Explainable AI and Human Computer Interaction

Dr. April Galyardt
Machine Learning Research Scientist
CERT Division
Research Review 2018

A brain teaser

I will sell you this rock for $200.

Will you buy it?

The detector says this is a piece of unobtainium.

If the detector is correct, it’s worth $1000.

https://ed.ted.com/lessons/can-you-solve-the-false-positive-riddle-alex-gendler
A brain teaser

There’s about a 1/11 chance that this rock is unobtainium.

Type
- False Positive
- Rock
- Unobtainium
Implications

Screening at the airport

An AI is predicting who is a threat.

The predictor says this person is a threat.

What’s the probability they’re actually a threat?

https://ed.ted.com/lessons/can-you-solve-the-false-positive-riddle-alex-gendler
Longstanding Science

Making *probabilistic judgements* is hard.

Depending on how probabilities *are presented* people make different choices.
Implications for Explainable AI

Current work in explainable AI is focused on *providing probabilities* to the end user.

That’s not enough.

The human-computer-interaction must provide support to help the user interpret those probabilities appropriately.