Shih, JL. et al. (Eds.) (2023). Proceedings of the 31st International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Explainable AI in the Real World: Challenges and Opportunities

Dora HORVAT[a], Ivica BOTICKI[a*], Peter SEOW[b] & Antun DROBNJAK[a]
[a]*Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia*
[b]*National Institute of Education, Nanyang Technological University, Singapore*
*ivica.boticki@fer.hr

Abstract: This paper presents the results of a systematic review of the research papers on the use of explainable AI in the real world. The present body of research indicates there is a huge drive from the academic society in pushing and exploring explainable AI across disciplines from a research perspective, and there is inherent need to design prototypes with increased complexity to tackle the numerous scientific and methodological issues in the process. The main conclusions of the review are that there exist serious methodological issues with the use of XAI in complex systems which reside on vast or layered information systems spanning across multiple organizational units with important data sometimes missing, potentially limiting the validity of the XAI approach used in practice. For XAI to work in the real-world context of education, the approaches to presenting explanations to the stakeholders such as teachers and students should be understandable by them to take appropriate actions or decisions. This would highlight the need to study of human-computer interaction between AI and users that would lead to better transparency, trust and personalization.

Keywords: Explainable AI, industry applications, education implications

1. Introduction

The main goal of the 1955 proposal for the Dartmouth Summer Research Project on Artificial Intelligence (AI), a workshop labeled as the birthplace of artificial intelligence was "...to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it." (McCarthy et al., 2006). Artificial intelligence was then defined as "making a machine behave in ways that would be called intelligent if a human were so behaving", which despite the plethora of definitions remains relevant today. Even though AI definitions today emphasize the broad reach of artificial intelligence, the formulation given by its creators still stands at its core.

Since its beginnings, AI has continuously been developing at a fast pace. In 1964, scientists at MIT developed ELIZA, one of the first language processing computing systems which was able to imitate a Rogerian therapist. In 1997, IBM's Deep Blue computer beat the world champion Gari Kasparov in chess, thus going down in history as the first computer to win a game against a world champion under tournament time conditions. A decade later, a yearly tournament for chess engines began, where engines compete against each other. Today, digital AI-based assistants such as Siri or Cortana are widely implemented and used in everyday life in language processing tasks.

In recent years in particular, the field of artificial intelligence has made major progress in almost all its main sub-areas, including computer vision, speech recognition, natural language processing, expert systems and decision making (Michael L. Littman, 2021). AI is in some form ingrained into all major fields of human work – finance, security, healthcare and medicine, criminal justice, transportation, marketing, telecommunications. With that, AI is no longer observed exclusively in the context of accuracy and model optimization; it is the

role of humans that becomes a crucial factor. As decisions made by intelligent systems are affecting human lives in areas such as medicine or law, the need for understanding how these decisions are furnished by AI methods becomes imperative. The notion of explainable artificial intelligence (XAI) in most cases implies more than just understanding the model. Depending on the area of application and the intended users, XAI is expected to achieve goals other than 'just explanations'. For example, when AI is used for decision-making, it is crucial to ensure the model is fair and unbiased; for systems that work with sensitive and personal data, privacy must be preserved; and an AI system must be trustworthy – the user has to be confident that a model will act as intended.

This paper critically examines the potential of XAI in the real world uses, mapping out the studies which move further from the proof-of-concept efforts in order to illustrate the potential for the application of XAI in the real world. By doing that, the study is a pioneering effort in the field.

2. Background

Although explainability is the core of responsible and trustworthy AI, across multiple disciplines and areas (Guidotti et al., 2018), there is still no agreement in the literature as to what explainability actually is. When talking about explanations in AI, there is no one-size-fits-all solution – the complexity of the explanations and the amount of detail provided is entirely dependent on who the intended users are (Barredo Arrieta et al., 2020). Furthermore, since explainability is ultimately a human-agent interaction problem (Miller, 2019), the solution to explainable AI lies not just in 'more AI', but in considering multiple aspects of human understanding, drawing insights from psychology and social sciences (Barredo Arrieta et al., 2020). Although the uniform definition of explainability remains elusive, the term black-box problem is a defining aspect of XAI. The black-box problem is a well known phenomenon in the field of artificial intelligence and represents the main limitation of effectiveness and usage of machine and deep learning models. As pointed out in the DARPA's Explainable Artificial Intelligence (XAI) Program, there is an inherent tension between machine learning performance (predictive accuracy) and explainability; with often the highest performing methods (e.g., deep learning) being the least explainable, and the most explainable (e.g., decision trees) being less accurate (Gunning and Aha, 2019).

