



AI & Ethiek

Marijn Storm

12 november 2021

Agenda

1 Wat is AI?

2 Huidige verplichtingen: AVG

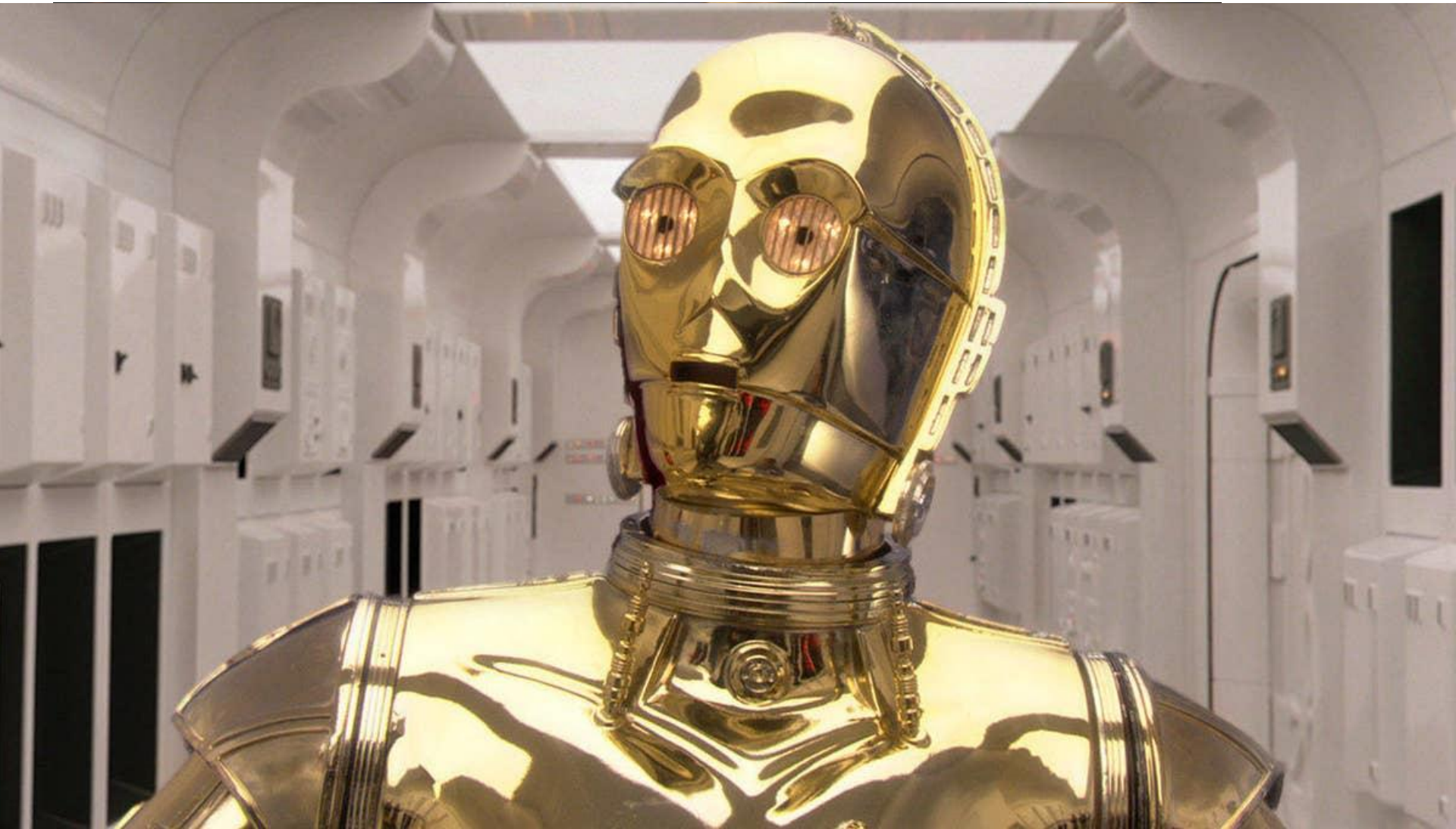
3 Uitdagingen voor AI

4 Lessen uit de praktijk

5 Toekomst: AI Verordening

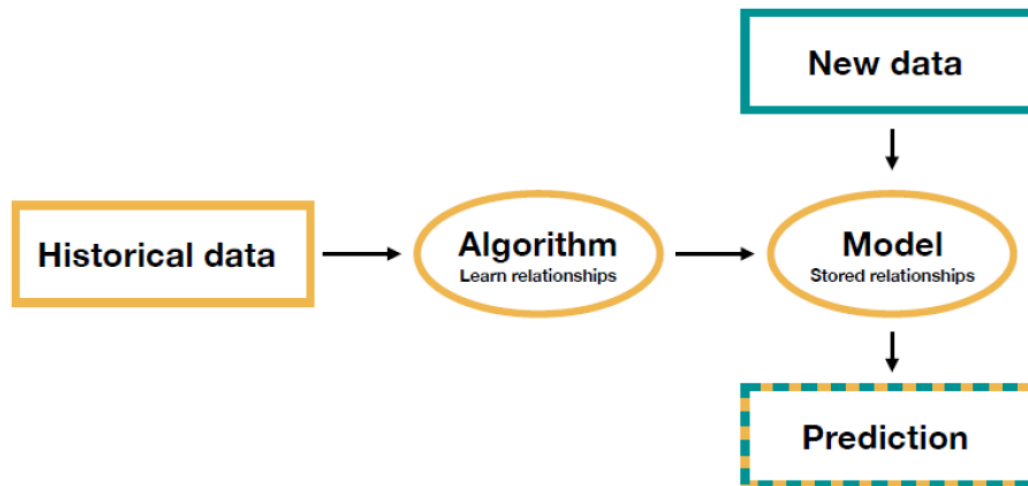


Wat is AI?



