

# New strategic design approach focuses on turning AI mistakes into user benefits

May 14 2024, by Nathan Deen

---



Credit: Pixabay/CC0 Public Domain

More and more often, automated lending systems powered by artificial intelligence (AI) reject qualified loan applicants without explanation.

Even worse, they leave rejected applicants with no recourse.

People can have similar experiences when applying for jobs or petitioning their health insurance providers. While AI tools determine the fate of people in difficult situations daily, Upol Ehsan says more thought should be given to challenging these decisions or working around them.

Ehsan, a Georgia Tech explainable AI (XAI) researcher, says many rejection cases are not the applicant's fault. Rather, it's more likely a "seam" in the design process—a mismatch between what designers thought the AI could do and what happens in reality.

Ehsan said "seamless design" is the standard practice of AI designers. While the goal is to create a process by which users get what they need without interruption or barriers, seamless design has a way of doing just the opposite.

No amount of thought or design input will keep AI tools from making mistakes. When mistakes happen, those impacted by them want to know why they happened.

Because seamless design often includes black-boxing—the act of concealing the AI's reasoning—answers are never provided.

But what if there were a way to challenge an AI's decisions and turn its mistakes into benefits for [end users](#)? Ehsan believes that can be done through "seamful design."

In his latest [paper](#), "Seamful Explainable AI: Operationalizing Seamful Design in XAI," posted to the *arXiv* preprint server, Ehsan proposes a strategic way of anticipating AI harms, learning their reasonings, and leveraging mistakes instead of concealing them.

## Giving users more options

In his research, Ehsan worked with loan officers who used automated lending support systems. The seams, or flaws, he discovered in these tools' processes impacted applicants and lenders.

"The expectation is that the lending system works for everyone," Ehsan said. "The reality is that it doesn't. You've found the seam once you've figured out the difference between expectation and reality. Then we ask, 'How can we show this to end users so they can leverage it?'"

To give users options when AI negatively impacts them, Ehsan suggests three things for designers to consider:

- **Actionability:** Does the information about the flaw help the user take informed actions on the AI's recommendation?
- **Contestability:** Does the information provide the resources necessary to justify saying no to the AI?
- **Appropriation:** Does identifying these seams help the user to adapt and appropriate the AI's output in a way that is different from the provided design but helps the user make the right decision?

Ehsan uses the example of someone who was rejected for a loan despite having a good credit history. The rejection may have been due to a seam, such as a flawed discriminating algorithm, in the AI that screens the applications.

A post-deployment process is needed in cases like this to mitigate damage and empower affected end users. Loan applicants, for instance, should be allowed to contest the AI's decision based on known issues with an algorithm.

## Against the grain

Ehsan said his idea for seamful design is outside of the mainstream vernacular. However, his challenge to current accepted principles is gaining traction.

He is now working with cybersecurity, health care, and sales companies that are adopting his process.

These companies may pioneer a new way of thinking in AI design. Ehsan believes this new mindset can allow designers to switch to a proactive mindset instead of being stuck in a reactive state of conducting damage control.

"You want to stay a little ahead of the curve so you're not always caught off guard when things happen," Ehsan said. "The more proactive you can be and the more passes you can take at your [design process](#), the safer and more responsible your systems will be."

Ehsan collaborated with researchers from Georgia Tech, the University of Maryland, and Microsoft. They will present their paper later this year at the 2024 Association for Computing Machinery's Conference on Computer-Supported Cooperative Work and Social Computing ([CSCW 2024](#)) in Costa Rica.

"Seamful design embraces the imperfect reality of our world and makes the most out of it," he said. "If it becomes mainstream, it can help us address the hype cycle AI suffers from now. We don't need to overhype AI's capacity or impose unachievable goals. That'd be a gamechanger in calibrating people's trust in the system."

**More information:** Upol Ehsan et al, Seamful XAI: Operationalizing Seamful Design in Explainable AI, *arXiv* (2022). [DOI](#):

[10.48550/arxiv.2211.06753](https://arxiv.org/abs/2211.06753)

Provided by Georgia Institute of Technology

Citation: New strategic design approach focuses on turning AI mistakes into user benefits (2024, May 14) retrieved 17 July 2024 from <https://techxplore.com/news/2024-05-strategic-approach-focuses-ai-user.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.