Instructions for use

Explainable AI (XAI) Design Cards

Designing Explainable AI User Interfaces

(1) AI Requirements Gathering
Find out about your users and what they need AI for.

(2) Explainable AI Requirements Gathering
Explore what explanations your users need about the AI they use.

(3) User Interface Design Strategies
Select an explanation strategy for each explanation requirement.

(4) User Interface Design Solutions
Choose a user interface design solution for each design strategy.

Rethinking Explainability and AI

Use the **Explainable AI Manifesto** cards to prompt rethinking about what Explainable AI means to you and your context.Speculate on alternative futures and strategies for Explainable AI in your context.
Use the manifesto cards to unblock stuck design thinking.

AI Requirements Gathering

R 01

Explainable AI
Design Cards

The big what

The big what

What are users doing that needs AI?

The big what

What is the AI model intended to do?

AI Requirements Gathering

R 02

Explainable AI
Design Cards

How would you use it

How would you use it

What should the AI do to be useful for you. And why?

How would you use it

How would you interact with the AI?

How would you use it

Who else would benefit from the AI? And why?

How would you use it

What should the output of the AI be? Text, or image, or audio, or visualisation, or something else?

How would you use it

How would AI usefully fit into your workflows?

Explainable AI Requirements Gathering

G 01

Explainable AI
Design Cards

Explainability

Explainability

Who is the explanation of this AI for?

Explainability

What questions does the AI need to answer for it to be trustworthy and effective for you. And why?

Explainability

What skills and knowledge do you bring?

Explainability

How would you want to interact with the AI's explanations? E.g. using text chat, or audio, or an interactive user interface, or something else

Explainable AI Requirements Gathering

G 02

Explainable AI
Design Cards

How the AI works

How the AI works

What and why do you want to know about...

How the AI works overall?

How the AI works

What and why do you want to know about...

How the AI was trained?

How the AI works

What and why do you want to know about...

How the AI was trained?

How the AI works

What and why do you want to know about...

Why the AI made a particular decision, or output, or prediction?

Explainable AI Requirements Gathering

G 03

Explainable AI
Design Cards

How well the AI performs

How well the AI performs

What and why do you want to know about...

What are the limitations of the AI?

How well the AI performs

What and why do you want to know about...

How the AI's performance compares to other approaches?

How well the AI performs

What and why do you want to know about...

How accurate the AI is in general?

How well the AI performs

What and why do you want to know about...

How reliable the AI is?

How well the AI performs

What and why do you want to know about...

What mistakes the AI typically makes?

Explainable AI Requirements Gathering

G 04

Explainable AI
Design Cards

How the AI uses data

How the AI uses data

What and why do you want to know about...

What data was used to train the AI?

How the AI uses data

What and why do you want to know about...

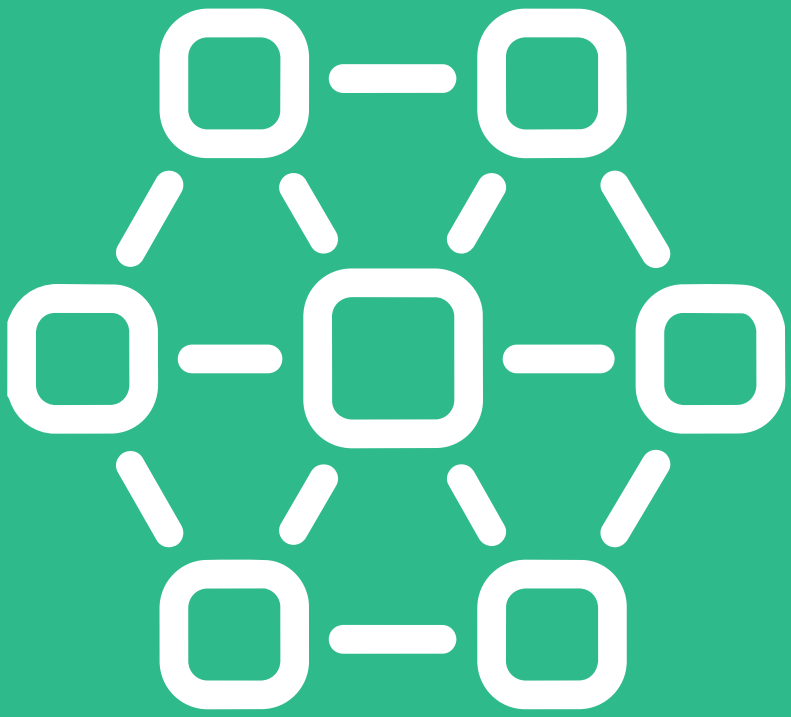
What data sources used by the AI are important for you you to know about when you use it?

