# Explainable Learning Analytics: challenges and opportunities of this emerging research line


Tinne De Laet
tinne.delaet@kuleuven.be
KU Leuven, LESEC
Leuven, Belgium

Tom Broos
tom.broo@kuleuven.be
KU Leuven, LESEC, AugmentHCI
Leuven, Belgium

Raphael Duorado
rasd2@cin.ufpe.br
Universidade Federal de Pernambuco
Brasil

Robin De Croon
Martijn Millecamp
Katrien Verbert
first.lastname@kuleuven.be
KU Leuven, AugmentHCI
Leuven, Belgium



**Figure 1: Example of Workshop outcome: opportunities and challenges of explainable learning analytics regarding the stakeholder dimension. The stickers are the result of an up- and downvoting procedure (green = upvoted challenge, gold = upvoted opportunity, yellow = veto).**


## ABSTRACT

In the last decade, we are witnessing a widespread adoption of artificial intelligence in a wide range of application domains. Learning analytics is no exception. Artificial Intelligence (AI) techniques and Machine Learning (ML) in particular are used to generate automatic predictions and recommendations regarding learning and teaching. A key challenge in the actual use and adoption of AI and ML is that they often operate as a 'black box', hereby impeding understanding and trust. The domain of Explainable Artificial Intelligence aims




*LAK20, March, 23–27,2020, Frankfurt, DE*



ACM ISBN XXX-X-XXXX-XXXX-X/xx/xx…$XX.00
https://doi.org/10.1145/1122445.1122456

at enhancing the transparency of AI techniques and therefore also holds substantial promise for the Learning Analytics domain.

This paper supports the shaping of the research line of Explainable Learning Analytics (XLA), by exploring the challenges and opportunities related to the data, stakeholders, communication, evaluation, and implementation & adoption of XLA. In particular, this paper reports on the outcomes of a 3-hour workshop with 44 international participants in which these challenges and opportunities were collaboratively defined. The obtained challenges and opportunities will form the basis for a deeper exploration, involving a wide range of stakeholders, of the promises of the XLA-field and the required points of focus for the next 10 years.

## CCS CONCEPTS

**• Security and privacy → Human and societal aspects of security and privacy**; **• Human-centered computing → Visual analytics**; *Human computer interaction (HCI)*; **• Computing methodologies → Artificial intelligence**; **• Applied computing → Education**.

## KEYWORDS

learning analytics, explainable learning analytics, explainable artificial intelligence, recommender systems, visual analytics

**ACM Reference Format:**

Tinne De Laet, Tom Broos, Raphael Duorado, Robin De Croon, Martijn Millecamp, and Katrien Verbert. 2018. Explainable Learning Analytics: challenges and opportunities of this emerging research line. In *LAK20: 10th International Learning Analytics and Knowledge Conference, March, 23–27,2020, Frankfurt, DE.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/1122445.1122456

## 1 INTRODUCTION

The domain of **Learning Analytics (LA)** has been established, at the background of the never-ending need for support of learners and teachers, under impulse of the growth of learning data, the development of algorithms and AI, and the research within learning science. LA is about 'collecting traces that learners leave behind and using those traces to improve learning' (Eric Duval). Obtaining actual improvement in learning and teaching does not come easy however. Verbert et al. [11] showed such improvements, which are related to 'impact' (level 4), can only be obtained after awareness of the data (level 1), self-reflection (level 2), and sense-making (level 3) have been obtained. Machine learning (ML) and Artificial Intelligence (AI) in general have provided plenty of algorithms that can support the translation from data to awareness, self-reflection, and sense-making. Visualization techniques in general and Learning Analytics Dashboards (LAD) in particular [12] have been shown to be able to provide a visual means of communication of the data and the outcomes of AI algorithms to stakeholders. The LA domain has also been shown to be an interesting application domain for AI and machine learning. At the same time, the domain of LA is maturing fast and is challenging the application of AI and ML techniques regarding the outputs generated. Algorithmic predictions and recommendations regarding learning and teaching have to be *interpretable* for and *explainable* to the involved stakeholders and have to be *translated to actionable recommendations.*

To be *interpretable and explainable*, the outcomes of the data analysis, visualization, and/or ML and AI algorithms have to be tailored to the particular stakeholders and end-users. While advanced visualization and/or ML techniques might create accurate and trustworthy insights and recommendations, they will not be trusted per se by the user. Opening the black-box of learning analytics to the user, in a user-tailored fashion is the first step towards obtaining interpretable insights and explainable recommendations. Including new approaches for obtaining transparency, trustworthiness, persuasiveness, and effectiveness is key.

