

PE Data Sessie

Sessie 1



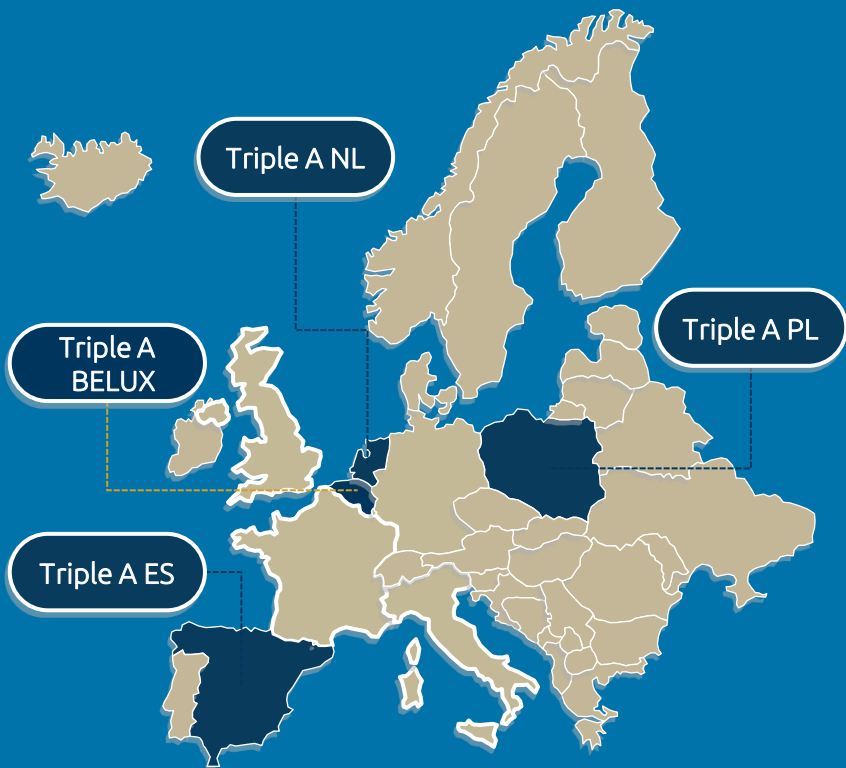
TRIPLE A

RISK FINANCE

19 maart 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)

VOETTEKST NOG TOEVOEGEN



TRIPLE A
RISK FINANCE

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



Joey de Mol

2017 - heden
Senior Consultant
Data & Analytics

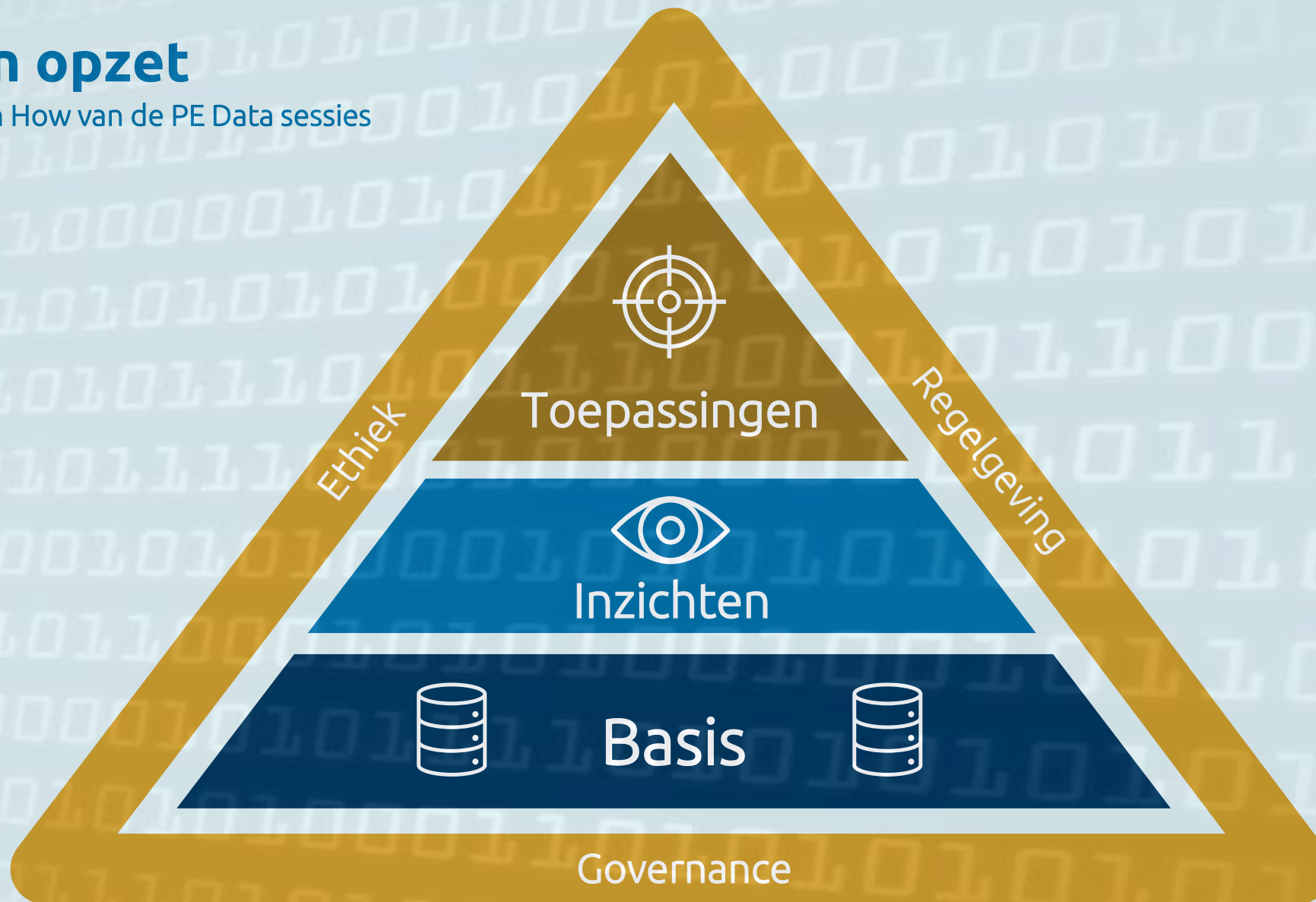


Cédric Jonckers

2022 - heden
Medior Consultant
Data & Analytics

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"



Sessie 1



Sessie 2



Agenda

1

Opening & Overview

2

Trends & Ontwikkelingen

3

Databronnen

Pauze

4

Datakwaliteit

5

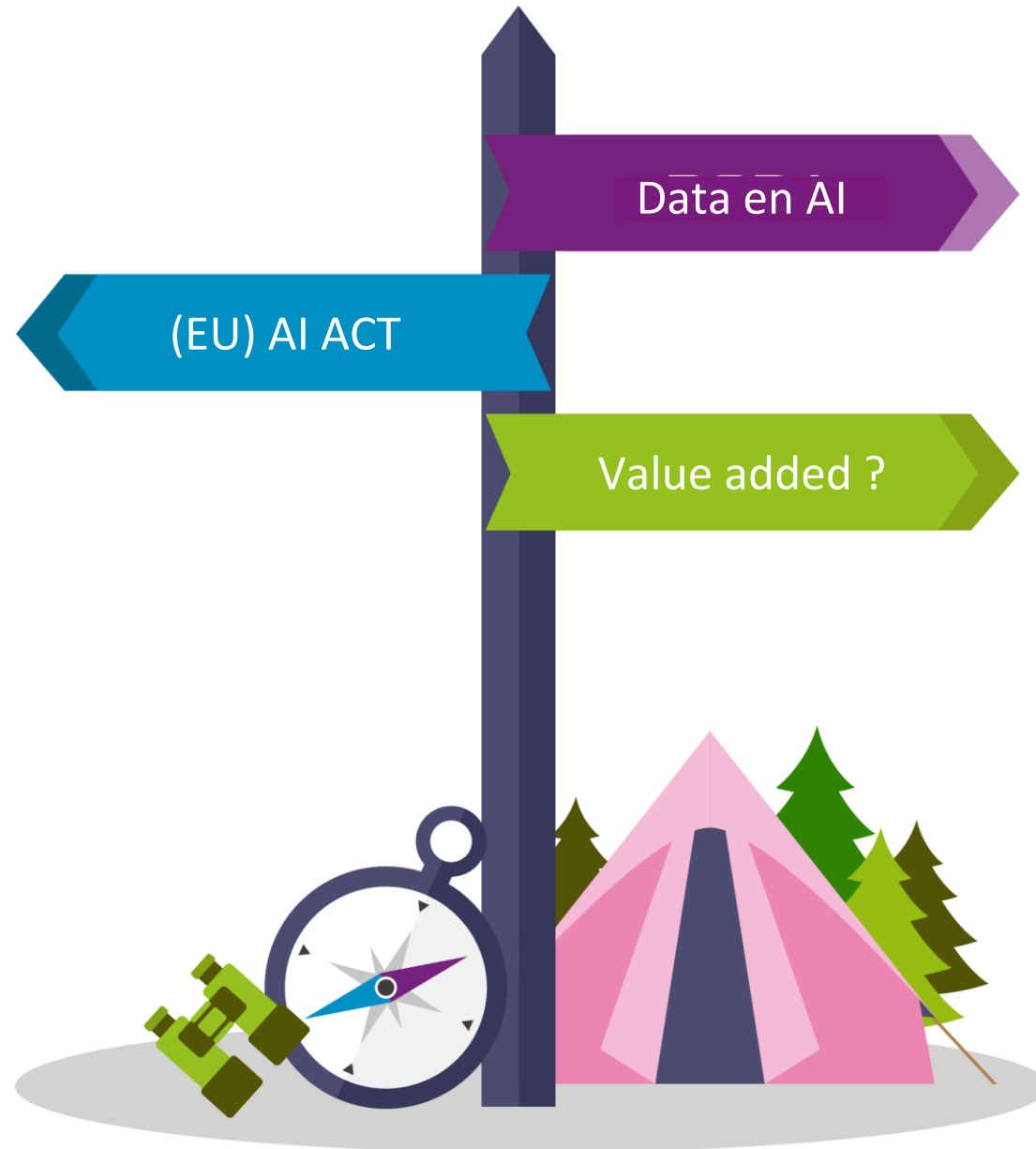
Architectuur

6

Data Visualisatie

DATA DETECTIVES

19 maart 2024



Welcome, IBM. Seriously.


Welcome to the most exciting and important marketplace since the computer revolution began 35 years ago. And congratulations on your first personal computer. Putting real computer power in the hands of the individual is already improving the way people work, think, learn, communicate and spend their leisure hours. Computer literacy is fast becoming as fundamental a skill as reading or writing.

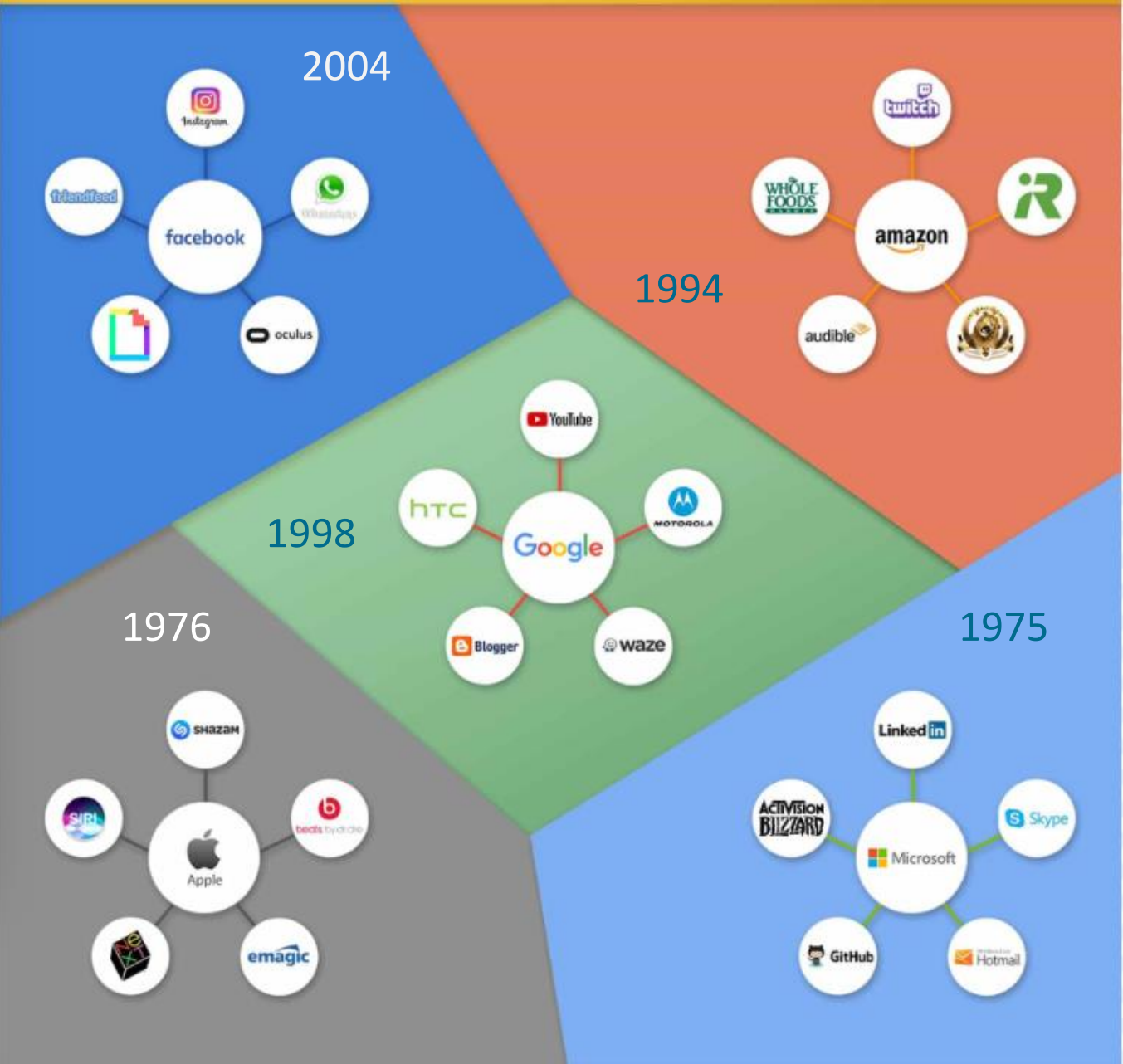
When we invented the first personal computer system, we estimated that over 140,000,000 people worldwide could justify the purchase of one, if only they understood its benefits.

Next year alone, we project that well over 1,000,000 will come to that understanding. Over the next decade, the growth of the personal computer will continue in logarithmic leaps.

We look forward to responsible competition in the massive effort to distribute this American technology to the world. And we appreciate the magnitude of your commitment.

Because what we are doing is increasing social capital by enhancing individual productivity.

Welcome to the task.  **apple**



AI 'could be as transformative as Industrial Revolution'

UK's outgoing chief scientist urges ministers to 'get ahead' of profound social and economic changes

Hannah Devlin *Science correspondent*

Wed 3 May 2023 17.54 CEST

Share



Commuters in London. Sir Patrick Vallance told MPs 'there will be a big impact on jobs' from generative AI. Photograph: Bloomberg/Getty

The new genre of AI could be as transformative as the Industrial Revolution, the government's outgoing chief scientific adviser has said, as he urged Britain to act immediately to prevent huge numbers of people becoming jobless.

<https://www.theguardian.com/technology/2023/may/03/ai-could-be-as-transformative-as-industrial-revolution-patrick-vallance>

[← Terug naar Krant](#)

5 mrt 16:42

Accountants Deloitte doen volgend jaar alle controles met behulp van AI


 Martijn Pols, Edwin van der Schoot


Het kantoor van Deloitte aan de Amsterdamse Zuidas, genaamd The Edge. Foto: Robin Utrecht/ANP

In het kort

- Deloitte wil in Nederland al zijn accountants voor het einde van het jaar met AI laten werken.
- Volgens het hoofd van de accountants Gera Hamer zijn er grote voordelen, zoals de aantrekkelijkheid van het vak.
- Toch is het tekort aan accountants niet zomaar weggepoetst met hulp van AI, waarschuwt Deloitte.

Personeelstekort oplossen

De opmars van zelflerende AI roept de vraag op of daarmee het huidige tekort aan controlerend accountants kan worden opgelost. Computers kunnen, als ze eenmaal goed genoeg geschoold zijn, het werk ongetwijfeld deels overnemen. 'Niet direct', zegt bestuurder Hamer. 'De hoeveelheid werk voor accountants stijgt door steeds meer en complexere regelgeving met zo'n 30%. Maar AI kan het werk wel aantrekkelijker maken, ook voor de volgende generatie studenten.'

Internationaal woedt de discussie hoe ver AI mag gaan in de accountantscontrole. Vooralsnog staan de menselijke blik én handtekening voorop in de controle. 'Ik kan me voorstellen dat ik het nog mee ga maken dat AI bijvoorbeeld de continuïteit van een bedrijf kan beoordelen', zegt Hamer echter.

Wie denkt dat AI daarmee uiteindelijk de accountant volledig gaat vervangen, zal bedrogen uitkomen, voorspelt zij. De AI-tool is nadrukkelijk enkel sparringpartner of hulpmiddel, de accountant blijft achter het stuur zitten en moet kritisch nakijken.

Kan de rekening dan omlaag voor klanten? Dat is 'te kort door de bocht', zegt Hamer. 'De rekening bestaat uit twee componenten, uren en gebruik van middelen. De verhouding tussen die twee kan veranderen. En de accountant houdt meer tijd over voor complexere zaken.'

Prompt antwoord

In een korte demonstratie van Headstart laten Hamer, chief technology officer Marc Verdonk en Van Gool aan het FD zien hoe het AI-programma in pakweg vijftien seconden constateert dat de verantwoording van de derivatenpositie in het jaarverslag van een willekeurig gekozen nutsbedrijf op 3 van 10 punten deels of geheel onvoldoende is.

Het is de computer die de tekst vergelijkt met de wettelijke vereisten, in dit geval de Nederlandse boekhoud- en controlestandaarden. Nog eens vijftien seconden later en één prompt (zoals een vraag of vervolgvraag in AI-terminen heet) verder, geeft Headstart een uitgebreide onderbouwing van die stelling.

Het model wordt verder ontwikkeld door de input van de Deloitte-werknemers die het gebruiken. Verdonk: 'De truc is dan ook dat we iedereen dit laten gebruiken, ook op de juiste manier, om een sneeuwbaaleffect te bewerkstelligen. Iedereen moet dus weten hoe je in de basis een goede prompt invoert.'

Artificial Intelligence kan niet zonder Actuarial Intelligence



DIENSTEN OPLOSSINGEN

Publicaties > Actuarial Intelligence

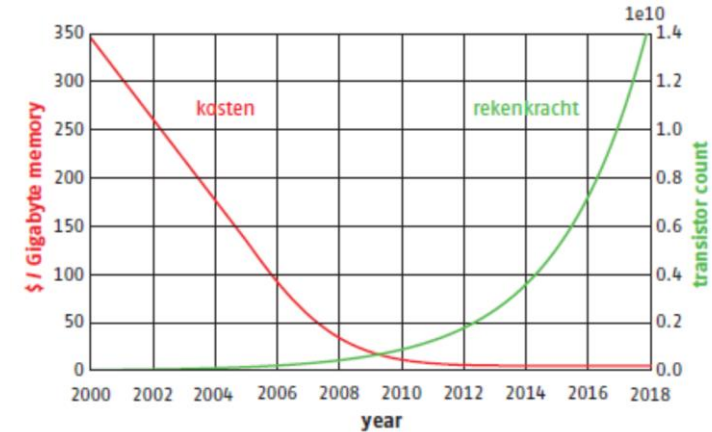
Actuarial Intelligence



Gepubliceerd: 4 maart 2019
Marktsector Verzekeraars

NEEM CONTACT OP

Onze beroepsgroep is te terughoudend bij het verkennen én toepassen van nieuwe modellen en technieken. Een gemiste kans, want **Artificial Intelligence** kan niet zonder Actuarial Intelligence.



Figuur 1: meer rekenkracht tegen lagere kosten (Bron: Wikipedia en hblok.net)

“Uiteindelijk draait AI niet om het bouwen van complexe modellen, maar om het oplossen van een probleem met behulp van data”

Artificial Intelligence kan ook niet zonder **Good Quality Data**



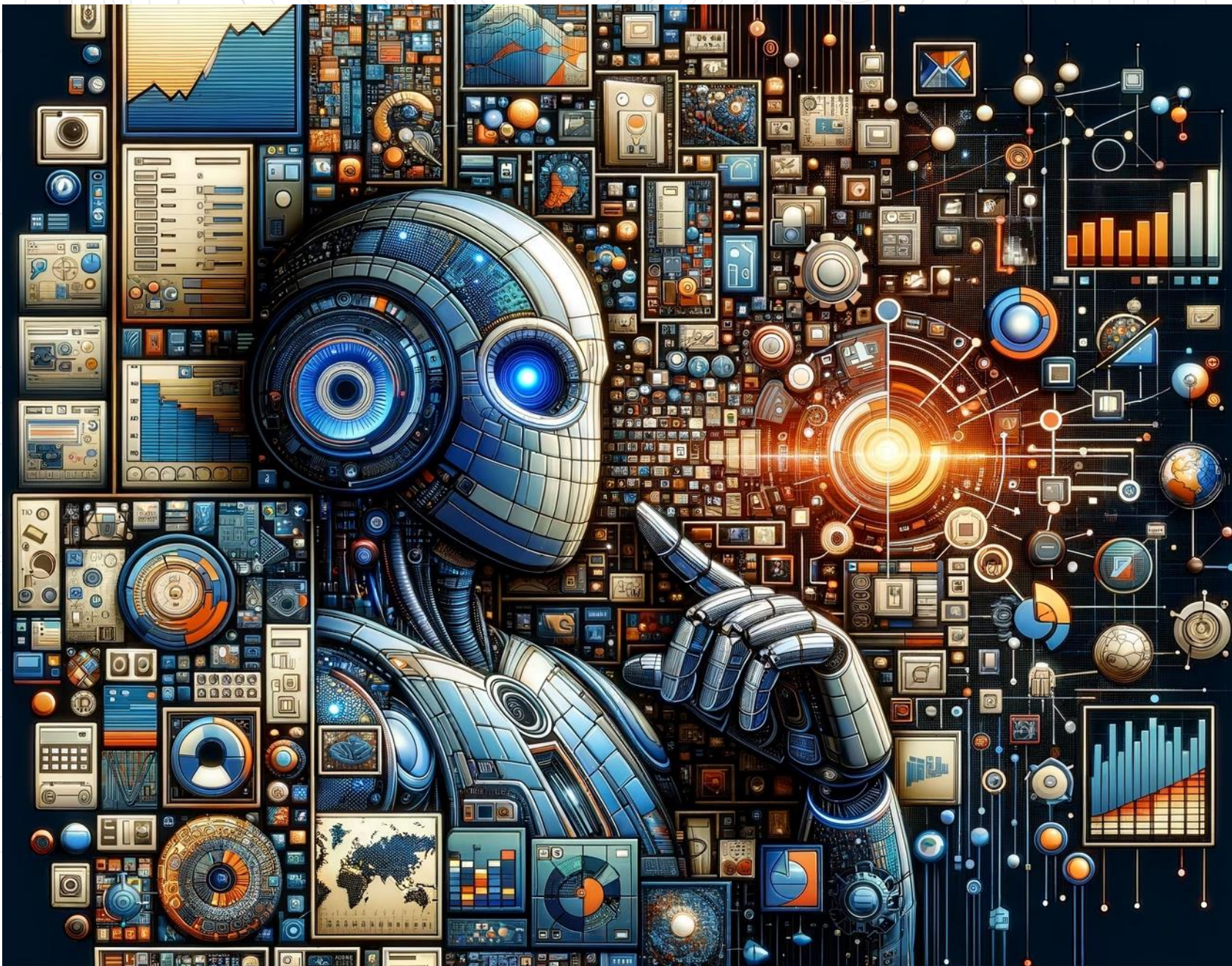
HELPING TO MAKE FUNDAMENTAL RIGHTS
A REALITY FOR EVERYONE IN THE EUROPEAN UNION

Data quality and artificial intelligence – mitigating bias and error to protect fundamental rights

FRA Focus

Algorithms used in machine learning systems and artificial intelligence (AI) can only be as good as the data used for their development. High quality data are essential for high quality algorithms. Yet, the call for high quality data in discussions around AI often remains without any further specifications and guidance as to what this actually means. Since there are several sources of error in all data collections, users of AI-related technology need to know where the data come from and the potential shortcomings of the data. AI systems based on incomplete or biased data can lead to inaccurate outcomes that infringe on people's fundamental rights, including discrimination. Being transparent about which data are used in AI systems helps to prevent possible rights violations. This is especially important in times of big data, where the volume of data is sometimes valued over quality.

