

The background features a series of concentric circles in a light gray color, centered on the page. In the four corners, there are stylized circuit-like lines in a light blue color, with small circles at the end of the lines, resembling a network or data flow.

DATA SCIENCE

JENS BAETENS

The background features a series of concentric, light gray circles centered on the slide. Overlaid on these are stylized, light blue circuit-like lines with small circles at their endpoints, located in the corners of the slide.

WAT IS DATA SCIENCE?



<https://www.menti.com/d7mr61nrs5>



WELKE VOETBALPLOEG GAAT WINNEN?

Eerste stap om deze vraag te antwoorden is om data te gaan zoeken

Naar wat voor data moet er gezocht worden?

Over welke periode moet er data gezocht worden?

Welke databronnen kunnen gebruikt worden?

VERWERKING VERZAMELDE DATA

15 → Huis

Spoor fouten op: verkeerde ingave, labelling, ruis..

Zorg dat de privacy in de dataset gegarandeerd is

Reduceer het volume door onnodige data, duplicaten, ... te verwijderen

Zoek naar reeds bestaande patronen in de data.

1/1/2001 | Gent
1-1/01 | Gonol
02-01/2001 | Gent
... | Gent
... | Gent

MODEL OPBOUWEN VOOR DE DATA

Kies 1 of meerdere machine learning technieken

Train met de verzamelde data

Valideer je resultaten en vergelijk verschillende technieken.

Test error

COMMUNICEER EN VISUALISEER

Hoe goed werkt ons model?

Wat kunnen we eruit leren?

Op maat van je doelpubliek:

- Grote publiek
- Bedrijfsleiders
- Politici

Landen	Odds*
Engeland	7.50
Belgie	8.00
Frankrijk	4.50
Nederland	15.00
Spanje	8.50
Duitsland	11.00
Portugal	8.50
Italie	7.50

