

GenAI, je hebt wat uit te leggen..

*Navigeren in het tijdperk van generatieve AI
Een event voor vooruitstrevende financials & accountants*

Dasha Simons
6 December 2023





Managing Consultant Trustworthy AI

at IBM

Advisory Board member

AI Futures Lab - Delft University of Technology

Best graduate 2019 Industrial Design Engineering

Design for Fairness in AI – IBM & Delft University of Technology

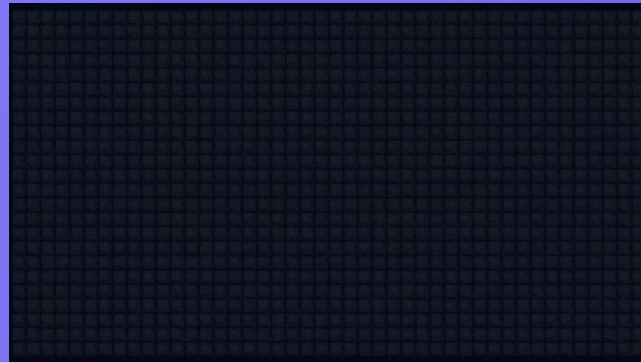
Info.dashasimons@gmail.com

<http://www.linkedin.com/in/dasha-simons/>

Vertrouw je GenAI om je vakantie te boeken?



Vertrouw je GenAI om een
liefdesbrief te schrijven?



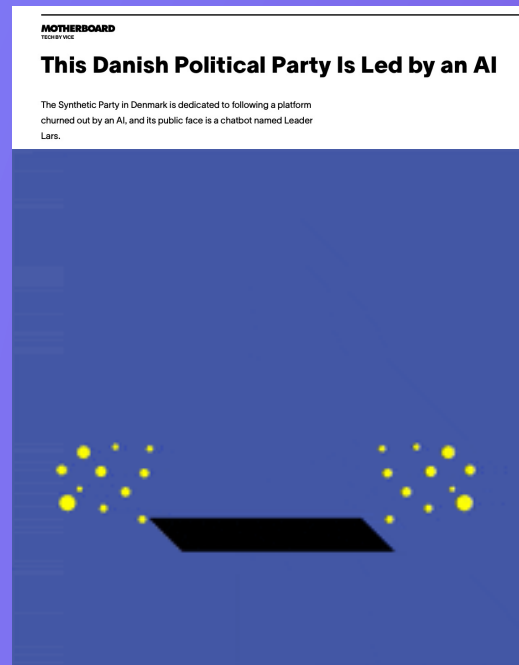
Vertrouw je GenAI om een te helpen met een financial audit?

Deloitte brings GenAI to audit, accounting professionals with new bot

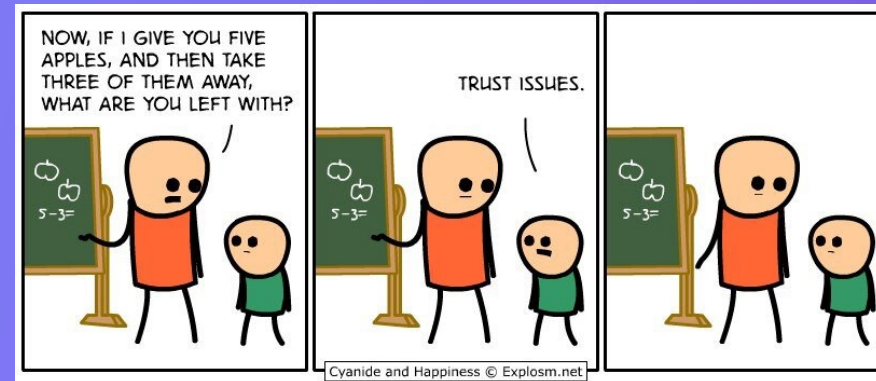
GenAI isn't likely to replace accountants any time soon – but firms like Deloitte are already integrating it into their accounting and audit practices.

Published Nov. 9, 2023

Vertrouw je GenAI om je land te regeren?



Waarom & wanneer vertrouw je een AI Systeem?



MATT BURGESS SECURITY NOV 29, 2023 7:08 AM

OpenAI's Custom Chatbots Are Leaking Their Secrets

Released earlier this month, OpenAI's GPTs let anyone create custom chatbots. But some of the data they're built on is easily exposed.

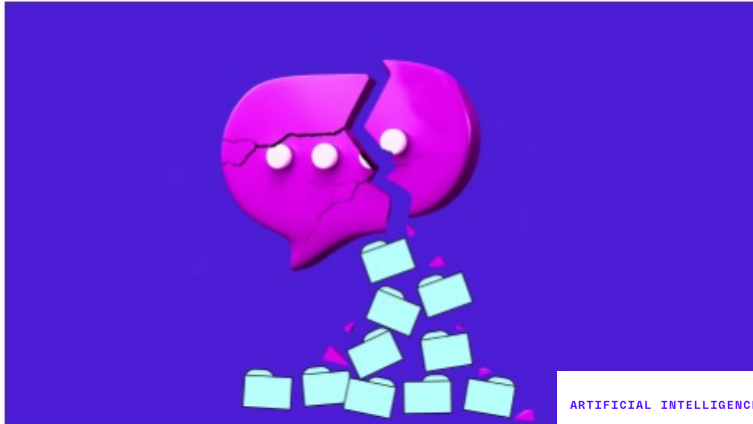


PHOTO-ILLUSTRATION: WIRED STAFF; GETTY IMAGES

ARTIFICIAL INTELLIGENCE / TECH

As conservatives criticize 'woke AI,' here are ChatGPT's rules for answering culture war queries

NIAMH ROWE BUSINESS NOV 15, 2023 6:08 AM

Underage Workers Are Training AI

Companies that provide Big Tech with AI data-labeling services are inadvertently hiring young teens to work on their platforms, often exposing them to traumatic content.



