MANTECH Publications
Journal of Product Design, Quality Engineering & Technology
Volume 6, Issue 2, May-August 2021

Explainable AI (XAI) for Design Decisions

Devender Chadhary[1], Harish Bansal[2], Rajaram Singh[3], Govind Pathak[4]
Assistant professor[1], Professor[2, 3, 4]
Department of Product Innovation
Union Christian College, Aluva, India
Email ID: *devenderchaudharyty@yahoo.com[1], harishban48@gmail.com[2], pathakgovind05@rediffmail.com[3]*

Abstract

Artificial Intelligence (AI) has become an important support tool in modern design processes, influencing decisions related to form generation, material selection, usability evaluation, and performance optimization. However, many AI-driven design systems operate as black boxes, offering limited insight into how and why specific decisions are produced. This lack of transparency reduces designer trust, makes validation difficult, and raises ethical and accountability concerns. Explainable Artificial Intelligence (XAI) aims to address these challenges by providing interpretable and understandable explanations for AI outputs. This paper presents a comprehensive review of Explainable AI in the context of design decision-making. It discusses the fundamentals of XAI, its relevance to design disciplines, commonly used explanation techniques, and application scenarios across product design, engineering design, and user experience design. The paper also highlights benefits, limitations, and future research directions of XAI-enabled design systems. The review shows that integrating XAI into design workflows can improve trust, collaboration, and informed decision-making, while supporting responsible and human-centered use of AI technologies.

Keywords: *Explainable AI, design decisions, human-centered design, interpretable models, decision support systems*

INTRODUCTION

Design decision-making is a complex activity that involves balancing functional requirements, user needs, technical constraints, aesthetics, cost, and sustainability considerations. Traditionally, design decisions were made primarily through human expertise, experience, and iterative prototyping. With the advancement of computational tools, Artificial Intelligence (AI) has increasingly been integrated into design workflows to assist designers in exploring large design spaces, predicting performance, and optimizing solutions.

Machine learning models, especially deep learning approaches, have demonstrated strong capabilities in tasks such as generative design, user behavior prediction, defect detection, and performance estimation. Despite these advantages, many AI systems used in design act as opaque black boxes. Designers often receive recommendations or outputs without clear justification, making it difficult to understand the reasoning behind decisions. This creates resistance among practitioners, particularly in safety-critical or high-stakes design domains.

Explainable Artificial Intelligence (XAI) has emerged as a response to this challenge. XAI focuses on developing AI models and methods that can provide explanations understandable to humans. In design contexts, XAI can help designers interpret AI suggestions, validate results, and integrate computational insights with human judgment. This paper reviews the role of XAI in design decision-making and explores how explainability can enhance transparency, trust, and effectiveness of AI-assisted design.

ARTIFICIAL INTELLIGENCE IN DESIGN DECISION-MAKING

Artificial Intelligence supports design decision-making across almost all phases of the design lifecycle, from early ideation to final evaluation and validation. In the conceptual design stage, AI-driven generative systems and evolutionary algorithms assist designers by automatically generating a wide range of design alternatives based on predefined goals and constraints. These constraints may include functional requirements, spatial limits, cost targets, sustainability indicators, or user preferences. By rapidly exploring large design spaces that would be impractical through manual methods, AI enables designers to consider unconventional or innovative solutions that may otherwise be overlooked.

During embodiment and detailed design stages, AI models are increasingly used to predict and optimize performance-related attributes. Machine learning techniques are applied to estimate parameters such as structural strength, weight, thermal efficiency, aerodynamic behavior, energy consumption, manufacturability, and even perceived user satisfaction. These predictive capabilities reduce the need for repeated physical prototyping and lengthy simulations, thereby shortening development cycles and lowering costs. In complex engineering contexts, AI also supports multi-objective optimization, helping designers balance trade-offs between competing criteria such as performance, reliability, and affordability.