Wat is AI?

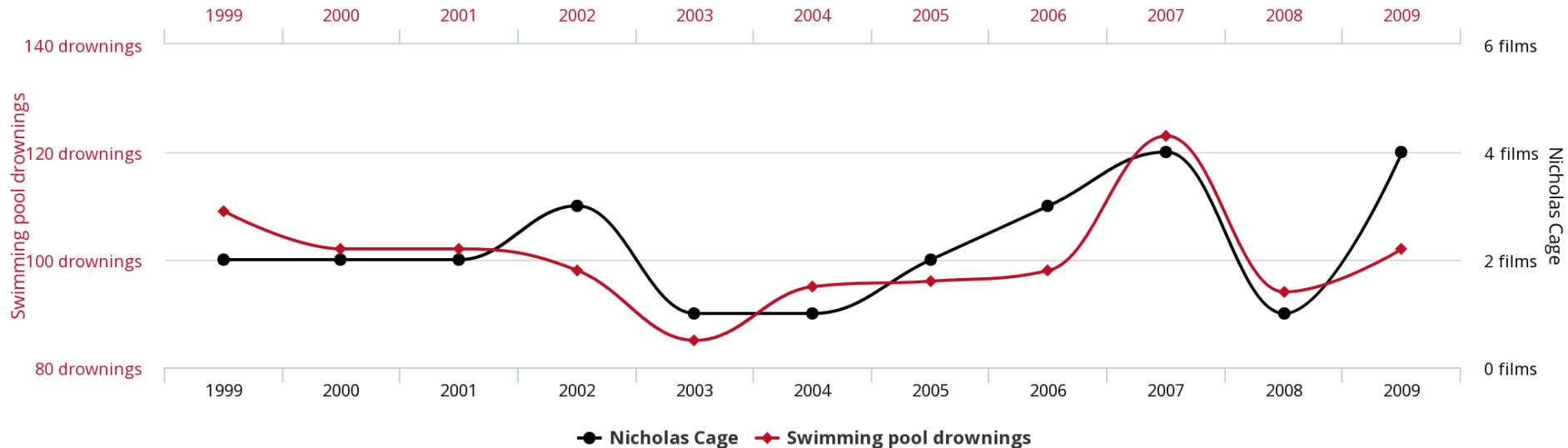
- AI stelt machines in staat om capaciteiten uit te voeren die normaal gesproken worden geassocieerd met menselijke intelligentie
 - **Redeneren**: Leren en conclusies vormen op basis van imperfecte data
 - **Begrijpen**: Interpretieren van de betekenis van data (incl. tekst, geluid, afbeeldingen)
 - **Interactie**: Interacties met anderen op natuurlijke wijze



AI is wiskunde

- AI is gebaseerd op correlaties, maar niet iedere **correlatie** is een **oorzakelijk verband**

Number of people who drowned by falling into a pool
correlates with
Films Nicolas Cage appeared in



Cijfers van [Centers for Disease Control & Prevention](#) en [Internet Movie Database](#)

tylervigen.com

Wat is AI?

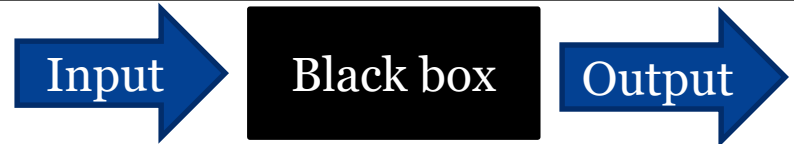
- **Voorbeeld:**

- Elektronicawinkel wil voorspellen of websitebezoekers een computer zullen kopen
- Je hebt een historische dataset:

rec	Age	Income	Student	Credit rating	Buys-computer
r1	<=30	High	Yes	Fair	No
r2	<=30	High	Yes	Excellent	No
r3	31-40	High	No	Fair	Yes
r4	>40	Medium	No	Fair	Yes
r5	>40	Low	No	Fair	Yes
r6	>40	Low	No	Excellent	No
r7	31-40	Low	Yes	Excellent	Yes
r8	<=30	Medium	No	Fair	No
r9	<=30	Low	Yes	Fair	Yes
r10	>40	Medium	Yes	Excellent	Yes
r11	<=30	Medium	Yes	Excellent	Yes
r12	31-40	Medium	No	Fair	Yes
r13	31-40	High	Yes	Fair	Yes
r14	>40	Medium	No	Excellent	No