A second challenge for LA, after interpretability and explainability, is to translate predictions and recommendations into feasible 'actions'. This is also referred to as the *actionability of the feedback.* To highlight the challenge of this actionability, let's consider the following example. Educational data mining techniques may discover that male students on average are more likely to fail in higher education. While such information can be interesting for researchers and policy makers, it does not provide a directly actionable recommendation towards an individual (male) student on how to improve his learning skills or study success. If actionable insights and recommendations can be created withing LA and if they can be tailored to the involved stakeholders, they will have the potential to create impact [11]. User-centered design involving the stakeholders and the actual integration of LA into actual educational practices and the pedagogy underlying these educational practices will support the actionability of the insights and recommendations. For example; if instructors cooperate in the integration of LA in their course design, the automatic recommendation of resources to students in the context of this class, informed by the course design, will improve.

The evolution towards actionable insights and explainable recommendations is urgent, as recent data protection and privacy regulations like the EU General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) stipulate that transparency is a fundamental right. Even more, they state that users have the right to withdraw themselves from automatic decision making and profiling. Human mediation in automatic decision making and profiling is a promising approach to accommodate ethical use, but also reinforces the need for transparency towards the human actor.

The domain of **explainable Artificial Intelligence (XAI)** has been developing and growing fast in the last years. AI is a part of a new generation of AI technologies called the third wave AI, including among other ambitious goals, the development of algorithms that can explain themselves [2]. The XAI research field aims at improving trust and transparency of AI-based systems, which can both concern automatic predictions and recommendations. AI algorithms often from the so-called 'black-box' phenomenon, indicating that it is hard for users, including domain experts, to get insights in the internal mechanisms underlying the algorithms and the outcomes produced by these algorithms. This is also referred to as algorithm opacity [2]. The problem of opacity has been growing together with the development of novel machine learning algorithms, such as deep learning and random forest, which itself was supported by the fastly growing computational power. Algorithm opacity can

however impede trust in predictions and recommendations provided by these ML algorithms and AI techniques, preventing their actual adoption and deployment in real-world scenarios.
The XAI field by itself is fastly maturing as shown in the XAI survey of Adadi and Berrada [2] and by even more recent contributions focusing on the trends within XAI [1], on the sub-domain of explainable recommendations [8, 14], on the evaluation of XAI [6], and visual interpretability [4, 9, 13, 15].
The survey of Adadi and Berrada [2] also recognizes the potential for XAI in different application domains: transportation, healthcare/medical [5, 10], legal, finance, and military. The number of application domains is still growing fast as shown by recent research dedicated to e.g. robotic agents [3]. While attention for XAI is also growing within the domain of LA, it still remains to be determined what the main research directions should be and to what level general XAI findings can be applied to the LA domain, and to what level specific developments have to be made.

The goal of this paper is to contribute to the creation and shaping of the exciting and promising domain of *Explainable Learning Analytics (XLA)*, focusing on the application domain-specific developments of XAI within LA. In particular, this paper aims at contributing to the discovery of the main opportunities and challenges of XLA. To this end, this paper reports on a workshop involving more than 40 international stakeholders to identify the main challenges and opportunities of XLA.

## 2 METHODOLOGY

Stakeholder input regarding the challenges and opportunities of XLA was collected during a 3-hour workshop at the 2019 European Conference on Technology Enhanced Learning (EC-TEL 2019), in Delft, the Netherlands. Beforehand, we identified five themes of focus for the workshop: data, stakeholders, communication, evaluation, and implementation & adoption. Table 1 provides an overview of the themes and how they were presented during the workshop.

During the workshop the following protocol was used:

- **Welcome and ice-breaker activity** (10 minutes)
- **Introduction** regarding LA and explanaibilty and interpretability of predictions and recommendations (15 minutes),
- **Idea generation round** in small groups (60 minutes, Figure 2),
- **Synthesis round** where all input per theme is collected and prioritized using grouping of input and up and downvoting, (30 minutes, see the teaser image (Figure 1) in the beginning of the paper for an example),
- **Plenary discussion** to finalize the identified challenges and opportunities (65 minutes).

