Ph.D. in Information Technology
Thesis Defense

May 30th, 2025
at 15:00 pm
"Emilio Gatti" Conference Room – building 20

Andrea TOCCHETTI – XXXV Cycle

Model Explainability through Human Knowledge and Crowdsourcing

Supervisor: Prof. Marco Brambilla

Abstract:

The ongoing development of incrementally complex and high-performing ML and AI models has made them more opaque than ever, making it hard to understand their behaviour and decision-making process. Such complexity calls for a fundamental property that models must have, i.e., they must be explainable, allowing human interpreters to understand their decision-making process. Consequently, the research community has focused on developing approaches to provide explanations for such black-box models faithfully. While several efficient methods have been developed, human interpretability still represents the most critical aspect. Humans might play various roles in this context, and their knowledge is fundamental to achieving such an objective. This PhD dissertation focuses on human-centred approaches in the context of Explainable AI, analyzing and developing techniques to involve humans, collect and structure their knowledge, and employ it towards explaining model behaviour. Besides XAI, crowdsourcing and gamification are essential to driving human involvement and behaviour.
In this dissertation, fundamental literature on such topics of interest is provided. Then, the role of humans in XAI and their contribution towards model robustness are described. Explainability approaches are then developed in the context of Natural Language Processing (NLP) and Computer Vision (CV). In NLP, a formalization to organize human knowledge in various tasks is defined, and data is collected through crowdsourcing. In CV, an approach to describe the decision-making process of black-box models using human knowledge was developed and tested against state-of-the-art methods, reporting on its effectiveness. This dissemination focuses on humans as the core element to achieve high interpretability, driving model trustworthiness. Several perspectives and human-driven approaches demonstrate the fundamental need to engage humans in XAI, highlighting the relevance of such a research topic.

PhD Committee

Prof. **Maristella Matera, Politecnico di Milano**

Prof. **Ujwal Gadiraju, TU Delft**

Prof. **Gabriele Tolomei, Università Sapienza di Roma**