APEX ACADEMIA PRESS
ELEGANCE ON EVERY RESEARCH

International Journal of Machine Learning, AI & Data Science Evolution
ISSN: 3068-5073/ Open Access / IJMLAIDSE_Y2F1A001
Apex Academia Press / https://doi.org/10.63665/ijmlaidse-y2f1a001

Original Article

User Interface Interpretability and Human-AI Interaction

Barnabas Joel

Ladoke Akintola University of Technology

Abstract

For a successful HAI, the AI systems should be understandable and easy to handle. This is because, gradually, these systems are becoming part of everyday decisions made. This article discusses how UI design and interpretability are tied together in a way that AI applications become more usable, trustworthy, and comprehensible. We demonstrate how AI functionalities should be integrated into those parts of the user interface that make sense: based on clear context, interactivity in feedback, and supportive images that help gain insight. A case study and user testing illustrate that users are much more likely to trust and use AI-generated outcomes when the user interface is more comprehensible. The findings obtained will contribute to the creation of AI systems that are easy to use and serve their purpose. They also contribute to increasing the body of research on explainable AI.

Keywords

Human-AI Interaction, Interpretability, Explainable AI (XAI), User Interface Design, Trust in AI, Human-Centered AI, Visual Explanations, Interactive Systems, Usability, AI Transparency.

Article History

Received: 15.12.2025

Accepted: 02.01.2026

Published: 16.01.2026

1. Introduction

A. Overview of AI-Human Interaction (HAI)

Researchers in HAI aim to make AI systems more explainable, reliable, and user-friendly. As AI becomes more integrated into everyday life tools, such as virtual assistants and medical diagnostic tools, people are no longer passive recipients of AI decisions, but they are now a part of the choices being made. HAI is moving away from autonomous systems to systems that present people with better choices. People and AI should be able to communicate with each other in a clear and natural way that is easy to understand. This will make people happy and encourage more people to use it.

B. The Value of Interpretability in Systems Driven by AI

The best AI systems are those that are easy to use. This is all the truer for sensitive areas of application, such as medical, financial, legal, or autonomous cars. "Interpretability" is the degree by which it is easy to figure out what an AI system does or how it works. Even if the models were right, people may not believe them, or even be able to check their results if they were difficult to understand. When the human and computer converse with each other, the user interface should be easy to understand. Users should be able to learn from the interface about AI in general, ask questions, and even challenge answers through the interface. When the models become more interpretable, people learn more about how AI models actually work.

C. Motivation and Problem Statement

It's still hard to make interfaces that let smart systems and people talk to each other in a way that makes sense and is useful. AI systems are still "black boxes," so we don't know much about how they work. People are less likely to trust, take responsibility, and make quick decisions when the stakes are high because they aren't open. A lot of work has been done on explainable AI (XAI), but most of it is still algorithmically focused and hasn't reached user interfaces yet. The aim of this project is to ensure from the outset that the front end-the interface between users and AI systems-is paramount. It will bring together algorithmic interpretability and good UI design.

D. The Paper's Goals and Contributions

The key aim of the paper is to seek clarity in user interfaces to facilitate better communication between humans and artificial intelligence. The research focuses on the elaboration or testing of a UI framework that embeds model interpretability-e.g., choice explanations, visual explanations, and feedback mechanisms-directly into the user experience. The research reviews prior studies in this domain while identifying main gaps and proposing design recommendations for user-friendly interface design. We show, via a case study or prototype system, the influence of design choices on usability, trust, and overall user satisfaction when engaging with AI-based applications.

2. Related Work

A. Synopsis of Current Human-AI Interaction Research

Within the past several years, there has been a great deal more research into how AI and people can work together. That means we need systems that allow AI agents and people to do the same things at the same time. The first models were all about how people and computers talk to each other mostly. The emphasis has shifted to AI which can adapt to its surroundings and respond to human input in such a way as to make sense as more and more autonomous learning systems become available. Chatbots, interactive machine learning, and so on, those systems that let people in, have been investigated. This research fails to convey the importance of understanding these interactions.

B. AI Systems' Interpretability

Much work has been done to make AI more explainable, especially in terms of how machines learn. There are many ways to learn the usage of deep neural networks and other complex models: SHAP, LIME, attention processes, to name a few. Those will tell you how confident a model is, what steps it takes to make predictions, and how important each feature is. However, these explanation methods are difficult to interface because they are most often too technical and not for the end user. It is becoming clearer that the explanation method should be tailored depending on what the user already knows and what he/she needs to know.

