Where Is the Evidence? A Plugin for Auditing Moodle's Learning Analytics

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

Keywords: Auditability, Artificial Intelligence, Learning Analytics, Moodle, Plugin Development.

Abstract: The paper presents the work-in-progress development of a Moodle plugin to improve the auditability of Moodle's Learning Analytics component. Future legislation, such as the EU AI Act, will require audits and "conformity assessments" of "high-risk" AI systems. Educational applications can be considered high-risk systems due to their important role in individual life and career paths. Therefore, their correctness, fairness, and efficiency must be assessed. However, auditing of the Learning Analytics functions in Moodle is limited. No suitable test-data is available, models and configurations are not persistent and only aggregated quality metrics are returned that are insufficient to assess fairness. The plugin addresses these issues and provides a data interface to extract data for audits. The plugin allows to a) upload and select data for the audit, b) clearly differentiate between model configuration and trained models, c) keep trained models, their configuration and underlying data for future inspections and comparisons, and finally, d) the plugin saves raw predictions for further analysis. The plugin enables the audit of Moodle's Learning Analytics and its underlying AI models and contributes to increased fairness and trustworthiness of Learning Analytics as well as its legally compliant application.

1 INTRODUCTION

Learning Analytics play an increasingly crucial role in shaping the learning experience in today's educational landscape (Ouhaichi et al., 2023; Kaddoura et al., 2022). AI-based Learning Analytics components utilize methods of machine learning to process learning data for analysis and predictions in educational contexts (Alam, 2023, 572). However, these AI-based systems can suffer from bias in models and datasets, reproduce inequalities and discrimination and thus, risk their trustworthiness (Rzepka et al., 2022). Therefore, it is imperative to ensure the correctness, fairness, and efficiency of the underlying AI models (Simbeck, 2023). More so, ongoing legislative efforts will make audits mandatory for high-risk AI systems to ensure their quality and legality, which will potentially affect educational AI-based systems (European Commission, 2021, no. 35). Audits verify that Learning Analytics components perform as intended and align with ethical values (Springer and Whittaker, 2019). Auditing is not just a safeguard, it's also a pathway to improve Learning Analytics components, to enhance their quality, and to foster trust and acceptance (Bose et al., 2019). To ensure that audits are applicable to AI-based Learning Analytics, the systems need to be auditable and accessible to third parties.

The open-source application Moodle is a widely used tool in teaching and learning environments. In addition to its learning management functions, Moodle integrates a Learning Analytics component implemented in `PHP` (Monllaó Olivé et al., 2020). This AI-based Learning Analytics component utilizes historical and current behavior data to train Logistic Regression models, predict students' performance factors, and identify students at risk of failing or dropping out of courses (Monllaó Olivé et al., 2020). It aims to assist educators in making informed decisions and identifying students who may require additional support. However, similar algorithms have been found to be biased in the past (Rzepka et al., 2022; Hu and Rangwala, 2020). Thus, Moodle's Learning Analytics must undergo "conformity assessments" (European Commission, 2021, 13) to evaluate statements about its functionality and fairness. In the case of Moodle, however, such audits are limited, because the auditability

[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.
Where Is the Evidence? A Plugin for Auditing Moodle's Learning Analytics.
DOI: 10.5220/0012689800003693

In *Proceedings of the 16th International Conference on Computer Supported Education (CSEDU 2024) - Volume 2*, pages 262-269
ISBN: 978-989-758-697-2; ISSN: 2184-5026

of Moodle's Learning Analytics component is unsatisfactory (Fernsel et al., tbd).

Here, our paper ties in: we extend the auditability of Moodle's Learning Analytics component by developing a plugin to facilitate evidence-based audits. The plugin solves various bottlenecks for Moodle's assessment of fairness and thus increases the overall auditability of Moodle's Learning Analytics component. Our research questions are as follows:

RQ1. What challenges do audits of Moodle's Learning Analytics face?

