

AI als assistent van de mens

De ontwikkeling van artificiële intelligentie gaat razendsnel, in die mate dat ze zelfs specialisten als professor Luc De Raedt verbaast. En ook wel wat beangstigt. 'We moeten goed nadenken om misbruik en onoordeelkundig gebruik van AI tegen te gaan. Dat is onder meer wat we doen in ons instituut Leuven.AI. Want AI is een fantastische technologie die de mens kan assisteren in tal van domeinen.'

Luc De Raedt is professor aan de afdeling Computerwetenschappen van de KULeuven en directeur van Leuven.AI. Dat instituut bestaat een drietal jaar en telt intussen om en bij de honderd professoren. Luc De Raedt is ook gastprofessor aan de universiteit van het Zweedse Örebro, waar hij deelneemt aan een omvangrijk AI-programma gefinancierd door een stichting van de befaamde Wallenbergfamilie.

Artificiële intelligentie is overal.

Wat is het precies?

'Onder AI verstaan we alle machines die enig intelligent gedrag vertonen zoals converseren in een natuurlijke taal, een wiskundig probleem oplossen, een tekst vertalen, iets ontwerpen, iemand assisteren op de werkvloer, met de auto rijden enzovoort. AI-tools helpen de mens bij allerlei beslissingen.'

Het lijkt alsof AI nieuw is, maar is dat ook zo?

'Alan Turing, die wordt beschouwd als de vader van de informatica, zag al de leermogelijkheden

van de computer. John McCarthy gebruikte de term artificiële intelligentie voor het eerst, in 1956. Het document dat toen door een twintigtal wetenschappers uit verschillende disciplines werd opgesteld, is nog altijd zeer actueel. Veel dingen die ze toen wilden bestuderen, worden nu werkelijkheid zoals natuurlijke taal, lerende machines, neurale netten.'

Waarom heeft AI zo lang op zich laten wachten? Was het vooral een zaak van technologie?

'In het begin was dat zeker een probleem, er zijn pas nu op grote schaal data en computerkracht beschikbaar. Maar ook de wiskundige en algoritmische basis van de technieken is in de loop van de decennia veel verfijnd. Het oorspronkelijke idee was om het menselijke brein te imiteren. Nu zijn we daarvan afgestapt. Het verschil tussen menselijke en artificiële intelligentie kunt u vergelijken met dat tussen vogels en vliegtuigen. Ze vliegen allebei maar op een andere manier, het ene is natuur, het andere engineering. Hetzelfde geldt voor het

menselijke brein en AI. Er is een analogie maar de basisprincipes zijn verschillend. Bij AI gaat het over algoritmische en wiskundige bouwstenen, niet over biologie.'

Waarom werd drie jaar geleden beslist Leuven.AI op te richten?

'AI boomt, meer en meer mensen zijn ermee bezig. We wilden intern aan de universiteit tot meer interdisciplinaire samenwerking komen en meer gezamenlijk naar buiten treden. Van het instituut maken filosofen, ethici, juristen, psychologen, astronomen, medische specialisten, ingenieurs, wiskundigen, computerwetenschappers deel uit. Dus mensen die de AI-technieken ontwikkelen, anderen die de toepassingen maken en nog anderen die reflecteren over het maatschappelijke belang. Dat is goed. De ethische discussie bijvoorbeeld is van recente datum omdat de mogelijkheden van AI nu veel duidelijker zijn dan pakweg tien jaar geleden.'

Begrijpt u de angst, de weerstand bij mensen die vrezen dat AI het van ons zal overnemen?

