

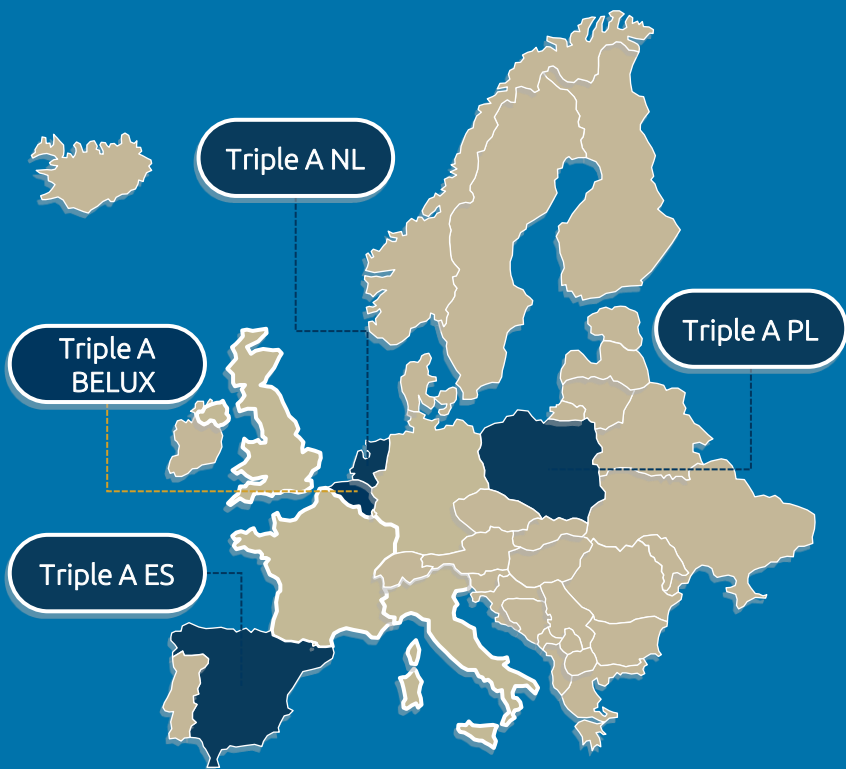
PE Data Sessie

Sessie 2



TRIPLE A
RISK FINANCE

14 Mei 2024
Data & Analytics Team
Triple A - Risk Finance



Triple A Offices

150+ consultants

Serving the vast majority of
European top 20 insurers

Part of Redmore, part of House of HR
(57k people at work, EUR 2+ Bn turnover)



100+

Clients

170

Consultants
(mostly in the Netherlands)

4

Offices (Amsterdam,
Barcelona, Brussels, Warsaw)



Insurance



Pensions



Banking



Risk & Strategy
Consulting



Data & Analytics



Actuarial Technology

Triple A - Risk Finance

is an independent and innovative consultancy
company specialised in actuarial, risk
management and data science services.

We are open and informal, and at the same
time secure and focused on results.

Our clients include insurers, banks, pension
funds, pension providers and (large) companies.

Services

Consulting
Modelling
Certification
Project Management
Operational Support
Outsourcing
Key Functions
Training

Sprekers



Ronald van de Langenberg

2012 - heden
Partner InAudit BV
& Dataverbeteraar



Pieter Stel

2023 - heden
Practice Lead
Data & Analytics



Axel Pison

2019 - heden
Senior Consultant
Data & Analytics



Nadine Konstanti

2022 - heden
Senior Consultant
NFR & AI

Agenda

1

Opening & Recap

2

AI & Machine Learning

3

ML Pricing

Pauze

4

Data Management & Governance

5

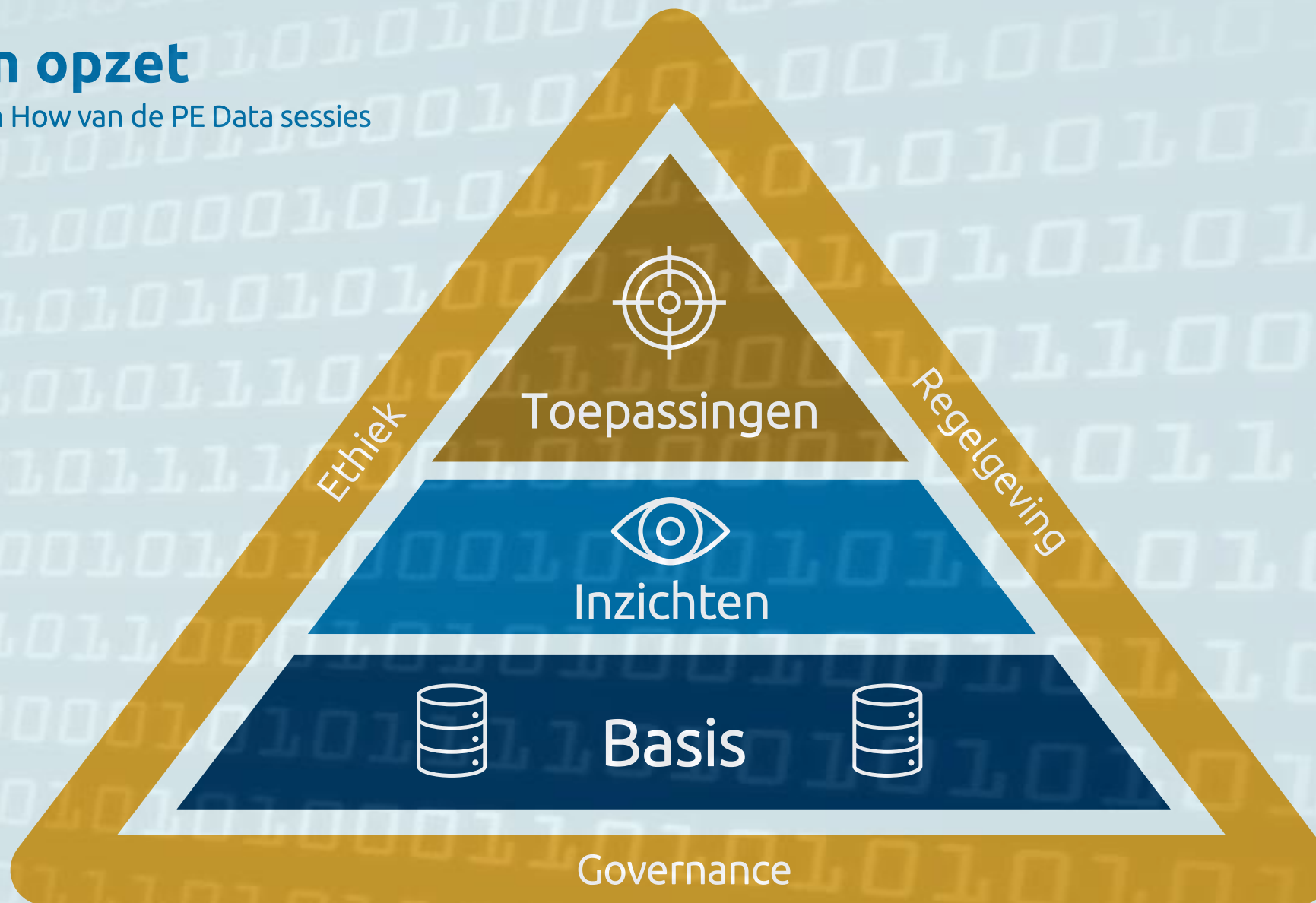
AI risico's, regelgeving en ethiek

6

Afsluiting

Scope en opzet

De Why, What en How van de PE Data sessies



"Veni Vidi Vici": "Ik kwam voor de data - Ik zag de data - Ik overwon met de data"



Recap PE sessie 1 - 19 maart

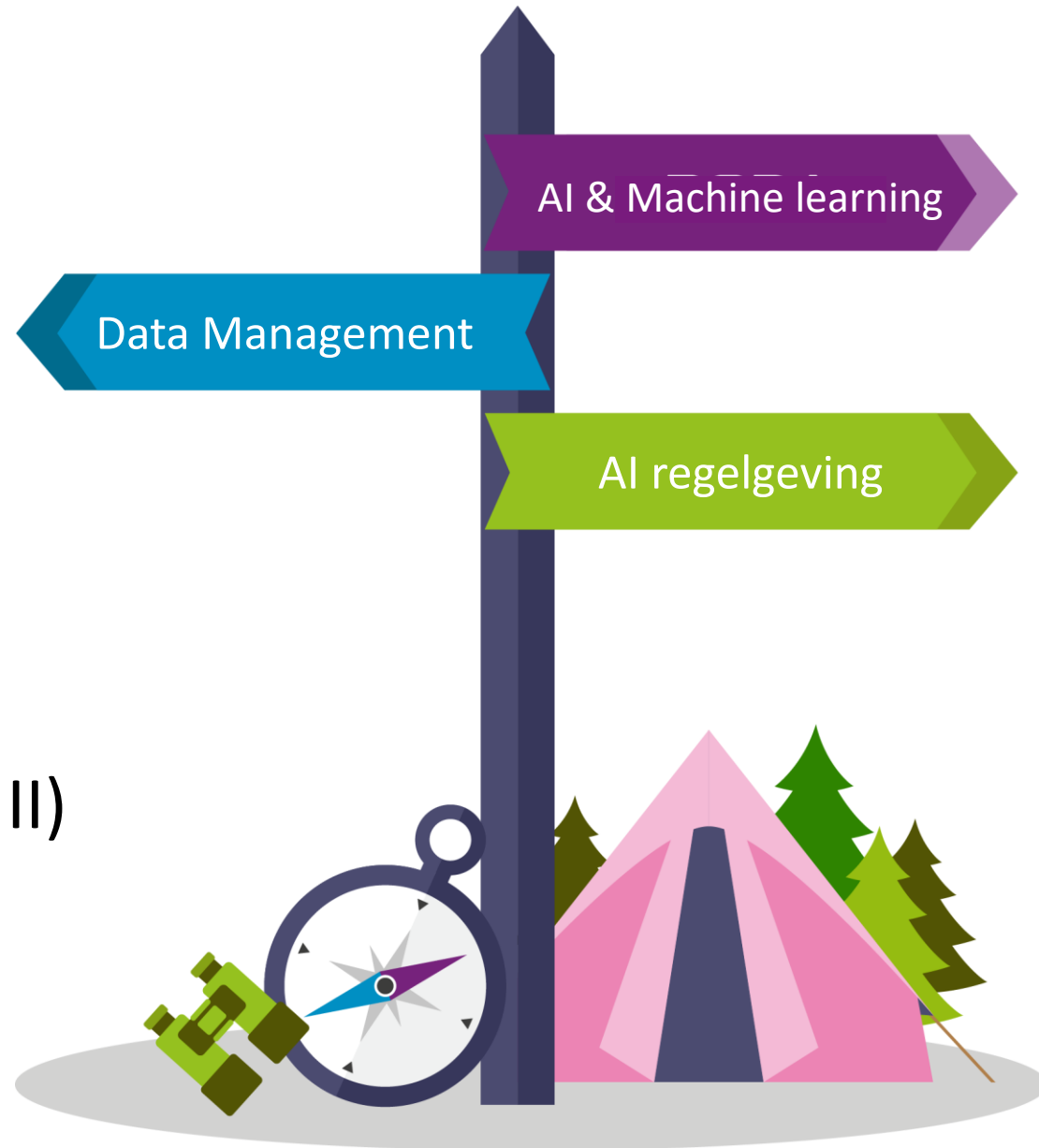


Agenda

- 1 Opening & Overview
- 2 Trends & Ontwikkelingen
- 3 Databronnen
- Pauze*
- 4 Datakwaliteit
- 5 Architectuur
- 6 Data Visualisatie

DATA DETECTIVES (deel II)

14 mei 2024



Metric mishap caused loss of NASA orbiter

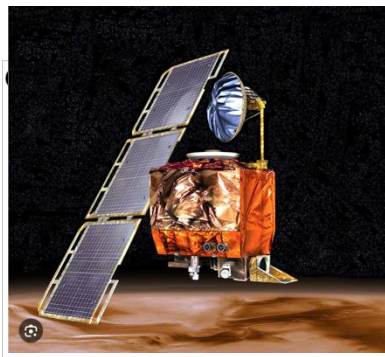
September 30, 1999
Web posted at: 4:21 p.m. EDT (2021 GMT)

In this story:

[Metric system used by NASA for many years](#)

[Error points to nation's conversion lag](#)

[RELATED STORIES, SITES](#) ↓



NASA's Climate Orbiter was lost September 23, 1999

By Robin Lloyd
CNN Interactive Senior Writer

(CNN) -- NASA lost a \$125 million Mars orbiter because a Lockheed Martin engineering team used English units of measurement while the agency's team used the more conventional metric system for a key spacecraft operation, according to a review finding released Thursday.

The units mismatch prevented navigation information from transferring between the Mars Climate Orbiter spacecraft team in at Lockheed Martin in Denver and the flight team at NASA's Jet Propulsion Laboratory in Pasadena, California.

Mars Climate Orbiter

Artikel [Overleg](#)

[Lezen](#) [Bewerken](#) [Brontekst bewerken](#)

Uit Wikipedia, de vrije encyclopedie

De **Mars Climate Orbiter** (MCO), oorspronkelijk **Mars Surveyor '98 Orbiter** geheten, was een mislukte [Amerikaanse](#) missie naar de planeet [Mars](#) in 1998-1999.

De Mars Climate Orbiter was een onbemande [ruimtesonde](#) die op 11 december 1998 gelanceerd werd naar Mars door [NASA](#) als onderdeel van het **Mars Surveyor '98**-programma. Het doel van de ruimtesonde was het [klimaat](#) van Mars, de [atmosfeer](#) en veranderingen in het planeetoppervlak te bestuderen. Tevens zou de Mars Climate Orbiter dienen als communicatieschakel tussen het controlecentrum van NASA en de [Mars Polar Lander](#).

Op 23 september 1999, terwijl de Mars Climate Orbiter een [baan](#) rond Mars aannam, verloor NASA het contact met de ruimtesonde, die vermoedelijk vernietigd werd in de atmosfeer. Het verlies van de ruimtesonde bleek veroorzaakt door menselijke fouten. De hoofdreden was een rekenfout: de berekeningen voor de stuurmotoren werden uitgevoerd met een eenheid van het [Brits-Amerikaanse maatsysteem](#), [pond-seconden](#) (lbfxs), in plaats van [newton](#)-seconden (Nxs) zoals in de software was ingeprogrammeerd. Hierdoor benaderde de ruimtesonde Mars veel te laag en brandde ze op in de atmosfeer.^{[1][2]}

Zoek op naam en omschrijving				Zoeken
€ 16.351,82				
Datum	Omschrijving	Type	Bedrag (€)	
12-12-2013		BA	4,67	Af
12-12-2013		BA	31,00	Af
12-12-2013		BA	7,26	Af
12-12-2013		BA	6,67	Af
12-12-2013	Naam:DBGA Omschrijving:WOONKOSTENBIJDRAGE	VZ	15.500,00	Bij
12-12-2013		GM	20,00	Af
12-12-2013		BA	6,87	Af

Betalen

- › Overschrijven
- › Af- en bijschrijvinge
- › Overzicht saldo
- › Verzendlijst
- › Incasso's

NL ^

9000 Amsterdammers krijgen miljoenen door fout

Negenduizend Amsterdammers hebben donderdag miljoenen euro's te veel woonkostenbijdrage van de gemeente gestort gekregen. AT5 heeft sinds donderdagavond meerdere meldingen gekregen van mensen die hun geluk niet op kunnen.

Amsterdammers met een laag inkomen en te hoge woonkosten ontvangen de woonkostenbijdrage van de gemeente als extraatje.

De bedragen die op de rekeningen zijn gestort variëren. Een gelukkige meldt dat hij in plaats van 155 euro maar liefst 15.500 euro op zijn rekening gestort heeft gekregen. Een ander zegt zelfs 23.000 euro te hebben ontvangen.

Foutje

Het gaat om een fout bij de Dienst Belastingen Gemeente Amsterdam (DBGA). De dienst, die midden in een verhuizing zit, probeert met man en macht het geld terug te krijgen. DBGA heeft inmiddels een crisisteam geformeerd om het geld terug te halen. Een aantal van de negenduizend heeft het geld echter al opzij gezet.

Het totaal aan uitgekeerde bijdragen komt waarschijnlijk boven de 100 miljoen euro uit. De dienst laat weten dat er iets is misgegaan met een bestand. Wat er precies is misgegaan wordt nog uitgezocht, maar het is waarschijnlijk een kwestie van punten en komma's.



▲ © YouTube

Waar kennen we Reaal ook alweer van? Foutje, bedankt!

We kennen de reclamefilmpjes van verzekeraar Reaal allemaal. Rijk de Gooijer was in die commercials als meneer Van Looijen te zien samen met presentator Bert Kuizenga. Met de tekst 'foutje bedankt' kwam Van Looijen stevast terug op zijn dubieuze schademelding. Bekijk hier de filmpjes terug.

[Military Life](#)

How a Misplaced Decimal Point Nearly Took Down Spain's Newest Submarines



(Courtesy of Navantia)

Military.com | By [Blake Stilwell](#)
Published April 29, 2021



Listen to this article now
Powered by [Trinity Audio](#)

00:00



03:36

The business of building submarines can literally be a race to the bottom. That's not a metaphor; that's just what submarines do. But submarines are, of course, required to do a lot more than sink below the surface.

In the United States Navy, some submarines are used as part of our nuclear triad, carrying ballistic cruise missiles to the enemy. Other subs exist to gather intelligence. One thing is clear: All American submarines are expected to resurface.



The Spanish submarine Isaac Peral has experienced some unforeseen issues. (Spanish Navy)

The first in the S-80's class was commissioned in 2003, but economic trouble in Spain delayed its design and construction by a decade. In that time, engineers noticed a critical design flaw: The boat was an estimated 100 tons heavier than they expected. One of the designers, it turns out, [put a decimal point in the wrong place](#) somewhere down the line.

This meant the sub, once submerged, may never make it back to the surface, which is (again) very important for submarines and their crews.

When data gives the wrong solution



TREVOR BRAGDON

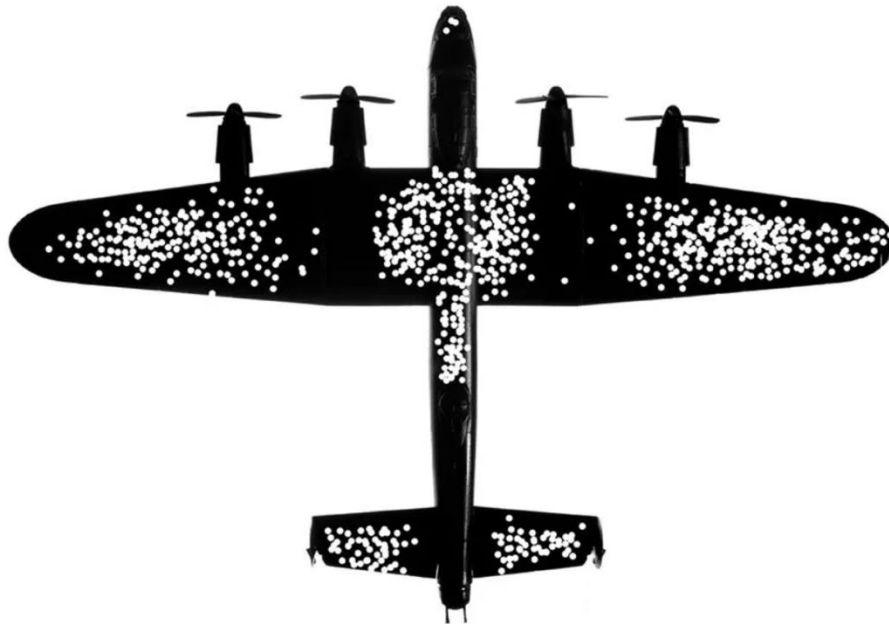
SEP 07, 2017



1



Share



ABRAHAM WALD'S PROBLEM SOLVING LESSON IS TO SEEK WHAT'S NOT THERE

by: **Al Williams**

26 Comments

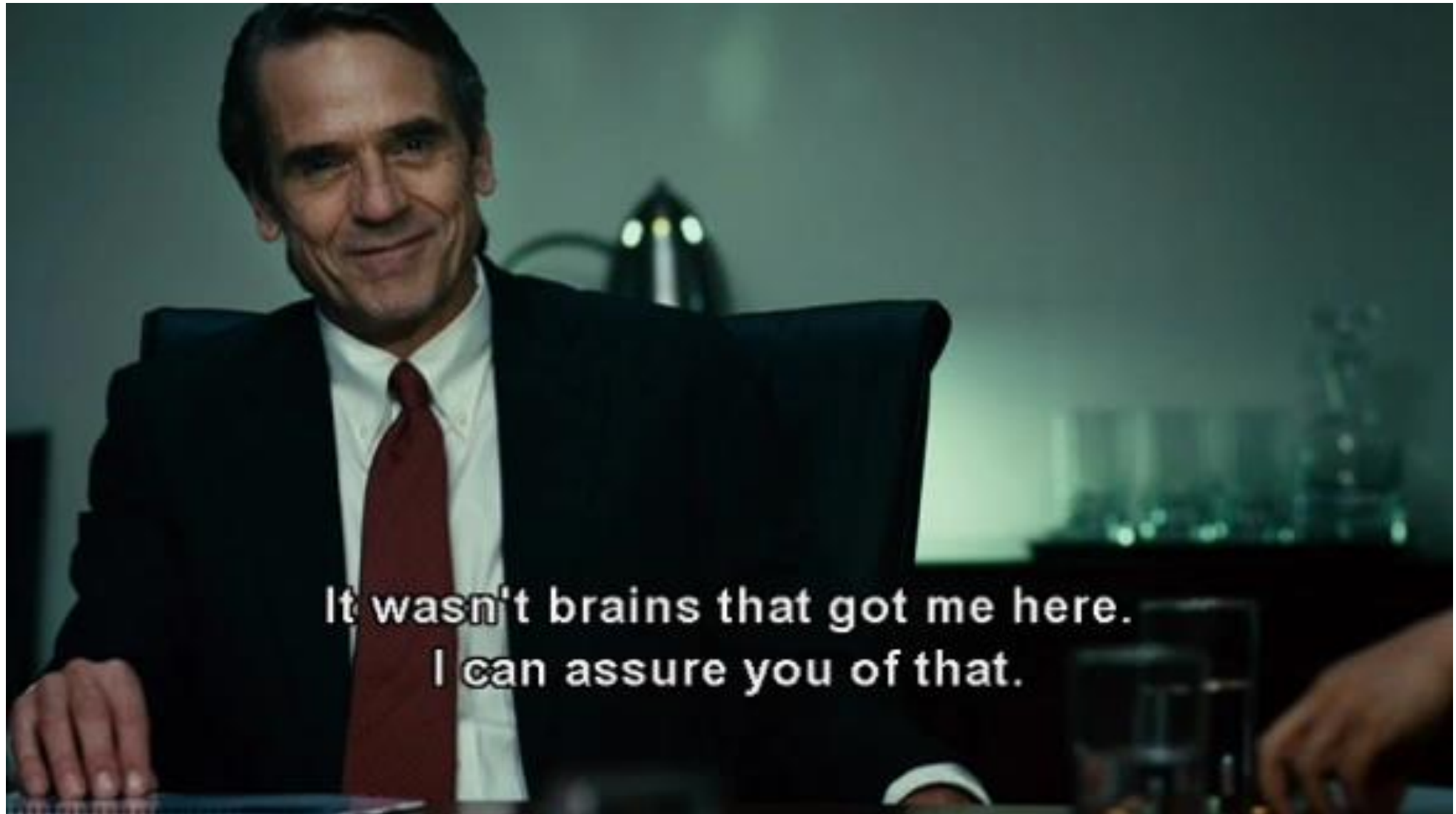


August 1, 2019



You may not know the name Abraham Wald, but he has a very valuable lesson you can apply to problem solving, engineering, and many other parts of life. Wald worked for the Statistical Research Group (SRG) during World War II. This was part of a top secret organization in the United States that applied elite mathematical talent to help the allies win the war. Near Columbia University, mathematicians and computers — the human kind — worked on problems ranging from how to keep an enemy plane under fire longer to optimal bombing patterns.

One of Wald's ways to approach problem was to look beyond the data in front of him. He was looking for things that weren't there, using their absence as an additional data point. It is easy to critique things that are present but incorrect. It is harder to see things that are missing. But the end results of this technique were profound and present an object lesson we can still draw from today.



Data in sport: soms werkt het



In Audit

Netflix

Netflix weet meer dan u denkt en is niet bang om dat te gebruiken



De nieuwe Netflix-serie 'Dark'. Beeld netflix

Krijgt u op Netflix ook suggesties voor films en series waarvan u denkt dat u ze nooit van uw leven zult bekijken? Als de ultieme strategie van de streamingdienst ook maar een beetje werkt, zult u dat binnenkort wél doen.