RQ2. How can plugins extend Moodle's auditability?

The position paper is structured as follows: section 2 outlines requirements for audits in general and particularly for Moodle's Learning Analytics function. Section 3 evaluates the auditability of Moodle as a challenge to generate suitable evidence for audits – especially regarding data availability, model testing, model comparisons, and metrics to assess fairness or biases. Section 4 outlines the function of the plugin "LaLA" (abbr. for Let's audit Learning Analytics) to counteract said shortcomings, while section 5 outlines an audit of Moodle's Learning Analytics using the plugin. Section 6 discusses the potential of interfaces and plugins for auditability, especially for open-source software.

2 BACKGROUND: PREREQUISITES FOR AUDITS

According to the AI-Act (European Commission, 2021, no. 35), AI systems in education or vocational training are potentially high-risk applications that impact individual learning and educational trajectories. Thus, audits to assess correctness, fairness, efficiency, and adherence to ethical standards control and support the functioning of such systems as well as promote acceptance and trust in their correctness (Mökander and Floridi, 2021; Simbeck, 2023). Periodical auditing, therefore, is necessary and Moodle's Learning Analytics provides no exception.

We call any system auditable when external auditors can review and verify it independently (Williams et al., 2022; Wolnizer, 2006). For this, external reviews need information on *claims* how relevant values *should* be used or produced. Further, an audit requires the system to provide *evidence* on how relevant values *actually are* used or produced. Audits then, validate claims on grounds of the provided evidence (Weigand et al., 2013; Mökander and Floridi, 2021; Fernsel et al., tbd).

Claims are normative statements about the functioning of a system. They usually are defined by the system provider (Stoel et al., 2012), can be derived from laws, regulations, or standards. *Evidence* are records of relevant information to back respective normative claims about the functioning of an AI system (Alhajaili and Jhumka, 2019). Means of *validation* enable auditors to access and validate the provided evidence. Therefore, "designing for auditability" (Zook et al., 2017, 7) implies that any AI system should reflect and accentuate its inherent claims and enable the collection of suitable evidence to facilitate validations by first or third-party assessments (Hutchinson et al., 2021; Awwad et al., 2020; Stoel et al., 2012).

A wide variety of challenges and limitations to audits exist and affect the auditability of Learning Analytics in general and Moodle's AI-integrating Learning Analytics in particular (Toreini et al., 2022; Raji et al., 2020). A challenge lies in defining verifiable claims. The absence of binding guidelines forces auditors in practice to decide consciously on ethical values, which therefore are highly subjective and situational (Rzepka et al., 2022; Landers and Behrend, 2022). A second challenge is access to suitable evidence to validate these claims. Typically, neither a system nor its raw sources (program code, model weights, data used for training and testing) are accessible to auditors. This holds especially true for proprietary or security-sensitive software systems (Raghavan et al., 2020; Alikhademi et al., 2022). Under these circumstances, auditors can only conduct data-based audits and imitate models, which makes an audit less conclusive (Alikhademi et al., 2022).

3 EVALUATION: AUDITING MOODLE

Any audit to ascertain the proposed claims based on evidence from Moodle's Learning Analytics component and the algorithms themselves relies on the system's onboard tools, as well as on available documentation, source code, logs and data (Fernsel et al., tbd). When evaluating Moodle concerning correctness, fairness, and efficiency of Learning Analytics components, three major obstacles emerge. The assessment relies on realistic test-data that includes both majority and minority groups. Further, the audit requires possibilities to evaluate the underlying models. Finally, fairness assessments need appropriate metrics to allow informed statements about their reliability. However, the lack of realistic test-data, insufficient possibilities for model evaluation and limited metrics inhibit Moodle's auditability (Fernsel et al., tbd). More precisely, the problem is that Moodle does not provide sufficient evidence to validate respective claims (Raghavan et al., 2020).