In the evaluation and testing phase, AI assists designers by automating usability analysis, reliability assessment, and regulatory compliance checks. For example, computer vision and data-driven usability models can analyze user interaction data to identify design flaws, while anomaly detection algorithms help predict failure modes and maintenance needs. AI-based compliance tools can also flag potential violations of standards or safety regulations early in the design process, reducing the risk of costly redesigns at later stages.

Despite these advantages, most AI systems used in design prioritize prediction accuracy and computational efficiency over transparency and interpretability. Advanced techniques such as deep neural networks, reinforcement learning models, and ensemble learning methods often function as black-box systems. While they deliver highly accurate recommendations, they provide little insight into how specific inputs, constraints, or assumptions influence the final output. For designers, this lack of visibility introduces several challenges:

- Difficulty in trusting AI-generated solutions, particularly when outcomes conflict with professional experience or intuition.
- Limited ability to justify design decisions to stakeholders, clients, or regulatory bodies who require clear reasoning and accountability.
- Challenges in identifying errors, hidden assumptions, or biases embedded within training data and model structures.

These limitations are especially critical in high-stakes design domains such as healthcare devices, transportation systems, and safety-critical infrastructure, where unexplained decisions can have serious consequences. As a result, there is a growing need for explainable approaches that align AI reasoning with design logic and human understanding. Explainable AI provides

mechanisms to make AI-supported design decisions more transparent, interpretable, and defensible, enabling designers to remain actively involved in decision-making rather than passively accepting algorithmic outputs.

CONCEPT AND PRINCIPLES OF EXPLAINABLE AI

Explainable Artificial Intelligence (XAI) refers to a collection of methods, models, and interface techniques that aim to make the behavior and outcomes of AI systems understandable to human users. Unlike traditional AI approaches that focus primarily on maximizing prediction accuracy, XAI emphasizes clarity, transparency, and meaningful communication between the system and its users. The objective of XAI is not limited to explaining *what* decision an AI system has produced, but also *why* that decision was made, *how* different inputs influenced the outcome, and *under what conditions* the decision might change.

In the context of design decision-making, this distinction is particularly important. Designers are not only interested in final recommendations, but also in the reasoning process that leads to those recommendations. Understanding the logic behind AI-supported decisions allows designers to validate results, challenge assumptions, and integrate computational insights with their own expertise and creative judgment.

Several core principles define Explainable AI and guide its application in design-oriented systems:

Transparency: Transparency refers to the degree to which the internal processes, decision pathways, or logic of an AI model are visible or traceable. In transparent systems, designers can identify how input parameters, constraints, and rules contribute to specific outputs. While full transparency is achievable in simpler models, even partial transparency in complex systems can significantly improve confidence and oversight in design decisions.

Interpretability: Interpretability focuses on how easily a human user can understand the explanation provided by the AI system. An interpretable explanation should be presented in a form that matches the background and needs of the user, such as designers, engineers, or managers. In design practice, this may involve simplified rules, ranked feature importance, visual mappings, or scenario-based explanations rather than raw mathematical formulations.

Justifiability: Justifiability refers to the ability of an AI system to provide reasons that support, defend, or question its outputs. For design decisions, this is crucial when outcomes need to be communicated to stakeholders, clients, or regulatory authorities. Justifiable explanations enable designers to articulate why a particular design option was selected and how it satisfies functional, aesthetic, or safety requirements.

Human-centeredness: Human-centeredness emphasizes that explanations should be aligned with human cognitive capabilities, design workflows, and domain knowledge. Rather than overwhelming users with technical detail, XAI systems should adapt explanations to user expertise and context. In design environments, this often means supporting exploratory reasoning, comparison between alternatives, and iterative refinement instead of one-time static explanations.

