Audits for Trust: An Auditability Framework for AI-Based Learning Analytics Systems

Linda Fernsel[a], Yannick Kalff[b] and Katharina Simbeck[c]
Computer Science and Society, HTW Berlin University of Applied Sciences, Treskowallee 8, 10318 Berlin, Germany
{*linda.fernsel, yannick.kalff, simbeck*}*@htw-berlin.de*

Keywords: Audit, Auditability, Artificial Intelligence, Learning Analytics.

Abstract: Audits contribute to the trustworthiness of Learning Analytics (LA) systems that integrate Artificial Intelligence (AI) and may be legally required in the future. We argue that the efficacy of an audit depends on the auditability of the audited system. Therefore, systems need to be designed with auditability in mind. We present a framework for assessing the auditability of AI-integrating systems in education that consists of three parts: (1) verifiable claims about the validity, utility and ethics of the system, (2) evidence on subjects (data, models, or the system) in different types (documentation, raw sources and logs) to back or refute claims, (3) means to validate evidence such as technical APIs, monitoring tools, or explainable AI principles must be accessible to auditors. We apply the framework to assess the auditability of the Learning Management System Moodle, which supports an AI-integrating dropout prediction system. Moodle's auditability is limited by incomplete documentation, insufficient monitoring capabilities, and a lack of available test data.

1 INTRODUCTION

Artificial Intelligence (AI) significantly impacts the field of Learning Analytics (LA). LA itself is gaining relevance in higher education (Baek and Doleck, 2023), K-12 classes (Paolucci et al., 2024), pre-school settings (Crescenzi-Lanna, 2020), and generally for virtual education (Elmoazen et al., 2023; Heikkinen et al., 2023). AI adds to the utility of LA elements of educational data mining (Baek and Doleck, 2023; Romero and Ventura, 2020), deep learning capabilities, machine learning (Ouyang et al., 2023), predictive and prescriptive analytics (Sghir et al., 2022; Xiong et al., 2024), or multi-modal models capable of processing complex physical behavioral data and stimuli (Crescenzi-Lanna, 2020). AI in LA offers data-driven insights into learning processes and students' behavior to predict learning success, risks of failure or drop-out, and to prescribe proactive measures (Susnjak, 2024). Above that, AI technologies promise opportunities to improve learning situations and outcomes for all students, especially those disadvantaged and struggling (Khalil et al., 2023).

However, AI technologies in LA entail ethical issues (Rzepka et al., 2022; Rzepka et al., 2023) and open questions about their utility or maturity in education (Drugova et al., 2024). Especially the potential threat to equality and equity principles in education raised research and practitioner interest in mitigating discriminatory elements of AI models in LA (Simbeck, 2024; Rzepka et al., 2023). Above that, ethical concerns create the urgency for adequate legislation to counter negative effects or prevent biased systems before they harm. AI and AI-driven LA face regulation, such as the European AI Act (European Union, 2024), or frameworks to impose ethical requirements on AI products (Toreini et al., 2022; Fjeld et al., 2020; Slade and Tait, 2019), and to mitigate potential negative effects and discriminatory biases (Baker and Hawn, 2022; Prinsloo and Slade, 2017). For this case, the "AI Act" (European Union, 2024) aims to regulate "high-risk" AI systems that could violate the "health and safety or the fundamental rights of persons" (European Union, 2024, Recital 52). AI-based LA systems can be considered "high-risk" because of their impact on personal educational success, which directly affects the individual "ability to secure [one's] livelihood" (European Union, 2024, Recital 56).

Regulatory frameworks like the AI Act require audits of AI systems that certify their legal and ethical compliance (Berghoff et al., 2022; Toreini et al., 2022). The AI Act mandates two types of audits for high-risk AI systems: conformity assessments before deployment (European Union, 2024, Art. 43) and post-market

[a] https://orcid.org/0000-0002-0239-8951
[b] https://orcid.org/0000-0003-1595-175X
[c] https://orcid.org/0000-0001-6792-461X

Fernsel, L., Kalff, Y. and Simbeck, K.
Audits for Trust: An Auditability Framework for AI-Based Learning Analytics Systems.

In *Proceedings of the 17th International Conference on Computer Supported Education (CSEDU 2025) - Volume 2*, pages 51-62
ISBN: 978-989-758-746-7; ISSN: 2184-5026

monitoring after system deployment (European Union, 2024, Art. 72). Audits provide accountability and transparency, which is also necessary for establishing trust in AI systems (Toreini et al., 2022; Williams et al., 2022; Bose et al., 2019; Springer and Whittaker, 2019). On a practical level, audits allow stakeholders, such as system providers or deploying institutions, regulators, and subjects of the systems' decisions, to understand how the system decides and to identify and correct biases or errors (Springer and Whittaker, 2019; Nushi et al., 2018; Rzepka et al., 2023). However, audits struggle with systems that are inaccessible, opaque, or proprietary (Fernsel et al., 2024b).

We argue that AI-based LA systems must be auditable for any audit to achieve its results. We define AI systems as software systems that implement methods of machine learning. Machine learning is "the computational process of optimizing the parameters of a model from data, which is a mathematical construct generating an output based on input data" (European Parliament, 2023, Recital 6a). AI-based LA systems implement machine learning methods to leverage learning data for analysis, predictions, and prescriptions in educational contexts (Baek and Doleck, 2023; Romero and Ventura, 2020; Ouyang et al., 2023).

2 AUDITABILITY OF AI

2.1 Audits of AI Systems

Technical, legal, and ethical reasons make audits of AI systems necessary to ensure accountability for accurate, compliant, and fair technological systems (Raji et al., 2020; Falco et al., 2021; Ayling and Chapman, 2022). For this purpose, auditing techniques, for example, from the field of finance, are adapted for AI systems (Mökander et al., 2022). However, there are no standards for audit quality (Alagić et al., 2021), and residual risks remain uncertain (Knechel et al., 2013).

An audit analyses if an AI system complies with legal regulations, organizational standards, or ethical values. It compares claims made by stakeholders, like developers or deployers of AI-based LA tools, to the system's actual behavior. Claims concern an AI LA system's validity, utility, and ethics: i.e., its models, components, data sets, or scopes are suitable for the intended purpose (validity); the system does fulfill its intended use (utility); ethical, moral, or legal standards are taken into consideration (ethics) (Minkkinen et al., 2024; Ayling and Chapman, 2022). Auditors then recover "auditable artifacts" (Ayling and Chapman, 2022) to validate whether an AI system is implemented and operating as claimed. AI's functionality, application field, and associated risks require interdisciplinary competencies and skills in mandating and conducting audits (Landers and Behrend, 2023). Above that, AI systems' design and implementation principles affect audits and make processes of assessing claims, actual behavior, and auditable evidence feasible (Li and Goel, 2024; Ayling and Chapman, 2022; Falco et al., 2021).

