

(Hoe) kan Farma AI vertrouwen?



Daan Di Scala



Simon van de Fliert

TNO innovation
for life



Trust**LLM**

NVFG
REGNED



Overzicht



- I. Uitdaging vertrouwen AI
- II. Wat doet TNO aan AI
 1. **TrustLLM:**
Taalmodel & Bijsluiterbot
 2. **Toegepast onderzoek:** Hybride AI
- III. Conclusie en take-aways

Interactief: Zou je AI vertrouwen...

- Assisteren met het schrijven van een e-mail?
- De weg te wijzen naar de NVFG Voorjaarsbijeenkomst?
- Een leuke grap, goede titel of grappige naam te verzinnen?
- Aanraders voor welke film vanavond te kijken op Netflix?
- Relatieadvies?
- Advies over wat je moet stemmen?
- Leren van nieuwe skills?
- Medisch advies?
- Hulp met juridische vragen?
- Assistentie met financiële kwesties?

**“High-Stakes
Beslisondersteuning”**



Uitdaging: AI is nuttig, maar problematisch...



New York lawyers sanctioned for using fake ChatGPT cases in legal brief

Airline held liable for its chatbot giving passenger bad advice - what this means for travellers

Bloomberg

AI-Powered World Health Chatbot Is Flubbing Some Answers



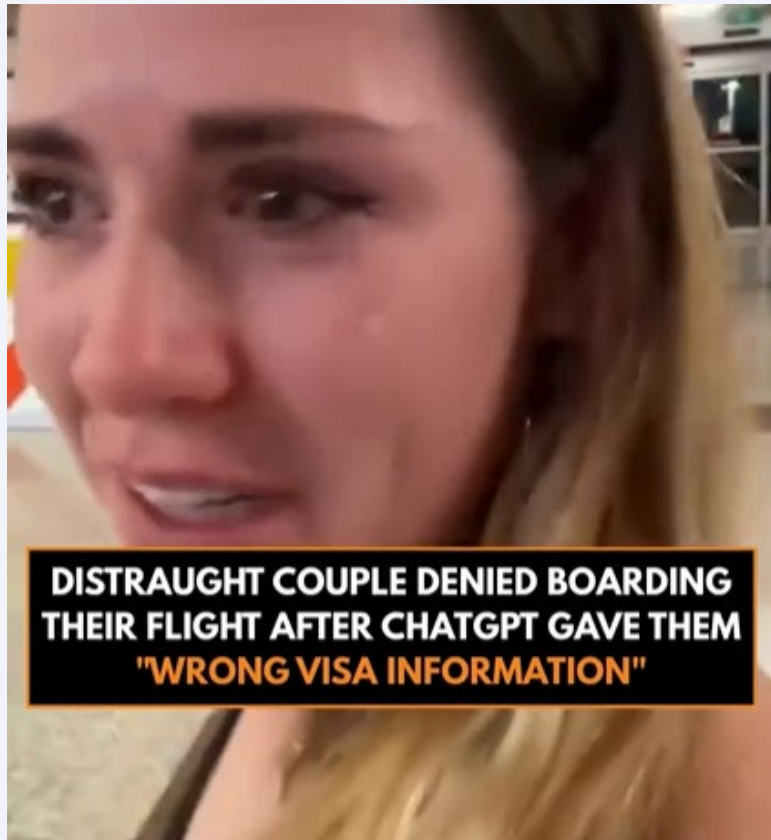
CTRL-ALT-DELOITTE

Consultants Forced to Pay Money Back After Getting Caught Using AI for Expensive "Report"

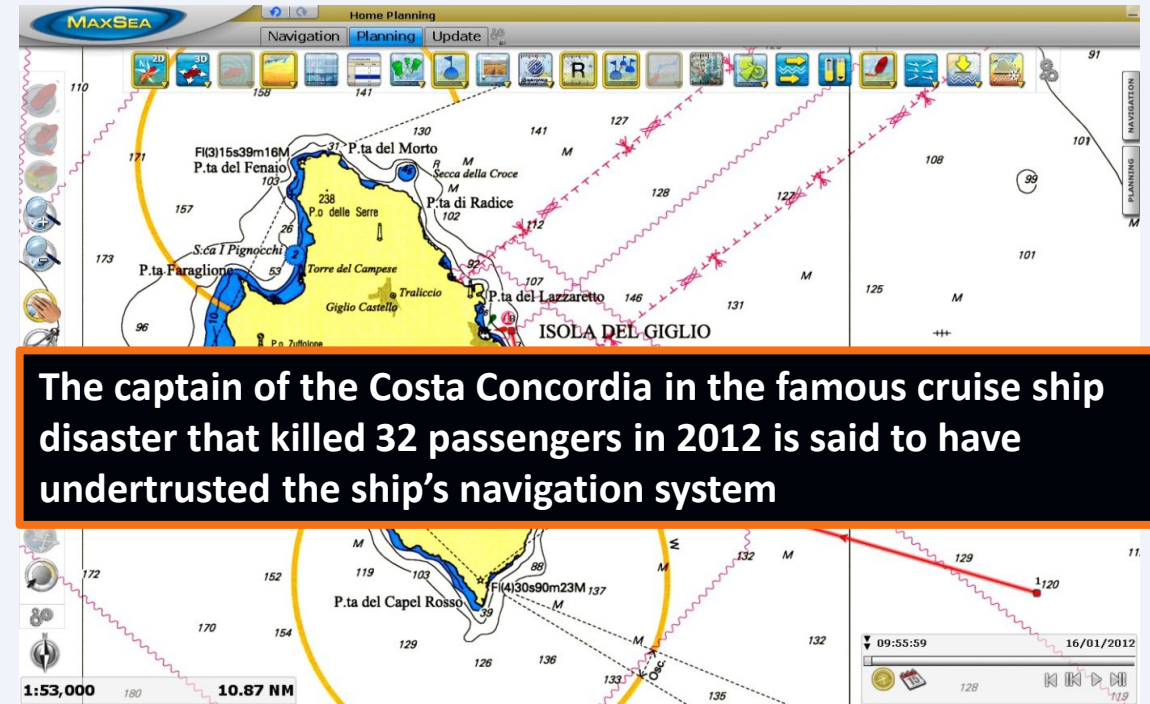
AI-gebruik in balans



Te veel vs. te weinig vertrouwen in AI:

