding through explaining algorithmic errors. *Computers and Education: Artificial Intelligence* (2022).

[26] Joel A. Howell, Lynne D. Roberts, Kristen Seaman, and David C. Gibson. 2018. Are We on Our Way to Becoming a "Helicopter University"? Academics' Views on Learning Analytics. *Technology, Knowledge and Learning* (April 2018). https://doi.org/10.1007/s10758-017-9329-9

[27] IBM. 2024. AI ethics. https://www.ibm.com/impact/ai-ethics

[28] JISC. 2023. Code of practice for learning analytics. https://repository.jisc.ac.uk/9204/1/code-of-practice-for-learning-analytics.pdf

[29] Thashmee Karunaratne. 2021. For Learning Analytics to Be Sustainable under GDPR—Consequences and Way Forward. *Sustainability* (Oct. 2021).

[30] D Lesli. 2019. Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. https://doi.org/10.5281/zenodo.3240529

[31] Lin Li, Lele Sha, Yuheng Li, Mladen Raković, Jia Rong, Srecko Joksimovic, Neil Selwyn, Dragan Gašević, and Guanliang Chen. 2023. Moral machines or tyranny of the majority? A systematic review on predictive bias in education. In *LAK23: 13th International Learning Analytics and Knowledge conference*.

[32] Min-Jia Li, Shun-Ting Li, Albert CM Yang, Anna YQ Huang, and Stephen JH Yang. 2024. Trustworthy and Explainable AI for Learning Analytics.. In *LAK Workshops*. 3–12.

[33] Marco Lünich and Birte Keller. 2024. Explainable Artificial Intelligence for Academic Performance Prediction. An Experimental Study on the Impact of Accuracy and Simplicity of Decision Trees on Causability and Fairness Perceptions. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3630106.3658953

[34] Mehul Mahrishi, Asad Abbas, and Mohammad Khubeb Siddiqui. 2024. Global Initiatives Towards Regulatory Frameworks for Artificial Intelligence (AI) in Higher Education. (2024). https://doi.org/10.1145/3672462

[35] Anuradha Mathrani, Teo Susnjak, Gomathy Ramaswami, and Andre Barczak. 2021. Perspectives on the challenges of generalizability, transparency and ethics in predictive learning analytics. *Computers and Education Open* (2021).

[36] Microsoft. 2022. https://blogs.microsoft.com/on-the-issues/2022/06/21/microsofts-framework-for-building-ai-systems-responsibly/

[37] Andy Nguyen, Ha Ngan Ngo, Yvonne Hong, Belle Dang, and Bich-Phuong Thi Nguyen. 2023. Ethical principles for artificial intelligence in education. *Education and Information Technologies* (2023).

[38] Eirini Ntoutsi, Pavlos Fafalios, Ujwal Gadiraju, Vasileios Iosifidis, Wolfgang Nejdl, Maria-Esther Vidal, Salvatore Ruggieri, Franco Turini, Symeon Papadopoulos, Emmanouil Krasanakis, et al. 2020. Bias in data-driven artificial intelligence systems—An introductory survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* (2020).

[39] University of Edinburgh. 2017. Learning Analytics Principles and Purposes. https://www.ed.ac.uk/files/atoms/files/learninganalyticsprinciples.pdf

[40] Chris Patterson, Emily York, Danielle Maxham, Rudy Molina, and Paul Mabrey. 2023. Applying a Responsible Innovation Framework in Developing an Equitable Early Alert System: A Case Study. *Journal of Learning Analytics* (March 2023). https://doi.org/10.18608/jla.2023.7795

[41] Paul Prinsloo and Sharon Slade. 2013. An evaluation of policy frameworks for addressing ethical considerations in learning analytics. Belgium.

[42] Paul Prinsloo, Sharon Slade, and Mohammad Khalil. 2022. The answer is (not only) technological: Considering student data privacy in learning analytics. *British Journal of Educational Technology* (2022).

[43] Joel R. Reidenberg and Florian Schaub. 2018. Achieving Big Data Privacy in Education. *Theory and Research in Education* (Nov. 2018).

[44] Irina Rets, Christothea Herodotou, and Anna Gillespie. 2023. Six Practical Recommendations Enabling Ethical Use of Predictive Learning Analytics in Distance Education. *Journal of Learning Analytics* (2023).

[45] Lynne D Roberts, Vanessa Chang, and David Gibson. 2017. Ethical considerations in adopting a university-and system-wide approach to data and learning analytics. *Big data and learning analytics in higher education: Current theory and practice* (2017).

[46] Lynne D Roberts, Joel A Howell, Kristen Seaman, and David C Gibson. 2016. Student attitudes toward learning analytics in higher education:"The fitbit version of the learning world". *Frontiers in psychology* (2016).

[47] Cristobal Romero and Sebastian Ventura. 2020. Educational data mining and learning analytics: An updated survey. *Wiley interdisciplinary reviews: Data mining and knowledge discovery* (2020).

[48] Maren Scheffel, Yi-Shan Tsai, Dragan Gašević, and Hendrik Drachsler. 2019. Policy Matters: Expert Recommendations for Learning Analytics Policy. In *Transforming Learning with Meaningful Technologies: 14th European Conference on Technology Enhanced Learning, EC-TEL 2019*.

