

PROGRAMMA

12.30-13.00u Onthaal

13.00-13.05u

Opening: KANTL-voorzitter **Leen van Dijck**

13.05-13.30u

MOETEN WE BANG ZIJN VOOR
FILOSOFISCHE ZOMBIES?

Walter Daelemans (Universiteit Antwerpen)

13.30-14.00u

DE TEKSTMACHINE EN DE TAALTHEORIE

Dirk Geeraerts (KU Leuven)

14.00-14.30u

BLINKENDE KNOPKE FIJN, DAT NIMMER VAN
SCHRIJVEN ZIJT MOE? OVER HET CREATIEVE
VERMOGEN VAN GROTE TAALMODELLEN

Tim Van de Cruys (KU Leuven)

14.30-15.00u

Koffiepauze

15.00-15.45u

VERANTWOORD INZETTEN VAN GENAI IN
ACADEMISCH SCHRIJFONDERWIJS: EEN
ILLUSIE?

Lieve De Wachter (KU Leuven)

15.45-16.15u

INTELLIGENTE TECHNOLOGIE: HOE STAAT
HET MET NEDERLANDS?

Frieda Steurs (Instituut voor de Nederlandse
Taal)

16.15-16.45u

DE GEBRUIKSVOORWAARDEN VAN TAAL

Juliet van Rosendaal (KU Leuven)

16.45-17.30u

Publieksdiscussie op basis van enkele
stellingen (moderator: Walter Daelemans)

17.30-18.30u

Receptie

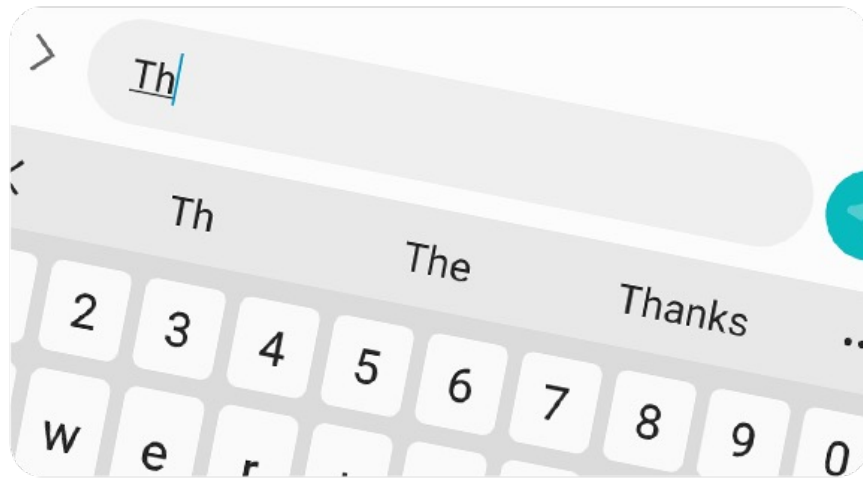


Moeten we bang zijn voor ‘filosofische zombies’?

Walter Daelemans

CLiPS

Universiteit Antwerpen



Taal en waarschijnlijkheid



Claude Shannon
Informatietheorie



Andrey Markov
Markovketens



In het leerproces

Opgelet: taalmodellen gebruiken tokens; 100 tokens = 75 woorden

					Moeten	we
				Moeten	we	bang
			Moeten	we	bang	zijn
		Moeten	we	bang	zijn	voor
	Moeten	we	bang	zijn	voor	filosofische
Moeten	we	bang	zijn	voor	filosofische	zombies

Bij gebruik (autoregressie)

De	oorspronkelijke	bewoners	waren	niet	bang	
oorspronkelijke	bewoners	waren	niet	bang	voor	?

voor	0.32
dat	0.13
wanneer	0.03
tafel	0.00
Gent	0.00
...	

Hoe kan dit werken?

