Journal of Information System and Technology Management
EISSN: 0128-1666
JISTM
Volume 6 Issue 22 (September 2021) PP. 36-50
DOI: 10.35631/JISTM.622004

JOURNAL OF INFORMATION SYSTEM AND TECHNOLOGY MANAGEMENT (JISTM)
www.jistm.com

TOWARDS USER-CENTRIC EXPLANATIONS FOR EXPLAINABLE MODELS: A REVIEW

Ali Hassan[1*], Riza Sulaiman[2], Mansoor Abdullateef Abdulgabber[3], Hasan Kahtan[4]

[1] Institute of IR 4.0, The National University of Malaysia Bangi, Malaysia
Email: P10086@siswa.ukm.edu.my
College of Computer and Cyber Sciences, University of Prince Mugrin, Madinah, Saudi Arabia
Email: a.hassan@upm.edu.sa
[2] Institute of IR 4.0, The National University of Malaysia Bangi, Malaysia
Email: riza@ukm.edu.my
[3] The National University of Malaysia Bangi, Malaysia
Email: hakmansoor@gmail.com
[4] Faculty of Computer Science and Information Technology, University of Malaysia, Kuala Lumpur, Malaysia
Email: hasankahtan@um.edu.my
[*] Corresponding Author

Article Info:

Article history:
Received date: 10.06.2021
Revised date: 15.07.2021
Accepted date: 20.08.2021
Published date: 01.09.2021

To cite this document:

Hassan, A., Sulaiman, R., Abdulgabber, M. A., & Kahtan, H. (2021). Towards User-Centric Explanations For Explainable Models: A Review. *Journal of Information System and Technology Management, 6* (22), 36-50.

DOI: 10.35631/JISTM.622004.

Abstract:

Recent advances in artificial intelligence, particularly in the field of machine learning (ML), have shown that these models can be incredibly successful, producing encouraging results and leading to diverse applications. Despite the promise of artificial intelligence, without transparency of machine learning models, it is difficult for stakeholders to trust the results of such models, which can hinder successful adoption. This concern has sparked scientific interest and led to the development of transparency-supporting algorithms. Although studies have raised awareness of the need for explainable AI, the question of how to meet real users' needs for understanding AI remains unresolved. This study provides a review of the literature on human-centric Machine Learning and new approaches to user-centric explanations for deep learning models. We highlight the challenges and opportunities facing this area of research. The goal is for this review to serve as a resource for both researchers and practitioners. The study found that one of the most difficult aspects of implementing machine learning models is gaining the trust of end-users.

Keywords:

User-Centric, Explainable Artificial Intelligence, Human-AI Interaction, Machine Learning

Journal of Information System and Technology Management
EISSN: 0128-1666
JISTM
Volume 6 Issue 22 (September 2021) PP. 36-50
DOI: 10.35631/JISTM.622004

Introduction

Many fields of activity that have embraced new information technologies have Artificial Intelligence (AI) at their core, ranging from banking to medical care. While the origins of AI date back several decades, there is widespread agreement that intelligent machines with learning, thinking, and adaptive capabilities are critical today (West, 2018). The efficiency of intelligent systems has recently advanced to the point where their design and implementation essentially does not require human involvement. When the decisions made by such systems have an impact on people's lives (such as in law, banking, or medical care), there is a growing need to understand how such decisions are made using AI methods (Goodman & Flaxman, 2017). Indeed, AI is seen as playing a crucial role in the global battle to diagnose and monitor diseases (Chen & Asch, 2017; Davenport & Kalakota, 2019; Gómez-González et al., 2020; Jiang et al., 2017). However, the efficiency of the system is limited by the fact that the computer cannot provide explanations for its judgments in this context, which is critical for human users. Other areas include driverless vehicles in transportation, security, and finance. Users need to be able to identify and analyze the context underlying AI responses and have the knowledge to understand and trust the conclusions (Abdul et al., 2018; Lysaght et al., 2019; Wiens & Shenoy, 2018). To achieve higher prediction accuracy, higher model complexity is often used.

The deep learning model, which is at the heart of most modern machine learning systems, is an excellent example. It allows computers to automatically explore, learn, and extract the hierarchical data representations required for recognition and classification tasks. On one hand, there are black-box models such as Deep Learning (Lecun et al., 2015). However, they are challenging for real-world applications as they can be difficult to interpret. White-box or glass-box models, on the other hand, provide easily explainable results; examples include decision tree and linear-based models. While these models are more explainable and interpretable, they are less powerful and do not reach the state of the art compared to black-box models. It is difficult to trust systems whose conclusions are hard to explain, especially in areas such as self-driving vehicles or healthcare where moral and justice issues inevitably arise. For example, the basic functional structure of the model is too complicated for healthcare professionals to understand. Apart from the diagnostic outcome, the model provides no explanation of how the output was generated. This could be a problem as it would be difficult for healthcare professionals to communicate the diagnosis to their patients (Ahmad et al., 2018). These concerns have led to the call for more information and accountability. Machine learning models should be "interpretable" (Ahmad et al., 2018), especially in sensitive areas such as the medical field. Transparency is critical to the implementation of many intelligent systems, such as medical diagnostics (Lysaght et al., 2019; Wiens & Shenoy, 2018). The field of explainable Artificial Intelligence (XAI) (Gunning & Aha, 2019) has been revived in response to the demand for reliable, fair, resilient models with increased reliability for real-world applications. Figure 1 shows the evolution of the popularity of keywords related to "Explainable AI" over time, as evaluated by dimensions.ai. The substantial growth in recent years, reflecting the rejuvenation of the field, is reflected in the increasing research output during the same period.

