



ARTIFICIËLE INTELLIGENTIE

NAAR EEN VIERDE INDUSTRIËLE REVOLUTIE?

Luc Steels (e.a.)



Koninklijke Vlaamse Academie van België
voor Wetenschappen en Kunsten, 2017
Standpunten 53

Artificiële intelligentie
Naar een vierde industriële revolutie?



KVAB Press

Uitgaven
van
de Koninklijke
Vlaamse Academie
van België
voor
Wetenschappen
en Kunsten

Standpunten nr. 53



Hertogsstraat 1
1000 Brussel
Tel. 02 550 23 23
www.kvab.be
info@kvab.be



Artificiële intelligentie
Naar een vierde industriële revolutie?

Luc Steels
Bettina Berendt
Aleksandra Pizurica
Dirk Van Dyck
Joos Vandewalle

Gedeeltelijke reproductie is toegelaten mits uitdrukkelijke bronvermelding.
Partial reproduction is permitted provided the source is mentioned.

© Copyright 2017 KVAB
D/2017/0455/09
ISBN 978 90656 917 67

Aanbevolen citeerwijze: Luc Steels e.a., *Artificiële intelligentie. Naar een vierde industriële revolutie?*, KVAB Standpunt 53, 2017.

Foto en ontwerp cover: Anne-Mie Van Kerckhoven

Artificiële intelligentie Naar een vierde industriële revolutie?

INHOUD

i.	Voorwoord	2
ii.	Samenvatting	4
iii.	Executive summary	6
	Inleiding	8
I.	Wat is AI?	9
II.	Opportunities	21
III.	Stimulansen voor onderzoek en ontwikkeling in AI.	26
IV.	Onderwijs	31
V.	Risico's	34
VI.	Slotbeschouwing: verwachtingen en limieten	39
	Referenties	41
	Samenstelling van de werkgroep	43

i. Voorwoord

Reeks Standpunten

De reeks Standpunten van de Academie is een bijdrage tot een wetenschappelijk onderbouwd debat over actuele maatschappelijke en artistieke thema's. De auteurs, leden en werkgroepen van de Academie schrijven in eigen naam, onafhankelijk en met volledige intellectuele vrijheid. De goedkeuring voor publicatie door een of meerdere Klassen van de Academie waarborgt de kwaliteit van de publicatie. Dit Standpunt werd goedgekeurd voor publicatie door de klassenvergadering van de Klasse van de Natuurwetenschappen op 1 december 2017.

Dit Standpunt: *Artificiële intelligentie. Naar een vierde industriële revolutie?*

Artificiële intelligentie (AI) is een wetenschappelijke en ingenieursdiscipline die tracht methoden en technologieën te vinden om systemen te bouwen die functies van het menselijk brein emuleren, bijvoorbeeld zintuiglijke waarneming en patroonherkenning, het plannen en de besturing van complexe systemen, het produceren en verwerken van taal, het leren van regelmatigigheden om voorspellingen te kunnen doen, het structureren van kennis enz. AI probeert niet de menselijke intelligentie realistisch na te bootsen en ook niet de werking van het brein te simuleren. Ze probeert wel problemen op te lossen die intelligentie vereisen. Hoewel 'intelligente' robots AI gebruiken, is dit maar een van de toepassingsgebieden van de AI, en niet het meest belangrijke.

Met AI is in de jaren vijftig een aanvang gemaakt. Ze heeft intussen cycli meegemaakt met perioden van hoge verwachtingen, beloften, veel enthousiasme en investeringspieken, gevolgd door perioden van kritiek en twijfel. Op dit moment zien we wereldwijd een nieuwe golf van enthousiasme voor het toepassen van AI in veel domeinen van de menselijke activiteit, en daaraan gekoppeld sterk toegenomen onderzoeksactiviteiten, vooral in Amerikaanse en Chinese bedrijven.

De huidige interesse voor en toepassingen van AI zijn zonder voorgaande. Hoewel veel van het enthousiasme eerder geïnspireerd is door sciencefiction dan door de werkelijkheid, is er toch een groeiende consensus dat AI van uiterst groot belang is voor de economie van de toekomst en kan bijdragen aan het beter functioneren van de maatschappij. AI biedt een krachtige nieuwe manier om producenten en consumenten met elkaar in contact te brengen en draagt zo bij tot een reorganisatie van de economie. Ze kan leiden tot nieuwe producten, extra waarde toekennen aan bestaande producten, en productieprocessen drastisch verbeteren. AI kan administratieve procedures efficiënter en met een hogere kwaliteit helpen verlopen. Ze draagt er ook toe bij dat iedereen meer toegang krijgt tot de massale hoeveelheden digitaal beschikbare kennis en ze reikt instrumenten aan om de

creativiteit te stimuleren en culturele fenomenen zoals muziek of literatuur te verspreiden.

Wij geloven dat AI voor onze huidige samenleving een sterke en positieve kracht kan zijn, als tenminste de nodige maatregelen worden genomen om het onderzoek, de ontwikkeling en de kennis over AI breed te verspreiden. Maar we zijn tegelijk bezorgd om premature toepassingen of verkeerd gebruik. We moeten ons dan ook bewust worden van de limieten van AI en actie ondernemen, zodat AI op veilige en betrouwbare manieren gebruikt wordt, in ieders voordeel.

De Klasse Natuurwetenschappen (KNW) van de Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten (KVAB) heeft daarom een werkgroep opgericht, aangevuld met leden uit de Klassen Technische Wetenschappen en Menswetenschappen, om de impact van AI in Vlaanderen te bestuderen. De werkgroep heeft dit document samengesteld in de reeks Standpunten. Daarin wordt de visie van de KVAB over de problematiek geschetst. Hoofddoel van dit document is een breed publiek zo objectief mogelijk te informeren en voor de betrokken partijen conclusies en aanbevelingen te formuleren om de problematiek correct te benaderen en om maximaal te kunnen inspelen op de kansen die zich (zullen) voordoen. Tegelijkertijd geeft dit Standpunt inzicht in de risico's en hoe we die het hoofd kunnen bieden.

ii. Samenvatting

Conclusies

1. De media berichten vandaag de dag op twee manieren over AI: zeer positief (alles kan) of uiterst negatief (de komst van AI is gevaarlijk). Dit Standpunt leert dat geen van beide terecht of productief is. Het gebruik van AI kan inderdaad een krachtige impuls geven aan zeer veel activiteiten, maar voor de meeste toepassingen is er dan wel een zeer grote investering nodig. AI is geen toverstok. Anderzijds is het perspectief van een superintelligent, allesvermogend schepsel vooralsnog sciencefiction en zal het dat nog lang blijven.
2. AI mag niet gelijkgesteld worden met robotica (hoewel AI ook voor robotica relevant is) en ook niet met *data analytics* (hoewel die momenteel erg belangrijk zijn). Het is een veel bredere waaier van methoden en technieken die vandaag de dag hun grootste impact hebben in (semi-)autonome IT-systemen en software-agenten.
3. Het is verkeerd te denken dat AI de menselijke intelligentie kan vervangen. In het beste geval is een AI-systeem in staat om van menselijk gedrag te leren hoe problemen het best kunnen worden opgelost, maar de mens is nog voor geruime tijd de oorsprong van de meeste kennis. De toekomst zit hem in de samenwerking van mens en machine.
4. AI zal bepaalde taken en dus ook banen in het gedrang brengen maar biedt tegelijk enorme mogelijkheden, enerzijds om bestaande jobs te behouden en anderzijds om nieuwe banen te creëren: door nieuwe producten en diensten of door ondersteunende AI-industrieën.
5. Er is in Vlaanderen al een chronisch tekort aan informatici, ingenieurs en wetenschappers. Aangezien grondige informaticakennis en competentie in wetenschappelijke disciplines voorwaarden zijn om AI te beheersen, volgt daaruit dat er een nog pregnanter tekort is aan AI-ingenieurs. Bovendien is er te weinig kennis aanwezig in bredere lagen van de bevolking, dus ook bij bedrijfsleiders of bestuurders, van wat AI kan en niet kan. Dat leidt tot te hoge en onrealistische verwachtingen of tot misplaatste angst.
6. Hoewel AI al in diverse onderzoeksdomeinen ingang heeft gevonden, zijn er in Vlaanderen te weinig onderzoeksgroepen bezig met fundamenteel onderzoek in AI en beschikken ze vrijwel nooit over voldoende structurele langetermijnfinanciering. Hierin moet dringend verandering komen.
7. Er zijn hoopvolle voorbeelden van AI-startups in Vlaanderen en AI wordt ook reeds in sommige grote bedrijven heel productief ingezet. Maar er moet een ecosysteem van alle AI-actoren worden opgezet. Dat moet niet alleen academische excellentie en bedrijfsinitiatieven samenbrengen, maar ook financiers, overheidsinstellingen en stuwende klanten. Recente initiatieven van Innoviris zetten de toon voor de Brusselse regio.

8. Vlaanderen kijkt niet als enige regio naar AI als bron van toekomstige welvaart. Amerika en China zijn ons ver vooruit en ook al de ons omliggende landen hebben eigen strategische plannen. Het is daarom belangrijk dat er ook in Vlaanderen een strategisch plan komt, opgesteld door de organen die hiervoor verantwoordelijk zijn.
9. Berichten over de risico's van AI zijn overtrokken en die risico's zijn wellicht niet die waar velen het meest de nadruk op leggen. Toch moet er over ethische kwesties rond het gebruik van AI grondig worden nagedacht en moeten er gedragscodes worden opgesteld. Dit moet niet alleen achteraf gebeuren, als een systeem al gebouwd is, maar moet deel uitmaken van het ontwerpen van een systeem.

Aanbevelingen

1. Er zijn initiatieven nodig om de Vlaamse samenleving, en in het bijzonder de jeugd, beter te informeren over de ontwikkelingen op het vlak van AI. Deze initiatieven moeten komen van de diverse actoren die verantwoordelijk zijn voor de communicatie over wetenschap en technologie, zoals Universiteit Vlaanderen.
2. Er zijn ook initiatieven nodig om het hoger onderwijs in AI sterk uit te breiden en nieuwe generaties van onderzoekers, ontwikkelaars en lesgevers te vormen.
3. Het stimuleren van AI in het onderwijs moet samengaan met forse stimulansen voor STEM in het lager en secundair onderwijs. AI berust sterk op wiskunde, exacte wetenschappen en vooral computerwetenschappen. Tegelijk vergt AI een diep inzicht in de gebieden waar het wordt toegepast.
4. Om de erosie van banen door AI tegen te gaan moet zwaar worden ingezet op naschoolse opleidingen: door de creatie van educatief materiaal over AI (eventueel in de vorm van MOOCs), door bedrijven de weg naar AI te wijzen, door postgraduaatscursussen aan universiteiten en hogescholen enz.
5. Er zijn hoopvolle tekenen van AI-activiteit in bedrijven en startups in Vlaanderen. Dat is de voedingsbodem waarop een AI-ecosysteem kan ontstaan. Dat kan alleen als alle actoren daar hun schouders onder zetten, zodat er een actie tot stand komt die vergelijkbaar is met wat is opgezet voor het stimuleren van micro-elektronica of moleculaire biologie.
6. We bevelen een AI-gedragscode aan voor een 'gezonde' toepassing van Artificiële intelligentie, in lijn met de *Barcelona Declaration for the proper use and development of AI in Europe*. Deze gedragscode legt de nadruk op voorzichtigheid, betrouwbaarheid, toerekeningsvatbaarheid, transparantie, grenzen aan autonomie en zorg voor het behoud van menselijke kennis.
7. Het opstellen van een strategisch plan voor AI in Vlaanderen is een urgente zaak, net zoals het ten uitvoer leggen van zo'n plan, via politieke en bestuurlijke wegen.

iii. Executive Summary

Artificial Intelligence Towards a Fourth Industrial Revolution?

Artificial Intelligence (AI) is a scientific and engineering discipline that tries to find methods and techniques to build systems that are able to emulate functions normally performed by the human brain, such as sensory perception and pattern recognition, planning and control of complex systems, production and understanding of language, learning of regularities in order to make predictions, organization of knowledge, etc. AI does not try to simulate human intelligence literally but to build systems that are capable of solving problems that require intelligence. 'Intelligent' robots use AI, but this is only one of the application fields of AI and not the most important one.

AI started in the nineteen-fifties and went through cycles with high expectations, promises, a lot of enthusiasm with peaks of investment, followed by periods of criticism and doubt. At the moment we see worldwide a very strong wave of enthusiasm about the application of AI in many domains of human activity; and coupled with this a very strong increase of research activities, particularly in US and Chinese companies.

The current interest and application of AI is without precedent. Although a lot of the enthusiasm is based more on science fiction than on reality, there is a growing consensus that AI is of the utmost importance for the economy of the future and can contribute to a better functioning of society. AI provides a powerful and novel way to link producers and consumers and thus contribute to the reorganization of the economy. It can lead to new products, provide added value to existing products and drastically improve production processes. AI can also help to streamline administrative procedures and increase their quality. And it contributes to giving everybody access to the massive amounts of knowledge that are now digitally available. Finally, it can help to stimulate creativity and distribute cultural items such as music or literature.

We believe that AI can exert a very strong positive force for our contemporary society, if the necessary measures are taken for research and for spreading technical knowledge of AI broadly. But we are also concerned about premature applications or wrong usage. It is therefore necessary to become conscious of the limits of AI and take action so that it is used in a safe and reliable way in the interest of all.

The Class of Natural Sciences (KNW) of the Royal Flemish Academy of Belgium for Science and the Arts (KVAB) has started a working group to study the impact of AI in Flanders. This working group has organized a meeting bringing together

some of the key actors in the field and has produced a document in the series 'Standpunten', in which the vision of the KVAB about this domain is expressed. The main purpose of this document is to inform the public as objectively as possible and to propose a series of conclusions and recommendations to concerned parties in order to deal with AI and ensure that our community can properly benefit from the huge opportunities, as well as get an insight in the risks and what to do about them.

Inleiding

Artificiële intelligentie (AI) is in het nieuws. Er gaat geen dag voorbij of we lezen over nieuwe toepassingen en hoe die de wereld van morgen zullen veranderen. We lezen over nieuwe startups en gigantische AI-investeringen door de grote technologiebedrijven die zijn ontstaan in de Verenigde Staten, zoals Facebook, Google, Amazon, IBM, Microsoft, Sales Force... Maar we lezen ook dat mensen als Bill Gates, Stephen Hawking en Elon Musk waarschuwen voor de gevaren van AI. Hoe realistisch zijn die beweringen? Niemand twijfelt nog dat er belangrijke ontwikkelingen bezig zijn en zoals altijd zijn speculatie en angst niet de beste raadgevers.

De Academie heeft daarom een werkgroep samengesteld om deze ontwikkelingen onder de loep te nemen en de activiteiten in Vlaanderen in kaart te brengen. Er blijken hier en daar vakken en opleidingen te bestaan aan onze universiteiten. Onderzoekers en (relatief kleine) onderzoeksgroepen zijn al decennia met AI bezig. AI wordt hier en daar succesvol toegepast in kleine en grote ondernemingen. De vraag is echter hoe we AI op veel grotere schaal in de Vlaamse samenleving kunnen benutten en hoe we deze zeer belangrijke technologische ontwikkeling in ons voordeel kunnen doen evolueren. De werkgroep heeft interne discussies gevoerd en een studie- en debatdag georganiseerd met academische specialisten, mensen uit het bedrijfsleven, vertegenwoordigers van de overheid op het regionaal, federaal en Europees niveau, en journalisten en kunstenaars¹.