In XAI-related literature, terms explainability and interpretability are often used interchangeably (Tjoa and Guan, 2021). Much like with explainability, there is no general agreement on the definition of interpretability, as well as no clear distinction between the two. However, there exist domains in which they do not convey the same concepts (Ehsan et al., 2018). In general, interpretability refers to the ability to explain or present a model in terms understandable to humans. It denotes passive characteristics of a model – how easy it is for humans to make sense of and identify relations between model's inputs and outputs (Barredo Arrieta et al., 2020) (Došilović, Brčić and Hlupić, 2018). Even though explainability is related to interpretability, explainability remains is associated with internal logic behind an intelligent system. In contrast to interpretability, explainability denotes an active characteristic of a model – how deep human's understanding of a model's inner workings is (Linardatos, Papastefanopoulos and Kotsiantis, 2021).

When differentiating between the two main approaches to generating explanations, the literature makes a clear distinction between post-hoc explainability and intrinsic explainability. Post-hoc explainability refers to methods which are applied to models that are not interpretable by design to improve their interpretability (Barredo Arrieta et al., 2020). They do not directly explain the inner workings of the model, instead they offer 'approximate' explanations such as textual, visual, or example-based explanations. Intrinsic explainability refers to models which are interpretable on their own due to their simple structure, such as linear regression or tree-based models. However, these models, which often stand to as an

interpretable alternative to complex black-box models such as neural networks, are not unconditionally interpretable. Taking high-dimensionality or heavy pre-processing and feature engineering into account, they are often not more intrinsically interpretable than black-box models (Lipton, 2018).

3. Methodology

Since explainable AI is an emerging concept in terms of applicable methods and techniques in machine and deep learning, the focus of this review was to get an insight into the state of explainable AI, i.e. – are explainable models and frameworks being implemented and validated and if so, to what degree. Since explainability is not an easily measured component of AI, studies evaluating explainability of the system on real users were an important point of interest. In addition, studies validating models on real scenarios were also taken into consideration. For the medical and healthcare field, clinicians or experts had to be involved in some way, either via design or through validation.

The search for relevant studies on real-life applications of XAI was conducted on Web of Science electronic database in the period between December 2021 and March 2022. The search query used is a general one in order to gather studies for a wide overview of explainable artificial intelligence in a range of domains, which is the main aim of this review. At the time of the search, the query returned 1777 results from Web of Science database Publication year after 2017 was selected as an additional filter in order to gain focus on the current state of applications of XAI.

The 1777 records from Web of Science were obtained for title and abstract screening using the inclusion and exclusion criteria agreed-upon during the inter-rater process. Application of inclusion and exclusion criteria resulted in the set of 163 studies on which the full-text screening was conducted. The full-text screening was more qualitative in contrast to the title and abstract screening, which had precisely defined inclusion and exclusion criteria. During the full-text screening, focus was on studies that implemented a usable explainable system and evaluated explainability of the system on human subjects through user studies. Following that, 144 papers were excluded, leaving 19 papers in the final list (Figure 1).

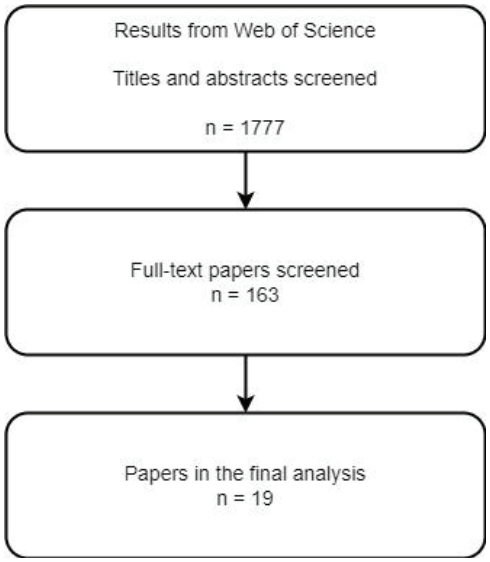

Figure 1. Screening the paper for final analysis

4. Inclusion and exclusion criteria

Final inclusion and exclusion criteria are formed as shown in Table 1. Agreed upon inclusion and exclusion criteria between the two raters after the 2nd inter-rater.