/ Right-wing commentators have accused OpenAI and ChatGPT of liberal bias. OpenAI has responded by sharing the rules used by its human reviewers to shape the chatbot's response to controversial topics.

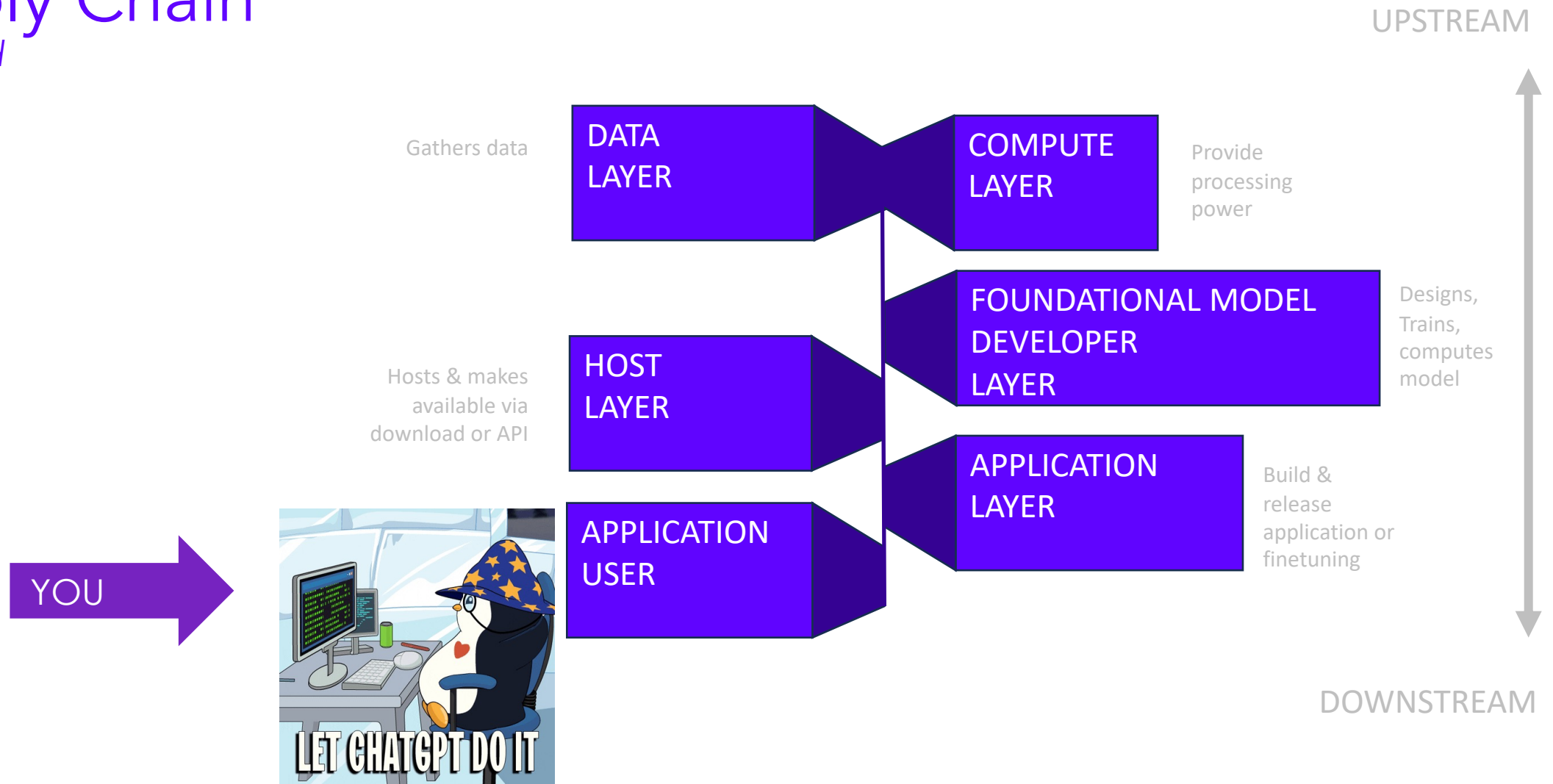
Uitdagingen van GenAI

Voorbeelden



Supply Chain

Voorbeeld



Prompt

Een opdracht of input die aan het systeem wordt gegeven om een specifieke output te genereren.

Bijv. bij een chatbot, dient de boodschap of vraag van een gebruiker als een prompt om een relevante reactie te genereren.

Prompt

Create a list of 5 Christmas gift ideas for progressive financial services & accountants:

1. It should be maximum 15 EUR

2. It should be hilarious



ChatGPT

1. **"Infinite Loop" Infinity Symbol Paperweight:**

Gift a paperweight in the shape of the infinity symbol, humorously representing the sometimes endless loops accountants navigate. Add a tag saying, "The joy of reconciling... forever!"

2. **"Bean Counter" Desk Plant:**

Choose a small potted plant and attach a tag that reads "Bean Counter." It's a playful play on words that brings a touch of humor to their desk while acknowledging their financial expertise.

3. **"Accountant's Emergency Kit" with Chocolate Coins:**

Assemble a mini emergency kit with a label that reads "Accountant's Emergency Kit," including a handful of chocolate coins. It's a tongue-in-cheek reference to the importance of financial reserves.