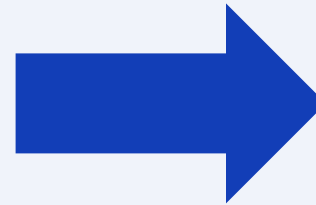
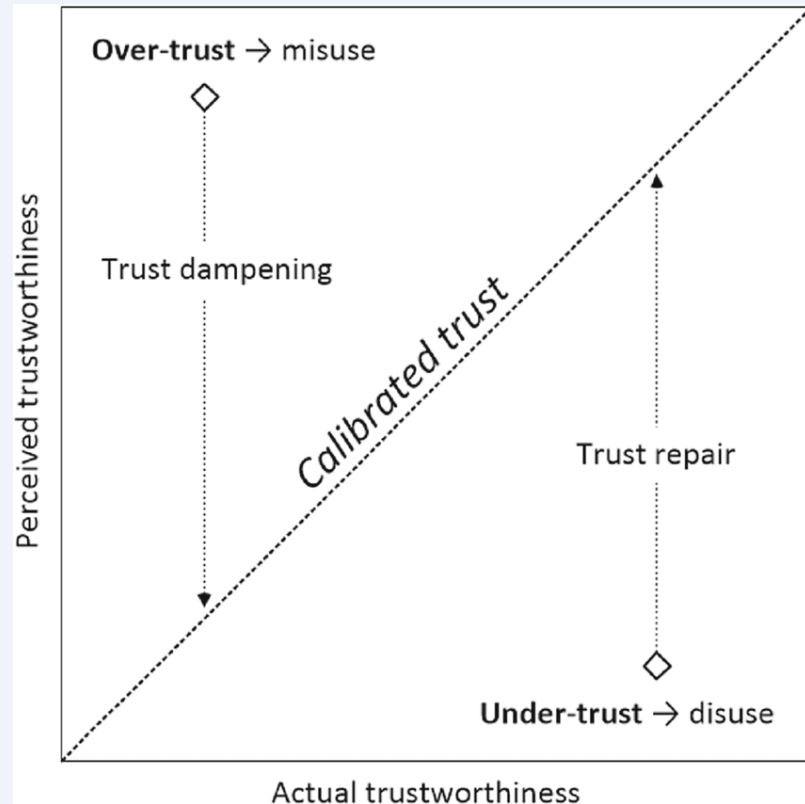


Overtrusting AI vermijden
→ kan leiden tot overreliance



Undertrusting AI vermijden
→ kan leiden tot dis-/misuse

De uitdaging: precies genoeg vertrouwen → precies genoeg gebruiken



Menselijke regie & Controle (Transparantie)
Kritisch & Kwalitatief
Verantwoord gebruik
Effectief (Impact)

Maar, hoe komen we daar? Drie W's



Werken aan modellen & systemen
(functioneel)



Wetten, normen & afspraken
(normatief)



Waakzaamheid, kritiek denken & AI-geletterdheid
(educatie)

Maar, hoe komen we daar? Drie W's



Werken aan modellen & systemen
(functioneel)



Wetten, normen & afspraken
(normatief)



Waakzaamheid, kritiek denken & AI-geletterdheid
(educatie)



Focus van vandaag

TNO Data Science Department werkt aan...:

Transparency of AI

Communications
between AI and
humans

Hybrid AI

Fairness / Bias

Sensitive and
synthetic data

Search and Explain
(RAG)

