Human Factors in Software and Systems Engineering, Vol. 94, 2023, 1–10
https://doi.org/10.54941/ahfe1003764

Explaining Algorithmic Decisions: Design Guidelines for Explanations in User Interfaces

Charlotte Haid, Alicia Lang, and Johannes Fottner

School of Engineering and Design, Technical University Munich, 85748 Garching, Germany

ABSTRACT

Results from Artificial Intelligence (AI) algorithms emerge without an explanation for the outcome being provided. The field of Explainable AI (XAI) has set out to develop algorithms that run before or after the original algorithms to make the results explainable. These include, for example, the LIME or the SHAP algorithm. However, the explanations of these new algorithms are very mathematical and expert knowledge is needed to understand the explanations. They are not suitable for every audience of an AI application. We show guidelines for the creation of understandable explanations and use the application example of AI-based shift planning in logistics to show how explanations can be used to address the end user with a user interface (UI).

Keywords: Explainable AI, User interface, Transparency, Explainability, Artificial intelligence

INTRODUCTION

Digitization and automation of logistics processes are leading to an ever-increasing amount of available data in production logistics. This data can be used as a basis for decision-making through data analysis tools or algorithms, for example Artificial Intelligence algorithms. AI-based decision support systems are increasingly in use in logistics (Owczarek, 2021). In addition to many advantages like a fast analysis or a wide pattern recognition, AI-based algorithms also have a disadvantage: they cannot explain their own behaviour and the decision made is often not comprehensible (Yampolskiy, 2020).

The non-provision of information to the user can result in limited effectiveness and a lack of transparency. The systems are not able to explain the decisions and actions to the user (Adadi and Berrada, 2018). This problem is addressed by the research field of Explainable Artificial Intelligence (XAI), which seeks to improve transparency and trust in AI-based systems (Adadi and Berrada, 2018).

However, this research rarely examines explanatory user interfaces and user interactions (Chromik et al., 2021). There is a lack of practical methods and examples on how to incorporate the human factor into the development process of AI-generated explanations (Schoonderwoerd et al., 2021). Thus, it is crucial to include the human-computer interaction (Adadi and

Berrada, 2018). Transparency, usability, and trust are the main requirements for a successful incorporation of AI-based systems in companies. All of these three can be addressed by a strong explanation in the user interface of a system.

In this paper, we first clarify the concept of Explainable AI as well as the requirements for an implementation of AI in industry. We then present existing guidelines for creating explanations in the context of AI. We demonstrate the application of these guidelines to the prototypical design of a shift scheduling system for logistics, which we use as an example for AI based decision support in logistics. In the final recommendations, we supplement these guidelines with our own recommendations that we applied during our development process.

BACKGROUND

AI research faces the challenge of making decisions of algorithmic systems explainable. To generate explanations, new algorithms have been designed to explain the decisions of the first algorithms post hoc. This subfield is called explainable artificial intelligence. With huge amounts of data being increasingly available, AI-based decision support is becoming a growing issue in manufacturing and logistics. Users like workers or managers, but also works councils in industry, demand transparency and fairness in the use of AI (Haid et al., 2021). Besides transparency, usability plays an important role in this context (Amershi et al., 2019; Hentati et al., 2016). It represents one of the most important quality factors of a user-friendly design for user interfaces. In addition to usability, trust in the system is an important criterion for the successful incorporation of AI in companies (Abdelkafi et al., 2019; Glikson and Woolley, 2020). For many people, it is difficult to trust an AI. Therefore, systems should be designed transparently to increase user trust. Various studies confirm this connection (Bussone et al., 2015; Weitz et al., 2019).

Explainable AI (XAI). The term refers to efforts that have emerged in response to AI transparency and AI trust issues (Adadi and Berrada, 2018). The research field of XAI is concerned with developing techniques that explain models while maintaining the same level of performance, thus making AI more transparent. The goal is to increase the comprehensibility of AI systems to humans (Adadi and Berrada, 2018). With the help of XAI, it is possible for systems to explain their procedures and thus make their behaviour more understandable (Arai, 2019).