Twee soorten AI

- Black box algoritmes



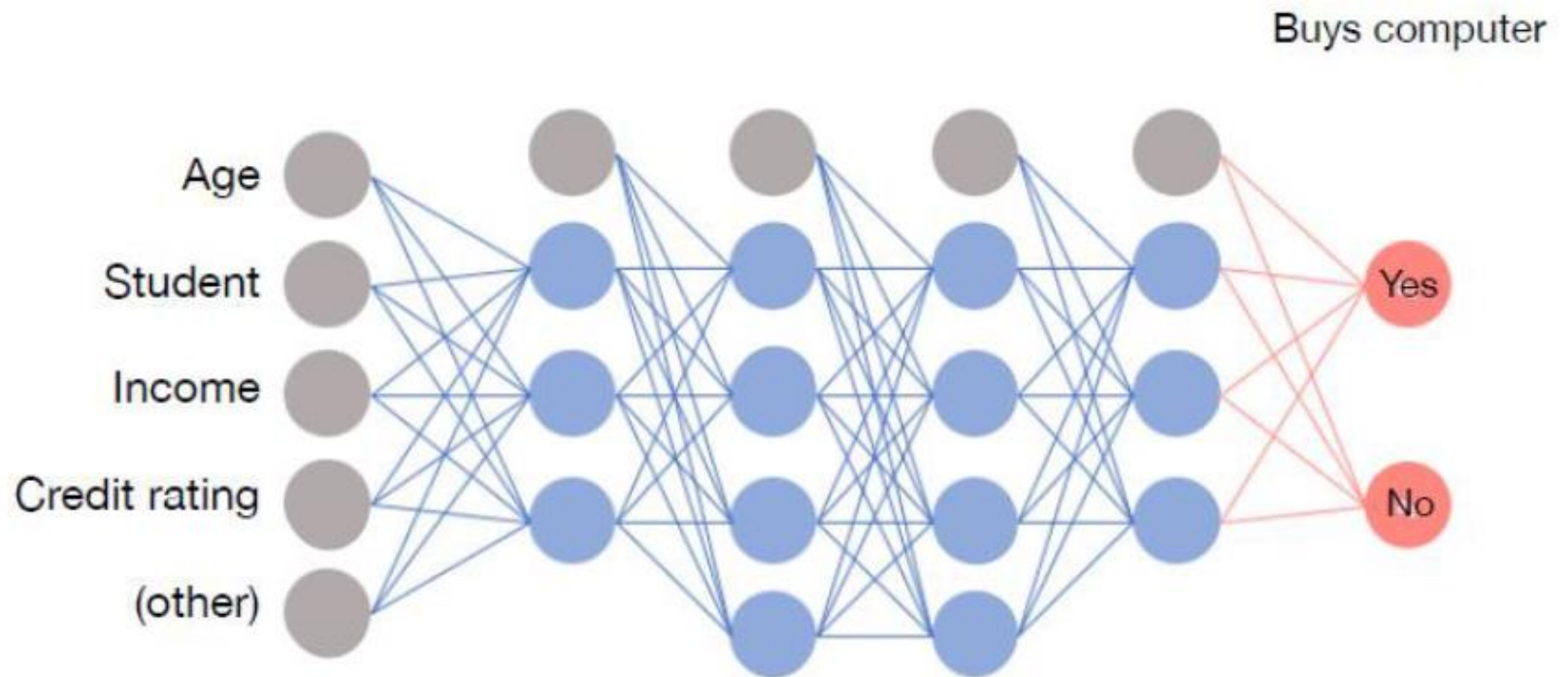
- Black box algoritmes halen automatisch correlaties uit verschillende data elementen, en blijven leren en evolueren tijdens het gebruik
- Kan leiden tot verrassende en onverklaarbare correlaties
- Uitdagingen met betrekking tot transparantie en recht op uitleg

- White box algoritmes

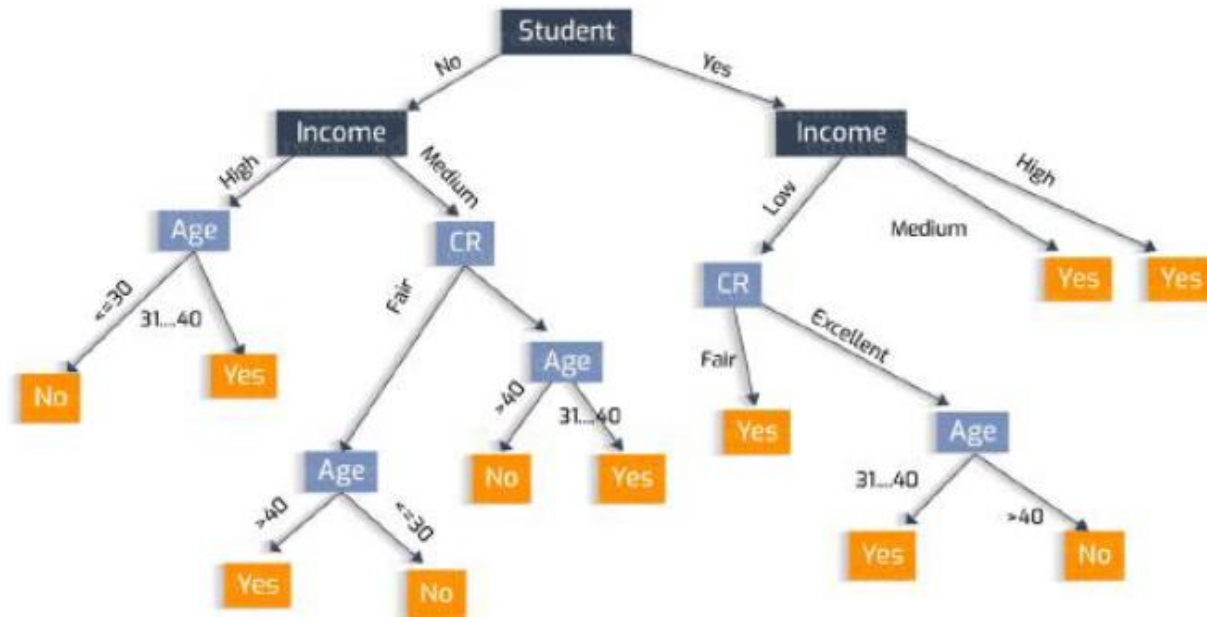


- Menselijk toezicht bij het trainen van en leren door het algoritme
- Mogelijk meer lineair en statisch dan black box algoritmes, maar kan nog steeds geavanceerde voorspellingen geven
- Over het algemeen een uit te leggen proces met consistente resultaten

Black box algoritme



White box algoritme



Buys computer

Huidige verplichtingen



Algemene beginselen

Artikel 5 AVG, verwerking van persoonsgegevens moet *onder andere* voldoen aan:

- **Rechtmatigheid, behoorlijkheid en transparantie**
 - Geen discriminatie of bias, verplichting tot uitleg
- **Doelbinding**
 - Gebruik van persoonsgegevens voor AI training?
- **Dataminimalisatie**
 - Hoeveel gegevens gebruiken voor vermindering van discriminatie of bias?

