Meena Kaushik
Mu-Ti Huang
Victor Grajski

uxai

A Visual Introduction to Explainable AI for Designers

Project Goals

Explainable Artificial Intelligence (XAI) is increasingly important given the popularity of black box models and use of these models in high stakes decision making scenarios such as criminal justice, medicine, and finance. With increasing societal pressure for explanations and legislation like GDPR's right to explanations, it's no longer a question of 'if we should explain AI models', but how and when we explain. However little is known about how to put techniques from Explainable AI (XAI) research into practice. The role of creating effective explanations often falls to UX design practitioners who bridge the gap between the people using the product ("users") and the product team (including data science and engineers). Their challenge is to create design solutions accounting for demands and constraints coming from users, the product team, and other stakeholders.

The need for a resource that can not only connect academic research to industry but also provide more actionable insights is evident in research and voiced in our early interviews with designers who work on AI products or guidelines. Therefore our primary goal for the website was to create a resource that informs and empowers UX designers when designing explainable interfaces. In deciding what content to feature on the website, we synthesized findings from a wide array of academic and industry resources and emphasized common threads. We leveraged the granularity of academic research and the practicality of industry resources to create an accessible and usable resource for UX designers and AI product teams to experiment with, break, and collaborate on creating explainable interfaces.

Discussion of Related Work

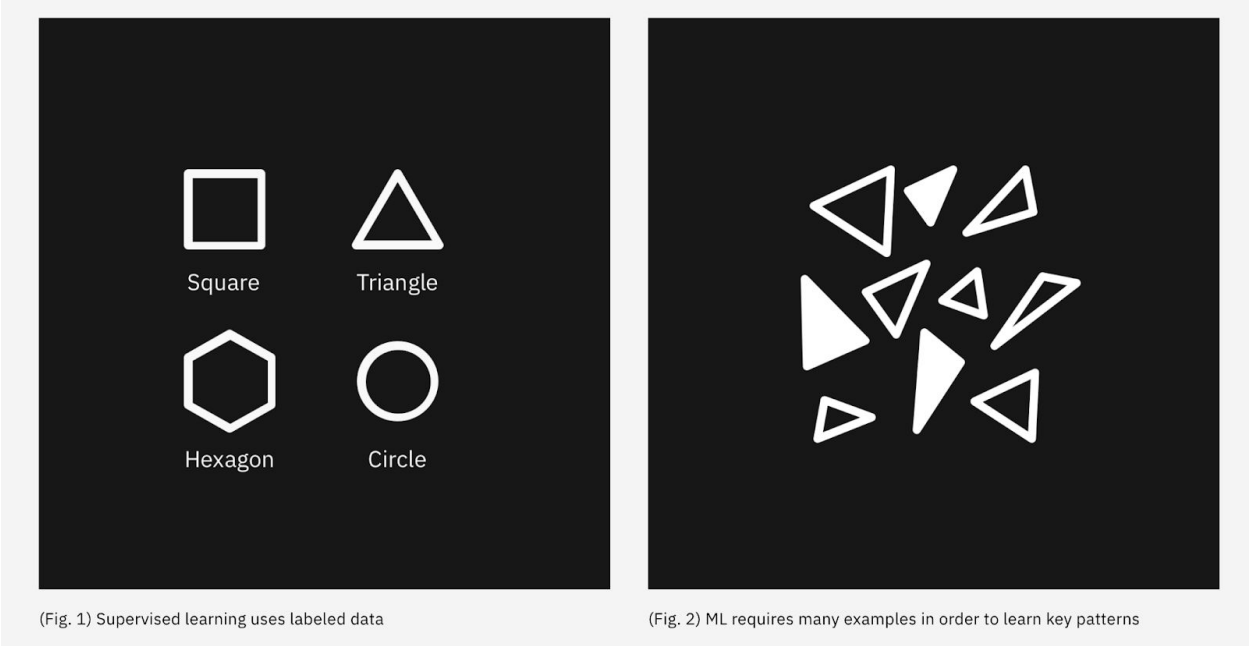
Google PAIR
https://pair.withgoogle.com/guidebook/

Google's People + AI Research (PAIR) guidebook offers guidance for developing AI in 'human-centered' ways. It is divided into six chapters, each focusing on a specific consideration (i.e. explainability + trust, mental models, etc.). The guidebook is principle-led, and seems intentionally broad so that product teams can take the concepts and apply them to specific use-cases. However, the guidebook seeks to make these principles more tangible for the end user by providing examples from Google's product suite, as well as short worksheets after every section. While this resource is

heavily text-based, with few visuals interspersed, we gained insight on important topics we should address in our resource and were motivated by the breakdown of content on this website.

IBM Design for AI
https://www.ibm.com/design/ai/

The intended audience of IBM's Design for AI resource appears primarily to be design practitioners for the purpose of educating them broadly on what artificial intelligence is, and the implications this holds for design. This guide has an extremely broad approach to the field; starting with a high-level introduction and how to use design thinking within a team working on an AI system followed by fundamentals of artificial intelligence. We were inspired by the breadth of content covered in this resource as well as the manner in which compex topics were communicated to non-technical readers through clear writing and visual explanations.