How the AI uses data

What and why do you want to know about...

What are the limitations or biases of the training data used?

User Interface Design Strategies

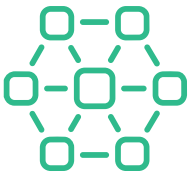


S 01

Explainable AI
Design Cards

When to explain

User Interface Design Strategies



For each identified Explainable AI Requirement...

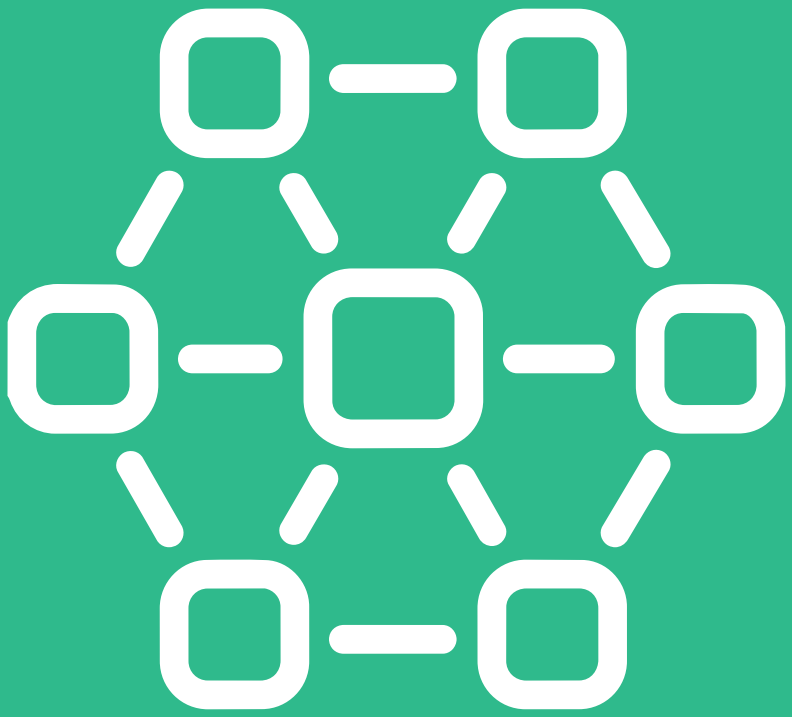
When to explain?

Post-hoc explanation after the AI has been used

Or

Real-time explanation whilst you are using the AI?

User Interface Design Strategies

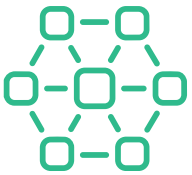


S 02

Explainable AI
Design Cards

What to explain

User Interface Design Strategies

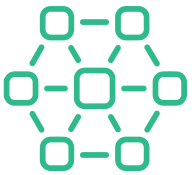


For each identified Explainable AI Requirement...

AI Model explanation

Explaining how an AI model works overall.

User Interface Design Strategies

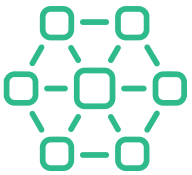


For each identified Explainable AI Requirement...

Transparent model

The AI model itself is interpretable.

User Interface Design Strategies

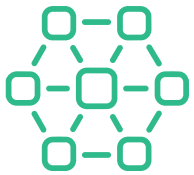


For each identified Explainable AI Requirement...

Model inspection

Explaining one aspect of an AI model for a specific outcome.

User Interface Design Strategies

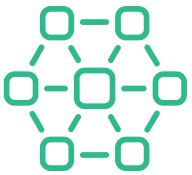


For each identified Explainable AI Requirement...

Outcome explanation

Explaining how a specific decision, prediction, or outcome is arrived at.

User Interface Design Strategies



For each identified Explainable AI Requirement...

Data explanation

Explaining which data is used for a specific outcome and why.

User Interface Design Solutions

D 01

Explainable AI
Design Cards

User Interface Design Solutions

For each selected User Interface Strategy...

