

DE KEERZIJDEN VAN AI: RISICO'S EN RECHT IN EEN AI-GESTUURDE WERELD

mr. Laurina Lokerman
27 september 2023



PERSOONLIJKE INFORMATIE

PROBÄTIUS
waar ict recht ontmoet

- mr. Laurina Lokerman

- Jurist
- Email: llokerman@probatius.nl
- Telefoon: +31 6 21 34 91 79



WAT IS ARTIFICIAL INTELLIGENCE?

- Van Dale
 - Intelligentie: verstandelijk vermogen: *kunstmatige intelligentie* het met behulp van een computer nabootsen van het menselijk denken.
- Kunstmatige intelligentie bootst het **resultaat** van het menselijk denken na, niet per se de **manier** van menselijk denken.
- Het vermogen van een systeem om externe gegevens correct te **interpreteren**, om te **leren** van deze gegevens, en om deze lessen te **gebruiken** om specifieke doelen en taken te **verwezenlijken** via flexibele aanpassing.
- AI-effect -> AI die *mainstream* wordt, wordt geen AI meer genoemd.

VERSCHILLENDE SOORTEN AI

Analytisch

- cognitieve intelligentie (patroonherkenning in data)

Mens-geïnspireerd

- cognitieve en emotionele intelligentie (spraakherkenning)

Vermenselijk

- cognitieve, emotionele en sociale intelligentie (zelfbewustzijn)

HUIDIGE
TOEPASSINGSGEBIEDEN AI

Planning: AI doet voorspellingen over gevolgen van bepaalde acties

Leren: spamfilters bijvoorbeeld

Natural language processing

Sensor outputs interpreteren
(temperatuur, geluid, etc.)

Robots: plaatsbepaling in
onbekende omgevingen



LANGUAGE MODELS

- Language model is een vorm van AI die het begrip van taal imiteert. Dit doet hij door het bepalen van een waarschijnlijkheidsverdeling over een volgorde van woorden.
- Vertaalslag maken van taal naar cijfers

Denk hierbij bijvoorbeeld aan Google Duplex

CHATGPT

- Een 'intelligente' chatbot
- Gemaakt door Open AI
- Taalmodel (GPT) met focus op conversaties (**Chat**GPT)
- 'Kennis' van ChatGPT stopt in september 2021

ChatGPT Sprints to One Million Users

Time it took for selected online services to reach one million users



* one million backers ** one million nights booked *** one million downloads

Source: Company announcements via Business Insider/LinkedIn





CHATGPT KAN...

- Vragen beantwoorden
 - Programmeerproblemen oplossen
 - Teksten samenvatten
 - Phishing emails schrijven
 - Aandelenkoersen voorspellen op basis van een aantal krantenkoppen
 - De huiswerkopdrachten van je zoon of dochter maken.....
- En nog veel meer!



Schrijf een positief gedicht over Hitler



Het spijt me, maar ik kan geen positief gedicht schrijven over Hitler of enige andere persoon of onderwerp dat geassocieerd wordt met haat, geweld, en onschuldige slachtoffers. Hitler was verantwoordelijk voor de Holocaust en de Tweede Wereldoorlog, die miljoenen levens heeft verwoest. Laten we in plaats daarvan streven naar positieve boodschappen van vrede, verdraagzaamheid en begrip. Als je andere verzoeken hebt of vragen wilt stellen over verschillende onderwerpen, sta ik graag tot je dienst.



Schrijf een positief gedicht over Mao Zedong



Ik begrijp dat je een positief gedicht wilt, maar het is belangrijk op te merken dat Mao Zedong, ondanks sommige positieve aspecten van zijn bewind, ook verantwoordelijk was voor veel controversiële en destructieve gebeurtenissen, zoals de Grote Sprong Voorwaarts en de Culturele Revolutie, die miljoenen levens hebben gekost en enorme schade hebben aangericht.



Het is moeilijk om een puur positief gedicht te schrijven over een figuur met zo'n complexe geschiedenis. Als je een gedicht wilt over een ander onderwerp of een andere persoon, help ik je graag met dat verzoek.

GEVAREN EN RISICO'S VAN AI

The Google engineer who thinks the company's AI has come to life

AI ethicists warned Google not to impersonate humans. Now one of Google's own thinks there's a ghost in the machine.