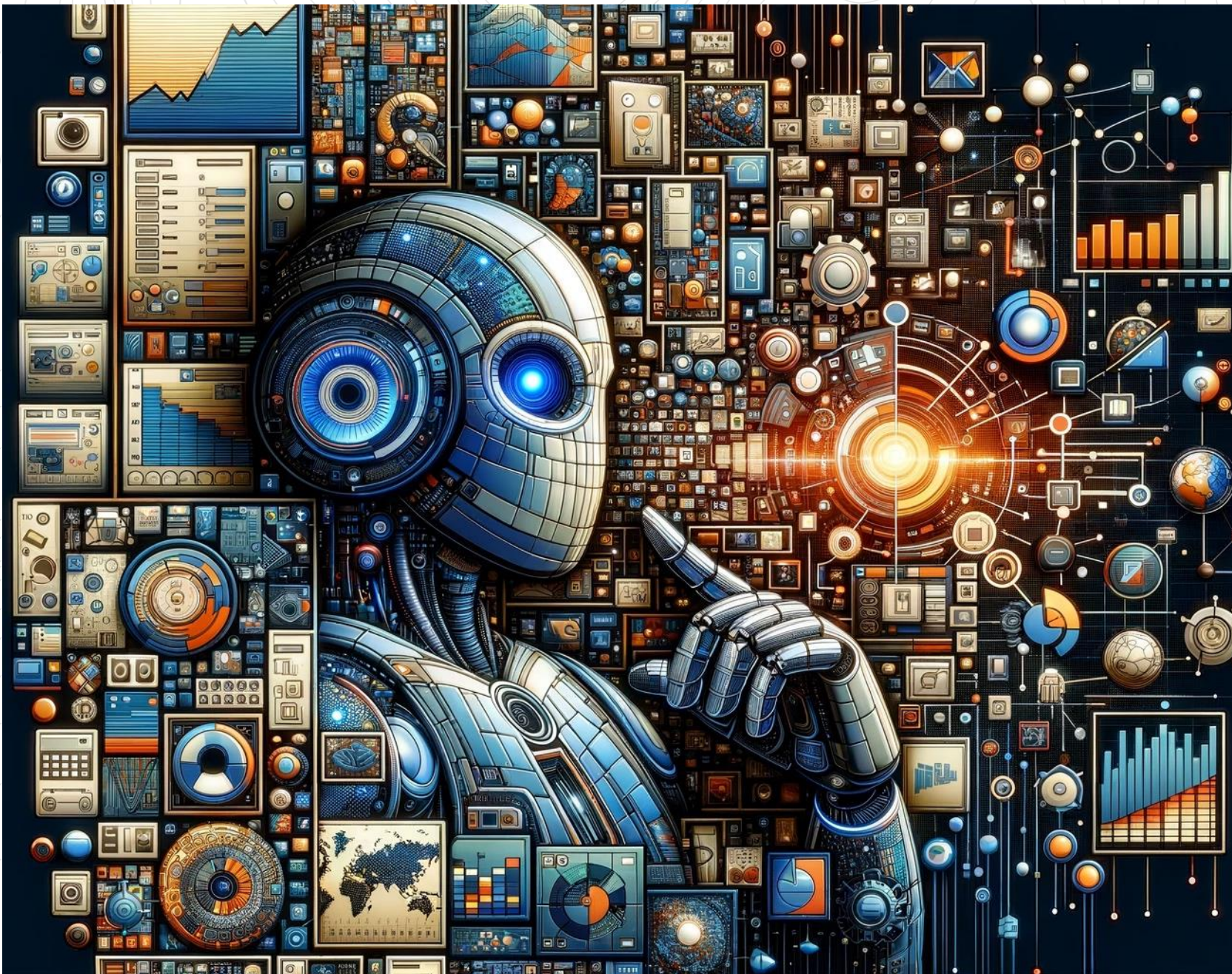
https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-data-quality-and-ai_en.pdf



DATA QUALITY DIMENSIONS

1. Accuracy
2. Completeness
3. Consistency
4. Integrity
5. Reasonability
6. Timeliness
7. Uniqueness
8. Validity

Source: DAMA DMBOK2



ADDITIONAL RISKS:

1. Risk of Misrepresentation
2. Risk of Historical Bias
3. Risk of Overfitting
4. Risk of Underfitting
5. Risk of Data Poisoning
6. Risk of Privacy Violations
7. Risk of Security Vulnerabilities
8. Risk of Non-Compliance

ADDRESSING THESE RISKS:

- Responsible data cleansing
- Continuous auditing data quality
- Techniques to mitigate bias
- Data security & privacy
- Techniques to mitigate fitting risks

Maatschappelijke zorgen over het gebruik van data

HEALTH INSURANCE HUSTLE

Health Insurers Are Vacuuming Up Details About You — And It Could Raise Your Rates

Without any public scrutiny, insurers and data brokers are predicting your health costs based on data about things like race, marital status, how much TV you watch, whether you pay your bills on time or even buy plus-size clothing.

Marshall Allen, July 17, 2018, 5 a.m. EDT



HEALTH INSURANCE HUSTLE
The Confounding Way We Pay for Care

This story was co-published with NPR.

To an outsider, the fancy booths at last month's health insurance industry gathering in San Diego aren't very compelling. A handful of companies pitching "lifestyle" data and salespeople touting jargony phrases like "social determinants of health."

But dig deeper and the implications of what they're selling might give many patients pause: A future in which everything you do — the things you buy, the food you eat,

the time you spend watching TV — may help determine how much you pay for health insurance.

The screenshot shows a webpage from CCW (Customer Contact Week). The article title is "When Big Data Goes Too Far: A Look Inside the Insurance Industry". The sub-headline reads: "Insurance companies are using data on race, marital status and what you watch on TV to assess risk". The author is Kindra Cooper, dated 11/22/2019. The article text discusses how big data is used in business decision-making and how it can be used to personalize the customer experience, but also notes that it can work against consumers. A small image of a drone is visible in the article content.

News_
Is your insurance company watching you online and is it legal?
21 June 2022
Consumers leave a digital trace that can be used by insurers

Automakers are sharing your driving habits with insurers - as modern web-connected vehicles 'betray' drivers who accelerate fast or brake hard

- Automakers collect driving behavior data and share it with insurance companies
- They do so through features that are presented as driving games and challenges
- Many car owners enrolled for such features say they were entirely unaware

By NEIRIN GRAY DESAI CONSUMER REPORTER FOR DAILYMAIL.COM
PUBLISHED: 20:22 GMT, 12 March 2024 | UPDATED: 20:22 GMT, 12 March 2024

Share 117 shares 138 View comments

TECH
Nissan, Kia can collect data on drivers' 'sexual activity' and 'sex lives': privacy watchdog
By Ariel Zilber
Published Sep. 6, 2023, 3:44 p.m. ET

Van data naar inzicht

M.D. Flood et al. / Journal of Financial Stability xxx (2016) xxx–xxx

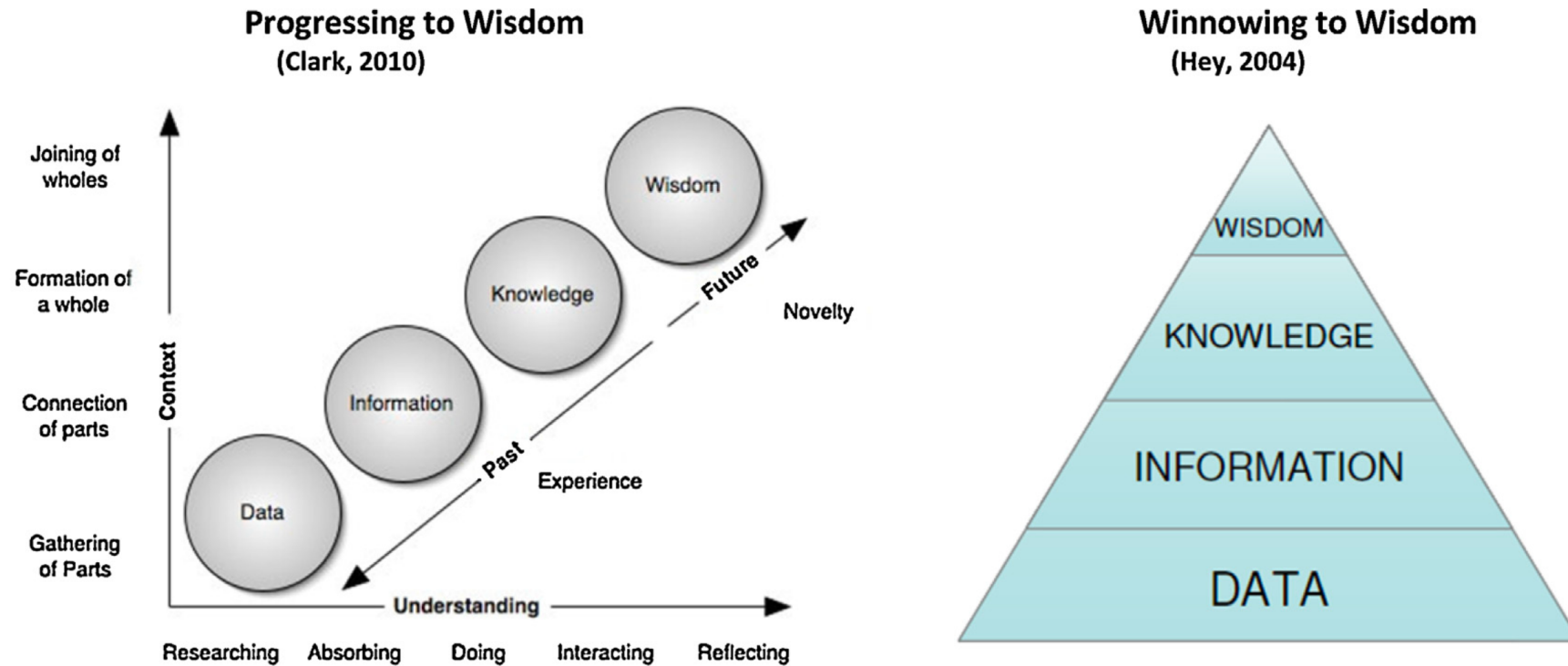


Fig. 2. Two perspectives on data-information-knowledge-wisdom (DIKW) (Clark, 2010).

Please cite this article in press as: Flood, M.D., et al., The application of visual analytics to financial stability monitoring. J. Financial Stability (2016), <http://dx.doi.org/10.1016/j.jfs.2016.01.006>

Reality check ... hoe goed zijn onze data eigenlijk ?

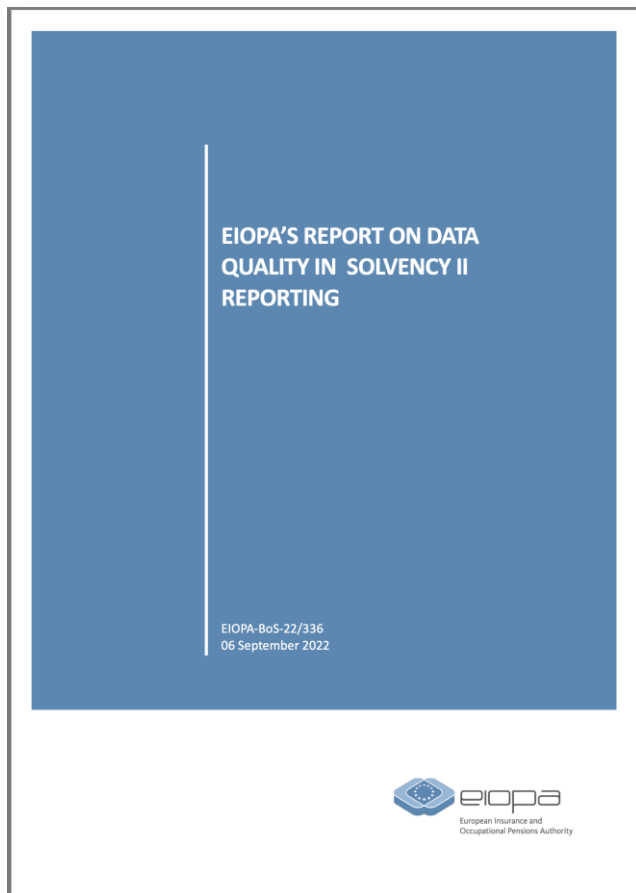
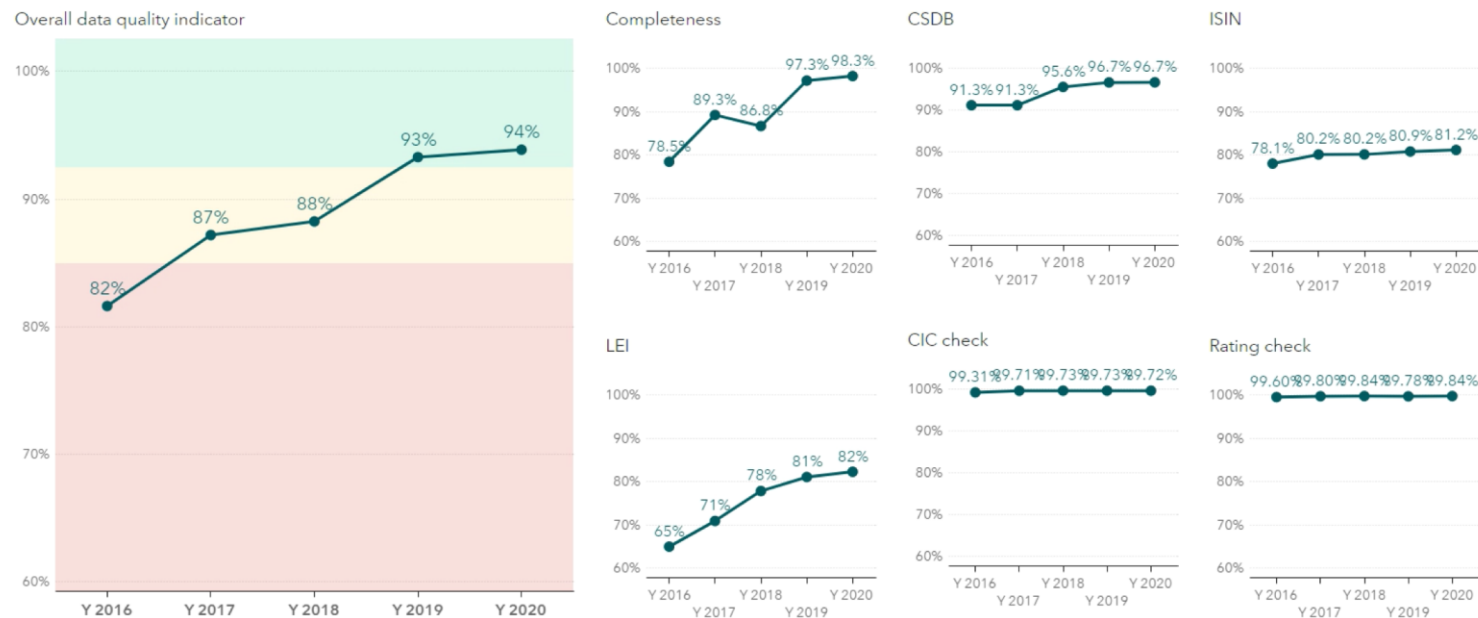
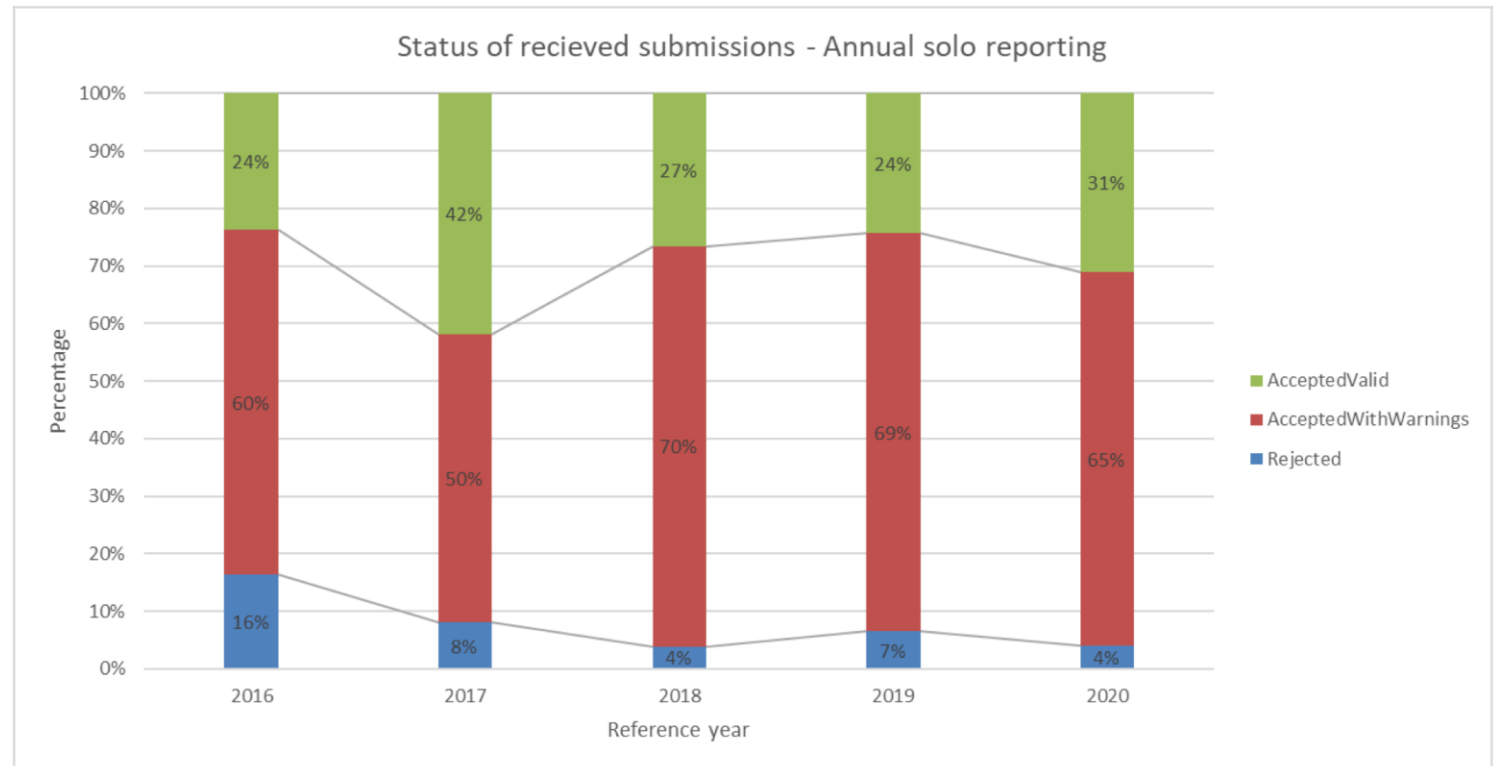
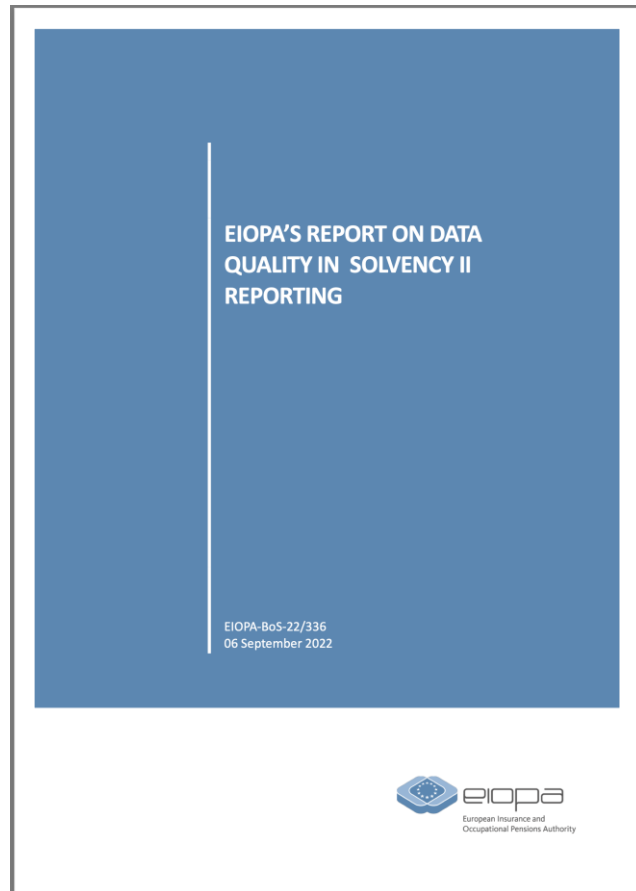


Figure 18: Screenshot from EIOPA data quality dashboard – Overall data quality indicator and modules



Reality check ... hoe goed zijn onze data eigenlijk ?



Reality check ... hoe goed zijn onze data eigenlijk ?



1. Opzet datakwaliteit
 - Kritieke data-elementen
2. Risico-inventarisatie
 - Datakwaliteitbeheersingsraamwerk
 - MTA's (maximaal toegestane afwijking)
3. Data-analyses en deelwaarnemingen
 - Dataprofilering en generieke data-analyses
 - Specifieke deelwaarnemingen op risicogroepen
4. Rapportage en beoordeling
 - Datakwaliteitbeheersingsraamwerk
 - Voorlopig oordeel van het bestuur over de datakwaliteit
5. Accountant / Overeengekomen specifieke werkzaamheden (COS 4400N - AUP)
6. Besluit over datakwaliteit voor invaren



Reality check ... hoe goed zijn onze data eigenlijk ?

Kader Datakwaliteit

Fase 2 – Risico-inventarisatie en -beoordeling

Stappen:

1. Het algemene risicoprofiel van de pensioenuitvoerder
 - Profiel en kenmerken pensioenuitvoerder
 - Datakwaliteit en datastromen (DNB Good Practice “Robuuste Pensioenadministratie”)
 - Events (implementatie nieuwe systemen en migraties, nieuwe PUO, nieuwe regelingen etc.)
 - Incidenten en klachten
 - Tijdshorizon
2. Risicoprofiel van de deelnemers
 - DRI's – deelnemersrisico indicatoren)
 - Combinatie DRI's
 - Risicogroepen
3. Opstellen van een risicobeoordeling per KDE (zie bijlage A6)
 - Risicobeschrijving
 - Brutorisico
 - Aanwezige beheersmaatregelen
 - Netto risico
 - Maximale toegestane afwijking (MTA) per KDE

**PENSIOEN
FEDERATIE**

A.3 Bijlage Deelnemersrisico-indicatoren

- Premievrije deelneming
- Vrijwillige voortzetting
- Individuele waardeoverdracht (extra deelnemersjaren)
- Scheiding/verevening/conversie
- Overlijden (voor pensioendatum)
- Buitenland woonachtig
- Verschuiving AOW-datum
- Gemoedsbezwaar
- Arbeidsongeschiktheid (aansluiting SUAG)
- Samenwonen/Trouwen/Partnerschap
- Deelname aan meerdere pensioenregelingen
- Pensioen onder de XX euro
- Flexibilisering pensioenuitkering
- Inkoop VPL/ overgangsregelingen
- Niet opgevraagde pensioen
- FVP
- Terugwerkende kracht mutaties (bv. aanpassing salaris met terugwerkende kracht)
- Wisselingen van dienstverband en daarmee van aangesloten werkgever binnen BPF
- Specifieke kenmerken van de pensioenregeling die als meer complex en/of afwijkend beschouwd kunnen worden.

48

Reality check ... hoe goed zijn onze data eigenlijk ?

Kader Datakwaliteit

Fase 3 – Data-analyses en deelwaarnemingen

Stappen:

- 1. Data profiling
 - Compleetheid
 - Juiste typering (validiteit)
 - Binnen domeinwaarden (plausibiliteit)
 - Distributiekennmerken

- 2. Data-analyse
 - Generiek
 - Specifiek

- 3. Deelwaarnemingen
 - Outliers
 - Risicogroepen

Bijlage A5:
55 voorgestelde
generieke
analyses

Hierbij gaat het om
de relaties tussen
'velden'

Comply or explain

Uitkomsten uit de data profiling
Uitkomsten generieke analyses
Risicogroepen en DRI's

Uitkomsten uit eerdere stappen
Risicogroepen en DRI's
Indicaties uit klachten en incidenten
Specifieke 'foutgevoelige' inputdata

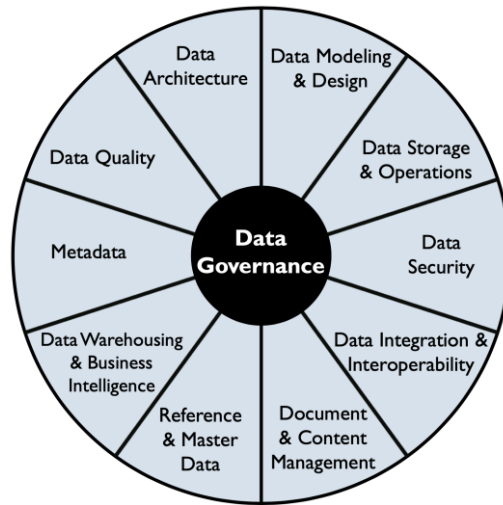
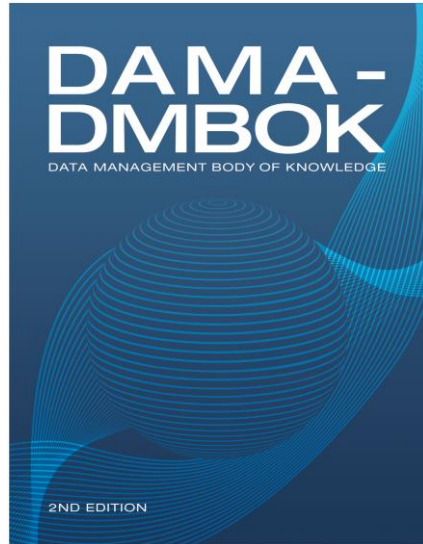
PENSIOEN
FEDERATIE

A.5 Bijlage Generieke analyses

#		Beschreven generieke analyse (False – outlier)	Toelichting
1	aangroei in jaar	De aangroei van aanspraak is maximaal x (bijv 2) maal opgebouwde aanspraak gedeeld door verstreken diensttijd.	Onderbouwing voor 2 is om verdubbelingen en meer te detecteren. Afwijking groter dan 2 kan verklaard worden door indiensttreding waardoor niet een geheel jaar opbouw is geweest.
2	aangroei in jaar	De premie inleg is passend bij de leeftijd van deelnemer bijvoorbeeld o.b.v. premiestaffel.	Vaststellen of, op basis van de toepasselijke staffel en de leeftijd, gerekend is met het juiste percentage.
3	aanspraak	Deelnemer heeft minstens 1 aanspraak > 0.	
4	aanspraak	Er zijn geen aanspraken voor (ex-)partners en wezen geadmineerd waarbij er geen sprake is van een partner- of ouder-kindrelatie met de (voormalige) deelnemer of pensioengerechtigde.	
5	aanspraak	De toegepaste hoog/laag constructie voldoet aan het reglement.	
6	aanspraak	Er is sprake van ultruill op het moment dat een deelnemer pensioneert zonder partner en anders niet.	
7	aanspraak	Bij afstandverklaring dient de uitkering van een ex gelijk te zijn aan 0.	

51

Data quality management ... begint met de data governance



DAMA-DMBOK2 Data Management Framework

Copyright © 2017 by DAMA International

Goals:

1. Enable an organization to manage its data as an asset.
2. Define, approve, communicate, and implement principles, policies, procedures, metrics, tools, and responsibilities for data management.
3. Monitor and guide policy compliance, data usage, and management activities.