Agents

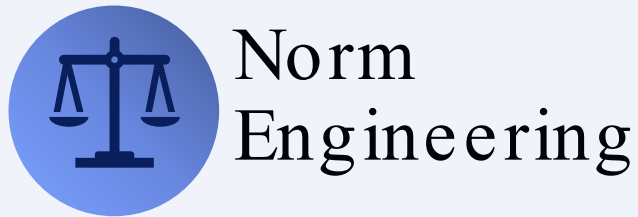
Semantics and
Interoperability

LLMs

Trustworthy AI

Use Cases

TNO AI Projecten



HEDGE-IoT



DEFRAUDify

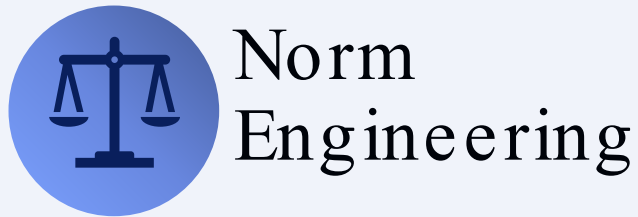
STARLIGHT

Trust**LLM**



MediSpeech

TNO AI Projecten



HEDGE-IoT



DEFRAUDify

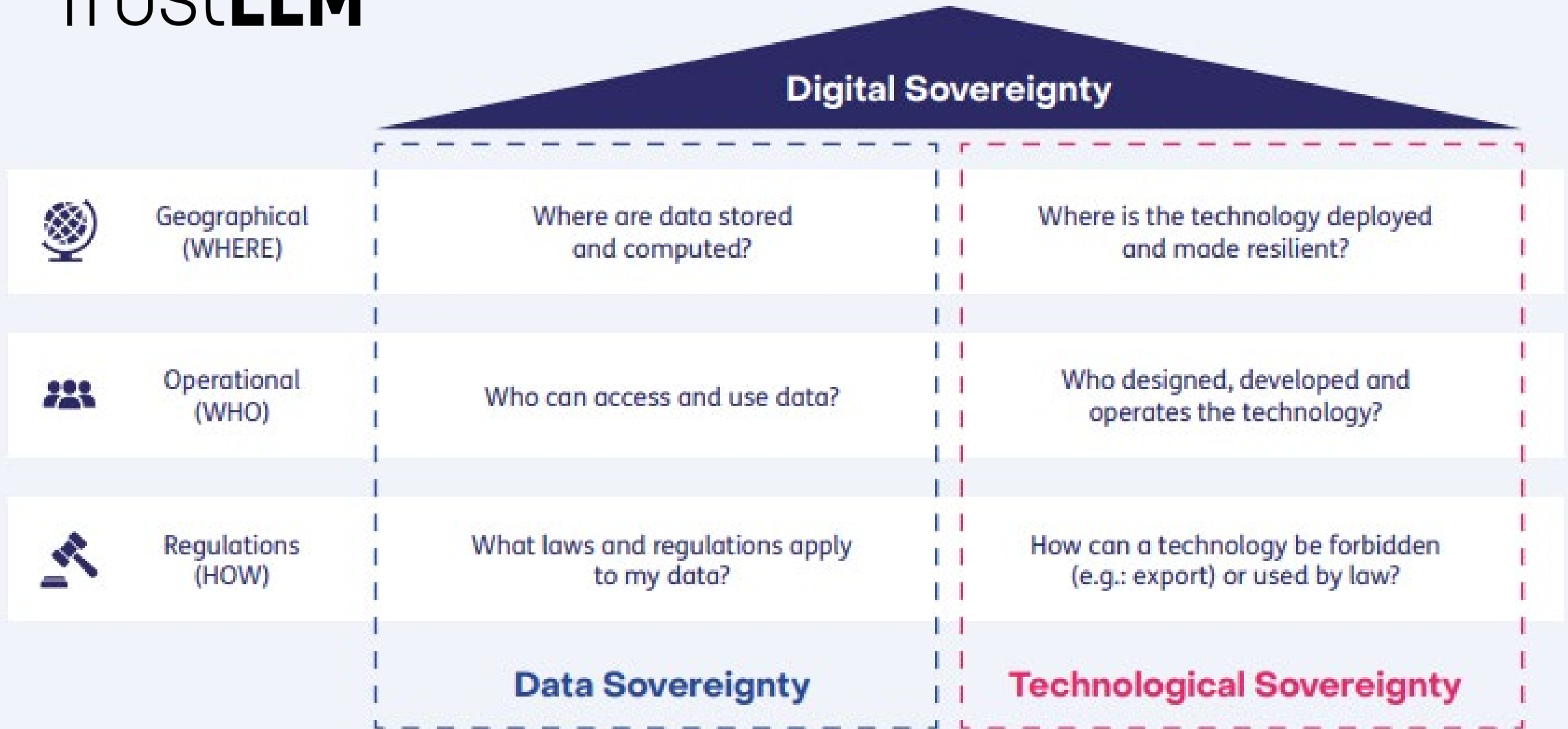
STARLIGHT

TrustLLM



MediSpeech

TrustLLM



Toegang tot data



United States v. Microsoft Corp., 584 U.S. ___ (2018)

CLOUD Act

DEEPSEEK UNMASKED:
EXPOSING THE CCP'S LATEST TOOL
FOR SPYING, STEALING, AND
SUBVERTING U.S. EXPORT CONTROL
RESTRICTIONS



NOS Nieuws • Dinsdag 13 augustus, 11:49

Stichting Brein haalt grote hoeveelheid illegale data voor trainen AI offline

"Impossible": OpenAI admits ChatGPT can't exist without pinching copyrighted work

AI • ANTHROPIC

Anthropic claims 3 Chinese companies ripped it off, using its AI tools to train their models: 'How the turn tables'

ChatGPT, Grok, Gemini and other AI chatbots are spewing Russian misinformation. study finds

Published on 18/06/2024 - 16:57 GMT+2



Jonathan Turley ✓
@JonathanTurley

...I learned that ChatGPT falsely reported on a claim of sexual harassment that was never made against me on a trip that never occurred while I was on a faculty where I never taught. ChatGPT relied on a cited Post article that was never written and quotes a statement that was never made by the newspaper.

[Post vertalen](#)

3:03 p.m. · 6 apr. 2023 · 91,3K Weergaven

Opkomst van digitale soevereiniteit

NOS Nieuws • Dinsdag 13 augustus 2024, 11:49

Stichting Brein haalt grote hoeveelheid illegale data voor trainen AI offline

PARESH DAVE BUSINESS MAY 31, 2023 7:00 AM

ChatGPT Is Cutting Non-English Languages Out of the AI Revolution

AI chatbots are less fluent in languages other than English, threatening to amplify existing bias in global commerce and innovation.

Chinese organisations launched 79 AI large language models since 2020, report says

Große KI-Modelle

FÜR DEUTSCHLAND

Machbarkeitsstudie 2023

LEAM:AI

KI BUNDESVERBAND



Why do we need a large GPT for Swedish?

What are the advantages of building a large language model for Swedish, and what should we look out for?



Magnus Sahlgren · Follow

Published in AI Sweden · 6 min read · Jul 14, 2022

◆ WSJ NEWS EXCLUSIVE

Europe to ChatGPT: Disclose Your Sources

Proposed legislation requires developers to list copyright material used in generative AI tools

Voorbeelden van EU Model Projecten

TrustLLM

GPT-NL

FACILITEIT VOOR EEN SOEVEREIN NEDERLANDS TAALMODEL

Why do we need a large GPT for Swedish?