3.1 Lack of Test-Data

The primary issue is concerned with the type of evidence required to substantiate the claims about the system. A simple examination of documentation, source code, or logs proved to be insufficient. Therefore, a more in-depth approach was necessary that consisted of evaluative tests of the applied models with coherent test-data. The audit process requires diverse and realistic data, representative of both majority and minority groups to make clear statements on potential biases. Predictions generated by the production model have to be scrutinized for biases and compared to a model that is trained with representative data. Statements about underrepresentation in an AI model's decisions require insights in the statistical population. Only then, one can comprehensively validate the underlying claims of Moodle's fairness.

However, executing a model test is challenged by the lack of suitable, openly accessible test-data for conducting model tests. Test-data was limited to data of production systems that requires permissions to be utilized for privacy reasons. Above that, such data usually is pre-biased: the data is not representative on a societal level because of, for example, unevenly distributed access chances to higher education (Suresh and Guttag, 2021; Mihaljević et al., 2023). Additionally, the intricacies of Learning Analytics models demand complex, logically structured sequential behavioral data, making it nearly impossible to mock test-data on scales necessary for model testing. Further, on a technical level, Moodle does not offer an interface to directly import external test-data into the system. Data-wise, any audit is restricted to the available data from running instances, which usually is imbalanced, has opaque statistical populations or biased distributions. Above that the data must be anonymized beforehand.

3.2 Limits of Moodle's Evaluation Mode

Moodle features an "evaluation mode" that trains models on part of the available data and tests them on the remaining data. However, this feature comes with limitations that further hindered the validation process. This mode exclusively evaluates *model configurations* rather than existing trained models. The standard procedure allows auditors to inspect the models' underlying configurations including indicators (features) and used predictions processors (e.g. the `php` machine learning backend). However, even if balanced test-data could be imported into the Moodle instance, the audit could not assess the trained models for fairness or biases. This renders any audit more of an approximation rather than a direct assessment of the models in production use.

Another limitation of Moodle's evaluation mode is its lacking ability to keep data in between evaluations. Models generated during the evaluation mode are not persisted beyond that evaluation phase, which impedes a more detailed analysis of concrete models employed on the platform. Especially the comparison of different models with test-datasets proved to be impossible, hindering any effort to audit Moodle's Learning Analytics component.

3.3 Metrics for Fairness Assessment

The third challenge pertained to the insufficient evaluation mode as well. It particularly concerns fairness assessments. Moodle's evaluation mode provides limited information, primarily in the form of aggregated metrics, which are unsatisfactory to validate claims related to fairness (Castelnovo et al., 2022). The mode does not provide raw predictions but only two simple metrics: an F1 score and its standard deviation in ten rounds of training and testing a model. However, Moodle refers to this metric as accuracy, which it is not: the F1 score is the harmonic mean of precision and recall (Jeni et al., 2013, 248). These metrics lack the granularity required for robust validation, especially in cases that involve group-based comparisons, which are essential for fairness-based claims. To assess the fairness of Moodle's AI-based Learning Analytics, additional metrics besides accuracy are necessary: precision, recall, specificity, false negative rates or false positive rates (Verma and Rubin, 2018). These metrics are fundamental for detailed assessments of the model performance and group-based comparative statistics for fairness audits.

3.4 Evaluation Results and Solutions

The three challenges, lacking and unmockable test-data, a restricted evaluation mode that cannot assess trained models from production usage or maintain data in-between evaluations, as well as insufficient fairness metrics restrict Moodle's auditability. Due to Moodle's open-source nature, these problems can be countered by software engineering. A plugin can provide a suitable extension of Moodle's onboard auditing capabilities and offer a more convenient approach to assess Learning Analytics models or datasets.