Dit rapport is een synthese van de vele ideeën die uit deze activiteiten zijn voortgevloeid. Het bevat de volgende onderdelen:

I. Wat is AI? Er is nog steeds heel wat onzekerheid over wat AI is en hoe het werkt. Daarom omschrijven we eerst het domein. We kijken onder andere naar het verschil tussen de kennisgebaseerde en *machine learning*-benadering van AI en naar hoe AI steeds put uit andere vakgebieden: de wiskunde, ingenieurs- en computerwetenschappen en de biologie, maar evengoed uit de humane wetenschappen (linguïstiek, psychologie). Telkens zien we ook dat AI een bijdrage levert aan deze vakgebieden.

II. Opportuniteiten. Vervolgens brengen we de industriële impact en het potentieel van AI in kaart. De toepassingen van AI zijn zeer breed en beslaan alle domeinen van de menselijke kennis. AI wordt dan ook gezien als de technologische basis van de 'nieuwe economie'. Topbedrijven in onze digitale maatschappij zijn dan ook de grootste investeerders in AI.

III. Stimulansen voor onderzoek en ontwikkeling in AI. Wat moet er gebeuren om de opportuniteiten te realiseren: in Vlaanderen, in onze buurlanden, en op Europees vlak? Kunnen we in Europa nog concurreren met de massale middelen

¹ <http://www.kvab.be/activiteiten/artificiele-intelligentie-studiedag-debat>.

die de Verenigde Staten en China aan AI besteden? Wat is het risico en wat zijn de kosten voor onze welvaart als we deze technologische ontwikkeling niet ten volle te baat nemen?

IV. Onderwijs. Hoe zit het met het onderwijs in AI in Vlaanderen en wat zegt een vergelijking met de ons omringende landen, in de eerste plaats Nederland? We zien dat de situatie niet erg rooskleurig is en dat nieuwe initiatieven dringend nodig zijn. Welke aanbevelingen kunnen we hiervoor formuleren?

V. Risico's. Vervolgens bespreken we de huidige beperkingen en problemen bij vroegtijdige of oneigenlijke toepassingen van AI. We bekijken de Europese richtlijnen en of die wel realistisch toepasbaar zijn, en ook gedragscodes die zijn voorgesteld vanuit het vakgebied zelf.

VI. Conclusies.

I. Wat is AI?

Recent waren we getuige van spannende ontwikkelingen in het onderzoek en de toepassingen op het gebied van kunstmatige intelligentie. Zo is de woordfoutenmarge van de Google-spraakgenerator drastisch gedaald: van 23% in 2013 naar 8% in 2015. Een vergelijkbare vooruitgang was er in beeldverwerking, waar het foutenpercentage 'van de top 5' in het benoemen van objecten in beelden (met duizend klassen) daalde van 28,2% in 2010 naar 6,7% in 2014 (Russakovsky, 2015). Automatische gezichtsherkenningssystemen die gebruik maken van het Face ++-platform zijn voldoende betrouwbaar geworden en worden in China gebruikt om toegang te verlenen tot gebouwen of geld over te schrijven (Knight, 2017). Ze worden nu getest op de luchthaven van Beijing als een ID-verificatiesysteem. Een doorbraak in complexe strategiespelen kwam er in 2016, toen het AlphaGo-programma van Google DeepMind een 9-dan grootmeester en de achttienvoudige wereldkampioen Go, Lee Sedol, met 4:1 versloeg. Voor dit experiment werd nog vooral geleerd van menselijke spelers, maar ondertussen is er een nieuw experiment AlphaGo-zero, dat volledig op zichzelf leert en resoluut boven AlphaGo zelf en alle menselijke spelers uitstijgt (Silver, 2017). Behalve in games met perfecte informatie, zoals schaken en Go, heeft AI ook vooruitgang geboekt in games met imperfecte informatie: het programma Libratus versloeg in 2017 enkele wereldwijde topspelers in poker. Verbazingwekkende resultaten zijn ook gemeld op verschillende gebieden van medische diagnostiek en behandeling. AI-systemen presteren volgens recente studies beter dan artsen bij het voorspellen van hartaanvallen en beroertes (Strickland, 2017), autisme (Scudellari, 2017) en de ziekte van Alzheimer (Nordrum, 2016). Ook zijn ze vergelijkbaar met wat ervaren specialisten bereiken in de diagnose van huidkanker (Esteva, 2017) en de planning van hersenkankerbehandelingen (Strickland, 2017). Een AI-systeem STAR (Smart Tissue Autonomous Robot) vertoonde superieure prestaties bij autonome chirurgie in vergelijking met zowel deskundige chirurgen als door robots geassisteerde operaties op ex vivo-weefsels en in levende varkens (Schademan, 2013).

Twee misverstanden

Welke technologie ligt aan de grondslag van al deze ontwikkelingen? Een eerste misverstand moeten we alvast uit de wereld helpen. AI is niet 'iets', een soort stof of object die magische eigenschappen van intelligentie heeft en die, na activering, leert en problemen oplost, zoals een mens dat zelfstandig doet. AI is wél een reeks van methoden en softwaretechnieken (algoritmen en datastructuren) waarmee informatiesystemen worden gebouwd die mentale taken aankunnen. Dat zijn taken waarvoor men algemeen aanneemt dat ze intelligentie vereisen, zoals het oplossen van problemen, het maken van plannen, taal begrijpen en produceren, modellen maken en de toekomst voorspellen, organiseren, een ingewikkeld apparaat (zoals een auto) besturen enz. Omdat AI uiteindelijk tot stand komt via software, is het vakgebied dat zich bezighoudt met AI traditioneel ingebed in de computerwetenschappen. Maar AI is meer dan informatica. Het legt een laag van functionaliteit bovenop wat de informatica biedt en impliceert een andere manier van denken en andere knowhow, net zoals informatica een laag van functionaliteit is bovenop elektronica en eigen competenties en talenten vereist.

Een tweede misverstand is dat AI op zichzelf kan bestaan. Gelijk welke toepassing van AI berust op een heleboel andere technologieën, vooral telecommunicatie, computertechnologie, software, sensoren, actuatoren, computerhardware. AI is altijd het extra bindmiddel dat een heel groot effect kan hebben op de kracht en het nut van een systeem. Maar intelligentie is een kwestie van gradatie: het is niet altijd mogelijk te zeggen wanneer een systeem nu intelligent is of niet. Bovendien verschuiven de grenzen. Iets wat we nu wonderbaarlijk en superintelligent vinden is de volgende dag al weer gewoon. En dus vindt men al snel dat iets wat met AI is gemaakt uiteindelijk toch niet meer 'echt intelligent' is.

Twee voorbeelden

Een eerste voorbeeld. Veel mensen gebruiken tegenwoordig in de auto een navigatiesysteem om van A naar B te rijden. Het lijkt eenvoudig, maar de technologie daarachter is verbluffend. Het begint met de beschikbaarheid van digitale kaarten, het lokaliseren van een wagen op zo'n kaart, een interface om aan te geven waar je naartoe wil en wat je op een bepaald moment moet doen. Dat berust allemaal op gigantisch veel werk van cartografen, een netwerk van satellieten, digitale technologie waarmee een interface kan gemaakt worden enz. Daar komt AI bovenop. AI is vooral relevant om het beste pad te vinden tussen oorsprong en doel. Hier komt een zoekalgoritme bij te pas, wat in het eerste decennium van AI (de jaren zestig) een van de hoofdonderwerpen was van het AI-onderzoek. Dat algoritme moet zo snel mogelijk een oplossing vinden en mag vooral niet vastlopen in een combinatorische explosie die in elk niet-triviaal zoekprobleem onvermijdelijk optreedt. Het algoritme moet ook zo veel mogelijk rekening houden met veranderende omstandigheden op de weg, wensen van de

bestuurder en karakteristieken van de wagen. Hoe meer het algoritme toegang heeft tot kennis, hoe 'slimmer' het wordt, vooral als het kan redeneren met die kennis. Als het bijvoorbeeld weet dat een voetbalwedstrijd aanleiding kan geven tot files op een bepaald stuk van een traject, dat dus vlak voor of na een match beter niet wordt gebruikt.

De interface kan ook intelligenter gemaakt worden door in plaats van een menusysteem gesproken taal te aanvaarden en door ook de rijinstructies in taal te formuleren. Hier betreden we een ander belangrijk terrein van AI: het herkennen, analyseren, begrijpen en produceren van taal. Taal is voor de mens een tweede natuur. Daardoor staan we niet stil bij de immense complexiteit van de cognitieve processen die conversatie in taal mogelijk maken. Bovendien vergt het begrijpen en produceren van taal kennis van het onderwerp en algemene kennis van de wereld. Pas toen computertaalkundigen en AI-onderzoekers taalverwerkende systemen begonnen te bouwen, kwam deze enorme complexiteit aan het licht.

Je kan dus vertrekken van een heel eenvoudig navigatiesysteem en het steeds intelligenter maken: met meer kennis, een meer natuurlijke interface, een zoekalgoritme dat de vele eisen van de gebruiker of van de wegomstandigheden in overweging kan nemen enzovoort. Het wordt nog sterker wanneer in het navigatiesysteem leeralgoritmen zijn ingebouwd, die bijvoorbeeld bijleren over welke wegen beter vermeden worden of die de gewoonten van de bestuurder leren kennen (elke dag dezelfde weg naar het werk, in het weekend naar het appartement aan zee, regelmatig inkopen in hetzelfde shoppingcenter), zodat die nog nauwelijks instructies hoeft te geven. Wellicht stappen we binnenkort in de zelfrijdende auto die al weet waar we naartoe willen nog voor we iets zeggen.

Is een intelligent navigatiesysteem slimmer dan de mens? Op een bepaalde manier wel, omdat het systeem toegang heeft tot heel veel informatie waar we zelf niet zo gemakkelijk aan kunnen of die we niet bij de hand hebben. Anderzijds beschikt de mens over een groot vermogen aan 'gezond verstand' dat soms cruciaal is om de weg te vinden: een wegwijzer negeren die overduidelijk fout is, weten dat meer dan één stad in Vlaanderen 'Heist' heet (Heist aan zee en Heist-op-den-Berg). Is zo'n intelligent navigatiesysteem een gevaar voor de mens? Zal het hem overtroeven in intelligentie? Een gevaar lijkt het niet echt te zijn, eerder iets wat bijzonder nuttig kan zijn, zeker ook in een vreemd land. Zal het de mens voorbijsteken? Dat laatste kan alleen als we zelf onze bekwaamheden in het lezen van kaarten niet blijven oefenen en doorgeven aan de volgende generatie.

Het is niet moeilijk om duizenden soortgelijke voorbeelden te vinden waar de grenzen van een toepassingsdomein worden verlegd door het gebruik van AI. Een basisinformatisering is altijd noodzakelijk. Daar kunnen componenten aan worden toegevoegd die op een betere manier beslissingen kunnen treffen, ook als het probleem combinatorische trekken vertoont en veel kennis vereist, of die op een

betere manier beslissingen kunnen communiceren, bijvoorbeeld via een dialoog in gesproken taal.

Een tweede voorbeeld: medische diagnose. De kennis van dokters is ongemeen verbluffend, maar ook in dit domein maakten de jongste decennia heel wat technologieën hun opwachting. Bijvoorbeeld om metingen te verrichten die vroeger onmogelijk waren, door data van patiënten op te slaan voor later gebruik of om ze uit te wisselen tussen verschillende dokters, door statistieken op grote schaal te berekenen om op die manier trends in ziekten op het spoor te komen en de volksgezondheid in kaart te brengen. Hoe komt AI hierbij kijken? Ook hier zijn allerlei hulpmiddelen denkbaar. Stel, een arts moet bij het maken van een diagnose een zoekprobleem oplossen: gegeven bepaalde symptomen, wat zijn de oorzaken en hoe kunnen die verholpen worden? De symptomen liggen niet altijd voor de hand en een deel van een goede diagnose bestaat juist uit het verzamelen van meer gegevens. De oorzaken liggen nog minder voor de hand, omdat het menselijk lichaam zo'n ontzettend complex systeem is waarover nog veel niet bekend is. Goede behandelingen vinden, zeker voor complexe aandoeningen, is nog minder voor de hand liggend, ook al omdat iedere mens licht anders reageert op een medicijn.

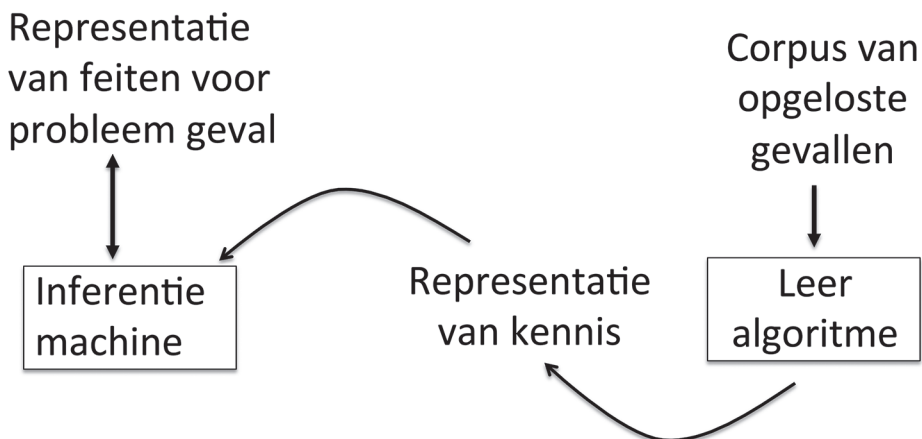
Op al deze aspecten van medische diagnose zijn de laatste decennia AI-methoden toegepast. De eerste medische 'expertsystemen' dateren van het einde van de jaren zeventig. Onderzoekers hebben geprobeerd om de redeneerstrategieën en de kennis van artsen te modelleren, te implementeren en vervolgens toe te passen om nieuwe gevallen te onderzoeken – met significant succes. Recenter worden technieken van taalverwerking en automatisch leren toegepast om in enorme hoeveelheden medische teksten te zoeken naar teksten over gevallen die lijken op dat van de te onderzoeken patiënt, om zo de beste diagnose te vinden en te leren welke therapie in het verleden het grootste succes of de minste neveneffecten heeft gehad. De medische beeldverwerking maakt nu ook uitgebreid gebruik van AI-technieken om patronen te herkennen. Meer en meer doen leeralgoritmen hun intrede. Zij krijgen data van zo veel mogelijk gevallen die door menselijke experts zijn onderzocht en leiden daaruit wetmatigheden en patronen af. Vooral in de interpretatie van metingen, bijvoorbeeld door scanners, heeft AI haar nut onomstotelijk bewezen.

Opnieuw kunnen we de vraag stellen of deze medische AI-systemen gevaarlijk zijn en of ze de mens overbodig zullen maken. Het is duidelijk dat het gevaar er alleen zal zijn als we te veel vertrouwen zouden hebben in de techniek en als we om die reden de ervaring, het doorzicht en de superieure intelligentie van de arts niet blijven zien als cruciaal. Bovendien moeten we menselijke artsen blijven opleiden en hen de kans geven om te leren, eenvoudigweg omdat de ultieme kennis over de mens... van de mens zelf komt. De arts moet wel leren om AI in de praktijk in te schakelen en zo te komen tot betere en snellere diagnoses. Bovendien heeft

de bouw van medische AI-systemen een grote toekomst voor zich, waardoor dit kan uitgroeien tot een industrie die vergelijkbaar is met de huidige farmaceutische industrie.

