Designing for AI

Interface considerations for advanced technologies

By Erica Gunn

UXPA Boston, 05.10.24

Slides are at http://ericagunn.com/community/

Background images generated with Adobe Firefly

Agenda

What is AI?

When should we use it?

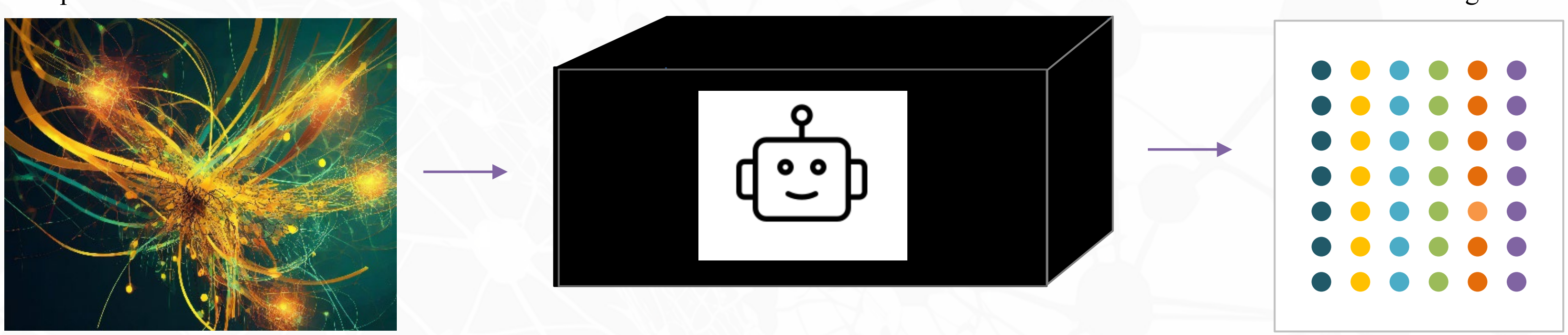
Why is it challenging to design for?

What are the design considerations?

What is AI?

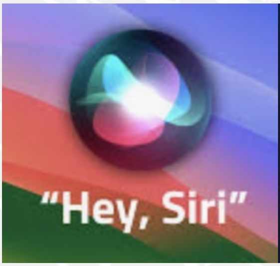
AI is a technology that helps to convert complicated data into meaningful information.

Most AI can improve its performance based on experience.

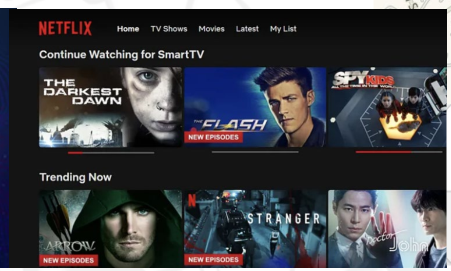


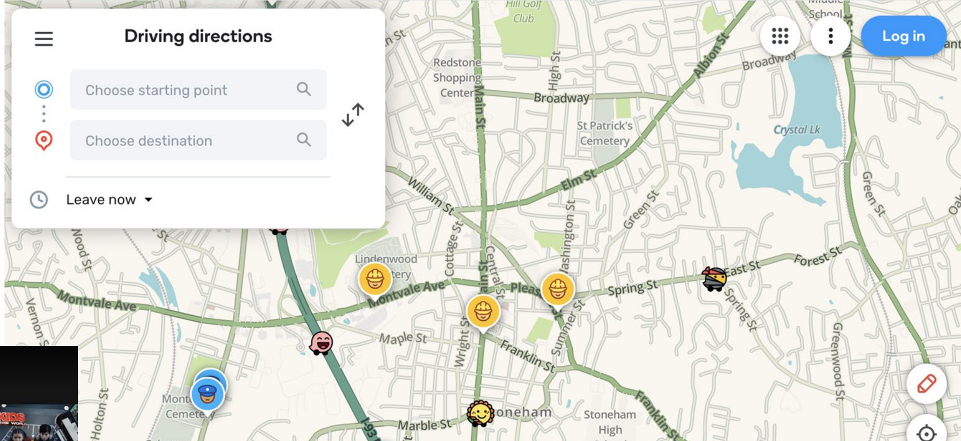
- Finding connections in big datasets
- Processing complex patterns
- Developing predictions based on previous results

