

Western University

Scholarship@Western

Inspiring Minds – A Digital Collection of
Western's Graduate Research, Scholarship and
Creative Activity

Inspiring Minds

November 2022

Folk Theories and Explainable Artificial Intelligence

Michael Ridley

western, mridley2@uwo.ca

Follow this and additional works at: <https://ir.lib.uwo.ca/inspiringminds>

Citation of this paper:

Ridley, Michael, "Folk Theories and Explainable Artificial Intelligence" (2022). *Inspiring Minds – A Digital Collection of Western's Graduate Research, Scholarship and Creative Activity*. 370.

<https://ir.lib.uwo.ca/inspiringminds/370>

Folk Theories and Explainable Artificial Intelligence

Michael Ridley

PhD Candidate, Library and Information Studies

Faculty of Information and Media Studies

If artificial intelligence is so smart, why can't it explain itself? Machine learning systems, the leading-edge of AI, are powerful, complex, consequential, opaque, and not very forthcoming. Systems like Amazon, Google, and others are part of our everyday experience. We need them to explain themselves. "I can't tell you" or "I don't know" aren't acceptable answers. The field of "explainable AI" (XAI) is about getting machine learning systems to tell us what they did, why they did it, and why they didn't do something else instead. How can we help machine learning systems make better explanations? Enter folk theories. Folk theories are the beliefs people hold about how machine learning systems work. They may be not fully accurate, but they are functional. They help people use the system. Having folk theories and XAI work together allows machines to better understand humans and humans to better understand machines. Win-win.