Profilering & geautomatiseerde besluitvorming

“Profilering” betekent

- **Geautomatiseerde verwerking** van persoonsgegevens
- **Evaluatie van persoonlijke aspecten** van een persoon
- Met name met de bedoeling om beroepsprestaties, economische situatie, gezondheid, persoonlijke voorkeuren, interesses, betrouwbaarheid, gedrag, locatie of verplaatsingen te **analyseren of voorspellen**

“Geautomatiseerde besluitvorming” betekent

- **Besluit** gemaakt met gebruik van persoonsgegevens
- Gebaseerd op **uitsluitend** een automatische verwerking (zoals profilering) **zonder betekenisvolle menselijke tussenkomst**
- Waaraan voor de betrokkene **rechtsgevolgen** zijn verbonden of de betrokkene **anderszins in aanmerkelijke mate treft**

AI & Geautomatiseerde besluitvorming

Intuïtieve voorbeelden:

- Geautomatiseerde afwijzing van een lening aanvraag o.b.v. een credit check algoritme
- Geautomatiseerde aanname van kandidaten o.b.v. een HR algoritme
- Geautomatiseerde weigering van toegang tot een gebouw op basis van gezichtsherkenning

Minder intuïtieve voorbeelden:

- Het uitgeven van leningsaanbevelingen zonder menselijke controle
- Geautomatiseerde ranking van top kandidaten voor een interview
- Gebruik van een algoritme om reclame voor kortlopende leningen met hoge rente te tonen aan groepen met lage inkomens

Geautomatiseerde besluitvorming

Stikte vereisten, alleen toegestaan:

- Met de **uitdrukkelijke toestemming** van de betrokkene;
- Waar noodzakelijk voor totstandkoming/uitvoering **overeenkomst**; of
- Waar toegestaan door **EU of lidstatelijk recht**

Betrokkene moet worden geïnformeerd over:

- **Bestaan** van de geautomatiseerde besluitvorming;
- **Consequenties** van de geautomatiseerde besluitvorming; en
- **Onderliggende logica** van de geautomatiseerde besluitvorming

Betrokkenen hebben recht op:

- Menselijke tussenkomst;
- Kenbaar maken standpunt;
- Uitleg over een genomen besluit; en
- Aanvechten genomen besluit

A close-up photograph of a white robotic hand with blue joints, positioned over a wooden Go board. The board is covered with numerous black and white Go stones. The robotic hand is in the process of moving a white stone. The background is softly blurred, focusing attention on the hand and the game pieces.

AI uitdagingen

The **AI** did it
is not an acceptable excuse.

Algorithmic Accountability

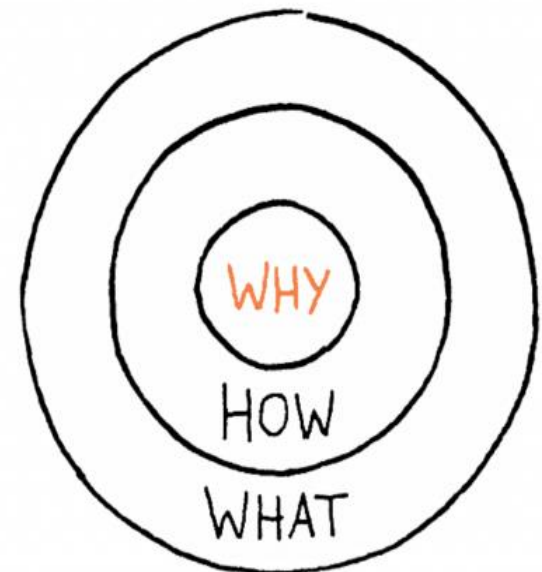
implies an obligation to report and justify algorithmic decision making and to mitigate any negative social impacts or potential harms.

Doel: uitleg van de rationale

Om betrokkenen in staat te stellen hun rechten te gebruiken, moeten ze weten:

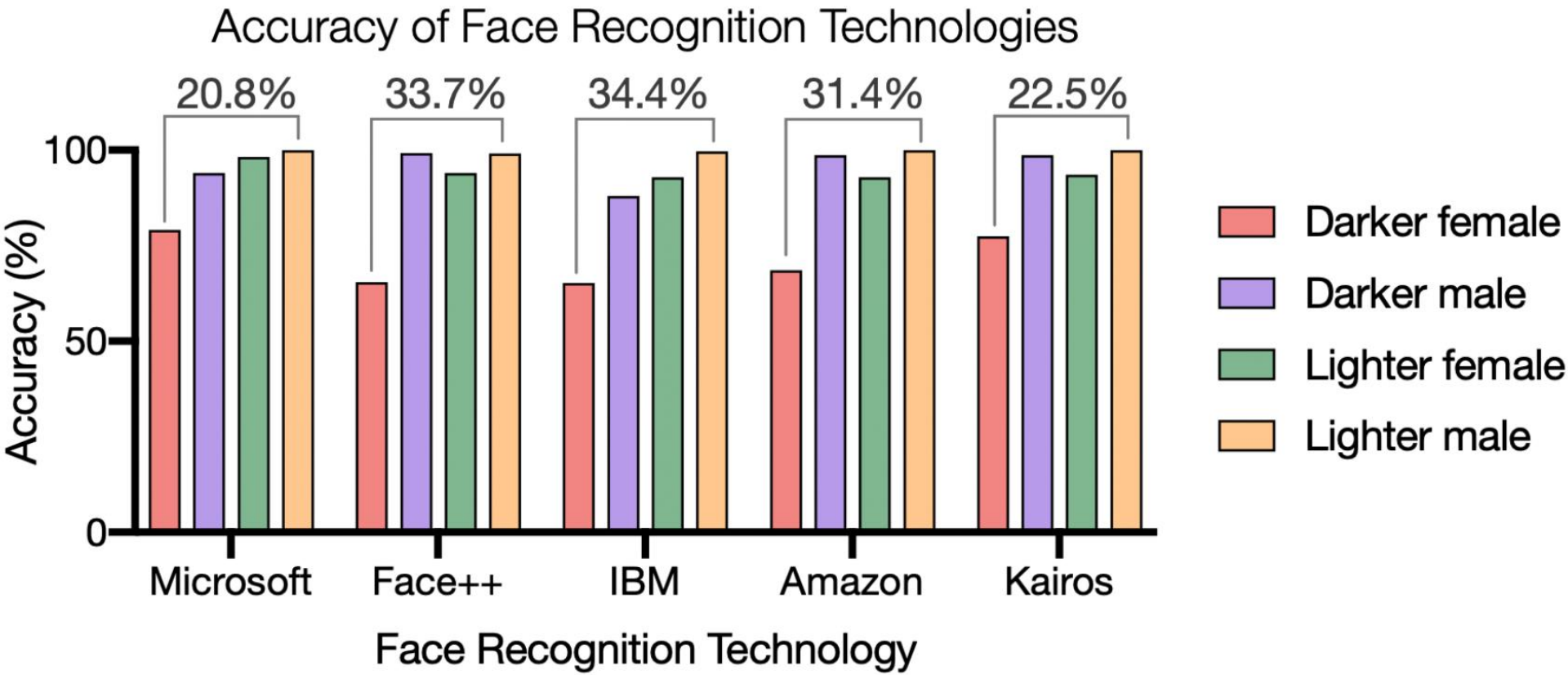
- **Wat** er is gebeurd (er is een geautomatiseerde beslissing genomen)
 - Privacyverklaring verstrekken
- **Hoe** de beslissing is genomen
 - Nuttige informatie over onderliggende logica
- **Waarom** de beslissing is genomen
 - Uitleg van de specifieke beslissing (waarom het X is, en niet Y)
- Zodat betrokkenen in staat zijn
 - Hun standpunt te bepalen
 - De beslissing te betwisten
 - Hun rechten af te dwingen

The Golden Circle



Training data is vaak ongelijk verdeeld

- AI is wiskunde: **ongelijke input = ongelijke output**



Verbod verwerken bijzondere persoonsgegevens

Algorithms can reduce discrimination, but only with proper data

Nov 16, 2018 Save This



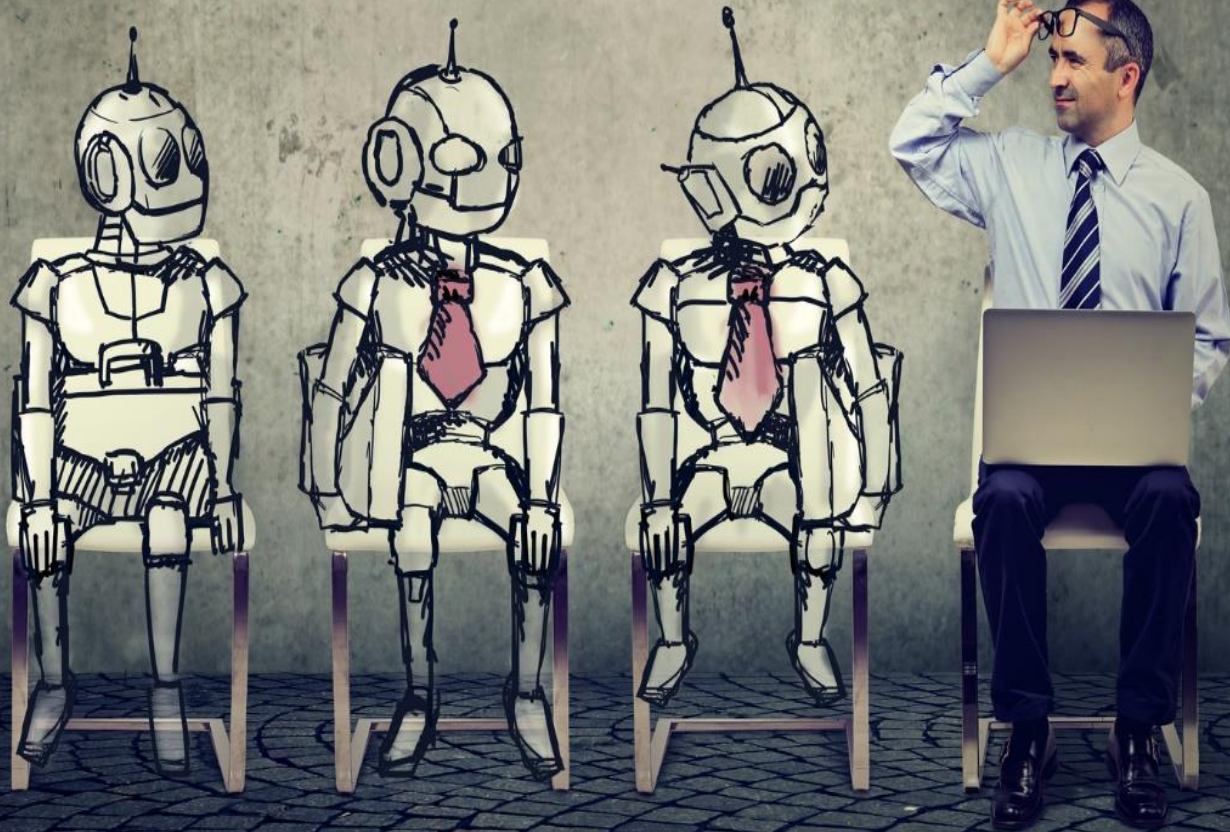
Lokke Moerel
IAPP Member Contributor

If we want to develop fair algorithms, we must get rid of the taboo of collecting ethnic data

