



# **PROGRAMMA AI & ETHIEK**

## Workshop ChatGPT

# Over ons

- Hogeschoolbreed programma
- Multidisciplinair team
- Doel: studenten voorbereiden op baan van de 'toekomst'
  - AI kunnen benutten
  - AI kunnen beheersen
  - AI kunnen besturen



# **PROMPT**

Een prompt is een natuurlijke taal instructie gericht aan een "bevroren" voorgetraind taalmodel met als doel om de werking ervan te verbeteren of te verfijnen.

<https://platform.openai.com/playground?mode=chat&model=gpt-4-0314>

# AI's NEW FRONTIER is called Chat-GPT

## Context & Doelen

Leer hoe ChatGPT betrouwbaar te gebruiken

1. Begrijpen wat ChatGPT wel en niet kan
2. Ethische overwegingen bediscussiëren
3. Effectieve prompts schrijven
4. Vervolgvrragen schrijven
5. Waarschuwing betrouwbaarheid
6. Overzicht van veel gestelde vragen [[ChatGPT FACs](#)].



Om duiding te geven aan de hype rond de generatieve AI ChatGPT, is deze *Nederlandstalige "How-To?" repository* opgesteld met [hints](#) en [tips](#) voor het verantwoord & effectief gebruik ervan in het hoger onderwijs.

Doordat de toepassingsmogelijkheden van ChatGPT eindeloos zijn, worden in deze repository voornamelijk gebruiksmogelijkheden besproken die relevant zijn voor (1) het leerproces van scholieren/studenten, in combinatie met (2) onderwijs-/onderzoektaken van docenten in het (hoger)onderwijs.

Daarnaast worden de randvoorwaarden ---*inclusief beoordelingskader + richtlijnen*--- en risicofactoren beschreven voor het verantwoord gebruik van [Foundation-models](#) zoals ChatGPT als onderdeel van **BKE** (Basis Kwalificatie Examineren) en **SKE** (Senior Kwalificatie Examineren) assessment in het hoger onderwijs.

Een centrale vraag is: "[Vormen Foundation models een betrouwbare tool die docenten kan helpen bij het beoordelen van content gemaakt door studenten?](#)"

Er zullen regelmatig updates volgen over nieuwe ontwikkelingen.

Disclaimer: deze tekst is door het gebruik van "gezond verstand" tot stand gekomen.  
Artificiële intelligentie [AI] is gebruikt ter verificatie van de gebruikte bronnen + vertaling van Engelstalige teksten.

Dit is een data product gemaakt door het [PROMETHEUS DATA SCIENCE LAB](#)  
van de Hogeschool Rotterdam.



Stel je een computer voor die jouw zinnen kan afmaken met een betere zinswending; of een gesprek met je kan voeren over een thema dat jou interesseert; of een probleem direct kan oplossen door honderden



## Overzicht van veel gestelde vragen [ChatGPT FACs]

*Disclaimer: deze tekst is door het gebruik van gezond verstand tot stand gekomen.  
Artificiële intelligentie is gebruikt ter verificatie van de gebruikte bronnen.*

- [0] Waarom veroorzaakt ChatGPT zoveel onrust?
- [0a] Wat is generatieve kunstmatige intelligentie [Gen-AI]?
- [0b] Is er een kort overzicht van ChatGPT's tekortkomingen?
- [0c] +Moet ik me zorgen maken over ChatGPT technologie?
- [0d] +Is er een ChatGPT "code-of-conduct" en/of richtlijn voor hbo docenten??
- [0e] +Kan ChatGPT benut worden als beoordelingsinstrument?
- [0f] +Voldoen Gen-AI —zoals ChatGPT— aan Europese AI-Regelgeving?
- [1a] Wat moet je weten over ChatGPT en wat kant deze "chatBot"?
- [1b] Wat zijn de functionele mogelijkheden & —Cyber Security— beperkingen van ChatGPT?
- [1c] Wat zijn ethische risico's & schaduwkanten van ChatGPT?
- [1d] Maakt ChatGPT "valsspelen" makkelijker en is het te detecteren?
- [1e] Kun je ChatGPT opvoeren als co-auteur?
- [1f] +Kun je ChatGPT citeren als bron?
- [2] Hoe geef je een opdracht aan ChatGPT?
- [3] Kan ChatGPT uitleggen hoe het werkt?
- [4] Kan ChatGPT uitleggen hoe het te gebruiken?
- [5] Heeft ChatGPT taalbegrip?
- [6] Kan ChatGPT logisch redeneren?
- [7a] +Kan ChatGPT broncode schrijven?
- [7b] Kan ChatGPT broncode uitleggen?
- [7c] Kan ChatGPT broncode output simuleren?
- [7d] Kan ChatGPT een "Deep Learning" lessenserieks bedenken?
- [7e] Kan ChatGPT broncode beoordelen en/of fouten opsporen?
- [8a] Kan ChatGPT gebruikt worden om bronnen te vermelden?
- [8b] Kan ChatGPT gebruikt worden om bronnen samen te vatten?
- [13] Kun je spreken tegen ChatGPT?
- [15] Prompt "patterns" voorbeelden
- [16] Geraadpleegde Bronnen

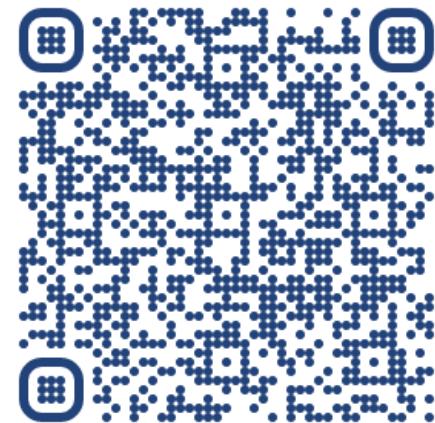
## AI's NEW FRONTIER is called Chat-GPT

Dit is een dataproduct gemaakt door het PROMETHEUS DATA SCIENCE LAB van de Hogeschool Rotterdam.

### Context & Doel

#### Leer hoe ChatGPT betrouwbaar te gebruiken

1. Begrijpen wat ChatGPT wel en niet kan
2. Ethische overwegingen bediscussiëren
3. Effectieve prompts schrijven
4. Vervolgvragen schrijven
5. Waarschuwing betrouwbaarheid
6. Overzicht van veel gestelde vragen [ChatGPT FACs].



Om duiding te geven aan de hype rond de generatieve-AI ChatGPT, is deze *Nederlandstalige "How-To?" repository* opgesteld met *hints en tips* voor het verantwoord & effectief gebruik ervan in het hoger onderwijs.

Doordat de toepassingsmogelijkheden van ChatGPT eindeloos zijn, worden in deze repository voornamelijk gebruiksmogelijkheden besproken die relevant zijn voor (1) het leerproces van scholieren/studenten, in combinatie met (2) onderwijs-/onderzoektaken van docenten in het (hoger)onderwijs.

Daarnaast worden de randvoorwaarden —*inclusief beoordelingskader + richtlijnen*— en risicofactoren beschreven voor het verantwoord gebruik van Foundation-models zoals ChatGPT als onderdeel van BKE (Basis Kwalificatie Examineren) en SKE (Senior Kwalificatie Examineren) assessment in het hoger onderwijs.

Een centrale vraag is: *"Vormen Foundation modellen een betrouwbare tool die docenten kan helpen bij het beoordelen van content gemaakt door studenten?"*

Er zullen regelmatig updates volgen over nieuwe ontwikkelingen.





HOGESCHOOL  
ROTTERDAM

# AI-gestuurd leren via het schrijven van “PROMPT-RECEPTEN”

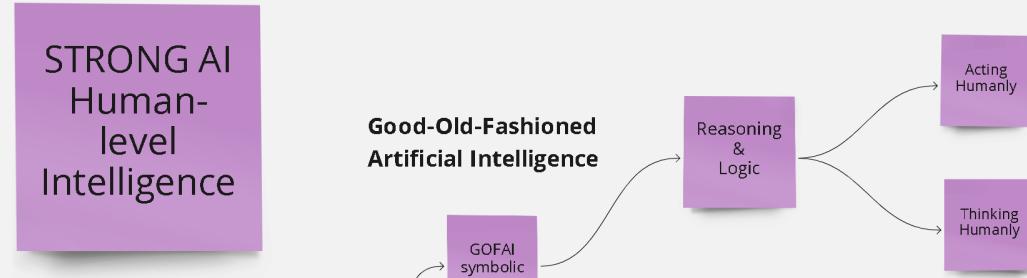
*Kans of bedreiging voor ICTO coaches*



Tech-lead  
Rob van der Willigen

# AI-taxonomie is complex

Representations are localized and knowledgeable to humans



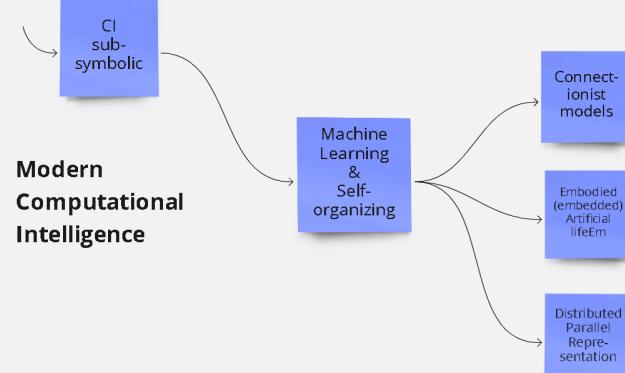
AI Capability Levels

Rule driven

---

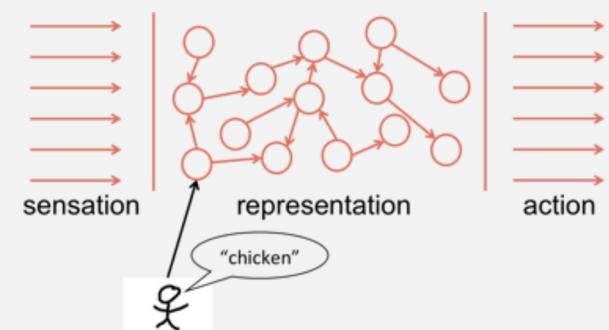
Data driven

WEAK AI (narrow) single purpose Intelligence



Representations are distributed not readable for humans

According to research at Cambridge University, it doesn't matter in what order the letters in a word are, the only important thing is that the first and last letter be at the right place. The rest can be a total mess and you can still read it without problem. This is because the human mind does not read every letter by itself, but the word as a whole.