(Fig. 1) Supervised learning uses labeled data

(Fig. 2) ML requires many examples in order to learn key patterns

`IBM Design for AI visual explanation of ML`

Microsoft Guidelines for Human-AI Interaction
https://www.microsoft.com/en-us/research/project/guidelines-for-human-ai-interaction/

Microsoft's Guidelines for Human-AI Interaction takes two forms: a research paper and printable or interactive cards that include the guidelines and examples in industry. The 18 guidelines are divided between stages of user interaction in which each is most relevant (i.e. initial interaction, during a single interaction, when something goes wrong, and across interactions over time), with accompanying examples from products (i.e. voice assistants, e-commerce, autocomplete, recommendations, etc). The guidelines are coupled with examples of how they may be applied to various products. We were particularly interested in the card tool, which turned the research insights into a brainstorming tool that helps people understand and apply the guidelines.

Microsoft Human-AI Interaction Guideline Cards

Questioning the AI: Informing Design Practices for Explainable AI User Experiences (Liao et al. 2020)
https://arxiv.org/pdf/2001.02478.pdf

Liao et al. explored explainability from a user's perspective by creating a question bank and interviewing current UX professionals on their experience encountering these questions from users. Their findings provided concrete strategies for exploring various types of questions and explanations that can be used to aid explainability in AI products. They also produced a table that illustrated a taxonomy of explanation methods as well as definitions and relevant question types. We used the data provided in their taxonomy to create the brainstorming cards.

Category of Methods	Explanation Method	Definition	Algorithm Examples	Question Type
Explain the model (**Global**)	Global feature importance	Describe the weights of features used by the model (including visualization that shows the weights of features)	[41, 60, 69, 90]	**How**
	Decision tree approximation	Approximate the model to an interpretable decision-tree	[11, 47, 52]	**How**, Why, Why not, What if
	Rule extraction	Approximate the model to a set of rules, e.g., if-then rules	[26, 93, 102]	**How**, Why, Why not, What if
Explain a prediction (**Local**)	Local feature importance and saliency method	Show how features of the instance contribute to the model's prediction (including causes in parts of an image or text)	[61, 74, 83, 85, 101]	**Why**
	Local rules or trees	Describe the rules or a decision-tree path that the instance fits to guarantee the prediction	[39, 75, 99]	**Why**, **How to still be this**
Inspect counterfactual	Feature influence or relevance method	Show how the prediction changes corresponding to changes of a feature (often in a visualization format)	[8, 33, 36, 51]	**What if**, How to be that, How to still be this
	Contrastive or counterfactual features	Describe the feature(s) that will change the prediction if perturbed, absent or present	[27, 91, 100]	**Why**, **Why not**, **How to be that**
Example based	Prototypical or representative examples	Provide example(s) similar to the instance and with the same record as the prediction	[13, 48, 50]	**Why**, How to still be this
	Counterfactual example	Provide example(s) with small differences from the instance but with a different record from the prediction	[37, 55, 66]	**Why**, **Why not**, How to be that

Taxonomy of XAI methods mapping to user question types

IDEO AI Ethics Tool
https://www.ideo.com/blog/ai-needs-an-ethical-compass-this-tool-can-help

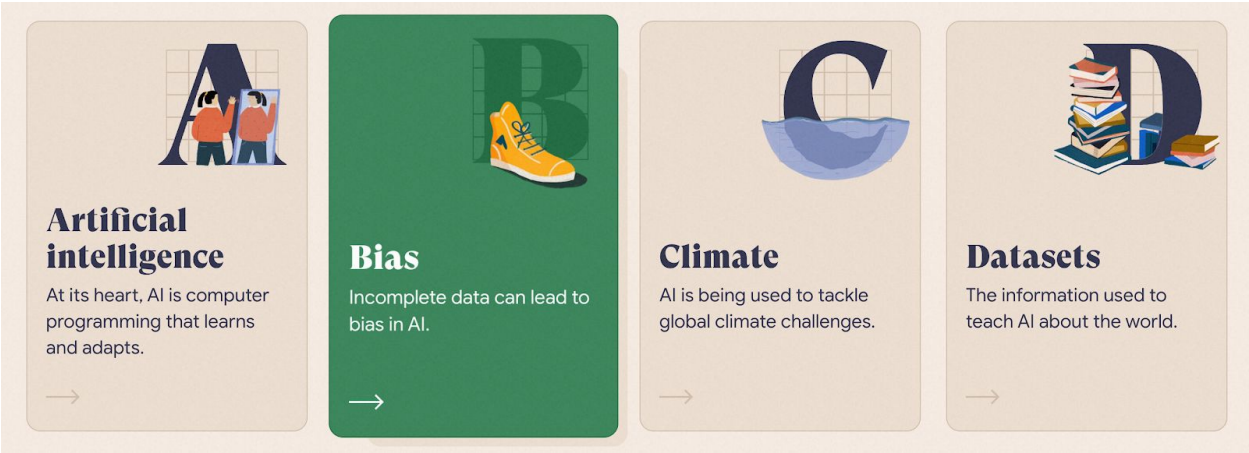
Designers often use tools from design agencies such as IDEO and frog to facilitate design explorations. IDEO created a tool specifically tackling AI ethics guidelines during the collaboration between designers, data scientists, and other internal stakeholders. Though the intent and the visual design of the tool are aligned with the vision for our project, we thought this particular form might not apply to a wide range of scenarios. The amount of text and the size of cards suggests that the people are most likely to use this tool in workshops rather than daily practice. A UX designer in usability testing also mentioned the pain point of being unable to utilize the brainstorming tool in their design practice in the industry. Therefore, we wanted to develop a practical brainstorming tool that AI product teams would like to keep referring to when they are developing explainability for their product.

