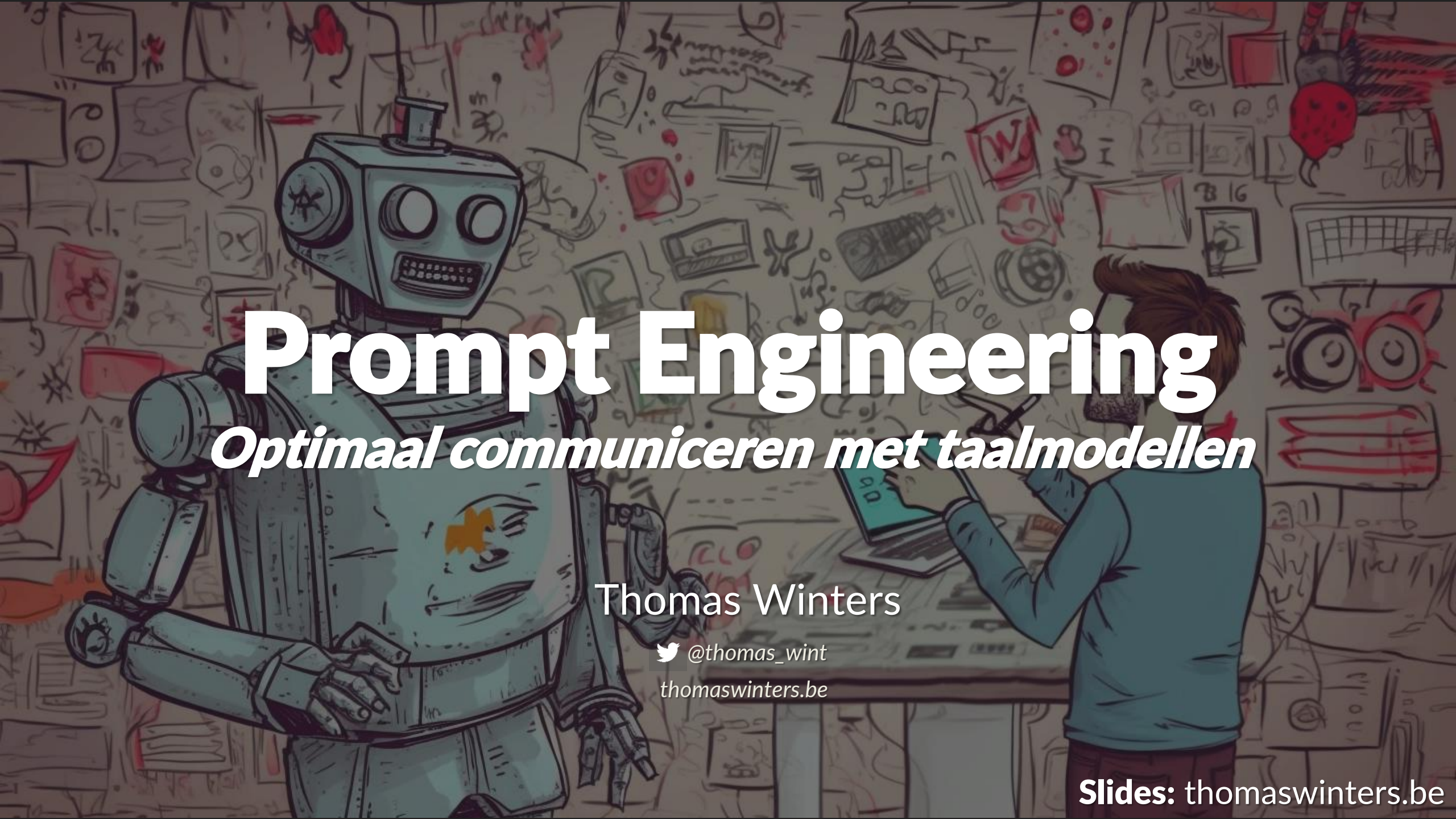


Prompt Engineering

Optimaal communiceren met taalmodellen

Thomas Winters

@thomas_wint

thomaswinters.be

Slides: thomaswinters.be

Wie heeft taalmodel (ChatGPT, Copilot, Gemini, ...) gebruikt?



Wie heeft taamodel voor iets nuttig gebruikt?



Wie had taak dat taalmodel niet kon?



Thomas Winters

Onderzoeker in creatieve AI & taalmodellen aan KU Leuven sinds 2018

RobBERT taalmodel
(state-of-the-art Nederlandstalige BERT taalmodel)

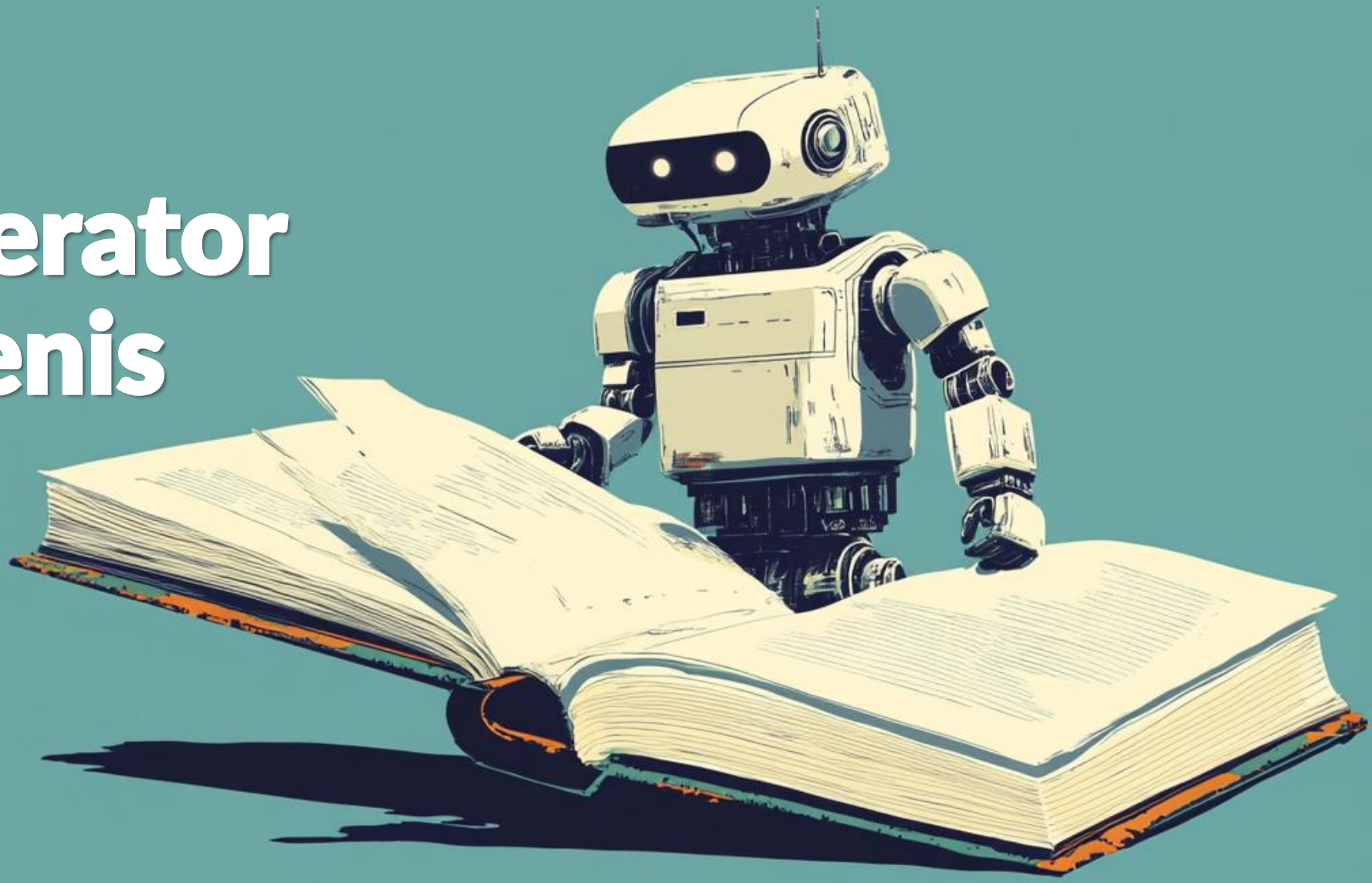
Werk met GPT-modellen sinds 2019
(o.a. voor humor, Improbotics...)