In the **introduction** a plenary presentation was provided with the general background of explainable AI, LA, and some examples of XLA. The presentation also focused on the concepts of predictions, recommendations, explainability and interpretability and why these are important. This introduction ensured that each participant had a basic understanding of the field of XAI and LA. Additionally, everyone was made aware of the protocol used in the workshop.

The **idea generation round** was organized such that pairs of participants discussed around each of the five themes. Tables were prepared such that at each table the five themes could be discussed

**Figure 2: Picture of the idea generation round during the workshop at the EC-TEL 2019 conference.**

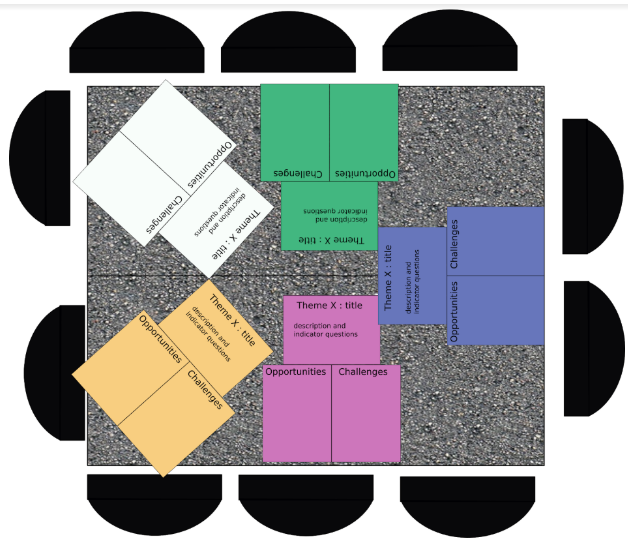


**Figure 3: Table lay-out for the idea generation round. At each table the five themes are discussed by pairs of participants.**

(as illustrated in Figure 3). Pairs of participants would discuss on the challenges and opportunities related to a particular theme during around twelve minutes using a push-through procedure. They added the output of their discussion using post-its to the discussion notes on the table. Each twelve minutes, the pairs progressed to the next theme at the same table and would add their findings to the already existing output. This round ended as soon as each pair of participants had addressed each theme once.

In the **synthesis round** the input from the different tables was grouped thematically, i.e. according to each of the five themes. First, the participants were asked to group the input (post-its) by sticking similar ideas together. Next, participants were invited to individual

**Table 1: The five different themes of focus of explainable learning analytics and the teaser questions that were provided to the participants.**

| Theme | Explanation verbally provided to participants |
|---|---|
| Data | What data can be useful to be explained?<br>What data about the user can be used to generate a prediction or a recommendation?<br>Is the data readily available? |
| Stakeholders | Who are the stakeholders of explainable learning analytics?<br>Why would they use learning analytics that needs explanations?<br>In which situation would they need explanations?<br>When would they see/use these explanations? |
| Communication | How would you communicate explanations to the user? (visual, text, mix, audio, ...).<br>How would you adapt the explanation? Based on personal/situational characteristics? Is it ethical to adapt interfaces?<br>Why would you trust or not trust a system? Do explanations help? |
| Evaluation | How would you evaluate explanations? Qualitative? Quantitative?<br>What end goal would you evaluate? |
| Implementation & adoption | What steps are needed to implement XLA?<br>Where do you see the domain of XLA in 10 years? What is needed to reach that? What can prevent XLA? What can stimulate XLA? For what purpose would you use XLA? |

dot-voting: each participant received eight stickers that he/she could use to highlight the most urgent or important challenges an opportunities: four stickers had to be used for challenges and four for opportunities. Additionally, each participant received two veto stickers: one for an opportunity and one for a challenge, to indicate their disagreement. Participants were requested to put their initials on the veto sticker such that they could be prompted for more explanation during the plenary discussion round.

The goal of the **plenary discussion** was to use the input from the discussion and synthesis round to define the most important challenges and opportunities of explainable learning analytics for each of the five themes. A list of most important challenges and opportunities was assembled based on the number of dot votes. Selected items for each of the themes were discussed one by one and participants were invited to bring forward the findings and elaborate on this. Other participants were invited to comment and discuss. This discussion was recorded using audio recording devices.

