

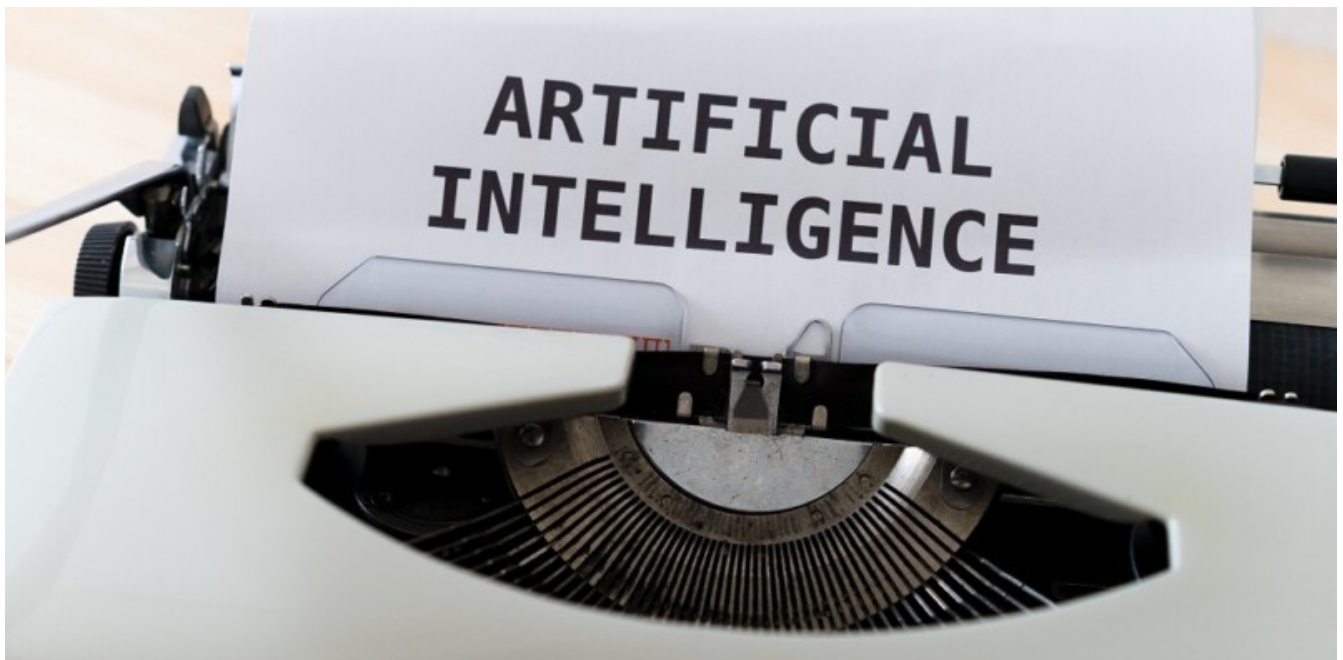
ARTIKELN | 31 JULI 2023

BEGROET 'HET NIEUWE AI' ONS MET EEN JUDASKUS?

Auteur: B.ICT Etienne Poeder RE CISA CDPSE

Beeld: Joshua Woroniec - Markus Winkler - Pavel Danilyuk

Leestijd: 8 min



Het begrip 'Judaskus' kennen velen uit de Bijbel. De apostel Judas gaf schijnbaar als blijk van goede intenties zijn meester een kus. In werkelijkheid was het een signaal voor de joodse leiders om Jezus gevangen te nemen op een kwetsbaar moment. Wij staan aan de vooravond van een nieuwe generatie artificiële-intelligentietoepassingen die resourcetechnisch en economisch als een geschenk uit de hemel lijken te komen.

Toepassingen van artificial intelligence (AI) kunnen leiden tot nieuwe banen en rollen, grotere afzetmarkten en hogere winstmarges. Hiermee vergaren wij 'capital', de som van assets. Maar hoe zit het dan met human capital? Zal AI met de goede intentie binnenkomen om uiteindelijk de feilbare mens voorbij te streven en deze overbodig te maken?