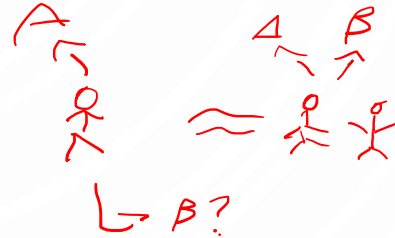
DEFINITIE

Data Science is het vermogen om de juiste data

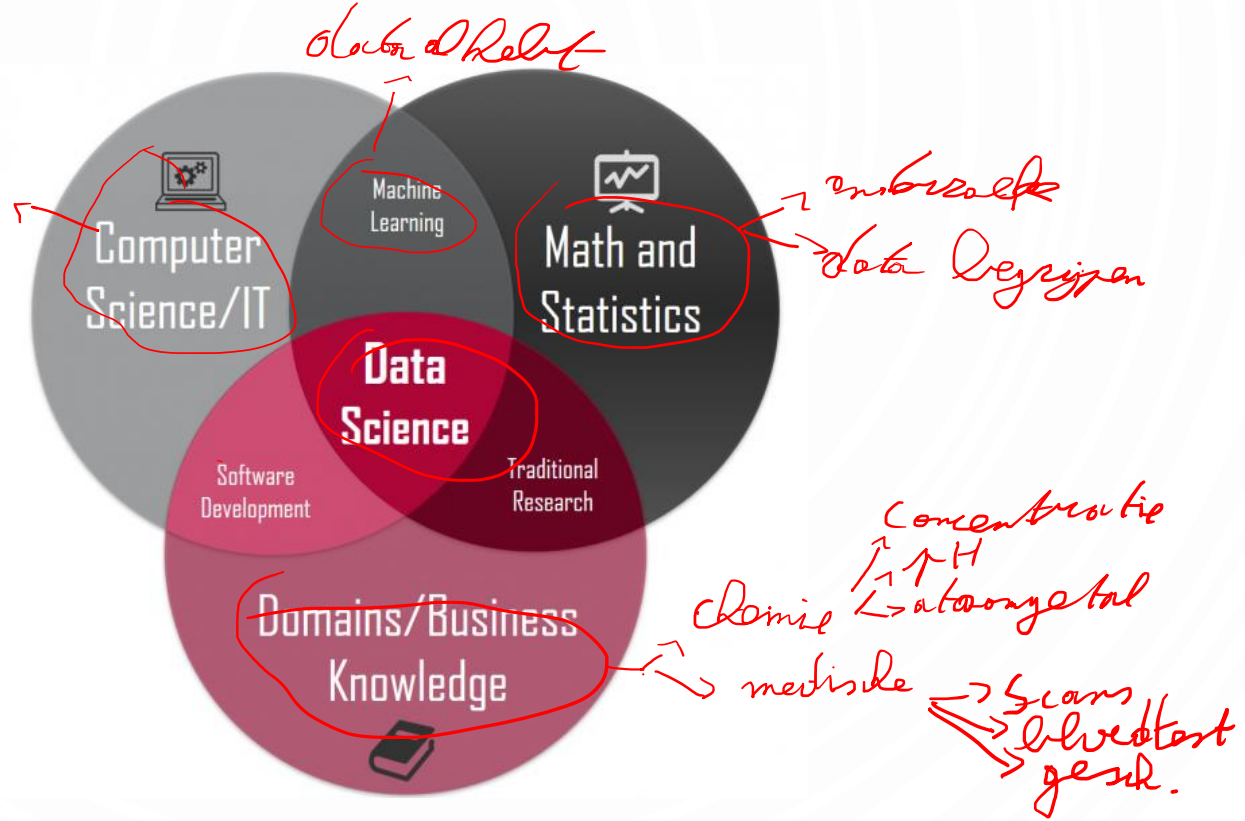
- te selecteren
- te begrijpen
- te verwerken
- te visualiseren

met als doel de waarde uit de data te halen.

↳ onderhoud → Aanraden van films
↳ planning

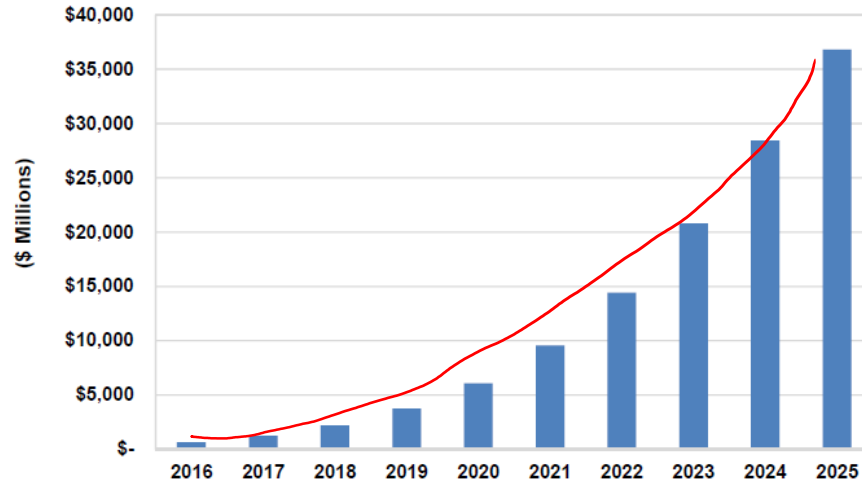


python
SQL
NoSQL
{
++
}
:
:



NOOD AAN DATASCIENCE?