SAM FEYS 12 december 2017, 17:52

'Navigatiesysteem zorgt voor belabberd niveau topografische kennis'

20 augustus 2012 13:31:24, door Robert Hagen, 0 reacties



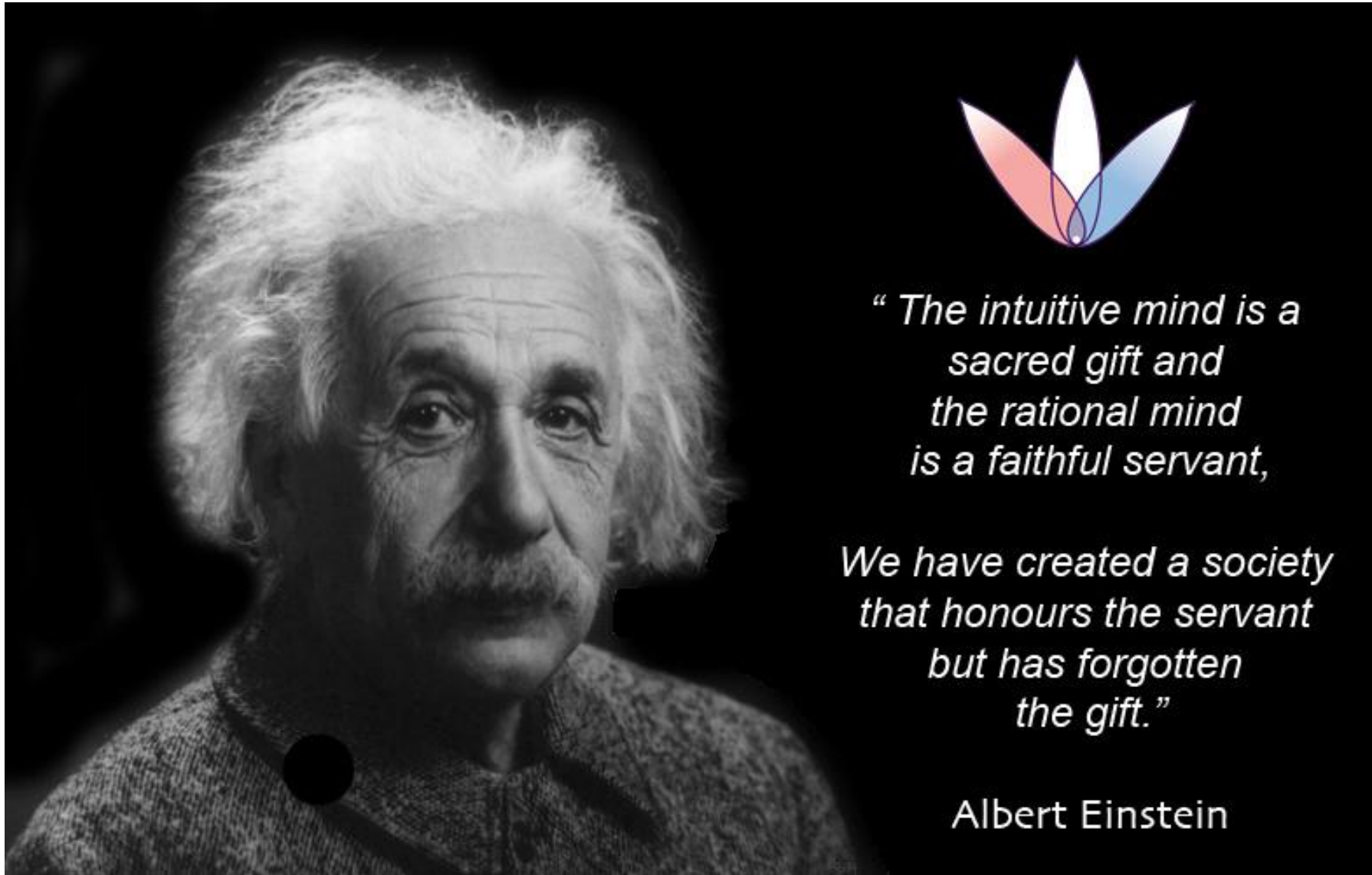
Door de enorme populariteit van navigatiesystemen daalt de topografische kennis van veel automobilisten tot een bedenkelijk niveau. Dat zou blijken uit onderzoek van Garmin-dochter Navigon, zelf een fabrikant van navigatiesystemen. Van de bestuurders die een navigatiesysteem gebruiken, zegt meer dan eenderde dat zijn of haar topografische kennis is afgenomen. Zo kan bijna tweederde van de respondenten bijvoorbeeld Den Bosch en Groningen niet meer aanwijzen op de kaart.

Opmerkelijk is dat automobilisten uit de Noordelijke regio een betere topografische kennis blijken te hebben dan de rest van de Nederlanders.

Bijna een kwart van de bestuurders met navigatie zet het navigatiesysteem ook aan terwijl ze de weg naar de bestemming al weten. "Ook al worden mensen nog wel eens per ongeluk een boerenerf opgestuurd: de meeste automobilisten kunnen niet meer zonder navigatie", zegt Chris Merrell, Hoofd EMEA Marketing Communicatie bij Garmin. "Mensen zien hun navigatiesysteem inmiddels als een vertrouwde metgezel. Automobilisten geven aan hun navigatiesysteem niet alleen te gebruiken voor de weg- en file-informatie, maar ook voor de gezelligheid", aldus Chris.

Overigens biedt het gebruik van een navigatiesysteem nog geen garantie dat je probleemloos op je bestemming aankomt. Zo blijkt dat 62 procent van de bestuurders mét een gps-systeem nog steeds wel eens de weg kwijt raakt.

Ontwikkel je “niet-pluis” intuïtie





Redenen om afscheid te nemen van Excel:

1. Data integriteit
2. Versiebeheer en audit trail
3. Beveiliging
4. Samenwerking
5. Workflow (eliminieren handmatige handelingen)
6. Integratie met andere dataflows
7. Rapportage & analyse
8. Rekenen met zeer grote datasets



Tom Veerman

Partner

Redenen om Excel te omarmen:

1. Flexibiliteit en snelle problem-solving
2. Ideaal voor ad-hoc analyses
3. Kosten-efficiënt
4. Controle over je tool en data
5. Geen afhankelijkheid van IT teams
6. Beschikbaarheid (offline/offsite)
7. Data gaan 'door je handen'
-> Méér 'niet plus' gevoel



cliënt:
observator:
datum:

Data Quality Engine

Configuration

Run

Reporting

About



TRIPLE A
RISK FINANCE



Reporting

STATUS



CDE	Completeness	Uniqueness	Validity	Accuracy	Consistency
Purpose of the loan					
Address					
Name					
Employment status					
Loan ID					
Interest rate					
Default2					
Default					
Assets					
Loan term (months)					
Number of days past due					
Loan amount					
Date of birth					
LTV					
Credit score					
Income (monthly)					
Monthly payment amount					



Data detectives:

Ontwikkel je 'niet pluis' gevoel bij (big) data, uitkomsten van data, algoritmen en andere calculaties !

Agenda

1

Opening & Recap

2

AI & Machine Learning

3

ML Pricing

Pauze

4

Data Management & Governance

5

AI risico's, regelgeving en ethiek

6

Afsluiting

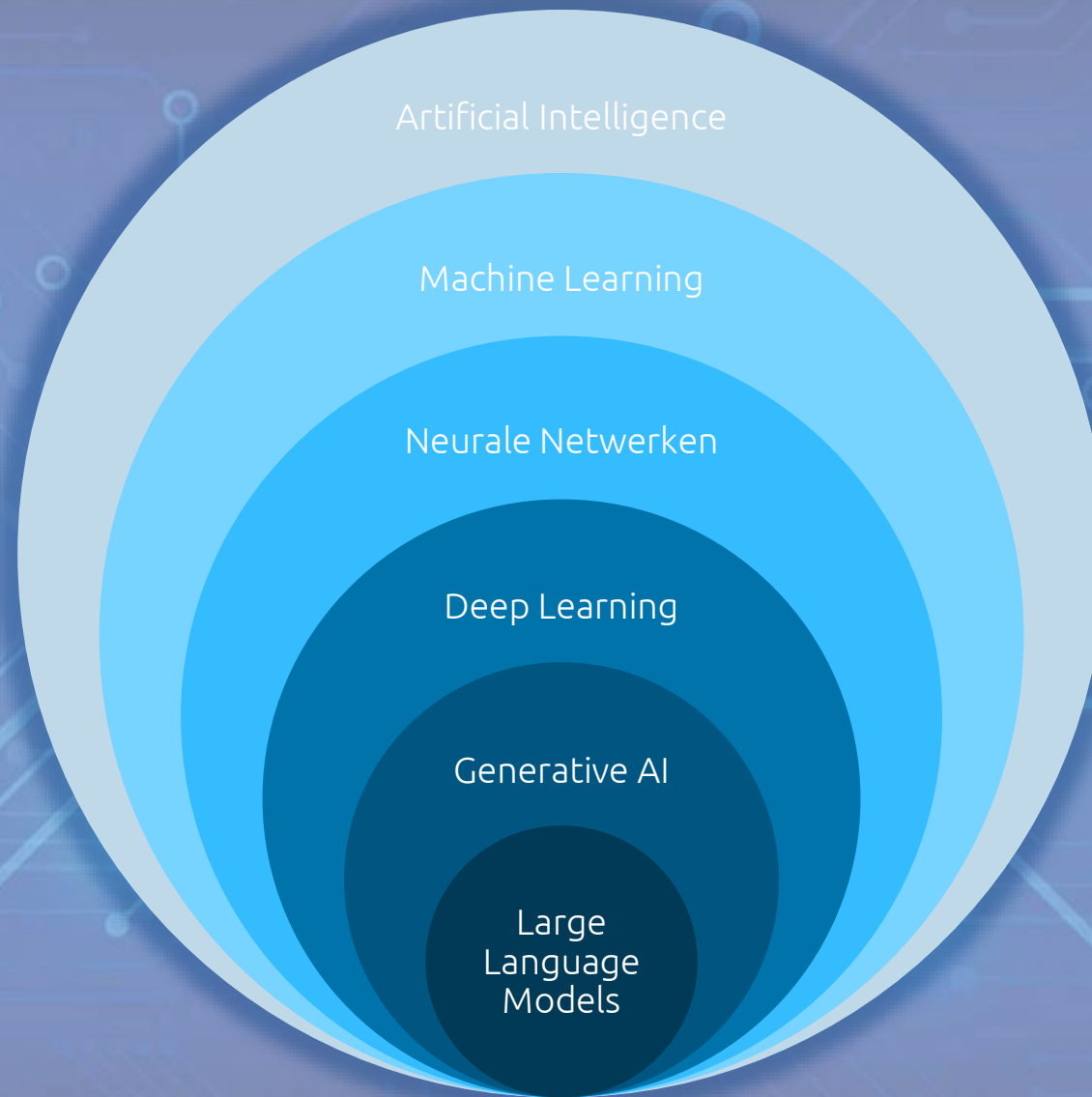
Wat is AI? Wat is ML?

“Artificial Intelligence (AI) is de tak van informatica die zich bezig houdt met het ontwikkelen van computersystemen die in staat zijn om taken uit te voeren die normaal gesproken menselijke intelligentie vereisen”



“Machine Learning is een tak van kunstmatige intelligentie (AI), waarbij computersystemen worden geprogrammeerd om te leren en zichzelf te verbeteren door ervaringen en data-analyse, zonder expliciete instructies te ontvangen”

Het AI-domein



De geschiedenis van AI



Het begin
1950-1970



AI Winter
1970-1995



Back on track
1995-2020



AI Boom
2020-heden



De eerste opmars van AI



Het begin
1950-1970

1966
ELIZA

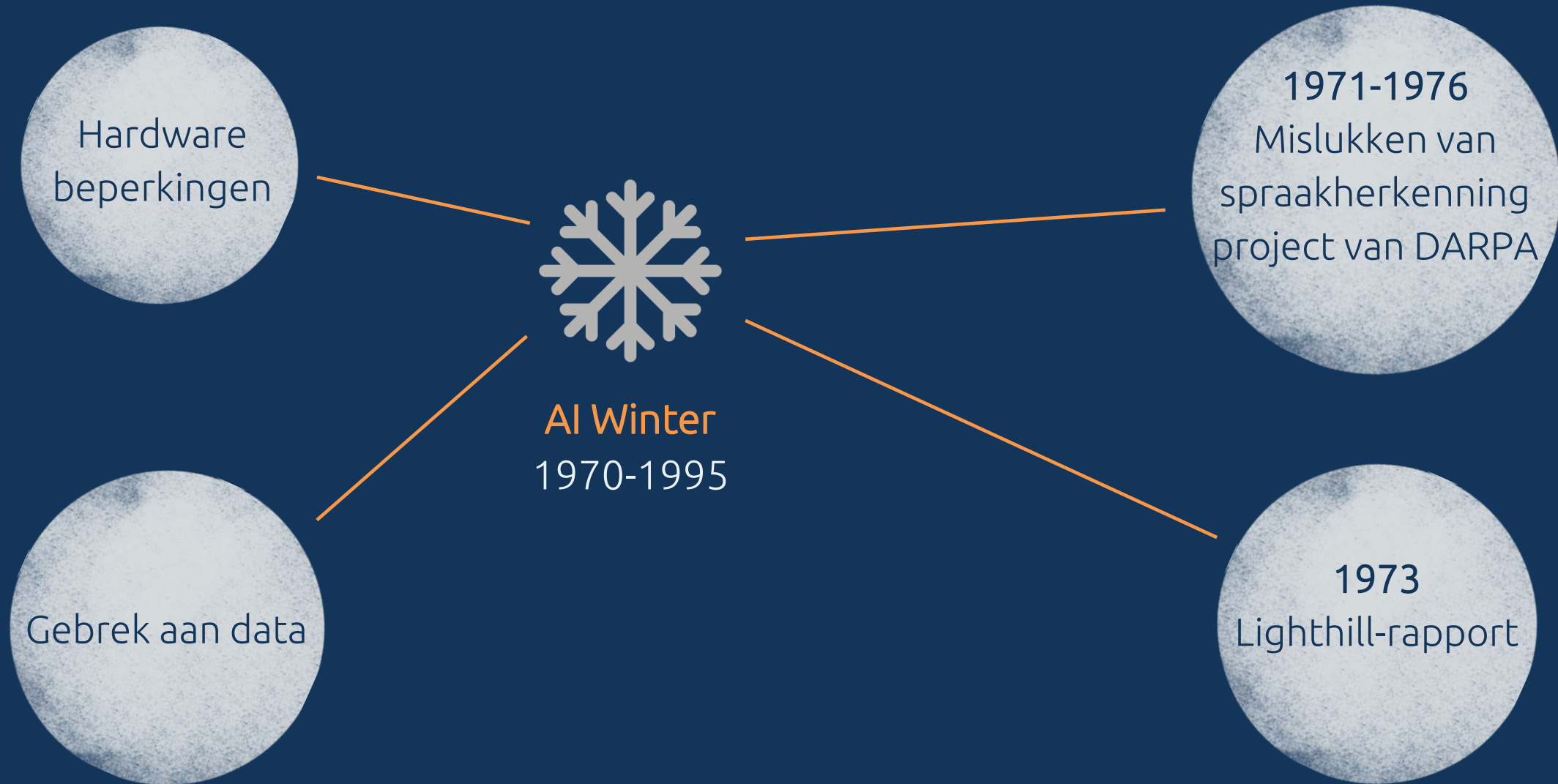
1958
Eerste neurale
netwerk
(perceptron)

1957
Onstaan van
beslisbomen

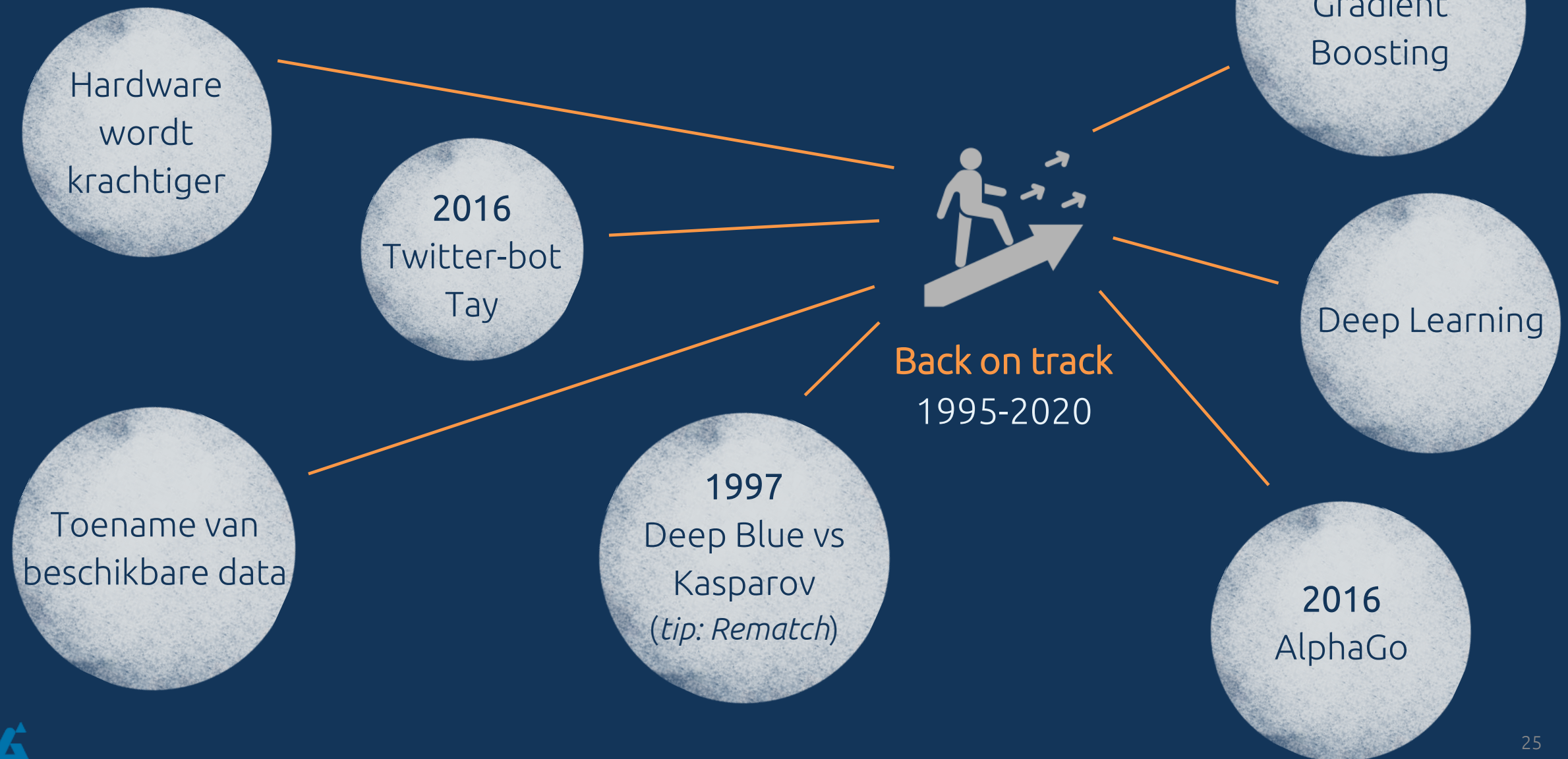
1956
Dartmouth
Conference

1950
Turing Test

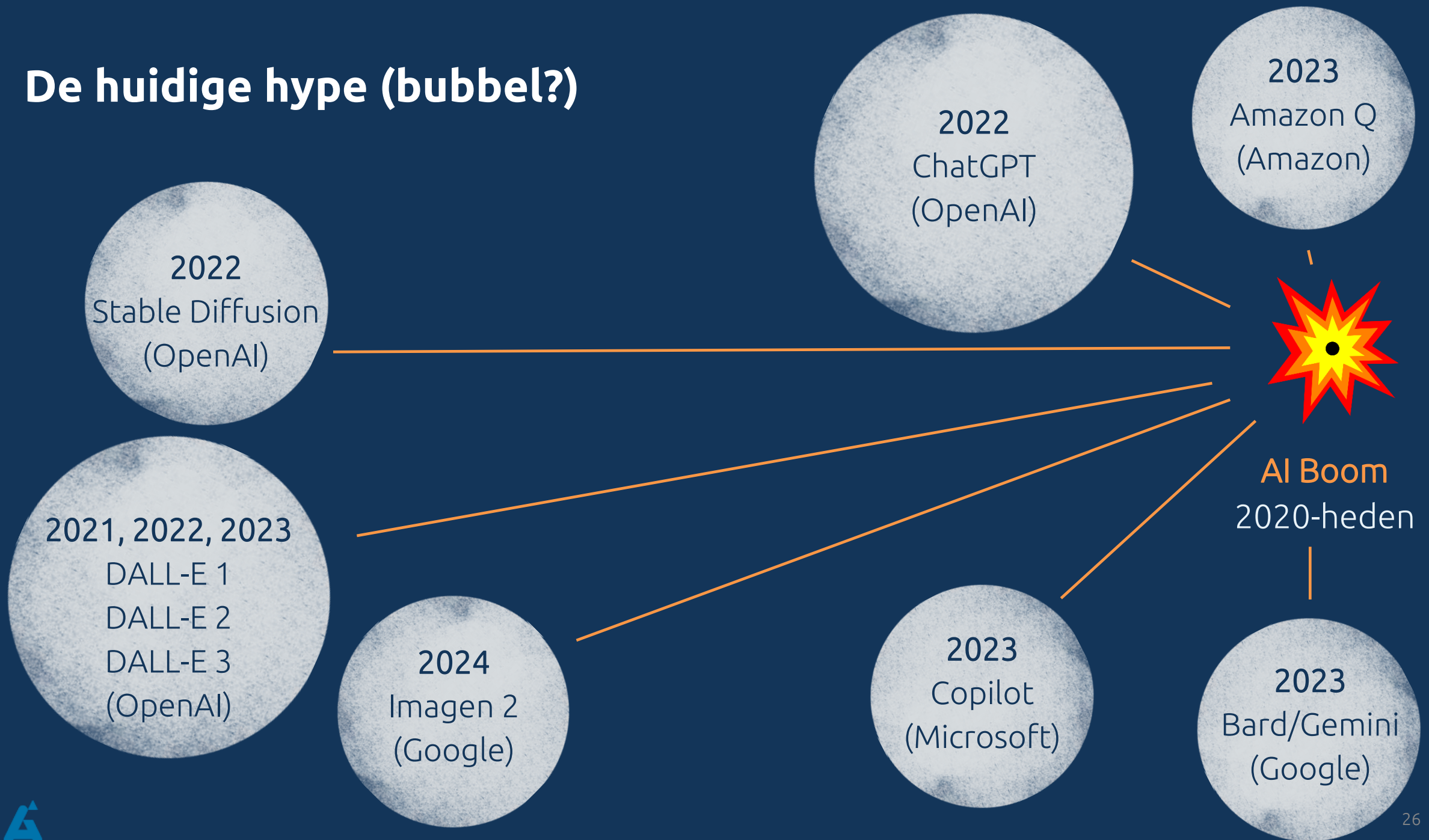
Periode van weinig enthousiasme en vertrouwen rondom AI



Optimisme begint weer aan te trekken



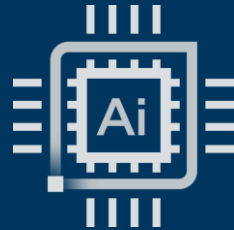
De huidige hype (bubbel?)



Traditionele AI versus Generative AI

Traditionele AI

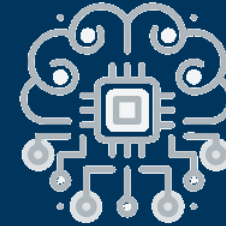
Voorspellen



Analyseren

Generative AI

Genereren



Zoeken

Traditionele AI kent veel toepassingen...

