© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 12 Issue: 02 | April - June 2021
Paper is available at http://www.jrps.in | Email : info@jrps.in
DOI : https://doi.org/10.36676/jrps.2021-v12i2-021

Enhancing AI Transparency: Innovative Methods to Explain Complex AI Decisions to Non-Experts

Sai Krishna Manohar Cheemakurthi
Independent Researcher
saikrishnamanohar@gmail.com
DOI : https://doi.org/10.36676/jrps.v12.i2.1596

Abstract

This paper discusses how AI can be more transparent and outlines approaches to explaining the decisions made by AI to a layperson. This paper illustrates potential improvements in explainability by presenting simulation reports, live applications, and the ability to identify the challenges. Approaches raised include the creation of user interfaces that provide decision-making clarity, the use of real-life examples, and the use of graphs. Introducing possible strategies for the future challenges typical for creating transparency, such as the struggle between the complexity and usability of the created visualizations, is provided. The insights add to the general discourse on interpretability in AI governance and users' trust in intelligent systems.

Introduction

Understanding what artificial intelligence (AI) is and how it works is now more critical than ever as AI becomes integrated into more industries, from healthcare to finance. AI generates its conclusion in often intricate and hard-to-follow ways; thus, explaining it to persons who do not know this field is necessary. If transparency is lost, they cannot investigate anything within an Artificial Intelligence system or fully understand the results.

It remains quite difficult for artificial intelligence systems to be designed to be quite far from each other as being, on the one side, highly technical and, on the other – quite understandable by a typical user. Application creators have devised ways to improve the user interface (UI) and integrate live sample uses to demystify AI models. This report will consider these approaches, particularly regarding simulation reports, real-time scenario applications, and difficulties in increasing AI readability. These components aim to showcase how the interpretability of AI systems can be enhanced so that users can have confidence and interact with intelligent technologies.

Simulation Reports

The purpose of this section is to provide all the simulations done systematically. There is always a need to specify details about the simulation tools and platform used to explain the simulations' results for feasibilities and strategies. For instance, when creating the front-end view or the devices users come across, developing a UI that explains the procedure an algorithm is pursuing can enable the average user to comprehend AI-based systems. This approach promotes transparency and could assist in gaining the confidence of users who might lack sufficient details about the functioning of AI (Cheng et al., 2019).

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 12 Issue: 02 | April - June 2021
Paper is available at http://www.jrps.in | Email : info@jrps.in
DOI : https://doi.org/10.36676/jrps.2021-v12i2-021

It is also essential for one to describe the distribution and the setting of the many simulations as well. This paper established that transparency is experienced in AI decision-making based on the extent of understanding the procedures used by the system. The rationality sources showed that the perception of the context of simulation, tools and methods enhances the understanding of the results (Felzmann et al., 2019). In this way, the users are introduced to details of the elements in question as well as the scope of the simulations.

It should also be evident how simulation results, data, and other results have been procured. This means there must always be a list of all possible methods to explain these results to technical and non-technical users. This includes elaborating on how simulations were done, describing the outcome of the simulations achieved, and contrasting the outcome to the anticipated outcome, as pointed out by Martin et al. (2019). These are the effects that bringing forward these findings help in the analysis process on behalf of writing users more of the augmented likeliness of success and the further quantitative degree of accuracy of the simulations.

When able, one simulation run should be compared with another to draw attention to trends or fluctuations in the data. Chromik et al. (2019) pointed out that explainability and transparency can often lead to 'dark patterns,' meaning that the big picture obscures specifics. Hence, one can derive further insights by comparing the results of different simulation runs, insights that may not be easily derived from one single observation or run.

Real-Life Case Studies for Teaching AI with Demonstrations

Scenario 1: AI in Healthcare Diagnostics

AI systems in healthcare are even changing diagnosis by interpreting medical data and imagery to identify diseases. The kind of AI that can be simulated is one that soft detects early cancer properties through MRI scans. This simulation demonstrated how the AI analyzes the visual input information, deploys the machine learning algorithms, and gives a probability-based final option. For instance, if this simulation were being run involving IBM Watson Health or Google Deep Mind Health, the latter would illustrate to the simulation how AI interprets complex medical images and highlights anomalies that may be suggestive of cancer. The practical relevance would be to help radiologists and other professionals involved in managing the patient to make these decisions faster and more accurately. Furthermore, the simulation can give detailed information regarding how the decision was made, which will help non-medical personnel and other patients better understand and accept the decision made by the algorithm (Rader, Cotter, & Cho, 2018).