Wat is AI?

Hoewel AI vele hoedanigheden kent in de psychologie, software engineering, programmeertalen, wiskunde en algoritmen, is AI het makkelijkst samen te vatten als het resultaat van de zoektocht naar ultieme intelligentie met de mens als meest geschikte 'leermeester'. De mens lijkt anderzijds ook uiterst gefascineerd te zijn door AI als 'onze protegé'. Dit blijkt ook uit de emerging trends van Gartner. Sinds 2016 is een opmars in AI-toepassingen zichtbaar: natural-language question

answering, AI for Marketing, augmented intelligence, explainable AI, emotion AI, autonomous driving level 4, Edge AI, AI PaaS, AI Augmented Design, AI-driven Innovation en Generative AI.

Waar komt de angst voor AI vandaan?

Een voormalig Apple- en een Google-ontwikkelaar voor AI-toepassingen voorspelden dat binnen twaalf jaar 40% van de werkende bevolking zijn baan zal verliezen aan AI-toegepaste technologieën. We zien steeds vaker dat zowel ambachtelijke als intellectuele taken moeiteloos uitgevoerd kunnen worden met toegepaste AI-technologieën. Het 'gewone volk' met zowel ambachtelijke als intellectuele competenties, lijkt hiermee in de schaduw komen te staan van onze vermeende betere digitale versies. Hebben wij hier te maken met de voortekenen van een nieuwe vorm van eugenetica, waarbij de huidige mensheid het onderspit delft aan een 'boze' autonome supercomputer? Niet gehinderd door alle gepubliceerde ontwikkelingen zie ik het nog niet zo'n vaart lopen bij de toepassing van AI ter verbetering van menselijke competenties.

Waar AI nagenoeg onuitputtelijk informatie kan verwerken, loopt de menselijke intelligentie en performance veelal aan tegen fysieke grenzen

Hoe 'dreigend' is AI voor human capital?

Toegegeven, het helpt niet dat ik bij de drogisterij bij het afrekenen te woord word gestaan door een foto van een vermeende caissière met ingesproken tekst. Bij de supermarkt is menselijke interventie inmiddels afgeschaald tot een eerstelijnscontrole door een goedlachse supermarktconducteur die aselectief kritische deelwaarnemingen uitvoert op afwijkingen tussen de inhoud van mijn boodschappentas en de gescande boodschappen. Verder weg – maar eng dichtbij – zijn de hogere generaties ziekenhuissystemen die medisch specialisten ondersteunen in hun besluitvorming en door middel van micro/nano-robots of exo-skeletons bij chirurgische ingrepen ons opereren.

Supersnelle verwerking

Een artikel in *Harvard Business Review* ontzenuwt voor ons de prangende vraag of AI menselijke intelligentie, de essentie van human capital, zal vervangen.¹ Waar AI met een constante stroomvoorziening en rekenkracht supersnel en nagenoeg onuitputtelijk informatie kan verwerken, loopt de menselijke intelligentie en performance veelal aan tegen fysieke grenzen zoals beperkte concentratie, de invloed van emotie en vermoeidheid. Dit kan leiden tot afwijkende verwerking van informatie. Door algoritmen te blijven aanpassen en voeden met nauwkeuriger modellen die menselijk gedrag nog beter kunnen nabootsen, lijkt de mens op den duur onvermijdelijk vervangbaar. Echter, is deze conclusie wel juist?

Valkuilen bij vergelijken van AI-toepassingen met de mens

Er is een wezenlijk verschil in vergelijkbaarheid van de resultaten tussen de mens en AI. Als AI onbedoeld afwijkt van een geprogrammeerde instructie is het altijd een fout. Voor trouwe science fiction-fans is dit een overtreding van het tweede gebod van Isaac Asimov's klassieker 'Runaround'. Namelijk: een robot moet commando's gehoorzamen, behalve wanneer dit een bedreiging is voor de mens. De programmeur past dan simpelweg de fout in de code aan, totdat de toegepaste AI de gewenste output (of gewenst gedrag) vertoont.

