AUDIT
MAGAZINE

ARTIKELEN | 21 JUNI 2021

VERTROUWEN IN AI – EEN AANPAK

Auteur: Hans Roelfsema - Michiel Krol - Theo-Jan Renkema
Beeld: Andy Kelly - Ryan Stone
Leestijd: 5 min

Artificial intelligence (AI) krijgt steeds meer invloed op onze besluitvorming. Controle op de werking van de intelligente algoritmes die onze data interpreteren, is essentieel om het vertrouwen in deze oplossingen te behouden. In dit artikel wordt een nieuw framework voor algorithm assurance beschreven.

De hoeveelheden data die organisaties produceren en verzamelen, nemen exponentieel toe. Steeds meer organisaties ontsluiten de 'schatten' die in die data verborgen liggen. Of doen een poging hiertoe. Data-analyse levert kostbare inzichten op, waarmee bedrijven hun klanten beter kunnen bedienen, accuratere voorspellingen kunnen doen, risico's beter kunnen beheersen, de kans op fraude kunnen verkleinen en nieuwe patronen kunnen ontdekken die een

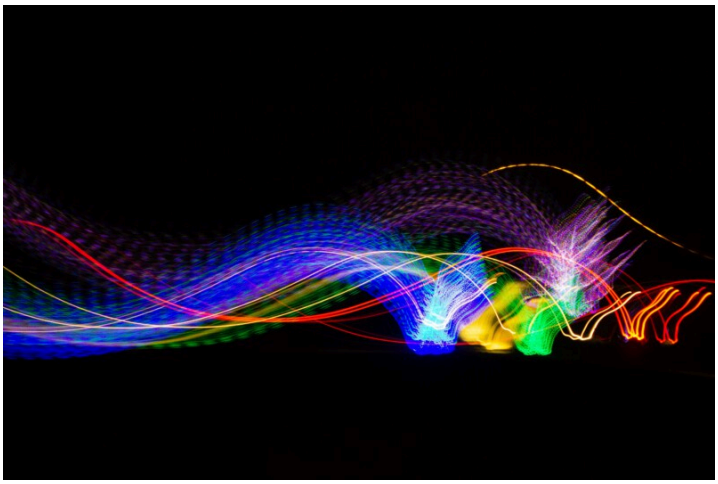
voedingsbodem zijn voor nieuwe businessmodellen.

Het is essentieel dat het vertrouwen in AI-modellen behouden blijft nu algoritmes steeds meer invloed krijgen op beslissingen die directe gevolgen kunnen hebben voor organisaties, mensen en de maatschappij

Geen handwerk

De analyse van grote hoeveelheden data is al lang geen handwerk meer. Voor geavanceerde data-analyse zijn organisaties aangewezen op automatisering. Algoritmes nemen de analyse over. Artificial intelligence en machine learning helpen om analyses voortdurend te verbeteren en inzichten steeds effectiever toe te spitsen op de grootst mogelijke waarde voor de business.

Dergelijke 'intelligente' algoritmes worden in de praktijk steeds vaker toegepast in productieomgevingen, waar hun inzichten en bevindingen directe gevolgen kunnen hebben voor klanten, werknemers, toeleveranciers en de business als geheel. Naarmate de potentiële impact van die algoritmes groeit, wordt het ook belangrijker goede methoden te ontwikkelen voor de beheersing omtrent de werking van de algoritmes. Organisaties willen zeker weten dat geautomatiseerde operationele processen zich houden aan alle beleidsregels en ook inderdaad de beoogde resultaten blijven leveren. Klanten, gebruikers en andere partijen die mogelijk afhankelijk kunnen zijn van de conclusies, willen erop kunnen vertrouwen dat de gebruikte algoritmes onbevooroordeeld en veilig zijn en dat ze voldoen aan alle wet- en regelgeving.



Bestaande assurance-aanpak is ontoereikend

Om mensen en organisaties vertrouwen te geven in IT, bestaan al jaren IT-assuranceprogramma's. Helaas blijken de methoden die voor dergelijke assurance worden gebruikt niet geschikt om algoritmes te beoordelen. Dat heeft te maken met een aantal zaken.

