



Advancing Explainability Through AI Literacy and Design Resources

Insights

- Explainability is best served when information about AI is incorporated into the entire user journey and AI literacy is built continuously throughout a person's life.
- We should encourage product teams to think about all the moments and ways they can help people understand how AI operates and makes decisions.

Britni expertly navigates her kayak down the narrow channel, creating a new route to share online (Figure 1). Navigator, her boating app, lets her post routes and gives her a percentage of the advertising revenue. As Britni rounds a bend, a notification comes in from Navigator: "Your route 'Marshy Inlet Trek' has been suspended due to safety concerns and will be hidden temporarily from users." Shocked that her most popular and lucrative (and in her experience very safe!) route has been suspended, Britni begins the process of investigating and contesting the suspension...

Navigator is a fictional app we

designed to teach good practices for explainability. Explainability, put simply, provides human-understandable reasons and context for decisions made by an AI system [1,2]. In so doing, explainability benefits users and society by helping individuals make informed decisions about their use of AI systems, empowering civic engagement with AI, informing policymakers about the impacts of AI, and more. For these reasons, as AI has become more central in people's lives, the public, the tech industry, regulators, and others have increasingly recognized explainability as a key aspect of



Figure 1. An excerpt from a visual story about a user's experience with a fictional boating app.

responsible innovation [3,4]. Recent advances in generative AI have further heightened interest in providing explanations.

Many researchers and practitioners are doing important work on explainability and related topics. Some focus on interpretability, creating tools and techniques that make it easier to precisely identify specific factors that influence a decision. Others focus on providing what we call *in-the-moment explanations* [5] delivered right when AI makes a decision (e.g., a medical diagnosis, whether someone qualifies for a loan, or which video to

recommend next). In-the-moment explanations should be helpful, especially when carefully designed, by providing relevant, specific context for the decision that was just made.

AI systems are often so complex, however, that we can't rely on in-the-moment explanations entirely. It's just too much information to pack into a single moment. Because of this, we think it's critical to move beyond in-the-moment explanations to a more holistic view of how users think about AI decisions and inferences. *Our approach to explainability emphasizes educating people about AI and how it affects their lives, not just at the moment*

of a decision but also throughout their experiences with AI systems and more broadly throughout their lives. From a tactical perspective, this means making thoughtful choices not just about what information people need about AI, but also when and how to present this information. Informed by our experiences working with product teams and our research on comprehension and attitudes toward AI [6], we focus on three placements: in-the-moment, throughout the product, and beyond the product.

First, construct well-designed *in-the-moment explanations*. While much work on explainability has

focused on in-the-moment explanations, in current practice, these explanations are often poorly executed from a user experience perspective. For example, they can contain unnecessary details, be overly technical, or repeat similar information for every decision even when the decisions and reasons are very different. They often also fail to take into account that a user may be distressed or in a hurry at the moment of a decision; this is not a good time to overload users with extra information or force them to develop an understanding of how the product operates in order to respond to an important decision that affects their welfare. Better design can greatly improve in-the-moment explanations. One of the first steps to improving them is identifying information that is meaningful and helpful in the current moment, and removing information that isn't critical at the moment of decision and moving it elsewhere in the user experience. Another important step is to equip people to act if needed; for example, to give them a way to contest a suspension. (The alert reader may note that these steps sound like standard good practice for human-centered design, and we agree. In fact, one of our overarching points is that good UX practices are not applied often enough in explainability aspects of the UI.)

Second, increase *explanatory information throughout the product*, not just at the moment of decision. When people use products that rely on AI, everything from online services to medical devices, we want to include a lot of chances for people to learn about the role AI plays, as well as its benefits and limitations. Much of the information we recommend removing from in-the-moment explanations (or information that was missing entirely) can be



Figure 2. An example of how Navigator, a fictional boating app, explains its behavior.

placed nicely elsewhere throughout AI products. These additions can be made not just during initial onboarding but throughout the entire product experience. Many moments in the product experience, such as the first time the user encounters a particular feature, are a great opportunity to educate the user. These moments give the user more time to absorb information and be better prepared to receive in-the-moment explanations when a specific decision is made at a later time. For example, if people are told early about what kinds of mistakes AI-powered products are likely to make, then they are better positioned to understand and remedy situations that might arise.

Third, provide information *beyond the product experience*. Information about a given AI product can be shared in many forms, such as help center articles, ethical assessments, or videos in the style of public service announcements explaining how a product uses AI. These help users and the public better understand how AI systems make decisions, and

interpret the meaning and consequences of these decisions. We have noticed that when these useful resources exist, they often are not (or are poorly) integrated into the product experience, so another tip is to think carefully about how best to connect in-product experiences with other resources. Beyond information about specific products, information about AI can be shared in education settings, curricula, and classrooms. Together, information about specific products and AI in general build AI literacy as well as a strong foundation for people to receive explanations.

In fact, we argue that *AI literacy is key to explainability*. AI literacy helps people become critical consumers of AI-powered technologies and prepare for civic participation, as well as potentially prepare for AI-related careers [7]. AI literacy can be built continuously throughout a person's life, for example, beginning in primary and secondary education and supported over time by AI education in product journeys.

