



Remain Relevant!



WAOP congres

*Datakwaliteit: het fundament voor
invaren en de toekomst*

AethiQs – Aron Jeurinck

PGGM – Jeroen Hermsen

19 juni 2024





Even voorstellen



Achtergrond

- Econometrie, actuariële wetenschappen en data science
- Senior Management consultant bij AethiQs
- Business analytics binnen het thema Data & AI
- Begeleiding van diverse OPF en BPF fondsen bij het uitvoeren van de werkzaamheden van het Kader Datakwaliteit, incl. opmaken AUP dossier en DNB.



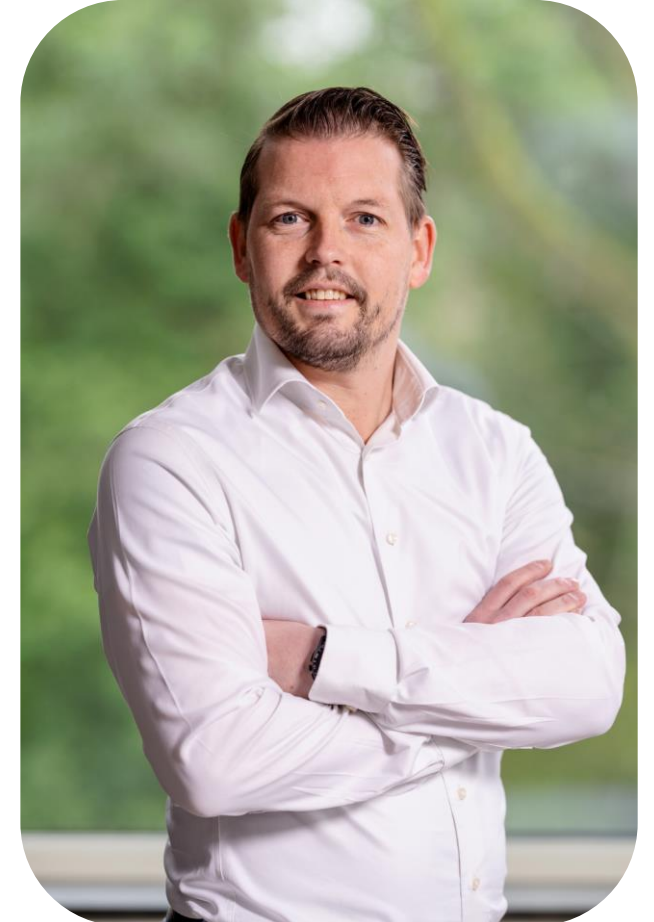


Even voorstellen



Achtergrond

- Accountancy, Financial Accounting, Business Control, Controle modellen en Data Science
- Manager Business Control – PGGM
- Lid werkgroep – Raad voor de Jaarverslaggeving (RJ 610) Pensioenfondsen
- Schrijfgroep – Pensioenfederatie voor Kader Datakwaliteit
- Schrijfgroep - Pensioenfederatie voor Toetsmoment 2
- Verantwoordelijk voor de implementatie van Kader Datakwaliteit bij PGGM voor alle klanten en het AUP dossier t.b.v. Fase 5.





Vandaag

- 1. Vraag maar raak: u bepaalt de agenda**
- 2. Praktijk voorbeelden**
- 3. Inspiratie huidige en toekomstige ontwikkelingen**



Over welk onderwerp wilt u meer weten vandaag?



Was datakwaliteit altijd al top of mind bij u of heeft dit hogere prioriteit gekregen door Wtp?



