Layered Explainability for Advisor Support: A Visual-Conversational Interface for Predictive Learning Analytics

Grzegorz Meller[1,2,3,4], Cédric Kestens[1] and Tinne De Laet[3,5,6]

[1]KU Leuven, Department of Computer Science, B-3001 Leuven, Belgium
[2]Augment, an imec research group at KU Leuven
[3]Leuven Engineering and Science Education Center (LESEC), KU Leuven
[4]Research Centre for Information Systems Engineering (LIRIS), KU Leuven
[5]Leuven Educational Research Institute (LIVO), KU Leuven
[6]Leuven.AI, KU Leuven Institute for Artificial Intelligence

Abstract

Predictive learning analytics dashboards can identify at-risk students, but their reliance on complex machine learning models often makes them difficult for study advisors to interpret. This paper presents a layered explainability approach that enhances transparency through three complementary levels: feature-importance explanations, directive data-centric explanations, and a conversational interface powered by a Large Language Model (LLM). A preliminary qualitative study with study advisors indicates that while feature-importance explanations improve transparency, directive data-centric explanations and conversational interface powered by an LLM are particularly effective in helping advisors interpret predictions and prepare student interventions. Findings suggest that explainability in learning analytics should prioritize student-advisor dialogue rather than solely exposing model internals.

Keywords

Explainable Artificial Intelligence, Learning Analytics, Human-Computer Interaction

1. Introduction

Predictive learning analytics dashboards (LADs) are increasingly adopted in higher education to support the early identification of at-risk students [1]. Although machine learning (ML) models can provide accurate predictions [2], their internal logic and sense-making of the features remain opaque to study advisors. The lack of understanding and trust hinders the adoption of predictive ML models in advising practice [3]. A key challenge is that ML models rely on data representations that can be difficult for end users to understand, particularly when trained on engineered behavioral features or, even more opaquely, raw log data. Explainable AI (XAI) offers techniques to improve transparency, but the effectiveness of these techniques ultimately depends on how explanations are communicated to the people who use them [4]. In advisor-facing settings, explanations must balance detail with cognitive load, connect abstract features to domain concepts, and support conversations with students [5].

This paper introduces **layered explainability** as a human-centered approach for LADs, employing three levels of progressive explanation that gradually move from model-centric transparency to advisor-centric sensemaking. The approach combines (1) feature-importance explanations that communicate how the model arrives at a specific prediction, (2) directive data-centric explanations that ground abstract features in observable student behavior, and (3) a conversational interface powered by a Large Language Model (LLM) that supports clarification, reflection, and dialogue preparation.

The goal of this study is to investigate how layered explainability can help advisors interpret complex predictive advising models and support the preparation of advisory conversations with students. Methodologically, this work follows a Design Science Research approach, in which we designed an

XAI-Ed 2026: Demystifying AI in Education and Learning Analytics through Explainability, Agency, and Transparency Workshop (XAI-Ed@LAK26), 27 April, 2026, Bergen, Norway
*Corresponding author.
grzegorz.meller@kuleuven.be (G. Meller)
0000-0003-3626-6962 (G. Meller)

advisor-facing LAD that implements layered explainability on top of an existing student success prediction model. The student success prediction model applies ML trained on learning analytics data to estimate whether a student is likely to pass or fail a given course. The prototype was evaluated through an initial qualitative user study with study advisors, employing think-aloud sessions and semi-structured interviews.

2. Layered Explainability Approach for predictive advising model

To support study advisors in interpreting the student success prediction model outputs, we conceptualize explainability as a process consisting of three consecutive layers. Each layer builds on the previous one, progressively moving from local model reasoning toward richer contextualization and interactive sense-making. Our conceptualization builds on two complementary strands of prior work. From the LADs perspective, it is grounded in the **Learning Analytics Process Model** by Verbert et al. [6], which conceptualizes how learning analytics progress from awareness (layer 1), through reflection by contextualized interpretation (layer 2), and sense-making (layer 3), towards impact. Although this model does not explicitly address the explainability of ML models, it provides an established theoretical lens for understanding how analytical outputs are interpreted and acted upon by its users. From a technological perspective, our work is inspired by the **visual–conversational interface** introduced by Samimi et al. [7]. While this work is situated in clinical risk assessment, it offers a pattern that could be transferred to the LAD context. Their findings confirmed that the combination of feature-importance explanations (layer 1), with directive data-centric explanations (layer 2), and conversational interactions (layer 3) helped professionals build a clearer understanding of model assessments. Although Samimi et al. [7] do not explicitly conceptualize their interface as a layered explainability, it integrates multiple explanation components that can be mapped onto our approach.