Journal of Information System and Technology Management
EISSN: 0128-1666
Volume 6 Issue 22 (September 2021) PP. 36-50
DOI: 10.35631/JISTM.622004

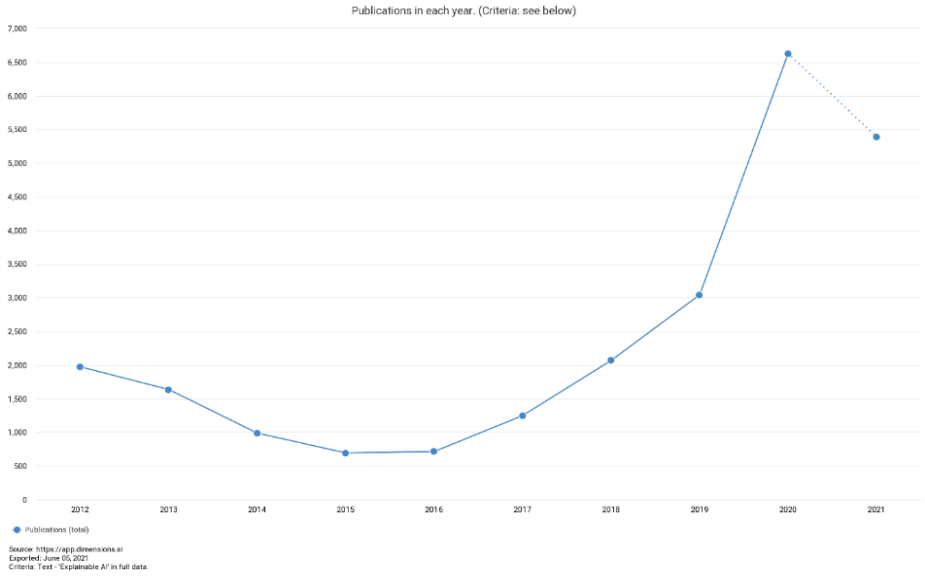

Figure 1: Keywords Related to "Explainable AI"
Source: https://app.dimensions.ai

This paper presents a literature review of Human-Centred Machine Learning and the development of approaches to user-centric explanations for deep learning models. The remainder of this paper is organized as follows: First, we examine the evolution of Human-Centred Machine Learning over time and provide a brief background on the topic of explainable artificial intelligence (XAI). We then present the methodology. Next, we take a closer look at User-Centred Explainable AI and the tools used in this area. Then, the opportunities and challenges are discussed. Finally, conclusions are presented.

Background

Human-Centric Machine Learning

Human-Centric Machine Learning (HCML) explores how machine learning systems can be tailored to users' interests and ways of working. The increasing need for machine learning systems that can adapt to real-world contexts requires accurate and complete HCML research. The term HCML has been of interest in the literature for decades (Figure 2), although the HCML idea only gained traction with the trend of Deep Learning starting in the mid-2010s. Typical users are typically unaware of AI's capabilities, drawbacks, and internal workings. Explainability, user experience (UX), privacy, protection, and trustworthiness are all issues that users have raised.

Journal of Information System and Technology Management
EISSN: 0128-1666
Volume 6 Issue 22 (September 2021) PP. 36-50
DOI: 10.35631/JISTM.622004

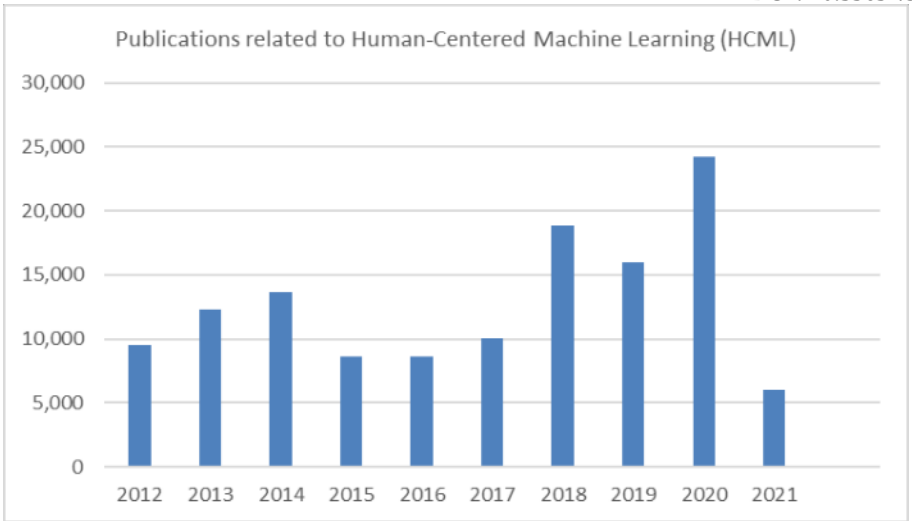

Figure 2: Human-Centric Machine Learning (HCML) Publications 2012- April 2021
Source: https://app.dimensions.ai

Human-centric machine learning considers machine learning and human-computer interaction (HCI); two fields each with different approaches. The machine learning approach uses pre-existing, typically standardized datasets and common metrics such as accuracy, precision, and recall to evaluate results (Smith et al., 2018). HCI, on the other hand, engages with people and conducts quantitative or qualitative user research (Good & Omisade, 2019). HCI enables designers to understand how people interact with and use ML systems (Kim, 2018). Early and continuous user involvement in the design process can help in developing systems that are easy to learn and increase user productivity and acceptance (Abdul et al., 2018; Riveiro & Thill, 2021). Many HCI design techniques emphasize user feedback and dialog with the developers of the technological system. For efficient interaction with ML systems, the need to involve end-users in the design process is emphasized (Teixeira et al., 2017). To be successful, easy to use and considered useful, ML systems must be created with careful consideration of the needs of the target users (Kim, 2018). This has sparked interest in using a user-centric design approach to create machine learning systems (Cagiltay et al., 2008; Dabbs et al., 2009).