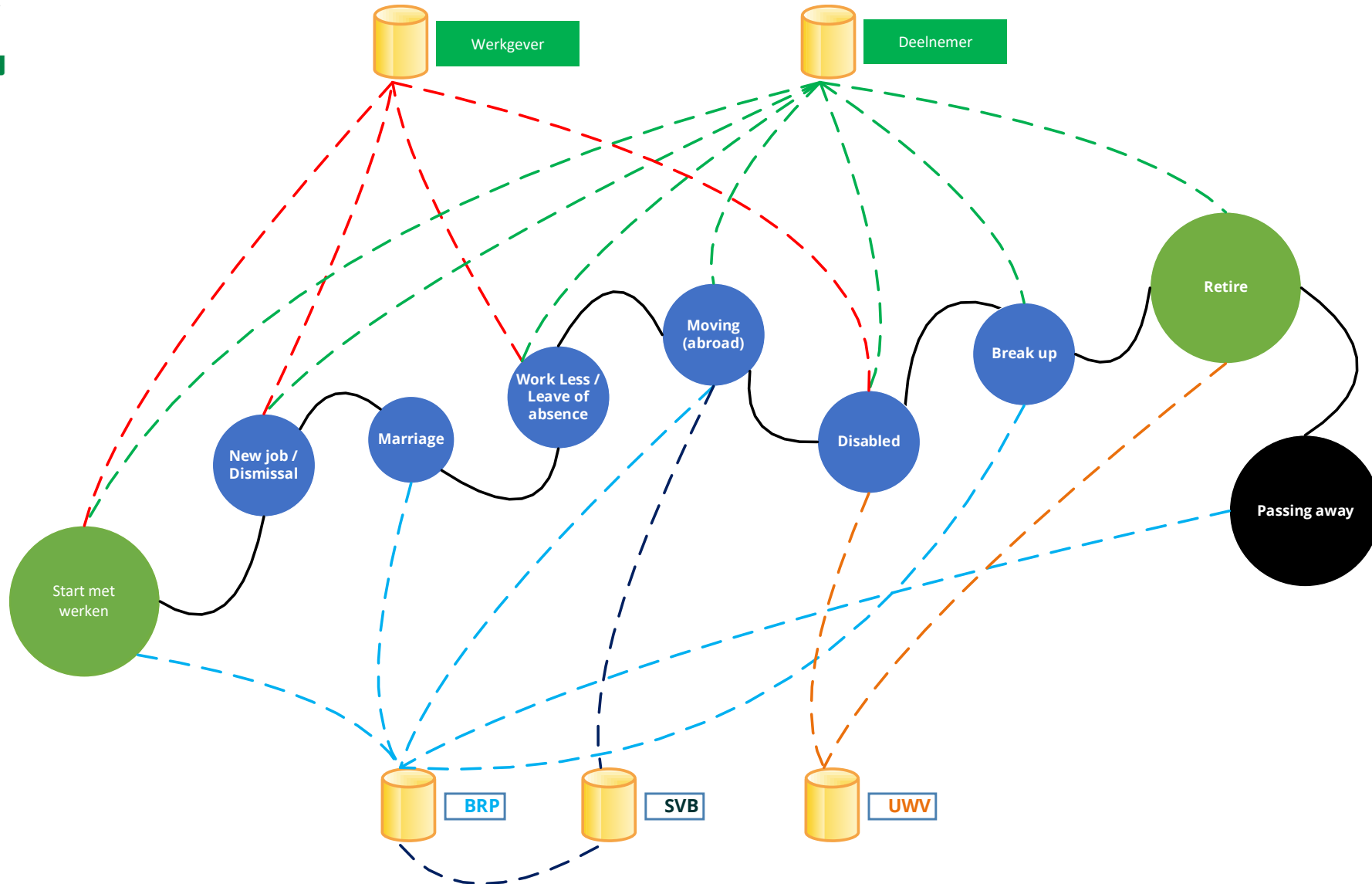
In welke fase van het Kader Datakwaliteit bevindt u zich?



Wat is op dit moment uw grootste datakwaliteit risico?



Het toepassen van **Kader** **Datakwaliteit** – met praktijk- voorbeelden



Duiding

- Gedreven vanuit levensgebeurtenissen van deelnemers
- Afhankelijkheid met externe bronnen, zoals de Basis Registratie Persoonsgegevens (BRP), Sociale Verzekeringsbank (SVB) en het UWV.

Wat is de rol van data en hoe belangrijk is beheersing om de kwaliteit te bevorderen?