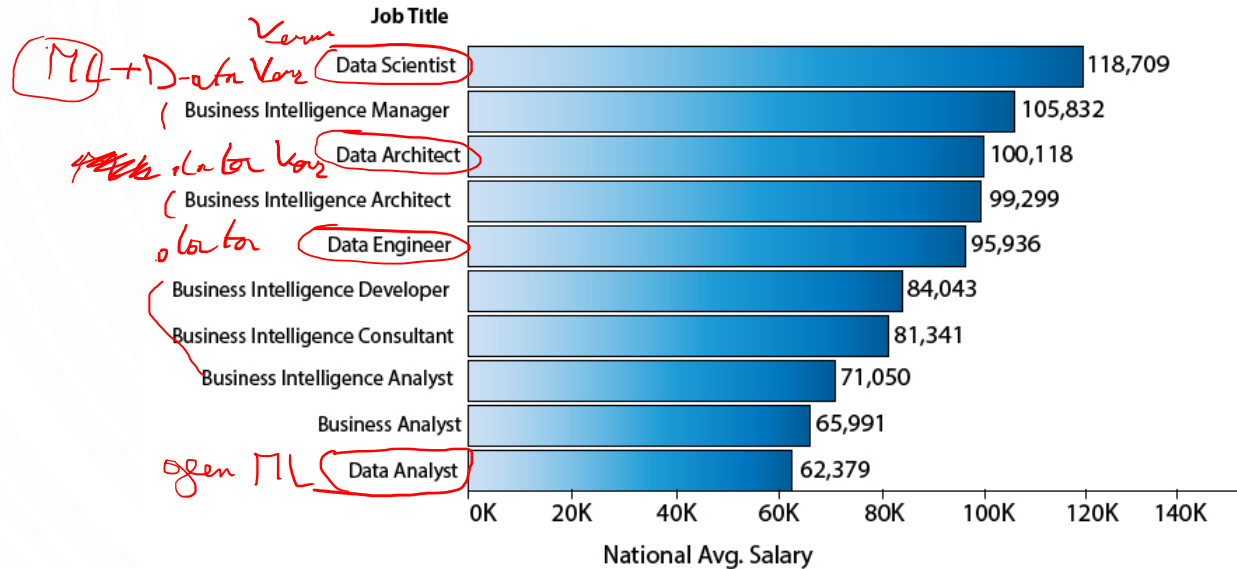
Chart 1.1 Artificial Intelligence Revenue, World Markets: 2016-2025



(Source: Tractica)

Source: <https://www.top500.org/news/market-for-artificial-intelligence-projected-to-hit-36-billion-by-2025/>

NOOD AAN DATASCIENCE?



Source: <https://data-flair.training/blogs/why-learn-data-science/>

lener worabrigs Contract analysis

auto's (Object detection and classification -
avoidance, navigation



(Object identification, detection,
classification, tracking from geospatial
images

Automated geophysical feature detection

(Text query of images

(Content distribution on social media

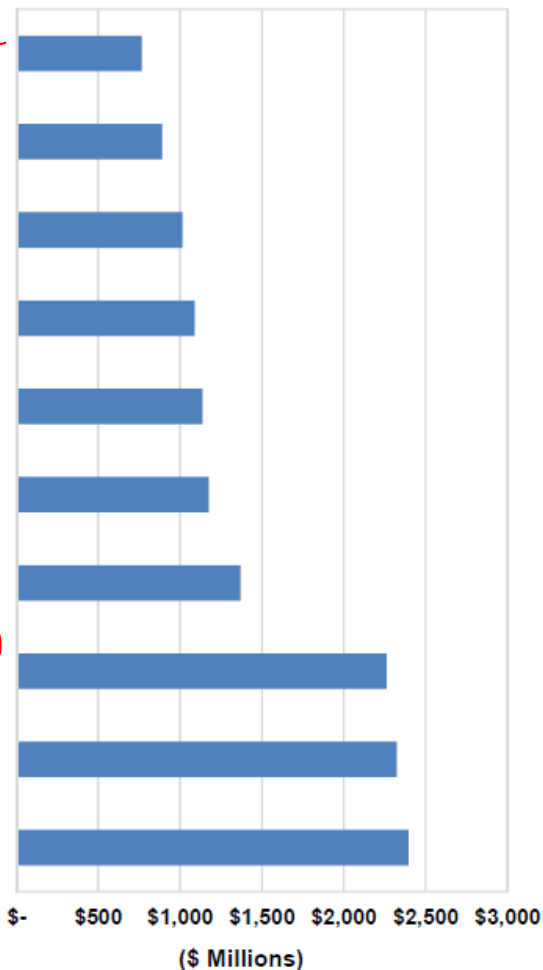
(Predictive maintenance

(Efficient, scalable processing of patient data)