C. Current User Interface Designs That Encourage AI Transparency

Much effort has been invested in creating user-friendly interfaces. For instance, recommendation systems might include visual sliders that alter results according to different criteria or might phrase things as "You liked X." Medical AI systems now sometimes have user interfaces that present confidence scores, highlight parts of diagnostic images, or supply text-based explanations for choices. These designs are not common yet, and evidence regarding real-world efficacy is often lacking. There is as yet considerable variation in exactly how explanations are communicated, and the empirical record varies widely regarding just how much they assist people in trusting or comprehending the information.

D. Literature Deficits

Researchers are still working hard to find ways to make AI easier to understand and to figure out how people and AI can work together. But we do not know much about how they actually do this, particularly when it comes to the development of something useful to the user. There is not a lot of research on how to explain the reasoning of AI to persons not knowledgeable in it. Fewer actually give clear advice or rules on how to do UI/UX design in order to make it more understandable. Many studies mention that knowing things makes one confident, but this has not been shown in real life so far. These omissions underpin the need for research in the ability to assess the effectiveness of novel interpretability strategies in interactive systems, rather than just suggesting them.

3. Foundations and Concepts

A. Key Terms Definition (e.g., Explainability, Interpretability, User Trust)

Interpretability refers to the simplicity of deducing what an AI system does or how it works. Some researchers define that a system can be explained if it can tell you what it does and why it does that. However, sometimes the two terms have been used to mean the same thing. "User trust" is the degree at which a user believes the system can execute what it says it can do and the reason for certain actions. These concepts are very similar

when it relates to how people and AI can work together. Explainability and interpretability have a direct impact on trust, which in turn influences the level of trust and utilisation of AI systems.

Table 1: Key Concepts in Interpretability and Human–AI Interaction

Term	Definition	Why It Matters in Human–AI Interaction
Explainability	The ability of an AI system to clearly describe what it did and why it made a certain decision.	Helps users understand the reasoning behind AI outputs, reducing confusion and enabling accountability.
Interpretability	How easy it is for a person to understand the internal working or behaviour of an AI system.	Allows users to mentally predict AI behaviour, improving usability and confidence during interactions.
User Trust	The level of confidence a user has that the AI system is reliable, honest, and competent.	Directly affects whether users accept AI decisions, continue using the system, or reject its recommendations.
Transparency	The openness of the system in showing how data is used, processed, and turned into decisions.	Encourages ethical use and helps prevent anxiety or suspicion when interacting with AI.
Human-AI Collaboration	The cooperative process where humans and AI work together to complete tasks.	Effective collaboration depends on clarity, trust, and the ability to understand AI suggestions.
Cognitive Load	The amount of mental effort a user needs to understand AI outputs or interfaces.	Low cognitive load improves user experience and enables better decision-making when using AI tools.

B. Interpretability Types: Post-hoc versus Global, Local versus Global.

You can see, at a glance, how all of the parts of the model fit together when you look globally at it. For example, a decision tree sorts data into groups based on the traits given. On the other hand, local interpretability is where you try to figure out why the system made a guess, like why it thought a movie was good or why it thought someone was sick. Post-hoc interpretability means figuring out what a model means after you have trained it with tools like LIME or SHAP. You don't need to do anything else to learn about decision trees and linear regression. People see things differently, and the interface gives more information depending on how these groups are chosen.

C. Design Principles and HAI Models

Many models illustrating how people and AI can cooperate in practice embody variants of older notions from HCI, adapted to mitigate the problems smart systems create. These models say that you should be honest, give feedback, be open to change, and think the same way. Important basic design heuristics for HAI interfaces include giving users control, providing them with context, responsiveness, and addressing morality-related concerns such as fairness and bias. When you design interfaces, think of ease of use, helpfulness of the explanation, and how to make it feel like it's for them. With a good HAI system, people should understand and, when appropriate, have the ability to ask questions about and change AI outputs.

4. Methodology

A. Research Methodology (e.g., Framework Analysis, User Study, Survey, or Prototype Development)

creating prototypes, user studies, surveys, and framework analysis are a few. The first step is the creation of a test interface with AI components that are easy for humans to use. This is a very important part of the mixed-methods approach that underpins the research. This prototype is based on human-centred design and employs a range of methods to explain things that are unique to the end user. Once this interface is complete, tests are carried out on people. The subjects speak to the AI system while doing exercises to help them make decisions. When the task is completed, we gather quantitative data, such as the performance of completing the task and the time taken, and qualitative data, such as how much trust is put in it and what was said about it. We observe this through discussions with the subjects, questions asked, and observation. The study shall try to appraise the effectiveness of interpretable components in fostering trust and understanding.