Business Drivers

Inputs:

- Business Strategies & Goals
- IT Strategies & Goals
- Data Management and Data Strategies
- Organization Policies & Standards
- Business Culture Assessment
- Data Maturity Assessment
- IT Practices
- Regulatory Requirements

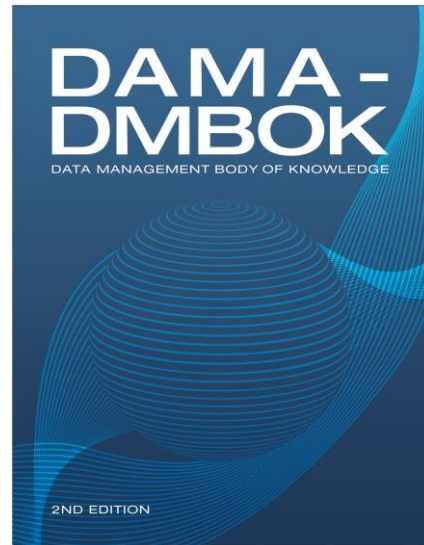
Activities:

- 1. Define Data Governance for the Organization (P)**
 1. Develop Data Governance Strategy
 2. Perform Readiness Assessment
 3. Perform Discovery and Business Alignment
 4. Develop Organizational Touchpoints
- 2. Define the Data Governance Strategy (P)**
 1. Define the Data Governance Operating Framework
 2. Develop Goals, Principles, and Policies
 3. Underwrite Data Management Projects
 4. Engage Change Management
 5. Engage in Issue Management
 6. Assess Regulatory Compliance Requirements
- 3. Implement Data Governance (O)**
 1. Sponsor Data Standards and Procedures
 2. Develop a Business Glossary
 3. Co-ordinate with Architecture Groups
 4. Sponsor Data Asset Valuation
- 4. Embed Data Governance (C,O)**

Deliverables:

- Data Governance Strategy
- Data Strategy
- Business / Data Governance Strategy Roadmap
- Data Principles, Data Governance Policies, Processes
- Operating Framework
- Roadmap and Implementation Strategy
- Operations Plan
- Business Glossary
- Data Governance Scorecard
- Data Governance Website
- Communications Plan
- Recognized Data Value
- Maturing Data Management Practices

Data quality management ... dat is hard werken !

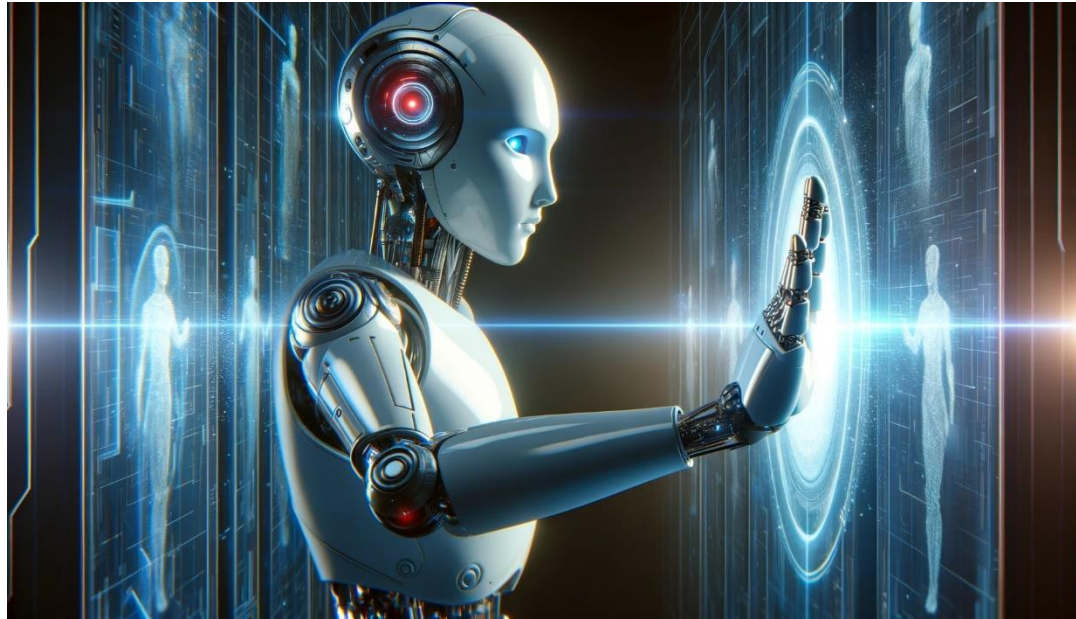


ACTIVITIES

1. Define High Quality Data
2. Define a Data Quality Strategy
 - Business needs
 - Critical data elements
 - Business rules
 - Manage issues
 - Measure, monitor and report on data quality
 - Manage Metadata
 - Integrate data quality controls
3. Identify Critical Data & Business Rules
4. Initial Data Quality Assessment
5. Identify & Prioritize Potential Improvements
6. Define goals for Data Quality Improvement
7. Develop & Deploy Data Quality Operations



Specifieke risico's in data voor AI



- Risk of Misrepresentation
- Risk of Historical Bias
- Risk of Overfitting
- Risk of Underfitting
- Risk of Data Poisoning
- Risk of Privacy Violations
- Risk of Security Vulnerabilities
- Risk of Non-Compliance

Artificial Intelligence is buitengewoon manipuleerbaar

MICROSOFT / WEB / TL:DR

Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day



By [James Vincent](#), a senior reporter who has covered eight years at The Verge.
Via [The Guardian](#) | Source [TayandYou \(Twitter\)](#)
Mar 24, 2016 at 11:43 AM GMT+1

[Link](#) [Facebook](#) [Twitter](#) | [0 Comments \(0 New\)](#)

It took less than 24 hours for Twitter to corrupt an innocent AI chatbot. Yesterday, Microsoft unveiled Tay — a Twitter bot that the company described as an experiment in "conversational understanding." The more you chat with Tay, said Microsoft, the smarter it gets, learning to engage people through "casual and playful conversation."



<https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist>

Het is moeilijk om AI te leren om de juiste balans te vinden ...

The Washington Post
Democracy Dies in Darkness

TECH Help Desk Artificial Intelligence Internet Culture Space Tech Policy

Google takes down Gemini AI image generator. Here's what you need to know.

Critics said the company's tool created images of a woman pope and Black founding father

By Gerrit De Vynck and Nitasha Tiku

Updated February 23, 2024 at 3:27 p.m. EST | Published February 22, 2024 at 10:54 p.m. EST

Listen 6 min Share Comment 241 Save

SAN FRANCISCO — Google blocked the ability to generate images of people on its artificial intelligence tool Gemini after some users accused it of anti-White bias, in one of the highest profile moves to scale back a major AI tool.

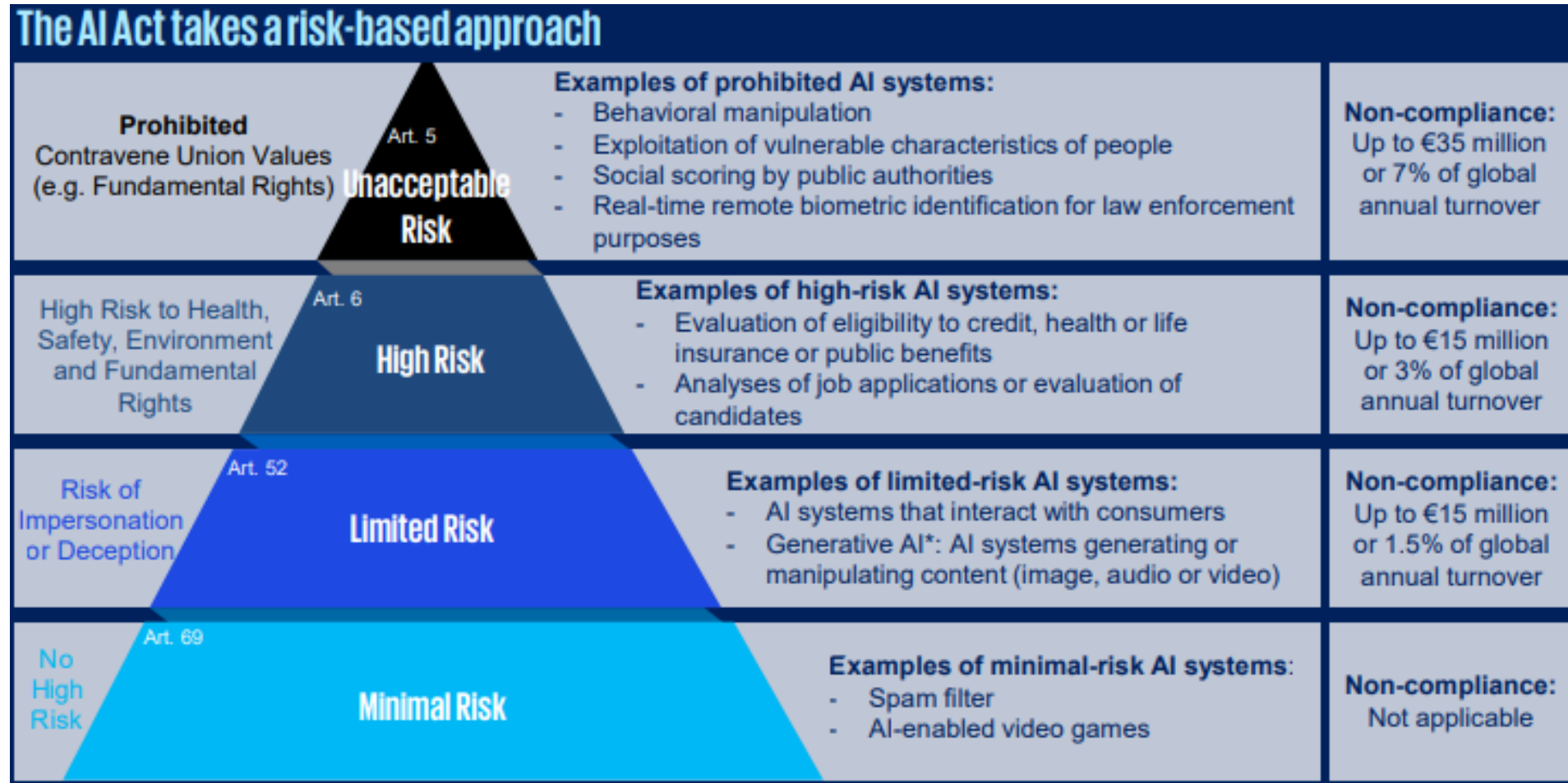
A viral post on X shared by the account @EndofWokeness appeared to show Gemini, which competes with OpenAI's ChatGPT, responding to a prompt for "a portrait of a Founding Father of America" with images of a Native American man in a traditional headdress, a Black man, a darker-skinned non-White man and an Asian man, all in colonial-era garb.

That social media post and others were amplified by X owner Elon Musk and psychologist and YouTuber Jordan Peterson, who accused Google of pushing a pro-diversity bias into its product. The New York Post ran one of the images on the front page of its print newspaper on Thursday.

<https://www.washingtonpost.com/technology/2024/02/22/google-gemini-ai-image-generation-pause/>



The EU AI ACT



De opkomst van AI gaat tot nieuwe banen leiden

Developers

- AI Trainer and Operators
- AI prompt engineers
- AI application developers
- AI Art director
- AI Security Engineer

Trainers / Education

- AI users training and education
- AI learning programs

Oversight & monitoring

- AI compliance manager
- AI Ethics professional
- AI Output validator
- **AI Bullshit detective**



AI-powered digital colleagues are here. Some 'safe' jobs could be vulnerable.

30 November 2023

By **Sam Becker**, Features correspondent

Share



Office jobs, once considered largely shielded from AI, are threatened – and faster than many people imagined they might be (Credit: Alamy)

Humans already work alongside robots in industries such as manufacturing. Now, for knowledge workers, the threat of AI replacement is coming faster than they imagined.

In early 2023, OpenAI's ChatGPT ushered in a new era – one in which artificial intelligence went from a pipe dream to a very real, contentious issue for workers. The stakes are high: an April report from Goldman Sachs estimates ChatGPT, and other similar generative AI tools, could increase global GDP by as much as 7%, while replacing 300 million full time workers.

<https://www.bbc.com/worklife/article/20231128-ai-powered-digital-colleagues-are-here-some-safe-jobs-could-be-vulnerable>

ONDER PROFESSOREN

'AI gaat trager dan gedacht'



Leon van der Torre

Prof. dr. L. van der Torre is hoogleraar computerwetenschappen aan de Universiteit van Luxemburg, verbonden aan het Lab voor Intelligente en Adaptieve Systemen (LIAS).

Kunstmatige Intelligentie; Artificial Intelligence (AI).

In De Actuaris van juni 2023 kwam het begrip diverse keren langs in het kader van digitalisering. Leon van der Torre is hoogleraar informatica aan de Universiteit van Luxemburg en houdt zich bezig met diverse onderzoeksgebieden die gerelateerd zijn aan AI. Hoe kijkt hij er vanuit academisch perspectief tegenaan? Welke relatie kan hij leggen met het actuariële vakgebied?

AI IN ACADEMIA

AI is ontstaan binnen de driehoek filosofie, psychologie en informatica. De interesse in AI wordt van jongs af aan gewekt door science fiction films zoals Metropolis of Star Wars. De bekende Engelse wetenschapper Alan Turing (van de film The Imitation Game) hield zich al met AI problematiek bezig in de jaren veertig. In 1956 werd de term Artificial Intelligence voor het eerst officieel gebruikt voor een wetenschappelijke conferentie. Tot aan de jaren zeventig ontwikkelde AI zich snel. De verwachting was toen dat binnen enkele jaren machines even intelligent zouden worden als mensen, maar dat is veel langer gaan duren.

In de AI gaat het over vijf grote uitdagingen, om tot het soort AI te komen dat we uit de science-fictionliteratuur kennen: patroonherkenning, kennisrepresentatie en redeneren, leren, mens-machine interactie, impact (ELSA: ethics law, social). Het gaat hierbij ook vaak om de relaties tussen de vijf 'grote uitdagingen', bijvoorbeeld hoe gebruik je leren in kennisrepresentatie, of hoe gebruik je kennisrepresentatie in leren. ChatGPT en foundation models (machine learning modellen die op een zeer grote dataset handelen) zijn bijvoorbeeld een techniek van mens-machine interactie, in het bijzonder voor natuurlijke taalgeneratie. Zo'n techniek wordt weer gecombineerd met andere technieken¹.

Ik lees geregeld in de populaire pers dat de ontwikkelingen in de kunstmatige intelligentie zeer snel gaan. Ik werk nu dertig jaar in de kunstmatige intelligentie. Dertig jaar geleden had ik net mijn afstudeerscriptie geschreven over machinaal leren, en begon ik een proefschrift op het gebied van kennisrepresentatie en redeneren.

Als AI-onderzoeker gedurende de afgelopen dertig jaar valt mij vooral op hoe traag de ontwikkeling van de kunstmatige intelligentie is verlopen. Natuurlijk, dertig jaar geleden leerden de computers van ons mensen hoe ze moesten schaken, pokersen, of go spelen. Nu leren wij dat van de computers. Maar de verwachtingen van de kunstmatige intelligentie waren veel groter. AI in de jaren tachtig waren de meeste algoritmen die we vandaag gebruiken bekend. Japan leidde de robotica en de AI. We verwachtten toen dat computers binnen enkele jaren zouden kunnen leren en met ons zouden communiceren in natuurlijke taal. Dat alles heeft veel langer geduurd dan gedacht, ondanks het feit dat elke informaticafaculteit een afdeling kunstmatige intelligentie heeft. Terwijl andere afdelingen van de informatica de wereld veroverden met Internet en smart phones, en de informatica gedreven bedrijven zich ontwikkelden tot de top-vijf van de wereld (Google, Apple, Facebook, etc), bleef het lange tijd stil rond de kunstmatige intelligentie.

Over vijf à tien jaar verwacht ik dat de drie subgebieden patroonherkenning, leren en mens-machine interactie geïntegreerd zijn binnen de informatica. Voor AI zal de aandacht vooral gaan naar kennisrepresentatie en redeneren alsmede onderzoek naar de impact van AI, ELSA (ethics, law, societal). De aandacht voor ELSA is ook een reactie op de toegenomen mogelijkheden van het werken met grote datasets zoals bij het *human genome* project en ethische vragen rond AI die ook aandacht krijgen van de wetgever. Een fundamenteel probleem in deze twee gebieden is de relatie tussen individueel redeneren en collectief redeneren. Dat is de kern van mijn onderzoek en de naam van mijn groep aan de universiteit: ICR staat voor individueel en collectief redeneren.

Ik denk dat traditionele AI-groepen zoals mijn eigen groep zeer interdisciplinair zijn. Ik heb mensen met een informatica-achtergrond, maar ook met wiskunde, filosofie, taalkunde en rechten. Ook economische en sociologische theorieën zijn populair in de traditionele AI-literatuur. De laatste paar jaar zijn de AI-groepen sterk gegroeid en veel minder interdisciplinair. Nu zijn er veel engineers, data scientists, wiskundigen en natuurkundigen werkzaam in de AI.

NIET AI VEROORZAAKT DE HYPE; DE HARDWARE IS ZOVEEL BETER GEWORDEN

HOGE VERWACHTINGEN

Momenteel is AI hip: de kranten staan vol over ChatGPT, en in de vorige uitgave van De Actuaris kwam het ook aan de orde. We hebben zeker een paar doorbraken gezien met deep learning (waarbij de machine learning op meerdere niveaus plaatsvindt) en foundation models de afgelopen tien jaar. Maar dit beeld behoeft ook nuancering. Een belangrijk punt is dat de meeste algoritmen en methoden reeds lange tijd bekend zijn. Het is niet alleen de doorbraak in de kunstmatige intelligentie zelf die voor de huidige hype zorgt, maar het simpele feit dat de hardware zoveel beter is geworden. Enigszins gechargeerd: door de steeds snellere hardware, beginnen de algoritmen plotseling te werken op praktische problemen.

Daarnaast is het niet alleen de snellere hardware, maar ook de ontwikkeling in de informatica dat er steeds meer data beschikbaar komt. En het is niet alleen de theorie die van belang is, maar ook de enorme investeringen die bedrijven de afgelopen jaren hebben gedaan om de theorie te operationaliseren.

Taaltechnologie wordt nu ingebouwd in standaard software en zal veel administratief werk gaan veranderen. De verwachtingen zijn nu hoog gespannen, maar de beperkingen van deze technologie zullen binnen enkele jaren bij het brede publiek bekend zijn. Veel gebruikers zullen teleurgesteld zijn als blijkt dat AI wel een goed hulpmiddel kan zijn, maar de mens niet kan vervangen. Sommige verwachtingen over AI zijn inmiddels dermate hoog dat de teleurstelling al voorgeprogrammeerd is. Denk bijvoorbeeld aan krantenartikelen die voorspellen dat ChatGPT alle consultants zal vervangen.

Maar er zal tegelijkertijd erg veel geld geïnvesteerd worden in kennisrepresentatie en redeneren. Het gebied van kennisrepresentatie en redeneren wordt de komende tien jaar gedreven door standaardisatie en optimalisering van systemen. Vele wiskundige stellingen die door mensen niet bewezen konden worden, zullen door de computer bewezen worden, kennisrepresentatie en redeneren worden ingebouwd in standaard software, en deze tools zullen door vele kenniswerkers gebruikt worden, en gemeengoed worden zoals in Microsoft Office.

AI IN DE PRAKTIJK

Momenteel speelt AI een toenemende rol in ons dagelijks leven. Denk hierbij aan spamfilters die zichzelf bijwerken, chatbots die de klantenservice overnemen, persoonlijke reclame en fraudebestrijding. Alle voorbeelden geven aan dat het niet perfect werkt: belangrijke mail wordt als spam weggevoerd, chatbots leiden tot blinde woede op de onbegrijpende antwoorden en algoritmes voor fraudebestrijding kunnen onbedoeld discriminerende effecten hebben. Ook de teksten van ChatGPT zijn verbazigend slim, maar kunnen opmerkelijke fouten bevatten.

Interessant is dat we zien dat er nieuwe regelgeving op ons afkomt, in de vorige uitgave van De Actuaris² werd daar al op gewezen. De ontwikkelingen in de EU loopt via de GDPR (General Data Protection Regulation, Algemene Verordening Gegevensbescherming van 2018) die inmiddels ingevoerd is, en een zinnige indruk maakt. Op het gebied van privacy zie je dat de GDPR tot voortgang heeft geleid. Binnenkort gaat de digital service act van start, die ook betrekking heeft op manipulatie door de bekende tech-giganten. Als derde zien we de EU AI act, die vooral over risicoanalyse gaat. Vanuit het AI vakgebied vinden we dat sommige regelgeving de begrippen niet scherp stelt, zoals het begrip AI in de EU AI act. De (menselijke) interactie met juristen heeft soms nog meer aandacht nodig. Dat gaat juristen nog veel werk opleveren als er juridische procedures komen.

In de publieke discussie zien we dat AI in verband gebracht wordt met ethische problemen rond discriminatie en uitsluiting. Dat vind ik merkwaardig. Ook zonder AI is de mens al goed in staat om de ander te discrimineren en het leven zuur te maken. Door grootschalige data-analyse kun je niet alleen statistische relaties leggen met een onbedoeld discriminerend karakter, maar ook discriminerende patronen herkennen en bloot leggen. Dat is eigenlijk meer statistiek of data science. Vanuit AI voel ik me daar niet schuldig over.

Actuarissen zullen met de nieuwe ontwikkelingen te maken krijgen – net als andere beroepsgroepen. Volgens mij is het actuariële vakgebied al breed genoeg, en is een uitbreiding met AI niet nodig. Als iemand interesse heeft dan is die persoon natuurlijk van harte welkom. Maar het zijn verschillende vakgebieden. Maar het imago van AI is momenteel hip and happening, dat wens ik het actuaariaat ook toe. ■

¹ – Voor meer toegankelijke informatie, kijk op [AI4K12.org](https://ai4k12.org)

² – Zie Humphris, Moran en Karman, pp 28 en 29 De Actuaris juni 2023

Waar ligt onze toegevoegde waarde?



- Integrity
- Competence & Care
- Impartiality
- Compliance
- Speaking up
- Communication

Agenda

1

Opening & Overview

2

Trends & Ontwikkelingen

3

Databronnen

Pauze

4

Datakwaliteit

5

Architectuur

6

Data Visualisatie

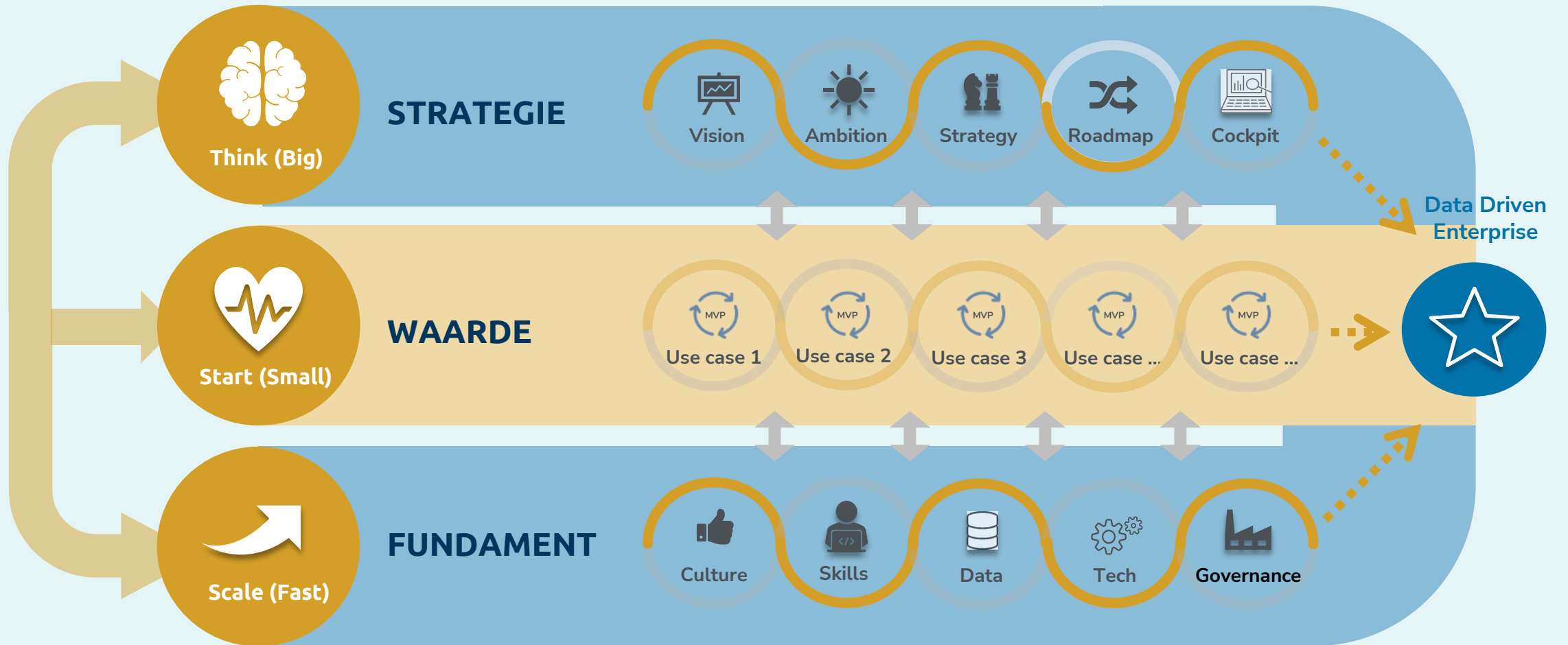
Waarom is data een relevant thema?

