Abstract

DARPA's In the Moment (ITM) Program: Human-Aligned Algorithms for Making Difficult Battlefield Triage Decisions

Matt Turek PhD

DARPA, Arlington, VA, USA

Abstract

DARPA's In the Moment (ITM) program seeks to develop algorithmic decision makers for battlefield triage that are aligned with key decision-making attributes of trusted humans. ITM also seeks to develop a quantitative alignment score (based on the decision-making attributes) as a method for establishing appropriate trust in algorithmic decision-making systems. ITM is interested in a specific notion of trust, specifically the willingness of a human to delegate difficult decision-making to an algorithmic system. While the AI community often identifies technical performance characteristics (e.g., error rate) as trust factors for autonomous systems, ITM focuses on human attributes and characteristics (e.g., risk tolerance, rule following, or other personality characteristics; subject matter expertise; and human values to name a few) that could be encoded into algorithmic systems. This presentation will provide an overview of ITM program, including the quantitative alignment framework that will produce an alignment score between the human trustor and algorithmic trustee, as well as the evaluation planned to assess the contribution of alignment to the willingness to delegate.

Learning Objectives: Define how difficult decisions are understood in the context of the In the Moment program.

Describe the role of trust and decision-maker alignment for the In the Moment program.

Discuss the elements of the In the Moment evaluation, including the role of human delegation of difficult decisions.

CAMBRIDGE
UNIVERSITY PRESS