Explainable artificial intelligence and its practical applications

Yijun Liu

Jiangxi's Agriculture University, Nanchang, China, 330000

yijun-liu02@163.com

Abstract. With the continuous development of the times, the artificial intelligence industry is also booming, and its presence in various fields has a huge role in promoting social progress and advancing industrial development. Research on it is also in full swing. People are eager to understand the cause-and-effect relationship between the actions performed or the strategies decided based on the black-box model, so that they can learn or judge from another perspective. Thus the Explainable AI is proposed, it is a new generation of AI that allows humans to understand the cause and give them a decision solution, so that every outcome has its own basis for decision. Although some considerable results have been achieved in the application of explainable AI, it is still at the beginning stage and there are still some challenges to be solved. From the transportation industry, which facilitates people's access to autonomous driving, to the medical industry, which saves people's lives, to the financial industry, which is a huge industry, and even in the education industry, which is accessible to all people, it has a presence. This paper talks about the current situation and problems of explainable AI through its application in various aspects. Explainable AI can serve not only developers but also users by satisfying their interest-related needs. The transparency of explainable AI is important when it is used in socially relevant applications, which is why we have conducted extensive research on explainable AI.

Keywords: XAI's theory, XAI's applications, Explainable Artificial Intelligence, Explainability, Interpretability, Explanations.

1. Introduction

XAI is an emerging branch of artificial intelligence that provides the appropriate reasons for decisions made by artificial intelligence so that they can be better understood by humans [1]. It improves transparency, reliability, causality, stability, security, and universality on the basis of ordinary artificial intelligence. The goal is to build a bridge between humans and machines so that their users can trust the decisions they make. In contrast to ordinary AI, the decisions made by explainable AI can be understood not only by computers but also by humans. While the black box of artificial intelligence cannot be understood even by coders as to why it makes decisions, explainable AI can provide explanations so that the humans who use it feel more confident that the decisions given are trustworthy and thus use them [2].

From the transportation industry that facilitates autonomous driving to the medical industry that saves the people and benefits the people, and the financial industry that has a huge industry, it is all there. In the current context of explainable AI, the goal and scope of data collection will become clearer, the design of decision solutions will follow a new explainable logic, the selection of decision solutions will have a new explainable basis, and the implementation and evaluation of decision solutions will become

more transparent. Therefore, it is important to address how the decision-making paradigm will affect the user's use of AI technologies under the influence of explainable AI, and how the management decision problems associated with each domain will change, which are questions that explainable AI can address [3]. Nowadays, every field of life, be it education, finance or healthcare, is constantly moving forward based on data. Artificial intelligence ensures that data is analyzed with a high degree of accuracy, thus continuously enhancing these fields. The role of AI is rapidly increasing in areas such as solving critical tasks and providing good judgment, and human mastery of AI is gradually expanding. However, AI is unable to explain to the end user the reasons behind the process that leads to its output after its input. This is not only a matter of trust, but also raises questions about the fairness of AI as well as its security. Therefore, it is very necessary to make AI more reliable. With the introduction of XAI, humans may be able to grasp the cause and effect relationship between for behind the results.

The concept of explainable AI has been a strong driving force in all aspects of society, and its demand is gradually increasing in both academia and industry. This paper focuses on the emergence of explainable AI, summarizes the current status of explainable AI in life and its dynamics from several aspects, and analyzes the shortcomings of the current applications of explainable AI in life. The paper will contribute to the development of explainable AI in various fields. It also provides several directions for the research of explainable AI.

2. Analysis of XAI's application in various aspects of life

2.1. The current status of XAI in medical applications

In modern medicine, physicians rely heavily on information provided by data for disease analysis. Artificial intelligence can be considered somewhat stronger than humans for data analysis, and AI can assist physicians in diagnosis and can provide clinicians with supporting clinical decisions. For example, current cancer detection can detect tumor-infiltrating lymphocytes and cancer cells in histological images, provide exact heat map visualization to explain classifier decisions, and explainable AI can be used by assessing the association between morphological and molecular cancer features [4]. Interpretable AI can also synthesize diagnostic scores from clinical cases, histological as well as molecular features, and thus facilitate basic cancer research as well as medical precision. XAI can also be used to analyze Alzheimer's disease, using images of the human brain at various stages as input, and using interpretable AI to identify the stages of Alzheimer's disease in a layer-by-layer step-by-step manner. This project will enable the detection of Alzheimer's disease in its early stages, thus helping to prevent and treat more patients [5].