4 PLUGIN DESIGN AND IMPLEMENTATION

A software engineering approach responded to the limitations to the auditability of Moodle's Learning Analytics components. "LALA" is a plugin for Moodle to retrofit essential functionalities to audit and assess Moodle's Learning Analytics component and, thus, enhance its auditability. The following subsections delineate the strategies to develop the LALA plug-in and to mitigate the identified challenges. The plugin seeks to present a more robust auditing experience that adheres to a general audit framework of claims, evidence and validation (Fernsel et al., tbd).

4.1 Technical Details of the Plugin

LALA is a Moodle "admin tool" plugin written in `php` (Fernsel, 2024). Figure 1 displays the most important implemented `php` classes. The diagram omits the various helper classes implemented for data processing tasks as well as classes that implement Moodle functionalities such as an event definition, a privacy provider and output renderers.

The plugin clearly differentiates between the configurations, which are generated from the available logistic regression models' configurations in the Learning Analytics components, and model versions which are created from the model configuration. Database tables are created for both, the model configurations and the model versions (Figure 2). During creation, the model version collects and stores evidence step by step, including the trained logistic regression model itself and different types of datasets produced in the creation process. Meta data of each piece of evidence is stored in a third database table. A dataset object contains all features calculated from the gathered data. The training and test datasets contain each a split of these features. The predictions dataset contains the results from using the trained logistic regression model on the test dataset. Related data is recursively gathered from data tables referenced by the subjects of the prediction. E.g. when predictions are created per student enrolment, relevant rows and columns of Moodle's `user_enrolments` table are returned. Because that table references the `user` and `enrol` tables, the relevant contents of those are returned as well. Because `enrol` references `course` and `role`, those also count as related data. Due to the dependence on production data, related data is not available when using test-data imported via LALA. The plugin uses the anonymized versions of the evidence classes, except when uploading own test-data.

4.2 Data Persistence

Moodle's Learning Analytics evaluation mode does not save models. LALA stores and retrieves these models to preserve evidence. This makes different trained models and their configurations comparable. Further, in its original state, updates to a model configuration did not persist former versions. LALA adapts versioning of newly created model configurations, keeping old configurations, even if they are deleted in the Moodle Learning Analytics component. Persisting evidence for future audits with available comparisons of models greatly improves the auditability of Moodle's Learning Analytics. Persisting evidence allows to reproduce and compare results and to validate the system's claims.

To achieve data persistence, the plugin serializes each dataset collected during the model version creation process and stores it as a `csv` file on the Moodle server. The location of the file is saved in the evidence database table and enables the download of the file for auditors. In the dashboard, new options to download datasets, models, training data, and so on, appear (Figure 3). The downloadable data includes the used test-data as well. This way, it can be imported into a different instance of Moodle for an audit or re-uploaded for future assessments as described in subsection 4.4.

4.3 Model Configuration vs. Trained Model

Moodle in its original state allows auditors to assess only model configurations. Those are not the actual trained models used for making predictions. Therefore, they are not suitable for in-depth audits of biases or fairness that might arise from the training data and model training. Model configurations allow, however, a first impression about the included indicators (i.e. features of the model) and if they are sensible. The evaluation mode, when used in a production system, can help estimate for which courses which model configurations lead to meaningful predictions.

LALA clearly distinguishes between model configurations and trained models. While sane configurations are important to create functioning models, the overall performance and fairness of a model cannot be derived by the included indicators alone. For any audit it is critical to have access to the final trained models, because any biases or misrepresentations in training data will lead to misaligned models that reproduce discriminatory decisions, categorizations or predictions and that are error prone to misrepresented groups.

