

Building Explainable and Ethical AI: Grounding, Reasoning, Explanation, and Ethical Alignment

Datum 12.10.2026
Subjekt Université Toulouse Capitole
Region Francúzsko
Kategorie Technológia, Univerzita, Veda a výskum

Abstract: Artificial intelligence has achieved remarkable capabilities in perception, prediction, and generation, yet major challenges remain in factual reliability, reasoning, explainability, and ethical responsibility.

This talk discusses four principles for building more trustworthy and human-compatible AI: factual grounding, controlled reasoning with structured formalisms, verified trustworthiness through explanation, and ethical alignment by design. I will briefly contrast symbolic and data-driven approaches to AI and discuss how logic, argumentation, learning, and causal reasoning can complement one another in addressing these challenges. Selected examples from our recent work will illustrate approaches to fact-checking for large language models, interpretable and explainable reasoning, computational ethics, causal reasoning, and multi-stakeholder normative decision-making. Together, these approaches suggest a path toward AI systems whose conclusions and actions are more reliable, explainable, compliant, and accountable.

<https://www.ut-capitole.fr/accueil/universite/membres/composantes/faculte-dinformatique/seminaire-s-irit-ut-capitole/building-explainable-and-ethical-ai-grounding-reasoning-explanation-and-ethical-alignment>