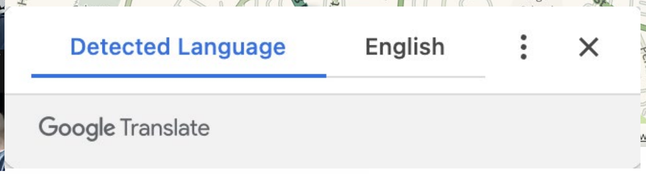
AI in action

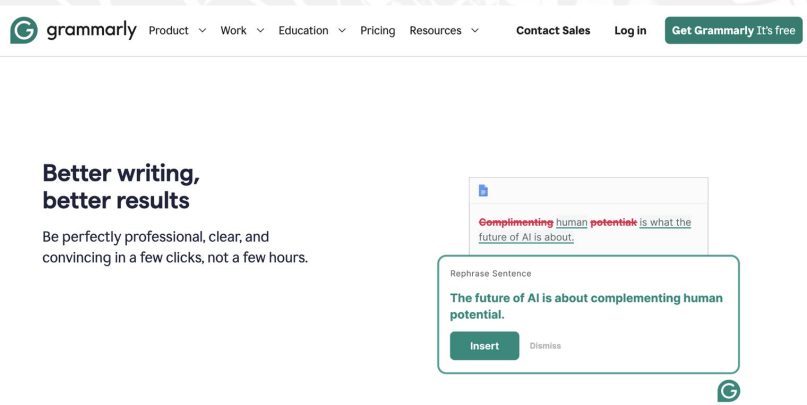

"Hey, Siri"

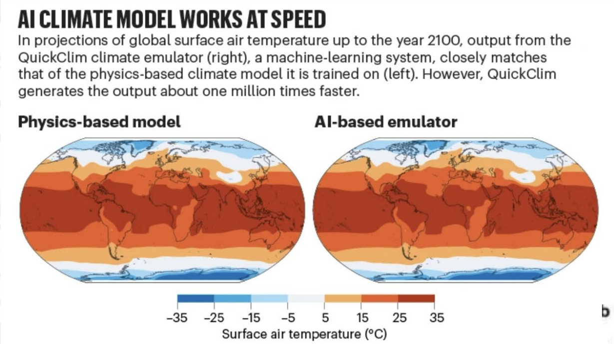

alexa

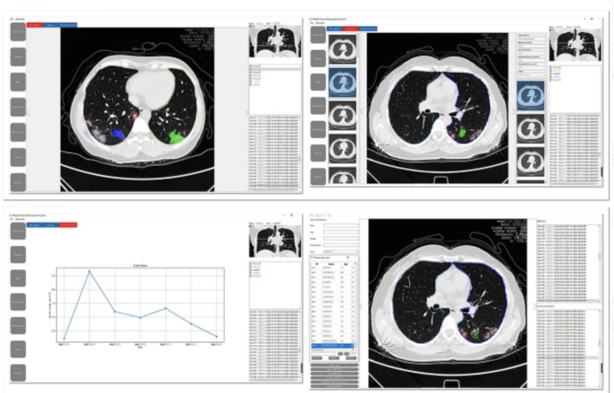

NETFLIX
Home
TV Shows
Movies
Latest
My List
Continue Watching for SmartTV
Trending Now