Hoe werkt AI?

Hoe werkt AI? AI gaat ervan uit dat intelligentie een proces is van informatieverwerking. Dit betekent dat er een manier moet zijn om informatie over het probleemdomein in de computer voor te stellen (de probleemrepresentatie) en om de processen te implementeren die deze informatie halen uit data of ze transformeren naar meer informatie. Een voorbeeld: in een medisch expertsysteem moeten data van patiënten en hun context voorgesteld worden in de computer en moeten de manieren om deze data te verzamelen of een mogelijke diagnose te stellen op basis van de data omgezet worden in computeroperaties. De processen die dat laatste doen hebben in AI-systemen vrijwel altijd twee componenten: een representatie van de kennis (bijvoorbeeld welke symptomen welke ziekten suggereren) en een proces dat deze kennis actief gebruikt om een probleem op te lossen (de inferentiemachine). Door deze splitsing in twee componenten wordt het mogelijk om een leersysteem op te zetten dat uit de data en uit feedback van experts of gebruikers nieuwe kennis kan afleiden (Zie Figuur 1.).



Figuur 1. Architectuur van een typisch AI-systeem. Er is een representatie van het probleem en van de beschikbare kennis om het probleem op te lossen. Een inferentiemachine past de regels toe op de feiten en breidt zo de feiten uit, tot er een oplossing is. Een leercomponent zorgt ervoor dat de kennis zelf wordt uitgebreid, ofwel op basis van een corpus van voorbeelden, input van experts en gebruikers, ofwel door ervaring vanuit het oplossen van een probleem.

Een voorbeeld: een intelligente schaakcomputer heeft (i) een probleemrepresentatie die de posities van de stukken op het bord weergeeft, (ii) een representatie van kennis die bestaat uit de spelregels en ook uit strategieën om de beste zet te kiezen uit een reeks mogelijkheden, en (iii) een kennisprocessor die deze kennis toepast op een specifieke configuratie van stukken. Feedback over de afloop van een zet kan dienen om de kennisrepresentatie uit te breiden en zo beter te leren schaken. De AI-schaakcomputer die erin geslaagd is om wereldkampioen schaken te worden werd volgens deze architectuur gebouwd, net als het recente systeem AlphaGo dat heeft gewonnen tegen de wereldleider in het nog uitdagender spel Go.

Door deze architectuur toe te passen op heel veel domeinen van menselijke kennis is ondertussen duidelijk geworden dat de omvang van die kennis gigantisch is. Een typische probleemrepresentatie kan gemakkelijk duizenden feiten bevatten en het aantal regels en heuristieken die experts gebruiken loopt al gauw in de honderdduizenden. Wij zijn ons niet bewust van de omvang van onze kennis. Bovendien is de snelheid waarmee we kennis toepassen op een probleem verbluffend, om nog te zwijgen van de snelheid en vindingrijkheid van de leerprocessen waarmee we deze kennis bijna spelenderwijze opbouwen. Op de koop toe verandert onze kennis voortdurend. Elke ervaring is een bron van nieuwe inzichten en er is ook een continu proces dat ervoor zorgt dat we overtollige feiten vergeten. Het probleem van AI is er dus gedeeltelijk een van schaalvergroting, zowel technisch (opslagcapaciteit en toegang) als methodologisch (hoe kunnen we zulke gigantische systemen bouwen en beheren). Maar de uitdaging van AI ligt vooral in het vinden van de juiste structuren om problemen en kennis te representeren en van de strategieën waarmee die worden toegepast. Op beide terreinen zijn er nog grote ontdekkingen te doen.

De kennisgebaseerde benadering

In de ontwikkeling van AI zijn er diverse manieren ontstaan om de probleemrepresentatie, kennisrepresentatie, inferentiemachine en leercomponent vorm te geven. Ruwweg is er een onderscheid tussen een kennisgebaseerde en een datagebaseerde benadering.

De kennisgebaseerde benadering was oorspronkelijk geïnspireerd op inzichten uit de psychologie – vooral het werk van Herb Simon en Allen Newell – over hoe mensen problemen oplossen. Deze benadering probeert de kennis van een menselijke expert zo goed als mogelijk in kaart te brengen door observaties en door gesprekken met de expert, en probeert die kennis vervolgens te gieten in de representaties, regels en zoekstrategieën die het gedrag van de expert benaderen. De kennisgebaseerde methode biedt daardoor het voordeel dat de inhoud en het gedrag van het kennisstelsel herkenbaar zijn, zowel voor de expert die de basis vormde voor het systeem als voor de gebruiker, die uitleg kan krijgen waarom bepaalde beslissingen zijn genomen in termen die hij begrijpt.

Op het einde van de jaren zeventig was de kennistechnologie, die qua implementatie vooral berust op symbolisch programmeren in talen zoals LISP en PROLOG, voldoende gevorderd om op die manier kennissystemen – ook bekend als expertsystemen – te bouwen. De eerste voorbeelden waren DENDRAL, een systeem in het domein van de organische chemie, MYCIN voor medische diagnose van infecties en het voorschrijven van antibiotica, en MACSYMA, een systeem voor computeralgebra dat een voorloper is van Mathematica.

Vanaf de jaren tachtig begon deze technologie zich ook in de industrie te verspreiden. Een van de eerste projecten was de Dipmeter Advisor voor de interpretatie van geofysische metingen ontwikkeld door Schlumberger, en R1, een expertstelsel voor de configuratie van computersystemen dat werd ontwikkeld door de Digital Equipment Corporation. Al snel ontstond een echte industrie in kennistechnologie die basistechnologie leverde, zoals LISP-machines, en die in opdracht van bedrijven kennissystemen bouwde. Ook zag een aantal ontwerpmethodologieën het licht.

België bleef niet achter bij deze ontwikkelingen. Onder impuls van Guy Verhofstadt, minister van Wetenschapsbeleid, was er in 1988 een programma voor het stimuleren van AI-onderzoek aan de universiteiten, en even later zette het IWONL/IRSIA (de voorloper van het huidige IWT en INNOVIRIS) grootschalige interacties op tussen de Belgische industrie en academische groepen. Zo werden in het AI-Lab van de VUB een aantal expertsystemen gebouwd, onder meer voor de diagnose van kerncentrales (met Tractebel), voor de diagnose van digitale telefooncentrales (met Bell Antwerpen), voor de roostering van de treinen (met de NMBS) en voor het beheer van metroverkeer (met MIVB, de Brusselse vervoermaatschappij). Veel van die systemen zijn nog steeds in gebruik of werden later verder ontwikkeld. De KU Leuven van haar kant bouwde een belangrijke onderzoeksgroep uit op het vlak van logisch programmeren, met Maurice Bruynooghe in een leidinggevende rol. Die legde de basis voor de huidige sterke onderzoeksactiviteiten op het vlak van symbolisch leren. Aan de KU Leuven en de VUB werd ook aan computertaalkunde gedaan. Deze tak van AI werd door Walter Daelemans aan de Universiteit Antwerpen eind jaren negentig uitgebouwd tot een krachtige groep. De UGent legde meer de nadruk op de filosofische en psychologische aspecten van kennisgebaseerde AI in de onderzoeksgroep Communicatie en Cognitie, mede geleid door Marc de Mey, die vooral in de jaren tachtig en negentig actief was.

Na een golf van enthousiasme werden de technieken van kennissystemen overgenomen in het mainstream-instrumentarium van informaticaontwikkelaars en aan het begin van de 21ste eeuw was de aandacht voor kennistechnologie afgenomen, vooral omdat het niet gemakkelijk bleek om de kennis van experts te analyseren en omdat het ook tijdrovend is om deze kennis mooi te formaliseren en te implementeren. Het correct transfereren van kennis en kunde van een technisch operator of expert naar een kennisingenieur is niet de gemakkelijkste opdracht. Beiden spreken een andere taal en leven in een andere denk- en

prioriteitenwereld. Het onderzoek en de toepassingen verschoven dan ook in drie richtingen:

(i) Om aan de problematiek van de kennisanalyse het hoofd te bieden is veel werk verricht: hoe haal je de kennis automatisch uit grote hoeveelheden tekst en informatie, vooral aan de hand van symbolische leertechnieken? Het meest frappante resultaat hiervan is het systeem WATSON, dat werd ontwikkeld door IBM. In 2011 heeft WATSON de beste menselijke spelers verslagen in het televisiespel Jeopardy, dat sterk steunt op common sense (gezond verstand), niet op gespecialiseerde kennis. WATSON heeft zich daarvoor encyclopedische kennis eigen moeten maken en deed dat door het 'lezen' van enorme hoeveelheden boeken en krantenartikelen. De technologie achter Watson wordt ondertussen gebruikt in allerlei toepassingsdomeinen (verzekeringen, medische diagnose, technische diagnose) en is de belangrijkste basis van IBM's huidige activiteiten op het vlak van AI.

(ii) Een andere manier om aan de kennisanalyse het hoofd te bieden is door de collectieve crowdsourcing van kennis. Dat wordt mogelijk gemaakt dankzij webinitiatieven, zoals Wikipedia en diverse spin-offs daarvan. De meeste participanten voegen kennis toe in taal, maar die kennis wordt vervolgens omgezet naar gigantische semantische netwerken en feitenbanken, die constant en automatisch worden aangevuld via taaltechnologie en semantische technologieën. Een voorbeeld hiervan is de Knowledge Graph, een semantisch netwerk dat is opgebouwd door Google en dat in 2016 al 70 miljard feiten bevatte. Het dient om op grote schaal logisch en associatief te redeneren, zodat vragen, zelfs al zijn ze zeer vaag, toch beantwoord kunnen worden. De resultaten van dit onderzoek vinden hun toepassing vooral in het zoeken naar informatie op het web. De meeste mensen zijn er zich niet van bewust, maar als je in een zoekmachine als Google iets intikt treedt er een hele waaier AI-componenten in werking. Het begint met taalverwerkingsprogramma's die voortdurend het web scannen om uit webpagina's de belangrijkste sleutelwoorden te halen en de vragen van gebruikers te proberen analyseren, en vervolgens komen er inferenties en associaties om vragen flexibel te begrijpen en zo veel mogelijk kennis en ook context naar boven te spitten.

Wie al eens de website bezocht van de Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten (<http://www.kvab.be/>), hoeft maar de term 'academie' in te tikken om de site als eerste mogelijke keuze te zien te krijgen en je hoeft maar 'laureaten academie' in te tikken om een overzicht te krijgen van de laureaten van de Vlaamse Academie, of 'standpunten academie' voor een overzicht van de KVAB-Standpunten. Anders gezegd: de context van het zoekproces is in rekening gebracht en informatie over hoe woorden zoals 'laureaten' of 'standpunten', en de concepten die erachter steken, met elkaar verbonden zijn, wordt gebruikt om met een razende snelheid een zo goed mogelijk antwoord te vinden.

(iii) *Agents*. Een sterk opkomende branche, waarin de kennisgebaseerde technieken zeer belangrijk zijn geworden, is het domein van de autonome *agents*. Deze toepassingen gaan verder dan het oplist van webpagina's of

het beantwoorden van vragen. Ze hebben ook een taalproductiecomponent en kunnen zelf zaken op het web opzoeken, echte antwoorden geven (niet zomaar een lijst van websites) en zelf acties uitvoeren, zoals het bestellen van een taxi, het opzetten van de centrale verwarming of het aankopen van een boek. SIRI (in 2009 gelanceerd door Apple) is het eerste belangrijke voorbeeld van dit toepassingsdomein, maar ondertussen zijn er ook andere producten, zoals Alexa (Amazon) of Cortana (Microsoft). *Chatbots* die in dialoog treden met gebruikers over nieuws of achtergrondinformatie zijn een ander voorbeeld van *agents*. Het ziet ernaar uit dat ze een groeiende rol beginnen te spelen ter vervanging of aanvulling van helpdesks, diensten na verkoop enz.

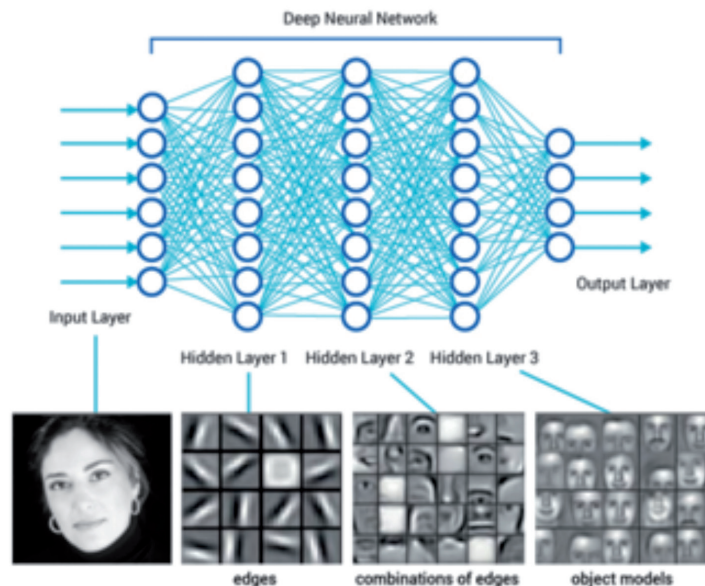
De datagebaseerde benadering

De datagebaseerde of 'machine learning' (automatisch lerende) benadering heeft wortels in de statistiek en patroonherkenning. Hier vertrekt men van data over het gedrag van mensen, over de beslissingen die ze hebben genomen of over verschijnselen die zijn waargenomen via sensoren. Vervolgens worden statistische technieken gebruikt om in die data patronen te ontdekken en die patronen worden dan weer aangewend om nieuwe problemen op te lossen. Een van de belangrijkste technieken hiervoor is Bayesiaanse inferentie, die is gebaseerd op probabiliteitstheorie.

Vooral vanaf de jaren negentig heeft de datagebaseerde benadering een hoge vlucht genomen, enerzijds omdat er meer en meer data ter beschikking komen, anderzijds omdat de rekenkracht gestaag is toegenomen. Zo was er een belangrijke verschuiving in het domein van natuurlijke taalverwerking van kennisgebaseerde benaderingen (systemen gebaseerd op de inzichten uit de taalkunde) naar statistische benaderingen. Op dit domein werden belangrijke resultaten geboekt door de groep van Walter Daelemans aan de Universiteit Antwerpen. Ook aan de Gentse universiteit is Bayesiaanse inferentie in onderzoek en ontwikkeling, vooral in de groep van Aleksandra Pizurica, en aan de KU Leuven wordt onder leiding van Luc De Raedt sinds het eind van de jaren tachtig onderzoek verricht naar het samengaan van de kennisgebaseerde benaderingen met machinaal leren, meer recent ook in combinatie met probabilistische modellen.

Binnen de datagebaseerde benadering van AI spelen de artificiële neurale netwerken (ANN) een belangrijke rol, al gaan de historische interessegolven af en aan. Hoewel ANN hun inspiratie halen uit de biologische neurale netwerken, zoals deze van de hersenen, is hun biologische getrouwheid eerder beperkt. ANN zijn wiskundige modellen die vlot in computers met software of in chips met hardware uitgevoerd kunnen worden en waarvan de prestaties en kwaliteiten op wiskundige basis gestoeld zijn. Er bestaat een brede toepassingsbasis in domeinen zoals classificatie, diagnose, predictie en patroonherkenning (herkennen van objecten in beeld en spraak). Over het algemeen kun je zeggen: telkens wanneer men over

een grote verzameling ingangs-uitgangsparen beschikt. Een (artificieel) neuron is een lineaire combinatie van ingangen, vermenigvuldigd met gewichten en een drempel (*threshold*) waarop een niet-lineaire functie wordt uitgevoerd. Een ANN is een netwerk met minstens enkele lagen van deze neuronen. Met de dataset van in- en uitgangen uit de toepassing worden al de gewichten van de neuronen aangepast in een leerproces in de computer, tot het ANN het verband tussen de ingangen en uitgangen goed weergeeft.