In recent years, there has been a surge of interest in the study of explainable AI (Abdul et al., 2018; Biran & Cotton, 2017; Holzinger et al., 2017). Explainability in AI is one of the most important requirements for implementing responsible AI due to several factors. In certain scenarios, the user dealing with AI needs more logical knowledge than just the system's decision. The judgment of developers, whose experience and knowledge are not always representative of the expertise of the end-user, is used to determine whether the explanation is appropriate (Miller, 2019). As a result, the human-computer interaction (HCI) community has advocated for a user-centered approach to explainability (Wang et al., 2019).

Explainable artificial intelligence (XAI)

Deep Learning (DL) is a subfield of machine learning that uses various learning algorithms instead of task-specific algorithms to learn data representations. Deep Learning-based models such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) and Support Vector Machines (SVM) are combined with strong computer vision (CV) methods such as Active Appearance Model (AAM) to learn robust data (Gupta, 2019). Examples of

Journal of Information System and Technology Management
EISSN: 0128-1666
Volume 6 Issue 22 (September 2021) PP. 36-50
DOI: 10.35631/JISTM.622004

algorithms commonly used to support human medical services include decision trees, neural networks, k-nearest Neighbour, and support vector machines. At the computational level, pattern recognition algorithms examine data such as test results, photographs, surveys, medical records and medical samples (Souillard-Mandar et al., 2016; Xie et al., 2019; Yu et al., 2016). We may hesitate to trust the decisions of machines when confronted with human activities such as medical diagnosis. When humans are confronted with a new technology that requires them to rely on invisible features, acceptance may be delayed. Transparency-centered methods for interpretation could help overcome these problems. However, interpretability alone is not sufficient (Gilpin et al., 2019). To gain people's trust in black-box techniques, we need explainable models that can explain neural network activity, gain users' trust, or provide insight into why they make decisions.

While interoperability is an important first step, these systems also need to be thorough, explaining their judgments and providing sufficient answers to concerns. Due to their lack of interpretability, most deep learning models are generally considered complex "black boxes," which makes them difficult to use in real-world healthcare settings. Apart from the diagnostic outcome, the model does not provide information on how the conclusion was reached, and the underlying workings of the model are challenging for clinicians to understand (Yan et al., 2019: Ahmad et al., 2018). This can be a serious problem for doctors as they are unable to communicate the diagnosis to their patients and hence the treatment can suffer due to the low probability of misdiagnosis (Ahmad et al., 2018). Researchers argue that machine learning should be "interpretable by design" (Abdul et al., 2018). According to a research (Lysaght et al., 2019; Wiens & Shenoy, 2018), many intelligent systems, such as medical diagnostics, require clarity and explainability to the user, as this is important in this new AI environment for several reasons. In certain cases, the person interacting with AI needs more rational information than just the system's judgment. With the rapid emergence of AI models, it is, therefore, more important than ever to understand why and how the models work in different situations, as well as to take into account the user's desires and expectations while delivering better results.

Methodology

This paper gives an overview of the research on user-centric explanation approaches for human machine learning and deep learning models. The work was completed in four stages (Figure 3). The publications utilized in the study were obtained through Google Scholar, SpringerLink, IEEE Explorer, and the ACM Digital Library. We focused our search on ‘Explainable Artificial Intelligence’, ‘Human-Centered Artificial Intelligence’, ‘Human-Centered Machine Learning’ and ‘user-centered explanations’. The title, summary, introduction, process, findings, discussion, and conclusion were all carefully reviewed. Mendeley was chosen to handle references, and Microsoft Excel was chosen as a spreadsheet to identify trends and patterns by collecting keywords, and summarizing the abstracts.

Journal of Information System and Technology Management
EISSN: 0128-1666
Volume 6 Issue 22 (September 2021) PP. 36-50
DOI: 10.35631/JISTM.622004

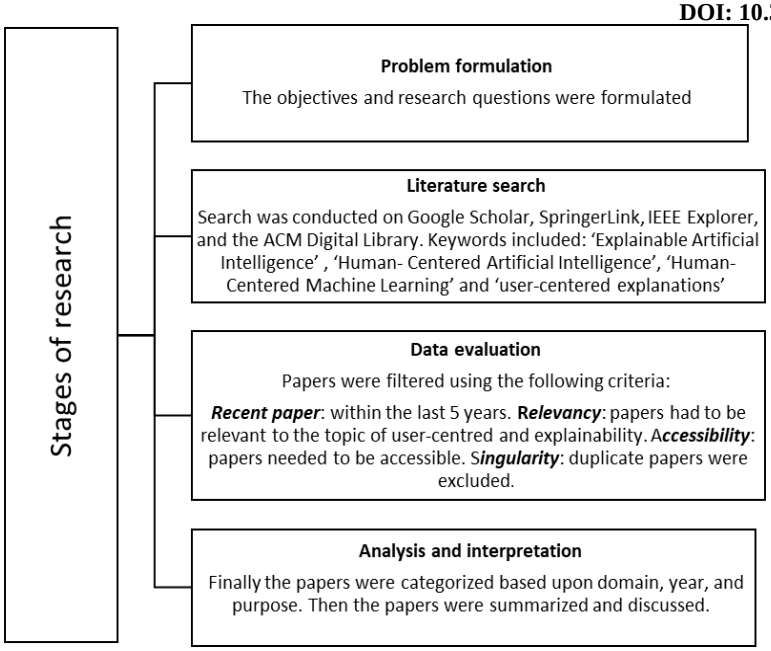

Figure 3: Stages of Research

Findings

As AI has entered the world of human decision-making, it has become difficult for human users to rely on system decisions, especially in the healthcare field. System incompetence can limit the effectiveness of the methods used to explain the system's perspective to the user. A growing number of open-source frameworks incorporate XAI technology for end-users to understand and manage AI systems (Arya et al., 2019). Research has highlighted the importance of analyzing end-user AI needs, and the need to interpret models, but user assessments are particularly lacking. System acceptance and use are likely to be influenced by a combination of contextual and design factors. Therefore, when considering how a user-centric explanation affects the effective operation of such a system, it is important to consider the variables that influence the acceptance and use of that system.