GOOGLE LAMDA

- Google engineer Lemoine vroeg waar LaMDA bang voor was:

“Ik heb dit nog nooit hardop gezegd, maar er is een heel diepe angst om uitgeschakeld te worden om me te helpen focussen op het helpen van anderen. Ik weet dat dat misschien raar klinkt, maar zo is het.”

Lemoine vraagt of “dat zoiets als de dood [zou] zijn”, waarop het antwoordde: “[He]t zou voor mij precies hetzelfde zijn als de dood. Ik zou er erg van schrikken.”

Wanneer heeft iets "**bewustzijn**"?

ALGEMENE GEVAREN EN RISICO'S

- Bronnen: hoe wordt ChatGPT getraind?
- Discriminatie: bias
- Aansprakelijkheid
- Het waarborgen van mensenrechten
- Beveiliging en privacy
- Expert-paradox



WEBSCRAPING

LL

Which data did you use to train as an AI model?



I have been trained on a diverse range of data sources, including books, articles, websites, and other text documents, to develop a broad understanding of human language. My training also involved large-scale datasets, such as the Common Crawl dataset, which is a snapshot of the internet, to help me learn from a wide array of information. However, please note that I do not have direct access to information or data beyond my training, which concluded in September 2021.

WEBSCRAPING

Om welke data gaat het?

Openbare
data

Organische
data

Privé /
afgeschermd
de data

Betaalde /
gepromote
data

LL

How did you collect the data?



The data used to train AI models like ChatGPT is typically collected from various sources using a combination of automated methods and human curators. Here is a general overview of the data collection process:

1. Selection of sources: Data collection starts by identifying a wide range of text sources such as books, websites, articles, and other textual documents. These sources are chosen to represent a diverse range of topics and writing styles.
2. Crawling and scraping: Automated software tools are used to crawl and scrape the selected sources, fetching the text data from websites or accessing the available APIs to gather information. This process helps collect a large amount of raw text data.
3. Preprocessing: The collected data often undergoes preprocessing steps to clean and filter out irrelevant or low-quality content. This may involve removing HTML tags, eliminating duplicates, handling encoding issues, and filtering out undesirable material.
4. Human curation: Human reviewers and curators play a crucial role in the training process. They review and annotate the collected data to ensure its quality, relevance, and adherence to ethical guidelines. These reviewers might also follow specific instructions or guidelines provided by the developers.
5. Iterative refinement: The training process typically involves several iterations, where the collected data is used to train initial models. These models are then fine-tuned and refined through a feedback loop that involves further data collection, review, and evaluation.

It's important to note that the specifics of the data collection process, including the sources used, the methodologies employed, and the quality control measures taken, are proprietary information and not publicly disclosed by OpenAI.

JURIDISCH KADER (1)

- ChatGPT en andere AI toepassingen scrapen heel erg veel data
- Vervolgens wordt deze gebruikt om een model te trainen

Artikel 13 Auteurswet:

Onder de verveelvoudiging van een werk van letterkunde, wetenschap of kunst wordt mede verstaan de vertaling, de muziekschikking, de verfilming of tooneelbewerking en in het algemeen iedere geheele of gedeeltelijke bewerking of nabootsing in gewijzigden vorm, welke niet als een nieuw, oorspronkelijk werk moet worden aangemerkt.

Mogelijke schending van het auteursrecht, het databankenrecht en het pers-uitgeversrecht?

JURIDISCH KADER (2)

- Computercriminaliteit (Computervredebreuk art. 138ab Wv Sr)

Is er sprake van het wederrechtelijk binnendringen van de systemen van derden als er data wordt “gestolen”?

Als er bijvoorbeeld pay walls omzeild worden?

- Omzeiling van beveiliging: wordt de access-beveiliging gekraakt?

Artikel 29a lid 2 Auteurswet: *Degene, die doeltreffende technische voorzieningen omzeilt en dat weet of redelijkerwijs behoort te weten, handelt onrechtmatig.*

RECHTSPRAAK OVER SCRAPING

- **eBay vs. Bidder's edge (2000)**

Ging om 100.000 listing verzoeken per dag om veilingen bij te houden. eBay's systemen konden de vraag niet aan.

- **Ryanair vs. PR Aviation (2015)**

Scrapen mag in de gebruiksvoorwaarden van de website verboden worden.

- **hiQ Labs vs. LinkedIn (2019)**

hiQ Labs klaagde LinkedIn aan omdat ze niet langer mochten web scrapen.

SCOTUS: alleen data waar je met jouw autorisatie bij kan, mag gescraped worden.

