ICT23-030 - Acquiring and explaining norms for AI systems

Abstract

Artificial Intelligence (AI) systems have become pervasive in our daily lives, influencing decisions ranging from purchases and employment to social connections, and even impacting the well-being of our children and elderly. Consequently, it is imperative for AI systems to adhere to the legal, social, and ethical norms of the societies in which they operate. The field of machine ethics addresses this imperative, aiming to develop AI systems capable of embodying normative competence. A central challenge in this field is the acquisition and representation of normative information in a format compatible with machine implementation. Manually encoding such information in a formal language can indeed be impractical, while applying machine learning methods (ML) introduces uncertainty about the precise learning outcomes, and it hampers the justification of decisions made based on the acquired normative information. For centuries, law and philosophy have engaged with norms, but their methodologies lacked formal specifications and alignment with machine-oriented approaches. To address the challenge of norm acquisition, the AXAIS project ("Acquiring and explaining norms for AI systems") advocates for an interdisciplinary approach that combines methodologies from Natural Language Processing (Large Language Models), Logic, and Legal Reasoning. Led by project PIs Ciabattoni (Logic), Horty (Philosophy & Legal Reasoning), and Mateis (Symbolic AI and ML), the project leverages their diverse expertise to automate the acquisition of normative information, with a focus on ensuring the explainability of decision-making processes guided by these norms. The AXAIS project will introduce a comprehensive framework capable of automatically translating extensive norm codes into symbolic representations with clear meaning. The envisioned framework will promote explicable reasoning, and will enable the acquisition of complex normative information from simple decisions, akin to the practice of case-based reasoning in legal contexts. Ultimately, the framework will contribute to the development of AI systems that operate in accordance with societal norms while maintaining transparency in their decision-making processes.

Scientific disciplines:
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: Ongoing (01.12.2024 - 30.11.2028)
GrantID: 10.47379/ICT23030

Further links to the persons involved and to the project can be found under
https://www.gmbh.wwtf.at/funding/programmes/ict/ICT23-030/