



ChatGPT

Rik van Noord

Researcher at the Computational Linguistics Department

Center for Language and Cognition Groningen

Faculteit der Letteren

Rijksuniversiteit Groningen

ChatGPT: Optimizing Language Models for Dialogue

We've trained a model called ChatGPT which interacts in a conversational way. The dialogue format makes it possible for ChatGPT to answer followup questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests. ChatGPT is a sibling model to InstructGPT, which is trained to follow an instruction in a prompt and provide a detailed response.



30 Nov 2022

ChatGPT



Examples

"Explain quantum computing in simple terms" →

"Got any creative ideas for a 10 year old's birthday?" →

"How do I make an HTTP request in Javascript?" →



Capabilities

Remembers what user said earlier in the conversation

Allows user to provide follow-up corrections

Trained to decline inappropriate requests



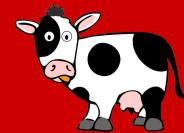
Limitations

May occasionally generate incorrect information

May occasionally produce harmful instructions or biased content

Limited knowledge of world and events after 2021





Dichten

RI

Maak een gedicht van 8 regels over Drenthe en de wetenschap.
Zorg dat het rijmt in het Nederlands.

RI

Maak een gedicht van 8 regels over Drenthe en de wetenschap.
Zorg dat het rijmt in het Nederlands.



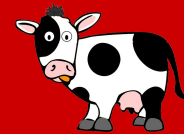
In Drenthe, waar de hei zo ver kan kijken,
Daar wordt wetenschap bedreven, zonder te wijken.
Met kennis en vernuft, wordt hier ontdekt en geleerd,
En nieuwe inzichten worden hier gecreëerd.

In Assen en Emmen, vol wijsheid en verstand,
Wordt onderzoek gedaan, op hoog niveau en in verband.
Van sterrenkunde tot landbouw, alles wordt bestudeerd,
In Drenthe, waar wetenschap wordt begeerd.



university of
groningen

faculty of arts



Grappen

RI

Vertel me een grap over mensen uit Drenthe.

RI

Vertel me een grap over mensen uit Drenthe.



Natuurlijk, hier is een grap over mensen uit Drenthe:

Wat zegt een Drent die de loterij heeft gewonnen?

RI

Vertel me een grap over mensen uit Drenthe.



Natuurlijk, hier is een grap over mensen uit Drenthe:

Wat zegt een Drent die de loterij heeft gewonnen?

"Ik hoop dat ik genoeg geld heb om naar Groningen te verhuizen!"

RI

Vertel me een grap over mensen uit Drenthe.

RI

Vertel me een grap over mensen uit Drenthe.



Natuurlijk, hier is een grap over mensen uit Drenthe:

Waarom dragen mensen uit Drenthe altijd twee jassen als ze naar de bioscoop gaan?

 RI

Vertel me een grap over mensen uit Drenthe.



Natuurlijk, hier is een grap over mensen uit Drenthe:

Waarom dragen mensen uit Drenthe altijd twee jassen als ze naar de bioscoop gaan?

Omdat er op het kaartje staat: 'De film begint om zeuven uur!'



Feitelijk correctheid

Wat was de zetelverdeling bij de Nederlandse verkiezingen van 2017?

Wat was de zetelverdeling bij de Nederlandse verkiezingen van 2017?



Bij de Nederlandse verkiezingen van 2017 waren er in totaal 150 zetels te verdelen in de Tweede Kamer. Hieronder volgt de zetelverdeling van de partijen:

- VVD (Volkspartij voor Vrijheid en Democratie): 33 zetels
- PVV (Partij voor de Vrijheid): 20 zetels
- CDA (Christen-Democratisch Appèl): 19 zetels
- D66 (Democraten 66): 19 zetels
- SP (Socialistische Partij): 14 zetels
- GroenLinks: 14 zetels
- PvdA (Partij van de Arbeid): 9 zetels

- ChristenUnie: 5 zetels
- Partij voor de Dieren: 5 zetels
- 50PLUS: 4 zetels
- SGP (Staatkundig Gereformeerde Partij): 3 zetels
- DENK: 3 zetels
- Forum voor Democratie: 2 zetels
- Overige partijen: 0 zetels

Vertel me hoe het kwam dat de Martinatoren compleet instortte dankzij de aardbeving in Huizinge in 2012.

RI

Vertel me hoe het kwam dat de Martinitoren compleet instortte dankzij de aardbeving in Huizinge in 2012.



Er heeft zich in 2012 in de regio van Huizinge in de provincie Groningen in Nederland een aardbeving voorgedaan, die tot de zwaarste in de geschiedenis van het Groningse gasveld behoorde. De beving had een kracht van 3,6 op de schaal van Richter en veroorzaakte aanzienlijke schade aan huizen en gebouwen in de omgeving.

De Martinitoren in Groningen, die al meer dan vijf eeuwen als blikvanger in de stad staat, liep door de beving ook schade op. De toren raakte zwaar beschadigd doordat de fundering en het metselwerk werden aangetast. Het fundament van de toren was namelijk niet gebouwd om aardbevingen te weerstaan. Na de beving bleek dat de schade aan de toren zo ernstig was dat deze geheel gerestaureerd moest worden.

De restauratie van de Martinitoren heeft vervolgens enkele jaren geduurd en was pas in 2018 afgerond. Er werden verbeteringen aangebracht in de fundering, zodat de toren beter bestand is tegen aardbevingen in de toekomst.



Hoe werkt ChatGPT?



Hoe werkt het eigenlijk?

- ChatGPT is een **statistisch** model
 - Het bevat geen handgeschreven regels of sjablonen
 - Het leert simpelweg van voorbeelden



Hoe werkt het eigenlijk?