We consider auditability as given *when a system is reviewable independently* (Williams et al., 2022; Wolnizer, 2006). (Weigand et al., 2013) conceptualize auditability as a) the system provides information on how relevant values should be used or produced (*claims*), b) the system generates information on how relevant values are used or produced (*evidence*), and c) stakeholders can *validate* these claims based on the provided evidence. The complexity of AI systems creates special requirements for audits and, thus, for a system's auditability. (Li and Goel, 2024) propose a framework for auditability that focuses on training data, underlying models, and organizational governance processes: "AI auditability demands more comprehensive information about the nature, process, quality assurance, and governance of training data, detailed process and governance information about AI model commissioning, development, deployment, and long-term monitoring, and the governance structure relevant to developing and managing the AI system" (Li and Goel, 2024). (Berghoff et al., 2022) approach auditability of AI systems from a cyber security perspective where increased system complexity impedes system auditability. (Raji et al., 2020) propose a joint internal audit process of auditors and auditees, who provide claims and artifacts as evidence. The joint product of an audit process should be a remediation plan to mitigate risks (Raji et al., 2020).

Claims are normative statements on a system's functionality, scope, and purpose. System providers and system deployers define claims in system standards, targeted fields of application, scope and use cases, or as part of the source code documentation (Stoel et al., 2012). Another set of claims stems from ethical or moral standpoints, laws, regulations, and standards that guide software implementation and use (Brundage et al., 2020).

In the context of AI-based LA, *evidence* is "relevant information about its execution" (Alhajaili and Jhumka, 2019) that allows us to analyze and trace errors in decisions. Auditees should enable evidence collection by organizational structures and processes that document a system's operation (Awwad et al., 2020; Stoel et al., 2012). Additionally, organizational processes, system logs, or data provide insights into AI-based LA's functioning and institutional setting.

Auditors with system access and evidence can verify whether an AI-integrating system meets the derived

claims. However, AI-based LA systems present challenges when designing test cases and selecting test data. Unlike other software, AI can handle a broader range of input data and assume more possible states (Berghoff et al., 2022). Therefore, a diverse and substantial amount of appropriate test data is required, but scarce (Tao et al., 2019; Fernsel et al., 2024b). Especially balanced test data representing marginalized student groups is scarce (Fernsel et al., 2024b). Further, in pre-deployment audits, predicting all *de facto* use cases and, therefore, all possible claims (Tao et al., 2019) is difficult. After deployment, models can be updated by learning from new training data or through feedback loops, and introduce bias in the process (Berghoff et al., 2022; Awwad et al., 2020). Therefore, the tests implemented for an AI are not necessarily realistic and make repeated audits after deployment necessary (Eitel-Porter, 2021; Mökander and Floridi, 2021).

2.2 Enabling Auditability of AI-integrating Systems

Because of the limited auditability of AI-integrating systems, some audits relied on self-audits and required auditees to answer a set of questions about the design principles of the system and the measures undertaken to guarantee functionality and compliance (Raji et al., 2020). While this is a valid approach, we argue that future AI-integrating systems, including AI-based LA systems, must be designed with auditability in mind to enable independent assurance. Even though these systems are inherently complex to audit, system providers and deploying institutions can take steps to enable independent auditability. Auditable AI-integrating systems require planning, documentation, the implementation of specific functionalities, such as logs, API, monitoring tools or explanations, and sometimes access to the system sources, such as program code, model configuration, and data.

Planning for Auditability. As AI-integrating LA systems are very complex, sufficient auditability will only be reached if it is planned for during the system design process. To help the completeness of evidence, "accountability plans" outline what and how information should be captured (Naja et al., 2022). Plans should also determine applicable definitions of ethical standards and the available data for their evaluation (Galdon-Clavell et al., 2020). (Slade and Tait, 2019) and (Kitto and Knight, 2019) discuss relevant ethical standards in LA. Based on the accountability plan, the system's organizational processes (project and risk management, design and development processes) must be adjusted for auditability. Workflow mechanisms to increase auditability include logging of model training and validation results, storing of model metadata, and continuous monitoring (Kreuzberger et al., 2022).

Documentation. The auditability of an AI-integrating system is further influenced by the completeness of documentation. The AI Act requires documentation for high-risk AI systems on the system in general, the models, and the relevant data (European Union, 2024, Art. 11). System-related documentation should include functionality and limitations of the system (European Union, 2024, Recital 66). The system auditability can be increased by documenting design and implementation choices, including the policies, external requirements, and organizational processes like project and risk management (Raji et al., 2020; Beckstrom, 2021). Model-related documentation should include information on algorithms for training, testing, and validation. Model parameters should also be documented, and documentation should also elaborate on the model performance (Beckstrom, 2021; Mitchell et al., 2019) (European Union, 2024, Art. 13, 3b), which includes providing complex evaluation metrics like ROC curves or a model-specific "measure of confidence" for each output (Ashmore et al., 2022). Data-related documentation should contain the data structure (Beckstrom, 2021; European High-Level Expert Group on AI, 2019) and information on data provenance, including the data acquisition method, data transformations, and data processing (Beckstrom, 2021; Gebru et al., 2021), e.g., for labeling and feature calculation. Documenting provenance contributes to reproducibility and helps to discover where biases originate and which data operations (e.g., data processing steps) influence them (Toreini et al., 2022). Aspects of data quality such as the balance of classes in the training, validation, and verification data sets (Beckstrom, 2021), and data completeness (Ashmore et al., 2022) should be documented as well. For data sets, suggested standards for documentation are Data Sheets (Gebru et al., 2021) and Dataset Nutrition Labels (Holland et al., 2018).

Providing Sources. AI-based LA results often lack reproducibility (Haim et al., 2023). Auditability can be increased further by providing the raw sources of an AI-integrating system, including the system source code (Tagharobi and Simbeck, 2022; Beckstrom, 2021), the model itself, including its model weights and the training and test data for evaluation purposes (Beckstrom, 2021). Under some circumstances, raw data cannot be provided due to privacy issues. For such cases, the auditor could be enabled to collect or create (synthesized) test data for the audit (El Emam et al., 2020). Synthetic data can also be helpful when data is scarce, for exam-

ple, to protect student privacy (Dorodchi et al., 2019) or for underrepresented minorities (El Emam et al., 2020).

Implementing Auditability. Enabling auditability requires specific system functionalities for externalizing system information, such as logging (Eitel-Porter, 2021; Bose et al., 2019), secure access to the system for auditors (Awwad et al., 2020), monitoring tools (Ashmore et al., 2022; Bharadhwaj et al., 2021; Eitel-Porter, 2021; Alhajaili and Jhumka, 2019), and explanations for model behavior (Brundage et al., 2020; Shneiderman, 2020; Guidotti et al., 2018).

