AI, Data science en programmeren

Deel 1/x

Inleiding

Serie van praatjes over AI, Data Science, Machine learning en programmeren.

Deze avond een introductie:

- Beschrijving AI: geschiedenis, huidige landschap en ontwikkelingen
- Terminologie
- Relatie AI met Data Science
- Inleiding Machine Learning (kern van de volgende bijeenkomst)

Enkele definities voor AI

“De wetenschap die zich bezighoudt met het creëren van een artefact dat een vorm van intelligentie vertoont.” (nl.wikipedia.org)

“AI is de mogelijkheid van een machine om mensachtige vaardigheden te vertonen - zoals redeneren, leren, plannen en creativiteit.” (Europees parlement)

“Een verzamelnaam voor machines, computers en software die taken kunnen doen waar eerder menselijke intelligentie voor nodig was.” (npokennis.nl)

AI: Vierde fase industriële revolutie?

Sommige bronnen noemen AI de vierde fase van de industriële r

Industriële revolutie: het vervangen van handmatig naar machina

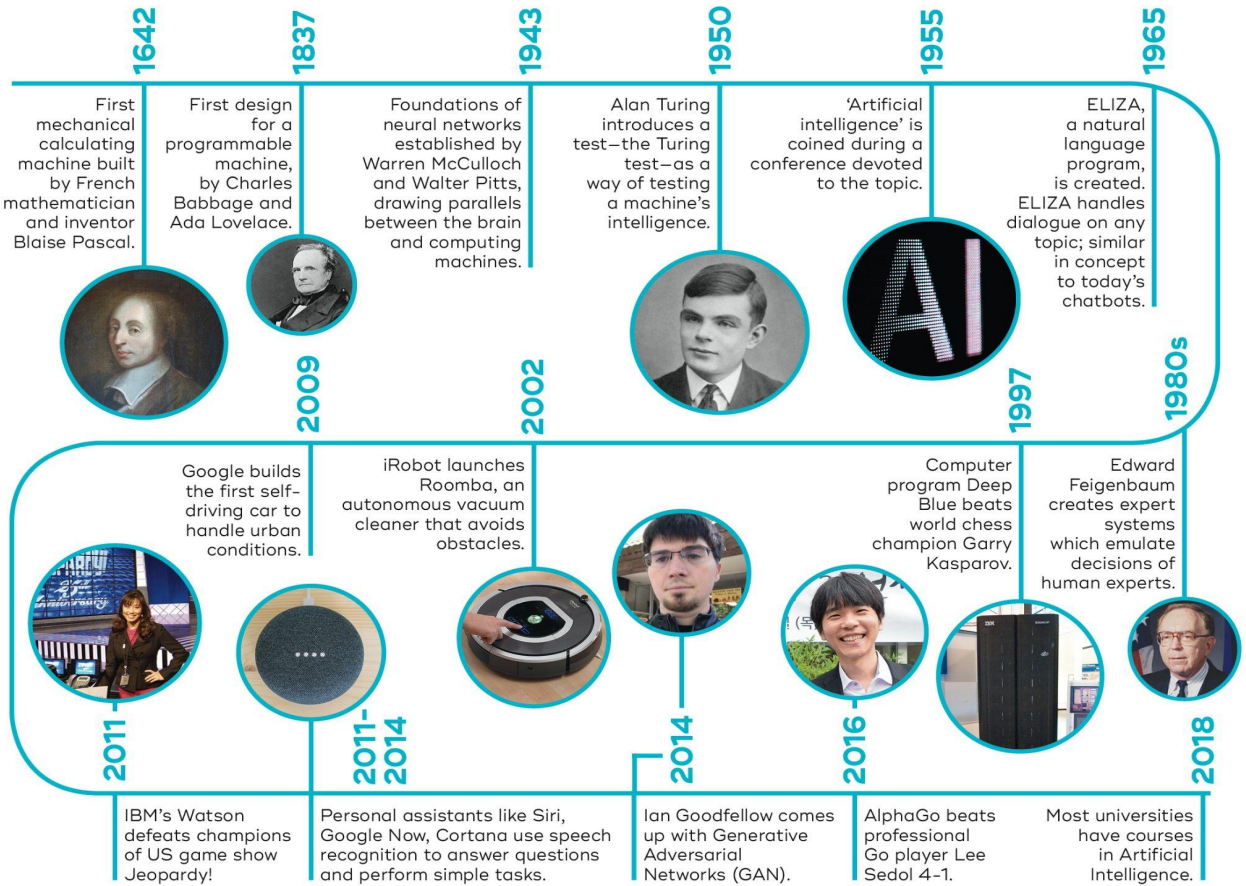
De fasen van de industriële revolutie zijn dan:

- fase 1: eind 18de eeuw: de stoommachine (en stoommotoren)
- fase 2: eind 19de eeuw: elektriciteit! Met name de uitvinding van de wisselstroommotor (transport!)
- fase 3: 2e helft 20ste eeuw: computer en internet
- fase 4: begin 21ste: AI. Richting overname denkprocessen ipv handmatige processen.