Table 1. Agreed upon inclusion and exclusion criteria between the two raters after the 2nd inter-rater

Inclusion criteria	Exclusion criteria

• Include papers proposing XAI models or frameworks and:
  ○ Demonstrating application and validating it on a real scenario or a simulated real scenario (use-cases)
  or
  ○ Performing user-studies or user-evaluations
• Include papers about the use of XAI in healthcare, medical and biomedical field if:
  ○ Clinicians/experts are involved in the study design
  ○ User-studies with clinicians/experts are performed
	• Exclude reviews and overviews of XAI field and methods
	• Exclude papers conducting surveys about explainability and trustworthiness of AI and XAI
	• Exclude papers discussing design patterns and principles
	• Exclude papers discussing ethical issues, bias, transparency and trust in the context of AI and XAI

5. Results

A total of 14 studies were included in this review (Table 2). The biggest proportion of studies belongs to the healthcare domain, as is the case in a systematic review of explainable AI application domains (Islam et al., 2022). Besides six studies from the healthcare domain, three studies describe the use of XAI in the domain of human-computer interaction in the form of collaborative games or human performance improvement; one study belongs to the telecommunications domain; one to the industry domain in the form of supply chain planning support; and one to the energy domain through recommendations for energy efficiency. Most of the studies are from 2020 to 2021, and they conduct user studies with 5 to 60 participants to evaluate their systems and approaches (except for (Irarrázaval et al., 2021)).

Gradient-boosted decision trees along with the Shapley values for generating explanations are used in two studies, (Chromik, 2021) and (Melançon et al., 2021). In (Deperlioglu et al., 2022) , (Xu et al., 2021) and (Wang and An, 2021) neural networks are used as the machine learning model while CAM (along with Grad-CAM and DeconvNet) method is used for providing visual explanations. Studies (Xie et al., 2020) and (Sardianos et al., 2021) do not specify the inner workings of their systems in terms of machine learning and explainable models; the focus is on iterative design of the system by following user requirements and user feedback. In (Xie et al., 2020) a Clinical Decision Support System is constructed. The system has multiple features for generating explanations, such as producing contrastive examples, outputting probabilities and the ability to show the most significant observations.

In (Sardianos et al., 2021) a recommendation system for energy efficiency with explainable and persuasive recommendations is designed. In (Khodabandehloo, Riboni and Alimohammadi, 2021) decision tree learning algorithm is used as an intrinsically explained model; the decision tree is then parsed to produce natural language explanations. In (Samuel, Abdullah and Raj, 2021) granular computing is used to interpret SVM's classification and construct syllogisms which are then transformed into natural language explanations. In (Gao et al., 2020) a spatial-temporal causal And-Or graph (STC-AoG) is used as representation of a robot's knowledge and inferred user's mental state. By parsing this graph and applying the proposed explanation generation framework, the robot can generate explanations with an aim to correct sub-optimal human behavior in human-robot collaboration tasks. In (Das and Chernova, 2020) a Rationale Generating Algorithm is proposed; it produces rationales (natural language expressions) which aim to aid the user's decision-making process and consequently increase the user's understanding of the performed task. In (Ehsan et al., 2018) an encoder-decoder network is used to translate between state and action information and natural language rationalizations, which serve as rationalizations for describing agent behavior. In (Sabol et al., 2020) semantically explainable fuzzy classifier CFCMC explains the decision by giving a semantic explanation on the possibilities of misclassification, and visual explanations by showing the training sample most responsible for a given prediction as well as training samples from other, conflicting classes. And finally, in (Irarrázaval et al., 2021) data is first clustered into groups using an unsupervised learning approach, after which a CART algorithm is used to construct a decision tree from which a set of rules is extracted.

Table 2. The final set of selected studies after the application of the literature review steps

Author and year	Domain	ML model	Explainable method	User study participants
(Xie et al., 2020)	Healthcare	-	-	6
(Xu et al., 2021)	Healthcare	ANN	Grad-CAM, Guided Grad-CAM	9
(Deperlioglu et al., 2022)	Healthcare	CNN	CAM	15
(Samuel, Abdullah and Raj, 2021)	Healthcare	SVM	Granular Computing	5
(Khodabandehloo, Riboni and Alimohammadi, 2021)	Healthcare	Decision Tree	-	8
(Sabol et al., 2020)	Healthcare	Fuzzy Model	Cumulative Fuzzy Class Membership Criterion (CFCMC)	14
(Irarrázaval et al., 2021)	Telecommunications	Clustering, CART	-	-
(Wang and An, 2021)	Education*	CNN	CAM, DeconvNet	30
(Gao et al., 2020)	Human-computer interaction	-	Rationalization	29
(Ehsan et al., 2018)	Human-computer interaction	Encoder-decoder NN	Rationalization	53