2.2. Current status of XAI applications in finance

With the explosive development of artificial intelligence in the past few years, AI is playing an increasingly significant role in finance. There are many aspects of the role in the financial sector, including analysis of transaction data, verification of customer eligibility, detection of financial fraud, application of anti-money laundering, risk control, automatic analysis of legal provisions, etc., providing a comprehensive range of decision aids [6]. Since the competition in the financial industry is also very intense, any decision is subject to certain risks and losses are uncertain. Therefore, the choice of artificial intelligence is also very careful, through the analysis of big data, the use of different formulas, different financial products also need different perspectives to observe to understand, and finally combined with a variety of factors to get the best solution, the reasons for which must also be very clear, so that can explain the importance of artificial intelligence in the financial level is also evident.

2.3. The current status of XAI in education

Explanation is a crucial part of the education industry, and the explanation of things is fundamental to human education. It is also what human cognition has been following for education. Explainable AI stands out by the word "explanation", so explainable AI can help the learning process at the educational level and achieve efficient knowledge transfer. Explainable AI generates educational decisions that can

be presented in a scientific and human-understandable way, allowing educators to expand the scope of education through such decisions and to know what they know and why they know it. Some examples of explainable AI that are already being promoted are: intelligent tutor systems, explainable educational recommendation systems, and explainable learning analytics systems, to name a few. Explainable educational AI can create an educational environment where things are done for a reason, fair and reasonable, respectful and equal, and technology for good [7].

2.4. The current status of XAI in language and culture

Explainable AI is also used in language translation, both for small languages and for cultural differences in different languages, which can be understood by explainable AI. Currently, explainable AI is also used in the recognition of hate speech, whether for gender, ethnicity and other verbal attacks in cyberspace is also often occurring, we can use interpretable AI to analyze the user's speech, so as to achieve the effect of screening. Of course, we do not use it only to screen hate speech, but sometimes we also use it for misunderstood words, eliminating the possibility of them being judged as offensive by contextualizing them to arrive at the right decision [8]. By using interpretable AI in hate speech recognition it is possible to better restrain its proliferation and control its development more easily. The use of explainable AI in this direction also allows regulators to better understand the reasons why these statements are judged as hate speech and thus better understand the meaning of these discourses, thus also strengthening their own ability to screen the hate speech. To support the stability of society.

2.5. Current status of XAI applications in transportation

In the evolving transportation industry, from the previous manual autonomous driving to the current emerging autonomous driving technology, are derived from the direction of artificial intelligence. In some specific occasions autonomous driving technology has replaced manual autonomous driving, such as specific streets, fixed routes have the emergence of autonomous driving figures. However, in some more complex traffic conditions, artificial intelligence has decision risks and safety risks [9]. The emergence of explainable AI can make up for this shortcoming as much as possible by providing reasons for each decision, making it easier for users to understand, trust, and manage the process of human-vehicle interaction, whether it is to investigate the reasons after an accident or to make judgments based on the explanations provided during driving. This greatly enhances the safety of autonomous driving and improves the transparency and trust of users.

3. Shortcomings of XAI in existing applications

Since everyone's understanding is different, it is a very subjective personal perception, so there is not a relatively comprehensive scientific evaluation system for explainable AI research, and developers need to be more in line with the user's evaluation of the product, so that a new perspective to improve the interpretability of the shortcomings, but if all based on the user's perspective to complete the interpretability has a strong subjective However, if the interpretability is done from the user's perspective, it will be very subjective, and if the interpretability is measured from the developer's perspective, it will be very theoretical. There is a certain contradiction between the two, and the current explainable AI does not balance the two well enough to achieve a win-win situation.

A good explainable AI is not only the study of data, but also the interaction between the machine and the user, which means that the user can have a channel to adjust, refute, and give feedback after receiving a reasonable explanation, so as to improve the quality of interpretability. As various fields are trying to use explainable AI, due to the different audiences in each field, developers should set up interpretable systems with different groups and specific scenarios, and provide specific interpretations to achieve personalized interpretations. A communication bridge needs to be established between users and explainable AI systems so that explainable AI can be applied in more domains [10].

The current explainable AI mainly adopts the post hoc explanation method, but the post hoc explanation is the analysis and explanation of the known result and does not reflect the actual logic of the decision and the reasons for its execution, in fact, if the different decisions and reasons can be

provided before the occurrence of the event and choose the better one to execute or provide the choice to the user to assist them in the decision, which is also called This is also known as the predictive power, which is a strong advantage for explainable AI with strong predictive power, thus increasing its accuracy [11]. Of course, these are not the only explanation methods, but in the future, we can try to make the best use of the advantages of various explanation methods, or by studying the similarities between each method, we can combine them and use them to take the essence and remove the dross. The focus is on improving the accuracy, reliability, safety, fairness, transparency, causality, and universality of explainable AI.