Usability of XAI algorithms. AI algorithms usually do not provide explanations on their results. In recent years, several methods emerged on how to explain AI models. Here, a large part deals with post hoc methods, giving an explanatory layer to the fully trained models (Laugel et al., 2019). The smaller part deals with ante hoc methods, where explanatory mechanisms are included in training (Longo et al., 2020). Global explanations, explaining the entire AI model and local explanations describing specific model behaviour, exist (Hammond et al., 2021).

LIME (Local Interpretable Model-Agnostic Explanations) is an algorithm that can explain the predictions of any classifier by learning an interpretable model around the prediction locally. In the case of image recognition, for example, a LIME algorithm can highlight the image areas based on which the algorithm arrived at its decision. LIME can be used for text, tabular data and more and gives diagrams or graphs as visualization of the explanation (Molnar, 2019).

SHAP (SHapley Additive exPlanations), a game theoretic approach that can be applied to the output of any machine learning model, connects optimal credit allocation with local explanations. It uses Shapley values for the allocation similar to game theory. The explanation of SHAP shows feature attributions as forces, whereas each feature value is a force that decreases or increases the predicted value (Molnar, 2019).

Both algorithms have in common that only the software developer has the expertise and background to understand the visualized curves or values. Laymen who know neither the original nor the explaining algorithm in detail cannot get much out of the explanation. Current research trends show that in addition to an algorithm-based explanation, an end-user-centric explanation of the decision is also necessary. Here, there is still a lack of practical examples and guidelines on how explanations can also be made comprehensible for the end user.

Conclusion. In summary, AI algorithms lack explanations to explain their results. Therefore, other algorithms have been developed that provide explanations as diagrams or graphs for the results of the first algorithms. LIME and SHAP are among the best known of these. However, the explanations lack user-friendliness. Often, only experts like software developers can understand the visualizations of the explanation. Therefore, it is important to additionally involve the end user, make it possible for them to understand the explanations and to create guidelines for good explanations.

GUIDELINES FOR EXPLANATIONS FROM LITERATURE

In the research of XAI, explanatory UIs and user interactions have hardly been studied (Chromik et al., 2021). There is a lack of practical methods and examples of how to include the human factor into the development of AI-generated explanations (Schoonderwoerd et al., 2021).

One of the most crucial factors to make a model understandable through explanations is the involvement of humans in XAI. To explain the complex system to humans, human-computer interaction skills are needed in addition to technical expertise (Adadi and Berrada, 2018). A user interface represents the interface of human-computer interaction and communication with a device or system. Explanations are an interface between the user and AI. This interface is understandable to the user and accurately represents the AI (Hamon et al., 2020; Sayed-Mouchaweh, 2021). However, this topic is poorly addressed in the literature. Most explanations are based on researchers' intuition. Thus, XAI should build on existing research regarding explanations in the fields of philosophy, psychology, or cognitive science (Miller, 2017).

What is a good explanation? A good explanation should be contrastive, selected, social, truthful, focus on the abnormal, be general and probable as well as be consistent with prior beliefs of the explainee. A contrastive explanation design means to explain why event A happened instead of another event B, rather than just emphasizing why event A occurred (Miller, 2017), (Molnar, 2019). In addition, it is important that explanations are limited to only one or two causes and are thus formulated selectively. Abnormal explanations, which use unlikely causes to explain an event, are good explanations, according to Molnar. Truthful explanations are those that predict the event as truthfully as possible (Molnar, 2019).

Properties of explanations by Molnar. Properties of explanations in terms of results from explainable AI algorithms are presented: accuracy, reliability, consistency, stability, understandability, certainty, novelty, and representativeness. Accuracy means how well the explanation predicts the data. If an explanation comes very close to the prediction of the black box model, it is considered reliable. If explanations are very similar between models that have been trained for the same task and provide similar predictions, they are consistent. High stability means that small variations in the characteristics of an instance do not change it significantly. To formulate explanations that people can follow, they should be formulated understandable. Safety means that the explanations reflect the safety of the model. Here, there is a connection with novelty. The higher the novelty, the more likely the model has a low certainty because data are missing. Accordingly, novelty denotes whether a data instance is far from the distribution of the training data. Finally, representativeness describes whether explanations refer to the whole model or only to individual predictions. Also emphasized is that people prefer short explanations (Molnar, 2019).