Auditors can use logs to understand the data flow through the system (Falco et al., 2021). Logs record the production process of any result like predictions or data sets (Kale et al., 2022) and, thus, enhance the auditability of AI systems (Brundage et al., 2020; Shneiderman, 2020). Secure system access for external auditors, e.g., via APIs, is a prerequisite for an audit (Awwad et al., 2020). (Springer and Whittaker, 2019) show that APIs allow systematic tests of scenarios based on the system's claims. An API can also enable secure third-party access to logs (Alla and Adari, 2021). Monitoring tools help to analyze performance, detect model behavior changes, and recognize violations of (ethical) constraints (Eitel-Porter, 2021). It involves tracking various aspects of the system, such as model input, the environment of use, internal model properties, and model output (Ashmore et al., 2022). Constant monitoring after deployment is a regular part of the AI life cycle (Alla and Adari, 2021) and post-market monitoring is a requirement for high-risk AI systems under the AI Act (European Union, 2024, Art. 72). Explanations, like feature importance explanations or counterfactual explanations (Bhatt et al., 2020), which are provided as part of the user interface, can help users or auditors understand the system's output (Shneiderman, 2020).

3 A FRAMEWORK FOR ASSESSING AUDITABILITY OF AI SYSTEMS

Based on our discussion of audits and auditability of AI systems and methods to enhance the auditability of AI-integrating systems, we propose a framework to assess and identify opportunities to improve the auditability of AI-integrating systems. Figure 1 visualizes the framework for assessing the auditability. Any audit process has three steps displayed from the bottom to the top: First, auditors designate verifiable claims about the system. Then, auditors identify, generate, and collect suitable evidence. Finally, the auditors validate claims based on the evidence they retrieve from the AI-based LA system.

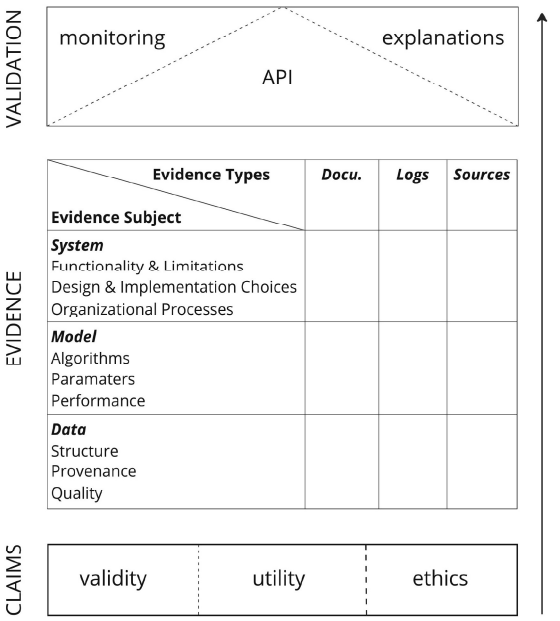

Figure 1: A framework for assessing auditability of AI systems.

Verifyable Claims. Developers or deploying organizations ensure the properties of the AI-based LA system and the processes in which it is applied. Auditors can derive verifiable claims from such assurance statements, which form the benchmark for the actual functioning of AI-based LA. Therefore, claims are the foundation for any audit (Brundage et al., 2020). Claims can concern *validity*: are methods correctly applied in the system, and is the system output correct? *Utility* describes if the system's functionality can be considered helpful in its use case. Finally, *adherence to underlying ethical principles* summarizes claims that the audited system complies with applicable law (GDPR) or latent social, organizational, or societal norms, e.g., corporate culture, accessibility, or diversity, equity, and inclusion (DEI) goals that attribute to fairness or objectivity (Landers and Behrend, 2023).

Evidence. Once auditors define the claims, they must identify, create, and collect evidence (Raji et al., 2020). Evidence comprises different subjects: system, model and data. It can take various forms as documentation, raw sources like source code, model weights, and raw data, and finally, logs (Raji et al., 2020; Brundage et al., 2020; Tagharobi and Simbeck, 2022; Beckstrom, 2021); (European Union, 2024, Art. 11, Art. 12).

Different evidence subjects assess specific aspects of AI-based LA systems. For the *system*, evidence

should prove the system's functionality and its limitations (European Union, 2024, Recital 66). Evidence must legitimize the underlying design and implementation choices and organizational processes (Beckstrom, 2021; Raji et al., 2020; European Union, 2024). Evidence for the *implemented models* contains the algorithms in use, model parameters, and model performance indicators (Beckstrom, 2021; Mitchell et al., 2019; European Union, 2024). Evidence on *data* informs about structure, provenance, and quality of test, training, or production data (Beckstrom, 2021; Ashmore et al., 2022; Gebru et al., 2021; Toreini et al., 2022; European Union, 2024).

Means of Validation. In the validation step, auditors access and assess any evidence about the AI-based LA systems to validate the claims about the system's validity, utility, and ethics. Means to validate claims can be integrated into the AI-based LA system—either as an interface to access raw data via APIs for further testing, in the form of monitoring tools to observe system output and parameters and deliver readily interpretable results or as explainable AI principles on dashboards (Fernsel et al., 2024b; Eitel-Porter, 2021; Bharadhwaj et al., 2021). The availability and ease of access to the evidence to validate claims determine the auditability of AI-based systems. It can further be an indicator of a system's transparency.

Utilizing the Framework. The framework aims to assess the auditability of an AI-based LA system and facilitate the design of auditable systems. We discussed several aspects of auditability, and their respective relevance varies with every audit situation. The identified claims dictate the required evidence subjects (system, model, data) and evidence types (sources, documentation, logs). The evidence types, in turn, specify the technical means of validation. When utilizing our framework to assess a system's auditability, auditors must consider which evidence is necessary to prove a stated claim and judge whether the evidence is sufficiently available and accessible.

Gathering claims requires a heuristic search, document analysis, and Q&A interviews with responsible positions to determine relevant claims and their hierarchy. Gathering evidence depends on subject and type and is closely related to the technical means of validation. Documentation is the most essential evidence, as it is the easiest to access and understand (Beckstrom, 2021). Arguably, the most challenging evidence could be source codes or logs of proprietary or security-sensitive systems (Alikhademi et al., 2022). However, evidence in the form of sources and logs must not be neglected: it may be necessary to complete information from the documentation or establish the credibility of the documentation (Beckstrom, 2021).

For the auditing practice, the framework can be used to define ex-ante responsibilities in the audited organization for providing claims and evidence. Further, it can be operationalized as a checklist to control the auditability of a system as an initial audit step or in the development cycle of a system. Since system development is an ongoing process, the framework assists in assessing the auditability on the developers' side and offers guardrails for quality assurance measures.