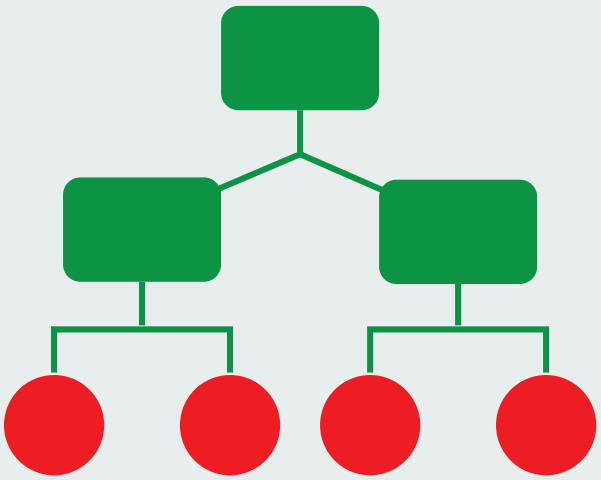
Decision trees and rules

+ Good for model inspection.

+ Good for explaining AI models and outcomes.

+ Good for explaining classification tasks.

- Can oversimplify complex system.



User Interface Design Solutions

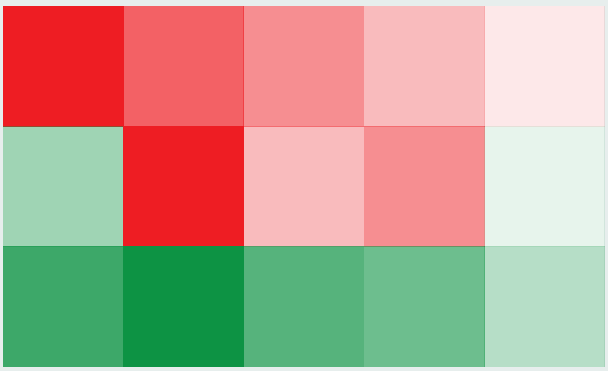
For each selected User Interface Strategy...

Heatmaps

+ Good for explaining outcomes.

+ Good for explaining classification.

- Meanings of colours can be unintuitive and ambiguous.



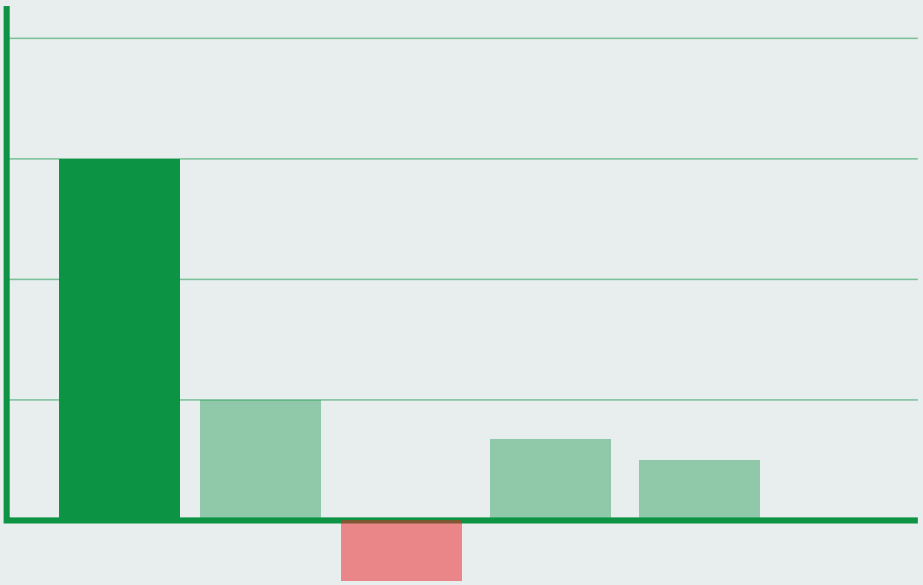
User Interface Design Solutions

For each selected User Interface Strategy...

Feature importance

+ Good for explaining how specific AI model features contribute to outcomes.

- Lists of features can become confusing.



User Interface Design Solutions

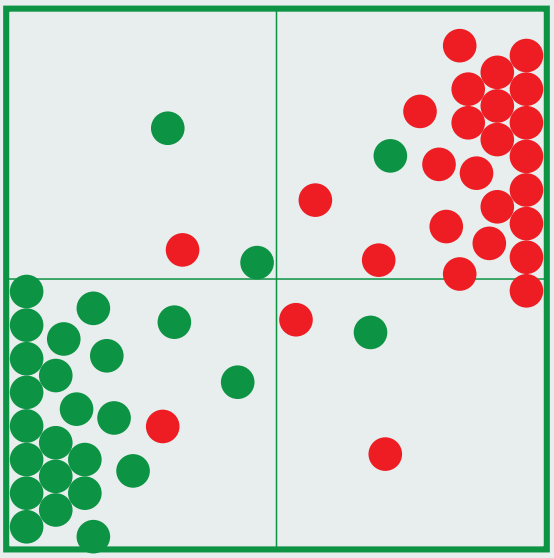
For each selected User Interface Strategy...