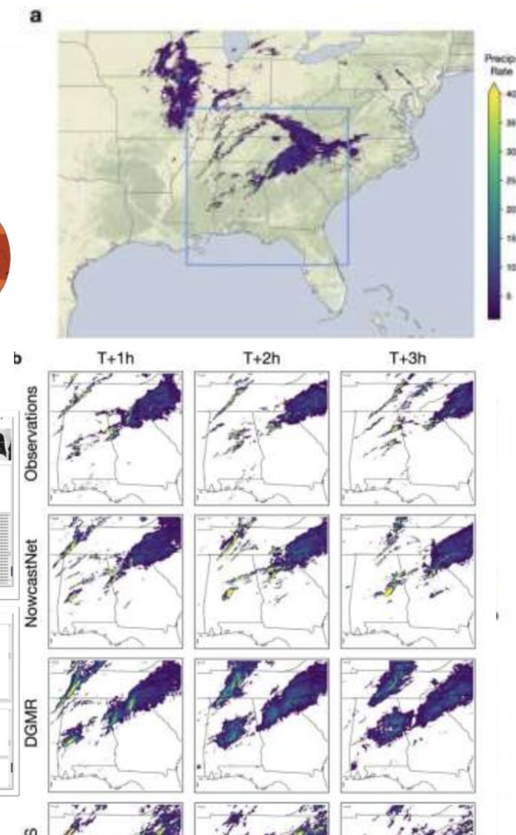

Driving directions
Choose starting point
Choose destination
Leave now
Log in

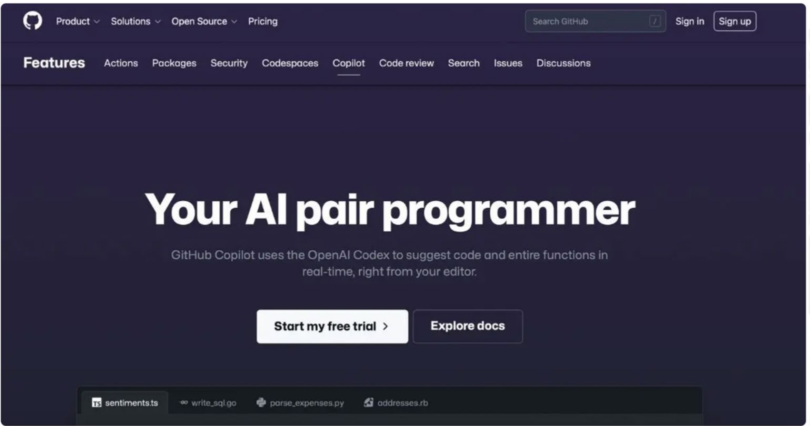

Detected Language
English
Google Translate


grammarly
Product
Work
Education
Pricing
Resources
Contact Sales
Log in
Get Grammarly It's free
Better writing, better results
Be perfectly professional, clear, and convincing in a few clicks, not a few hours.
Complimenting human potentiak is what the future of AI is about.
Rephrase Sentence
The future of AI is about complementing human potential.
Insert
Dismiss


AI CLIMATE MODEL WORKS AT SPEED
In projections of global surface air temperature up to the year 2100, output from the QuickClim climate emulator (right), a machine-learning system, closely matches that of the physics-based climate model it is trained on (left). However, QuickClim generates the output about one million times faster.
Physics-based model
AI-based emulator
−35 −25 −15 −5 5 15 25 35
Surface air temperature (°C)

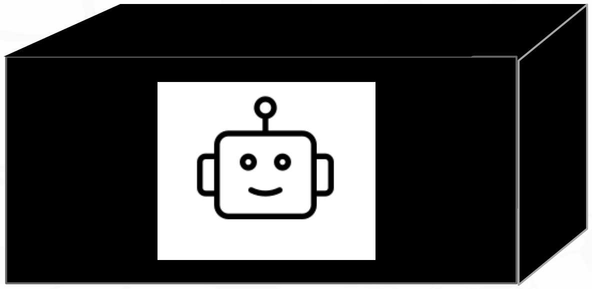



a
b
T+1h
T+2h
T+3h
Observations
NowcastNet
DGMR


Product
Solutions
Open Source
Pricing
Sign in
Sign up
Features
Actions
Packages
Security
Codespaces
Copilot
Code review
Search
Issues
Discussions
Your AI pair programmer
GitHub Copilot uses the OpenAI Codex to suggest code and entire functions in real-time, right from your editor.
Start my free trial
Explore docs

When do we use AI?