Design principles by Chromik et al. Four important design principles for explanations are addressed (see Figure 1). The first states that explanations should be written in natural language. Often, visual explanations of an AI can only be understood by experts, so wording should be based on a user's speech style. In addition, they can be linked to visual cues to facilitate understanding (Chromik et al., 2021). Yu et al., for example, implemented a switch that allows the user to convert a visual explanation into a detailed text (Yu et al., 2019). Second, the user should be given the opportunity to respond, for example with follow-up questions. Accordingly, explanations should be structured in such a way that the content is presented to the user progressively in order to avoid overwhelming the user (Chromik et al., 2021). Springer and Whittaker emphasize a progressive disclosure of the explanations so that only the most important information is revealed to the user and further details are

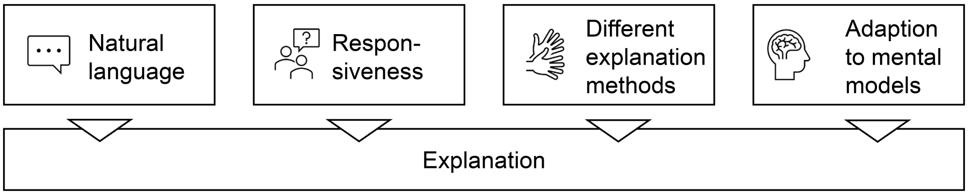


Figure 1: Design principles for explanations by [Chr-2021].

faded in below if necessary (Springer and Whittaker, 2020). For example, a "Why?" button can be included to provide the user with the opportunity to obtain visual information (Millecamp et al., 2019).

Offering different explanation methods is another design principle, as the user can understand the facts via different ways. The research recommends local and global explanations to give the user an overview of the system but also at the same time the possibility to get information about specific facts (Chromik et al., 2021). Last, explanations should relate to the user's mental models. For example, Xie et al. employ a button which can be used by physicians to display only high confidence and low complexity explanations (Xie et al., 2020).

Additions by Mueller et al. Mueller et al. take up the inclusion of the user's mental models in their research on the characteristics of an AI explanation system. They criticise that many XAI systems do not refer to these models and the user's goals or knowledge but reveal a blanket explanation. Explanations should be relevant and timely to the user's goals. For example, if a critical situation like an error occurs, the user can be supported by an explanation. If an explanation is iterative and interactive, this can help the user answer their own questions (Mueller et al., 2021).

Explanations should not only act as a tool of an algorithm but can only be effective if the user can understand these explanations. It is important that the explanations refer to the user's working context and goals as well as tasks and are supported by instructions, tutorials, and comparisons. Additionally, it is recommended to consider use cases, user models, up-to-dateness, and attention as well as distraction limits when creating explanations. It is important to consider the consequences of an explanation on trust. Another approach is the use of repeated and re-explanations in dynamic AI systems to show the user the changes in the algorithm (Mueller et al., 2021).

In summary, building on the above literature, interdisciplinary collaboration in this topic can provide important advances, as expertise from AI, psychology, or human-computer interaction can be combined. This can advance XAI research tremendously (Abdul et al., 2018). In addition, the field of interaction design, which has not been systematically analyzed so far, can be included to make an XAI effectively explainable (Chromik et al., 2021). Given guidelines for explanations should be respected.

APPLICATION IN USE CASE

Based on the recommendations for good and transparent explanations from the literature, we created a user interface that prototypically represents the user interface of a shift scheduling system. The shift scheduling system called "iPlan" is designed to assign employees to workstations in logistics. In particular, the preferences of the employees for certain workstations are taken into account in this system (Haid et al., 2022). The algorithm for the assignment is an AI-based algorithm from the class of constraint programming algorithms. It provides no detailed explanation on its results.