Complexiteit

Allereerst heeft dat te maken met de complexiteit van algoritmes. Ze bevatten vaak mechanismen die niet te vangen zijn in logische business rules en daarmee zijn deze mechanismen voor het menselijk brein moeilijk navolgbaar.

Omstandigheden

Daarnaast heeft het te maken met de omstandigheden waaronder de huidige algoritmes zijn ontstaan. Omdat de

belofte van data-analyse groot is, staat er veel druk op organisaties om de potentiële waarde snel te ontsluiten. De focus ligt dan ook vooral op snelheid van innovatie, wat ten koste gaat van beheersbaarheid. Vaak is niet duidelijk wie in de organisatie precies verantwoordelijkheid draagt voor zo'n algoritme.
Het risico is bijvoorbeeld dat organisaties algoritmes al in gebruik nemen voordat alle mechanismen om de correcte werking te waarborgen volledig zijn ingericht. 'Design for assurance', waarbij beheersingsmechanismen al in de ontwerpfase in de algoritmes worden meegenomen, is nog geen standaardpraktijk. Dat kan onder meer betekenen dat auditors achteraf moeten proberen te ontleden hoe de algoritmes precies werken (post-assurance). Bestaande assurance frameworks zijn daar over het algemeen niet op ingericht.

Ethische vragen

Ten slotte wordt assurance voor algoritmes nog ingewikkelder doordat ethische vragen een rol gaan spelen. Dat gaat niet alleen over bekende factoren als veiligheid en privacy, maar bijvoorbeeld ook over de vraag hoe algoritmes omgaan met mogelijke vooringenomenheid (bias), vervuiling of manipulatie in de datasets. In de meeste bestaande assurance frameworks voor algoritmes wordt wel degelijk nagedacht over bovenstaande aspecten, maar blijft het over het algemeen bij principiële aanbevelingen, zonder aan te geven wat in de praktijk moet gebeuren om het probleem op te lossen.

Naarmate de potentiële impact van die algoritmes groeit, wordt het ook belangrijker goede methoden te ontwikkelen voor de beheersing omtrent de werking van algoritmes

Algorithm assurance en AI assurance support

Nu algoritmes steeds meer invloed krijgen op belangrijke beslissingen die directe gevolgen kunnen hebben voor organisaties, mensen en de maatschappij, is het essentieel dat het vertrouwen in AI-modellen behouden blijft. De Rabobank en PA Consulting hebben daarom samen een nieuw en degelijk framework voor algorithm assurance ontwikkeld, dat (in tegenstelling tot vele bestaande frameworks) de volledige levenscyclus van AI-modellen in beschouwing neemt. Dit framework stelt auditors in staat niet alleen een vinger op de zere plek te leggen, maar ook concreet aan te geven wat er moet veranderen in bijvoorbeeld processen, de organisatie of in de techniek om het probleem te verhelpen. Het framework is inmiddels toegepast op algoritme gerelateerde assurance cases, waardoor het zich al in de praktijk heeft bewezen.

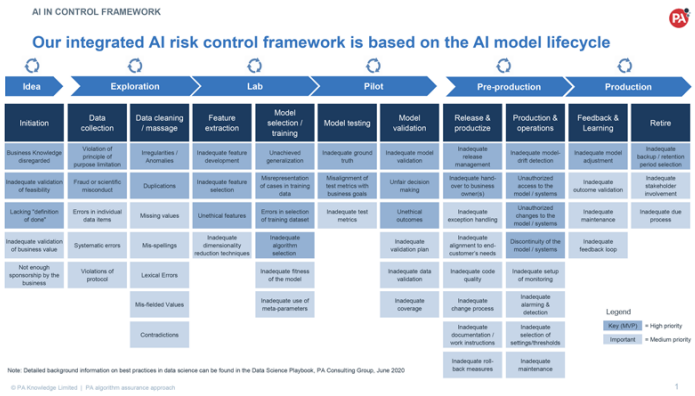
Nieuw raamwerk

Een essentieel onderdeel van deze algorithm assurance-aanpak is een nieuw raamwerk voor risicobeheersing. Met dit AI Risk Control Framework worden de risico's die gepaard gaan met het gebruik van AI-modellen, zorgvuldig en volledig in kaart gebracht. Het raamwerk is gebaseerd op de levenscyclus van algoritmes vanaf initiatie en ontwikkeling naar productie, tot uiteindelijk stopzetting en juiste afhechting van gebruikte modellen en data. De robuustheid van het algoritme wordt geadresseerd, maar ook de organisatie en governance eromheen, de vereisten aan de productieomgeving en datamanagementaspecten. Daarnaast wordt specifiek aandacht besteed aan risico's op het gebied van 'vertrouwen', zoals dataprivacy, veiligheid en ethiek.