4 CASE STUDY

In this section, we apply the proposed auditability framework to assess the auditability of the dropout prediction system in Moodle 4.3. Table 1 lists the derived claims, and table 2 summarizes the results of desirable and available claims.

Moodle's dropout prediction system aims to prevent students from dropping out of a course (Monllaó Olivé et al., 2018). The software ships with an un-trained machine learning model (a model configuration) that, once trained on a particular Moodle platform, predicts whether a student is likely to drop out of a course (Moodle, 2023a; Monllaó Olivé et al., 2018). A model configuration can be tested in an "evaluation mode" before going live (Moodle, 2023a).

We chose this use case because Moodle is a commonly used open-source learning management system with potentially "high-risk" AI-based LA components under the AI Act. Additionally, dropout prediction models have repeatedly been shown to work better for majority groups better represented in training data (Gardner et al., 2019; Rzepka et al., 2022). Therefore, there is a risk that some groups of students benefit less from the AI-based LA module than others.

4.1 Claims

We consult the documentation of Moodle's student dropout prediction system and additional literature to identify claims. The first claim *v1* is that "[t]he accuracy and recall of the presented prediction model for predicting at-risk students are good for a production system" (Monllaó Olivé et al., 2018). Since the dropout prediction model design is based on the "Community of Inquiry" framework (Garrison et al., 1999), a second claim *v2* is that cognitive depth (metric applied in Moodle for "cognitive presence" of a student) and social breadth (metric applied in Moodle for "social presence" of a student) are valid indicators for dropout prediction (Moodle, 2023a).

Table 1: Overview of claims made for Moodle dropout prediction system by type. **Bold blue**: sufficient evidence available across types and subjects to validate claim; else: claim cannot be verified, even with some available evidence.

Type	Claim
Validity	(v1) sufficiently good predictions (v2) cognitive depth and social breadth are valid indicators
Utility	(u1) reduced dropout in online courses
Ethics	**(e1) AI-created predictions are marked as such** **(e2) stakeholders can decide whether to use the system** **(e3) GDPR conformity** (e4) equal efficiency for learners of different locations and financial backgrounds

Table 2: Overview of desirable and available evidence for each claim by evidence type and subject. ✓: required evidence of type and subject is fully available. □: incomplete evidence of type and subject. **Bold blue**: sufficient evidence available across types and subjects to validate claim; else: claim cannot be verified, even with some available evidence.

Evidence type \ Evidence subject	Documentation	Logs	Sources
System			
Functionality & Limitations	**e1**✓, **e2**✓, **e3**□, e4□		**e3**✓, e4✓
Design & Implementation choices	v1□, v2□, u1□, **e3**□, e4□		
Organizational processes	v1✓, u1□, **e3**□, e4□		
Model			
Algorithms	v1✓, **e3**□, e4□		**e3**✓, e4✓
Parameters	v1□, e4□		v1✓, e4✓
Performance	e4□	v1□, u1□, e4□	
Data			v1□, v2□, u1□, e4□
Structure	v1✓		
Provenance	v1✓, v2□, e4□	v2□, e4□	
Quality	v1□, e4□		

Furthermore, broad references can be found to the utility of the dropout prediction system (is the system functionality useful in its specific context?). Moodle's LA system should "not only predict events, but change them to be more positive" (Moodle, 2023a). The Moodle documentation asserts that the dropout prediction system is most useful for courses that run entirely online due to features that rely on Moodle activities (Moodle, 2023a). As utility claim *u1*, we can formulate that the dropout prediction system reduces dropout rates in online courses.

MoodleHQ, the organization leading the development of Moodle, explains which ethical principles drive the implementation and use of AI in Moodle: Users should always know when AI is used, stakeholders should be able to decide which AI components to use, AI components should preserve users' data privacy and security, and AI components should be efficient for all learners, "regardless of their location or financial situation" (Moodle, 2023b). Four ethics-related claims can be derived from these principles to audit Moodle's dropout prediction system. Firstly, dropout predictions are marked as being calculated by an AI (*e1*). Secondly, stakeholders can decide whether to use the dropout prediction system (*e2*). This implies that institutions can activate the feature, and learners can opt in or out. Third, student data collection, processing, and storage by the dropout prediction system follows the EU General Data Protection Regulation GDPR (*e3*). And lastly, the dropout prediction model is equally efficient for all learners, regardless of their geographical location and financial background (*e4*).

4.2 Required Evidence

We established that evidence in the form of documentation (Raji et al., 2020; Beckstrom, 2021), raw sources (Tagharobi and Simbeck, 2022; Beckstrom, 2021) and logs (Brundage et al., 2020) is suitable to verify claims. Evidence can concern aspects of Moodle's dropout prediction system (the system), the dropout prediction model, and the underlying data. As noted before, not all evidence must be available to validate all claims. This section examines which evidence is required to validate which claim. The claims and evidence subjects are indicated in cursive.

Validity. To prove *v1* (sufficiently good predictions), the most reliable way would be to reproduce the quality

assessment conducted by (Monllaó Olivé et al., 2018). To evaluate the dropout prediction configuration, auditors require access to a Moodle system with test data *(data sources)* to calculate the *model performance*. We call this type of system a "test system". In the absence of openly available test data, additional documentation on *data quality* requirements is helpful for the acquisition of suitable test data. In this use case, data can only be acquired by exporting data from a Moodle platform, not through synthesis. This is firstly because of the lack of seed data. Secondly, even if sufficient information on data properties was available, auditors cannot import data for model input. Instead, the model requires meaningful related data, which cannot be synthesized (Fernsel et al., 2024b).

If suitable test data cannot be obtained, at least the reliability of the evaluation conducted by Monllaó Olivé can be judged. For that, the auditor needs to know the details of the quality evaluation *(organizational processes)* and the properties of the used training and test data *(data structure, provenance, and quality)*. Additionally, information on *model algorithms* (training and testing, including feedback loops) and which *model parameters* were chosen and why *(system design and implementation choices)* could help to identify erroneous implementations that lead to invalid evaluation results.

To prove *v2* (cognitive depth and social breadth are valid indicators), auditors need to verify whether this claim is scientifically sound and supported by studies. This information can be expected in the documentation on the translation of the "Community of Inquiry" framework into cognitive depth and social breadth indicators *(system design and implementation choices)*. Trust can be increased by examining the importance of each indicator on the predictions made on a Moodle instance that is already using the dropout prediction model *(data provenance)*. We call such systems "production systems".

Utility. The utility-related claim (reduced dropout in online courses) requires evidence that indicates the system's impact on student behavior in online courses. This information could be found in the documentation on the scientific foundation of the dropout prediction system and in conducted studies *(design and implementation choices)*, as well as in applied evidence-based design methods *(organizational processes)*. If such evidence is unavailable, auditors may verify the system's utility by analyzing the feedback given by humans for predictions; i.e., did a student drop out and did Moodle predict this correctly *(model performance)*—provided that they have access to a production system.

