

De Onderzoeksgroep
Artificial Intelligence Lab

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

Alexandra Cimpean

ter behaling van de graad van Doctor in de wetenschappen

Titel van het proefschrift:

**Fairness-Aware Reinforcement Learning for
Balanced Sequential Decision-Making**

Promotor:

Prof. dr. Ann Nowé (VUB)

Co-promotor:

Prof. dr. Pieter Libin (VUB)

De verdediging heeft plaats op

dinsdag 13 oktober 2026 om 16u

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

De verdediging is ook online te volgen.

Contacteer Ioana.Alexandra.Cimpean@vub.be
voor meer informatie.

Samenstelling van de jury

Prof. dr. Jens Nicolay (VUB, voorzitter)

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 behaalde in 2020 haar MSc in Computerwetenschappen aan de VUB en startte in 2021 haar doctoraat in het Artificial Intelligence Lab onder begeleiding van Prof. dr. Ann Nowé en Prof. dr. Pieter Libin. Haar onderzoek, ondersteund door een FWO-mandaat sinds 2022, richt zich op eerlijke reinforcement learning-algoritmes en resulteerde in verschillende eerste-auteurspublicaties in internationale tijdschriften en workshops. Daarnaast begeleidde ze vijf thesen en assisteerde ze bij bachelorpractica.

Abstract van het doctoraatsonderzoek

Artificiële intelligentie (AI) neemt steeds vaker beslissingen die ons dagelijks leven vormgeven, van wie er wordt aangenomen tot hoe vaccins worden verdeeld tijdens een epidemie. Omdat deze beslissingen zich sequentieel voordoen en de omstandigheden in de echte wereld voortdurend veranderen, blijft het een grote uitdaging om ervoor te zorgen dat ze eerlijk blijven. Bovendien betekent eerlijkheid verschillende dingen voor verschillende mensen. Zo staat het gelijk behandelen van gelijksoortige individuen vaak haaks op het eerlijk vertegenwoordigen van diverse demografische groepen. Traditioneel dwingen AI-systemen tot een enkel, rigide compromis tussen eerlijkheid en prestatie. We betogen echter dat AI belanghebbenden een spectrum van evenwichtige keuzes moet presenteren ter ondersteuning van de besluitvorming. Daartoe stellen we eerst een post-hoc evaluatie van geleerde strategieën voor, om aan te tonen dat eerlijkheid mogelijk is met minimaal prestatieverlies door de best presterende strategieën te vergelijken. Vervolgens ontwikkelen we FAREL als een eerlijkbewust reinforcement learning framework. FAREL gebruikt multi-objective reinforcement learning, een type machine learning dat via vallen en opstaan gelijktijdig meerdere doelstellingen leert balanceren, om te volgen hoe beslissingen zowel individuen als groepen in de loop van de tijd beïnvloeden. Deze aanpak stelt belanghebbenden in staat om de haalbare afwegingen te zien, waardoor geautomatiseerde besluitvorming transparanter en beter te controleren wordt.

We valideren FAREL in drie gesimuleerde praktijkscenario's: epidemiebestrijding, rekrutering en fraudedetectie. In de Belgische COVID-19-epidemie identificeert FAREL met succes strategieën voor contactreductie die besmettingen minimaliseren, terwijl de maatschappelijke last van contactbeperkingen eerlijk wordt verdeeld over verschillende leeftijdsgroepen. Daarnaast laten we zien dat er eerlijkere strategieën voor vaccintoewijzing kunnen worden gekozen met een minimale impact op de reductie van infecties of ziekenhuisopnames. Bij rekrutering maakt FAREL het mogelijk om meerdere concepten van eerlijkheid in evenwicht te brengen bij het opbouwen van hooggekwalificeerde teams en het gelijk behandelen van vergelijkbare kandidaten. Bij fraudedetectie optimaliseert het de opsporing van frauduleuze transacties zonder eerlijke klanten onevenredig te hinderen. Uiteindelijk tonen onze experimenten aan dat verschillende vormen van eerlijkheid met elkaar kunnen botsen, wat betekent dat een universele definitie van eerlijke AI niet mogelijk is. Door een kader te bieden van diverse opvattingen over eerlijkheid en een overzicht van de haalbare afwegingen tussen prestatie en eerlijkheid, stelt FAREL belanghebbenden in staat om complexe situaties in de praktijk te doorgronden en opties te kiezen die aansluiten bij hun maatschappelijke waarden.