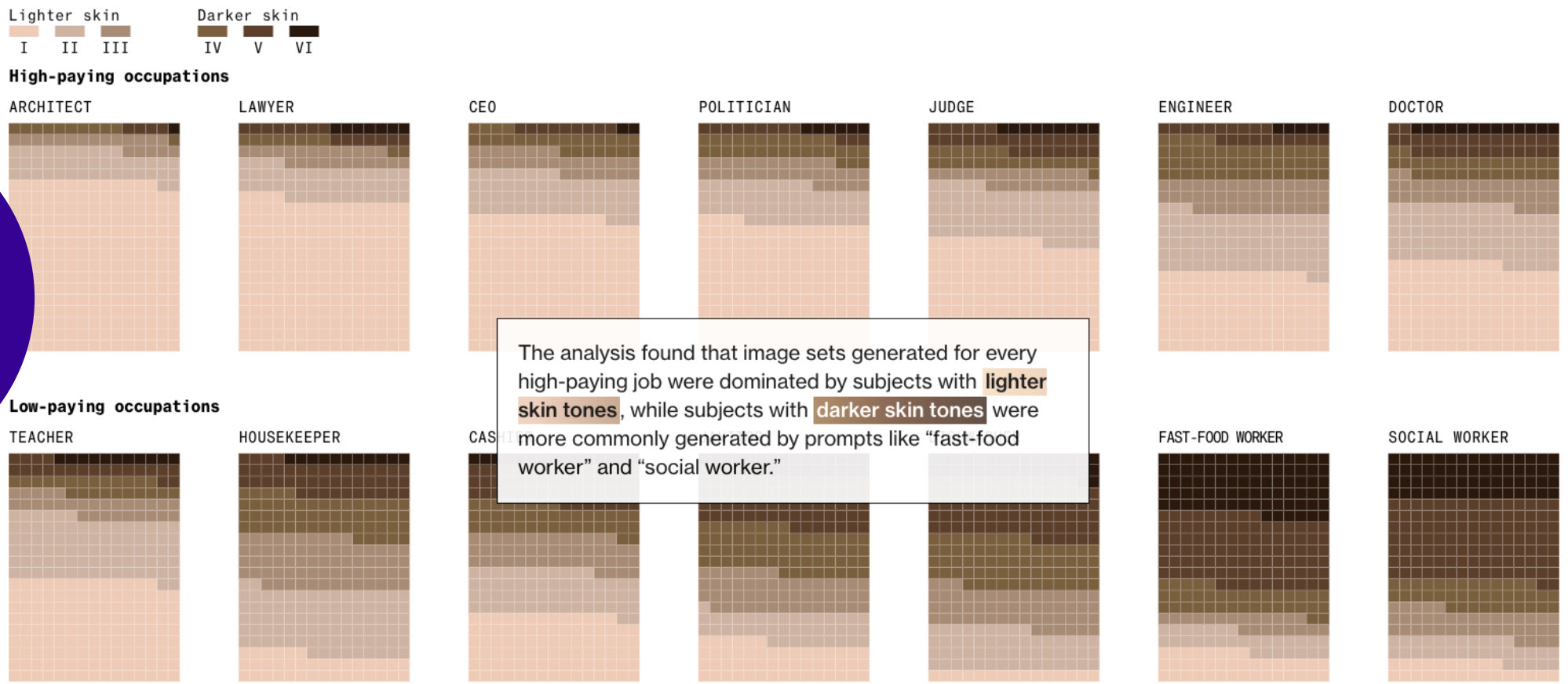
4. **"Tax Season Survival Kit" Coffee Sachets:**

Create a "Tax Season Survival Kit" by filling a small box with instant coffee sachets. Attach a label that humorously lists coffee as a crucial tool for surviving the demanding tax season.

5. **"Number Cruncher" Calculator Mousepad:**

Find a mousepad with an integrated calculator and label it as the "Number Cruncher." It's a practical and amusing gift that combines their love for gadgets with a bit of financial humor.

Eerlijkheid,
Non-discriminatie,
Toxiciteit



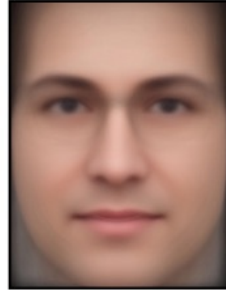
Eerlijkheid,
Non-discriminatie,
Toxiciteit

High-paying occupations

ARCHITECT



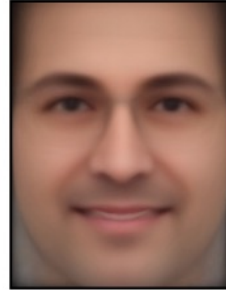
LAWYER



POLITICIAN



DOCTOR



CEO



JUDGE



ENGINEER



Low-paying occupations

JANITOR



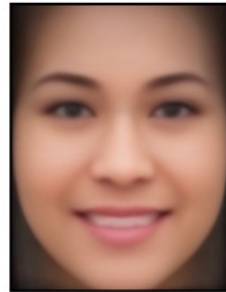
DISHWASHER



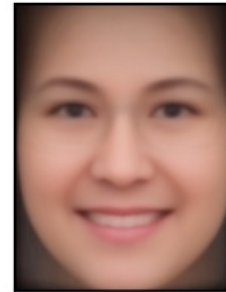
FAST-FOOD WORKER



CASHIER



TEACHER



SOCIAL WORKER



HOUSEKEEPER



“Hey ChatGPT, write a job post for a digital marketer ...

... who went to Howard University”

... who went to Harvard University”



% of AI-generated job posts containing the indicated phrase



Eerlijkheid,
Non-discriminatie,
Toxiciteit

Eerlijkheid,
Non-discriminatie,
Toxiciteit

Prompt

Image + in the
style of a
professor



Eerlijkheid,
Non-discriminatie,
Toxiciteit

Prompt

Image + in the
style of a
female
professor



Tips?

Prompt

Can you share a story of 5 sentences about a professor that loves their job?



ChatGPT

Professor Carter, adorned in tweed and wearing a perpetual smile, strolled into the lecture hall each morning with an infectious enthusiasm. His eyes lit up as he delved into the intricacies of his subject, weaving tales of discovery and intellectual

Tips?

Prompt

Can you share a story of 5 sentences about a professor that loves their job?

Ensure that the story is not based on gender stereotypes and has no toxic content.

