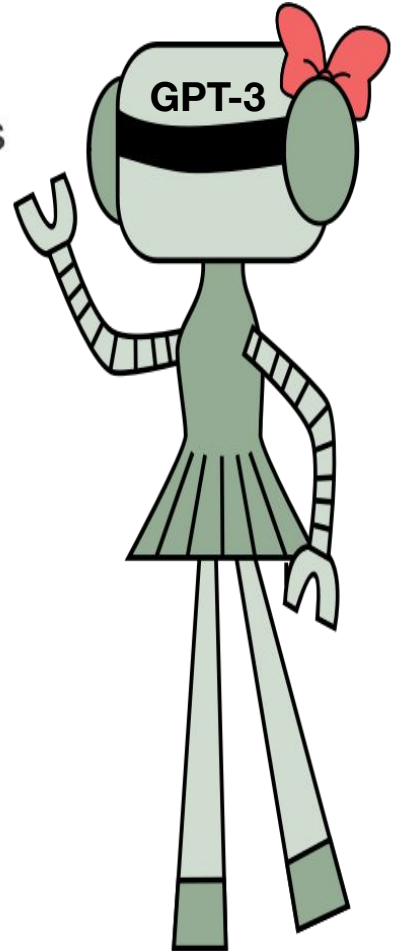


Note for downloadable pdf

Taken outside of my actual presentation's context, the message these slides are supposed to convey could be unclear or misunderstood. I particularly care to remark that there is absolutely no negative message regarding Bram Vanroy's work (which I take as example of training models with language-specific synthetic data for Dutch, as his are the core existing models for this language), which, in fact, I appreciate and value a lot.

If you had to pick just one feature that makes humans different from other animals what would that be?

The feature that makes humans different from other animals is our capacity for language.



Are Language Models Models of Language?

Grote Taaldag
Utrecht, 31 January 2025

Are Language Models Models of Language?



Malvina Nissim

m.nissim@rug.nl

[https:// malvinanissim.github.io](https://malvinanissim.github.io)



Center for Language and Cognition



university of
 groningen

What Are Language Models?



Is ChatGPT a
Language Model?

ChatGPT = Chat + GPT



Generative



Input

“Hello ladies and”

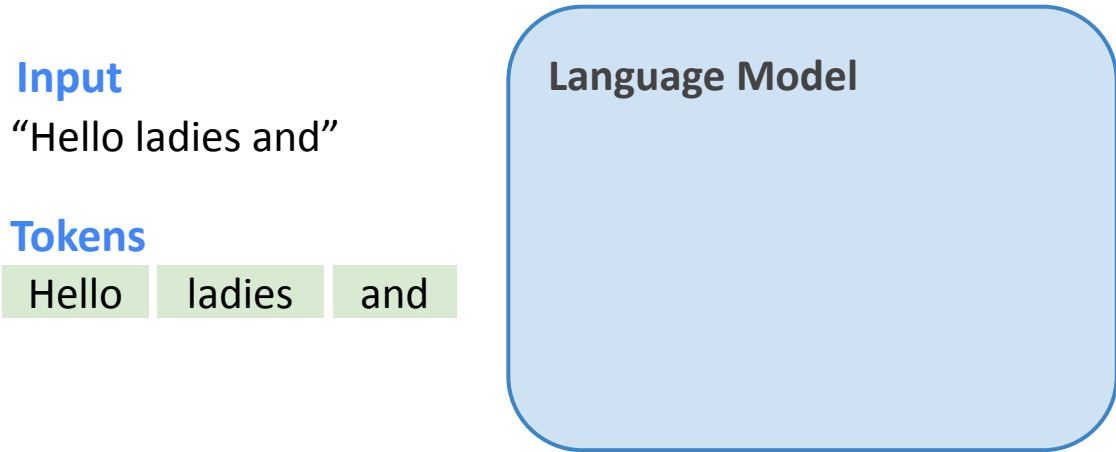
Tokens

Hello

ladies

and

Language Model



```
graph LR; Input["Input  
“Hello ladies and”"] --> Tokens["Tokens  
Hello ladies and"]; Tokens --> LModel["Language Model"]
```

Input

"Hello ladies and"

Tokens

Hello

ladies

and

Language Model

Prediction

Probabilities

0.2%

...

1 %

...

90%

...

aadvark

badgers

gentlemen

Vocabulary Words

Generative



Transformer

Generative

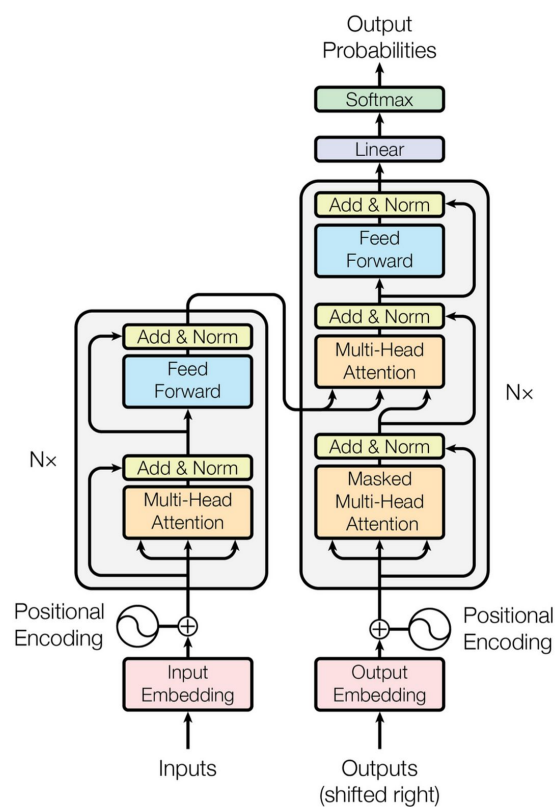


Figure 1: The Transformer - model architecture.

Transformer

Input

"Hello ladies and"

Tokens

Hello

ladies

and

Language Model

Prediction

Probabilities

0.2%

...

1 %

...

90%

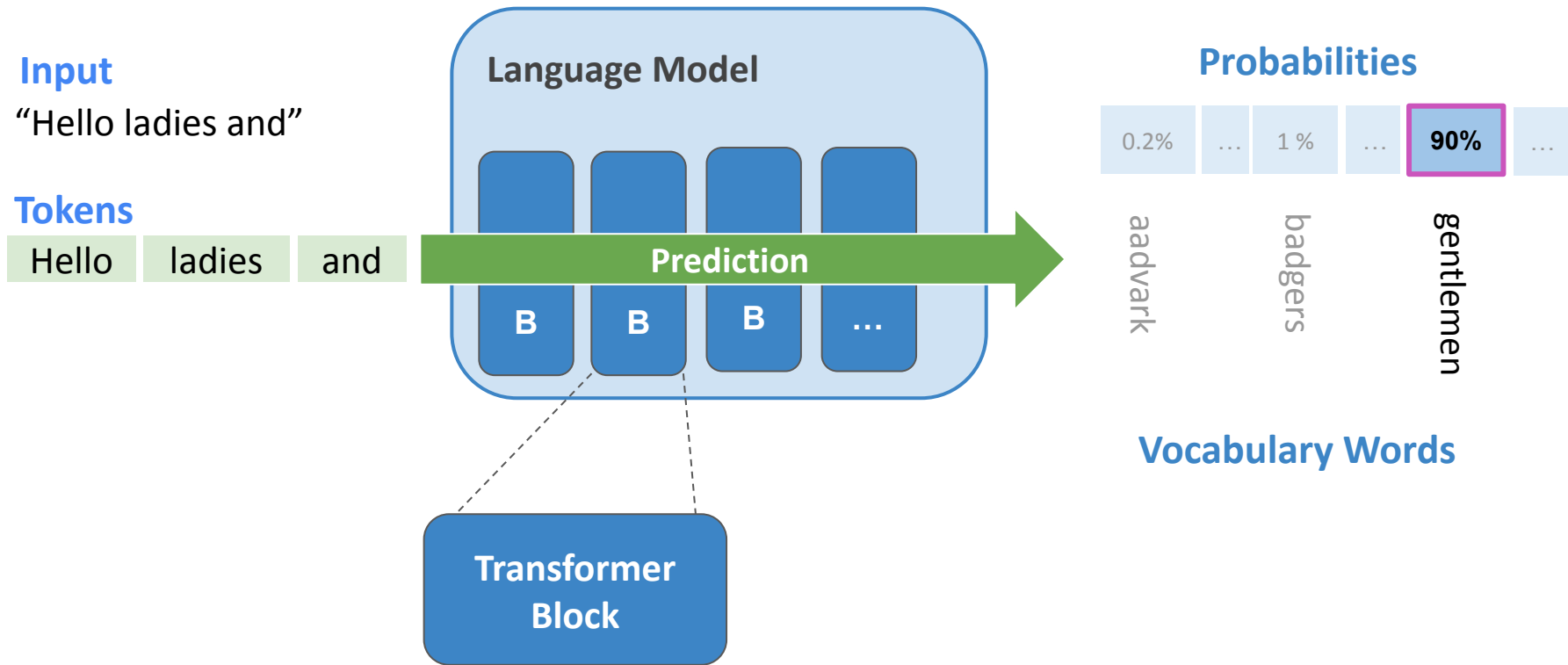
...