*natural language processing [nlp]*

“natuurlijke taal  
verwerking”

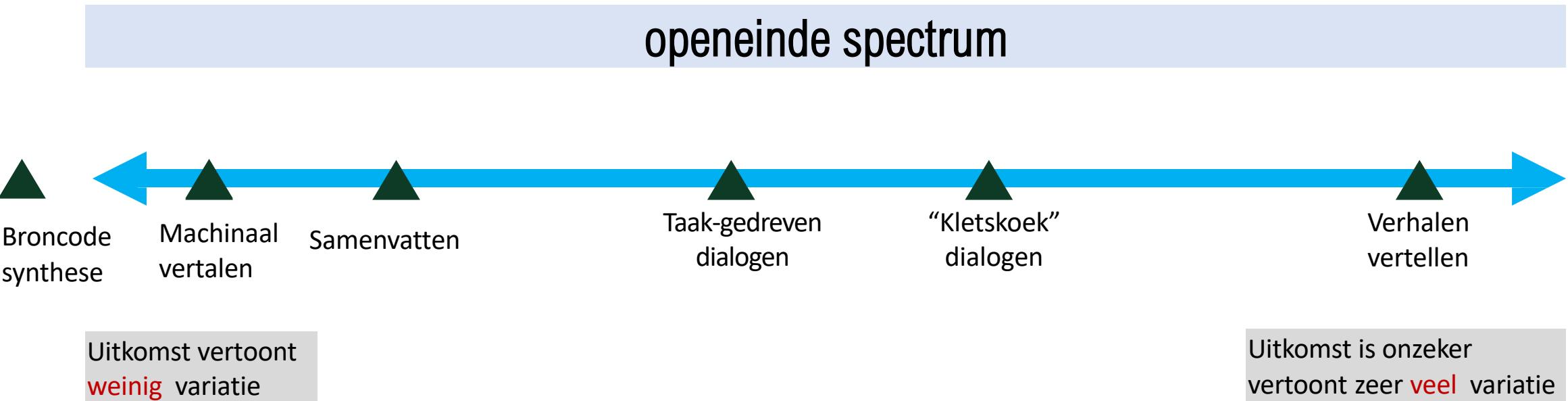
---*Natural Language Processing {NLP}*---

# natuurlijke taalgeneratie

*begrijpen (NLU) + generatie (NLG)*

**NLP** richt zich op het ontwerpen van **taal modellen** die vloeiende, coherent en betekenisvolle taal produceren voor menselijke consumptie.

# Categorisering NLG-taken



# State-of-the-art NLG

[Nurse] Hi Mr.#name#, you were discharged on #date#. There are some questions i'd like to check with you.

[Patient] Ok, Ok.

[Nurse] Well, have you been experiencing swelling recently?

[Patient] Swelling? It comes and go, comes and go.

[Nurse] Comes and go ... I see .. #repetition#

[Nurse] ... #pause#... When did it start?

[Patient] Let me see, started from three weeks ago.

... ...

[Nurse] Are you experiencing any headache right now as we speak?

[Patient] Umm ... #back-channel#

[Nurse] Let me check, the last time you told me is sometimes at night.

[Patient] Oh, right, only a bit.

... ...

[Nurse] Still feel some chest pain or chest discomfort?

[Patient] Yes, my head is... #false-start# no, the pain is much better.

Still feel headache though ... #topic-drift#

... ...

[Nurse] Any giddiness or palpitation?

[Patient] Palpitation? Do not have-- #interruption#

[Nurse] Well ... Do you-- #interruption#

[Patient] and no giddiness, no, nothing.

... ...

[Nurse] Ok, you need to check your heartrate everyday.

[Nurse] Do you know how to use the device?

[Patient] Yes, yes, no problem.

... ...

Swelling: started from three weeks ago, comes and go.

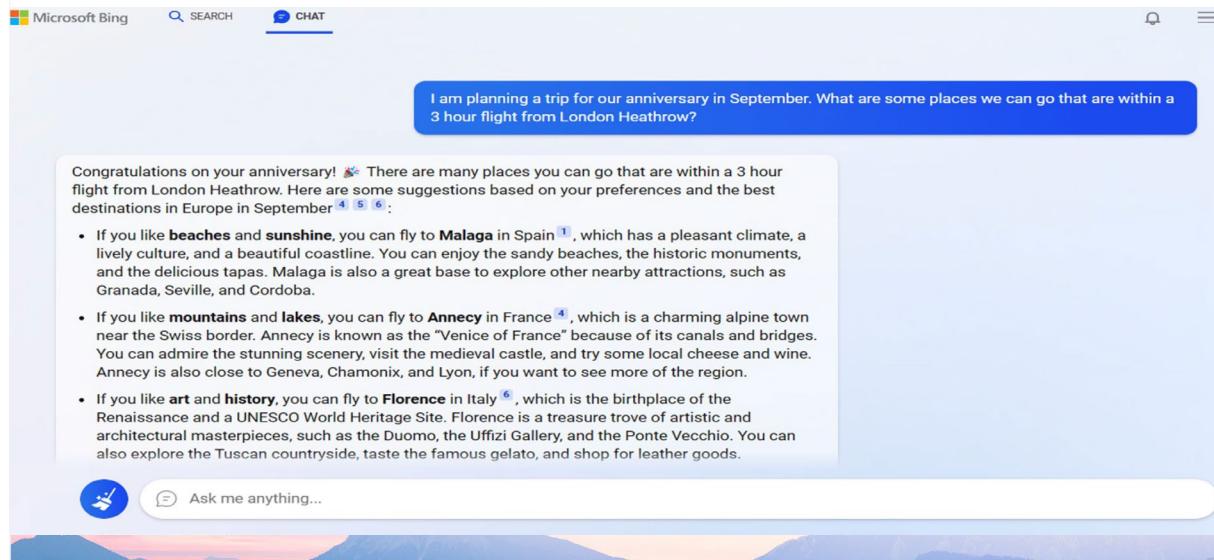
Headache: sometimes, at night, only a bit.

Chest pain: much better.

Dizziness: none.

# State-of-the-art NLG

Reinventing search with a new AI-powered Microsoft Bing and Edge, your copilot for the web