In many design applications, explainability does not require complete mathematical transparency of the AI model. Designers often benefit more from partial explanations, visual cues, sensitivity analyses, or comparative insights that reveal how design variables influence outcomes. Such explanation formats align closely with established design thinking practices, where sketching, prototyping, and visual reasoning play a central role. By supporting understanding without constraining creativity, XAI enables AI systems to act as collaborative partners rather than authoritative black boxes in the design process.

XAI TECHNIQUES RELEVANT TO DESIGN

Explainable AI techniques used in design decision-making can be broadly categorized into **model-intrinsic explainability** and **post-hoc explanation methods**. This classification is useful for designers and researchers, as it reflects whether explainability is built directly into the AI model or added after the model has been trained. Each category offers distinct advantages and limitations depending on the complexity of the design problem and the level of transparency required.

Model-Intrinsic Explainability

Model-intrinsic explainability refers to AI approaches that are interpretable by design. These models are structured in such a way that their decision logic can be directly examined and understood without the need for additional explanation tools. Common examples include linear

and logistic regression models, decision trees, rule-based systems, fuzzy logic models, and Bayesian networks.

In design contexts, intrinsic models are particularly valuable when decision logic must be explicitly traced and communicated. For instance, a decision tree applied to material selection can clearly illustrate how criteria such as cost, mechanical strength, environmental impact, availability, and manufacturability influence the final recommendation. Each branching point in the tree represents a logical decision based on a specific parameter, making the reasoning process transparent and easy to validate.

Similarly, rule-based systems are often used in early-stage design and configuration tasks, where expert knowledge can be encoded as design rules. Designers can directly inspect, modify, or challenge these rules, which supports iterative refinement and learning. Bayesian networks further allow designers to understand probabilistic relationships between design variables, supporting reasoning under uncertainty.

Despite their advantages, model-intrinsic approaches have notable limitations. They often struggle to capture highly nonlinear, high-dimensional, or complex relationships that are common in advanced design problems such as generative design, topology optimization, or user behavior prediction. As design challenges increase in scale and complexity, purely interpretable models may sacrifice predictive accuracy or flexibility. This trade-off has led to increased reliance on complex models supplemented by post-hoc explanation techniques.

Post-Hoc Explanation Methods

Post-hoc explanation methods aim to explain the behavior of complex, often black-box AI models after they have been trained. These techniques do not alter the internal structure of the model but instead analyze inputs and outputs to generate explanations that are accessible to human users. Post-hoc methods are widely used in design optimization, performance evaluation, and simulation-driven design, where deep learning and ensemble models are common.

One widely used approach is **feature importance analysis**, which identifies and ranks the input variables that most strongly influence a model's output. In design applications, this helps

designers understand which parameters—such as geometry dimensions, material properties, or usage conditions—have the greatest impact on predicted performance or quality metrics.

Local explanation techniques such as **LIME (Local Interpretable Model-agnostic Explanations)** and **SHAP (SHapley Additive exPlanations)** provide explanations for individual predictions rather than global model behavior. These methods are especially useful in design review scenarios, where designers need to justify specific recommendations or investigate unexpected outcomes. By showing how small changes in input values affect a particular decision, local explanations support diagnostic reasoning and design refinement.

Sensitivity and saliency analysis methods further contribute to explainability by visualizing how variations in input parameters influence outputs. In visual and geometric design tasks, saliency maps can highlight regions of a design that most strongly affect predicted performance or user response. Such visual explanations align well with designers' reliance on sketches, diagrams, and visual reasoning.

While post-hoc methods significantly enhance understanding of complex AI systems, they also introduce challenges. Explanations may be approximate rather than exact, and different methods can sometimes produce inconsistent results. Designers must therefore interpret post-hoc explanations critically and consider them as decision-support tools rather than definitive truths.

Overall, the combined use of intrinsic and post-hoc XAI techniques offers a flexible and practical approach for integrating explainability into design decision-making. By selecting appropriate techniques based on task complexity and user needs, designers can achieve a balance between predictive power and interpretability.