Static image recognition, classification, and
tagging

leuwan

(Algorithmic trading strategy performance
improvement



MORE EXAMPLES OF USE CASES

Financiële sector

- Fraude detecteren
- Risico's inschatten
- Beleggen op de beurs



MORE EXAMPLES OF USE CASES



Gezondheidszorg

- Bij een aantal ziektes betere diagnoses dan menselijke dokters
- Remote opvolgen door middel van chatbots, remote sensors, ...
- Ontlast de dokter voor behandeling en opvolging patiënten



MORE EXAMPLES OF USE CASES

Marketing

- Gepersonaliseerde reclame
 - Klanten binding
 - Type klant detecteren en anders behandelen
- soort reclame, opvolging

email . . .
telefonisch . . .
whatsapp



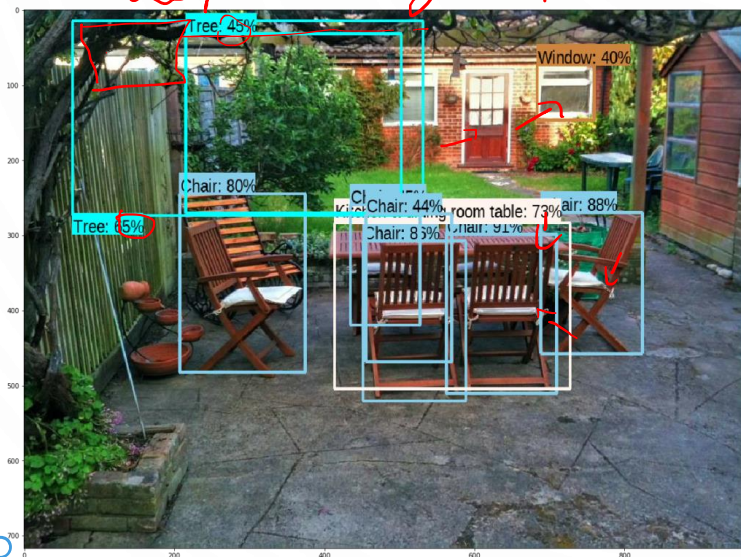
COMPUTER VISION

CUDA \rightarrow GPU

\rightarrow ML training



Deep learning - CNN



A person riding a motorcycle on a dirt road.



A group of young people playing a game of frisbee.



A herd of elephants walking across a dry grass field.

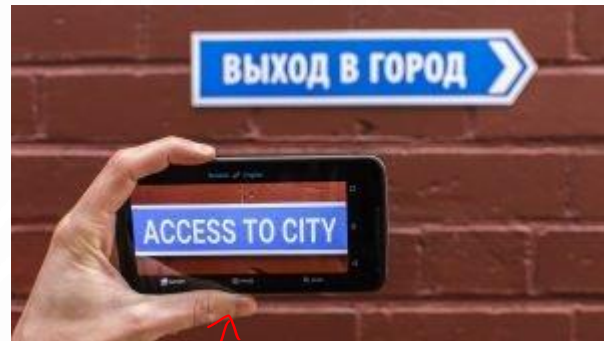


Photo \rightarrow Text (Translate) Model 1
Text \rightarrow photo Model 2
Model 3

Machine learning use cases in retail



(Demand Prediction)



Price Formation



Logistics



(Merchandizing ,



(Personalized Offers)



Fraud Detection

Wholesale → Normal → Retail
→ minder adv



(Churn Prediction



Location
Optimization



Sentiment Analysis



Document Work
Automation

↳ Tweets
↳ Facebook
↳ Klacht?
↳ Delen ...