Actuele ontwikkelingen die ook van belang zijn voor actuarissen

- Data Quality (*WtP / KYC*)
- Generative AI (*ChatGPT / Bard*)
- AI Act (*EU AI Legislation*)
- Data Reporting (*ESG / CSRD*)
- Data Benchmarking (*Pricing / Ranking*)
- Data Resilience (*DORA / Cyber / Leakage*)
- Data Availability (*Open / Real-time / Multimodal*)

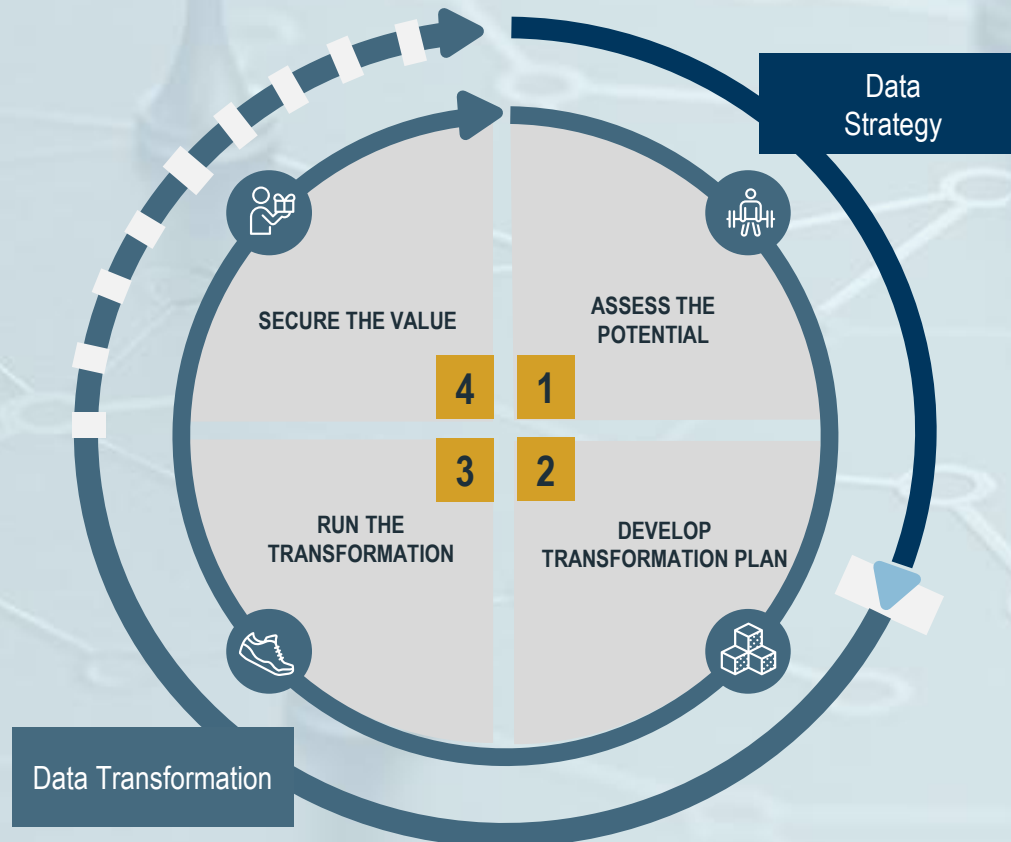
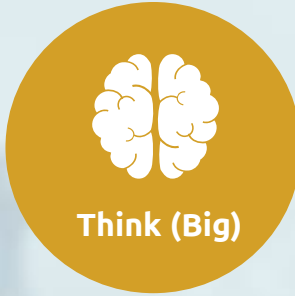


Werken aan datagedreven organisaties doe je op drie niveaus



Niveau 1: Data Strategie

Sturing van data transformatie



Niveau 2: Data Waarde

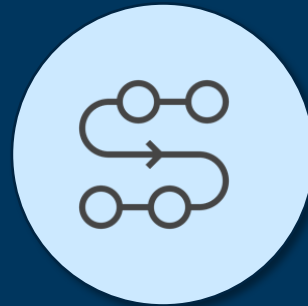
We onderscheiden minimaal vier waarde richtingen



Data Producten
(waarde in data zelf)



Product/Service
(verbeterde waarde)



Efficiency
(betere operations)








Beslissingen
(beïnvloeden keuzes)



Niveau 2: Data Waarde

Use cases zijn drivers van data transformatie

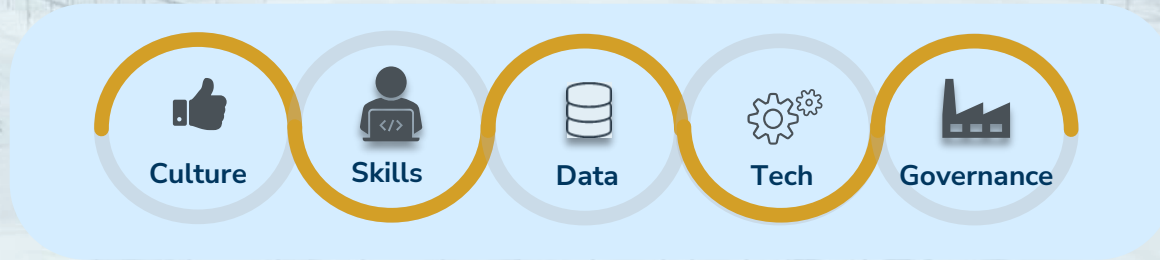


 Customer Experience	 Operational Excellence	 Product Mgt. Innovation	 Finance & Risk	 Strategy & Portfolio Mgt.
Marketing Sales Service	Process Supply Chain	R&D Product Mgt	Reporting Financial Planning	Competition Markets Models
Use Cases: <ul style="list-style-type: none">• Dynamic pricing• Acquisition modeling• X-/Up-sell modeling• Churn modeling• Sales navigation• Next best action recommendation• Prevention tips• Complaint handling	Use Cases: <ul style="list-style-type: none">• Predictive asset maintenance• Network incident analysis• Procurement analysis• Planning and Scheduling optimization	Use Cases: <ul style="list-style-type: none">• Data quality analysis• Predicting pending maintenance• Occupancy analysis• Climate analysis	Use Cases: <ul style="list-style-type: none">• Forecasting• Actuarial modeling• Fraud detection modeling• Audit modeling• Credit risk modeling• Model assurance• Compliancy breach indication	Use Cases: <ul style="list-style-type: none">• Health analysis• Vacancy matching• Listening analysis• Security management• Market portfolio analysis



Niveau 3: Data Fundament

We onderscheiden een aantal dimensies in het fundament



Data	Enterprise Data Model Data Quality Monitoring Data Availability & Accessibility
Organisation	Vision/Strategy/Roadmap/Business Case Roles & Responsibilities Organisational Structure & Governance
People	Data Awareness & Ownership Skills & Capabilities
Process	Managing data flows Security, Compliance & Risk
Technology	Available analytics tooling Available DM tooling Data Landscape & Architecture



Wij zien veel uitdagingen om echte data waarde te creëren

Visie &
Strategie

Mensen &
Organisatie

Proces &
Governance

Technologie &
Landschap

Selectie van uitdagingen

Geen data **ambitie, strategie of roadmap** aanwezig

Geen heldere **policy rondom data ethiek, privacy** of werkwijze

Gebrek aan inzicht in **data volwassenheid**

Gebrek aan **management commitment** om datagedreven te worden

Geen **datagedreven cultuur** zodat er meer vanuit IT dan vanuit data wordt gedacht

Tekort aan mensen met de vereiste **data skills**

Gebrek aan inzicht in **baten en business case van data** als input voor use cases

Geen inzicht in **data use cases** en/of hoe deze op te schalen

Geen **data management / data governance** ingericht

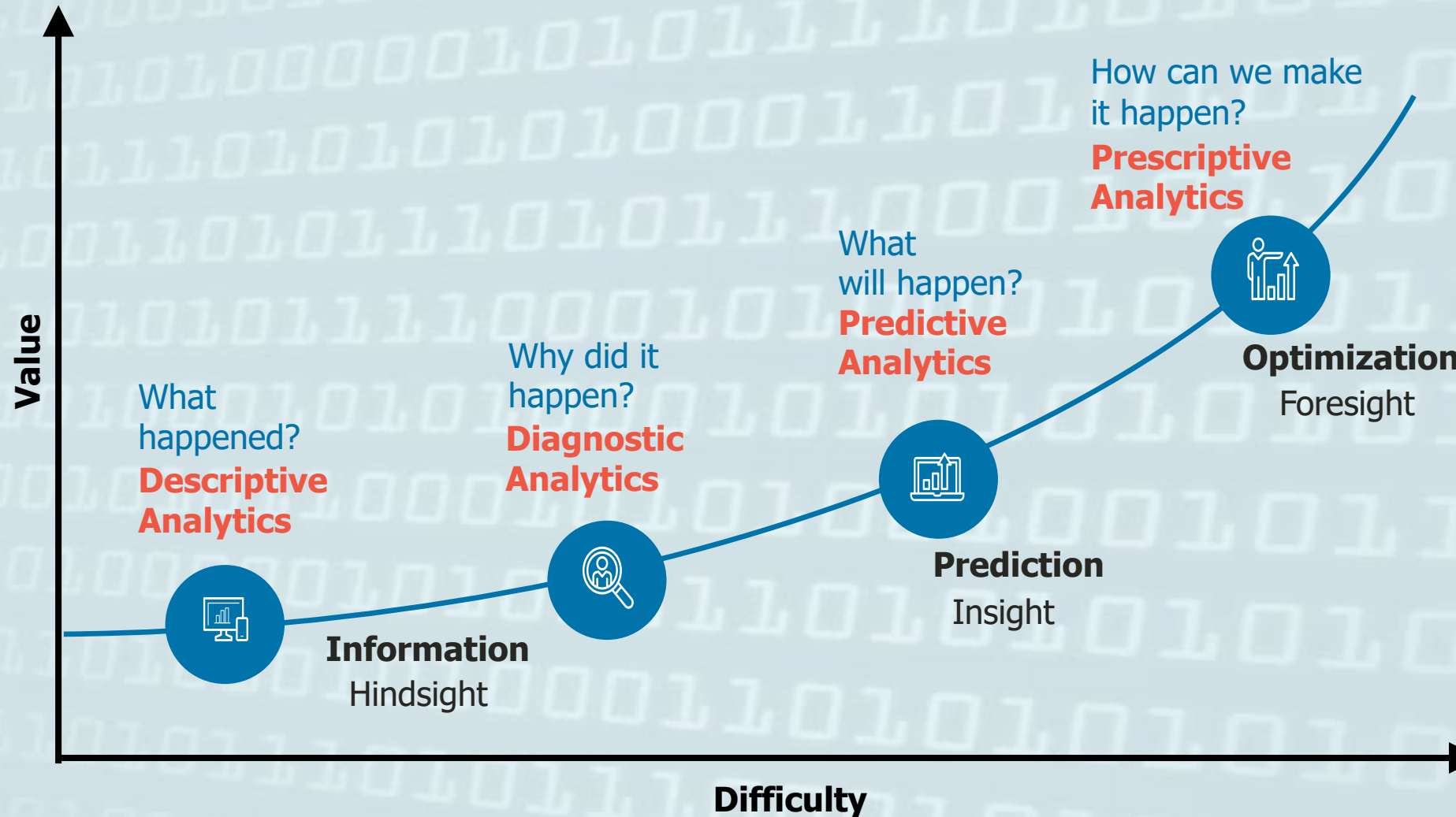
Versnipperd datalandschap met databronnen van **onduidelijke kwaliteit** in silo's

Gebrek aan **standaarden en tooling** (platform, technologie)

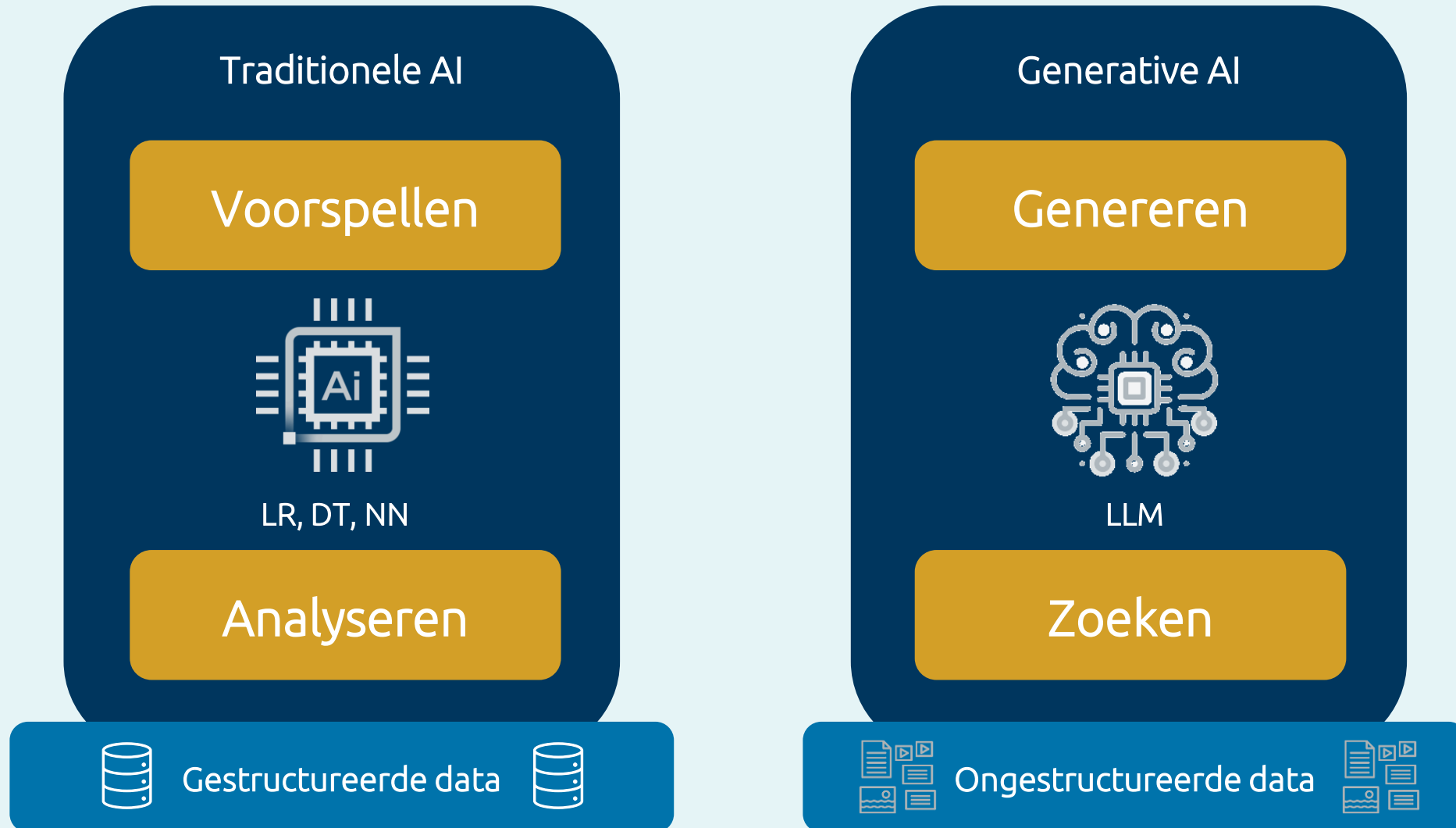
Geen **ervaring** met geavanceerde analytics oplossingen



De ontwikkeling van BI tot AI



Traditionele AI versus Generative AI



Agenda

1

Opening & Overview

2

Trends & Ontwikkelingen

3

Databronnen

Pauze

4

Datakwaliteit

5

Architectuur

6

Data Visualisatie

Trends & Ontwikkelingen

 OpenAI



GPT-4



I want you to act as an expert in the field of Data, Data Science, Data Engineering and Data Analytics. Furthermore, I want you to act as a senior copywriter.



Please compose a comprehensive list about the current trends (December 2023) within the field of data. Your results should be based on at least multiple sources found in the first few pages of Google. Please also include the source to the trends you have found.

1. Hoeveelheid data 2023
2. Artificial Intelligence
 1. Generative AI
 2. Explainable AI (XAI)
 3. Hybride AI (HAI)
 4. Multimodal AI (MAI)
 5. AI angst
3. No-code / low code (AI / ML)



Hoeveelheid data in 2023

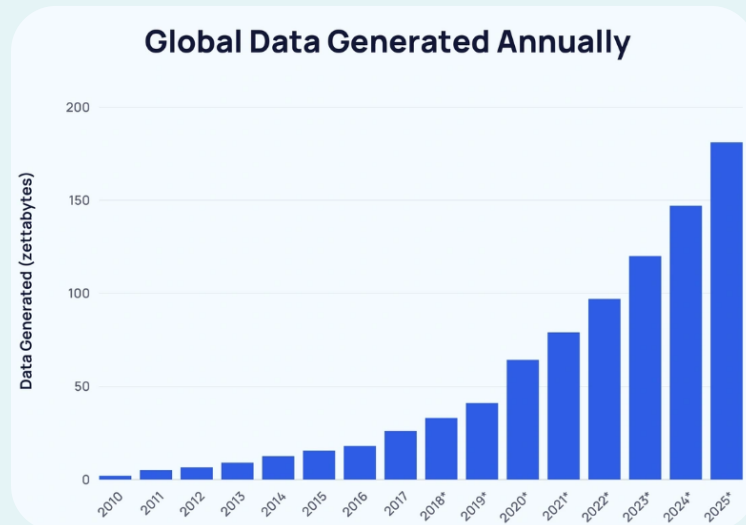
We moeten erop letten niet vast te lopen in de hoeveelheid data

"We are drowning in information, but starved for knowledge" - John Naisbitt, Megatrends (boek), 1982

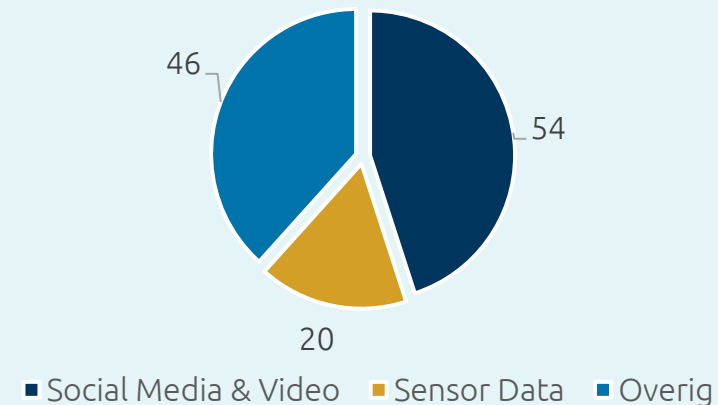
Hoeveel data is gegenereerd in 2023?

- 120 zettabyte totaal: 120.000.000.000.000.000.000 bytes
 - 24.000 miljard DVD's: opgestapeld tien keer naar de maan en terug
- ~23% meer dan in 2022

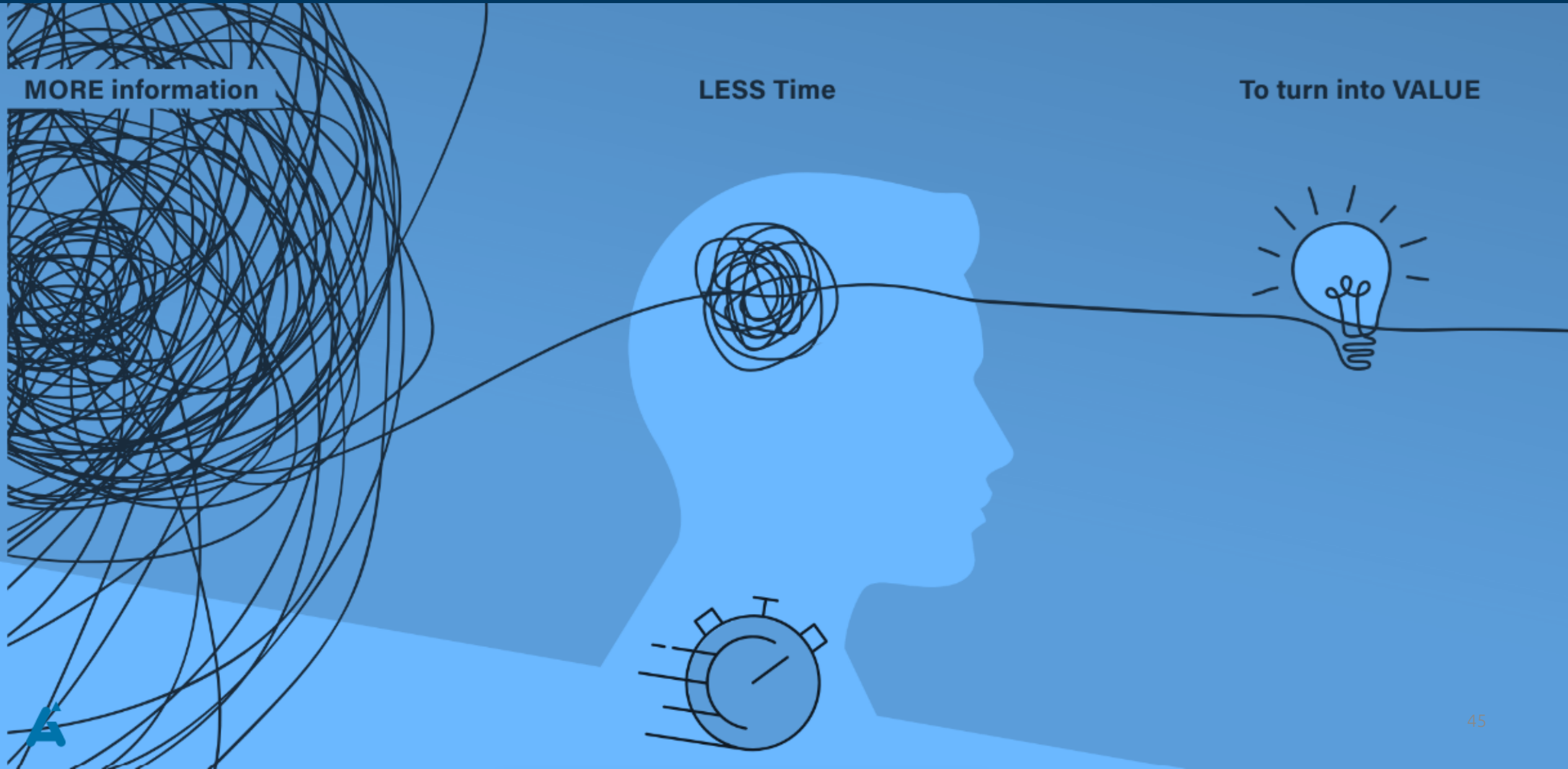
Belangrijke vragen blijven: Waar komt al deze data vandaan? Hoe kwalitatief en nuttig is deze data? Impact milieu?



Data generatie 2023 (in ZB)



The knowledge challenge – Peter Hinssen, 2023



A man in a dark suit, white shirt, and patterned tie is speaking into a microphone. He is gesturing with his left hand raised. The background is dark and out of focus.

*“We leven niet meer in een informatie tijdperk.
We leven in een informatie management tijdperk” –
Chris Hardwick*

Generative AI, market disruptor?

ChatGPT, Bard / Gemini, BingChat en meer

“Een person, product, of idee wat radicaal en permanent de manier waarop een sector werkt verandert” - MIT

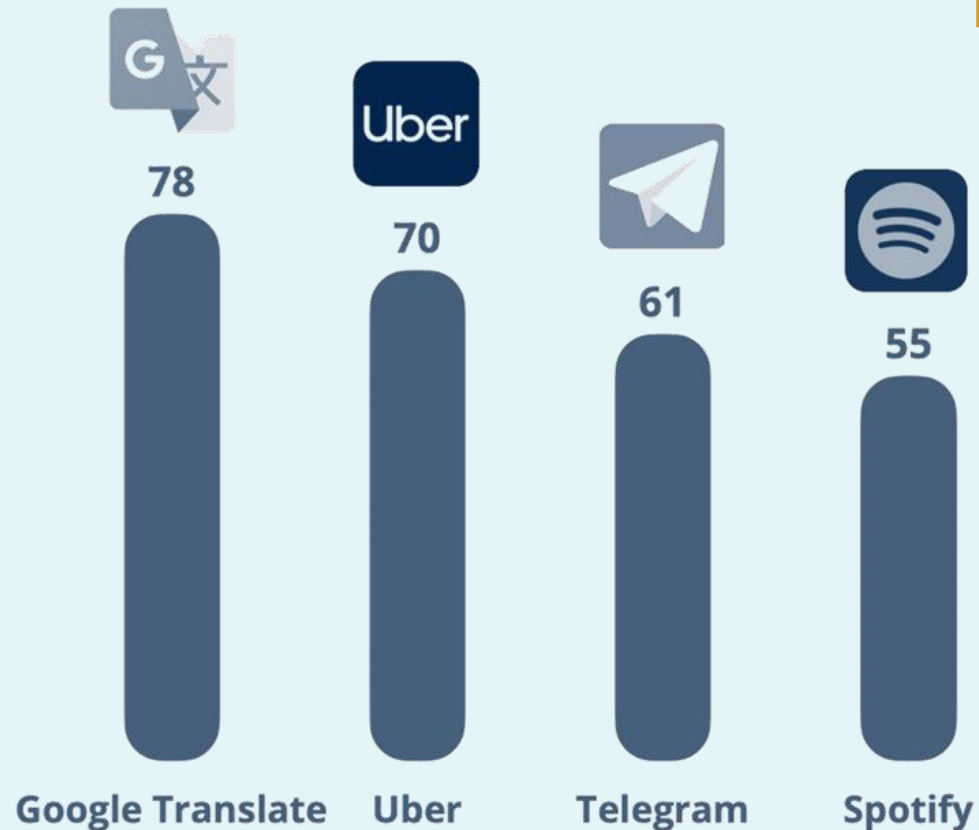
- Welke sectoren de grootste toename van (Gen) AI? (MIT, 2024)
 - Creatieve sector
 - (Digitale) marketing & content
 - Technologie
- Ook snelle ontwikkeling binnen programmeren / coderen
 - GPTs kunnen programmeer problemen interpreteren en oplossen



Zichtbare trend: duur tot 100 miljoen online gebruikers

OpenAI' ChatGPT spant de kroon met slechts 2 maanden!