Ethics. To validate *e1* (AI-created predictions are marked as such) and *e2* (stakeholders can decide whether to use the system), documentation on *system functionality and limitations* can be helpful. To assess *e3* (GDPR compliance) documentation on *system functionality and limitations*, *design and implementation choices*, *algorithms* and *organizational processes* could show whether data privacy and security mechanisms have been included. Analyzing the source code could provide detailed information about the system's behavior.

Several pieces of evidence can help verify *e4* (equal model performance across groups). The documentation on *system functionality and limitations*, *design and implementation choices*, and *organizational processes* could reveal structural issues that might lead to a biased system. It could also contain information on how risks are handled. Evidence on the *model algorithms* and *parameters* can uncover further potential for ethical issues (Tagharobi and Simbeck, 2022). Documentation of the *model performance* by risk group could indicate the equality of prediction quality for different groups. Properties of the training and test data *(data provenance, data quality)* must be known to ensure the validity of the performance evaluation. Trust can be increased if *data* is available for reproducing or extending quality measurements.

4.3 Means of Validation

Where evidence is not directly available, it should be made accessible through means of validation. We argued that evidence can be made accessible through APIs, monitoring, or explainable AI mechanisms. In this subsection, we assess to what extent Moodle's dropout prediction system implements interfaces to access and collect evidence for validating claims.

API. Moodle does not provide an API for secure third-party access to the dropout prediction system. However, the internal "Analytics API" may be used to access and extend the machine learning capabilities of Moodle with a plugin (Monllaó Olivé et al., 2018), e.g., to add further monitoring functions. In a different publication, we used this approach to increase the auditability of Moodle successfully (Fernsel et al., 2024b).

Monitoring. The primary monitoring capability is the "evaluation mode" for evaluating model configurations or models trained on other Moodle instances (Moodle, 2023a). An auditor needs access to a test system to use this monitoring capability. The "evaluation mode" trains a new model on some of the data from finished courses on the platform and then tests it against the remaining

data (Moodle, 2023a). A model trained on a different platform is evaluated by testing it against the data on the new platform. When doing so, the model trained for evaluation is not retained. The "evaluation mode" returns two values per selected analysis interval: the weighted F1-score and the standard deviation (Moodle, 2023a). Depending on the chosen machine learning module, more values like the Matthews' correlation coefficient may be returned (Monllaó Olivé et al., 2018).

Smaller monitoring capabilities for production systems are also available. Auditors can monitor which courses cannot be used by the model, which students have been classified as at risk of dropping out, which indicators have been calculated for which student, and which human feedback has been given for the dropout prediction model: correct, "not applicable" or "incorrectly flagged" (Moodle, 2023a).

Explanations. Moodle integrates explanations for AI outputs in production systems. As mentioned above, Moodle monitors the results of the dropout prediction model together with the calculated indicators per student. Moodle highlights influential indicators to explain the model result (Moodle, 2023a).

4.4 Evidence Accessibility

Validity. The preferred way to prove *v1* is to reproduce the dropout prediction model performance assessments from (Monllaó Olivé et al., 2018), which appears to have been made with Moodle's "evaluation mode" for LA models (Moodle, 2023a). The performance assessment requires a test system to obtain logs of the *model performance*. However, as previously mentioned, *data sources* for the test system are not publicly available.

If no test system is available, the validity of the quality evaluation may be estimated by reviewing information on the properties and production of the system. Basic information on the *structure and provenance* of the data used for MoodleHQ's quality assessment describe (Monllaó Olivé et al., 2018). They also elaborate on the data quantity. The model training algorithms—logistic regression and a feed-forward neural network—*(model algorithm)* and the model evaluation methods are documented as well *(organizational processes)*. Auditors will need to analyze relevant parts of the source code: the values for fixed *model parameters* (like the number of training epochs, learning rate, or batch size) are neither documented nor logged but can only be found in the source code. Also, a source code analysis (Tagharobi and Simbeck, 2022) found an undocumented 500MB limit for training data.

To conclude, evidence is insufficiently available and accessible to fully validate claim *v1* that the dropout prediction model correctly predicts dropout risks. The documentation on the cognitive depth and social breadth indicators *(system design and implementation choices)* offers a starting point for auditors to validate *v2* (Moodle, 2023a; Monllaó Olivé et al., 2018). MoodleHQ does not provide studies that support their indicator definitions. If an auditor can access a production system, she could review the explanations logged for the model's predictions *(data provenance)* and evaluate the soundness of the chosen indicators. Since this type of access could be challenging, we deem the available evidence insufficient to effectively assess claim *v2* that cognitive depth and social breadth are valid indicators.

Utility. To assess the utility-related claim *u1*, the first step is to review the available documentation. The scientific theory behind the choice of model features is explained thoroughly (Monllaó Olivé et al., 2018), but no studies on the model's impact are documented *(design and implementation choices)*. Project management, design, and development processes are not documented either *(organizational processes)*. Production system-specific utility may be analyzed by viewing aggregated information about the feedback given by humans for predictions *(model performance)*. In summary, the available evidence does not validate the claim *u1* that the dropout prediction system reduces dropout rates in online courses.

Ethics. To validate ethics-based claim *e1*, documentation on *system functionality and limitations* can be reviewed. The screenshots displayed in the Moodle documentation show that users viewing dropout predictions are made aware of the uncertainty of predictions (Moodle, 2023a). However, users are not explicitly informed that the AI-based LA system calculates the predictions. The available evidence allows to reject claim *e1*.

Documentation on *system functionality and limitations* could also help validate claim *e2*. The documentation shows that teachers can decide how to use the predictions, and administrators can turn the dropout prediction system on or off (Moodle, 2023a). Students do not appear able to opt in or out of being classified by the dropout prediction model. In conclusion, the available evidence indicates that claim *e2*—stakeholders can decide whether to use the dropout prediction system—can only partly be confirmed.

Concerning *e3*, the documentation on *system functionality and limitations*, *design and implementation choices*, and *organizational processes* does not explicate actions to comply with the GDPR, except that exportable data is anonymous and access to insights can be managed (Moodle, 2023a). We conclude that only a source code analysis *(algorithms, system func-*

tionality, and limitations) can assess the claim that the dropout prediction system is GDPR compliant.