[49] Nabila Sghir, Amina Adadi, and Mohammed Lahmer. 2023. Recent advances in Predictive Learning Analytics: A decade systematic review (2012–2022). *Education and information technologies* (2023).

[50] Allyson Skene, Laura Winer, and Erika Kustra. 2022. Clouds in the silver lining of learning analytics: ethical tensions for Educational Developers. *International Journal for Academic Development* (Aug. 2022). https://doi.org/10.1080/1360144X.2022.2099208

[51] Sharon Slade and Paul Prinsloo. 2013. Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist* (2013).

[52] Sharon Slade and Alan Tait. 2019. *Global guidelines: Ethics in Learning Analytics*.

[53] Christina M. Steiner, Michael D. Kickmeier-Rust, and Dietrich Albert. 2016. LEA in Private: A Privacy and Data Protection Framework for a Learning Analytics Toolbox. *Journal of Learning Analytics* (April 2016). https://doi.org/10.18608/jla.2016.31.5

[54] Elham Tabassi. 2023. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. Gaithersburg, MD. https://doi.org/10.6028/NIST.AI.100-1

[55] Yi-Shan Tsai and Dragan Gasevic. 2017. Learning analytics in higher education — challenges and policies: a review of eight learning analytics policies. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*. ACM, Vancouver British Columbia Canada.

[56] Yi-Shan Tsai, Alexander Whitelock-Wainwright, and Dragan Gašević. 2020. The privacy paradox and its implications for learning analytics. In *Proceedings of the tenth international conference on learning analytics & knowledge*.

[57] Yi-Shan Tsai, Alexander Whitelock-Wainwright, and Dragan Gašević. 2021. More than figures on your laptop:(Dis) trustful implementation of learning analytics. *Journal of Learning Analytics* (2021).

[58] Yi-Shan Tsai, Alexander Whitelock-Wainwright, and Dragan Gašević. 2020. The privacy paradox and its implications for learning analytics. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge (LAK 20)*. https://doi.org/10.1145/3375462.3375536

[59] Dimitrios Tzimas and Stavros Demetriadis. 2021. Ethical issues in learning analytics: a review of the field. *Educational Technology Research and Development* (April 2021). https://doi.org/10.1007/s11423-021-09977-4

[60] The Open University. 2023. Data Ethics Policy 2023. https://help.open.ac.uk/documents/policies/ethical-use-of-student-data/files/287/Data%20ethics%20policy%20NEW%20July%202023_.pdf

[61] Hristina Veljanova, Carla Barreiros, Nicole Gosch, Elisabeth Staudegger, Martin Ebner, and Stefanie Lindstaedt. 2022. Towards Trustworthy Learning Analytics Applications: An Interdisciplinary Approach Using the Example of Learning Diaries *(Communications in Computer and Information Science)*. https://doi.org/10.1007/978-3-031-06391-6_19

[62] Hristina Veljanova, Carla Barreiros, Nicole Gosch, Elisabeth Staudegger, Martin Ebner, and Stefanie Lindstaedt. 2023. Operationalising Transparency as an Integral Value of Learning Analytics Systems–From Ethical and Data Protection to Technical Design Requirements. In *International Conference on Human-Computer Interaction*. Springer.

[63] Deborah West, Henk Huijser, and David Heath. 2016. Putting an Ethical Lens on Learning Analytics. *Educational Technology Research and Development* (Oct. 2016).