An iterative approach with intermediate feedback from potential users was chosen to create the user interface (Figure 2). For user-friendly handling, we

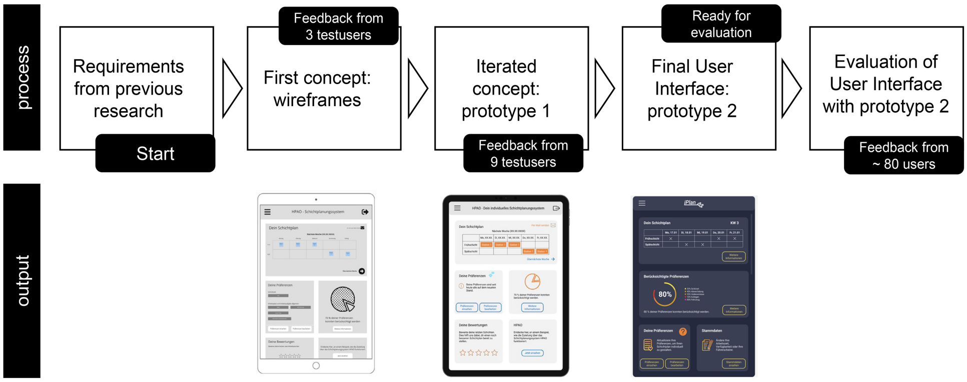
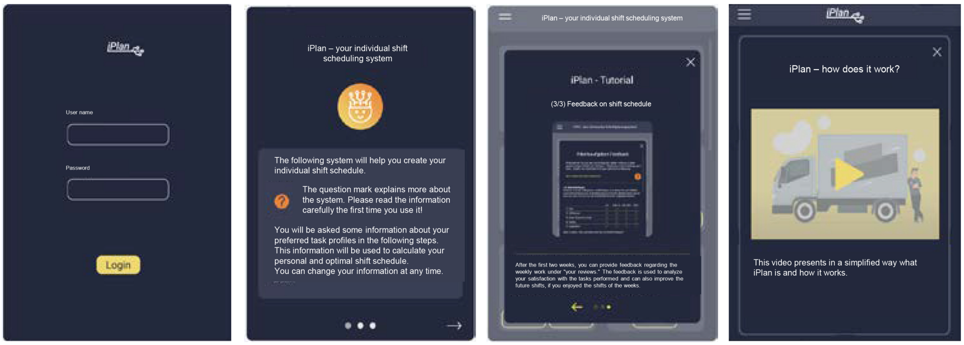

process
Requirements from previous research
Start
Feedback from 3 testusers
First concept: wireframes
Iterated concept: prototype 1
Feedback from 9 testusers
Ready for evaluation
Final User Interface: prototype 2
Evaluation of User Interface with prototype 2
Feedback from ~ 80 users
output

Figure 2: Approach for the creation and evaluation of the user interface.

selected an app for tablet or smartphone as target design and created the prototypical designs in Figma. Based on requirements of an initial research, a first design was created as wireframes. This design was evaluated with three test persons. With the feedback, a second design was created, prototype 1, and again evaluated. Subsequently, two versions of prototype 2 were created, a transparent version with many explanations and a less transparent version without explanatory aids. Both versions were tested with a group of 40 people (80 people in total: 87% male, 11% female, 2% diverse; 64% of age 15 – 35, 36% of age 36-65), with the test participants having to work through various scenarios and then fill out a questionnaire based on the system usability scale (SUS). All test persons were potential end users from the area of production logistics. The transparently designed version of the user interface was found to be very explanatory and user-friendly. However, it did not have significantly better usability compared to the group that was presented the non-transparent UI. This may be due to the fact, that both versions already had a high usability and acceptance among users (mean value for usability with SUS score for UI A: 86.4, UI B 83.8).

Excerpts from our user interface for the preference-based shift scheduling system are shown in Figures 3 and 4. The dark background is intended to create a passive and calm atmosphere, yellow and orange highlight control elements. Start screen is a login page (Figure 3 (1)). After the login, the user is given a brief introduction to the symbols used (Figure 3 (2)). With excerpts from the main system, the most important pages are briefly explained (Figure 3 (3)). The user can always access the tutorial pages via the main menu. To give different explanation methods, there is a video available as well, which explains the purpose and the advantages for the users of the shift scheduling system iPlan in 90 seconds.