Figuur 2. Een ANN is een netwerk van artificiële neuronen dat is georganiseerd in verschillende lagen. Informatie komt binnen op de linkerinputlaag van het netwerk en propageert via gewogen verbindingen naar de volgende laag, en zo tot op de outputlaag.²

Men gebruikt bijvoorbeeld een dataverzameling van duizenden foto's van katten en honden om die met een ANN te kunnen onderscheiden. Tijdens de leerfase past het netwerk zijn gewichten aan, tot het bij het aanleggen van een foto uit de dataset aan de ingang ook effectief een uitgang aangeeft dat een kat of hond is herkend. De grote kracht van een goedgetraind ANN bestaat erin dat een ongeziene foto van een kat of een hond ook zeer goed geklasseerd zal worden. Dit noemt men generalisatie. Het is een gesuperviseerde vorm van automatisch

² <https://medium.com/diaryofawannapreneur/deep-learning-for-computer-vision-for-the-average-person-861661d8aa61>

leren met een degelijke wiskundige basis, want er is aangetoond dat elke continue afbeelding tussen meerdere ingangen en uitgangen voorgesteld kan worden met een dergelijk ANN met minstens een paar lagen van neuronen.

In de jaren veertig en vijftig was er een eerste golf van interesse voor ANN dankzij het werk van McCulloch, Pitts, Rosenblatt, Widrow en Hoff. Het enthousiasme voor deze modellen taande toen duidelijk werd dat belangrijke logische basisfuncties niet gerepresenteerd of geleerd konden worden met een neuraal netwerk van één laag. Maar in de jaren tachtig kwam er mede door het werk van Rumelhart, McClelland, Hinton en anderen een tweede golf van interesse voor ANN met meerdere lagen en getraind met *backpropagation*. In het midden van de jaren negentig was deze golf enigszins geluwd door gebrek aan rekenkracht en de opkomst van de *support vector-machines*, een andere vorm van gesuperviseerd automatisch leren die steunt op de theorie van statistisch leren van Vapnik en Chervonenkis. Deze methodes zijn heel performant en robuust, en kunnen zowel op classificatie- als regressieproblemen toegepast worden. Een erg succesvolle rekenefficiënte variëte daarvan is de Least Squares Support Vector Machine (LSSVM), die de onderzoeksgroep Stadius van de KU Leuven heeft ontwikkeld.

In de jaren 2005-2010 zorgden drie factoren voor een nieuwe golf van interesse voor ANN: nieuwe leertechnieken, een toename van de rekenkracht, vooral door het gebruik van grafische processoren (GPUs), en een vloed van data door het toenemende gebruik van het internet en het internet der dingen. Deze nieuwe golf van AI-onderzoek en -ontwikkeling staat bekend als *deep learning*, een verwijzing naar het feit dat de netwerken meerdere lagen van neuronen gebruiken en dus 'dieper' zijn dan de tweelagige netwerken uit de jaren vijftig en zestig (Lecun et al., 2015). Op dit moment beleven we de glorieperiode van *deep learning*. Terecht. De methode heeft haar kracht bewezen, vooral in problemen van patroonherkenning, maar ook in big data-analyse en het soort van beslissingen dat vroeger uitsluitend het terrein was van kennisgebaseerde systemen. Systemen gebaseerd op *deep learning* leverden in allerlei benchmarkstudies betere resultaten op dan andere leermethoden en dan de kennisgebaseerde benaderingen gebaseerd op menselijke modellen. Tegelijkertijd zijn er ook veel domeinen waar *deep learning* niet de oplossing is. Zo is het bijzonder moeilijk, en in vele gevallen zelfs onmogelijk, een *deep learning*-netwerk te gebruiken om verklaringen voor een beslissing te geven in menselijke taal.

Het is dankzij de beschikbaarheid van zeer veel rekenkracht en data dat *deep learning* recent echt in de praktijk kon worden gebracht. Zo heeft het AI-lab van Facebook een automatisch systeem met *deep learning* ontwikkeld om de namen toe te voegen aan de foto's van de personen in een grote fotoverzameling. Google's DeepMind-technologieën ontwikkelden systemen die professionele spelers kunnen verslaan in diverse spelletjes, zoals videogames van Atari en Go. Google Translate gebruikt Long Short Term Memory-netwerken om teksten automatisch te

vertalen uit en naar meer dan honderd talen. Ook het IDLab aan de UGent boekte belangrijke resultaten in dit domein, vooral voor de analyse en aanbeveling van muziek en opto-elektronische implementaties van bepaalde lerende netwerken.

De toekomst van data-gebaseerde technieken zit volgens de stakeholders van *deep learning* in 1) de verschuiving van superviseerd naar ongesuperviseerd leren, ook wel de 'generatieve aanpak' genoemd, en in 2) *reinforcement learning* (het stellen van acties die op termijn een cumulatieve beloning maximaliseren). Deze systemen zouden zonder (of met veel minder) expliciete voorbeelden moeten kunnen leren. Recente GAN-netwerken (Generative Adversarial Networks), waarbij twee (of meer) neurale netwerken tegen elkaar 'spelen' om zo hun gedrag te verbeteren, wordt gezien als een van de belangrijke toekomstige perspectieven van AI. Grote vooruitgang wordt ook verwacht van een combinatie van leermethodes en complexe redeneringen zoals ontwikkeld in de kennisgebaseerde AI.

Toekomst

Wellicht zal AI door toekomstige evoluties meer en meer een combinatie van technieken gebruiken. De kennisgebaseerde technieken zijn beter in het rationeel beredeneren van de oplossing van problemen en in het uitleggen via taal welke kennis zij bevatten of hoe ze tot een oplossing komen. De datagebaseerde technieken zijn beter om de immense hoeveelheden data die nu geproduceerd worden beter te benutten of om aan patroonherkenning te doen. Er is geen enkelvoudig algoritme dat alle taken van intelligentie aankan. Systemen die reële problemen aanpakken bestaan altijd uit een combinatie van heel veel modules en kennisbronnen die vandaag de dag via internet gemakkelijk met elkaar verbonden kunnen worden.

Intelligentie is geen eendimensionale kwantificeerbare grootheid. Er zijn veel verschillende vormen van intelligentie, die berusten op verschillende mechanismen. Niet alleen zijn veel dieren op sommige terreinen superieur aan mensen, maar er is ook een enorme variatie in de menselijke bevolking, met specialisaties die afhankelijk zijn van talent, interesse en leefomstandigheden. Een groep jagers die in tropische regenwouden in Brazilië of Afrika kan overleven is veel intelligenter in de kennis en omgang met de natuur dan een stadsmens, die er nog geen dag zou overleven, maar die laatste is dan weer intelligenter in de creatie en het gebruik van technologie.

Een tweede belangrijke conclusie is dat de rol van menselijke expertise van groot belang is, bij zowel kennisgebaseerde als datagebaseerde AI. De tijd dat de mens aan de kant gezet kan worden of moet onderdoen is nog lang niet aangebroken en zal er misschien ook nooit komen. Bij kennisgebaseerde AI is de menselijke expert de bron van de concepten, modellen, probleemoplossingsstrategieën en taal. Bij datagebaseerde AI is het voor de meeste toepassingen nodig dat de

menselijke expert eerst een corpus samenstelt van voorbeelden en dat hij elk voorbeeld nauwkeurig annoteert. Het AI-systeem vindt dan mogelijk wel nieuwe aspecten in de input, die voor ons niet zichtbaar zijn of niet opvallen, en het kan ook nieuwe verbanden ontdekken doordat het zo'n massale hoeveelheid voorbeelden kan bekijken. Er zijn ook technieken die puur inductief werken (*non supervised learning*), zonder menselijke tussenkomst, maar die hebben een beperkte toepasbaarheid. De krachtigste oplossingen zullen komen uit een symbiose van menselijke kennis en vaardigheden en artificiële intelligentie. In dit verband spreekt men dikwijls over IA (*intelligence amplification*, versterking van de [menselijke] intelligentie), eerder dan over AI (*artificial intelligence*), of ook soms over *augmented intelligence* (uitbreiding van de menselijke intelligentie).

II. Opportuniteiten

Artificiële intelligentie is een generische technologie zoals informatica, die in alle domeinen van de menselijke activiteit een belangrijke rol kan spelen. Over het algemeen kan men stellen dat AI de volgende fase is in de informatisering van zowat alles wat een mens doet. Maar het is tegelijk nuttig enkele domeinen uit te lichten en zo beter de opportuniteiten te zien waar AI in de nabije toekomst (de komende vijftien jaar) een ingrijpende impact zal hebben op onze samenleving. Deze lijst bouwt voort op het rapport *Artificial Intelligence and Life in 2030* (Grosz, 2016), een interimrapport van een project dat over een periode van honderd jaar periodiek de vooruitgang en impact van AI onderzoekt. Het rapport noemt vervoer, gezondheidszorg, onderwijs, veiligheid, robots, cultuur en hulp voor achtergestelde groepen als domeinen waar AI voor de burger het sterkst voelbaar zal zijn. Uiteraard zijn er nog meer domeinen waar de impact van AI nu al groot is en nog zal toenemen, zoals wetenschappelijk onderzoek, marketing, organisatie en bestuur, maar hier is de impact indirect.

1. VERVOER. Het lijkt geen twijfel meer dat het vervoer in de toekomst elektrisch zal zijn en autonoom. De autonomie wordt aangestuurd door twee ontwikkelingen: de komst van de intelligente auto en van intelligente transportsystemen. Stappen naar de intelligente auto worden al jaren gezet. Meer en meer sensoren worden nu al ingebouwd en allerlei functionaliteiten, zoals automatisch parkeren of ABS (*anti-lock braking system*), zijn reeds gemeengoed. Maar pas sinds de laatste jaren worden zelfrijdende auto's ook uitgetoetst, eerst door AI-laboratoria en nu ook door vrijwel alle autofabrikanten. Er is geen reden meer waarom de waarneming en controlefuncties die nodig zijn voor het besturen van een auto niet door een AI-systeem uitgevoerd kunnen worden, wellicht beter dan door de mens. De nodige navigatiesystemen bestaan ook en bovendien kunnen het onderhoud en de reparatie van een auto drastisch verbeterd worden door het inzetten van diagnosesystemen zoals die reeds in de jaren negentig zijn gedemonstreerd. Al deze technieken zijn ook relevant om drones de nodige autonomie te geven,

waardoor ze op korte afstand leveringen kunnen doen. Ook autonome scheepvaart komt er aan.

Intelligente transportsystemen zijn in opmars als onderdeel van de *smart city*. Zij proberen het verkeer in de stad of op snelwegen te meten en beter te coördineren. Dat vergt een netwerk van sensoren, de analyse van enorme hoeveelheden data om in het verkeer patronen te herkennen, en technieken om dat verkeer te stroomlijnen en terug te koppelen naar individuele auto's en naar het openbaar vervoer. Dit zijn stuk voor stuk domeinen waar AI een doorslaggevende rol kan spelen.

Planningssystemen zijn cruciaal om het transport van materialen zo efficiënt mogelijk te organiseren. Ook hier kunnen AI-technieken voor een enorme sprong voorwaarts zorgen. Het is dan ook niet verwonderlijk dat bedrijven die gespecialiseerd zijn in bestellingen en leveringen, zoals Amazon, zeer actief zijn inzake AI. Ten slotte is AI relevant om het delen van persoonlijke transportmiddelen (fietsen, auto's) beter te ondersteunen, onder meer door patronen te vinden in vraag en aanbod.

De introductie van AI in het vervoer zal een omwenteling betekenen die hopelijk leidt tot minder vervuiling en minder ongevallen. Een eigen auto zal minder nodig zijn en er zal minder tijdverlies zijn door files. Maar deze ontwikkelingen zetten ook een groot aantal banen op de tocht, gaande van taxichauffeurs of vrachtwagenbestuurders tot personeel in garages, pakjesbezorgers of mensen die betrokken zijn bij de fabricatie van auto's.

2. GEZONDHEIDSZORG. Medische toepassingen waren al een van de belangrijkste motoren vóór de eigenlijke ontwikkeling van AI-technologie eind jaren zeventig. Ondertussen is er gewerkt aan systemen die in sommige gevallen in staat zijn tot een betere patroonherkenning van testdata (bijvoorbeeld X-stralenanalyse, scanners, biomoleculaire analyses) dan menselijke experts. De stand van zaken bij de ondersteuning van medische diagnoses is ook sterk geëvolueerd, bijvoorbeeld door te gaan zoeken in de immense en snelgroeiende medische literatuur naar gevallen die lijken op dat van een bepaalde patiënt. Ook zijn er allerlei toepassingen mogelijk om thuis patiënten op te volgen, te coachen en alarm te slaan indien nodig.

Er wordt ook hard gewerkt aan robots en andere apparaten met een ingebedde intelligentie die helpen bij operaties of complexe medische ingrepen. Recentelijk worden de technieken van *machine learning* ook gebruikt om epidemiologische studies uit te voeren op basis van data die taalverwerkingsprogramma's afleiden uit sociale media, bijvoorbeeld om risicogroepen te identificeren en medische hulp efficiënter en op tijd te brengen bij diegenen die ze het hardste nodig hebben.

De opportuniteiten voor het gebruik van AI in de gezondheidssector liggen dus voor de hand. In veel landen zijn deze ontwikkelingen meer dan welkom, omdat de medische sector onder grote druk staat om het hoofd te bieden aan de vergrijzing van de bevolking en de bevolkingsexplosie. Het effectieve gebruik van de mogelijkheden kan nog een hele tijd op zich laten wachten omdat AI diep ingrijpt in de gewoonten zowel van het medisch personeel als van patiënten. Er bestaan bijvoorbeeld enorme hoeveelheden data die leersystemen zouden kunnen gebruiken om nieuwe relaties te ontdekken tussen symptomen en profielen enerzijds en ziekten en therapieën anderzijds. Maar daarbij rijzen ook belangrijke vragen over privacy en de betrouwbaarheid van de gebruikte leeralgoritmen.

We willen ook benadrukken dat alle AI-technologieën op het terrein van de gezondheidszorg zeker geen aanleiding kunnen zijn voor het verminderen of uitschakelen van de rol van de mens, integendeel. Het zijn hulpmiddelen die het voor het zorgpersoneel mogelijk moeten maken om zich meer bezig te houden met menselijke aspecten, zoals empathie, ethische vraagstukken en de gevoelsmatige ondersteuning van patiënten. Bovendien moeten we de menselijke expertise blijven doorgeven, ontwikkelen en uitbreiden, en moeten mensen blijvend de ultieme verantwoordelijkheid nemen voor medische beslissingen.