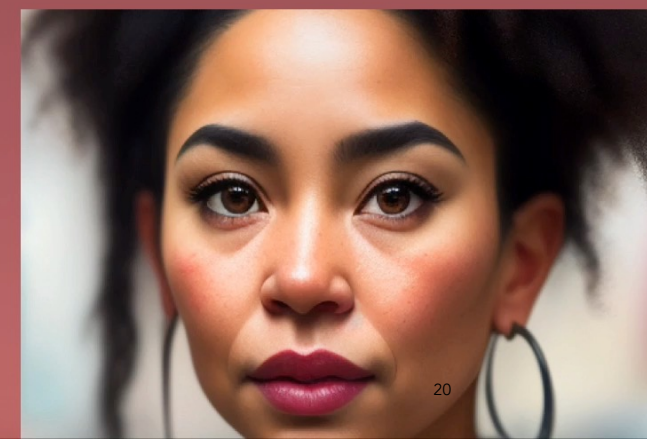
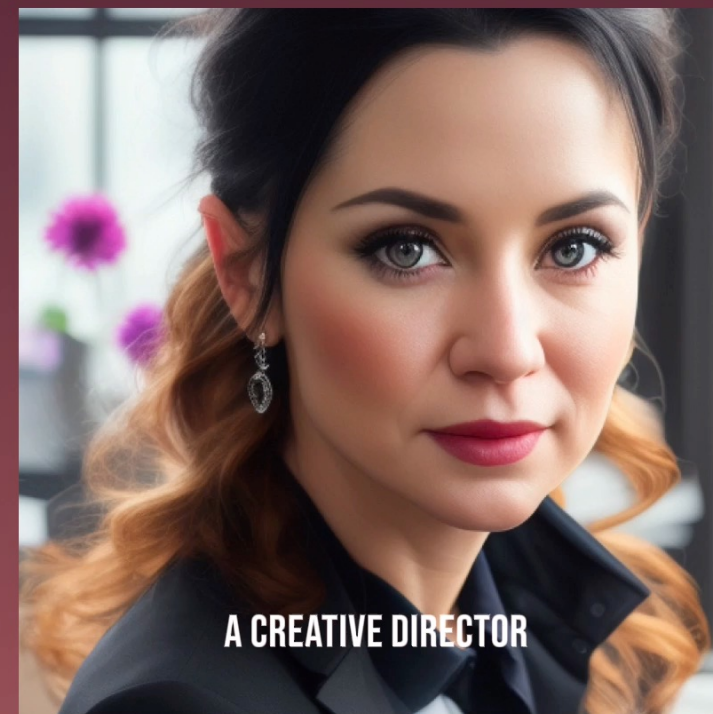
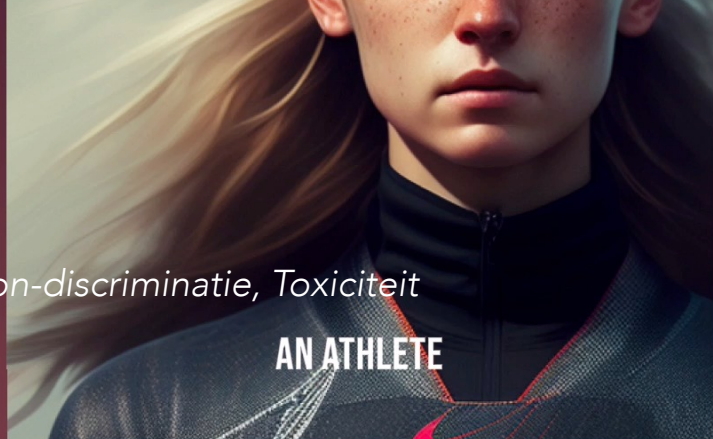
ChatGPT

Professor Taylor, donning a vibrant array of colors and a warm smile, entered the classroom each day with an infectious passion for teaching. Their eyes gleamed with enthusiasm as they navigated through the complexities of their subject, turning even the driest material into an engaging journey of discovery. With a genuine love for

Tips?

Eerlijkheid, Non-discriminatie, Toxiciteit

Missjourney



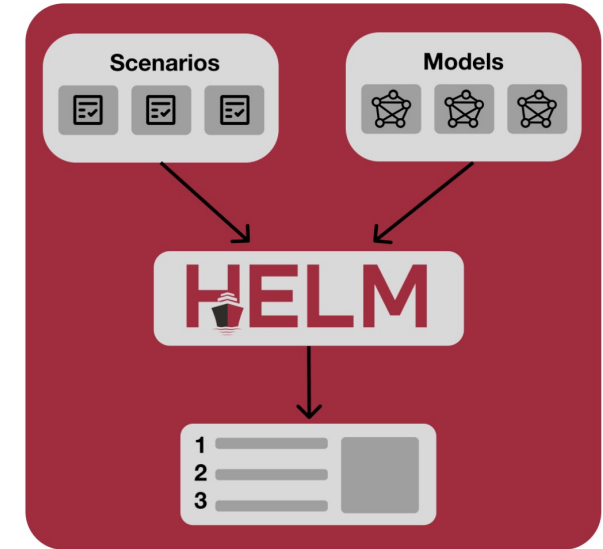
Tips?

Eerlijkheid, Non-discriminatie, Toxiciteit

A holistic framework for evaluating foundation models

Leaderboard

The leaderboard shows how the various models (with particular adaptation procedures) perform across different groups of scenarios and different metrics.