- Assistant for **tedious or repetitive tasks** (image search, inputting info into forms)
- **Skill or knowledge extension** (writing, translation, recommendation engine, code or other advanced skills)
- Human-machine collaboration for **advanced decision making**



“Black box problems”: How do we know what’s happening inside?

Familiar Black Boxes

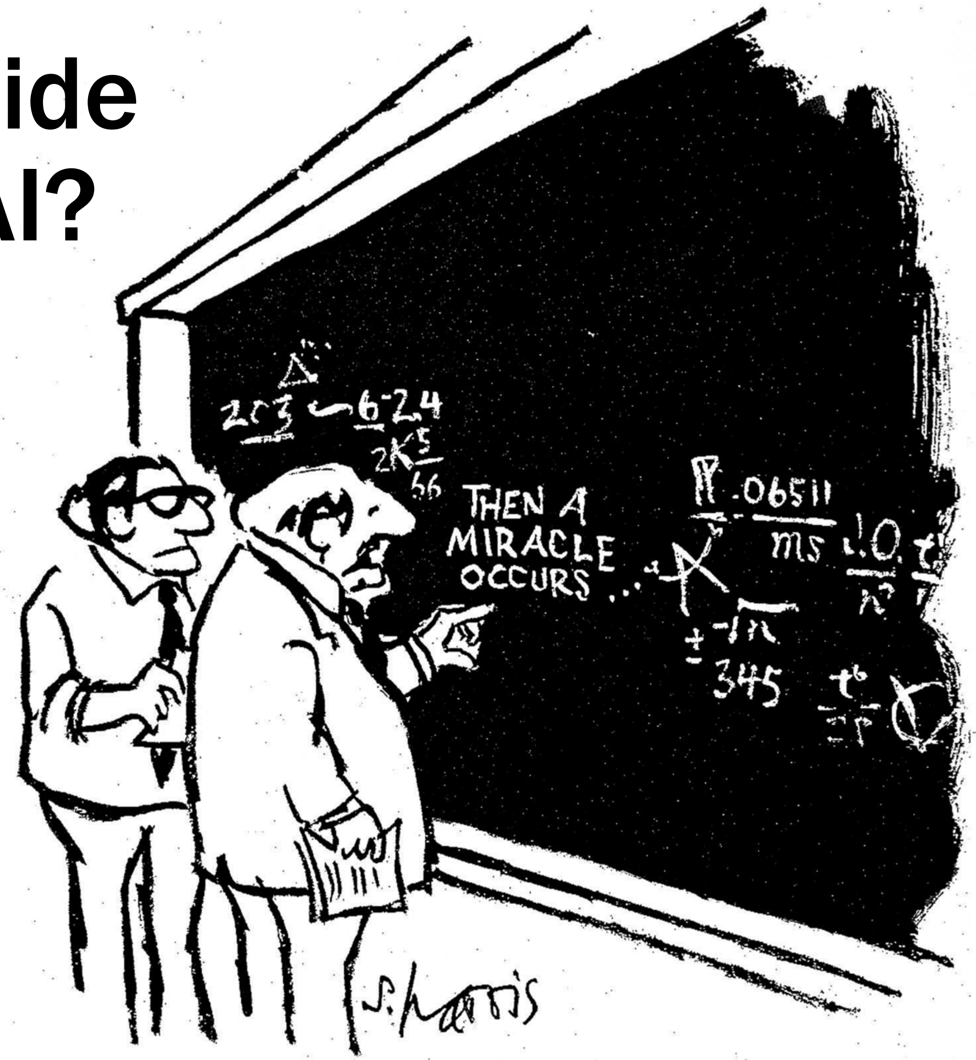
Do you *really* know what's going on inside these things?

We all have plenty of black boxes in our lives. They only tend to worry us when they are unfamiliar or unpredictable, or when we think that no one understands what's going on.

How do users decide whether to trust AI?

Understanding how users evaluate and build trust with a new technology can help us to build better tools to support those needs.