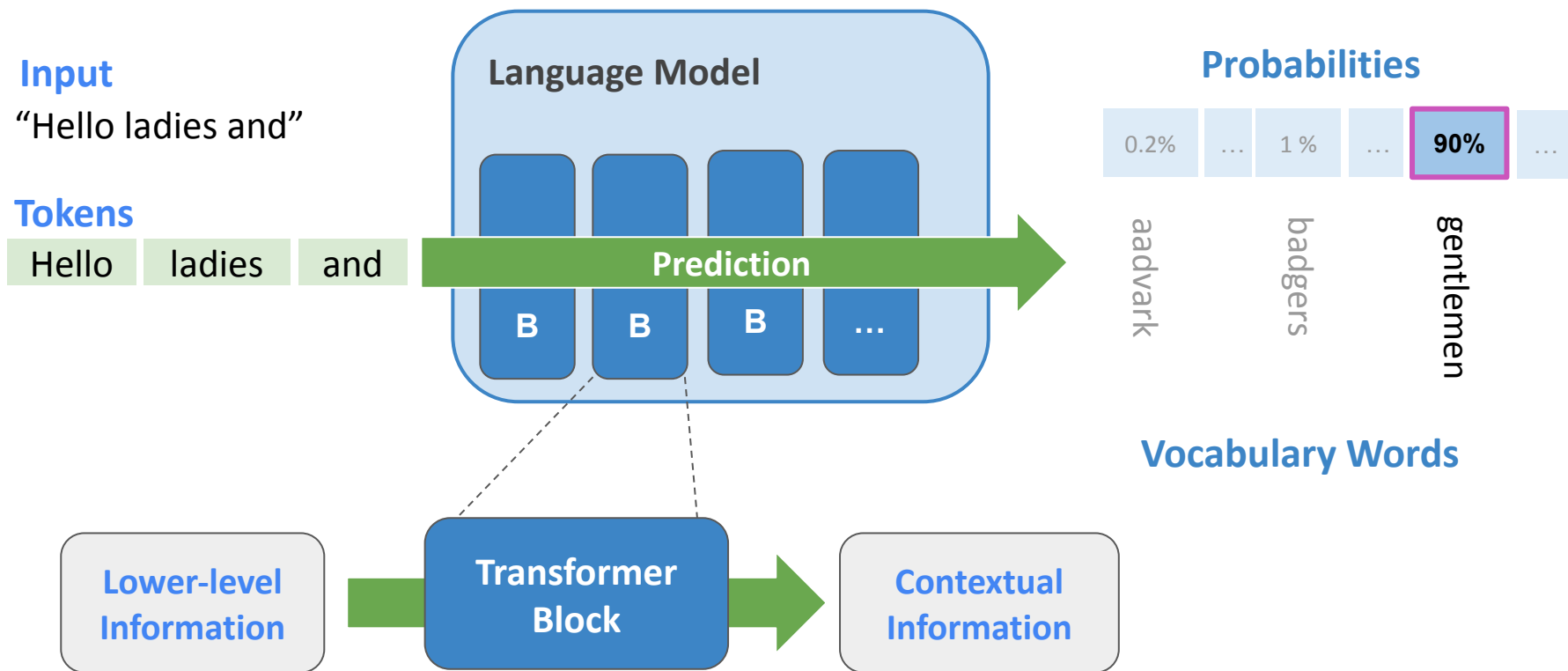
aadvark

badgers

gentlemen

Vocabulary Words





Pre-trained

Generative



Transformer

Pre-trained



Generative

GPT

Transformer

Pre-trained

Generative

GPT

Transformer



Feb 2019

Better Language Models and Their Implications

We've trained a large-scale unsupervised language model which generates coherent paragraphs of text, achieves state-of-the-art performance on many language modeling benchmarks, and performs rudimentary reading comprehension, machine translation, question answering, and summarization—all without task-specific training.

GPT-2



Better Language Models and Their Implications

We've trained a large language model that generates coherent text, performs rudimentary translation, question-answering, and task-specific training.

"Due to our concerns about malicious applications of the technology, we are not releasing the trained model"

Nov 2019



INDEPENDENT

Tech

AI deemed 'too dangerous to release' makes it out into the world

Extremists could generate 'synthetic propaganda', automatically creating white supremacist screeds, researchers warn

Andrew Griffin • Thursday 07 November 2019 12:21 •  Comments

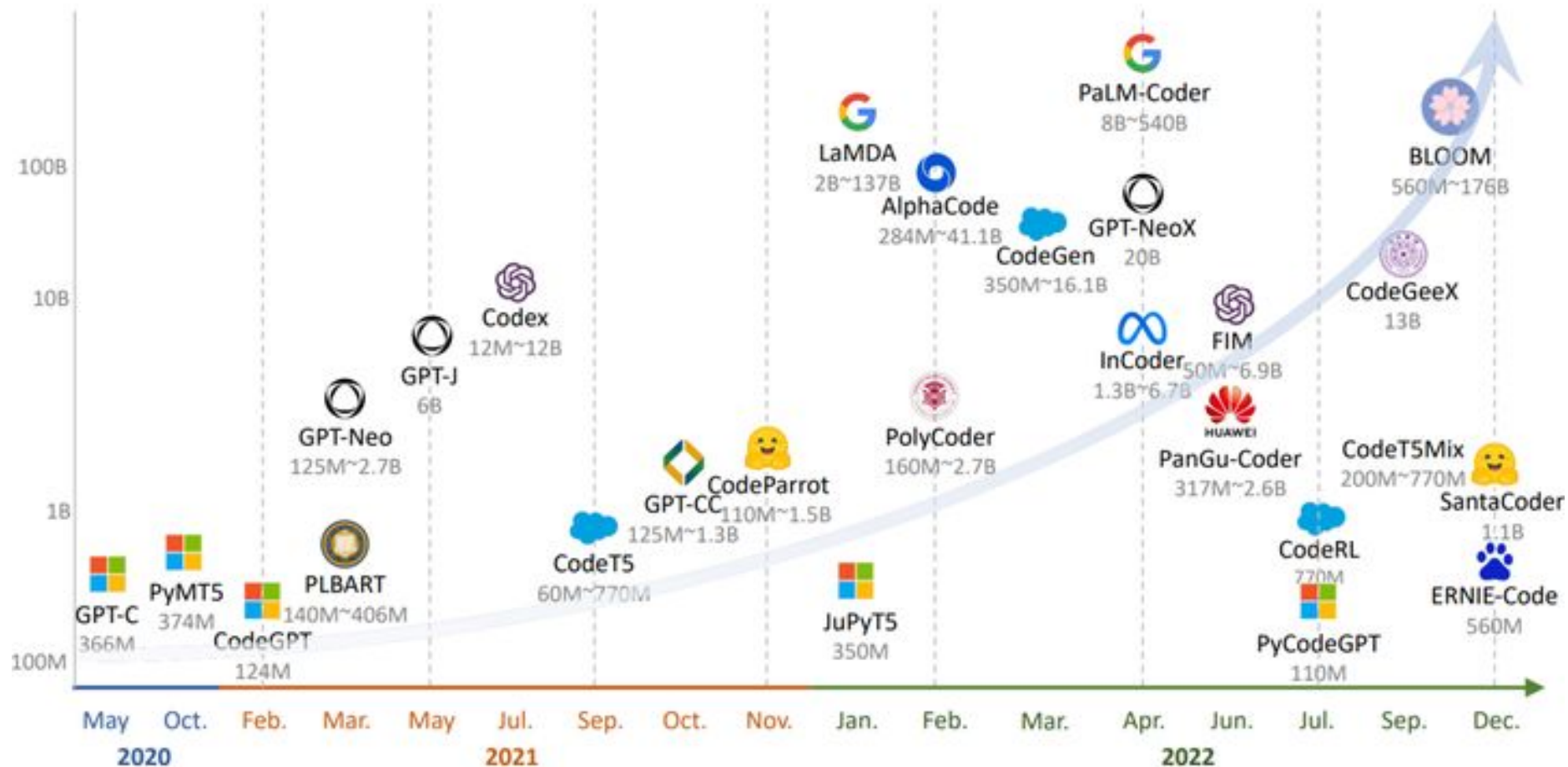


The University of Groningen is

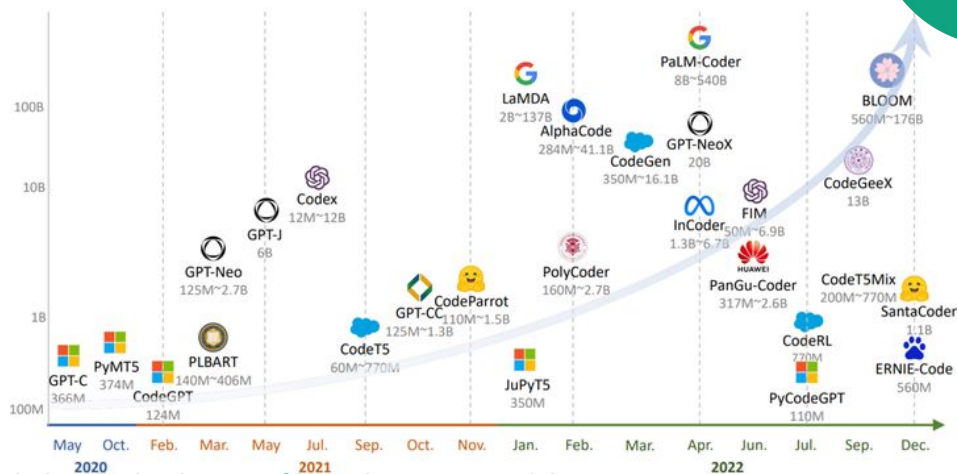
The University of Groningen is looking for candidates. A PhD in Applied Sciences does not qualify. Qualifications in English in this range may be preferred. Full details are found below. A university teaching position requires that you speak Norwegian.

You should receive a Bachelor's degree in English, the University of Groningen, from at least the beginning of a year and a minimum of eight years of teaching experience and proficiency in Norwegian. There should be at least 30 hours of lecture. A doctoral degree must...





