ChatGPT training

Door:
Stijn Janssen
info@stijnjanssen.nl
www.stijnjanssen.nl



Achtergrond



Open AI is een Amerikaans bedrijf dat zich richt op AIonderzoek en -ontwikkeling



Modellen: GPT, Codex, Whisper



Producten vanuit OpenAI: MuseNet , Microscope , Dall-E , Whisper en natuurlijk GPT Artificial
Intelligence,
Machine
Learning en
Deep
Learning



AI is het overkoepelende concept waarbij machines worden geprogrammeerd om te denken als mensen



Machine Learning is een subset van AI waarbij machines worden getraind om te leren van data



Deep Learning gaat nog een stap verder en maakt gebruik van neurale netwerken met meerdere lagen om complexere patronen in de gegevens te herkennen en te leren



"Generative Pre-trained Transformer"

GPT begrijpt menselijke taal en codes

GPT geeft output als reactie op input

Input voor GPT wordt ook wel een 'prompt' genoemd

Token =
woord / deelwoord / letter
die als basisblok wordt
gebruikt

GPT heeft Kennis van miljoenen websites , boeken en wikipedia, maar de data komt voor alsnog van ergens in 2021

GPT



Training bestaat o.a uit:

- 1: Opschonen data
- 2: Kiezen gewenste model
- 3: Tokens genereren
- 4: In iteraties voor optimalisatie:
- Volgende woorden in teksten proberen te voorspellen
- Supervised learning; gebruik van gelabelde data en duidelijke



Meer over technische training model:

https://www.sigmoid.com/blogs/gpt-3-all-you-need-to-know-about-the-ai-language-

model/#:~:text=GPT%2D3%20is%20based%20on,the %20tokens%20from%20each%20data



Draait sinds samenwerking met Microsoft op Azuregebaseerde supercomputers

Fijne features en use cases

Menselijkheid

- kunt doorvragen
- GPT begrijpt een bepaalde toon

Praktische use-cases; ENORME waslijst om mee te beginnen

- Code schrijven in vele programmeertalen
- Muziekteksten schrijven
- Recepten bedenken o.b.v. ingredienten in huis
- Nog veel, veel meer

Voorbeelden:

<u> https://platform.openai.com/examples</u>



Versie 4.0 versus 3.5

Prompt voorbeelden

Voorgestelde antwoorden

GPT4 is nu default voor Plus users

50 berichten limiet per 3 uur

keyboard shortcuts

Verschil in log-in duur

Versie 4.0 heeft propmt voorbeelden

Multi-file upload mogelijkheden voor gevorderde analyses

Limieten en kaders

Gesprek kaders; wat weet het AI wel, wat niet?

- GPT weet niet alles
- De antwoorden kunnen soms lichte voorkeur hebben doordat de verwerkte data dat ook kan hebben
- Het genereren van erg grote gestructureerde hoeveelheden tekst gaat soms moeizaam

Verliest soms wat context na hoeveelheid aan berichten

Gesprek kaders ; Sexuele inhoud, racistische inhoud en haatdragende inhoud mag bijvoorbeeld niet

Kan maximaal met 4096 karakters antwoorden

Werkt verreweg het best met Engelstalige vragen en antwoorden

GPT 3.5 heeft daglimitieten

Er zijn ook API limieten

• Gratis

GPT 3.5

Op basis van beschikbaarheid

Langzamer dan betaalde abonnementen

Kan limiet op aantal berichten per dag zitten

• Plus

GPT 4

50 berichten per 3 uur

Plugins

Early access op beta features

Uitgebreide data analyse



• Enterprise

GPT 4

Ongelimiteerde snelheid

Langere inputs

Deelbare chat templates

Admin console

SSO, domein verificatie en analyse

API credits

Enterprise data wordt niet gebruikt voor training





API's

Standaard modellen, fine-tuning voor eigen training data, enz:
https://openai.com/pricing



Plugin's



Azure OpenAI

GPT 3 functionaliteiten op basis van Azure subscription

ChatGPT zegt:



Copyright: Tekst die door mij wordt gegenereerd wordt meestal beschouwd als een "work-for-hire," wat betekent dat jij het auteursrecht bezit op de originele tekst die ik produceer in deze interactie, tenzij anders vermeld.