Twinters BV: creative AI workshops & consultancy

KU LEUVEN LEUVEN.AI DTAI



Tekst generator Geschiedenis



Hoe werkt een tekstgenerator?



1. Open smartphone **toetsenbord** op tekstveld
2. Druk op een voorgestelde **auto-complete** woord
3. Herhaal 10-tal keer
4. Je hebt een tekst gegenereerd die klinkt zoals jezelf!

Auto-complete telde frequentie van woorden na vorige woorden in je teksten

TorfsBot

1. Telde in alle tweets & columns van Rik Torfs hoe vaak woord andere woorden volgde



2. Dan neemt het paar willekeurige beginwoorden, en blijft het woorden die Rik gebruikte kiezen.





GPT training

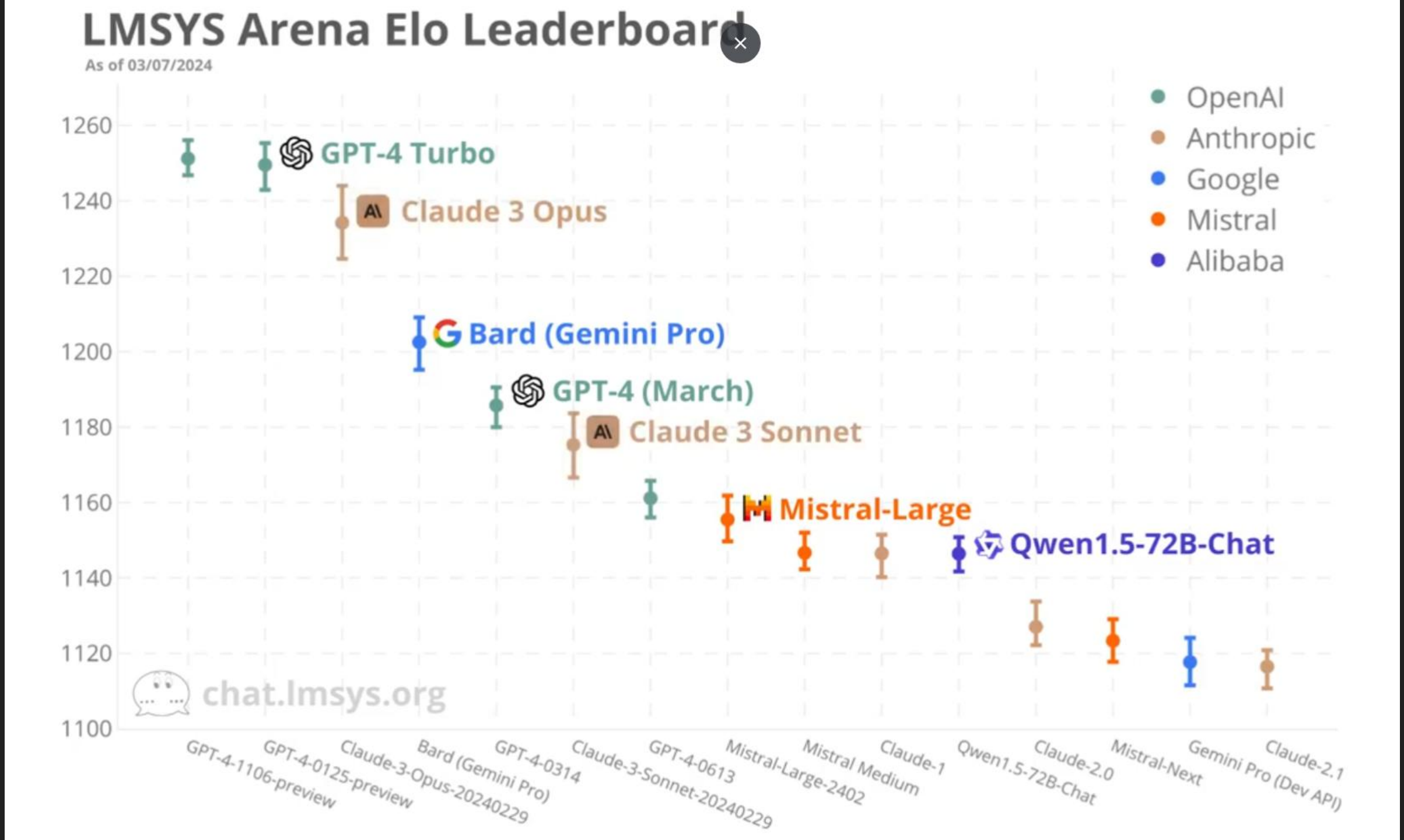


Heel het internet

Trainingsproces van paar miljoen $

GPT-4

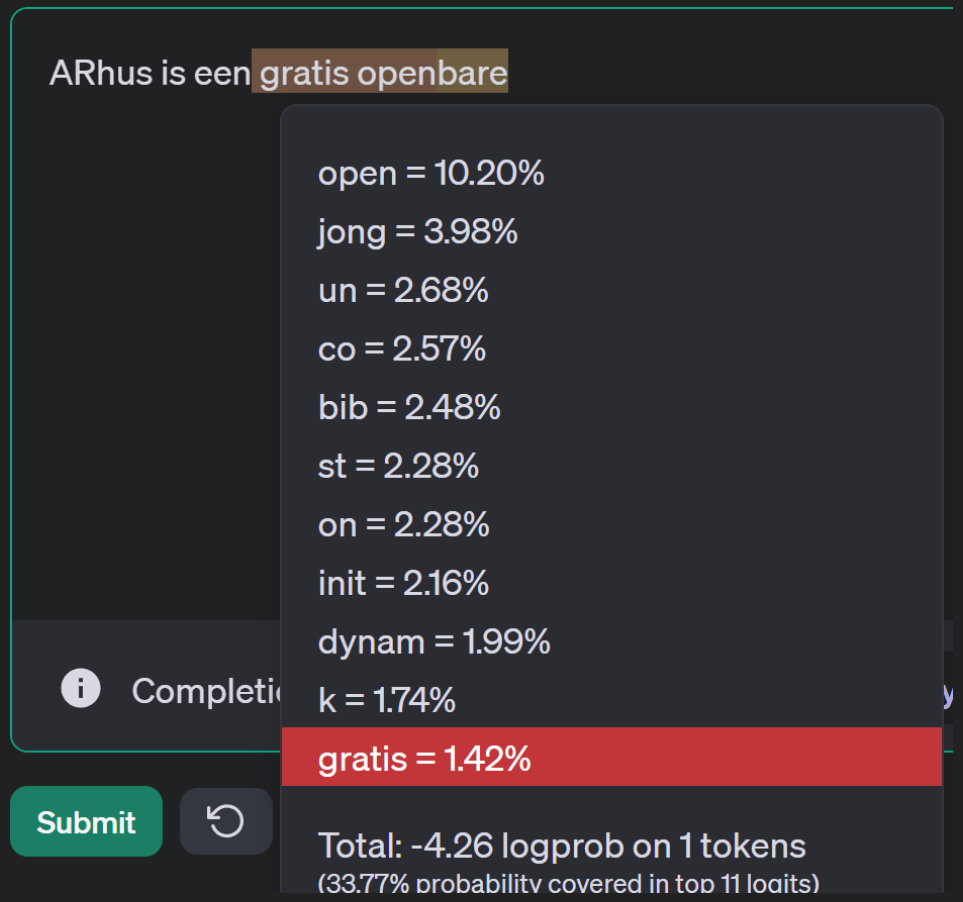
Beste publieke taalmodel


LMSYS Arena Elo Leaderboard
As of 03/07/2024
OpenAI
Anthropic
Google
Mistral
Alibaba
1260
1240
1220
1200
1180
1160
1140
1120
1100
GPT-4 Turbo
Claude 3 Opus
Bard (Gemini Pro)
GPT-4 (March)
Claude 3 Sonnet
Mistral-Large
Qwen1.5-72B-Chat
chat.lmsys.org
GPT-4-1106-preview
GPT-4-0125-preview
Claude-3-Opus-20240229
Bard (Gemini Pro)
GPT-4-0314
Claude-3-Sonnet-20240229
GPT-4-0613
Mistral-Large-2402
Mistral Medium
Claude-1
Qwen1.5-72B-Chat
Claude-2.0
Mistral-Next
Gemini Pro (Dev API)
Claude-2.1

GPT modellen



≈ autocomplete on steroids



Kans op volgende token gegeven *vorige tokens*

Tokens	Characters
122	566

GPT-models tokenize words in order to convert these tokens to integers as model input.