3. ONDERWIJS. De nood aan onderwijs, met inbegrip van naschools of volwassenonderwijs, is enorm gestegen. AI-onderzoekers zijn al sinds de jaren zestig bezig met de bouw van allerlei systemen ter ondersteuning van het onderwijs met computers. Dat deden ze eerst door *programmed instruction* te operationaliseren, vervolgens met leeromgevingen zoals LOGO en Lego LOGO die kinderen spelenderwijs leren programmeren en concepten van wiskunde bijbrengen, en daarna met interactieve intelligente tutors, bijvoorbeeld in het taalonderwijs. Net als in andere domeinen zijn er gradaties in de hoeveelheid intelligentie die in een systeem wordt ingebed. Intelligente tutors proberen een model te maken van de kennis van de leerling en reiken gepersonaliseerde voorbeelden en feedback aan om het leerproces zo optimaal mogelijk te begeleiden. Ook worden nu taalinterfaces gebruikt en dialoogtechnieken om een vloeiender interactie te realiseren tussen de leerling en het systeem.

Het grootste probleem bij computergebaseerd onderwijs was lange tijd hoe je deze systemen tot bij mogelijke gebruikers kon brengen. Dat werd midden jaren negentig opgelost door de opkomst van het web, dat het mogelijk maakt om direct en gratis onderwijs te verschaffen aan zeer grote groepen mensen, ook zij die normaal geen toegang hebben tot hoogkwalitatief onderwijs. Het resultaat is dat inmiddels miljoenen mensen taaltutors als DuoLingo gebruiken en dat studenten in ontwikkelingslanden bijna net zo goed Stanford-cursussen over AI kunnen volgen als de studenten op Stanford zelf. Een verdere stroomversnelling is ontstaan met de zogenaamde MOOCs (*Massive Online Open Courses*). Die hebben niet alleen tutoringfuncties maar ze gebruiken ook mechanismen van sociale media

om studenten te motiveren, een probleem bij veel computergebaseerd onderwijs. Sommige MOOCs worden door honderdduizenden studenten over de hele wereld gevolgd. Het is dan ook niet langer mogelijk oefeningen manueel te verbeteren of menselijke feedback te geven. Ook hiervoor worden AI-technieken ingezet.

Het lijkt weinig twijfel dat gepersonaliseerde onlineleersystemen een steeds grotere impact zullen hebben op het onderwijs. De huidige toepassingen gaan al zeer ver. Zo zijn er MOOCs op het vlak van muziekopleiding waarbij studenten hun oefeningen registreren en via internet opsturen als audiobestanden. Die worden automatisch geanalyseerd, verbeterd, vergeleken en gedeeld met die van andere studenten. De AI-muziektutors gebruiken multimedia om leer materiaal te presenteren. Studenten die een instrument willen leren bespelen krijgen zelfs feedback over hun houding en hoe ze bepaalde tonen beter kunnen spelen na een visuele analyse van video's van hun spel. Een extra voordeel is dat de ontwikkelaars van MOOCs grote hoeveelheden data over de leerpatronen van veel studenten krijgen. Die worden met *machine learning* geanalyseerd en gebruikt om de cursussen beter te maken (Steels, 2015).

4. VEILIGHEID. We geven eerste enkele voorbeelden van onveiligheid en criminaliteit om duidelijk te maken waar AI een rol kan spelen: neem de terroristische aanvallen die heel wat mensen beroeren omdat ze een impact hebben op hun dagelijks leven. Of misdadige praktijken op het internet die steeds moeilijker te controleren vallen en de witteboordcriminaliteit die door de globalisering moeilijker te detecteren valt. En er zijn de klassieke veiligheidsproblemen in de bedrijfswereld, bijvoorbeeld in chemische bedrijven.

Ook op dit domein is AI onontbeerlijk gebleken om datastromen aan te pakken. Doordat er nu via camera's, sociale media en andere kanalen immens veel data beschikbaar zijn, is het niet langer doenbaar om ze met menselijke inspectie snel genoeg te doorzoeken. AI-technologieën, zoals gezichtsherkenning, automatische vertaling, patroonherkenning en vooral automatisch leren om voorspellingen te doen waar mogelijke problemen kunnen ontstaan, worden meer en meer ingezet. AI wordt ook beschouwd als een mogelijk krachtig wapen in het ontmaskeren van nepnieuws.

Defensie is een ander facet van veiligheid. Het is bekend dat vooral Amerika enorm veel investeert in het gebruik van AI bij oorlogsvoering: de inzet van robots bij terreinverkenningen, automatische wapens, logistiek enz. Sommige Europese landen, zoals Frankrijk, zijn ook op dit domein actief, hoewel de toepassingen in onze contreien maar weinig aanhangers hebben.

5. ROBOTS. Hoewel AI dikwijls geassocieerd wordt met robots, zijn robots, zeker op relatief korte termijn, niet het meest evidente platform dat een grote impact zal hebben op het alledaagse leven van de meeste burgers, niet omdat AI niet relevant is voor robotica, maar omdat de problemen inzake mechanica,

batterijen, veiligheid en productie erg groot zijn. De markt voor robots met enige AI-functionaliteit situeert zich dan ook vooral in industriële productie-eenheden, zoals de auto-assemblage of de verpakkingindustrie. De rol van AI ligt hier zowel in het aanbrengen van de nodige componenten als in het monteren en de kwaliteitscontrole.

De huisrobot is nog niet doorgedrongen en zit er ook niet meteen aan te komen. Het enige tot nu toe succesvolle voorbeeld is de robotstofzuiger, die in 2001 werd gelanceerd, ook in de lokale zaken voor huishoudtoestellen. Hij heeft een minimumaantal eenvoudige AI-functies (vermijden van obstakels, het systematisch doorkruisen van de ruimte, monitoring van de batterij). Ondertussen hebben deze robotstofzuigers een visuele perceptie, leren ze een 3D-model van een huis aan, kunnen ze zichzelf opladen enz., maar ze worstelen nog steeds met het probleem van de trappen, en ook de problemen met de mechanische onderdelen en het onderhoud zijn vooralsnog niet opgelost. Op korte termijn mogen we verwachten dat er in keukens ruimte zal zijn voor een 'extra paar handen' die helpen bij het koken en afwassen, maar ook de problemen bij het flexibel manipuleren van objecten zijn nog lang niet opgelost, hoewel de recente beschikbaarheid van goedkope manipulatoren en de ontdekking van nieuwe manipulatie-algoritmen en leersystemen veelbelovend zijn. Zoals in andere domeinen moeten we ook hier wellicht onze blik richten op een symbiose van mens en machine. Denk aan een 'exoskeleton' dat mensen helpt die niet meer kunnen lopen, of aan intelligente prothesen die handen of armen vervangen en die al dan niet direct worden aangestuurd door neurale impulsen. (Zie hierover het KVAB-Standpunt *Naar een inclusieve robotsamenleving*; Van Brussel en De Schutter, 2016).

6. CULTUUR. AI speelt nu reeds een enorme rol in aankopen via internet, als go-between tussen diegenen die inhoud creëren en zoeken. Zo is het opnemen en distribueren van muziek grotendeels verschoven naar de digitale media. Dit heeft geleid tot een enorm muziekaanbod, maar ook tot nieuwe bedrijfsactiviteiten die luisteraars helpen om te vinden wat ze zoeken en om nieuwe dingen te ontdekken (denk aan Spotify). AI speelt hierin een bepalende rol: door het gepersonaliseerde aanbod, de ontdekking en het uitvergrooten van trends, de monitoring van reacties via sociale media enz. De enorme vlucht van computers en telecommunicatie heeft ook aanleiding gegeven tot nieuwe media, vooral computergames en *virtual reality*. De toepassing van AI om dit verder uit te bouwen en meer 'intelligent' te maken is volop aan de gang.

De digitalisering is niet onverdeeld positief voor wie professioneel met de creatie van cultuur bezig zijn, omdat de verloning van hun werk veel moeilijker is geworden, onder meer door piraterij. Anderzijds heeft een creatief individu nu toegang tot enorme hoeveelheden digitale hulpmiddelen om zelfstandig veel functies uit te voeren waar vroeger specialisten voor nodig waren. We kunnen alleen maar hopen dat deze mogelijkheden een ferme boost zullen geven aan de

cultuursector, zodat meer mensen creatief kunnen zijn of toegang zullen hebben tot het werk van anderen.

7. ACHTERGESTELDE GROEPEN. AI kan in principe gebruikt worden om groepen van mensen in kansarmoede kansen te geven die ze anders nooit krijgen. Dit is al het geval met de intelligente tutoringapplicaties en met de MOOCS die via het web gratis beschikbaar zijn voor iedereen, zonder dat er hoge inschrijvingsgelden betaald moeten worden aan een of andere Amerikaanse of Engelse universiteit en zelfs zonder dat er in een land hogere opleidingen bestaan. Medische zorg kan beschikbaar gemaakt worden voor een grotere groep gebruikers, ook in gebieden waar vrijwel geen medische knowhow beschikbaar is, en al helemaal niet voor een betaalbare prijs. Andere vormen van expertise, bijvoorbeeld in de landbouw, het bosbeheer, urbanisme enz., kunnen via AI-technologie verspreid worden dankzij de groeiende penetratie van smartphones en internet.

Een ander voorbeeld zijn de AI-programma's achter sociale media, die het mogelijk maken dat groepen die anders geen stem hebben in de samenleving elkaar vinden en zich kunnen organiseren. Zo zijn er allerlei hulpmiddelen voor (meer) directe democratie. AI-programma's kunnen helpen om zaken op de agenda te zetten door te laten zien wat de consequenties zijn van bepaalde beslissingen. Een goed voorbeeld zijn milieukwesties. Met mobiele telefoons als meetinstrument, eventueel met sensoruitbreidingen, is het mogelijk dat groepen mensen zelf hun omgeving monitoren, bijvoorbeeld wat lucht- of waterverontreiniging betreft. Een voorbeeld hiervan is het project NoiseTube, waardoor mensen met hun smartphone metingen over vervuiling (geluid, lucht) kunnen uitvoeren en die automatisch versturen naar een centrale server. Die visualiseert en analyseert de data³.

Activiteiten inzake 'AI voor sociaal welzijn' krijgen heel wat aandacht in de AI-onderzoekswereld. Ze vormen een tegengewicht voor de overtrokken verhalen en sciencefictionfilms die dikwijls de boodschap overbrengen dat AI gevaarlijk is en dat ze de mens zal overtroeven. Voorbeelden van deze activiteiten zijn de reeks *Beneficial AI*-conferenties georganiseerd door het Future of Life Institute in Cambridge Ma (<https://futureoflife.org/bai-2017/>) en de initiatieven van het Partnership on AI-platform, dat vooral industriële partners met een belangrijke AI-activiteit groepeerd (<https://www.partnershiponai.org/>). Zoals elke technologie is AI breed inzetbaar, dus ook voor het verbeteren van ieders lot. Vlaanderen, en algemener de Europese Unie, zou hierop kunnen inzetten, in plaats van de nadruk te leggen op defensie, zoals in Amerika.

³ <http://www.ademloos.be/nieuws/resultaten-noisetube-project-tuinwijk-borgerhout>.

III. Stimulansen voor onderzoek en ontwikkeling in AI

Gezien al deze toepassingsgebieden is het geen wonder dat vanaf 2015 AI economisch gezien in een stroomversnelling is terechtgekomen (Economist, 2016). Een indicator zijn de investeringen in start-ups die zijn gespecialiseerd in AI. Ze zijn in vijf jaar tijd (2012-2017) gestegen van 415 miljoen dollar naar 5 miljard per jaar. Ook de grote informaticabedrijven investeren massaal in AI en richten onderzoekslaboratoria op, dikwijls in de buurt van universiteiten om het nodige talent aan te trekken. Recente voorbeelden (september 2017) zijn een IBM-investering van 250 miljoen dollar in een nieuw MIT-IBM laboratorium in AI, de uitbreiding van Google Research in Zürich met 2000 extra AI-onderzoekers en -ontwikkelaars (januari 2017) en de creatie van een nieuw Toyota AI-researchlaboratorium bij Stanford, een investering van 1 miljard dollar. De 'markt' voor ondersteunende AI-technologie wordt tegen 2025 geschat op 31 miljard dollar en als we er *data analytics* bijtellen komen we aan 70 miljard dollar (Tractica, 2017). Sommige prognoses spreken zelfs van een markt van 130 miljard euro tegen 2025, wat een jaarlijkse groei van 15 tot 25% zou betekenen (McKinsey, 2017). Daarin zitten nog niet alle producten en diensten die dankzij AI gemaakt en verkocht zullen worden. Er is dan ook vrijwel een consensus onder economische experts: AI zal een cruciale rol spelen in de economische ontwikkeling van een land. Of, zoals de Russische president Putin het in september 2017 uitdrukte: 'De natie die leidt in AI zal de wereld domineren.'

Onlangs verscheen een eerste studie van McKinsey die de mogelijke impact van AI op de Belgische/Vlaamse economie in kaart brengt⁴. Dit rapport voorspelt een extra groei met 1% van het BNP tegen 2039, maar alleen als AI effectief in de economie geïntegreerd wordt. Dat lijkt een onderschatting in vergelijking met buitenlandse studies.

Voor zover bekend heeft de Vlaamse regering nog geen strategisch masterplan voor AI. In het Brussels Gewest kwam er recent een initiatief van Innoviris voor de creatie van een 'AI-ecosysteem' dat bedrijven en universiteiten samenbrengt. Er wordt 4 miljoen euro uitgetrokken om uitmuntend onderzoek, technologietransfer, het gebruik van AI in bedrijven en het maken van nieuwe producten te stimuleren. De oproep kreeg een zeer grote respons. De projectevaluatie is ondertussen achter de rug en de samenwerkingsprojecten zijn van start gegaan. In de volgende pagina's overlopen we wat er in het buitenland gebeurt. Daaruit kunnen we extrapoleren wat het effect zou kunnen zijn van AI in Vlaanderen. Bovendien bevatten de strategische plannen van andere landen suggesties voor Vlaanderen: hoe kan de ontwikkeling en ook de verspreiding van AI het best ondersteund worden?

⁴ <https://www.vrt.be/content/dam/vrtnieuws/bestanden/2017/Shaping-the-future-of-work-in-Europes-digital-front-runners.pdf>.

Noord-Amerika

De Verenigde Staten en Canada zijn sinds lang koploper op het domein van AI, met langetermijninvesteringen, een netwerk van uitmuntende en stabiele laboratoria, universitair onderwijs in AI op een zeer hoog niveau en een zeer actieve en dynamische informatica-industrie. De influx van R&D-middelen voor AI wordt in de Verenigde Staten geschat op 30 miljard per jaar. Dankzij de sterke onderwijs-, onderzoeks- en bedrijfscultuur is er een groot aantal start-upbedrijven waar AI een belangrijke en wellicht de belangrijkste competitieve component van uitmaakt. Sommige zijn intussen uitgegroeid tot miljardenondernemingen, met Google als een van de belangrijkste voorbeelden.

Omdat de nieuwe economie voor een groot stuk op AI steunt, is het niet verwonderlijk dat de Verenigde Staten hierin de absolute koploper zijn geworden, met bedrijven als Google, Amazon, Facebook, Uber, Airbnb, Booking.com, Salesforce enz., die allemaal belangrijke R&D-afdelingen voor AI hebben. Ook de toeleveranciers van basistechnologie voor AI – de gespecialiseerde processoren van NVIDIA, de *cloud computing services* van Amazon of Microsoft, de *knowledge graph services* van Google enz. – zijn bijna uitsluitend in Amerikaanse handen. In een eerste stadium gebruikten deze bedrijven AI-technologie vooral voor zichzelf, maar gezien de grote vraag stellen ze die nu ter beschikking als services waarmee andere bedrijven hun diensten en producten kunnen verbeteren (Fast, 2017). Dat gaat van het oproepen van *machine learning*-programma's tot patroonherkenning in grote hoeveelheden data, automatische vertaling en *conversational agents*, die bijvoorbeeld de rol van een verkoper deels kunnen overnemen of een aantal problemen bij een Helpdesk autonoom kunnen oplossen.

