DIGHUM college
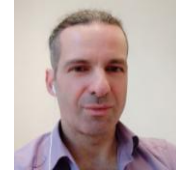
Ethics and governance of artificial intelligence

Yoann Morello
TU Wien

Supervisors
Agata Ciabattoni, TU Wien

AXAIS
Acquiring and explaining norms for AI systems

The Project

Norms shape society. As Artificial Intelligence advances, ensuring it follows legal, social, and ethical norms is vital.
But how do we encode these norms in a way machines can understand? Manual encoding is not feasible for large corpora of norms, and machine learning lacks explainability and formal guarantees.

Interdisciplinary Approach + Research Methods

We

- combine **Machine Learning, Logic and Legal Reasoning** techniques
- plan to **integrate LLMs** for automatically acquiring norms and cases, and Logic for their formalization, verification, and explanation.

Inspired by common law, we will use case decisions to guide future judgments via the doctrine of precedent.

Project Contribution

We aim to
develop tools to translate large corpora of norms (e.g. street codes) from **natural language to machineprocessable formats, enabling explainable AI decisions guided by these norms.**

Our tools will support the design of normsensitive agents and systems for automated legal compliance checks.