Feature plot

+ Good for explaining role of AI model features.

- Can be hard to interpret.

- Work best for a small number of features.



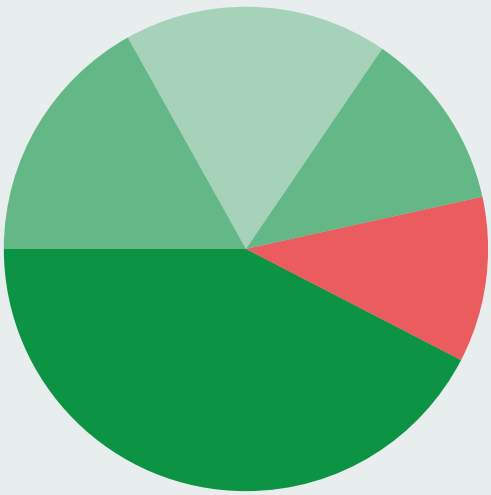
User Interface Design Solutions

For each selected User Interface Strategy...

Feature explanation

+ Good for explaining models and data.

- Can be hard to interpret.



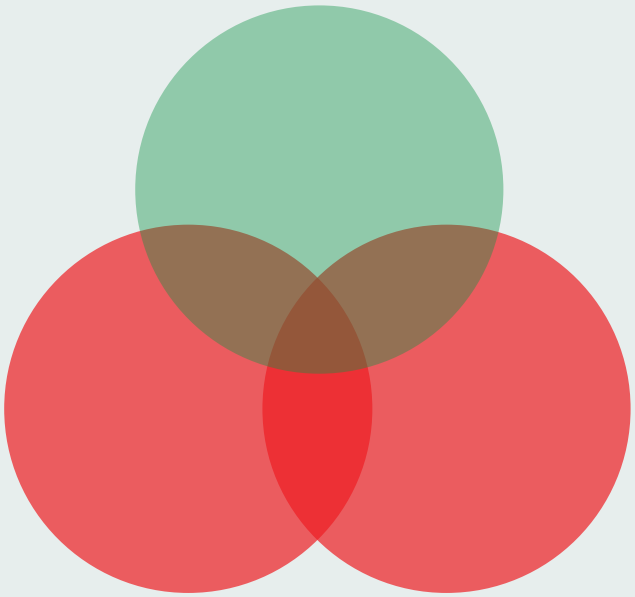
User Interface Design Solutions

For each selected User Interface Strategy...

Data based explanation

+ Good for explaining outcomes and data.

- Reliance on text/ numerical data.



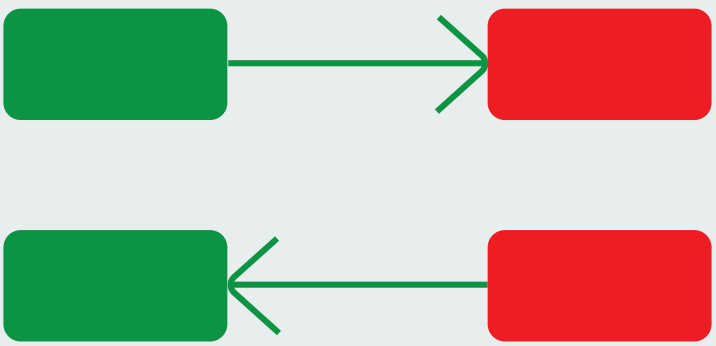
User Interface Design Solutions

For each selected User Interface Strategy...

Contrastive explanation

+ Good for explaining individual outcomes.

- Relies on reasonable contrastive examples.



User Interface Design Solutions

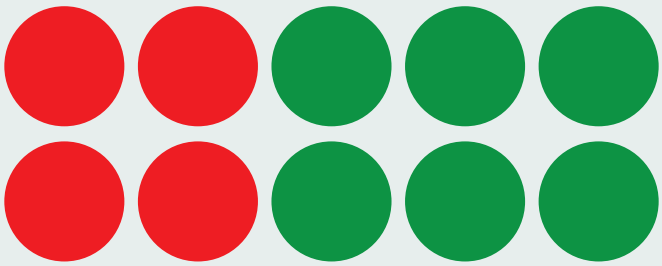
For each selected User Interface Strategy...

Example based explanation

+ Good for explaining individual outcomes.

+ Good for interpreting models.

- Doesn't provide an explanation why.



