

Big data zijn grotendeels 'a big bubble'

Noelsspeak

Geert Noels

In Noelsspeak spreekt econoom en Econopolis-stichter Geert Noels zich maandelijks onomwonden uit. Vandaag heeft hij het over de valse beloftes van big data.

'Data zijn de nieuwe olie', titelde het Britse magazine The Economist in 2017. Olie was de belangrijkste grondstof en drijver van de oude economie, maar in de nieuwe economie zijn dat data. Net als olie moeten data worden geraffineerd om echt waarde te kunnen leveren. Maar voor de hoogtechnologische motoren van vandaag zijn veel data helaas onbruikbaar.

Charles Babbage, de Britse geleerde uit de 19de eeuw, wordt weleens de uitvinder van de computer genoemd. Hem werd gevraagd: 'Mister Babbage, als je de verkeerde cijfers in de machine steekt, komen er dan correcte antwoorden uit?' Uiteraard was het antwoord nee. Toen al wist men dat het antwoord van een machine maar zo slim kan zijn als de input die ze heeft gekregen.

Het GIGO-principe was geboren: Garbage In, Garbage Out. Software en besturingssystemen hangen af van de kwaliteit van de informatie waarmee mensen ze voeden. De term werd pas echt bekend toen computers doorbraken. Dat gebeurde veel later, met de komst van de mainframecomputer. Ene George Fuechsel, een IBM-programmeur, zou de term GIGO hebben gebruikt toen hij les gaf op de 305 RAMAC-computer voor klanten in New York.

Zestig jaar later is de technologie sterk geëvolueerd. De processoren zijn oneindig veel sneller dan in de tijd van Babbage en Fuechsel, en er is een massa aan nieuwe mogelijkheden geopend. Naast data zijn algoritmes belangrijker geworden. Daarom denken we vandaag dat computers erg slim zijn geworden. Machines zouden zelfs zelflerend kunnen zijn: ze worden slimmer door de resultaten van hun eigen spinsels te evalueren en het verwerkingsproces continu te verbeteren. De slimste computers hebben de mens geklopt in uitdagende spelletjes: schaken, Go en de aartsmoeilijke quiz Jeopardy.

Welke bewijzen willen we nog meer? Is de GIGO-wet nu verbroken? Kunnen we de computer voeden met junkdata en verwachten dat hij clever genoeg is om slimme antwoorden te genereren?

Helaas weten algoritmes niet of ze te maken hebben met kwalitatieve data of met rommel. Daarbij komt dat niet alleen in data fouten kunnen sluipen, maar ook in algoritmes. Elke gebruiker van een spreadsheet kent het gevaar. Een Excel is maar zo goed als de kwaliteit van de data en de formules die in het rekenblad werden ingevoerd.

Niets is zo gevaarlijk als een spreadsheet met foute formules, verkeerde links en enkele fouten in de basisgegevens. Het rekenblad schotelt je binnen de milliseconde een antwoord voor dat er accuraat uitziet maar tot het negende cijfer na de komma fout is berekend. Hoeveel schade zou Excel al hebben veroorzaakt in de bedrijfswereld? En hoeveel keer erkende de consultant of de financieel directeur vervolgens zijn fout?

Een spreadsheet is toegankelijk voor veel mensen, waardoor fouten nog af en toe worden ontdekt. Maar bij de big data en de zelflerende systemen van vandaag wordt dat steeds moeilijker. Hoe complexer een uitkomst oogt, hoe meer mensen ze blindelings willen vertrouwen. In de film 'War Games' uit 1983, met Matthew Broderick in de hoofdrol, wordt een nucleaire oorlog maar nipt vermeden omdat de doldraaiende supercomputer zijn eigen inconsistenties net op tijd inziet.

En wilt u nu eens wat weten? Onlangs opperde een Amerikaanse legergeneraal de nucleaire codes toe te vertrouwen aan een algoritme.

Hoog bullshitgehalte

Begrijp me niet verkeerd. Artificiële intelligentie, big data en allerlei 'slimme systemen' openen wel degelijk enorme mogelijkheden. Maar ik heb een hekel aan het geloof in de onfeilbaarheid en aan het hoge bullshitgehalte waarmee ze worden gepresenteerd. Want ze openen ook enorme mogelijkheden voor charlatans, om hun mooi verpakte rommel te verkopen als een intelligent en betrouwbaar systeem. In de big data zit nog altijd veel junk, waarin dataspecialisten lang moeten speuren naar bruikbare brokken. Ik zie veel 'slimme systemen' die me al snel vervelen met domme antwoorden, foute verbanden en goed ogende maar onverstandige aanbevelingen.

Data zijn dus niet de nieuwe olie. Ze zijn net als olie een grondstof die moet worden geraffineerd. Maar in tegenstelling tot olie zijn veel data geen smeermiddel van goede beslissingen. Integendeel, ze creëren smog die verblindt. Bedrijven die op een goudmijn denken te zitten en gouddelvers - excuseer, dataspecialisten - binnenhalen om in hun kelders vol historische data te gaan mijnen, zouden weleens klatergoud kunnen ontdekken

Big data zijn grotendeels 'a big bubble'. Net zoals de spreadsheet uw financiën niet gezond maakt, kunnen big data van een slecht ondernemingsmodel geen goed maken. 'Garbage In, Garbage Out' geldt nog altijd. Het is zelfs 'Big Garbage In, Gigantic Garbage Out'. Artificiële intelligentie en big data zijn de toekomst, maar ze kunnen nooit kritische en creatieve mensenhersen vervangen.

Geert Noels

- Econoom en medeoprichter van Econopolis