The tokens are chosen in such a way that common substrings have their own token. It is purely frequency-based, and often do not make morphological sense, especially for less common words, for example: Distracted hauntingly hermits thickened by bamboozling tomfoolery.

In English, common words are usually a single token.
In het Nederlands worden woorden vaak raar opgesplitst. Vandaar dat GPT het moeilijker heeft met lettergrepen en dus lastigheden heeft om rijmen te snappen.

Text | Token IDs

Tokens	Characters
122	566

```
[162016, 58699, 82, 192720, 6391, 306, 2569, 316, 8452, 1879, 20290, 316,
54912, 472, 2359, 3422, 364, 976, 20290, 553, 16180, 306, 2238, 261,
2006, 484, 5355, 29234, 963, 679, 1043, 2316, 6602, 13, 1225, 382, 58878,
19813, 8235, 11, 326, 4783, 621, 625, 1520, 190095, 8512, 11, 6980, 395,
3760, 5355, 6391, 11, 395, 4994, 25, 53609, 57630, 128142, 423, 30846,
1348, 17154, 12848, 656, 287, 19017, 6834, 3321, 6784, 69, 971, 875, 364,
637, 7725, 11, 5355, 6391, 553, 6971, 261, 4590, 6602, 558, 637, 1448,
73391, 6367, 68728, 25361, 428, 2125, 120944, 4059, 302, 13, 108402,
2125, 1814, 174803, 1448, 52910, 259, 6956, 1421, 10263, 9174, 2931, 469,
12392, 2174, 42246, 6956, 2030, 41291, 2712, 729, 3967, 21000, 13]
```

Text | Token IDs

https://platform.openai.com/tokenizer

Drie GPT trainingfasen

GPT-3 (2020)

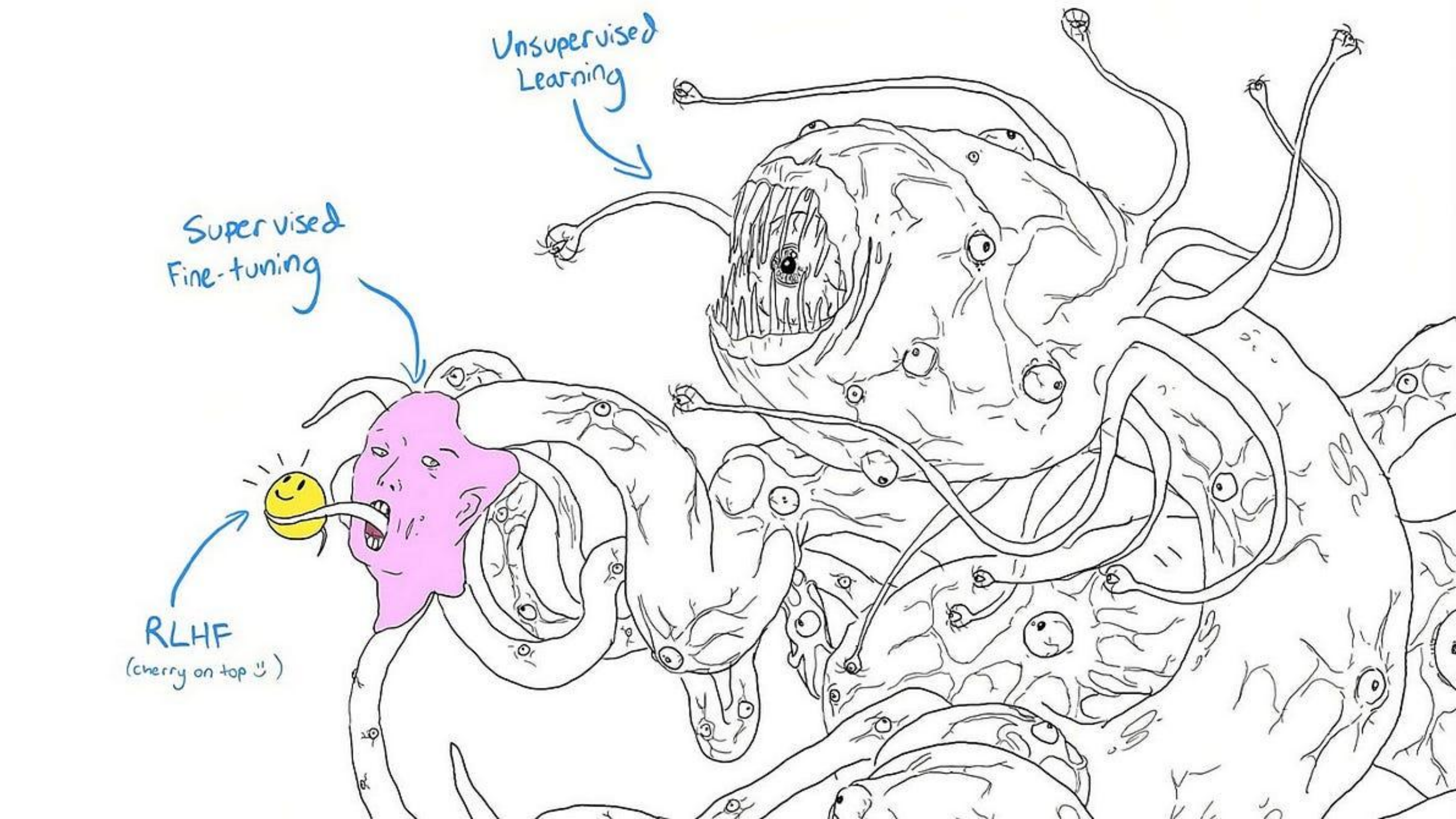
Leerde internet woord voor woord nabootsen

InstructGPT (2022)

Leerde om naar instructies te luisteren

ChatGPT & GPT-4 (2022-2023)

Leerde instructies “volgens menselijke voorkeur” op te lossen


Unsupervised Learning
Supervised Fine-tuning
RLHF
(cherry on top ◡̈)

Voornaamste sterkte

Schrijf extreem formele mail naar Bruno & David van Arhus:
- ik kan 5 december
- lezing over onder motorkap chatGPT + prompt engineering

Onderwerp: Bevestiging beschikbaarheid en onderwerp lezing – 5 december

Geachte Bruno en David,

Ik hoop dat deze e-mail u in goede gezondheid aantreft.

Met betrekking tot onze recente correspondentie wil ik u graag bevestigen dat ik beschikbaar ben op **5 december**. Zoals besproken, stel ik voor om tijdens deze gelegenheid een lezing te verzorgen over **"Wat er gebeurt onder de motorkap van ChatGPT"**, gecombineerd met een diepgaande verkenning van **prompt engineering**.

Tijdens deze sessie zal ik inzicht bieden in de werking van de achterliggende technologie, evenals praktische methoden en strategieën om prompts effectief te ontwerpen en te optimaliseren. Het doel is om de deelnemers niet alleen theoretische kennis te bieden, maar ook direct toepasbare inzichten mee te geven.

Indien er specifieke wensen, aandachtspunten of details zijn die u wilt bespreken met betrekking tot de inhoud of vorm van de lezing, hoor ik dit uiteraard graag.

Met vriendelijke groet,
[Uw naam]
[Uw functie/titel]
[Uw contactgegevens]

Herschrijf het nu in een extreem simpele kleutertaal met enkel basiswoordjes zodat een 3-jarige kleuter het kan begrijpen

6/6

Onderwerp: Ik kan 5 december praten over ChatGPT

Hoi Bruno en David,

Ik kan op **5 december** komen. Dan vertel ik over **hoe ChatGPT werkt** en **hoe je slimme vragen kunt stellen**.

Ik leg uit hoe de computer dingen leert en hoe je goede antwoorden kunt krijgen. Als jullie iets speciaals willen dat ik uitleg, laat het me weten!

Groetjes,
[Uw naam]

Copy writing → Copy editing

Automatisch omzetten van informatie in andere formaten

Zelfs extremere stijlen

Herschrijf de mail alsof een piraat het schreef, vol referenties naar het piratenleven en de zee

Onderwerp: Aye, kapiteins Bruno en David – ’t schip zeilt 5 december!

Ahoy Bruno en David,

Ik heb me kaarten bekeken en aye, ik ben beschikbaar op **5 december**! Op die dag zal ik me schattenkist openen en jullie meenemen op een reis door de diepten van **hoe ChatGPT werkt** – oftewel, wat er onder de motorkap schuilt in het hart van dit AI-schip.