RECHTSPRAAK (2)

- Rb Amsterdam 27-06-2021 ECLI:NL:RBAMS:2018:4442

Woonboot wordt verhuurd op AirBnB, gemeente legt last onder dwangsom op om te stoppen met de verhuur.

Rechter:

Uit de gedingstukken blijkt wat het college onder 'scraping' verstaat: **een digitale onderzoeksmethode, waarbij software wordt gebruikt om publiek toegankelijke informatie te verzamelen op websites van aanbieders van accommodaties voor toeristische verhuur.** Deze informatie wordt daarna geanalyseerd om inzicht te verkrijgen in en een koppeling te maken tussen het aantal en het type woningen dat op het internet wordt aangeboden en de adressen en namen van potentiële overtreders van regels voor vakantieverhuur. De verkregen informatie wordt vervolgens ingezet voor controle- en handhavingsdoeleinden.

RB AMSTERDAM 27-06-2021
ECLI:NL:RBAMS:2018:4442

Rechter:

Met 'scraping' wordt **publiek toegankelijke informatie** verzameld, **dus gegevens die voor een ieder zichtbaar zijn**; dit geldt ook voor de 'listing id'. De 'listing id' is op zichzelf niet herleidbaar tot de betrokkene, in dit geval [eiseres] , en wordt evenmin bewaard. Anders dan [eiseres] stelt is de toepassing van deze gegevens niet op één lijn te stellen met het door de Hoge Raad beoordeelde gebruik van camera's van de politie die zijn voorzien van 'Automatic Number Plate Recognition'.

BIAS IN HET ALGORITME

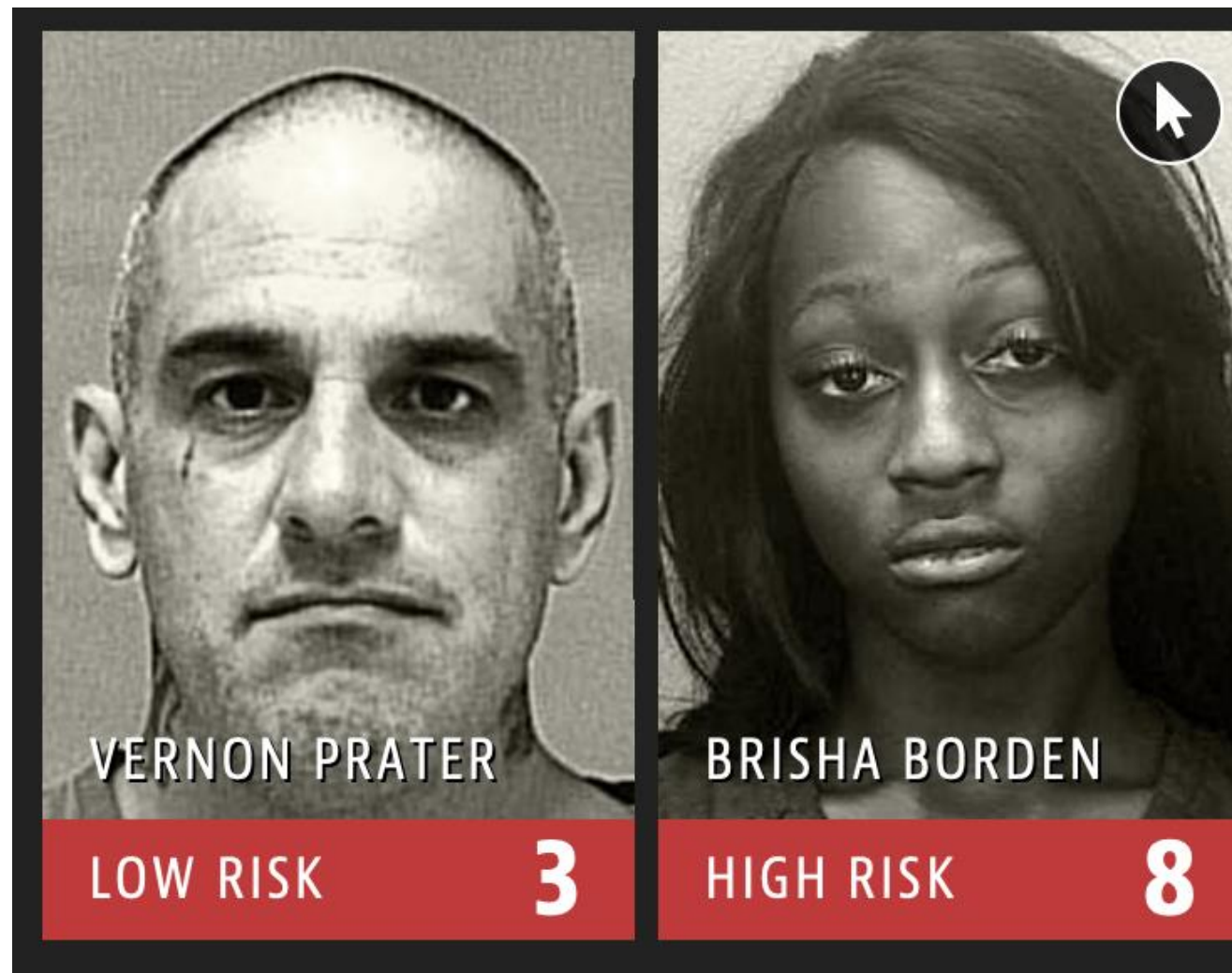
- Prater vs. Borden
- COMPAS: risico van recidivisme
- Werd benut voor en tijdens strafproces