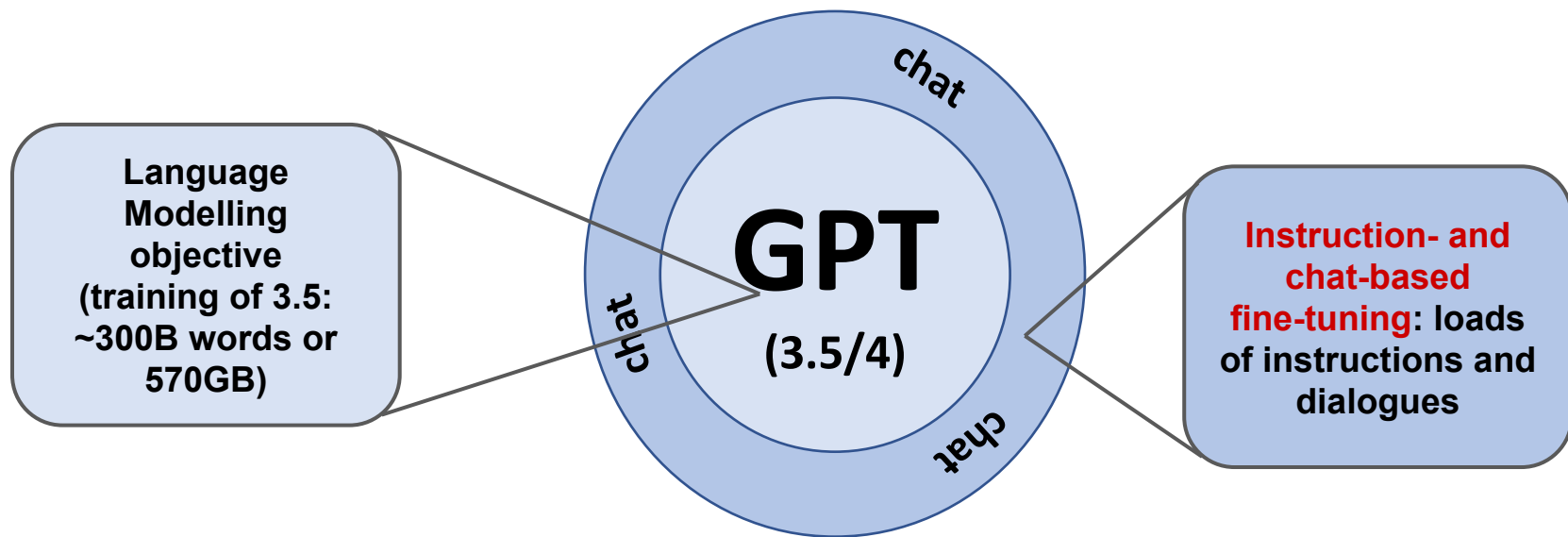
30 November 2022

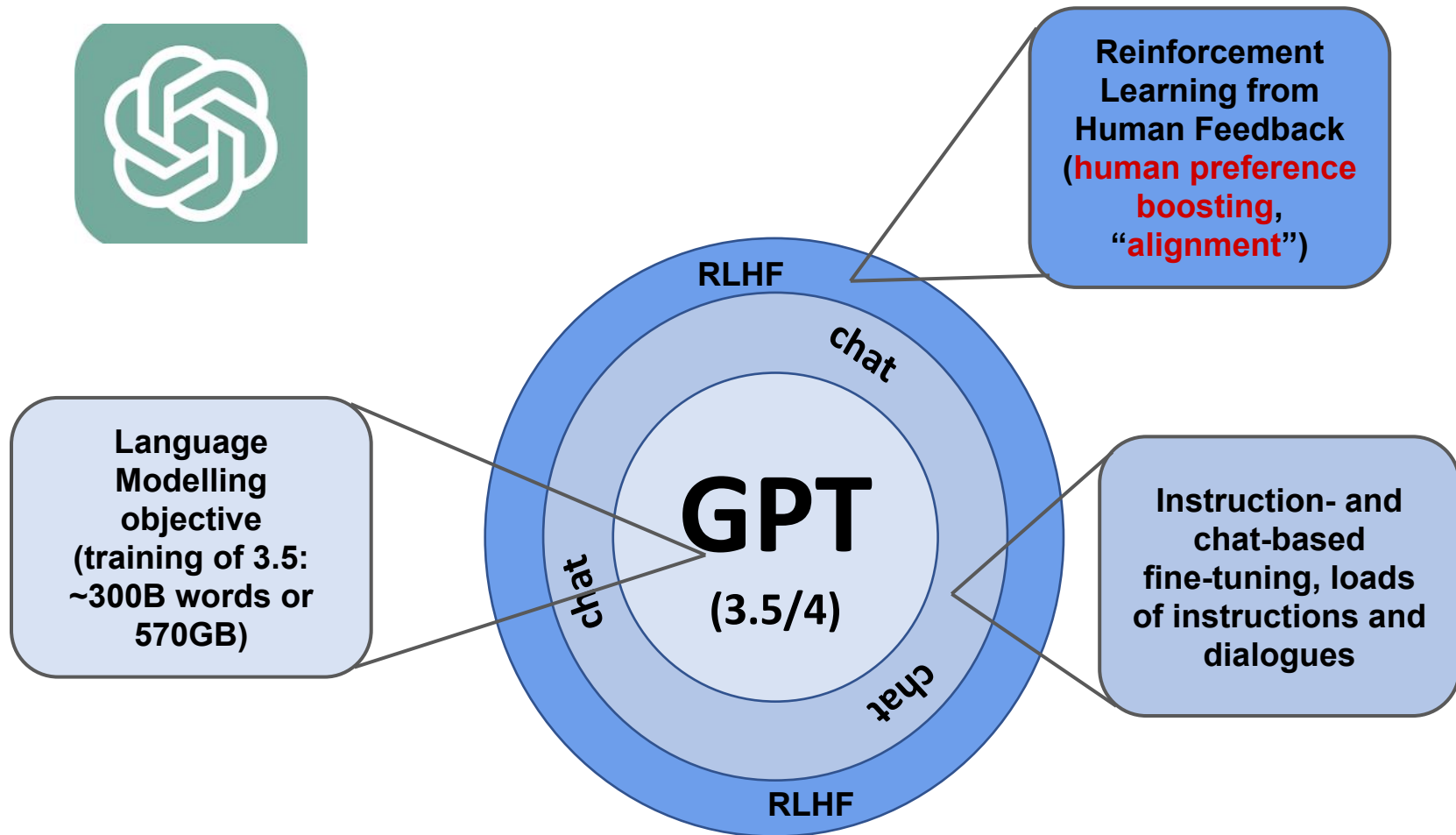




**Language
Modelling
objective**
(training of 3.5:
~300B words or
570GB)

GPT
(3.5/4)





What is a good (or better) model?

What is a good (or better) model?

Automatic Evaluation

on downstream tasks

General Multiple choice questions
and reasoning (MMLU + SQUAD)

Linguistic acceptability

Sentiment analysis

...

What is a **good (or better)** model?

Automatic Evaluation
on downstream tasks

General Multiple choice questions
and reasoning (MMLU + SQUAD)

Linguistic acceptability

Sentiment analysis

...

Human Judgement
on model outputs

“Does it sound natural to
you?”

“How likely is it that it was
written by a machine?”

“Which one do you prefer?”