Algorithms can reduce discrimination, but only with proper data ([link](#))

Lokke Moerel / IAPP, 2018

Lessen uit de praktijk

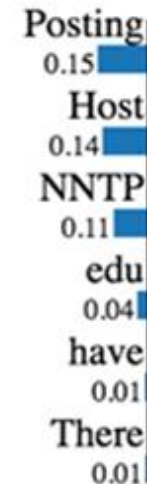


Uitleggen van AI: XAI

Prediction probabilities



atheism christian



Text with highlighted words

From: johnchad@triton.unm.edu (jchadwic)

Subject: Another request for Darwin Fish

Organization: University of New Mexico, Albuquerque

Lines: 11

NNTP-Posting-Host: triton.unm.edu

Hello Gang,

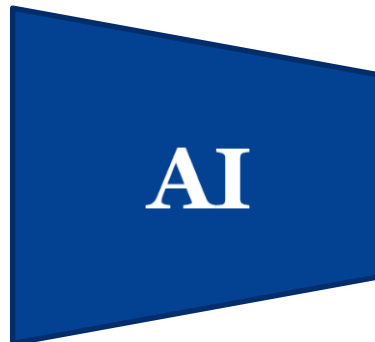
There have been some notes recently asking where to obtain the DARWIN fish.

This is the same question I have and I have not seen an answer on the

net. If anyone has a contact please post on the net or email me.

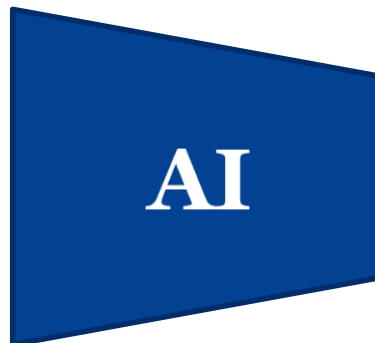
Gevoelige gegevens verbergen werkt averechts

#	Leeftijd	Opleidings-niveau	Etniciteit
1	<=30	HBO	Westers
2	<=30	WO	Niet-westers
3	31-40	WO	Westers
4	>40	HBO	Westers
5	>40	MBO	Westers
6	>40	WO	Niet-westers
7	31-40	HBO	Westers
8	<=30	HBO	Westers
9	<=30	MBO	Westers
10	>40	WO	Niet-westers
11	<=30	WO	Niet-westers
12	31-40	WO	Westers



#	Leeftijd	Opleidings-niveau	Etniciteit
1	<=30	HBO	Westers
3	31-40	WO	Westers
4	>40	HBO	Westers
7	31-40	HBO	Westers
8	<=30	HBO	Westers
12	31-40	WO	Westers

#	Leeftijd	Opleidings-niveau	Etniciteit
1	<=30	HBO	
2	<=30	WO	
3	31-40	WO	
4	>40	HBO	
5	>40	MBO	
6	>40	WO	
7	31-40	HBO	
8	<=30	HBO	
9	<=30	MBO	
10	>40	WO	
11	<=30	WO	
12	31-40	WO	