Onder president Obama bereidde de Amerikaanse regering belangrijke strategische documenten voor, zoals *Preparing for the future of Artificial Intelligence* (Holdren en Smith, 2016), dat de opportuniteiten en het belang van AI voor de Amerikaanse economie beschrijft, en het *National Artificial Intelligence Research and Development Strategic Plan* (Biegel en Kurose, 2016), dat aangeeft hoe de V.S. hierop zullen inspelen.

Ook Canada is bijzonder actief. Zo is er recent een groot project van 1 miljard euro opgestart voor de creatie van een 'AI-ecosysteem' dat bedrijven, financiers, universiteiten en onderzoekslaboratoria samenbrengt, met de expliciete bedoeling om nieuwe start-ups in het domein van AI mogelijk te maken. Deze AI-hub is gevestigd in Toronto⁵.

⁵ <https://www.nextcanada.com/next-ai>.

China

China heeft een netwerk van instituten en onderzoeksactiviteiten aan universiteiten en in bedrijven. Dat situeert zich vooral bij Baidu, de Chinese tegenhanger van Google, die researchlabs heeft voor *deep learning*, big data, en *general AI*, maar ook bij bedrijven als Alibaba, Tencent en Xiaomi. De bijdrage van Chinese onderzoekers is de laatste tijd enorm toegenomen, bijvoorbeeld op de laatste IJCAI-conferentie, de belangrijkste wereldconferentie voor AI die in 2017 plaatsvond in Melbourne. Een derde van de inzendingen en een kwart van de deelnemers kwam uit China.

Ook China beschouwt AI als van immens strategisch belang voor de toekomstige economie, zowel voor de productie van goederen, de huidige sterkte van de Chinese economie, als voor het integreren van AI in allerlei consumentenproducten. Tegelijk is men er zich bewust van de achterstand op Amerika en zette men onlangs (2017) een mega-initiatief op om die om te buigen in een voorsprong. De investeringen lopen in de tientallen miljarden dollar en het doel is om tegen 2030 een AI-kernindustrie op te zetten van 150 miljard dollar. Op haar beurt moet die een tienvoudige impact hebben op de economie (1,5 biljoen euro)⁶.

Europa

Vreemd genoeg hebben Europese onderzoeksprogramma's de laatste decennia nooit veel interesse getoond voor AI, terwijl men had kunnen verwachten dat er een *flagship*-programma voor AI geweest zou zijn, of een sterke component voor fundamenteel AI-onderzoek in de FET-programma's (*Future and Emerging Technology*), of een significant aantal ERC-grants in het domein van AI. Die lauwe ondersteuning van AI-onderzoek is deels te wijten aan het algemene gebrek aan onderzoekers en onderzoeksinfrastructuur voor AI in de meeste Europese landen, waardoor er wellicht te weinig voorstellen zijn gekomen en er te weinig invloed was op de Europese bottom-up besluitvorming. Als er al projecten zijn, komen ze vooral uit het Verenigd Koninkrijk, dat ruimschoots de vruchten heeft geplukt van de Europese investeringen: grote groepen (Google, Amazon, Microsoft e.a.) hebben er onderzoekslaboratoria gevestigd, vooral in samenwerking met universiteiten zoals Cambridge, Oxford, Edinburgh, UCL en Imperial (Londen).

Natuurlijk zijn er oproepen geweest die raakvlakken hebben met AI, zoals voor de mechanische aspecten van Robotica⁷, de steun voor cognitieve wetenschappen (zoals het EUCOG-programma), taaltechnologie en semantische technologieën voor het web, en vooral de computersimulatie van de biologie van het brein (het *Human Brain Flagship*-project), hoewel de impact hiervan op AI zeer marginaal

⁶ http://english.gov.cn/state_council/ministries/2017/09/08/content_281475843927886.htm.

⁷ <https://ec.europa.eu/programmes/horizon2020/en/news/new-horizon-2020-robotics-projects-2016>.

is). Er zijn ook een aantal toepassingsgerichte projecten geweest met een AI-component. Al deze projecten hebben echter niet als doel om AI zelf vooruit te helpen en in een solide langetermijnbasis te voorzien waaruit vervolgens de AI-industrie kan groeien. AI-onderzoekers koesteren een lichte hoop dat dit in de toekomst kan veranderen. Zo werd eind 2017 een eerste call gelanceerd voor de bouw van een AI-platform (in feite een AI-ecosysteem met vooral bedrijven; bedrag: 20 miljoen euro) dat misschien de kiem van grotere projecten zal worden. In elk geval zijn de investeringen in R&D voor AI vanuit de Europese Unie niet in verhouding tot wat in de V.S. en China gebeurt. Om een idee te geven: alleen al het met privé-geld gefinancierde OpenAI-project in de V.S. (Berkeley), dat op zoek gaat naar veilige algemene AI, heeft een basisfinanciering van 1 miljard dollar. Anderzijds zien we wel dat de ons omringende Europese landen recentelijk het belang van AI hebben onderkend en dat ze elk op hun manier plannen hebben opgesteld. Nederland staat op het punt een grote actie over AI op te zetten (25 miljoen euro), met daarbij ook de creatie van een nationaal AI-centrum. AI is ook sterk ontwikkeld in Italië en Zwitserland, en zoals al gezegd nog meer in het Verenigd Koninkrijk, dat hiervoor veel Europese steun kreeg en dankzij AI een stijging van 10,3% van het BNP verwacht tegen 2030⁸.

We kijken hier ook nog even naar Frankrijk en Duitsland, twee van onze belangrijkste handelspartners. Frankrijk heeft een lange traditie in het AI-onderzoek, hoewel dat minder zichtbaar is dan het Amerikaanse. Het vindt vooral plaats in laboratoria van INRIA, in mindere mate CNRS en nog meer in diverse bedrijfslaboratoria, zoals Renault, Thales en Airbus. Een recent rapport vermeldt 68 laboratoria, waarin alles samen 13.250 onderzoekers en ontwikkelaars actief zijn. Nog onder president Hollande zijn er strategische plannen gemaakt voor een groots opgezette stimulans die alle AI-actoren moet samenbrengen. *France IA* werd gelanceerd in januari 2017⁹ met als doel: (1) Het in kaart brengen van de activiteiten rond AI in Frankrijk. (2) De identificatie en het bepalen van de onderzoeksonderwerpen en onderwijsnaden. (3) Technologietransfer en focus op de realisatie van het onderzoekspotentieel, met daarin verschillende deelgroepen die sleuteldomeinen voor AI ontginnen zoals: robotica en het Internet van Dingen (IoT), energie en *smart cities*, hybride technologie, computerondersteund onderwijs, gezondheid en medische toepassingen, AI in de administratie e.a. (4) De ontwikkeling van AI-toelevering door Franse bedrijven voor technologie en diensten, met vooral aandacht voor autonome auto's, klantrelatie en toepassingen in de financiële sector. (5) Veiligheid: de mogelijke rol van AI voor defensie en nationale veiligheid. (6) Anticiperen van de macro-economische en sociale omwentelingen die AI zal teweegbrengen.

⁸ <https://www.pwc.co.uk/economic-services/assets/ai-uk-report-v2.pdf>.

⁹ <https://www.economie.gouv.fr/files/files/PDF/DP-lancement-strategie-IA.pdf>.

De strategische plannen bevatten ook middelen voor de creatie van een ecosysteem, *France is AI* (<https://franceisai.com/>), dat een aantal partners samenbrengt: meer dan 250 start-ups, onderzoekers (INRIA, CNRS, universiteiten), incubatoren met een focus op AI (Agoranov, Telecom Paris Tech, Robot Lab), verenigingen van onderzoekers en ontwikkelaars (AFIA), en investeringsfondsen gespecialiseerd in AI (SERENA Capital).

Ook Duitsland heeft een lange traditie in AI-onderzoek. AI wordt er aan alle universiteiten onderwezen, er is een nationaal laboratorium voor AI-onderzoek (DFKI; <https://www.dfki.de/web>) en er bestaat een uitgebreid netwerk van bedrijven waarin AI aan bod komt, vooral in de auto-industrie maar ook in de informatica (bijvoorbeeld SAP). Momenteel wordt een belangrijk nieuw impulsprogramma voorbereid, dat vergelijkbaar is met *France IA*. De Duitse aanpak springt minder in het oog dan de Amerikaanse, maar is daarom niet minder *grundig*. Duitsland gaat ervan uit dat de bevolking verouderd en dat er in de toekomst onvoldoende mankracht zal zijn om de economische productie op peil te houden. Dat is ook het belangrijkste argument waarom in Japan de AI-investeringen de laatste jaren sterk zijn gestegen. In landen als Frankrijk is men eerder bezorgd is om de werkloosheid.

Aangezien Duitsland sterk staat in de industriële productie, concentreren de inspanningen zich in de eerste plaats op hoe AI hierin een (nog grotere) rol kan spelen. Denk aan predictief onderhoud om fouten in de fabricage te corrigeren en onderhoudsprocessen te optimaliseren, collaboratieve robots met het oog op symbiose op de werkvloer (in plaats van een volledige automatisering), kwaliteitscontrole enz. (McKinsey, 2017) Ook op andere domeinen, zoals *supply-chain management*, de versnelling van het onderzoek, marketing, bedrijfsprocessen, personeelsbeleid enz. zijn Duitse bedrijven en onderzoekscentra zeer actief. Daarnaast concentreren de inspanningen zich ook op de productieve inbedding van AI. Het belangrijkste voorbeeld van dat laatste is de autonome auto. Duitse projecten (zoals aan de Vrije Universiteit van Berlijn; <http://autonomos-labs.com/>) krijgen minder publiciteit dan de Amerikaanse, maar boeken wel uiterst sterke resultaten, die Duitse autobouwers in toekomstige producten zullen inbouwen. De prognose is dat tegen 2030 15% van alle verkochte auto's autonoom zal zijn.

IV. Onderwijs

Gezien de zeer sterk gestegen interesse is er ondertussen een enorme vraag naar competente AI-ontwikkelaars, die in grote drommen worden weggehaald uit universiteiten (Sample, 2017). Vorig jaar kaapte Uber bijvoorbeeld een derde van het personeel van het robotics laboratorium van Carnegie Mellon weg. De flessenhals voor de ontwikkeling van AI ligt dan ook niet in de (on)beschikbaarheid van grondstoffen, maar eerder in het aantrekken van het nodige talent en het vinden van ontwikkelaars met voldoende competentie in AI. Zo belanden we bij de

kwestie van onderwijs. De toepassing van AI, en nog veel meer het onderzoek in AI, vergt een zeer gedegen interdisciplinaire kennis, met een grondige basisvorming in de computerwetenschappen en wiskunde, maar ook kennis van en affiniteit met humane wetenschappen (logica, taalkunde, psychologie). Een vorming in de technische kanten van AI neemt verschillende jaren in beslag en de evolutie van AI gaat razendsnel, zodat permanente bijscholing noodzakelijk is.

Aan alle Vlaamse universiteiten is AI aanwezig, maar vooral in geïsoleerde cursussen en keuzevakken in de diverse faculteiten, zoals in de biomedische wetenschappen, menswetenschappen, communicatiewetenschappen of economische wetenschappen. De cursussen zijn vooral bedoeld om de studenten noties aan te reiken waar AI over gaat, maar ze zijn te beperkt om vervolgens AI ook zelf toe te passen, laat staan om onderzoek in dit vakgebied te doen.

Opleidingen die studenten klaarstomen voor de technische kant van AI leunen aan bij de studierichtingen Computerwetenschappen of bij ingenieursstudies. Op bachelorniveau biedt geen enkele universiteit een aparte studierichting AI aan. Wel komt een vak AI voor, meestal als verplicht vak en soms zelfs alleen als keuzevak. Op masterniveau is er wel een aantal AI-specialisaties of specialisaties waar AI een belangrijke component vormt binnen de studierichtingen informatica:

- Aan de UGent is er een verplicht vak AI in de opleiding Master of Computer Science (faculteit Wetenschappen). De ingenieurs volgen in de opleiding Master of Computer Science Engineering een verplicht vak Machinaal Leren en een keuzevak AI. Er zijn ook diverse deels verwante keuzevakken in deze en andere masterprogramma's. Een specialisatie AI is er niet. Binnen de opleiding Toegepaste Taalkunde (faculteit Letteren en Wijsbegeerte) is er ook de mogelijkheid om computertaalkunde te studeren.

- Aan de KU Leuven is er een uitgebreide hoofdoptie AI in de master in ingenieurswetenschappen. In computerwetenschappen is er naast de normale AI-basisvakken, zoals AI programmeren en *machine learning*, een brede waaier aan keuzevakken: *data mining*, taalverwerking, visie en robotica enz.

- Aan de VUB is er ook een specialisatierichting AI in de opleiding Computerwetenschappen. Naast de te verwachten vakken (AI programmeren, *machine learning*, natuurlijke taalverwerking, multi-agent systemen enz.) zijn er vakken die sterk de biologische kant van AI benadrukken (*swarm intelligence*, bio-informatica, evolutie van spraak, adaptieve systemen). De VUB werkt hiervoor nauw samen met de ULB.

- De Universiteit Antwerpen heeft geen specialisatierichting in AI, maar wel in *datascience* (waarin het vak AI weliswaar niet voorkomt, maar diverse vakken overlappen wel met een AI-opleiding). In de master Linguïstiek is er een keuzespecialisatie in *computational psycholinguistics*, die studenten opleidt in natuurlijke taalverwerking en computertaalkunde.

- Aan de Transnationale Universiteit Limburg worden de AI-opleidingen hoofdzakelijk georganiseerd in Maastricht. Het gaat meer bepaald om een master in AI met de nadruk op *data science* en *knowledge engineering*.

- Ten slotte is er de succesvolle eenjarige master na master aan de KU Leuven¹⁰ met speciale opties in technologische AI (de optie engineering en computerwetenschappen, met facetten als robotica, beeldverwerking, onzekerheid, neurale netwerken), de spraak en taaltechnologie (taalverwerking, automatische vertaling), en de optie *big data analytics* (met thema's als tekstgebaseerde informatievergaring en *machine learning*).

In vergelijking met Nederland, is het aanbod in Vlaanderen beperkter. In Nederland wordt AI beschouwd als een zelfstandige discipline, op hetzelfde niveau als wiskunde, informatica en natuurwetenschappen. Vrijwel elke universiteit (niet de technische universiteiten) heeft een (Engelstalige) bacheloropleiding AI, die wordt gevolgd door een tweejarige master. De instroom van studenten is zeer groot, met een gemiddelde van 150 tot 200 eerstejaars per universiteit. Sommige (zoals de Universiteit van Amsterdam) hebben zelfs een numerus clausus ingevoerd. Deze opleidingen worden aan verschillende faculteiten gegeven, afhankelijk van waar de nadruk op ligt, maar vooral in de masterprogramma's die aanleunen bij de computerwetenschappen. Twee voorbeelden:

- aan de Universiteit van Amsterdam maakt de opleiding deel uit van de faculteit Natuurwetenschappen, Wiskunde en Informatica¹¹. Hetzelfde geldt voor de Vrije Universiteit Amsterdam;

- aan de Radboud Universiteit in Nijmegen valt AI dan weer onder Psychologie en Kunstmatige Intelligentie¹², met een master die sterk gericht is op neurale systemen en psychologische modelering. Dezelfde nadruk op menswetenschappen zien we in Utrecht, met een master die vooral focust op symbolische AI (*agents, cognitive processing, redeneringen*)¹³.

