VUB SCIENCES & BIOENGINEERING SCIENCES

The Research Group
Artificial Intelligence Lab

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

Inês Terrucha

to obtain the degree of Doctor of Sciences

Joint PhD with Universiteit Gent

Title of the PhD thesis:

Delegation of conflict-of-interest decisions to autonomous agents

Promotors:
Prof. dr. ir. Pieter Simoens (UGent)
Prof. dr. Tom Lenaerts (VUB)

The defence will take place on

Monday, August 26, 2024 at 4.30 p.m. in auditorium P Jozef Plateau, Gent

The defence can also be followed through a live stream: https://teams.microsoft.com/l/meetup-join/19%3ameeting_YmE3ODdjYzMtNTgzZi00YjUwLWE2NTUtZGI4MTlhNDQ2NzQ2%40thread.v2/0?context=%7b%22Tid%22%3a%22d7811cde-ecef-496c-8f91-a1786241b99c%22%2c%22Oid%22%3a%22c0e01f4e-fa5d-4f70-8f6e-8a9b2c75d2bc%22%7d

Members of the jury

Prof. dr. Hennie De Schepper (UGent, chair)
Prof. dr. Ann Nowé (VUB, secretary)
Prof. dr. Tony Belpaeme (UGent)
Prof. dr. Fernando Santos (University of Amsterdam, The Netherlands)
Prof. dr. Jeremy Pitt (Imperial College London, UK)

Curriculum vitae

Inês Terrucha enrolled as a joint PhD student between UGent and VUB to do research under the supervision of Prof. Pieter Simoens and Prof. Tom Lenaerts.
During her PhD, Inês co-authored 4 papers in international journals.
When Inês is not at the office, she is preparing for an Ironman triathlon.

Abstract of the PhD research

More and more, humans rely on autonomous agents to aid them in the most varied range of activities. However, it is still unclear how the introduction of autonomous agents in our strategic decisions might affect the collective outcomes of conflict-of-interest situations. This thesis explores this question through an interdisciplinary lens that combines behavioral experiments and evolutionary game theoretical models. Through the experimental setting, it is found that those who delegate exhibit more prosocial behavior. However, this does not necessarily translate into higher success rates in avoiding collective disaster. Assuming that this might be due to errors in programming their agents, a model is proposed where delegation is distinguished from no-delegation through the timing at which mistakes may occur: before the game starts (delegation) vs. during the game (no-delegation). This model allows for the study of the long-term implications of delegation to our society, leading to the vision of human-AI hybrid teams tackling social dilemmas together. Overall, this thesis finds that delegation to autonomous agents has the potential to benefit our society, though we advocate for the controlled deployment of such agents to guarantee this positive outcome.