What are the advantages of building a large language model for Swedish, and what should we look out for?



Magnus Sahlgren · Follow

Published in AI Sweden · 6 min read · Jul 14, 2022

A P E R T V S

EPFL

ETH zürich

CSCS



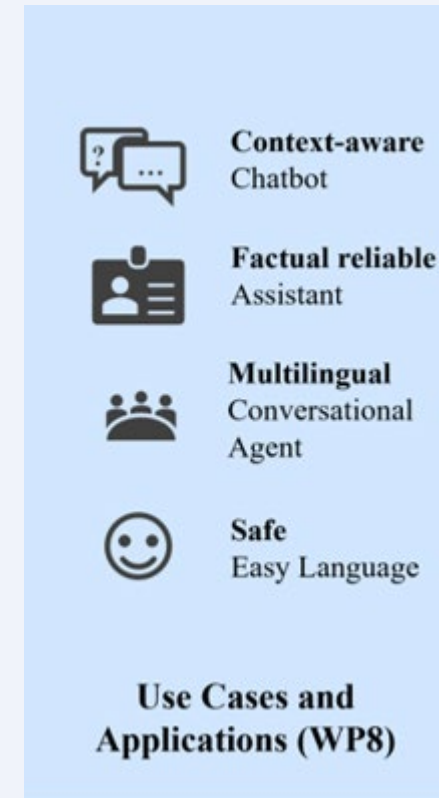
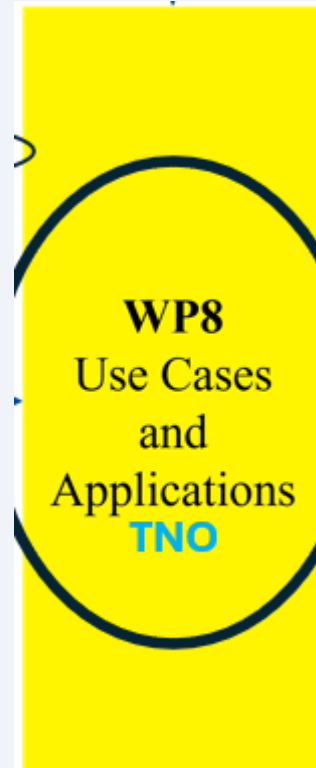
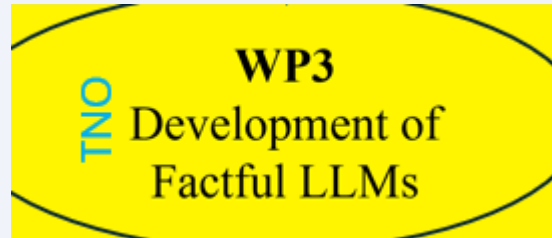
TrustLLM

- Het hoofddoel van TrustLLM is de ontwikkeling van een open, betrouwbaar en feitelijk LLM, met als eerste doelgroep de Germaanse talen.
- De focus op de Germaanse talen kan dienen als blauwdruk voor toekomstige activiteiten in andere taalfamilies.
- Men hoopt de bruikbaarheid van talen met weinig middelen/gegevens te vergroten door te leren van talen met meer middelen/gegevens.



Funded by
the European Union

TNO's focus op onderzoek en implementatie



Use case: Bijsluiterbot


- Medische bijsluiters bevatten cruciale informatie, maar patiënten hebben moeite om deze te lezen.
- Vooruitgang in AI verandert de manier waarop [gebruikers met informatie omgaan](#).
- Hoe kunnen LLM's patiënten ondersteunen bij het begrijpen van medische bijsluiters?
 - Zijn LLM's in staat om medische vragen te beantwoorden?
- Doel: een proof of concept maken dat het TrustLLM-model kan gebruiken om vragen over medische bijsluiters te beantwoorden.



Select a Model for your Chat

Please choose the underlying model you want to use. This selection will create a new session.

Start Chatting

 Hello! I can help you in understanding medical information related to your chosen leaflet. Choose a leaflet on the right and lets discuss its contents!

● Choose your leaflet

No details to show.

Select a medical leaflet to show the details here!

Send 



When do i have to replace the respimat medicine?

 Hide steps (2) ^

 **Model thinking...**

Okay, the user is asking when they need to replace the Respimat medicine. First, I should check the provided leaflet, h26896_smpc_en.pdf, for information on replacing Respimat. Respimat is a dry powder inhaler, so the instructions might be about when to change the device. I remember that inhalers sometimes have a dose counter or a specific number of doses. The user might be referring to replacing the medication in the Respimat, not the device itself. But sometimes people confuse the device with the medication. Wait, Respimat is a device that holds the medication, so replacing the medicine would mean refilling or changing the dose. But the leaflet might mention when to replace the medication based on usage or expiration. Alternatively, the user might be asking about replacing the inhaler device, like when it's no longer working. But the question says "replace the respimat medicine," so it's more about the medication inside. I need to check the leaflet for instructions on when to replace the medication, possibly related to the number of doses used or expiration date. Looking at the possible sections in the leaflet: "How to use" might have info on when to replace the medication. Also, "Storage" could mention expiration dates. The answer should guide the user to check the dose counter or expiration date as per the leaflet.

 **Tool: final_result**

Args: "{\\"question\\":\\"When do I have to replace the Respimat medicine?\\"}"

EVIDENCE

[c1] h26896_smpc_en.pdf **Highlight**

- □ When the dose indicator enters the red area of the scale the patient needs to get a new prescription; there is approximately medication for 7 days left (28 puffs).

[c2] h26896_smpc_en.pdf **Highlight**

- □ Once the dose indicator reaches the end of the red scale, BERODUAL RESPIMAT locks automatically - no more doses can be released.

