Gary Klein Ph.D.
Seeing What Others Don't

The Discovery Platform

A tool for exploring intelligent systems.

Posted April 27, 2021 Reviewed by Jessica Schrader

How do you examine things that are so complex and opaque that even their designers don't understand how they work?

Recently, my colleague Shane Mueller and I had a chance to do just that. This post describes our project as a case study and hopefully as a source of ideas that can be used in different settings.

Here's what happened. Shane and I, together with Robert Hoffman, the third member of our team, have been working on a DARPA (Defense Advanced Research Projects Agency) program called Explainable Artificial Intelligence (XAI). I've described it in an earlier post. The XAI program arranged for 11 international groups of AI experts who have been trying to make it easier for people using their systems to understand how the systems arrive at recommendations and decisions. And this is tricky. The designers of the AI systems don't fully understand their outputs because the systems rely on machine learning, which means that they absorb hundreds of thousands, sometimes millions, of examples and tune themselves to digest these examples, but the tuning is invisible to the designers.

I was discussing this problem with Bill Ferguson, one of the leaders of the BBN/Raytheon project. Bill was trying to figure out the kinds of images and questions his AI system did well with, and the kinds it failed at. So he went through his database and identified three rules of thumb. Once Bill told people using his AI system about these rules of thumb, these users improved their performance.

When I heard about that, I figured I could use Bill's approach, and his database, and apply it to other AI systems. I would have a Discovery Platform—a basis for making discoveries about AI systems. It seemed very straightforward, a clear path to success.

Except that it was a bad idea. Bill disliked the database and interface he was using. It was designed for analyzing performance data, not for making discoveries. So Bill's system was an example of what *not* to do, rather than a prototype for a Discovery Platform.

But this bad idea was also an opportunity. Knowing some of the limitations of Bill's system gave us a chance to design something we think is better. Something that would help designers, and users, get a better understanding of how a specific system works, where it fails, why it fails, and perhaps even workarounds to overcome the failure.

Interviews with Ferguson captured some of the features needed for a Discovery Platform:

Commonalities and patterns. Bill wanted to examine commonalities to spot general themes. "Hmmm, my AI system is getting location questions right—oh, I see, it is relying on extra cues, like a kitchen usually has a sink, a refrigerator, a stove.

Exceptions. It had to make it easier to find exceptions, anomalies, and outliers and show the actual images so that the designers could perhaps notice something important

Failures. It should make it easy to pull out failures—cases the user got wrong, so that the designers could diagnose the reasons for these failures; e.g., "I notice that a key object is obscured in most/all of these photographs."

Contrasts. These might be cases the AI system failed at that it usually got right—e.g., Ferguson studied photographs of soccer (which his AI typically nailed) that were mis-labeled and noticed that they were all indoor soccer games. You can also contrast cases that were AI successes with cases that were AI failures. Bill wanted to have better ways to easily set up contrasts.

Confusions. Bill wanted to be able to look at high confusion classes because something might be brewing there.

Representations and instances. Thumbnails. Showing the photographs instead of hiding them. Ferguson needed to study individual photographs.

Shuttling. Bill wanted to easily shuttle back and forth between a statistical view and the specific instances.

This all seemed too complicated to ever achieve, but that's one reason I like to work with Shane. In short order, Shane had built a system that did these things. When we demonstrated it to Bill Ferguson and his colleagues, they felt how "sticky" it was, in that it was hard for them to stop playing around with it and trying new things. It was hard for them not to use it to make discoveries.

You can watch the YouTube video here.

You can also access the system itself, and play around with it [obereed.net:3838/mnist/]. And the code is open source and so if you are an AI developer who wants to use it on your own classification system, you can download it here.

I hope that the principles of the Discovery Platform can apply more generally, beyond AI systems, to support speculative thinking and exploration in other contexts.

Acknowledgment and Disclaimer: This material is approved for public release. Distribution is unlimited. This material is based on research sponsored by the Air Force Research Lab (AFRL under agreement FA8650-17-2-7711. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsement, either expressed or implied, of AFRL or the U.S. Government.

About the Author

Gary Klein, Ph.D., is a senior scientist at MacroCognition LLC. His most recent book is *Seeing What Others Don't: The remarkable ways we gain insights.*