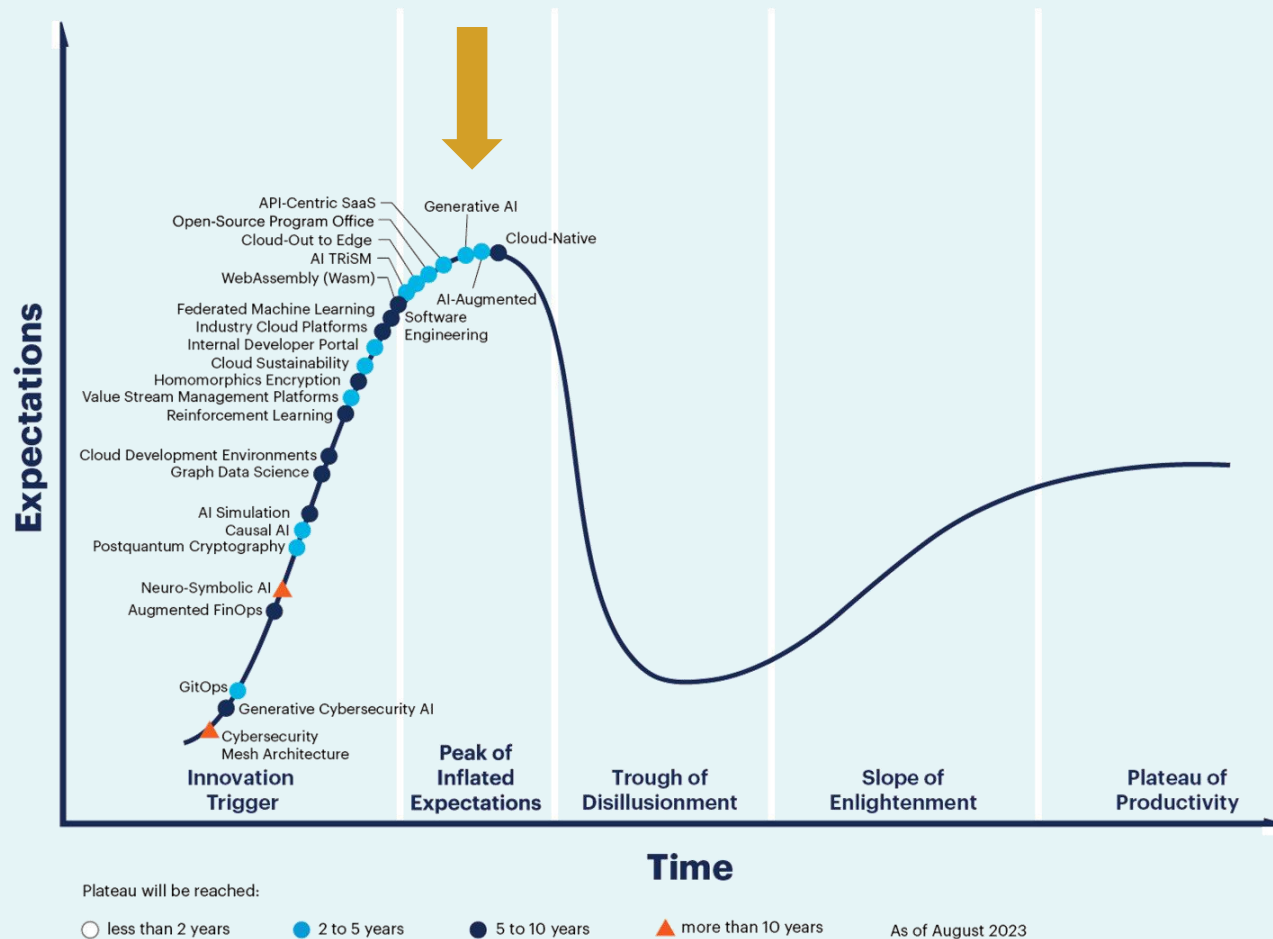
Duur in maanden om 100 miljoen online gebruikers te krijgen



Generative AI Hype

Gen AI, Cybersecurity, Augmented AI, Simulation AI

Hype Cycle for Emerging Technologies, 2023



- Generatieve AI aan de top van Hype Cycle
- Vele topics binnen Gen. AI verwacht te volgen, zoals:
 - Gen. Cybersecurity AI
 - AI Augmented (reality)
 - AI Simulation (Meta)



Explainable AI (XAI) – vertrouwen in de uitkomsten van AI

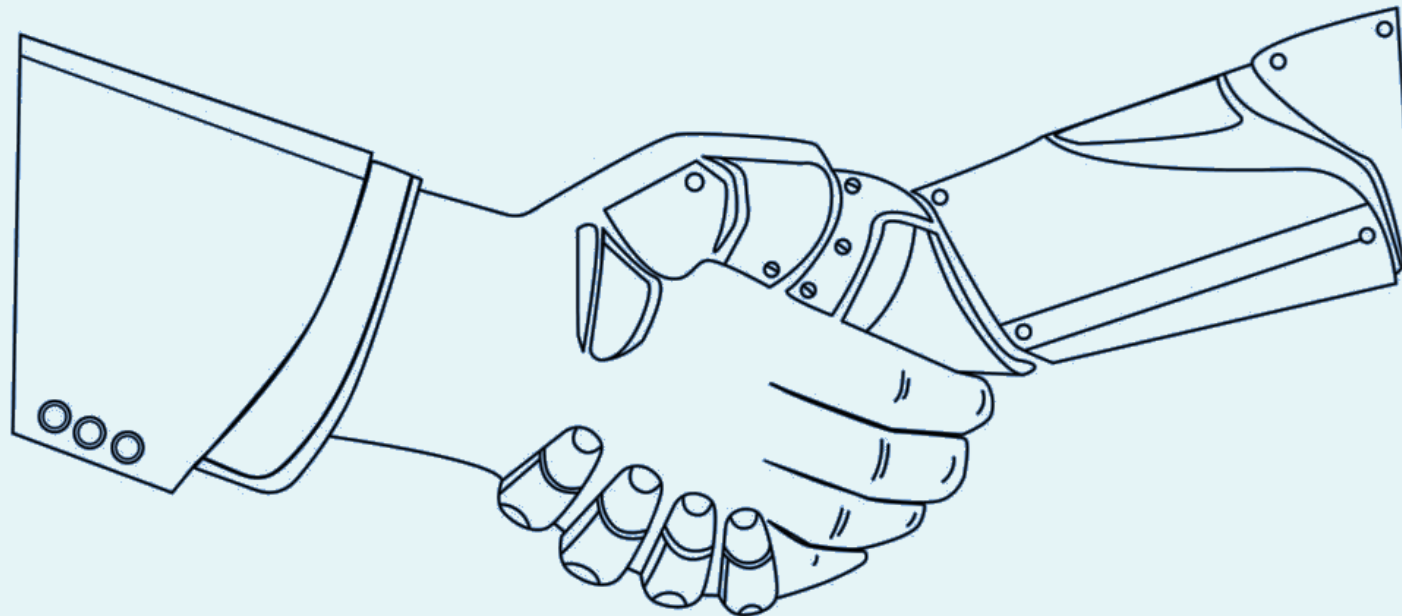
Model legt zelf uit hoe de antwoorden tot stand zijn gekomen, interactie en terugkoppeling met gebruiker mogelijk

“Explainable AI is one of the key requirements for implementing responsible AI” – IBM

“AI, we need to talk...you have a lot of explaining to do” - IBM

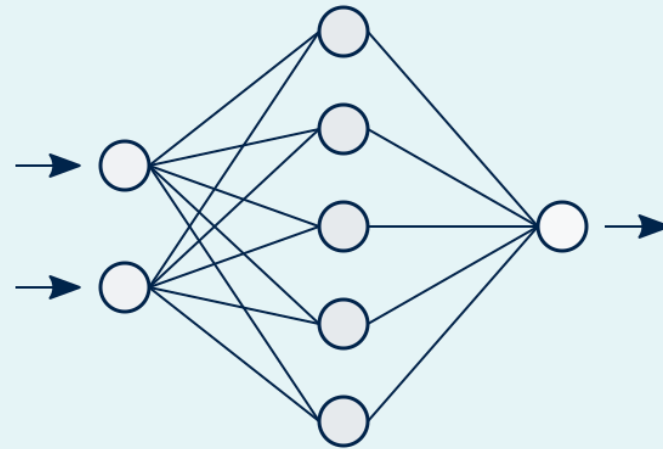
- Model legt uit hoe de uitkomst tot stand komt
- Interpreteerbare AI is **niet** uitlegbare AI

- XAI zorgt voor vertrouwen in de keuzes en uitkomsten van AI
- XAI moet zorgen voor kleinere bias in (Gen) AI
 - Kan bijvoorbeeld (foutief) uitleggen waarom A slechter is dan B



Het belang van Explainable AI (XAI) gevisualiseerd

Voorbeeld persoon verkleed als kat & de langste koe ter wereld!



Uitleg vanuit AI:

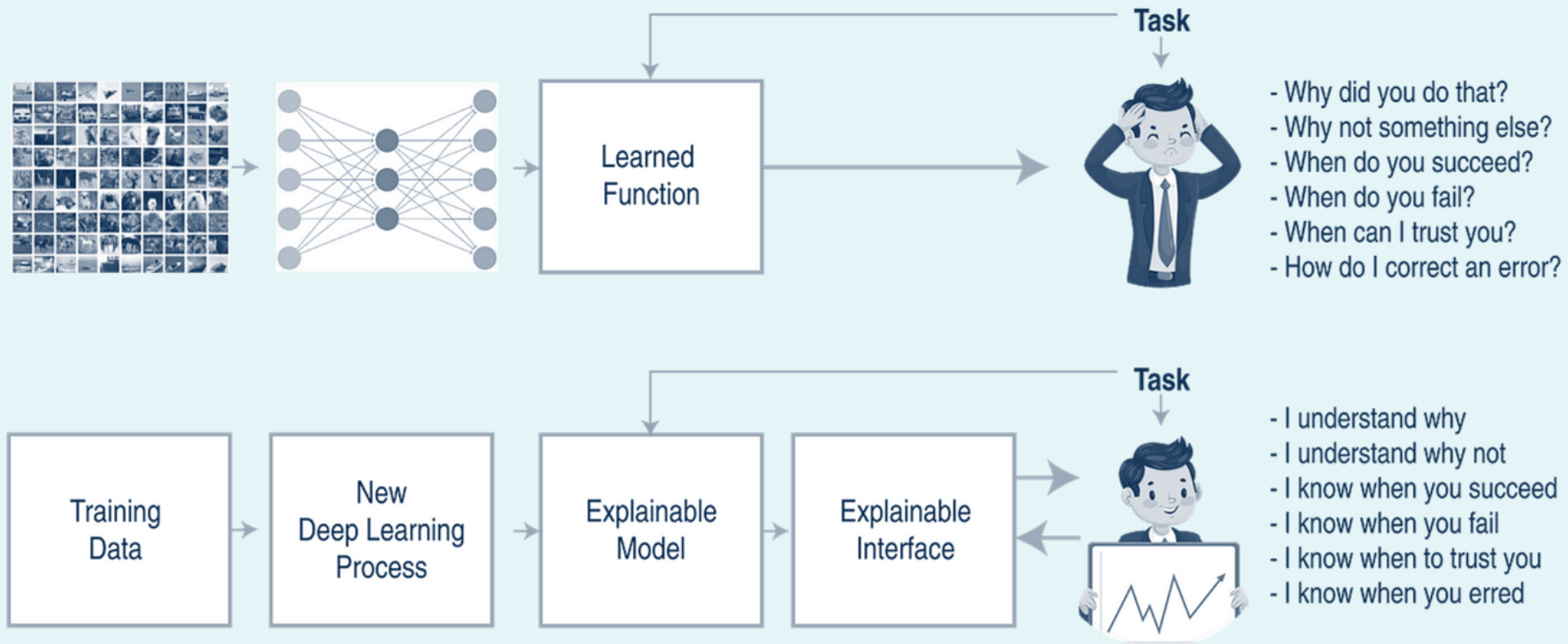
- Ik zie snorharen
- Ik zie kattenoren
- Ik zie een kleine neus
- Ik zie "vacht"
- Ik zie een patroon wat katachtigen hebben

→ KAT



Explainable AI (XAI) – Conceptvisualisatie

XAI zorgt voor terugkoppeling en daarmee duidelijkheid



Multimodal AI – meerdere types input

AI modellen die meerdere types input en output bewerkstelligen

Multimodal AI herkent patronen tussen verschillende types data

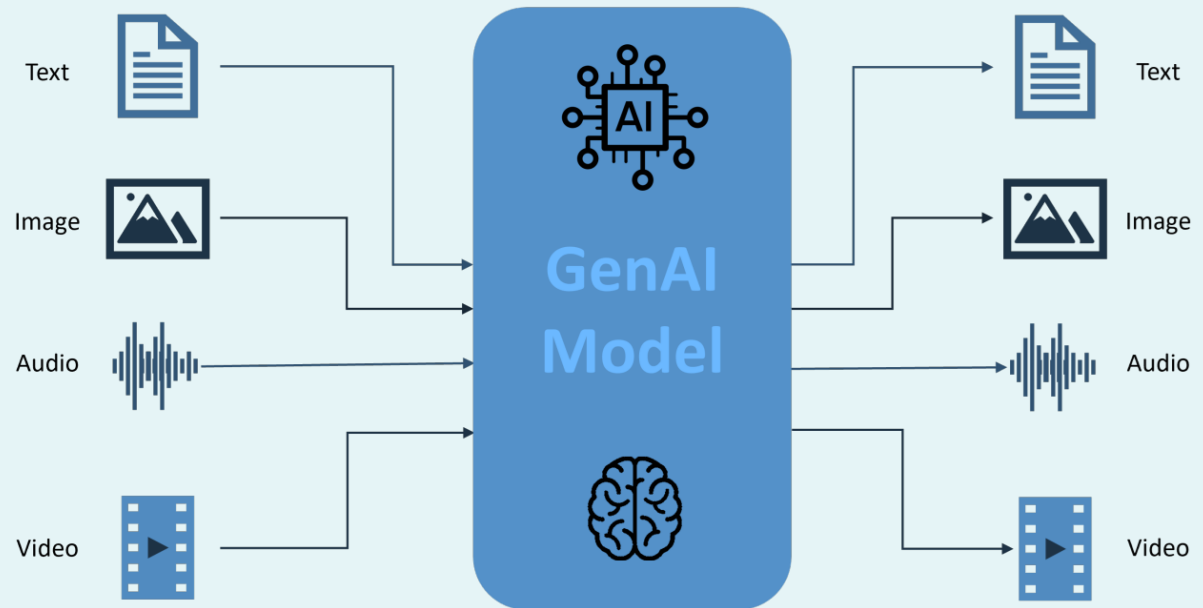
Één multimodal model bestaat uit een complexe samenvoeging van meerdere unimodale modellen

Meerdere types input tegelijk mogelijk

Zelfs input als:

- Geur, PH waarde, temperatuur, (lucht)druk etc
- Integratie met sensordata
- Chemische sensoren en meer

Grootste voordeel: **context**



T&O: Multimodal AI – use case autoschade

De toekomst van schadebehandeling?

Betrokkenen schetsen op basis van tekst, spraakmemo, video, tekening en meer de situatie.

Multimodale AI gebruikt alle inputs ter inschatting situatie

Explainable AI (XAI) kan uitkomst Multimodale AI verklaren

AI Angst: Professionele angst

Neemt AI mijn baan over?

- Neemt AI over? Zo ja, wat?
- Baanzekerheid, creativiteit, authenticiteit?
- Geen glazen bol, geen antwoorden, maar...

"AI won't replace Human Connection" – Sam Altman, OpenAI (ChatGPT)

- In 1997 schaakgrootmeest Garry Kasparov verslagen door AI DeepBlue
- Schaken niet minder populair, zelfs populairder dan ooit
- Dezelfde moraal geldt voor bijvoorbeeld klantenservice



AI Angst: Persoonlijke angst

Deepfakes en openbare data

*"We used publicly available data and licensed data" – Mira Murati,
CTO OpenAI*

Openbaar \neq openbaar domein

Data van sociale media zijn gebruikt voor het trainen van de modellen

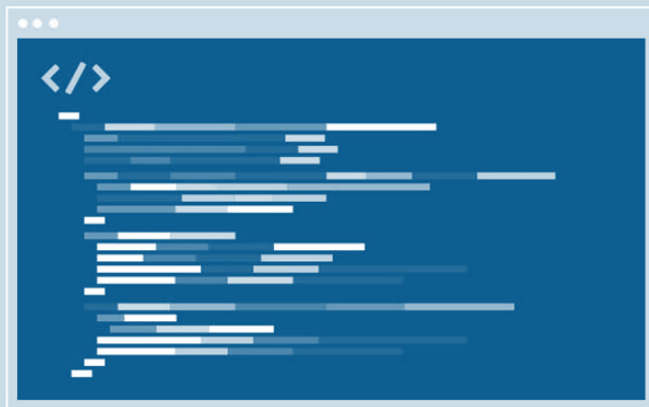
Ook data van onwetende personen

(No-)code & low-code

Overschaduwd door opkomst van AI, maar in forse opmars!

“Low code technologies revenues have doubled in the past three years” – Gartner

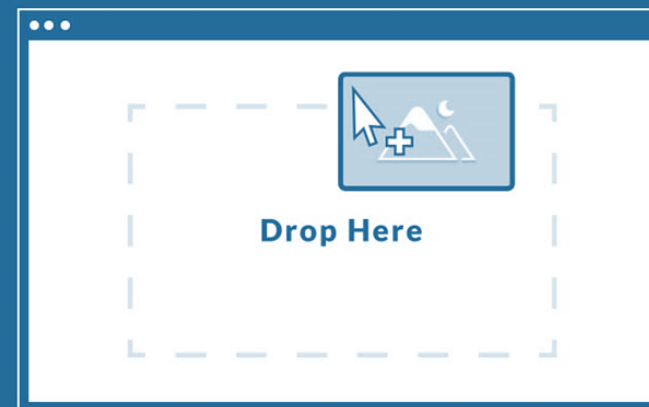
- Iedereen gebruikt no-code / low-code in het dagelijks leven, **zoek kansen waar mogelijk!**
- Gartner (IT consultancy) voorspelt dat in 2025 meer dan 70% van de applicaties in no-code / low-code gebouwd zal zijn



- Geschikt voor: niet-technische gebruikers, snelle prototypes, start-ups met minder middelen en meer

Voordelen:

- Makkelijk in gebruik
- Snelle development van tools, dashboards etc mogelijk
- Meer automatisering
- Lagere kosten (vooral in uurloon)
- Makkelijke data integratie
- Flexibiliteit



Agenda

1

Opening & Overview

2

Trends & Ontwikkelingen

3

Databronnen

Pauze

4

Datakwaliteit

5

Architectuur

6

Data Visualisatie

Databronnen

Een greep uit tal van databronnen

Interne data

Klantdata



Financiële transacties



Productdata



Websites



Sales



Marketing



Rapportages



Klantenservice



Social media



HR data



Externe data

Openbare data



Semi-openbare data



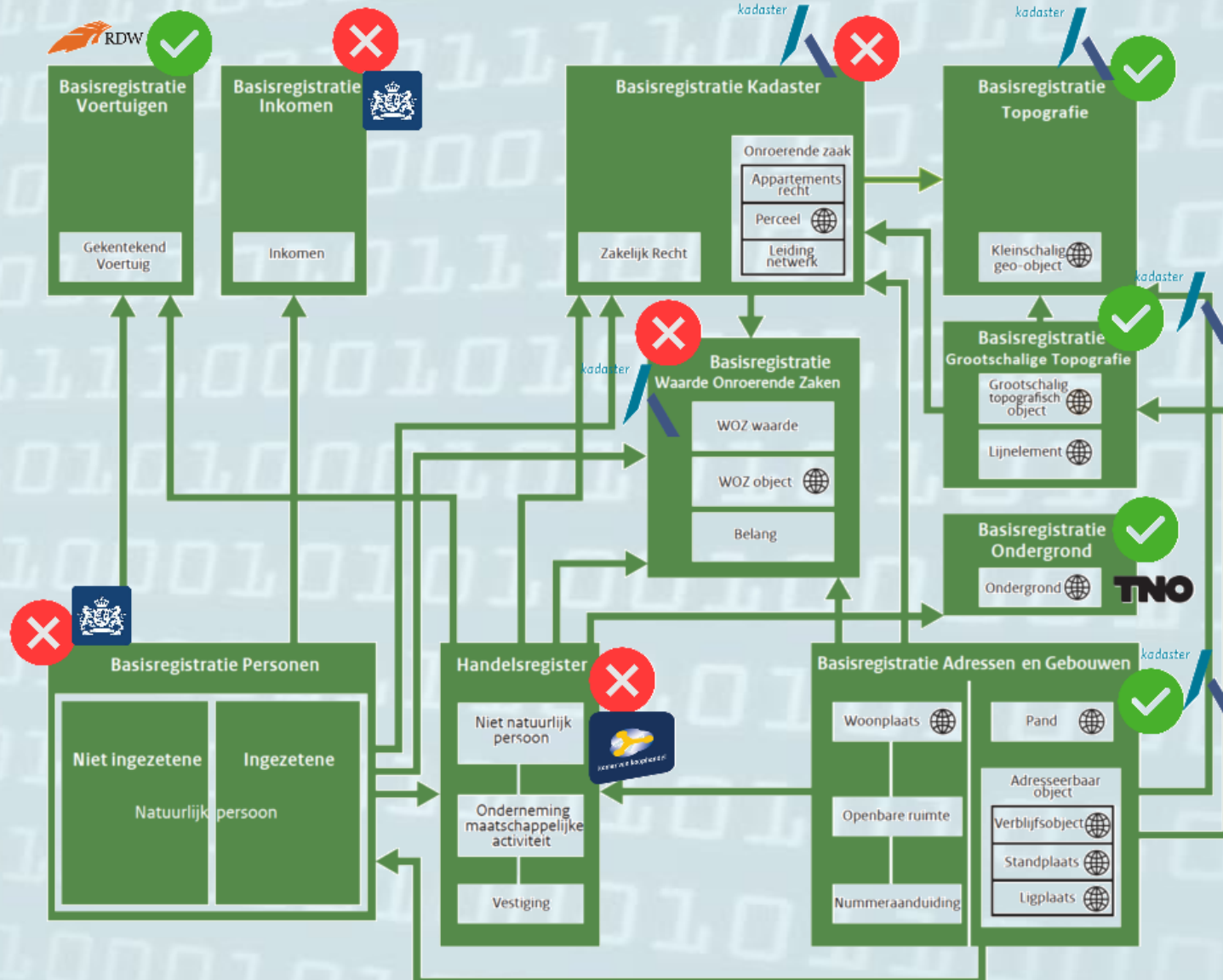
Belastingdienst

Betaalde data



Databronnen

De basisregistraties van de overheid



Databronnen

Gestructureerd vs semi-gestructureerd vs ongestructureerd

Gestructureerde data

Georganiseerd op een vooraf bepaalde manier en structuur, met vaste velden en datatypes.



Semi-gestructureerde data

Georganiseerd met enige mate van structuur, maar niet zodanig als in bijvoorbeeld een database.



Ongestructureerde data

Niet georganiseerd op een vooraf bepaalde manier/structuur.



Databronnen



Financiële transacties



Databronnen

kadaster



Databronnen



Websites



Rapportages



Klantenservice



Databronnen

Het blijft haast oneindig toenemen



Databronnen: the sky is the limit

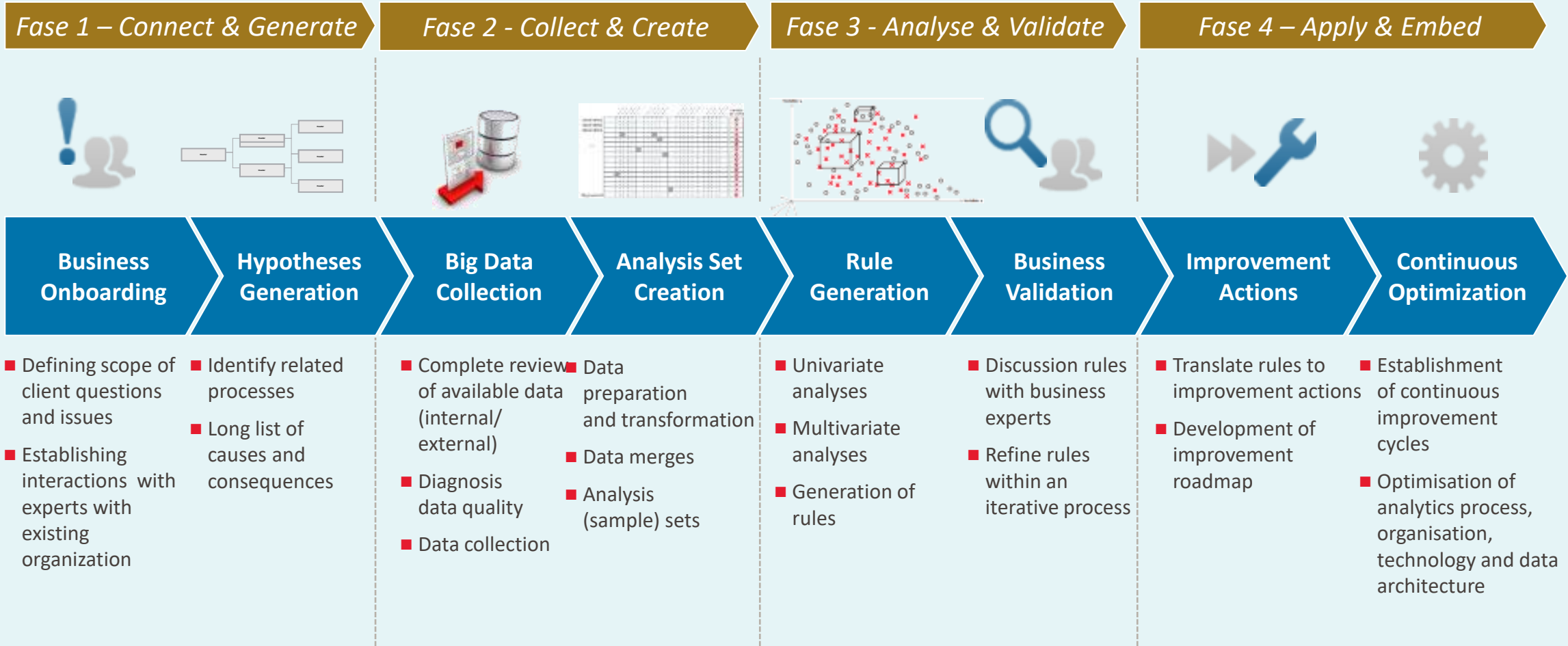


Stel: je hebt toegang tot alle data ter wereld.
Welke data zou je dan het liefst willen gebruiken?



Van hypotheses tot data collectie tot analyse tot verbetering

Hoe data uit zeer verschillende bronnen kunnen bijdragen aan het oplossen van business problemen



Van hypotheses tot data collectie tot analyse tot verbetering

Drie succesvolle use case voorbeelden uit verschillende sectoren



Agenda

1

Opening & Overview

2

Trends & Ontwikkelingen

3

Databronnen

Pauze

4

Datakwaliteit

5

Architectuur

6

Data Visualisatie

Datakwaliteit

Het wezenlijke belang van datakwaliteit



Datakwaliteit: de business case

Het belang van datakwaliteit

IBM: de totale kosten die gepaard gaan met slechte datakwaliteit in 2016 in de VS alleen: **\$3.1 triljoen** (= \$3.100 miljard = \$3.100.000 miljoen = \$3.100.000.000.000)

Uit onderzoek van **Anaconda** bleek in 2021 dat Data Scientists ongeveer **45%** van hun tijd kwijt zijn aan het prepareren van data.