Groeimodel

Het framework is een groeimodel, dat rekening houdt met nieuwe inzichten in het vakgebied die direct in het framework verwerkt kunnen worden. In dit framework zijn de risico's opgenomen die zich voordoen gedurende de levenscyclus van een AI-model. Voor elke stap in de levenscyclus zijn de potentiële risico's in kaart gebracht, samen met de beheersmaatregelen om deze risico's te beperken en, voor auditdoeleinden, het bewijs dat nodig is om aan te tonen dat

deze beheersmaatregelen aanwezig zijn. Dit framework is geschikt voor algorithm assurance ongeacht welk AI-model (zie *figuur 1*).



Figuur 1. AI risk control framework

Ook aan de achterkant, waar de output van de AI-modellen moet worden getest, is een nieuwe oplossing ontwikkeld (het AI assurance supportmodel), die in staat is de resultaten van AI-modellen te testen, zelfs als het onderliggende algoritme zelf niet toegankelijk is (een black-boxbenadering, zoals bijvoorbeeld bij SaaS-oplossingen het geval kan zijn). Dit supportmodel maakt op zijn beurt gebruik van AI voor de controle van de ingaande en uitgaande datastromen en de logica van het AI-model dat getest moet worden. Rabobank en PA trekken drie belangrijke conclusies uit deze nieuwe oplossingen voor algorithm assurance in vergelijking met conventionele (IT-)auditbenaderingen.

Klanten, gebruikers en andere partijen die mogelijk afhankelijk kunnen zijn van de conclusies, willen erop kunnen vertrouwen dat de gebruikte algoritmes onbevooroordeeld en veilig zijn en dat ze voldoen aan alle wet- en regelgeving

Conclusies

Ten eerste blijkt het voor auditors essentieel om de werelden van assurance, IT en data science bij elkaar te brengen. Een ervaren data scientist in het auditteam kan bijvoorbeeld de juiste vragen stellen over het AI-model, zodat een goed beeld ontstaat van hoe het is ontworpen en gebouwd en waarom bepaalde keuzen zijn gemaakt. Ook om de AI assurance supportmodellen te kunnen ontwikkelen is veel ervaring met data science een vereiste. Een gezamenlijke aanpak en proces zijn een voorwaarde om tot betrouwbare algorithm assurance te kunnen komen.

Ten tweede blijkt uit de ervaring met het AI risk control framework en AI sssurance support dat de toevoeging van een data scientist ook het conventionele auditproces verandert. De beste manier om een statistisch relevant AI assurance supportmodel te ontwikkelen is in een iteratief proces, waarbij de data scientist het model voortdurend aanscherpt om tot een optimaal resultaat te komen. Auditors zullen die iteratieve aanpak in hun methodiek moeten opnemen om ook dat proces zorgvuldig te kunnen bewaken.

Het derde en belangrijkste verschil met de conventionele benadering is dat algorithm assurance daadwerkelijk aantoont of AI-modellen echt doen wat ze moeten doen. Dat neemt zorgen en wantrouwen weg bij het management, werknemers, klanten en toezichthouders. Dankzij de toepassing die door Rabobank en PA is ontwikkeld, leidt algorithm assurance in de praktijk tot de positieve zekerheid dat algoritmes die gebruikmaken van artificial intelligence daadwerkelijk doen wat ze moeten doen.

Over
Hans Roelfsema is data transformation lead bij PA Consulting.

Michiel Krol is head of Audit Data Excellence bij de Rabobank.

Theo-Jan Renkema is chief IT & Digital Audit bij de Rabobank en hoogleraar data analytics & audit aan de Tilburg University.

Tags: Artificial intelligence, RPA

Bron url: https://auditmagazine.nl/artikelen/vertrouwen-in-ai-een-aanpak/