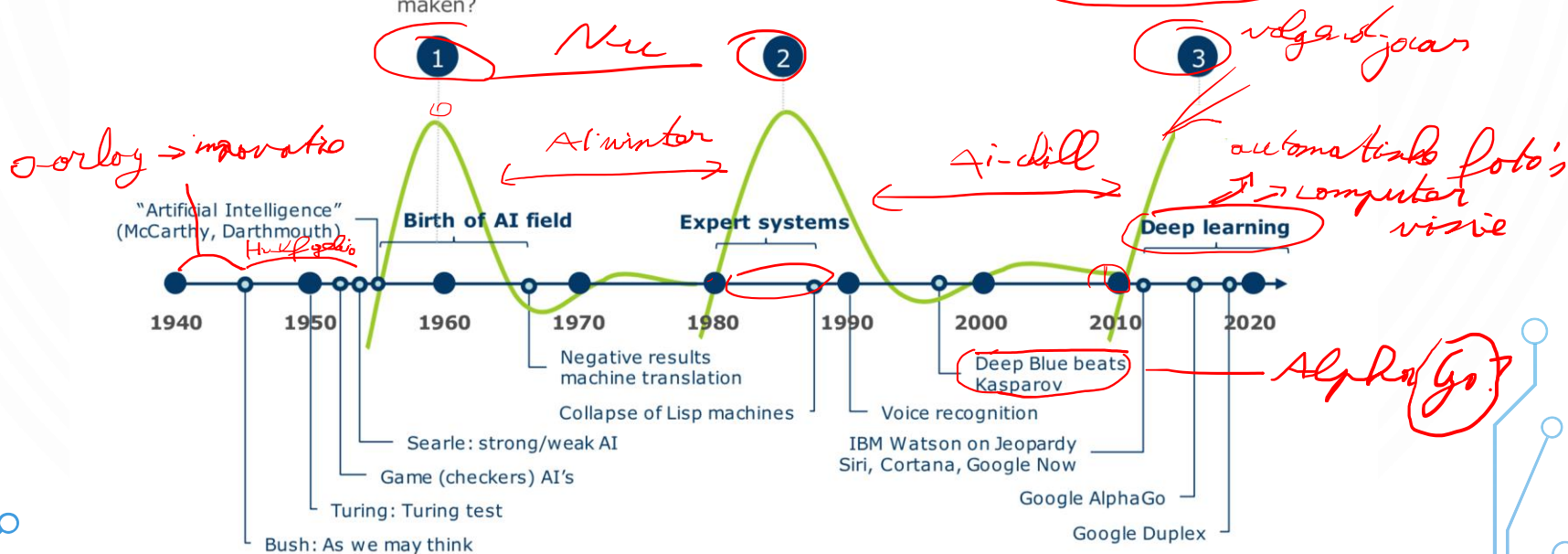
GESCHIEDENIS

als een mens
(AI vanuit psychologisch /
neurologisch perspectief:
kunnen we een
mechanisch brein
maken?)

*Computers - Data is
↑ hot problem
doelgericht*

AI vanuit cognitief
perspectief: kunnen we
een machine maken die
kan redeneren?

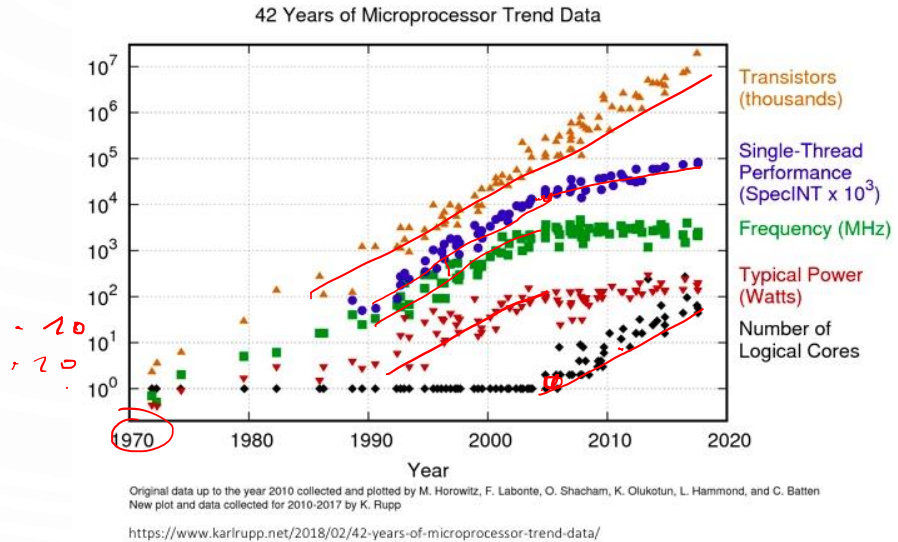
AI vanuit dataperspectief:
kunnen we een machine
taken leren op basis van
grote hoeveelheden data?