Kader Datakwaliteit – volgordelijkheid



Bestuur stelt de kaders vast
(risicobereidheid,
correctie/herzieningenbeleid,
kritische data-elementen,
maximale toegestane afwijking)



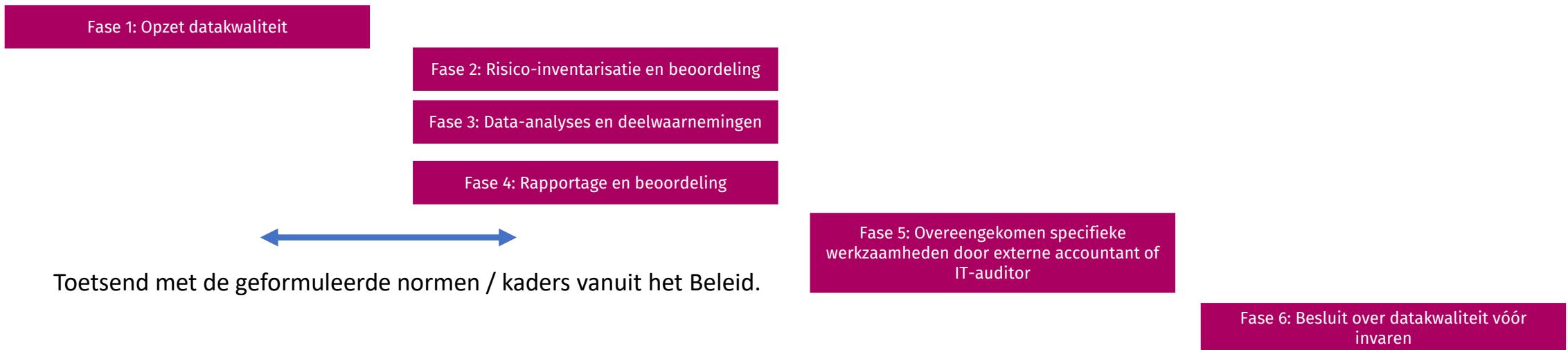
Uitvoering risico-
analyses (inventarisatie
en beoordeling) en
uitvoeren van de
geïdentificeerde
aanvullende
werkzaamheden



Controle op proces via Agreed
Upon Procedures.
(door externe accountant of IT-
auditor)



Bestuur stelt
datakwaliteit vóór
invaren vast.



Fase 1: Opzet datakwaliteit

Fase 1 Opzet Datakwaliteit

- Datakwaliteitsbeleid
- KDE's relevant voor invaren

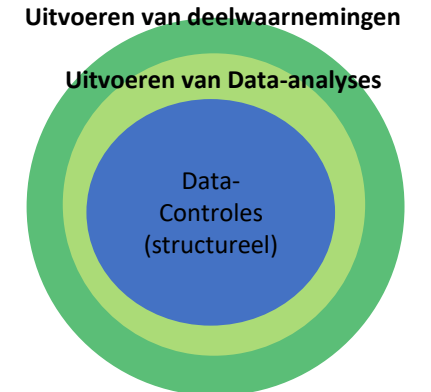
Identificatie van de Kritische Data Elementen die ten grondslag liggen aan de aanspraak-berekening en/of het opgebouwde kapitaal van deelnemers

Fase 2: Risico-inventarisatie en beoordeling

		Beheersing (De mate van beheersing is)				
		Zeer hoog	Hoog	Midden	Laag	Zeer laag
Risico hoogte (Het bruto risico is)	Zeer hoog	0	0	0	0	0
	Hoog	0	26	20	5	4
	Midden	1	8	20	15	1
	Laag	2	0	2	6	1
	Zeer laag	0	0	0	0	0

Identificeren van de beheersmaatregelen (IST) op de geïdentificeerde Kritische Data-elementen voor Invaren, met daarbij de koppeling naar de risicobereidheid van het fonds.