To validate *e4*, documentation on the choice of model features and their underlying principles *(design and implementation choices)*, as well as limiting technical factors *(system functionality and limitations)* hint at existing or absent bias in the dropout prediction system (Monllaó Olivé et al., 2018; Moodle, 2024; Moodle, 2023a). Source code analysis is required to complement the documentation. However, it cannot rule out any bias (Tagharobi and Simbeck, 2022). Evaluating the *model performance* per group could provide additional evidence on model fairness. Such a quality assessment is not documented and thus needs to be conducted by the auditor. Access to a production system (including the database) and *data sources*, including demographic data, is necessary. No evidence could be found that the risk of model bias was considered in the design and development (*organizational processes*). No information on the *data provenance* (e.g., information on data acquisition and pre-processing) or relevant *data quality* (e.g., information on representativeness) for MoodleHQ's quality assessment is available (Monllaó Olivé et al., 2018). We conclude that insufficient evidence is available and accessible for an efficient audit of claim *e4* that model performance is equally high across groups.

5 DISCUSSION

We have demonstrated that our auditability assessment framework is helpful for AI-based LA systems by successfully applying it to Moodle's dropout prediction feature.[1] Through the structured approach of claims, evidence, and means of validation analysis, we especially predicted challenges that would await an audit of Moodle's dropout prediction feature. This can inform the development of suitable fixes and features that retrofit auditability (Fernsel et al., 2024b).

Although Moodle is open source, sufficiently documented, and includes a comprehensive logging system with explanations, only three of seven identified claims are effectively auditable. The lack of documentation depth primarily constrains auditability. More documentation is needed on system design and implementation choices to justify the validity and ethical design of the system. When documentation is incomplete or not trustworthy enough, additional evidence for an audit of the dropout prediction system must be collected from the system, e.g., by monitoring. Two significant challenges hinder this approach. The first challenge is the inadequacy of Moodle's model monitoring capability. Predictions are not preserved when evaluating a model configuration and are inaccessible to the auditor. Thus, the auditor cannot verify the model's performance and has to rely on minimal metrics returned by Moodle. Implementing an API to execute individual evaluation steps and retrieve intermediate data or extend monitoring capabilities prevents this shortcoming. The second challenge is the current absence of publicly available test data. Therefore, data-based audits of Moodle's dropout prediction model are very resource-intensive.

The assessment of Moodle shows that future LA systems need to provide system access to third-party auditors, e.g., by creating "auditor" roles, recording data (anonymized training data, predictions), and enabling auditors to control evaluation parameters. The auditability assessment framework bears one caveat when applied to any AI-integrating system. It is time and effort-consuming to assess and improve auditability if claims and evidence must be prepared and the technical means for validation are not readily accessible. We are confident the improved audit quality will justify these additional costs. The auditability framework assists in mitigating challenges and in procuring the ongoing development of more robust and ethically fair software for "high-risk" application fields.

6 CONCLUSION

AI in education continues to gain in importance. Regular auditing is essential for sustainable learning success that achieves fair, non-discriminatory applications. Despite the increasing demand for auditing AI systems, auditability is a neglected design requirement for most AI systems. For this reason, we have sought to define auditability to improve transparent and traceable audits of AI-based Learning Analytics in development and deployment. Lacking auditability negatively impacts independent audits, which include lack of documentation, restricted access to the system and its raw sources (code, model weights, or data), and incomprehensible system output (Berghoff et al., 2022; Mökander and Floridi, 2021; Alikhademi et al., 2022; Tagharobi and Simbeck, 2022). Additionally, system-independent factors, such as heterogeneous ethical standards (Mökander and Floridi, 2021) and difficulty achieving test coverage for AI-integrating systems (Berghoff et al., 2022; Tao et al., 2019) diminish auditability.

Following a review of auditability in general, AI audit challenges, and factors enabling AI auditability, we suggest a framework for a systematic approach to assess and ensure specific requirements for the auditability of AI-based LA. Our framework is based on three pillars: claims, evidence, and means of validation. To make

[1] We demonstrated the applicability of the framework to prototype AI-based LAs in (Fernsel et al., 2024a).

AI systems auditable, system providers and deployers must provide certifiable claims about utility, validity, and ethics (Landers and Behrend, 2023; Brundage et al., 2020). Depending on the claims and the audit procedure, substantial evidence must be made available to auditors: *evidence types* include documentation, raw sources, and logs (Brundage et al., 2020; Tagharobi and Simbeck, 2022; Beckstrom, 2021; Raji et al., 2020). *Evidence subjects* are the overall system, models and data (European Union, 2024). AI-integrating systems should provide APIs (Springer and Whittaker, 2019), monitoring tools (Ashmore et al., 2022; Bharadhwaj et al., 2021; European Union, 2024; Eitel-Porter, 2021; Alhajaili and Jhumka, 2019) and explanations (Brundage et al., 2020; Shneiderman, 2020; Guidotti et al., 2018) to enable the validation of evidence. Audit requirements and standards for AI audits are being developed. However, legislators and standardization bodies must consider auditability requirements as well. We see this as an important leverage point where our framework can be applied to derive process requirements for external audits, implement auditability by design in the QA of system development, and give stakeholders a way to insist on consistent audits. Finally, the framework supports developing and maintaining robust, trustworthy AI-based LA systems that foster acceptance among students and teaching professionals.

We conclude that the proposed framework is useful for auditors and system providers to prepare for an audit and determine how much an AI-integrating LA system is auditable. Moreover, developers of AI-integrating systems can benefit from the framework by identifying areas for improving the auditability of their products. We appeal to developers of AI-integrating systems to consider auditability right from the start when designing their systems to ensure trustworthy, ethical, and future-fit products that comply with current and upcoming legislation, such as the European AI Act. Considering that LA can potentially enhance learning outcomes (Lang et al., 2022), increasing the auditability of LA systems ultimately leads to an improved learning experience for a broader audience.

ACKNOWLEDGEMENTS

This publication is part of the research project “Fair Enough? Investigating the fairness of learning analytics systems”, which was funded by the German Federal Ministry of Education and Research (BMBF) Grant No.: 16DHB4002/3. The authors would like to thank the various reviewers who provided valuable comments at different stages of the paper.

REFERENCES

Alagić, A., Turulja, L., and Bajgorić, N. (2021). Identification of Information System Audit Quality Factors. *Journal of Forensic Accounting Profession*, 1(2):1–28.

Alhajaili, S. and Jhumka, A. (2019). Auditability: An Approach to Ease Debugging of Reliable Distributed Systems. In *2019 IEEE 24th Pacific Rim International Symposium on Dependable Computing (PRDC)*, pages 227–2278, Kyoto, Japan. IEEE.

Alikhademi, K., Drobina, E., Prioleau, D., Richardson, B., et al. (2022). A review of predictive policing from the perspective of fairness. *Artificial Intelligence and Law*, 30(1):1–17.

Alla, S. and Adari, S. K. (2021). What Is MLOps? In *Beginning MLOps with MLFlow: Deploy Models in AWS SageMaker, Google Cloud, and Microsoft Azure*, pages 79–124. Apress, Berkeley, CA.