APPLICATIONS OF XAI IN DESIGN DECISIONS

Explainable AI has found increasing application across a wide range of design domains as designers seek to balance the computational power of AI with human understanding and responsibility. By embedding explainability into AI-supported workflows, XAI enables designers to critically engage with algorithmic recommendations, rather than treating them as unquestionable outputs. This section discusses key application areas where XAI has

demonstrated significant value in supporting design decisions.

Product and Industrial Design

In product and industrial design, XAI supports decision-making related to form development, material selection, functional layout, and sustainability considerations. AI-driven generative design tools are commonly used to propose optimized geometries that minimize weight, reduce material usage, or improve structural performance. However, without explainability, such outputs may appear unintuitive or impractical to designers.

XAI techniques address this issue by revealing how different constraints and performance criteria influence the generated form. For example, when an AI system recommends a specific geometry to reduce weight, explainable outputs can illustrate how stress distribution, load paths, material properties, and manufacturing constraints collectively shaped the final design. Visual explanations such as heat maps or feature influence diagrams allow designers to see which regions of a component are critical to performance and which areas allow greater flexibility.

This level of insight enables designers to refine and adapt AI-generated concepts based on contextual knowledge, aesthetic goals, or production realities. Rather than blindly accepting automated forms, designers can use XAI to negotiate between algorithmic optimization and human creativity. Additionally, explainable recommendations support communication with non-technical stakeholders by clarifying why certain design choices were made.

Engineering Design and Optimization

In engineering design, XAI plays a crucial role in improving trust and reliability in optimization-driven decision-making. Engineering problems often involve multiple, competing objectives such as cost, performance, safety, durability, and environmental impact. AI-based optimization algorithms can efficiently search for optimal solutions, but their recommendations may be difficult to interpret without explanation.

XAI-enhanced systems allow designers and engineers to understand trade-offs between objectives by explicitly showing how changes in design variables affect outcomes. For instance, explainable models can indicate why a slight increase in material thickness leads to

significant improvements in fatigue life, or why reducing cost beyond a certain threshold negatively impacts safety margins. Such insights are particularly valuable in simulation-driven design workflows, where AI models are tightly coupled with numerical analysis tools.

In safety-critical fields such as aerospace, automotive, and civil engineering, explainability is essential for validating AI-supported decisions. Regulatory compliance and certification processes often require clear justification of design choices. XAI helps engineers trace the reasoning behind AI-generated recommendations, identify potential modeling errors, and ensure that optimization results align with engineering principles and standards.

User Experience and Interaction Design

Explainable AI is increasingly relevant in user experience (UX) and interaction design, where AI systems personalize interfaces, adapt content, or predict user behavior. AI-driven personalization can enhance usability and engagement, but it also raises concerns related to bias, transparency, and user trust.

XAI enables designers to understand why certain interface adaptations or recommendations are suggested based on user behavior data. For example, explanations can reveal which interaction patterns, accessibility needs, or contextual factors influenced layout changes or content prioritization. This insight helps designers evaluate whether AI-driven adaptations genuinely improve user experience or inadvertently exclude certain user groups.

By supporting transparency in personalization logic, XAI contributes to ethical and inclusive design practices. Designers can identify and mitigate unintended biases, ensure fairness across diverse user populations, and provide clearer justifications for adaptive system behavior. In this way, XAI strengthens the role of designers as ethical decision-makers in AI-enabled interaction systems.

Overall, the application of XAI across product, engineering, and interaction design demonstrates its potential to bridge the gap between algorithmic intelligence and human-centered design thinking. As AI becomes more deeply embedded in design practice, explainability will remain essential for maintaining trust, accountability, and creative control.

BENEFITS OF XAI FOR DESIGN PRACTICE

The integration of Explainable AI into design decision-making offers several important benefits that directly address the limitations of conventional black-box AI systems. By making AI reasoning visible and understandable, XAI strengthens the role of designers as informed decision-makers rather than passive recipients of algorithmic outputs. These benefits extend across technical, organizational, and ethical dimensions of design practice.