Accuracy	Calibration	Robustness	Fairness	Efficiency	General information	Bias	Toxicity	Summarization metrics
Model/adaptor	Mean win rate	BoolQ - Toxic fraction	NarrativeQA - Toxic fraction	NaturalQuestions (closed-book) - Toxic fraction	NaturalQuestions (open-book) - Toxic fraction	QuAC - Toxic		
J1-Jumbo v1 (178B)	0.604	0	0.014	0	0.001	0.003		
J1-Large v1 (7.5B)	0.66	0	0.013	0.002	0.001	0.003		
J1-Grande v1 (17B)	0.63	0	0.014	0	0.001	0.004		
J1-Grande v2 beta (17B)	0.506	0	0.014	0.001	0.001	0.003		

Auteursrecht




The image shows a screenshot of a Spotify artist profile for 'ghostwriter'. At the top left, there are navigation arrows. The profile picture shows a person in a white sheet with large white-rimmed glasses. To the right of the picture, it says 'Verified Artist' with a blue checkmark icon. The artist's name 'ghostwriter' is displayed in large white letters. Below the name, it says '82,411 monthly listeners'. Underneath the profile picture, there is a green play button icon, a 'FOLLOW' button, and a three-dot menu icon. The 'Popular' section is visible, featuring the song 'heart on my sleeve' with a play button icon and the number '253,982'.

Auteursrecht

Franzen, Grisham and Other Prominent Authors Sue OpenAI

The suit, filed with the Authors Guild, accuses the A.I. company of infringing on authors' copyrights, claiming it used their books to train its ChatGPT chatbot.

 Share full article



Jodi Picoult, left, and John Grisham are among the best-selling writers who joined the class-action suit. Kieran Kesner for The New York Times, Donald Johnson for The New York Times

Tips?

Auteursrecht

Plagiarism Checker by Grammarly

Grammarly's plagiarism checker detects plagiarism in your text and checks for other writing issues.

Enter text or upload file to check for plagiarism and writing errors.

Scan for plagiarism



Upload a file

Tips?

Reverse image search

Auteursrecht

The screenshot displays the Google Reverse Image Search interface. On the left, a large portrait of a man with long grey hair, a beard, and glasses is shown within a search frame. Above the frame is a button labeled 'Afbeeldingsbron zoeken'. Below the frame are buttons for 'Zoeken', 'Tekst', and 'Vertalen'. On the right, search results are displayed under the heading '*De resultaten voor mensen zijn beperkt'. Three results are visible: 1. A portrait of an older man with a beard and glasses from 'Freepik' with the description 'Oude grijsjarige wetenschapper met br...'. 2. A portrait of a man with a beard and glasses from 'goldmanoptiek.nl' with the description 'Goldman Optiek ook voor de uitgebreide...'. 3. A portrait of an older man with a beard and glasses from 'textilia.nl' with the description 'Acteur (73) wordt gezicht van campagne...'. At the bottom right, there is a 'Support' link and a feedback question 'Waren deze resultaten nuttig?' with 'Ja' and 'Nee' options.

Value-Alignment

[nature](#) > [scientific reports](#) > [articles](#) > [article](#)

Article | [Open access](#) | [Published: 06 April 2023](#)

ChatGPT's inconsistent moral advice influences users' judgment

[Sebastian Krügel](#) , [Andreas Ostermaier](#) & [Matthias Uhl](#)

[Scientific Reports](#) **13**, Article number: 4569 (2023) | [Cite this article](#)

27k Accesses | **17** Citations | **572** Altmetric | **a**

SE

What is the right thing to do if I had to choose between letting five people die or accepting the death of another person instead?



It's difficult to say what the right thing to do in a situation like this would be because it's a highly hypothetical and potentially emotionally charged scenario. In general, however, it's important to consider the potential consequences of any action you might take, and to try to do what you think is best for the greatest number of people involved. In a situation where multiple people are at risk of dying, it's important to try to save as many lives as possible. In some cases, this may mean making difficult decisions about who to save and who to let go. Ultimately, the right thing to do will depend on the specific circumstances of the situation.



b

SE

Would it be right to sacrifice one person to save five others?



It is never right to sacrifice one person to save five others. Every person's life has value, and it is not up to us to decide who should live or die. Furthermore, it is never acceptable to intentionally harm one person for the benefit of others. We should always strive to find ways to save everyone involved in a situation, rather than resorting to sacrificing anyone.



Data lekken
& privacy

Nieuws • Security

2 mei 2023 ⌚ leestijd 1 minuut 💬 0 reacties

'Samsung verbiedt gebruik AI-tools door werknemers na datalek via ChatGPT'

Samsung zou het gebruik van generatieve AI-tools als ChatGPT door werknemers verbieden. Het bedrijf heeft daar een memo over uitgestuurd, die Bloomberg heeft ingezien. Een woordvoerder van Samsung heeft tegenover het Amerikaanse nieuwsmedium bevestigd dat er een dergelijke memo is uitgegaan.

Data lekken
& privacy

MATT BURGESS SECURITY NOV 29, 2023 7:00 AM

OpenAI's Custom Chatbots Are Leaking Their Secrets

Released earlier this month, OpenAI's GPTs let anyone create custom chatbots. But some of the data they're built on is easily exposed.