IDEO AI Ethics Tool

The A-Z of AI
https://atozofai.withgoogle.com/

One of the examples of dividing complex concepts into multiple segments for a more approachable learning experience is "The A-Z of AI." Designed by Oxford Internet Institute and Google, the website successfully simplifies the basic concepts and applications of AI in a narrative that the general public

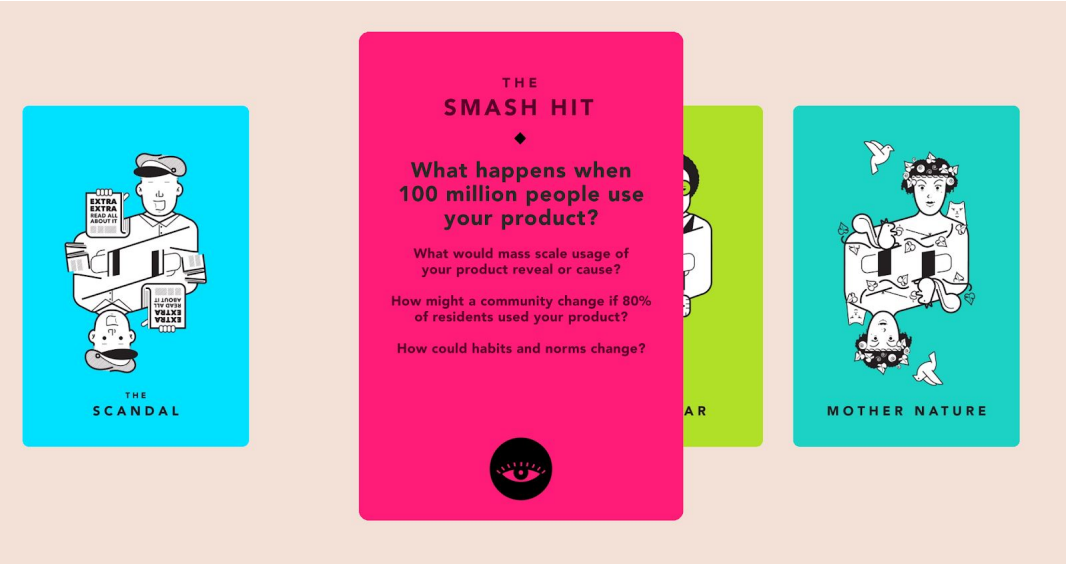
is familiar with. Moreover, the use of hand-drawn illustrations inspires us to explore more ways to increase the exposure of humanity throughout our storytelling of UXAI. The digital card format and interaction is also an influence of our brainstorming tool on the website.



Examples of illustrated thumbnails from The A-Z of AI

Tarot Cards of Tech
http://tarotcardsoftech.artefactgroup.com/

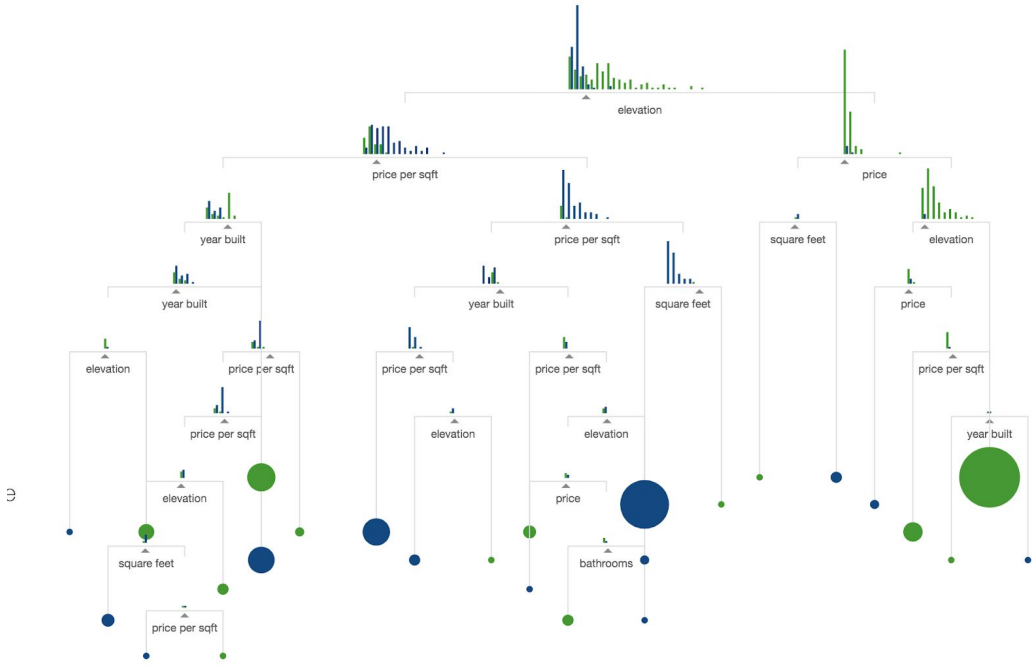
Tarot Cards of Tech demonstrates the power of digital card learning with animated illustrations for various scenarios. Though the browsing experience of the website is a little bit confusing, the visuals of the artifacts attract users' attention at first sight. However, the amount of text on the back of the cards after being selected by the user was not as attractive as the front. Thus, we also provided visual examples to describe the detailed information and examples of each card. We also referred to the interaction pattern of these cards in building our brainstorming tool.