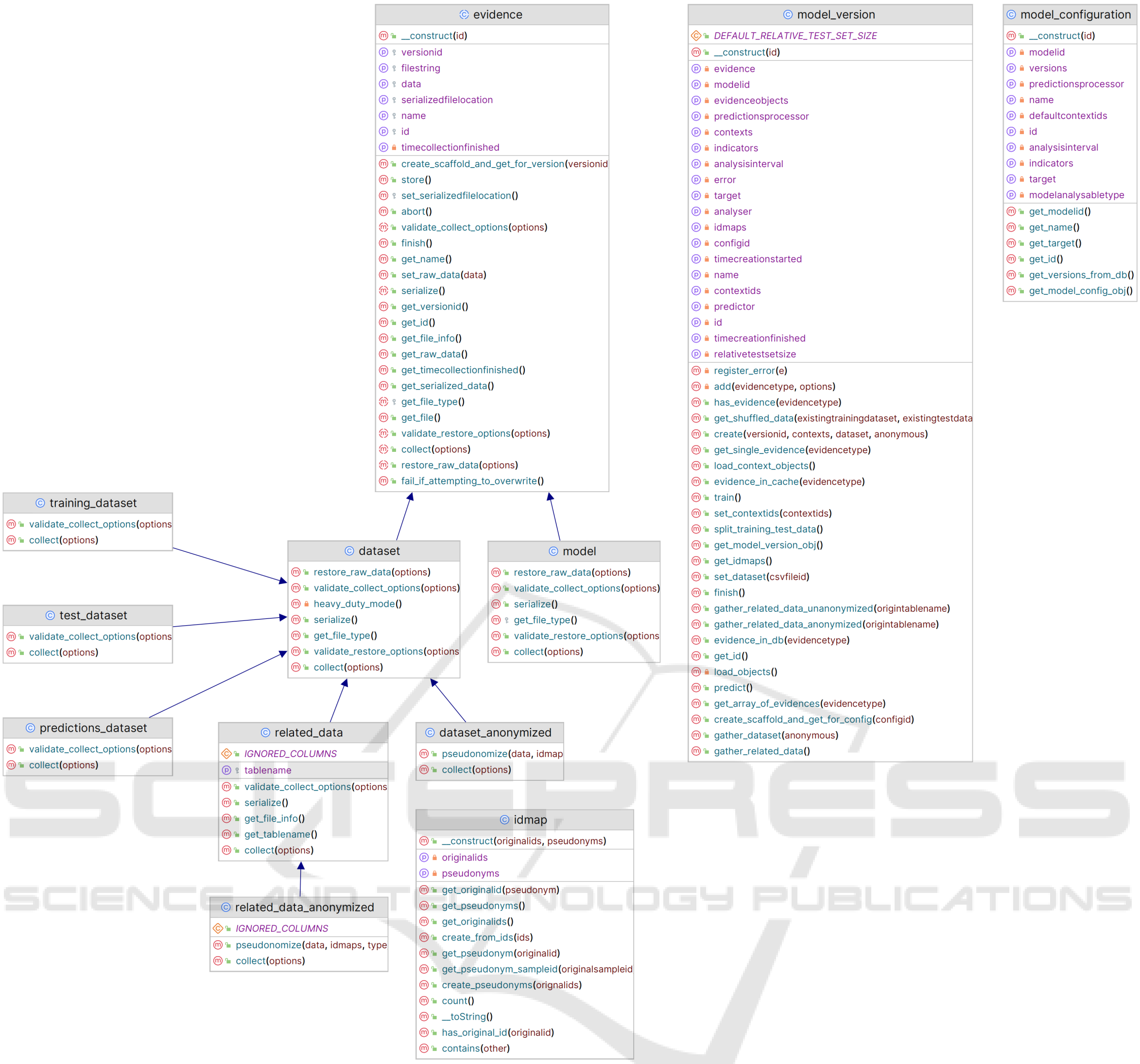

Figure 1: Important LALA classes and their hierarchy, exported from PHPStorm and edited.

4.4 Data Selection

The plugin adds an interface to import datasets into Moodle. Auditors can then, for example, upload and select balanced datasets to audit trained models. This gives substantially more control over the process and the targets of an audit, especially for assessing a model's fairness. The persisted trained models can be, for example, audited with different data-sets or different trained models can be benchmarked with the same test-data.