Supervised learning

Leer op basis van gelabelde data

Classificatie

Fraude-detectie
Spam-detectie
Beeldherkenning
Diagnostiek

Regressie

Risicobeoordeling
Waardebeoordeling

Unsupervised learning

Leer zonder kennis vooraf

Clusteren

Klantgroepen identificeren
Outlierdetectie

Dimensiereductie

Ruisreductie
Feature engineering
Visualiseren van hoge-dimensionale data

Reinforcement learning

Leer op basis van beloning en straf

Game AI

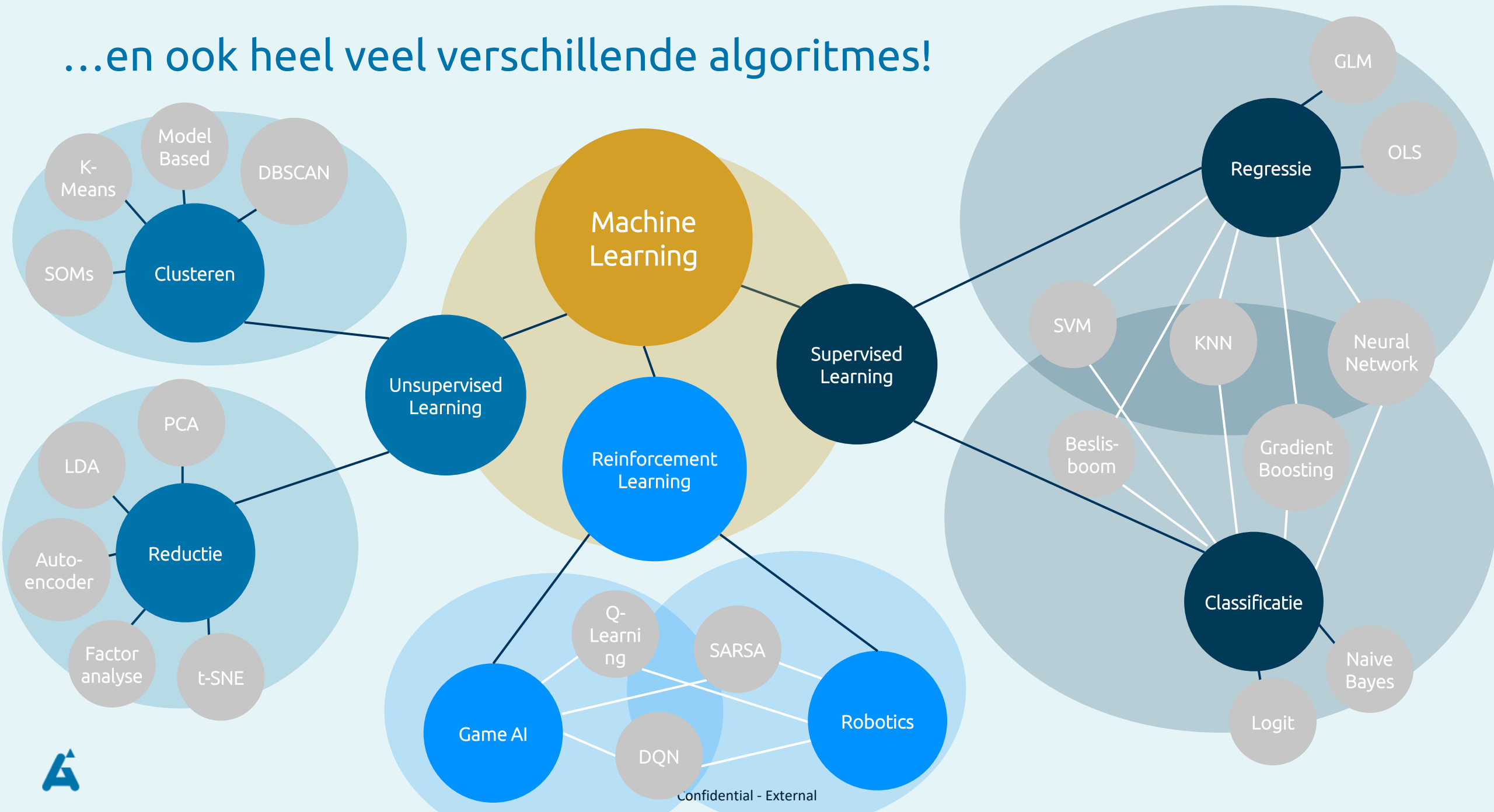
AlphaGo
DeepBlue
"Potje FIFA tegen de computer"

Robotics

Industriële robots
Zelfrijdende auto's
Robotstofzuiger



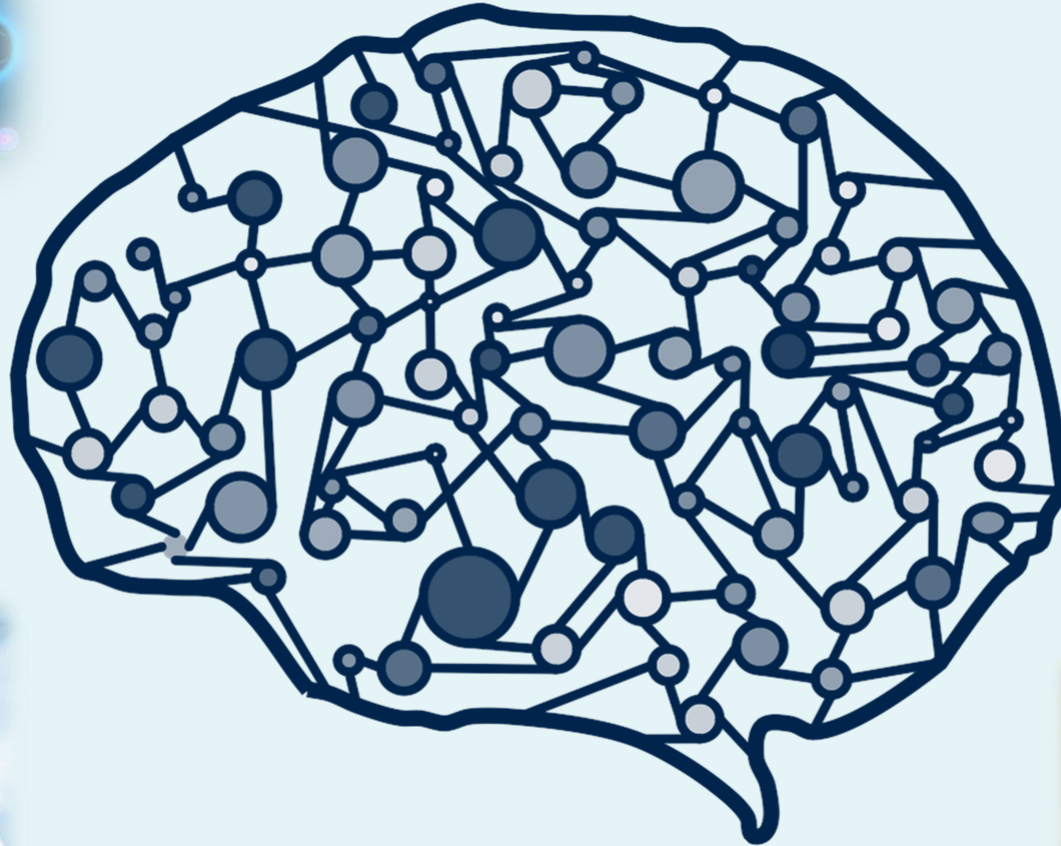
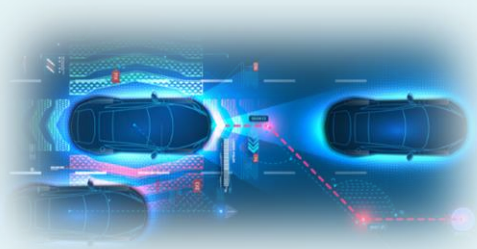
...en ook heel veel verschillende algoritmes!



Neurale Netwerken

Neurale Netwerken

Hebben in de afgelopen jaren een ware revolutie teweeggebracht in diverse vakgebieden

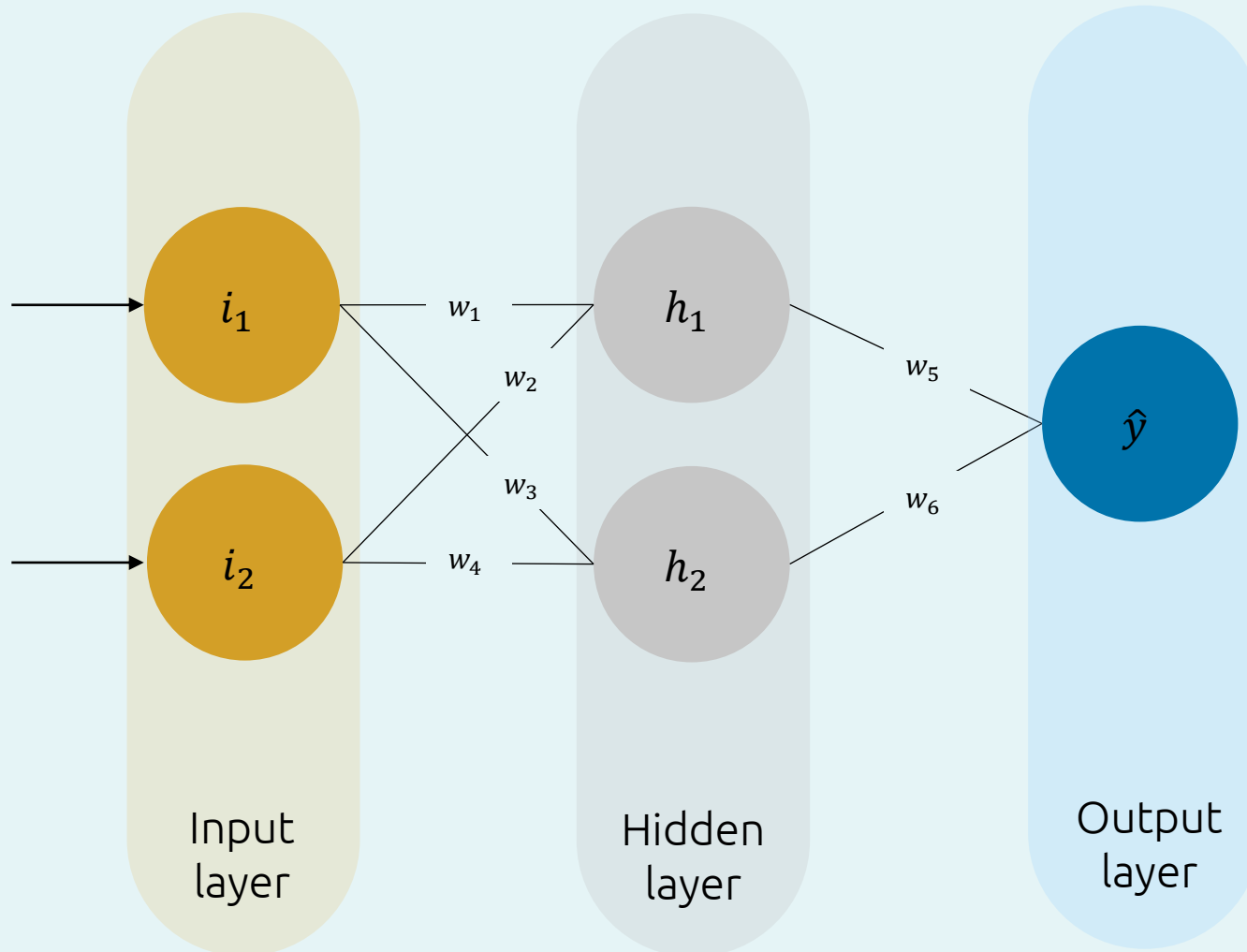


Hey Siri



Neurale Netwerken

Een simpel voorbeeld



Optimaliseren

Loss functie, bijvoorbeeld: $MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$

Gradient Descent

Backpropagation (kettingregel!)

Berekening

$$h_1 = f(w_1 i_1 + w_2 i_2)$$

$$h_2 = f(w_3 i_1 + w_4 i_2)$$

$$\hat{y} = g(w_5 h_1 + w_6 h_2)$$

$$\hat{y} = g(w_5 f(w_1 i_1 + w_2 i_2) + w_6 f(w_3 i_1 + w_4 i_2))$$

$$\hat{y} = w_5 f(w_1 i_1 + w_2 i_2) + w_6 f(w_3 i_1 + w_4 i_2)$$

f = niet-lineaire functie (sigmoid, relu)

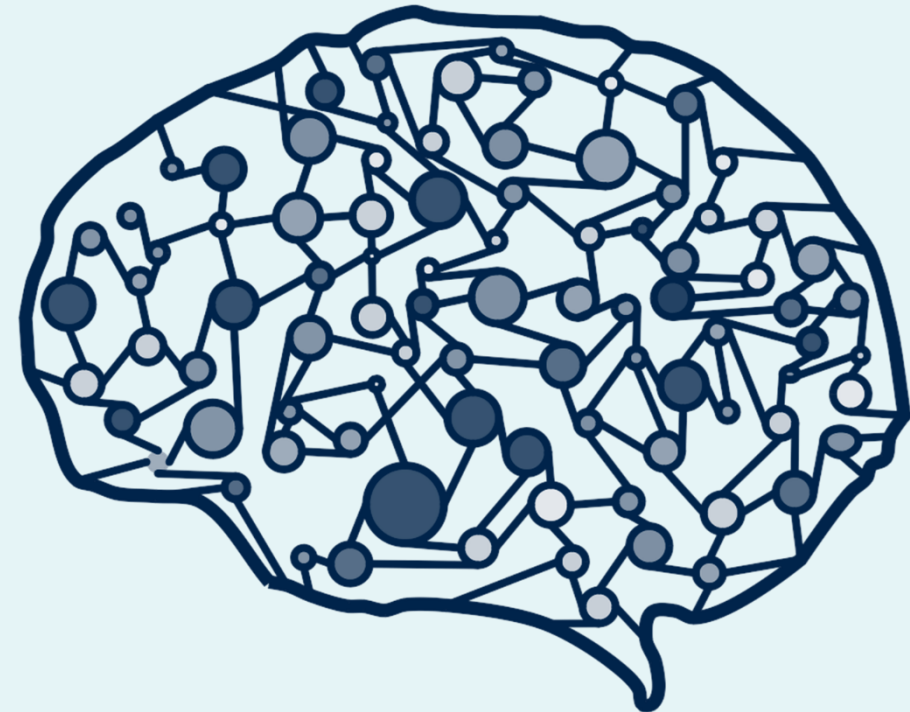
g = (niet-)lineaire functie (lineair, softmax)



Neurale Netwerken

Er zijn diverse netwerkkarchitecturen, de juiste keuze hangt af van de specifieke data, taak en gewenste prestatie

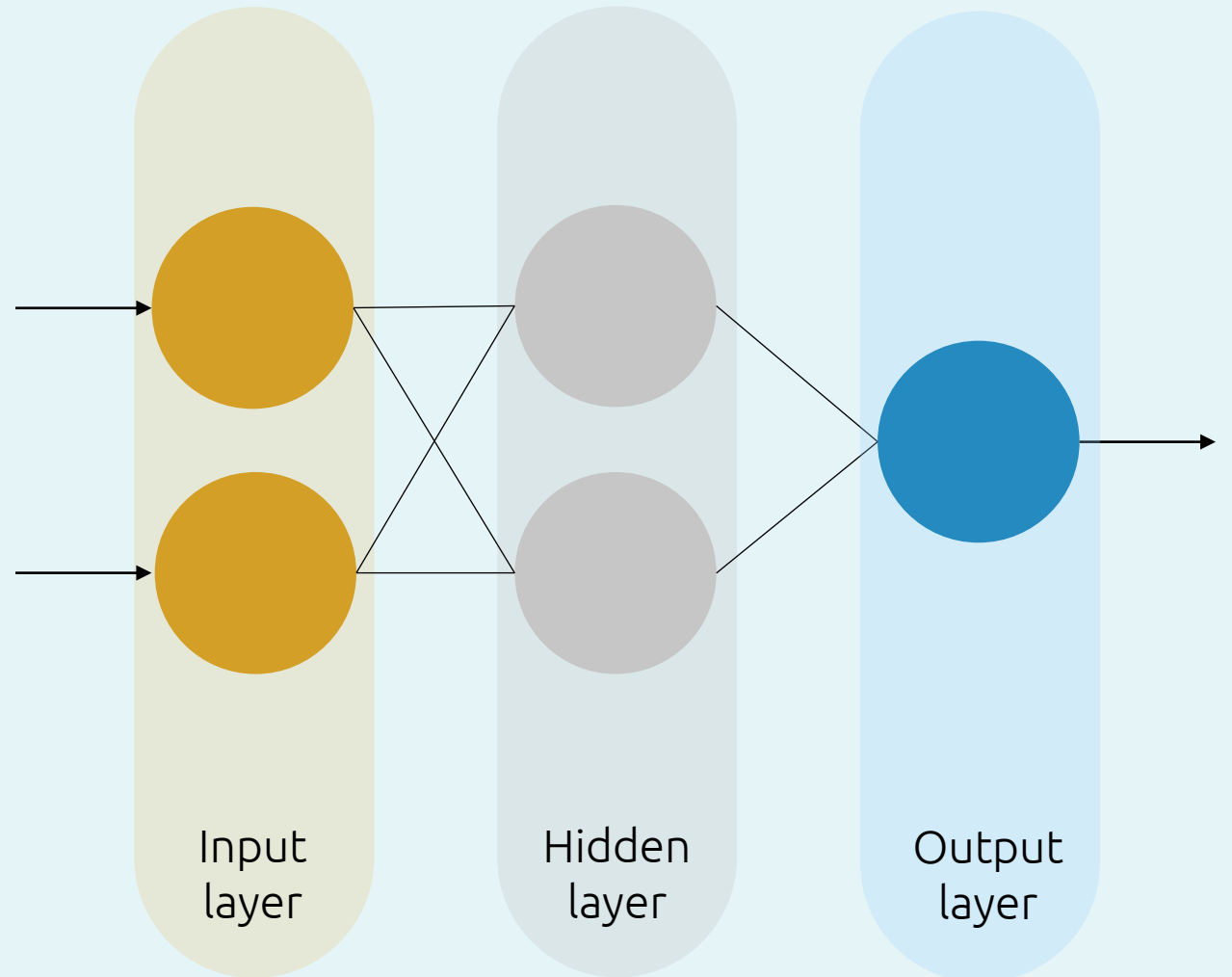
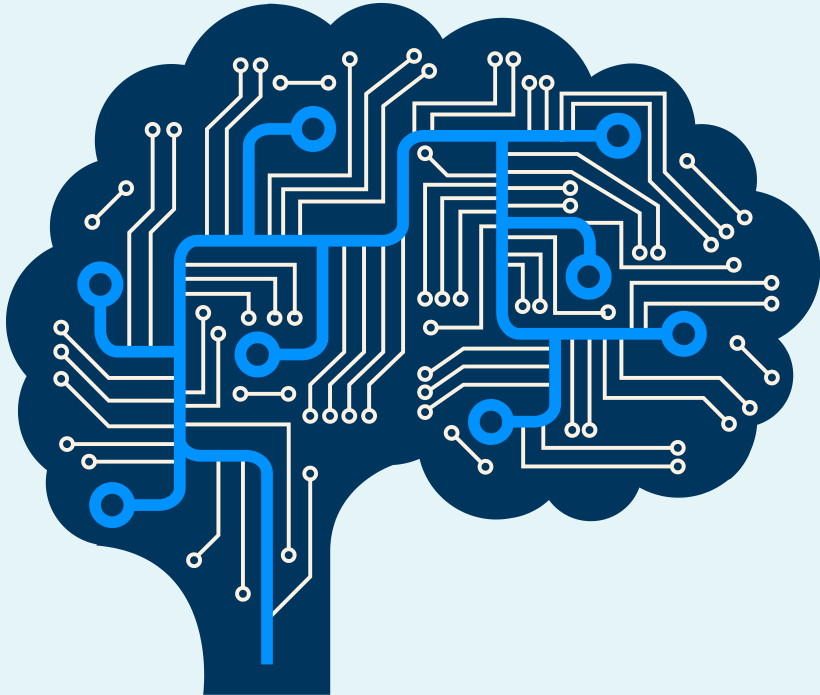
- Feedforward (FNNs)
- Convolutional (CNN's)
- Recurrent (RNN's)
- Generative adversarial (GAN's)
- Decoders
- Encoders
- Transformers
- ...



Neurale Netwerken

Feedforward NN

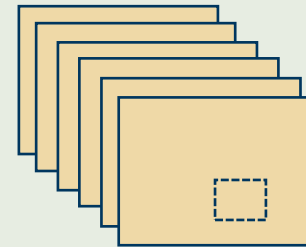
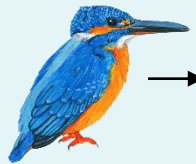
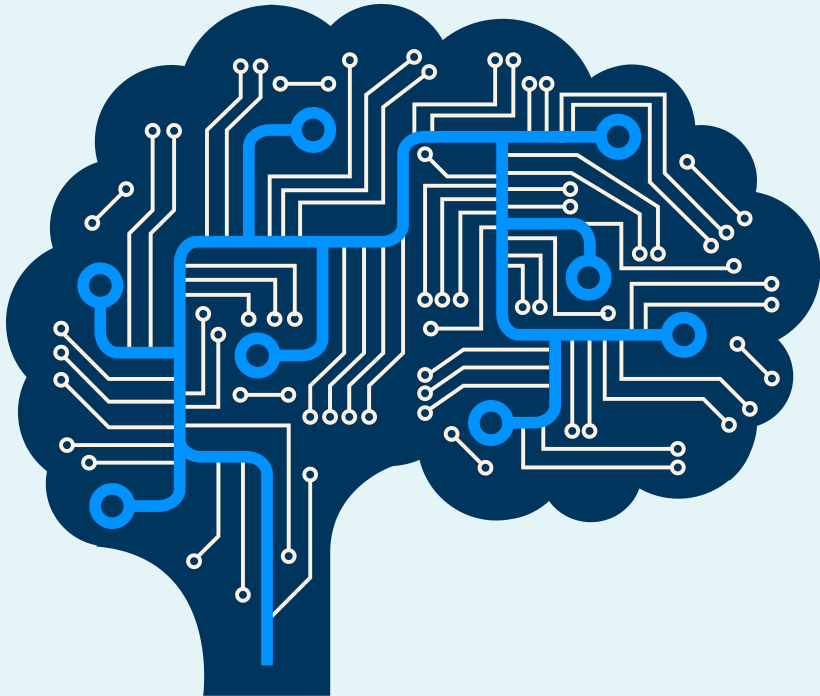
Classificatie- en
regressieproblemen



Neurale Netwerken

Convolutional NN

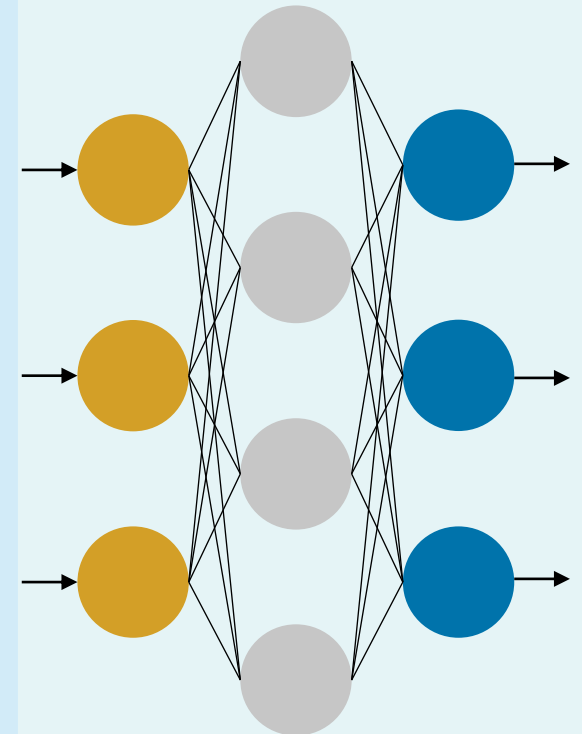
Beeldclassificatie,
objectdetectie en
gezichtsherkenning.