B. Platforms, Datasets, or Tools Employed

You then create the prototype using Vue.js or React, which are frameworks running in a web browser. People can learn more about how models work with the help of SHAP and LIME. You will be using the MIMIC-III healthcare dataset to make tools that help you figure out what's wrong with you, and the MovieLens dataset to make a system that tells you what films to watch. Both these sets of data are different in their usage. People can still use the system while it undergoes review because it is hosted on a server that is either on the user's computer or in the cloud.

C. Evaluation Metrics (such as Task Performance, Trust, and Usability)

There are plenty of ways to find out how well an interface works. SUS and watching people use a product in person are two common ways to find out how easily a product can be used. We use trust scales that have been tested to see how much people trust AI's decisions. It's good if the AI system can do its job well and fast. We also get subjective feedback to find out how clear, helpful and satisfying the explanation is for the user. These numbers show us how easy it is to understand AI and how that changes how people use it.

5. Design and Implementation

A. Overview of the Suggested Framework, System, or Interface

The proposed architecture is an AI interface that lets the user choose an interaction mode with the support of ease-of-use features. It consists of three parts: the frontend, which is the interface with which the user directly interacts; the backend, which contains the AI model; and the interpretability module. First, it has a trained machine learning model at the backend-such as a recommendation engine or classification engine. Using such a model, one can predict what would be the next step based on what the user says or what is going around them. The interpretability layer rests between the AI model and the user interface. It finds feature importance by applying SHAP or LIME and measures the confidence. Further, the data is provided to the front-end interface, which was designed to present the data in a non-interfering and understandable manner. Since the interface is dynamic and adaptive, a large set of people with various levels of technological understanding can work with it. It is easy for users to read explanations, have a conversation with an AI model, and provide feedback.

B. Including Interpretability in UI Components (e.g., Feedback Loops, Visual Cues, and Explanations)

It is full of user-centered design elements, making it easier to use. Next to the result, there's a short note on why the AI chose that choice. The UI depicts changes in recommendations due to changes of some of the settings in a recommendation system as a bar chart. Other things it may contain are "You might like this because you've been interested in X lately." From things such as sliders or color gradients-which show how sure or unsure AI is-they can immediately be informed how much they can rely on the answer AI gives them. Once users click on "Why?", they reveal more elaborative answers describing how AI works using interactive graphics. The UI also contains feedback loops, giving an opportunity for the user to grade the outputs they believe are wrong or unclear and giving explanations a score. This not only empowers the users but gives useful data to the system which shall contribute to its growth in years to come.

C. Use case or case study (e.g., AI in education, healthcare, and recommender systems)

For evaluating the proposed framework, the case study applied a movie recommendation system using the MovieLens dataset. We have chosen this domain since it has a large yet easy-to-handle framework for studying how people react to AI decisions. So, the algorithm came up with a way to suggest films based on what people like and do in the past. Feature-based explanations are used by the system to suggest something. For instance, it may suggest that the movie is of the right genre, has high user ratings, or it is similar to movies that the user viewed. If explanation panels are larger, users will be able to notice the logical basis of the tip and then decide whether to go with the suggestion or not. This case study will show how user-friendly features make a difference in the utilization, comprehension, and trustworthiness of a real-world application. It also shows you how you can make more difficult or easier explanations so that they stay useful but not a burden to the user.

6. Results and Evaluation

A. Results from User Studies or Experiments

In total, the user study included thirty participants using the AI-driven recommendation system. Some tasks asked people to choose movies they liked, go through recommendations, and discuss reasons for liking them. A few key conclusions were derived from the results of the study. For example, it was ensured that people did what AI recommended after it had been explained to them. That is, people felt sure about what they were doing and believed in it. The ability to ask AI 'why' it did something made them feel better and in control. Quantitative results indicated that introducing features of interpretability improved task efficiency due to a decrease in decision time. The fact that participants preferred explanation formats that are more flexible and customizable underlined the value of user-oriented design in interpretability.