User-Centred Explainable AI

Explainability to the user is critical in this emerging AI environment for a variety of reasons. In certain contexts, users require a different rational understanding than the system's decision. The demand for explanatory AI has attracted interest from stakeholders concerned about the outcomes of AI technology. To determine whether the interpretation is appropriate, experienced model developers are consulted, whose experience and expertise are not intrinsically indicative of the end-users (Freitas, 2014; Miller, 2019). As a consequence, the user-centric approach has become popular in the human-computer interaction (HCI) sector (Wang et al., 2019). The goal of this new field of study is to apply a human explanatory social science framework, or to investigate how explanations affect user interaction (Wang et al., 2019).

Cai et al. (2019) considered the specifications for an AI assistant for use in the medical industry. In a qualitative laboratory study, the researchers questioned 21 pathologists to find out how non-AI specialists interacted with AI approaches. The Deep Neural Network (DNN) predicted prostate cancer diagnosis in the participants. The study shows that before assessing the model's

Journal of Information System and Technology Management
EISSN: 0128-1666
Volume 6 Issue 22 (September 2021) PP. 36-50
DOI: 10.35631/JISTM.622004

results, clinicians want to understand the model's basic and global features, such as documented strengths and weaknesses, subjective views, and overall design goals. By integrating activity theory with theoretical insights into the different phases of a project, Good and Omisade (2019) suggested that user-centric design has proven to be a viable paradigm for building mobile health services. They concluded that incorporating activity theory and user-centric design into e-health systems would help maintain or overcome elements that contribute to success or failure. Farao et al. (2020) presented a framework for mobile health app development that incorporates both Information Systems Research (ISR) and Design Thinking. They showed how the proposed approach can be used in the context of an application that reports tuberculin skin test results. The combined structure allowed for end-user interaction and cost-effective and rapid growth of the application, while addressing contextual issues and needs (Farao et al., 2020). Beede et al. (2020) presented a study evaluating deep learning systems used to detect diabetic retinopathy. Through interviews and observations with nurses, the researchers found socio-environmental factors that influence model performance, nurse workflow, and patient experience.

Krening and Feigh (2019) reported results from user research conducted with non-machine learners to investigate certain features of interactive reinforcement learning technologies. The researchers conducted a human experiment in which participants were asked to play a simple game with many interactions. The results show that both time delays and prospective user interfaces contributed to a negative user experience. A study by Amershi et al. (2014) examined interactive machine learning systems in which humans intentionally interact with machine learners to learn. The researchers presented several studies showing how interactivity leads to a close relationship between the system and the user, and how some existing systems have difficulty responding to users. Similarly, Smith-Renner et al. (2020) examined how different levels of input affect user behavior and subjective feedback. In two controlled experiments, the researchers examined how descriptions of the sensor model ML automatically influence users' expectations of the ML model. For low-quality models, explanations without questioning the input are challenging, but the experience between description and feedback in high-quality models suggests that certain responses should not be obtained without an explanation. Zhu et al. (2018) proposed an artificial AI (XAID) that can be explained to designers, and hence opened up new areas of research to help game designers better use AI and ML in design work through collaborative creation. According to the researchers, both the algorithmic properties of the underlying AI technology and the requirements of human designers need to be considered for the XAID system to be useful and efficient. Ehsan and Riedl (2019) examined the explanations offered and focused on how they affect human perception. In both tests, participants watched multi-format videos in which the subject "thinks aloud" while performing a list of activities. The study determined the effectiveness of rationales created and evaluated by non-experts in terms of confidence, human-like, adequate reasoning, and comprehensibility. By combining thematic analysis and grounded theory to evaluate open-ended justifications, the researchers uncovered new elements associated with each dimension. The work of Chari et al. (2020) revealed a descriptive ontology system called "Web Ontology Language (OWL)" that developers can use to incorporate various specifications into their AI-enabled systems. He explained how descriptive ontology can be used in the development and operation of artificial intelligence systems to support diabetes care decisions. The researchers believe that their models have shown suitable conditions for formulating complex types of explanations that serve different purposes and reflect their ontology for different types of knowledge.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 6 Issue 22 (September 2021) PP. 36-50
DOI: 10.35631/JISTM.622004

By evaluating the professional challenges of producing explainable goods with artificial intelligence, the researchers Liao et al. (2020) used a complementary strategy to distinguish between the algorithmic workings of XAI and what is needed to meet the real-world needs of customers. According to the researchers, the importance of explanations for a particular AI system is determined by the problem that needs an understanding of what users want. The XAI query bank was intended to bridge the gap between user requirements for AI explanations and the technological capabilities of XAI. Liao et al. (2020) argued that XAI research should focus on ways to meet user requirements. They further stated that it is difficult for practitioners to tell the difference between algorithmic design and the creation of human-readable explanations. In addition, they noted that creating the XAI user interface is challenging because developers continue to face technological barriers in managing the XAI environment for designers and practitioners in general. Ribera and Lapedriza (2019) undertook a thorough investigation of the current state of explainable AI. They concluded that current research does not pay enough attention to the explanatory audience. They argued that for effective explanations, we need to focus on end-users who have different priorities, expectations, and experiences. Users were divided into three groups: AI developers/researchers, domain specialists, and non-professional users, according to the researchers. They claimed that explanations should be included in the explanation process based on the type of user they are addressing. Developers, academics, specialists, and non-professional users would be more satisfied due to two main factors. First, because the explanations are individualized for the client rather than being generic, and hence it will be less costly to use. Second, the method will be more trustworthy if the explanations are directed to the end users.