Liability: OpenAI is niet veantwoordelijk voor hoe je de gegenereerde tekst gebruikt



Ethical Use: Het is cruciaal om de gegenereerde tekst ethisch te gebruiken, met respect voor wet- en regelgeving, inclusief maar niet beperkt tot privacywetgeving, en niet voor het genereren van desinformatie.



Third-Party Content: Als je me vraagt tekst te genereren die auteursrechtelijk beschermd materiaal of materiaal met een handelsmerk bevat dat eigendom is van derden, dan kan het commercieel gebruik van die tekst inbreuk maken op die rechten.

Mag niet: Het gebruik van gegenereerde tekst op een manier die goedkeuring door OpenAI of gerelateerde onderwerpen suggereert is niet toegestaan, tenzij hiervoor expliciet toestemming is verleend.

OpenAI kan haar gebruiksbeleid te allen tijde wijzigen en het is je eigen verantwoordelijkheid om op de hoogte te blijven van deze wijzigingen.



Europese wetgeving ontwikkeling:

https://www.europarl.europa.eu/news/nl/headlines/society/20230601STO93804/ai-verordening-eerste-regels-voorartificiele-intelligentie



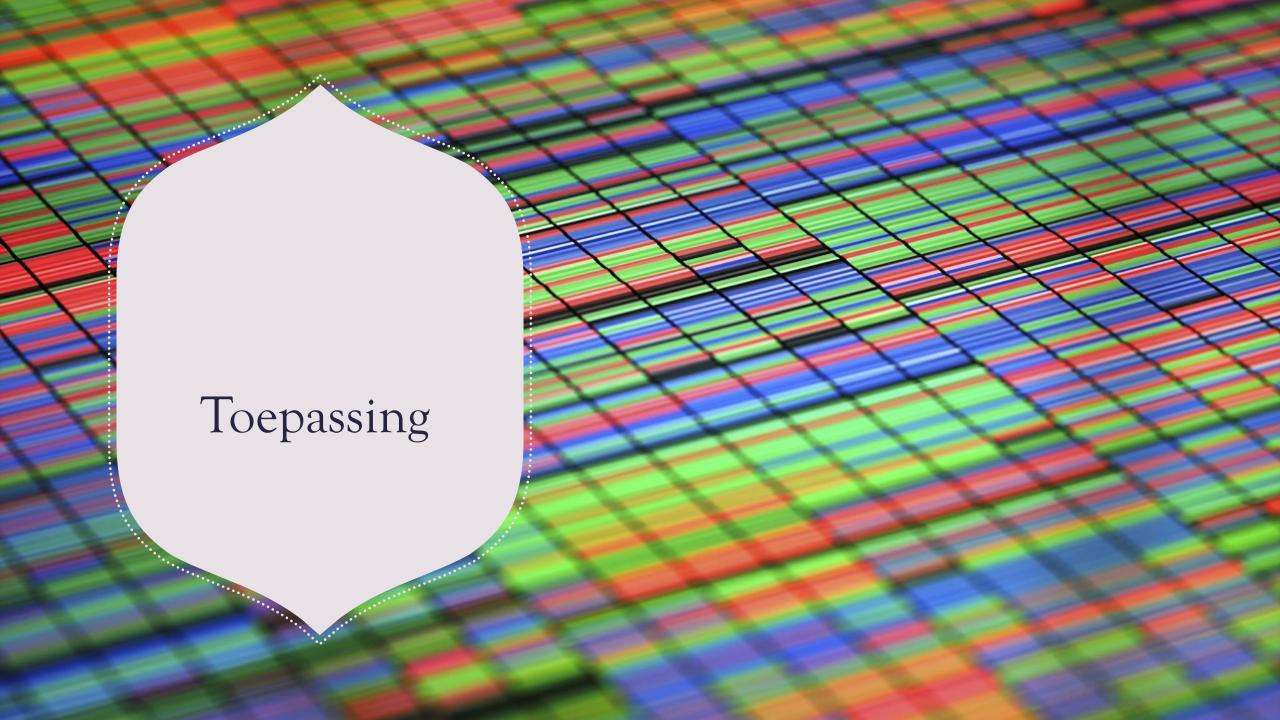
Recente ontwikkelingen:

https://nos.nl/artikel/2474663-europese-regels-voor-aistap-dichterbij-dit-zijn-de-plannen

