

Marian de Joode, Jan Kwint, Ward van Breda

# Selecteren en ontwikkelen met behulp van Artificial Intelligence

Onze wereld verandert, en onze organisaties veranderen in rap tempo mee. Dat hebben we als mensheid zelf veroorzaakt: door globalisering en technologische ontwikkelingen zijn er ongekennde mogelijkheden beschikbaar gekomen. Data, artificial intelligence, robotisering, gamificatie: we hebben een werkelijkheid gecreëerd die ook voor werving, selectie en ontwikkeling een totaal nieuw speelveld toegankelijk maakt. Technologie staat centraal, we maken de beweging van High Touch naar High Tech.



**Marian de Joode** (m.dejoode@ltp.nl)

is innovatiemanager bij assessmentbureau LTP Business Psychologen

**Jan Kwint** (j.kwint@ltp.nl)

is CEO van assessmentbureau LTP Business Psychologen

**Ward van Breda** (ward@sentimentics.com)

combineert de cognitieve psychologie met een carrière in de artificiële intelligentie en is betrokken bij het promotieonderzoek AI in het selectie-interview dat de VU samen met LTP uitvoert.

Mz | zomer 2020

## Selecteren en ontwikkelen op basis van Artificial Intelligence

Door het inzetten van algoritmes en het beschikbaar maken van data, kunnen selectie- en ontwikkelinstrumenten voor veel bredere vraagstellingen worden ingezet dan voorheen, ter ondersteuning van de gehele *employee journey*. Daarnaast zorgen technologische mogelijkheden voor vernieuwingen in de wijze van aanbieden van assessments, bijvoorbeeld door het verkorten van vragenlijsten door machine learning, of door het inzetten van gamificatie toepassingen. Dit vergroot een positieve *candidate experience*, een belangrijk element in het huidige selectieproces.

Bij het selecteren of het in kaart brengen van het ontwikkelpotentieel van een medewerker, hebben we decennialang assessmentonderdelen ingezet waarin heel weinig is veranderd: persoonlijkheid, capaciteiten en motieven van mensen werden en worden nog steeds met behulp van gevalideerde vragenlijsten en tests gemeten. Dat is op zich niet verwonderlijk: de klassieke persoonlijkheids-vragenlijsten en intelligentietests hebben zich bewezen als betrouwbare en valide instrumenten om een voorspelling te doen van individueel functioneren. De interpretatie van de uitkomsten van die instrumenten werd en wordt nog steeds vaak gedaan door een psycholoog/adviseur of HR-professional. Maar, met de toenemende digitalisering wordt *aandacht* een schaars goed. Dat is echter wel weer belangrijk voor de *candidate experience*: voor een deelnemer is het belangrijk zich gehoord en gezien te voelen en mensen vinden het moeilijk te accepteren als ze afgewezen worden door een formule.

Het inzetten van Artificial Intelligence ten behoeve van selectiebeslissingen ligt echter voor de hand, het gaat tenslotte om het toepassen van beslisregels. Algoritmes worden natuurlijk al jarenlang gebruikt in online assessments waarbij de rapportage automatisch wordt gegenereerd, zij vervangen de interpretatie van de psycholoog. Die algoritmes in online assessments zijn meestal nog gebaseerd op expertkennis die in systemen geprogrammeerd was: de *psycholoog in de machine*. Alles nog op basis van vragenlijsten en testen.

Voor de meeste vraagstellingen is dit een prima oplossing. Sterker nog, uit onderzoek blijkt dat deze *statistische methode* tot een be-

ter advies leidt dan de interpretatie door een adviseur, de *klinische methode*<sup>1</sup>. Met de statistische methode wordt bias vermeden en zijn de uitkomsten consistent en transparant. In het kader van het bevorderen van *fairness* en diversiteit is dit een vooruitgang.

We bevinden ons nu in de volgende fase: die waarin data met behulp van machine learning de basis vormen voor zelflerende assessmentssystemen. Dit biedt een heleboel nieuwe kansen, en roept voor ons professionals ook veel nieuwe vragen op: wie is er bijvoorbeeld verantwoordelijk als systemen beslissingen nemen? En wat voor data gebruiken we daar eigenlijk voor? Hoe kunnen we de algoritmes controleren? Wat is de impact op mensen, organisaties en de maatschappij?