**Ethical approval** was obtained from the ethical commission of BLINDED. Workshop participants that subscribed to the workshop received an explanation of the research performed beforehand by email, including a notice that they would be invited to sign a consent form if they would participate in the plenary discussion. During the workshop, all participants were informed about the protocol and invited to sign the consent forms if they agreed their audio was recording during the plenary discussion. Participants not consenting could still participate in all parts of the workshop, except the final audio-recorded plenary discussion. At several stages during the workshop pictures were taken to support the processing of results, and participants were asked for their permission for this.

## 3 RESULTS

The workshop organized at the EC-TEL 2019 conference in Delft, the Netherlands was attended by 44 participants. All these participants participated in the idea generation round. After a short break, 17 participants consented with the recording of their voices during the final discussion, and therefore participated in both the idea generation round, the synthesis round, and the plenary discussion. Below, we elaborate on the main opportunities and challenges that were identified during the synthesis round and the plenary discussion (as illustrated in Table 2), grouped by each of the five themes.

### 3.1 Opportunity

*Data.* The main identified opportunity for data for XLA is to use data regarding *course/learning design* and *reusing teachers' previous data* (6 opportunity stickers, no challenge stickers, no veto). During the discussion, participants elaborated that augmenting systems with human-generated data might alleviate the explainability issue. After all, human-generated data could be useful to explain computer-generated data. The participants recognized the opportunity that well-modelled learning design can influence predictions and explanations. Feedback on these predictions and explanations can in turn improve the underlying models. An example of such improvements is elaborated on in the paper of Mothilal et al. [7], where the explanatory technique of LIME is used to obtain explanations of the prediction of first-year engineering student success, which improved the model of required starting competences. However, it is still an open research challenge which data can and should be collected to support explainability. Nonetheless, the participants agreed that the expectations should be clear beforehand and that annotations on the data, if used with care, can support the explainability. A final warning regarding the data of course/learning design

**Table 2: Most voted challenges and opportunities.**

| Theme | Item | #opportunity | #challenge | #veto |
|---|---|---|---|---|
| | *Opportunities* | | | |
| Data | Course/learning design, re-using teachers' previous data | 6 | 0 | 0 |
| Stakeholders | Teachers reflect/understand own effectiveness (by visualising features) | 4 | 1 | 0 |
| Evaluation | Evaluate impact of the explainations | 3 | 1 | 0 |
| Communication | Data storytelling | 7 | 0 | 0 |
| Implementation & adoption | Adapt to the target groups | 5 | 2 | 0 |
| | *Challenges* | | | |
| Data | Include recency of data | 0 | 4 | 0 |
| Stakeholders | Community building | 1 | 4 | 0 |
| Evaluation | Added value for user (pre/post)? | 1 | 3 | 0 |
| Communication | Time-based LA, splitting explanation per phase/step | 0 | 6 | 0 |
| Implementation & adoption | Support (technical, pedagological) when system is deployed | 1 | 3 | 0 |

is that it can and should be context-specific, complicating wider use.

*Stakeholders.* The main identified opportunity of XLA for stakeholders is for *teachers* when they can use XLA to *reflect upon or understand their own effectiveness.* (4 opportunity stickers, 1 challenge sticker, no veto).
XLA will definitely provide an opportunity for teachers as they can disclose understandable explanations and recommendations to this target audience regarding their teaching effectiveness. If stakeholders provide different perspectives, one should somehow prioritize or weight priorities; it is however which metrics should be used in combining these perspectives?

*Communication.* The main opportunity for communication of XLA is *data storytelling* (7 opportunity stickers, 0 challenge stickers, no veto)
Recent research and technological advancement have identified opportunities to automate data storytelling [**?** ]. It remains to be researched, however, to what level this automation is feasible and when a data scientist should still be in the loop. Storytelling, both manual and automatic, unlocks the opportunity of personalization. This immediately raises additional concerns related to ethics. For example, can personalized explanations trigger different interpretations issues? Storytelling also has the opportunity to emphasize some part of the data, hereby providing an answer to the data abundance problem. One should be careful, however, not to 'obscure' the data: one should be transparent on which data is emphasized and which is hidden.