Explainable AI Manifesto

M 01

Explainable AI
Design Cards

Valuing Artistic Practice for Explainability

Valuing Artistic Practice for Explainability

Aesthetic-Driven Research

Promote research that focuses on first-person accounts and explorations of AI art's aesthetic and experiential aspects rather than purely technical explanations.

Valuing Artistic Practice for Explainability

Collaborative workshops

Host workshops encouraging artists to explore new creative avenues with Explainable AI, pushing the boundaries of traditional art forms and methods.

Valuing Artistic Practice for Explainability

AI Art Exhibitions

Host exhibitions showcasing artworks that explain AI concepts, accompanied by workshops and talks to engage the audience in understanding the underlying technology and its implications in art making and creative practices.

Valuing Artistic Practice for Explainability

Artist-in-Residence Programs

Establish artist-in-residence programs within AI research labs to prioritise artistic exploration and control, and practice-led research on Explainable AI. Organise residencies that bring artists and AI researchers together to create works that make AI models understandable through sound, visual, and interactive art.

Valuing Artistic Practice for Explainability

Demystifying AI

Develop educational materials, such as interactive installations, online platforms, videos, and articles, that explain the computational materiality of AI in accessible language and engaging formats.

Explainable AI Manifesto

M 02

Explainable AI
Design Cards

Openness

Openness

Open Communities

Support community-driven projects that leverage Explainable AI to create novel and impactful artworks, promoting a collaborative approach to AI art. Proactively reach out to underrepresented art groups to build more inclusive communities around AI and the Arts.

Openness

Transparency and Surprise

Promote projects that embrace glitches and imperfections in AI as artistic elements, revealing the inner workings and boundaries of the technology through unexpected and playful outcomes instead of pragmatic explanations.

Openness

Open Collaborations

Facilitate collaborations between artists, technologists, and other disciplines to foster interdisciplinary innovation in AI art, and incorporate Explainable AI into existing art practices.

Openness

Open Access

Support open access to datasets and models with transparency around data sources and use. Reduce output homogenization, and enable broader creative experimentation in Explainable AI. Promote open source AI projects that let artists and researchers inspect and modify underlying code and data.

Explainable AI Manifesto

M 03

Explainable AI
Design Cards

Hacking and Glitches

Hacking and Glitches

Hackathons

Regularly organise hackathons and workshops that encourage collaboration and experimentation with Explainable AI, fostering a culture of innovation and creativity. For example, building on the Hacker's Guide to using generative AI as an artistic material.

Hacking and Glitches

Transparency and Surprise

Promote projects that embrace glitches and imperfections in AI as artistic elements, revealing the inner workings and boundaries of the technology through unexpected and playful outcomes instead of pragmatic explanations.

Explainable AI Manifesto

M 04

Explainable AI
Design Cards

Empowerment, Inclusion, and Fairness

Empowerment, Inclusion, and Fairness

Fair AI Practices

Advocate for ethical Explainable AI practices recognising the inherent bias in AI, prioritising the well-being and empowerment of all users, particularly marginalised groups, building a risk-register of inequity, and gaining consent for any data that is used in AI training and inference.

Empowerment, Inclusion, and Fairness

Ethical Guidelines

Develop and promote ethical guidelines for the use of AI in art, emphasising the importance of explainability contributing to doing no harm and enhancing human well-being.

Empowerment, Inclusion, and Fairness

Transparency and Surprise

Promote projects that embrace glitches and imperfections in AI as artistic elements, revealing the inner workings and boundaries of the technology through unexpected and playful outcomes instead of pragmatic explanations.

Empowerment, Inclusion, and Fairness

Community Engagement.

Engage with underrepresented communities to gather input and feedback on AI tools, datasets, and their bias, to inform future Explainable AI ensuring they meet the needs of a broad range of users.

Empowerment, Inclusion, and Fairness

Accessible and Inclusive Design

Design Explainable AI tools and platforms which incorporate actionable mechanisms to advance accessible, ethical, and inclusive AI practices. For example, building on studies of the use of generative AI tools by people with sensory impairments, and designing for artists with a range of technical literacies.

Empowerment, Inclusion, and Fairness

AI Literacy

Launch campaigns to educate creatives on Explainable AI's benefits, challenges, and ethics while advocating for artist control. Share case studies showing AI's support for creative integrity. Provide training and AI literacy programs that help artists use AI confidently and maintain control through transparent, empowering Explainable AI approaches.

Empowerment, Inclusion, and Fairness

Accessibility Audits

Conduct regular explainability audits and build risk-registers of AI Arts tools and platforms to ensure they meet accessibility standards and address the needs of diverse user groups.