In our setup, the first layer focuses on communicating how the ML model reached a specific, local prediction. This is commonly achieved using model-agnostic techniques like Local Interpretable Model-agnostic Explanations (LIME) [8], which provide **feature-importance explanations** for individual predictions. In the LADs domain, Scheers and De Laet [5] demonstrated the value of interactive feature-importance explanations for study advisors.

The second layer, to provide contextualization of the features, moves from the feature importances to the actual learning behaviour of students underlying the model feature. To this end, **directive data-centric explanations** [9] provide insights based on the statistical properties of the log data underlying the model feature. For instance, a feature indicating *"low online activity"* as an important factor in an ML prediction can be grounded by a chart presenting the student's activity over the course of the semester.

The third and final layer addresses in-depth, complex, and unscripted user questions through a **conversational interface powered by an LLM**. The utility of LLMs in delivering contextualized explanations is rising across high-stakes fields. In clinical decision-making, the RetCare workflow introduced by Wang et al. [10] integrates LLMs with feature-importance scores computed by LIME, and provides easy-to-understand, trust-enhancing explanations. Within LADs, Susnjak [11] demonstrated the LLM's capability to generate personalized and conversational prescriptive feedback for students based on their feature-importance values. This research indicates that LLMs can serve as a dynamic and contextual layer for interpreting feature-importance explanations.

While these above mentioned components are individually recognized, their synergistic combination in a single, advisor-facing LAD remains a research gap.

3. LAD Design

The LAD design, illustrated in Figure 1, represents an operational instantiation of the layered explainability approach. The main purpose of the dashboard is to support study advisors at KU Leuven in advising students in the Bachelor's programs in the fields of business and economics, within an open-admission system characterised by high dropout rates and heterogeneous student backgrounds. The system is explicitly designed to support advisor–student dialogue, not to automate decisions or to be used independently by students. The dashboard is grounded in data from the ALPACAS (Adaptive Learning Paths for ACtivation and Assessment of Students) project at the Faculty of Economics and Business at KU Leuven. The dataset consists of learning management system (LMS) interaction logs, capturing students' behavioral traces such as logins, clicks, and submissions across courses, together with academic outcomes [12]. Building on this data, prior work by Tiukhova et al. [2] performs feature engineering to transform low-level interaction events into higher-level behavioral indicators, such as measures of consistency, proportions, and temporal patterns reflecting aspects of self-regulated learning behavior, which are then used to train and evaluate several ML models. In this study, we adopt the Support Vector Machine (SVM) model from this prior work.

The dashboard is based on a three-column layout, where the left column presents the local model prediction for an individual student, including the predicted risk category and associated probability. This is accompanied by a LIME-based **feature-importance explanation (layer 1**) that visualizes how specific features contributed positively or negatively to the prediction. Recognizing the interpretability challenges posed by highly engineered features, individual feature attributions are grouped into higher-level behavioral constructs such as Consistency, Effort, and Engagement. These groupings function as a semantic translation layer, aiming to reduce cognitive load and align model explanations with concepts familiar to advisors. The presentation of explanations follows a progressive disclosure [13] where higher-level behavioral constructs are shown by default, while more detailed information, such as underlying features, can be revealed on demand through unfolding.

The central column of the dashboard provides **directive data-centric explanations (layer 2)** that ground abstract feature-importance scores in observable student behavior. This layer visualizes raw or lightly processed activity data that correspond to the features emphasized by the prediction model. Examples include time-series charts of weekly online activity, comparisons against benchmark profiles (e.g., a *"perfect student"*), and summary statistics related to assessment or forum engagement. These visualizations enable advisors to validate whether the model's assessment is consistent with the student's actual behavior and to identify concrete patterns that can serve as discussion points. By anchoring explanations in data that advisors can readily interpret, this layer aims at supporting sensemaking and reducing reliance on feature-importance explanations alone.

