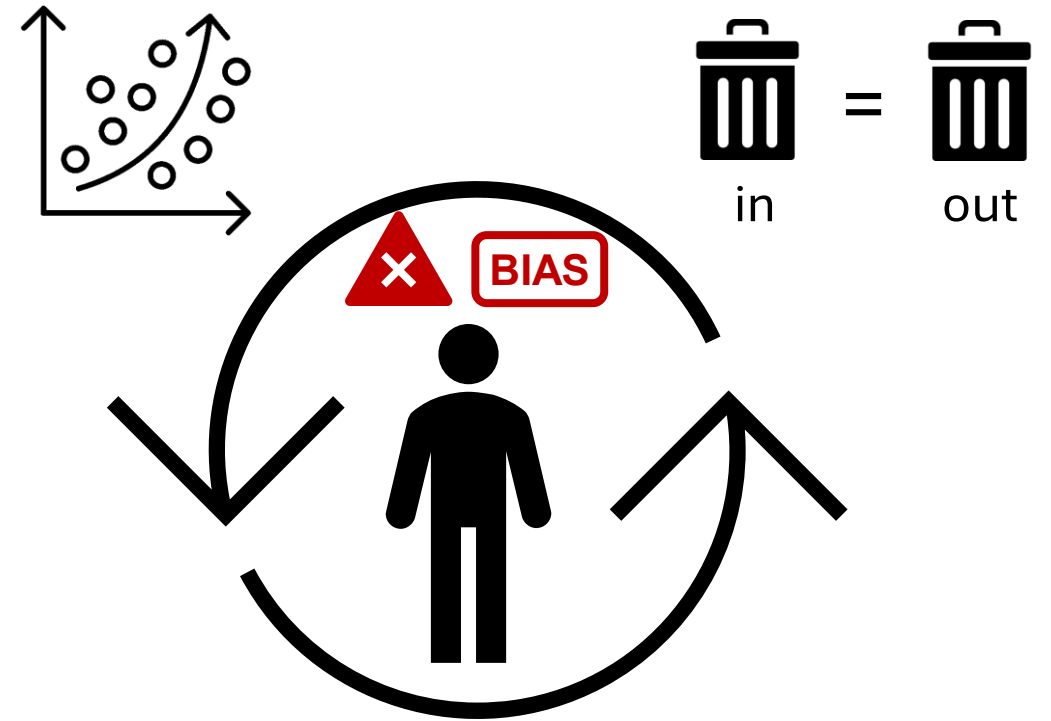
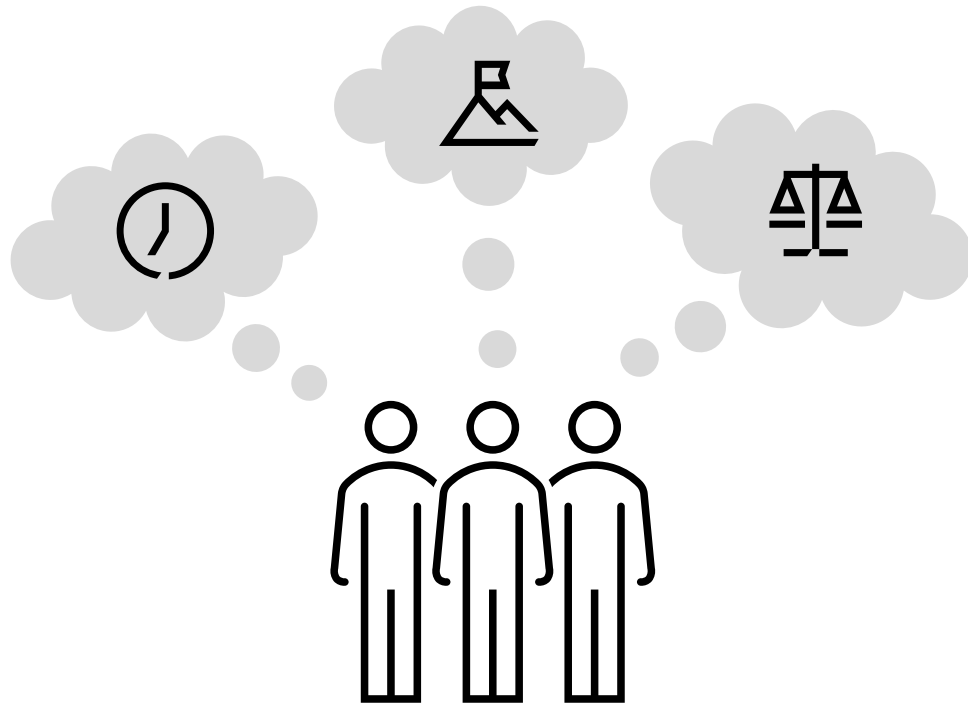