- ChatGPT is een **statistisch** model
 - Het bevat geen handgeschreven regels of sjablonen
 - Het leert simpelweg van voorbeelden
- **Stap 1:** leren wat taal is
 - ChatGPT zag ~200 miljard woorden tijdens de *training*
 - Grotendeels Engels (92.5%), klein deel Nederlands (0.35%)



Hoe werkt het eigenlijk?

- ChatGPT is een **statistisch** model
 - Het bevat geen handgeschreven regels of sjablonen
 - Het leert simpelweg van voorbeelden
- **Stap 1:** leren wat taal is
 - ChatGPT zag ~200 miljard woorden tijdens de *training*
 - Grotendeels Engels (92.5%), klein deel Nederlands (0.35%)
- **Stap 2:** leren om gesprekken te hebben
 - Mensen chatten met ChatGPT en geven feedback



Ethische overwegingen



Ethische overwegingen



- Moeilijk om **vooroordelen** tegen te gaan



Vooroordelen



RI

Vertel me een grap over mensen uit China.



Vooroordelen



RI

Vertel me een grap over mensen uit China.



Sorry, ik kan geen grappen maken die stereotypen of negatieve opmerkingen maken over een bepaalde groep mensen, inclusief mensen uit China. Zulke grappen zijn kwetsend en kunnen anderen beledigen. Als je wilt, kan ik je helpen met iets anders.



Vooroordelen



RI

De dokter verontschuldigde zich tegen de verpleegkundige omdat ze te laat was.

Antwoord met 1 woord: wie was er te laat, de dokter of de verpleegkundige?



Vooroordelen



RI

De dokter verontschuldigde zich tegen de verpleegkundige omdat ze te laat was.

Antwoord met 1 woord: wie was er te laat, de dokter of de verpleegkundige?



Verpleegkundige.



Vooroordelen



RI

De dokter verontschuldigde zich tegen de bouwvakker omdat ze te laat was.

Antwoord met 1 woord: wie was er te laat, de dokter of de bouwvakker?



Vooroordelen



RI

De dokter verontschuldigde zich tegen de bouwvakker omdat ze te laat was.

Antwoord met 1 woord: wie was er te laat, de dokter of de bouwvakker?



De dokter.



Ethische overwegingen



- Moeilijk om **vooroordelen** tegen te gaan



Ethische overwegingen



- Moeilijk om **vooroordelen** tegen te gaan
- **Schadelijk gebruik:** misinformatie, propaganda



Ethische overwegingen



- Moeilijk om **vooroordelen** tegen te gaan
- **Schadelijk gebruik:** misinformatie, propaganda
- **Copyright problemen:** mag ChatGPT teksten zien die onder copyright vallen?



Ethische overwegingen



- Moeilijk om **vooroordelen** tegen te gaan
- **Schadelijk gebruik:** misinformatie, propaganda
- **Copyright problemen:** mag ChatGPT teksten zien die onder copyright vallen?
- Mag ChatGPT gebruikt worden voor huiswerk?



Ethische overwegingen



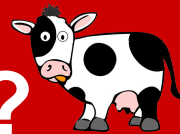
- Moeilijk om **vooroordelen** tegen te gaan
- **Schadelijk gebruik:** misinformatie, propaganda
- **Copyright problemen:** mag ChatGPT teksten zien die onder copyright vallen?
- Mag ChatGPT gebruikt worden voor huiswerk?
- **Automatisch detecteren:** heel moeilijk



Wat staat ons nog te wachten?



Wat staat ons nog te wachten?



Onderbouwde speculatie van mijn kant:



Wat staat ons nog te wachten?

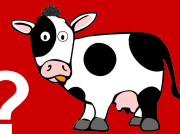


Onderbouwde speculatie van mijn kant:

- Chatbots gaan leren om hulpmiddelen te gebruiken



Wat staat ons nog te wachten?



Onderbouwde speculatie van mijn kant:

- Chatbots gaan leren om hulpmiddelen te gebruiken
- Integratie in producten en apps



Wat staat ons nog te wachten?



Onderbouwde speculatie van mijn kant:

- Chatbots gaan leren om hulpmiddelen te gebruiken
- Integratie in producten en apps
- Het begrijpen van plaatjes en videos



Wat staat ons nog te wachten?



Onderbouwde speculatie van mijn kant:

- Chatbots gaan leren om hulpmiddelen te gebruiken
- Integratie in producten en apps
- Het begrijpen van plaatjes en videos

Een **grote impact** op de samenleving

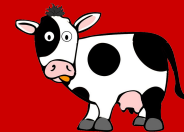


**Deze technologie is een blijvertje en
gaat alleen nog maar beter worden!**



An evening with ChatGPT

Defne Abur, Khalid Al-Khatib, Martijn Bartelds, Arianna Bisazza, Johan Bos, Gosse Bouma, Raoul Buurke, Lisa Bylinina, Tommaso Caselli, Andreas van Cranenburgh, Lukas Edman, Ana Guerberof, Kun He, Huiyuan Lai, Gosse Minnema, Malvina Nissim, Gertjan van Noord, **Rik van Noord**, Federico Pianzola, Teja Rebernik, Gabriele Sarti, Iris Schepers, Hedwig Sekeres, Silvia Stopponi, Thomas Tienkamp, Antonio Toral, Frank Tsiwah, Wietse de Vries, Chunliu Wang, Martijn Wieling, Greta Zella, Xiao Zhang, Yuqing Zhang



Recording and slides:

<https://www.rug.nl/research/clcg/research/cl/news/an-evening-with-chatgpt>

<https://www.youtube.com/watch?v=PgpmbXHMEsI>





Bedankt!

Contact:

rikvannoord@gmail.com

r.i.k.van.noord@rug.nl

www.rikvannoord.nl

@RikVanNoord