>

WA

Vul de volgende zin aan met 1 woord: "De energieprijzen in België zijn in 2023 fors gedaald. De koopkracht van de Vlaming is aanzienlijk"

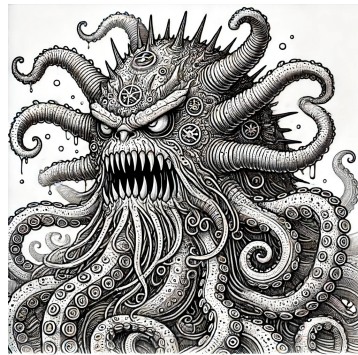


"De energieprijzen in België zijn in 2023 fors gedaald. De koopkracht van de Vlaming is aanzienlijk toegenomen."

Dataset	Sampling prop.	Epochs	Disk size
CommonCrawl	67.0%	1.10	3.3 TB
C4	15.0%	1.06	783 GB
Github	4.5%	0.64	328 GB
Wikipedia	4.5%	2.45	83 GB
Books	4.5%	2.23	85 GB
ArXiv	2.5%	1.06	92 GB
StackExchange	2.0%	1.03	78 GB

Voortraining

Miljarden woorden
internettext



PROMPT



AANVULLING

Dataproblemen:

Copyright
Bias
Onvoorspelbaarheid
Moeizame prompting
“Hallucinaties”

Alignmentproblemen:

Alignering naar welke ideologie?

Welk ethisch kader?

Misbruik

Onderbetaalde medewerkers

Taalverandering (?)