LALA solves one part of the test-data problem: Moodle's shortcoming in its original state is that only datasets from past courses are available as test-data. Now, for example, external standardized test-datasets extend the capabilities to audit Moodle's Learning Analytics. However, the other part of the problem – that is generating standardized and balanced test-datasets – lies outside of Moodle and the plugin's scope. We will come back to this issue in the discussion for further work.

4.5 Raw Predictions

LALA makes the raw predictions of the models available and generates outputs in the CSV format. The raw data allows more in-depth information than single metrics like an F1 score and its standard deviations. Auditors can then run suitable individual statistical transformations, analysis and tests on the raw data. The audit gains more detailed knowledge of the models. Additionally, fairness assessments and the detec-

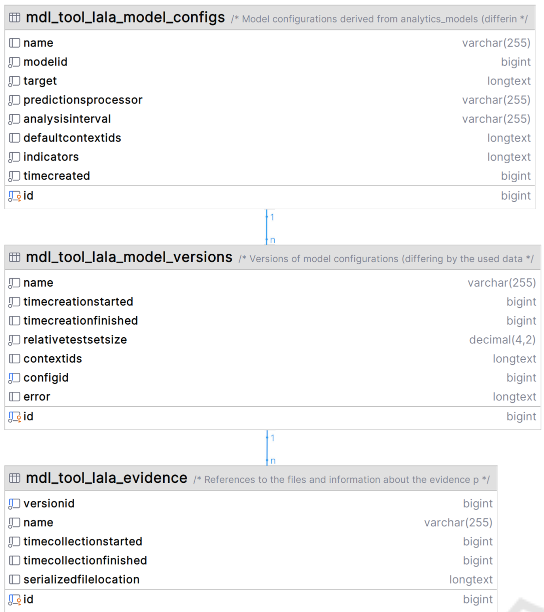

Figure 2: LALA data tables diagram, exported from PHP-Storm and edited.

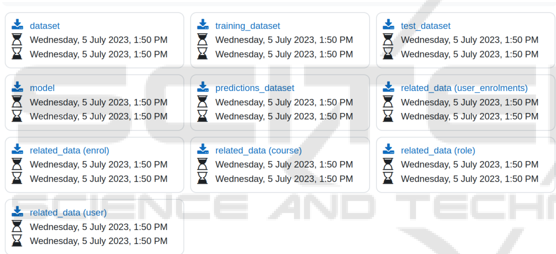

Figure 3: Evidence provided by LALA for download.

tion of discriminatory biases require more nuanced metrics alongside metrics for accuracy. The plugin contributes to an easy availability of raw predictions to derive additional suitable metrics that allow more substantial statements on Moodle's Learning Analytics fairness and ethical value adherence.

5 AUDITING WITH LALA AND FURTHER CONCERNS

The plugin facilitates a straightforward auditing process. After installing the plugin in a running Moodle instance, new options become available. Auditors assume a dedicated role in the Moodle environment and see a page of all current and older versions of the model configurations in the Learning Analytics component. They can then automatically or manually create a new model from a configuration. The manual mode allows to define which data should be gathered for the model or alternatively to upload own data. Once created, the model version information is displayed along with the evidence produced during model training and testing. Once the plugin has completed the data collection, auditors can download the evidence. This includes predictions and related raw data, which allow extended statistical testing for model fairness and biases. This evidence can then be used to validate or refute claims about fairness and trustworthiness.

Especially privacy and security concerns were considered during the plugin development. Security-wise, necessary features are bound to the dedicated auditor-role and do not intervene with permissions guidelines and rule-sets for existing user roles. This separates sensitive tasks from daily use. The role of auditor is assigned by the instance's administrator group. The auditor role is permitted to display specific pages, to download evidence, and to create new model versions. To achieve anonymization, datasets and related data are pseudonymized and only used if user-related data contains at least three distinct `IDs` (e.g. there need to be three distinct course enrolments) and concerns at least three different users. Otherwise, the evidence collection is aborted and the collected and pseudonymized data is deleted to comply with GDPR (European Parliament, 2016).