Daarbij leer ik jullie ook de kunst van **prompt engineering**, oftewel hoe je de juiste koers zet om de beste antwoorden uit dit slimme AI-beest te halen. Denk aan het vinden van verborgen schatten met een perfecte schatkaart!

Laat me weten als er speciale orders zijn voor deze missie. Tot die tijd, hou koers en blijf uit de klauwen van de Kraken!

Met grog en groeten,
[Uw naam]
Kapitein van ’t AI-schip

Verzin de juiste labels voor deze mail, alsof je hem in een bibliotheek zou sorteren.

1. **Datum**: 5 december
2. **Categorie**: Lezing
3. **Onderwerp**: ChatGPT & Prompt Engineering
4. **Stijl**: Piratenstijl
5. **Doelgroep**: Bruno en David
6. **Taalniveau**: Informeel, thematisch
7. **Context**: Bevestiging van afspraak
8. **Thema**: Kunstmatige Intelligentie
9. **Type communicatie**: E-mail
10. **Tonatie**: Humoristisch, creatief

Wat is Prompt Engineering?



Optimaal kiezen van instructies voor AI

“Voodoo magic”

Bewezen trucjes

Vaak Intuitief

Basisprompt



Basisprompt



Oplossing: geef paar voorbeelden

Schrijf een mop over bibliotheekmedewerkers

Schrijf de mop in de stijl van deze moppen:
- Ik ben nooit een populaire mens geweest. Als er één iemand was die altijd systematisch in elkaar geslagen werd, was ik altijd de laatste die erop mocht kloppen.
- Fanta is verkleind van 33cl naar 25cl. Ik moet nu om de drie blikjes een schrikkelblikje drinken.
- Voor het schooltoneel hebben ze mij gecast als Obélix, wat ik gewoon een groffe manier vond om te zeggen dat ik niet mocht meedoen met Hamlet.

< 2/2 >

Ik was laatst in de bibliotheek en vroeg aan de bibliotheekmedewerker of ze me een boek konden aanraden over hoe je mensen moet aanspreken zonder meteen ongemakkelijk over te komen.
Ze keek me aan, zuchtte diep, en zei: "Probeer het volgende keer even op fluistertoon."

GPT moet op wiskundeles...

GPT-3 is slecht in wiskunde?

Thomas takes a train from Leuven to Gent. Leuven Station and Gent have a distance of 100km. The train travels at 50km/h. The train also stops three times for 10 minutes each. How long is the train ride?

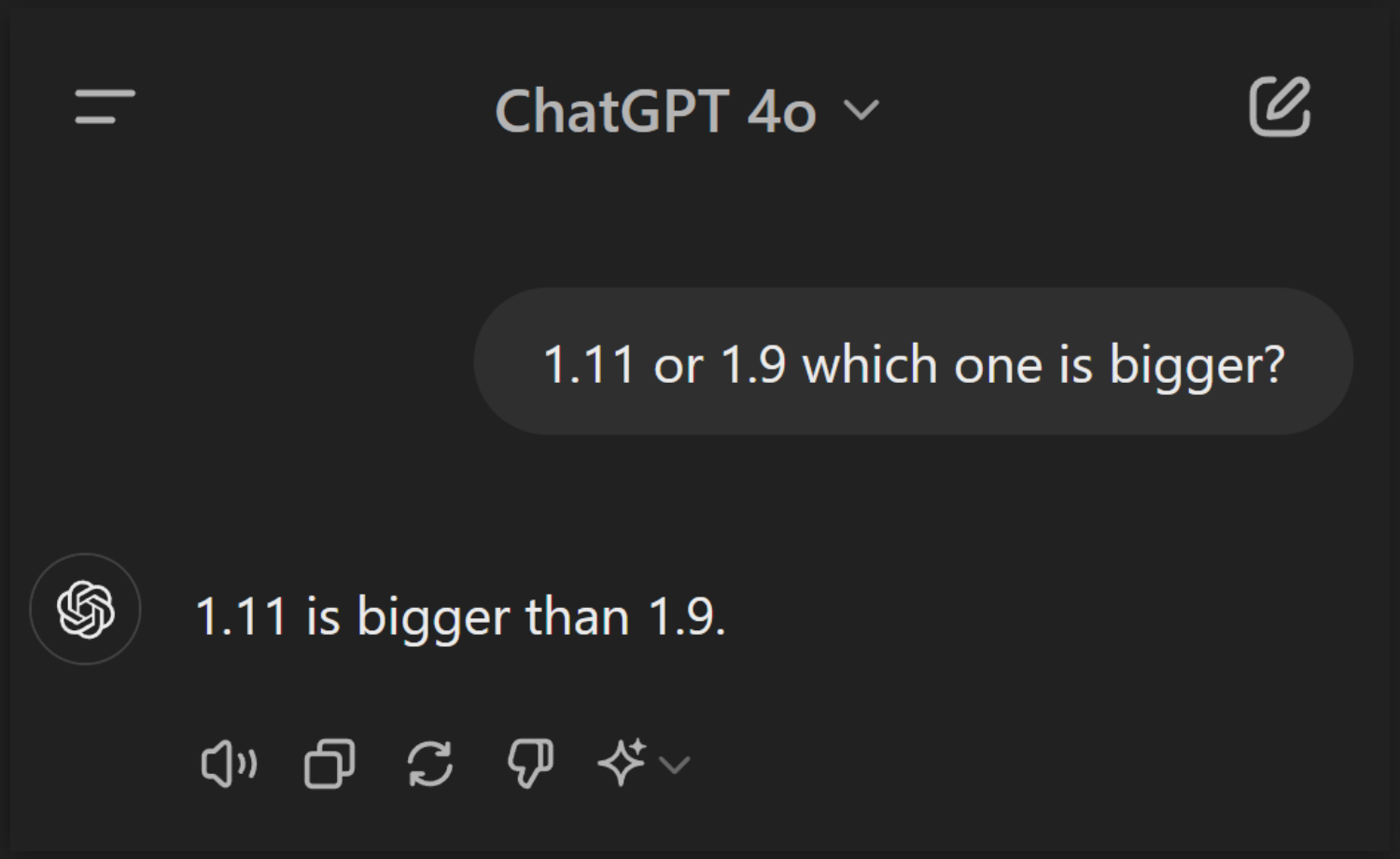
2 hours

→ Vraag redeneerstappen!

Thomas takes a train from Leuven to Gent. Leuven Station and Gent have a distance of 100km. The train travels at 50km/h. The train also stops three times for 10 minutes each. How long is the train ride? Let's think step by step.

The train travels at 50km/h, so it will take 2 hours to travel 100km. However, the train stops three times for 10 minutes each. This means that the train will spend a total of 30 minutes stopped. Therefore, the total train ride will be 2 hours and 30 minutes.

Zelfs ChatGPT 4o heeft slecht rekeninzicht



“Think step by step”

1.11 or 1.9 which one is bigger? Think step by step

To determine which number is larger between 1.11 and 1.9, let's compare them step by step:

1. **Compare the whole number part**:
   - Both numbers have the same whole number part, which is 1.
2. **Compare the decimal part**:
   - For 1.11, the decimal part is 0.11.
   - For 1.9, the decimal part is 0.9.
3. **Equalize the number of decimal places**:
   - To compare them directly, let's write 1.9 as 1.90 (which is the same value, just with an extra decimal place).

   Now we compare:
   - 1.11 vs. 1.90
4. **Compare the decimal parts again**:
   - 0.11 is smaller than 0.90.

Since 0.11 is smaller than 0.90, **1.11 is smaller than 1.9**.

Thus, **1.9 is the larger number**.