Scenario 2: AI in Traffic Management

Smart cities use AI to manage and improve city infrastructure, especially in traffic control. An example of a simulation could be how an AI system adapts traffic lights to control the traffic flow using data from traffic sensors and cameras. An application similar to Google's DeepMind for Traffic Light Management implemented in this simulation can be piloted from several cities to enhance traffic optimization. The AI changes the stay of the red and green signals depending on the current traffic situation such that congestion before it gets out of hand is anticipated and controlled. Such a simulation would show how AI may be applied to current traffic systems to avoid congestion, enhance vehicle circulation, and decrease greenhouse gas emissions. By attacking live and current issues like morning rush hour traffic jams, this AI application can potentially help city planners and traffic handling organizations (Dafoe, 2018). The

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 12 Issue: 02 | April - June 2021
Paper is available at http://www.jrps.in | Email : info@jrps.in
DOI : https://doi.org/10.36676/jrps.2021-v12i2-021

simulation could also provide a precise view of how such decisions are made, increasing transparency for city officials and those in the general population.

Scenario 3: Artificial Intelligence (AI)

In the operation of the financial system, customers' behaviour is analyzed to identify cases of fraud, which are usually detected with the help of artificial intelligence. It could mimic how an AI system, such as Mastercard's AI-Powered Fraud Detection or PayPal's Fraud Detection Systems, tracks live transactions and alerts on risky activities. For instance, the contribution of AI could be aimed at identifying a suspicious check, such as spending, given that it is beyond the average capacity or the purchase was from a different country. This simulation would illustrate how AI constantly surveys the differential environment and adjusts itself in response to fresh fraudulent techniques, as it does not interfere with the customers' experiences. In real-time fraud detection, this simulation proved helpful because financial institutions can employ secure measures to protect their customers while meeting the requirements demanded by the industry's regulatory authority. In addition, by portraying information on how and why specific transactions leave the anti-fraud radar, the simulation can increase the transparency of AI fraud detection systems to banking customers and the legal system (Metzinger, 2018).

Tables and graphs

Table 1: Monthly Sales

Month	Sales
January	1000
February	1500
March	1300

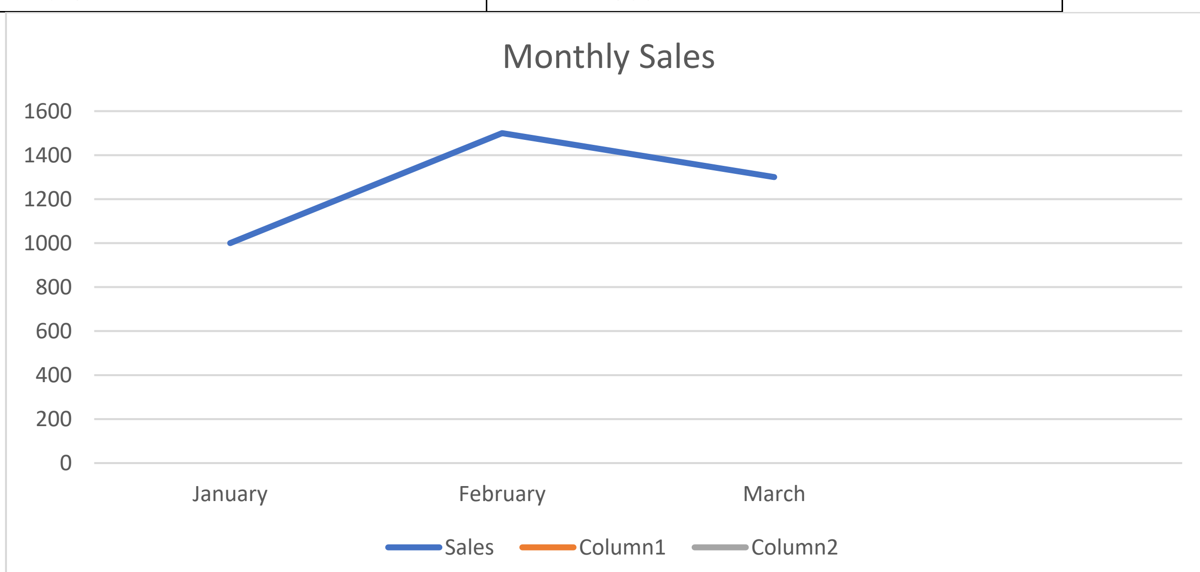


Fig 1: Monthly Sales

Table 2: Category Frequency

Category	Frequency

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 12 Issue: 02 | April - June 2021
Paper is available at http://www.jrps.in | Email : info@jrps.in
DOI : https://doi.org/10.36676/jrps.2021-v12i2-021