(Das and Chernova, 2020)	Human-computer interaction	-	Rationalization	60
(Chromik, 2021)	General*	XGBoost	SHAP	16
(Melançon et al., 2021)	Industry	XGBoost	SHAP	*
(Sardianos et al., 2021)	Energy	-	Explainable Recommendation	8
(Westerski et al., 2021)	Industry	ADM	Rationalization	-
(ten Broeke et al., 2021)	Healthcare	-	Explainable Decisions	15
(Chakraborti et al., 2019)	IT	-	Explainable Recommendation	-
(Gonzalo-Cristóbal et al., 2021)	Education	Monte Carlo + ANN	Explainable Recommendation	12
(Mirchi et al., 2020)	Healthcare	SVM	Adaptive Explanation	50

6. Discussions

There is no doubt that Explainable AI presents one of the most challenging topics in contemporary research. With the rise of artificial intelligence, the importance of explaining its inner workings is ever so important across disciplines, with the healthcare domain being most prominent, especially in the context of medical diagnosis and decision-support tools (Islam et al., 2022; Tjoa & Guan, 2021). The prospects and opportunities for medical XAI are at first glance many: the improvement of medical diagnosis, help with allocation of resources, reduction of bias, further AI development and increased adoption. Nevertheless, practical implications of the use of explainable AI across disciplines remain and are therefore explored as part of the study presented in this review paper.

Designing explainable AI often comes with difficult design decisions lying at the intersection of AI technology and its practical applications. The reviewed studies indicate that the ambiguity of the term explainable presents a great challenge. The concept of explainability is abstract and must be observed in the context of the user and its environment. Consequently, explanations should not be regarded just as just a product, but a multidisciplinary process (Khosravi et al., 2022). To achieve this, the human must be put in the focus of the design of explainable systems. Just as in decision-support systems, the human-centric approach and consideration of context should be the focus when developing XAI platforms. Such an approach can be noted in (Mirchi et al., 2020), where the experts' opinion altered the importance of metrics (features) that was determined by the AI model, to ensure that the framework was in concordance with the current guidelines of neurosurgical education. In the design of the explanations, the authors followed several cognitive and learning models and theories, while also ensuring that the given feedback mimics real-life experiences through textual, audio and video-based instructions. This type of feedback is claimed to improve self-guided learning and develop responsibility, which is highlighted as highly beneficial (Winne, 2021). The claims about benefits and relevance of the framework are however not verified, as the conducted user study only validates the technical workings of the model.

Once the explainable AI designs and solutions are in place, they are trialed and tested in real-life environments, but the testing is done in a laboratory fashion. The whole experiential setup is arranged with maximum support given to the party applying XAI solutions, often neglecting the idiosyncrasies of the real world and oftentimes not addressing the realistic performance of XAI. The vast majority of screened papers from the medical field only report

results in the form of technical metrics such as accuracy and precision, whereas the explainable part of the machine learning system often appears to be implemented in order to formally justify the XAI label, without much thought or validation on whether these explanations are useful in real-life settings. This is also noted in (Liao & Varshney, 2021), where the “disconnect between technical XAI approaches and supporting user’s end goals in usage context” is identified as one of the pitfalls of developing explainable models. The disconnect can be observed in two related aspects; one aspect refers to the lack of cross-disciplinary research and studies on users’ needs and preferences. The other aspect refers to the lack of performance and relevance evaluation of XAI models in real-life settings. (Liao & Varshney, 2021) identify the absence of studies providing evidence that incorporation of explainable components in AI systems and solutions improves realistic user performance in judgment and decision making, while (Adadi & Berrada, 2018) report the lack of XAI models evaluation no assessment of their relevance to the user. Some studies warn that giving up predictive power in favor of transparency and explainability should be carefully considered and properly justified upon (Lipton, 2018).