Fase 3: Data-analyses en deelwaarnemingen



Uitvoeren aanvullende werkzaamheden (get clean) en het implementeren van de structurele beheersmaatregelen (stay clean)

A blurred background image of a meeting table. Several documents are spread out, some with colorful sticky notes (pink, green, yellow) attached. A person's hand is visible on the left side, resting on the table. In the background, there are office supplies like a printer and a laptop.

Stelling:

Het vaststellen van de risicobereidheid van het fonds is eenvoudig en goed uitlegbaar

A blurred background image of a meeting table. Several documents are spread out, some with colorful sticky notes (pink, green, yellow) attached. A person's hand is visible on the left side, resting on the table. In the background, there are office supplies like a printer and a laptop.

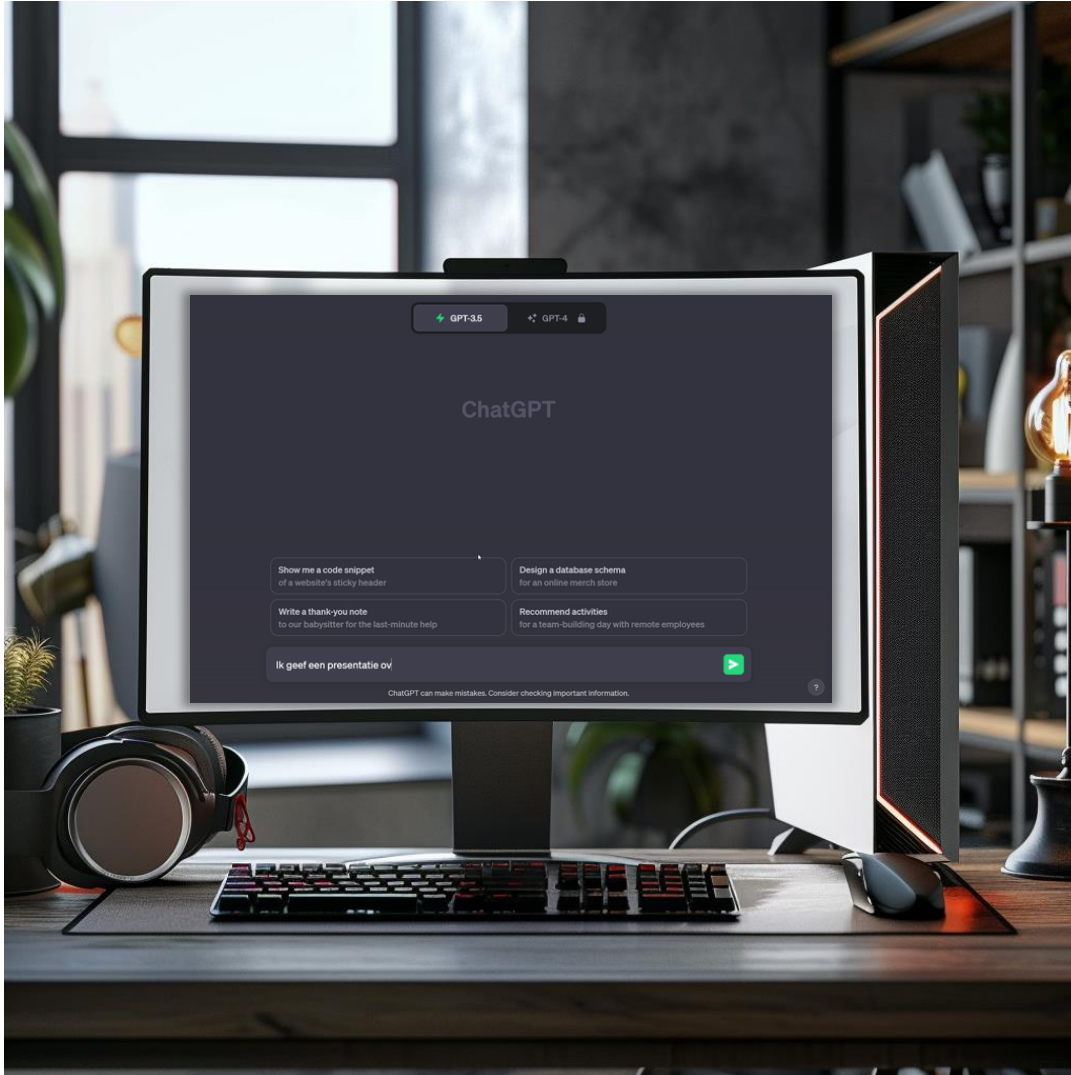
Stelling:

Onderdelen uit het Kader Datakwaliteit zijn herbruikbaar voor continue monitoring op datakwaliteit in NPS



Inspiratie

De ontwikkeling van data en datakwaliteit



De techniek

- **NLP: Natural Language Processing**
- **LLM: Large Language Models**
- **GPTs: Generative Pre-trained Transformers**

In Italië eet ik graag... **Pizza (90%)**
In Italië eet ik graag... **Friet (5%)**
In Italië eet ik graag... ... **(5%)**

NOS Nieuws • Zaterdag 1 juli, 20:46 • Aangepast zaterdag 1 juli, 22:12



Twitter beperkt aantal tweets dat je per dag mag lezen

Twitter heeft een limiet gesteld aan het aantal tweets dat gebruikers mogen lezen. Volgens Twitter-baas Elon Musk is dat nodig om het platform goed te laten draaien en gaat het om "een tijdelijke noodmaatregel".

Volgens Musk is het zogenoemde *scraping* een groot probleem voor Twitter.

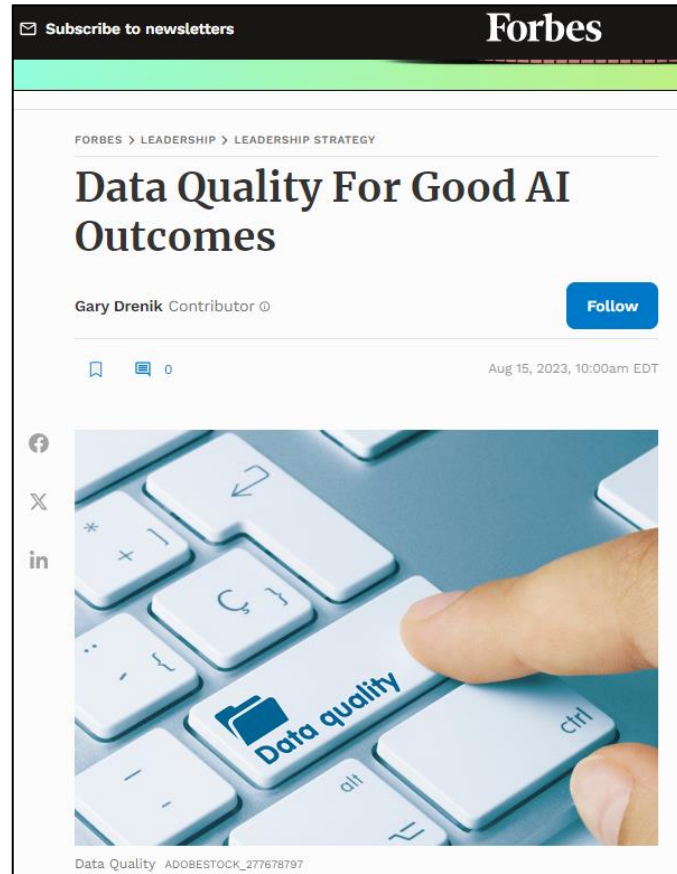
Hierbij gebruiken bedrijven software om geautomatiseerd data te vergaren op Twitter voor het ontwikkelen van modellen voor kunstmatige intelligentie. Dat kan er volgens de topman toe leiden dat Twitter trager of onbetrouwbaarder wordt.

Bronnen: NOS en FD

- Partijen die taalmodellen ontwikkelen, maken afspraken met organisaties die goede data beschikbaar hebben.
- Niet degene met het juiste model, maar degene met de juiste data heeft een goede concurrentiepositie.