Although LALA offers a significant step forward, some challenges remain. Audits need more openly available test-data. The import feature allows auditors to use designated test-data, yet, specific datasets with sufficient anonymity and respectable group sizes to test potential discriminatory effects are not readily available. The problem is aggravated because test-data is quite complex and not mockable. Currently, Moodle only provides complete non-anonymized course backups. Privacy-compliant data-sets require to omit specific information that is crucial for model training when exporting. Therefore, the ability to export pseudonymized databases and automatically remove only privacy-relevant data fields would benefit audits, simplify the use of the plugin, and minimize privacy risks. The ongoing development process is concerned with reducing storage and processing requirements of the plugin. Further, multiple machine learning implementations and direct predictions with trained models need to be implemented to add to the feasibility and usability for audits to validate Moodle's Learning Analytics claims even further.

6 CONCLUSION

Machine learning models have often been found to be unfair, for example, when they discriminate against certain groups or are error-prone in their predictions or classifications (Rzepka et al., 2022). To mitigate unfairness and biases in Moodle's Learning Analytics and to guarantee the trustworthiness and acceptance of Learning Analytics models, it is crucial to audit them before deployment and continuously during their utilization. However, an audit of Moodle's Learning Analytics currently faces challenges stemming from a lack of auditability (Fernsel et al., tbd), which means, Moodle is not sufficiently accessible for audits that test claims by collecting and assessing evidence to validate its propagated features (Williams et al., 2022; Wolnizer, 2006). Specifically, Moodle does not store and make available evidence that is necessary to prove or refute fairness claims, Moodle only allows to inspect model configurations instead of trained models, and Moodle only outputs insufficient metrics to assess fairness (RQ1).

To address this lack of evidence, the developed plugin enables auditors to train and test Learning Analytics model configurations while also storing the intermediate results and providing these datasets as downloads. The stored raw predictions can be used for more in-depth inferential statistics to assess the overall fairness of the underlying models. Therefore, the plugin LALA extends Moodle's auditability (RQ2).

By enabling fairer Learning Analytics models and increasing trust in their predictions, we hope to reach more learners and to maximize the potential benefits of these models. The Moodle case study shows that auditability is not a given for open-source applications. Open source applications must also be designed with auditability in mind (Zook et al., 2017). Nevertheless, the Moodle example in particular shows that possible solutions can be retrofitted for open-source software to meet the requirements of scientifically sound audits that validate the claims made by the system through evidence.

ACKNOWLEDGEMENTS

We would like to thank the constructive remarks of several reviewer that helped to refine and improve our argument.

REFERENCES

Alam, A. (2023). Harnessing the Power of AI to Create Intelligent Tutoring Systems for Enhanced Classroom Experience and Improved Learning Outcomes. In Rajakumar, G., Du, K.-L., and Rocha, Á., editors, *Intelligent Communication Technologies and Virtual Mobile Networks*, pages 571–591. Springer Nature, Singapore.

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., Purves, D., and Gilbert, J. E. (2022). A review of predictive policing from the perspective of fairness. *Artificial Intelligence and Law*, 30(1):1–17.

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

Bose, R. P. J. C., Singi, K., Kaulgud, V., Phokela, K. K., and Podder, S. (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.

Castelnovo, A., Crupi, R., Greco, G., Regoli, D., Penco, I. G., and Cosentini, A. C. (2022). A clarification of the nuances in the fairness metrics landscape. *Scientific Reports*, 12(11):4209.

European Commission (2021). Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending vertain Union Legislative Acts.

European Parliament (2016). General Data Protection Regulation: GDPR.

Fernsel, L. (2024). Let's audit Learning Analytics, version 4.0.1, released under GNU GPL 3, https://github.com /LiFaytheGoblin/moodle-tool_lala/.

