ICT23-030 - Acquiring and explaining norms for AI systems

Zusammenfassung

Artificial Intelligence (AI) systems have been permeating various facets of our daily life. They influence our purchasing choices, employment decisions, social connections, and even impact the well-being of our children and elderly. As such, it becomes imperative for AI systems to adhere to the legal, social, and ethical norms of the societies in which they operate. Addressing this imperative, the field of machine ethics is dedicated to crafting AI systems capable of embodying normative competence. A central open problem of this field is the acquisition and representation of normative information in a form that allows for machine implementation. This endeavor necessitates an interdisciplinary approach, which is offered by the AXAIS project. The project PIs: Ciabattoni (Logic), Horty (Philosophy & Legal Reasoning), and Mateis (AI), are leveraging their diverse expertise to acquire norms for use in AI systems, with a focus on ensuring the explainability of decision-making processes guided by these norms. Our approach combines methodologies from Natural Language Processing, Logic, and Legal Reasoning. Through this synthesis, we aim to create a framework capable of automatically translating extensive norm codes, while providing symbolic representations with a clear meaning. The envisioned framework champions explicable reasoning, and allows complex normative information to be acquired from simple decisions, akin to the practice of case-based reasoning in legal contexts.

Wissenschaftliche Disziplinen:
Mathematical logic (35%) | Artificial intelligence (50%) | Legal theory (10%) | Philosophy of law (5%)

Keywords:
Deontic Logic; Large Language Models; Normative Reasoning; Answer Set Programming; Common law; AI and Law

Principal Investigator:	Agata Ciabattoni
Institution:	TU Wien
Co-Principal Investigator(s):	John Horty (University of Maryland) Cristinel Mateis (AIT - Austrian Institute of Technology)

Status: Vertrag in Vorbereitung

Weiterführende Links zu den beteiligten Personen und zum Projekt finden Sie unter
https://wwtf.at/funding/programmes/ict/ICT23-030/