FONDAZIONE BRUNO KESSLER
CENTER FOR DIGITAL HEALTH AND WELLBEING

unibz
Freie Universität Bozen
Libera Università di Bolzano
Università Liedia de Bulsan

Topic Description – PhD in Computer Science

Research Topic FBK_1: Non-Anthropocentric Ethical Artificial Intelligence for Decision Making

Supervisor: Dr. Mauro Dragoni

Topic Description: Artificial Intelligence (AI) has traditionally been designed and evaluated within an anthropocentric framework, where ethical considerations are largely modeled after human values, cognition, and social structures. This approach, which often assumes AI systems should either replicate human ethical reasoning or adhere to predefined rules reflecting human moral intuitions, raises fundamental philosophical and practical challenges. Centering AI ethics solely on human norms risks reinforcing biases, overlooking alternative ethical paradigms, and constraining AI's capacity to engage with the broader ecological and systemic contexts in which it operates. Furthermore, the assumption that ethical reasoning can be fully captured by static rules or optimization functions limits the adaptability and responsiveness of AI in dynamic, real-world scenarios. As AI systems are increasingly deployed in critical domains—ranging from healthcare and environmental monitoring to governance and autonomous decision-making—rethinking AI ethics beyond human-centric constraints and towards a more flexible, self-adaptive paradigm becomes imperative.

This PhD focusses on designing, developing, and validating a non-anthropocentric ethical framework for AI that moves beyond rigid, human-centered conceptions of ethical behavior. Instead of grounding AI ethics in pre-established human norms, the objective is to explore the potential for AI to engage in ethical reasoning as an emergent and context-sensitive process. The research aims to build upon concepts from formal logic, probabilistic reasoning, and active inference, aiming to develop an AI system capable of dynamically integrating ethical considerations through interaction with its environment. Central to this approach will be the idea that AI should not merely minimize its own uncertainty but should also consider the broader implications of its actions on other agents and the environment. This principle, framed in terms of expected free energy minimization, aims to introduce a novel ethical constraint that fosters relational, adaptive decision-making rather than fixed rule-following. By doing so, AI can develop ethical strategies that evolve over time, reflecting both immediate situational contexts and long-term systemic effects.

Required mandatory skills: Degree in Philosophy and/or Human-Centered Artificial Intelligence (LM-55-R).

Technical skills:

- Development and implementation of machine learning algorithms applied to human-centric domains;
- Knowledge representation and reasoning (symbolic formalisms, formal logics, deriving consequences from explicit knowledge bases);
- Affective computing (measurement and analysis of behavioral and physiological affective signals, design of affective agents in real-world application contexts);
- bayesian models and active inference applied to cognition and AI.

Humanities and interdisciplinary skills:

- Moral philosophy and ethics applied to AI, with critical reasoning skills on conceptual and technical solutions to specific ethical problems, including the use of symbolic formalisms to develop ethical behaviors in AI systems;
- Analysis of issues of justice, equity, and diversity raised by the spread of AI;
- A non-anthropocentric perspective in the study of AI, applying the conceptual tools of anthropology to the debate on the social and cultural impacts of intelligent systems;
- Understanding of human metacognitive capabilities and their limitations in relation to human-machine integration in decision-making processes;
- Professional experience in AI data ethics.

Desirable (optional) skills:

- Experience with biologically and cognitively inspired computational approaches (predictive/Bayesian brain, active inference);
- Familiarity with the philosophical and scientific debate on non-human agency and non-anthropocentric ethics;
- Knowledge of European and international frameworks on human rights and AI governance;
- Certifications in International Relations and International Protection of Human Rights.


FONDAZIONE
BRUNO KESSLER
CENTER FOR DIGITAL
HEALTH AND WELLBEING


unibz
Freie Universität Bozen
Libera Università di Bolzano
Università Liedia de Bulsan