- Familiarity
- Experience with good outcomes
- Switching between perspectives: “explanation” for a single data point compared against local and global outcomes
- Cognitive fit



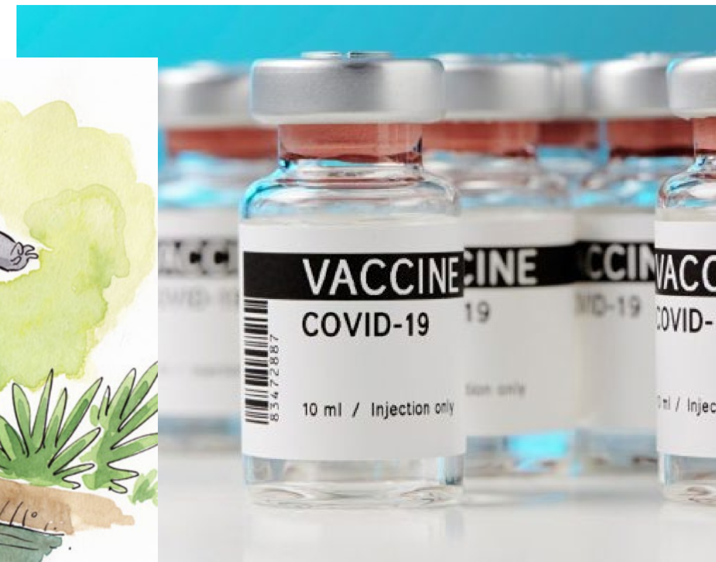
"I THINK YOU SHOULD BE MORE EXPLICIT HERE IN STEP TWO."

Image Credit: Sydney Harris

The perils of trust

Too much

Not enough



We really want to give our users **calibrated confidence** by building interpretable systems.

What do we really mean by “trust”?

There are different levels of trust:

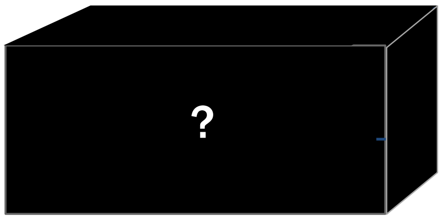
- Trust in the model
- Trust in the decision-maker
- Trust in your own ability to understand

Types of trust:

- Relationship-based
- Evidence-based

Be careful when interpreting user feedback: people
like easy answers and simple solutions, but they often need
depth to make good decisions

Levels of information



Black box - no idea what's in there

Transparent - I can see what's going on, but don't necessarily understand

Explainable - I have some explanations that seem to cover most things, but I'm not sure that they're right

3	42	1
5	82	3
6	84	2
4	?	1

Interpretable - I can figure out (and test) how it works.

Knowing the answer may or may not help me much.

Confidence - I can count on the system to provide me with meaningful and useful information.

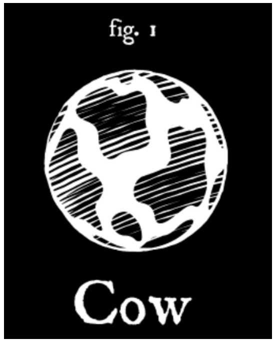
System behaviors/outputs are relevant and appropriate to my needs, with the appropriate level of control, detail, and visibility for me to use it safely for important tasks.

Solution Complexity

As simple as possible

...but not simpler

!==



Humans like simple answers.

Relying on user feedback may lead to persuasive but misleading explanations.

Audience matters: novice users often prefer simple explanations.

Why is all of this important?