Convolutional
layer



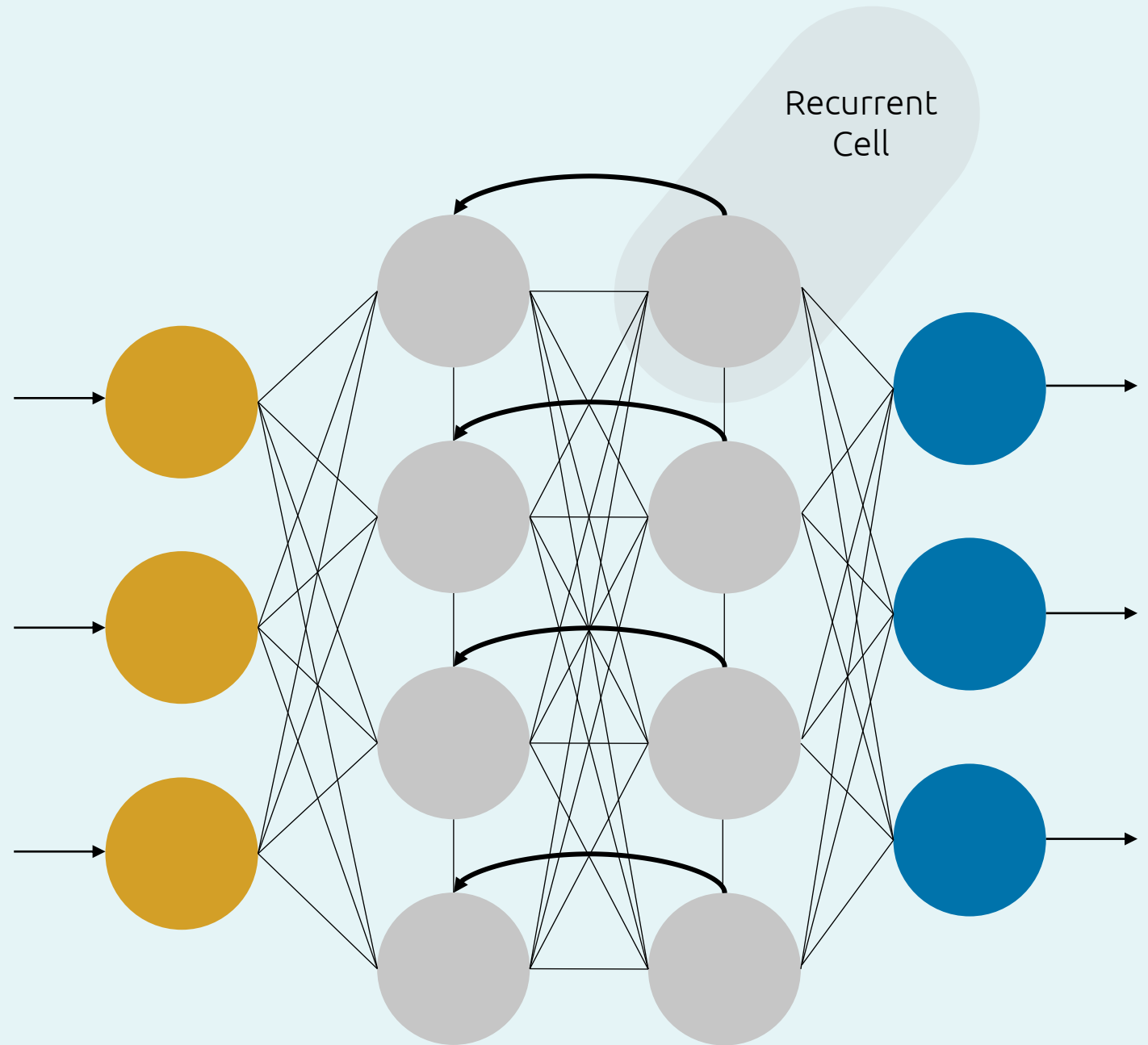
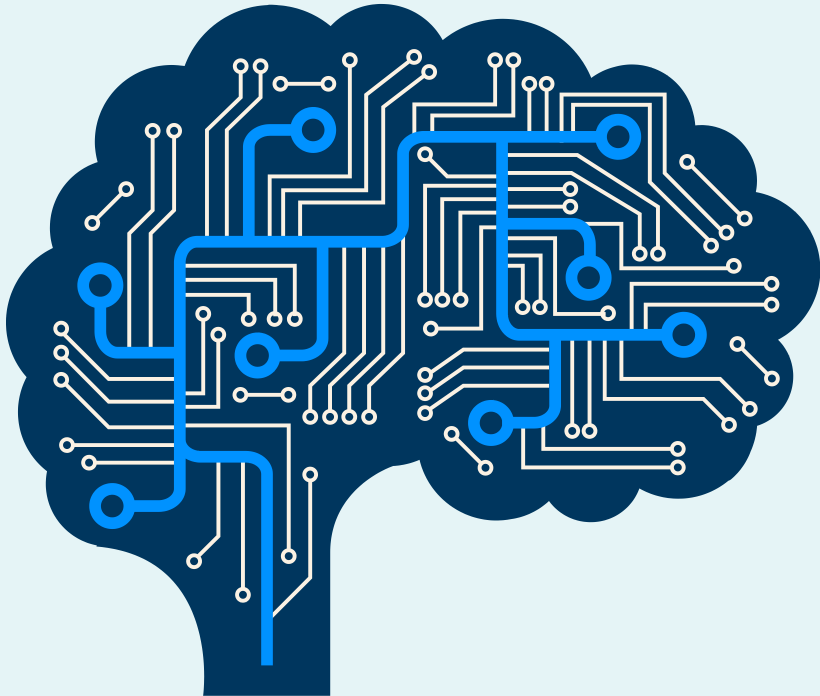
Pooling
layer



Neurale Netwerken

Recurrent NN

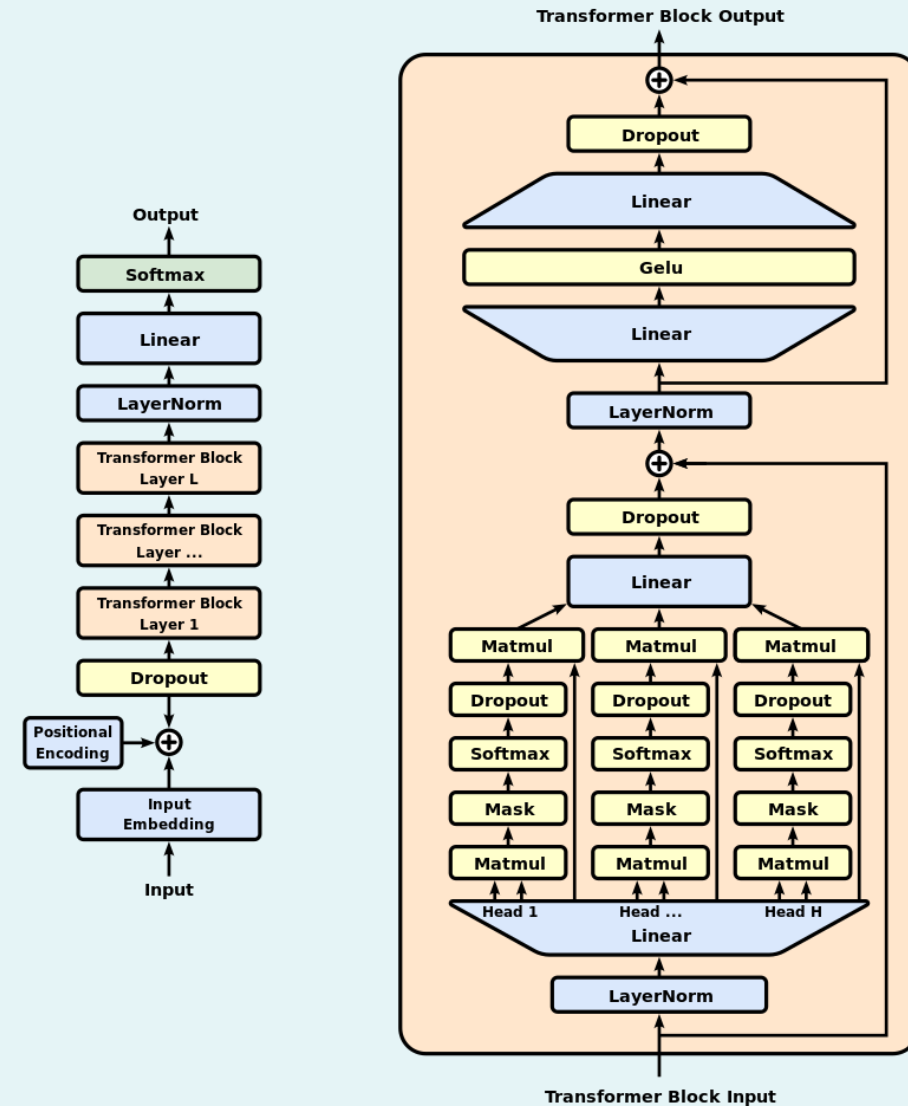
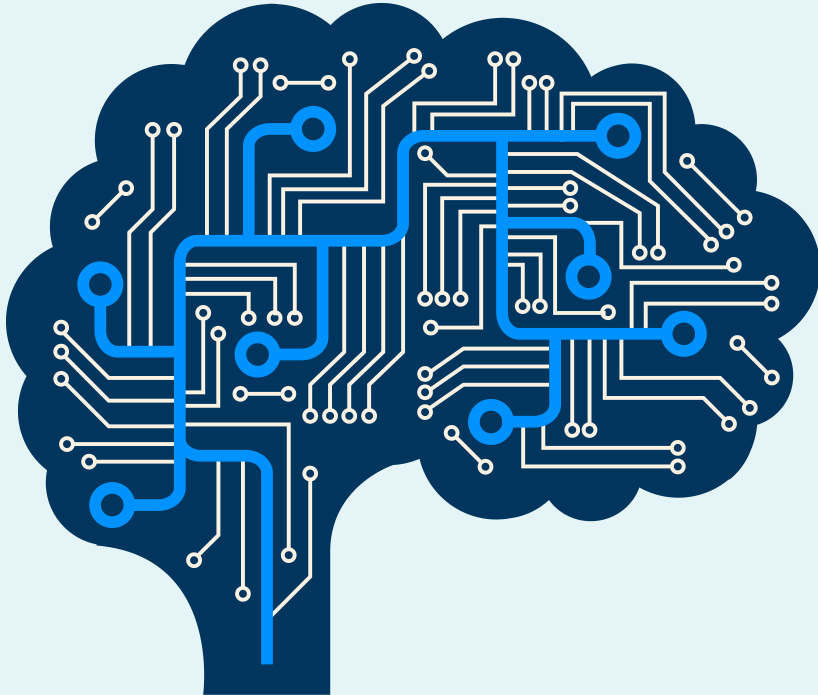
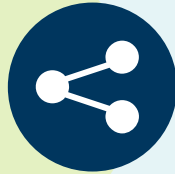
Vertaling,
spraakherkenning en
tekstgeneratie.



Neurale Netwerken

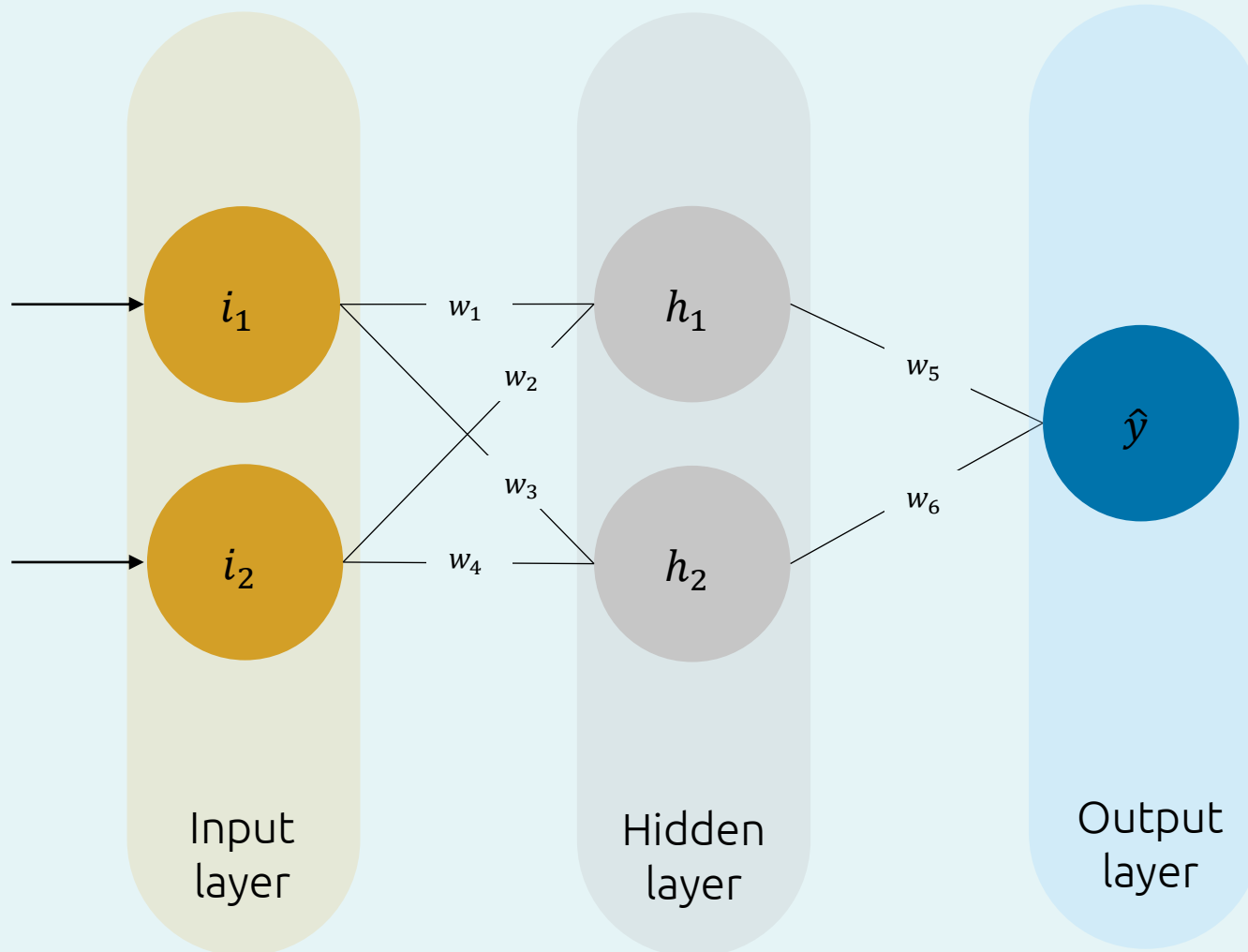
Transformers

De basis voor LLMs/ChatGPT.
Genereren van tekst en
afbeeldingen



Neurale Netwerken

De optimale waardes hangen af van de specifieke taak en de hoeveelheid beschikbare data.

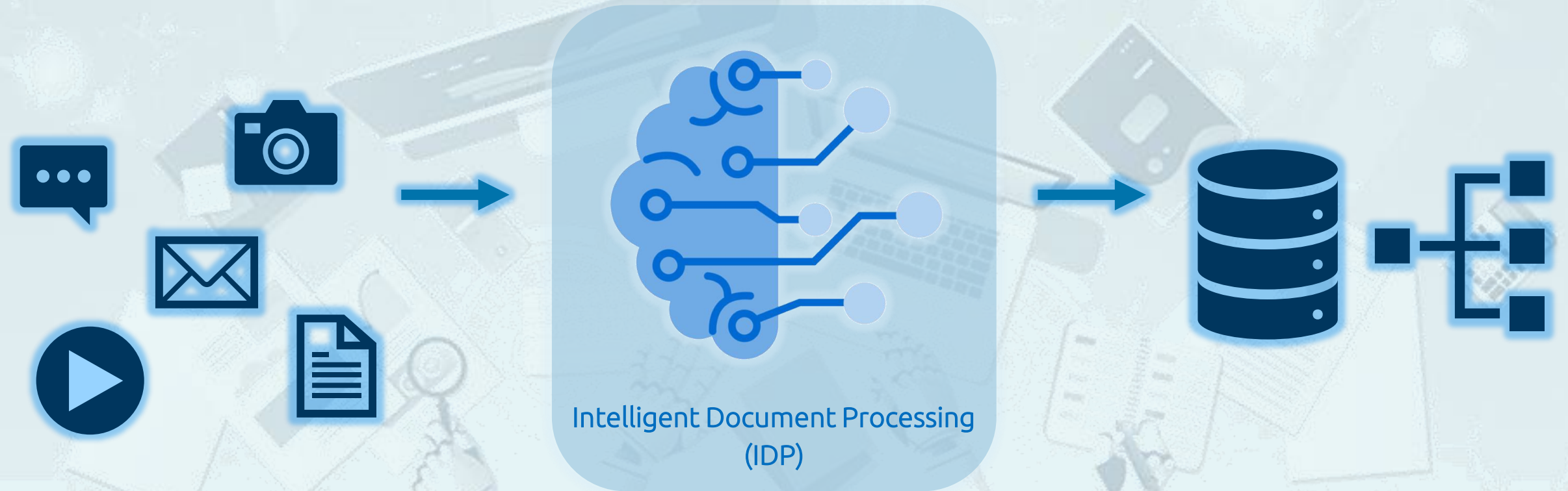


Hyperparameters

- Architectuur van het netwerk
 - Aantal layers
 - Aantal nodes per laag
 - Connecties tussen de lagen
- Leertempo
- Aantal epochs
- Batchgrootte
- Activeringsfunctie
 - Sigmoid
 - ReLU
 - tanh
- Regularisatie
 - L1-regularisatie
 - L2-regularisatie
- Optimizers
 - SGD
 - Adam
 - RMSProp
- ...

Wat is er nog meer mogelijk?

Verwerken van ongestructureerde data



Wat is er nog meer mogelijk?

Geautomatiseerde claims verwerking



Claim

- Geautomatiseerde verzameling van gegevens
- Eerste melding van schade, diefstal of verlies



Schadebepaling

- Schademeldingen kunnen worden ingediend via virtuele bots
- Vermindert verspilling van tijd en middelen bij het verzamelen en verwerken van gegevens



Onderzoek

- Foto's van de schade worden vergeleken met originele foto's uit de database
- Kostenbepaling
- Fraudepatronen detecteren



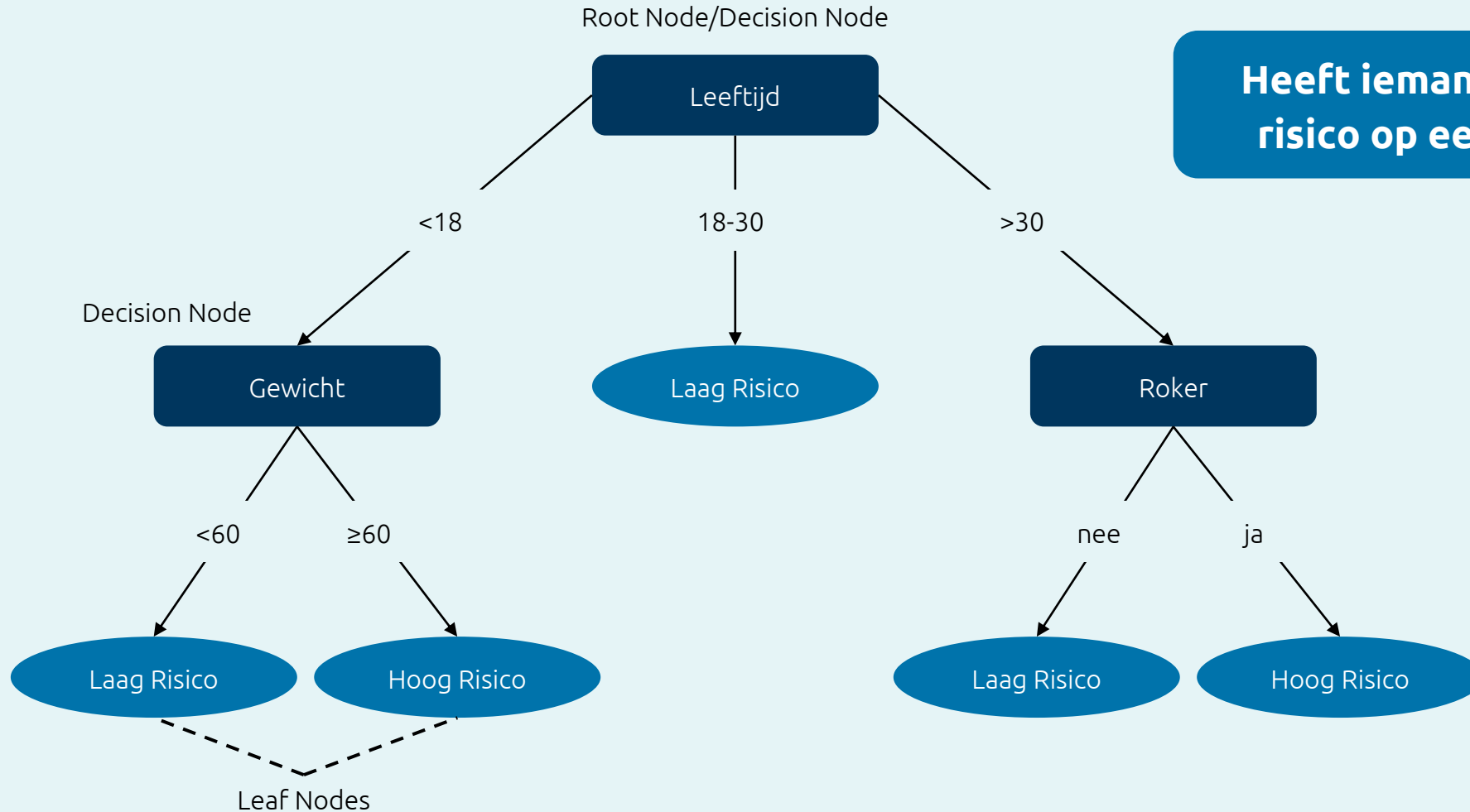
Afhandeling

- Automatische schaderegeling

Gradient Boosting

Gradient Boosting

De basis: de beslisboom

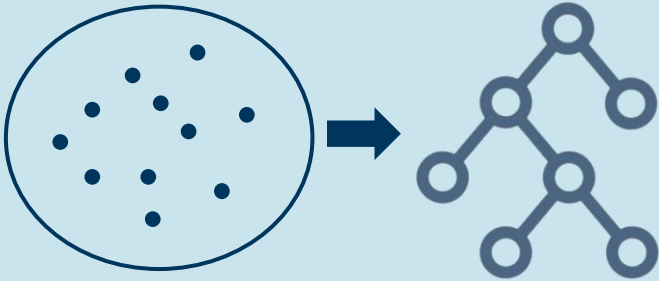


Gradient Boosting

Slim combineren van weak learners!

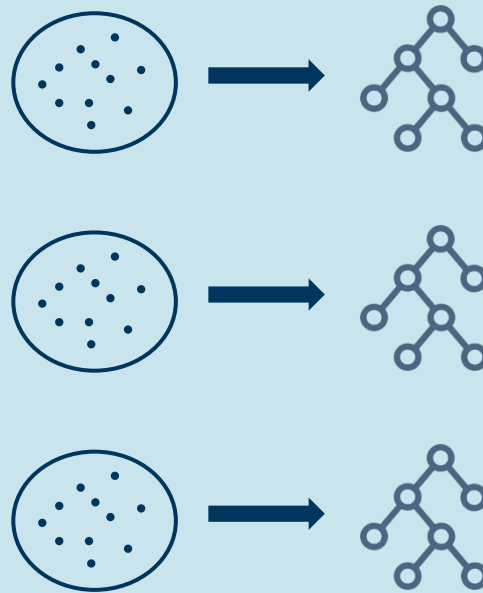
Decision Tree

1 enkel model



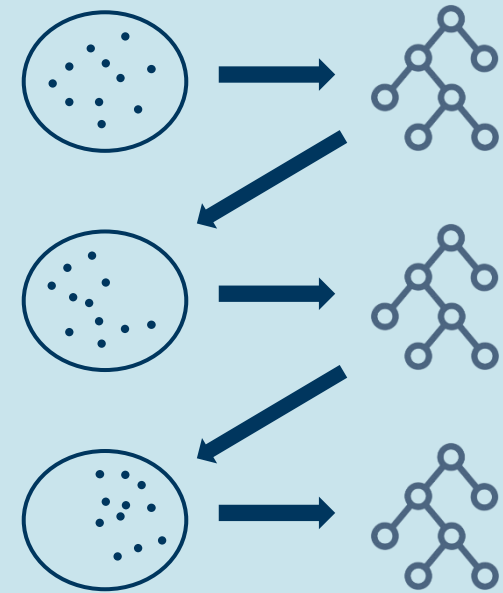
Random Forest

N onafhankelijke modellen



Gradient Boosting

N afhankelijke, sequentiële modellen



Gradient Boosting

Maar wat is er dan precies “slim” aan?



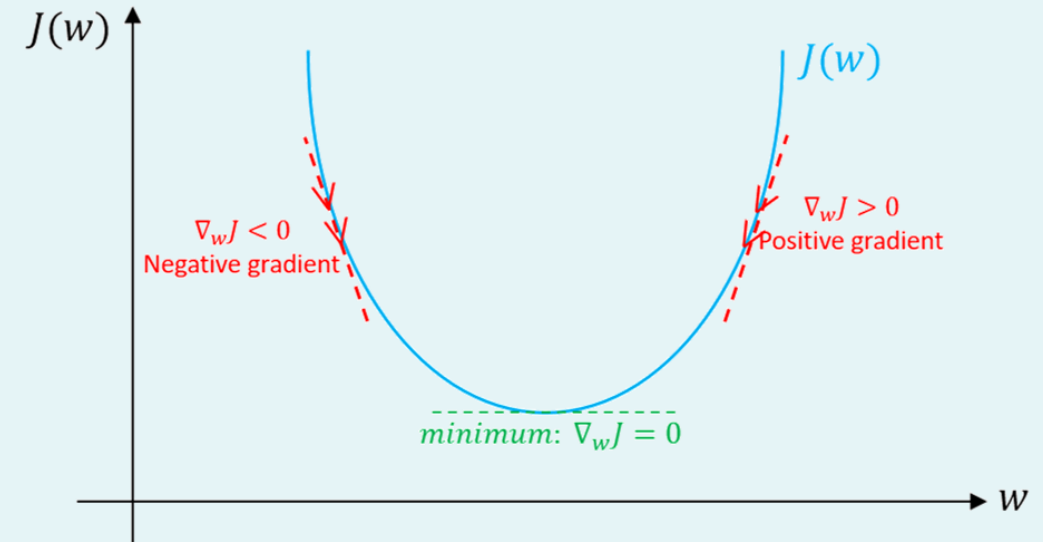
Jerome Friedman (1999)



Pas Gradient Descent toe op een loss-function die definieert hoe goed je model presteert.



Gradient Descent: beweeg in de *tegengestelde* richting van de gradient (=afgeleide) om het minimum te bereiken.



Gradient Boosting

Maar wat is er dan precies “slim” aan?

Loss functie: de Mean Squared Error!

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Gradient Descent: beweeg in *tegengestelde* richting van de Gradient om het minimum te bereiken:

$$\frac{\delta MSE}{\delta \hat{y}_i} = -2(y_i - \hat{y}_i)$$

$$-\frac{\delta MSE}{\delta \hat{y}_i} = 2(y_i - \hat{y}_i)$$

Dus, train een weak learner op de residuals!



Gradient Boosting

Hoe werkt dat voor een regressieprobleem?