## People Analytics

Waar assessment in het verleden een redelijk geïsoleerd meetmoment was, is het nu het startpunt van data-gestuurde HR. De data die verzameld worden, vormen het begin van de *employee journey*, ze geven inzicht in het potentieel en de ontwikkelrichting van de medewerker. Deze psychometrische data kunnen gecombineerd worden met real-time gegevens zoals performance gegevens, feedbackdata en gegevens over welbevinden. Deze gegevens kunnen bijvoorbeeld via een app verzameld worden.

Deze data zijn de basis voor People Analytics. Met behulp van dashboards kunnen real-time analyses gedaan worden: wat zijn bijvoorbeeld de meest bepalende factoren voor high performance op individueel en teamniveau? Dan kan de selectie daarop aangepast worden. Wie beschikken er in de organisatie over leiderschapspotentieel in het kader van onze talentmanagementprogramma's? Waar zitten in onze eigen organisatie de mensen die over de kwaliteiten beschikken die schaars worden, die we tot nu toe over het hoofd hebben gezien, of die onvoldoende kansen hebben gekregen? Hoe gaan we het meeste rendement uit ons opleidingsbudget halen?

Aangezien het potentieel en de veranderkracht van de mensen nog steeds de cruciale factor is voor het onderscheidende succes van een organisatie, is het inzetten van people analytics van strategisch belang. Mits op een verantwoorde wijze gerealiseerd, kan dit een essentiële nieuwe strategische rol voor MD en HR inluiden. Zie de recente inventarisatie onder CHRO's in het kader op de volgende pagina.



Een compacte inventarisatie onder een aantal CHRO's van grote bedrijven laat zien waar zij mogelijkheden zien voor het inzetten van data en algoritmes:

- Het inzetten van data en algoritmes ten behoeve van het **objectiever maken van recruitment en selectie**. Gevolg: betere selectiebeslissingen en meer diversiteit. Laat het algoritme maar (mee)beslissen!
- **Retentie**: houd je (goede) mensen langer vast, bijvoorbeeld door vaker te meten hoe het met ze gaat en data te gebruiken om arbeidssatisfactie te voorspellen en daarop in te grijpen.
- **Job rotatie**: gebruik data om te bepalen wie van de medewerkerspool ook andere werktaken kan vervullen, of in een ander team beter tot zijn/haar recht komt
- **Strategische personeelsplanning en Talent Development**. Laat AI op de langere termijn voorspellen wie de high potentials/leiders zijn van de toekomst zijn.
- **Samenstelling high performing teams**. Gebruik algoritmes om de meest effectieve samenstelling van een team te voorspellen.
- **Development, met name het kruisvlak van recruitment en L&D**: het in kaart brengen van leerafstand. AI kan suggesties doen over mensen en beroepen met een overbrugbare leerafstand tot een andere functie (Talent navigatie, up- en reskilling).
- Tenslotte: **acceptatie en bewustwording van de nieuwe werkelijkheid**. Het geleidelijk inzetten van robots die routinematig werk overnemen kan helpen om de organisatie op een positieve manier kennis te laten maken met nieuwe technologie zoals AI.

## Naast individuele ook team-assessments

De inhoud en manier van samenwerken op het werk is sterk aan verandering onderhevig, en daarmee ook de kwaliteiten die nodig zijn om succesvol te zijn. Gezien het feit dat we meer in teams samenwerken wordt het steeds interessanter om niet alleen naar de geschiktheid van een individuele medewerker te kijken. Met een team-assessment beoordelen we ook de beste samenstelling van mensen in het team als geheel, door de data van de beoogde teamleden met elkaar te combineren. Hiermee worden afbreukrisico's geïdentificeerd die kunnen ontstaan door negatieve teamdynamieken, of door een slechte chemie tussen leidinggevende en teamleden. Zo kan veel ellende voorkomen worden. Ook voor het snel samenstellen van wisselende projectteams is dit een uitkomst. Met behulp van een dashboard met assessmentdata van de teamleden kan de - ge-

zien de opdracht van het team - beste samenstelling worden gerealiseerd, door op basis van de combinatie van de afzonderlijke profielen van de teamleden de beste match te laten zien.

