Gary Klein Ph.D.
[Seeing What Others Don't](#)

AIQ: Artificial Intelligence Quotient

Helping people get smarter about smart machines.

Posted July 1, 2020

Are you ready to work with software tools that have artificial intelligence (AI) capabilities embedded in them? Are you ready to become perpetually confused about how these tools are working and whether to take their recommendations seriously?

Apparently, many people aren't.

Even the computer scientists developing machine learning and deep neural nets don't fully understand how their creations think. The reason is that once the algorithms take over, massaging enormous data sets, correlating and back-propagating like crazy, even the people who designed and developed the systems no longer fully understand how they arrive at their answers. The developers can guess, but it is just a guess. And most of the people using their systems don't even know how to guess.

For the past few years I have been working with Robert Hoffman and Shane Mueller to address this problem. We have been part of a large program called XAI, which stands for explainable artificial intelligence. This program is sponsored by DARPA, the Defense Advanced Research Projects Agency. Most of the teams working on XAI are building impressive systems to use AI to make their own AI systems more readily understood.

In contrast, Robert, Shane, and I have been trying to develop straightforward tools to increase explainability. We call our set of tools AIQ, artificial

intelligence quotient, and our goal is to help the people using AI systems to get smarter about how their own systems work. We want to raise their IQ about the AI systems they're wrestling with.

We want them to have better mental models about their AI systems (see Borders, Klein & Besuijen, 2019). Having a better mental model isn't just about understanding how a system works. Better mental models are also about appreciating how a system doesn't work—how it fails and what are its limitations. And better mental models are about diagnosing why these failures and limitations occur. Finally, better mental models depend on learning workarounds for the failures and limitations.

So far we've assembled nine different tools but I expect that we'll keep expanding the tool kit. Here are just three of the tools in the kit.

Cognitive Tutorial. One of the tools is an up-front tutorial to provide users with a better mental models (how the system works, how it fails, why it fails, and how to adapt). Actually, Shane and I created an earlier version of this Cognitive Tutorial almost a decade ago (Mueller & Klein, 2011) and conducted some successful trial applications for complex logic systems. We've updated that version to handle AI systems.

Explainability Scales. We have developed and validated several measurement scales): trust in an AI system, explanation goodness, explanation satisfaction, and mental model adequacy, along with methods for analyzing user mental models of the AI (Hoffman, Mueller, Klein & Littman, 2018.

Self-Explaining Scorecard. This scorecard is an ordinal scale for gauging the power and sophistication of AI techniques to improve explainability; we have applied it to several ongoing XAI efforts (Klein, Hoffman &Mueller, 2020).

We recently received funding from a government organization to apply, evaluate, and further extend the AIQ tool kit.

----

interpreted as necessarily representing the official policies or endorsement, either expressed or implied, of AFRL or the U.S. Government.

References

Borders, J., Klein, G., & Besuijen, R. (2019). An operational account of mental models: A pilot study. Proceedings of the 2019 International Conference on Naturalistic Decision Making, San Francisco, CA.

Hoffman, R.R., Mueller, S.T., Klein, G., & Litman, J. (2018). "Metrics for Explainable AI: Challenges and Prospects." Technical Report, Explainable AI Program, DARPA, Washington, DC. [https://arxiv.org/abs/1812.04608]

Klein, G., Hoffman, R.R., & Mueller, S.T. (2020). "Scorecard for Self-Explaining Capabilities of AI Systems." Technical Report, Explainable AI Program, DARPA, Washington, DC.

Mueller, S.T. & Klein, G. (2011). Improving users' mental models of intelligent software tools. IEEE Intelligent Systems, 26(2), 77-83.

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.*