WAAROM NU?

Veel snellere hardware

- Custom designed GPU's



¹⁰ B - ¹⁰ B - ⁶ GB - ¹² TB - ¹⁵ PB - ¹⁸ ExB - ²¹ ZB

WAAROM NU?

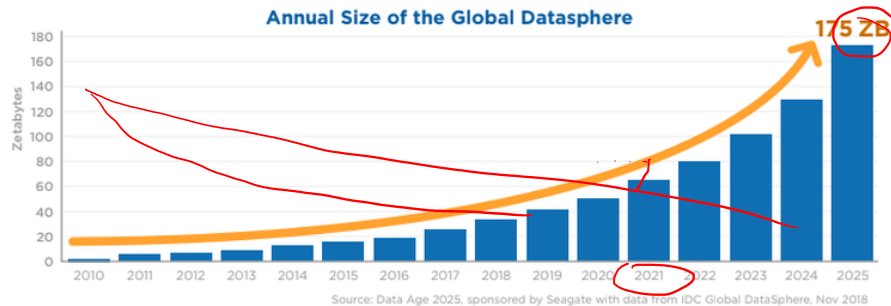
Veel meer data beschikbaar

- Internet of Things

- Social Media

- Prijs van dataopslag gezakt

Figure 1 - Annual Size of the Global Datasphere



Source: Data Age 2025, sponsored by Seagate with data from IDC Global DataSphere, Nov 2018

2021 This Is What Happens In An Internet Minute



Created By:
@LoriLewis
@OfficiallyChadd

WAAROM NU?

Betere algoritmes

Open source frameworks

- Python, PyTorch, Tensorflow, ... *Sklearn* → *ML-technieken*
Neurale Netwerken

ENGAGEMENT PROCESS

Step 1: Build the Data Model



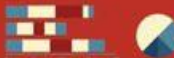
Step 2: Define The Report



Step 3: Generate SQL commands



Step 4: Create Report



Data Warehouse

The data warehouse is a "schema-on-load" approach because the data schema must be defined and built prior to loading data into the data warehouse. Without an underlying data model, the BI tools will not work.

Business Intelligence Questions

What happened?

Descriptive Analysis

Standard Reporting



Business Intelligence Analyst

VS

Data Science Questions

Why? What will happen?
What should I do?

Predictive Analysis

Prescriptive Analysis



Data Scientist

DIFFERENCE BETWEEN BUSINESS INTELLIGENCE AND DATA SCIENCE

CHARACTERISTICS

Focus	Reports, KPIs, trends	Patterns, correlations, models
Process	Static, comparative	(Exploratory, experimentation, visual)
Data Sources	Pre-planned, added slowly	On the fly, as-needed
Transform	Up front, carefully planned	In-database, on-demand, enrichment
Data quality	Single version of truth	(“Good enough,” probabilities)
Data model	Schema on load	Schema on query
Analytics	Retrospective, Descriptive	<u>Predictive</u> , Prescriptive, Preventative

ENGAGEMENT PROCESS

Step 1: Define Hypothesis to Test



Step 2: Gather Data



Step 3: Build Data Model



Step 4: Explore the Data



Step 5: Build and Refine Analytic Models



Step 6: Ascertain Goodness of Fit



repeat