WETENSCHAPPEN &
BIO-INGENIEURSWETENSCHAPPEN

De Onderzoeksgroep
Artificial Intelligence Lab

nodigt U graag uit op de openbare verdediging van het proefschrift van

Willem Röpke

ter behaling van de graad van Doctor in de wetenschappen

Titel van het proefschrift:

**Thinking in Trade-Offs:
Training Agents to Balance Conflicting Objectives
under Uncertainty**

Promotor:
Prof. dr. Ann Nowé (VUB)
Co-promotor:
Prof. dr. Roxana Rădulescu (VUB, Universiteit Utrecht, NL)
Dr. Diederik M. Roijers (VUB, gemeente Amsterdam, NL)

De verdediging heeft plaats op

woensdag 18 februari 2026 om 16u

Campus Etterbeek VUB, Pleinlaan 2, Elsene in auditorium I.0.03

Samenstelling van de jury

Prof. dr. Beat Signer (VUB, voorzitter)
Prof. dr. Bart Bogaerts (VUB)
Prof. dr. Marie-Anne Guerry (VUB)
Prof. dr. Guillermo Alberto Pérez (UAntwerpen)
Prof. dr. Giorgia Ramponi (University of Zürich, CH)

Curriculum vitae

Willem Röpke verwierf een FWO-beurs en vervoegde in 2021 het AI Lab van de Vrije Universiteit Brussel om zijn doctoraat over multi-objective reinforcement learning te vervolgen. Hij publiceerde in toonaangevende tijdschriften en conferenties, en leverde bijdragen aan open-sourceprojecten. Hij was gastonderzoeker aan de Universiteit van Oxford en de Universiteit van Galway, was onderwijsassistent voor machine learning, en organiseerde workshops, waaronder MODeM 2024 en EWRL 2023.

Abstract van het doctoraatsonderzoek

Elke dag staan mensen voor keuzes die meerdere, vaak tegenstrijdige doelstellingen omvatten: kosten tegenover kwaliteit, veiligheid tegenover snelheid, of persoonlijk voordeel tegenover maatschappelijk belang. Zulke trade-offs leveren zelden één “juiste” oplossing op, en worden nog uitdagender wanneer meerdere beslissers betrokken zijn. Reinforcement learning biedt een krachtig raamwerk voor het construeren van kunstmatige agenten die autonoom handelen in complexe, onzekere omgevingen, en via vallen en opstaan leren hoe zij doeltreffende beslissingen kunnen nemen. Toch richt het merendeel van de bestaande benaderingen zich op één enkele doelstelling, en negeert daarmee de trade-offs die eigen zijn aan echte besluitvorming. Dit proefschrift pakt die leemte rechtstreeks aan. Het centrale thema is hoe we agenten kunnen ontwerpen die *in trade-offs kunnen denken*: agenten die over meerdere doelstellingen kunnen redeneren en optimaal kunnen handelen onder onzekerheid.

De bijdragen ontvouwen zich in drie samenhangende delen. Eerst overbruggen we single- en multi-objective reinforcement learning door te laten zien hoe decompositietechnieken het mogelijk maken om gevestigde single-objective methoden uit te breiden naar het leren van het Pareto front, een klassieke oplossingsverzameling die efficiënte trade-offs vastlegt voor specifieke beslissers. Op deze basis introduceren en analyseren we vervolgens alternatieve oplossingsconcepten die de voorkeuren van beslissers directer weerspiegelen, waarbij we formele theoretische garanties ontwikkelen en hun praktische relevantie aantonen. Ten slotte richten we ons op multi-agent systemen en formuleren we een nieuwe reductie van multi-objective naar single-objective games, die niet alleen nieuwe theoretische inzichten oplevert maar ook de overdracht van krachtige algoritmen tussen domeinen mogelijk maakt.

Door sterke verbindingen te leggen tussen multi-objective en single-objective paradigma's, legt het proefschrift een fundament om toekomstige vooruitgang te versnellen, zodat ontwikkelingen in het ene veld direct kunnen worden vertaald naar het andere. Deze vooruitgang brengt ons dichter bij systemen die hun gedrag kunnen aanpassen aan verschillende belanghebbenden, conflicterende doelstellingen transparant kunnen afwegen, en verantwoordelijk kunnen opereren in veiligheidskritische omgevingen in de echte wereld.