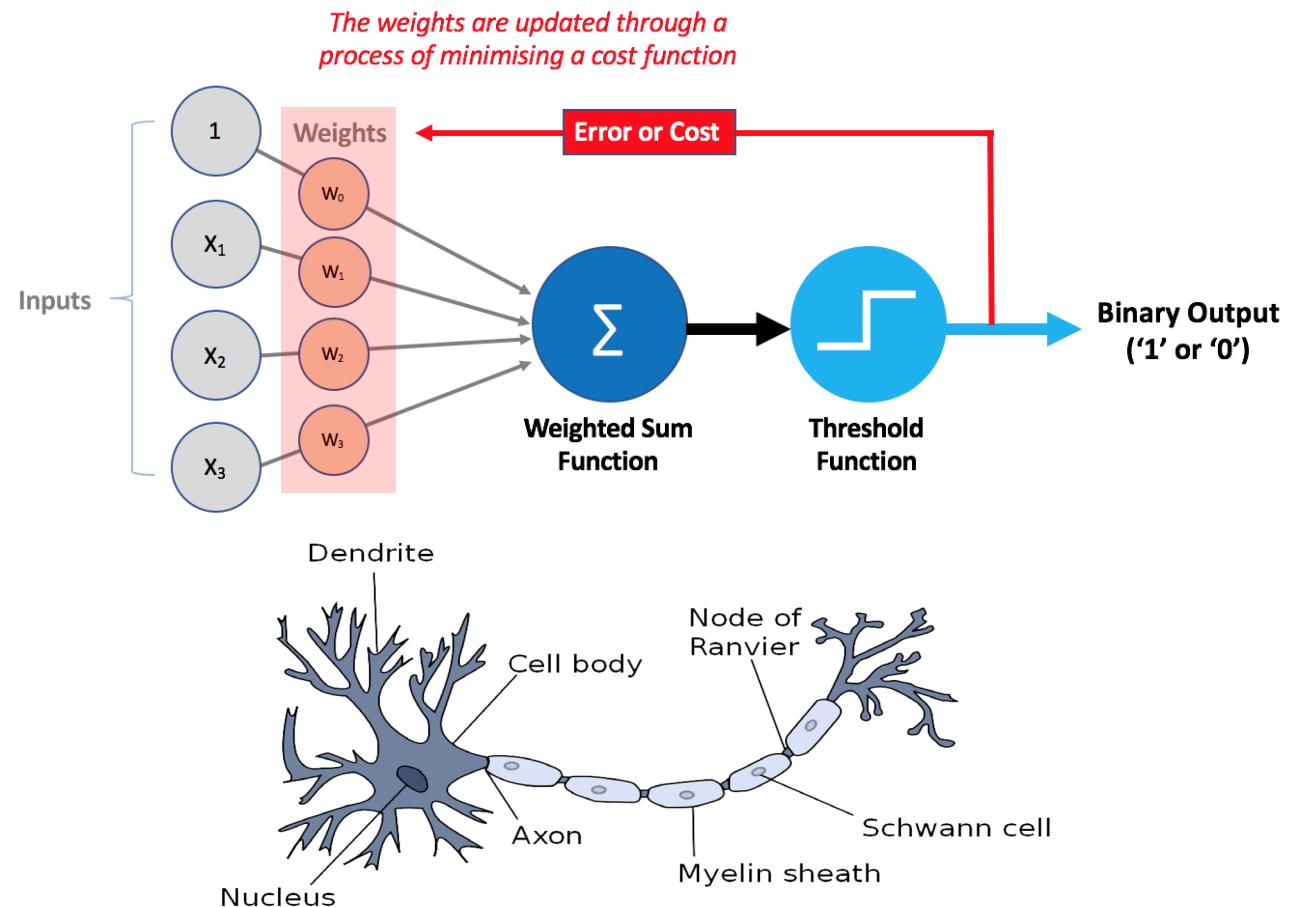
*---neuraal netwerk ---*

*“Hoe ziet een taalmodel  
er vanbinnen uit?”*

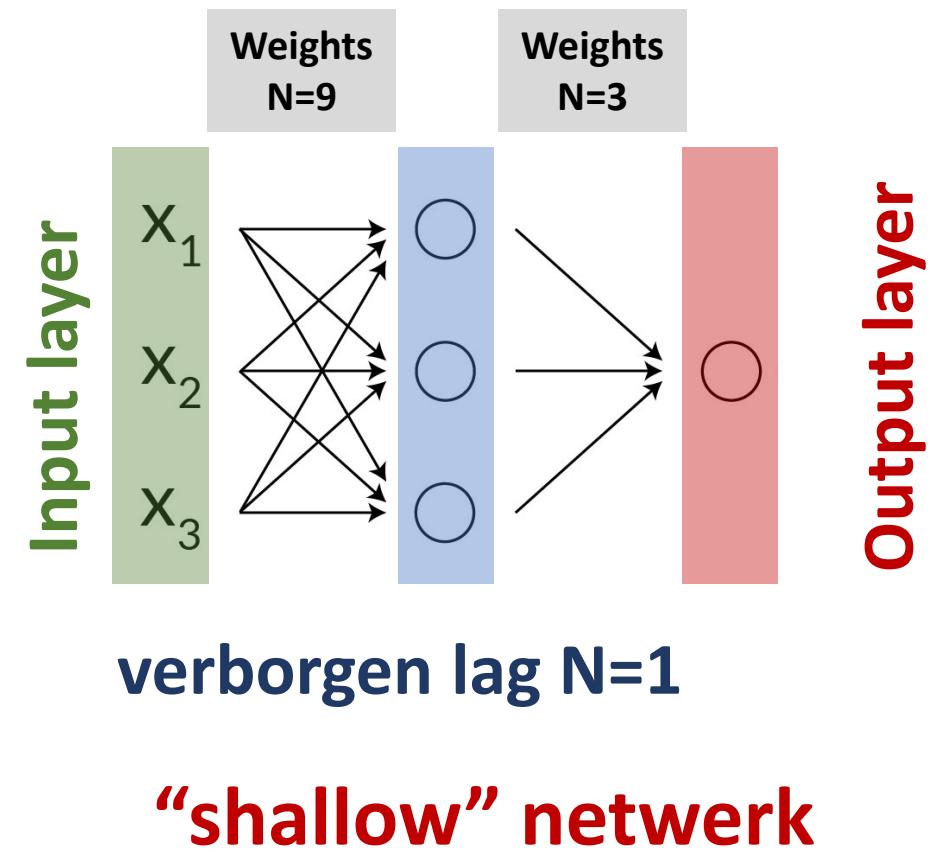
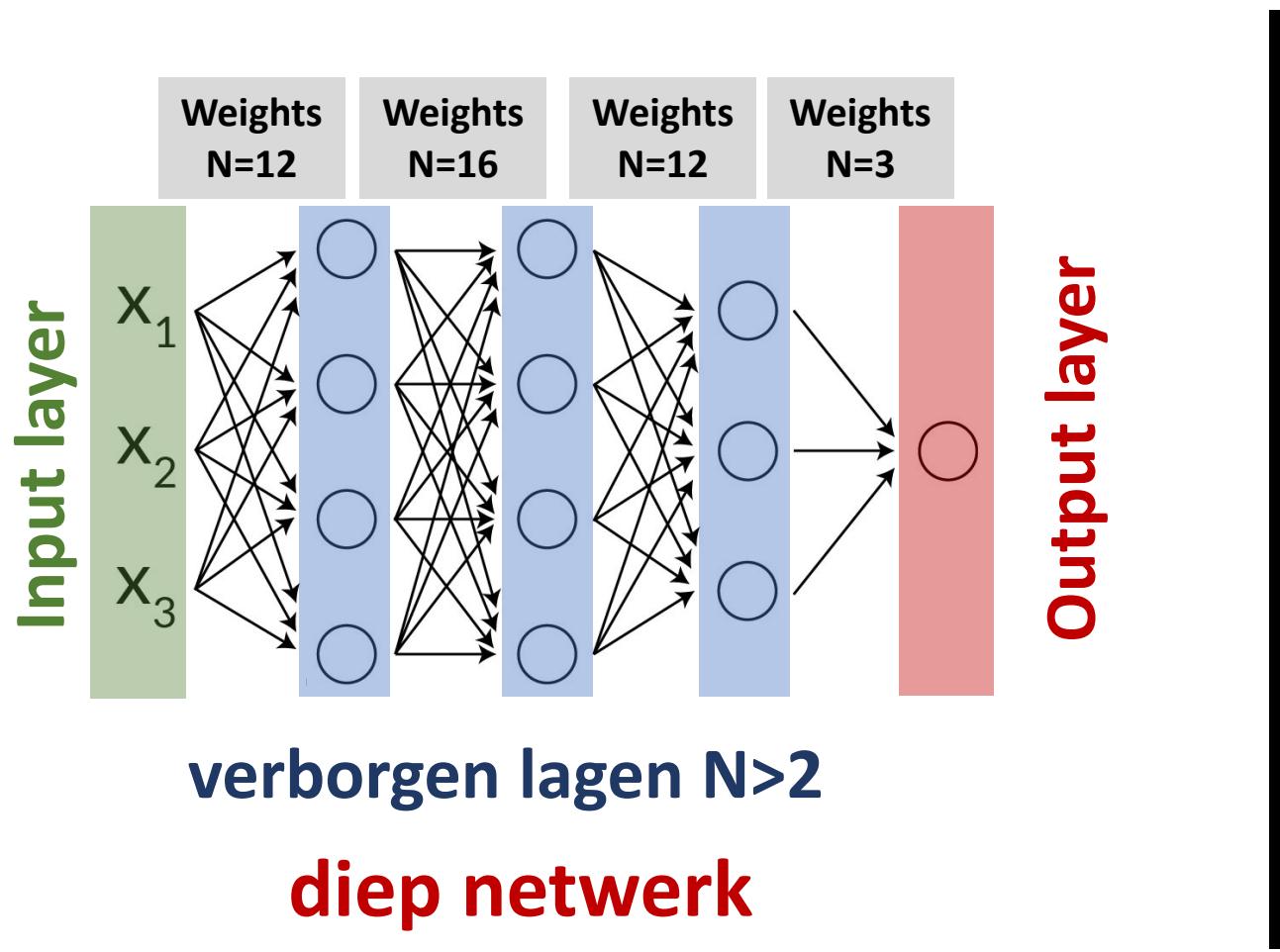
# Kunstmatige Intelligentie

**Names for Artificial Neurons**

- {unit}
- {cell}
- {node}
- {perceptron}



# Neuraal Netwerk modellen



# Diep neuraal netwerk

## What are the limits of deep learning?

The much-hyped artificial intelligence approach boasts impressive feats but still falls short of human brainpower. Researchers are determined to figure out what's missing.

M. Mitchell Waldrop, Science Writer

There's no mistaking the image: It's a banana—a big, ripe, bright-yellow banana. Yet the artificial intelligence (AI) identifies it as a toaster, even though it was trained with the same powerful and oft-publicized deep-learning techniques that have produced a white-hot revolution in driverless cars, speech understanding, and a multitude of other AI applications. That means the AI was shown several thousand photos of bananas, slugs, snails, and similar-looking objects, like so many flash cards, and then drilled on the answers until it had the classification down cold. And yet this advanced system was quite easily confused—all it took was a little day-glow sticker, digitally pasted in one corner of the image.

This example of what deep-learning researchers call an "adversarial attack," discovered by the Google Brain team in Mountain View, CA (1), highlights just how far AI still has to go before it remotely approaches human capabilities. "I initially thought that adversarial examples were just an annoyance," says Geoffrey Hinton, a computer scientist at the University of Toronto and one of the pioneers of deep learning. "But I now think they're probably quite profound. They tell us that we're doing something wrong."

That's a widely shared sentiment among AI practitioners, any of whom can easily rattle off a long list of deep learning's drawbacks. In addition to its vulnerability

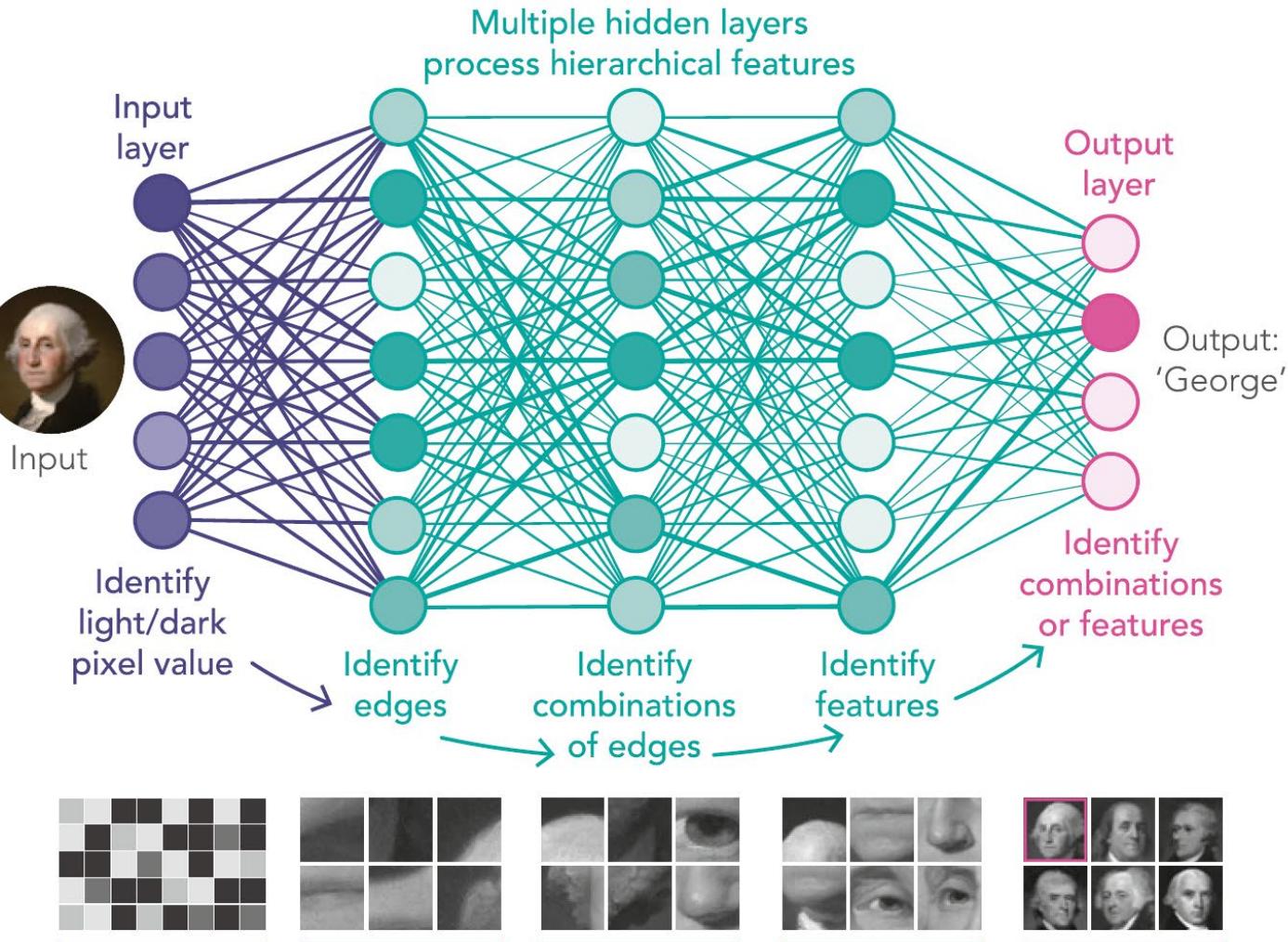


Apparent shortcomings in deep-learning approaches have raised concerns among researchers and the general public as technologies such as driverless cars, which use deep-learning techniques to navigate, get involved in well-publicized mishaps. Image credit: Shutterstock.com/MONOPOLY919.

Published under the PNAS license.