## Nieuwe kwaliteiten meten

Naast het onderzoek of de persoonlijkheden met elkaar matchen, zijn er andere kwaliteiten dan de klassieke competenties die het verschil gaan maken, nu en in de toekomst. Het gaat dan om eigenschappen als veranderkracht, empathie, sociale wendbaarheid, oplossingsgerichtheid, kunnen omgaan met *information overload*, moreel kompas; de factoren die bepalend zijn voor het bouwen van een wendbare organisatie. Deze kwaliteiten kunnen gedeeltelijk gemeten worden met behulp van klassieke assessmentmethoden, zoals vragenlijsten en capaciteitentests. Maar door technologische ontwikkelingen kunnen we nu ook wezenlijk nieuwe assessmentvormen hanteren, zoals gamificatie of het met algoritmes interpreteren van taalgebruik en emoties, en het meten van neurologische en fysiologische reacties, bijvoorbeeld hartslag. Het zijn allemaal disrupties in het selectievak die andere relevante menselijke eigenschappen meetbaar kunnen maken; vaak vormen ze ook een verbetering van de *candidate journey*, en voor HR en MD een bron van veel nieuwe relevante inzichten over de geschiktheid en het potentieel van medewerkers. Voorwaarde is wel dat deze nieuwe toepassingen steeds kritisch beoordeeld worden op hun wetenschappelijke kwaliteit en de beleving van de deelnemer, voordat ze worden ingezet voor de beoordeling van mensen. Dat is nu nog vaak onvoldoende gewaarborgd. Hierover later meer.

## Voorbeeld uit de wetenschap: AI in het selectie-interview

Reinout de Vries, hoogleraar aan de VU, onderschrijft dat de toepasbaarheid van AI-toepassingen in de selectiepsychologie nog in de kinderschoenen staat. Hij doet al enige jaren onderzoek naar het inzetten van AI bij het meten van persoonlijkheid. De eerste uitkomsten daarvan zijn veelbelovend, maar het onderzoek is complex. Op dit moment wordt bij de Vrije Universiteit in Amsterdam samen met LTP in een PhD-project gewerkt aan een op AI gebaseerd selectie-instrument. Centraal hierin staat het interview, nog steeds het meest gebruikte selectiemiddel. De kwaliteit van de voorspelling van werksucces op basis van het klassieke selectie-interview is vaak discutabel. Door hier AI op toe te passen kan wellicht een verbetering van de kwaliteit van de selectiebeslissing gerealiseerd worden. De bronnen die in dit onderzoek uit het interview worden gedestilleerd zijn: woordgebruik, stemgebruik en

gezichtsexpressies. Bij woordgebruik worden twee benaderingen in de algoritmes ingezet: open (data als lerend systeem) en gesloten (expert-model). Het verschil wordt hieronder weergegeven.

## Open vocabulary

- Bottom up
- “Free’ algorithm: Machine/statistical/deep/neural network learning (e.g. bag-of-words, N-grams, word2vec)
- Identificatie van (combinaties van) woorden die mensen gebruiken om hun persoonlijkheidskenmerken optimaal weer te geven

## Open vocabulary

- Top Down
- “Fixed’ algorithm: word counts or parameters/loadings from regression/factor analysis
- Gebruik van theoretisch / empirisch afgeleide (combinaties van) woorden die door mensen worden gebruikt om hun persoonlijkheidskenmerken optimaal weer te geven

Het verschil tussen beide benaderingen wordt goed duidelijk gemaakt door een voorbeeld van een analyse op Facebook van de schaal *Agreeableness* – de factor *Vriendelijkheid* uit het Big Five model (Park et al., 2015). Dit illustreert meteen waarom in het VU-onderzoek een combinatie van beide benaderingen gekozen wordt.



## Praten met interviewrobot Sigmund

In het onderzoek worden de data voor de algoritmes verzameld met behulp van een video interview via internet. Daarnaast heeft Ward van Breda, de ontwikkelaar van de algoritmes in het onderzoek, deze ook beschikbaar gemaakt via *Sigmund*, een “menselijke” robot.

In zijn algoritmes richt hij zich op het meten van persoonlijkheidsprofielen en op het in kaart brengen van emoties. Het interessante van het inzetten van *Sigmund* is dat deze robot – naast het registreren van emoties – er ook zelf met passende emotie op kan reageren. Een aspect van het onderzoek is dan ook of met kandidaten in de interactie met deze “menselijke” robot meer laten zien dan in de interactie met de toepassing op het internet.<sup>2</sup> Ter illustratie hieronder een voorbeeld van één van de algoritmes dat – aan de hand van het HEXACO-model (Big 5 uitgebreid met *integriteit*) – persoonlijkheid meet:

Trait Type	Trait value	Confidence value	Trait Label	Confidence Label
Conscientiousness	0.5	0.1	moderately positive	no confidence
Emotionality	0.28	0.51	moderately positive	pretty conflicting
Agreeableness	0.31	0.54	moderately positive	pretty conflicting
Emotionality	0.42	0.31	moderately positive	moderately conflicting
Openness	0.38	0.27	moderately positive	slightly conflicting
Integrity	0.41	0.22	moderately positive	moderately conflicting