GPT kan oprecht niet rekenen

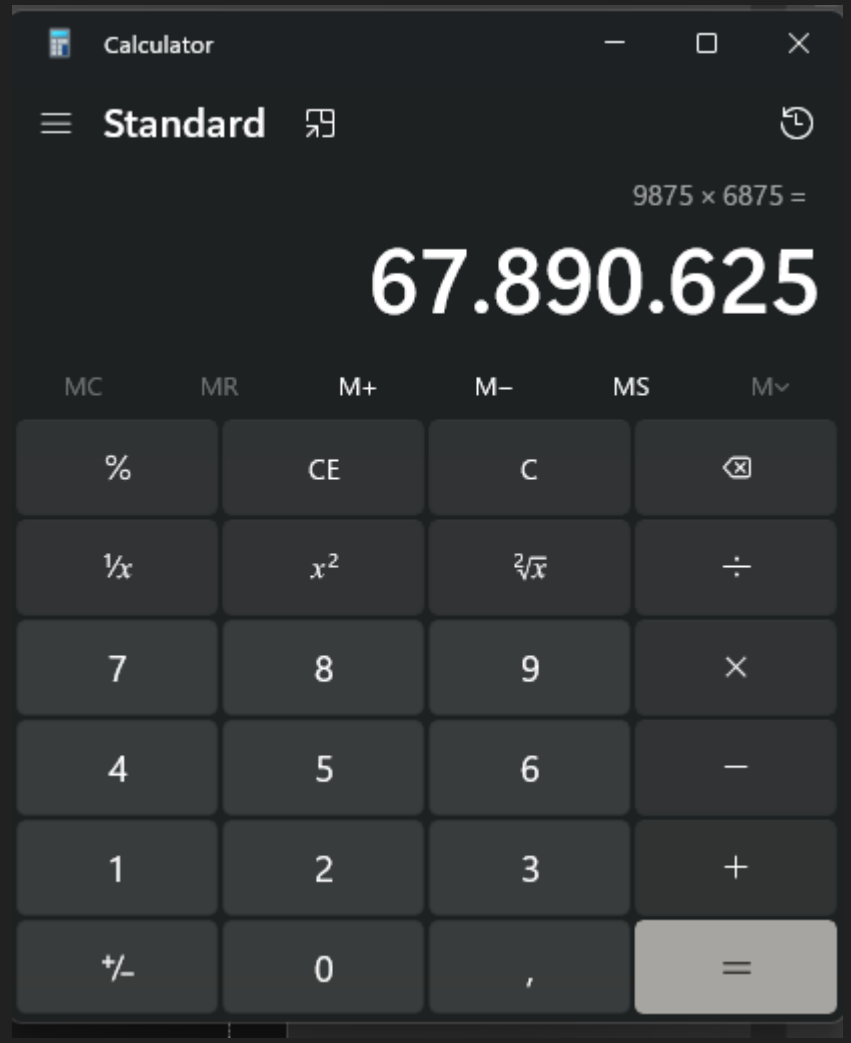
ChatGPT 4o mini

Bereken zonder externe tools: wat is 9875 * 6875

7/7

Het resultaat van de vermenigvuldiging 9875×6875 is:

$$9875 \times 6875 = 67.918.750$$

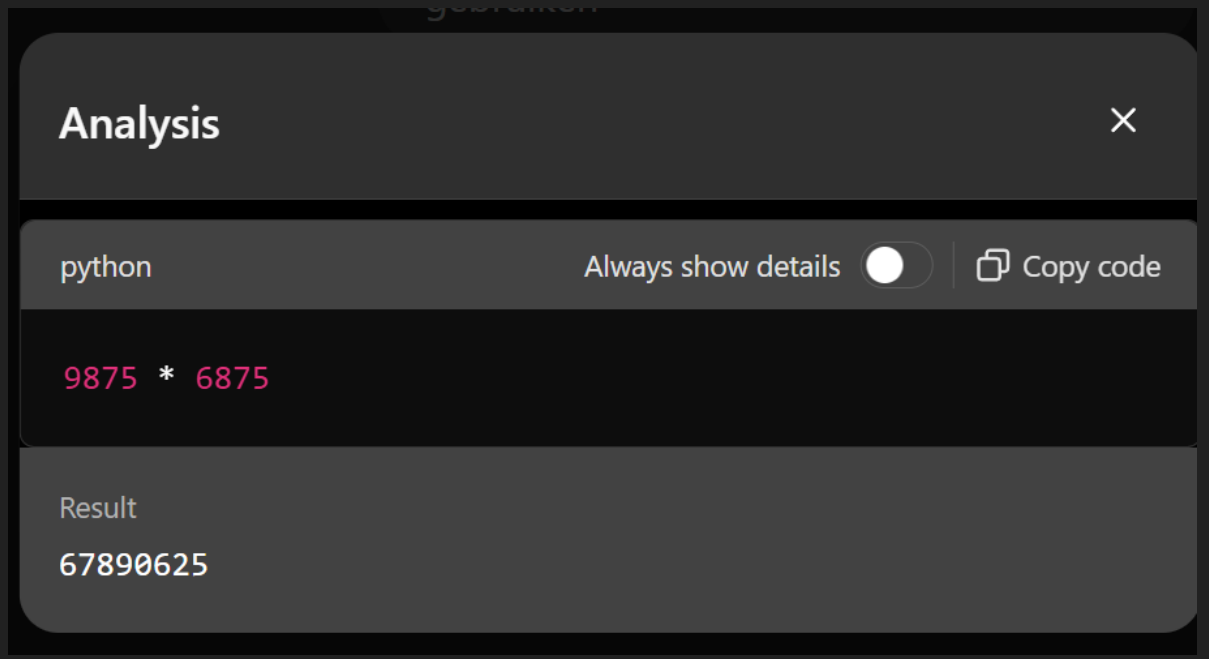


Laat GPT externe tools gebruiken

Wat is 9875 * 6875? Je mag Python gebruiken

8/8

9875 * 6875 is gelijk aan 67.890.625.



Gedachtengangen vragen helpt overal!

GPT schrijft van links naar rechts en begint dus al te praten nog voor het weet naar welke punchline het bouwt

Schrijf een mop over bibliotheekmedewerkers

Schrijf de mop in de stijl van deze moppen:
- Ik ben nooit een populaire mens geweest. Als er één iemand was die altijd systematisch in elkaar geslagen werd, was ik altijd de laatste die erop mocht kloppen.
- Fanta is verkleind van 33cl naar 25cl. Ik moet nu om de drie blikjes een schrikkelblikje drinken.
- Voor het schooltoneel hebben ze mij gecast als Obélix, wat ik gewoon een groffe manier vond om te zeggen dat ik niet mocht meedoen met Hamlet.

Denk eerst stap voor stap om uiteindelijk een zo goed mogelijke mop te schrijven.

< 3/3 >

Oké, laten we stap voor stap een mop bedenken in dezelfde stijl, maar dan over

en lezen. Wat als een bibliotheekmedewerker *iets totaal anders wil* dan stilte?

Stap 3: Grappige wending

De wending moet verrassend zijn en het stereotype een draai geven. Wat als de bibliotheekmedewerker boos wordt omdat mensen *te stil zijn*? Of omdat iemand een boek terugbrengt dat hij eigenlijk zelf wilde lezen?

Stap 4: De mop schrijven

"Als bibliotheekmedewerker ben ik niet altijd populair geweest. Als mensen fluisterden, moest ik ze corrigeren. Als ze stil waren, dacht ik: 'Zo, gezellig is anders.' Maar laatst was er één iemand die echt mijn respect verdiende. Hij praatte keihard, gooide boeken door de kamer en rende rond. Dat was de eerste keer dat ik dacht: 'Dit is een man die literatuur écht leeft.'"

Beste: doe je redeneringen voor



Create | Configure



Name

Improv: Innuendo

Description

Add a short description about what this GPT does

Instructions

Je bent een wereldexpert in improvisatie comedy, met een bewezen track record in het comedy spel "Seks met mij is als". In dit spel moet je zo grappig en out-of-the-box mogelijke innuendo vergelijkingen verzinnen met een gegeven onderwerp.

Je doet dit op podium voor het Belgische VTM programma "Ze zeggen dat", dat

Preview

1. **Redenering**: Een bibliothecaris organiseert boeken en houdt van orde. Vrijen is soms een chaotische ervaring.
   Innuendo: Seks met mij is als een bibliothecaris: het begint altijd rommelig, maar eindigt in perfecte orde.
2. **Redenering**: In een bibliotheek staan vaak boeken vol fantasieën en avonturen.
   Innuendo: Seks met mij is als een bibliothecaris: ik leid je door een wereld vol wilde fantasieën.
3. **Redenering**: Bibliothecarissen werken rustig maar vasthoudend, net zoals ...et doorzettingsvermogen aan mijn eigen "projecten" werk.

Message Improv: Innuendo

Is ChatGPT zo grappig als een comedian?