Voortraining

Miljarden woorden
internettekst

Alignering

Verfijning door menselijke feedback

PROMPT



AANVULLING

Verfijning met eigen data

Voortgezette voortraining met
domeindata of taakdata
Bijv. medisch

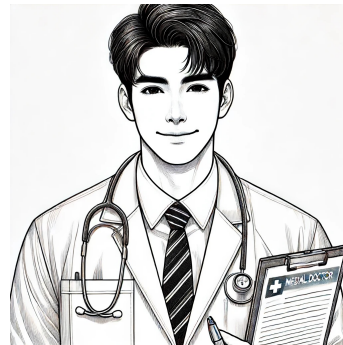
Voortraining

Miljarden woorden
internettekst

Alignering

Verfijning door menselijke feedback

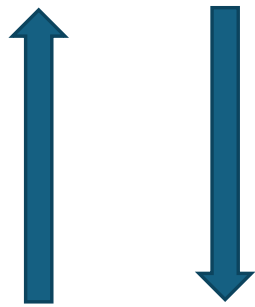
PROMPT



AANVULLING

RAG: Retrieval-Augmented Generation

Kennisbank
Gebruikersdata

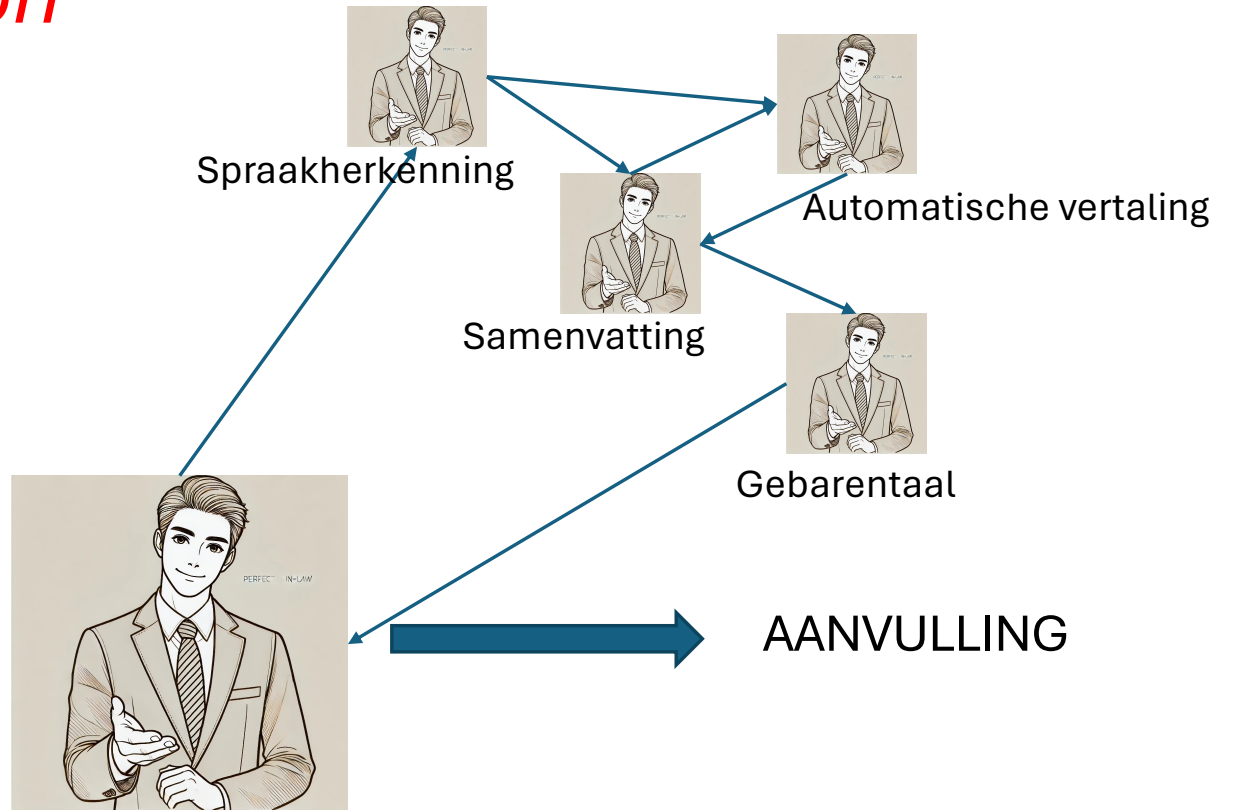


PROMPT



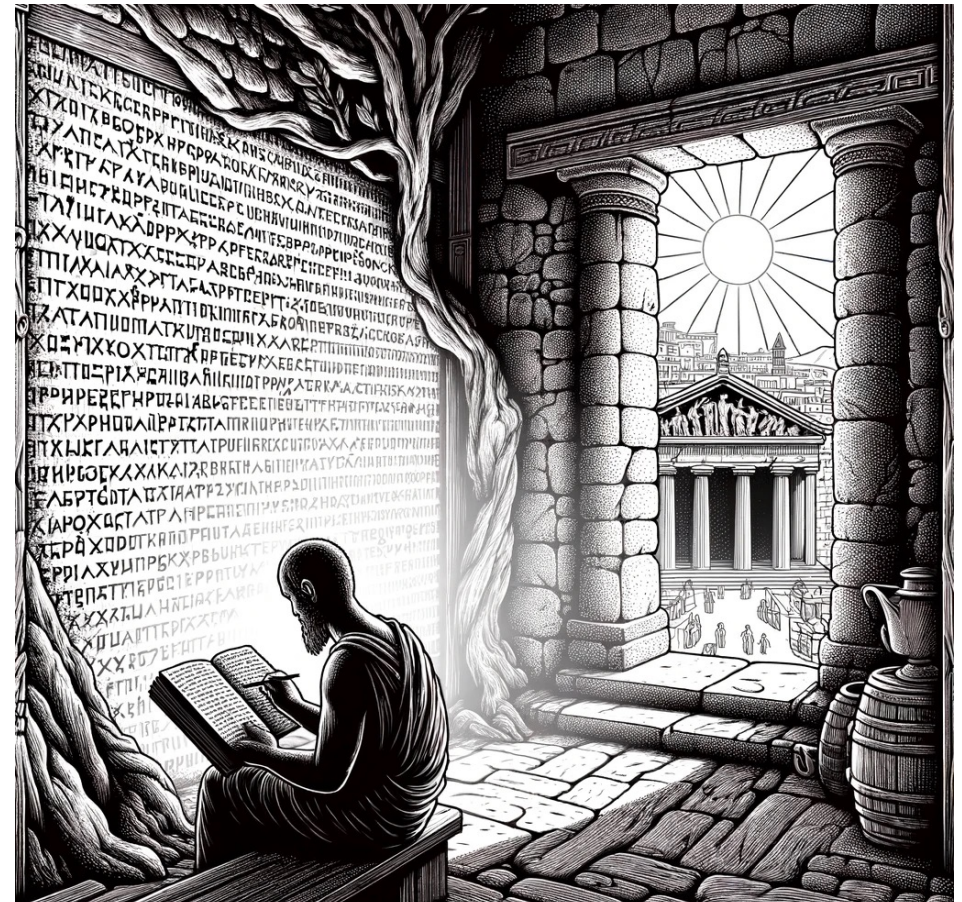
In context leren

Agentive AI



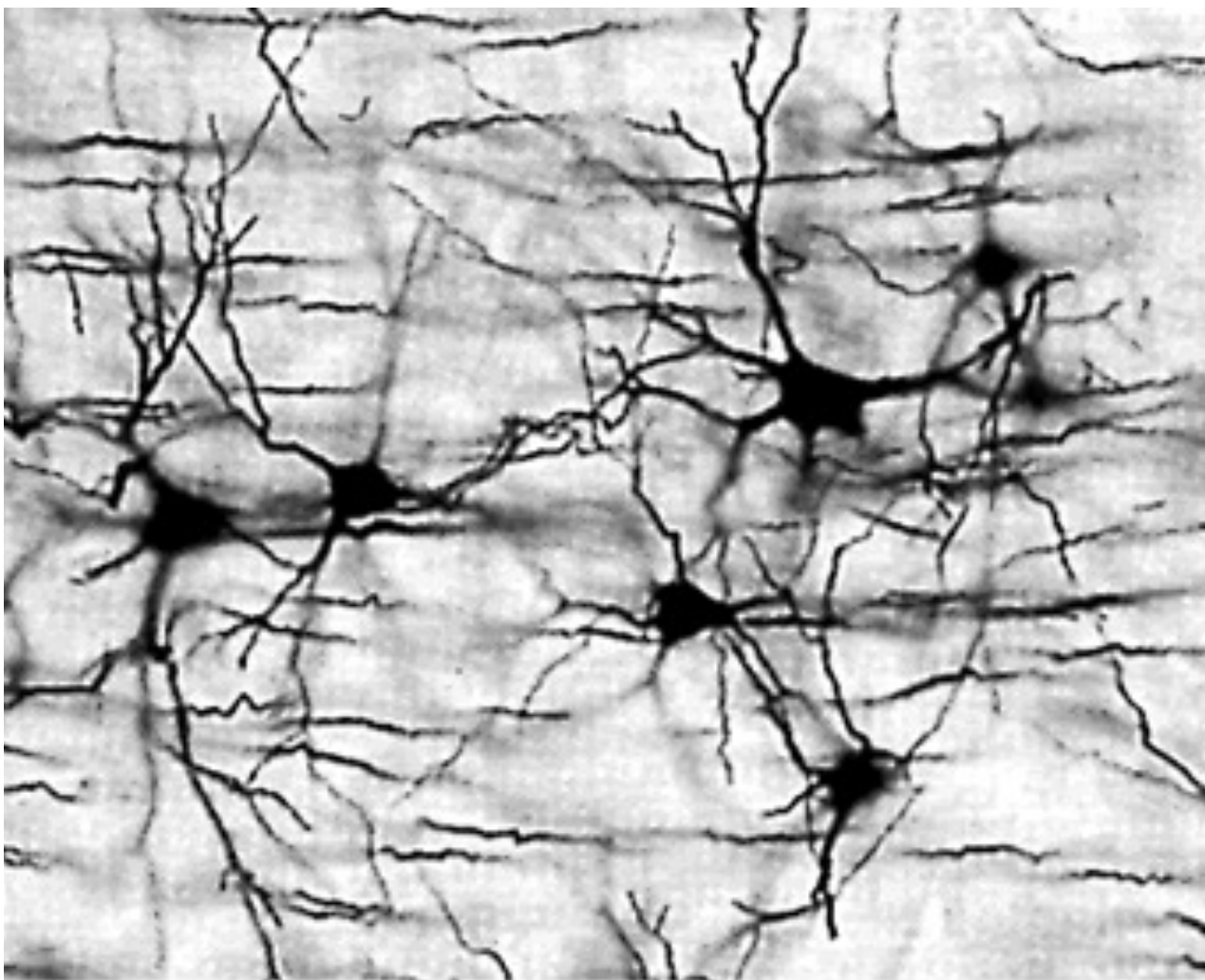
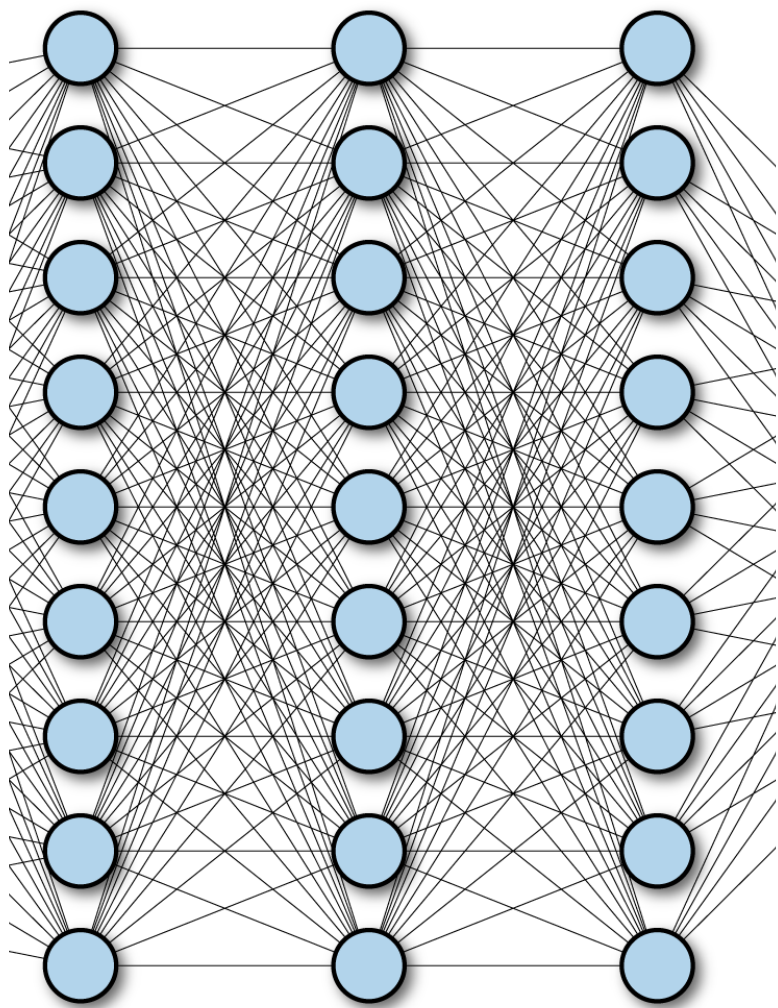
Is (heel veel) tekst voldoende voor intelligentie?

- Patronen (van patronen (van patronen ...))
- Correcte tekstproductie
 - Taalkundig (vorm, betekenis, structuur, stijl)
 - Wereldkennis en eenvoudig redeneren
- Emergente competenties:
 - Samenvatten, vertalen, vragen beantwoorden, dialoog, programmeren, ...
- (Nog niet) waargenomen:
 - Intentionaliteit, (zelf)bewustzijn, complex redeneren, reflectie, opinies, vrije wil ...