The back of a card from Tarot Cards of Tech when selected

A Visual Introduction to Machine Learning
http://www.r2d3.us/visual-intro-to-machine-learning-part-1/

We were inspired by the way the authors were able to break down a complex concept like a decision tree and make it accessible to a non-technical audience. Our team had a similar goal in making Explainable AI accessible to UX professionals. We were particularly struck with the way the authors integrated "scrollytelling" to make both the narrative and visual elements digestible. We emulated this approach by creating a sequence of narratives interspersed with digestible, interactive visualizations to convey the complex concepts we were explaining.



Visual Introduction to Machine Learning Tree Diagram

Who Funds the World Health Organization
https://flowingdata.com/2020/04/30/who-funds-the-world-health-organization/

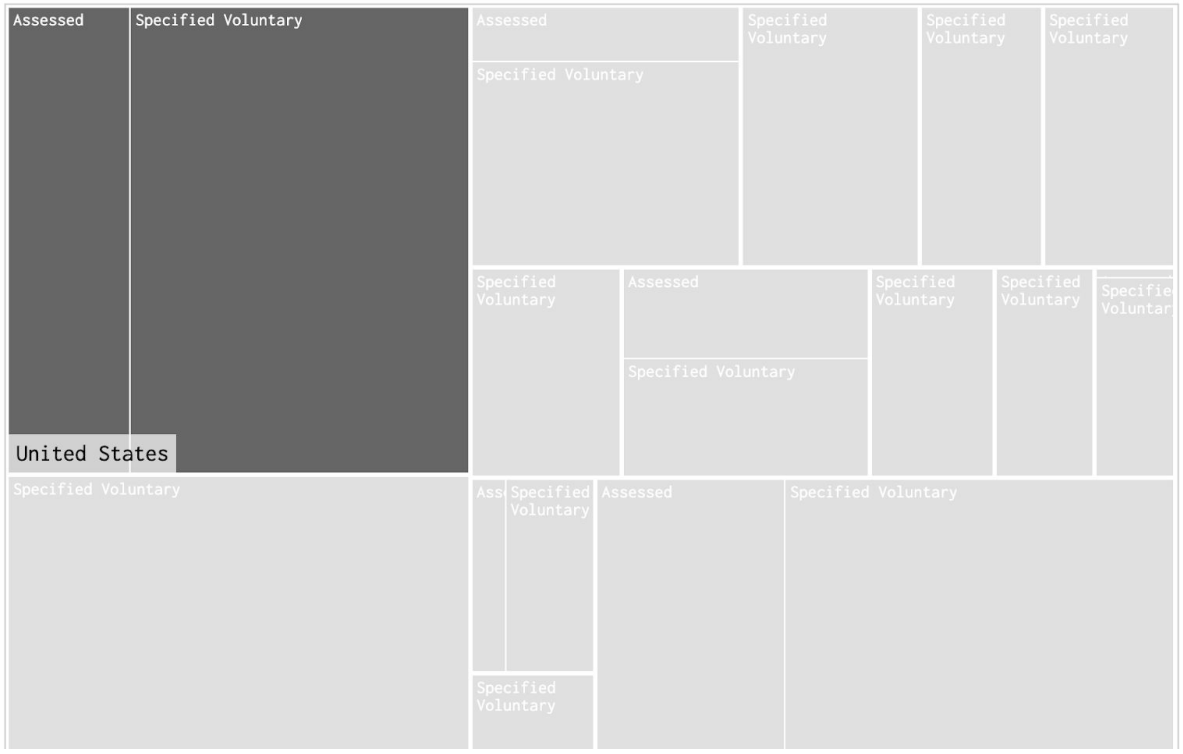
We were inspired by Nathan Yau's use of a tree map to break down the hierarchy of how the WHO is funded. In our first iteration, we emulated this approach when breaking down the different AI explanation types. However, our usability tests showed this was not the appropriate use case for a tree map. This is because the data in Nathan's project has both categorical and numerical data whereas our AI explanation types dataset was solely categorical. We then settled on a collapsible tree visualization because it maintained the interactive hierarchy we were initially inspired by in Yau's work while accurately conveying this categorical dataset.

FUNDING WITH AND WITHOUT THE UNITED STATES

Biennial budget for 2018-2019.