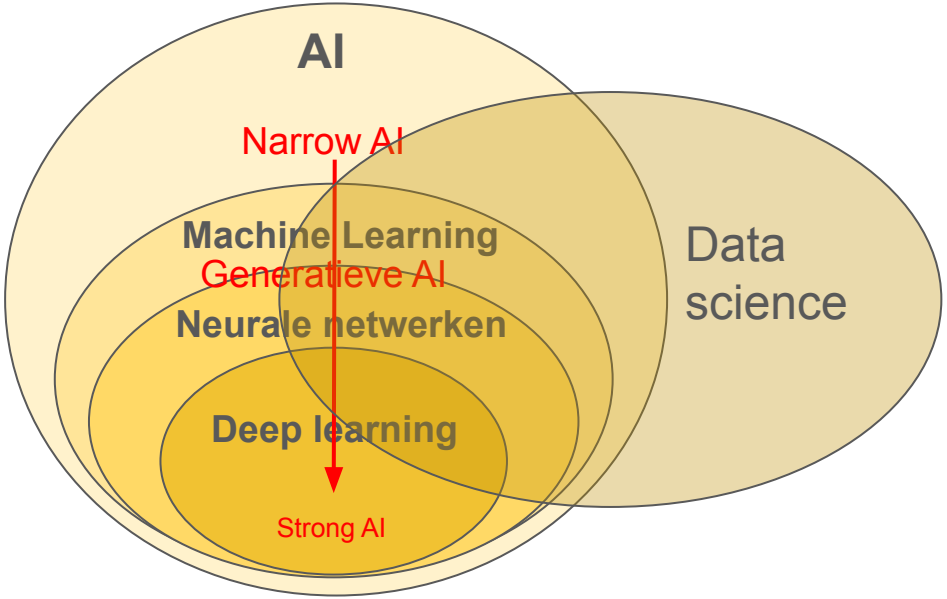
Bijdrage groei AI:

- veel data beschikbaar (internet)
- toename rekenkracht
- ontwikkeling algoritmes

Geschiedenis



AI - Deelgebieden



AI: Onderscheid zwakke (“narrow”) en sterke (“strong”) AI

Narrow AI

- **Brede** aanduiding van AI ;-)
- Ook wel “Weak” AI genoemd
- Veelal een specifieke taak waarvoor het getraind is
- Sneller en preciezer dan mensen
- Soms zelf-lerend
- Soms creërend (*generatieve* AI)
- Niet de denkwijze van een mens, geen “Theory of mind”

You
How do you feel today?

ChatGPT
As an AI, I don't have feelings or emotions like humans do. However, I'm here and ready to assist you with any questions or tasks you have! How can I help you today?

Voorbeelden Narrow AI

- robots (auto industrie)
- schaakcomputers
- spraak-, image-, handschrift- gezichts-herkenning
- search engines
- fraudeopsporing (belastingdienst, Chinese “overtreders”)
- spamfilters
- chat robots
- zelfrijdende auto’s

Narrow AI - generatieve AI

Binnen Narrow AI heb je de **Generatieve** AI.

Gaat om nieuwe content ***genereren***

Voorbeelden:

- OpenAi's ChatGPT, Google's Gemini (opvolger Bard), DeepSeek (code)
- OpenAI's Sora (text naar video)
- Suno (text naar muziek)
- Virtuele assistenten (Siri, Alexa, Copilot)
- Persoonlijke aanbevelingen (Netflix, Spotify, …)

Strong AI

Ook wel **General AI** (“AGI”) of **Deep AI** genoemd.

Hier zijn we nog niet. Enkele kenmerken:

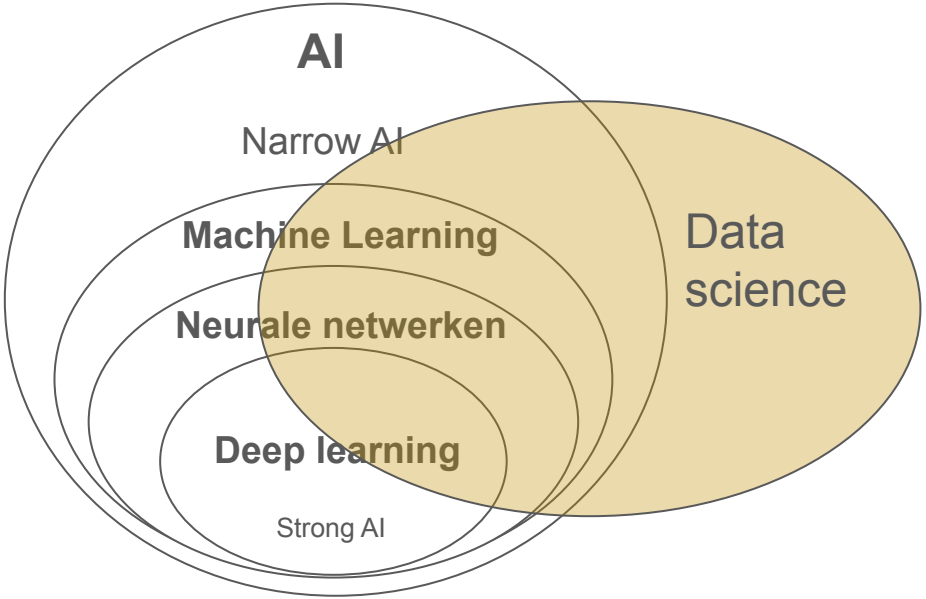
- Laat machines dingen doen en leren net als mensen
- Heeft “Theory of mind”

Voorbeelden:

- Terminator, iRobot ;-)

Verder bestaat ook de term: “Super AI” (beter dan mensen)

Data Science



Creëert waarde uit data.

Eigen vakgebied met meerdere disciplines, waaronder wiskunde, statistiek, informatica en informatiekunde

Data Science

Data Science zoekt patronen in data en kan voorspellingen doen. Daardoor:

- Beter inzicht in je data
- Betere beslissingen
- Betere planning (wat is de volgende stap?)

Enkele voorbeelden waar data science een rol kan spelen:

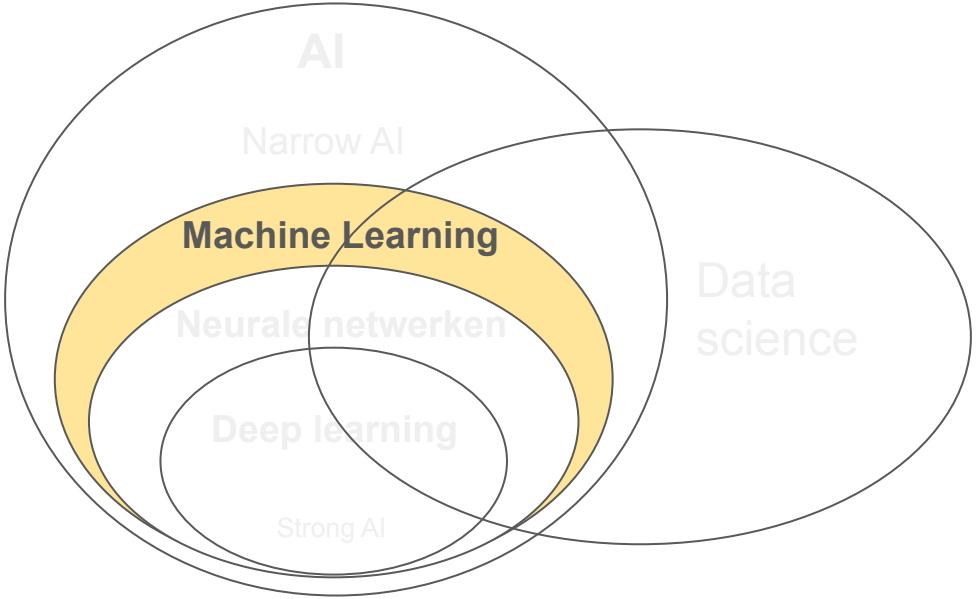
- voorspellen van verkiezingen
- route planning (de beste route vinden)
- winst/verlies verwachting
- vraag naar producten, voorraadbeheer

AI kan niet zonder Data Science

Data Science vs AI

	Data Science	AI
Wat is het?	Verzamelen, analyseren en produceren van bevindingen uit (veel) data	Implementatie van een voorspellend model wat een vorm van intelligentie vertoond (veelal om toekomstige gebeurtenissen of trends te produceren).
Doelen	Het vinden van patronen die verborgen zijn in (veel) data en deze gebruiken om beslissingen te maken.	Leer machines en computers om op een intelligente, mensachtige manier te handelen.

Machine learning



“Onderzoeksveld binnen AI dat zich bezig houdt met de ontwikkeling van technieken waarmee computers kunnen leren en voorspellen.”

Veel overlap met technieken gebruikt bij Data science.

Wat is Machine learning

Wanneer een algoritme geautomatiseerd leert van data spreek je van Machine learning.

Tevens kan met de opgedane kennis voorspellingen gedaan worden.