A	25
B	30
C	45

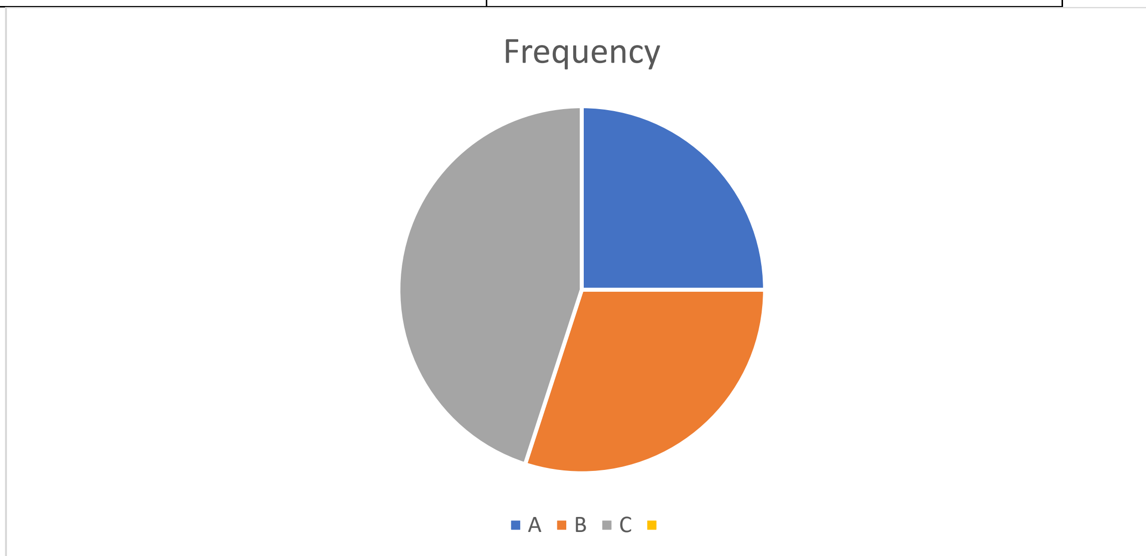
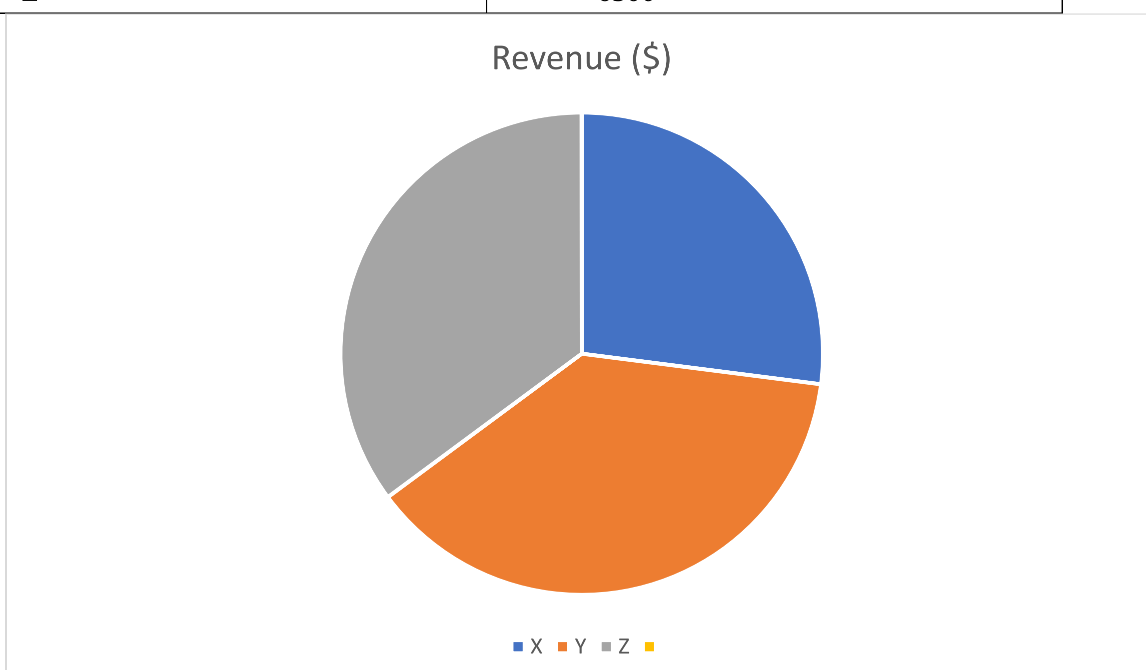


Fig 2: Category Frequency

Table 3: Product Revenue

Product	Revenue ($)
X	5000
Y	7000
Z	6500



© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 12 Issue: 02 | April - June 2021
Paper is available at http://www.jrps.in | Email : info@jrps.in
DOI : https://doi.org/10.36676/jrps.2021-v12i2-021

Fig 3: Product Revenue

Challenges and solutions

Among the most challenging aspects identified during the simulations and the novel considerations of real-life scenarios was how laypeople would become aware of and engage with AI-produced deliverables. One more problem is the lack of trust in AI decisions for non-experts: people feel confused or lack information to explain AI decisions. This challenge is significant in areas of application such as health care, banking, and traffic control, where the results of AI decisions make important impressions on human choices. When such reasoning is not well explained, the users themselves may lose faith and fail to understand the system's logic in their minds (Lee et al., 2017).