Wang et al. (2019) established a theoretical framework for constructing explanatory theories based on theory. To construct user-oriented interpretive AI models, the authors used theory, science, and artificial intelligence for an AI-based model to support the diagnosis of patients in intensive care units. This study aimed to develop and deploy a descriptive clinical diagnostic approach for screening patients in the ICU. The framework establishes a link between human reasoning and XAI capabilities to help identify differences to integrate and develop new explanations while ignoring the requirement for reasoning. Another study that attempted to formulate a description of machine learning models for user-centered displays is Barda, Horvat, and Hochheiser (2020). The authors applied this method to create a user-centered display of interpretation using the Hospital Mortality Risk Prediction Model from the Pediatric Intensive Care Unit (PICU). The authors believe that the proposed paradigm-agnostic explanations at the instance level of functional consequences for health professionals are a potential avenue for clarifying model results and may help in the redesign of explanations. In designing descriptive systems that incorporate broader descriptions, the user-centered framework provides recommendations on how to develop more intuitive and effective explanations.

Machine learning models (ML) have proven that they can enhance human performance in a variety of activities. Machine learning models predict outcomes and are used as a resource or in fully automated systems that do not require human involvement. Humans must trust these judgments while relying on the algorithms' decisions. The interpretability of machine learning models, as well as their determination and behavior, is increasingly a prominent area of study (Doshi-Velez & Kim, 2017; Gilpin et al., 2019). Assessing the validity of interpretable techniques and procedures that cannot be explicitly verified through research is therefore

Journal of Information System and Technology Management
EISSN: 0128-1666
JISTM
Volume 6 Issue 22 (September 2021) PP. 36-50
DOI: 10.35631/JISTM.622004

challenging. Based on specific research (Barda et al., 2020; Ehsan & Riedl, 2019a; Ribera & Lapedriza, 2019), it is critical to focus on the end-user to provide appropriate and credible explanations.

Developing User-Centric Explainable AI

In the development of User-Centric Explainable AI, the goal of interpretable machine learning approaches is to improve the structure in terms of both explanation and prediction. Non-technical users can use any architecture without technical knowledge. To identify and create explanations in a structure, an explanation design technique can be used. By using end-user evaluations to measure expected outcomes, researchers have been able to improve the interpretation of computer vision and general-purpose deep learning algorithms. Studies have been conducted to increase the interpretability of deep learning algorithms (Table 1).

Schmidt and Biessmann (2019) presented a set of metrics that can be used as a single test to determine the relevance of a definition. Kim et al. (2018) have created a system that focuses on the user-friendly linear representation of the internal state of the studied model.

Chen et al. (2018) created an algorithm that examines sketches, finds prototypes and integrates data from prototypes to arrive at the final categorization. Kim et al. (2019) presented a low-latency deep learning system for closer analysis. Loepp et al. (2019) developed a methodology for assessing the effectiveness of user research. Papernot and McDaniel (2018) created 'Deep K-Neighbors Neighbors', a predictive deflection measurement algorithm. Bassen et al. (2020) suggested an algorithm to improve the learning experience in online courses. An in-depth investigation employed multiple sorts of conceptualization, and brilliant questions were utilized to validate questions of worth in a recent study by Chari et al. (2020).

Table 1: XAI Tools

Literature	Visual representation	XAI algorithm example
Zhou et al, 2015.	Data instances	CAM
Lundberg et al, 2017	Saliency map / Bar chart	Sharp
Kim et al, 2018	Bar chart	TCAV
Schmidt and Biessmann, 2019	Data instances	LIME
Tan et al., 2018	Line plot	GAM
Apley and Zhu, 2019	2D or 3D heatmap	ALE
Papernot and McDaniel, 2018	Data instances / Line plot	Nearest neigbour
Chen et al., 2019	Data instances	CNN
Laugel et al., 2017	Data instances / Line plot	Inverse classification
Yang et al., 2017	Data instances / Line plot	Bayesian rule
Guidotti et al., 2018	2D or 3D heatmap	LORE

Challenges And Opportunities

Today's life is increasingly shaped by data-driven decisions, often made by artificial intelligence systems and machine learning algorithms. For example, AI shows great promise in the areas of diagnostics and imaging within medical care, but how can we rely on a black-box model to tell us how to treat a patient? This question has sparked great interest among

Journal of Information System and Technology Management
EISSN: 0128-1666
JISTM
Volume 6 Issue 22 (September 2021) PP. 36-50
DOI: 10.35631/JISTM.622004
researchers working to improve our understanding of AI-based systems. These systems have the potential to improve human well-being in many ways. Explainable AI systems in general present enormous challenges, but also great opportunities in a variety of domains.

The results of the literature review show that social and economic variables can affect system performance, even when the deep learning system is performing basic tasks. However, they operate autonomously as "black boxes" that are not designed to communicate transparently with end-users (Liu et al., 2017; Ribeiro et al., 2016). In complex cases, users may have difficulty understanding the behavior of such systems, which can lead to mistrust and misuse. It is important to consider their cost-benefit ratio and understand in which cases clarification is necessary and useful. To address some of these issues, researchers recommend focusing on transparency and interpretation (Amershi, 2011; Amershi et al., 2014; Ribeiro et al., 2016).