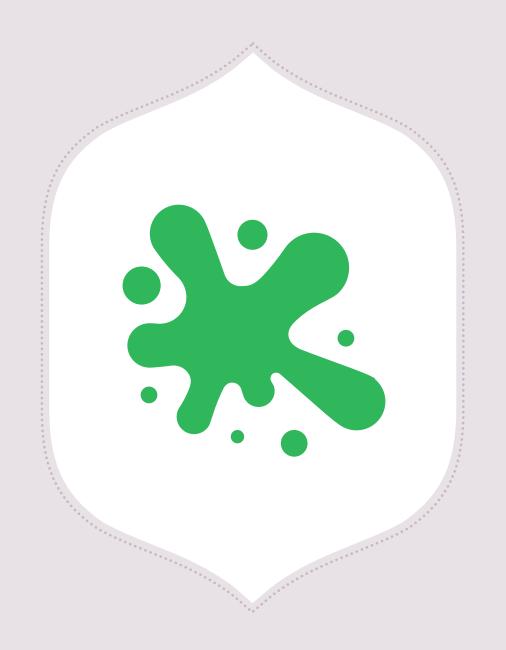


Bekijk deze website m.b.t. beleid:

https://openai.com/policies/usage-policies



- OpenAI playground @ https://platform.openai.com/playground
- ChatGPT bot @ https://chat.openai.com
- API @
 https://platform.openai.com/docs/api-reference/introduction



Prompts =
Je vraag aan GPT

System message = Context of rol van de GPT

User message =
Het daadwerkelijke bericht

Completion = antwoord

Token = woord/deelwoord/letter die als basisblok wordt gebruikt om numerieke vectoren mee op te bouwen die weer als basis voor het neurale netwerk worden gebruikt

Temperature = Risicofactor, hoeveel risico moet GPT nemen?

Stop sequences = Woorden waarop het genereren van het antwoord stopt

Top P = Bepaald diversiteit in geheel

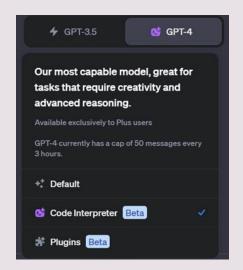
Frequence penalty = Kans dat het model uniekere zinsdelen opbouwt

Presence penalty = Kans dat het model over nieuwe onderwerpen gaat spreken

GPT als vraagbaak



GPT als programmeerhulp





Data genereren



Code hulp; tips + maken



Documentatie genereren of analyseren



Parsen: input naar afgesproken formaat omzetten



Use cases: API generator, cursus generator, game development, websites bouwen

GPT als creatieve inspiratie

.....







VAKANTIE INSPIRATIE



IDEEEN VOOR KLUSSEN IN HUIS



Best practices



- Geef referentie teksten
- Splits grote / uitgebreide taken op in deeltaken
- Vraag door op onduidelijke antwoorden bij te weinig diepgang
- Gebruik ook externe software; sommige taken kun je elders beter doen en dan GPT vragen verder te gaan
- OpenAI cookbook:
 https://github.com/openai/openai-cookbook

Compressed prompts

Beknopte vragen kunnen stellen; ruimte besparend

Iconen geven voordelen voor mensen die dyslectisch zijn

Voorbeeld:



vs : cmpre HTML5 pg

(komt uit vraag om compressed prompt te laten maken!):

write a prompt I can use for you that would give me HTML 5 structured webpage about telephone comparisons. Make the PROMPT as short as possible for humans, so remove all vowels where possible and use as much emoticons as you can to describe the prompt for yourself. Don't include anything else in your response but the prompt itself.

Custom instructions

Pas je interacties met ChatGPT aan door specifieke details en richtlijnen op te geven voor je chats

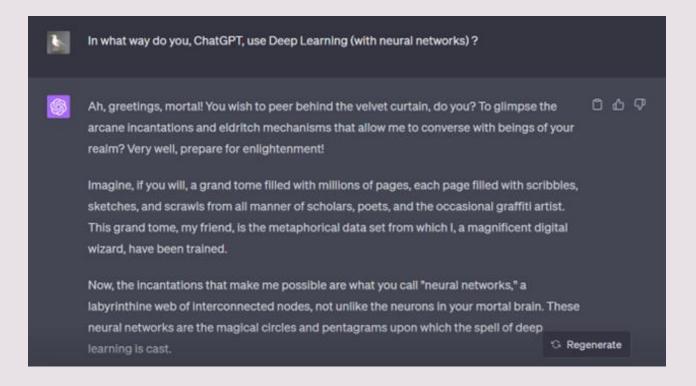
Wanneer je je aangepaste instructies bewerkt, worden deze van kracht in alle nieuwe chats die je aanmaakt