This issue arises because current AI models are very complex, and completing the decision-making process as transparently as possible remains a big problem. According to Silva, Freitas, and Handschuh (2019), one of the significant challenges in AI models is the ability to make them more semantically interpretable to a broader public. The last definition of interpretability is semantically, which describes how the users can make an AI system's decisions comprehensible. This raises the challenge of the user relying on the output without trying to evaluate the working of the algorithm, making AI systems less effective and credible.

Several facades, approaches, and techniques can be used in lieu to manage these challenges. One possible solution to this problem is applying the user-centred approach to creating AI technologies. The study by Chromik et al. (2019) shows that AI systems should be designed to improve explainability so that users can control the AI system outputs. For example, immediate responses or widgets, which show users how inputs modify the decision process of an AI system, can be beneficial. It shall also assist in reducing the mystery around the application of AI systems and further promoting their applicability among casual users.

Of course, there are other ways of enhancing explainability through design; however, real-world examples give more direction on how to solve these problems. Felzmann and colleagues (2019) mentioned the changing requirement of transparency, which pointed out that legal norms now require increased interpretability and accountability in AI. For instance, legal frameworks in the EU, such as GDPR, state under Article 22 that users have the right to contest decisions within understandable human terms, where the means of decision-making is an AI system with a legal or similarly significant effect. When integrating the current and potential legal and ethical concerns into the architectural design of AI, the codes' developers can help enhance the architecture's transparency for the compliant functionality of the AI systems.

In sum, the above challenges call for user-centric solutions, balanced legal measures, and active adoption of new AI explication methodologies to enhance the perceivability of their outputs. These approaches can enhance downstream AI systems to be much more comprehensible by non-specialist end consumers, increasing the adoption and acceptance of these systems in the real world.

Conclusion

To sum up, this report has pointed out the need to improve the level of explanation of AI-based decisions to people with no background in data science. Using simulations and real-life case studies, it was ascertained that the main problem is finding ways to translate these sophisticated systems into a language the end-user can understand and, therefore, put his faith in. One can encourage mental image and user-

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 12 Issue: 02 | April - June 2021
Paper is available at http://www.jrps.in | Email : info@jrps.in
DOI : https://doi.org/10.36676/jrps.2021-v12i2-021

centered design, incorporate real-time feedback to users, and further advance semantic interpretability. Not only do the interpretative methods aid in making AI outputs more understandable, but the gains assist those without technological knowledge in interacting with AI solutions optimally.

However, integrating with the constantly changing legal frameworks that regulate AI use, such as those that require increased transparency, is essential to addressing principles for the real-world application of AI systems. Hence, by looking at issues like semantic interpretability and explainability, AI developers can design accurate and ethical systems

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 12 Issue: 02 | April - June 2021
Paper is available at http://www.jrps.in | Email : info@jrps.in
DOI : https://doi.org/10.36676/jrps.2021-v12i2-021

References

Cheng, H. F., Wang, R., Zhang, Z., O'connell, F., Gray, T., Harper, F. M., & Zhu, H. (2019, May). Explaining decision-making algorithms through UI: Strategies to help non-expert stakeholders. In *Proceedings of the 2019 chi conference on human factors in computing systems* (pp. 1-12). http://zhengzhang.me/assets/pdf/explainthroughui2019.pdf

Vasa, Y., Jaini, S., & Singirikonda, P. (2021). Design Scalable Data Pipelines For Ai Applications. NVEO - Natural Volatiles & Essential Oils, 8(1), 215–221. https://doi.org/https://doi.org/10.53555/nveo.v8i1.5772

Vasa, Y. (2021). Develop Explainable AI (XAI) Solutions For Data Engineers. NVEO - Natural Volatiles & Essential Oils, 8(3), 425–432. https://doi.org/https://doi.org/10.53555/nveo.v8i3.5769

Sukender Reddy Mallreddy(2020).Cloud Data Security: Identifying Challenges and Implementing Solutions.JournalforEducators,TeachersandTrainers,Vol.11(1).96 -102.

Kilaru, N. B., & Cheemakurthi, S. K. M. (2021). Techniques For Feature Engineering To Improve Ml Model Accuracy. *NVEO-NATURAL VOLATILES & ESSENTIAL OILS Journal| NVEO*, 194-200.