Datakwaliteit is essentieel voor het juist functioneren van AI



Artificial Intelligence

Slechte data zijn als junkfood voor je AI

Onderzoek: 'Datakwaliteit een van de grootste uitdagingen voor adoptie AI'

Nieuws april 22, 2020

<https://informatieprofessional.nl/onderzoek-datakwaliteit-een-van-de-grootste-uitdagingen-voor-adoptie-ai/>
<https://www.salesforce.com/nl/blog/data-centric-ai/>
<https://www.forbes.com/sites/garydrenik/2023/08/15/data-quality-for-good-ai-outcomes/?sh=6ab87a441184>



4 sept 12:17

AI ontketent een enorm gevecht om data

The Economist

Bouwers van AI-modellen zijn op zoek naar nieuwe, betere bronnen om hun razende honger naar data te stillen. Ondertussen bekijken bedrijven die over enorme hoeveelheden data beschikken hoe ze die het best te gelde kunnen maken.

Bronnen: FD

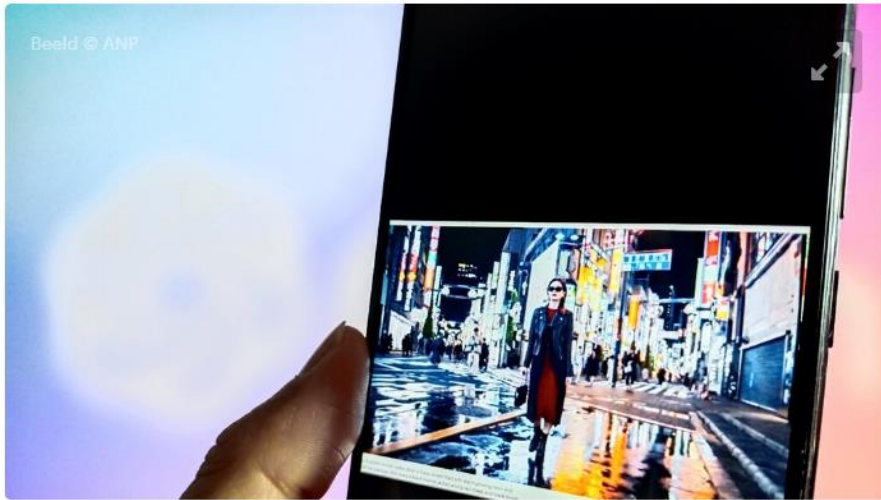
- Niet meer zelf modellen trainen, maar voortborduren op grote modellen van tech partijen (voornamelijk voor ongestructureerde data).
- De mogelijkheden lijken eindeloos, rekenkracht is aanwezig. Is er een tekort aan **goede** data?

Zorgen om AI van Sora: 'Golf aan desinformatie komt op ons af'



Door Harm Teunis

20 februari 2024 12:14 · Aangepast 20 februari 2024 12:14



Een stijlvolle vrouw die wandelt door Tokio, terwijl neonlicht weerspiegeld wordt door plassen op de grond. Het is een van de video's die ChatGPT-maker OpenAI online heeft gezet. De video is gemaakt door kunstmatige intelligentie met een nieuw programma dat Sora heet. Knap, zeker. Maar niet zonder gevaar, waarschuwen experts. "Het was eerst best ingewikkeld om te maken. Maar de rem die er was, is er nu af."