1. NAME OF THE MEDICINAL PRODUCT

Berodual Respimat, oplossing voor inhalatie 20 microgram/ 50 microgram

2. QUALITATIVE AND QUANTITATIVE COMPOSITION

The delivered dose (the dose that leaves the mouthpiece of the Berodual Respimat) is 20 microgram ipratropium bromide monohydrate (equivalent to 19 microgram ipratropium bromide anhydrous) and 50 microgram fenoterol hydrobromide per puff.

Excipient with known effect: This medicine contains 1.12 microgram benzalkonium chloride in each actuation.

For the full list of excipients, see section 6.1.

3. PHARMACEUTICAL FORM

Inhalation solution Clear, colourless, inhalation solution

EVIDENCE

[c1] h26896_smpc_en.pdf

Highlight

- □ When the dose indicator enters the red area of the scale the patient needs to get a new prescription; there is approximately medication for 7 days left (28 puffs).

When to get a new Berodual Respimat

- □ Berodual Respimat inhaler contains 120 puffs (120 doses) if used as indicated.
 - □ When the dose indicator enters the red area of the scale the patient needs to get a new prescription; there is approximately medication for 7 days left (28 puffs).

I got a headache after taking respimat, but also was taking ozempic which made me nauseus

Hide steps (1) ^

Model thinking...

Okay, the user is asking about experiencing a headache after taking Respimat and also taking Ozempic, which caused nausea. Let me check the provided leaflet for Respimat, which is h26896_smpc_en.pdf. I need to see if there's any information about side effects related to headaches or interactions with Ozempic. First, I'll look for sections in the leaflet that mention side effects. Common sections are "Side effects" or "Adverse reactions." If there's a list of possible side effects, I should check if headache is listed. Also, I need to see if there's any mention of drug interactions, especially with Ozempic (which is semaglutide). If the leaflet doesn't specifically address interactions between Respimat and Ozempic, I should note that. Also, since the user is experiencing both side effects, it's possible that the headache is a side effect of Respimat, the nausea from Ozempic, or maybe a combination. But without specific info on interactions, I can't confirm that. I should advise the user to consult their healthcare provider if the symptoms persist or worsen, as they might need to adjust the medication or dosing. Also, mention that the leaflet doesn't provide details on this specific interaction, so professional medical advice is necessary.

The information provided in the Respimat (formoterol and montelukast) leaflet does not specifically mention interactions with Ozempic (semaglutide) or address headaches as a side effect. However, common side effects of Respimat may include headache, while Ozempic is known to cause nausea as a common adverse effect.

Since drug interactions or combined side effects are not detailed in the provided leaflet, it is important to consult your healthcare provider for personalized advice. They can assess whether the symptoms are related to either medication, an interaction, or another underlying factor.

For further guidance, refer to the full leaflet or consult a healthcare professional.

Review Side Effect

Validate your effects with medicine.

User Summary

The user asked about the replacement schedule for Respimat medicine and mentioned experiencing a headache after taking it. They also noted that they were taking Ozempic, which caused nausea.

Medicine

ozempic

respimat

+ Add Link

Side Effect

nauseus

headache

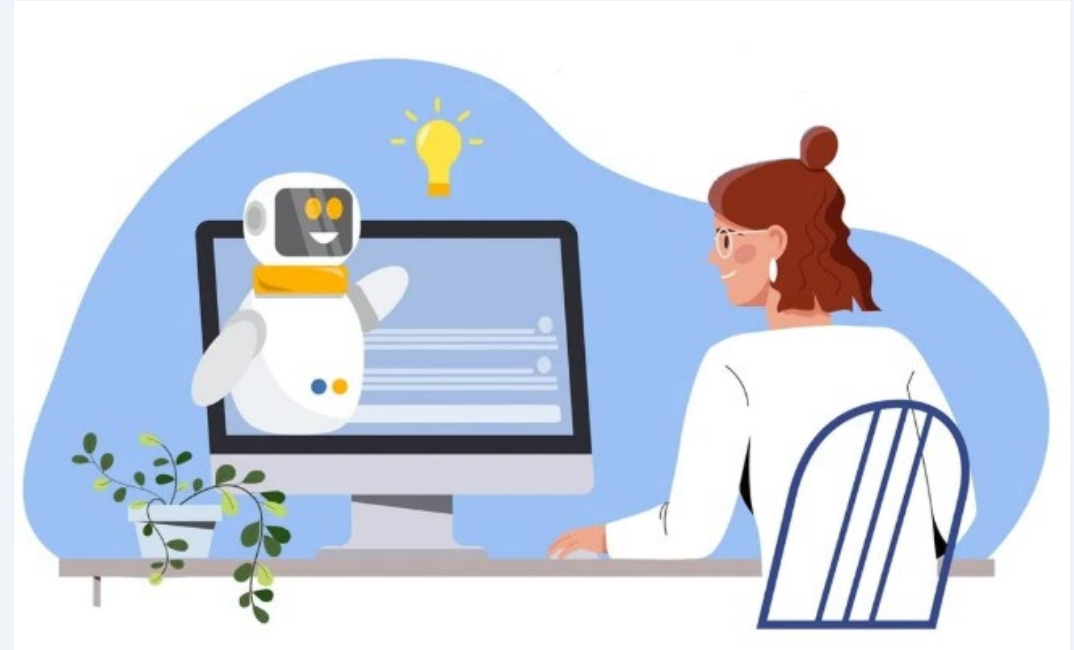


Cancel

Save

Bijsluiterbot

- Een mooie proof of concept, maar er is meer werk nodig om het naar productie te brengen.
- Er is meer testing nodig met echte gebruikers.
- Dit legt de basis voor toekomstig onderzoek naar het gebruik van LLM's in het medische domein.