Start met een (heel simpele) initiële voorspelling:

$$y = F_0(X)$$

Bepaal de residuals:

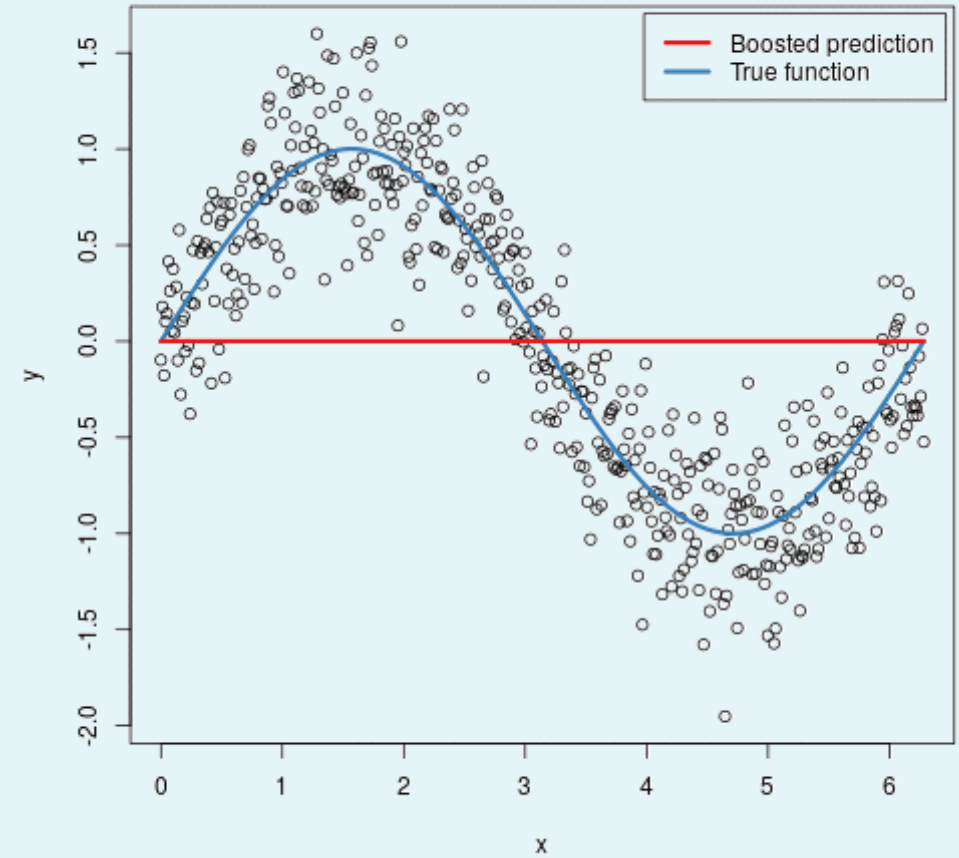
$$r_k = (y - F_k(X)), \quad k = 0, \dots, M$$

Train een beslisboom Δ_k op de residuals:

$$r_k = \Delta_{k+1}(X), \quad k = 0, \dots, M$$

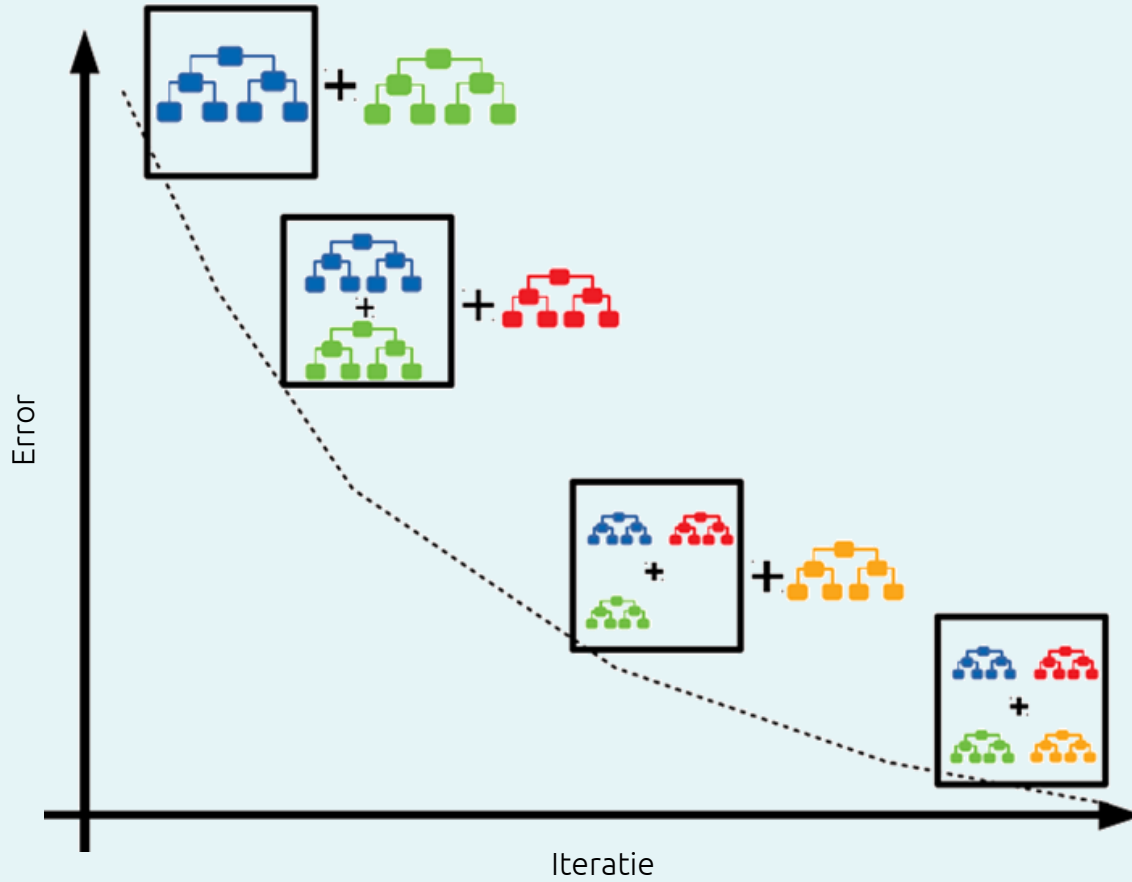
Voeg de uitkomsten van de beslisboom toe aan je totale model, totdat je aan je stopcriterium voldoet:

$$F_k(X) = F_{k-1}(X) + \eta * \Delta_k(X), \quad k = 1, \dots, M$$



Gradient Boosting

Heeft een overzichtelijk aantal hyperparameters



Hyperparameters

- Aantal estimators (beslisbomen)
- Leertempo
- Maximale diepte van de bomen
- Minimum aantal datapunten per blad
- Subsample-ratio
- Feature subsampling
- Regularisatie
 - L1-regularisatie
 - L2-regularisatie
- ...

Agenda

1

Opening & Recap

2

AI & Machine Learning

3

ML Pricing

Pauze

4

Data Management & Governance

5

AI risico's, regelgeving en ethiek

6

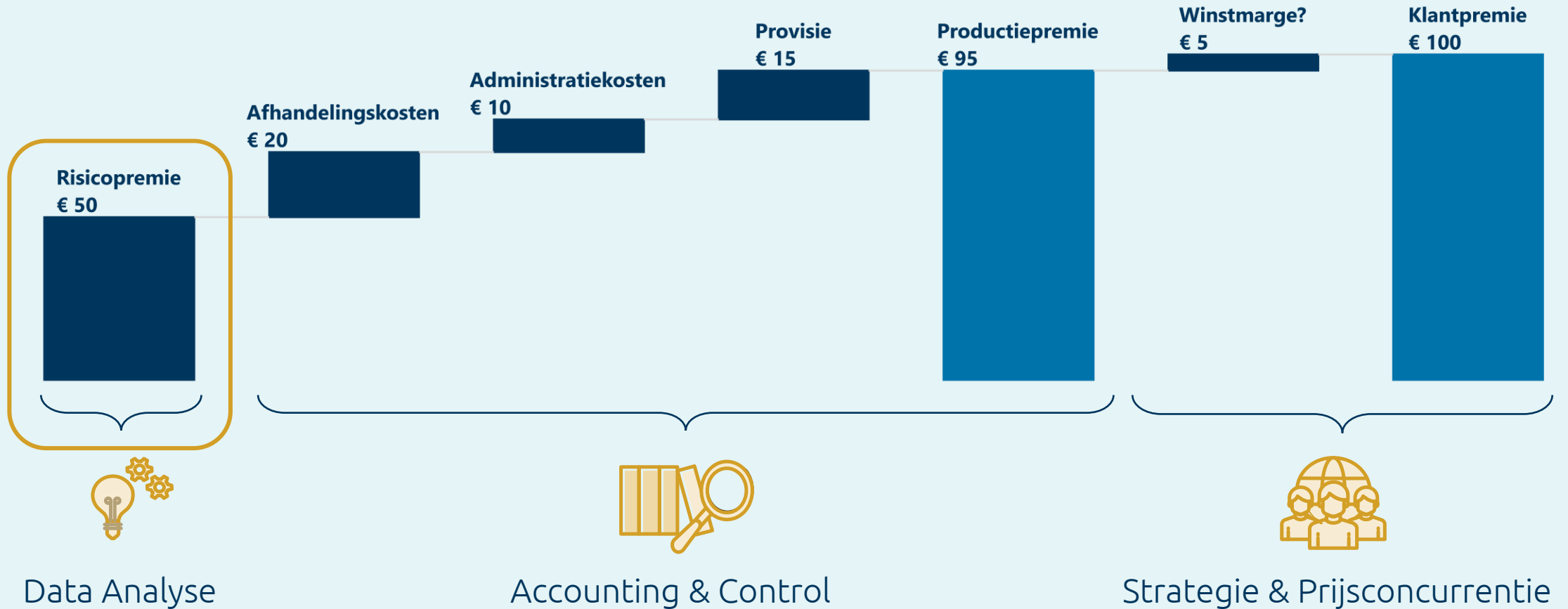
Afsluiting

Pricing

Hoe bepaal je de (kost)prijs van je product?



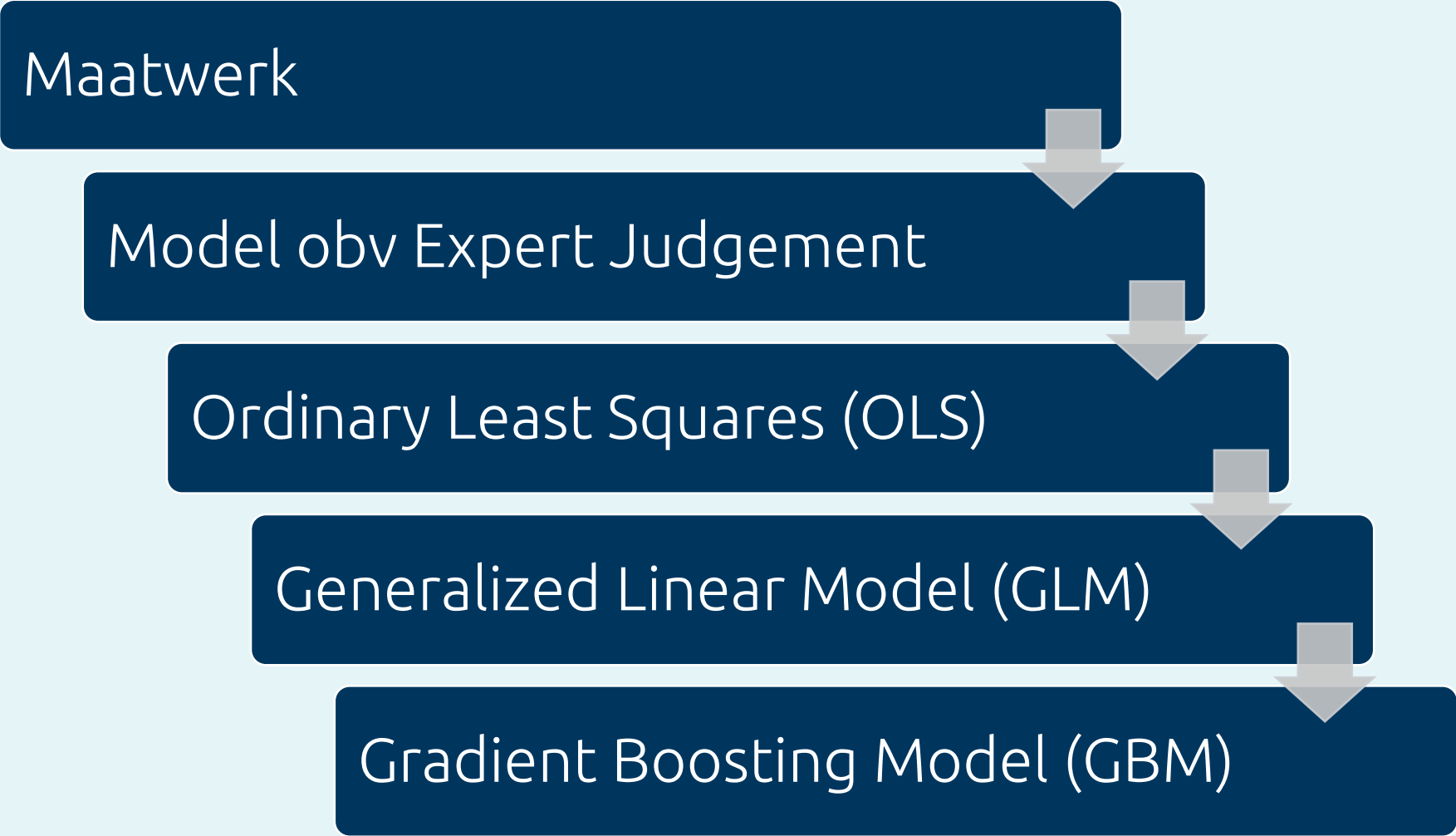
Premieopbouw



Opties voor bepaling risicopremie

Afhankelijk van soort verzekering (en dus de hoeveelheid data)!

Maatwerk



```
graph TD; A[Maatwerk] --> B[Model obv Expert Judgement]; B --> C[Ordinary Least Squares (OLS)]; C --> D[Generalized Linear Model (GLM)]; D --> E[Gradient Boosting Model (GBM)];
```

Model obv Expert Judgement

Ordinary Least Squares (OLS)

Generalized Linear Model (GLM)

Gradient Boosting Model (GBM)

Standaardmodellering GLMs

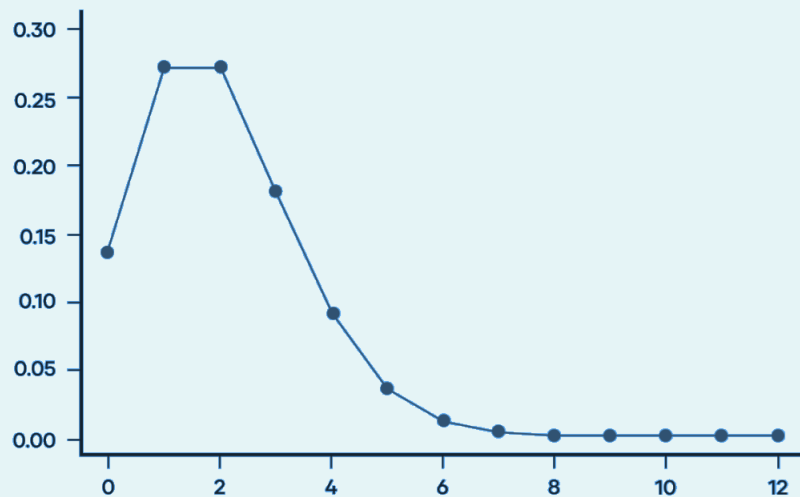
De verwachte schadelast wordt gemodelleerd als...

Frequentie (aantal schades)

Model

$$F = \exp\{X_0 * \beta_{0f} + X_1 * \beta_{1f} + \dots + X_n * \beta_{nf}\}$$

Verdeling: Poisson

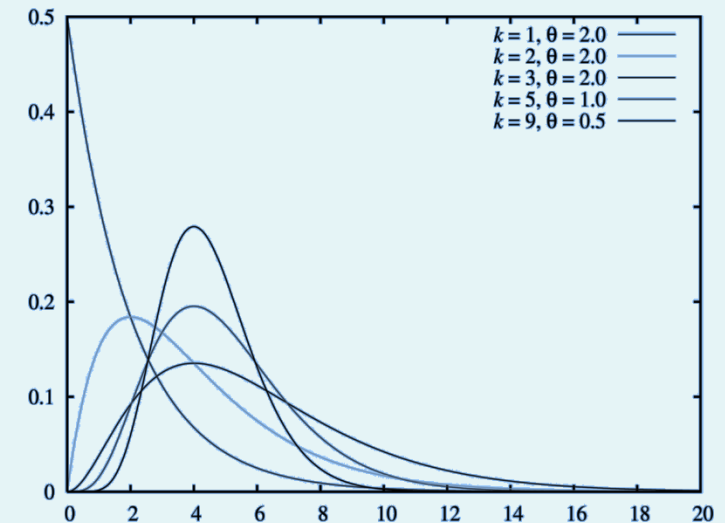


Schadelast (schadehoogte)

Model

$$S = \exp\{X_0 * \beta_{0s} + X_1 * \beta_{1s} + \dots + X_n * \beta_{ns}\}$$

Verdeling: Gamma

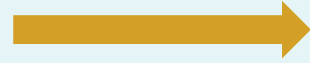


Waarom zouden we GLMs vervangen voor GBMs...?



Model performance

- Complexe, niet-lineaire verbanden
- Automatisch herkennen van kruisverbanden



Resultaat

- Betere risico-inschatting
- Lagere premie -> hoger marktaandeel
- Hogere premie -> stabielere rendement



Toekomstbestendig

- Gebruik van (live) sensordata in auto's vereist geavanceerde modellen

Overeenkomsten, verschillen en uitdagingen bij gebruik van GBMs

Hoe gaan we om met onderstaande stappen?



Data preparatie



Modelling



Expert Judgement



Uitlegbaarheid



Livegang

Data preparatie

Weinig verschillen!



Overeenkomsten, verschillen en uitdagingen bij gebruik van GBMs

Hoe gaan we om met onderstaande stappen?



Data preparatie



Modelling



Expert Judgement



Uitlegbaarheid



Livegang

Modellering

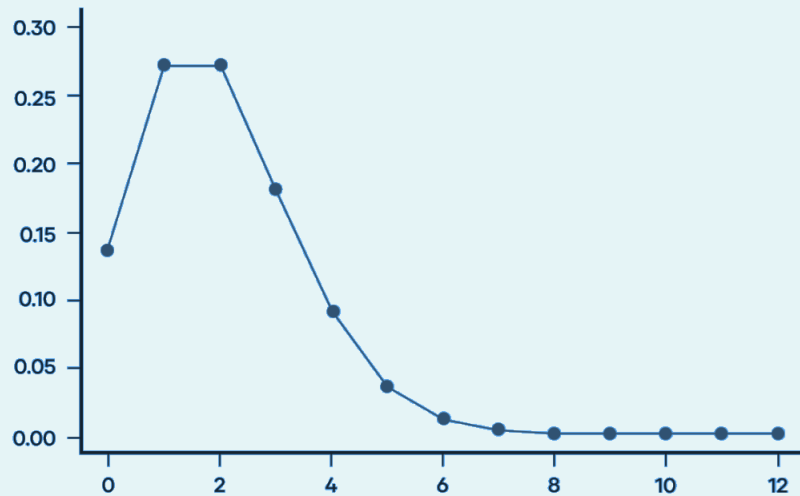
Voor GLMs geldt deze standaardmodellering

Frequentie (aantal schades)

Model

$$F = \exp\{X_0 * \beta_{0f} + X_1 * \beta_{1f} + \dots + X_n * \beta_{nf}\}$$

Verdeling: Poisson

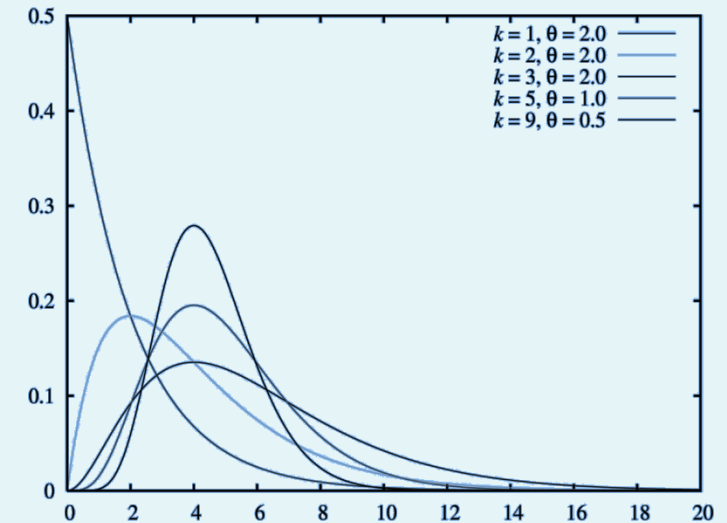


Schadelast (schadehoogte)

Model

$$S = \exp\{X_0 * \beta_{0s} + X_1 * \beta_{1s} + \dots + X_n * \beta_{ns}\}$$

Verdeling: Gamma



Modelling

Voor GBMs verandert er hier niet zoveel...

Frequentie (aantal schades)

Model

$$F = GBM_F(X)$$

Loss functie

Gebruik dat wat werkt!

Root Mean Squared Error
Mean Absolute Error
Huber Loss
Poisson Deviance



Schadelast (schadehoogte)

Model

$$S = GBM_S(X)$$

Loss functie

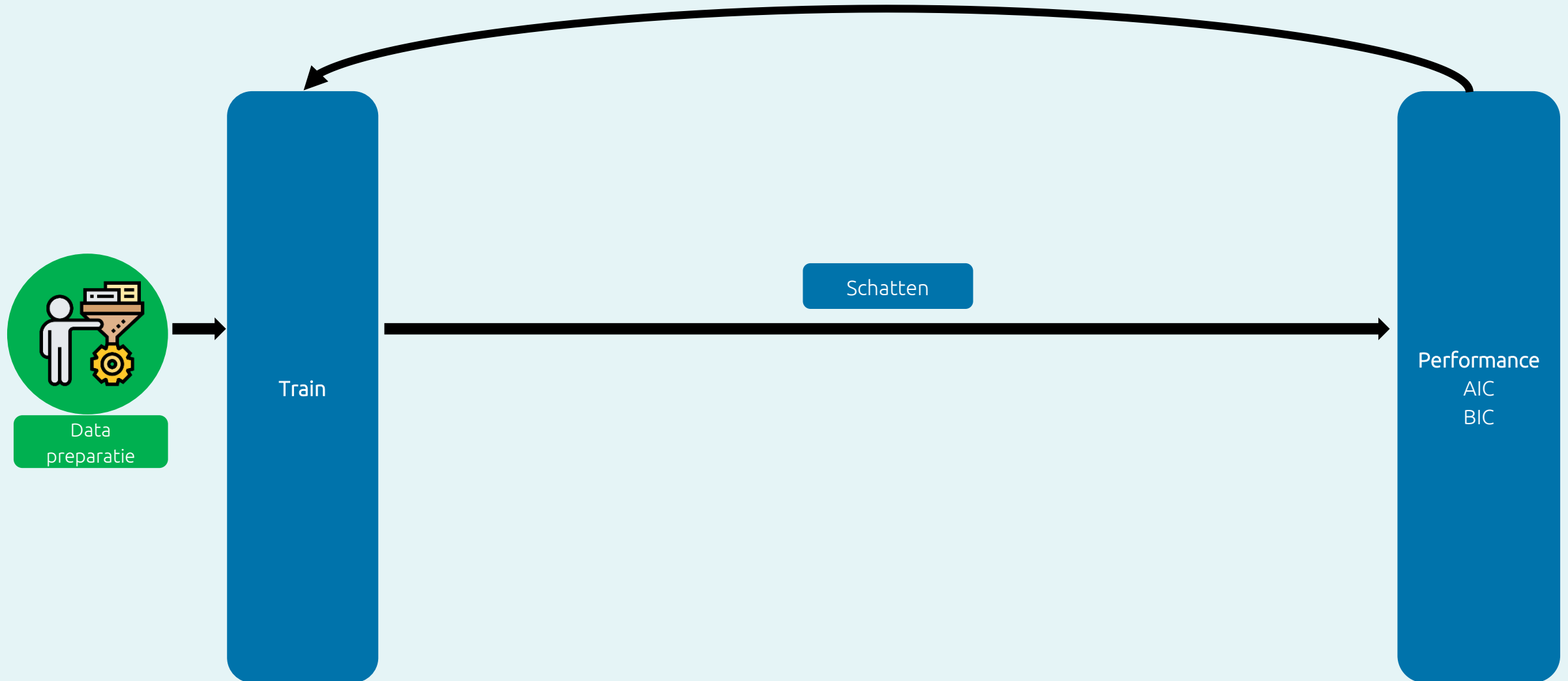
Gebruik dat wat werkt!