We emphasize that the goal of AI literacy (and explanations) isn't to get everyone to understand all of the technical details—it's to make sure people understand the parts that matter to them. A good metaphor here is financial literacy (or similarly, nutrition literacy). While we may not need to know every detail of what goes into interest rate hikes or the intricacies of financial markets, it's important to know how they affect us—from paying off credit cards to

While much work on explainability has focused on in-the-moment explanations, in current practice, these explanations are often poorly executed from a user experience perspective.



Figure 3. An excerpt from “Discover AI in Daily Life” that shows a farmer using an AI app to diagnose a problem with her plants.

buying a home or paying for student loans. In the same way, AI literacy isn’t about understanding every technical aspect of a machine learning algorithm—it’s about knowing how to interact with it and how it affects our daily lives.

OUR RESOURCES

Below, we highlight resources that help AI practitioners—developers, designers, researchers, educators, students, and others—learn effective, creative ways to incorporate AI explainability in product design. At Google, we use our AI Principles (ai.google/principles) to guide responsible technology development, and in accordance with AI Principle No. 4—Be accountable to people—we encourage product teams to think about all the moments and ways they can help people understand how AI operates and makes decisions. During our consultations, we saw some challenges AI practitioners face and created resources to help them. We iteratively developed and tested these resources in engagements at Google and elsewhere, during design workshops, trainings, office hours, launch reviews, and more.

Rubric. We developed a rubric to help product teams identify explanatory information that is meaningful and helpful to users. While we are cautious about rubrics and oversimplification, best practices around explainability are sufficiently evolved that we believe some crisp guidance is useful [8]. We created our rubric iteratively, drawing on an

analysis of real and fictional examples of AI systems to create an initial version and then testing and refining the rubric on a wide variety of real products over time.

The rubric lays out 22 different pieces of relevant information that may be useful to include at some point in the person’s product journey or broader education (explainability, withgoogle.com/rubric). For example, *human-in-the-loop* recommends describing whether humans monitor and/or override decisions made by AI. Per *business model*, users should have a basic understanding of the financial incentives and business model of the product. And *low quality results* recommends describing how the system identifies and annotates low quality results, for example, decisions that are based on missing or out-of-date information. Taken together, the 22 items in this rubric serve as a useful guide to improve explainability across a wide range of circumstances and products.

Navigator. Many of the AI practitioners we have engaged with have asked for concrete examples of good explanations. To meet this need, we designed Navigator (Figure 2), a fictional AI app for casual boaters that demonstrates good UX practices for AI education and explanations. We use Navigator in workshops and trainings as a storytelling device to visually illustrate key concepts about explainability, asking practitioners to think first about Navigator (a product in which they are not invested, and about which they do not have

preconceptions) and then gradually transitioning to reflection on their own products.

We’ve found this method to be an engaging and effective way to help AI practitioners think about explainability. Workshop participants say the materials drive home the importance of explainability, challenge how they make product decisions, and give them a critical lens to assess their products. Participants also say the stories build empathy for how users experience explanations (or their absence) and help them design experiences that empower users.

Model-U case studies. We launched a series of case studies in the PAIR Guidebook to help AI practitioners think through difficult issues and hone their skills for explaining AI [9]. The case studies revolve around an imagined high-tech new car—the Model-U—and cover five different hypothetical situations that intentionally include questionable or problematic explainability practices. For each situation, participants can discuss how and why to improve the design of the AI systems and their explanations. Many of the design challenges involve choosing good temporal placements for specific pieces of information. Workshop participants share that these materials give them a new perspective on explainability. One of their favorite activities is a community engagement

Consider how you can best support AI understanding and explanation at the following moments:

- **In-the-moment explanations:** Provide and thoughtfully present information that is most needed to interpret and respond to a given decision.
- **Explanations in product:** Leverage additional in-product moments to explain AI systems, such as:
 - Onboarding
 - First use of a feature
 - Errors
 - Data voids
 - Positive user feedback (explain what went *right*)
 - When the user contests
 - Decision summaries (e.g., a weekly review).
- **Beyond the product experience:** Broaden the scope of AI education beyond products.

scenario in which they role-play different stakeholders.

Discover AI in Daily Life. Efforts to support K-12 AI education are in early stages, with experts calling for contributions to standards, research, curricula, and more [7]. In this context, we offer “Discover AI in Daily Life” (Figure 3), a short, lightweight lesson in Google’s Applied Digital Skills curriculum that can be used on its own or as part of a larger curriculum [10]. The video-based lesson is available free online at g.co/DiscoverAI. It is designed for middle school students, and includes engaging animations and hands-on activities, while also supporting high school and adult learners.

The lesson focuses on AI literacy, especially helping learners recognize where and how AI touches their lives and the lives of others, and understand broad capabilities and challenges with AI. For example, it includes simple, nontechnical explanations of how a machine can “learn” from patterns in data, and why it’s important to train AI responsibly and avoid unfair bias. Resources like this support people’s lifelong engagement with AI and explanations of it.