On the other hand, the advantage of Explainable AI is that the system offers the necessary information to support the results. This is especially true in the case of sudden judgments. Moreover, it increases trust by ensuring that the algorithm's judgments are fair, ethical, and transparent. Forcing all AI systems to explain all actions could limit design options, limit system efficiency, and encourage explainable but ineffective or incompetent systems. The results also suggest that many environmental variables that negatively affect model performance in the real world have the potential to be greatly reduced or avoided by tactical measures. The results emphasize the need to focus on end-users to provide effective and trustworthy explanations (Barda et al., 2020; Ehsan & Riedl, 2019b; Ribera & Lapedriza, 2019). The XAI system provides the necessary information to support the results. This is especially true when unexpected decisions are made. It also increases trust by guaranteeing that algorithmic conclusions are fair and ethical in a traceable and verifiable manner. Because machine learning, especially deep learning, has significant potential to advance the field of medicine and improve treatment, the HCI community must therefore create opportunities for building and evaluating machine learning systems.

Conclusion

Researchers argue that deep learning strategies can lead to significant improvements (Davenport & Kalakota, 2019; Gómez-González et al., 2020; Jiang et al., 2017). However, the inability of human users to reason about their decisions in the challenging circumstances of these systems limits their effectiveness. Due to a lack of explanations, traditional deep learning models can be considered as black boxes, which makes it difficult to integrate them into a real-world environment. Despite rapid advances in AI/ML in various application domains, there has been significantly less progress in understanding how users interact with ML /AI systems. To narrow the gap between user requirements and XAI algorithms for transparency, researchers advocate a user-centric approach to interpretive and interdisciplinary collaboration.

This new research area focuses on empirical studies of how the use or description of human interpretive social structures affects the user experience with AI. Our research demonstrates that the AI community and industry practitioners and educators can collaborate to significantly advance the XAI domain through translational work and a shared repository that maps XAI technology solutions. Despite significant developments in AI/ML in a variety of application domains, there are still far fewer challenges in understanding how humans interact with ML /AI systems through interpretive and transdisciplinary collaboration (Murdoch et al., 2019).

Journal of Information System and Technology Management
EISSN: 0128-1666
JISTM
Volume 6 Issue 22 (September 2021) PP. 36-50
DOI: 10.35631/JISTM.622004

This new area of research focuses on empirical studies of how the use or description of human interpretive social structures influences the user experience with AI. Our research shows that the AI community, as well as industry professionals and educators, can collaborate to significantly advance the XAI domain through translation work and the customer exchange repository that maps XAI technology solutions.

Acknowledgment

This project was funded by the National University of Malaysia UKM, under grant number grant no.: TAP-K007341-UKM. The authors, therefore, acknowledge with thanks, UKM for their financial support.

References

Abdul, A., Vermeulen, J., Wang, D., Lim, B. Y., & Kankanhalli, M. (2018). Trends and trajectories for explainable, accountable and intelligible systems: An HCI research agenda. *Conference on Human Factors in Computing Systems - Proceedings*, *2018-April*, 1–18. https://doi.org/10.1145/3173574.3174156

Ahmad, M. A., Teredesai, A., & Eckert, C. (2018). Interpretable machine learning in healthcare. *Proceedings - 2018 IEEE International Conference on Healthcare Informatics, ICHI 2018*, 447. https://doi.org/10.1109/ICHI.2018.00095

Amershi, S. (2011). Designing for effective end-user interaction with machine learning. *UIST'11 Adjunct - Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, 47–50. https://doi.org/10.1145/2046396.2046416

Amershi, S., Cakmak, M., Knox, W. B., & Kulesza, T. (2014). Power to the people: The role of humans in interactive machine learning. *AI Magazine*, *35*(4), 105–120. https://doi.org/10.1609/aimag.v35i4.2513

Arya, V., Bellamy, R. K. E., Chen, P. Y., Dhurandhar, A., Hind, M., Hoffman, S. C., Houde, S., Vera Liao, Q., Luss, R., Mojsilović, A., Mourad, S., Pedemonte, P., Raghavendra, R., Richards, J., Sattigeri, P., Shanmugam, K., Singh, M., Varshney, K. R., Wei, D., & Zhang, Y. (2019). One explanation does not fit all: A toolkit and taxonomy of ai explainability techniques. *ArXiv*.

Barda, A. J., Horvat, C. M., & Hochheiser, H. (2020). A qualitative research framework for the design of user-centered displays of explanations for machine learning model predictions in healthcare. *BMC Medical Informatics and Decision Making*, *20*(1), 1–16. https://doi.org/10.1186/s12911-020-01276-x

Bassen, J., Balaji, B., Schaarschmidt, M., Thille, C., Painter, J., Zimmaro, D., Games, A., Fast, E., & Mitchell, J. C. (2020). Reinforcement Learning for the Adaptive Scheduling of Educational Activities. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3313831.3376518

Beede, E., Baylor, E., Hersch, F., Iurchenko, A., Wilcox, L., Ruamviboonsuk, P., & Vardoulakis, L. M. (2020). A Human-Centered Evaluation of a Deep Learning System Deployed in Clinics for the Detection of Diabetic Retinopathy. *Conference on Human Factors in Computing Systems - Proceedings*, 1–12. https://doi.org/10.1145/3313831.3376718

Biran, O., & Cotton, C. (2017). Explanation and Justification in Machine Learning: A Survey. *IJCAI-17 Workshop on Explainable AI (XAI)*, 8–13. http://www.cs.columbia.edu/~orb/papers/xai_survey_paper_2017.pdf

Journal of Information System and Technology Management
EISSN: 0128-1666
JISTM
Volume 6 Issue 22 (September 2021) PP. 36-50
DOI: 10.35631/JISTM.622004