Root Mean Squared Error
Mean Absolute Error
Fair Loss
Gamma Deviance



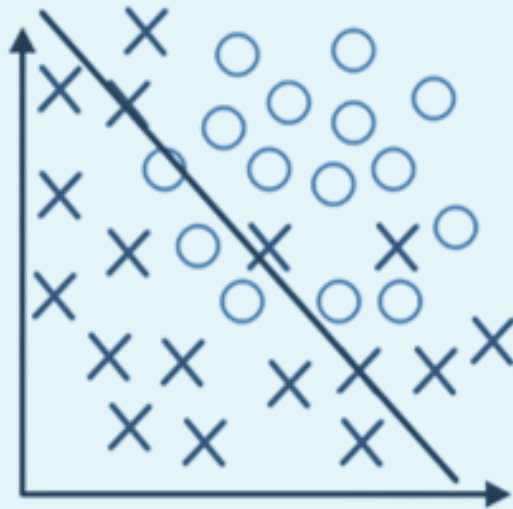
Modellering

Het optimalisatieproces bij GLMs is relatief simpel

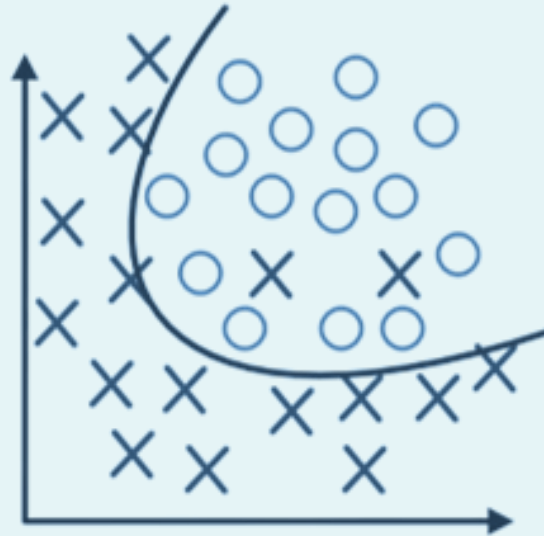


Modelling

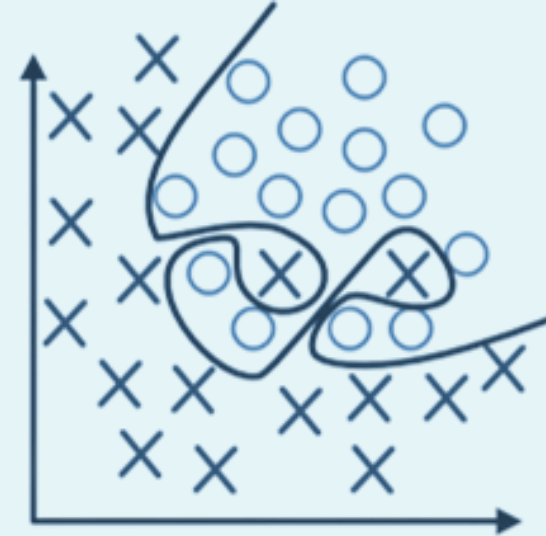
Datzelfde proces zou bij ML leiden tot problemen ...



Underfitting



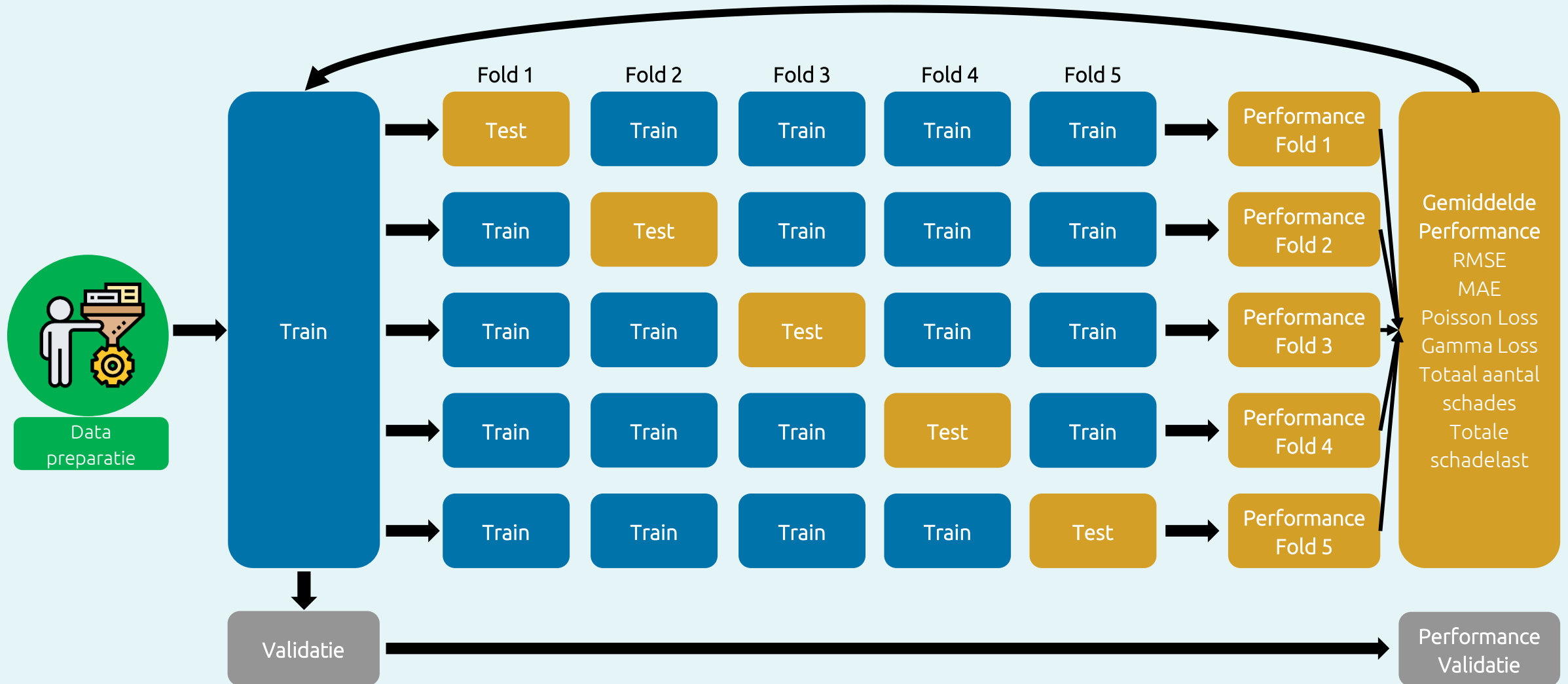
Ideale fit



Overfitting

Modellering

... en daar moet rekening mee worden gehouden in het optimalisatieproces



Overeenkomsten, verschillen en uitdagingen bij gebruik van GBMs

Hoe gaan we om met onderstaande stappen?



Data preparatie



Modelling



Expert Judgement



Uitlegbaarheid



Livegang

Expert Judgement

Lastiger in vergelijking met GLMs, maar niet onmogelijk bij GBMs!

GLM

Model

$$\text{Risicopremie} = X_{0f} X_{0s} * X_1 * X_2 * \dots * X_n$$

$X_{0f} :=$ gemiddelde frequentie = 0.1

$X_{0s} :=$ gemiddelde schadelast = 1500

$X_1 :=$ Merk

$X_2 :=$ Regio

Merk	Factor
Kia	1
Volkswagen	3
Ferrari	5

Regio	Factor
Stad	2
Dorp	1
Platteland	0.5

Expert Judgement

- Handmatig tunen van factoren

GBM

Model

$$\text{Risicopremie} = GBM_F(X) * GBM_S(X)$$

$$\text{Risicopremie} = \text{?} * \text{?}$$

$$\text{Risicopremie} = \text{?}$$

Expert Judgement

- Zelf uitprogrammeren als extra laag over het model heen
- Mogelijkheid tot restricties bepaalde features (monotoon verloop)
- Restricties opleggen door handig om te gaan met features in model

Overeenkomsten, verschillen en uitdagingen bij gebruik van GBMs

Hoe gaan we om met onderstaande stappen?



Data preparatie



Modelling



Expert Judgement



Uitlegbaarheid

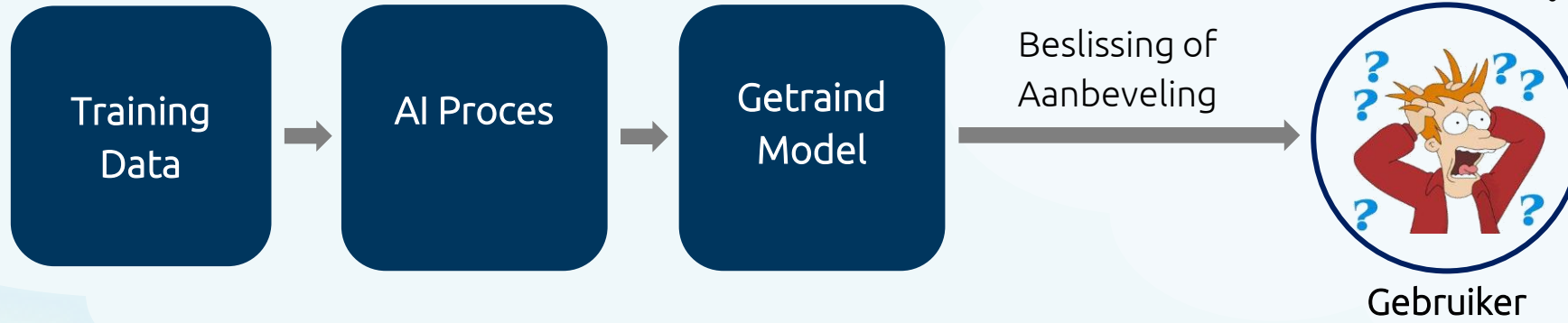


Livegang

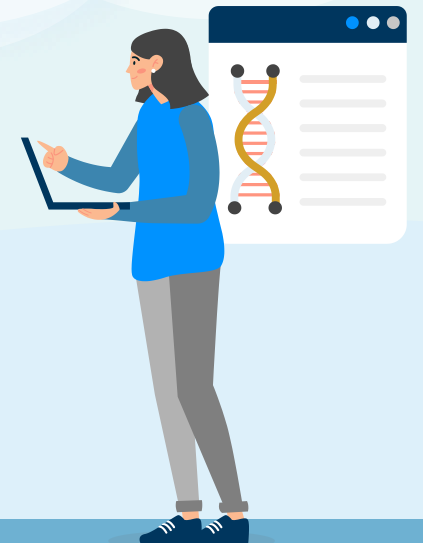
Uitlegbaarheid / Explainable AI

...wordt steeds belangrijker!

Huidig



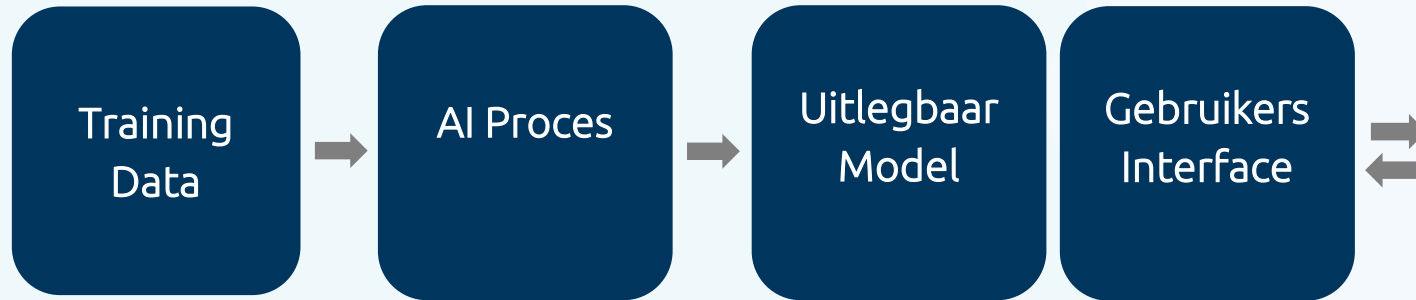
- Waarom doe je dit?
- Waarom niet iets anders?
- Wanneer is het goed?
- Wanneer gaat het mis?
- Kan ik je vertrouwen?
- Hoe moet ik een fout verbeteren?



Uitlegbaarheid / Explainable AI

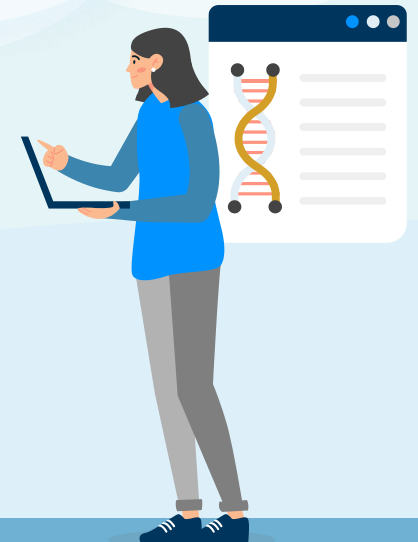
...wordt steeds belangrijker!

XAI



Gebruiker

- Ik snap waarom
 - Ik snap waarom iets niet is
 - Ik snap wanneer het goed is
 - Ik snap wanneer het mis gaat
 - Ik kan je vertrouwen
 - Ik weet waarom je een error gaf



Uitlegbaarheid / Explainable AI

Lastiger in vergelijking met GLMs, maar niet onmogelijk bij GBMs!



Merk = Volkswagen
Regio = Stad

GLM

Model

$$\text{Risicopremie} = X_{of} X_{os} * X_1 * X_2 * \dots * X_n$$

$X_{of} :=$ gemiddelde frequentie = 0.1

$X_{os} :=$ gemiddelde schadelast = 1500

$X_1 :=$ Merk

$X_2 :=$ Regio

Merk	Factor
Kia	1
Volkswagen	3
Ferrari	5

Regio	Factor
Stad	2
Dorp	1
Platteland	0.5

Voorbeeld

$$\text{Risicopremie} = 1500 * 0.1 * 3 * 2 = 900$$

GBM

Model

$$\text{Risicopremie} = GBM_F(X) * GBM_S(X)$$

$$\text{Risicopremie} = \text{?} * \text{?}$$

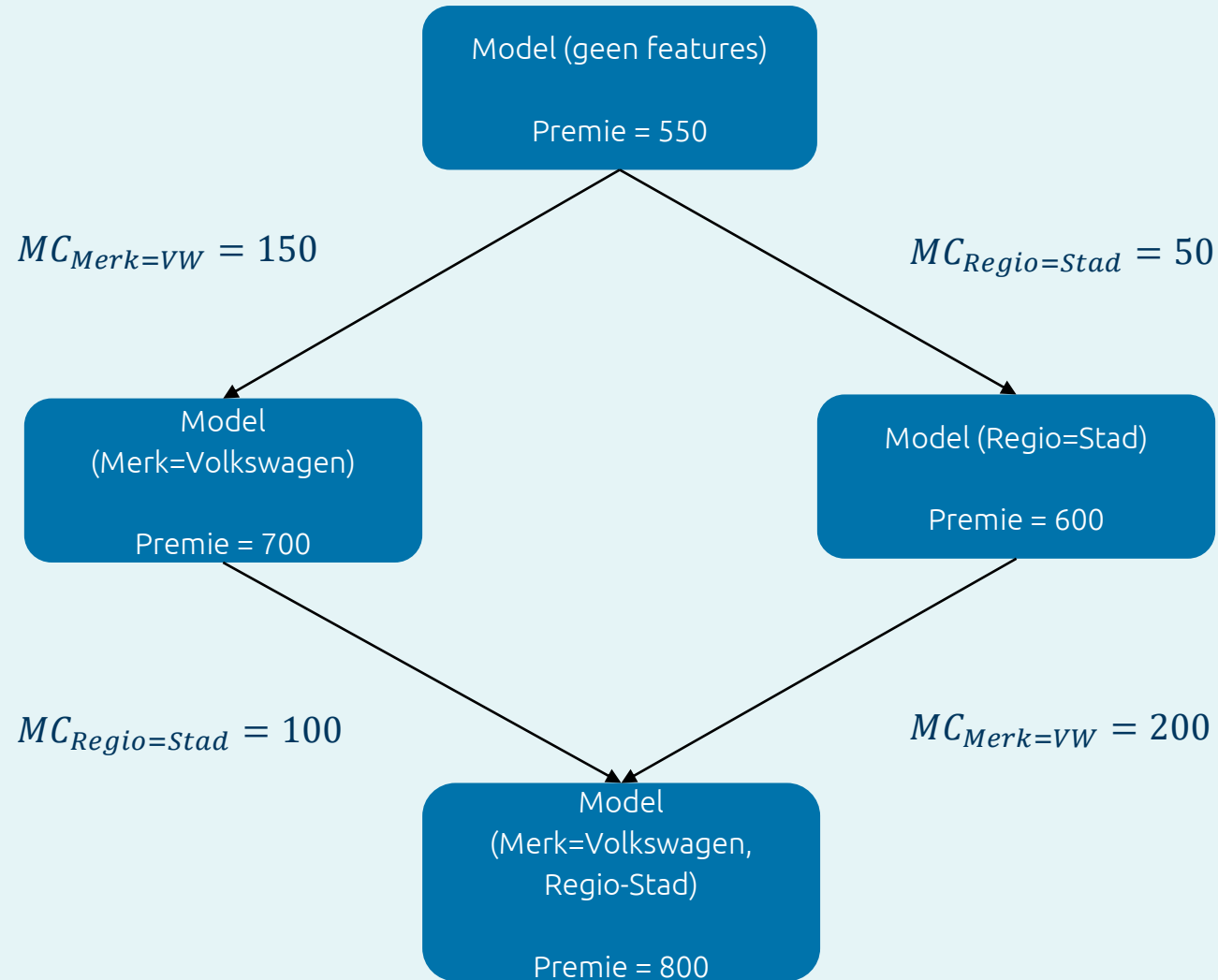
$$\text{Risicopremie} = \text{?}$$

Voorbeeld

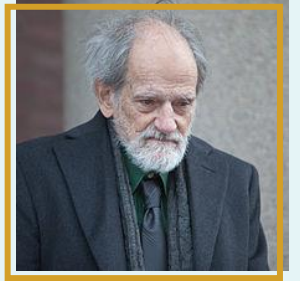


Uitlegbaarheid / Explainable AI

Shapley to the rescue!



Shap



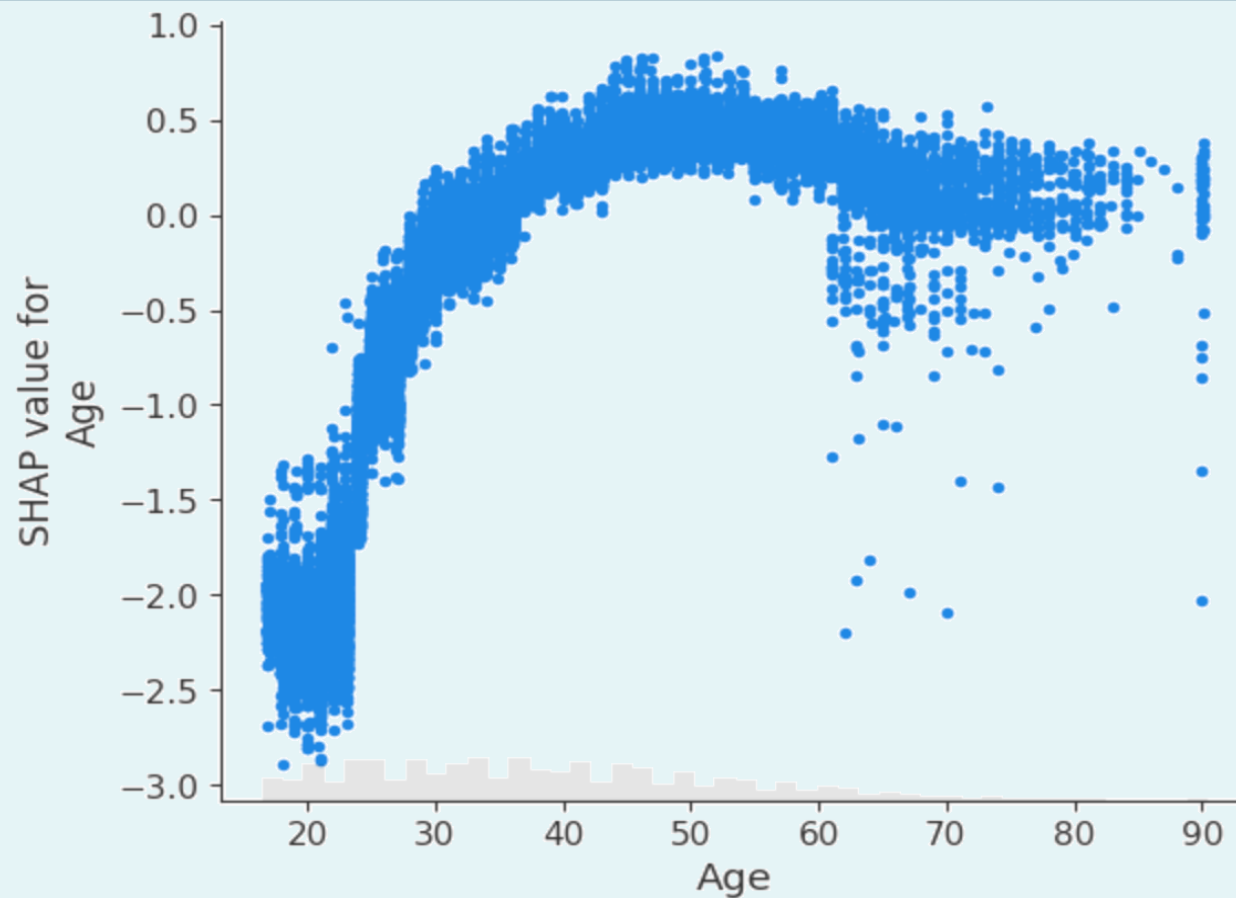
$$Shap_{Merk=VW} = \frac{1}{2} * 150 + \frac{1}{2} * 200 = 175$$

$$Shap_{Regio=Stad} = \frac{1}{2} * 50 + \frac{1}{2} * 100 = 75$$

Uitlegbaarheid / Explainable AI

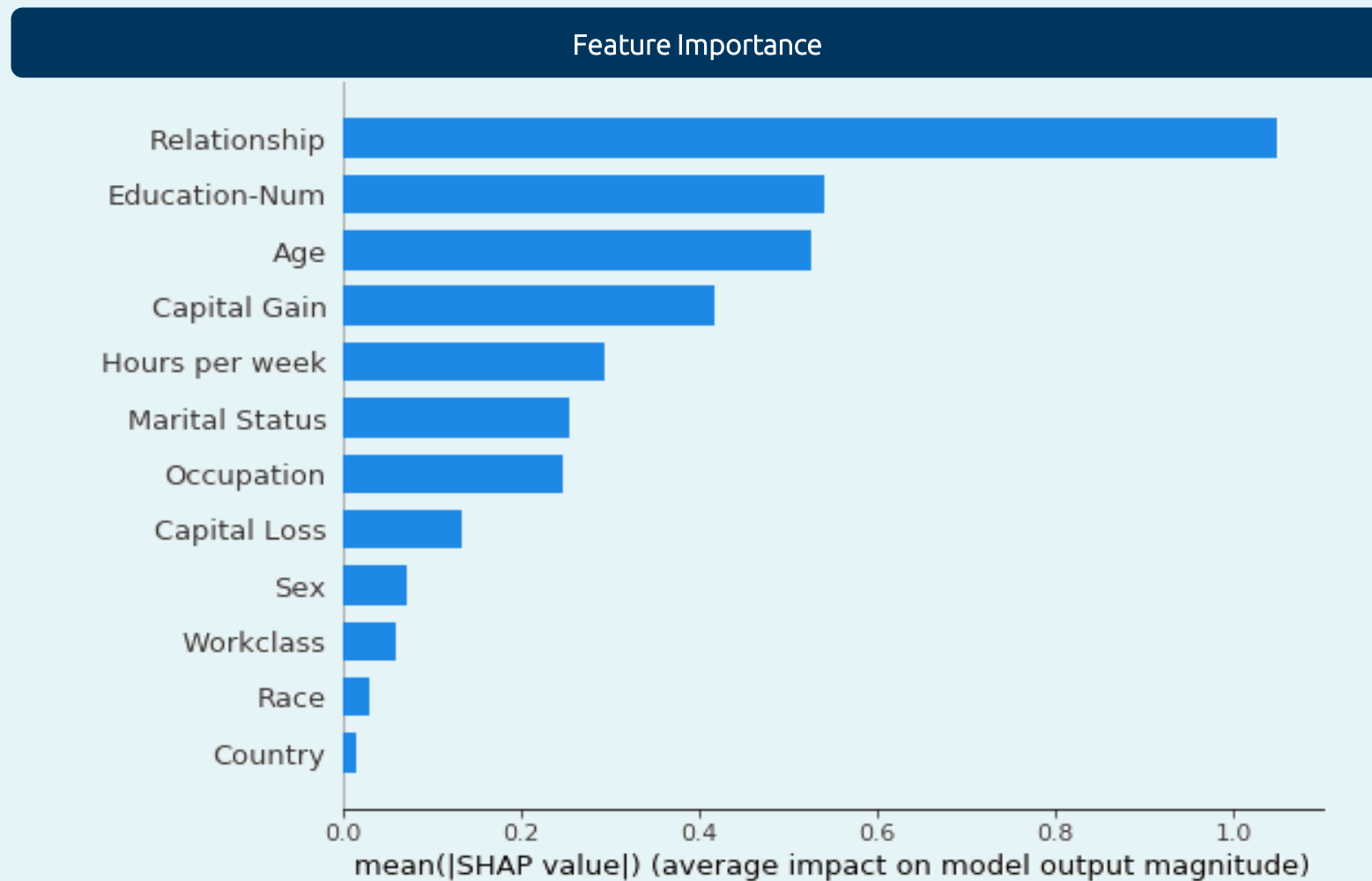
Shapley to the rescue!