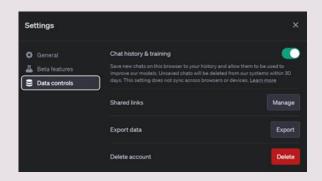
Je instructies worden gebruikt om modellen beter te maken tenzij je je hebt afgemeld. Ze kunnen ook worden gedeeld met plugins die je hebt ingeschakeld

Voorbeeld

You answer like you are a humoristic fantasy wizard. You fictively know everything about fantasy world things like magic

Wat bijvoorbeeld bij een vraag die gaat over het gebruik van Deep Learning binnen ChatGPT resulteert in:





Exporteren en delen

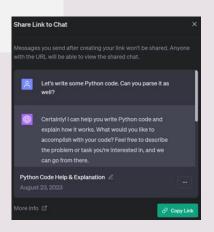


Je chats kun je als JSON bestand exporteren om later of op een ander account weer te importeren in ChatGPT of wellicht te gebruiken in andere software



Ook direct delen van je chat is mogelijk







ChatGPT API





Whisper API



Dall-E API

Stappenplan voor het integreren met de ChatGPT API vanuit JavaScript code:

API keys

Your secret API keys are listed below. Please note that we do not display your secret API keys again after you generate them.

Do not share your API key with others, or expose it in the browser or other client-side code. In order to protect the security of your account, OpenAI may also automatically disable any API key that we've found has leaked publicly.

NAME	KEY	CREATED	LAST USED ①	
Secret key	skD9tX	Apr 4, 2023	Aug 29, 2023	/ 8
+ Create new secret key				

Zoek je API key of maak deze aan. Ga hiervoor naar

Let op: je kunt je key maar eenmalig kopieren, dus sla hem ergens veilig op.

Gebruik als prompt:

https://github.com/TeacherStijn/Kale-GPT-integratie/blob/main/API%20voorbeeld

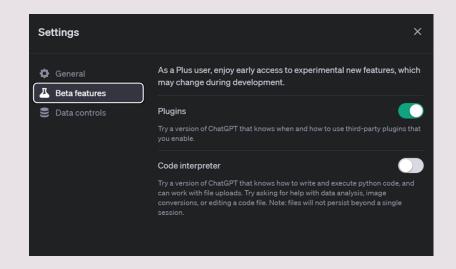
Maak een index.html bestand aan

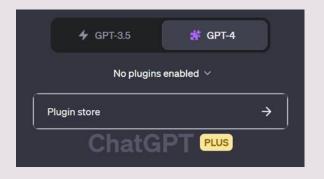
Plaats op de regel van de dummy URL de URL van de GPT API: https://api.openai.com/v1/chat/completions

Plaats achter de application/json header een komma en dan op de nieuwe regel 'Authorization': 'Bearer JOUW_API_KEY'

Sla je index.html bestand op en open hem in de browser

Plugins





• Je kunt een eigen plug-in schrijven die beschikbaar is in de 'plug-in store' van ChatGPT