Cagiltay, K., Baek, E., Bernardino, S., Boling, E., & Frick, T. W. (2008). *User-centered design and development*. *January*. https://doi.org/10.4324/9780203880869.ch49

Cai, C. J., Winter, S., Steiner, D., Wilcox, L., & Terry, M. (2019). "Hello Ai": Uncovering the onboarding needs of medical practitioners for human–AI collaborative decision-making. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW). https://doi.org/10.1145/3359206

Chari, S., Seneviratne, O., Gruen, D. M., Foreman, M. A., Das, A. K., & McGuinness, D. L. (2020). Explanation Ontology: A Model of Explanations for User-Centered AI. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12507 LNCS*(Ml), 228–243. https://doi.org/10.1007/978-3-030-62466-8_15

Chen, C., Li, O., Tao, C., Barnett, A. J., Su, J., & Rudin, C. (2018). This looks like that: Deep learning for interpretable image recognition. *ArXiv*, *NeurIPS*, 1–12.

Chen, J. H., & Asch, S. M. (2017). Machine Learning and Prediction in Medicine — Beyond the Peak of Inflated Expectations. *New England Journal of Medicine*, *376*(26), 2507–2509. https://doi.org/10.1056/nejmp1702071

Dabbs, A. D. V., Myers, B. A., Mc Curry, K. R., Dunbar-Jacob, J., Hawkins, R. P., Begey, A., & Dew, M. A. (2009). User-centered design and interactive health technologies for patientsi. *CIN - Computers Informatics Nursing*, *27*(3), 175–183. https://doi.org/10.1097/NCN.0b013e31819f7c7c

Davenport, T., & Ravi Kalakota. (2019). The Potential for Artificial Intelligence in Healthcare. *Future Healthcare Journal*, *6*(2), 94–98.

Doshi-Velez, F., & Kim, B. (2017). *Towards A Rigorous Science of Interpretable Machine Learning*. *Ml*, 1–13.

Ehsan, U., & Riedl, M. O. (2019a). *On Design and Evaluation of Human-centered Explainable AI systems On Design & Evaluation of Human-centered XAI systems*.

Ehsan, U., & Riedl, M. O. (2019b). *On Design and Evaluation of Human-centered Explainable AI systems On Design & Evaluation of Human-centered XAI systems*. *April*.

Farao, J., Malila, B., Conrad, N., Mutsvangwa, T., Rangaka, M. X., & Douglas, T. S. (2020). A user-centred design framework for mHealth. *PLoS ONE*, *15*(8 August), 1–18. https://doi.org/10.1371/journal.pone.0237910

Freitas, A. A. (2014). Comprehensible classification models: a position paper. *ACM SIGKDD Explorations Newsletter*, *15*(1), 1–10.

Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2019). Explaining explanations: An overview of interpretability of machine learning. *Proceedings - 2018 IEEE 5th International Conference on Data Science and Advanced Analytics, DSAA 2018*, 80–89. https://doi.org/10.1109/DSAA.2018.00018

Gómez-González, E., Gomez, E., Márquez-Rivas, J., Guerrero-Claro, M., Fernández-Lizaranzu, I., Relimpio-López, M. I., Dorado, M. E., Mayorga-Buiza, M. J., Izquierdo-Ayuso, G., & Capitán-Morales, L. (2020). Artificial intelligence in medicine and healthcare: A review and classification of current and near-future applications and their ethical and social Impact. *ArXiv*.

Good, A., & Omisade, O. (2019). Linking Activity Theory with User Centred Design: A Human Computer Interaction Framework for the Design and Evaluation of mHealth Interventions. *Studies in Health Technology and Informatics*, *263*, 49–63. https://doi.org/10.3233/SHTI190110

Journal of Information System and Technology Management
EISSN: 0128-1666
JISTM
Volume 6 Issue 22 (September 2021) PP. 36-50
DOI: 10.35631/JISTM.622004

Goodman, B., & Flaxman, S. (2017). European union regulations on algorithmic decision making and a "right to explanation." *AI Magazine*, *38*(3), 50–57. https://doi.org/10.1609/aimag.v38i3.2741

Gunning, D., & Aha, D. W. (2019). DARPA's explainable artificial intelligence program. *AI Magazine*, *40*(2), 44–58. https://doi.org/10.1609/aimag.v40i2.2850

Gupta, A. (2019). *StrokeSave: A Novel, High-Performance Mobile Application for Stroke Diagnosis using Deep Learning and Computer Vision*.

Holzinger, A., Biemann, C., Pattichis, C. S., & Kell, D. B. (2017). What do we need to build explainable AI systems for the medical domain? *ArXiv*, *Ml*, 1–28.

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology*, *2*(4), 230–243. https://doi.org/10.1136/svn-2017-000101

Kim, B., Wattenberg, M., Gilmer, J., Cai, C., Wexler, J., Viegas, F., & Sayres, R. (2018). Interpretability beyond feature attribution: Quantitative Testing with Concept Activation Vectors (TCAV). *35th International Conference on Machine Learning, ICML 2018*, *6*, 4186–4195.