US Funding | No US Funding

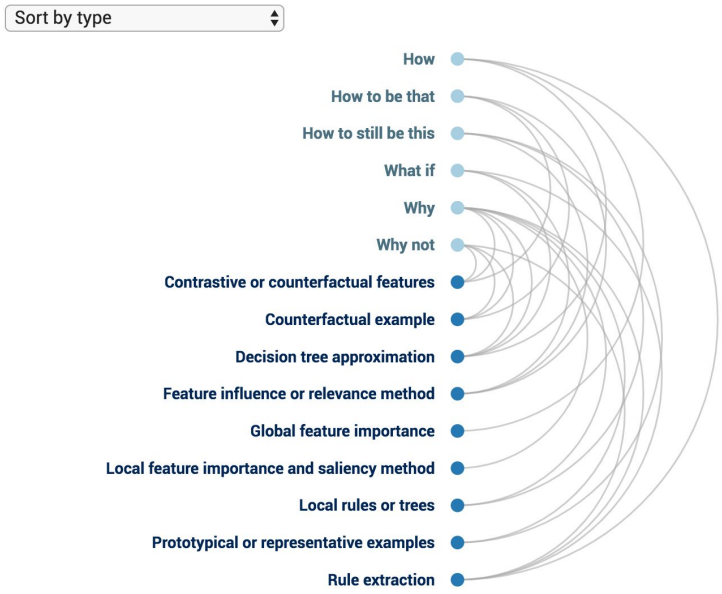
$5,623,603,000 Total



Who Funds the WHO Treemap

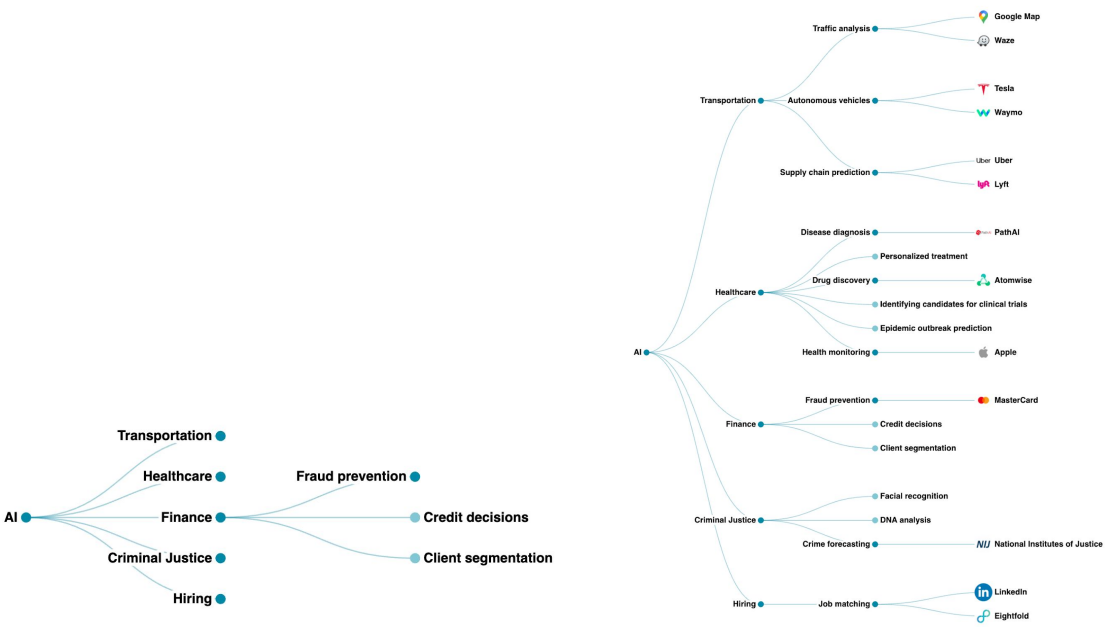
Visualization

Arc Diagram
In order to illustrate the connections between questions and explanation methods, we used an arc diagram. We included a filter so that users can sort by type, name, and number of connections. There is also a hover effect; when the user hovers over a node, its paths are highlighted to increase legibility of the diagram. By default the nodes are grouped by type: questions, and answers. These types are also coded by color.



`Connecting Questions to Explanation Methods`

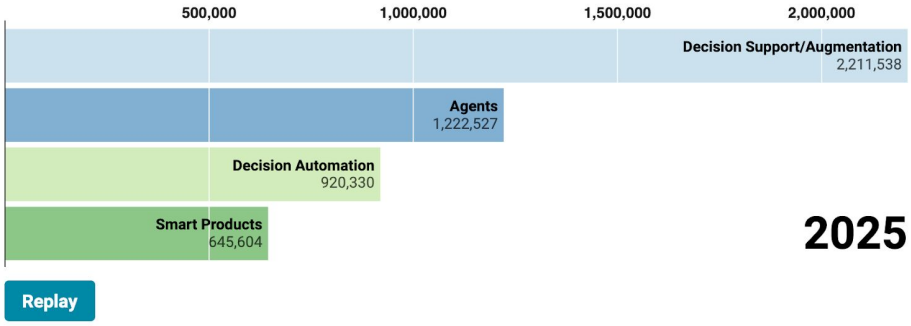
Collapsible Tree Diagrams
To encourage the audience to learn the basic knowledge of AI and explainable AI, we decided to present the information in multiple layers. By dividing the presented information through a collapsible tree diagram, we simplify the amount of data. Users can learn as they click through each node to create a more engaging experience. We hide the logos of the companies that utilized AI in their products until the user clicks on the last nodes to stimulate the feeling of getting rewards at the end of the journey.



Common AI Industry and Applications Collapsible Tree

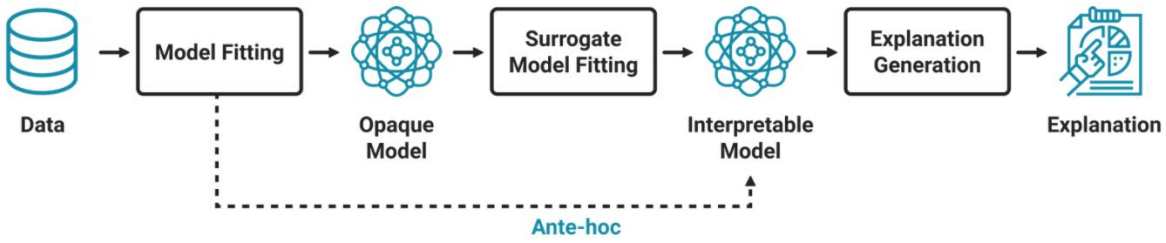
Racing Bar Graphs

Though the racing bar graph does not maximize the value of its form, we decide to keep the visualizations for engagement in the large amount of content in the website. Rather than a static bar graph showing the change of the predicted business value of AI, we wanted to use the animation of the racing bar graph to emphasize that AI-assisted decision will be the most valuable type of AI, where explainable AI is extremely important. Also, the change of visualizations to vary the visual narrative style of the website.