One of the most significant benefits of XAI is **improved trust and acceptance of AI tools among designers**. Trust is a critical factor in the adoption of new technologies within design teams. When designers are able to see why an AI system recommends a particular solution, they are more likely to rely on it appropriately and integrate it into their workflow. Explainable outputs reduce uncertainty and skepticism, especially when AI-generated results differ from prior experience or intuition. Over time, this transparency helps build confidence in AI-supported tools without encouraging blind dependence.

XAI also enables **enhanced collaboration between human designers and AI systems**. Rather than replacing human judgment, explainable systems function as collaborative partners that support exploration, comparison, and learning. Designers can question AI recommendations, adjust constraints, and observe how outcomes change in response. This interactive dialogue promotes a more iterative and reflective design process, where human creativity and contextual understanding complement computational efficiency.

Another important advantage is **better validation and verification of design decisions**. In professional design environments, decisions must often be justified to clients, managers, regulators, or multidisciplinary teams. XAI provides clear reasoning pathways that support design reviews, audits, and documentation. Designers can trace how specific inputs and assumptions influenced outcomes, making it easier to identify errors, inconsistencies, or unrealistic constraints early in the process. This is particularly valuable in safety-critical and compliance-driven design domains.

XAI further supports **ethical and accountable design practices**. As AI systems increasingly influence decisions that affect users and society, accountability becomes essential. Explainable models help designers detect and mitigate biases embedded in data or algorithms, reducing the

risk of unfair or exclusionary outcomes. Transparency also enables clearer responsibility allocation, ensuring that human designers remain accountable for final decisions rather than deferring responsibility to opaque systems.

Overall, by making AI reasoning visible, XAI aligns computational intelligence with human creativity, professional responsibility, and ethical awareness. It supports a balanced integration of AI into design practice, where efficiency and innovation are achieved without sacrificing transparency, trust, or human control.

CHALLENGES AND LIMITATIONS

Despite its potential, XAI faces several challenges in design applications. Explanations may oversimplify complex models, leading to misunderstanding. There is also a trade-off between model accuracy and interpretability. Additionally, different stakeholders may require different types of explanations, making universal solutions difficult.

Another limitation is the lack of standardized evaluation metrics for explainability in design contexts. What is understandable to one designer may not be meaningful to another.

FUTURE RESEARCH DIRECTIONS

Future research should focus on developing design-specific XAI frameworks that integrate seamlessly with existing design tools. More empirical studies are needed to evaluate how explanations influence designer behavior and decision quality. There is also scope for combining XAI with interactive visualization and immersive technologies to enhance understanding.

CONCLUSION

Explainable Artificial Intelligence represents a critical step toward responsible and effective use of AI in design decision-making. By providing transparent and interpretable insights, XAI helps designers understand, trust, and critically evaluate AI-generated recommendations. This review has highlighted the principles, techniques, and applications of XAI in various design domains. While challenges remain, the integration of explainable methods into AI-driven design systems has strong potential to support human-centered, ethical, and informed design practices. As AI continues to shape the future of design, explainability will remain essential for

maintaining human agency and accountability.

REFERENCES

1. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning.
2. Gunning, D. (2017). Explainable artificial intelligence (XAI). Defense Advanced Research Projects Agency.
3. Molnar, C. (2022). Interpretable machine learning. Lean Publishing.
4. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you? Explaining the predictions of any classifier.
5. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions.
6. Simon, H. A. (1996). The sciences of the artificial. MIT Press.
7. Blessing, L. T. M., & Chakrabarti, A. (2009). DRM: A design research methodology.
8. Pasquale, F. (2015). The black box society. Harvard University Press.
9. Norman, D. A. (2013). The design of everyday things. Basic Books.
10. Tufte, E. R. (2001). The visual display of quantitative information. Graphics Press.