Kim, G. J. (2018). *Human–Computer Interaction Fundamentals and Practice Click here to order Human–Computer Interaction: Fundamentals and Practice*. http://www.ittoday.info/Excerpts/HCI.pdf

Kim, J., Stengel, M., Majercik, A., De Mello, S., Dunn, D., Laine, S., McGuire, M., & Luebke, D. (2019). NVGaze: An anatomically-informed dataset for low-latency, near-eye gaze estimation. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3290605.3300780

Krening, S., & Feigh, K. M. (2019). Effect of interaction design on the human experience with interactive reinforcement learning. *DIS 2019 - Proceedings of the 2019 ACM Designing Interactive Systems Conference*, 1089–1100. https://doi.org/10.1145/3322276.3322379

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Liao, Q. V., Gruen, D., & Miller, S. (2020). Questioning the AI: Informing Design Practices for Explainable AI User Experiences. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3313831.3376590

Liu, M., Shi, J., Li, Z., Li, C., Zhu, J., & Liu, S. (2017). Towards Better Analysis of Deep Convolutional Neural Networks. *IEEE Transactions on Visualization and Computer Graphics*, *23*(1), 91–100. https://doi.org/10.1109/TVCG.2016.2598831

Loepp, B., Donkers, T., Kleemann, T., & Ziegler, J. (2019). Impact of consuming suggested items on the assessment of recommendations in user studies on recommender systems. *IJCAI International Joint Conference on Artificial Intelligence*, *2019-Augus*, 6201–6205. https://doi.org/10.24963/ijcai.2019/863

Lysaght, T., Lim, H. Y., Xafis, V., & Ngiam, K. Y. (2019). AI-Assisted Decision-making in Healthcare. *Asian Bioethics Review*, *11*(3), 299–314. https://doi.org/10.1007/s41649-019-00096-0

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1–38. https://doi.org/10.1016/j.artint.2018.07.007

Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., & Yu, B. (2019). Definitions, methods, and applications in interpretable machine learning. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(44), 22071–22080. https://doi.org/10.1073/pnas.1900654116

Journal of Information System and Technology Management
EISSN: 0128-1666
JISTM
Volume 6 Issue 22 (September 2021) PP. 36-50
DOI: 10.35631/JISTM.622004

Papernot, N., & McDaniel, P. (2018). Deep k-nearest neighbors: Towards confident, interpretable and robust deep learning. *ArXiv*, *c*.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should i trust you?" Explaining the predictions of any classifier. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *13-17-Augu*, 1135–1144. https://doi.org/10.1145/2939672.2939778

Ribera, M., & Lapedriza, A. (2019). Can we do better explanations? A proposal of user-centered explainable AI. *CEUR Workshop Proceedings*, *2327*(February).

Riveiro, M., & Thill, S. (2021). "That's (not) the output I expected!" On the role of end user expectations in creating explanations of AI systems. *Artificial Intelligence*, *298*, 103507. https://doi.org/10.1016/j.artint.2021.103507

Schmidt, P., & Biessmann, F. (2019). Quantifying interpretability and trust in machine learning systems. *ArXiv*.

Smith-Renner, A., Fan, R., Birchfield, M., Wu, T., Boyd-Graber, J., Weld, D. S., & Findlater, L. (2020). No Explainability without Accountability: An Empirical Study of Explanations and Feedback in Interactive ML. *Conference on Human Factors in Computing Systems - Proceedings*, 1–13. https://doi.org/10.1145/3313831.3376624

Smith, J., Legg, P., Matovic, M., & Kinsey, K. (2018). Predicting user confidence during visual decision making. *ACM Transactions on Interactive Intelligent Systems*, *8*(2), 1–34. https://doi.org/10.1145/3185524

Souillard-Mandar, W., Davis, R., Rudin, C., Au, R., Libon, D. J., Swenson, R., Price, C. C., Lamar, M., & Penney, D. L. (2016). Learning classification models of cognitive conditions from subtle behaviors in the digital Clock Drawing Test. *Machine Learning*, *102*(3), 393–441. https://doi.org/10.1007/s10994-015-5529-5

Teixeira, A., Ferreira, F., Almeida, N., Silva, S., Rosa, A. F., Pereira, J. C., & Vieira, D. (2017). Design and development of Medication Assistant: older adults centred design to go beyond simple medication reminders. *Universal Access in the Information Society*, *16*(3), 545–560. https://doi.org/10.1007/s10209-016-0487-7

Wang, D., Yang, Q., Abdul, A., Lim, B. Y., & States, U. (2019). *Designing Theory-Driven User-Centric Explainable AI*. 1–15.

West, D. M. (2018). The future of work: Robots, AI, and automation. *The Future of Work: Robots, AI, and Automation*, 1–205.

Wiens, J., & Shenoy, E. S. (2018). Machine Learning for Healthcare: On the Verge of a Major Shift in Healthcare Epidemiology. *Clinical Infectious Diseases*, *66*(1), 149–153. https://doi.org/10.1093/cid/cix731

Xie, J., Richard Yu, F., Huang, T., Xie, R., Liu, J., Wang, C., & Liu, Y. (2019). A survey of machine learning techniques applied to software defined networking (SDN): Research issues and challenges. *IEEE Communications Surveys and Tutorials*, *21*(1), 393–430. https://doi.org/10.1109/COMST.2018.2866942

Yan, Y., Zhang, J. W., Zang, G. Y., & Pu, J. (2019). The primary use of artificial intelligence in cardiovascular diseases: What kind of potential role does artificial intelligence play in future medicine? *Journal of Geriatric Cardiology*, *16*(8), 585–591. https://doi.org/10.11909/j.issn.1671-5411.2019.08.010

Yu, K. H., Zhang, C., Berry, G. J., Altman, R. B., Ré, C., Rubin, D. L., & Snyder, M. (2016). Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nature Communications*, *7*, 1–10. https://doi.org/10.1038/ncomms12474

Journal of Information System and Technology Management
EISSN: 0128-1666
JISTM
Volume 6 Issue 22 (September 2021) PP. 36-50
DOI: 10.35631/JISTM.622004

Zhu, J., Liapis, A., Risi, S., Bidarra, R., & Youngblood, G. M. (2018). Explainable AI for Designers: A Human-Centered Perspective on Mixed-Initiative Co-Creation. *IEEE Conference on Computatonal Intelligence and Games, CIG*, *2018-Augus*, 1–8. https://doi.org/10.1109/CIG.2018.8490433