In dit voorbeeld wordt gebruik gemaakt van een algoritme op basis van een *top down* strategie. Hierin worden woordenlijsten gebruikt met factorladingen die relateren aan de verschillende persoonlijkheidsstreken. Deze woorden worden bij een match binnen de grammaticale context geanalyseerd, waaruit vervolgens een score voortkomt tussen -1 en 1, waarbij -1 een sterke negatieve relatie, 0 een neutrale, en 1 een sterke positieve relatie vertegenwoordigd met de persoonlijkheidsstrek. Naast het doorontwikkelen van dit algoritme, worden er op dit moment ook alternatieve text analyse algoritmes ontwikkeld op basis van een *bottom up* strategie. Dit is ook het geval voor algoritmes die stemgebruik en gezichtsexpressie interpreteren. Welke algoritmes er uiteindelijk gebruikt gaan worden om persoonlijkheidsstreken te meten, is afhankelijk van de accuraatheid van metingen op een onafhankelijke testset.

## Responsible AI in HR

### Fixed algorithms

Kunnen we dan nu al op basis van alleen AI een zinvolle voorspelling doen over toekomstig functioneren? Uit onderzoek ten behoeve van onze eigen assessmentpraktijk, weten we dat de *fixed algorithms* (de top down benadering, algoritmes gebaseerd op theoretische modellen en expertoordelen) hun waarde bewezen hebben. Deze algoritmes zijn controleerbaar, het is geen black box. De algoritmes worden door veel assessmentpartijen op een verantwoorde wijze ingezet, met succes en bewezen voorspellende waarde. Het strikt hanteren van het statistische model laat in veel gevallen betere uitkomsten zien dan wanneer er een menselijke beoordelaar aan het beoordelingsproces wordt toegevoegd. Met de menselijke beoordelaar sluipt onvermijdelijk bias in het oordeel. Maar die menselijke beoordelaar heeft op dit moment nog wel het beste zicht op de specifieke context van de selectie- of ontwikkelvraag. Daarom richten wij ons op de doorontwikkeling van een lerend systeem waarin wij de oordelen van de menselijke assessor toevoegen als onderdeel van de algoritmes. Alle uitkomsten krijgen een weging op basis van hun theoretische voorspellende waarde (Schmid, 2016). Input en weging kunnen per functie en context verschillen en worden in overleg met de opdrachtgever ingeregeld. Aan dit systeem worden ook performance data toegevoegd. Met behulp van continue analytics worden de algoritmes verder verfijnd, waardoor het algoritme door de tijd heen steeds beter wordt.





## Selecteren en ontwikkelen met behulp van Artificial Intelligence

### Free algorithms

Een ander verhaal zijn de *free algorithms (bottom up)*. Dit zijn de algoritmes die louter gebaseerd zijn op bijvoorbeeld machine learning technieken, waarbij de (steeds veranderende) gevonden patronen oncontroleerbaar tot een oordeel leiden. Ook op dit gebied doen wij onderzoek, juist omdat wij vinden dat de kwaliteit van selectiehulp-middelen aan de hoogste eisen moet voldoen, gezien de impact die zij kunnen hebben. Voorlopig zetten wij duidelijke vraagtekens bij het inzetten van deze vorm van AI. Wij vinden dat de wetenschappelijke basis nog te dun is en er op dit moment nog een te groot risico is dat de algoritmes “vervuild” raken met bias, zonder dat dat controleerbaar is (we hebben tenslotte nog steeds een voorkeur voor witte middelbare mannen als het gaat om het vervullen van veel posities, dat zullen de data bevestigen).

Het feit dat deze *free algorithms* een black box zijn en het gebrek aan onderbouwing van hun voorspellende waarde, hebben recent geleid tot een aantal reacties van beroepsorganisaties. Zo heeft het NVP een nieuwe sollicitatiecode geformuleerd waarin strenge richtlijnen voor selectie via algoritmes zijn opgenomen. Eén van de bepalingen luidt: “bij inzet van data-gedreven recruitment-methoden met behulp van AI of algoritmes, dienen deze middelen transparant en gevalideerd te zijn. De mogelijke risico’s en tekortkomingen daarvan moeten ook duidelijk zijn voor de kandidaat.” Eind vorige jaar heeft een mensenrechtenorganisatie in de VS zelfs een klacht ingediend tegen een AI-driven assessment organisatie en hen beschuldigd van ‘*unfair en deceptive practices*’<sup>3</sup>. Voorlopig zien wij het dus nog niet gebeuren dat AI-toepassingen de recruiter of HR-professional zal gaan vervangen. Het is realistisch om AI in te zetten als IA: *Intelligence amplified*; als zeer waardevolle *tool* voor de recruiter, en niet *in plaats van* de recruiter.