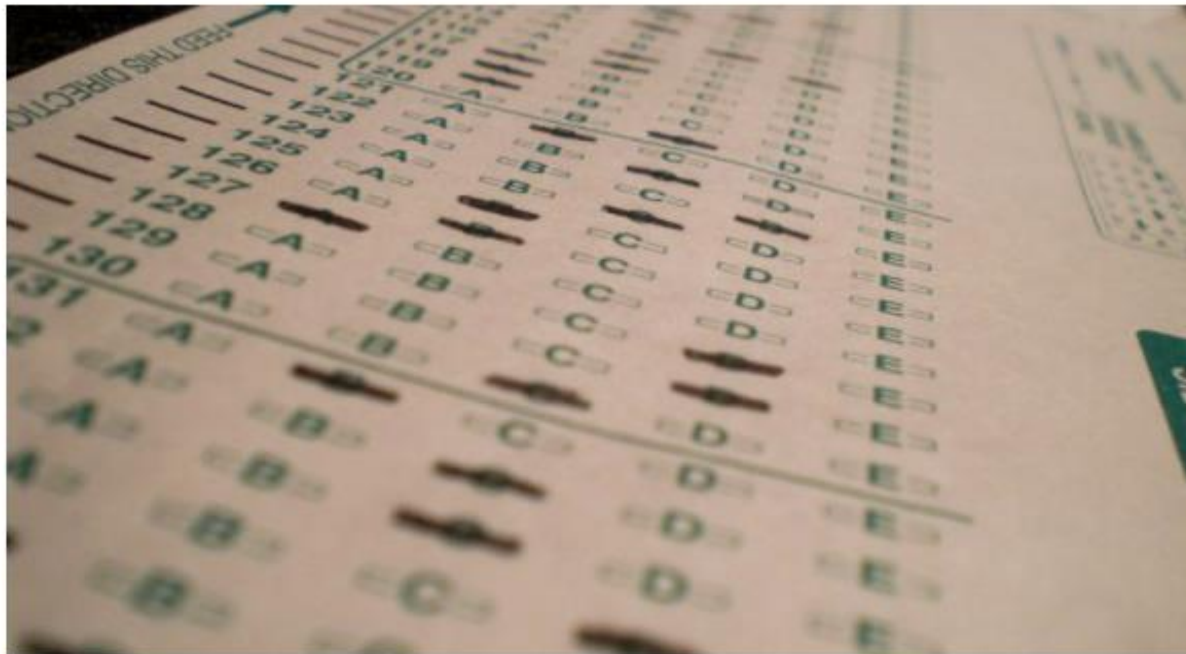


#	Leeftijd	Opleidings-niveau	Etniciteit
1	<=30	HBO	
3	31-40	WO	
4	>40	HBO	
7	31-40	HBO	
8	<=30	HBO	
12	31-40	WO	

The Highest-Performing Women Are (Still) Scoring Lower Than Men on the SAT

2/15/2018 by **PHYLLIS ROSSER**

Women's average scores on the SAT have always been lower than men's—even though they receive higher average grades in all courses in high school and college. The SAT gender gap was first reported in *Ms.* in 1985; although the difference in average scores between the sexes has narrowed since then from 59 to 20 points, it has never been eliminated. 53 percent of SAT test takers are women, but the gate-keeping college entrance exam is underestimating their ability to succeed.

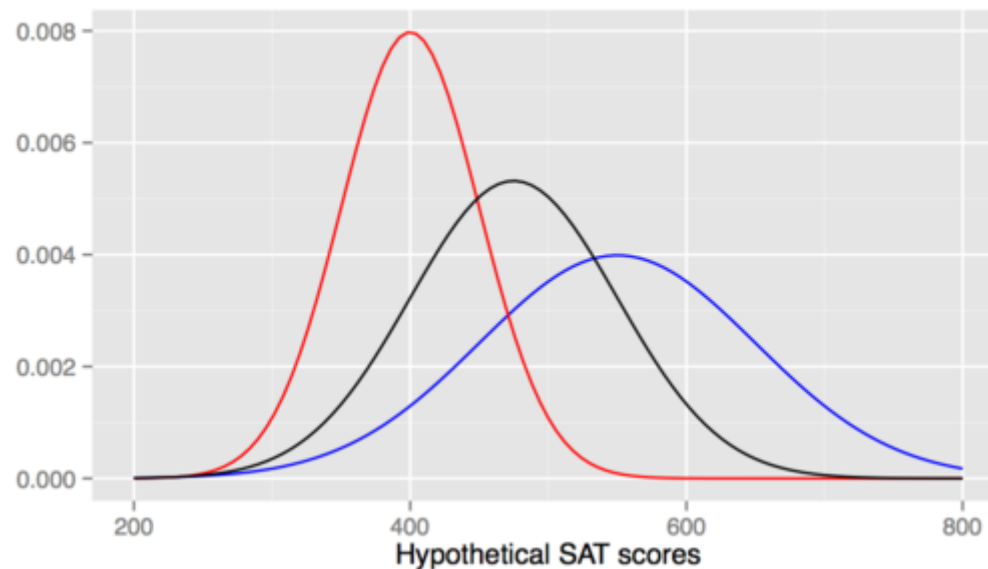


[timlewisnm](#) / Creative Commons

Bron: <https://msmagazine.com/2018/02/15/highest-performing-women-still-scoring-lower-men-sat/>