MMLU

MEASURING MASSIVE MULTITASK LANGUAGE UNDERSTANDING

Dan Hendrycks
UC Berkeley

Collin Burns
Columbia University

Steven Basart
UChicago

Andy Zou
UC Berkeley

Mantas Mazeika
UIUC

Dawn Song
UC Berkeley

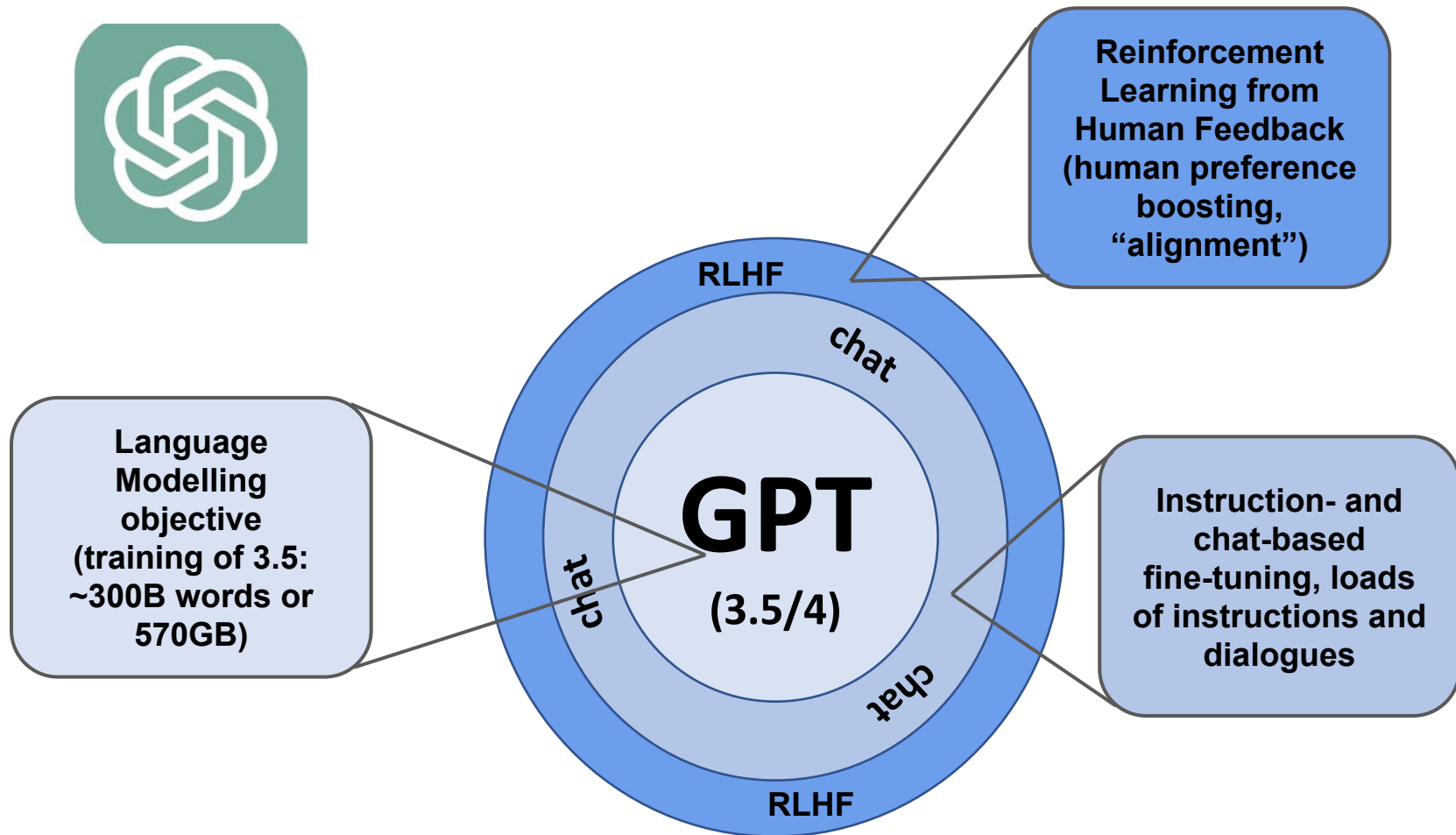
Jacob Steinhardt
UC Berkeley

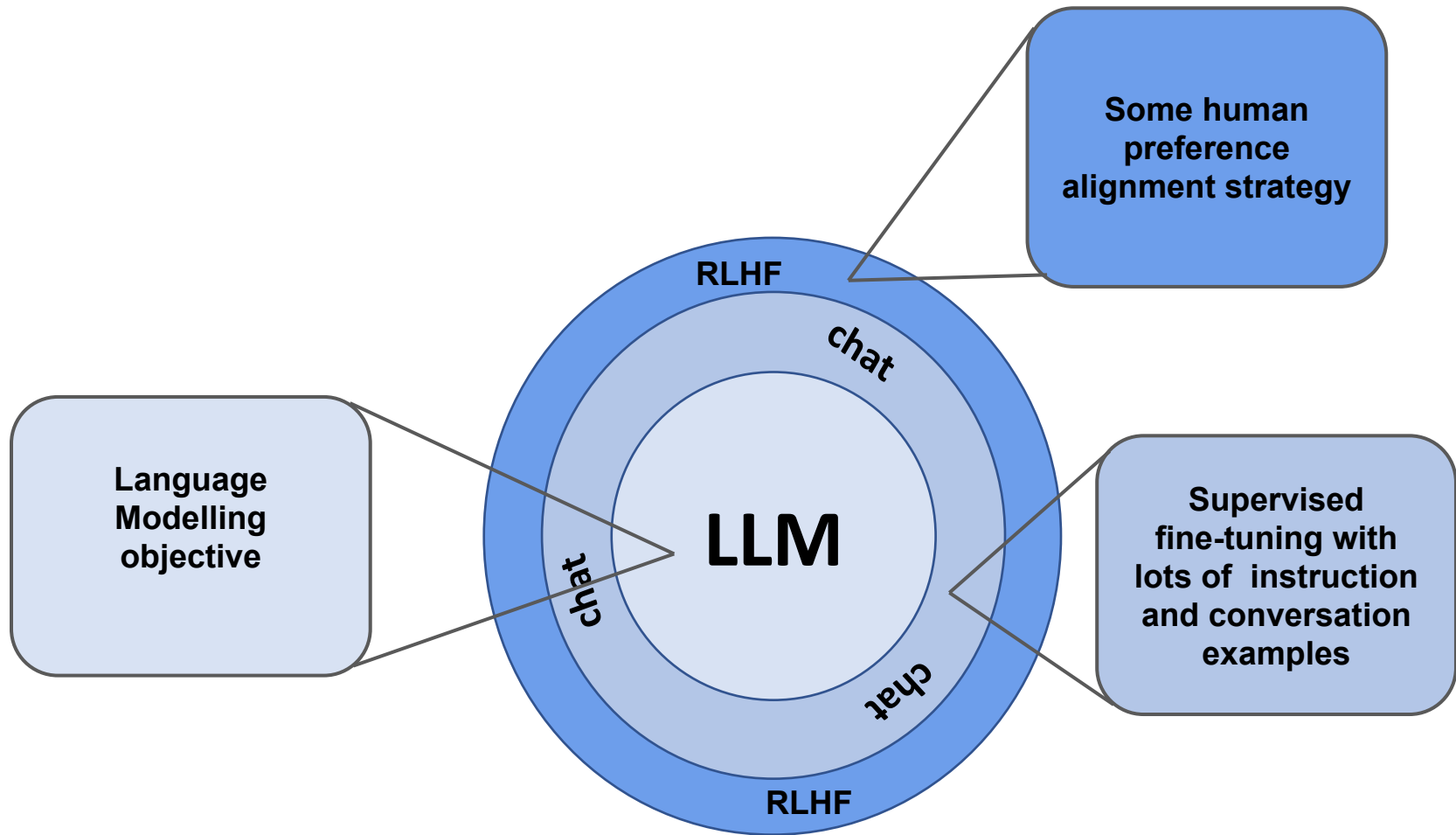
57 (factual/reasoning) tasks covering multiple topics:

- Maths
- US history
- Law
- Computer Science
- ...

SQUAD Dataset - Questions / Answers on given topics

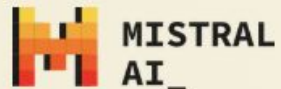
title string	context string	question string	answers sequence
Beyoncé	Beyoncé Giselle Knowles-Carter (/bi:'jɒnsər/ bee-YON-say) (born September 4, 1981) is an American singer, songwriter, record producer and actress. Born and raised in Houston, Texas, she performed in various singing and dancing competitions as a child, and rose to fame in the late 1990s as lead singer of R&B girl-group Destiny's Child. Managed by her father, Mathew Knowles, the group became one of the world's best-selling girl groups of all time. Their hiatus saw the release of Beyoncé's debut album, <i>Dangerously in Love</i> (2003), which established her as a solo artist worldwide, earned five Grammy Awards and featured the Billboard Hot 100 number-one singles "Crazy in Love" and "Baby Boy".	When did Beyonce start becoming popular?	{ "text": ["in the late 1990s"], "answer_start": [269] }
Beyoncé	Beyoncé Giselle Knowles-Carter (/bi:'jɒnsər/ bee-YON-say) (born September 4, 1981) is an American...	What areas did Beyonce compete in when she was growing up?	{ "text": ["singing and dancing"], "answer_start": [207] }
Beyoncé	Beyoncé Giselle Knowles-Carter (/bi:'jɒnsər/ bee-YON-say) (born September 4, 1981) is an American...	When did Beyonce leave Destiny's Child and become a solo singer?	{ "text": ["2003"], "answer_start": [526] }
Beyoncé	Beyoncé Giselle Knowles-Carter (/bi:'jɒnsər/ bee-YON-say) (born September 4, 1981) is an American...	In what city and state did Beyonce grow up?	{ "text": ["Houston, Texas"], "answer_start": [166] }
Beyoncé	Beyoncé Giselle Knowles-Carter (/bi:'jɒnsər/ bee-YON-say) (born September 4, 1981) is an American...	In which decade did Beyonce become famous?	{ "text": ["late 1990s"], "answer_start": [276] }
Beyoncé	Beyoncé Giselle Knowles-Carter (/bi:'jɒnsər/ bee-YON-say) (born September 4, 1981) is an American...	In what R&B group was she the lead singer?	{ "text": ["Destiny's Child"], "answer_start": [220] }