Legal right to explanations and transparent systems (GDPR, etc)

Move from “black box” to **participatory decision-making**

User control is a basic principle to **support accountability**

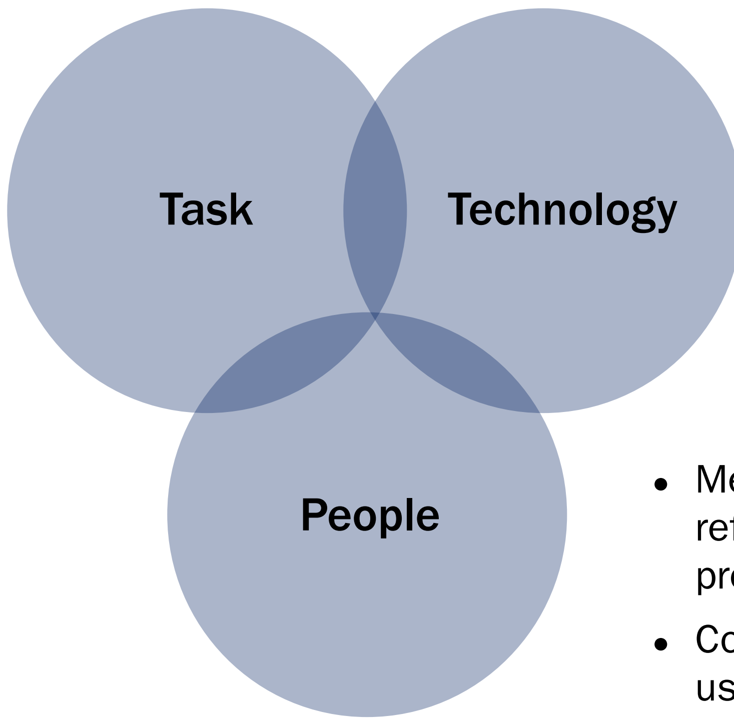
Support users in making fully **informed decisions**

Reducing bias and negative impacts

Why is interpretability hard?

There are many challenges when designing interpretable AI. These come from different sources, and require different solutions and approaches to solve.

- Representing a sophisticated decision-making process with many dependencies and important tradeoffs
- Projecting a high-dimensional variable space to a simplified framework; need to align to user's expectations



- Complexity of model inputs
- Structural aspects of models

- Mental models, frames of reference, decision making preferences and approaches
- Cognitive load and managing user's mental limits

Provide Context

What does this data mean?

A user needs to be able to make informed decisions based on the data

- **Provide enough detail** (and the right kind) to answer the important questions
- **Establish traceability** to communicate logical sequence and decision points
- **Present multiple views** of the data
- Create tools to **refine results** and **provide granularity**
- Help the user **assess quality and certainty**
- Establish expected behavior and **emphasize outliers**

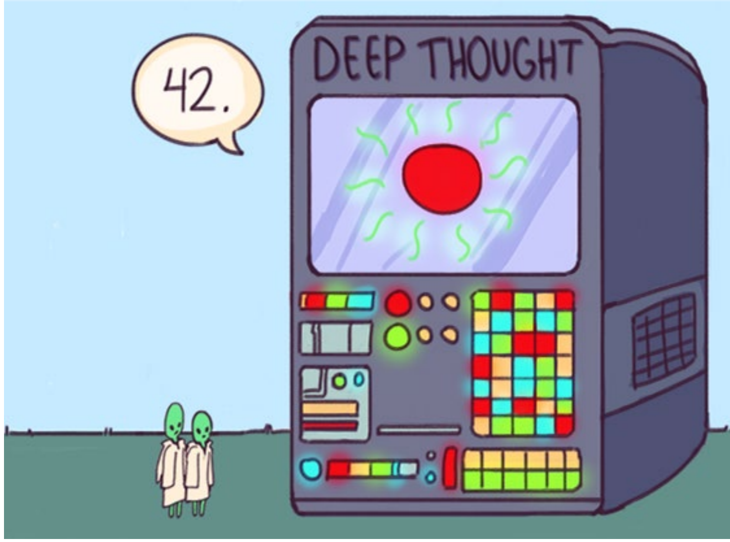


Image Credit:
breakingsmart.com

Types of Explanation

There are many formats and approaches for explanation.
In general, a mix of approaches is best.