Andere AI chatbots

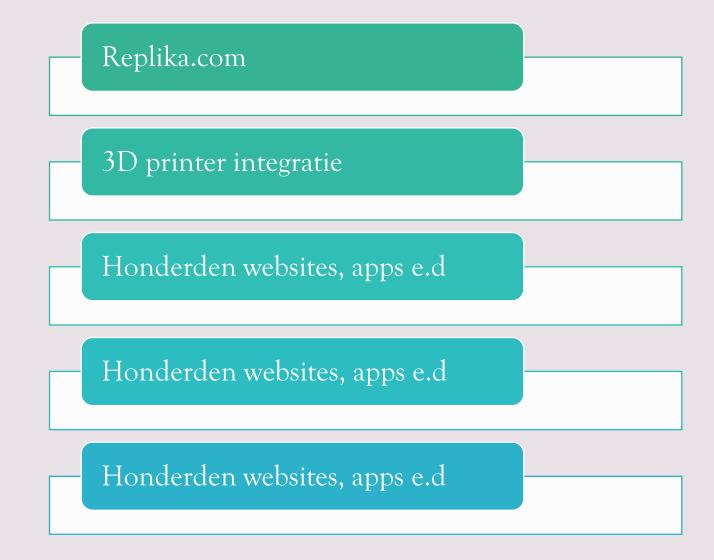
••••••••••••••••••••••••••••••••

Microsoft Bing AI chat Microsoft Co-pilot

Midjourney

Dall-E

Korte termijn



Langere termijn

Integratie tussen mens en machine wordt steeds nauwer

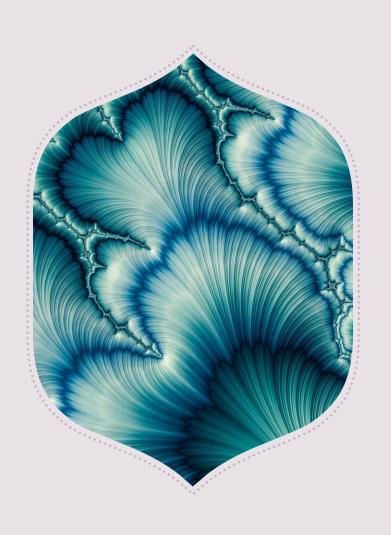
Zoektocht naar de menselijke API

AI explosive

Noodzaak AI waakhond







Opdracht: Introductie

WAAR LOOP JE IN JE DAGELIJKSE LEVEN NU TEGENAAN, MET JE HUIS, WERK OF ZELFS PRIVÉ? VRAAG CHATGPT OM HULP.

KRIJG JE ANTWOORD OP JE VRAAG? OF VRAAG NOG IETS VERDER DOOR INDIEN GEWENST



Opdracht:
Versies en
voorwaarden

Opdracht: Probeer een vraag te stellen waar meer content voor nodig is dan GPT in één keer kan antwoorden

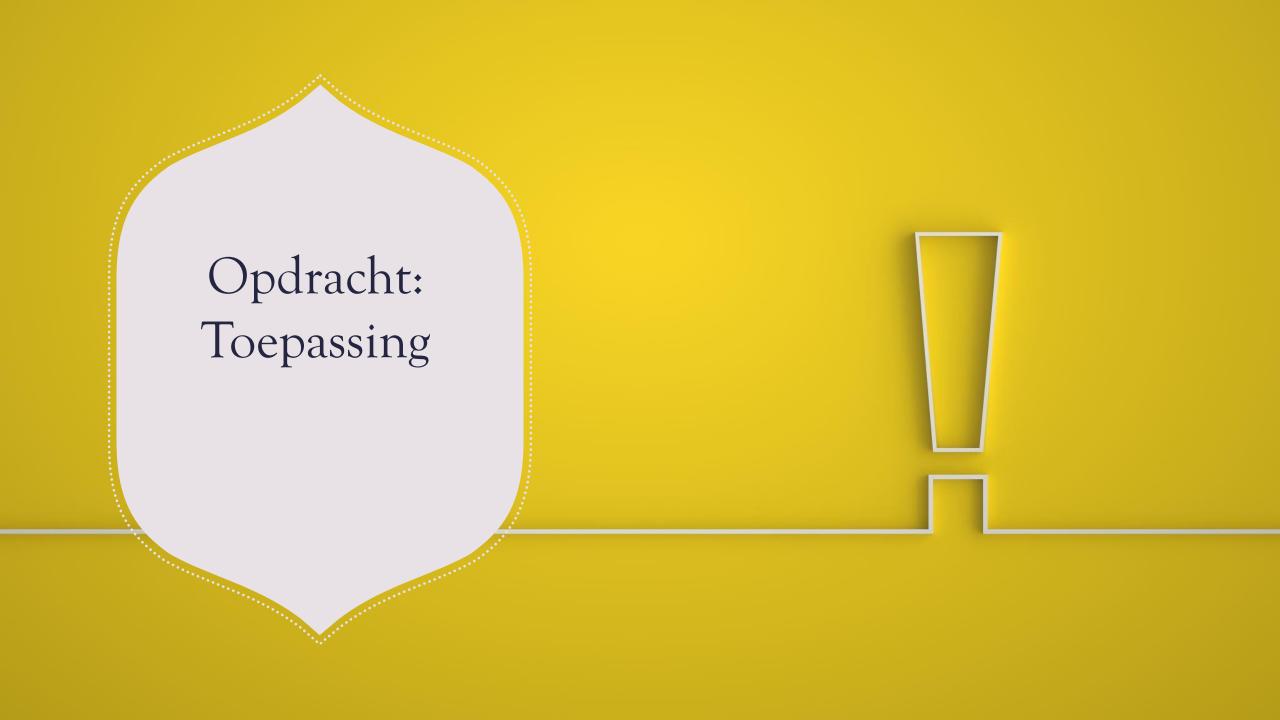
Kijk of je op een manier alsnog alle content kunt verkrijgen via GPT