Drie W's richting betrouwbare AI



Werken aan modellen & systemen
(functioneel)



Wetten, normen & afspraken
(normatief)



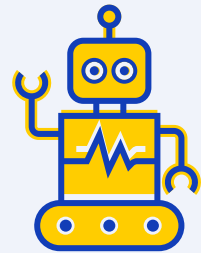
Waakzaamheid, kritiek denken & AI-geletterdheid
(educatie)



Twee aanpakken:

1

Eigen LLM
(+ data)

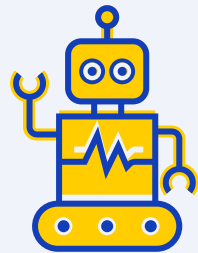


Betrouwbaar taalmodel



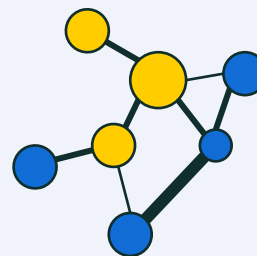
2

Hybride AI



Data AI systeem,
Chatbot

+



Kennismodel
(graaf / ontologie)



EU-FarmBook

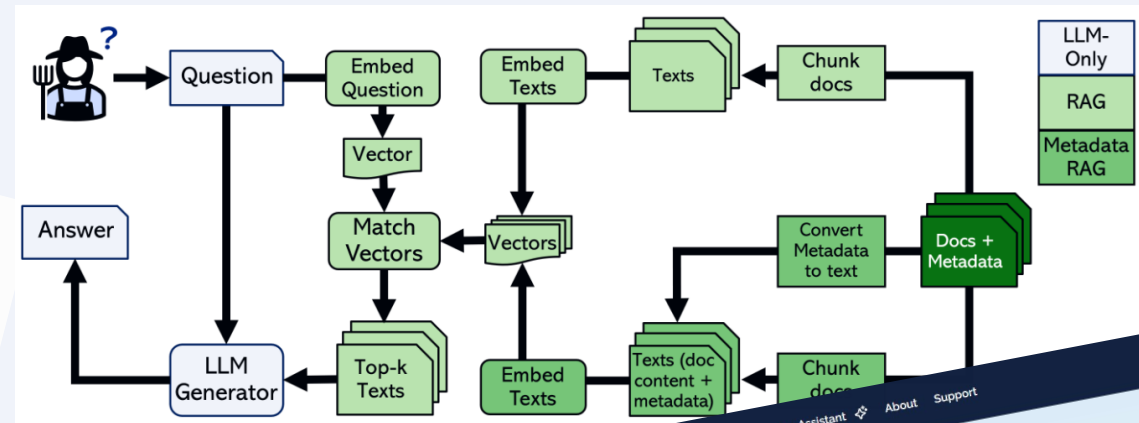
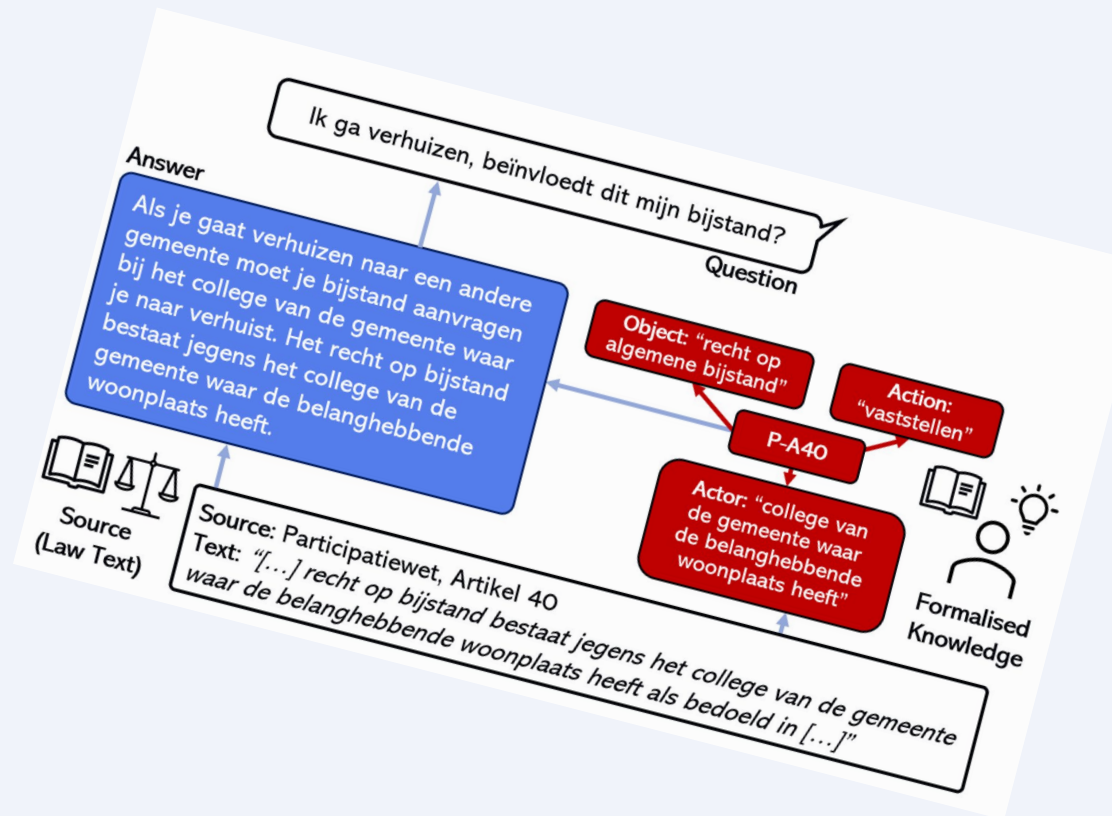


TrustLLM



Toegepast Onderzoek: ontwikkeling van Hybride AI assistenten

- Veel onderzoek over RAG: Retrieval Augmented Generation (“zoek en leg uit”)
- Binnen TNO DS werken we aan kennis graaf-gebaseerde RAG
- Verschillende vormen en maten (allerlei GraphRAG aanpakken, allerlei domeinen met voorkeuren)
- Wetenschappelijke focus op Betrouwbare AI → Hoe evalueren we de state-of-the-art methoden?
- Toepassingen in high-stakes domeinen (agrarisch advies, juridisch advies, ..., bijsluiterbot!)