Worldwide Business Value Forecast by AI Type Racing Bar Graph

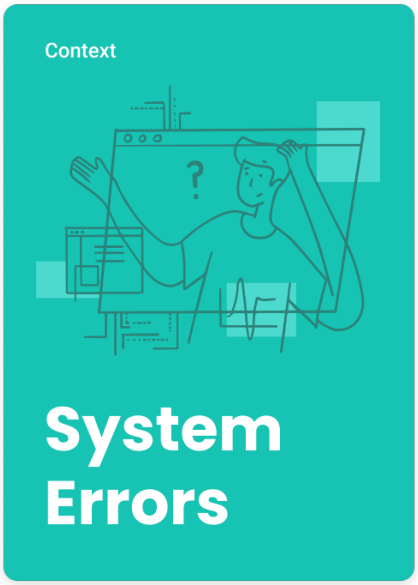
Iconography
To engage our target audience, we utilized iconography to not only illustrate different terms that explain the concepts of AI and explainability but also generate a dynamic flow during the reading experience.

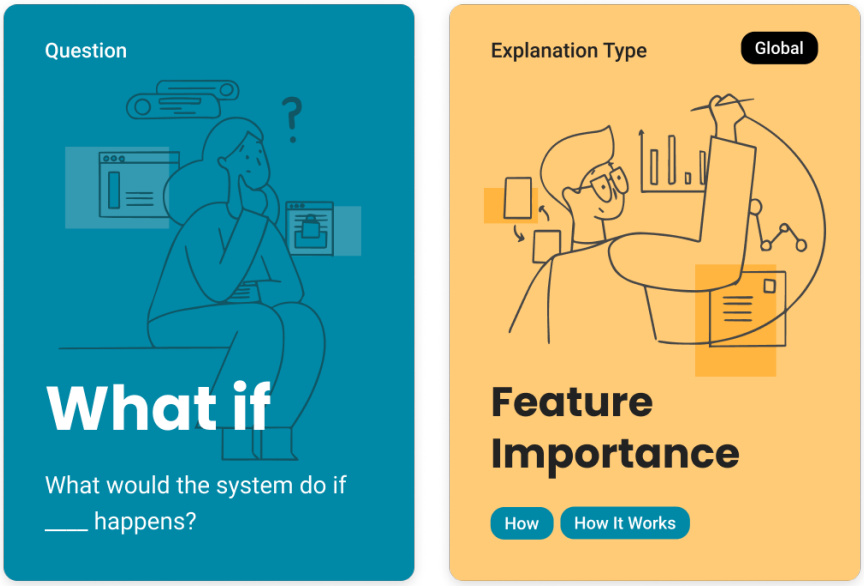


`Visual depiction of an ante-hoc explainability approach`

Brainstorming Cards

Illustration
For these cards that often have technical terminology on the front, we decided to create illustrations that add a touch of humanity to the front of the cards. For this iteration, the illustration demonstrates the scenario for the corresponding categories in addition to the color difference. However, we plan to create distinctive illustrations for each card to attract users to implement in their work.