Anders is het als een mens in een voorgeschreven instructie afwijkt van een verwerking van informatie. Dit hoeft namelijk geen fout te zijn. Menselijke afwijkingen kunnen het gevolg zijn van beïnvloeding door emoties, intuïtie, improvisatie of culturele sensitiviteit. In dit geval betreft dit dus een weloverwogen wilsbesluit op basis van aspecten die wezenlijk hebben bijgedragen aan de keuze en handelingen die hieruit voortvloeien.



De mens heeft een vrije wil en dit is precies de kracht en de blauwdruk van de mens. Een computer heeft een voorgeprogrammeerde instructie. Zelfs het random number generator (RNG) dat ten doel heeft 'random' waarden te genereren, volgt een voorgeschreven instructie en is dus wezenlijk anders dan een mens die random getallen kiest. Het personificeren van AI-toepassingen omdat het leeuwendeel van de trainingsdata gebaseerd is op menselijk handelen en redeneren, leidt hiermee tot kromme analogieën tussen de functionaliteit van AI en de competenties van de mens.

Onderlinge afhankelijkheid de mens en AI?

Het vermogen van de mens om af te wijken op basis van nieuw geïntroduceerde variabelen is niet per se onverenigbaar met AI. De Turingtest – een bekende test om vast te stellen of AI 'goed genoeg' is – is in feite niet meer dan een test waarmee wordt vastgesteld hoe goed AI zich kan presenteren als zijnde een mens zonder dat een menselijke gesprekspartner dit merkt. AI wordt hiermee getraind om beter te presteren met de mens als norm. Het doel is hier niet om aan te tonen dat AI beter is. AI kan de mens juist heel goed aanvullen. Men noemt dit ook wel 'augmented intelligence'. De mens verbetert hiermee zijn vermogen door (deels) te steunen op AI bij de verwerking van input om tot betere besluitvorming en uitvoering van taken te komen.

Hedendaagse toepassingen

We zien dit ook in de hedendaagse toepassingen, zoals de nieuwe generatie ziekenhuisinformatiesystemen die helpt bij diagnostiek en besluitvorming inzake behandelplannen, het remsysteem van auto's dat inschat wanneer er autonoom ingegrepen moet worden, de 'Lane & Park'-assist die helpt bij het inparkeren, en gepaste content voorgeschoteld krijgen op basis van onze zoekprofielen. Het gaat hier om het verbeteren van het menselijk vermogen dat ontstaat in een samenspel tussen de mens en AI. Het betreft hier dus niet een mens die in haar bestaansrecht bedreigd wordt door een 'slimmer' systeem.

Het echte gevaar zit in het onkritisch en gedwee steunen op dergelijke AI-toepassingen. Een zorgwekkend voorbeeld is een autobestuurder die al slapend 110 kilometer per uur reed op een snelweg met behulp van de autopilotfunctie. Een tragisch voorbeeld is een Belgische jongeman die na zes weken intensief chatten met een geavanceerde chatbot geadviseerd werd zelfmoord te plegen. Helaas volgde hij 'het advies' van de chatbot op en dat is hem noodlottig geworden.

Er is en blijft voor elke AI-toepassing een functionele noodzaak en rol voor de mens

Wie is er verantwoordelijk als het misgaat?