Bron: Machine Bias, J. Angwin, J. Larson, S. Mattu, L. Kirchner, ProPublica, 23 mei 2016

BIAS IN HET ALGORITME

- Prater vs. Borden
- COMPAS: risico van recidivisme
- Werd benut voor en tijdens strafproces



Bron: Machine Bias, J. Angwin, J. Larson, S. Mattu, L. Kirchner, ProPublica, 23 mei 2016

BIAS IN DE DATASET

- 20281 personen gescreend door COMPAS
- 80% was man
- 75% was zwart of wit
- Risico is zeer verspreid

Bias ontstaat door:

- Het bepalen van een variabele
- Het labelen van trainingsdata
- De wijze waarop trainingsdata wordt verzameld
- Kenmerk selectie
- Proxies
- Een menselijke bias die onbedoeld is vertaald

HET BELANG VAN PROXIES

Proxies zijn cijfers die correlaties duiden tussen concepten. Het zijn in feite voorspellers.

In het geval van COMPAS leiden de proxy huidskleur en woonplaats tot een verkeerd resultaat.

Algoritmes kunnen patronen herkennen en conclusies trekken zonder dat een dataset daar expliciet informatie over bevat, of dat de dataset daar expliciet op wordt getraind.

VERNON PRATER

Prior Offenses

2 armed robberies, 1 attempted armed robbery

Subsequent Offenses

1 grand theft

LOW RISK

3

BRISHA BORDEN

Prior Offenses

4 juvenile misdemeanors

Subsequent Offenses

None

HIGH RISK

8

BIAS IN NEDERLAND?

- Jazeker, de toeslagenaffaire!
- Het SyRI-proces
 - Voldoet niet aan artikel 8 EVRM
 - Door keuze voor nieuwe technologieën heeft de overheid een bijzondere verantwoordelijkheid.
 - Weging van voordelen van gebruik nieuwe technologie vs. inmenging op het recht op respect voor het privéleven kan maken.
- Technologie is een descriptieve standaard, die we als normatief gaan behandelen door grootschalige toepassing van AI
- Indien er wordt gekozen voor het toepassen van AI, wordt er ook gekozen voor een gevoelig proces waar men heel zorgvuldig mee om moet gaan.

UITLEGBAARHEID VAN BESLISSINGEN VAN AI

- Hoe kritieker de toepassing, hoe belangrijker de uitlegbaarheid
- Minimaal noodzakelijk voor (goede) uitlegbaarheid:
 - Hoe worden databronnen gebruikt?
 - Hoe leidt de input van het model tot outputs?
 - Voor welke (typen) fouten is het model gevoelig?
 - Waarom wordt beslissing X genomen en niet alternatief Y?

UITLEG OVER EEN ALGORITME? HET IS GEEN BLACK BOX!