Fernsel, L., Kalff, Y., and Simbeck, K. (tbd). Assessing the Auditability of AI-integrating Systems: A Framework and Learning Analytics Case Study.

Hu, Q. and Rangwala, H. (2020). Towards Fair Educational Data Mining: A Case Study on Detecting At-risk Students. In Rafferty, A. N., Whitehill, J., Cavalli-Sforza, V., and Romero, C., editors, *13th International Educational Data Mining Society, Paper presented at the International Conference on Educational Data Mining (EDM)*, page 7, Online. ERIC.

Hutchinson, B., Smart, A., Hanna, A., Denton, E., Greer, C., Kjartansson, O., Barnes, P., and Mitchell, M. (2021). Towards Accountability for Machine Learning Datasets. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 560–575, Online. ACM.

Jeni, L. A., Cohn, J. F., and de La Torre, F. (2013). Facing Imbalanced Data Recommendations for the Use of Performance Metrics. *International Conference on Affective*

Computing and Intelligent Interaction and workshops : [proceedings]. ACII (Conference), 2013:245–251.

Kaddoura, S., Popescu, D. E., and Hemanth, J. D. (2022). A systematic review on machine learning models for online learning and examination systems. *PeerJ. Computer science*, 8:e986.

Landers, R. N. and Behrend, T. S. (2022). Auditing the AI auditors: A framework for evaluating fairness and bias in high stakes AI predictive models. *American Psychologist*, 78(1):36.

Mihaljević, H., Müller, I., Dill, K., Yollu-Tok, A., and von Grafenstein, M. (2023). More or less discrimination? Practical feasibility of fairness auditing of technologies for personnel selection. *AI & Society*, pages 1–17.

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., and Wiese, D. (2020). A supervised learning framework: using assessment to identify students at risk of dropping out of a MOOC. *Journal of Computing in Higher Education*, 32(1):9–26.

Ouhaichi, H., Spikol, D., and Vogel, B. (2023). Research trends in multimodal learning analytics: A systematic mapping study. *Computers and Education: Artificial Intelligence*, 4:100136.

Raghavan, M., Barocas, S., Kleinberg, J., and Levy, K. (2020). Mitigating bias in algorithmic hiring: evaluating claims and practices. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* '20, pages 469–481. ACM.

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., and Barnes, P. (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.

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.

Simbeck, K. (2023). They shall be fair, transparent, and robust: auditing learning analytics systems. *AI and Ethics*.

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.

Suresh, H. and Guttag, J. (2021). A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. In *Equity and Access in Algorithms, Mechanisms, and Optimization*, ACM Digital Library, pages 1–9, New York,NY,United States. Association for Computing Machinery.

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.

Verma, S. and Rubin, J. (2018). Fairness Definitions Explained. In *2018 IEEE/ACM International Workshop on Software Fairness (FairWare)*, pages 1–7, Gothenburg, Sweden. IEEE.

Weigand, H., Johannesson, P., Andersson, B., and Bergholtz, M. (2013). Conceptualizing Auditability. In Deneckère, R. and Proper, H. A., editors, *Proceedings of the CAiSE'13 Forum at the 25th International Conference on Advanced Information Systems Engineering (CAiSE)*, page 8, Valencia, Spain. CEUR.

Williams, R., Cloete, R., Cobbe, J., Cottrill, C., Edwards, P., Markovic, M., Naja, I., Ryan, F., Singh, J., and Pang, W. (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, Sydney.

Zook, M., Barocas, S., Boyd, D., Crawford, K., Keller, E., Gangadharan, S. P., Goodman, A., Hollander, R., Koenig, B. A., Metcalf, J., Narayanan, A., Nelson, A., and Pasquale, F. (2017). Ten simple rules for responsible big data research. *PLoS computational biology*, 13(3):e1005399.