*Evaluation.* The main identified opportunity of the evaluation of XLA is related to the evaluation of the *impact of the explanations* (3 opportunity stickers, 1 challenge sticker, 0 veto).
While the problem of evaluating XLA is new, there are consolidated techniques (from evaluation as a whole) that can be adapted ranging from perceived utility evaluation to impact on, e.g., advising and decision making. Evaluation of XLA can start with a usability evaluation, but can and should go beyond usability studies. One should take care to not only have separate evaluations with separate stakeholder groups, but to use the opportunity of mixed-group evaluations. Evaluation should moreover focus on both subjective and more objective indicators: one should not only rely on subjective statements but also attempt to look for objective/quantitative measures, such as the impact on learning gain.

*Implementation & adoption.* The main opportunity for the implementation and adoption of XLA is the adaptation to different target groups (5 opportunity stickers, 2 challenge stickers, no veto)
Different target groups might need different explanations and interpretations of learning analytics predictions and recommendations. Each stakeholder might have particular needs and therefore, the explanations and interpretations should be personalized to the particular group of stakeholders.

## 3.2 Challenges

*Data.* The main identified challenge for data in XLA is to include information about and take into account the *recency of data* (0 opportunity stickers, 4 challenge stickers, 0 veto stickers)
It is challenging to find a threshold that could uniquely define what 'recent' and 'old' data are. There is a tension between how valuable is old data (to obtain enough data to train the models or to show historical evolution) and new data (more representative of current state). Additionally, attention should be paid to how to explain to users which data is used in these models and that the predictions and recommendations rely on past data to evaluate the current state. Finally, deploying XLA can, and most likely will, influence the data itself as it is expected to have an impact on actual learning and teaching.

*Stakeholders.* The main challenge for the stakeholders is to build a *strong community.* (1 opportunity sticker, 4 challenge stickers, no veto)
For the entire field of learning analytics it is a challenge to build a strong community that could support explanations of predictions and recommendations. However, if done well, the explanations

have the opportunity to foster trust among different stakeholders, such as students and teachers in a MOOC.

*Communication.* The main challenge for communication within XLA is to consider the *time dimension of learning analytics* (0 opportunity stickers, 6 challenge stickers, 0 veto). Longitudinal data is challenging to handle within learning analytics. XLA should be able to provide explanations for the different phases over time. Moreover, these explanations should be tailored to the particular phases and contexts they are provided in.

*Evaluation.* The main challenge for a good evaluation of XLA is to identify the *added value for the stakeholders* (1 opportunity sticker, 3 challenge stickers, 0 veto).
The evaluation of XLA should focus on identifying the added value of explanations for different stakeholders, and in particular should be able to show how the explanations contribute to what the stakeholders already know (e.g., using a pre/post test setup).

*Implementation & adoption.* The main challenge for implementation and adoption of XLA are both *technical and pedagogical support* during deployment (1 opportunity sticker, 3 challenge stickers, 0 veto)
The actual implementation and adoption of XLA will provide ample challenges, especially when deployments at scale are considered. These issues are not only from a technical nature, but also pedagogical: how can the explanations be used appropriately during the learning process?

## 4 DISCUSSION AND CONCLUSION

This short paper called upon the input of more than 40 stakeholders to shape the domain of Explainable Learning Analytics (XLA), which aims at developing LA-specific advancement regarding Explainable Artificial Intelligence (XAI). In particular, this paper reports on the opportunities and challenges of XLA as identified by this group of international stakeholders collected during a 3-hour workshop at the 2019 European Conference on Technology Enhanced Learning (EC-TEL). Opportunities and challenges were collected regarding five main themes: data, stakeholders, communication, evaluation, and implementation & adoption. With these findings this paper is providing a small, but important contribution to the development of XLA.