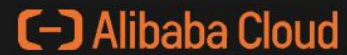
Llama



Mistral

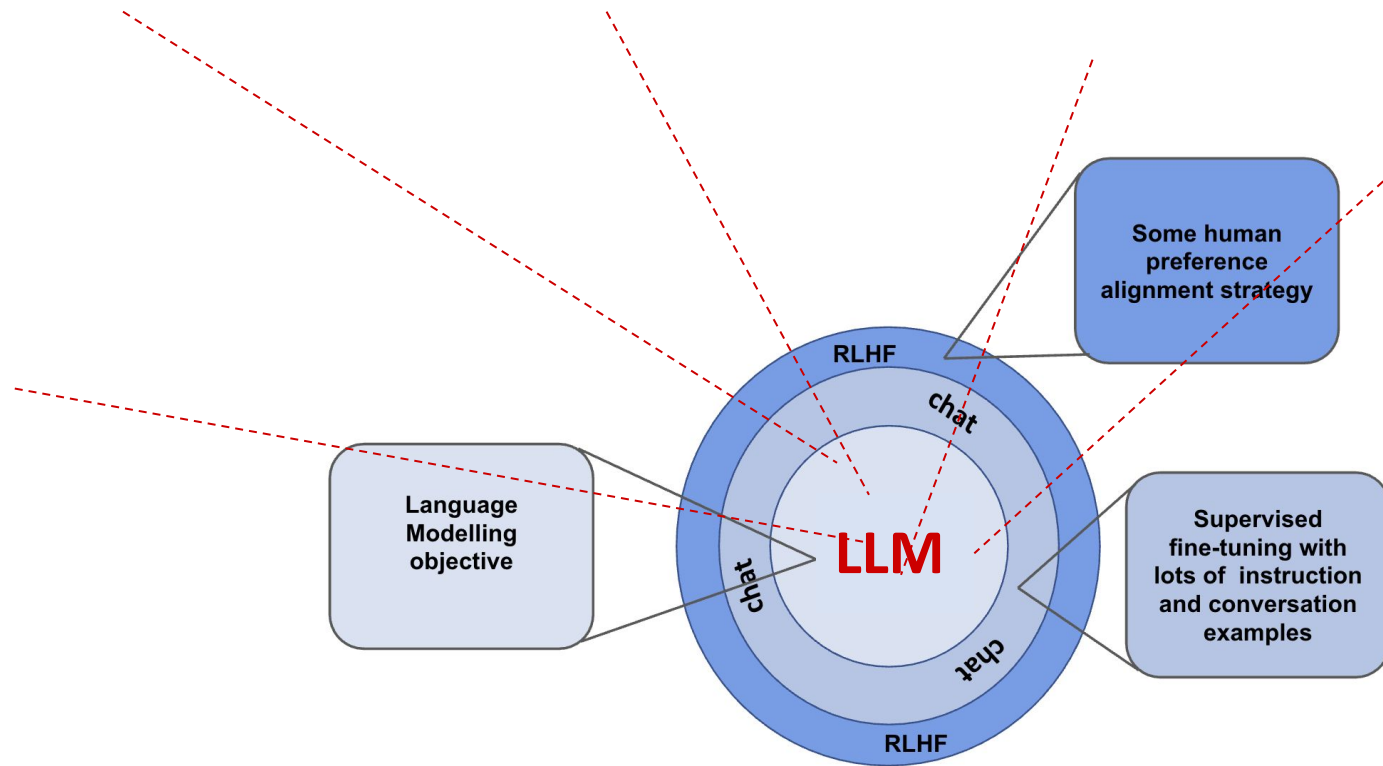


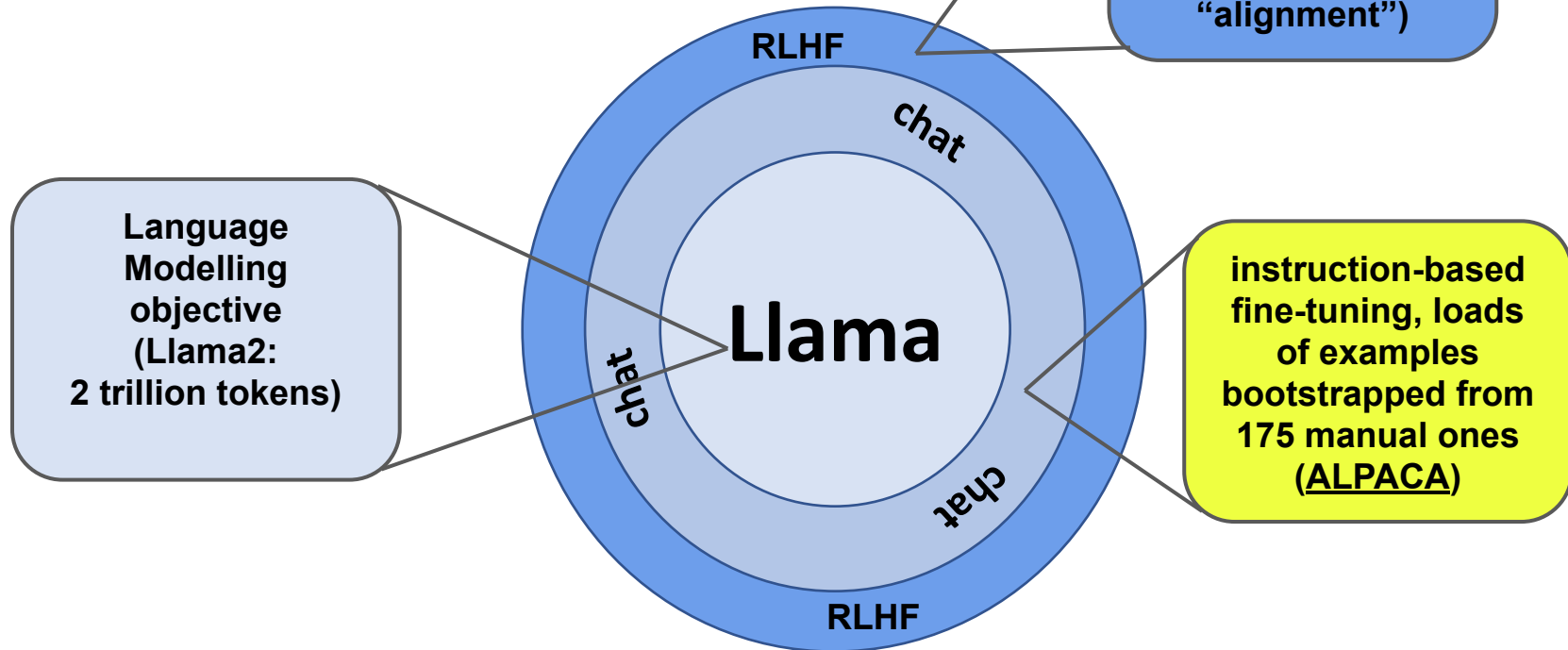
Phi



Qwen

...





2023

Camoscio: an Italian Instruction-tuned LLaMA

Andrea Santilli¹, Emanuele Rodolà¹

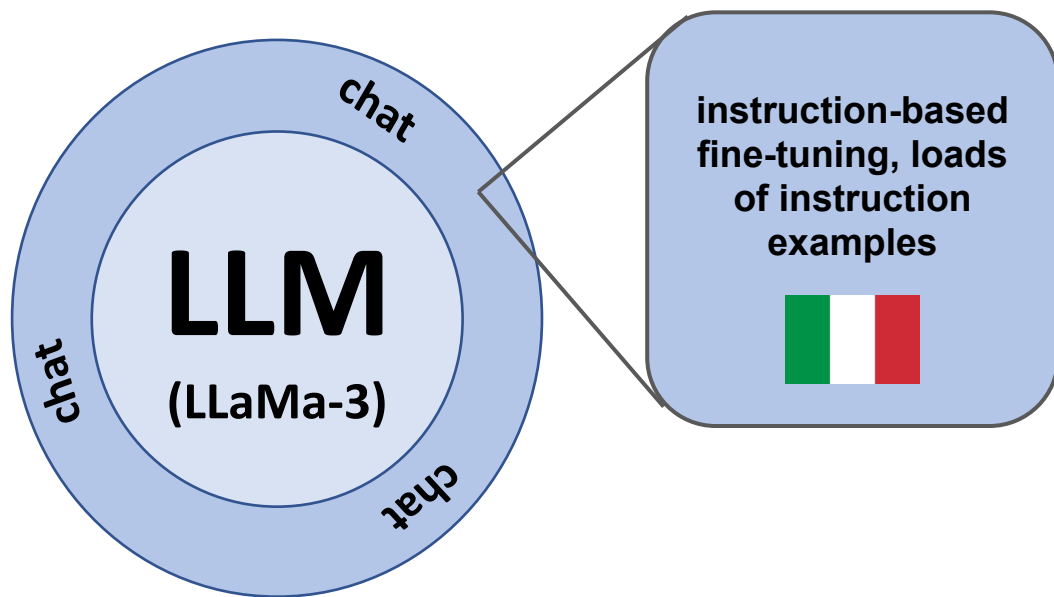
¹Sapienza University of Rome - Computer Science Department

2023

Camoscio: an Italian Instruction-tuned LLaMA

Andrea Santilli¹, Emanuele Rodolà¹

¹*Sapienza University of Rome - Computer Science Department*

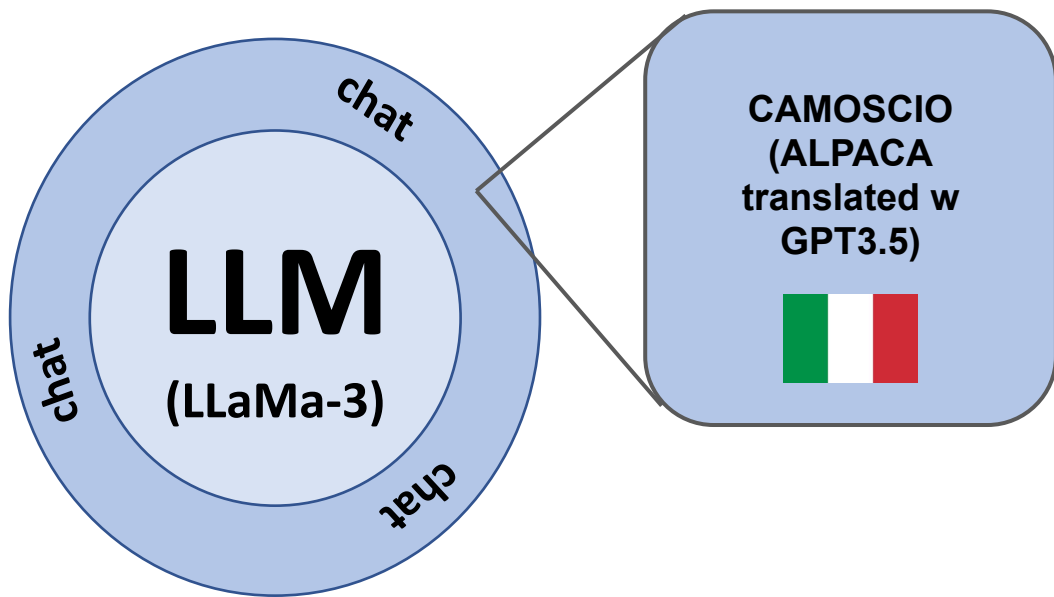


2023

Camoscio: an Italian Instruction-tuned LLaMA

Andrea Santilli¹, Emanuele Rodolà¹

¹Sapienza University of Rome - Computer Science Department



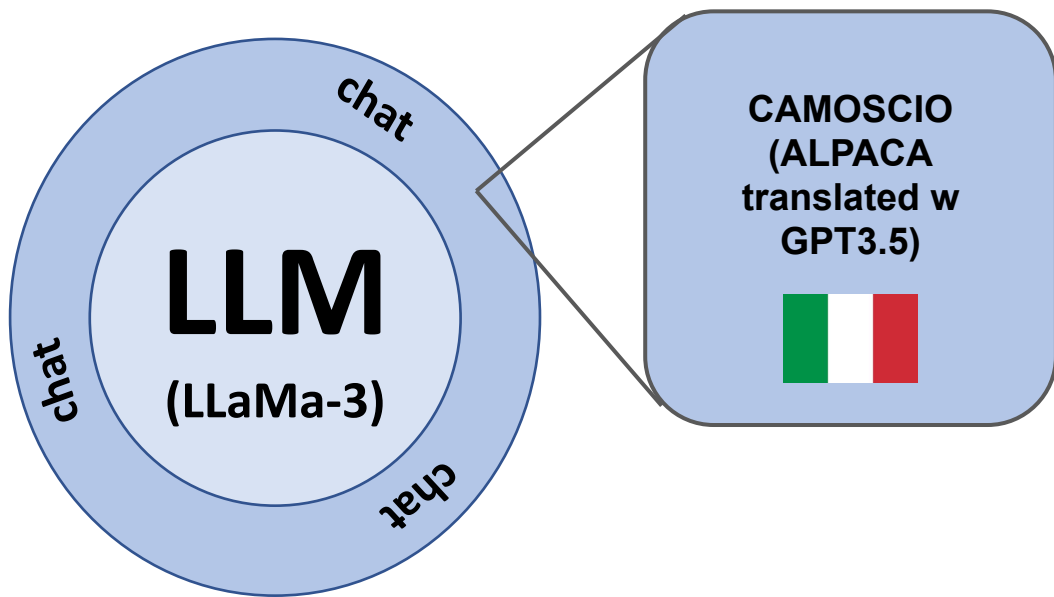
- We introduce an instruction-tuning dataset for the Italian language, stemming from the Stanford Alpaca [10] dataset, translating it to Italian.