The right column of the LAD incorporates a **conversational interface powered by an LLM (layer 3)**. It serves two primary functions, namely clarifying dashboard elements and supporting the preparation of advisory conversations. Suggested questions, meant to overcome articulation barriers [14], are formulated to help advisors explore trends, request plain-language explanations of visualizations, or generate student-facing questions. Importantly, the LLM does not replace human judgment or provide prescriptive decisions. Instead, it acts as an interpretive and conversational aid, helping advisors articulate concerns, reflect on evidence, and structure dialogue with students. In our prototype, we simulated this functionality using Microsoft Copilot, without applying any domain-specific fine-tuning. Responses were generated through prompting, where the model was provided with relevant dashboard context, including data and feature descriptions, while excluding any identifying information.

4. Initial User Study

The study employed a qualitative approach utilizing think-aloud protocols and semi-structured interviews to gather in-depth user feedback. Participants interacted with a high-fidelity LAD prototype, presented as a clickable Figma [15] mock-up, verbalizing their reasoning. The session concluded with an interview focusing on the tool's clarity, usability, and relevance for academic advising. Ethical approval was obtained from the ethical committee of KU Leuven with the number G-2025-9450. Three study advisors participated in the study, where each took approximately 45 minutes. Each study was recorded and transcribed. Inductive thematic analysis was performed on the collected transcripts.

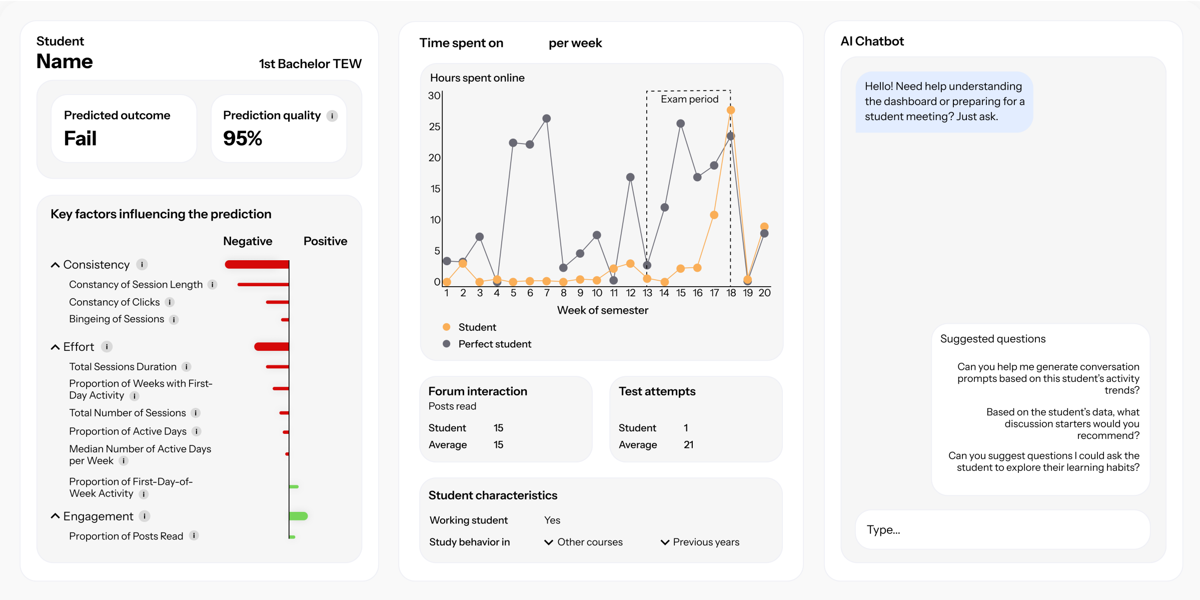

Figure 1: LAD design incorporating the layered explainability approach with a visual-conversational interface.

All participants interacted with the same student case, which was based on authentic data from the ALPACAS dataset. It was fully anonymized to prevent identification.

The **feature-importance explanation (layer 1)**, while appreciated for its attempt to provide transparency, introduced several concerns. Participants reported that the level of detail required considerable time to interpret and translate into actionable insights. As Participant 3 (P3) stated: *"It asks quite a lot from me to go through these different factors and to get an idea of what they mean. I have to read everything very well"*. Furthermore, a challenge was to understand several feature-importance variables. Participants struggled to grasp both what these features represent and why they were relevant for the ML model's prediction, hindering their ability to confidently use the information in a conversation with a student. P1 expressed the following concern: *"What does it mean 'Proportion of Weeks with First-Day Activity'? Why would this be interesting?"*.