The input from stakeholders is undoubtedly valuable for the advancement of XLA. Therefore, this contribution can only be a small step towards a more profound integration of the different stakeholders in the development of the domain, as this paper has some important limitations. First, the workshop was held at the EC-TEL 2019 conference hereby causing a biased sample of the stakeholder population. The involved stakeholders were mainly researchers active in Technology Enhanced Learning. Future stakeholder consultation should more heavily involve practitioners and end-users of LA.
A positive element of the stakeholder population is that they represented the wider domain of Technology Enhanced Learning, of which LA is only a sub-domain. On the negative side however, hereby introducing the second limitation, this meant that some attendants were not very acquainted with the specifics of the LA domain, while others were considered experts. The same holds for XAI: some attendants were experienced researchers or users, while others were not familiar with the domain. For future stakeholder consultations we recommend to set up a protocol that aims at better handling such differences in expertise, both regarding the LA and the XAI domain.
A third obvious limitation is the short duration of the workshop, which limited both the width and the depth of the discussion. We therefore aim for this paper to stimulate further discussion.

To conclude, we can state that this paper contributes to the development of the XLA domain by the identification of challenges and opportunities regarding data, stakeholders, evaluation, communication, and implementation & adoption.

## ACKNOWLEDGMENTS

We thank the participants of the EC-TEL 2019 XLA workshop for their contribution. This work was funded by the EU LALA project (grant no. 586120-EPP-1-2017-1-ES-EPPKA2-CBHE-JP). This project has been funded with support from the European Commission. This publication reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

## REFERENCES

[1] Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y. Lim, and Mohan Kankanhalli. 2018. Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 582, 18 pages. https://doi.org/10.1145/3173574.3174156
[2] Amina Adadi and Mohammed Berrada. 2018. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* 6 (2018), 52138–52160.
[3] Sule Anjomshoae, Amro Najjar, Davide Calvaresi, and Kary FrÃďmling. 2019. Explainable Agents and Robots: Results from a Systematic Literature Review. *Proceedings Of The 18th International Conference On Autonomous Agents And Multiagent Systems*, 1078–1088.
[4] Jaegul Choo and Shixia Liu. 2018. Visual Analytics for Explainable Deep Learning. (2018).
[5] Bum Chul Kwon, Min-Je Choi, Joanne Taery Kim, Edward Choi, Young Bin Kim, Soonwook Kwon, Jimeng Sun, and Jaegul Choo. 2019. RetainVis: Visual Analytics with Interpretable and Interactive Recurrent Neural Networks on Electronic Medical Records. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2019), 299–309.
[6] Sina Mohseni, Niloofar Zarei, and Eric Ragan. 2018. A Survey of Evaluation Methods and Measures for Interpretable Machine Learning. *arXiv.org* (2018). http://search.proquest.com/docview/2139809046/
[7] Ramaravind Kommiya Mothilal and Tom Broos. 2018. Predicting first-year engineering student success: from traditional statistics to machine learning. In *Proceedings of the 46th SEFI conference*. Copenhagen, Denmark, 1–8. https://www.sefi.be/proceedings/?conference=copenhagen2018
[8] Sixun Ouyang, Aonghus Lawlor, Felipe Costa, and Peter Dolog. 2018. Improving Explainable Recommendations with Synthetic Reviews. (2018).
[9] Thilo Spinner, Udo Schlegel, Hanna Schäfer, and Mennatallah El-Assady. 2019. explAIner: A Visual Analytics Framework for Interactive and Explainable Machine Learning. *IEEE transactions on visualization and computer graphics* (2019).
[10] A. Vellido. 2019. The importance of interpretability and visualization in machine learning for applications in medicine and health care. *Neural Computing and Applications* (2019).
[11] Katrien Verbert, erik Duval, Joris Klerkx, Sten Govaerts, and JosÃľ Santos. 2013. Learning Analytics Dashboard Applications. *American Behavioral Scientist* 57 (09 2013). https://doi.org/10.1177/0002764213479363
[12] Katrien Verbert, Sten Govaerts, Erik Duval, Jose Luis Santos, Frans Van Assche, Gonzalo Parra, and Joris Klerkx. 2014. Learning dashboards: an overview and future research opportunities. *Personal and Ubiquitous Computing* 18, 6 (2014), 1499–1514.
[13] Quan-shi Zhang and Song-chun Zhu. 2018. Visual interpretability for deep learning: a survey. *Frontiers of Information Technology & Electronic Engineering* 19, 1 (2018), 27–39.
[14] Yongfeng Zhang and Xu Chen. 2018. Explainable Recommendation: A Survey and New Perspectives. (2018).

[15] Xun Zhao, Yanhong Wu, Dik Lun Lee, and Weiwei Cui. 2019. iForest: Interpreting Random Forests via Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2019), 407–416.