Taalmodellen en -technieken voor verbetering van datakwaliteit in een soeverein jasje

Floor van Lieshout

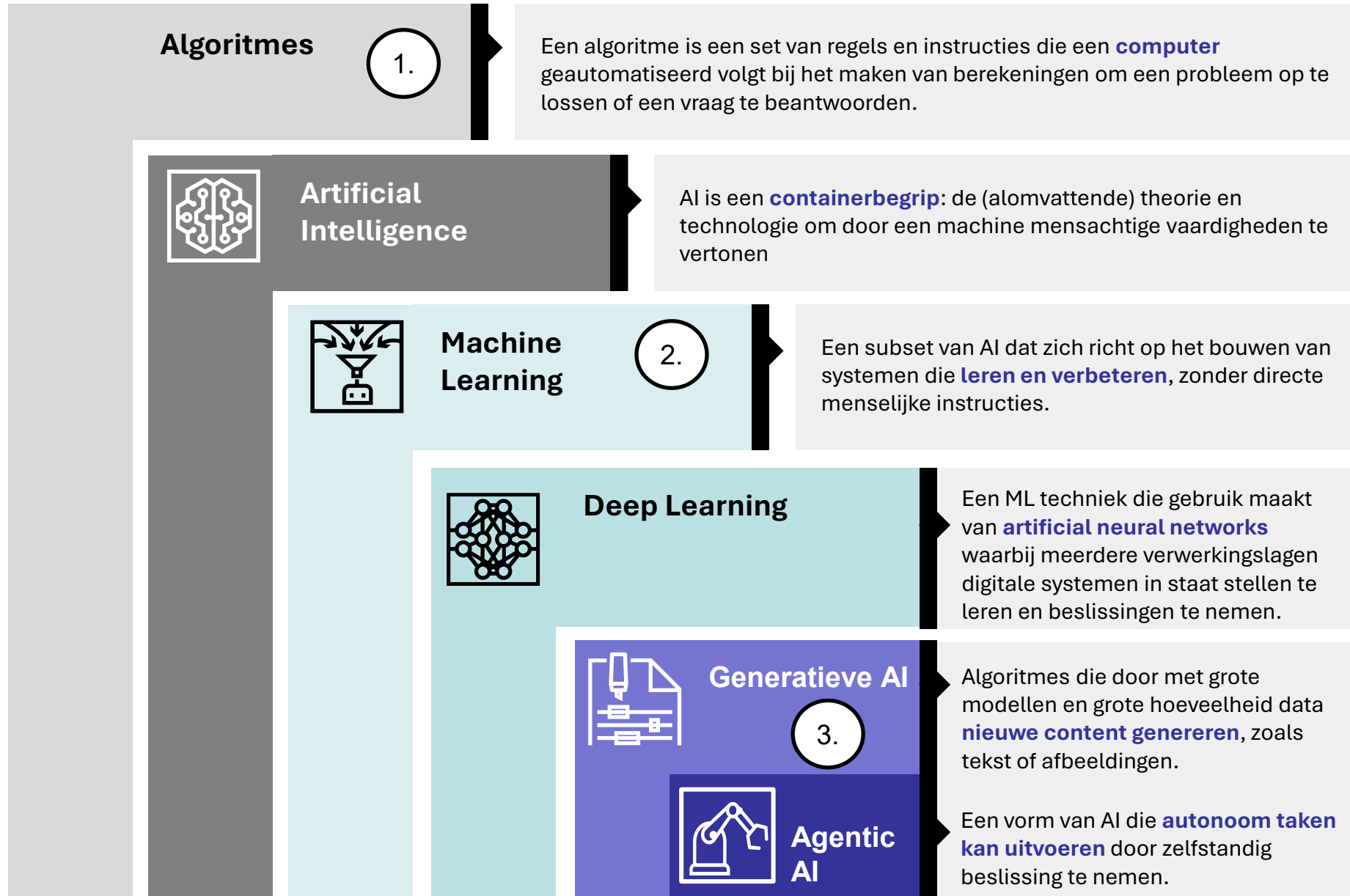




Hoe kunnen we AI systemen inzetten om onszelf te controleren?



Waar hebben we het over als we het over AI hebben?





Praktijkvoorbeeld 1: Algoritme

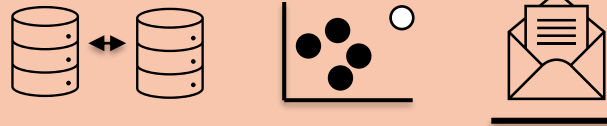
Hersteloperatie van de toeslagenaffaire

Setting

- Nieuwe compensatieroute voor aanvullende schade voor slachtoffers van de toeslagenaffaire
- $n=1$ vs. $n=1000$
- Datakwaliteit → ?
- Wat is de waarheid?

Aanpak

- Meerdere oplossingsrichtingen:



Uitdagingen

- 3 type documenten
- >70 verschillende varianten
- Gebruik van GenAI was een no-go



OCR +
taalalgoritme

Resultaten

- Van ~31.000 ouders de brieven kunnen uitlezen (79%)
- Bij ~6.300 ouders een verschil gevonden t.o.v. de data uit de systemen (4,6%)
- Steekproef van verschillen bekeken → vrijwel allemaal typfouten in de brieven

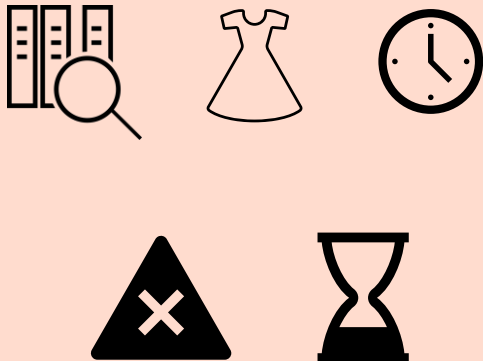


Praktijkvoorbeeld 2: Machine Learning

Assortimentsclassificatie

Setting

- Inkoopproces bij een grote publieke sectororganisatie waar 80% van de inkoopaanvragen doorheen loopt
- De beoordelaar van deze aanvragen moet het juiste assortiment selecteren



Uitdagingen

- (potentieel) vrij gevoelige data
→ Geen GenAI, wel Machine Learning
- Ook Machine Learning modellen maken fouten

Hoe zeker is het model van de voorspelling?

Wanneer is er menselijke interventie nodig?



Praktijkvoorbeeld 2: Machine Learning

Assortimentsclassificatie

Zekerheidsscore

Assortiment

Het assortiment wordt voorgesteld op basis van de aanvraag (AI-model).

VOORSPEL ASSORTIMENT **RESET VOORSPELLING**

Voorspelling actueel.

Assortiment zoeken *

Meubilair - Kantoor (Kantoorgerelateerde artikelen, inclusief IT-middelen)

Zekerheid (confidence) **Zeer zeker — 97.6%**

0.975609756097561

Top-3 voorspellingen

Assortiment

Het assortiment wordt voorgesteld op basis van de aanvraag (AI-model).

VOORSPEL ASSORTIMENT **RESET VOORSPELLING**

Voorspelling actueel.