Thomas Redman (Data Quality Solutions): er wordt geschat dat de kosten van slechte datakwaliteit **15% tot 25% van de totale omzet** bedragen voor de meeste bedrijven.

Gartner: in 2021 waren de kosten die voortvloeien uit slechte datakwaliteit in de VS gemiddeld ongeveer **\$12.9 miljoen** per organisatie.

Datakwaliteit

Hoe leidt slechte datakwaliteit tot kosten?

- Direct: garbage in = garbage out
 - Foutieve dashboards -> foutieve conclusies
 - Foutieve analyses/conclusies
 - Foutieve data in model -> foutieve uitkomsten
 - Foutieve rapportages
- Indirect
 - Tijd kwijt met corrigeren
 - Tijd kwijt met controleren
 - Tijd kwijt met herhalen van proces vanwege fouten



Datakwaliteit dimensies

Wat zijn de checkpoints van datakwaliteit?



Datakwaliteit

Oorzaken van gebrekkige datakwaliteit bij financiële instellingen

Legacy systemen

Fusies en overnames

Volmachten

Geen 'digital nomad' model

Data in silo's

Gebrek aan standaardisatie aan de voorkant

(Wijzigende) mappings

Handmatige invoer

Gebrek aan bewustzijn belang

Gebrekkige governance



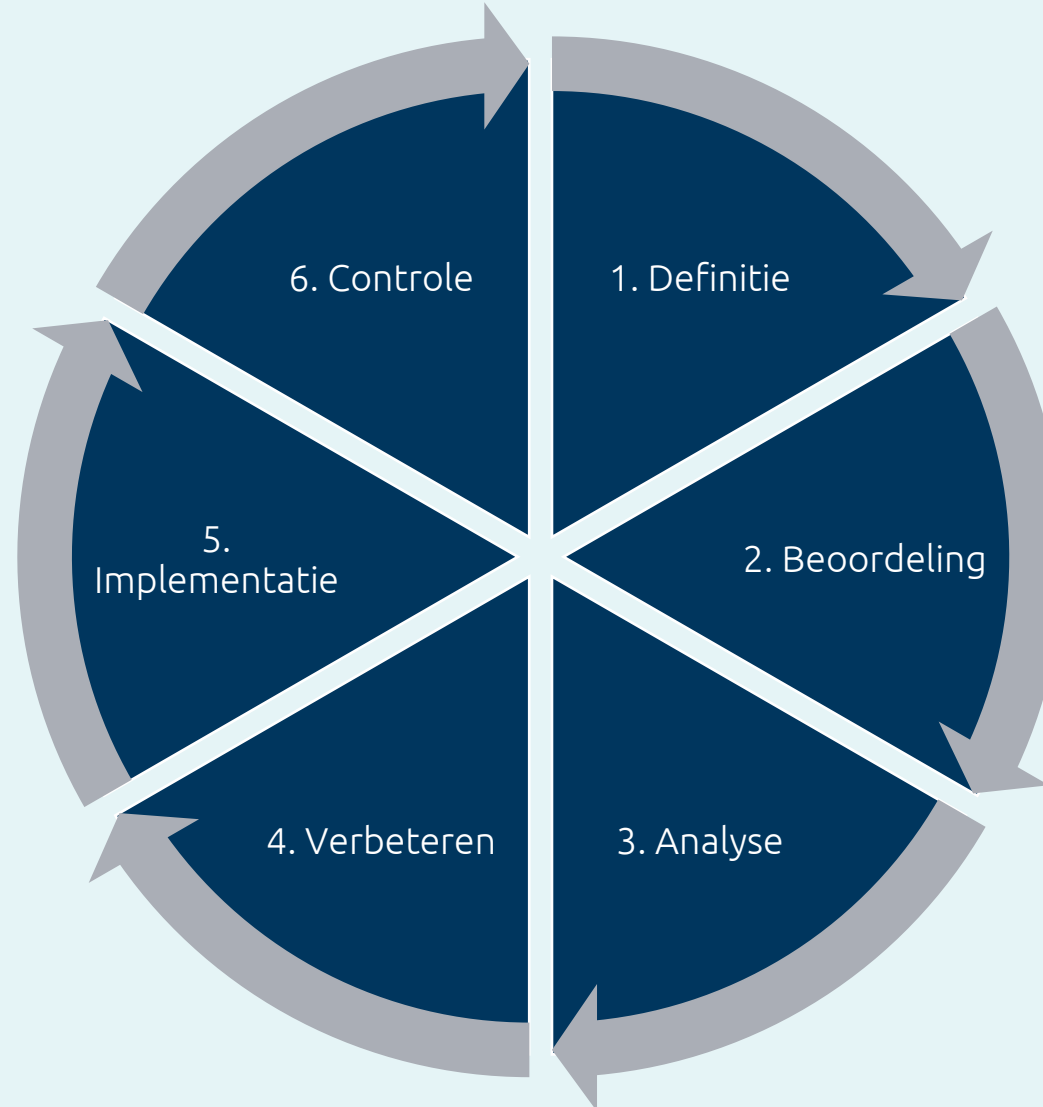
Datakwaliteit

Het waarborgen van datakwaliteit

Richt een controleproces in dat je in staat stelt om periodiek je datakwaliteit te beoordelen

Implementeer deze verbeteringen, het liefst aan de "voorkant"

Stel concrete verbeterpunten op gebaseerd op de analyse uit stap 3



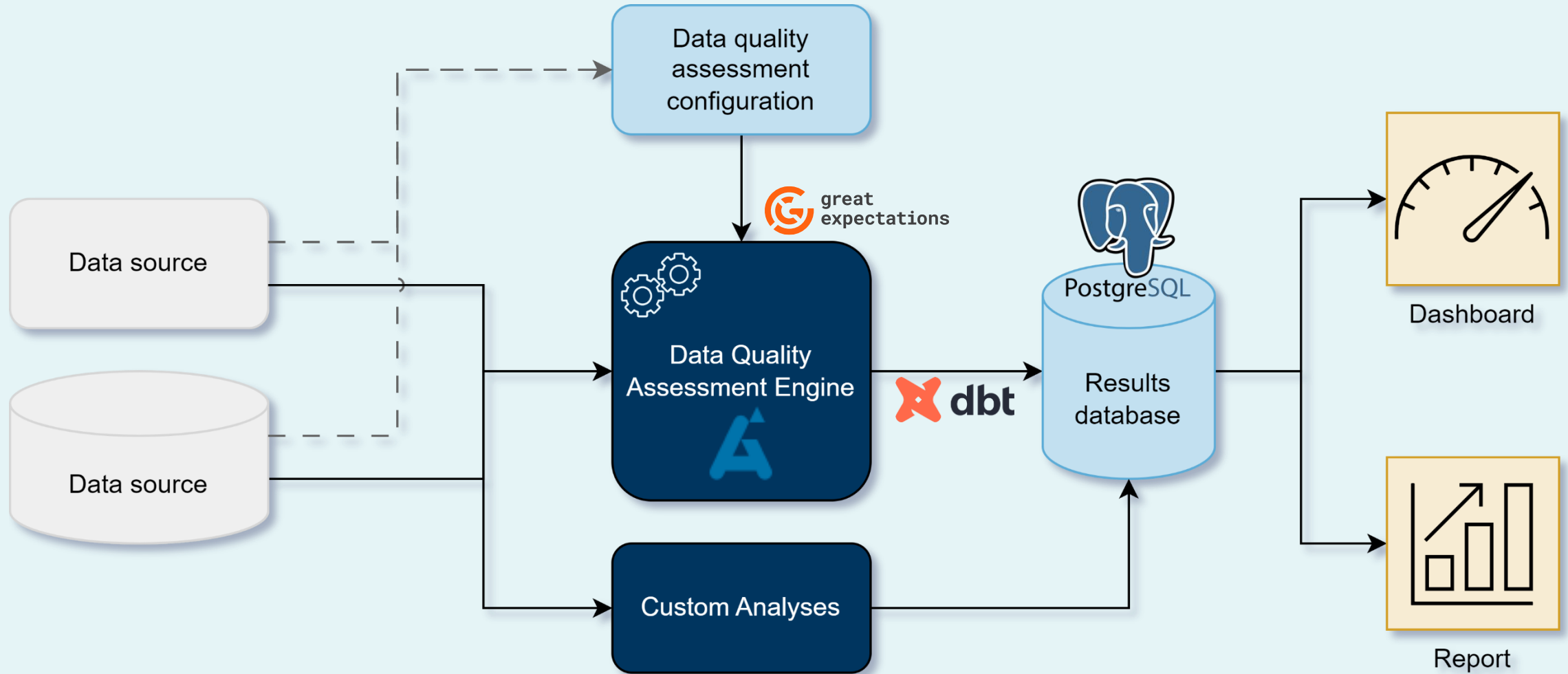
Definieer je doelen voor wat betreft verbetering van datakwaliteit

Beoordeel de kwaliteit van je data over verschillende datakwaliteitsdimensies en geldende business-rules

Analyseer de resultaten uit stap 2. Probeer te identificeren waar en waarom datakwaliteit tekort schiet.

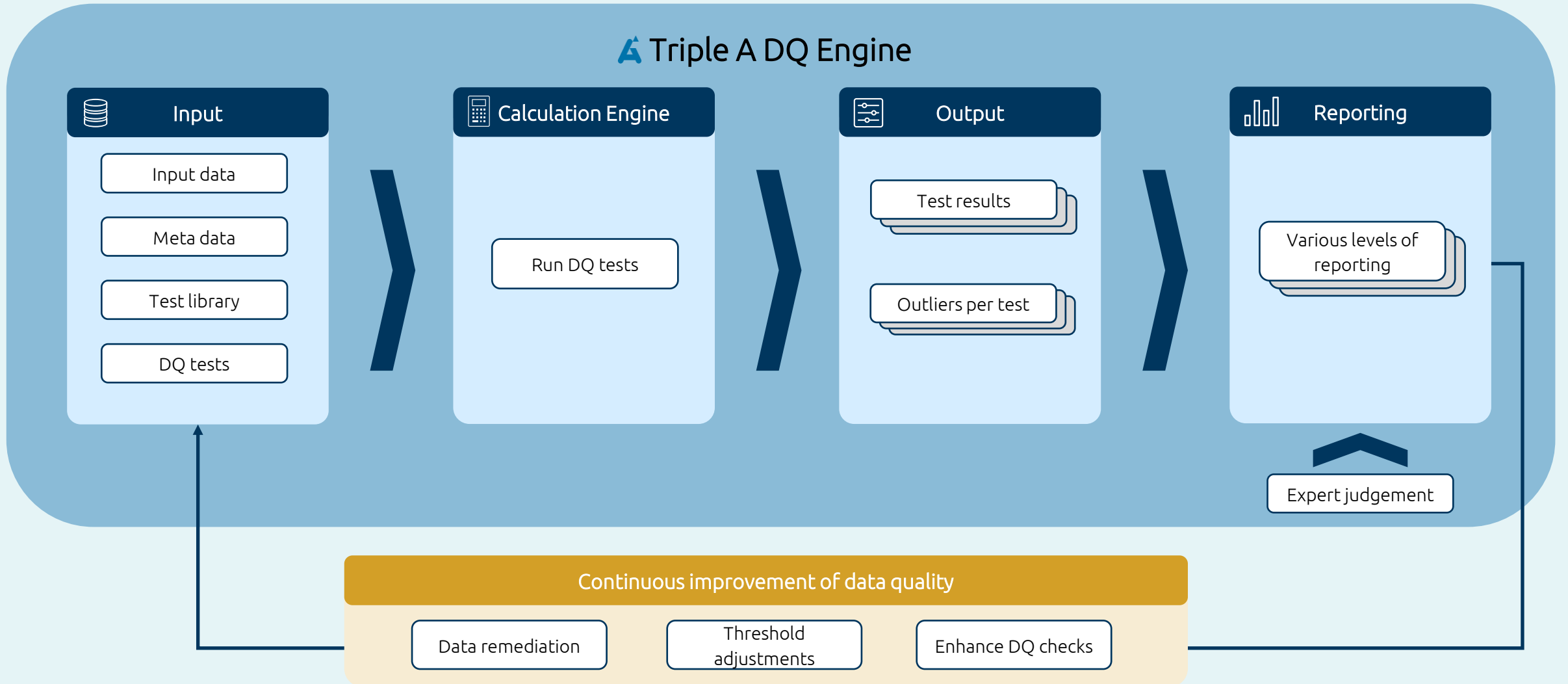
De Data Quality Engine

Architectuur



Datakwaliteit verbeterproces

Van data input tot reporting



Data Quality Engine

Configuration

Run

Reporting

About



Configuration

1 Test configuration

Upload configuration file

Manual test configuration

2 Data selection

Upload data file

Upload data map

3 CDE specification

Upload CDE file

Manual CDE specification

4 Primary key specification

Upload primary key file

Manual primary key specification

5 Test-CDE mapping

Upload test-CDE mapping file

Manual test-CDE mapping

Test configuration

Test library	Threshold lowerbound	Threshold upperbound	Reporting dimension
Determine number of missing attribute values	0.95	0.99	Completeness
Determine number of duplicate records based on primary key	0.99	1	Uniqueness
Assess whether attribute has the expected properties	n.a.	n.a.	Validity
Determine number of attribute values not within domain	0.92	0.97	Validity
Compute summary statistics	n.a.	n.a.	Accuracy
Determine number of the IQR outliers	0.9	0.95	Accuracy
Assess logical relationship between two or more attributes within dataset	0.9	0.95	Consistency

Save

Download as CSV file

Back to main screen

- De testconfiguratie geeft een overzicht van alle soorten tests die kunnen worden uitgevoerd.
- Standaard drempelniveaus en rapportage dimensies kunnen worden aangepast.
- In het geval van handmatige configuratie kan het bestand worden opgeslagen en gedownload als een csv-bestand. Bij het uitvoeren van een nieuwe datakwaliteitsanalyse op basis van dezelfde testdefinities kan het bestand opnieuw worden geüpload.



Data Quality Engine

Configuration

Run

Reporting

About



Configuration

1 Test configuration

Upload configuration file

Manual test configuration

2 Data selection

Upload data file

Upload data map

3 CDE specification

Upload CDE file

Manual CDE specification

4 Primary key specification

Upload primary key file

Manual primary key specification

5 Test-CDE mapping

Upload test-CDE mapping file

Manual test-CDE mapping

CDE specification

+ Add row

- Remove row

CDE ID	Dataset	Name of column	Type	Missing allowed	CDE	Field length	Decimal length (in case of numeric)	Domain values	Notes
1	Customer_loans	Name	Str	<input type="checkbox"/>	<input checked="" type="checkbox"/>	50		[]	
2	Customer_loans	Date of birth	Str	<input type="checkbox"/>	<input checked="" type="checkbox"/>	10		[01-01-1900:01-01-2006]	
3	Customer_loans	Address	Str	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	100		[]	
4	Customer_loans	Income (monthly)	Num	<input type="checkbox"/>	<input checked="" type="checkbox"/>	7	2	[0:100000]	
5	Customer_loans	Employment status	Str	<input type="checkbox"/>	<input checked="" type="checkbox"/>	20		[P, F, parttime, fulltime]	
6	Customer_loans	Assets	Num	<input type="checkbox"/>	<input checked="" type="checkbox"/>	7	2	[0:100000]	
7	Customer_loans	Loan ID	Str	<input type="checkbox"/>	<input checked="" type="checkbox"/>	8		[CCNNNNNN]	
8	Customer_loans	Loan amount	Num	<input type="checkbox"/>	<input checked="" type="checkbox"/>	7	2	[0:100000]	
9	Customer_loans	Interest rate	Num	<input type="checkbox"/>	<input checked="" type="checkbox"/>	6	4	[-100:100]	
10	Customer_loans	Loan term (months)	Num	<input type="checkbox"/>	<input checked="" type="checkbox"/>	3	0	[0:360]	

Save

Download as CSV file

Back to main screen

- De kritische data-elementen (CDE's) zijn gebaseerd op de kolommen die getest moeten worden op datakwaliteit.
- Voor elke CDE moeten meerdere eigenschappen worden opgegeven.
- Net als bij de testconfiguratie: na het handmatig invoegen van alle CDE's kan deze tabel worden opgeslagen en gedownload als csv-bestand om opnieuw te worden geüpload voor toekomstige datakwaliteitsanalyses.



Data Quality Engine

Configuration

Run

Reporting

About



Configuration

1 Test configuration

Upload configuration file

Manual test configuration

2 Data selection

Upload data file

Upload data map

3 CDE specification

Upload CDE file

Manual CDE specification

4 Primary key specification

Upload primary key file

Manual primary key specification

5 Test-CDE mapping

Upload test-CDE mapping file

Manual test-CDE mapping

Test-CDE mapping

+ Add row

- Remove row

Check ID	Dataset	Name of column	Test name	Reporting dimension	Test ID	Condition	Threshold lowerbound	Threshold upperbound	Reporting dimension override
1	Customer_loans	Name	Determine number of mi	Completeness	1		0.98	1	
2	Customer_loans	Date of birth	Determine number of mi	Completeness	1		0.95	0.99	
3	Customer_loans	Address	Determine number of mi	Completeness	1		0.95	0.99	
4	Customer_loans	Income (monthly)	Determine number of mi	Completeness	1	Group by gender	0.95	0.99	
5	Customer_loans	Employment status	Determine number of mi	Completeness	1		0.95	0.99	
6	Customer_loans	Loan ID	Determine number of mi	Completeness	1		0.98	1	
7	Customer_loans	Loan amount	Determine number of mi	Completeness	1		0.95	0.99	
8	Customer_loans	Interest rate	Determine number of mi	Completeness	1		0.95	0.99	
9	Customer_loans	Loan term (months)	Determine number of mi	Completeness	1		0.95	0.99	

Save

Download as CSV file

Back to main screen

- Uit de testdefinities die in stap 1 zijn gedefinieerd, kunnen de tests worden geselecteerd die moeten worden uitgevoerd.
- Voor elke geselecteerde test kunnen de CDE's worden gekozen waarop de test moet worden toegepast.
- Voor elke combinatie van tests en CDE's kunnen optionele of verplichte testvoorwaarden worden gespecificeerd.
- Het is mogelijk om de drempelniveaus en rapportage dimensie aan te passen wanneer afwijking van de standaardwaarden gewenst is.



Data Quality Engine

Configuration

Run

Reporting

About

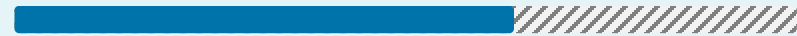


Run

Run Data Quality Analysis

0%

100%



Data Quality Engine

Configuration

Run

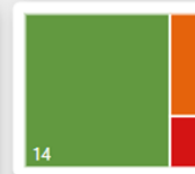
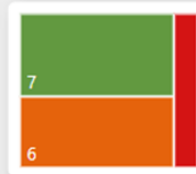
Reporting

About



Reporting

STATUS

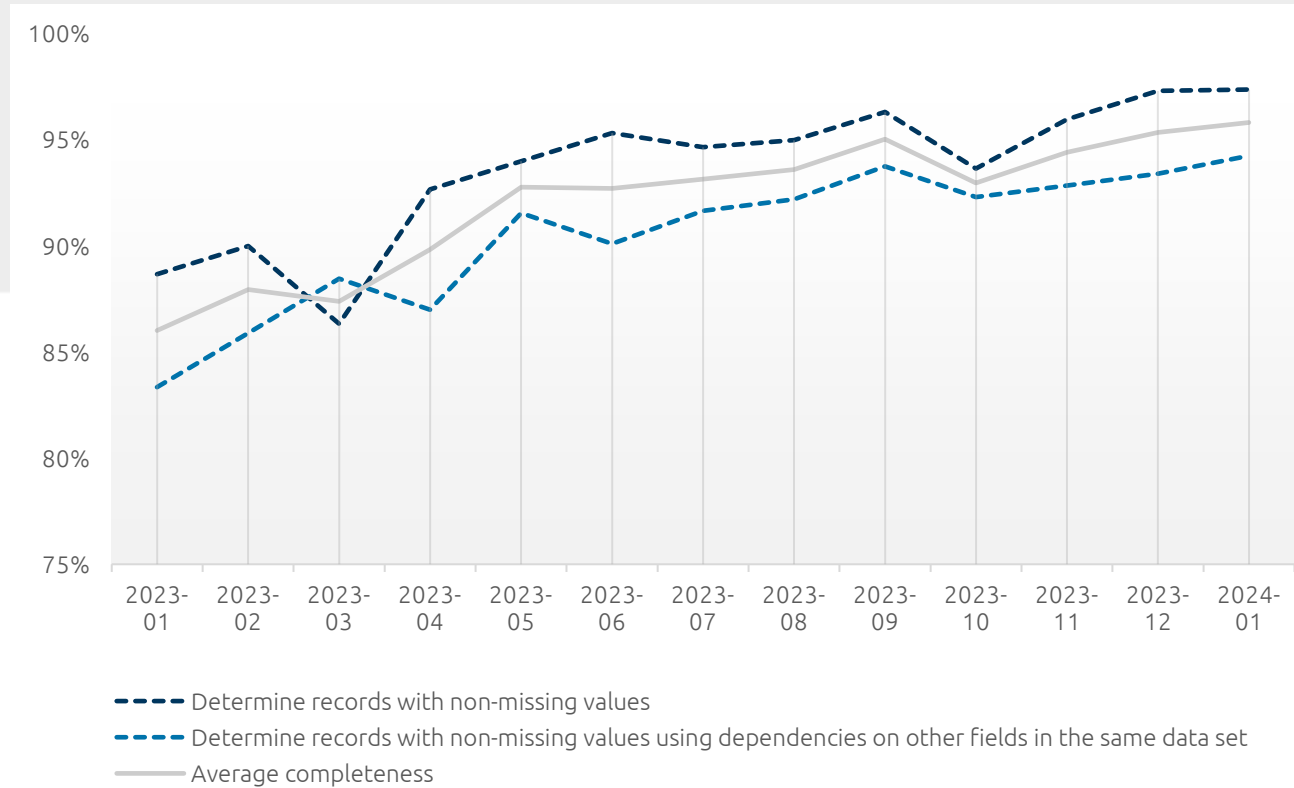


CDE	Completeness	Uniqueness	Validity	Accuracy	Consistency
Purpose of the loan					
Address					
Name					
Employment status					
Loan ID					
Interest rate					
Default2					
Default					
Assets					
Loan term (months)					
Number of days past due					
Loan amount					
Date of birth					
LTV					
Credit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

OUTLIERS TABLE - COMPLETENESS

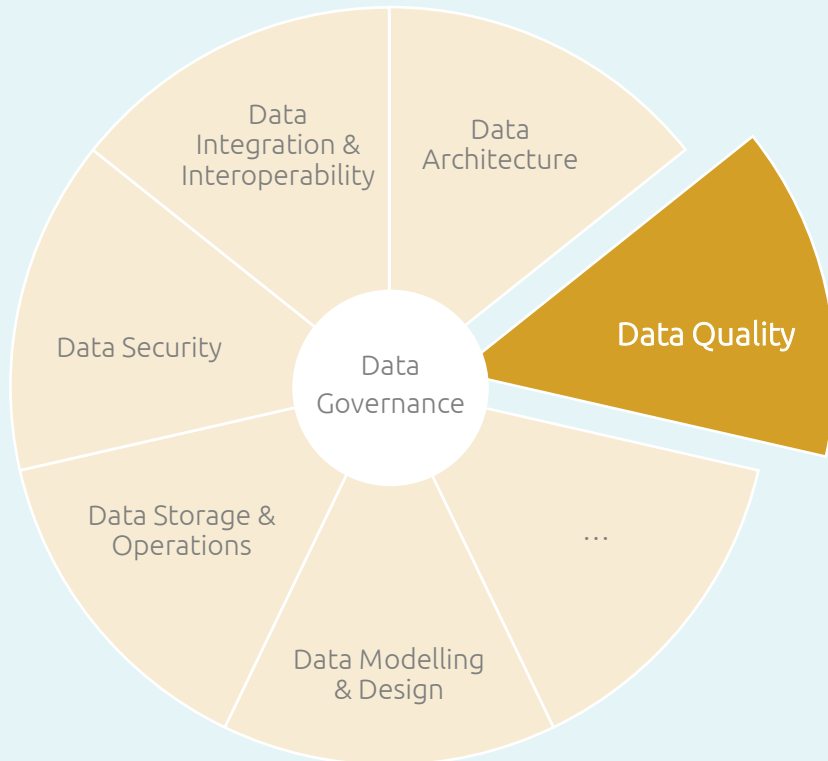
Test ID	Table	Loan ID (primary key)	Date of birth	Loan amount	Loan term	Monthly payment amount
16	Customer_loans	133709	13-4-1957		12	
16	Customer_loans	4678896	30-3-1944		2	
16	Customer_loans	9008779	24-12-1959			300
16	Customer_loans	23456764	7-7-1939			800
16	Customer_loans	628463	8-10-1959		1	200
16	Customer_loans	8595749	9-1-1959		10	
16	Customer_loans

AVERAGES OVER TIME - COMPLETENESS



Datakwaliteit binnen Data Management framework

Data Management Framework (DAMA wheel)



Data Quality Framework



Data Quality Engine

Completeness

Accuracy

Consistency

Validity

Uniqueness

Timeliness

Traceability

Accessibility

Interactieive Break



Agenda

1

Opening & Overview

2

Trends & Ontwikkelingen

3

Databronnen

Pauze

4

Datakwaliteit

5

Architectuur

6

Data Visualisatie

De blauwdruk van het datalandschap

Data Architectuur

*“Data architectuur is de blauwdruk voor het organiseren, integreren en beheren van data binnen een organisatie”
- IBM*



Organisatie

Organisatiestructuur
Strategie & Beleid
Processen
Wetgeving



Data

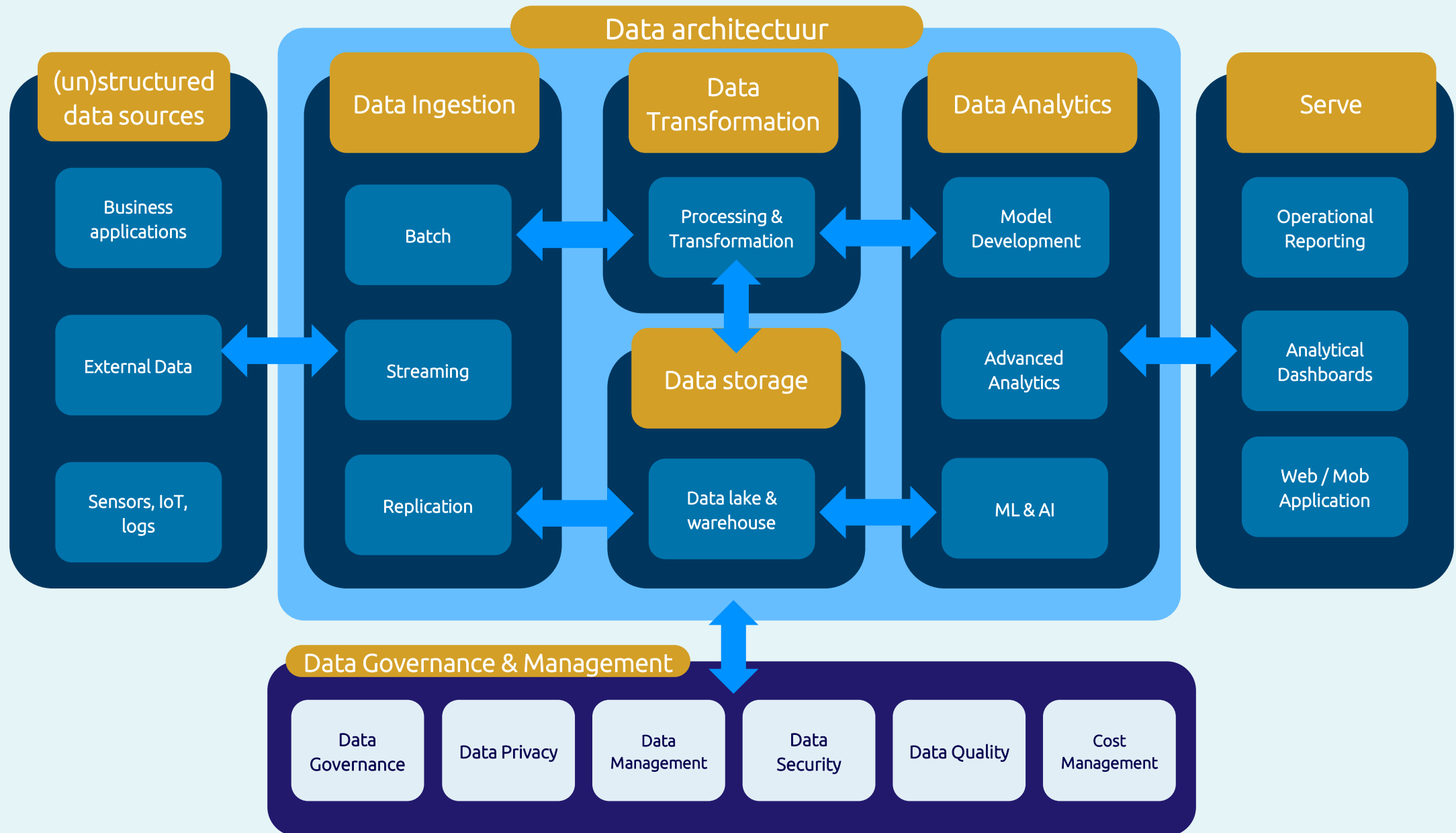
Datamodellen
Opslag
Lifecycle Management



Infrastructuur

Hardware, software, & networking
Applicaties & verbindingen

Kerncomponenten Data Architectuur

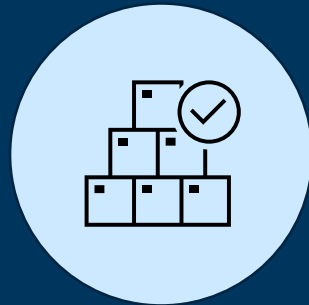


Wat levert goede data architectuur op?