Explanations of what the user should do - for example to click on something - are displayed with a small hand. For better usability, various icons that could stand for further explanations (for example, a speech bubble, a robot, and a question mark) were evaluated. The question mark was rated as the most fitting by the test users and is used in many places to provide further explanations of content (Figure 4 (1)). The information can be displayed by clicking on the question mark (Figure 4 (2)). The third image shows the


iPlan
User name
Password
Login
iPlan – your individual shift scheduling system
iPlan – your individual shift scheduling system
iPlan - Tutorial
(3/3) Feedback on shift schedule
iPlan
iPlan – how does it work?

The following system will help you create your individual shift schedule.

The question mark explains more about the system. Please read the information carefully the first time you use it!

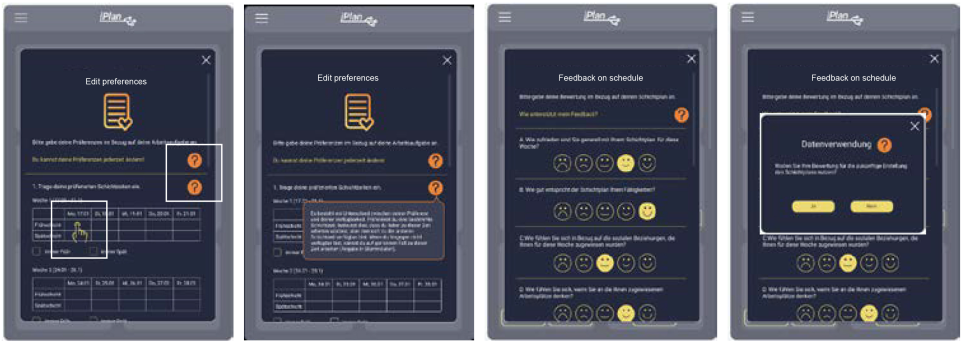
You will be asked some information about your preferred task profiles in the following steps. This information will be used to calculate your personal and optimal shift schedule. You can change your information at any time.

After the first two weeks, you can provide feedback regarding the weekly work under "your reviews." The feedback is used to analyze your satisfaction with the tasks performed and can also improve the future shifts, if you enjoyed the shifts of the weeks.

This video presents in a simplified way what iPlan is and how it works.

Start screen | Start tutorial after login | Tutorial pages | Video explanation

Figure 3: Sample pages from design prototype.



Explanation for action | Detailed information | Feedback on results | Question on data usage

Figure 4: Sample pages from design prototype.

option for feedback that our users can provide on their shift assignment if needed. Here we have paid particular attention to notes on the use of data: Only if the user explicitly agrees, their data will be used to improve the results of the AI algorithm (Figure 4 (3)). Otherwise, the data will be stored but not used further by the algorithm. For example, a manager could view the data and discuss the feedback directly with employees if necessary.

RECOMMENDATION OF DESIGN GUIDELINES

Reflecting the explanation methods and advises from literature as well as our own process of developing a transparent user interface, we recommend the following guidelines for designing explanations of your AI system.

Involve humans in development. Developing a user interface with explanations is combining expertise from different fields. AI developers, social science experts and human-computer interaction experts should be part of your development team. Consider working with use cases, user models, and prototypes in the development process. Follow an iterative process to get feedback from potential users in each stage of the prototype. We recommend having at least two iterations to get valuable feedback and use scenario techniques to let the user interact with the user interface. If you already know the people you are developing the system for, you should actively involve them in the development process as well as your team. Use regular meetings to present results and give the opportunity for questions and feedback.

Use high quality explanations. A good explanation is easy to understand and addresses the user's trust. Limiting the explanation to one or two causes instead of various causes rises the quality of your explanation. Make sure to be consistent with other explanations and underlying beliefs of your system.