The **directive data-centric explanations (layer 2)** with the graph *'Hours spent online'* was seen as intuitive and informative. Participants noted that the visual trend of weekly activity fluctuations over time was immediately effective in signaling changes in student engagement or disengagement. This visualization provided a clear, data-driven starting point for discussion without requiring extensive interpretation. As P2 stated: *"It really shows me what the student is doing or not doing. I can see, they did not do anything from Week 3 until Week 10, that would concern me as a student advisor"*. Regarding other directive data-centric explanations, participants expressed the need to bridge the gap between feature-importance scores and data-centric visualizations. Participants expressed confusion when data-centric graphs or tables did not directly correspond to the features highlighted in the left column of LAD. Participants advocated for an interactive design, specifically the ability to click on a feature-importance value and immediately access a corresponding data-centric visualization. P3 explained: *"I see the 'Test Attempts' in the middle column, but I don't see it in the left column. So then this variable is not taken into account to predict the student's outcome. As a study advisor, I find it valuable to see, but here is the question: Why is it not taken into account by the predictive model? [...] It might be nice that you could also click on these (feature-importance) factors and then, in the same way, get more details and visualizations about them"*.

The **conversational interface powered by an LLM (layer 3)** was highly appreciated for its dual use, namely clarifying visualizations and guiding the conversation with the student. Advisors valued its ability to offer clarifications for dashboard components and suggest useful questions, effectively serving as a conversation guide prior to and during an advisory conversation. P1 stated: *"It is really helping me* *to interpret what I see on the screen. I think it is really nice that you have some information about how to interpret the data"*, and P2: *"I like the suggested questions to get the conversation started with the student [...] I am likely to use it as a starting point that helps me guide the conversation"*. It was also strongly emphasized that the system should be used only to guide the advisory conversation, not replace it. Participants stressed that digital activity logs do not present the full story of student engagement, as a significant portion of study-related activities still occur offline. The system was thus claimed to be a good solution for initiating and structuring the conversation, ensuring the human advisor maintains a holistic view of the student's context.

5. Discussion

This paper explored how a layered explainability approach can support study advisors in interpreting the student success prediction model. By integrating feature-importance explanations, directive data-centric explanations, and a conversational interface powered by an LLM, the proposed design moves beyond isolated XAI techniques toward a more holistic, visual-conversational interface [7]. Findings from the initial qualitative study suggest that while feature-importance explanations are valued for transparency, they are often cognitively demanding and difficult to translate into advisory action on their own. It is in line with findings of Sîrbu et al. [16] who showed that detailed explanations can cause extraneous cognitive load, having negative effects both on trust and task performance. It is also in line with findings of Scheers and De Laet [5] who observed that adding feature-importance explanations in LAD can increase the information load.

In contrast, directive data-centric explanations, particularly the *'Hours spent online'* chart, were perceived as understandable and actionable, serving as an effective bridge between model output and real-world advising practice. It is in line with the findings of Bhattacharya et al. [9], who demonstrated that directive data-centric explanations enhance users' ability to interpret a dashboard, and derive actionable insights for informed guidance. An important observation concerns the role of these data-centric explanations in relation to feature-importance explanations. While participants initially encountered difficulties in interpreting abstract feature representations, the integration of data-centric visualizations appeared to support their understanding and increase participants' confidence in interpreting the system's outputs. This finding is consistent with prior work that combines feature-importance explanations with more concrete, directive representations. For example, Samimi et al. [7] report that visual representations of data-centric explanations enable quicker comprehension of feature-importance information. More broadly, this aligns with research emphasizing the importance of presenting supporting evidence alongside AI predictions, where evidence-based explanations can strengthen users' understanding and foster more informed engagement with AI systems [17]. In this sense, integrating multiple forms of explanation within a layered design shows promise for supporting appropriate user trust in AI-assisted decision-making contexts. These findings further suggest that extending the set of data-centric explanations to more systematically cover the feature-importance outputs by providing corresponding behavioral representations for individual features may enhance the overall interpretability and usefulness of the dashboard.