Ashmore, R., Calinescu, R., and Paterson, C. (2022). Assuring the Machine Learning Lifecycle: Desiderata, Methods, and Challenges. *ACM Computing Surveys*, 54(5):1–39.

Awwad, Y., Fletcher, R., Frey, D., Gandhi, A., et al. (2020). *Exploring Fairness in Machine Learning for International Development*. MIT D-Lab, Cambridge.

Ayling, J. and Chapman, A. (2022). Putting AI Ethics to Work. *AI and Ethics*, 2(3):405–429. PII: 84.

Baek, C. and Doleck, T. (2023). Educational Data Mining versus Learning Analytics: A Review of Publications From 2015 to 2019. *Interactive Learning Environments*, 31(6):3828–3850.

Baker, R. S. and Hawn, A. (2022). Algorithmic Bias in Education. *International Journal of Artificial Intelligence in Education*, 32(4):1052–1092.

Beckstrom, J. R. (2021). Auditing machine learning algorithms. A white paper for public auditors. *International Journal of Government Auditing*, 48(1):40–41.

Berghoff, C., Böddinghaus, J., Danos, V., Davelaar, G., et al. (2022). Towards Auditable AI Systems: From Principles to Practice.

Bharadhwaj, H., Huang, D.-A., Xiao, C., Anandkumar, A., et al. (2021). Auditing AI models for Verified Deployment under Semantic Specifications.

Bhatt, U., Xiang, A., Sharma, S., Weller, A., Taly, A., et al. (2020). Explainable machine learning in deployment. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* ’20, pages 648–657. ACM.

Bose, R. P. J. C., Singi, K., Kaulgud, V., Phokela, K. K., and et al. (2019). Framework for Trustworthy Software Development. In *2019 34th IEEE/ACM International Conference on Automated Software Engineering Workshop (ASEW)*, pages 45–48, San Diego, CA. IEEE.

Brundage, M., Avin, S., Wang, J., Belfield, H., et al. (2020). Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims.

Crescenzi-Lanna, L. (2020). Multimodal Learning Analytics Research with young Children: A systematic Review. *British Journal of Educational Technology*, 51(5):1485–1504.

Dorodchi, M., Al-Hossami, E., Benedict, A., and Demeter, E. (2019). Using synthetic data generators to promote open science in higher education learning analytics. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 4672–4675. IEEE.

Drugova, E., Zhuravleva, I., Zakharova, U., and Latipov, A. (2024). Learning Analytics driven Improvements in Learning Design in higher Education: A systematic Literature Review. *Journal of Computer Assisted Learning*, 40(2):510–524.

Eitel-Porter, R. (2021). Beyond the promise: implementing ethical AI. *AI and Ethics*, 1(1):73–80.

El Emam, K., Mosquera, L., and Hoptroff, R. (2020). *Practical Synthetic Data Generation*. O'Reilly Media, Inc.

Elmoazen, R., Saqr, M., Khalil, M., and Wasson, B. (2023). Learning Analytics in virtual Laboratories: A systematic Literature Review of empirical Research. *Smart Learning Environments*, 10(1).

European High-Level Expert Group on AI (2019). Ethics guidelines for trustworthy AI — Shaping Europe's digital future.

European Parliament (2023). Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts.

European Union (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence.

Falco, G., Shneiderman, B., Badger, J., Carrier, R., Dahbura, A., Danks, D., Eling, M., Goodloe, A., Gupta, J., Hart, C., Jirotka, M., Johnson, H., LaPointe, C., Llorens, A. J., Mackworth, A. K., Maple, C., Pálsson, S. E., Pasquale, F., Winfield, A., and Yeong, Z. K. (2021). Governing AI safety through independent audits. *Nature Machine Intelligence*, 3(7):566–571.

Fernsel, L., Kalff, Y., and Simbeck, K. (2024a). Assessing the auditability of ai-integrating systems: A framework and learning analytics case study. *arXiv*, 2411.08906.

Fernsel, L., Kalff, Y., and Simbeck, K. (2024b). Where is the evidence? In Poquet, O., Ortega-Arranz, A., Viberg, O., Chounta, I.-A., McLaren, B. M., and Jovanovic, J., editors, *Proceedings of the 16th International Conference on Computer Supported Education (CSEDU 2024)*, volume 2, pages 262–269.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., et al. (2020). Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. *Berkman Klein Center Research Publication*, (1).

Galdon-Clavell, G., Zamorano, M. M., Castillo, C., Smith, O., et al. (2020). Auditing Algorithms: On Lessons Learned and the Risks of Data Minimization. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pages 265–271, New York. ACM.

Gardner, J., Brooks, C., and Baker, R. (2019). Evaluating the Fairness of Predictive Student Models Through Slicing Analysis. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, LAK19, pages 225–234, New York. ACM.

Garrison, D. R., Anderson, T., and Archer, W. (1999). Critical Inquiry in a Text-Based Environment: Computer Conferencing in Higher Education. *The Internet and Higher Education*, 2(2):87–105.

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., et al. (2021). Datasheets for datasets. *Communications of the ACM*, 64(12):86–92.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., and Pedreschi, D. (2018). A Survey of Methods for Explaining Black Box Models. *ACM Computing Surveys*, 51(5):93:1–93:42.

Haim, A., Shaw, S., and Heffernan, N. (2023). How to open science: A principle and reproducibility review of the learning analytics and knowledge conference. In *LAK23: 13th International Learning Analytics and Knowledge Conference*, pages 156–164.

Heikkinen, S., Saqr, M., Malmberg, J., and Tedre, M. (2023). Supporting self-regulated Learning with Learning Analytics Interventions: A systematic Literature Review. *Education and Information Technologies*, 28(3):3059–3088.

Holland, S., Hosny, A., Newman, S., Joseph, J., and Chmielinski, K. (2018). The Dataset Nutrition Label: A Framework To Drive Higher Data Quality Standards.

Kale, A., Nguyen, T., Harris, Frederick C., Jr., Li, C., Zhang, J., and Ma, X. (2022). Provenance documentation to enable explainable and trustworthy AI: A literature review. *Data Intelligence*, pages 1–41.

Khalil, M., Slade, S., and Prinsloo, P. (2023). Learning Analytics in Support of Inclusiveness and disabled Students: A systematic Review. *Journal of Computing in Higher Education*, pages 202–219.

Kitto, K. and Knight, S. (2019). Practical ethics for building learning analytics. *British Journal of Educational Technology*, 50(6):2855–2870.

Knechel, W. R., Krishnan, G. V., Pevzner, M., Shefchik, L. B., and Velury, U. K. (2013). Audit Quality: Insights from the Academic Literature. *AUDITING: A Journal of Practice & Theory*, 32(Supplement 1):385–421.