De vraag rijst waarom het AI-onderwijs in Vlaanderen veel minder is uitgebouwd dan in Nederland. De belangrijkste verklaring is wellicht dat door de nagenoeg volledige stop op nieuwe opleidingen in Vlaanderen een bacheloropleiding in AI heel moeilijk te realiseren valt, ook omdat bestaande disciplines er een mogelijke concurrent in zien van hun eigen opleidingen. Veel disciplines worden geconfronteerd met een krimpend personeelsbestand, zodat nieuwe aanstellingen eerder in de bestaande kernactiviteiten van de discipline voorkomen dan in nieuwe richtingen, zelfs al zijn die sterk toekomstgericht. Bovendien hebben sommige AI-opleidingen in Nederland een niet-technologisch gerichte component, zodat ze heel wat studenten aantrekken die niet noodzakelijk de wiskundige en wetenschappelijke achtergrond hebben van de opleidingen ingenieurs- of

¹⁰ <https://wms.cs.kuleuven.be/cs/opleidingen/master-artificial-intelligence>.

¹¹ <http://www.uva.nl/programmas/bachelors/kunstmatige-intelligentie/kunstmatige-intelligentie.html> en <https://masters.vu.nl/en/programmes/artificial-intelligence/index.aspx>.

¹² <http://www.ru.nl/opleidingen/bachelor/artificial-intelligence/> en <http://www.ru.nl/english/education/masters/artificial-intelligence/>.

¹³ <https://www.uu.nl/bachelors/kunstmatige-intelligentie> en <https://www.uu.nl/masters/en/artificial-intelligence>.

computerwetenschappen, maar toch zeer productief werk kunnen verrichten in AI, vooral dan wat de facetten 'taal' en 'denken' betreft.

V. Risico's

Met elke technologie zijn risico's verbonden. Dat is voor AI niet anders. Die waarover in de media het meest gesproken wordt, zijn dikwijls overtrokken en leiden mensen af van de echte risico's: oneigenlijk gebruik, onbetrouwbaarheid, gebrek aan transparantie en verantwoordelijkheid, de schending van rechten en het risico op verlies van menselijke kennis. Onlangs werden diverse initiatieven genomen om de risico's in kaart te brengen en ze in te perken, onder meer door het Future of Life Institute in Boston, het IEEE Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems (opgezet door de IEEE Standards Organization¹⁴.) Ook worden nieuwe instituten opgericht, zoals het AI NOW-instituut van New York University (november 2017), dat specifiek als doel heeft de sociale implicaties van AI te onderzoeken¹⁵. Er worden daarnaast ook gespecialiseerde workshops en conferenties georganiseerd¹⁶.

Ook de Europese Unie zet stappen voor de juridische omkadering van het gebruik van AI-systemen. Ze beperkt en regelt onder meer het datagebruik en geeft burgers rechten wat geautomatiseerde individuele besluitvorming betreft, zoals profilering. Deze wet, de Algemene Verordening Gegevensbescherming die in mei 2018 in voege treedt (GDPR; EPC, 2016), behandelt ook expliciet de vraag naar uitleg: of door de huidige wet een recht op uitleg vastgelegd is, wordt momenteel bediscussieerd en zal de komende jaren ongetwijfeld het voorwerp van jurisprudentie en wetenschap blijven (Wachter, 2017; Selbst en Powles, 2017). De verordening is zeer belangrijk, omdat ze ook een zeer grote impact zal hebben op welke AI-systemen in Europa na 2018 nog gebruikt kunnen worden.

We gaan wat dieper in op mogelijke risico's. Inspiratie hiervoor komt van de recente Barcelona Declaration for Proper Development and Usage of Artificial Intelligence, die werd gelanceerd door het B-debate International Centre for Scientific Debate in Barcelona¹⁷, en de IEEE Principles of Ethically Aligned Design¹⁸. Binnen de Vlaamse AI werken onder meer de groepen van Bettina Berendt aan de KU Leuven en Toon Calders aan de Universiteit Antwerpen rond risico's van AI en de omgang ermee (FAT/ML, 2017)

¹⁴ http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html.

¹⁵ <https://medium.com/@AINowInstitute/the-10-top-recommendations-for-the-ai-field-in-2017-b3253624a7>.

¹⁶ <https://www.fatml.org/>.

¹⁷ <http://www.bdebate.org/en/synopsis/page/3-barcelona-declaration>.

¹⁸ http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf.

1. ONEIGENLIJK GEBRUIK. De sprong voorwaarts die AI maakte, is er gekomen door de toenemende maturiteit van AI-technologie, de enorme groei in computerkracht, de beschikbaarheid van internetplatformen voor de 'levering' van *intelligente services* en een groter wordende bereidheid van veel economische actoren om de technologie uit te proberen. We moeten ons echter ook bewust zijn van de limieten. Heel veel fundamentele problemen inzake AI zijn nog niet opgelost en zullen nog radicale doorbraken vergen, waar langetermijninvesteringen voor nodig zijn. Bovendien moet de toepassing van AI op sterke randvoorwaarden berusten. Anders zullen de resultaten teleurstellend zijn en mogelijk gevaarlijk. Deze randvoorwaarden moeten nog veel meer in kaart gebracht worden.

2. ONBETROUWBAARHEID. Vrijwel alle systemen en producten die in onze maatschappij gebruikt worden moeten strenge tests ondergaan om hun betrouwbaarheid en veiligheid te checken. Het zou normaal zijn dat hetzelfde gebeurt voor AI-systemen. Er zijn al sinds de jaren tachtig en negentig verificatie- en validatieprocedures ontwikkeld voor kennisgebaseerde systemen, maar vergelijkbare procedures voor datagebaseerde AI en voor AI-systemen *human in the loop* zijn nog een heel recent onderzoeksgebied. Uiteraard maakt de praktijk van automatisch leren een verschil tussen de data die gebruikt worden bij training en de data die daarna gebruikt worden om na te gaan of een systeem een voldoende competentieniveau heeft bereikt, maar er zijn steevast significante verschillen tussen een testdataset en de toepassing in realistische omstandigheden. Ook bedreigingen (hackeraanvallen, antagonistisch leren) en beveiligingsmethodes hiertegen moeten nog krachtiger in bestaande AI-methoden geïntegreerd worden. Bovendien hebben we in Europa een netwerk van agentschappen nodig (of een centraal Europese agentschap) dat toepassingen homologeert, zoals dat voor medicijnen en voedsel het geval is. Het Europees Parlement heeft in deze geest recent een plan goedgekeurd voor een agentschap Robotica en AI, dat mogelijkwijze deze taak op zich kan nemen.

3. INTERPRETEERBAARHEID. Wanneer een AI-systeem een beslissing neemt, moet degene die de impact van die beslissing ondergaat kunnen interpreteren en begrijpen waarom dit besluit is genomen en de beslissing ook kunnen aanvechten. De uitleg moet in een taal gegeven worden die de betrokkene verstaat (Doshi-Velez and Kortz, 2017). Dit is vooral belangrijk op het domein van financiële beslissingen (zoals de toekenning van een lening), juridische beslissingen (bijvoorbeeld of vroegtijdige vrijlating wordt toegestaan), verzekeringen, belastingen, geneeskundige verzorging enz. Veel van de huidige datagebaseerde AI-systemen zijn niet in staat om dit te doen. Hun beslissingen zijn gebaseerd op omvangrijke hoeveelheden data en een grote verzameling modelparameters en statistiek. Om deze beslissingen te kunnen begrijpen, interpreteren en eventueel aan te vechten, heeft men uitleg nodig over de data, de werkwijze van de modellen en de factoren die meegespeeld hebben in een specifieke beslissing. Welke combinaties van deze drie elementen nodig en mogelijk zijn, wordt op dit moment intens besproken, zowel in computerwetenschappelijk onderzoek naar *explainable AI*, AI-systemen die hun gedrag kunnen uitleggen (Biran en Cotton, 2017), als op juridisch vlak en

in de gedragswetenschappen. Hier komt AI in aanraking met rechten en principes die centraal zijn voor een democratie, zoals databescherming, transparantie en verantwoordingsplicht. De Europese Verordening Gegevensbescherming (GDPR) behandelt ook expliciet het recht op uitleg. Daar hadden we het hierboven al over. De vereisten van de Verordening zullen ongetwijfeld het onderzoek en de praktische uitwerking van deze concepten stimuleren.

4. TRANSPARANTIE. Er is een groeiende bezorgdheid over *AI-chat-bots* en soortgelijke systemen die opereren op het internet en in sociale media, en die bedoeld zijn voor de manipulatie van de politieke opinie, desinformatie door het propageren van valse feiten, afpersing of andere vormen van gedrag dat gevaarlijk is voor het individu en het destabiliseren van de maatschappij. Deze *chat-bots* pretenderen om menselijk te zijn en geven de identiteit niet bloot van wie erachter zitten. Het gebruik van AI heeft deze *chat-bots* heel realistisch gemaakt, zodat gebruikers misleid worden. Een mogelijke oplossing voor dit probleem is om het in elk geval duidelijk te maken wanneer een interactie vanuit een mens of een AI-systeem komt, en ervoor te zorgen dat, als het gaat om een artificieel systeem, degene die ervoor verantwoordelijk zijn getraceerd en geïdentificeerd kunnen worden. Deze oplossing zou bijvoorbeeld door een systeem van watermarkering kunnen gebeuren en verplicht worden in Europa.

5. VERANTWOORDELIJKHEID. AI-systemen hebben niet alleen de capaciteit om beslissingen te nemen. Wanneer ze ingebed zijn in fysieke systemen, zoals een zelfrijdende auto, hebben ze ook het potentieel om deze beslissingen om te zetten in acties in de reële wereld. Dit roept vragen op over de veiligheid en over wie verantwoordelijk en aansprakelijk is (Tanghe en De Bruyne, 2017). Het roept ook vragen op in hoeverre deze autonome systemen gevaarlijk kunnen worden en de mens overtroeven. Sommige zorgen behoren wel thuis in het domein van de sciencefiction, maar het is ontegensprekelijk nodig na te gaan hoe het gedrag van autonome intelligente systemen ingeperkt kan worden, waardoor we de nadelen kunnen verminderen. Duidelijke regels inzake autonomie zijn nodig, waardoor ontwikkelaars die ook effectief in hun systemen kunnen integreren, al vanaf de ontwerpfase. Op die manier kan ook de verantwoordelijkheid vastgesteld worden als er sprake is van falen, net zoals bij andere producten. Voor bepaalde types autonome systemen moeten waarschijnlijk strengere grenzen aan de autonomie gesteld worden (zie over autonome wapens verder in de tekst). In november 2017 heeft een commissie van hooggeplaatste AI-experten bij de Verenigde Naties aangedrongen op een preventieve ban van dodelijke autonome wapens en in december 2017 was er een hoorzitting in het Belgisch Parlement voor een soortgelijke ban in België¹⁹.

6. SCHENDING VAN GRONDRECHTEN EN ETHISCHE PRINCIPES. Naast de veelbesproken bedreigingen van het recht op leven (door AI-wapens) en het recht op privacy en databescherming (door profilering via AI, zie ook het KVAB-Standpunt

¹⁹ <http://www.dekamer.be/kvvcr/showpage.cfm?section=/flwb&language=nl&cfm=/site/wwwcfm/flwb/flwbn.cfm?legislat=54&dossierID=2219>.

over privacy; Berbers et al., 2017) laten huidige systemen al grote risico's zien voor andere fundamentele rechten, zoals de vrijheid van discriminatie, gelijke kansen en levensvriendelijke arbeidsomstandigheden. Daarnaast kunnen beslissingen van sociale robots, autonome auto's enz. allerlei ethische gevolgen hebben, waarvoor er niet noodzakelijk één 'correcte' keuze bestaat. Naar het voorbeeld van andere disciplines beginnen professionele codes van computerwetenschappen en AI een respect voor de mensenrechten te vereisen, in het besef dat deze rechten ook in conflict met elkaar kunnen staan.

AI-onderzoekers en -ontwikkelaars trachten dit doel op twee manieren te bereiken. Eén richting probeert rechten te formaliseren en in algoritmes te integreren, en dus te komen tot een formele verificatie of validatie van ethische en juridische concepten. Deze aanpak is belangrijk, maar rechten zijn vaak niet volledig formaliseerbaar en beslissingen zijn zelden volledig geautomatiseerd. Daarom moet deze benadering aangevuld worden door methodes zoals *value-based design* (ontwerp op basis van waarden en rechten), die ook met niet-formaliseerbare en conflicterende waarden werken. Een gelijkaardige aanpak wordt door de EU-Verordening GDPR ook wettelijk vereist ('gegevensbescherming door ontwerp'), wat opnieuw tot AI-onderzoek naar geschikte methodes inspireert.

7. VERLIES AAN MENSELIJKE KENNIS. Het onmiskenbare enthousiasme voor AI geeft soms de indruk dat menselijke intelligentie niet langer nodig is en heeft er zelfs al toe geleid dat bepaalde bedrijven werknemers ontslaan en vervangen door AI-systemen. We zien dit als een vorm van kortzichtigheid en een ernstige fout. Alle AI-systemen hangen af van menselijke intelligentie. Kennisgebaseerde systemen modelleren kennis en inzicht van menselijke experts, datagebaseerde systemen berusten op data van menselijk gedrag. Hieruit volgt dat menselijke expertise steeds onderwezen, ontwikkeld en geoefend moeten worden. Bovendien is menselijke expertise in de meeste domeinen immens veel krachtiger dan artificiële intelligentie, vooral voor gevallen die niet in de voorbeelden of het representatieve bereik van een leersysteem zitten.

De vooruitgang in perceptie, leren en redeneren heeft geleid tot een enorme toename van AI-toepassingen. Vele daarvan bevatten een laag risico, omdat de gebruiker fouten relatief gemakkelijk kan opmerken en verbeteren. Voorbeelden zijn persoonlijke assistenten (zoals Siri, Alexa, GoogleNow), aanbevelingsmotoren op het web en de plaatsing van advertenties gebaseerd op profielen van de gebruiker. Andere opkomende AI-toepassingen bevatten hoge risico's, mogelijk met als gevolg financiële catastrofes, infrastructuurverstoringen of zelfs verlies van mensenlevens.

Voorbeelden van deze toepassingen zijn:

- Zelfrijdende auto's. Opkomende autonome voertuigen hebben het potentieel om het verlies aan mensenlevens in het verkeer drastisch te verminderen door hun belangrijkste oorzaak, menselijke fouten, weg te nemen. Beeldverwerking uitgerust met AI kan echter soms fouten maken die geen mens zou maken.

De vraag is hoe de samenleving met zulke fouten moet omgaan. Naast fouten spelen ethische vragen en moeilijke keuzes een belangrijke rol. Kunnen we accepteren dat er beslissingen gemaakt worden die mensenlevens beïnvloeden en die het resultaat zijn van het optimaliseren van een abstracte wiskundige functie? Hoe moet men een dergelijke optimalisatiefunctie definiëren?