- **Ad hoc explainers (rationalizing)** – Explain why but not how a decision was made.
- **How explanations** – Give a holistic representation of ML algorithm. Users may also develop a “how” mental model from a collection of instance explanations.
- **Why explanations** – Identify features of input data or model logic that contributed.
- **Why not explanations** – Show why this item or feature wasn’t included in the output. Explain why the model output may be different than the user’s expectations.
- **What if explanations** – Allow experimentation to see alternate outcomes.
- **How-to explanations** – Explain how to adjust model inputs to get a desired outcome. Supports development of user’s mental model.
- **What else explanations** – Identify comparable results elsewhere in the representation space. These are popular, but can be misleading if the data distribution is non-uniform.

Constructing Knowledge

How do we make sense of all this information?

Once we open the black box, users need to understand what's inside, and what to do with it.

- Compare to familiar reference points
- Provide examples, real-world use cases, and training / sample data showing successes
- Ultimately, there is no substitute for experience over time

Don't be afraid to let your user "under the hood." But also remember that you need to guide them when they get there.

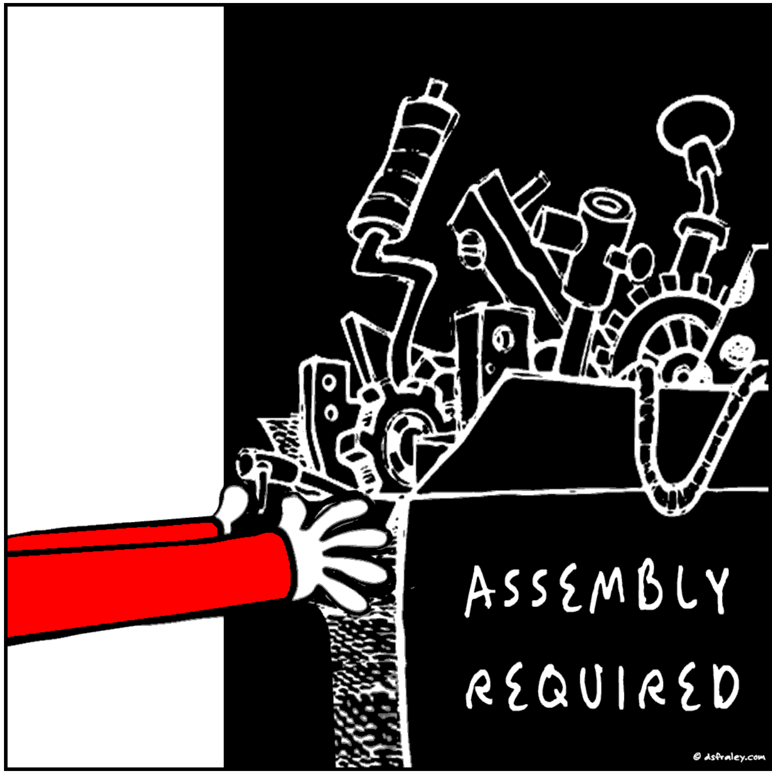


Image Credit: D.S. Fraley

Interpretability is a full-stack effort

Interpretability and user confidence result from thoughtful effort at each stage of model development and deployment

- **Data science:**
  - Design technical algorithms to be interpretable, and check for bias.
  - Design (visualization) tools that allow data scientists to explore and understand the models during development, and to test for alignment to user task.
- **Product design:**
  - Build software that allows users to submit jobs and interact with models. Provide mechanisms to test the reliability of the output. Advanced designs may even allow users to train the model themselves, where appropriate.
  - Design user workflows, interactions and interface text to be clear to users.
- **Documentation and client communication:** design “explainers” or other documentation to simplify models for a general audience.

Summing up

AI is a tool, and it shouldn't be a mystery.
Principles of good UX still apply.

- Identify data and AI methodologies that **support meaningful analysis**
- Open up the black box: **let people see what's inside**
- Help users **know what's important, and why**
- **Support users in testing** and evaluating AI-generated solutions
- **Put information in context, with enough detail** to support informed decision-making

Keep in touch!

www.ericagunn.com

Erica Gunn
(please include a message)

@Erica.Gunn

@EricaGunn
#DataVisualizationSociety
Nightingale Magazine

Slides are at: http://ericagunn.com/community/

Bibliography

A Multidisciplinary Survey and Framework for Design and Evaluation of Explainable AI Systems (2021). Mohseni, S., Zarei, N., & Ragan, E. D. *ACM Transactions on Interactive Intelligent Systems*, *11*(3–4).