'Zeker, er is angst voor wat men niet begrijpt. Maar je moet AI eerder bekijken als een technologie die mensen zal helpen bij routinezaken of bij dingen die veel tijd vragen. Alle onderzoek gaat in de richting dat de mens de controle moet behouden. Dat neemt niet weg dat er problemen zijn. De nieuwe taalprogramma's zoals ChatGPT zijn zo complex en werken met zoveel data en parameters, dat een mens onmogelijk kan nagaan waarom de machine nu precies dat welbepaalde antwoord geeft. Het systeem is goed getraind om correcte en sociaal acceptabele antwoorden te geven, dat is al veel verbeterd, maar er zijn geen garanties dat het dit altijd doet. Er is ook geen garantie op de correctheid van wat het genereert. En het geeft niet weer op basis van welke bronnen het iets maakt. Dat is best wel beangstigend, voor het onderwijs bijvoorbeeld. Collega's hebben aan ChatGPT al eens gevraagd om een computerprogramma te schrijven, zoals ze dat doen aan

**'Een basiskennis
AI voor iedereen
zou een goede
zaak zijn.'**

eerstejaarsstudenten. Wel, ChatGPT was geslaagd. Velen in de AI-wereld zijn verbaasd, een beetje gechoqueerd zelfs, dat ChatGPT zomaar publiek beschikbaar is gemaakt. Het is een erg krachtige tool die u kunt misbruiken of onoordeelkundig gebruiken. Daarom zijn ethici en ook juristen hierin heel erg geïnteresseerd.'

Dat zelfs specialisten niet weten hoe een systeem tot een bepaalde conclusie komt, is inderdaad wel beangstigend. Dat betekent toch dat u belangrijke beslissingen niet louter aan AI kunt overlaten?

'Voor beslissingen met een grote persoonlijke impact zoals het toekennen van een lening of iemand al dan niet toelaten tot een studie moet er inderdaad een verklaring zijn waarom het systeem tot die conclusie komt. De mens moet ernaast staan, AI beslist niet maar assisteert. Men ziet dat al vaak in de medische wereld. Het is niet de bedoeling dat de AI-systemen die een medisch beeld analyseren de beslissing nemen of de diagnose stellen, maar ze kunnen wel wijzen op de plekken waar artsen eens heel goed moeten naar kijken. AI als assistent, dat is meer en meer de richting die het uitgaat. Enkele jaren geleden zei een van de founding fathers dat radiologen in de toekomst minder nodig zouden zijn. Nu bekijken we dat anders. Als de technologie goed werkt en goedkoper wordt, zal ze meer ingezet worden. Daardoor kan de radioloog van veel routinewerk verlost worden en meer tijd maken voor wat echt telt.'



Waarvoor is AI, naast medische toepassingen, nog zeer geschikt?

'Voor routinematige taken die geen manipulatie vragen en waar het proces altijd hetzelfde is. Taken van bijvoorbeeld een loodgieter of een elektricien zijn zeer moeilijk te automatiseren omdat elk huis, elke klus anders is en er dus heel veel variabiliteit is. Handwerk zal in de toekomst

wellicht beter standhouden dan bepaalde soorten kantoorwerk. Ook de meer creatieve toepassingen komen in het vizier van AI: rapporten of artikels schrijven bijvoorbeeld.'

Zijn de stakende scenaristen in Hollywood terecht bezorgd?

'Ik denk het wel. De ontwikkeling en razendsnelle verbetering van taalmodellen zal doorgaan. Kijk, jongeren kunnen door de rekenmachine veel minder goed hoofdrekenen dan de oudere generaties, met schrijven is het gevaar dat het ook zo loopt. Waarom zou u nog correct leren schrijven als een tekst wordt gegenereerd wanneer u gewoon een paar ideeën intikt? Ook het genereren van beelden en video zal verbeteren. AI kan al vrij makkelijk logo's maken, al kunt u zich de vraag stellen hoe origineel het resultaat is. Striptekenaar en ingenieur Nix is hier doctoraatsstudent. Hij wil kijken of en hoe er automatisch strips kunnen worden gegenereerd.'

Kan AI dan ook creatief zijn?

'Als u ChatGPT iets laat schrijven in rapstijl, dan kunt u toch zeggen dat daar enige creativiteit in zit. Geniaal is het natuurlijk niet. Van echt creatieve genieën en kunstenaars is het denkkader zo uniek dat ze de grenzen verleggen. Maar het imiteren en combineren van ideeën ligt nu binnen de mogelijkheden van AI voor tekst en beeld. Video zal het volgende zijn.'