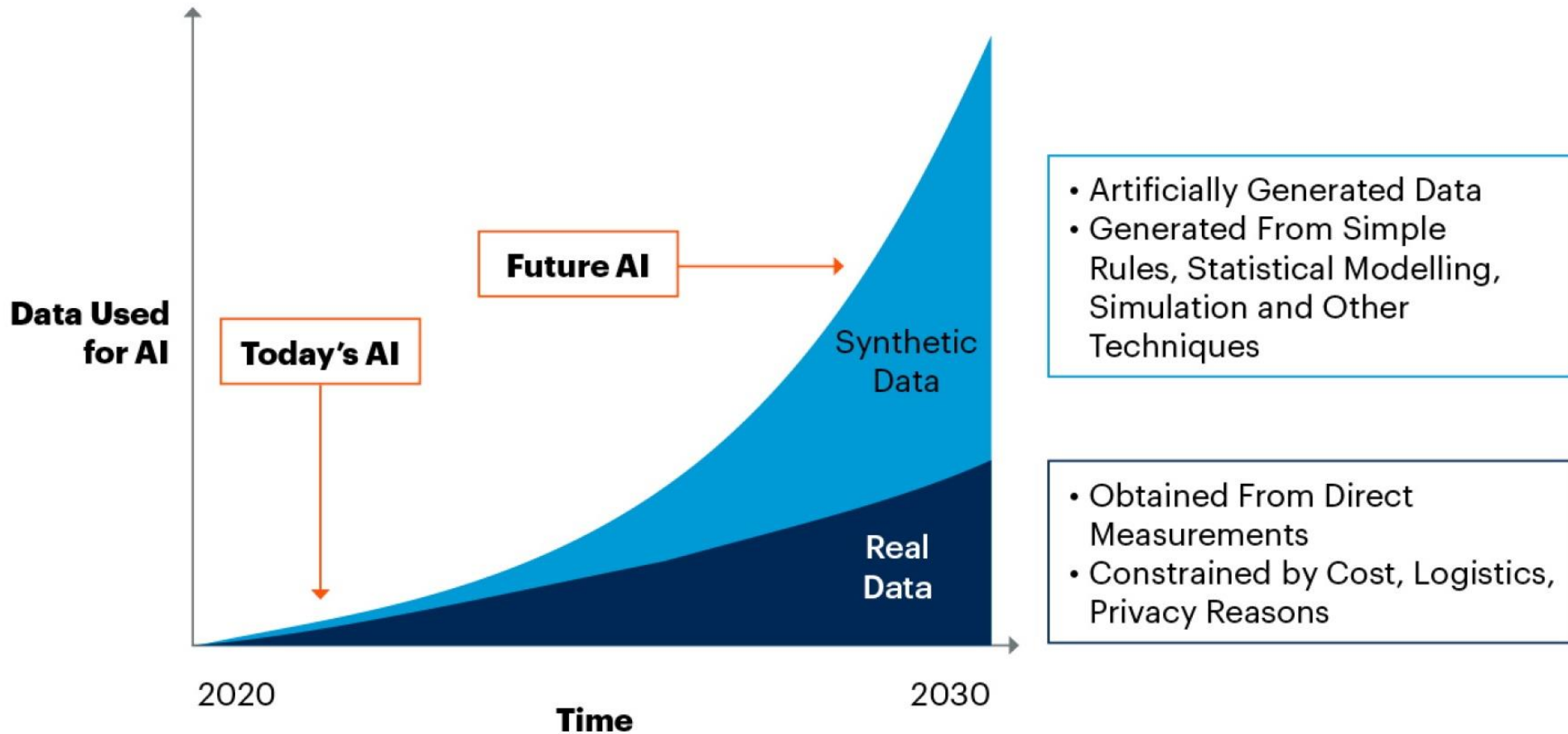
Data elementen repareren

- Het verschil in SAT scores kan worden gerepareerd:
 1. Splits de SAT scores van mannen en vrouwen in groepen, bijvoorbeeld per 5%
 2. Bereken de gemiddelde score per groep (van mannen en vrouwen gecombineerd)
 3. Ken deze gemiddelde score toe aan zowel de mannen in vrouwen in de groep
 4. Resultaat: de bias is verwijderd uit het data element



Training dataset: vul gaten met synthetische data

By 2030, Synthetic Data Will Completely Overshadow Real Data in AI Models



Forget About Your Real Data — Synthetic Data Is the Future of AI ([link](#))

Gartner / Maverick Research, 2021

Toekomst: AI Verordening

- Voorstel van de Europese Commissie op 21 april 2021 ([link](#))



Scope: AI Systemen

“software die is ontwikkeld aan de hand van een of meer van de technieken en benaderingen die zijn opgenomen in de [onderstaande lijst] en die voor een bepaalde reeks door mensen gedefinieerde doelstellingen output kan genereren, zoals inhoud, voorspellingen, aanbevelingen of beslissingen die van invloed zijn op de omgeving waarmee wordt geïnterageerd”

AI technieken en benaderingen:

- a) Benaderingen voor machinaal leren, waaronder gecontroleerd, ongecontroleerd en versterkend leren, met behulp van een brede waaier aan methoden, waaronder diep leren (“deep learning”)
- b) Op logica en op kennis gebaseerde benaderingen, waaronder kennisrepresentatie, inductief (logisch) programmeren, kennisbanken, inferentie- en deductiemachines, (symbolisch) redeneren en expertsystemen
- c) Statistische benaderingen, Bayesiaanse schattings-, zoek- en optimalisatiemethoden

Opzet van de AI Verordening



**Overig: geen
vereisten**

**Hoog risico: voldoen
aan vereisten**

**Onaanvaardbaar risico:
verboden**

Plus: specifieke vereisten voor enkele specifieke systemen, zoals voor deep fakes

Enkele voorbeelden

Onaanvaardbaar risico:

- Subliminale technieken die leiden tot fysieke of psychologische schade
- Gebruik maken van kwetsbaarheden (zoals leeftijd of handicap) met fysieke of psychologische schade tot gevolg
- Biometrische identificatie op afstand **voor rechtshandhaving**

Hoog risico:

- Biometrische identificatie op afstand
- Werving en selectie, specifiek:
 - bekendmaken van vacatures, het screenen of filteren van sollicitaties, de evaluatie van kandidaten tijdens interviews
 - promotie of ontslag, toewijzen van taken, toezicht op gedrag, evaluatie van personeel

Verplichtingen

Veel verplichtingen zijn vergelijkbaar met die van verwerkingsverantwoordelijken onder de AVG

- Maar nu ook voor de leverancier
- En ook als de AI buiten de EU wordt getraind
 - Zie voor meer informatie: *De privacywet en de nieuwe EU Artificial Intelligence Verordening: wat is er nieuw en wat niet?* ([link](#))

Enkele voorbeelden:

- Systeem voor risicobeheer (incl. mitigerende maatregelen)
- Specifieke verplichtingen m.b.t. data governance om bias en discriminatie te voorkomen
 - AI Verordening voorziet in uitzondering op Art. 9 & 10 AVG, voor monitoring, opsporing, en correctie van bias
- Ontworpen voor menselijk toezicht

Maar het duurt nog even

- Eerst beoordeling en mogelijkheid voor amendementen door Europees Parlement en Europese Raad
 - Geen vaste tijdslijnen
 - Resulteert waarschijnlijk in aanpassingen
 - Maar een lager beschermingsniveau is onwaarschijnlijk
- Vervolgens een overgangsperiode, nadat de AI Verordening is aangenomen door het Europees Parlement
- Huidig voorstel:
 - Implementatieperiode van **twee jaar**
 - Extra transitieperiode voor bestaande AI systemen van **een jaar**



Vragen?



MARIJN STORM

mstorm@mofo.com