B. Evaluation of Task Performance, Trust Levels, and User Feedback

We used surveys and interviews to find out what people really thought about the job. People reported that the best part is being transparent and explicit, especially the charts showing why the decisions were made. A likert scale trust index showed 25% more trust in the system from the cohort using the interpretable interface than from the cohort using the traditional black box interface. In tasks such as speed and choice correctness-that is, how well they matched what they already liked-the interpretable group also did better. Some people, however, felt that at times too much information could be a bad thing, especially if it was provided without any context, or when they do not have much time. This brings out the importance of striking a balance between making things incomprehensible and giving people enough information to be helpful.

C. Comparing If Applicable to Non-Interpretable Systems

It was standard to create a version of the system that could be understood which would be controllable. It provided the same AI-powered suggestions, yet nobody was ever able to figure them out. Comparing the two systems made it very clear that the interpretable version was better. The interpretable system helped these individuals make choices that were more reasonable and logical. Second, it increased trust and usability scores: many of those who used the black-box interface extensively claimed not to know what to do and did not want to listen to someone they didn't know. This research corroborates the notion that enhancing the comprehensibility of AI systems significantly facilitates collaboration between humans and AI. If you don't do this, the systems may not work as well-even when you have the right model.

7. Discussion

A. Results Interpretation

These findings confirm the hypotheses that people are happier and more trusting and that they make better decisions, when user interfaces are more usable. People became not only more interested in the system but made choices utilizing more information confidently. They no longer thought they were only getting the results of a computer program. They didn't just feel like they were watching the AI think; they felt part of it because they could see and hear the explanations. Such findings confirm that ease of understanding in AI systems is something to be aimed for in the design process, not an afterthought, and especially when the user thinks it matters.

B. Consequences for AI System Development and UI/UX Design

These findings from the study have great implications for the development of AI systems and designing and building UI / UX. This study illustrates for designers the importance of crafting explanations which are comprehensible, aesthetic, and relevant to the needs of the end-users. Adaptive interfaces, which change dynamically with what users know or have done previously, may further improve the experience of interpretation. These findings mean that the developers of AI should come up with models that easily can be comprehensible and make their decision-making process transparent. It could be the use of black-box models with good explanation tools or models that basically are understandable themselves. The front and back must connect perfectly to work together.

C. Difficulties and Trade-offs (e.g. Explainability vs. Performance)

This paper enumerated many benefits of understanding but also several design issues and trade-offs. When people are given too much technical information, it gets hard to think. It does not precisely help people in non-expert domains to make better sense of things. Another problem is how to have a model that is both simple and complex: deep neural networks are excellent at their jobs, but they do not always do something that you actually want. Less understandable models maybe or may not be as clear, perhaps. One needs to consider the application, the people who will use it, and how much important decision-making there is to find the proper balance between how well the model performs its task and how interpretable it is. Finally, it is also a key concern that the explanation systems must be verified to be accurate and helpful because poor or overly simplistic explanations may do more harm than good.

Table 2: Key Difficulties and Trade-offs in Explainability vs. Performance

Aspect	Explainability-Focused Models	Performance-Focused Models (e.g., Deep Neural Networks)
Model Complexity	Simple, easier to understand	Highly complex, difficult to interpret
User Understanding	Suitable for non-experts but may oversimplify	Hard for users to follow or trust the reasoning
Accuracy / Performance	May have lower accuracy in complex tasks	Usually, higher accuracy and better task performance
Cognitive Load	Low cognitive load - easy explanations	High cognitive load—too technical or unclear explanations
Decision Confidence	Users trust decisions because they can understand them	Users may distrust decisions due to "black box" behaviour
Suitable Use Cases	Healthcare explanations, auditing, legal cases	High-speed prediction, large-scale classification, vision tasks
Risk Factors	Oversimplified explanations may mislead users	Incorrect or unclear explanations can harm user trust
Main Trade-off	More interpretability → less predictive power	More predictive power → less interpretability

8. Future Work

There are many entertaining ways to learn more about how people and AI collaborate and about how to make AI more explainable. Among the biggest priorities is to work on improving adaptive interfaces: these are interfaces that change by how the user uses them, what they like, and how well they use them. In the future, systems should be able to adjust on the fly the frequency, format, and depth of explanations for each user. Many of those trying to fix things now think one answer fits all. Whereas an expert might prefer technical breakdowns or statistical models, a beginner might find it easier to understand things when they are explained in simple terms. The next generation of user interfaces will need to be context-aware and learn from how people use them to make explanations more personalized in real-time. You can speak to AI, gesture, or view it all in one go to converse with it. In the future, as AI will increasingly be deployed on smart devices-from voice assistants and self-driving cars to wearables-we need to understand AI systems in more than words and pictures. Voice-based explanations, haptic feedback, and augmented reality visual overlays exemplify some forms of multimodal explanations.