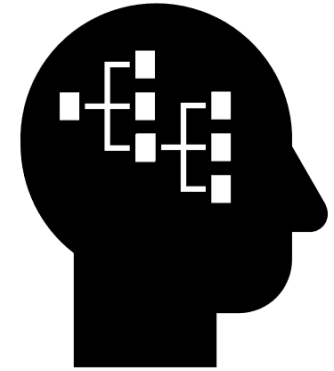
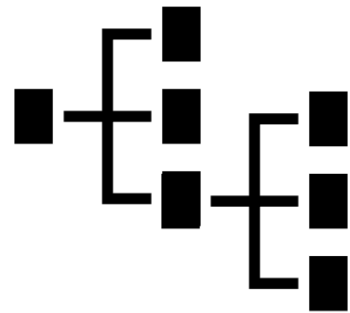
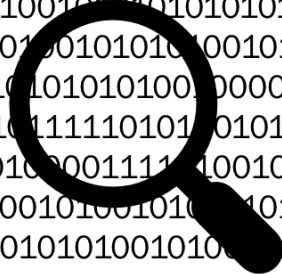


[1] QuALA-NL: Question & Answer with Legal Attribution in Dutch, Romy van Drie, Roos Bakker, Daan Di Scala, Maaïke de Boer, LREC 2026

[2] Harvesting Trust Indicators: A Method for Evaluating Trustworthiness of RAG-Driven Advice, Daan Di Scala, Maaïke de Boer, HHAI 2026

Hybride AI

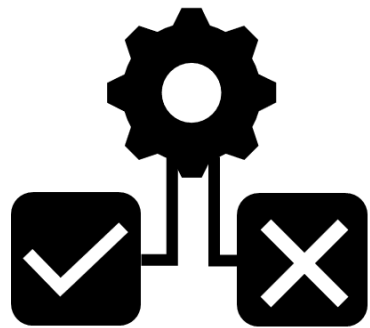
100101101100110100101
01010010101010101010
101010101010100101010
10101010101000000010
10101111101010101010
0101000111110010010
10100101001010101010
0100101010010101010
011110100101001010100



Data gedreven AI
grote taalmodellen (LLMs),
natuurlijke taalverwerking,
machine learning

Kennis gedreven AI
ontologieën,
kennis grafen
regels & redeneren

De Mens
als expert,
als controleur,
als beslisser

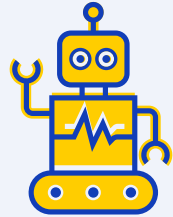


Het meest effectieve **hybride**
kunstmatige intelligentiesysteem
voor beslis ondersteuning

Een Hybride AI voorbeeld: EU-FarmBook project



I am looking for best practices
to water my crops!



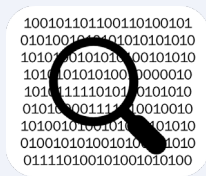
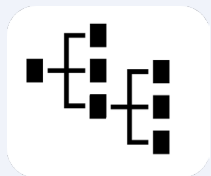
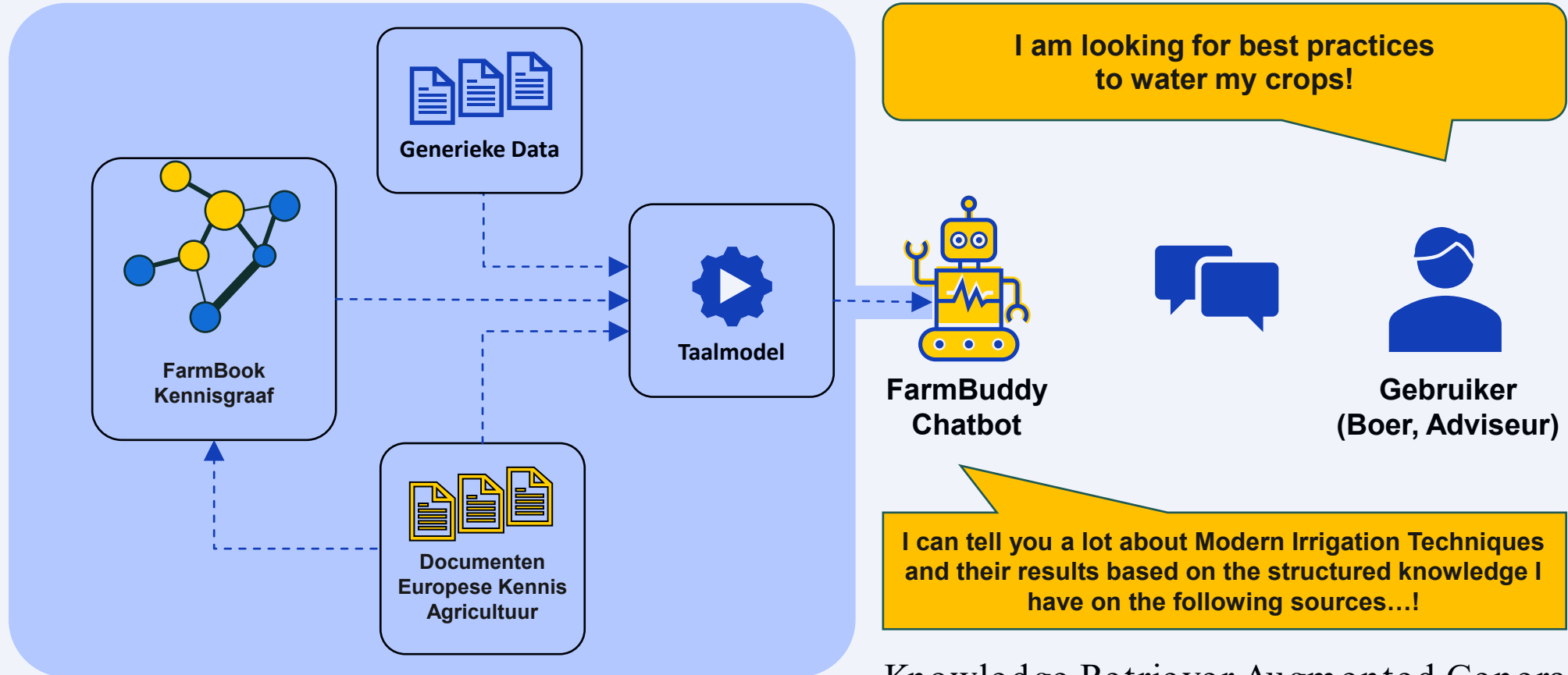
FarmBuddy
Chatbot



