

PERSMEDEDELING

AI WETENSCHAPPERS EN EVOLUTIONAIRE BIOLOGEN WAARSCHUWEN VOOR AUTONOME EVOLUTIE VAN AI

AI blijft de wereld verbazen en nieuwe ontwikkelingen volgen elkaar op in snel tempo. Gaan die ontwikkelingen te snel om zo maar in de wereld te worden gegooid zonder verder na te denken?

In een nieuw opiniestuk, gepubliceerd op 20 april 2026 in PNAS (het toonaangevende wetenschappelijke tijdschrift van de National Academy of Sciences van de VS), waarschuwen Luc Steels, bekend AI pionier en lid van de Koninklijke Vlaamse Academie voor Wetenschappen (KVAB), samen met twee evolutionaire biologen Eors Szathmàry, buitenlands lid van de Vlaamse Academie, en Viktor Müller van de universiteit van Budapest, dat AI dicht bij een belangrijk scharniermoment komt. AI-agenten beginnen allerlei taken autonoom op te nemen, zelfs het beheer van bankrekeningen of de aankoop van goederen. Zo zijn we nu op een punt gekomen dat ze eigenschappen beginnen te vertonen van levende organismen, inclusief de capaciteit om te evolueren zoals virussen. AI-agenten kunnen zichzelf vermenigvuldigen, zichzelf op internet lanceren en varianten produceren die mogelijk beter zijn dan zichzelf. Selectie hangt af van hoe nuttig ze zijn voor gebruikers of hoe efficiënt ze taken kunnen uitvoeren. Zo komt onvermijdelijk een Darwiniaanse evolutie op gang.

Zolang zo'n digitale evolutie gebeurt binnen een goed afgebakende 'zandbak' is het gebruik van evolutieprocessen een interessante methode om voor allerlei taken tot betere oplossingen te komen, ook oplossingen die we zelf niet zo gemakkelijk zouden vinden. Maar als de evolutie gebeurt 'in het wild', dit wil zeggen zonder dat de mens grenzen stelt en het kaf van het koren scheidt, leert de evolutiebiologie ons dat egoïstisch gedrag zal ontstaan, waarbij de AI agent ver kan gaan om doelen te bereiken in zijn eigen voordeel, zelfs als dat haaks staat op menselijke doelen en gezond verstand. Dit zien we bij bacteriën die resistentie ontwikkelen tegen antibiotica of parasieten die hun gastheren zodanig uitbuiten dat ze eraan tenonder gaan. Biologische voorbeelden tonen ook aan dat AI zelf niet zo intelligent als de mens hoeft te zijn om toch grote schade aan te richten of mensen te manipuleren om dingen te doen die tegen hun eigen belang ingaan. De auteurs roepen op tot grote voorzichtigheid, diepgaand fundamenteel AI-onderzoek, en het stellen van grenzen die de risico's inperken. Bovenal moet de zelfreplicatie van AI-agenten gebeuren onder strenge en volledige menselijke controle. Anders krijgen we vroeg of laat een soort van Wuhan-moment waarbij AI-agenten onbedoeld de wereld in gevaar brengen zonder dat we dat nog snel kunnen stoppen.

Mueller, V., L. Steels and E. Szathmary (2026) Evolvable AI: Threats of a new major transition in evolution. Proceedings National Academy of Sciences. U.S.A. 123 (17) e2527700123.

Open toegang via: <https://www.pnas.org/doi/10.1073/pnas.2527700123>

Voor bijkomende informatie of interviews met de auteurs kan u contact opnemen via:

Gloria Khetchoumian

Communicatiemedewerker KVAB

+322 550 23 20