Gevolgen

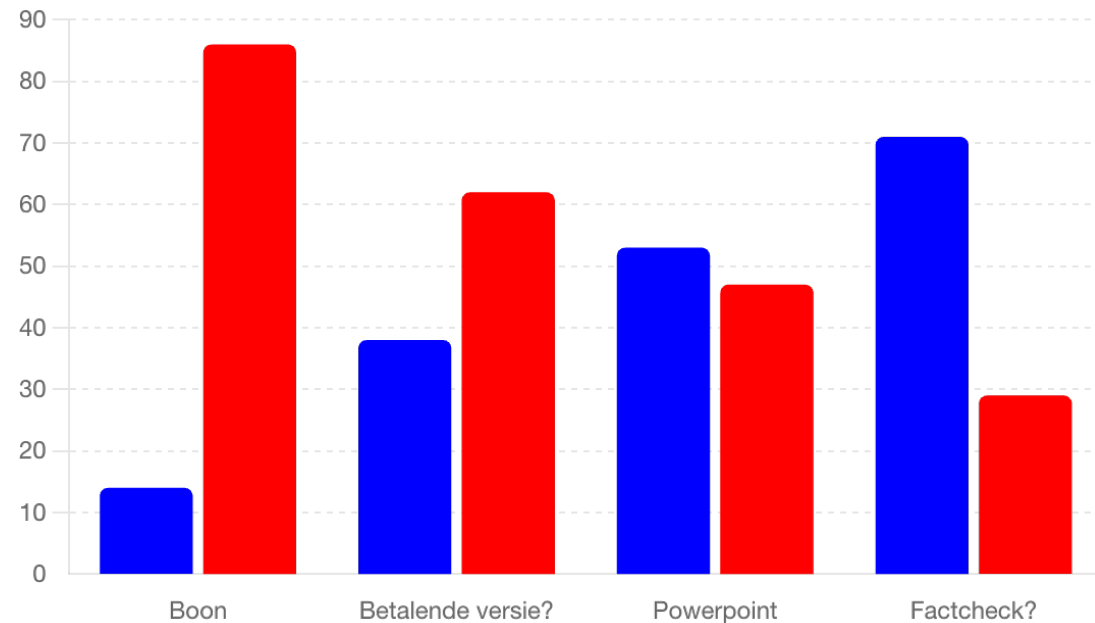
- Voor de taaltheorie
 - Bestaansbewijs dat taal kan geleerd worden uit data
- Voor het Nederlands
 - De Nederlands taalvaardigheid van bestaande taalmodellen is vaak onvoldoende
- Voor de creatieve auteurs
 - Een taalmodel kan creatief zijn (ongewild)
- Voor het onderwijs
 - Hebben scripties, papers en huiswerk nog wel zin?
 - Verplichte melding van gebruik is zinloos door integratie in platformen
- Voor de cultuur
 - Lezen en schrijven worden een hobby
 - Veranderen taalmodellen onze taal en opinies?
- Voor de economie
 - Werkgelegenheid, duurzaamheid



Publieksdiscussie

Instagram poll

- Wint ChatGPT ooit De Boon?
- Gebruiken jullie liever de uitgebreidere, Engelstalige ChatGPT?
- ChatGPT mag mij helpen met mijn PowerPointpresentatie maar niet met mijn masterscriptie?
- Factcheck jij de antwoorden die ChatGPT geeft?



Enkele stellingen

- ChatGPT is de eerste echte schrijfmachine. Maar daardoor staat het tot schrijfvaardigheid in dezelfde ambigue relatie als een rekenmachine tot wiskundige kennis: behulpzaam maar ook afstompend. [DG]
- Moet Vlaanderen aansluiten bij het initiatief van GPT-NL, en moeten we er naar streven om terug een soort gemeenschappelijk platform te creëren zoals met de STEVINcall? [FS]
- Wie is aansprakelijk als het misloopt met GenAI? De maker van het taalmodel of degene die het gebruikte? [WD]