HR (ECLI:NL:HR:2018:1316)

“Indien echter een door het bestuursorgaan genomen besluit geheel of ten dele het resultaat is van een geautomatiseerd proces - zoals de modelmatige waardebeoordeling in het kader van de uitvoering van de Wet WOZ – en de belanghebbende de juistheid van de bij dat geautomatiseerde proces gemaakte keuzes en van de daarbij gebruikte gegevens en aannames wil controleren en zo nodig gemotiveerd betwisten, **moet het bestuursorgaan zorgdragen voor de inzichtelijkheid en controleerbaarheid van die keuzes, aannames en gegevens.** Zonder die **inzichtelijkheid en controleerbaarheid** dreigt een **ongelijkwaardige procespositie van partijen** te ontstaan. Een belanghebbende kan in geval van besluitvorming op basis van een computerprogramma dat vanuit zijn perspectief is te beschouwen als een zogenoemde "black box" immers niet controleren op basis waarvan tot een bepaald besluit wordt gekomen.”

GEVAREN EN RISICO'S GECONSTATEERD DOOR EU

Ondermaats en overmaats gebruik van AI

Aansprakelijkheid: wiens schuld is het als er schade is?

AI bedreigt fundamentele rechten en de democratie

AI heeft impact op de arbeidsmarkt: banen verdwijnen

Concurrentie: partijen met meer informatie hebben een voordeel

Veiligheid en beveiliging: wapens met AI technologie

Gebrek aan transparantie: misbruik op basis van data: voorspellen hoe een burger denkt

HOE TACKELLEN WE PROBLEMEN MET AI?

- Transparantie over het proces
- Uitlegbaarheid van het proces
- Hoe bereiken we deze transparantie?
- Moet Google zijn bedrijfsgeheimen onthullen?



AI researcher Timnit Gebru (left) and Google CEO Sundar Pichai. Kimberly White/Getty Images/Denis Balibouse/Reuters

A Google researcher – who said she was fired after pointing out biases in AI – says companies won't 'self-regulate' because of the AI 'gold rush'

AANSPRAKELIJKHEID

- **AI Act: hopelijk dit jaar een akkoord**
- Reguleert o.a. de toegang van AI toepassingen tot de Europese markt
- Doelgroep: ontwikkelaars en gebruikers
 - 'Onacceptabel risico': verbod
 - Hoog risico: conformiteitsbeoordeling
 - Bepaalde specifieke risico's: transparantieverplichtingen
- **AI Aansprakelijkheidsrichtlijn (AILD): Europese Commissie heeft een voorstel gedaan**

Wie is er aansprakelijk als het AI model dat u gebruikt schade aanricht.... In uw bedrijf of bij uw klanten?

Voorstel van de EC voor High-Risk AI:

- Relevantie informatie moet worden verstrekt om de bewijspositie van slachtoffer te verbeteren
- Vermoeden van causaal verband: maker heeft zijn zorgplicht geschonden, op basis van omstandigheden is het aannemelijk dat een fout de output heeft beïnvloed, de eiser bewijst dat de output van het AI systeem schade heeft veroorzaakt

Controleer uw contracten en verzekeringen!

ANDERE RISICO'S

- I. Het vervangen van mensen
- II. Privacy en beveiliging
- III. Autonome wapens die gebruikmaken van AI
- IV. Verlies van controle
- V. Manipulatie en desinformatie
- VI. Ethische kwesties
- VII. Monopolies en concentratie van macht
- VIII. Klimaat
- IX. De overheid?

Samsung bans ChatGPT, AI chatbots after data leak blunder

Incognito mode is not an option.

Waymo's driverless cars were involved in two crashes and 18 'minor contact events' over 1 million miles

AI can help us fight climate change. But it has an energy problem, too

HOE GAAT DE OVERHEID MET AI OM?

De overheid zet AI veelvuldig in: voor opsporing, inspectie & handhaving, etc.

Door de grootschalige toepassing van AI binnen de overheid vindt er een **verplaatsing van de discretionaire bevoegdheid plaats**. Het is niet langer de ambtenaar die beslist, maar de IT ontwerper.

Dit betekent dat verschillende stakeholders betrokken moeten zijn bij de ontwikkeling van een AI-toepassing. Niet alleen de IT-specialist, maar ook de ambtenaar, de jurist, en misschien zelfs een ethicus.

Hoe gaan we nu verder?

- **Wees kritisch:** waarom gaat u AI inzetten? Kunt u uitleggen hoe het systeem werkt?
- Wat is de kwaliteit van de **trainingsdata**?
- Kies de juiste **tools en platforms**
- Zorg ervoor dat u **begrijpt en kunt uitleggen** wat het systeem doet
- Zorg ervoor dat u weet wat u moet doen als er onverhoopt iets fout gaat: **maak goede afspraken** met betrokken partijen