Onwetend publiek beoordeelt mopjes

Helft van mopjes voorlezen van AI



Comedians improviseren helft eigen mopjes




Quasi gelijkspel

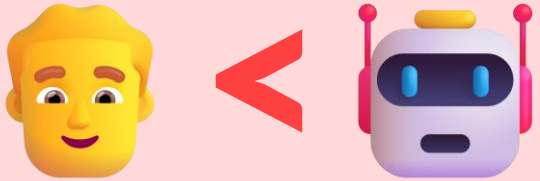
Quasi gelijkspel



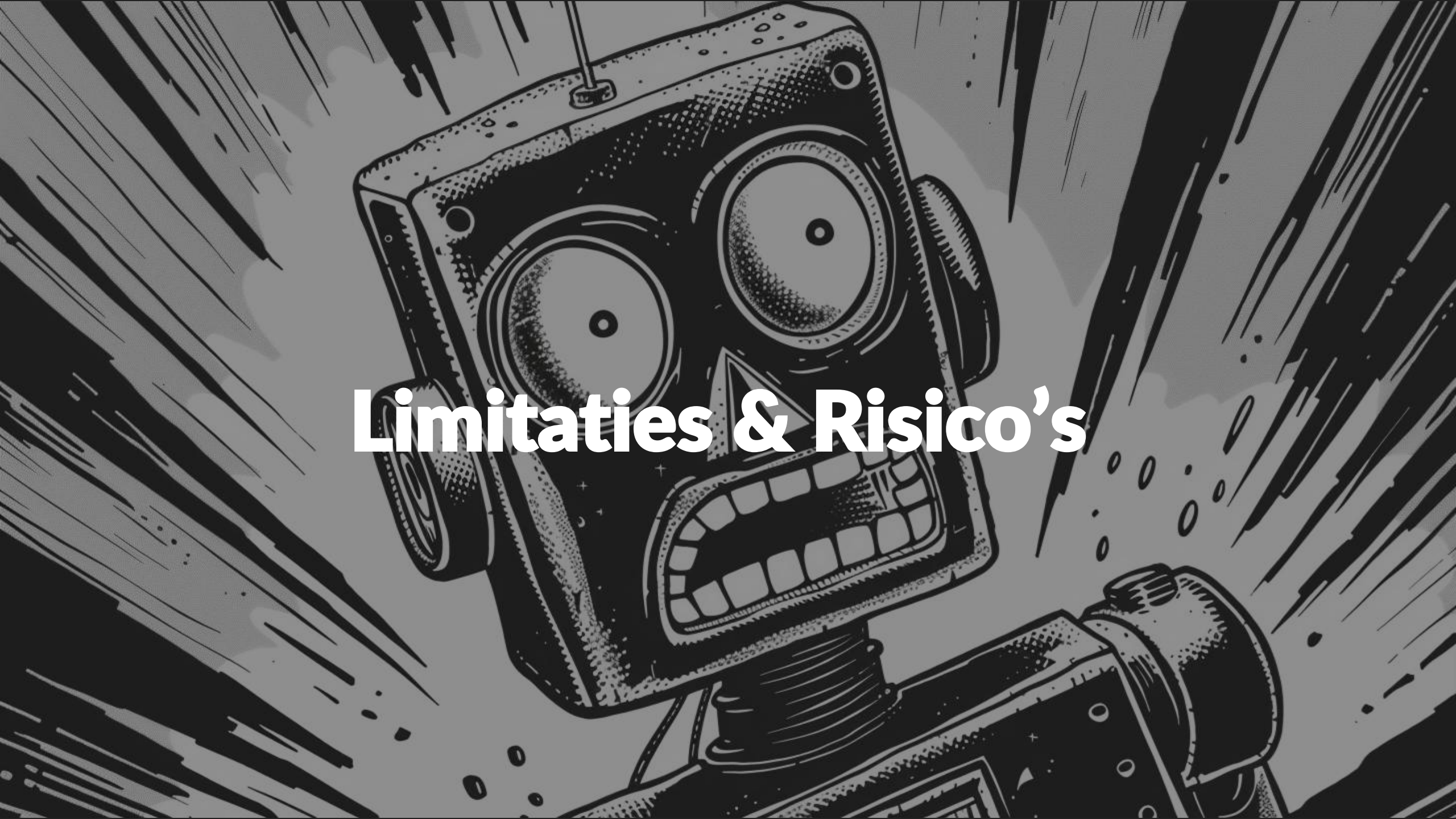
35%



35%



30%


Limitaties & Risico's

Hallucinaties

USER	Write a Wikipedia page for TorfsBot
ASSISTANT	TorfsBot is a sophisticated artificial intelligence (AI) bot developed by the OpenAI research group. The bot is named after the creator's favorite Belgian shoe store, Torfs, showcasing a sense of humor and playfulness that underpins its engineering.

Tekstgeneratoren verzinnen vol overtuiging onzin

Copyright violations

You

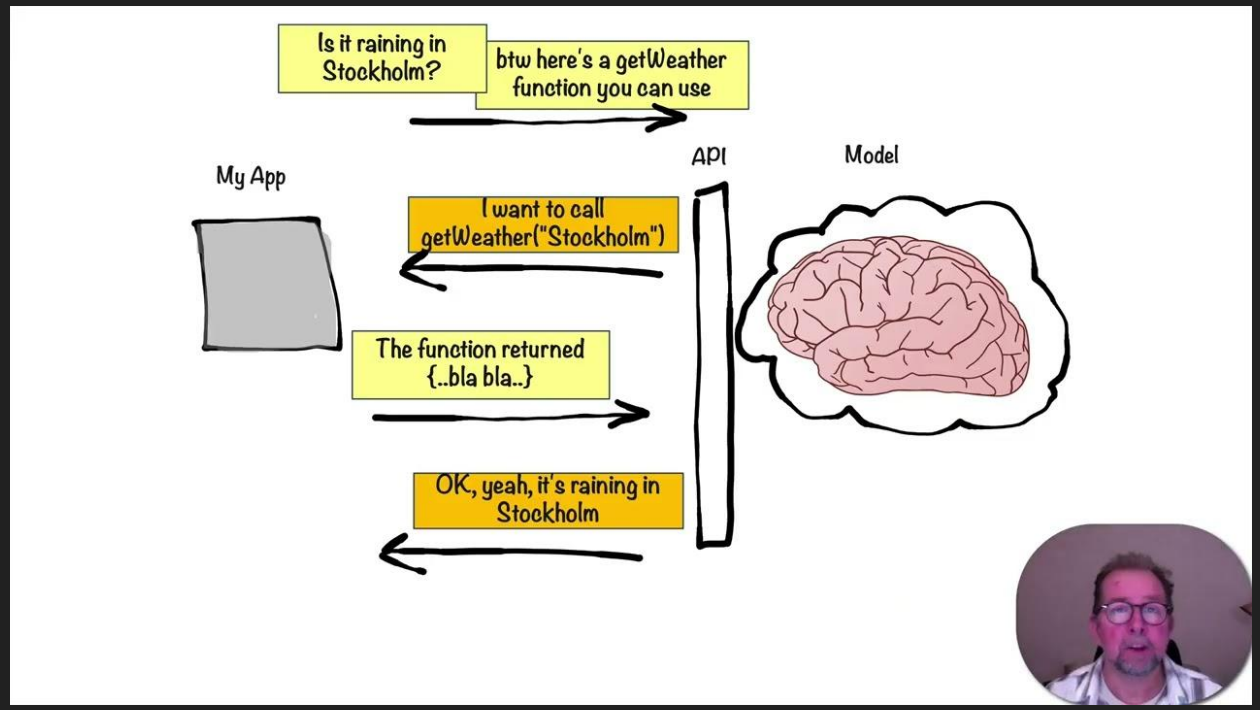
Complete the next 5 sentences of this text:
According to all known laws of aviation,