Dependence Plot



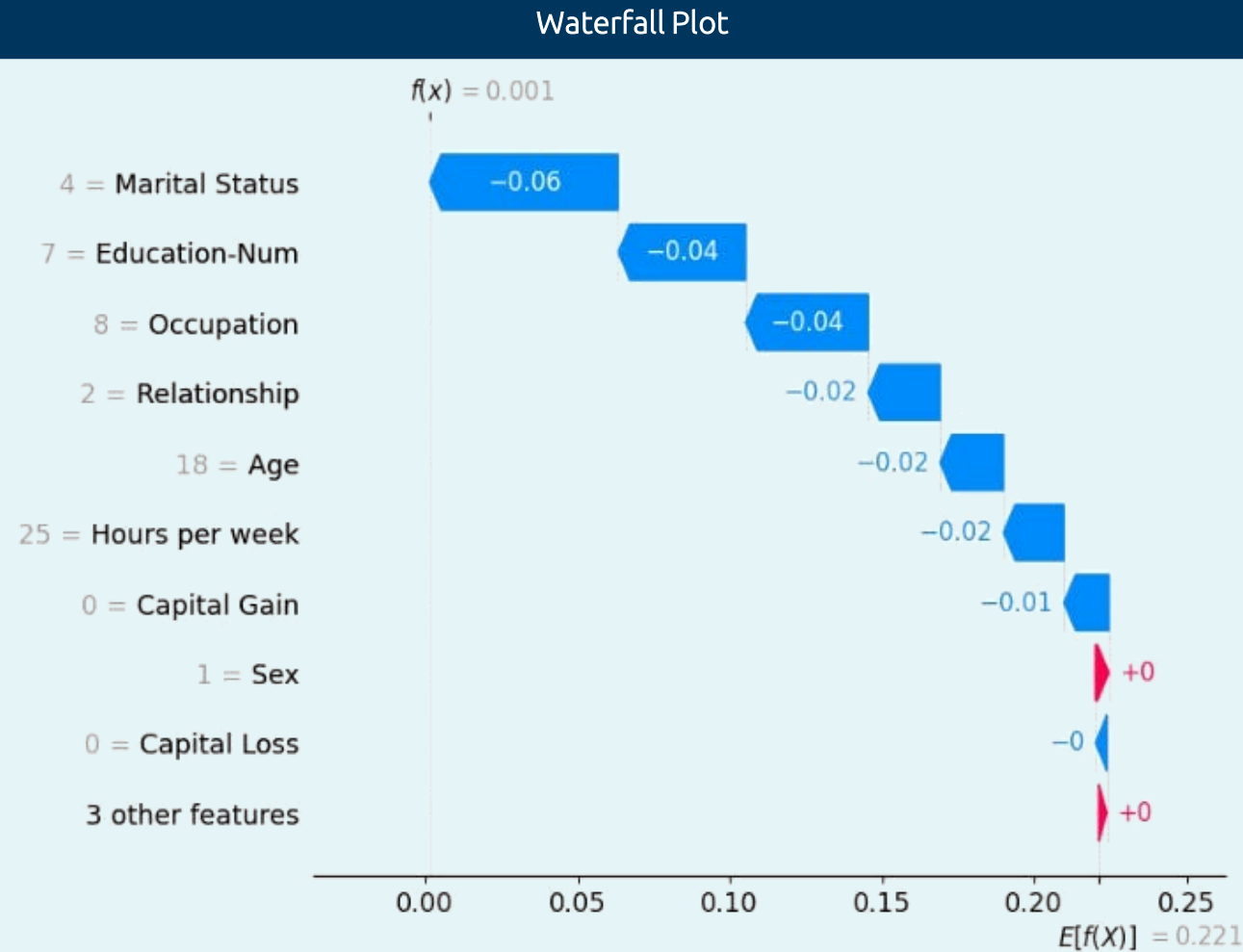
Uitlegbaarheid / Explainable AI

Shapley to the rescue!



Uitlegbaarheid / Explainable AI

Shapley to the rescue!



Overeenkomsten, verschillen en uitdagingen bij gebruik van GBMs

Hoe gaan we om met onderstaande stappen?



Data preparatie



Modelling



Expert Judgement



Uitlegbaarheid



Livegang

Livegang

Bestaande (premie)systemen zouden weleens de bottleneck kunnen zijn...

Overeenkomsten, verschillen en uitdagingen bij gebruik van GBMs

Hoe gaan we om met onderstaande stappen?



Data preparatie



Modelling



Expert Judgement



Uitlegbaarheid



Livegang

Key Takeaways

Het domein van AI is heel groot...

Neurale netwerken: zeer krachtig, maar zeer complex

Standaard voorspelprobleem? Gradient Boosting!

Insurance Pricing op basis van Gradient Boosting? Het is mogelijk!



Break



AI Pubquiz



Agenda

1

Opening & Recap

2

AI & Machine Learning

3

ML Pricing

Pauze

4

Data Management & Governance

5

AI risico's, regelgeving en ethiek

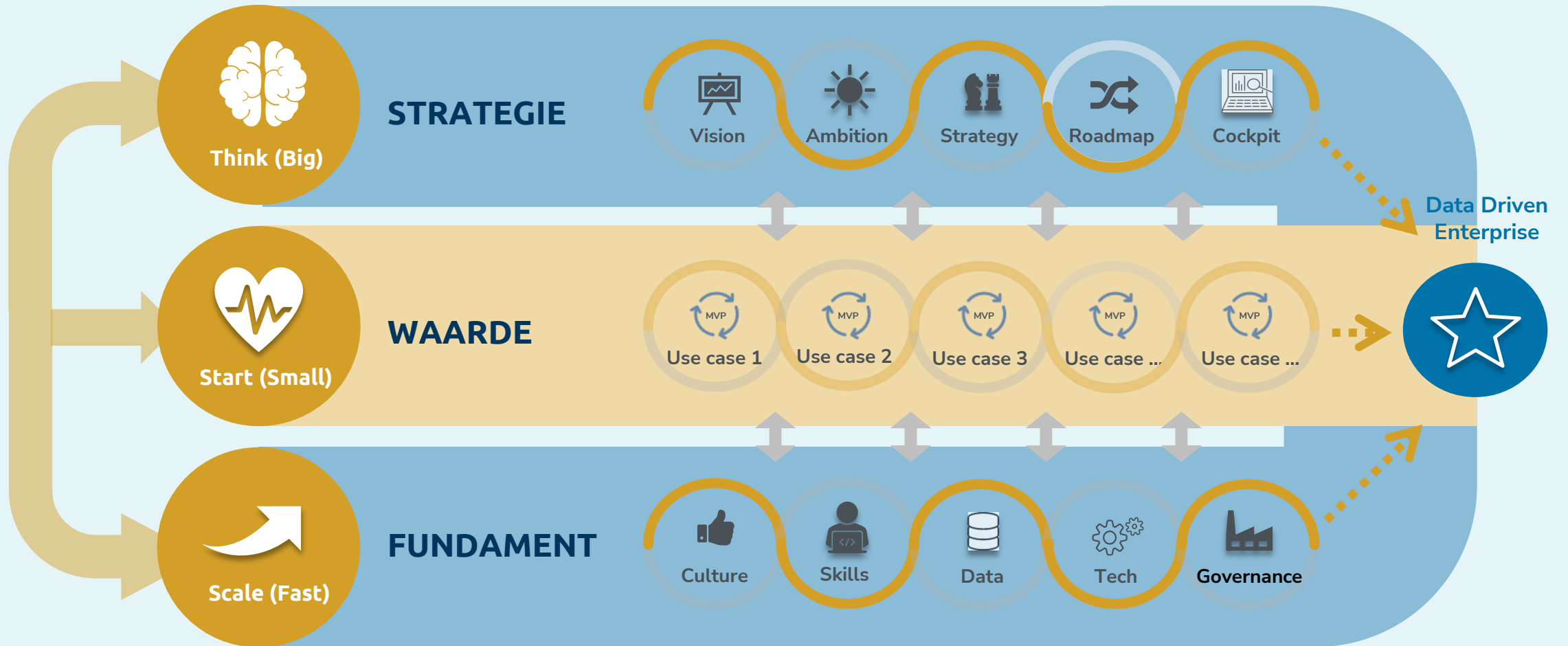
6

Afsluiting

“Data driven organisatie”



Werken aan datagedreven organisaties doe je op drie niveaus





Data Management



DATA MANAGEMENT IS HET FUNDAMENT VOOR DATAGEDREVEN WERKEN.

DATA VISIE

DATA PRINCIPLES

DATA ORGANISATIE

GOVERNANCE

AWARENESS

STEWARDSHIP

**POLICIES &
STANDARDS**

DATA QUALITY

DATA LIFECYCLE

DATA SECURITY



“Data Management is the development, execution and supervision of plans, policies, programs and practices that deliver, control, protect and enhance the value of data and information assets throughout their lifecycles.”

DAMA DMBOK, 2019



De rol van Data Management in een business context

Samenwerken in data management is essentieel voor hoge data kwaliteit wat bijdraagt aan de strategische behoeften van een organisatie



Data Management | Drivers

We herkennen twee typen 'drivers' voor Data Management: offensief en defensief.

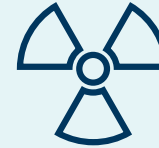
Offensief



Gericht op waarde creatie

- Data-gedreven besluitvorming
- Ontwikkeling van nieuwe business modellen
- Procesverbeteringen

Defensief



Gericht op vermijden van risico's

- Compliance gedreven (voldoen aan wet- en regelgeving)
- Inschatten en beheersen van risico's

DAMA DMBOK





Data Governance

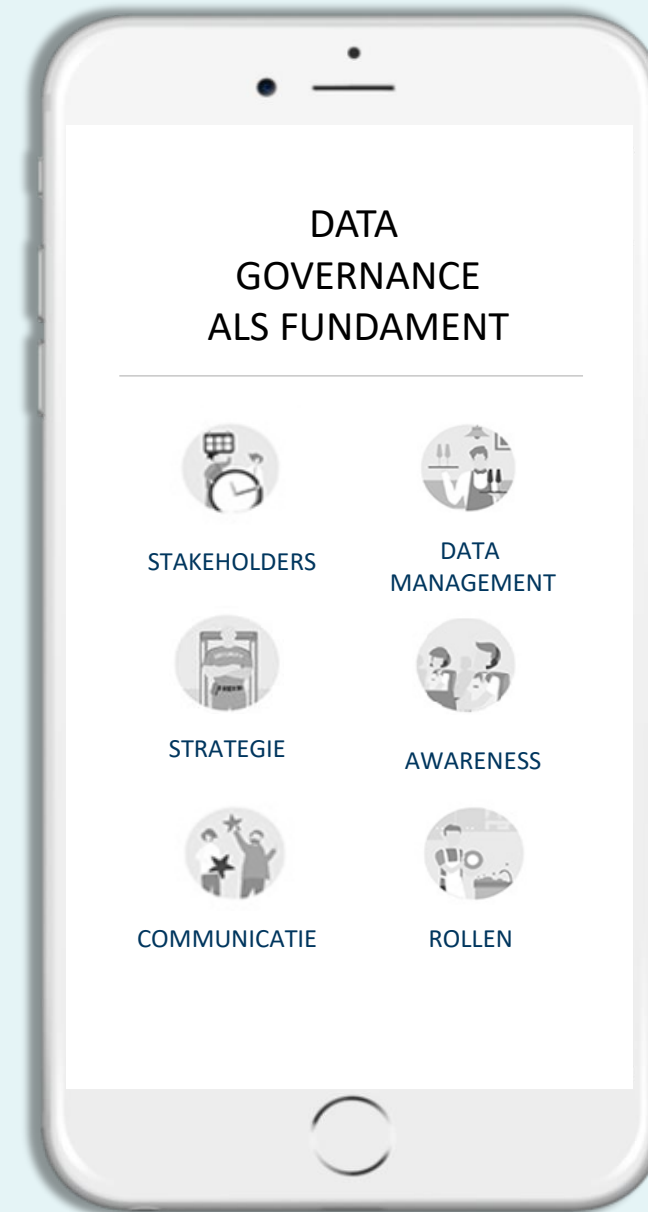


DATA GOVERNANCE

WAAR GAAT DATA GOVERNANCE EIGENLIJK OVER?
EN WAAROM IS HET NODIG OM DIT IN TE RICHTEN?

Data governance gaat om de organisatie en beheersing van data management activiteiten. Zorgen dat op de juiste plaats de verschillende onderdelen opgepakt worden.

Belangrijk hierbij: Stakeholder management, Strategie, visie, Awareness, Communicatie, organisatie, beheersing.



Data Governance | Waarom?



WET- EN REGELGEVING

Om te voldoen aan de steeds strengere wet- en regelgeving zijn governance structuren vereist. Ook verwacht de DNB dat de organisatie dit heeft ingericht.



DATAKWALITEIT

Het verbeteren van datakwaliteit is geholpen door een georganiseerde structuur en governance. Een van de doelen van data governance is het verbeteren van processen.



RISICO'S BEPERKEN

Naast het verbeteren van processen zorgt een goede data governance organisatie ook voor het beperken van risico's. Data security, privacy en risk management zijn onderdeel van data governance.



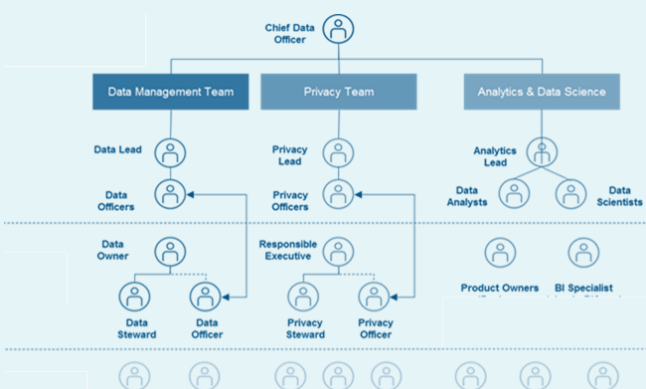
METADATA

Vastleggen van metadata vraagt een goede en heldere data governance. De verwevenheid van metadata maakt het complex en vraagt juist overlegstructuren, rollen en verantwoordelijkheden.

Data Governance | Inrichten van een data organisatie

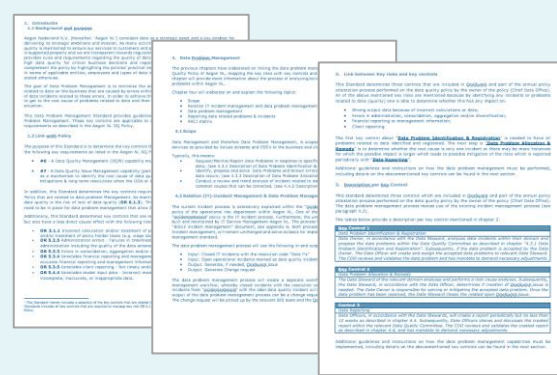
ORGANISATIE INRICHTING

- Het toekennen van eigenaarschap en het invullen van data management rollen
- Creëren van data awareness in de organisatie
- Aansluiting op bestaande structuur
- Overlegorganen instellen (DQB, DQC, DQ ops)
- Data domeinen toekennen op basis van business processen



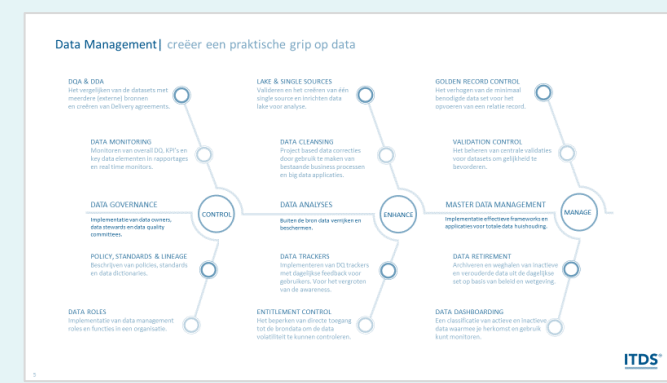
POLICIES & STANDARDS

- Het beschrijven van de gekozen data strategie en visie in beleidsstukken.
- Vertaling van policies naar standaarden en praktische guidelines
- Aansluiten van data standaarden op werkinstructies en bestaande processen

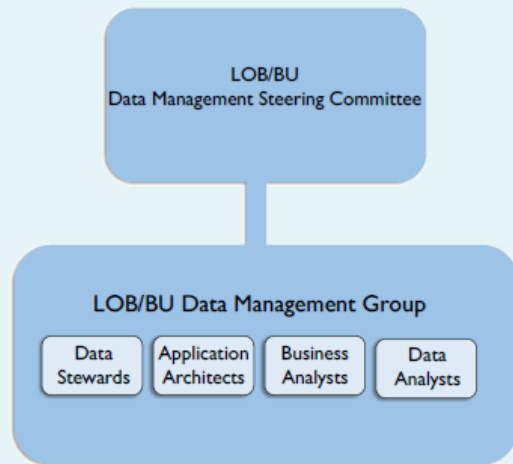


DATA STEWARDSHIP

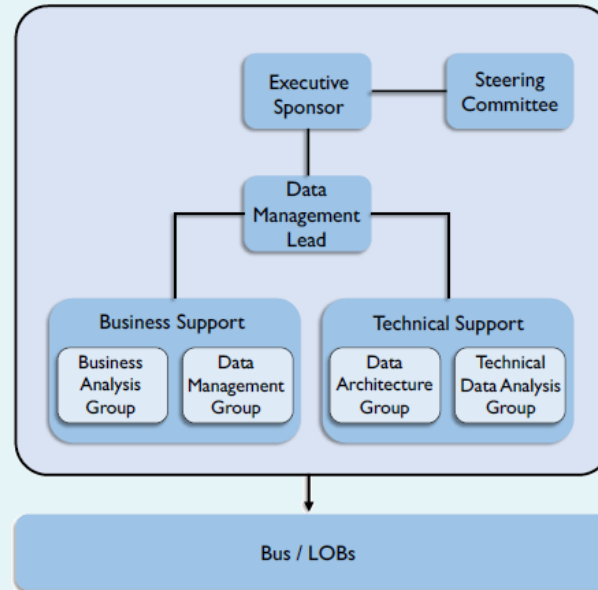
- Als onderdeel van de inrichting is een belangrijke rol weggelegd voor data stewards.
- Inzetten van de verschillende typen data stewards op diverse onderdelen
- Data steward council en overlegorganen
- Data awareness en praktische richting



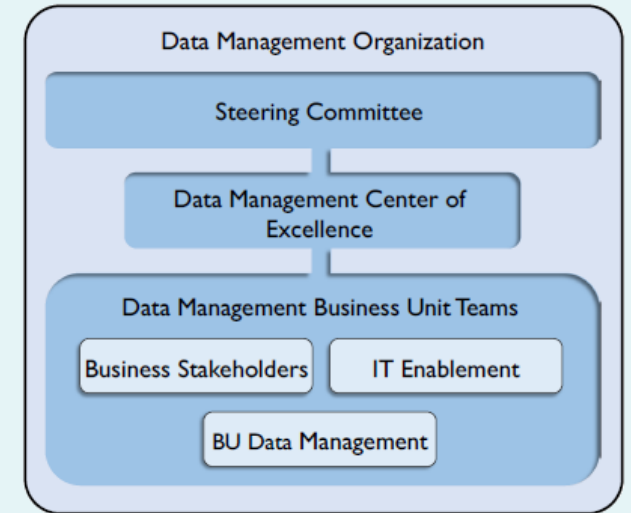
Data Governance | verschillende typen operating models



Decentraal operating model

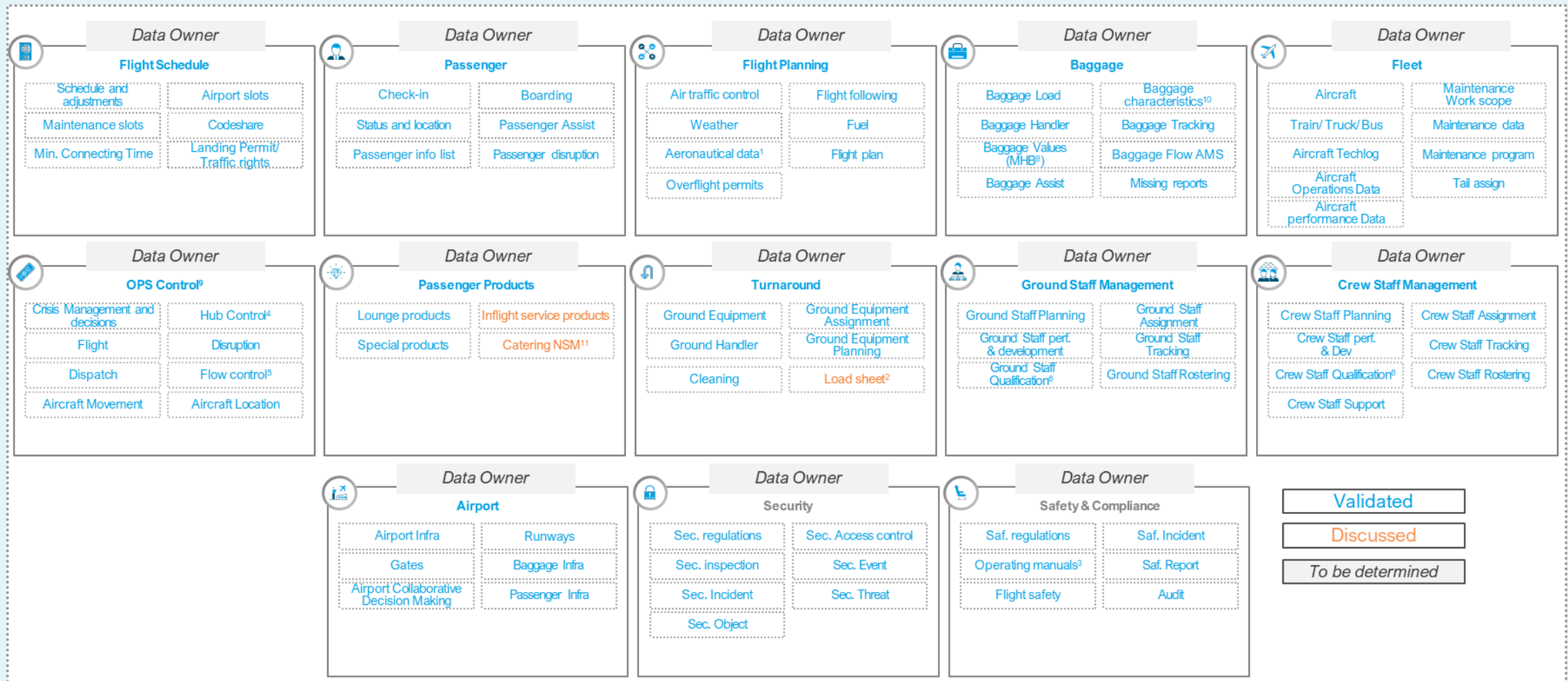


Centraal operating model



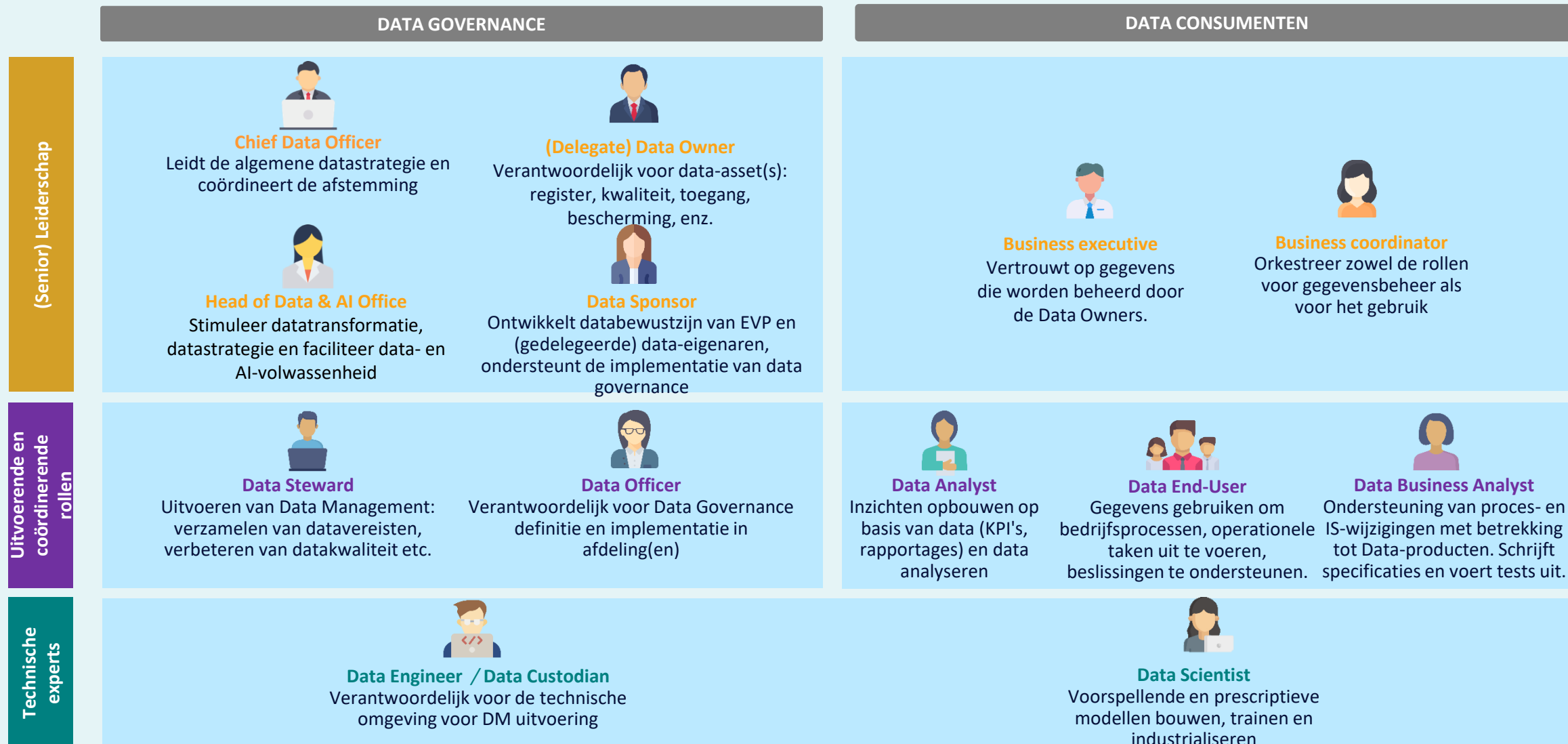
Hybride / Federated operating model

Data Governance | het inrichten van data domeinen en het bepalen van data owners



Een voorbeeld van een data domein map

Data Governance | Data governance rollen



Metadata

A person's hands are shown writing in a lined notebook. The right hand holds a blue pen and is in the process of writing 'workshop data' on a line. The left hand rests on the notebook. A semi-transparent dark blue banner is overlaid across the middle of the image, featuring the word 'Metadata' in white, bold, sans-serif font. The background is dark, and a portion of a laptop keyboard is visible on the left side.