Gebruiker
(Boer, Adviseur)

I can tell you a lot about Modern Irrigation
Techniques and their results!

Een Hybride AI voorbeeld: EU-FarmBook project



Kennis gedreven + Data gedreven

Knowledge Retriever Augmented Generation (KRAG)

Toegespitste antwoorden op eigen data

Verhoogd inzicht in eigen data

Verbonden, gestructureerde kennis verhoogt betrouwbaarheid

Conclusie + Takeaways

Belangrijk om AI te vertrouwen op de juiste momenten voor de juiste taken

(zowel te weinig als te veel vertrouwen voorkomen)

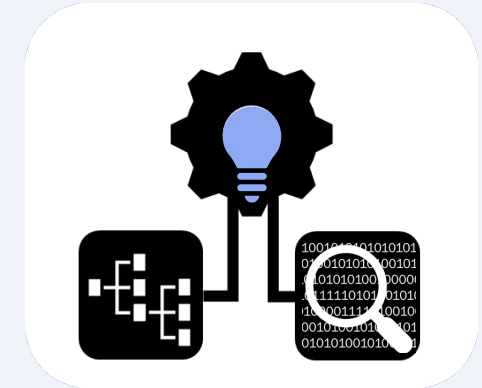
Bij TNO werken we mee aan:

- TrustLLM:
 - Een Europees taalmodel op gecureerde data
 - Use case: **Bijsluiterbot**, een AI assistent voor complexe vragen over bijsluiters
- Hybride AI, een koppeling tussen data, AI, gestructureerde kennis en mens

Takeaways:

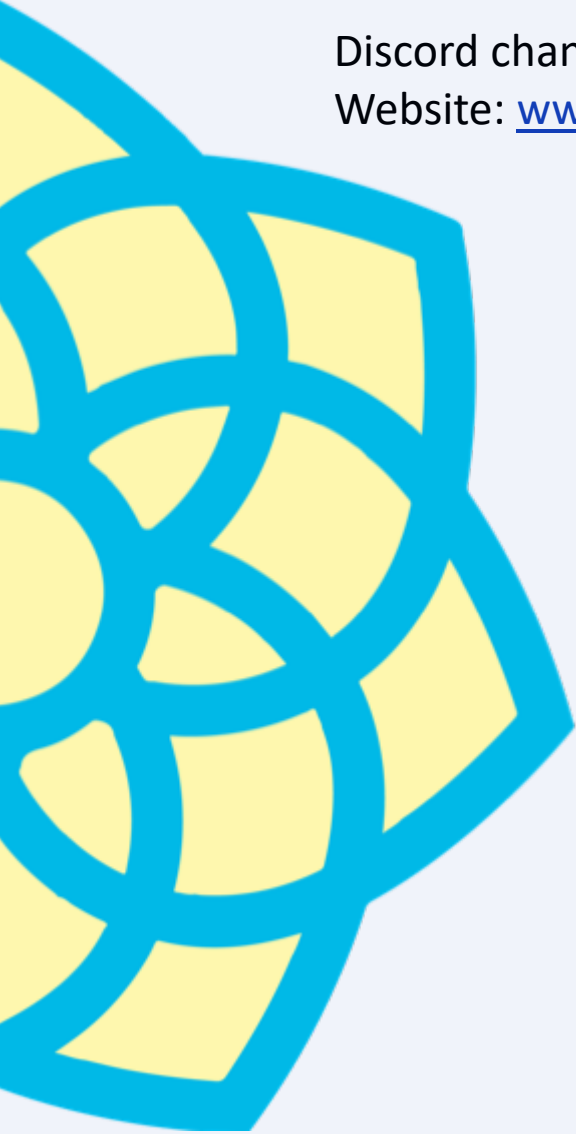
Een nauwkeurig, eerlijk en betrouwbaar systeem staat aan de basis van efficiënte ondersteuning van de professional

We **werken** aan betrouwbare AI-technieken (Trustworthy AI), er komen **wetten** die AI-gebruik gaan afbakenen, maar: Niet alles is wat het lijkt. Blijf **waakzaam** en kritisch!



More about the TrustLLM project

Discord channel (open to everyone) on the [AI Nordics server](#): #trust-llm-public
Website: www.trustLLM.eu



Funded by
the European Union

(Hoe) kunnen we AI vertrouwen?



Vragen?

Geïnteresseerd in ons werk?
Laten we in contact blijven:

Daan Di Scala, daan.discal@tno.nl

Simon van de Fliert, simon.vandefliert@tno.nl

Simon



Daan