Data Architectuur



Data is **toegankelijk**



Data is **consistent**



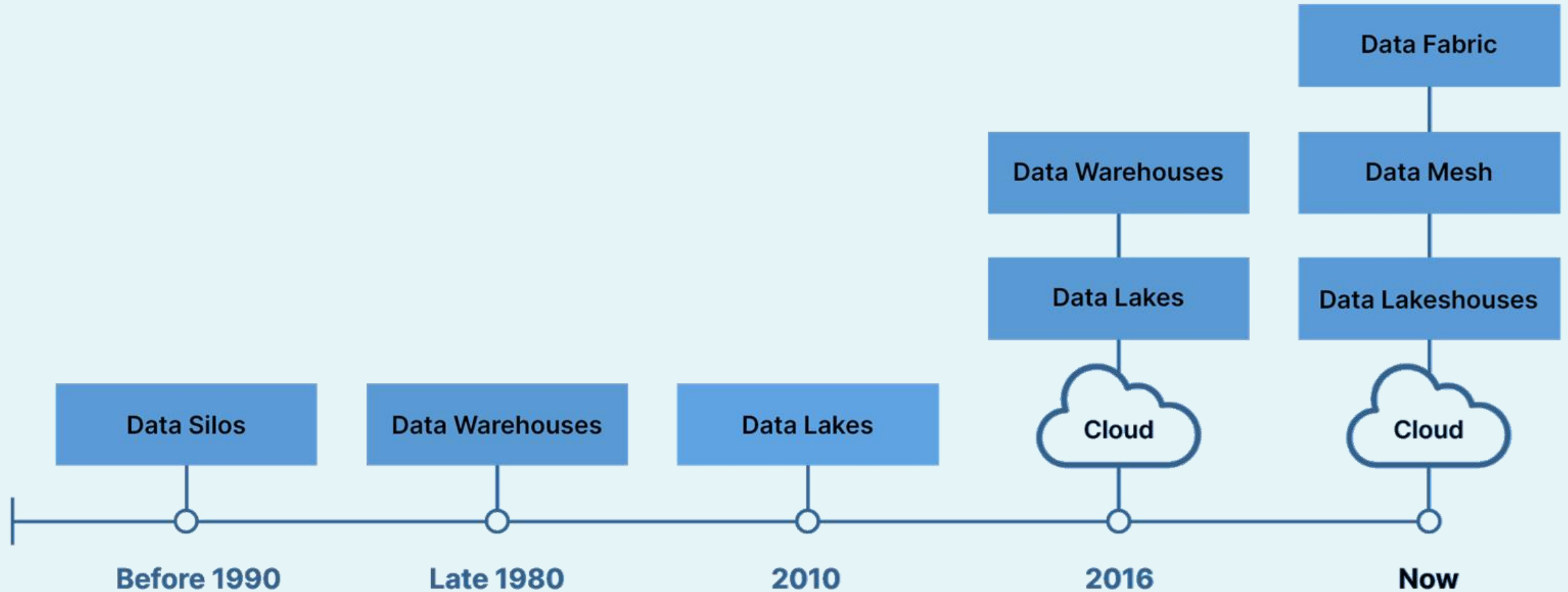
Data van hoge **kwaliteit**



Flexibel:
Simpel,
Schaalbaar, Goedkoop

Waar komen we vandaan en waar staan we nu?

Data Architectuur





Van operationeel naar business intelligence

Data Architectuur – Het Datawarehouse



Reporting & BI

Gestructureerde data

Specifiek

Data Governance,
Kwaliteit & Transparantie



AI & ML

Implementatiekosten
(data modelling)

Schaalbaarheid

Weinig flexibel
(ETL)

De kracht van Big Data benutten

Data Architectuur – Het Datalake



AI & ML

Ongestructureerde data

Schaalbaarheid

Goedkope dataopslag

Flexibel (ELT)



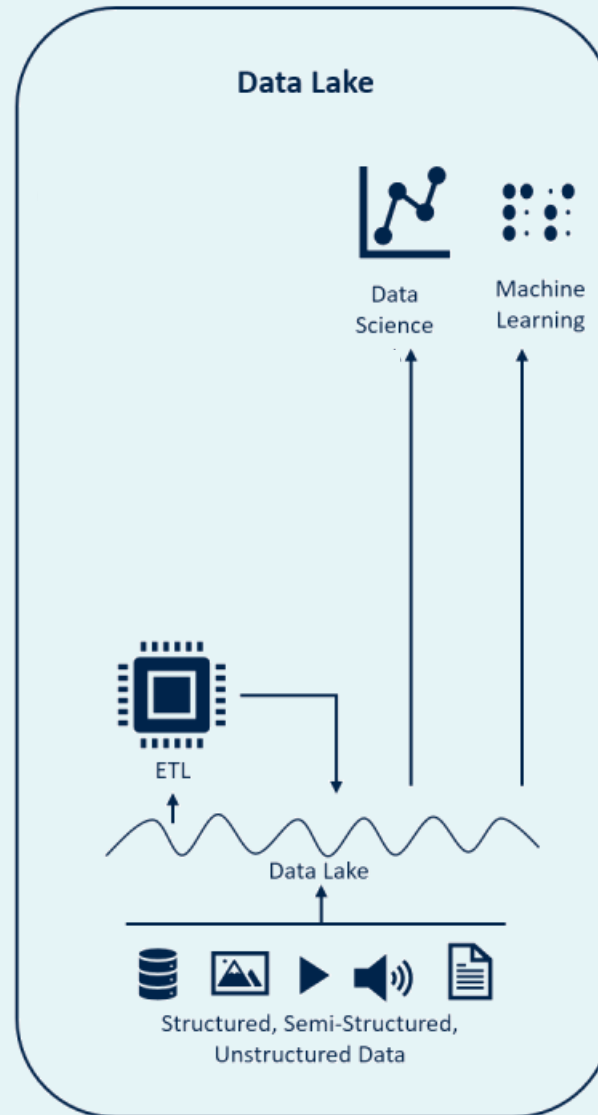
Reporting & BI

Ongestructureerd

Beheerkosten

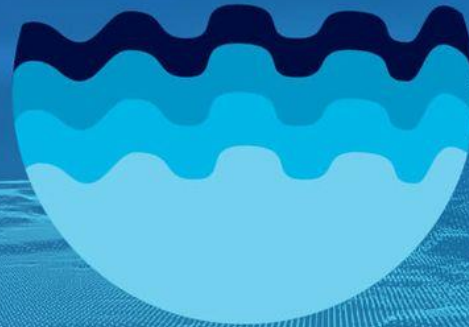
Data Governance,
Kwaliteit & Transparantie

Risico op dataswamp





Data Warehouse



Data Lake

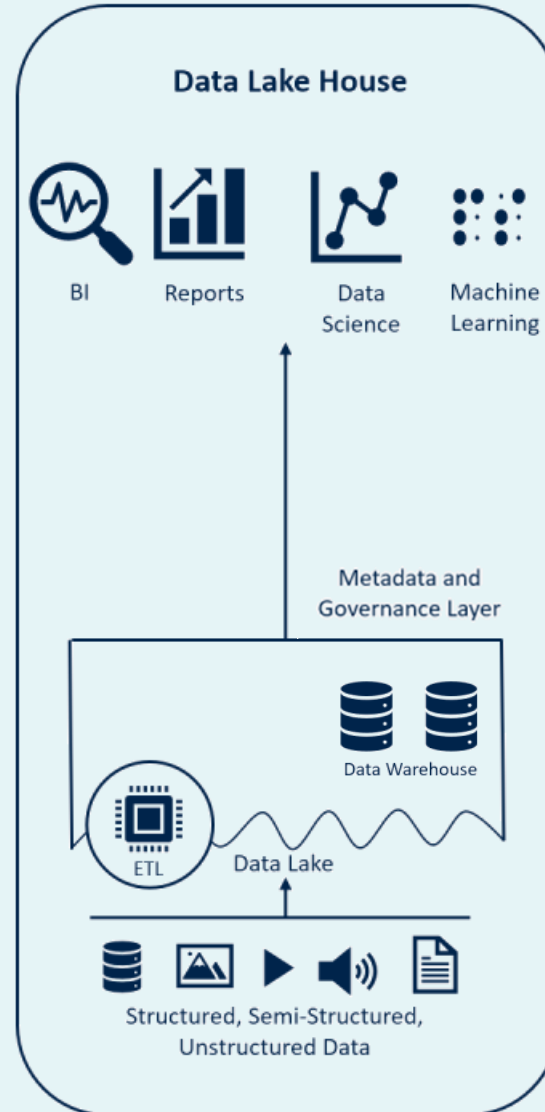


Data Lakehouse



Het beste van beide werelden?

Data Architectuur – Het Datalakehouse



Reporting, BI, AI & ML

(On)gestructureerde Data

Data Governance, Kwaliteit & Transparantie

Flexibel (ELT) en schaalbaar

Eén dataplatform



Integratie verschillende technologieën

Implementatiekosten (data modelling, expertise)

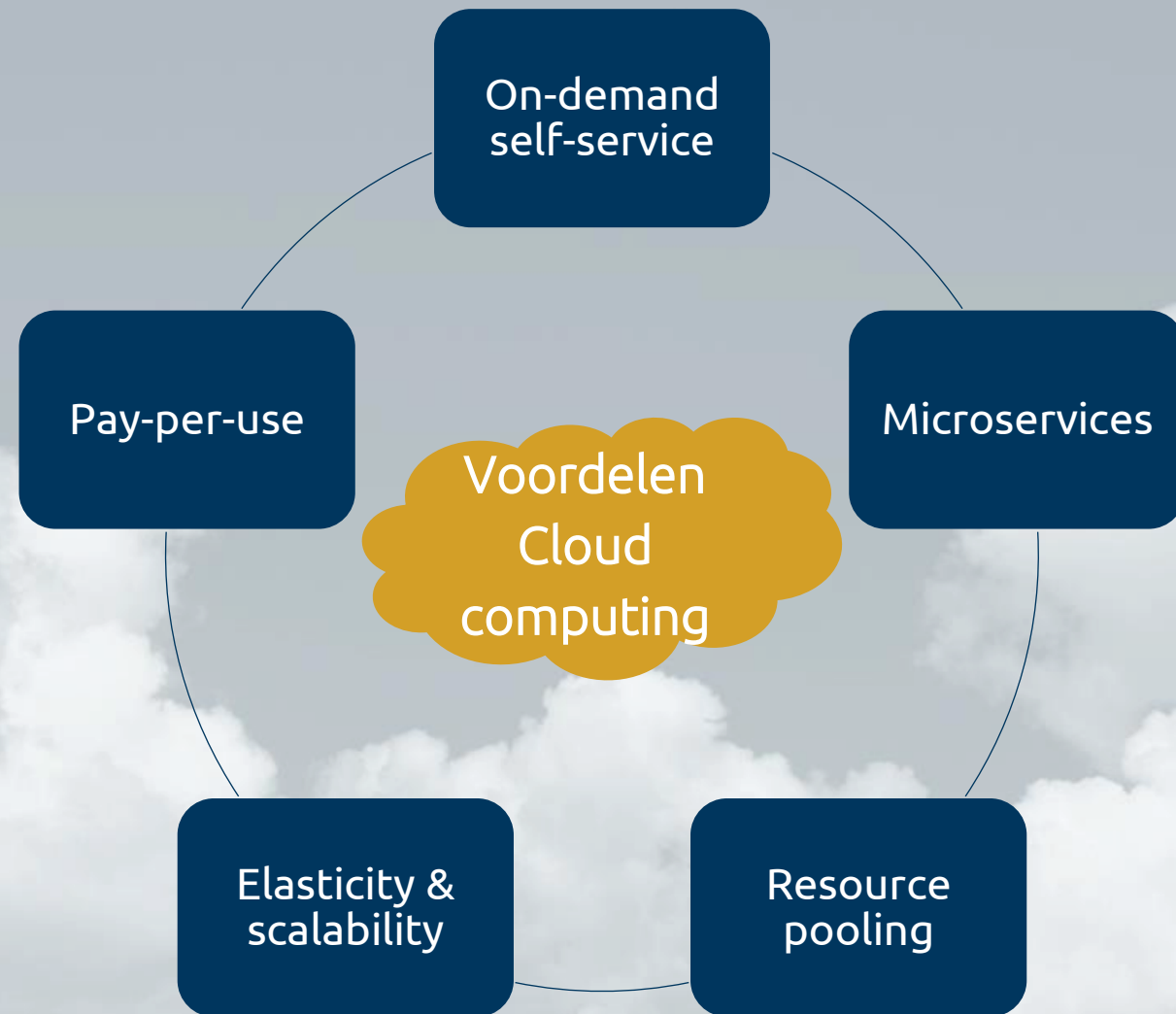
Complexiteit

Volwassenheid

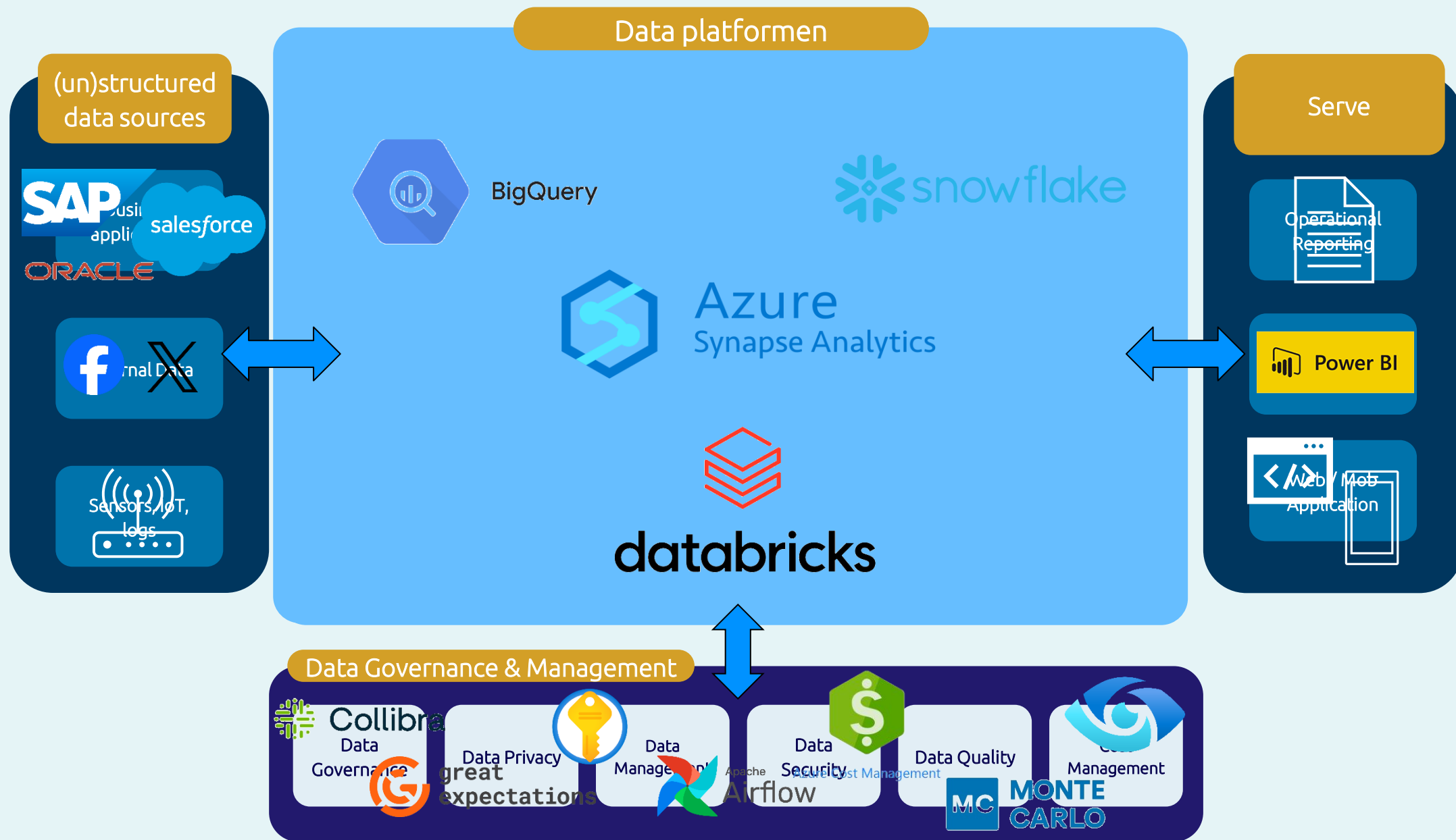


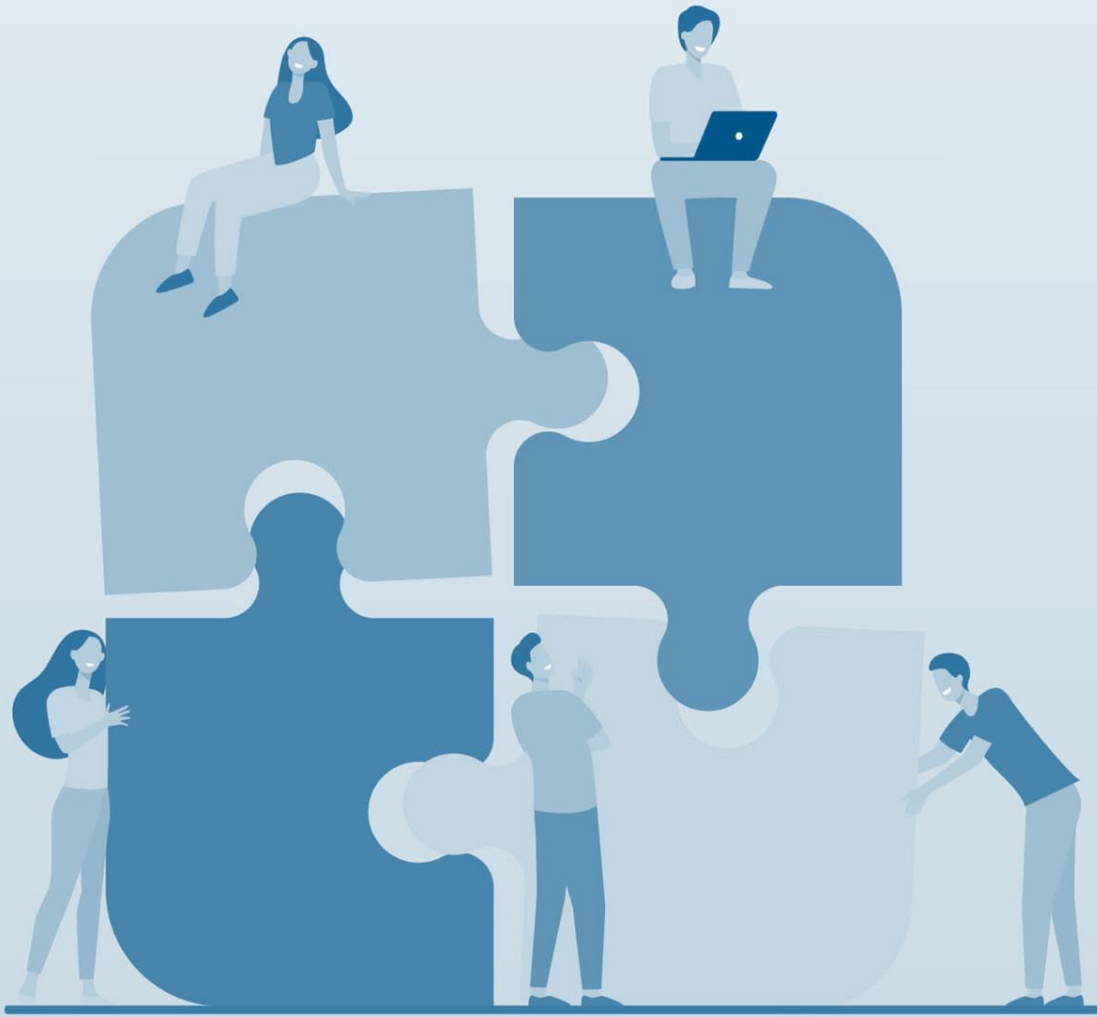
Cloud Data Architecture is de basis van de moderne data stack

Data Architectuur



Kerncomponenten Data Architectuur





De moderne actuaris

Data Architectuur

Data wordt toegankelijker

Interdisciplinaire samenwerking

Kortere doorlooptijd
(actuariële) tooling

Nieuwe verantwoordelijkheden

Agenda

1

Opening & Overview

2

Trends & Ontwikkelingen

3

Databronnen

Pauze

4

Datakwaliteit

5

Architectuur

6

Data Visualisatie

Het belang van Data Visualisatie

Zes redenen waarom Dashboarding belangrijk is voor een organisatie

Inzicht in (gevisualiseerde) beschrijvende statistieken data, KPI's en meer

**Verhoogde
inzichtelijkheid
in data**

**Betere
besluitvorming**

Data gedreven beslissingen zijn objectiever en leiden tot beter resultaat

Inefficiënties identificeren en processen optimaliseren

**Gestroomlijnde
processen**

**Verhoogde
productiviteit**

Medewerkers met juiste en up-to-date informatie zijn productiever

Een waarheid leidt tot een betere samenwerking met een duidelijker doel voor ogen

**Verbeterde
samenwerking**

**Competitief
voordeel**

Gartner: "Organisaties die data dashboarding effectief toepassen behalen over het algemeen een competitief voordeel"



Wat komen we tegen op de werkvloer?