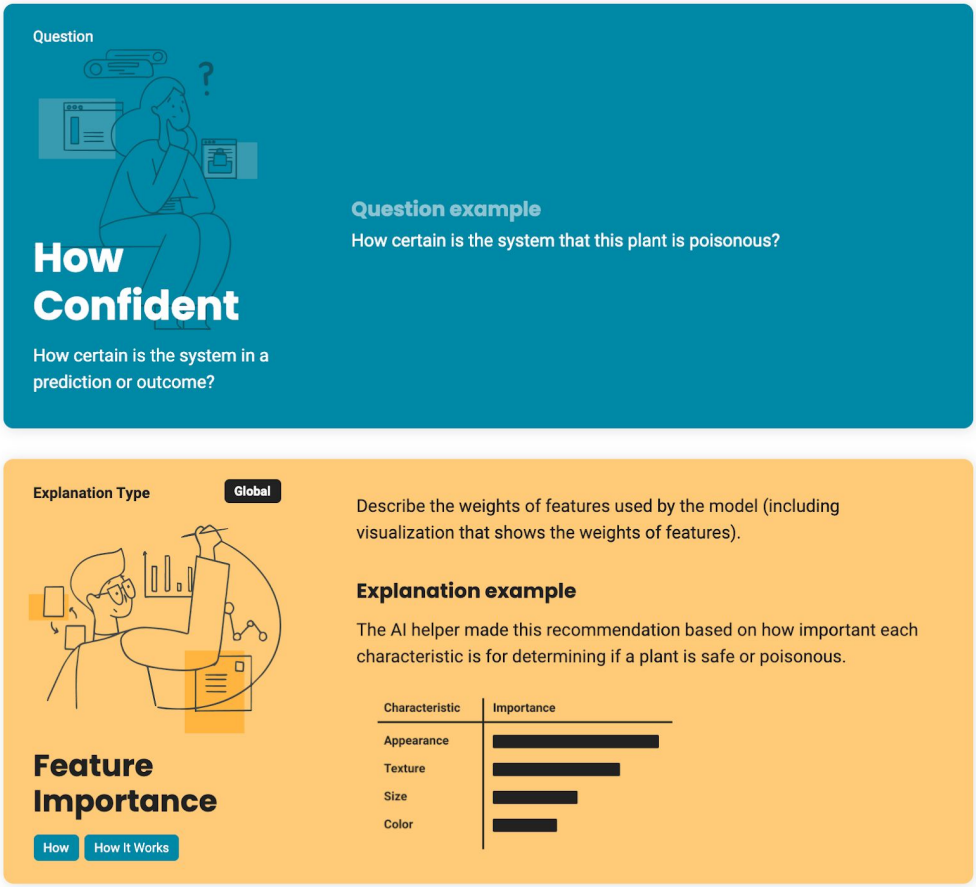




Brainstorming Cards

Interaction and Animation

To further convey how product teams can use the cards to brainstorm for different scenarios, we animated the cards so it will flip to the other side when selected. We also simplified the information presented at first glance and expanded into details when the cards are flipped to the back. Due to the high number of cards presented, we utilized a popping animation to emphasize which card is selected.



`Expanded cards`

Data

The information presented on the website was based on an extensive literature review of research papers and industry resources that discuss explainability in AI. The list of works used are featured on the website.

Tools

- D3 via Observable Notebook
    - We built our D3 visualizations in Observable and then used Jeremy Ashkenas' Handy Embed Code Generator to bring our visualizations into our Webflow website. One challenge we had to work through was getting our styles to work on Webflow since Observable exports its own stylesheet by default, overwriting any styles in our D3 code. We solved this by moving all our styles into separate cells from our visualizations which we then embedded in addition to our visualization cells.
- Illustrator

    - We used Illustrator to illustrate the scenario that people can apply the brainstorming cards to. It is also used to edit and recreate the graphs that we collected from our literature review to fit our narrative.
- Figma
    - We used Figma to collaboratively design the website and the brainstorming cards. We also conducted usability testing through the prototype on Figma.
- Webflow
    - Webflow is a website builder similar to Squarespace that we used to construct our website.

Usability Testing

Objective

The purpose of the study was to collect feedback on the usability interactive elements, the legibility of the content (appropriate and understandable terms, etc.), the logic of the information architecture, and whether or not participants would utilize the website as a tool in their own practice. Specifically, we wanted to understand if:

1. Designers could learn about AI and XAI through our resource
2. Our toolkit would be a useful addition to their workplace

Method

We employed qualitative usability testing as a means to address the rationale above. The focus of qualitative usability testing is to understand how participants perceive the user interface, and then deliberately probe to understand why. We asked each participant several simple prompts throughout each section of the website, then allowed participants to explore the prototype naturally, while thinking out loud, to discover how intuitive the features and language were.

Participants

Participants included 3 UX designers with experience designing AI products at large technology companies. One participant had additional experience of creating AI design guidelines, specifically for explainability. They were selected from the researchers personal networks.

Scenarios/Tasks

We had a single task scenario: we asked participants to imagine they were using our prototype for a project at work that involved designing for AI. We looked for points of confusion, clarification, excitement, and comments about usefulness.

Test Measures

For each section, we measured the following items:

- First impressions, and how participants initially make sense of the section
- What participants expected to see, why, and what they expected to see but was missing

- Terms and concepts that are unclear
- Understanding of the navigation, information architecture
- Recognition of interactive elements

Additionally, we measured the perceived usefulness of the overall website, as well as the anticipated usefulness of the toolkit.

Results

Overall, we found that the language and terms used throughout either needed to be defined or removed because they were too technical, and that we need to be more upfront about how designers will benefit from using our resource. For example, one participant said "I don't know what Explainable AI -- is that a noun?", which clearly indicated they didn't understand the terminology that is central to our resource. Another participant who works in the explainable AI field even said "if I'm brand new I don't know what 'Explainable AI' is."

All three participants wanted a combination of a hook, explaining why Explainable AI matters, as well as a clear understanding of how they will personally benefit from our website. This desire showed up early, with one participant saying "If i need to figure out 'how recommendation systems work' I don't know where to start", referring to the information architecture on the home page. The same participant noted that "I don't know why this is important or why it's a main consideration for me as an AI designer", referring to Explainable AI and wanted a case study or a similar hook that illustrates "this happens because systems are not explainable and crazy bad things happen". This sentiment was experienced by all three participants, with the other two asking for a case or scenario that had bad outcomes for the company as a way of showing why Explainable AI matters.

All participants were excited when they discovered the resource page, with the who, what, why, when, etc. as well as the cards. However, all participants mentioned that we should put this front and center since these tools are what matter most for designers. As one participant said, "you are burying the lead" by having the cards come last.

All participants understood how they could interact with the information visualizations. For the tree diagram, they expected that clicking on a node would expand the diagram. For the tree map, they expected that clicking on a section would reveal more detail, however it was not immediately intuitive. Furthermore, they were more interested in seeing all options upfront. Finally, for the arch diagram, they understood how to use the filter, and asked if clicking on a node would reveal more information.