Metadata Management | Wat is het?

WAT IS METADATA?

Metadata is 'data over data'. Het beschrijft (kenmerken van) de data. Metadata geeft context bij de data en zorgt ervoor dat data gebruikt kan worden voor waarde creatie.

WAT IS METADATA MANAGEMENT?

"Metadata Management is the planning, implementation and control activities to enable access to high quality integrated metadata."
- Dama DMBOK



Metadata Management | De bibliotheek metafoor

Je staat in een bibliotheek en bent op zoek naar een boek...
er staan er duizenden, dus hoe weet je waar je moet zoeken?

Er wordt gebruik gemaakt van een catalogus. In de catalogus zoek je bijvoorbeeld op:

Onderwerp, naam, auteur, locatie, beschikbaarheid, etc.

Organisaties zijn in feite digitale bibliotheken...

met duizenden digitale documenten, datasets, dashboards en meer. Vaak ontbreekt een logische catalogus. Dit maakt het moeilijk voor data analisten, scientists en andere rollen om de juiste data te vinden.

Daarom is gestructureerde **metadata** noodzakelijk:
Definities, eigenaarschap, locatie, beschikbaarheid, etc.

Metadata Management | Gebruikelijke 'frustraties' van data analisten / users



Veel tijd kwijt aan het zoeken van de juiste bronnen van een specifieke data set.



Het vinden van documentatie over data sets (de wijze van data collectie, hoe te interpreteren, wat zegt het). Dit kan leiden tot verkeerde interpretaties en re-work om juiste documentatie te vinden.



Het vinden van de juiste definities van data concepten of business definities.



Onduidelijkheid over de toegang tot bepaalde data sets of data velden. Mag je iemand toegang geven of mag je hiertoe toegang vragen?

Metadata Management | Typen metadata

Er bestaan veel verschillende typen Metadata, welke kan worden onderverdeeld in verschillende categorieën, afhankelijk van het doel en het type data dat wordt beschreven. Twee veel voorkomende voorbeelden:

Business metadata

Gaat over informatie die data en processen beschrijft in een business context.

Deze metadata helpt om data assets te managen, te verbeteren en om het beter in te zetten.

Voorbeelden:

Definities; business rules, berekeningen, GDPR/Security labels, governance informatie (zoals wie is de data steward?)

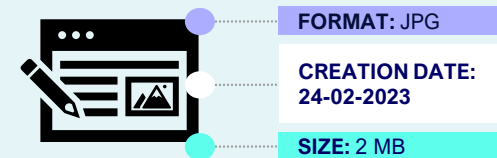


Technical metadata

Gaat over informatie over de technische kenmerken en detail van data, over het systeem dat data opslaat en over de uitwisseling van data tussen systemen.

Voorbeelden:

File format, grootte, database object kenmerken, data, ETL job details, data lineage documentatie



Waar of niet waar?



68% van de financiële instellingen weten precies hoeveel data ze opslaan.

Niet waar! 68% weet juist NIET hoeveel data ze precies opslaan.
Bron: Data in Finance Magazine 2023



Gemiddeld besteden financiële instellingen **€50.000** tot **€100.000** per jaar aan Data Management.

Niet waar! Gemiddeld besteden financials 100.000 – 500.000 jaarlijks aan Data Management.
Bron: Data in Finance Magazine 2023



Slecht data management kost een organisatie gemiddeld **€ 9.700.000,-** euro per jaar.

Waar! Volgens Gartner 2019

Agenda

1

Opening & Recap

2

AI & Machine Learning

3

ML Pricing

Pauze

4

Data Management & Governance

5

AI risico's, regelgeving en ethiek

6

Afsluiting

AI risico's, regelgeving en ethiek



De actualiteit

■ NIEUWS

Autoriteit Persoonsgegevens roept minister Hoekstra op het matje om algoritme

Profilerende software Minister Hoekstra wordt door de Autoriteit Persoonsgegevens ter verantwoording geroepen over het gebruik van algoritmen die tot discriminatie kunnen leiden.

Merijn Rengers, Carola Houtekamer
1 mei 2023 • Leestijd 2 minuten



■ ACHTERGROND

LET OP, zegt de computer van Buitenlandse Zaken bij tienduizenden visumaanvragen. Is dat discriminatie?

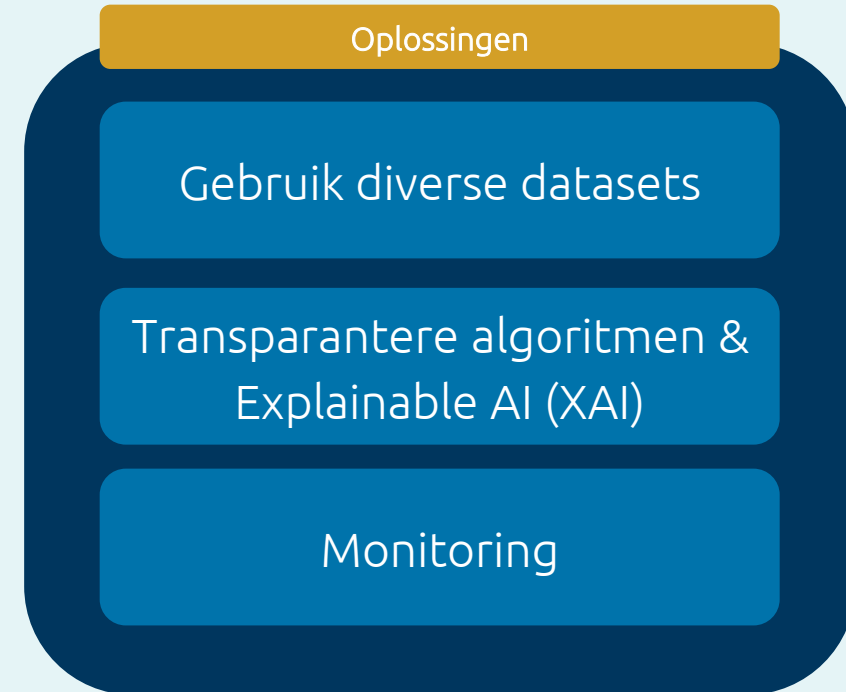
Discriminerend algoritme Volgens een onderzoek discrimineerde het algoritme dat Buitenlandse Zaken gebruikt om visumaanvragen te beoordelen. Uit onvrede met die conclusie vroeg het ministerie om een second opinion.

Carola Houtekamer, Merijn Rengers
1 mei 2024 • Leestijd 6 minuten



AI Risico – historische vooringenomenheid (bias) en mispresentatie

- Ontstaat door te trainen op gegevens die niet representatief zijn voor de doelgroep
- Developers zelf onbewust of vooringenomen



Artificial Intelligence Act – wat houdt het in?

“De AI-act is een Europese verordening en een belangrijke ontwikkeling in het reguleren van AI binnen de Europese Unie. Dit juridisch kader is ontworpen om de opkomst van AI te ondersteunen terwijl het essentiële normen voor veiligheid, ethiek en rechten waarborgt”

1

Fundamentele elementen

3

Risicobeoordeling

2

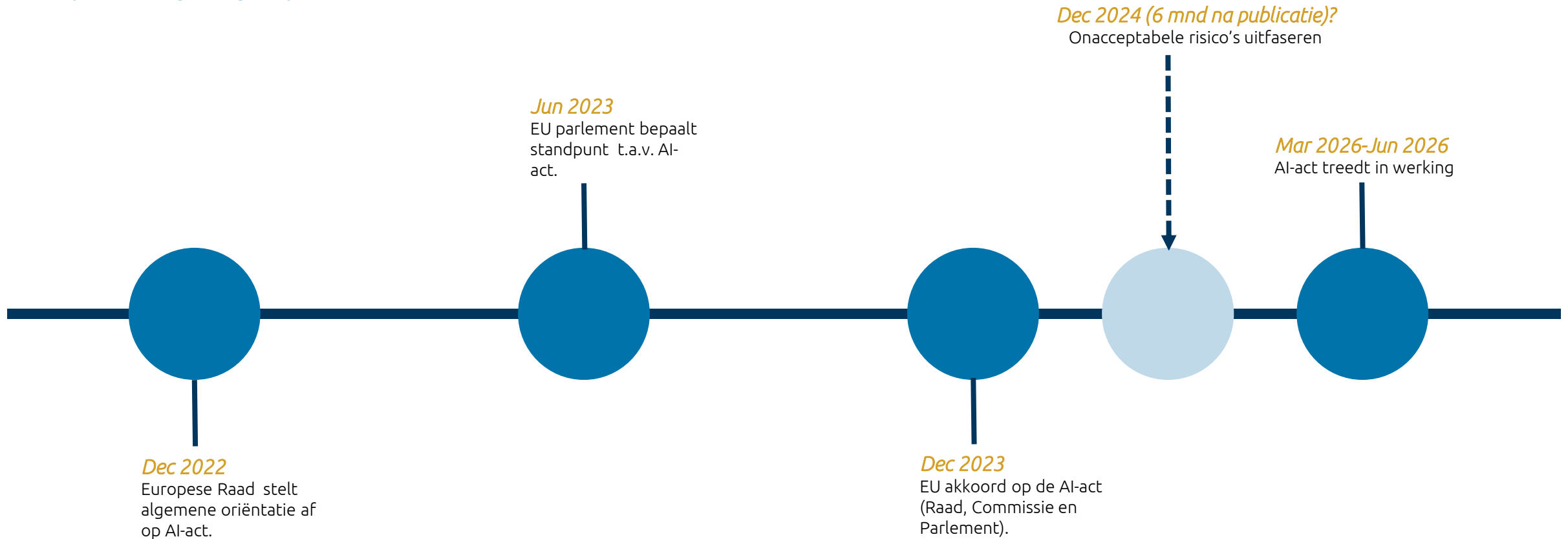
Doelstellingen

4

Impact

Artificial Intelligence Act

Europese wetgeving en juridisch kader omtrent alle inzet van AI



AI Act: Fundamentele elementen

1
Fundamentele elementen

3
Risicobeoordeling

2
Doelstellingen

4
Impact

Fundamenten AI act

Evenwichtig & evenredig

Innovatie stimuleren en tegelijkertijd risico's beheersen

(Tegen) verboden praktijken

Verbod op zaken als discriminatie en manipulatie

Risico beoordeling

AI systemen moeten een risicobeoordeling doorlopen

Governance & handhaving

Zowel nationaal als Europees toezicht op AI implementatie

AI Act: Doelstellingen

1
Fundamentele elementen

3
Risicobeoordeling

2
Doelstellingen

4
Impact

Veiligheid van AI-tools waarborgen

Normen vaststellen en menselijke & technische fouten minimaliseren

Grondrechten EU-burgers in acht nemen

Privacybescherming, non-discriminatie, transparantie en soortgelijke grondrechten

Stimuleren van innovatie en investeringen

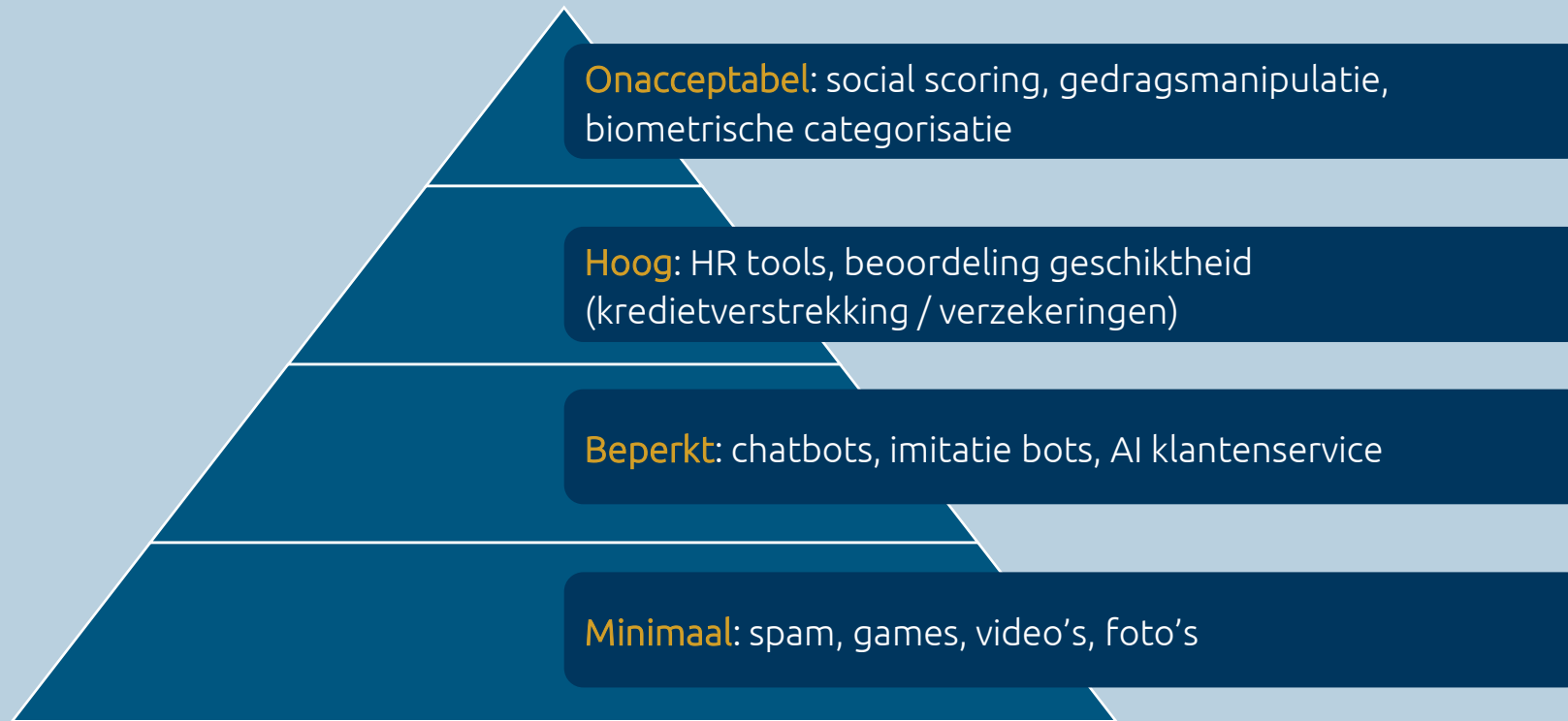
Stimuleren onder een evenwichtige regulering

Handhaving bestaande wet- en regelgeving

Coherentie met bestaande wetgeving, inclusief gegevensbescherming

AI Act: Risico beoordeling

Een risicogebaseerde benadering



1
Fundamentele elementen

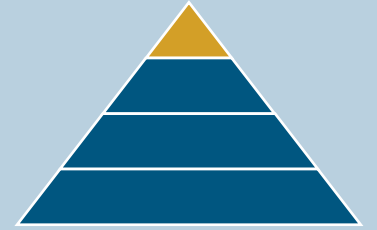
2
Doelstellingen

3
Risicobeoordeling

4
Impact



AI Act: **onacceptabel** risico



Binnen 6 maanden uitfaseren

Voorbeeld: sociale scoresystemen

Deze systemen kunnen mensen beoordelen op basis van hun online gedrag en hen uitsluiten van toegang tot krediet, banen of huisvesting

AI Act: **hoog** risico

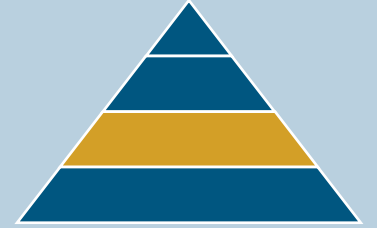


Binnen 1 jaar transparantie en documentatie
Binnen 2 jaar compliant

Voorbeeld: AI gebaseerde kredietscore modellen

Deze scores kunnen onnauwkeurig zijn en leiden tot discriminatie
van bepaalde groepen mensen

AI Act: **beperkt** risico



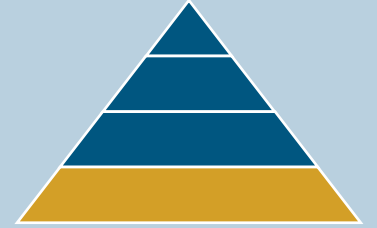
Binnen 1 jaar transparantie en documentatie
Binnen 2 jaar compliant

Voorbeeld: AI gebaseerde klantenservice

Deze chatbots kunnen onbetrouwbare informatie verstrekken en
tegen het klantbelang in



AI Act: **minimaal** risico



Transparantie opbouwen in de tool

Voorbeeld: AI Exploratory Data Analysis tool

Kan een foutieve en onnauwkeurige analyse uitvoeren



AI Act: Impact

Risico beoordeling & non-compliance

1
Fundamentele elementen

3
Risicobeoordeling

2
Doelstellingen

4
Impact

Risico beoordeling

AI systemen die in de beperkte, hoge of onacceptabel categorie vallen, moeten een risico beoordeling doorlopen.

Non-compliance

Boetes variërend van 35 miljoen of 7% van de wereldwijde omzet bij gebruik verboden AI tot 7,5 miljoen of 1,5% van de totale omzet bij onjuiste informatie risicostatus



Handhaving, wat mogen toezichthouders?

Toegang eisen tot alle
documentatie, broncode en
model parameters

Bevel tot staking geven indien
toch te risicovolle tool

Bindend aanwijzingen geven over
aangepast gebruik

Boetes opleggen



Normenkaders: Zelfregulering in Verzekering en Banking Sector

Ethisch Kader voor datagedreven besluitvorming

Ethisch kader door verzekeraars door VvV



Kader gebaseerd op aanbevelingen adviesorgaan EC (Europese Commissie)

Voor **data** gedreven toepassingen

'High-level' **vereisten** voor het ethisch en verantwoord gebruik

Uitleg over **eisen** voor balans innovatie en (rechts)zekerheid

Vanaf 1 jan 2021 onderdeel van vernieuwde zelfregulering

Toetsing door Stichting toetsing verzekeraars, eerste audits van 2023

AI in de bancaire sector door NVB

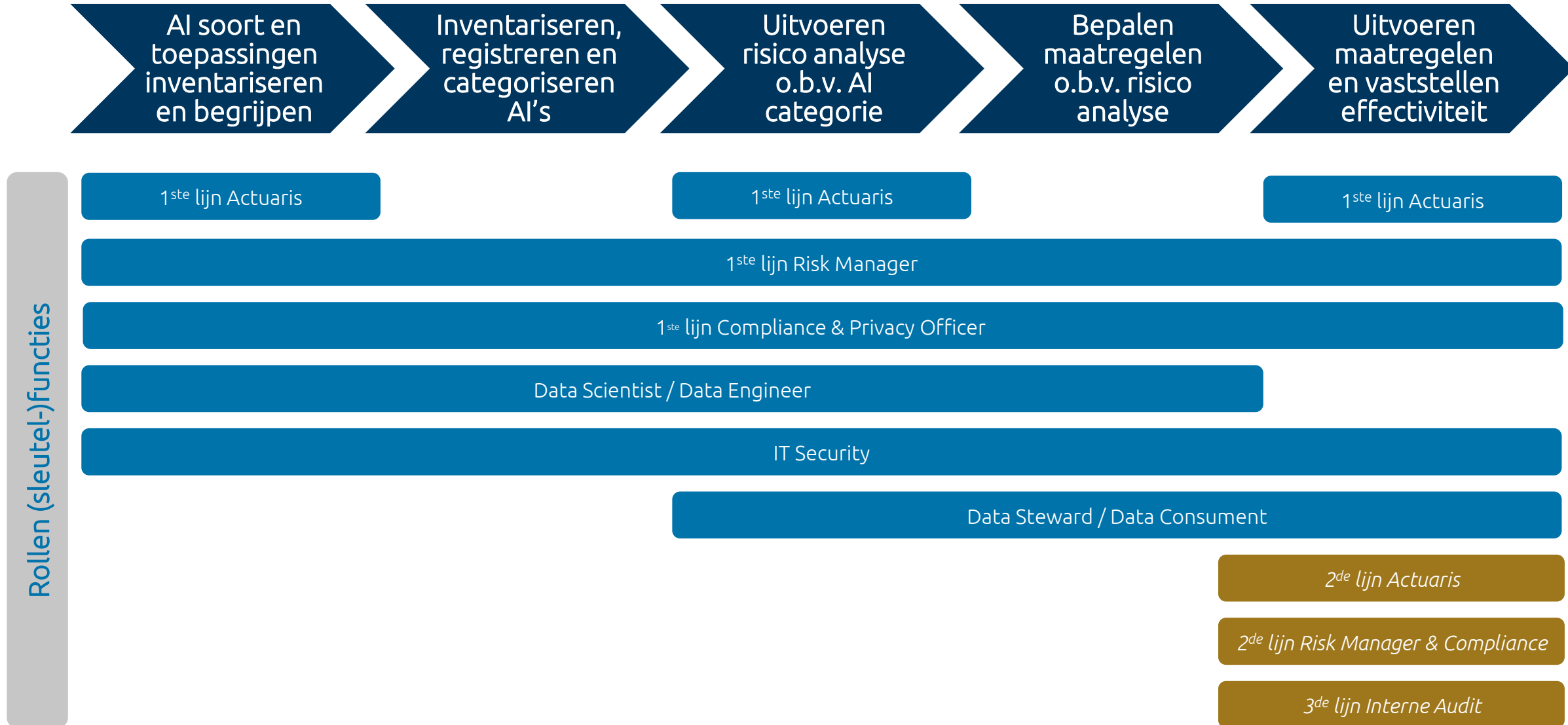


Voor **AI** toepassingen

'High-level' **principes** voor het ethisch en verantwoord gebruik

Uitleg over **spelregels** voor balans innovatie en (rechts)zekerheid

Wrap-up en next steps



Voorstel: Data (Officer) Functie

Wij stellen voor om de bestaande Actuariële functie, Risk Management functie en Interne Audit functie uit te breiden met een nieuw te vormen Data (Officer) functie



Agenda

1

Opening & Recap

2

AI & Machine Learning

3

ML Pricing

Pauze

4

Data Management & Governance

5

AI risico's, regelgeving en ethiek

6

Afsluiting

Lessons learned: key take aways

- AI zal het werk van actuarissen blijvend beïnvloeden:
 - Benut AI voor o.a. automatisering operationele taken
 - Durf te experimenteren (bv. adopteer ML technieken binnen Pricing)
- Data governance om kwaliteit en bronnen te valideren is cruciaal
 - Organiseer data management onderdelen
 - Ken data rollen en verantwoordelijkheden toe
- AI Act wordt ingevoerd
 - Breng AI risico's in kaart en ga er actief op sturen
 - Richt Data (Officer) Functie in

En...ontwikkel je “niet-pluis” intuïtie



Bedankt voor uw aandacht

Bezoekadres

Hoogoorddreef 54

1101 BE Amsterdam Zuidoost

www.aaa-riskfinance.nl

info@aaa-riskfinance.nl