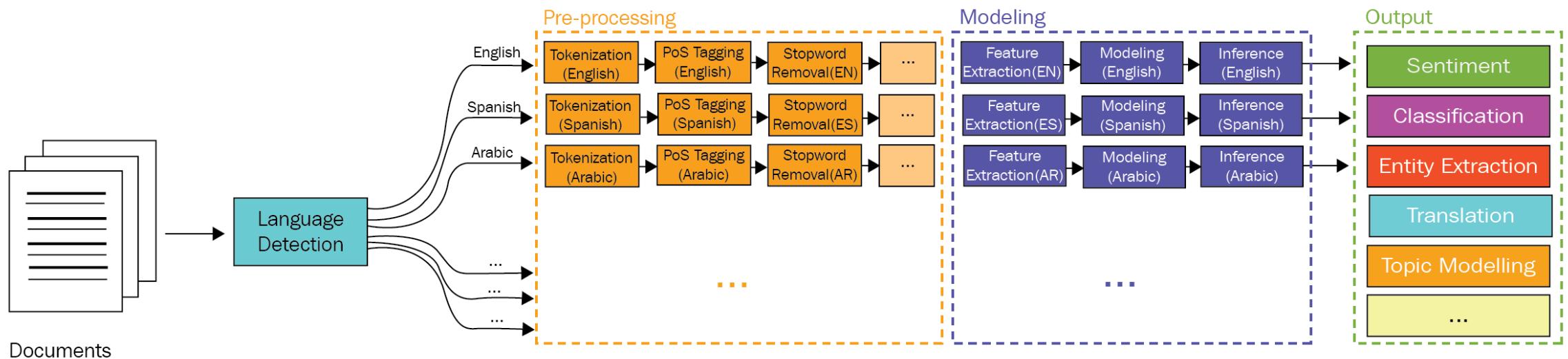
January 22, 2019 | vol. 116 | no. 4

[www.pnas.org/cgi/doi/10.1073/pnas.1821594116](http://www.pnas.org/cgi/doi/10.1073/pnas.1821594116)



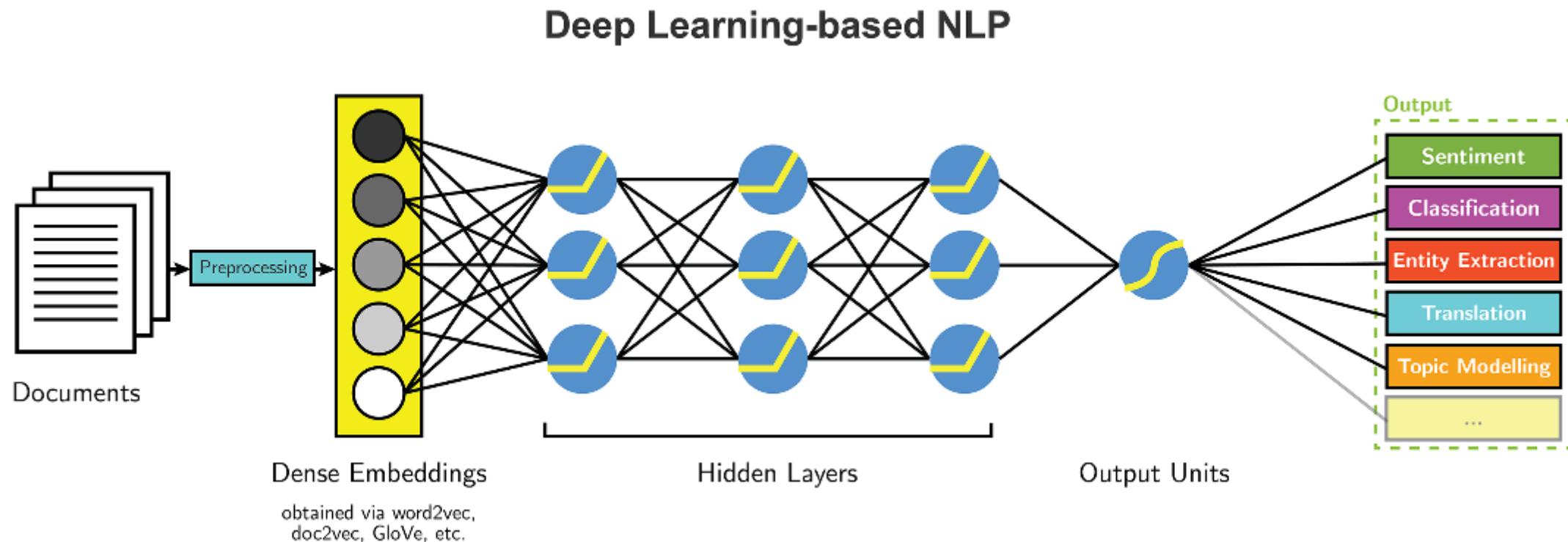
# Diep neuraal netwerk

## Classical NLP



<https://www.oreilly.com/library/view/python-natural-language/9781787121423/6f015f49-58e9-4dd1-8045-b11e7f8bf2c8.xhtml>

# Diep neuraal netwerk



<https://www.oreilly.com/library/view/python-natural-language/9781787121423/6f015f49-58e9-4dd1-8045-b11e7f8bf2c8.xhtml>

# “grote” *taalmodellen*

# ChatGPT

--- *[Autoregressive] Generative Pre-trained  
[neural network] Transformer---*

een heel groot diep neuraal netwerk

**begrijpen (NLU) + generatie (NLG) taalmodel**

autoregressie naar het proces waarbij tekst één token per keer wordt gegenereerd,  
waarbij elke token geconditioneerd is op de eerder gegenereerde tokens

# Transformer Training

President Franklin <M> born <M> January 1882.

Lily couldn't <M>. The waitress had brought the largest <M> of chocolate cake <M> seen.

Our <M> hand-picked and sun-dried <M> orchard in Georgia.

T5

D. Roosevelt was <M> in

believe her eyes <M> piece <M> she had ever

peaches are <M> at our

*Pre-training*

*Fine-tuning*

President Franklin D.  
Roosevelt was born  
in January 1882.

When was Franklin D.  
Roosevelt born?

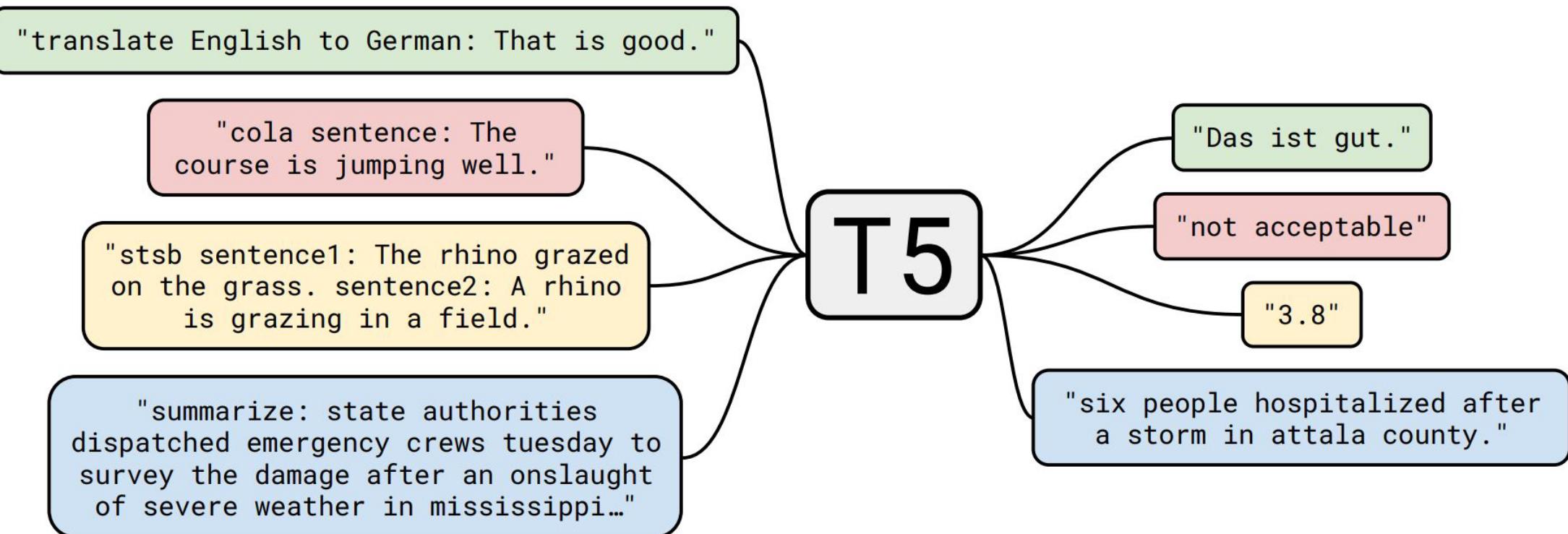
T5

1882

# Transformeren van woorden naar woorden



[https://ai.googleblog.com/  
2020/02/exploring-transfer-  
learning-with-t5.html](https://ai.googleblog.com/2020/02/exploring-transfer-learning-with-t5.html)



# Hoe groot is Large?

Huge “foundation models” are turbo-charging AI progress  
They can have abilities their creators did not foresee

The economist  
Jun 11th 2022



# The Economist

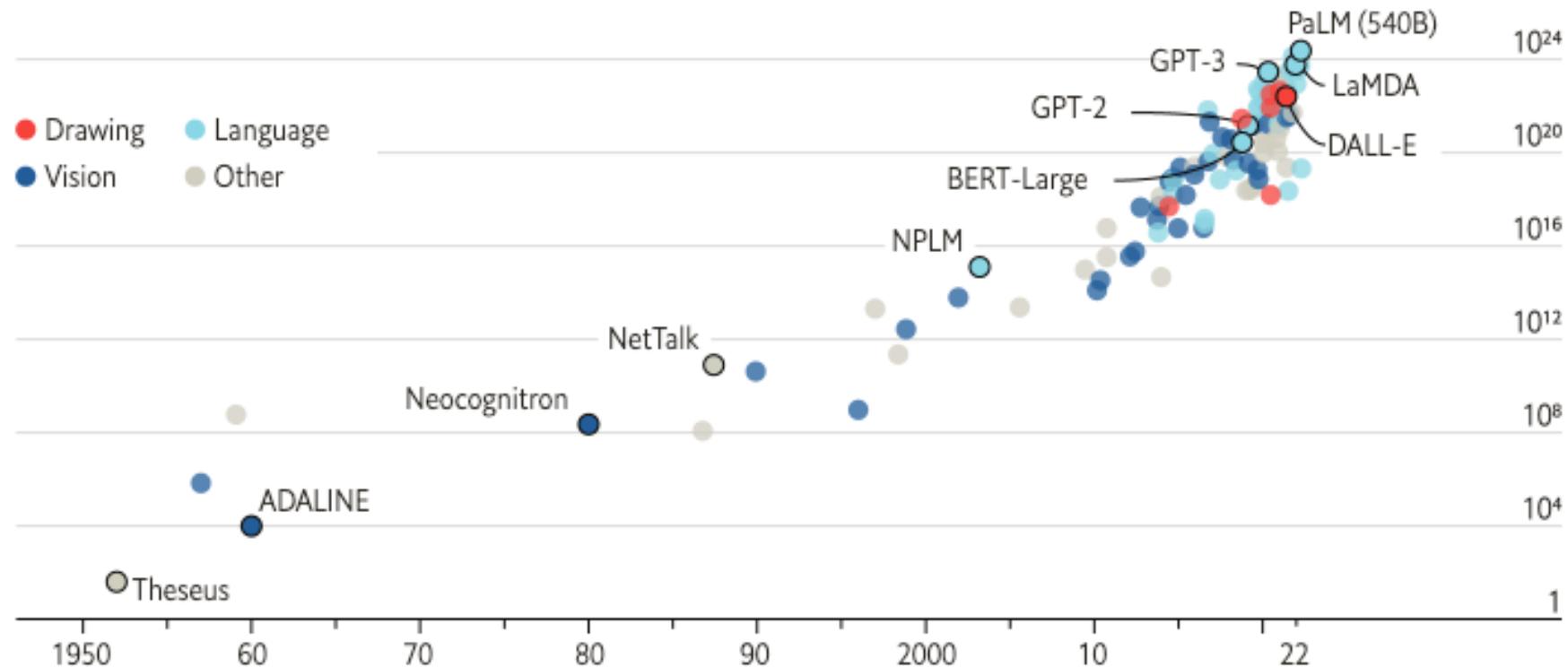


[https://www.economist.com/interactive/briefing/2022/06/11/  
huge-foundation-models-are-turbo-charging-ai-progress](https://www.economist.com/interactive/briefing/2022/06/11/huge-foundation-models-are-turbo-charging-ai-progress)