2023

Camoscio: an Italian Instruction-tuned LLaMA

Andrea Santilli¹, Emanuele Rodolà¹

¹Sapienza University of Rome - Computer Science Department



- We introduce an instruction-tuning dataset for the Italian language, stemming from the Stanford Alpaca [10] dataset, translating it to Italian.
- We train Camoscio on this dataset and evaluate its zero-shot performance on several downstream tasks for the Italian language (NewsSum-IT, SQuAD-IT, XFORMAL IT).

Automatic Evaluation
on downstream tasks

2023/
2024

GEITje 7B Ultra: A Conversational Model for Dutch

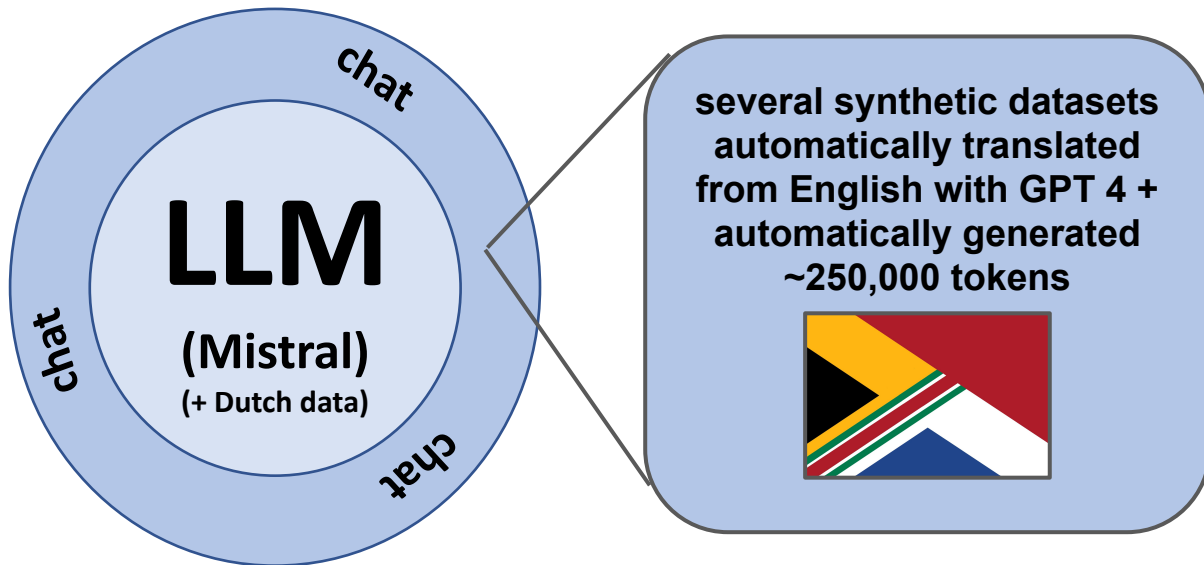
Bram Vanroy
KU Leuven, Dutch Language Institute
Leuven, Belgium
`bram.vanroy@kuleuven.be`



2024

GEITje 7B Ultra: A Conversational Model for Dutch

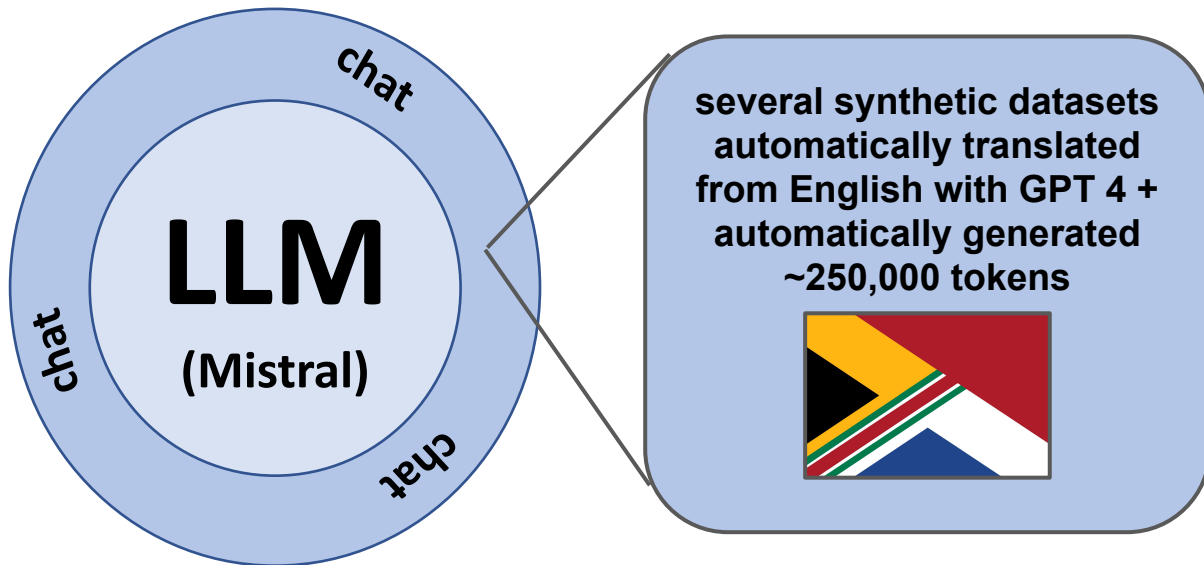
Bram Vanroy
KU Leuven, Dutch Language Institute
Leuven, Belgium
bram.vanroy@kuleuven.be



2024

GEITje 7B Ultra: A Conversational Model for Dutch

Bram Vanroy
KU Leuven, Dutch Language Institute
Leuven, Belgium
bram.vanroy@kuleuven.be



Automatic Evaluation on downstream tasks

- General Multiple choice questions (MMLU + SQUAD)
- Story continuation
- Sentiment analysis
-

(Synthetic + native)

2024

Fietje: An open, efficient LLM for Dutch

Bram Vanroy^{1,2}

BRAM.VANROY@KULEUVEN.BE

¹*KU Leuven, Blijde Inkomststraat 21, 3000 Leuven, Belgium*

²*Dutch Language Institute, Rapenburg 61, 2311 GJ Leiden, The Netherlands*

Fietje

An open, efficient LLM for Dutch



2023

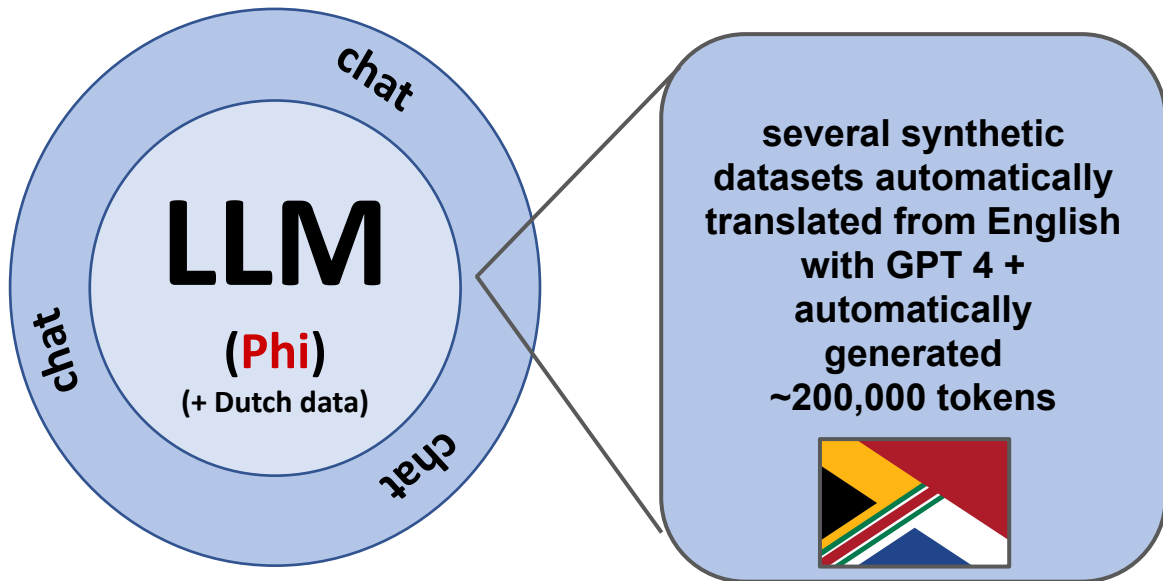
Fietje: An open, efficient LLM for Dutch

Bram Vanroy^{1,2}

BRAM.VANROY@KULEUVEN.BE

¹*KU Leuven, Blijde Inkomststraat 21, 3000 Leuven, Belgium*

²*Dutch Language Institute, Rapenburg 61, 2311 GJ Leiden, The Netherlands*



Automatic Evaluation on downstream tasks

- General Multiple choice questions (MMLU + ARC)
- Linguistic acceptability judgements
- Sentiment analysis
- ...

(Synthetic + native)

future

GPT-NL

TNO innovation
for life



Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

SURF

future

GPT-NL

Home

Contact



Over GPT-NL

Commitments

Samenwerken

Planning

Veelgestelde vragen

Nieuws ▾

Home

Open, transparant en soeverein

In Nederland wordt gewerkt aan de ontwikkeling van een eigen open taalmodel genaamd GPT-NL. Dit model is essentieel voor het versterken en behouden van onze digitale soevereiniteit.