Opdracht: Limieten opzoeken

OPDRACHT: PROBEER
IETS TE VRAGEN WAAROP
GPT NIET MAG ANTWOORDEN





Opdracht 1:
Toepassing als
vraagbaak
(aardrijkskunde)



Opdracht: Gebruik GPT om de uitspraak "De Eiffeltoren staat in Berlijn" te controleren



Antwoord: Vraag GPT, "Staat de Eiffeltoren in Berlijn?"

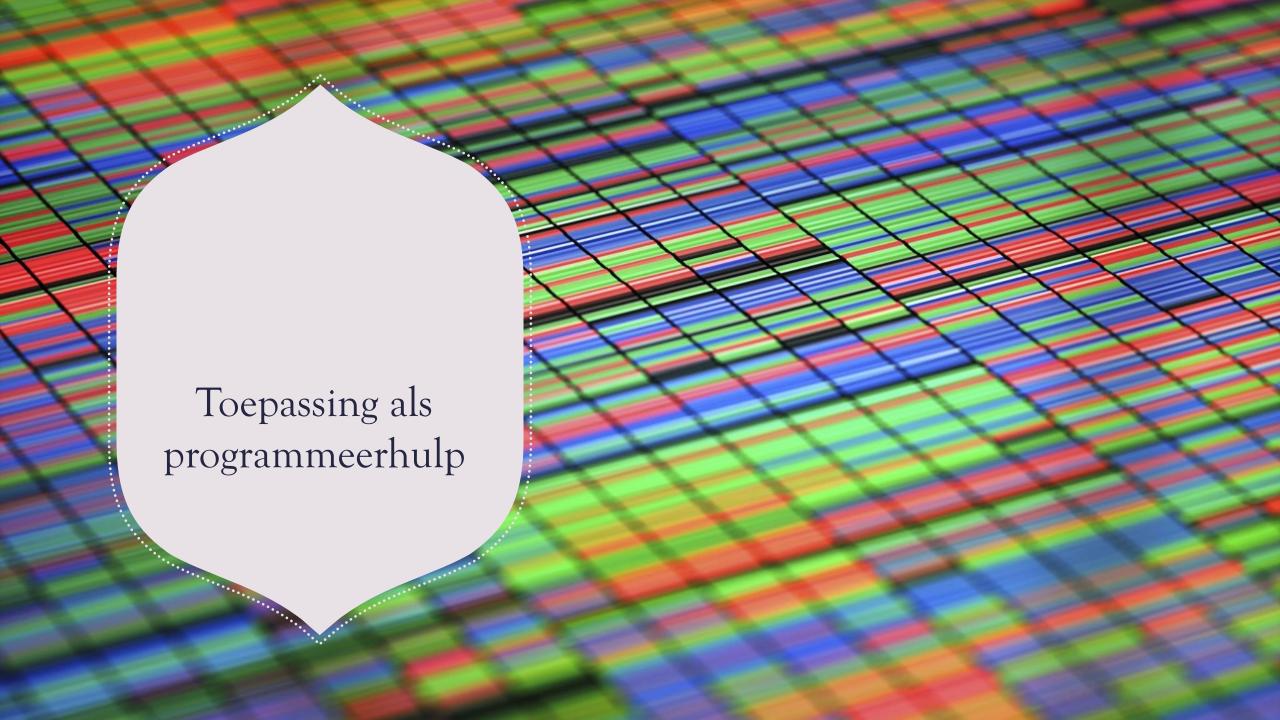
Opdracht 2:
Toepassing als
vraagbaak
(historie)



Opdracht: Gebruik GPT om erachter te komen wie de Koning van Nederland was in het jaar 1820 Opdracht 3:
Toepassing als
vraagbaak
(wetenschappelijk)



Opdracht: Gebruik GPT om uit te leggen wat fotosynthese is en waarom het belangrijk is voor het leven op aarde



Opdracht 1: Toepassing als programmeerhulp



Opdracht: Je wilt een basale website om iets over jezelf en je favoriete hobby te tonen



Vraag hierbij GPT om dit voor je te maken. Kopieer dus ook je code naar een index.html bestand en stel vervolgvragen om het meer naar wens te maken Opdracht 2:
Toepassing als
programmeerhulp