Het is onvermijdelijk dat AI-toepassingen met een rap tempo een functionele plaats verwerven in onze maatschappij en dat dit zelfs een onmisbare plaats kan worden. Echter, er is en blijft voor elke AI-toepassing een functionele noodzaak en rol voor de mens. Allereerst in de ontwikkeling van nauwkeurigere modellen en nieuwe variabelen zodat de AI-toepassingen beter worden. Anderzijds zie ik de mens als het geweten van AI-toepassingen. Iemand die aan de bel trekt omdat een AI-toepassing mogelijk leidt tot ethische bezwaren op langere termijn voor mens en maatschappij. AI heeft simpelweg geen geweten, geen bewustzijn en is niet in staat aansprakelijkheid en schuld te onderkennen als de uitkomst onwenselijk is.

Onwenselijke situatie

Wat als een AI-toepassing volgens de actuele trainingsdata wel de instructie opvolgt, maar deze instructie tot een onwenselijke situatie leidt? Hieruit blijkt dat de mens cruciaal blijft als eindverantwoordelijke. De AI-toepassing dient, uiteraard herleidbaar, onderhevig te blijven aan een 'management override'. Indien een fout getraceerd kan worden is de mens de eindverantwoordelijke, zelfs als een AI-toepassing de fout traceert en voorstelt hoe het op te lossen. Je kunt hierbij denken aan 'predictive maintenance' (wanneer is het beste moment voor nieuw onderhoud?) of 'price forecasting' (voorspellingen prijzen van ruwe materialen). Een mens moet uit hoofde van zijn functie en/of rol altijd en in elke fase een instrument hebben om autonoom te kunnen ingrijpen als AI-toepassingen in een besluit of taak lijken af te wijken van de beoogde menselijke doelstelling.

Levensgevaarlijk

Ingrijpen door een mens is nog prangender wanneer een perfect uitgevoerde instructie levensgevaarlijk is geworden. Twee tragische voorbeelden betreffen beide een assemblagemachine in de automotive-industrie. In beide gevallen werden werknemers onbedoeld geïdentificeerd en 'verwerkt' als fysieke objecten in het assemblageproces en hebben dit niet overleefd' (Volkswagenfabriek in 2015 en een onderleverancierfabrikant van Hyundai en KIA in 2016).



Hoewel de factor mens steevast als zwakste schakel wordt geportretteerd, blijft de rode draad de mens. Niet gehinderd door enige vorm van perfectie en onfeilbaarheid zijn mensen onvoorspelbaar en dat is nu juist de creatieve kracht die de mens doet ontwikkelen en de feilbaarheid doet herkennen in mensen en objecten. Dit vind ik een mooie, niet te programmeren en kostbare eigenschap. Hiernaast is zelfs aan te merken dat de creatieve 'kracht' van AI-toepassingen zoals Chat-GPT, DALL-E en Midjourney welliswaar in staat zijn om te 'creëren', maar in feite genereren deze in een notedop via 'reverse engineering' mutaties op bestaande werken. Toegegeven, plagiaat maakt AI-toepassingen dan wel weer heel menselijk.

Hebben human capital en AI een gedeelde toekomst?

Hoe het ook zij, wij kunnen er niet meer omheen. Creaties die ooit voorbehouden waren aan creatieve mensen moeten rekening houden met AI-toepassingen die zich op elk creatief gebied begeven. AI zal voorlopig door middel van een paar prompts in luttele seconden indruk blijven maken op mensen met essays, artikelen, tekeningen, foto's, video's, schilderijen en examenresultaten.

Is er dan nog wel plaats voor de creatieve maar feilbare mens? Hoe lang dulden de onfeilbare AI-algoritmen ons nog? Het antwoord van deze existentiële vragen dragen wij in onszelf. Wij zijn de norm om onze creaties te toetsen. Het is de mens die AI heeft ontwikkeld en wij zullen dit blijven doen en niet andersom. Wij moeten investeren in een collectief besef dat het wezen van de mens niet in 'nullen' of 'enen' is te bevatten. Wij kunnen en hoeven daarom ook niet elk probleem op te lossen met snellere, efficiëntere en effectievere geautomatiseerde volbrenging van taken.