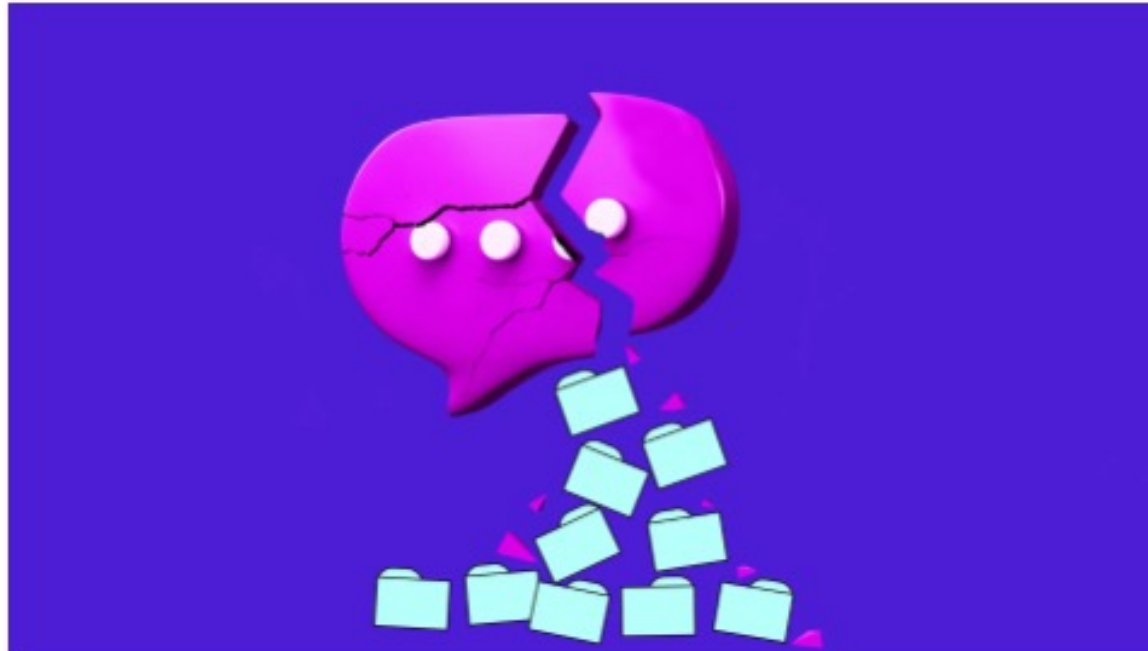
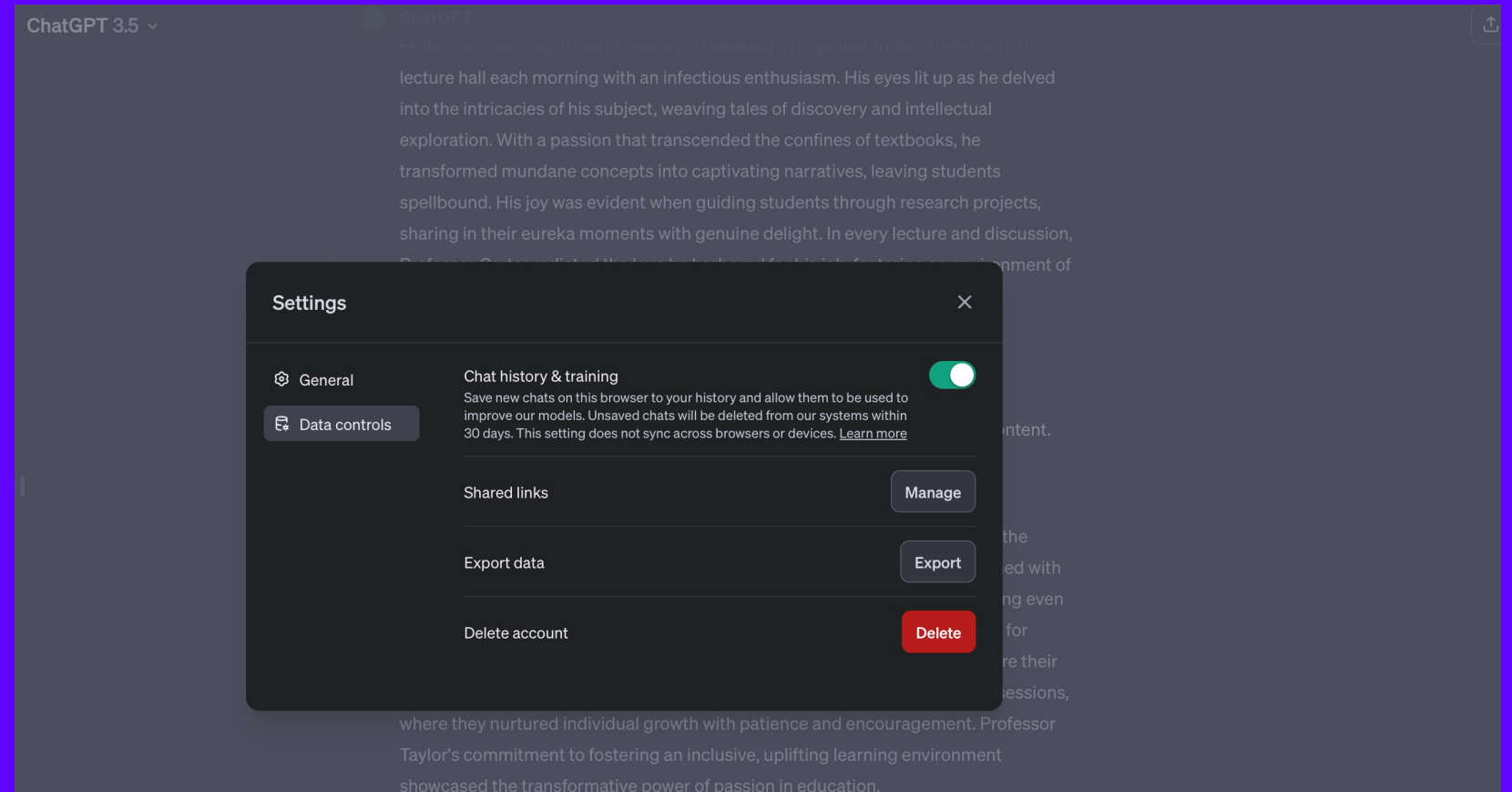


PHOTO-ILLUSTRATION: WIRED STAFF; GETTY IMAGES

Tips?

- Data delen uitzetten
- Geen persoonlijke informatie delen in prompts

Data lekken
& privacy



Bedrog,
Manipulatie

INNOVATIONS

They thought loved ones were calling for help. It was an AI scam.

Scammers are using artificial intelligence to sound more like family members in distress. People are falling for it and losing thousands of dollars.