Received 23 September 2024; revised 28 October 2024; accepted 22 November 2024

Principles	Requirements & data collection	Design	Development	Testing	Release & Monitoring
Accountability	- Define roles & responsibilities - Define consequences of misuse - Define auditing approach	- Define controls and guidelines within the LA system - Establish an escalation procedure for when breakdowns occur	- Translate defined controls and guidelines into software features - Ensure built-in audit logs that track actions, decisions, and system changes	- Test rights and responsibilities - Make sure that implemented controls and guidelines are functioning correctly and providing expected outcomes	- Perform periodic assessments - Monitor changes in roles and responsibilities - Monitor changes on consequences of misuse - Ensure actions are taken as result or intended or unintended harms
Explainability	- Define what elements will need explanations (data, algorithm, outputs, ...) - Define the type of explanations needed considering different elements and users of the LA system (e.g., novice vs expert users).	- Define how explanations will be generated (e.g., by incorporating explainable AI tools, like LIME, by using explainable ML models like decision trees) - Define how explanations will be presented to the user (e.g, via narratives, tables) - Define how explanations will be deliver to the user (e.g, via interactive elements of the system, documentation) - Define how the user will interact with such explanations (confirm they have read them, have an opportunity to challenge them) -Incorporate user-centered design principles to ensure explanations are accessible for various user groups	- Ensure that explanation methods (LIME, SHAP) are chosen and incorporated based on the model complexity and user needs - Translate the output of the model explanation methods into different system, documentation and interface features to ensure users can see them and interact with them accordingly	- Test the correctness of the explanations - Test the accessibility of the explanations (easy for users to find, and access) for different user groups - Test the comprehensibility of explanations (easy for users to understand) for different user groups	- Perform periodic assessments for correctness, accessibility and comprehensibility of explanations - Consider changes based on user feedback and assessment results
Fairness & bias	- Select a definition of fairness - Identify relevant sensitive attributes (also consider intersectionality) - Identify potential sources of bias (data bias, algorithmic bias, social bias) - Identify potentially applicable bias identification and mitigation methods	- Apply relevant pre-processing bias mitigation methods (e.g, manage data imbalances, errors or missing values) - Analyse and select algorithms/parameters that could minimise biases for the given LA system - Design system features that could minimise human biases (i.e., how humans interpret and act over the presented information) , in addition to algorithm-driven biases.	- Apply appropriate in-processing and post-processing bias mitigation methods - Ensure continual feedback mechanisms that assess biases arising from system use	- Select and apply appropriate fairness metrics to ensure not just the accuracy of the LA system, but also its fairness - When possible, test with different data to ensure robustness to data drifts. - Test the algorithms behind the LA system, but also the real-life impact of the LA system as a whole (e.g, impact on student outcomes)	- Continuously monitor the AL system for accuracy and fairness and the trade-off between them - Regularly update fairness algorithms and models based on changing social contexts and legal standards
Safety	- Define responsible design, development, test and deployment practices - Perform tool safety risk assessment - Define procedures to collect and maintain accurate data - Secure staff training on responsible development practices	- Define test scenarios to evaluate model and data accuracy - Identify methods and metrics to evaluate model accuracy (e.g., error rate, accuracy) - Define fail-safes in the system that will not display or revert predictions if accuracy thresholds are not met	- Ensure all software modules adhere to responsible safety protocols in coding and deployment. - Implement monitoring and audit approaches to detect safety vulnerabilities (e.g.,data poisoning)	- Test predictions/outputs using testing scenarios - Test for possible deviations from intended or expected functionality (e.g., test models data drifts)	- Communicate responsible use of system, tools, predictions to end users - Execute periodic safety assessment (e.g., monitor and detect data drifts, concept drifts) - Evaluate and monitor impact of LA and predictions in students and staff
Security	- Identify local and regional data security standards - Gather security requirements and identify assets - Perform a security risk assessment - Create an information security policy	- Use the security risk assessment to identify proper level of security measures against threats - Define role-based access controls to protect sensitive data	- Apply appropriate tools and measures to secure code, models and data - Ensure secure configuration of tools	- Build routine security test for all elements of the system - Test and review security controls (mitigate or eliminate risks)	- Monitor periodically compliance of security policies and improve them if necessary - Apply security and monitoring response program for security threats and vulnerabilities -Continuously update security protocols as new threats or vulnerabilities emerge
Transparency	- Define who needs to be informed (staff, students, expert/novice technical audiences) - Define what information needs to be delivered (which information is tracked by the LA system, when and how is data processed, which decisions are supported with such data) - Define mechanisms for effectively delivering information depending on the different roles and user types (short videos, weakly emails)	- Design software features to be accompanied by relevant information (e.g., i icons) - Design documentation of the LA system according to each relevant role (e.g., documentation for students vs. tutors, vs. software developers) - Determine how and when to deliver information to the different roles so that it is clear and understandable, and it will effectively reach users. - Design mechanisms so that users can request additional information and ask relevant questions - Define consent mechanisms that are understandable and provide clear, accessible information for different stakeholders - Co-design with users to understand which transparency mechanisms are most effective.	- Implement communication campaigns to ensure the designed documentation is effectively distributed across relevant roles - Incorporate explanations when appropriate within the LA system - Embed transparency mechanisms (e.g., "info" buttons) directly within the system's interface. - Implement mechanisms to enable users to ask further questions	- Assess whether users (students, tutors and other roles) are aware of all the different elements of the LA system (data, algorithm, etc.) and how these different elements affect them and other roles within and outside the HE Institution. Questionnaires, interviews, focus groups, and immediate feedback via the LA system are possible mechanisms to test how informed users are. - Test transparency features during real-time access.	- Conduct regular transparency audits to ensure all stakeholders remain informed as systems evolve. - Consider user feedback - Modify communication campaigns as needed to ensure more effective information mechanisms - Adapt the LA system to ensure different users can effectively process the provided information
Privacy	- Perform privacy impact and risk assessment - Determine nature of consent (opt-in/out) - Define consent-seeking procedure - Obtain consent for data collection - Define what, how, and why data is being collected	- Design methods for sharing knowledge on data literacy and protection (documents, workshops) - Specify data access roles - Specify anonymisation techniques	- Implement data privacy techniques, considering use of privacy enhancing technologies - Comply with transparency and communication regulations	- Ensure data access is appropriate for each stakeholder - Ensure initially specified and later modified data preferences are appropriately propagated	- Establish (independent) process for handling complaints regarding data access and use - Continually monitor data access as roles and personnel change - Deploy data literacy and protection training

CHECK CURRENT REGULATIONS

Fig. 1. Responsible AI Framework for Learning Analytics in HEI