Anderzijds hoeven wij ook niet alles zelf te doen als het beter kan. Laten wij niet vergeten dat het controleren van lijstwerk door accountants kortgeleden nog standaardwerk was en niet per se een alternatief. Dat er tegenwoordig met minder accountants hetzelfde werk gedaan kan worden, omdat data-analyses en vernuftige query's op veel data in een oogwenk bijdragen aan nauwkeurigere controles op maatregelen, is een welkom geschenk en geen bedreiging. Het heeft mijns inziens het vakgebied van auditing veel interessanter gemaakt, en is geen bedreiging voor het bestaansrecht van de accountant en interne IT-auditor.

Door AI kunnen op korte termijn het formuleren van bevindingen en het vervaardigen van auditrapportages geautomatiseerd worden

Data-analyse

Als IT-auditor zie ik de ontwikkeling van AI-mogelijkheden ook rooskleurig in als het gaat om data-analyse (100% populatiecontroles), process mining (compliance en performance) en AI-powered text processing. Deze technieken zullen eraan bijdragen dat auditors efficiënter, effectiever, gestructureerder en sneller bewijslast kunnen verzamelen en afwijkingen kunnen identificeren. Door AI kunnen op korte termijn het formuleren van bevindingen en het vervaardigen van auditrapportages geautomatiseerd worden.

Echter, er zijn tal van taken die niet weggelegd zijn voor AI-toepassingen. Denk hierbij aan het afnemen van interviews, waarbij non-verbale communicatie in zowel zenden als ontvangen van cruciaal belang is. Het doorvragen bij de auditee als het 'juiste antwoord' wel gegeven is, maar 'een onderbuikgevoel' ons juist dwingt om vragen te blijven stellen. Het beoordelen van compenserende maatregelen die niet eerder de revue gepasseerd zijn, lees: niet aangeboden zijn als trainingsdata aan de AI-toepassing.

Blijven afvragen

Achter de schermen moeten wij ons afvragen of wij het beheer van AI-algoritmen willen toevertrouwen aan een AI-toepassing? Hoe worden onethische overwegingen geïdentificeerd en geweerd uit de code? Hoe zit het met eigenaarschap, inzichtelijkheid, toerekeningsvatbaarheid en traceerbaarheid van alle data en verwerking die hebben

AUDIT

MAGAZINE

geleid tot een zekere besluitvorming ('straight through processing')? Ik heb mij in het 'BIV-AO-klasje' vrijwillig doch stelselmatig laten indoctrineren met het credo: 'zelfcontrole is geen controle!' Ik verstout mij dan ook in de veronderstelling dat wij de voorgaande problematiek niet willen oplossen met een nieuwe AI-toepassing.

Hoewel het indruist in de methodologische sferen van auditors omdat een norm bij voorkeur niet feilbaar mag zijn, mag wat mij betreft de feilbare mens als norm een creatieve uitzondering vormen. Wij zijn als feilbare mens wel degelijk in staat gebleken om vanuit een collectief geweten AI-toepassingen te ontwikkelen, te verbeteren en te toetsen of het van toegevoegde waarde is om onze auditwerkzaamheden beter te kunnen uitvoeren. Wellicht kunnen wij dan toch inspiratie putten uit het tweede gebod van het Nieuwe Testament van de Bijbel: je naaste liefhebben als jezelf. Naast de huidige inzet op AI-toepassingen is blijvend investeren in human capital als de 'onzichtbare leermeester' van AI-toepassingen, een veilige investering en verzekering voor zowel de toekomst van betere AI-toepassingen als human capital.

Noot

1. AI should augment human intelligence, not replace it (hbr.org)

Over

Etiënne Poeder is zelfstandig IT-auditor. Daarvoor bekleedde hij onder andere audit- en adviesfuncties bij BDO en namens Deloitte in onder andere de private sector en (semi)overheidsinstanties.

Tags: Cyber resilience

Bron url: <https://auditmagazine.nl/artikelen/begroet-het-nieuwe-ai-ons-met-een-judaskus/>