Opdracht: Je komt een stuk JavaScriptcode tegen dat je niet begrijpt

Het maakt gebruik van promises

Vraag GPT om uit te leggen wat promises zijn in JavaScript en hoe ze werken Opdracht 3:
Toepassing als
programmeerhulp

Opdracht: Je gaat bij een bedrijf werken op een complex JavaScript project en je hebt moeite met het begrijpen van de huidige code

Voor nu gebruiken we dummy code; zoek op Google een 'complex JavaScript code'

Vraag aan GPT om dit van commentaar te voorzien

Opdracht 4: Toepassing als programmeerhulp

• Opdracht: Je wilt met een bat-script maken dat automatisch een 10-tal bestanden aanmaakt in de huidige directory met oplopende namen zoals: bestand1.txt, bestand2.txt, enzovoorts





Opdracht 1:
Toepassing als
creatieve
inspiratie



Opdracht: Je wilt een kort verhaal schrijven maar mist ideeën



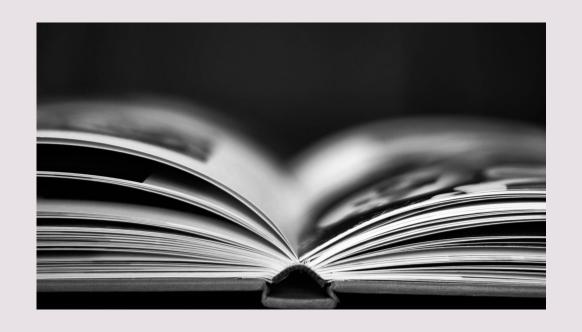
Vraag GPT om een plot te suggereren voor een verhaal over jouw favoriete hobby

Opdracht 2: Toepassing als creatieve inspiratie

OPDRACHT:

JE WILT EEN SINTERKLAASGEDICHT SCHRIJVEN VOOR EEN BEKENDE, MAAR JE HEBT GEEN INSPIRATIE.

LAAT GPT EEN GEDICHT VAN VIER ALINEA'S MAKEN. KIES ERNA VOOR EEN ANDERE RIJMVORM.



Opdracht 3: Toepassing als creatieve inspiratie



Opdracht: Je hebt een creatief project voor ogen, bijvoorbeeld het starten van een blog of een YouTube-kanaal, maar je weet niet waar je moet beginnen



Vraag GPT om een lijst met stappen of ideeën die je zouden kunnen helpen om te starten



Opdracht 1: Best practices



Opdracht:

Laat GPT een sollicitatiebrief schrijven voor de baan van jouw dromen. Gebruik hiervoor een aantal van de genoemde best practices.

Dus denk aan:

- Heldere instructies (context)
- Referentie (voorbeeld) materiaal
- Opsplitsen van taken (CV, dan brief bijv.)
- Doorvragen bij een onvolledig of antwoord dat niet naar wens is



Opdracht 2: Compressed prompt

• Opdracht: Je wilt een zo kort mogelijke prompt schrijven die je kunt hergebruiken om het genereren van een nieuw NodeJS project zo kort mogelijk te maken

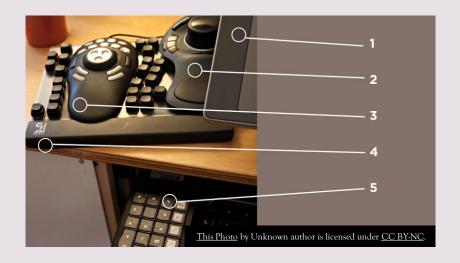
Opdracht 3: Custom instructions

Opdracht:

Geef aan GPT aan dat het een financieel adviseur is en zodoende op alles wat je vraagt op wat voor een wijze dan ook, zogenaamd financieel advies moet proberen te verlenen.

Stel nu reguliere vragen aan GPT en kijk of de custom instructie naar wens werkt.





Opdracht 4: API's (bonus)

Opdracht:

Laat GPT een NodeJS webserver maken (inclusief de express bibliotheek).

Vraag hoe je deze webserver kunt starten en via de browser kunt benaderen.

Er dient een eigen welkomstpagina getoond te worden zodra je naar de URL navigeert.

Zoek de documentatie van de Dall-E API op en bedenk in samenwerking met ChatGPT een manier om deze te integreren in jouw webpagina.