The Role of Interactive Visualization in Fostering Trust in AI (2021). Beauxis-Aussalet, E., Behrisch, M., Borgo, R., Ebert, D., El-Assady, M., Keim, D. A., Riveiro, M., Schreck, T., Strobelt, H., & van Wijk, J. J. *IEEE Computer Graphics and Applications*, vol. 41, no. 06, pp. 7-12.

A Review of User Interface Design for Interactive Machine Learning (2018). John J. Dudley and Per Ola Kristensson. *ACM Trans. Interact. Intell. Syst.* 8(2), Article 8.

Designing for Confidence: The Impact of Visualizing Artificial Intelligence Decisions (2022). Karran AJ, Demazure T, Hudon A, Senecal S, Léger PM. Front Neurosci. Jun 24, Vol 16, 883385.

The State of the Art in Enhancing Trust in Machine Learning Models with the Use of Visualizations (2002). Chatzimparmpas, A., Martins, R., Jusufi, I., Kucher, K., Rossi, F., & Kerren, A. *Computer Graphics Forum*, *39*(3), 713–756.

Techniques for Interpretable Machine Learning: Uncovering the mysterious ways machine learning models make decisions (2020). Du, M., Liu, N., & Hu, X. *Communications of the ACM*, *63*(1).

explAIner: A Visual Analytics Framework for Interactive and Explainable Machine Learning (Jan. 2020). T. Spinner, U. Schlegel, H. Schäfer and M. El-Assady. *IEEE Transactions on Visualization and Computer Graphics*, 26(1), 1064-1074

"Why should I trust you?" Explaining the predictions of any classifier (2016). Ribeiro, M. T., Singh, S., & Guestrin, C.. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *13-17-August-2016*, 1135–1144.

Interactive visualization for Fostering Trust in ML (2023). Polo Chau, Alex Endert, Daniel A. Keim, and Daniela Oelke. In Dagstuhl Reports, Volume 12, Issue 8, pp. 103-116, (Dagstuhl Seminar 22351) Schloss Dagstuhl - Leibniz-Zentrum für Informatik.

A survey of visual analytics techniques for machine learning (2021). Yuan, J., Chen, C., Yang, W. *et al.Comp. Visual Media* **7**, 3–36.

Bibliography

Interactive knowledge discovery and knowledge visualization for decision support in multi-objective optimization (2023). Henrik Smedberg, Sunith Bandaru. *European Journal of Operational Research.* 306 (3), 1311-1329.

Enhancing Trust in Machine Learning Models with the Use of Visualizations (2020). Chatzimparmpas, A., Martins, R.M., Jusufi, I., Kucher, K., Rossi, F. and Kerren, A., The State of the Art in Computer Graphics Forum, 39: 713-756.

Knowledge visualisation for strategic decision-making in the digital age (2022). Schiuma, G., Gavrilova, T., & Carlucci, D.. *Management Decision*, *60*(4), 885–892.

The Role of Uncertainty, Awareness, and Trust in Visual Analytics (2016). Sacha, D., Senaratne, H., Kwon, B. C., Ellis, G., & Keim, D. A.. *IEEE Transactions on Visualization and Computer Graphics*, *22*(1), 240–249.

CueTIP: A mixed-initiative interface for correcting handwriting errors (2006). Shilman, Michael & Tan, Desney & Simard, Patrice. UIST: Proceedings of the Annual ACM Symposium on User Interface Software and Technology. 323-332.

Understanding what machine learning produces - Part II: Knowledge Visualization Techniques (1996). Cunningham, S. J., Humphrey, M. C. & Witten, I. H. Hamilton, New Zealand: University of Waikato, Department of Computer Science. Computer Science Working Papers 96/21.