The conversational interface further enhanced this process, but in a manner that differs from prior findings. While Samimi et al. [7] report that conversational interfaces are primarily used to support the interpretation of existing explanations, our findings suggest that they can also play an equally important role in supporting advisors in preparing for advisory sessions with students. This indicates that conversational explainability may extend beyond interpretability toward supporting the practical use of AI outputs in advisory work, particularly in contexts where insights must be mediated through advisors which suggest that conversational interfaces can contribute to integrating AI insights into real-world advisory practices.

The results highlight that explainability in LAD should not be understood solely as exposing model internals, but as supporting sensemaking and dialogue in socio-technical decision contexts. Advisors emphasized the importance of using the dashboard as a conversation guide rather than a decision-making authority, reinforcing prior concerns about over-reliance on algorithmic predictions. This is both in line with findings of Scheers and De Laet [5], De Laet [3] who reported that explanations should provide a trigger for further conversation with the student, but also with the observations of Schemmer et al. [18] who showed that explanations can influence automation bias.

6. Limitations, and Future Work

Several limitations must be acknowledged. First, the study involved a very small number of participants and focused on a single context, which limits the generalizability. Second, the evaluation was limited to advisors' perceived experiences with the dashboard, rather than on observable changes in advisory decisions or student outcomes. Finally, the conversational interface powered by an LLM was not systematically evaluated for explanation fidelity, bias, or hallucination risks.

Future work will address these limitations. First, user studies will be expanded to involve a larger group of advisors, and will measure additional factors, including trust calibration [19], enabling more robust assessments. Second, design improvements will be implemented based on the current feedback from advisors. Finally, systematic evaluation of LLM-generated explanations will be integrated.

Acknowledgments

We thank Elena Tiukhova for her guidance on the dataset and the use of predictive models. We also extend our gratitude to the participants of our user study. This research was supported by KU Leuven's Special Research Fund (BOF) grant number IDN/24/004.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT for spelling checks. The content was subsequently reviewed and edited by the authors, who take full responsibility for the publication.

References

[1] G. Ramaswami, T. Susnjak, A. Mathrani, R. Umer, Use of predictive analytics within learning analytics dashboards: A review of case studies, Technology, Knowledge and Learning 28 (2023) 959–980. URL: https://doi.org/10.1007/s10758-022-09613-x. doi:10.1007/s10758-022-09613-x.

[2] E. Tiukhova, P. Vemuri, N. L. Flores, A. S. Islind, M. Óskarsdóttir, S. Poelmans, B. Baesens, M. Snoeck, Explainable learning analytics: Assessing the stability of student success prediction models by means of explainable ai, Decision Support Systems 182 (2024) 114229. URL: https://www.sciencedirect.com/science/article/pii/S0167923624000629. doi:https://doi.org/10.1016/j.dss.2024.114229.

[3] T. De Laet, Learning Dashboards for Academic Advising in Practice, Springer International Publishing, Cham, 2023, pp. 55–75. URL: https://doi.org/10.1007/978-3-031-27646-0_4. doi:10.1007/978-3-031-27646-0_4.

[4] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, F. Herrera, Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai, Information Fusion 58 (2020) 82–115. URL: https://www.sciencedirect.com/science/article/pii/S1566253519308103. doi:https://doi.org/10.1016/j.inffus.2019.12.012.

[5] H. Scheers, T. De Laet, Interactive and explainable advising dashboard opens the black box of student success prediction, in: Technology-Enhanced Learning for a Free, Safe, and Sustainable World: 16th European Conference on Technology Enhanced Learning, EC-TEL 2021, Bolzano,

Italy, September 20-24, 2021, Proceedings, Springer-Verlag, Berlin, Heidelberg, 2021, p. 52–66. URL: https://doi.org/10.1007/978-3-030-86436-1_5. doi:10.1007/978-3-030-86436-1_5.

[6] K. Verbert, E. Duval, J. Klerkx, S. Govaerts, J. L. Santos, Learning analytics dashboard applications, American Behavioral Scientist 57 (2013) 1500–1509. URL: https://doi.org/10.1177/0002764213479363. doi:10.1177/0002764213479363. arXiv:https://doi.org/10.1177/0002764213479363.