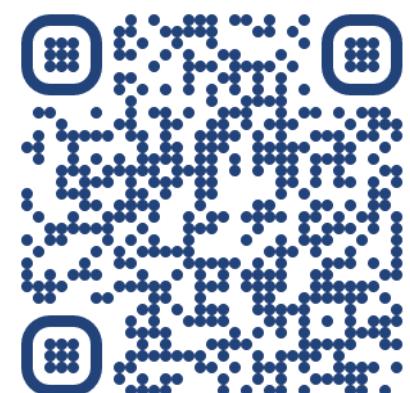
## The blessings of scale

AI training runs, estimated computing resources used

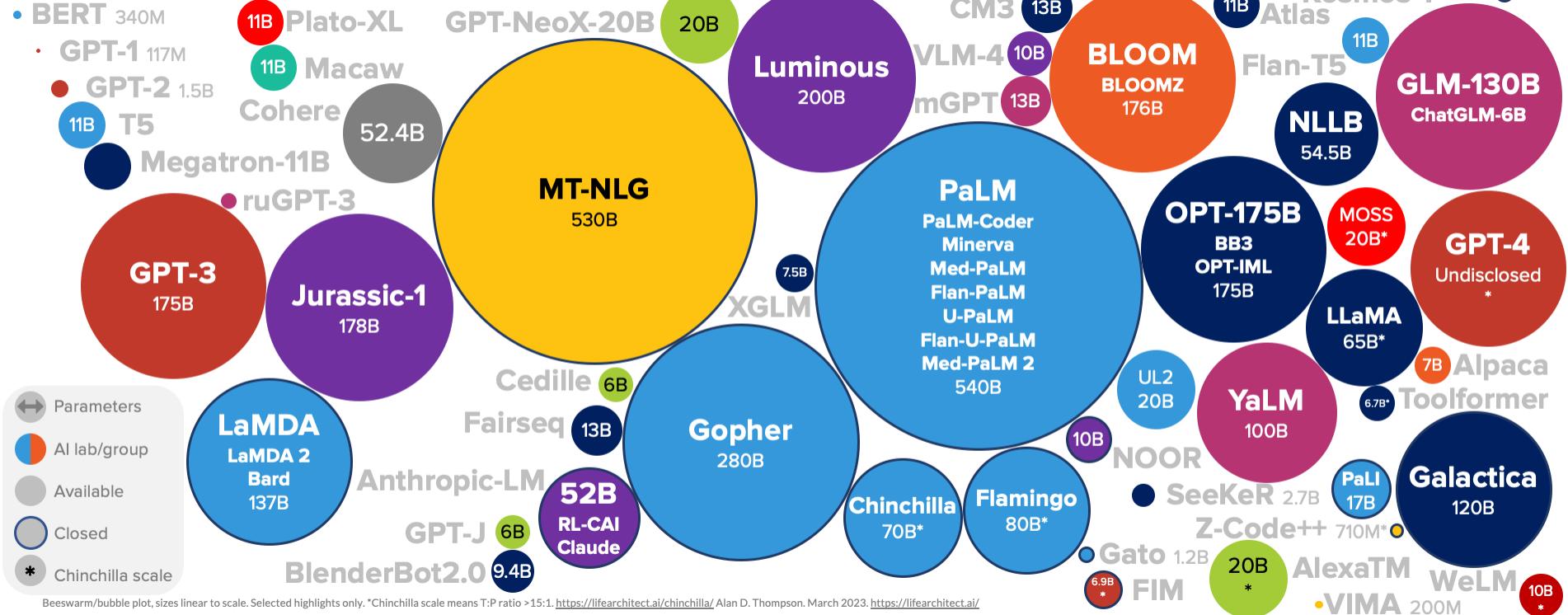
Floating-point operations, selected systems, by type, log scale



Sources: "Compute trends across three eras of machine learning", by J. Sevilla et al., arXiv, 2022; Our World in Data



# **Large Language Models: Het LLM-Landschap**



<https://s10251.pcdn.co/pdf/2023-Alan-D-Thompson-AI-Bubbles-Rev-7b.pdf>



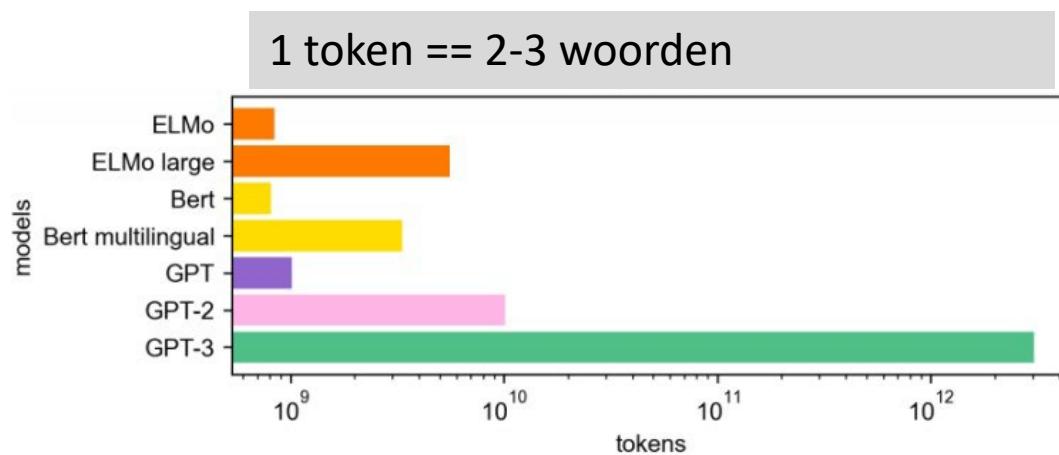
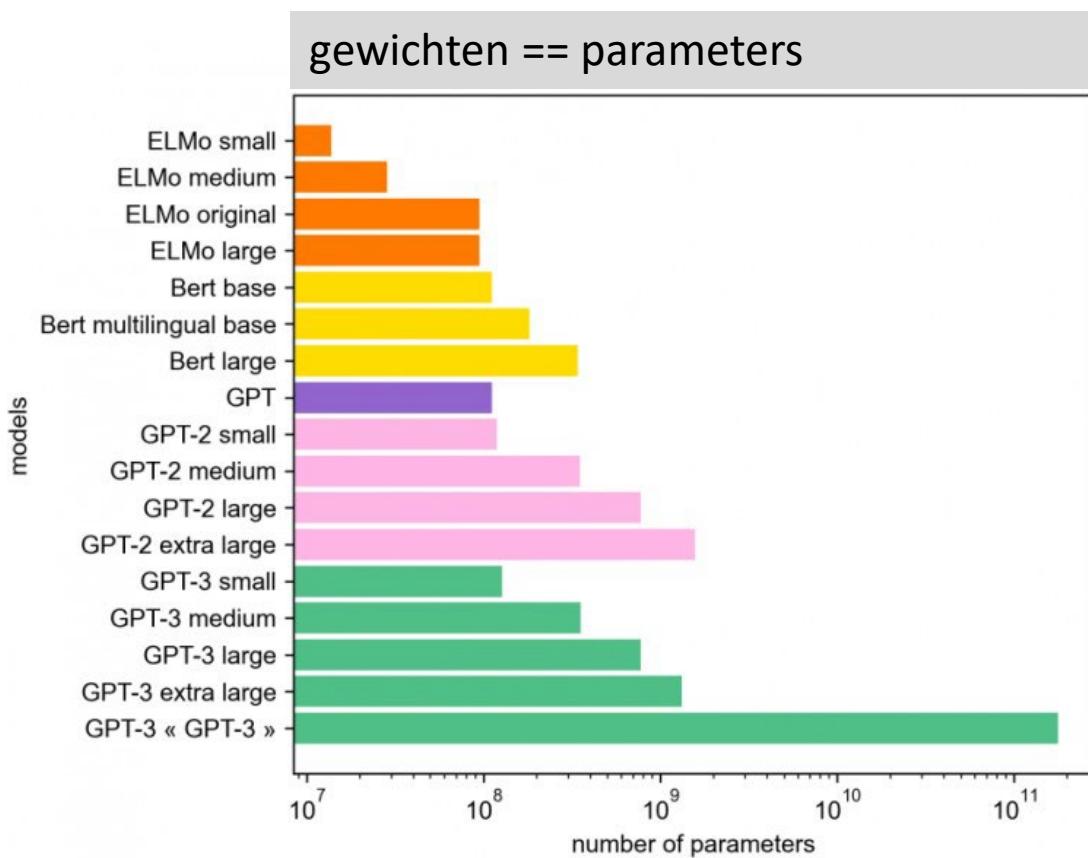
<https://docs.google.com/spreadsheets/d/105KVQW1Hx5ZAkcg8AIRjbQLQzx2wVaLi0SqUi-ir9Fs/edit#gid=1158069878>



<https://huggingface.co/spaces/adt/models-table>

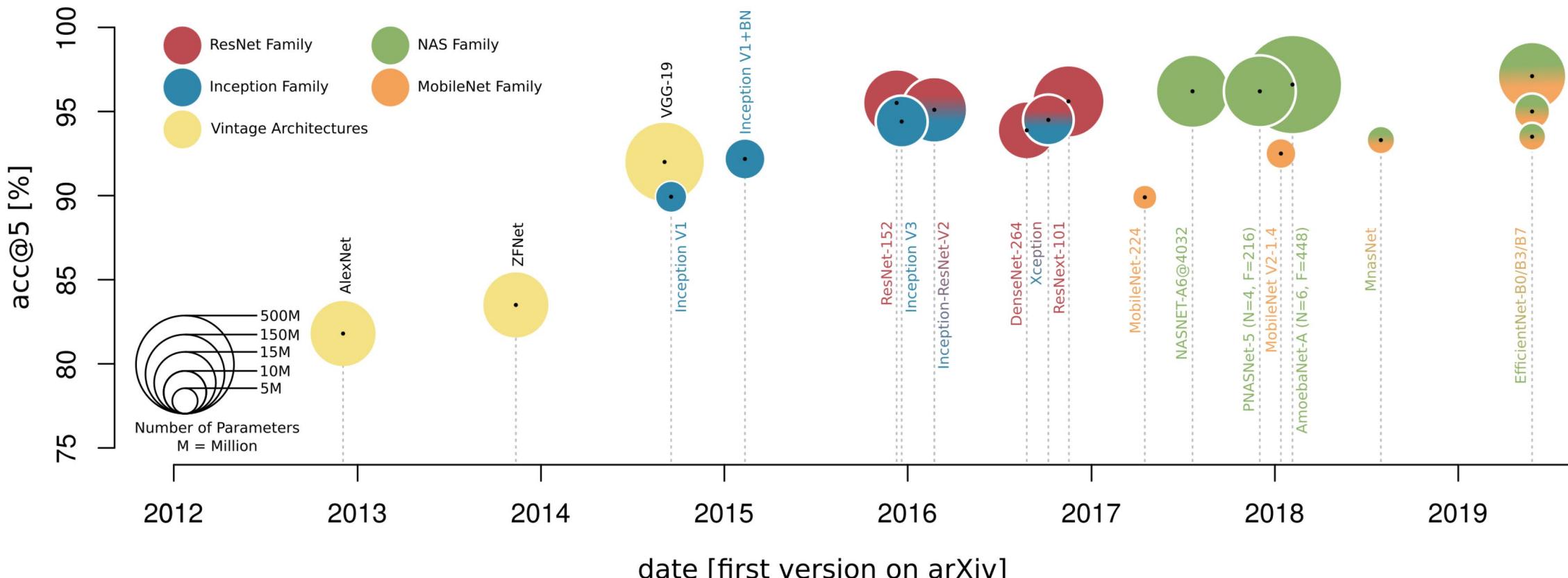
*“waarom zo  
groot?”*

# Grote training datasets vereist meer gewichten



[https://hellofuture.orange.com/  
en/the-gpt-3-language-model-  
revolution-or-evolution/](https://hellofuture.orange.com/en/the-gpt-3-language-model-revolution-or-evolution/)

# betrouwbaarheid neemt toe door meer gewichten



How to calculate the number of learnable parameters?