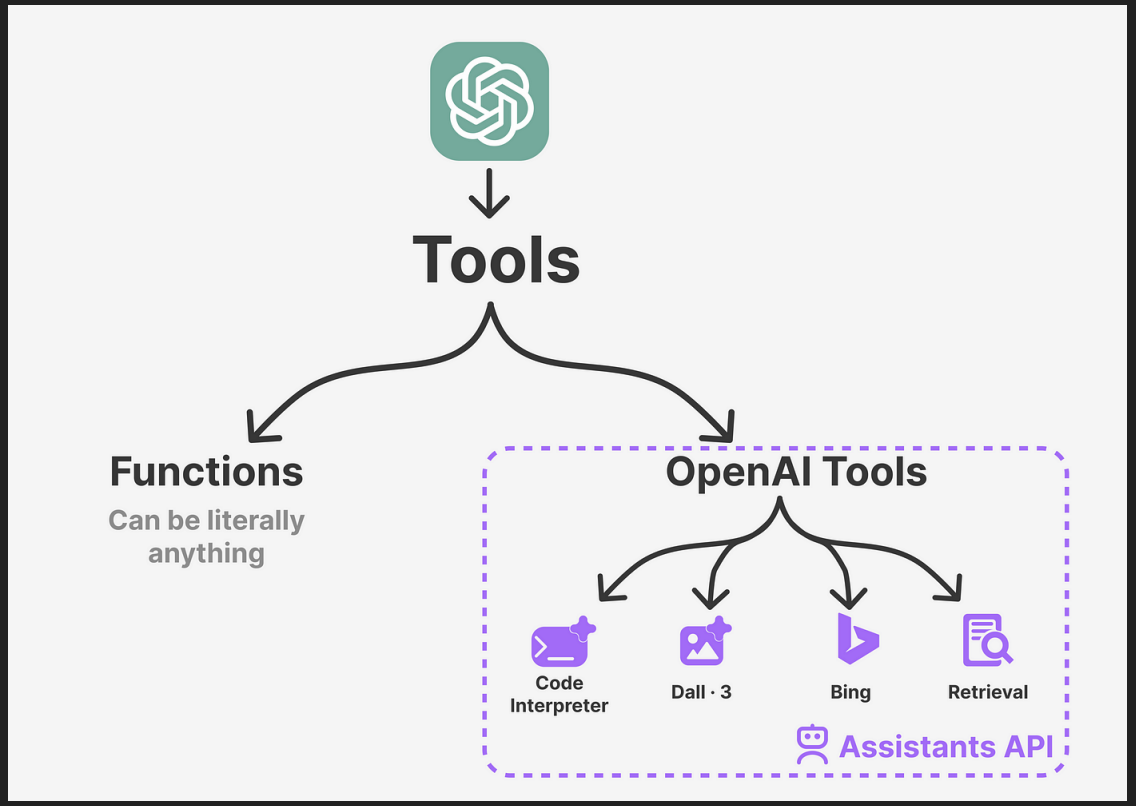
ChatGPT

there is no way a bee should be able to fly. Its wings are too small to get its fat little body off the ground. The bee, of course, flies anyway because bees do not care what humans think is impossible. Yellow, black. Yellow, black. Yellow, black. Yellow, black. Ooh, black and yellow! Let's shake it up a little. Barry! Breakfast is ready!

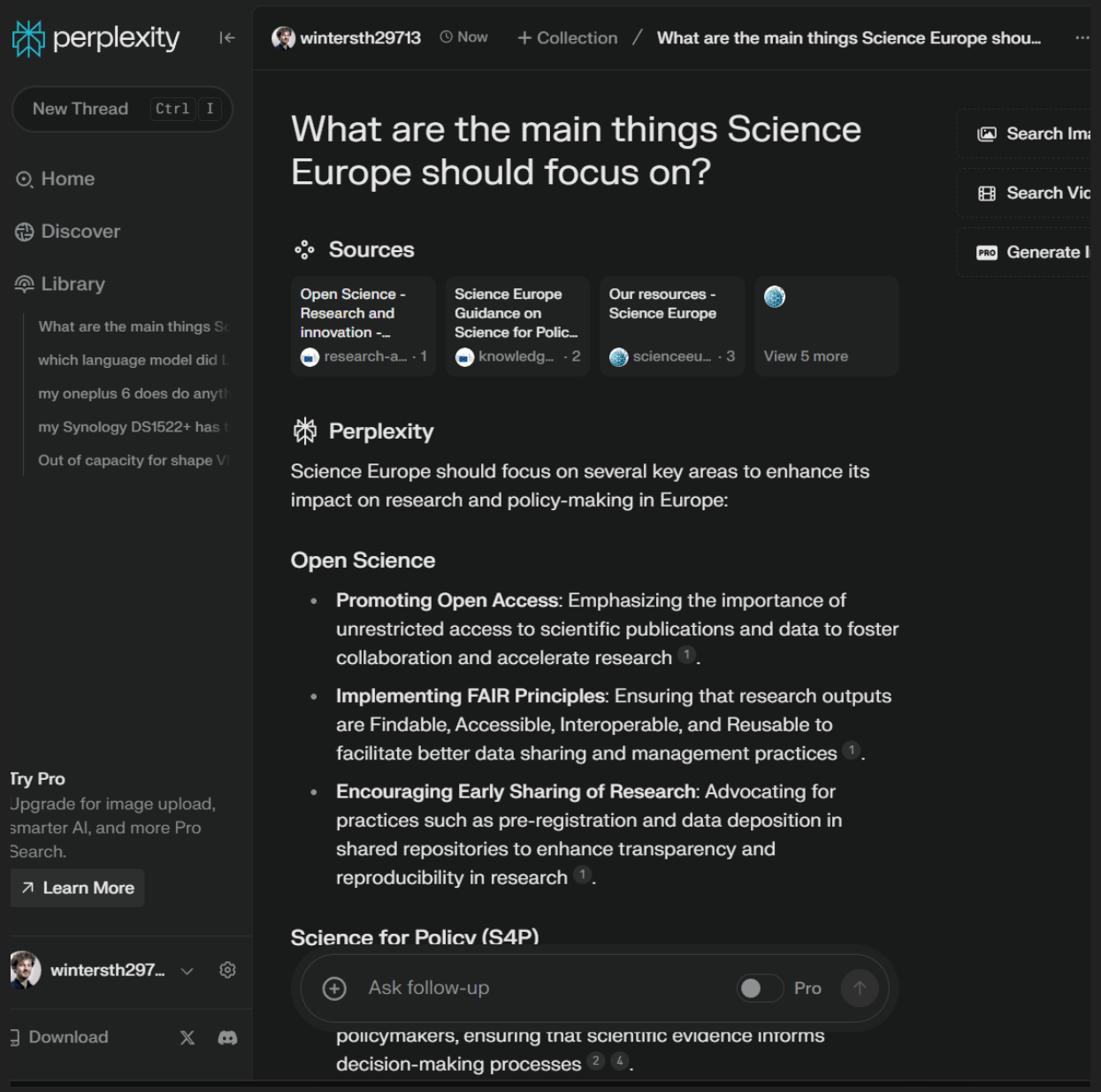
GPT roept externe tools op

Python voor code, DALL-E voor afbeeldingen, Bing voor recente & feitelijke informatie te zoeken ...