Modellen spelen een centrale rol: manieren om een computer van data te laten leren.

Zaken die een rol spelen bij een model:

- input data (zowel om het model te trainen als om te testen)
- output data (een voorspelling)
- het gekozen algoritme (bv. een lineair regressie algoritme (later meer))

Zaak is om voor jouw situatie (data) het juiste model (algoritme) te kiezen.

Waarom Machine learning? (waarom AI?)

Enkele redenen:

- Steeds meer data -> ML helpt structureren en overzicht verkrijgen
- Betere beslissingen
- Helpt bij patroonherkenning (fraudebestrijding) en trends (voorspellingen)
- Oplossen van complexe problemen
- Sneller en (veelal) accurater dan mensen

Zaken dus waar een mens tekort schiet

Veel raakvlakken met de technieken uit Data Science

- Verzamelen, opschonen, ordenen, formatteren, normaliseren van data
- Data distributie (“gaussian”, standaard deviatie, percentiel waarden, …)
- Regressie analyse - “schatten van waarden”
- Classificatie - “schatten van een (bekende) categorie” (man of vrouw, hond of kat, …)
- Clustering - “schatten van (onbekende) categorieën
- Model trainen / testen

Hoe verder:

- Niet beide (Data science en Machine Learning) uitwerken.
- We gaan kijken naar Machine Learning technieken/voorbeelden en leren dan tevens diverse aspecten van Data science kennen.
- “Learning by doing” (direct ook programmeer code gebruiken)

Drie deelgebieden binnen Machine Learning 1/3

Supervised (*begeleid*) learning

- *Gelabelde* data: van input is de gewenste output (“label”) bekend.
- **Doel**: een voorspellend model ontwikkelen dat de relatie tussen de input en output begrijpt en dit kan toepassen op nieuwe data
- Voorbeelden:
  - spam filteren
  - verkoopcijfers
  - taggen van facebook vrienden
  - diagnostiseren van ziektes, aandoeningen
- Modellen en algoritmen
  - **regressie** modellen (“schatten van waarden””)
    - Algoritmen: **lineaire regressie**, **polynome regressie**
  - **classificatie** modellen (“schatten van een (bekende) categorie” (man of vrouw, hond of kat, …”).
    - Algoritmen: **beslisbomen**, logistische regressie, Support Vector Machine (SVM), k-Nearest Neighbor (kNN)

Deelgebieden binnen Machine learning 2/3

Unsupervised (*onbegeleid*) learning

- Geen gelabelde data. Wel input, geen gewenste output.
- **Doel**: zelfstandig vinden van patronen, structuur in de data
- Je weet niet waar je naar moet zoeken in de data, geen gewenste uitkomst bekend
- Voorbeelden:
  - gerichte advertenties (webwinkels)
  - aanbevelingen (Netflix, Spotify)
  - fraudeopsporing
- Modellen en algoritmen
  - Clustering
    - Algoritmen: k-Means, Hierarchical Mean, Shift-Means
  - Patronen, associatie (“aanbevelingen”)
  - Reductie: versimpel complexe data

Deelgebieden binnen Machine learning 3/3

Reinforcement (*versterkend*) learning

- Zelfstandig leren om, mbv feedback (bv. “goed” of “fout”), een doel te bereiken
- Lijkt het meest op “menselijk leren”
- Kan beginnen zonder input
- Voorbeelden:
  - zelf rijdende auto’s
  - personaliseren van content (“click fuik” / “bubbels”)

Werkwijze

Geen eenduidig antwoord..

Afhankelijkheden van de beschikbaarheid en kwaliteit van de data, moeilijkheid van het probleem, onzekerheden, randvoorwaarden, gewenste resultaat

In het algemeen:

1. **Juiste meetmethode** in kaart brengen (onafhankelijke observaties?)
2. **Data**: Data bepalen, data verzamelen, data onderzoeken, data opschonen, data formatteren
3. **Splits** de data in training-data en test-data
4. **Train**: Kies model(len) (algoritme(s)) en pas toe met bekende data
5. **Evalueer**: Evalueer model(len) / algoritme(s), verbeter model
6. **Herhaal** train/evaluatie
7. **Test**: Pas model toe op test-data

Voorbeeld

Stel je hebt de volgende data:

```
students = {'hours':   [29,9,10,38, NULL,16,26,25,50,10,30,33,43,2,39,15,44 ,,29,41,15,  ,24,50],

            'results': [65,7, 8,76, NULL,23,56,  ,100,3,74,48,73,0,62,37,74 ,,40,90,42,56,58,100]}
```

Bepaal het lineaire (“rechtlijnige”) verband in deze data.

Dan bv. te “schatten” wat de y is als x = 25

Eerst maar opschonen

Vragen^H^H^H^H^H^HBier?