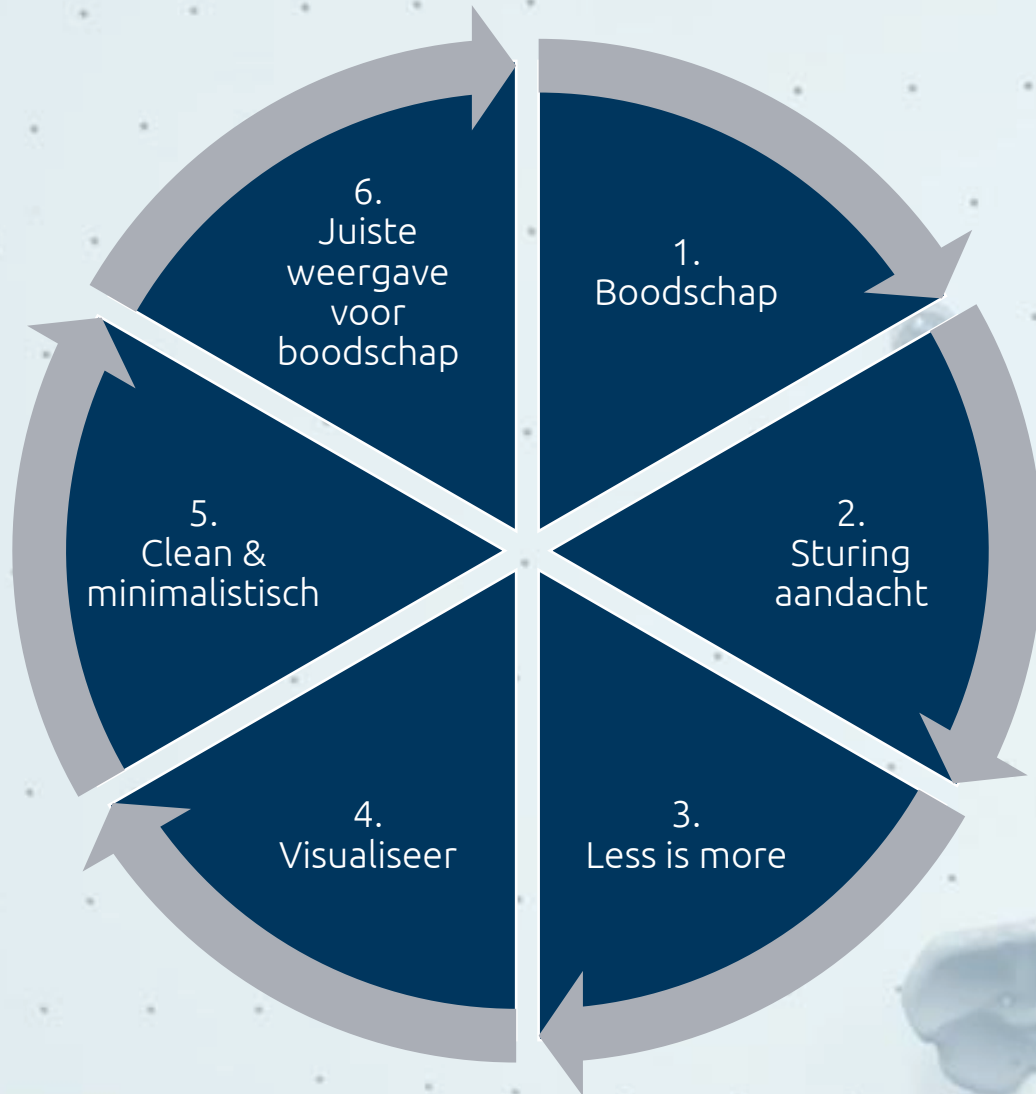
“Iedereen kijkt naar een eigen waarheid”

- Bij klanten steeds grotere vraag naar inzicht in data
- Veelal statistische rapportages
- BI afdelingen zien veel dezelfde type Ad Hoc vragen terugkomen naar aanleiding van opgeleverde (statistische) rapportages.
- Er wordt veel zelf “geklust” in Excel om extra inzichten te genereren

De ideale oplossing:

Een dynamisch dashboard waarin iedereen naar dezelfde data kijkt en direct kan doorklikken naar diepere niveaus wanneer hier behoefte aan is.

Dé zes visualisatie handvatten



Waarom is sterke visualisatie zo belangrijk?

“Een goede visualisatie brengt de kern van de boodschap in één keer over dankzij systeem 1 denken” – Daniel Kahneman, 2011

Picture Superiority Effect: het percentage personen wat de boodschap heeft onthouden is 6,5 keer zo hoog bij inzet visuals

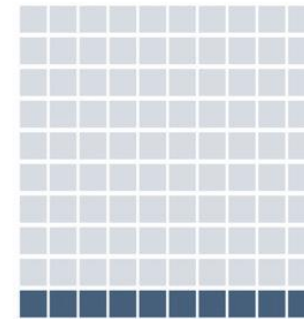
Visuals **60.000** keer sneller verwerkt dan tekst alleen

Learning retention rate **400-700%** hoger door visuals

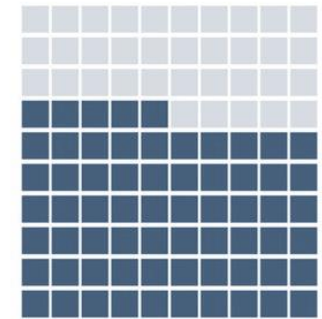
90% van de informatie die naar het brein wordt verstuurd is visuele informatie

Picture Superiority Effect

Memory retention after 3 days



10%
Text or Audio Only



65%
Text + Picture

Visualisatie krachtige tool dankzij systeem 1 denken

Systeem 1

Automatisch
Intuïtief
Impulsief
Snel
Kost weinig energie

95%



Systeem 2

Rationeel
Weloverwogen
Berekenend
Traag
Kost veel energie

5%

Wat zijn de mogelijkheden qua tooling?



Er zijn nog veel meer mogelijkheden

Voorkeur wordt vaak bepaald door budget, bestaande data architectuur en simpelweg gewoonte

Power BI en Tableau het populairst

Visualisatie DO's en DON'Ts

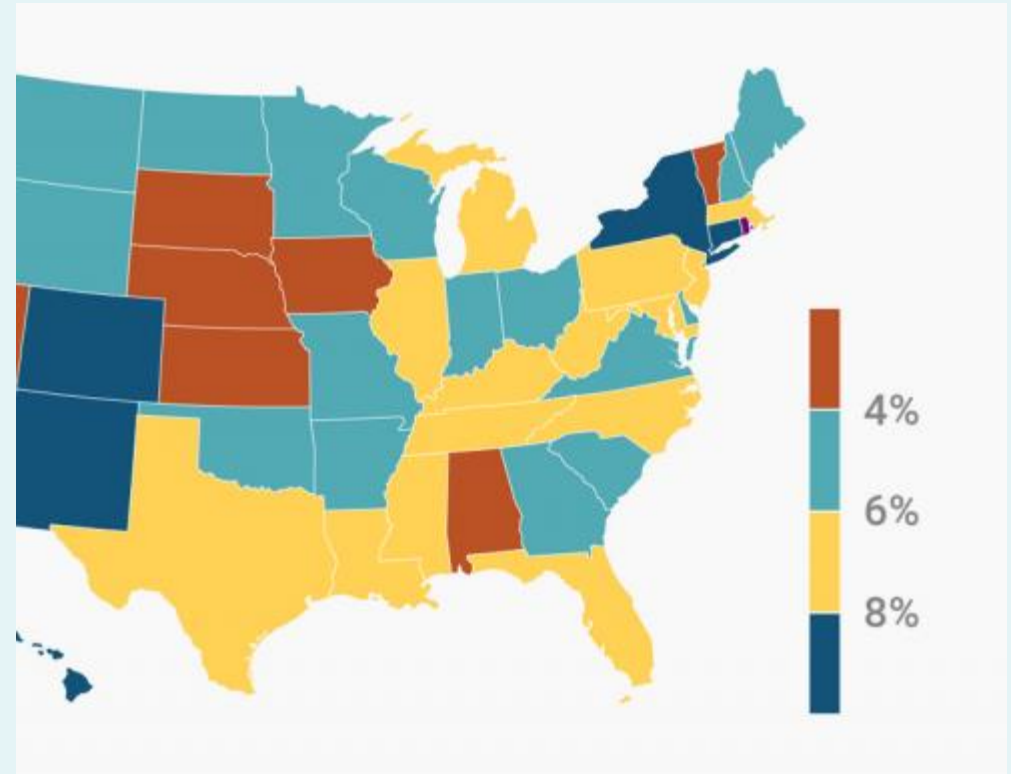
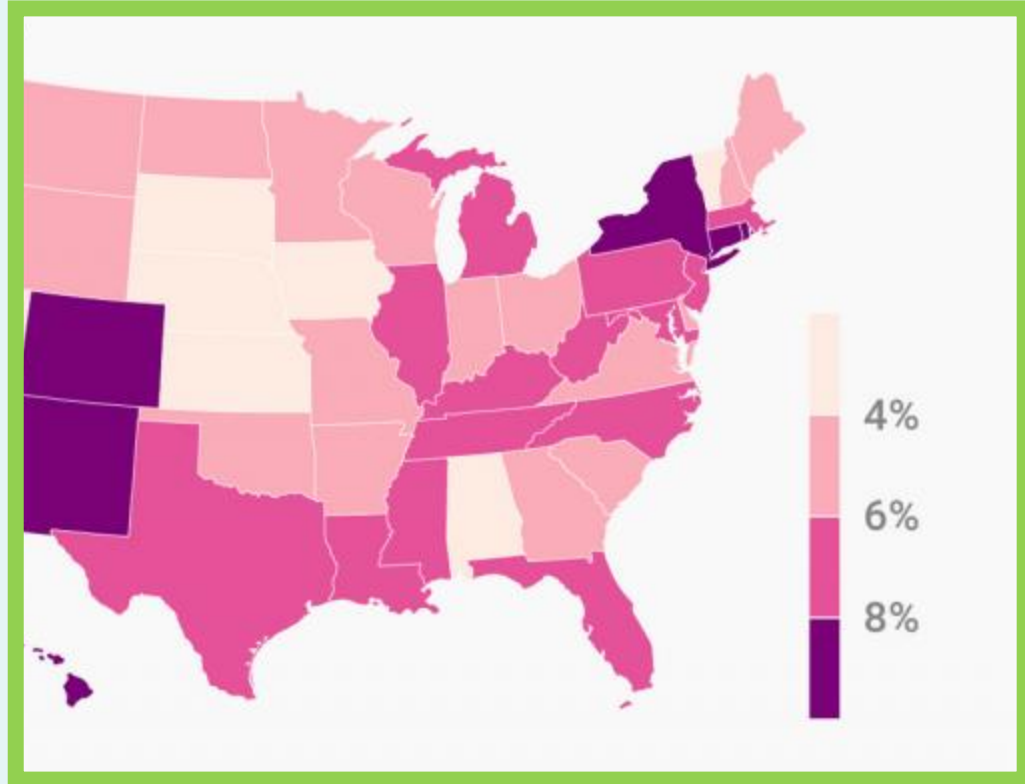


DOs



DON'Ts

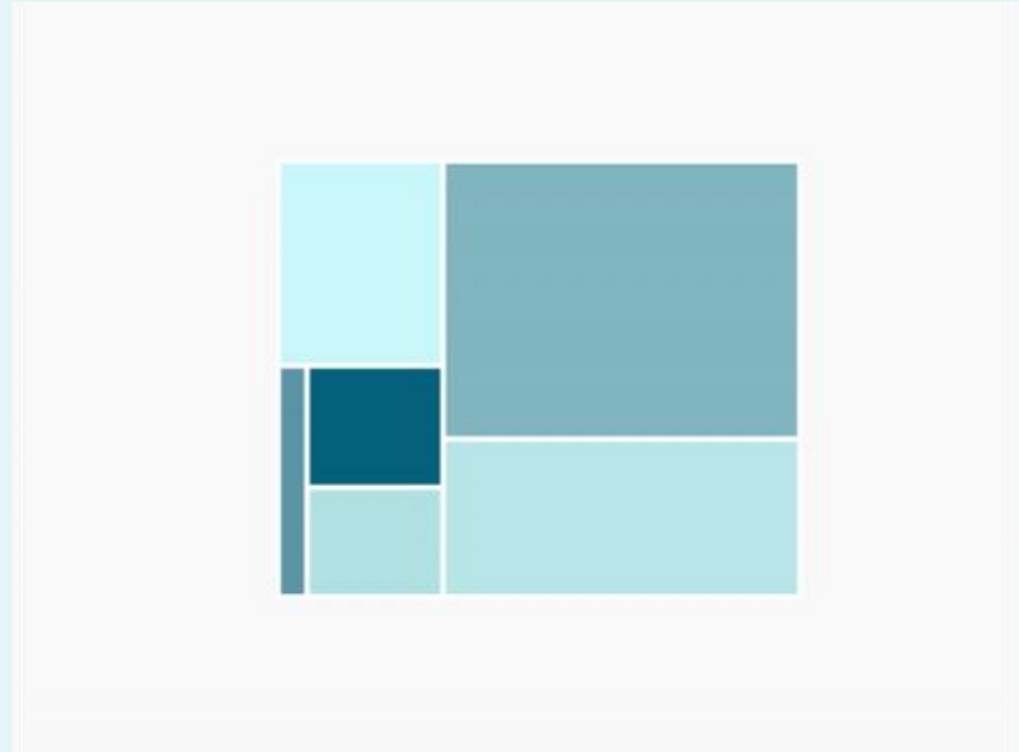
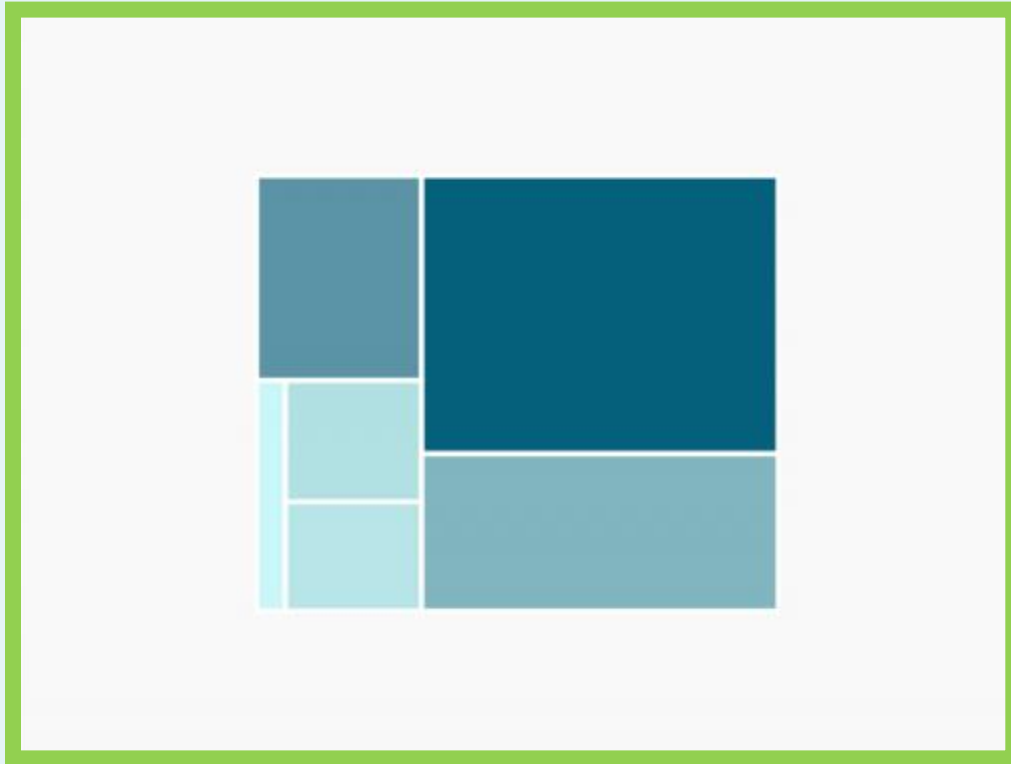
Voorbeeld 1



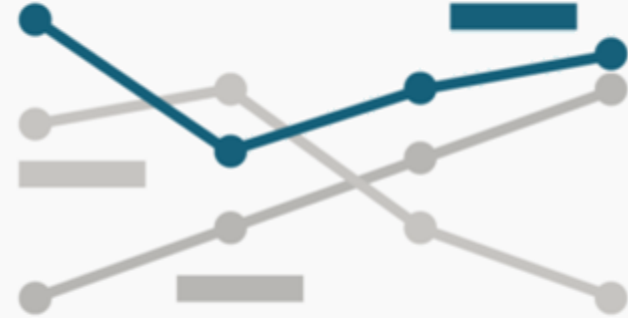
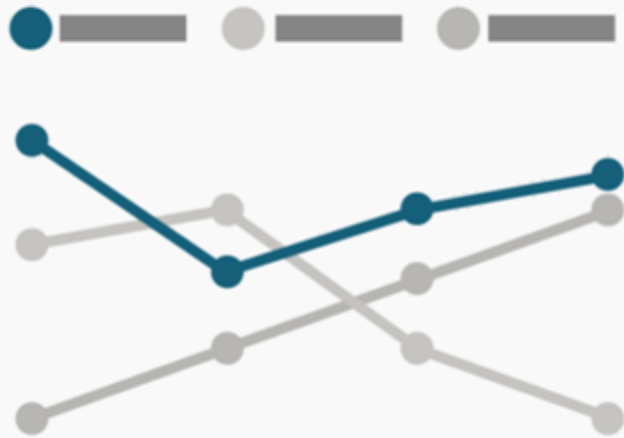
Voorbeeld 2



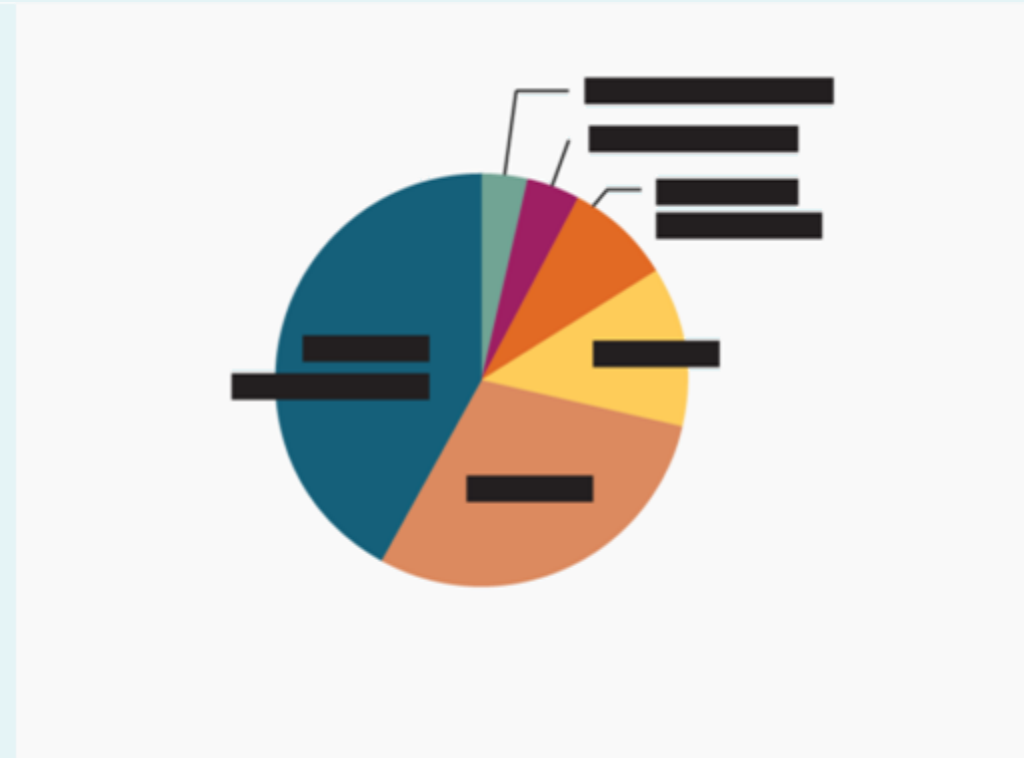
Voorbeeld 3



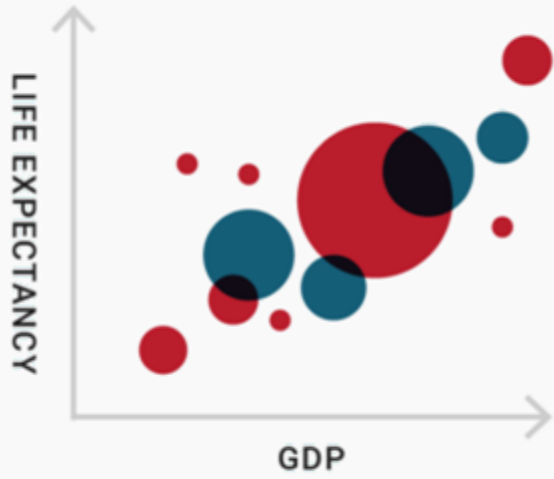
Voorbeeld 4



Voorbeeld 5



Voorbeeld 6





WERKNEEMERS



ACTIVITEIT



PENSIOEN



EXTRA PENSIOEN



BELEGGEN

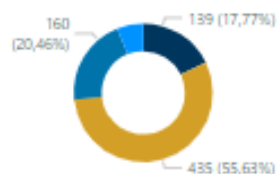


GEREEDHEID

AANTAL LOGINS

782

Totaal aantal unieke logins.



AANTAL LOGINS AFGELOPEN 12 MAANDEN

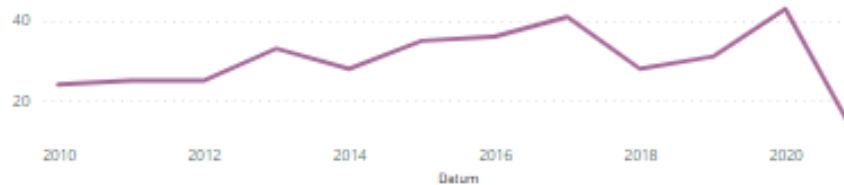
(Leeg)

Total unique count of logins in the past 12 months.

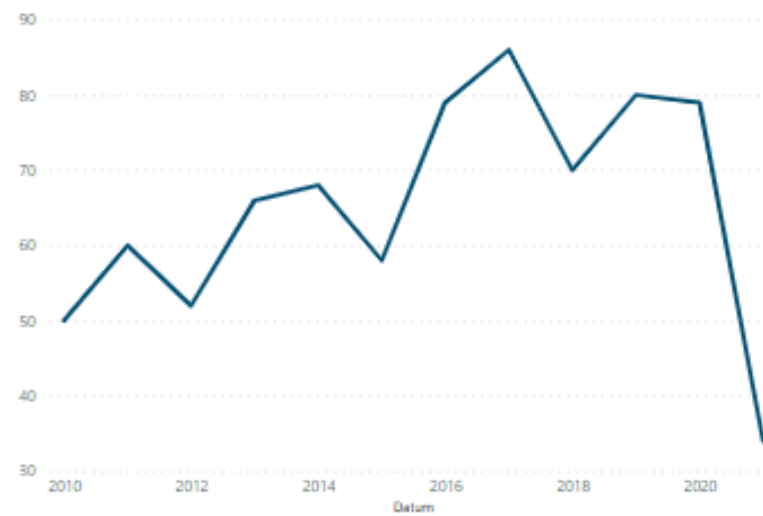
ACTIVITEIT



LAATST GEZIEN



LOGINS



Afsluiting – take aways

- Data & AI ontwikkelingen blijven razendsnel gaan
- Actuarissen ontkomen er niet aan om mee te bewegen

Zoek actief naar kansen, deze liggen met name in:

- benutten AI voor automatisering operationele taken
- adopteren AI technieken (nieuwe standaard)

Waak voor bedreigingen, let op:

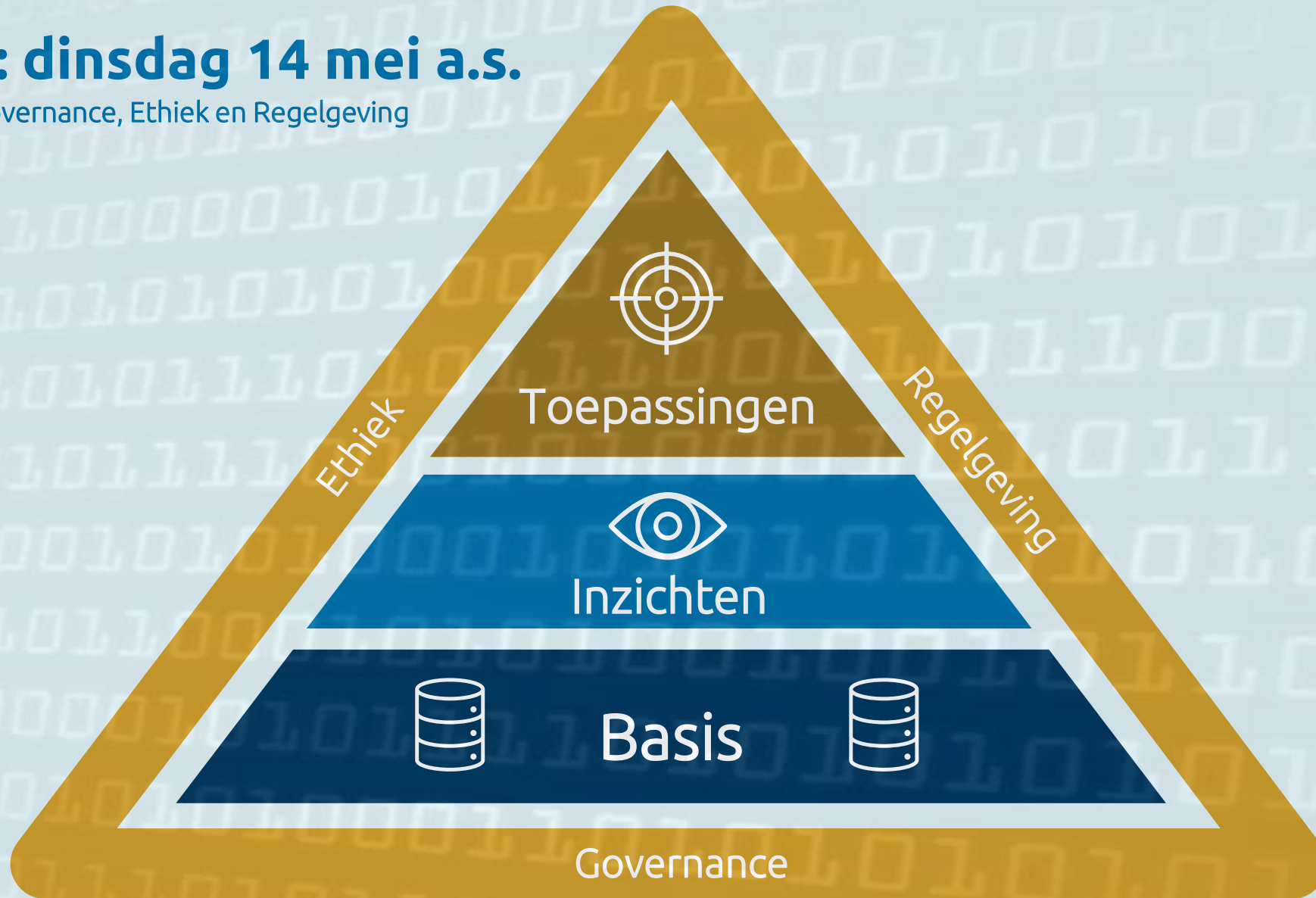
- valideren datakwaliteit en bronnen (wordt nóg belangrijker)
- grip houden op AI toepassingen
- oppakken controlerende 'data detective'/VAR rol

Data in sport: soms werkt het



Sessie 2: dinsdag 14 mei a.s.

AI/ML, Pricing, Governance, Ethiek en Regelgeving



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



Sessie 1



Sessie 2





Bedankt voor uw aandacht!

Bezoekadres

Hoogoorddreef 54

1101 BE Amsterdam Zuidoost

www.aaa-riskfinance.nl

info@aaa-riskfinance.nl