In the education context, XAI aims to address concerns related to fairness, accountability, transparency and ethics in educational interventions supported by AI algorithms (Khosravi et al., 2022). XAI can benefit teachers by gaining a better understanding of how AI systems work and make decisions, which can help them to better integrate AI tools into their teaching practice. For students, XAI can personalize their learning experiences by providing explanations that are tailored to their individual needs and characteristics. Explainable AI can support student’s self-regulation by providing transparency and interpretability of the predictions and recommendations, which can help students better understand their performance and take appropriate actions to improve it (Afzall et al., 2021). For XAI to work in the real-world context of education, the approaches to presenting explanations to the stakeholders such as teachers and students should be understandable by them to take appropriate actions or decisions. This would highlight the need to study of human-computer interaction between AI and users that would lead to better transparency, trust and personalization.

Serious methodological issues surface with the use of XAI in complex systems which reside on vast or layered information systems spanning across multiple organizational units, where data sharing is limited. Such inherent issues reflect on the quality of integration of XAI solutions heavily and raise the question of whether the totality of information contributing to the actual workings of the processes is well modeled and described. What is more, in some cases important data from certain section of a system might be totally missing, potentially limiting the validity of the XAI approach used in practice. (Melancon, 2021), in their cooperation with Michelin, points out the lack of data as a restricting challenge, due to the absence of a standard practice of archiving all data in detail. In the medical domain, despite continued improvements of electronic health records, data quality and availability still present an issue (Gerlings, 2022).

7. Conclusions

The complexity of XAI solutions presents a hurdle to their seamless user adoption. This does not come as a surprise, since there is a huge drive from the academic society in pushing and exploring explainable AI across disciplines from a research perspective, and there is inherent need to design prototypes with increased complexity to tackle the numerous scientific and methodological issues in the process. (Westerski, 2021) noted that, in their case, the introduction of a framework which is significantly different from users’ current habits had a negative impact on adoption of the system. To minimize those negative effects, users should be involved in the process of design and validation, with keeping in mind that the same models have very different results depending on organization profile.

Last but not the least, the issues of data collection processes and data privacy are tightly related to the implementation of XAI systems in real-life scenarios, as such systems heavily relying on a variety of data sources. Although the manipulation and processing of such data presents a challenge itself, data privacy remains an insurmountable issue, since explaining data inevitably leads to discovering certain sensitive bits and information. Ethical and trust issues regarding model transparency and the overall black-box problem of AI emerge across domains. As an example, the explanations are crucial for pedagogical effectiveness of a digital system, as well as gaining students' and teachers' trust in given decisions (Conati et al., 2018).

References

Afzaal, M., Nouri, J., Zia, A., Papapetrou, P., Fors, U., Wu, Y., X, Li, & Weegar, R. (2021). Explainable Ai For Data-driven Feedback and Intelligent Action Recommendations To Support Students Self-regulation. Frontiers in Artificial Intelligence, (4). https://doi.org/10.3389/frai.2021.723447

Barredo Arrieta, A. et al. (2020) "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," Information Fusion, 58. doi:10.1016/j.inffus.2019.12.012.

Došilović, F.K., Brčić, M. and Hlupić, N. (2018) "Explainable artificial intelligence: A survey," in 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 210–215. doi:10.23919/MIPRO.2018.8400040.

Ehsan, U. et al. (2018) "Rationalization: A Neural Machine Translation Approach to Generating Natural Language Explanations," in AIES 2018 - Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society. doi:10.1145/3278721.3278736.

Guidotti, R. et al. (2018) "A survey of methods for explaining black box models," ACM Computing Surveys, 51(5). doi:10.1145/3236009.

Gunning, D. and Aha, D. (2019) "DARPA's Explainable Artificial Intelligence (XAI) Program," AI Magazine, 40(2), pp. 44–58. doi:10.1609/aimag.v40i2.2850.

Linardatos, P., Papastefanopoulos, V. and Kotsiantis, S. (2021) "Explainable ai: A review of machine learning interpretability methods," Entropy. doi:10.3390/e23010018.

Lipton, Z.C. (2018) "The mythos of model interpretability," Communications of the ACM, 61(10). doi:10.1145/3233231.

Melo, E., Silva, I., Costa, D. G., Viegas, C. M. D., & Barros, T. M. (2022). On the Use of eXplainable Artificial Intelligence to Evaluate School Dropout. Education Sciences, 12(12), 845. MDPI AG. Retrieved from http://dx.doi.org/10.3390/educsci12120845.

McCarthy, J. et al. (2006) "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955," AI Magazine, 27(4), p. 12. doi:10.1609/aimag.v27i4.1904.

