Vertrouwen in een algoritmiserende samenleving

Hoog tijd om algorithm assurance op te pakken

17 april 2019

Mona de Boer

Stel dat u onverhoopt een keer de huisartsenpost moet bellen met een medisch probleem dat niet kan wachten tot uw eigen huisarts weer beschikbaar is. Normaal gesproken krijgt u een speciaal daartoe opgeleide deskundige (triagist) aan de lijn, die aan de hand van een gestructureerde Q&A in korte tijd de urgentie van uw situatie bepaalt en beoordeelt of u direct gezien moet worden. Stelt u zich nu eens voor dat u aan de andere kant van de lijn een triage-algoritme treft in plaats van een mens. Het algoritme analyseert met *Voice- en Speech Recognition* en *Natural Language Understanding* wat er gezegd wordt, en ook de intonatie van stemmen, en de omgevingsgeluiden. Binnen een minuut weet u of u richting het ziekenhuis moet.

Science fiction? Allerminst. Begin 2018 maakte een Deense start-up bekend meldkamermedewerkers te gaan ondersteunen bij het diagnosticeren van een hartstilstand met behulp van een virtuele kunstmatig intelligente assistent. Bij een hartstilstand is tijd van levensbelang. Iedere extra minuut voordat het slachtoffer hulp krijgt in een ziekenhuis, vermindert de kans op overleven. De virtuele assistent analyseert het gesprek, ademhalingsgeluiden en achtergrondgeluiden die kunnen duiden op een hartaanval. De medewerker in de meldkamer krijgt in realtime waarschuwingen van het systeem en suggesties voor vervolgvragen en -acties. Met de kunstmatige intelligentie bleek de meldkamermedewerker tijdens een test in 95 procent van de gevallen de juiste diagnose te stellen. Zonder hulp van het systeem was dat 73 procent. De *European Emergency Number Association* (EENA) kondigde vorig jaar een partnership met de Deense start-up aan.

Een ander voorbeeld van algoritmisering in het medische domein komt van Google. Google haalde in de zomer van 2018 de krantenkoppen met de onthulling dat zij een kunstmatige intelligentie heeft ontwikkeld die – naar eigen zeggen – betrouwbaarder dan artsen de sterftekans van een patiënt kan voorspellen. Het Google-algoritme claimt in een mum van tijd een medisch dossier door te kunnen pluizen op belangrijke data die indicatief zijn voor, positief gesteld, de overlevingskans. Waar of niet? Dat zal nog moeten blijken. Maar de realiteit is dat dergelijke algoritmes in ontwikkeling zijn en dat ze nauwkeuriger worden naarmate ze worden blootgesteld aan meer van de juiste data. Hoewel het algoritme qua nauwkeurigheid ogenschijnlijk te wensen overlaat, bleek het

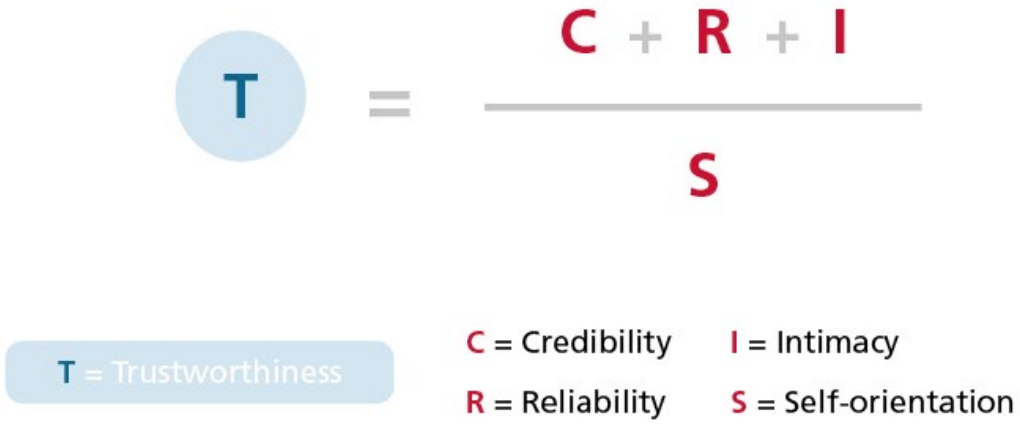
twee keer zo accuraat als de inschatting van de artsen. Naar verwachting zal dit nog verder toenemen naarmate het wordt blootgesteld aan meer data.

Niet alle voorbeelden van de algoritmisering van onze samenleving zijn zo wezenlijk van aard. Dat neemt niet weg dat ze dagelijks resulteren in belangrijke beslissingen over ons welzijn. Hoe zit het bijvoorbeeld met algoritmes die bepalen wie welke baan krijgt, wie welke woning krijgt, wat de prijzen zijn van onze basisbehoeften? Hoe weten we dat de beslissingen die ze ondersteunen of autonoom nemen te vertrouwen zijn en de spelregels van onze samenleving volgen? Welke eisen stellen we aan algoritmes en hoe gaan we de naleving daarvan toetsen?