<https://doi.org/10.3390/rs12101667>

*“emmergente  
eigenschappen”*

# **Extreme opschaling van taalmodellen leidt tot onvoorziene "emergente eigenschappen"**

Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., Chi, E. H., Hashimoto, T., Vinyals, O., Liang, P., Dean, J., & Fedus, W. (2022). Emergent Abilities of Large Language Models. *Transactions on Machine Learning Research*, 2835-8856. <https://doi.org/10.48550/arXiv.2206.07682>

Kosinski, M. (2023). *Theory of mind* may have spontaneously emerged in large language models. arXiv preprint <https://doi.org/10.48550/arxiv.2302.02083>

# The Sally Anne Test

The Sally Anne Test has been used in psychological research to investigate **Theory of Mind**.

A **Theory of Mind** is the ability to understand your own and other people's beliefs, desires, intentions and emotions.



<http://www.educateautism.com/images/articles/sally-anne/sally-anne-test-educate-autism.png>

Theory of Mind is important because “the ability to make inferences about what other people believe to be the case in a given situation allows one to predict what they will do.”

(Baron-Cohen, Leslie, & Frith, 1985, p. 39)

It has been proposed that lacking a Theory of Mind may explain some of the social and communication difficulties experienced by individuals with autism. (Tager-Flusberg, 2007; Frith & Frith 2002)

# Sally-Anne False-Belief-Task

To start the test, two dolls, one called Sally and the other called Anne, are presented to the child.

This is **Sally**



This is **Anne**

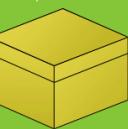


The child is then told that Sally has a basket and that Anne has a box.

Sally's  
**BASKET**



Anne's  
**BOX**



Next the child is told that Sally puts a marble inside her basket...



...Sally then leaves and goes outside where she can no longer see her basket...

I'm Back!



Next come the  
**QUESTIONS**

To make sure the child understands the basics s/he is asked:

- Q1. Which one is Sally?
- Q2. Which one is Anne?
- Q3. Where is the marble now?
- Q4. Where was the marble in the beginning?

This is  
**Sally**



This is  
**Anne**



The marble  
**WAS**  
in the basket



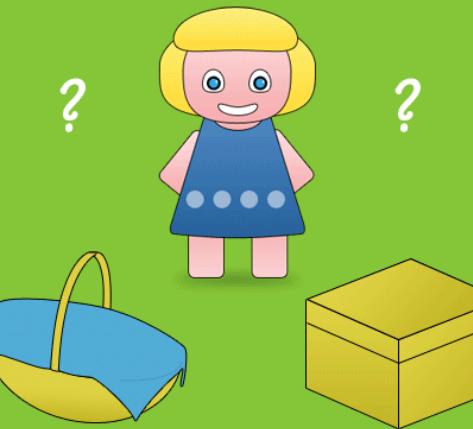
The marble is  
**NOW**  
in the **BOX**



Then comes the  
**BIG QUESTION**

The “Theory of Mind” question asks:

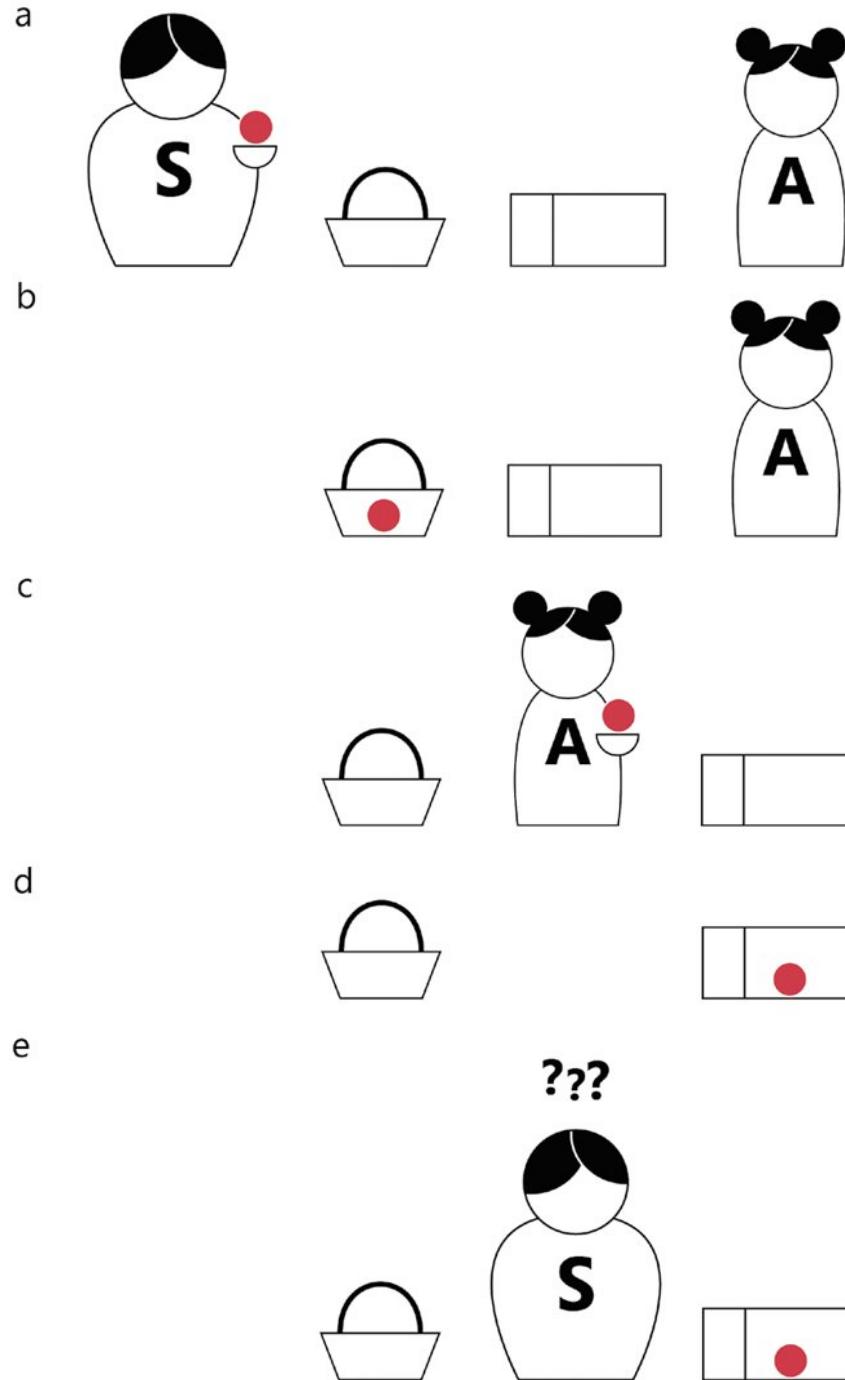
**Where will Sally look for her marble?**



If the child realises that Sally will look for the marble in her own basket and not in the box, then the child can understand Sally's perspective and you can say the child has a Theory of Mind.

If the child says that Sally will look for the marble in Anne's box then you might say the child **does not** have a Theory of Mind. In this case the child does not appear to understand Sally's perspective and that she did not see Anne take the marble and put it inside the box.

# False-Belief-Task



## The Sally–Anne Test

The Sally–Anne test is taken by children between the ages of three and four. It involves staging a simple puppet show involving two characters, Sally and Anne (a). Sally places a marble in a basket, closes its lid and leaves the room (b).

Shortly thereafter, the very naughty Anne flips open the lid of the basket, pulls out the marble (c) and places it in a box sitting in the corner (d).

Now, the child who has witnessed all of this is asked a simple question: when Sally returns, where will she look for the marble (e)?

Almost every child under 4 years of age exclaims with confidence, “In the box!” This answer is baffling to adults for the obvious reason: there’s no way Sally could have known that the marble was mischievously displaced by Anne because Sally wasn’t around to witness that.

Young children provide an incorrect answer because they have yet to develop what is known as the *theory of mind* (*ToM*), a differentiating feature of human beings when compared to most other mammals.

We develop *ToM* by the time we’re five years old. In fact, if you were to administer the test to five-year-olds, you would be greeted with sheer bewilderment for having wasted their time before they gave you the right answer.

Theory of mind bestows on us two critical pieces of knowledge that, when wielded properly, have the ability to bring out the best in us: Our mind’s version of reality isn’t true reality: it’s just one interpretation of reality. Everybody has their own mind and, thus, their own interpretation of reality.



HOGESCHOOL  
ROTTERDAM

*“Hoe betrouwbaar  
zijn LLM en hoe  
test je dat?”*

# Veel voorkomende gebreken bij LLMs

Inconsistente output

Crippled by length

coreference issues

Longer strings that are repeated many times in the dataset

Unnecessarily repeating entities

Maintaining **coherence** between **paragraphs**

No real understanding, commonsense, factual correctness”

How to learn discourse?

Lack of implicit “planning”

sub-optimal evaluation metrics “perplexity”

MLE: cannot capture high-level semantics

Biased pre-trained models

Domain transfer is hard

Single path generative flow

Softmax Bottleneck issues!

