SCIENCES & BIOENGINEERING SCIENCES

The Research Group
Artificial Intelligence Lab

has the honor to invite you to the public defence of the PhD thesis of

Alexandra Cimpean

to obtain the degree of Doctor of Sciences

Title of the PhD thesis:

Fairness-Aware Reinforcement Learning for Balanced Sequential Decision-Making

Supervisor:
Prof. dr. Ann Nowé (VUB)
Co-supervisor:
Prof. dr. Pieter Libin (VUB)

The defence will take place on
Tuesday, October 13, 2026 at 4 p.m.

VUB Etterbeek campus, Pleinlaan 2, Elsene, Promotiezaal D.2.01

The defence can be followed online. Contact Ioana.Alexandra.Cimpean@vub.be for more information.

Members of the jury

Prof. dr. Jens Nicolay (VUB, chair)
Prof. dr. Joost Vennekens (VUB)
Prof. dr. Vincent Ginis (VUB)
Prof. dr. Roxana Rădulescu (Universiteit Utrecht, NL)
Prof. dr. Patrick Mannion (University of Galway, IE)

Curriculum vitae

Alexandra obtained her MSc in Computer Science from the VUB in 2020 and began her PhD at the Artificial Intelligence Lab in 2021, supervised by Prof. Dr. Ann Nowé and Prof. Dr. Pieter Libin. Her research, supported by an FWO fellowship since 2022, focuses on fair reinforcement learning algorithms and led to several first-author journal and workshop publications. She also supervised five student theses and assisted with bachelor teaching.

Abstract of the PhD research

Artificial intelligence (AI) is increasingly making decisions that shape our daily lives, from who gets hired to how vaccines are distributed during an epidemic. As these decisions happen sequentially over time and real-world conditions constantly change, ensuring fairness remains a major challenge. Furthermore, fairness means and is defined as different things to different people. For instance, ensuring that similar individuals are treated equally often conflicts with ensuring that diverse demographic groups are represented fairly. Traditionally, AI systems force a single, rigid compromise between fairness and performance. Instead, in this work we argue that AI should present stakeholders with a spectrum of balanced choices to support decision-making. To this end, we first propose a post-hoc evaluation of learned policies, to highlight that fairness may be achievable at little to no cost to performance by comparing the top performing policies. Next, we develop FAReL as a fairness-aware reinforcement learning framework. FAReL uses multi-objective reinforcement learning, a type of machine learning that learns through trial and error to simultaneously balance multiple objectives, to track how decisions impact both individuals and groups over time. This approach allows stakeholders to see the attainable trade-offs, making automated decision-making more transparent and accountable.
We validate FAReL across three simulated real-world scenarios: epidemic mitigation, job hiring, and fraud detection. In the context of the Belgian COVID-19 epidemic, the framework successfully identifies contact reduction strategies that minimises infections while ensuring the societal burden of restricting social contacts is fairly shared among different age groups. Additionally, we show that more fair vaccine allocation strategies can be chosen at little to no loss in the reduction of infections or hospitalisations. In job hiring, FAReL allows balancing multiple fairness notions when selecting highly qualified teams with treating similar applicants equitably. In fraud detection, it optimises catching fraudulent transactions without disproportionately locking out or disrupting honest customers. Ultimately, our experiments show that different forms of fairness can clash, meaning a one-size-fits-all definition of fair AI fails. By providing a framework of diverse fairness notions and an overview of the attainable trade-offs, FAReL empowers stakeholders to navigate complex real-world settings and select options that align with their societal values.