Assortiment zoeken *

Distributie - Medicijnopslag (Geneeskundige dienst goederen)

Zekerheid (confidence) **Zeer zeker — 84.8%**

0.8477062721902754

Top-3 voorspellingen

#	Assortiment	Confidence
1	Distributie - Medicijnopslag (Geneeskundige dienst goederen)	84.8%
2	Verbruiksmaterialen - EHBO (Geneeskundige dienst goederen)	10.5%
3	Meetapparatuur - Diagnostiek (Geneeskundige dienst goederen)	4.7%



Praktijkvoorbeeld 3: Generatieve AI Hulplijn voor slachtoffers van online misbruik

Setting

- Hulplijn (chat & telefonie) voor slachtoffers van online misbruik
- Veel eigen documentatie met handleidingen, tips & tricks en partners om naar door te verwijzen
- Documentatie is slecht doorzoekbaar tijdens een gesprek

OFFLIMITS
Offlimits Hulplijn
Kennisassistent

Chats
Doxing
Doxing
Verschil tussen doxing schending

Alle gesprekken verwijderen

Nieuwe chat
Kennisbank

Jij
Wat is doxing?

Assistent

Doxing is het zonder toestemming verspreiden van persoonlijke informatie van iemand op internet, zoals naam, adres, telefoonnummer, sociale media-accounts, geloof of seksuele identiteit, met het doel te intimideren, bedreigen of vernederen.

Korte suggesties voor de hulplijnmedewerker:

- Vraag of de cliënt al bewijs heeft verzameld van de doxing-aanval.
- Vraag of de cliënt al contact heeft opgenomen met de betrokken platformen of instanties.

Gevonden documenten:
Doxing
Wetboek van Strafrecht Privaatrecht en EU verordening. Interpretatie en prioriteiten toegevoegd (geelblauw)
Template Doxing 18+

Stel je vraag...

AI gegenereerde antwoorden kunnen fouten bevatten, controleer dit altijd zelf op juistheid

Open source & open weights LLM

Volledig on-premise

Alleen gekoppeld aan eigen documentatie



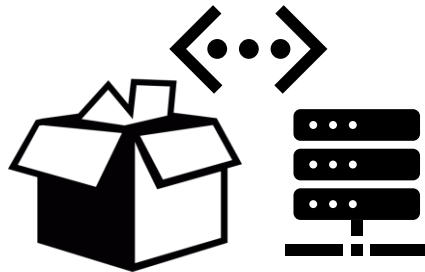
Praktijkvoorbeeld 3: Generatieve AI

Hulplijn voor slachtoffers van online misbruik

Proprietary models



Open-source models



*Open-weights models
Large*



*Open-weights models
Small*





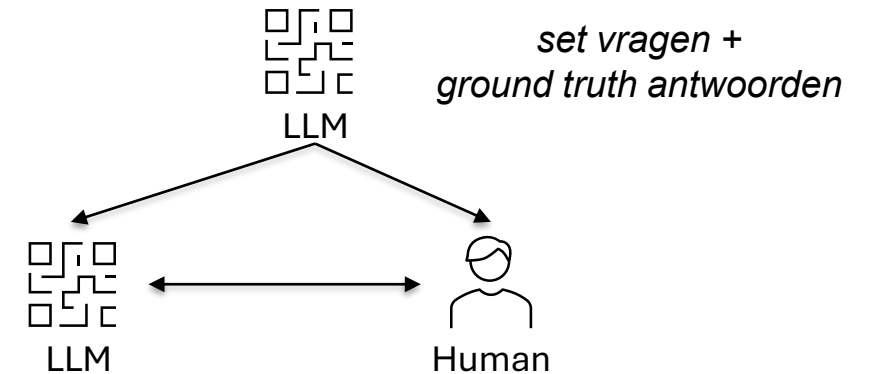
Praktijkvoorbeeld 3: Generatieve AI

Hulplijn voor slachtoffers van online misbruik

LLM as a judge

- Het inzetten van Large Language Models om de output van andere Large Language Models te toetsen
- Kan op heel veel dimensies toetsen:
 - Answer relevancy
 - Faithfulness
 - Tool correctness
 - Hallucination
 - Role adherence
- Meerdere frameworks die je hiervoor kunt gebruiken:
 - DeepEval
 - OpenAI Evals
 - LangSmith
 - JASP for audit

LLM and human as a judge



- Compleetheid
- Nauwkeurigheid
- Relevantie
- Semantische overeenstemming
- Duidelijkheid

