Collaborative intelligence between humans and intelligent systems relies heavily on the skills of humans and intelligent systems for reaching agreements. This requires complex dialogue processes, which include human reasoning based on common sense and goal-oriented decision-making performed by the intelligent systems.

This project explores these challenges for human-aware strategic decision-making in interaction between intelligent systems and humans, with a particular focus on applications for mental-health and wellbeing. In order to model human reasoning, cognitive theories are formalized in logic-based computational architectures using non-monotonic reasoning techniques such as abstract argumentation and answer set programming.

We implement proof-of-concept prototypes of the systems aiming to validate and verify our approach in real use-cases, by considering ethics guidelines for responsible artificial intelligence.

Motivation decision-graph

Variables: Expectations | Subjective Norm | Perceived Behavioral Control.
Values: Inhibiting behavior (N), Indifferent to behavior (M), Positive to behavior (P).

Different motivation models are explored. For instance, a behavior-change heuristics model is proposed that constrains an agent’s plan for motivating behavior. The constraints follow suggestions from experts in, e.g., psychology and occupational therapy. A decision-graph is comprised of 27 mental states defined by the individual’s expectations, the individual’s motivation to comply with norms, and the individual’s perceived behavioral control. We aim to formalize a computational Theory of Mind that captures particular human beliefs, utilized in an agent’s empathic interaction.

Ongoing work