A Collaborative Platform for Identifying Context-Specific Values

Enrico Liscio, Michiel van der Meer, Catholijn M. Jonker and Pradeep K. Murukannaiah

Values are abstract motivations that guide our opinions and actions. Engineering value-sensitive agents that learn and align their actions with human values is essential for robust and beneficial AI. What values should an agent elicit, learn, or align with?

Basic vs. Context-Specific Values

Basic values are:
- General and abstract;
- Applicable across contexts;
- Suitable for societal questions.

Context-specific values are:
- Applicable to a context;
- Defined within a context;
- Suitable for concrete usage.

Contributions

As context-specific values vary with contexts, we need an efficient and reusable approach to identify them. We propose:
- Axies: a methodology for identifying context-specific values [1];
- A collaborative web platform to support Axies [2];
- An evaluation of Axies via a user study involving 60 subjects [1].

Axies Methodology

Axies facilitates inductive reasoning and collaborative work. Axies exploits natural language processing and active learning techniques to guide annotation.

Experiments

1. Experiment 1: Value List Generation
   - Value Lists: Covid-G1, -G2, Energy-G1, -G2
   - Value Evaluators: (Policy Experts): n = 2

2. Experiment 2: Policy Expert Evaluation
   - Value Lists: Covid-G1, -G2, Energy-G1, -G2
   - Value Evaluators: (Group 2): n = 3

3. Experiment 3: Crowd Evaluation
   - Value Lists: Covid-G1, -G2, Energy-G1, -G2
   - Value Evaluators: (Proximate Crowd): n = 52

References