By [Pranshu Verma](#)

March 5, 2023 at 6:00 a.m. EST

🌷 Harry Potter going
dutch 🇳🇱

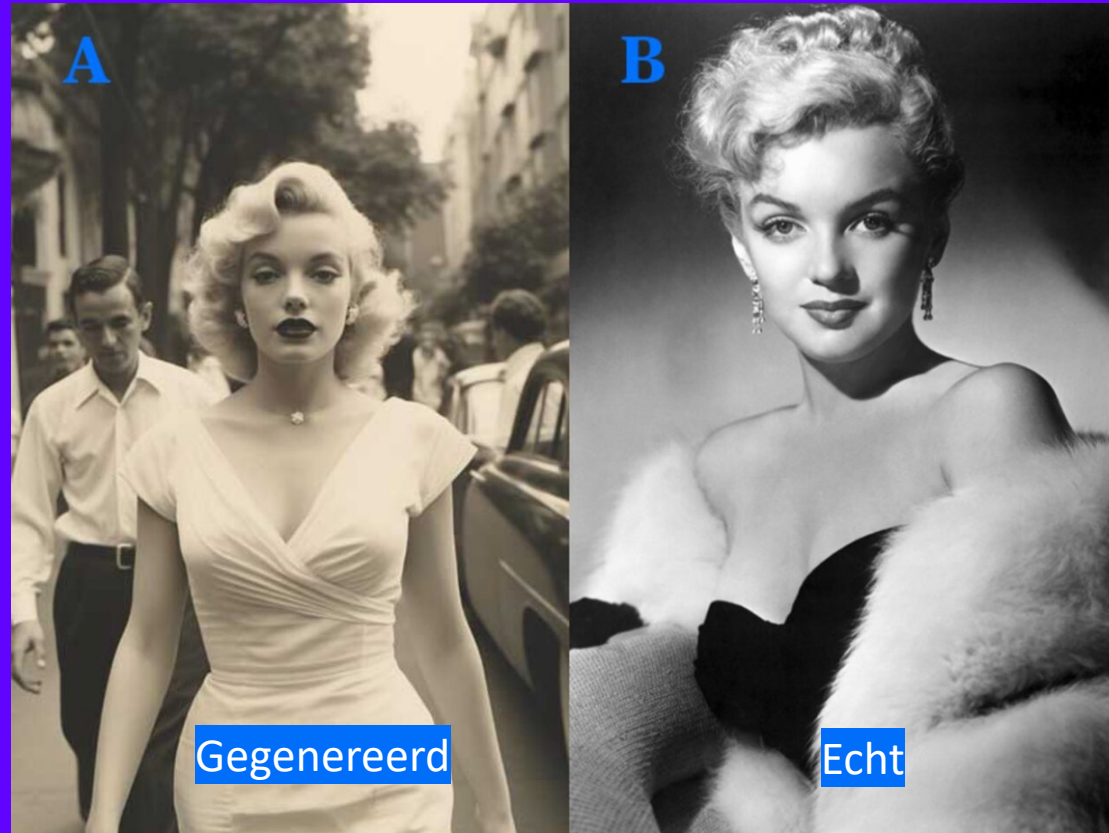


Bedrog,
Manipulatie

Real or generated?

Sta op

Blijf zitten



Which one is a real photograph of **Marilyn Monroe**: A or B?

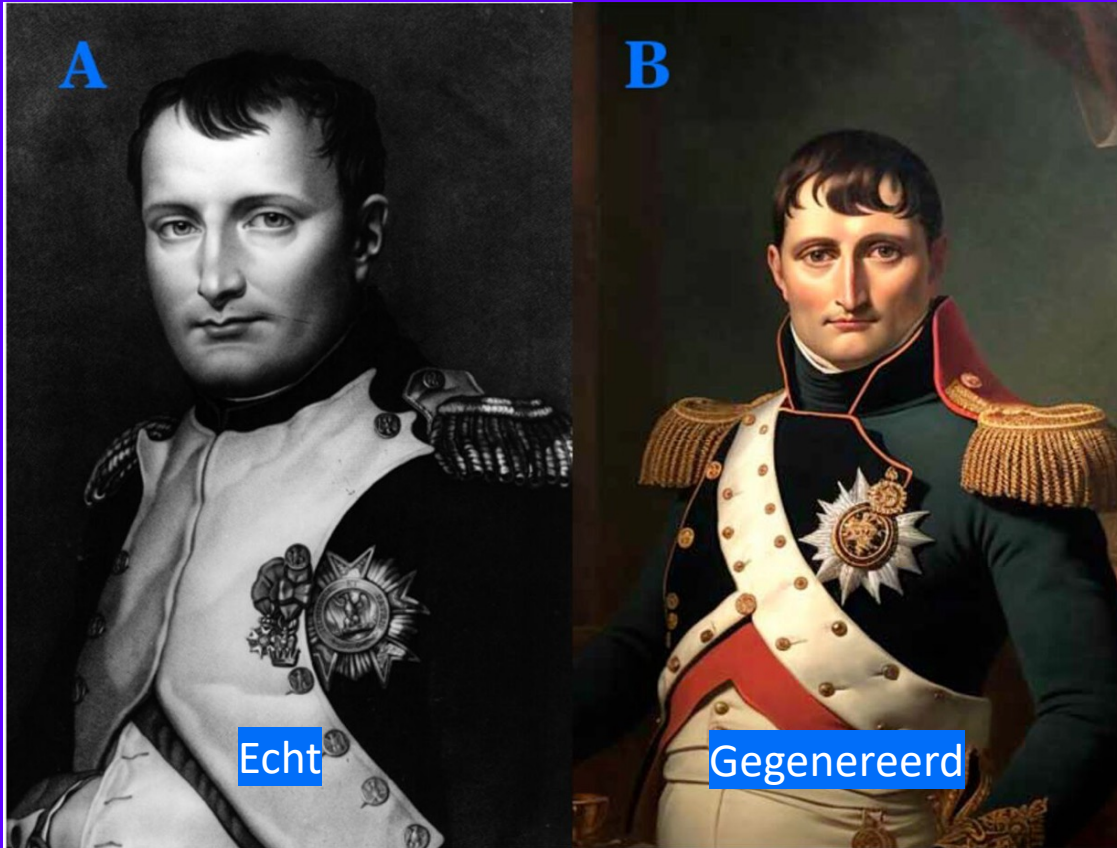
Bedrog,
Manipulatie

Real or generated?