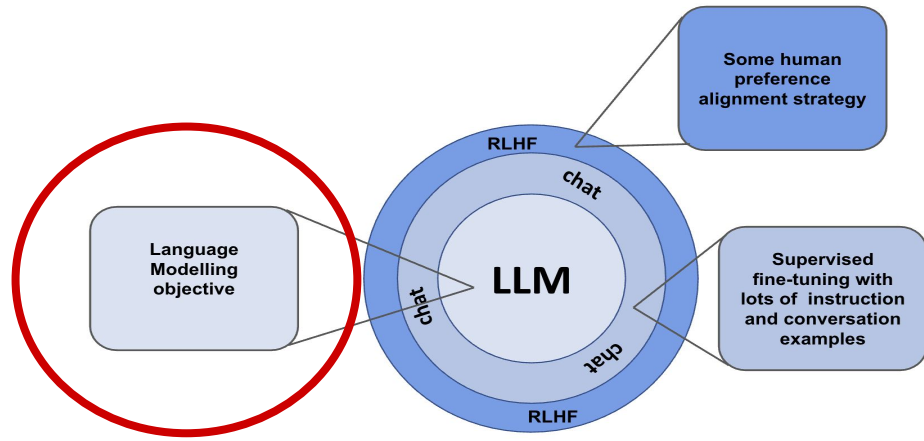
Wilt u meebouwen?





summary

	(pre-)training	instruction-tuning	evaluation
Early small GPTs (2020)	Dutch from scratch (Wikipedia + Web)	-	Human, <i>naturalness</i> of generated text
Camoscio (2023)	Llama-based	Machine-translated ITA instructions (Alpaca)	Downstream (Machine translated)
GEITje (2023?)	Mistral-7B + Dutch data (GigaCorpus and MADLAD-400 (10B tokens))	Ultrafeedback Dutch No Robots Dutch Alpaca Dutch (translated)	Downstream (translated versions of SQuaD, MMLU, HellaSwag, etc)
Fietje (2024)	Phi-based + CulturaX and Wiki data (28B tokens)	Ultrafeedback Dutch No Robots Dutch Belebele (Multi, Dutch selected)	Downstream (translated versions of MMLU, ARC, plus some native Dutch)
GPT-NL (?)	Dutch from scratch?		



pre-trained models

	(pre-)training	instruction-tuning	evaluation
Early small GPTs (2020)	Dutch from scratch (Wikipedia + Web)	-	Human, <i>naturalness</i> of generated text
Camoscio (2023)	Llama-based	Machine-translated ITA instructions (Alpaca)	Downstream (Machine translated)
GEITje (2023?)	Mistral-7B + Dutch data (+ GigaCorpus and MADLAD-400 (10B tokens))	Ultrafeedback Dutch No Robots Dutch Alpaca Dutch (translated)	Downstream (translated versions of SQuaD, MMLU, HellaSwag, etc)
Fietje (2024)	Phi-based + Dutch data (+ CulturaX and Wiki data (28B tokens))	Ultrafeedback Dutch No Robots Dutch Belebele (Multi, Dutch selected)	Downstream (translated versions of MMLU, ARC, plus some native Dutch)
GPT-NL (?)	Dutch from scratch?		



Llama2:

**2 trillion
tokens**



Meta Warns Its Latest Large Language Model 'May Not Be Suitable' for Non-English Use

19 Jul 2023

Llama2:

**2 trillion
tokens**

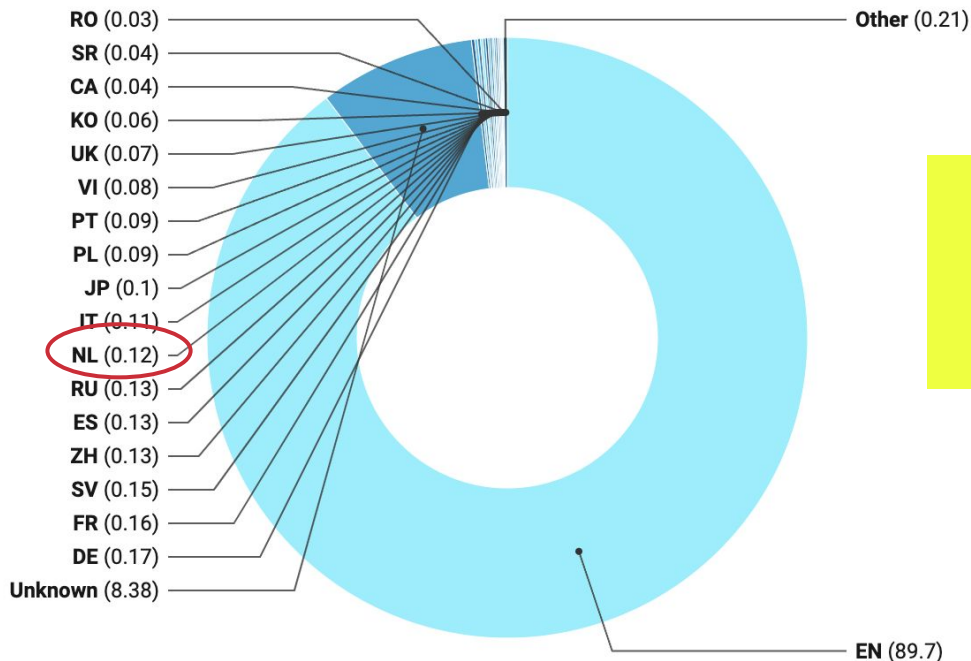


Meta Warns Its Latest Large Language Model 'May Not Be Suitable' for Non-English Use

19 Jul 2023

Llama 2's Language distribution in pretraining data with percentage

The distribution of languages in Llama 2's corpus, subsetting to those found in more than 0.005% of the documents.



Llama2:

2 trillion
tokens

NL: 0.1%



April 23, 2024

Tiny but mighty: The Phi-3
small language models with
big potential



April 23, 2024

Tiny but mighty: The **Phi-3**
small language models with
big potential





April 23, 2024

Tiny but mighty: The **Phi-3**
small language models with
big potential





Multi-lingual Capability

April 23, 2024

Tiny but mighty: The **Phi-3**
small language models with
big potential

Phi-3.5-mini is our latest 3.8B model update. The model used additional continual pre-training and post-training data leading to substantial gains on multi-lingual, multi-turn conversation quality, and reasoning capability. The model has been trained on selective set of languages listed here: Arabic, Chinese, Czech, Danish, **Dutch**, English, Finnish, French, German, Hebrew, Hungarian, Italian, Japanese, Korean, Norwegian, Polish, Portuguese, Russian, Spanish, Swedish, Thai, Turkish and Ukrainian.



April 23, 2024

Tiny but mighty: The Phi-3 small language models with big potential

*Phi
training
data*

To date, the standard way to train large language models has been to use massive amounts of data from the internet. This was thought to be the only way to meet this type of model's huge appetite for content, which it needs to "learn" to understand the nuances of language and generate intelligent answers to user prompts. But

Microsoft researchers had a different idea.



April 23, 2024

Tiny but mighty: The Phi-3 small language models with big potential

*Phi
training
data*

To date, the standard way to train large language models has been to use massive amounts of data from the internet. This was thought to be the only way to meet this type of model's huge appetite for content, which it needs to "learn" to understand the nuances of language and generate intelligent answers to user prompts. But

Microsoft researchers had a different idea.

"Instead of training on just raw web data, why don't you look for data which is of extremely high quality?" asked Sebastien Bubeck, Microsoft vice president of generative AI research who has led the company's efforts to develop more capable small language models. But where to focus?



April 23, 2024

Tiny but mighty: The Phi-3 small language models with big potential

*Phi
training
data*

To date, the standard way to train large language models has been to use massive amounts of data from the internet. This was thought to be the only way to meet this type of model's huge appetite for content, which it needs to "learn" to understand the nuances of language and generate intelligent answers to user prompts. But Microsoft researchers had a different idea.

"Instead of training on just raw web data, why don't you look for data which is of extremely high quality?" asked Sebastien Bubeck, Microsoft vice president of generative AI research who has led the company's efforts to develop more capable small language models. But where to focus?

Inspired by Eldan's nightly reading ritual with his daughter, Microsoft researchers decided to create a discrete dataset starting with 3,000 words – including a roughly equal number of nouns, verbs and adjectives. Then they asked a large language model to create a children's story using one noun, one verb and one adjective from the list – a prompt they repeated millions of times over several days, generating millions of tiny children's stories.

A yellow sun icon with a black outline, positioned in the top left corner of the slide.

future

GPT-NL

TNO innovation
for life

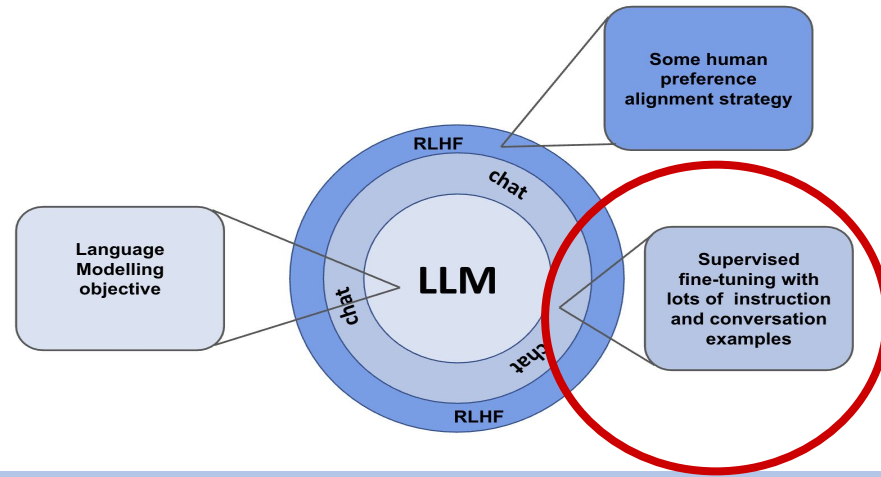


Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

SURF

Om een inschatting te maken van hoeveel data nodig is om een model vanaf nul te trainen, kijken we naar eerder ontwikkelde taalmodellen. Alle betere taalmodellen van de afgelopen jaren zitten tussen de 300 en 15.000 miljard tokens aan trainingsdata, dus 300 miljard teksttokens is zo ongeveer het minimum voor GPT-NL. Dat is een gigantische hoeveelheid tekst, ongeveer zo veel Nederlandse tekst als nu op het internet is te vinden. Als je uitgaat van 300 miljard tokens, dan zijn dat 3 miljoen Harry Potter boeken of een stapel papier van 10 km hoog - dat is nog hoger dan de Mount Everest.