Het gaat snel, maar anderzijds hadden we gedacht dat bijvoorbeeld de zelfrijdende auto er al zou zijn geweest.

'In de jaren zestig dacht Nobelprijswinnaar Herbert Simon dat een AI-systeem tien jaar later wereldkampioen schaken zou zijn. Dat is uiteindelijk pas veel later gelukt. Hetzelfde verhaal met de vertaalsoftware. Er zijn AI-zomers en -winters die elkaar afwisselen, dat zijn de befaamde Gartner hype cycles. In het begin is er een doorbraak, de verwachtingen zijn hoog en er wordt massaal geïnvesteerd. De verwachtingen bereiken zo'n

hoge piek dat ze niet in te lossen zijn. Van de piek gaat het naar een dal. In die periode leert men begrijpen hoe men de technologie nuttig kan inzetten en dan volgt het geleidelijk oplopende productieve deel van de curve. Met de zelfrijdende auto is het net zo. Er was een grote hype, daarna zakte de interesse weg en nu komt hij weer boven water. In San Francisco rijdt hij al rond.'

Wat is het grote probleem dat een echte doorbraak in de weg staat?

'Het onvoorspelbare, men kan geen data van alles hebben. Een collega vertelde me over een project in Freiburg om een robot van de campus naar het centrum van de stad te laten gaan. Er was drie jaar aan gewerkt. De demo ging goed tot er op een brug duiven wegvlogen. Het ding stopte en geraakte niet meer verder. Dezelfde problemen ziet men ook bij de zelfrijdende auto's in San Francisco. De camera is verblind, er is te weinig beeld, er duiken onvoorspelbare situaties op en de auto stopt. Het zou veel makkelijker zijn om in een nieuwe stad alleen maar zelfrijdende auto's toe te laten. Nu ik erover denk, misschien zijn wij wel meer het probleem dan de zelfrijdende auto. Daarnaast is ook rekening houden met de context een moeilijkheid. Het zal tijd vragen om de systemen nog veel robuuster te maken. En we moeten ook beseffen dat een honderd procent robuuste, zelfrijdende auto wellicht nooit haalbaar zal zijn. Maar ook wij zijn geen honderd procent robuuste chauffeurs, kijk maar naar de vele ongevallen. De vraag is dus ook hoe goed de zelfrijdende auto moet zijn.'

Omdat de vooruitgang van systemen als ChatGPT zo razendsnel gaat, riepen wetenschappers enkele maanden geleden op om de pauzeknop in te drukken. Is dat realistisch?

'Ik heb de brief ook ondertekend. In Wallonië had diezelfde dag iemand zelfmoord gepleegd mede onder invloed van ChatGPT. Dat is toch zeer zorgwekkend. De brief was een signaal om even



te stoppen met de verdere ontwikkeling van het volgende taalmodel. Met ChatGPT heeft iedereen een tool die teksten kan vervalsen. Stel u voor dat we iedereen een apparaat zouden geven om geld te vervalsen en zeggen: gebruik het niet voor grote coupures, maar voor briefjes van vijf of tien euro komt u er wel mee weg. Dat is eigenlijk wat we nu doen met de AI-systemen die binnen ieders bereik liggen. Er is een radicale impact op onderwijs, media, persoonlijke levens. Wat moeten we doen met onderwijstaken, met wetenschappelijke artikels, met krantenartikels, met sollicitaties, met fake news? Het gaat heel ver. Ik geloof dan wel niet in het doembeeld dat AI alles zal overnemen, maar men moet de technologie wel oordeelkundig gebruiken om fouten en misbruiken te voorkomen. Daar moeten we goed over nadenken.'

Dat is niet alleen een zaak van universiteiten, zeker de grote commerciële spelers moeten ook mee. Zit dat erin?