Als IT-auditor, data-specialist, voorzitter van de nieuwe NOREA werkgroep Algorithm Assurance, tevens kritisch burger van deze samenleving, heb ik mij de afgelopen tijd uitvoerig over deze vragen gebogen. Hoe pak je zo'n groots vraagstuk aan? Principles based of rules based? Top-down of bottom-up? Waar begin je?

Waarom niet zoals leiderschapsgoeroe Steven Covey voorstelt 'with the end in mind'? Stel, u zou een moment uw ogen sluiten en aan het triage-algoritme denken. In het fictieve voorbeeld heeft u, verkerend in een kwetsbare positie, als patiënt de huisartsenpost gebeld en hangt u zojuist de telefoon op in het vertrouwen dat het algoritme de beste beslissing voor uw gezondheid heeft genomen. Kunt u op dit fictieve beeld duiden wat u dat vertrouwen geeft?

Vertrouwen is een ingewikkeld concept omdat het zoveel dimensies heeft. En daar waar we er in theorie vat op denken te hebben, gaat het vaak over vertrouwen in mensen, niet in systemen. In 2000 introduceerde David Maister in zijn boek 'The Trusted Advisor' het begrip '*Trust Equation*'.



Figuur 1: David Maister's Trust Equation (bron)

Dit algoritme voor vertrouwen modelleert de mate van vertrouwen die in een zakelijke relatie tussen mensen kan worden bereikt. Iemand die vertrouwen wil scheppen,

kan vooral geen genoeg krijgen van geloofwaardigheid (weten waar je het over hebt), betrouwbaarheid (afspraken nakomen) en vertrouwelijkheid (zorgvuldig omgaan met de informatie die je krijgt). Tegelijkertijd helpt het om de focus buiten jezelf te leggen (de belangen van de ander te behartigen in plaats van de eigen).

Of een model voor intermenselijk vertrouwen een-op-een te projecteren is op mens-machine-interactie, zal moeten blijken. In de raamwerken voor verantwoorde kunstmatige intelligentie die momenteel het licht zien, zie je parallellen in de eisen die we stellen, bijvoorbeeld:

Geloofwaardigheid: hoe presteert het model/algoritme binnen het domein waarop het getraind is?
- Betrouwbaarheid: presteert het model/algoritme op het moment dat het ertoe doet?
- Vertrouwelijkheid: welke gegevens verwerkt het model/algoritme en worden daarbij privacyregels nageleefd?
- Zelf-oriëntatie: bevat het model/algoritme (on)bedoelde *biases*, die bepaalde doelgroepen benadelen?

Tegelijkertijd wijst de praktijk nog vaak genoeg uit dat mensen en systemen – bij gelijke omstandigheden – verschillend worden beoordeeld. Onderzoek van Amerikaanse wetenschappers naar taken op het gebied van het maken van voorspellingen (forecasting), toont aan dat mensen de voorkeur geven aan menselijke forecasters, ook als zij zien dat forecasting-algoritmes even goed of beter presteren. [DIET15] Dit fenomeen noemen de auteurs '*algorithm aversion*'. Deze aversie openbaart zich te meer als mensen algoritmes zien presteren, zelfs als ze voor eenzelfde taak beter presteren dan mensen. Waarom? Omdat mensen sneller het vertrouwen in algoritmes verliezen dan in mensen wanneer ze beiden dezelfde fouten hebben zien maken. Het is belangrijk de achterliggende oorzaken hiervan te begrijpen, omdat deze aversie in een rap algoritmiserende samenleving onnodige kosten met zich meebrengt.

Het is duidelijk dat we als samenleving voor een aantal uitdagende, maar buitengewoon interessante vraagstukken staan. Deze draaien om de verantwoorde inzet van kunstmatige intelligentie. Daarbij gaan we ook een hoop leren over onszelf. Want met het bepalen van wat verantwoord is voor intelligentie die de onze moet nabootsen, zeggen we ook iets over onszelf.

Hoe pak je zo'n groots vraagstuk aan? In ieder geval 'with the end in mind' en volgens Erica Dhawan, co-auteur van het boek 'Get Big Things Done' door gebruik te maken van 'connectional intelligence': multidisciplinair en door het mobiliseren van de 'wisdom of the crowd'. De werkgroep Algorithm Assurance pakt de handschoen graag op om beoordelingskaders, toetsingsnormen en onderzoeksmethoden voor IT-auditors te ontwikkelen, waarmee zij in de praktijk organisaties en hun belanghebbenden kunnen helpen vertrouwen te creëren in algoritme-gedreven beslissingen.

Heeft u ideeën of suggesties voor de werkgroep? Wij doen graag een beroep op u en uw gedachten (mona.de.boer@pwc.com).

Literatuur

[DIET15] Dietvorst, B., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144* (1), 114-126.

Drs. M (Mona) de Boer RE RA | Director Data Analytics bij PwC

Mona de Boer leidt binnen PwC de Algorithm Assurance praktijk, die zich richt op de validatie van geavanceerde (kunstmatige intelligentie gedreven) besluitvormingsmodellen en algoritmes. Mona is voorzitter van de Algorithm Assurance werkgroep van de NOREA, en als docent en promovenda verbonden aan de Universiteit van Amsterdam.