Kreuzberger, D., Kühl, N., and Hirschl, S. (2022). Machine Learning Operations (MLOps): Overview, Definition, and Architecture.

Landers, R. N. and Behrend, T. S. (2023). Auditing the AI Auditors. A Framework for Evaluating Fairness and Bias in high stakes AI predictive Models. *The American psychologist*, 78(1):36–49.

Lang, C., Siemens, G., Friend Wise, A., Gašević, D., and Merceron, A., editors (2022). *Handbook of Learning Analytics - Second edition*. Society for Learning Analytics Research (SoLAR), 2 edition.

Li, Y. and Goel, S. (2024). Artificial Intelligence Auditability and Auditor Readiness for Auditing Artificial Intelligence Systems. *SSRN Journal*.

Minkkinen, M., Niukkanen, A., and Mäntymäki, M. (2024). What about investors? ESG analyses as tools for ethics-based AI auditing. *AI & Society*, 39(1):329–343. PII: 1415.

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., and Gebru, T. (2019). Model Cards for Model Reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, pages 220–229, New York. Association for Computing Machinery.

Mökander, J., Axente, M., Casolari, F., and Floridi, L. (2022). Conformity Assessments and Post-market Monitoring: A Guide to the Role of Auditing in the Proposed European AI Regulation. *Minds & Machines*, 32(2):241–268.

Mökander, J. and Floridi, L. (2021). Ethics-Based Auditing to Develop Trustworthy AI. *Minds and Machines*, 31(2):323–327.

Monllaó Olivé, D., Du Huynh, Q., Reynolds, M., Dougiamas, M., et al. (2018). A supervised learning framework for learning management systems. In *Proceedings of the First International Conference on Data Science, E-learning and Information Systems*, DATA '18, pages 1–8, New York. Association for Computing Machinery.

Moodle (2023a). Documentation.

Moodle (2023b). Moodle and our AI principles.

Moodle (2024). Analytics API.

Naja, I., Markovic, M., Edwards, P., Pang, W., et al. (2022). Using Knowledge Graphs to Unlock Practical Collection, Integration, and Audit of AI Accountability Information. *IEEE Access*, 10:74383–74411.

Nushi, B., Kamar, E., and Horvitz, E. (2018). Towards Accountable AI: Hybrid Human-Machine Analyses for Characterizing System Failure. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, volume 6, pages 126–135, Zurich, Switzerland. AAAI Press.

Ouyang, F., Wu, M., Zheng, L., Zhang, L., et al. (2023). Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering course. *International Journal of Educational Technology in Higher Education*, 20(1):4.

Paolucci, C., Vancini, S., Bex Ii, R. T., Cavanaugh, C., Salama, C., and de Araujo, Z. (2024). A review of learning analytics opportunities and challenges for K-12 education. *Heliyon*, 10(4):e25767.

Prinsloo, P. and Slade, S. (2017). Ethics and Learning Analytics: Charting the (Un)Charted. In Lang, C., Siemens, G., Wise, A., Gasevic, D., and University of Edinburgh, U. K., editors, *Handbook of Learning Analytics*, pages 49–57. Society for Learning Analytics Research (SoLAR).

Raji, I. D., Smart, A., White, R. N., Mitchell, M., et al. (2020). Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* '20, pages 33–44. ACM.

Romero, C. and Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, 10(3).

Rzepka, N., Fernsel, L., Müller, H.-G., Simbeck, K., and Pinkwart, N. (2023). Unbias me! Mitigating Algorithmic Bias for Less-studied Demographic Groups in the Context of Language Learning Technology. *Computer-Based Learning in Context*, 6(1):1–23.

Rzepka, N., Simbeck, K., Müller, H.-G., and Pinkwart, N. (2022). Fairness of In-session Dropout Prediction. In *Proceedings of the 14th International Conference on Computer Supported Education (CSEDU)*, pages 316–326. Scitepress.

Sghir, N., Adadi, A., and Lahmer, M. (2022). Recent advances in Predictive Learning Analytics: A decade systematic review (2012-2022). *Education and Information Technologies*, pages 1–35.

Shneiderman, B. (2020). Human-Centered Artificial Intelligence: Three Fresh Ideas. *AIS Transactions on Human-Computer Interaction*, pages 109–124.

Simbeck, K. (2024). They shall be fair, transparent, and robust: auditing learning analytics systems. *AI and Ethics*, 4.

Slade, S. and Tait, A. (2019). *Global guidelines: Ethics in Learning Analytics*. ICDE.

Springer, A. and Whittaker, S. (2019). Making Transparency Clear: The Dual Importance of Explainability and Auditability. In *Joint Proceedings of the ACM IUI 2019 Workshops*, page 4, Los Angeles. ACM.

Stoel, D., Havelka, D., and Merhout, J. W. (2012). An analysis of attributes that impact information technology audit quality: A study of IT and financial audit practitioners. *International Journal of Accounting Information Systems*, 13(1):60–79.

Susnjak, T. (2024). Beyond Predictive Learning Analytics Modelling and onto Explainable Artificial Intelligence with Prescriptive Analytics and ChatGPT. *International Journal of Artificial Intelligence in Education*, 34(2):452–482. PII: 336.

Tagharobi, H. and Simbeck, K. (2022). Introducing a Framework for Code based Fairness Audits of Learning Analytics Systems on the Example of Moodle Learning Analytics. In *Proceedings of the 14th International Conference on Computer Supported Education (CSEDU)*, volume 2, pages 45–55. Scitepress.

Tao, C., Gao, J., and Wang, T. (2019). Testing and Quality Validation for AI Software–Perspectives, Issues, and Practices. *IEEE Access*, 7:120164–120175.

Toreini, E., Aitken, M., Coopamootoo, K. P. L., Elliott, K., Zelaya, V. G., Missier, P., Ng, M., and van Moorsel, A. (2022). Technologies for Trustworthy Machine Learning: A Survey in a Socio-Technical Context.

Weigand, H., Johannesson, P., Andersson, B., and Bergholtz, M. (2013). Conceptualizing Auditability. In Deneckère, R. and Proper, H. A., editors, *Proceedings of CAiSE'13*, page 8, Valencia, Spain. CEUR.

Williams, R., Cloete, R., Cobbe, J., Cottrill, C., et al. (2022). From transparency to accountability of intelligent systems: Moving beyond aspirations. *Data & Policy*, 4(2022).

Wolnizer, P. W. (2006). *Auditing as Independent Authentication*. Sydney University Press, Sydnesy.

Xiong, Z., Li, H., Liu, Z., Chen, Z., et al. (2024). A Review of Data Mining in Personalized Education: Trends and Future Prospects Current. *Frontiers of Digital Education*, 1(26-50).