Address the user with your explanation. To best pick up the user, use natural language in the explanations. Make sure to refer to the working context and the goals of the user and show only relevant explanations. You can use catchy symbols to facilitate understanding. To prevent the user from being overwhelmed, fade in explanations gradually and give the user the option of reading up on explanations as needed. Since not everyone likes to read, you should offer different explanation methods such as images or videos as an alternative. Consider attention and distraction limits of people while reading, listening, or watching. People prefer short explanations. In general, it is useful to alternate the explanation methods.

Interact with the user. Responsiveness is one of the main design principles by Chromik et al. Giving your user the opportunity to respond on your explanations makes him or her feel more being part of the system. The explanation should therefore be interactive and iterative. If you have a dynamic AI system, you can use repeated explanations to explain results and show changes in the algorithm. The user could for example change input data himself and see the differing results for a better understanding. Give local and global explanations: an overview of the system as well as information on specific facts should be included. Support your explanation by instructions, tutorials, or comparisons, to give the user even better opportunities to learn about the system and ask questions.

Consider consequences on trust. Have in mind that trust, usability, and transparency of a system are linked. High usability and transparency can increase user trust in a system. The importance of users developing trust in a system is particularly high in the field of AI. AI systems arouse a certain amount of scepticism among users in advance for various reasons, including media attention and reporting. It is therefore important to develop not only user trust in the system, but also to involve users in the development process through participatory concepts, the opportunity for questions and criticism, as well as a good support during the implementation.

CONCLUSION

This paper offers ideas on how explanations of AI systems can be presented to the user via user interfaces. Guidelines for the creation of explanations and the goodness of explanations were derived from the literature and supplemented by our own. Basically, it is important to involve later users in the development process of the AI system and the corresponding user interface. The user should feel addressed by the explanations and an interaction with the system should be possible. Good explanations that address different mental models of the users are essential. The impact of a user interface on the trust in an AI system should not be underestimated and is therefore important to consider. These guidelines can be applied not only to the development of a shift planning system but can be used in many ways for the development of AI systems.

ACKNOWLEDGMENT

The authors would like to thank Sarah Fürstberger and Stefano Ferrari for their efforts in developing and testing the user interface of our shift scheduling system. Furthermore, we thank Dr. Charlotte Unruh and Prof. Dr. Tim Büthe for their helpful feedback on the user interface during our common research project. The project was funded by the Institute for Ethics in Artificial Intelligence from Technical University Munich.

REFERENCES

Abdelkafi, N., Döbel, I., Drzewiecki, J. D., Meironke, A., Niekler, A. and Ries, S. (2019) *Künstliche Intelligenz (KI) im Unternehmenskontext*.

Abdul, A., Vermeulen, J., Wang, D., Lim, B. Y. and Kankanhalli, M. (2018) 'Trends and Trajectories for Explainable, Accountable and Intelligible Systems', *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems,* ACM.

Adadi, A. and Berrada, M. (2018) 'Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)', *IEEE Access*, vol. 6, pp. 52138–52160.

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R. and Horvitz, E. ([2019]) 'Guidelines for Human-AI Interaction', in *Proceedings of the 2019 CHI*, pp. 1–13.

Arai, K. (2019) *Intelligent Computing: Proceedings of the 2019 Computing Conference, Volume 2* [Online], Cham, Springer International Publishing AG. Available at https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=5813662.

Bussone, A., Stumpf, S. and O'Sullivan, D. ([2015]) 'The Role of Explanations on Trust and Reliance in Clinical Decision Support Systems', in *2015 IEEE International Conference on Healthcare Informatics*, pp. 160–169.

Chromik, M., Eiband, M., Buchner, F., Krüger, A. and Butz, A. (2021) 'I Think I Get Your Point, AI! The Illusion of Explanatory Depth in Explainable AI', *26th International Conference on Intelligent User Interfaces,* ACM, pp. 307–317.

Glikson, E. and Woolley, A. W. (2020) 'Human Trust in Artificial Intelligence: Review of Empirical Research', *Academy of Management Annals*.

Haid, C., Stohrer, S., Unruh, C., Büthe, T. and Fottner, J. (2022) 'Accommodating Employee Preferences in Algorithmic Worker-Workplace Allocation', *Human Factors in Management and Leadership*.