[7] R. Samimi, A. Bhattacharya, L. Gosak, G. Stiglic, K. Verbert, Visual-conversational interface for evidence-based explanation of diabetes risk prediction, in: Proceedings of the 7th ACM Conference on Conversational User Interfaces, CUI '25, Association for Computing Machinery, New York, NY, USA, 2025. URL: https://doi.org/10.1145/3719160.3736616. doi:10.1145/3719160.3736616.

[8] M. T. Ribeiro, S. Singh, C. Guestrin, "why should i trust you?": Explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, Association for Computing Machinery, New York, NY, USA, 2016, p. 1135–1144. URL: https://doi.org/10.1145/2939672.2939778.

[9] A. Bhattacharya, J. Ooge, G. Stiglic, K. Verbert, Directive explanations for monitoring the risk of diabetes onset: Introducing directive data-centric explanations and combinations to support what-if explorations, in: Proceedings of the 28th International Conference on Intelligent User Interfaces, IUI '23, Association for Computing Machinery, New York, NY, USA, 2023, p. 204–219. URL: https://doi.org/10.1145/3581641.3584075. doi:10.1145/3581641.3584075.

[10] Z. Wang, Y. Zhu, J. Gao, X. Zheng, Y. Zeng, W. Tang, Y. Wang, E. M. Harrison, C. Pan, L. Ma, L. Wang, Retcare: Towards interpretable clinical decision making through LLM-driven medical knowledge retrieving, in: Artificial Intelligence and Data Science for Healthcare: Bridging Data-Centric AI and People-Centric Healthcare, 2024. URL: https://openreview.net/forum?id=jqo1vk63qE.

[11] T. Susnjak, Beyond predictive learning analytics modelling and onto explainable artificial intelligence with prescriptive analytics and chatgpt, International Journal of Artificial Intelligence in Education 34 (2024) 452–482. URL: https://doi.org/10.1007/s40593-023-00336-3. doi:10.1007/s40593-023-00336-3.

[12] P. Vemuri, M. Snoeck, S. Poelmans, Adaptive release learning paths to motivate active learning and engagement in students, in: Proceedings of the 18th International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2021), ERIC, 2021, pp. 287–290.

[13] D. Lei, Y. He, J. Zeng, What is the focus of xai in ui design? prioritizing ui design principles for enhancing xai user experience, in: H. Degen, S. Ntoa (Eds.), Artificial Intelligence in HCI, Springer Nature Switzerland, Cham, 2024, pp. 219–237.

[14] J. Song, Z. Ashktorab, Q. Pan, C. Dugan, W. Geyer, T. W. Malone, Interaction configurations and prompt guidance in conversational ai for question answering in human-ai teams, Proc. ACM Hum.-Comput. Interact. 9 (2025). URL: https://doi.org/10.1145/3757486. doi:10.1145/3757486.

[15] Figma, Figma: The collaborative interface design tool, 2026. URL: https://www.figma.com/, accessed: 2026-04-15.

[16] A.-M. Sîrbu, T. C. Schelhorn, U. Gnewuch, Explanation provision strategies in llm-based data assistants: Impact on extraneous cognitive load, trust, and task performance, in: ECIS 2025 Proceedings, 2025, p. 6. URL: https://aisel.aisnet.org/ecis2025/hci/hci/6.

[17] A. Bhattacharya, J. Ooge, G. Stiglic, K. Verbert, Directive explanations for monitoring the risk of diabetes onset: Introducing directive data-centric explanations and combinations to support what-if explorations, in: Proceedings of the 28th International Conference on Intelligent User Interfaces, IUI '23, Association for Computing Machinery, New York, NY, USA, 2023, p. 204–219. URL: https://doi.org/10.1145/3581641.3584075. doi:10.1145/3581641.3584075.

[18] M. Schemmer, N. Kühl, C. Benz, G. Satzger, On the influence of explainable ai on automation bias, 2022. URL: https://arxiv.org/abs/2204.08859. arXiv:2204.08859.

[19] M. Wischnewski, N. Krämer, E. Müller, Measuring and understanding trust calibrations for automated systems: A survey of the state-of-the-art and future directions, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI '23, Association for Computing Machinery, New York, NY, USA, 2023. URL: https://doi.org/10.1145/3544548.3581197.