CONCLUSION

Explainability helps people understand and interact with the systems that make decisions and inferences about them. This should go beyond providing explanations at the moment of a decision; rather, explainability is best served when information about AI is incorporated into the entire user journey and AI literacy is built continuously throughout a person’s life. We have shared resources that can be used in both industrial and academic environments to encourage

We invite AI practitioners and educators to use the resources at explainability.withgoogle.com

Where to start:

Developers and designers:
Explainability Rubric and Case Studies

University educators and students:
Explainability Case Studies

K-12 educators and students:
AI Literacy Lesson

AI practitioners to think more broadly about what explanations can look like across products and ways to provide people with a solid foundation that helps them better understand AI systems and decisions.

ACKNOWLEDGMENTS

We thank our Google colleagues in Responsible Innovation, Trust & Safety, PAIR, and Learning Lab, as well as participants in our workshops, for their valuable contributions to these ideas and resources. We are grateful to our collaborators at Channel Studio for their inspired work on the Navigator fictional app, images, and storytelling.

ENDNOTES

1. Terms such as *explainability*, *explainable artificial intelligence (XAI)*, *transparency*, and *interpretability* are often used interchangeably or even inconsistently. For purposes of this paper, our definition of *explainability* is fairly close to many common definitions of transparency, involving notions such as comprehension, impact, process, and accountability. We orient to interpretability as an important property that supports explainability; for example, tools that provide post hoc explanation of specific outputs from black box models, counterfactual analyses to identify changes that would affect a decision, or the use of models such as linear regression and decision trees that facilitate easily tracing outcomes back to inputs.
2. We use the term *AI* for simplicity, but most of our comments also apply to automated decision-making systems and complex algorithmic systems, as well as to specific types of AI such as machine learning.
3. See, for example, Blueprint for an AI Bill of Rights: Making automated systems work for the American people. The White House Office of Science and Technology Policy, Oct. 2022; www.whitehouse.gov/ostp/ai-bill-of-rights; also Jobin et al. (<https://doi.org/10.1038/s42256-019-0088-2>).
4. At the same time, scholars have initiated a healthy dialogue around the limitations of transparency and explainability, e.g., Edwards and Veale (<https://scholarship.law.duke.edu/dltr/vol16/iss1/2/>) and Ananny and Crawford (<https://doi.org/10.1177/1461444816676645>). While we are broadly aligned with these critiques, here we focus on methods of creating thoughtful, well-designed explanatory experiences so that we may get as much out of explainability as we can.
5. Our discussion of “in-the-moment” and

“beyond-the-moment” connects with the notion of local and global explanations, but we focus more on the timing of the information delivery; for example, both local and global information can be (and often are) delivered at the moment of a specific decision. We share this interest in temporal placement with Cai et al. (<https://doi.org/10.1145/3359206>), which focuses on onboarding moments for medical experts, and with Dhanorkar et al. (<https://doi.org/10.1145/3461778.3462131>), which focuses on providing information to practitioners throughout the development cycle. Our own focus is on end users and creating activities that help designers explore information presentation at different moments within and beyond the product.

6. Kelley, P.G., Yang, Y., Heldreth, C., Moessner, C., Sedley, A., and Woodruff, A. “Mixture of amazement at the potential of this technology and concern about possible pitfalls”: Public sentiment towards AI in 15 countries. *IEEE Data Engineering Bulletin* 44, 4 (2021), 28–46.
7. Touretzky, D., Gardner-McCune, C., Martin, F., and Seehorn, D. Envisioning AI for K–12: What should every child know about AI? *Proc. of the AAAI Conference on Artificial Intelligence* 33, 1 (2019), 9795–9799; <https://doi.org/10.1609/aaai.v33i01.33019795>
8. Woodruff, A. 10 things you should know about algorithmic fairness. *Interactions* 26, 4 (2019), 47–51, 2019; <https://doi.org/10.1145/3328489>
9. Zevenbergen, B., Woodruff, A., and Kelley, P.G. Explainability case studies. *CSCW 2020 Workshop on Ethics in Design*; <https://doi.org/10.48550/arXiv.2009.00246>
10. Woodruff, A. et al. “Discover AI in Daily Life”: An AI literacy lesson for middle school students. *Proc. of the 54th ACM Technical Symposium on Computer Science Education* v.2, 2023; <https://doi.org/10.1145/3545947.3576224>

👤 **Patrick Gage Kelley** is a user experience researcher at Google focusing on security and privacy. Previously a professor of computer science at the University of New Mexico, he received his Ph.D. from Carnegie Mellon University. He has worked at Wombat Security Technologies, Intel Labs, and the National Security Agency.
→ patrickgage@acm.org

👤 **Allison Woodruff** is a user experience researcher at Google, currently working on explainability and AI literacy. Prior to Google, she worked at the Palo Alto Research Center and Intel Labs. She received her Ph.D. in computer science from the University of California, Berkeley, and is a member of the SIGCHI Academy.
→ woodruff@acm.org