Littman, M. L., Ajunwa, I., Berger, G., Boutilier, C., Currie, M., Doshi-Velez, F., ... & Walsh, T. (2022). Gathering strength, gathering storms: The one hundred year study on artificial intelligence (AI100) 2021 study panel report. arXiv preprint arXiv:2210.15767. Miller, T. (2019) "Explanation in artificial intelligence: Insights from the social sciences," Artificial Intelligence. doi: 10.1016/j.artint.2018.07.007.

Shubhendu, S. and Vijay, J. (2013) Applicability of Artificial Intelligence in Different Fields of Life. Available at: www.ijser.in.

Shukla Shubhendu, S. and Vijay, J. (2013) "Applicability of Artificial Intelligence in Different Fields of Life," International Journal of Scientific Engineering and Research (IJSER), 1(1).

Tjoa, E. and Guan, C. (2021) "A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI," IEEE Transactions on Neural Networks and Learning Systems, 32(11). doi:10.1109/TNNLS.2020.3027314.

Altman, D.G. (1990) Practical statistics for medical research. CRC press.

Chromik, M. (2021) "Making SHAP Rap: Bridging Local and Global Insights Through Interaction and Narratives," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). doi:10.1007/978-3-030-85616-8_37.

Das, D. and Chernova, S. (2020) "Leveraging rationales to improve human task performance," in International Conference on Intelligent User Interfaces, Proceedings IUI. doi:10.1145/3377325.3377512.

Deperlioglu, O. et al. (2022) "Explainable framework for Glaucoma diagnosis by image processing and convolutional neural network synergy: Analysis with doctor evaluation," Future Generation Computer Systems, 129. doi:10.1016/j.future.2021.11.018.

Ehsan, U. et al. (2018) “Rationalization: A Neural Machine Translation Approach to Generating Natural Language Explanations,” in AIES 2018 - Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society. doi:10.1145/3278721.3278736.
Gao, X. et al. (2020) “Joint Mind Modeling for Explanation Generation in Complex Human-Robot Collaborative Tasks,” in 29th IEEE International Conference on Robot and Human Interactive Communication, RO-MAN 2020. doi:10.1109/RO-MAN47096.2020.9223595.
Irarrázaval, M.E. et al. (2021) “Telecom traffic pumping analytics via explainable data science,” Decision Support Systems, 150. doi:10.1016/j.dss.2021.113559.
Islam, M.R. et al. (2022) “A Systematic Review of Explainable Artificial Intelligence in Terms of Different Application Domains and Tasks,” Applied Sciences (Switzerland), 12(3). doi:10.3390/app12031353.
Khodabandehloo, E., Riboni, D. and Alimohammadi, A. (2021) “HealthXAI: Collaborative and explainable AI for supporting early diagnosis of cognitive decline,” Future Generation Computer Systems, 116. doi:10.1016/j.future.2020.10.030.
Landis, J.R. and Koch, G.G. (1977) “The Measurement of Observer Agreement for Categorical Data,” Biometrics, 33(1). doi:10.2307/2529310.
Melançon, G.G. et al. (2021) “A machine learning-based system for predicting service-level failures in supply chains,” Interfaces, 51(3). doi:10.1287/INTE.2020.1055.
Sabol, P. et al. (2020) “Explainable classifier for improving the accountability in decision-making for colorectal cancer diagnosis from histopathological images,” Journal of Biomedical Informatics, 109. doi:10.1016/j.jbi.2020.103523.
Samuel, S.S., Abdullah, N.N.B. and Raj, A. (2021) “Interpretation of SVM to Build an Explainable AI via Granular Computing,” in Studies in Computational Intelligence. doi:10.1007/978-3-030-64949-4_5.
Sardianos, C. et al. (2021) “The emergence of explainability of intelligent systems: Delivering explainable and personalized recommendations for energy efficiency,” International Journal of Intelligent Systems, 36(2). doi:10.1002/int.22314.
Wang, C. and An, P. (2021) “A Mobile Tool that Helps Nonexperts Make Sense of Pretrained CNN by Interacting with Their Daily Surroundings,” in Extended Abstracts of MobileHCI 2021 - ACM International Conference on Mobile Human-Computer Interaction: Mobile Apart, Mobile Together. doi:10.1145/3447527.3474873.
Xie, Y. et al. (2020) “CheXplain: Enabling Physicians to Explore and Understand Data-Driven, AI-Enabled Medical Imaging Analysis,” in Conference on Human Factors in Computing Systems - Proceedings. doi:10.1145/3313831.3376807.