Haid, C., Unruh, C., Pröger, I., Fottner, J. and Büthe, T. (2021) 'Personaleinsatzplanung in der Logistik', *Zeitschrift für wirtschaftlichen Fabrikbetrieb*, vol. 116, no. 12, pp. 908–912.

Hammond, T., Verbert, K., Parra, D., Knijnenburg, B., O'Donovan, J. and Teale, P. (eds) (2021) *26th International Conference on Intelligent User Interfaces*, ACM.

Hamon, R., Junklewitz, H. and Sanchez, I. (2020) *Robustness and Explainability of Artificial Intelligence - From technical to policy solutions*.

Hentati, M., Ben Ammar, L., Trabelsi, A. and Mahfoudhi, A. ([2016]) 'Model-driven Engineering for Optimizing the Usability of User Interfaces', in *Proceedings of the 18th International Conference on Enterprise Information Systems*, pp. 459–466.

Laugel, T., Lesot, M.-J., Marsala, C., Renard, X. and Detyniecki, M. (2019) *The Dangers of Post-hoc Interpretability: Unjustified Counterfactual Explanations* [Online]. Available at https://arxiv.org/pdf/1907.09294.

Longo, L., Goebel, R., Lecue, F., Kieseberg, P. and Holzinger, A. (2020) 'Explainable Artificial Intelligence: Concepts, Applications, Research Challenges and Visions', vol. 12279, pp. 1–16.

Millecamp, M., Htun, N. N., Conati, C. and Verbert, K. (2019) 'To explain or not to explain', in *Proceedings of the 24th International Conference on Intelligent User Interfaces*, pp. 397–407.

Miller, T. (2017) *Explanation in Artificial Intelligence: Insights from the Social Sciences* [Online]. Available at https://arxiv.org/pdf/1706.07269.

Molnar, C. (2019) *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable* [Online], Morisville, North Carolina, Lulu. Available at https://christophm.github.io/interpretable-ml-book/.

Mueller, S. T., Veinott, E. S., Hoffman, R. R., Klein, G., Alam, L., Mamun, T. and Clancey, W. J. (2021 [2021]) 'Principles of Explanation in Human-AI Systems' [Online]. Available at https://arxiv.org/pdf/2102.04972.

Owczarek, D. (2021) *The Trend of AI in Logistics and Supply Chains - Applications, Advantages, and Challenges* [Online]. Available at https://nexocode.com/blog/posts/ai-in-logistics/.

Sayed-Mouchaweh, M. (ed) (2021) *Explainable AI Within the Digital Transformation and Cyber Physical Systems: XAI Methods and Applications*, Cham, Springer International Publishing; Imprint Springer.

Schoonderwoerd, T. A., Jorritsma, W., Neerincx, M. A. and van den Bosch, K. (2021) 'Human-centered XAI: Developing design patterns for explanations of clinical decision support systems', *International Journal of Human-Computer Studies*, vol. 154.

Springer, A. and Whittaker, S. (2020) 'Progressive Disclosure', *ACM Transactions on Interactive Intelligent Systems*, vol. 10, no. 4, pp. 1–32.

Weitz, K., Schiller, D., Schlagowski, R., Huber, T. and André, E. (2019) "'Do you trust me?"', *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents,* Association for Computing Machinery, pp. 7–9.

Xie, Y., Chen, M., Kao, D., Gao, G. and Chen, X. '. (2020) *CheXplain* [Online]. Available at https://arxiv.org/pdf/2001.05149.

Yampolskiy, R. V. (2020) 'Unexplainability and Incomprehensibility of AI', *Journal of Artificial Intelligence and Consciousness*, vol. 07, no. 02, pp. 277–291.

Yu, B., Yuan, Y., Terveen, L., Wu, Z. S., Forlizzi, J. and Zhu, H. (2019) *Keeping Designers in the Loop: Communicating Inherent Algorithmic Trade-offs Across Multiple Objectives* [Online]. Available at https://arxiv.org/pdf/1910.03061.