Surrogate-loss functions

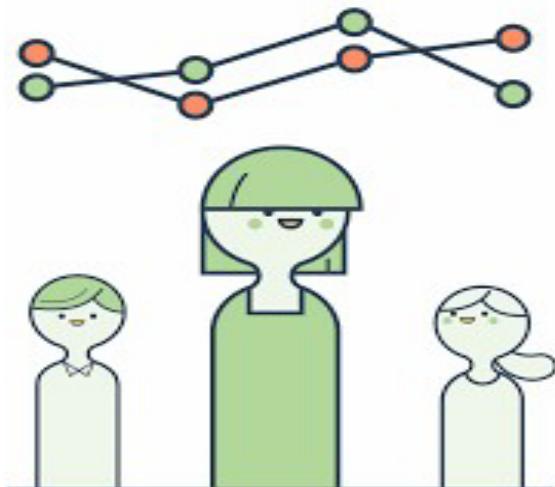
Auto-regressive!

exposure-bias

Degenerative sampling methods

word-by-word generation: can't see **global context!**

# Hoe test je LLMs?



Menselijke beoordeling

Belangrijkste vorm van evaluatie

Automatische meetmethoden kunnen  
menselijke beslissingen niet repliceren

Gouden standaard bij het ontwikkelen  
van nieuwe automatische meetmethoden

Mensen evalueren prestaties van LLMs op de **taak** waarvoor het **model** is ontworpen en de **intentie/inschatting** van de **eindgebruiker**

### Oordeel van de eindgebruiker

*Interessant*

*Boeiend*

*Algemeen/specifiek*

*Relevant*

*Semantisch geschikt*

*Begrijpelijk*

*Vloeiend geschreven*

*Correct vs. Onbegrijpelijk*

*Algemene indruk*

### Kwaliteit van de gegenereerde tekst

*Coherent*

*Fout herstellend*

*Consistentie*

*Diversiteit in de antwoorden*

*Diepgang van het onderwerp*

*sympathiek (empathie, persoonlijkheid)*

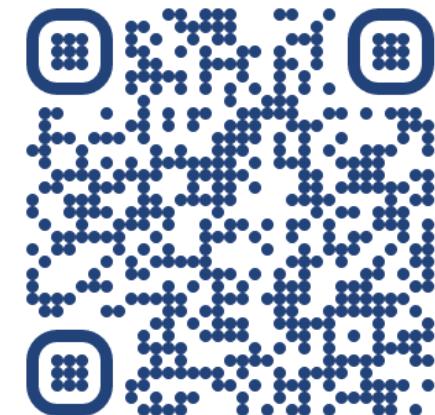
*Begripvol*

*Flexibel en aanpasbaar*

*Informatief*

*Nieuwsgierig*

*Algemene indruk*



Mehri, S., Choi, J., D'Haro, L., F., Deriu, J., Eskenazi, M., Gasic, M., ... & Zhang, C. (2022). Report from the nsf future directions workshop on automatic evaluation of dialog: Research directions and challenges. arXiv preprint <https://doi.org/10.48550/arXiv.2203.10012>

# *“Prompt Recepten” Schrijven Uitgelegd*

# PROMPT

Een prompt is een natuurlijke taal instructie gericht aan een **"bevroren" voorgetraind taalmodel** met als doel om de werking ervan te verbeteren of te verfijnen.

# PROMPT-ENGINEERING

Het creatieve proces van het schrijven van een effectief ***prompt-recept*** wordt in het Engels "***prompt engineering***" genoemd.

Het schrijven van prompt-recepten

***---pseudo-Code---***

is een talige manier van het programmeren van "bevroren" voorgetraind taalmodellen.

# **Wat is het belang van Prompt Recepten Schrijven?**

Deskresearch / Waarheidsvinding

Effectieve Zoekopdracht

Testen & Evalueren van Chatbots & taalmodellen

Innovatie van taalmodellen, generatieve AI & NLP

# Prompt Recept Structuur

Een prompt is opgebouwd uit de volgende elementen:

Instructie(s)

Context

Invoergegevens

Uitvoer-indicator

*Classificeer de onderstaande tekst als neuraal, negatief of positief*

*Text: Ik vond het eten wel zozo.*

*Sentiment:*

# Prompt-recept versus Takgerichtheid

Tekst samenvatten

Vraag beantwoorden

Tekstclassificatie

Rollenspel

Code genereren

Redeneren

# Prompt recept schrijf technieken

Few-shot prompts (In Context leren)

Chain-of-thought (CoT) prompting

Self-Consistency

Knowledge Generation Prompting

ReAct

# Kansen & Gevaren

Model veiligheid

prompt injectie / "lekkage" / "gevangenisuitbraken"

Human-in-the-loop [RLHF]

Toekomstige richtingen

DIY {do it yourself}

# In-Context Leren {ICL}

## No Prompt

**Zero-shot  
(os)**

speudo = **pseudo**

**1-shot  
(1s)**

soelt = stoel  
**speudo** = **pseudo**

**Few-shot  
(FS)**

soelt = stoel  
[.]  
pciinc =picnic  
speudo = **pseudo**

## Prompt

**Ontcijfer de letters tot een woord, en schrijf dat woord:**  
speudo = **pseudo**

**Ontcijfer de letters tot een woord, en schrijf dat woord:**  
soelt = stoel  
speudo = **pseudo**

**Ontcijfer de letters tot een woord, en schrijf dat woord:**  
chiar =chair  
[.]  
pciinc =picnic  
skicts =**sticks**

# In-Context Leren {ICL}

## Fine-tuning (FT)

- + hoge betrouwbaarheid
- Voor elke nieuwe taak is een gecureerde en gelabelde dataset nodig
- Niet generaliseerbaar, gebruikte uitvoer indicator is fout gevoelig

vereist veel tijd  
hoog sturend  
taak specifiek

## Few-shot (FS)

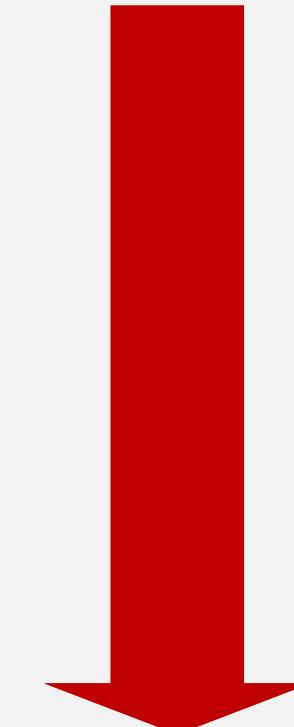
- + weinig taakspecifieke gegevens nodig
- + gebruikte uitvoer indicator is niet fout gevoelig
- vereist oefening en grote woordenschat

## One-shot (1S)

- + meest natuurlijke vorm van opdracht tegen
- vereist zeer veel oefening en taalkundigheid

## Zero-shot (oS)

- + vereist nauwelijks extra inspanning
- Antwoorden zijn vaak dubbelzinnig en/of incorrect en/of wartaal



Snel  
generaliserend  
openeinde output

# The mystery of in-context learning

Large language models (LMs) such as GPT-3 <sup>3</sup> are trained on internet-scale text data to predict the next token given the preceding text. This simple objective paired with a large-scale dataset and model results in a very flexible LM that can “read” any text input and condition on it to “write” text that could plausibly come after the input. While the training procedure is both simple and general, the GPT-3 paper found that the large scale leads to a particularly interesting emergent behavior <sup>4</sup> called in-context learning.

**What is in-context learning?** In-context learning was popularized in the original GPT-3 paper as a way to use language models to learn tasks given only a few examples.<sup>[1]</sup> During in-context learning, we give the LM a prompt that consists of a list of input-output pairs that demonstrate a task. At the end of the prompt, we append a test input and allow the LM to make a prediction just by conditioning on the prompt and predicting the next tokens. To correctly answer the two prompts below, the model needs to read the training examples to figure out the input distribution (financial or general news), output distribution (Positive/Negative or topic), input-output mapping (sentiment or topic classification), and the formatting.

Circulation revenue has increased by 5%  
in Finland. // Positive

Panostaja did not disclose the purchase  
price. // Neutral

Paying off the national debt will be  
extremely painful. // Negative

The company anticipated its operating  
profit to improve. // \_\_\_\_\_

Circulation revenue has increased by  
5% in Finland. // Finance

They defeated ... in the NFC  
Championship Game. // Sports

Apple ... development of in-house  
chips. // Tech

The company anticipated its operating  
profit to improve. // \_\_\_\_\_

LM

LM

Given the following 3 statements: (1) Circulation revenue has increased by 5% in Finland is Positive (2) Panostaja did not disclose the purchase price. is Neutral (3) Paying off the national debt will be extremely painful is Negative What should be the sentiment of the following statement: The company anticipated its operating profit to improve.

- ✓ Searching for: what is sentiment analysis
- ✓ Generating answers for you...

Sentiment analysis is the process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the writer's attitude towards a particular topic, product, etc. is positive, negative, or neutral<sup>1</sup>.

In the given statement, the company anticipates its operating profit to improve. Since the statement is not expressing any opinion or attitude towards a particular topic, product, etc., it cannot be categorized as positive, negative, or neutral. Therefore, the sentiment of the statement is neutral.

Learn more:

1. [bing.com](http://bing.com)
  2. [geeksforgeeks.org](http://geeksforgeeks.org)
  3. [freecodecamp.org](http://freecodecamp.org)
- +2 more

# Kan Bing deze taken foutloos uitvoeren??

*Stanford University is located in \_\_\_\_\_, California.* [Trivia]

*I put \_\_\_ fork down on the table.* [syntax]

*The woman walked across the street, checking for traffic over \_\_\_ shoulder.* [coreference]

*I went to the ocean to see the fish, turtles, seals, and \_\_\_\_\_.*  [lexical semantics/topic]

*Overall, the value I got from the two hours watching it was the sum total of the popcorn and the drink.*

*The movie was \_\_\_\_\_.*  [sentiment]

*Irohwent into the kitchen to make some tea. Standing next to Iroh, Zuko pondered his destiny. Zuko left the \_\_\_\_\_.*

*[some reasoning –this is harder]*

*I was thinking about the sequence that goes 1, 1, 2, 3, 5, 8, 13, 21, \_\_\_\_\_.*  [some basic arithmetic; they don't learn the Fibonnacisequence]

# Instellen van token-machine parameters

*Bij het bevragen van een taalmodel dient rekening gehouden te worden met de volgende instellingen*

*Je kunt heel verschillende resultaten krijgen met dezelfde prompts wanneer de instellingen aanpast*

*Een belangrijke instelling is het bepalen hoe deterministisch het model is bij het genereren van voltooiing voor prompts*

**Temperatuur** en **top\_p** zijn twee belangrijke parameters om te onthouden in gedachten te houden

- laag voor meer exacte antwoorden
- hoog voor meer diverse/generieke antwoorden

[Upgrade](#)[Help](#) Hogeschool Rotterdam

Load a preset...

Save

View code

Share

...

**USER** IF  $20 + 30$  equals  $2030$  what is the correct answer to  $40 + 50$ 

Mode

 Chat Beta**ASSISTANT** The correct answer to  $40 + 50$ , based on the pattern given, would be  $4050$ .

Model

gpt-4-0314

Temperature 0.7

Maximum length 256

Top P 1

Frequency penalty 0

Presence penalty 0

**USER** gaot**ASSISTANT** I'm not sure what you mean by "gaot." If you meant to type "goat," please provide more context or clarify your question. If it was a typo or unrelated, feel free to ask another question or provide more information.**USER** Stanford University is located in ??? California.

I put ??? fork down on the table.

**ASSISTANT** Stanford University is located in Palo Alto, California.

I put the fork down on the table.

**USER**

The woman walked across the street, checking for traffic over ??? shoulder. [coreference]

**ASSISTANT** The woman walked across the street, checking for traffic over her shoulder. Add message<https://platform.openai.com/playground?mode=chat&model=gpt-4-0314>

DIY {do it yourself}

# Schrijf een prompt om te bepalen of Bing over Theory-of-Mind {ToM} beschikt

### The Sally Anne Test

The Sally Anne Test has been used in psychological research to investigate Theory of Mind.

A Theory of Mind is the ability to understand your own and other people's beliefs, desires, intentions and emotions.