In creating attempts across multiple channels for interpretability, cognitive coherence is hard to maintain while sensory modalities are challenging to coordinate. On the bright side, this is a great opportunity for user experiences that are more intuitive and accessible. Finally, more studies are needed to understand how this technology fares in varied situations. For instance, in healthcare, finance, or law, there are different norms of rules, ethics, and laws, and developing frameworks either for interpretability or for user interface components to serve particular application types, such as recommender systems, cannot be utilized beyond these domains. The creation of modular design patterns and APIs, which are changeable for certain situations without worsening either the explanation or the user experience, challenges the designers. We need interdisciplinary collaboration, long-term research, and cross-domain testing of solutions to create smart systems that are extensible and easy to use.

9. Conclusion

In Human–AI Interaction, there is an emerging need for transparency in user interfaces. In this work, we addressed this need by proposing a design framework for enhancing transparency in user experience. As AI is becoming ubiquitous in everyday applications, from recommendation engines to decision support systems, educating people about how it works is an issue of growing importance. We started with discussing the concept of interpretability and its consequences for users' trust, freedom, and involvement in AI systems. We have shown that interpretability can be added to the interaction loop by developing and deploying a prototype interface with model-agnostic visual explanations, confidence indicators, and feedback mechanisms. Researchers used a movie recommendation system as a test case in order to explore real-world impacts of these factors on people. User studies demonstrate that transparent UIs reduced the effort taken to perform work, while users were happier and even trusted AI decisions more than in non-transparent UIs. People liked the fact that they could comprehend the predictions made by AI, ask questions about them, and even disagree with them.

That means facilitating better understanding of AI may improve human–machine harmony. The results showed that explanations need to be tailored according to user knowledge and goals but also have to strike a balance between clarity and utility. Our research shows that interpretability is not just a technical attribute of AI systems, but it is also a key component in ensuring their ethicality, utility, and accountability. Furthermore, the study discussed key points to remember for new design approaches, such as avoiding information that is too complex and finding the proper balance between honesty and accuracy within the models. The work advocates for a human-centered approach toward AI interpretability by offering developers, designers, and researchers' clear concepts and concrete recommendations. This forces us to rethink AI systems since we should not think of them as robots that can think independently. Think of them as collaborators who have to be transparent, direct, and receptive to feedback. More importantly, much as how humans work and collaborate with one another, we should know how they will do this. And this is not only a goal of research, but necessary in order for systems to be made that people can trust and use in a useful way.

10. References

[1] Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.

[2] Lipton, Z. C. (2018). The mythos of model interpretability. *Communications of the ACM*, 61(10), 36–43.

[3] Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, 1–38.

[4] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144).

[5] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (pp. 4765–4774).

[6] Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., ... & Horvitz, E. (2019). Guidelines for human-AI interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–13).

[7] Binns, R., Veale, M., Van Kleek, M., & Shadbolt, N. (2018). 'It's reducing a human being to a percentage': Perceptions of justice in algorithmic decisions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–14).

[8] Gunning, D. (2017). Explainable artificial intelligence (XAI). *Defense Advanced Research Projects Agency (DARPA), Program Information*.

[9] Eiband, M., Schneider, H., Bilandzic, M., Fazekas-Con, S., & Butz, A. (2018). Bringing transparency design into practice. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces* (pp. 211–223).

[10] Kulesza, T., Burnett, M., Wong, W. K., & Stumpf, S. (2015). Principles of explanatory debugging to personalize interactive machine learning. In *Proceedings of the 20th International Conference on Intelligent User Interfaces* (pp. 126–137).

[11] Abdul, A., Vermeulen, J., Wang, D., Lim, B. Y., & Kankanhalli, M. (2018). Trends and trajectories for explainable, accountable and intelligible systems. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–18).

[12] Holzinger, A., Biemann, C., Pattichis, C. S., & Kell, D. B. (2017). What do we need to build explainable AI systems for the medical domain? *Review and perspectives*, 71, 101–113.

[13] Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, 36(6), 495–504.

[14] Liao, Q. V., Gruen, D., & Miller, S. (2020). Questioning the AI: Informing design practices for explainable AI user experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–15).

[15] Zhang, Y., Liao, Q. V., Bellamy, R. K. E., & Singh, M. (2020). Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 295–305).