As the development of explainable AI becomes closer to human, other ethical issues will emerge. Replicating human behavior does not solve these moral and ethical issues, which requires developers to build an ethical standard and system framework to ensure the healthy development of explainable AI.

4. Conclusion

The design and development of explainable AI can integrate more knowledge in human culture, physics, mathematics and other fields, which can expand psychological models, social development research models and so on. High financial technology, intelligent medical care, modern logistics, e-commerce era, new retail, etc. are all current hot areas, and the decision making problems in them are to be explored by the developers of explainable AI. The decision management model supported by explainable AI, intelligent and convenient model, knowledge management model, data analysis model, information management model, etc. all show new opportunities.

Both academia, industry and government are very concerned about the development of explainable AI. In academia, explainable AI is a hot research topic in many fields, but it is still in the initial stage, so the development of explainable AI in academia will be deepened step by step. In the industry, the demand for explainable AI is also growing, so for the industry, explainable AI needs to be developed and used more closely to meet the needs of users and the interests of enterprises. For government departments, explainable AI brings reliability and transparency, and its decisions can be made to obtain the reasons for the decision, so as to effectively respond to people's concerns and questions, so in the explainable AI in the social field tend to be more humane, rationalization.

This paper summarizes some of the applications of explainable AI in society, analyzes some shortcomings and defects in the overall applications, discusses the concept of explainable AI and the necessity of its development, points out the key development directions of explainable AI in various fields, analyzes the challenges to be encountered in the development of explainable AI, and provides some pertinent basis and reference for the subsequent development of explainable AI. It provides some pertinent basis and reference for the subsequent development of explainable AI, and promotes the research and development of explainable AI in the future.

In this study, since the research on explainable AI is still in its initial stage, the examples given in this paper are applied in a shallow manner. Compared to all the fields where explainable AI has been applied, this paper only selects examples in the fields of finance, healthcare, education, transportation, and language. Only some representative applications in these five areas are selected for in-depth discussion and analysis. For the subsequent research on explainable AI, both application areas and theoretical concepts need to be studied in more depth and detail.

References

[1] Dr G.R. Karpagam, Aditya Varma, Samrddhi M. Understanding, Visualizing and Explaining XAI Through Case Studies, 2022 8th International Conference on Advanced Computing and Communication Systems, 2022.

[2] David Gunning, Mark Stefik, Jaesik Choi. XAI-Explainable artificial intelligence [J]. SCIENCE ROBOTICS, 18 Dec 2019, Vol 4, Issue 37. DOI: 10.1126/scirobotics.aay7120.

[3] Nakul Tanwar, Yasha Hasija. Explainable AI; Are we there yet? 2022 IEEE Delhi Section Conference, 2022.

[4] Alexander Binder, Michael Bockmayr, Miriam Hägele. Morphological and molecular breast

cancer profiling through explainable machine learning [J]. Nature Machine Intelligence, 2021.

[5] K. Muthamil Sudar; P. Nagaraj; S. Nithisaa; R. Aishwarya. Alzheimer's Disease Analysis using Explainable Artificial Intelligence (XAI). 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), 2022.

[6] Jurgita Černevičienė, Audrius Kabašinskas. review of Multi-Criteria Decision-Making Methods in Finance Using Explainable Artificial Intelligence [J]. Frontiers in artificial intelligence, 10 March 2022, doi: 10.3389/frai.2022.827584.

[7] Wang P., Tian S-Y, Qiaoyu Sun. Explainable educational artificial intelligence research: system framework, application value and case study [J]. Journal of Distance Education, Vol.6, 2021.

[8] Mastromattei M., Ranaldi L., Fallucchi F., Zanzotto F. M. Syntax and prejudice: ethically-charged biases of a syntax-based hate speech recognizer unveiled. Peer J Computer Science 8:e859, 2022. https://doi.org/10.7717/peerj-cs.859.

[9] W. W. Guo, Q. Wang. Explainable Interaction in Human-Autonomous Vehicle Interaction [J]. Packaging Engineering, Vol. 18, 2020.

[10] Wu D., Sun G.Y. Towards Explainable Interactive Artificial Intelligence: Motivations, Approaches, and Research Trends [J]. Journal of Wuhan University (Philosophy and Social Science Edition), 2021, (No. 5).

[11] F. K. Došilović, M. Brčić and N. Hlupić, Explainable artificial intelligence: a survey, 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2018, pp. 0210-0215. doi: 10.23919/MIPRO.2018.8400040.