Net bi

13:25 DN
hoe

13:22 Aja
ach
Lea

13:16 Bos
neg

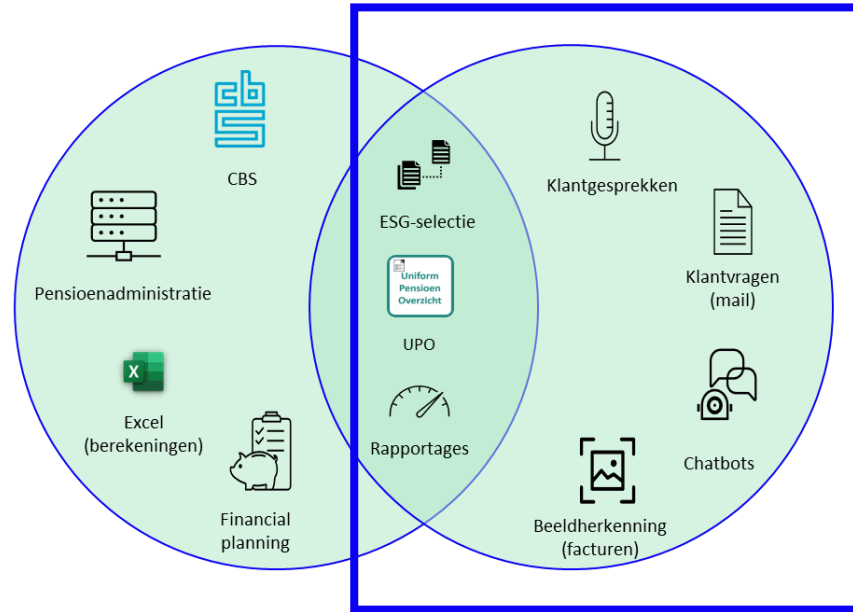
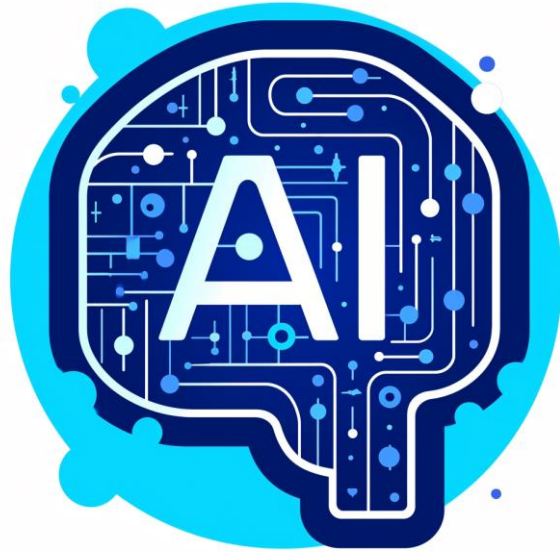
12:59 Mir
Ord
doe

12:57 Arn
sto
'be

- AI leidt tot mogelijke toename van fake news
- In de definitie van datakwaliteit worden andere dimensies relevant:
 - **Credibility:** De mate waarin de data eigenschappen heeft die als waar en geloofwaardig worden beschouwd in een specifieke context.
 - **Lineage:** Maatstaf of er feitelijke informatie bestaat over de bron van de data, hoe het bewerkt is en waar het voor gebruikt is.

<https://www.dama-nl.org/wp-content/uploads/2020/09/DDQ-Dimensions-of-Data-Quality-Research-Paper-version-1.2-d.d.-3-Sept-2020.pdf>

De ontwikkelingen vanuit AI hoek samengevat



Doorbraak in **AI: NLP, LLM and GPT's** and snelle ontwikkelingen op dit gebied

Veel **beschikbare ongestructureerde data** die eerst out of scope was

Datakwaliteit krijgt een bredere scope, zowel in type data als in datakwaliteit dimensies



***Maakt u al gebruik van data & AI
toepassingen?***

Welke mogelijkheden ziet u?



Stellingen



Stelling:

***Datakwaliteit draagt bij aan het vertrouwen in
Pensioenfondsen***



Stelling:


Datakwaliteit is het fundament voor het kunnen doen van een goede pensioenuitvoering (ook in de toekomst)



***Hoe wordt het bestuur bij uw fonds
meegenomen in het onderwerp datakwaliteit?***

In welke mate van detail?

En andere rollen/gremia?



In welke mate wordt u bij het datakwaliteit dossier ontzorgt door uw PUO?



Dank voor uw aandacht!



Jeroen Hermsen
jeroen.hermsen@pggm.nl

Aron Jeurinck
aron.jeurinck@aethiqs.nl