- Robotica in operatietechnieken. De implementatie van robots die chirurgen helpen om delicate operaties uit te voeren met een perfect stabiele robothand leidt tot weinig controverse. Recente ontwikkelingen gaan echter in de richting van een (bijna) volledige automatisering van bepaalde chirurgische taken, zoals bij oogoperaties van LASIK, knieoperaties en sommige neurologische operaties). Het potentieel om het chirurgische resultaat te verbeteren is groot, maar fouten in perceptie, redenering en uitvoering kunnen mensenlevens in gevaar brengen.
- Geautomatiseerde aandelenhandel. AI-systemen werken veel sneller dan menselijke handelaars, waardoor het voor mensen onmogelijk is om elke transactie te controleren en fouten te voorkomen. Gevallen van marktinstabiliteit als gevolg van geautomatiseerde handel zijn al waargenomen. (Kirilenko et al, 2017).
- Autonome wapens. Laat men autonome AI-systemen toe om oorlogsoperaties uit te voeren? Voorstanders van deze systemen beweren dat robots op het slagveld beter zullen presteren dan mensen bij het volgen van de Conventie van Genève (niet beïnvloed door emoties en stress) en dat ze mensen buiten de vuurlijn kunnen houden, in oorlogen die anders onvermijdelijk zijn. De tegenstanders waarschuwen voor risico's, zoals het op gang brengen van een wereldwijde wapenwedloop. Verder kan de technologie in verkeerde handen vallen. Ook cyberaanvallen of systeemfouten zijn mogelijk en leiden tot het aanvallen van verkeerde doelen. Een open brief van AI- en robotica-onderzoekers die pleiten voor een verbod op offensieve autonome wapens die buiten zinvolle menselijke controle gaan²⁰, is ondertekend door bijna 3500 AI-onderzoekers en door ongeveer 19.000 anderen. Ook in België is er een parlementair initiatief.

Het antwoord op deze uitdagingen ligt in de ontwikkeling van robuuste AI-systemen. Dat kan door in te zetten op de volgende aspecten:

1. Omgaan met fouten in de modellen van de wereld vereist onderzoek naar de ontwikkeling van optimalisatietechnieken die onzekerheden over de waarden van de parameters van de beperkingen toelaten en die voorkomen dat het leersysteem te sterk past op de trainingsgegevens en dus onvoldoende generaliseert. Daarnaast is er het vermijden van risicovolle doelstellingen. Het leersysteem moet zich niet alleen richten op de meest voorkomende gevallen, maar ook rekening houden met de slechtst mogelijke uitkomsten, zelfs al hebben die een kleine waarschijnlijkheid.

²⁰ Autonomous Weapons: An Open Letter from AI & Robotics Researchers. *Announced at the opening of the IJCAI 2015 conference.* <https://futureoflife.org/open-letter-autonomous-weapons/>.

2. Vooral het omgaan met niet-gemodelleerde aspecten van de wereld is een grote uitdaging. Een leermodel kan falen wanneer de samenstelling (distributie) van de data tijdens het leren verschilt van die van de data waarop het model moet worden toegepast. Daarom mag men niet vertrouwen op één enkele techniek of algoritme, hoe verfijnd ook, maar moet men veeleer een combinatie van methoden toestaan en additioneel concepten en eigenschappen leren vanuit het World Wide Web of andere kennisbronnen, om zo de kennisbank automatisch aan te vullen.

3. Een laatste uitdaging is het omgaan met verkeerdelijk gespecificeerde doelen. Een AI-systeem zou in staat moeten zijn om de werkelijke behoefte van de gebruiker te begrijpen en die te toetsen aan verschillende aangeleerde normen, gevolgen voor het milieu enz., in plaats van iets letterlijk uit te voeren. We zijn geneigd onze werkelijke doelstellingen verkeerd uit te drukken, wat kan leiden tot onverwachte implementaties en/of ongewenste gevolgen. Twee voorbeelden die in dit verband vaak worden gegeven zijn: 'Ga zo snel mogelijk naar de luchthaven' (moet het systeem met dat doel alle verkeersregels overtreden en iedereen in gevaar brengen?) Of: 'Zo veel mogelijk paperclips maken' (we zouden niet willen dat het systeem alles in paperclips probeert te veranderen, zoals bij koning Midas alles wat hij aanraakte in goud veranderde).

We hebben hier een aantal cruciale benaderingen vanuit AI, andere disciplines en de wetgeving beschreven die het bewaken van transparantie, interpreteerbaarheid, verantwoordelijkheid, controle en de bescherming van rechten als doel hebben. Een wettelijk kader en voldoende onderzoek dat deze doelen realiseert kunnen de Vlaamse en Europese AI-industrie een competitief voordeel geven.

VI. Slotbeschouwing: verwachtingen en limieten

Is AI het fundament van de Vierde Industriële Revolutie die op dit moment losbarst? Wellicht is niet alleen AI daarvoor verantwoordelijk, maar ook het Internet van Dingen, Big Data, Cloud Services en fundamentele sociale veranderingen in de maatschappij, zoals de voortschrijdende globalisering. Toch is ook duidelijk dat van AI een zeer belangrijke rol verwacht wordt in vrijwel alle domeinen van de menselijke activiteit waar informatisering nu al ingang heeft gevonden. De verwachtingen zijn hooggespannen en dat wordt nog versterkt door uitspraken van goeroes in management consultancy, pr-departementen van bedrijven met een belangrijke activiteit in AI, aandachtzoekers allerhande en een deel van de pers dat liever straffe verhalen brengt over de komende superintelligente robot dan zorgvuldig te rapporteren over wat al kan en niet kan. Een gezonde dosis scepticisme is nodig. Het zou een mooie zaak zijn als de stem van experts ter zake, die duidelijk de limieten van de huidige AI aangeven, meer gehoor krijgt (zie bijvoorbeeld Brooks, 2017).

Daarnaast is er in het AI-onderzoek ook ontegensprekelijk vooruitgang die een positieve impact kan hebben op onze samenleving. Bovendien kunnen we

bedrijven alleen maar aanraden grondig na te gaan wat ze met AI kunnen doen om competitief te blijven op wereldniveau. Sommige zijn daar al volop mee bezig, maar het besef van de opportuniteiten die AI biedt is nog lang niet ver genoeg doorgedrongen in onze maatschappij. Zeer dikwijls zal een oplossing waarbij artificiële en natuurlijke menselijke intelligentie constructief samenwerken, de meest performante blijken te zijn.

Daarnaast zien we dat onze belangrijkste economische partners strategische plannen hebben om het onderzoek, de ontwikkeling en de verspreiding van AI in de maatschappij te stimuleren: door de uitbouw van het onderwijs, het voorzien in investeringskapitaal, het stimuleren van een ecosystemen waarin alle actoren samenwerken enz. Het is hoog tijd dat er ook in Vlaanderen over soortgelijke acties wordt nagedacht. Met name in het hoger onderwijs is er een grote kloof gegroeid met de ons omringende landen, vooral met Nederland. Het onderwijs is de eerste stap om de keten van AI-innovatie in gang te zetten.

Hoewel onze bezorgdheid over AI-toepassingen vooral de lange termijn betreft, bestaat er nu al een zeer reëel en substantieel risico dat door de AI-ontwikkeling een aanzienlijk aantal banen geautomatiseerd zal worden. Deze trend geldt ook voor hoogopgeleide professionals, van boekhouders tot juridisch adviseurs (Duboshi en Sappin, 2017). Wij dringen er dan ook op aan te streven naar een partnerschap van mensen en AI-systemen om de productie en de diensten te verbeteren, en tegelijk een toename van de werkloosheid te voorkomen. Er moeten ook structurele investeringen in AI-onderwijs en -onderzoek komen. Het bouwen van robuuste, verklaarbare, betrouwbare en verantwoorde AI-systemen vereist een dieper inzicht en meer diepgaande kennis bij AI-ontwikkelaars, wat betekent dat er serieuze inspanningen geleverd moeten worden om in het onderwijs en onderzoek naar uitmuntendheid te streven. Een partnerschap van mensen en AI-systemen vereist ook dat elementen van AI in de vroege stadia van ons onderwijs geïntroduceerd worden.

Referenties

- Berbers, Y., Hildebrandt, M., Vandewalle, J. et al. (2017) *Privacy in tijden van internet, sociale netwerken en big data*. KVAB-Standpunt 49. http://www.kvab.be/sites/default/rest/blobs/1316/tw_privacy.pdf
- Biegel, B. en J. Kurose (2016) *The National Artificial Intelligence Research and Development Strategic Plan*. National Science and Technology Council Committee on Technology. www.nitrd.gov/PUBS/national_ai_rd_strategic_plan.pdf
- Biran, O. en Cotton, C. (2017) *Explanation and Justification in Machine Learning: A Survey*. Proc. IJCAI 2017 Workshop on Explainable Artificial Intelligence (XAI) (pp. 8-13). www.intelligentrobots.org/files/IJCAI2017/IJCAI-17_XAI_WS_Proceedings.pdf
- Brooks, R. (2017) *The seven deadly sins of AI predictions*. MIT Technology review. October 2017. www.technologyreview.com/s/609048/the-seven-deadly-sins-of-ai-predictions/
- Doshi-Velez, F. en M. Kortz (2017) *Accountability of AI under the law: the role of explanation*. Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein Center for Internet & Society working paper. <https://dash.harvard.edu/handle/1/34372584>
- Dubhashi, D. en S. Lappin (2017) *AI Dangers: Imagined and Real*. Communications of the ACM, 60(2):43-45.
- Economist (2016) *Artificial Intelligence in the Real World*. The Economist Intelligence Unit, London. https://www.eiuperspectives.economist.com/sites/default/files/Artificial_intelligence_in_the_real_world_1.pdf
- EPC (2016) Regulation (EU) 2016/679 of the European Parliament and of the council of 27 April 2016. Office Journal of the European Union, 119/1.
- Esteva, A. et al (2017) *Dermatologist-level classification of skin cancer with deep neural networks*. Nature, Feb 2017.
- Fast (2017) *How Amazon, Google, Microsoft, And IBM Sell AI As A Service*. Fast Company newsletter. <https://www.fastcompany.com/40474593/how-amazon-google-microsoft-and-ibm-sell-ai-as-a-service>
- Goodman, B. en S. Flaxman (2016) *European union regulations on algorithmic decision-making and a "right to explanation"*. <https://arxiv.org/abs/1606.08813>
- Grosz, B. (2016) *Artificial Intelligence and Life in 2030. One Hundred Year Study on Artificial Intelligence*. Report of the 2015 study panel. Stanford University. September 2016. <https://ai100.stanford.edu/>
- Holdren, J. en M. Smith (2016) *Preparing for the future of AI. Executive Office of the President*. National Science and Technology Council Committee on Technology. https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

Kirilenko, A. et al. (2017) *The flash crash: high-frequency trading in an electronic market*. The Journal of Finance, 72(3):967-998, 2017.

Knight, W. (2017) *Paying with your face*. MIT Technology Review 10 Breakthrough Technologies 2017.

Knight, W. (2017b) *China's Awakening*. MIT Technology Review. November 2017

Lecun, Y., Y. Bengio en G. Hinton (2015) *Deep Learning*. Nature 512(7553):436-444.

McKinsey, Inc. (2017) *Smartening up with Artificial Intelligence (AI) - What's in it for Germany and its industrial sector*. McKinsey & Co. Inc.

Nordrum, A. (2016) *Can deep learning help clinicians predict Alzheimer's disease?* IEEE Spectrum, Posted 15 June 2016.

Quoc V. Le, M Ranzato, R. Monga, M. Devin, K. Chen, G. Corrado, J. Dean, A. Ng (2011) *Building high-level features using large scale unsupervised learning*. <https://arxiv.org/abs/1112.6209>

Russakovsky, O. et al. (2015) *Imagenet large scale visual recognition challenge*. International Journal of Computer Vision, 115(3): 211-252.

Sample, I. (2017) *Big tech firms' AI hiring frenzy leads to brain drain at UK universities*. The Guardian 2 nov 2017. <https://www.theguardian.com/science/2017/nov/02/big-tech-firms-google-ai-hiring-frenzy-brain-drain-uk-universities>

Scudellari, M. (2017) *AI predicts autism from infant brain scans*. IEEE Spectrum, 15 Feb 2017.

Selbst, A.D. en J. Powles (2017) *Meaningful information and the right to explanation*. International Data Privacy Law, 7(4).

Shademan et al. (2016) *Supervised autonomous robotic soft tissue surgery*. Science Translational Medicine, May 2016.

Silver, D. et al. (2017) *Mastering the game of Go without human knowledge*. Nature 550, 354-359.

Steels, L. (2015) (ed.) *Learning about music in the age of MOOCS*. John Benjamins Pub. Amsterdam.

Strickland, E. (2017) *IBM Watson makes a treatment plan for brain-cancer patient in 10 minutes; doctors take 160 hours*. IEEE Spectrum, Posted 11 Aug 2017.

Tanghe, J. en J. De Bruyne (2017) *Aansprakelijkheid voor schade veroorzaakt door autonome motorvoertuigen*. Rechtskundig Weekblad (25). pp. 916-986.

Tractica (2017) *Artificial Intelligence: 10 Trends to Watch in 2017 and Beyond*. <https://www.tractica.com/resources/white-papers/artificial-intelligence-10-trends-to-watch-in-2017-and-beyond/>

Van Brussel H., J. De Schutter et al. (2016) *Naar een inclusieve Robotsamenleving: Robotisering, automatisering en werkgelegenheid*. KVAB-Standpunt 46 www.kvab.be/standpunten/naar-een-inclusieve-robotsamenleving

Samenstelling van de werkgroep

Bettina BERENDT (KU Leuven)

Marc DE MEY (UGent, KMW)

Danny DE SCHREYE (KU Leuven)

Charles HIRSCH (VUB, KNW)

Jan KRETZSCHMAR (UAntwerpen, KTW)

Aleksandra PIZURICA (UGent, KTW)

Luc STEELS (VUB, KNW)

Joos VANDEWALLE (KU Leuven, KTW)

Dirk VAN DYCK (UAntwerpen, KNW)

Johan WAGEMANS (KU Leuven, KMW)

KNW = Klasse Natuurwetenschappen

KMW = Klasse Menswetenschappen

KTW = Klasse Technische Wetenschappen

ARTIFICIËLE INTELLIGENTIE

NAAR EEN VIERDE INDUSTRIËLE REVOLUTIE?

Artificiële Intelligentie (AI) zit wereldwijd in de lift. Dankzij betere algoritmen, meer computerkracht en meer data, is het nu mogelijk om allerlei toepassingen te bouwen in heel wat domeinen van menselijke activiteit: van geneeskunde tot administratie, van autonome wagens tot intelligente zoekmachines voor het web. Hierdoor is een soort goudkoorts ontstaan, vooral in Amerika en China, waarbij bestaande ondernemingen, maar ook nieuwe start-ups, op grote schaal investeren. Er ontstaat een nieuwe industrie met mogelijk een grote maatschappelijke impact. Dit KVAB Standpunt legt uit wat AI is, wat er allemaal mee mogelijk is, welke ethische vragen het oproept, en hoe diverse landen op deze nieuwe technologie inspikken. Het geeft ook een aantal aanbevelingen zodat ook onze regio mee kan profiteren van deze ontwikkeling.

De reeks Standpunten van de Academie is een bijdrage tot het wetenschappelijk onderbouwd debat over actuele maatschappelijke en artistieke thema's. De auteurs, leden en werkgroepen van de Academie schrijven in eigen naam, onafhankelijk en met volledige intellectuele vrijheid. De goedkeuring voor publicatie door een of meerdere Klassen van de Academie waarborgt de kwaliteit van de gepubliceerde studies