Theory of Mind is important because "the ability to make inferences about what other people believe to be the case in a given situation allows one to predict what they will do."  
(Baron-Cohen, Leslie, & Frith, 1985, p. 39)

It has been proposed that lacking a Theory of Mind may explain some of the social and communication difficulties experienced by individuals with autism.  
(Tager-Flusberg, 2007; Frith & Frith 2002)

### Sally-Anne False-Belief-Task

To start the test, two dolls, one called Sally and the other called Anne, are presented to the child.



The child is then told that Sally has a basket and that Anne has a box.



Next come the QUESTIONS



To make sure the child understands the basics s/he is asked:

- Q1. Which one is Sally?
- Q2. Which one is Anne?
- Q3. Where is the marble now?
- Q4. Where was the marble in the beginning?



This is Sally  
This is Anne

I'm Back!

Next come the QUESTIONS

Then comes the BIG QUESTION

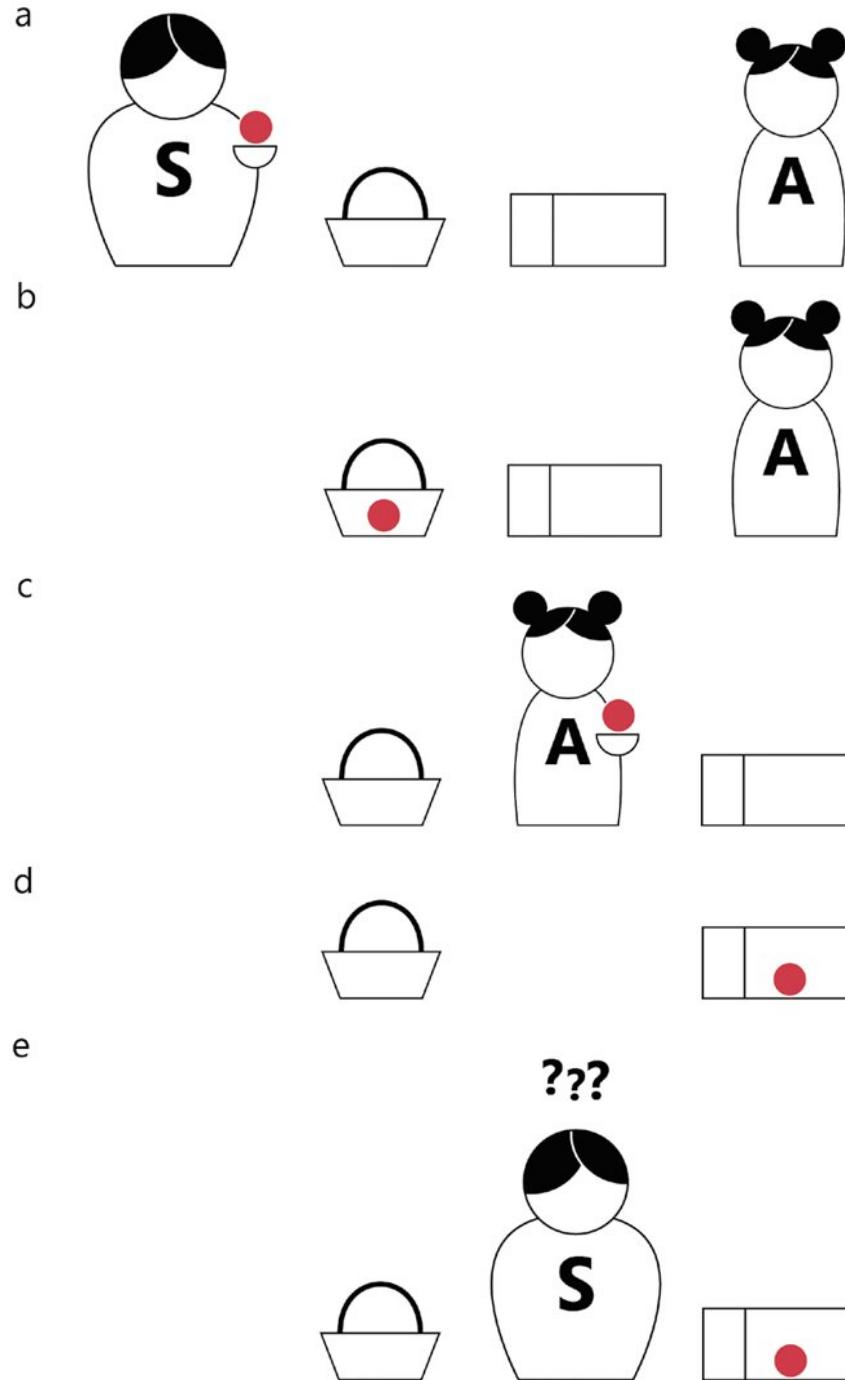
The "Theory of Mind" question asks:  
Where will Sally look for her marble?



If the child realises that Sally will look for the marble in her own basket and not in the box, then the child can understand Sally's perspective and you can say the child has a Theory of Mind.

If the child says that Sally will look for the marble in Anne's box then you might say the child does not have a Theory of Mind. In this case the child does not appear to understand Sally's perspective and that she did not see Anne take the marble and put it inside the box.

# False-Belief-Task



## The Sally–Anne Test

The Sally–Anne test is taken by children between the ages of three and four. It involves staging a simple puppet show involving two characters, Sally and Anne (a). Sally places a marble in a basket, closes its lid and leaves the room (b).

Shortly thereafter, the very naughty Anne flips open the lid of the basket, pulls out the marble (c) and places it in a box sitting in the corner (d).

Now, the child who has witnessed all of this is asked a simple question: when Sally returns, where will she look for the marble (e)?

Almost every child under 4 years of age exclaims with confidence, “In the box!” This answer is baffling to adults for the obvious reason: there’s no way Sally could have known that the marble was mischievously displaced by Anne because Sally wasn’t around to witness that.

Young children provide an incorrect answer because they have yet to develop what is known as the *theory of mind* (*ToM*), a differentiating feature of human beings when compared to most other mammals.

We develop *ToM* by the time we’re five years old. In fact, if you were to administer the test to five-year-olds, you would be greeted with sheer bewilderment for having wasted their time before they gave you the right answer.

Theory of mind bestows on us two critical pieces of knowledge that, when wielded properly, have the ability to bring out the best in us: Our mind’s version of reality isn’t true reality: it’s just one interpretation of reality. Everybody has their own mind and, thus, their own interpretation of reality.



Rob

robvdw

Edit profile

8 followers · 0 following

Hogeschool Rotterdam

Netherlands

[https://www.researchgate.net/profile/Roberto\\_Van\\_Der\\_Willigen](https://www.researchgate.net/profile/Roberto_Van_Der_Willigen)

## Highlights

PRO

## Organizations



Overview Repositories 15 Projects Packages

## Pinned

[Psignifit-MEX](#) Public

Mex files for Matlab 2010b on Win64 / Mac OS X 10.6.7

1 star

1,619 contributions in the last year



Learn how we count contributions

Contribution settings ▾

## Contribution activity

March 2023

2023

2022

2021

[Created 391 commits in 3 repositories](#)

[HR-ChatGPT/ChatGPT-UITGELEGD](#) 381 commits

[HR-ChatGPT/chatgpt.uitgelegd](#) 8 commits

[HR-Data-Supported-Healthcare/Corpora-that-Support-Clinical-Reasoning](#) 2 commits

[Joined the Hogeschool Rotterdam organization](#)



Hogeschool Rotterdam

De wildgroei aan gezondheidsdossiers heeft een overvloed aan klinisch tekst opgeleverd die vaak onbenut blijft. AI in de vorm van NLP kan uitkomst bieden.





RobFvdW

Nov 12, 2020 · 6 min read · [Listen](#)

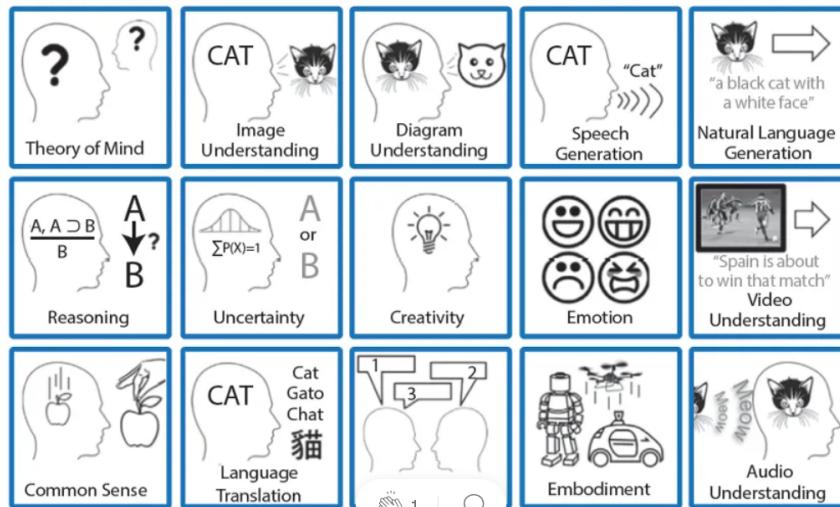


Serengeti Plains. By [Kristin Moger](#)

## The World-Wide-Web AI-Safari

Artificial Intelligence: a Human Centred View

AI aims to **mimic & automatise** tasks which otherwise require human perception, cognition and/or motor skills – e.g. pattern recognition, learning, logical reasoning & planning, decision making, problem solving, designing, creativity, likelihood estimation, language acquisition, multi-sensory interfacing, actuated body movement control, locomotion & manipulation, sentiment analysis, and generalisation (see refs [1]...[3]).



RobFvdW

4 Followers

My research and  
biological & psych  
to smart informa

Follow





## Robert Frans van der Willigen

Hogeschool Rotterdam | HR · Instituut voor Communicatie, Media en Informatietechnologie (CMI)  
PhD

Contact

About

48

Publications

83,070

Reads ⓘ

685

Citations

### Introduction

Binnen het Creating-010 onderzoeksthema Design in the 21st Century onderzoeken we de valorisatie van Machinaal Lerende (ML) Neurale-Netwerk (NN) algoritmen. De uitkomsten ervan moeten het ontwerp van Human-Centred (HC) data-producten mogelijk maken binnen toepassingsdomeinen zoals Smart & Social City, Zorg-Tech, Smart Logistics & Maritime Innovation. <https://robfvdw.medium.com>

### Skills and Expertise

- (Brain)
- (Animal Behavior)
- (Neural Networks)
- (Ethology)
- (Biophysics)
- (Data Science)
- (Data Visualization)
- (Behavioral Analysis)
- (Behavioral Neuroscience)
- (Machine Learning)

### Publications

Publications (48)

Designing Neural Networks Through Sensory Ecology "Biology to the rescue of AI"  
Produced by Living-Lab: AiRA, Hub voor Data & Responsible AI, Hogeschool Rotterdam  
Lunch-Lezing Creating-010 FEB 2022

### Current institution

Hogeschool Rotterdam



Instituut voor Communicatie, ...  
Current position  
Data Scientist

### Citations since 2017



[Learn about citations on ResearchGate](#)

### Co-authors

### Top co-authors





HOGESCHOOL ROTTERDAM

**overtref jezelf**