Ph.D. in Information Technology
Thesis Defense

March 3rd, 2026
at 15:30 pm
Room BIO1 – building 21

Chiara CRISCUOLO – XXXVIII Cycle

Ethics in Data Science: Exploration of Fairness in Data-Driven Decision Systems

Supervisor: Prof. Davide Martinenghi

Abstract:

This thesis examines fairness in data-driven decision-making systems, providing both theoretical foundations and practical tools for defining, measuring, mitigating, and communicating it. Unfairness in these systems can have significant societal consequences. For example, AI chatbots used in mortgage applications have been found to recommend denials for Black applicants more frequently than for identical white applicants, illustrating racial bias in financial decision-making. Similarly, a UK government welfare fraud detection system exhibited biases related to age, disability, marital status, and nationality, raising concerns about inequitable treatment in social services. These cases highlight the urgent need for fairness-aware approaches in data-driven systems to prevent discrimination and ensure equitable outcomes. Fairness is particularly challenging because it is not a universally measurable concept, but a complex, context-dependent requirement that interacts with multiple dimensions of system design and evaluation. Starting from three main research questions, the work develops a comprehensive framework that spans the entire lifecycle of machine learning systems. First, it introduces a unified taxonomy of fairness metrics, composed of statistical and causal ones, and proposes a Fairness Decision Tree to guide the contextual selection of appropriate metrics for the specific context of use. To support interpretability and accessibility of fairness metrics, the thesis also presents *FairnessFriend*, a chatbot designed to help non-expert users understand the fairness measurement process. Second, the thesis focuses on unfairness mitigation. It offers a structured catalog of mitigation techniques (for pre-, in-, and post-processing approaches), and introduces *FAIR-CARE*, a scalable framework for the comparative evaluation of these techniques across multiple datasets, models, and metrics. This framework is applied to real-world healthcare datasets, demonstrating how fairness can be systematically assessed and improved. The work also presents *LLM4Fairness*, a conversational agent based on large language models, which enables interactive exploration and explanation of fairness results to non-expert users. Third, the thesis explores fairness in relation to other non-functional requirements and application contexts. It investigates the trade-off between fairness and

data quality, analyzes intersectional fairness in ranking systems, and proposes *SoCRATe*, a novel framework for fairness in recommender systems with limited item availability, where users who are interested in the same things but have limited availability actually can receive equal treatment. Overall, the thesis contributes to the design of ethical and trustworthy artificial intelligence and machine learning systems, supports regulatory compliance (e.g., AI Act), and promotes user trust, offering actionable perspectives and tools for embedding fairness into the design, evaluation, and communication of data-driven decision systems.

PhD Committee

Prof. Maristella Matera, Politecnico di Milano

Prof. Donatella Firmani, Sapienza Università di Roma

Prof. Paolo Missier, University of Birmingham