Other key takeaways from the pilot study are below:

- Re-frame website: Lead with the toolkit and integrate throughout, why XAI matters, follow with the educational resources
    - Frame it as saving designers time, they don't have to read all the academic literature, etc).
- Rename sections to be more descriptive and intuitive
- Visuals need to stand on their own, try to reduce amount of text
- Keep stakeholders / steps 1-4 section but make skimmable and help designers think through prioritization / tradeoffs
- Re-order and refine visuals

- Toolkit: great for workshops with stakeholders, hone UX writing, is there a way to link the two types together, to see more concretely?

Discussion

From these results we plan on refining the information visualizations, reorganizing the information architecture, inverting the order so users know the value of the site immediately, and adding a case study or scenario upfront to demonstrate why users should care about Explainable AI. We will review the overall narrative to ensure a clearer distinction and smoother flow between sections. We will also explore ways to integrate our toolkit throughout the site and identify sections that can benefit from having visuals and/or further clarification. Finally, we intend to revisit the explanation taxonomy (presented as a zoomable treemap) to make it more intuitive and enable viewers to see all explanations upfront.

Future Work

Future work includes more usability testing with a broader audience. In our initial study we chose designers who work on AI products at large technology companies, however, we are curious to test our site with designers from smaller start-ups as well as designers with little to no experience with AI. We are also interested in hosting an ideation workshop with cross-functional teams to study the effectiveness of our brainstorming tool: what works, what doesn't work, and what is missing. Finally, we intend to refine the website to include more visual aids and interactivity and add scenario cards to the toolkit to facilitate learning.

Appendix

Links to Website and Observable Notebooks

UXAI Website: https://www.uxai.design/

XAI Question and Explanation Types Arc Diagram:
https://observablehq.com/@mkaushik92/info-247-d3-assignment

Common AI Industry and Applications Collapsible Tree:
https://observablehq.com/@mutihuang/common-ai-industry-and-applications-collapsible-tree

Worldwide Business Value Forecast by AI Type Racing Bar Graph:
https://observablehq.com/@mutihuang/worldwide-business-value-forecast-by-ai-type

Explanations for AI Tree:
https://observablehq.com/@victor-grajski/explanations-for-ai-tree/2

Individual Contributions

Names	**Tasks**
Meena	Literature Review (50+ academic papers and industry resources), Content Writing, Arc Diagram, Design support
Mu-Ti	Visual Design, Illustration, Website Construction, Collapsible Tree Diagram, Racing Bar Graph
Victor	Website Construction, Collapsible Tree Diagram, D3 Troubleshooting, Observable Notebook Implementation

Usability Testing Stimuli

A VISUAL
INTRODUCTION TO
EXPLAINABLE AI
FOR DESIGNERS

uxai

This resource was created to familiarize user experience (UX) designers with the field of Explainable Artificial Intelligence (XAI) and support a human-centered approach to designing AI products and services.

Introduction
Get familiar with the landscape of Artificial Intelligence (AI).

Explainable AI
Understand the background and key concepts of eXplainable AI (XAI).

Explanation Guide
Learn how to explain an AI system in order to improve the user experience.

Toolkit
Experiment and brainstorm different explainable AI scenarios.

Glossary
Key terms and concepts in explainable AI.

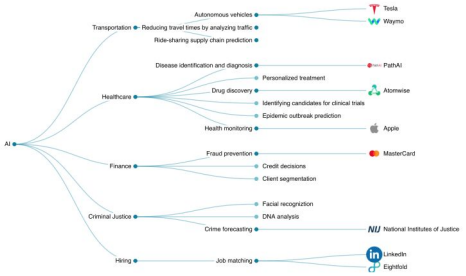
AI in the Industry

With the development of AI becoming more efficient and more reliable, countries and companies are increasingly investing in AI. Autonomous vehicles, chatbots, fraud detections, AI is being applied to a wide range of industry to automate and augment productions and decision making.

$2.9 Trillion
In 2021, artificial intelligence (AI) augmentation will create $2.9 trillion of business value

Business Value Forecast by AI Type

Common AI Decision Support Applications



Where do I get started?

Explore how different intelligibility and explanation types can apply to this scenario by browsing and selecting a card below. Note that some explanation types may be better suited for this scenario based on the user, their goals, and the underlying model.

You can also download these cards to explore how they may be applied to your work and/or to other scenarios.

All | Intelligibility Types | Explanation Types

INTELLIGIBILITY TYPE

What

What did the system do?

INTELLIGIBILITY TYPE

Why

Why did the system do ___?

INTELLIGIBILITY TYPE

Why Not

Why did the system not do ___?

INTELLIGIBILITY TYPE

What If

INTELLIGIBILITY TYPE

How

INTELLIGIBILITY TYPE

How To

Explanation Guide

Learn how to explain an AI system in order to improve the user experience.

Explanation Strategy

Based on industry and academic research, we propose the following strategy for determining what kind of explanation to use.

01. WHO are your users?

Identify the distinct groups of people who are interested in explanations from your AI system and understand the nuances within these groups. There will likely be varying degrees of factors such as expertise, attitudes towards AI, and self-confidence, all of which can influence trust and how people interact with the system. By identifying and understanding your users, you can ensure the explanation matches their complexity capability and is best tailored to their domain.

There are typically four distinct groups to consider:

Decision Makers

Decision makers are people who use the recommendations of an AI system to make a decision, such as a physician or loan officer.

Affected Users

Affected users are people who are impacted by the recommendations made by