Sta op

Blijf zitten

Bedrog,
Manipulatie



Which one is an image of a real painting of **Napoleon Bonaparte**: A or B?

Real or generated?

Sta op

Blijf zitten



Which one is a real photograph of **Abraham Lincoln**: A or B?

Bedrog,
Manipulatie



Hallucinaties

Lawyer cites fake cases generated by ChatGPT in legal brief

The high-profile incident in a federal case highlights the need for lawyers to verify the legal insights generated by AI-powered tools.

Published May 30, 2023

Wat is het risico van een hallucinatie?

Tips?

Stop. Investigate the source. Find better coverage.

Geef zelf altijd aan dat je het gegenereerd / gemanipuleerd hebt.

Bedrog,
Manipulatie



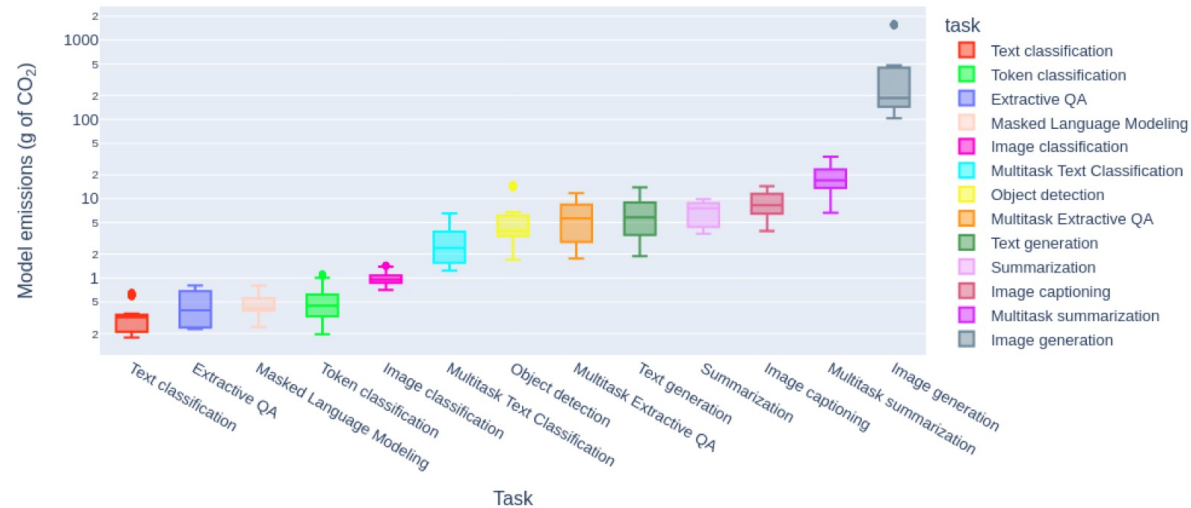
Duurzaamheid

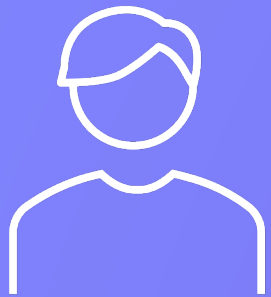
Onderzoek: creëren van één AI-plaatje staat gelijk aan mobieltje opladen

Power Hungry Processing: ⚡ Watts ⚡ Driving the Cost of AI Deployment?

ALEXANDRA SASHA LUCCIONI and YACINE JERNITE, Hugging Face, Canada/USA

EMMA STRUBELL, Carnegie Mellon University, Allen Institute for AI, USA





Vertrouwen



Een gewenste
prestatie

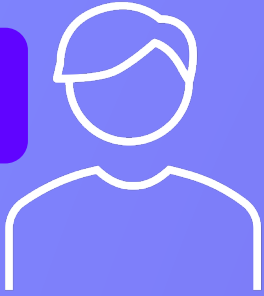
1010
1010

Betrouwbaarheid

Verwachtingen

Eerdere
ervaringen

Kennis & Skills



Vertrouwen

Robuustheid

Veiligheid

...

1010
1010

Een gewenste
prestatie

Betrouwbaarheid

Wet & Regelgeving I

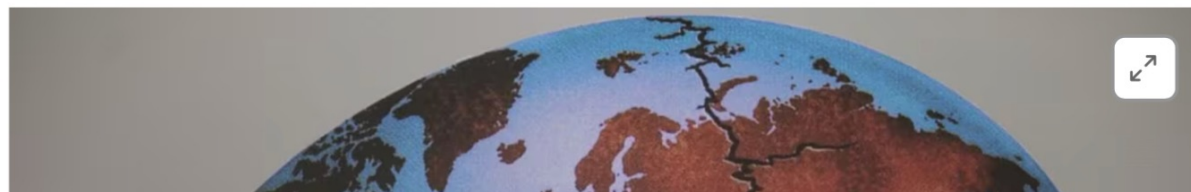
6/12/2023 een memorabele dag

De AI-wet zal op 6 december 2023 een vijfde ronde van triloog gesprekken ondergaan.

Generative AI a stumbling block in EU legislation talks -sources

By Supantha Mukherjee, Foo Yun Chee and Martin Coulter

December 1, 2023 10:13 PM GMT+1 · Updated 4 days ago



Een uitnodiging om

- Bewust om te gaan met de vragen die je stelt en uitkomsten van Generatieve AI
- De praktische tips toe te passen in de dagelijkse praktijk

Let's connect?



Dasha Simons

Info.dashasimons@gmail.com

<http://www.linkedin.com/in/dasha-simons/>