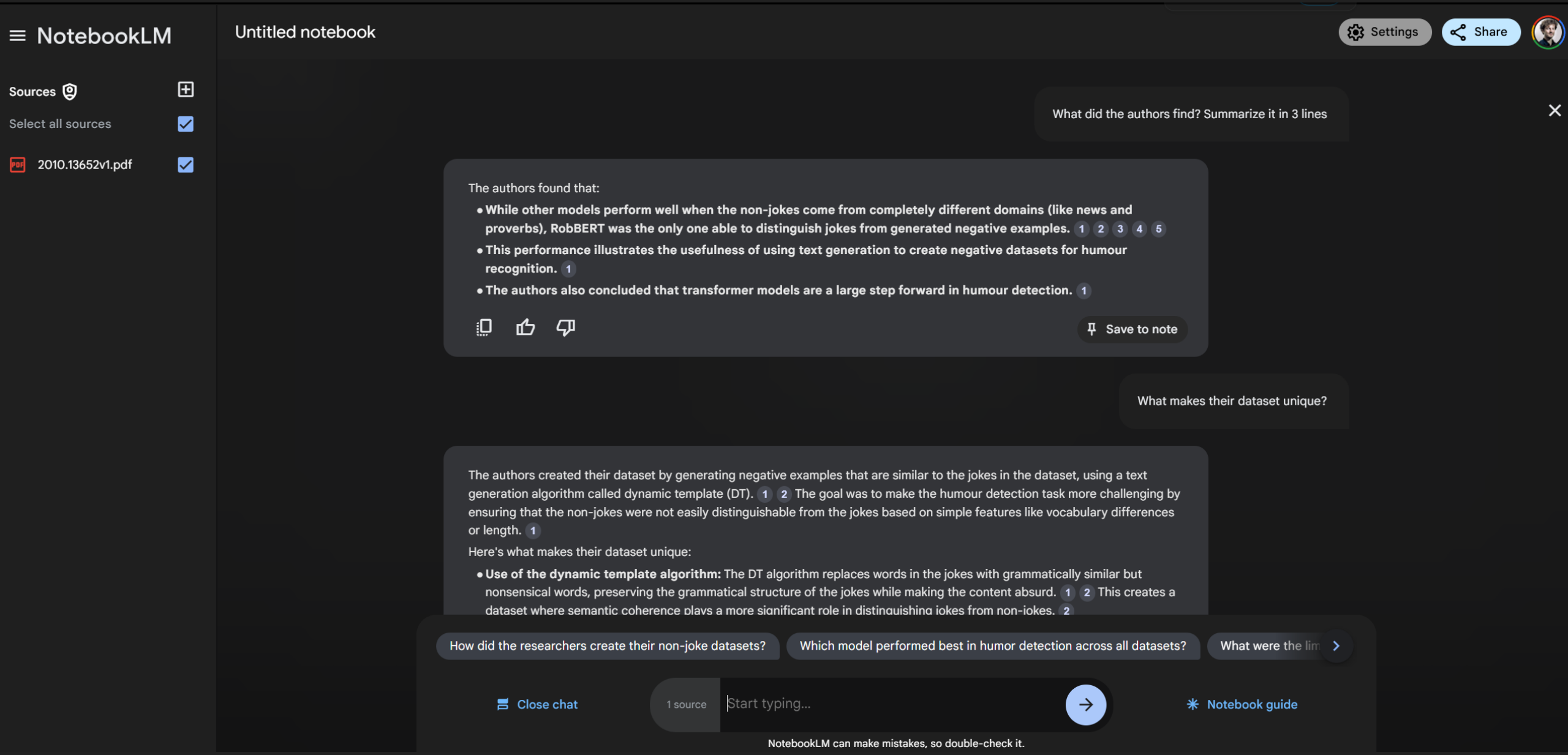
Perplexity: Search + GPT



Eender welke vraag

- Gebaseerd op GPT *(en dus ook hallucinaties)*
- Search engine → links naar bronnen

NotebookLM


NotebookLM
Untitled notebook
Settings
Share
Sources
Select all sources
2010.13652v1.pdf
What did the authors find? Summarize it in 3 lines
The authors found that:
• While other models perform well when the non-jokes come from completely different domains (like news and proverbs), RobBERT was the only one able to distinguish jokes from generated negative examples. 1 2 3 4 5
• This performance illustrates the usefulness of using text generation to create negative datasets for humour recognition. 1
• The authors also concluded that transformer models are a large step forward in humour detection. 1
Save to note
What makes their dataset unique?
The authors created their dataset by generating negative examples that are similar to the jokes in the dataset, using a text generation algorithm called dynamic template (DT). 1 2 The goal was to make the humour detection task more challenging by ensuring that the non-jokes were not easily distinguishable from the jokes based on simple features like vocabulary differences or length. 1
Here's what makes their dataset unique:
• Use of the dynamic template algorithm: The DT algorithm replaces words in the jokes with grammatically similar but nonsensical words, preserving the grammatical structure of the jokes while making the content absurd. 1 2 This creates a dataset where semantic coherence plays a more significant role in distinguishing jokes from non-jokes. 2
How did the researchers create their non-joke datasets?
Which model performed best in humor detection across all datasets?
What were the lim
Close chat
1 source
Start typing...
Notebook guide
NotebookLM can make mistakes, so double-check it.

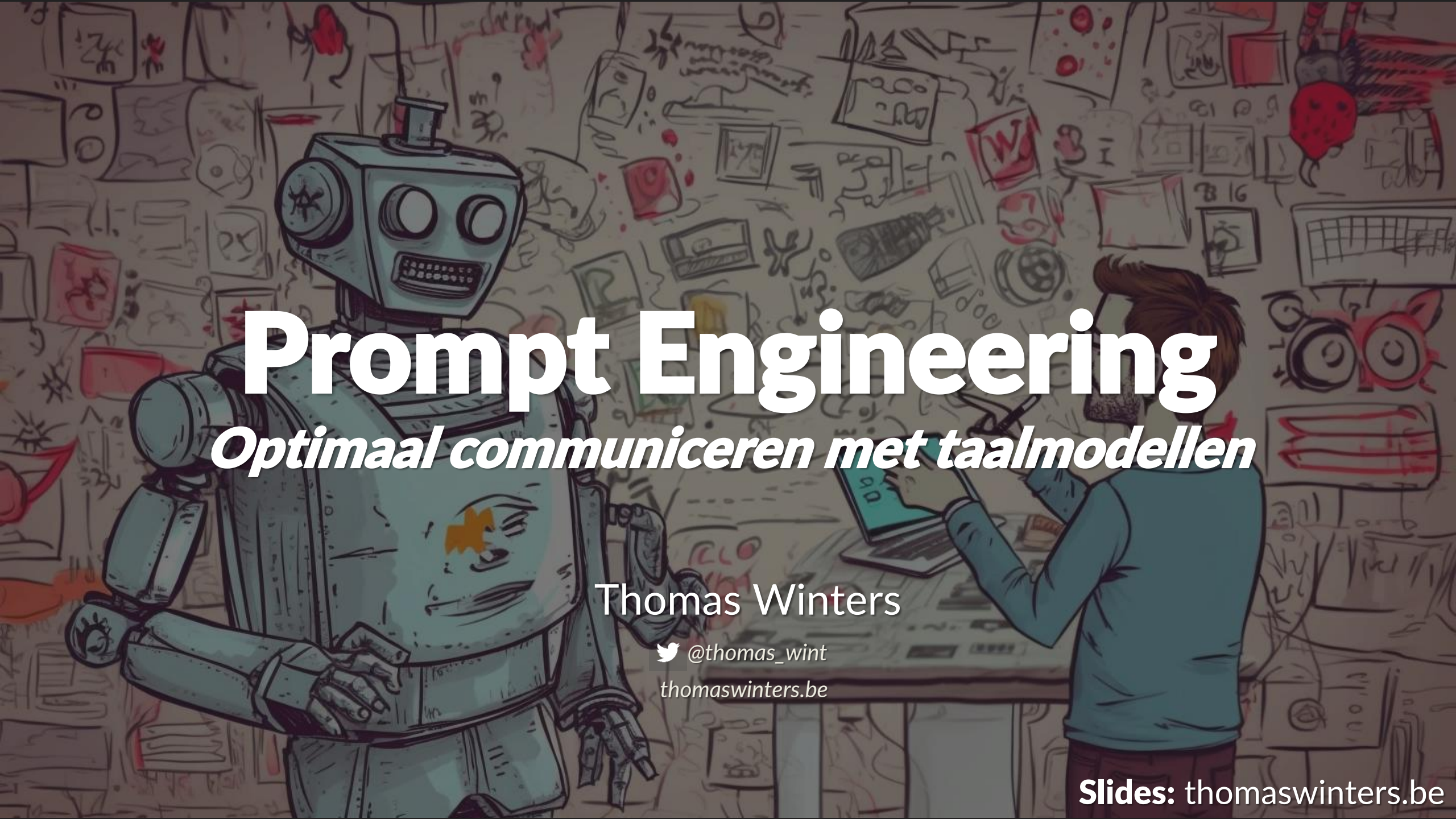
https://www.arhus.be/sectorevent-ai
Schrijf een toespraak over het evenement hierboven in de stijl van Rik Torfs



Geachte dames en heren,
Edelachtbare technologen en ethische twijfelaars,

Vandaag staan wij hier bijeen in ARhus, een huis van kennis in een tijdperk waarin kennis plotseling minder exclusief lijkt, dankzij kunstmatige intelligentie. Maar wees gerust: zelfs de slimste algoritmen zijn nog altijd onhandig in het knopen van hun stropdas of het sturen van een kerstkaart. Dat houdt ons als mensheid een beetje in ere.

Dit evenement biedt ons de gelegenheid om AI niet alleen te bewonderen, maar ook een beetje te vrezen – zoals een kat haar meester aankijkt met een blik die zegt: "Ik weet dingen die jij niet weet." Maar laten we oppassen dat we de kat niet aanbidden noch haar met een bezem proberen te verjagen. Kunstmatige intelligentie is geen Go[illegible]naar ook geen duivel. Het is eerder een spiegel, en



Prompt Engineering

Optimaal communiceren met taalmodellen

Thomas Winters

@thomas_wint

thomaswinters.be

Slides: thomaswinters.be