'De grote bedrijven zijn er zich ook van bewust dat er iets moet gebeuren, dat ze enige zelfregulering

zullen moeten toepassen. Overheidsregulering komt er sowieso aan met de Europese AI-act. Dat is dansen op een slap koord, want we willen de excessen eruit halen zonder AI overboord te gooien.'

Data zijn de basis voor AI-systemen, automatisch komen ook thema's als copyright en privacy op tafel.

'Dat is inderdaad de grote moeilijkheid met generatieve AI, zoals de taalmodellen en beeldgeneratoren. Ze zijn op zoveel data gebaseerd dat ze overal maar gaan pikken. Dat geeft problemen van copyright en licenties. Persoonlijke data zijn al beschermd door de Europese GDPR-regelgeving, maar voor burgers is de toepassing te ingewikkeld. Ook ik klik nog te vaak "ja" op cookiesmeldingen om ervan af te zijn. AI en de grote spelers weten zeer veel over ons, soms zelfs zaken die we zelf nog niet weten of beseffen. Dat zijn allemaal problemen die we nog moeten zien op te lossen. Er is nog een andere vraag die vaak gesteld wordt in verband met data. Als een model steeds meer data nodig heeft, dan gaat men het op de duur trainen met data die het zelf gegenereerd heeft en dan zal het niet veel nieuws meer bijleren. Ook in de data zijn de bronnen eindig, ik denk dat we die limiet zelfs al stilaan bereiken.'

Zal de versnelling die we nu zien dan niet vertragen?

'Neen, want de technieken om teksten, beelden en video te genereren, worden beter zodat men ook met dezelfde data nog grenzen kan verleggen. Een van de grenzen waar we in Leuven veel aan werken is het leren uit data. Vroeger kwam intelligentie neer op redeneerprocessen en stapsgewijs denken. Zoals men bij het plannen van een reis eerst de bestemming vastlegt en dan de trip verder verfijnt, pakte men ook een wiskundig vraagstuk aan of het schrijven van een verhandeling. Dat redeneren zijn we bij AI wat uit het oog verloren door de nadruk te gaan leggen op data en machine learning. Dat puur datagedreven

werken geeft soms foute of ongewenste resultaten. Enkele jaren geleden kocht ik een nieuwe auto die me zou vertellen hoe snel ik waar kon rijden. Hij kon de verkeerstekens wel lezen maar ging toch vaak de mist in omdat hij de context niet correct kon inschatten. In AI-onderzoek groeit er nu weer veel meer aandacht voor redeneerprocessen en context. Die zijn zeer belangrijk om goed werkende systemen op te zetten.'

Er is ook heel wat te doen over het energieverbruik van alles wat met AI te maken heeft. Worden op dat vlak stappen vooruit gezet?

'Het energieverbruik voor het trainen van taalmodellen is inderdaad gigantisch, maar de technologie verbetert en de modellen worden kleiner. Een model voor het spelen van het bordspel Go was aanvankelijk enorm groot omdat het op veel computers tegelijk draaide, maar enkele jaren later was het al veel kleiner. Ook daar zouden we eens goed moeten over nadenken. Aan de andere kant, soms moet men het eerst maken voordat men het efficiënter kan aanpakken. Ook daar zijn we mee bezig binnen het instituut.'

Moeten gewone burgers weten hoe AI werkt?

'Het zou een heel goede zaak zijn als iedereen een basiskennis van AI meekrijgt in het lager of secundair onderwijs, om te weten wat er met data gebeurt, hoe AI werkt, welke de kansen en de gevaren zijn. Er is een universiteitsbreed opleidingsonderdeel "AI voor iedereen", specifiek gericht op niet-technische studenten, ook gebaseerd op de vrij toegankelijke online cursus Elements of AI. Ook voor leerlingen van de laatste graad van het secundair onderwijs zou dat perfect geschikt zijn. De kennis over AI moet in elk geval veel meer doorsijpelen in het onderwijs.'