### Regelgeving

Het is goed dat er nu aandacht en regelgeving is om onverantwoorde AI-toepassingen te kunnen controleren en aan banden te leggen. Maar wij zouden het ook een gemiste kans vinden om de mogelijkheden van AI ten behoeve van selectie- en ontwikkeling van mensen te ontkennen en naast ons neer te leggen. Met voorbeelden uit de medische wetenschap, waar AI niet meer weg te denken is, is het aan ons om deze toepassingen ook voor de psychologische praktijk op een verantwoorde wijze toepasbaar te maken. We kunnen dit niet alleen aan technici en statistici overlaten, en hebben daarin als HR- en MD-professionals een belangrijke rol te vervullen.

Als wij het niet doen, doen anderen het voor ons. | Mz

### De pijlers van Responsible AI

Responsible Data Science / AI is een samenwerking tussen NL universiteiten waarin onderzoek wordt gedaan naar hoe je data op een verantwoorde manier kunt verwerken.

Dit zijn de aspecten waar zij op wijzen:

- **Fairness** (gelijkheid): machine learning gaat over systemen die leren van voorbeelden (datapunten) in plaats van software regels. De data doen het werk. Maar hoe goed zijn die voorbeelden? Geven die een evenwichtig beeld van de werkelijkheid? Zit er geen bias in? Je gebruikt geen data over cultuurbeleving in het Gooi om een film-middag voor jongeren in Amsterdam ZO te programmeren.
- **Accuracy** (nauwkeurigheid): er is een verschil tussen correlatie en causaliteit. De ontwikkeling van het aantal echtscheidingen in een bepaald gebied en periode kan gelijk zijn aan de ontwikkeling van het aantal uren dat gekeken wordt naar voetbal in hetzelfde gebied en dezelfde periode. Relatie betekent nog geen oorzakelijk verband.
- **Confidentiality** (vertrouwelijkheid): als je een oorzakelijk verband hebt gevonden, wil dat nog niet zeggen dat je die – zonder toestemming – mag gebruiken. Voorbeeld: vader krijgt postpakket met happy box (voor dochter), omdat zij op internet informatie heeft gezocht over zwangerschapskwalen; of een aanbieding van een sportschool als je een week lang pizza’s hebt besteld.
- **Transparency** (helderheid): computer says “no”. Bij wie moet je aankloppen als de computer/algorithmte je afwijst? Klopt het algoritme? Daarnaast noemt Jim Stolze in zijn boek *Algoritmisering, wen er maar aan* nog een vijfde factor:
- **Accountability**: accountants krijgen een nieuwe rol: de controle op (de uitkomsten van) algoritmes. Verschillen accountantskantoren zijn hier reeds actief mee bezig.

### Verder lezen

- Kwint, J. en de Joode, M. (2018), *Ik ben Erica, de psychologie van veranderkracht*, uitgeverij Kompas
- Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., Ungar, L. H., & Seligman, M. E. P. (2014), *Automatic personality assessment through social media language*, Journal of Personality and Social Psychology, 108(6), 934–952
- Schmidt, F.L (2016), *The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 100 Years of Research findings*, ResearchGate
- Stolze, Jim (2018), *Algoritmisering, wen er maar aan*, uitgeverij Boom, 2018

### Noten

<sup>1</sup> Niessen, Suzan, Rob Meijer, en Marvin Neumann, *Mis(ver)standen in de selectiepraktijk*, [www.tijdschriftdepsycholoog.nl/artikelen/misverstanden-in-de-selectiepraktijk](http://www.tijdschriftdepsycholoog.nl/artikelen/misverstanden-in-de-selectiepraktijk), 1 november 2019

<sup>2</sup> zie ook <https://www.ltp.nl/ltp-in-ad-robot-sigmund-kiest-beste-sollicitant/>

<sup>3</sup> <https://www.washingtonpost.com/technology/2019/11/06/prominent-rights-group-files-federal-complaint-against-ai-hiring-firm-hirevue-citing-unfair-deceptive-practices/>