Naast de private data, aangeleverd door de content board, moeten we de dataset voor GPT-NL met andere bronnen aanvullen. Hiervoor gebruiken we ook Engelstalige data, Germaanse data en code, verkregen via publiek beschikbare bronnen of synthetisch gegenereerd.



instruction data

	(pre-)training	instruction-tuning	evaluation
Early small GPTs (2020)	Dutch from scratch (Wikipedia + Web)	-	Human, <i>naturalness</i> of generated text
Camoscio (2023)	Llama-based	Machine-translated ITA instructions (Alpaca)	Downstream (Machine translated)
GEITje (2023/2024)	Mistral-7B (+ GigaCorpus and MADLAD-400 (10B tokens))	Ultrafeedback Dutch No Robots Dutch Alpaca Dutch (translated)	Downstream (translated versions of SQuaD, MMLU, HellaSwag, etc)
Fietje (2024)	Phi-based (+ CulturaX and Wiki data (28B tokens))	Ultrafeedback Dutch No Robots Dutch Belebele (Multi, Dutch selected)	Downstream (translated versions of MMLU, ARC, plus some native Dutch)
GPT-NL (?)	Dutch from scratch?		

Instruction Dataset Name	Nature	%
ultrachat 200k dutch	gpt-4-turbo; multi-turn; generated	85.42
no robots dutch	gpt-4-turbo; prompt translate, answer generated	2.20
stackoverflow-chat-dutch	gpt-3.5-turbo; multi-turn; code; translated; only 50% used	8.38
alpaca-cleaned-dutch	gpt-3.5-turbo; translated	2.62
dolly-15k-dutch	gpt-3.5-turbo; translated	1.39

Instruction Dataset Name	Nature	%
ultrachat 200k dutch	gpt-4-turbo; multi-turn; generated	85.42

dataset).² Using synthetic data, either translated or generated from-scratch with a different LLM, is common, even for English, because handcrafting conversations for large volumes of data is expensive and time-consuming – although not impossible (e.g., Rajani et al., 2023; Conover et al., 2023).

	translated; only 50% used	
alpaca-cleaned-dutch	gpt-3.5-turbo; translated	2.62
dolly-15k-dutch	gpt-3.5-turbo; translated	1.39

```
{"id": 10547, "instruction": "Ontdek de SMS-/teksttaal in het  
gegeven bericht en vertaal het naar formeel Engels.", "input":  
"Wag1 b snel, be4 we miss d trein!", "output": "Waarschuwing:  
Wees snel, voordat we de trein missen!"}
```

```
{"id": 10237, "instruction": "Maak een anagram van de  
woorden.", "input": "conform", "output": "fromc"}
```

```
{"id": 43856, "instruction": "Genereer een parodie kop met een  
woordspeling", "input": "", "output": "\"Kaas boos: Publiek  
woedend over stijgende zuivel prijzen\""}
```

```
{"id": 43829, "instruction": "Voorspel wat een klant zou kunnen bestellen in een caf\u00e9.", "input": "Het is 7:30 uur op een zaterdagochtend.", "output": "De klant zal waarschijnlijk een warm drankje zoals koffie bestellen, een ontbijtitem zoals bacon en eieren, toast of havermout, en een ontbijtgebak zoals een muffin of croissant. Ze kunnen ook iets kouds bestellen, zoals een smoothie of sap. Als het een etablissement is dat alcohol serveert, kunnen ze ook een mimosa of bloody mary bestellen."}
```

```
{"id": 11057, "instruction": "Genereer een gedetailleerde boodschappenlijst.", "input": "Huishouden van vier volwassenen.", "output": "- 4 gallon melk\n- 2 dozijn grote eieren\n- 4 pakken brood\n- 2 pond boter\n- 2 pond bacon\n- 2 pond rundergehakt\n- 4 paprika's\n- 2 pond broccoli\n- 1 pond uien\n- 2 pond wortelen\n- 1 pond bleekselderij\n- 4 blikken ananas\n- 4 blikken zwarte bonen\n- 1 grote doos ontbijtgranen\n- 1 grote zak appels\n- 1 grote zak sinaasappels\n- 2 pakken yoghurt\n- 4 pakken plakjes kaas\n- 1 pak tortilla's\n- 2 pakken bevroren groenten\n- 1 pot pindakaas\n- 1 fles olijfolie"}
```

```
[ { "content": "Kun je me wat thema's voorstellen die we kunnen overwegen voor het aanstaande verlovingsfeest van mijn vrienden?", "role": "user" }, { "content": "Natuurlijk! Een verlovingsfeest is een prachtige gelegenheid die de liefde en de toekomstige verbintenis van een paar viert. Hier zijn enkele thema-ideeën die je kunt overwegen voor het verlovingsfeest van je vrienden:\n\n1. **Tuinfeest**\n Een ontspannen feest in de buitenlucht, met veel bloemen, lichtsnoeren en een sfeer van een picknick. Dit kan zowel overdag als 's avonds worden gedaan.\n\n2. **Wijngaard Elegantie**\n Voor een meer verfijnde sfeer, host het feest bij een lokale wijngaard of creëer een wijngerelateerd thema waar wijnproeven centraal staat.\n\n3
```

Instruction Dataset Name	Nature	%
ultrachat 200k dutch	gpt-4-turbo; multi-turn; generated	85.42
no robots dutch	gpt-4-turbo; prompt translate, answer generated	2.20
stackoverflow-chat-dutch	gpt-3.5-turbo; multi-turn; code; translated; only 50% used	8.38
alpaca-cleaned-dutch	gpt-3.5-turbo; translated	2.62
dolly-15k-dutch	gpt-3.5-turbo; translated	1.39

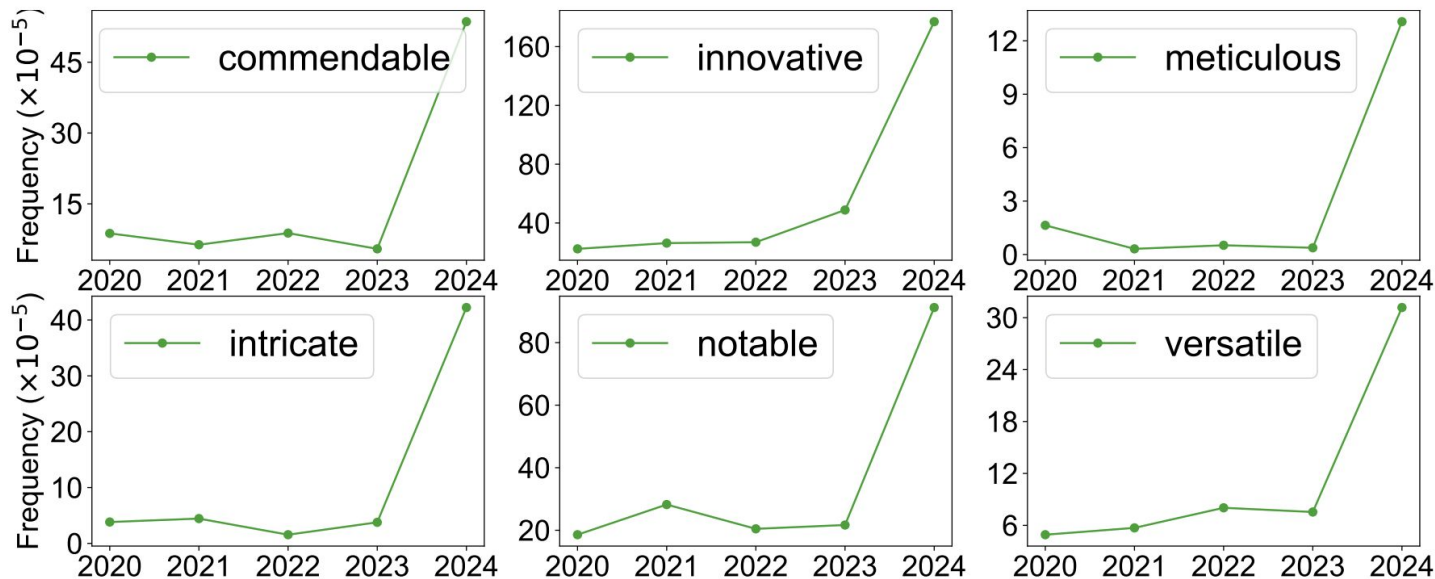


Figure 1: Shift in Adjective Frequency in *ICLR* 2024 Peer Reviews. We find a significant shift in the frequency of certain tokens in *ICLR* 2024, with adjectives such as “commendable”, “meticulous”, and “intricate” showing 9.8,

evaluation

	(pre-)training	instruction-tuning	evaluation
Early small GPTs (2020)	Dutch from scratch (Wikipedia + Web)	-	Human, <i>naturalness</i> of generated text
Camoscio (2023)	Llama-based	Machine-translated ITA instructions (Alpaca)	Downstream (Machine translated)
GEITje (2023?)	Mistral-7B (+ GigaCorpus and MADLAD-400 (10B tokens))	Ultrafeedback Dutch No Robots Dutch Alpaca Dutch (translated)	Downstream (translated versions of SQuAD, MMLU, HellaSwag, plus some native Dutch)
Fietje (2024)	Phi-based (+ CulturaX and Wiki data (28B tokens))	Ultrafeedback Dutch No Robots Dutch Belebele (Multi, Dutch selected)	Downstream (translated versions of MMLU, ARC, plus some native Dutch)
GPT-NL (?)	Dutch from scratch?		

SQUAD Dataset - Questions / Answers on given topics, in Dutch

title string · lengths	title_en string · classes	context string · lengths	question string · lengths	answers sequence
 1-19 70.6%	 Beyoncé 0.6%	 479-845 55.4%	 2-112 96.2%	
		het eind van de jaren negentig beroemd geworden als leadzangeres van R&B-meisjesgroep Destiny's Child. Onder leiding van haar vader, Mathew Knowles, werd de groep een van de best verkopende meisjesgroepen aller tijden. Hun hiaat leidde tot de release van Beyoncé's debuutalbum Dangerous in Love (2003), waardoor ze wereldwijd een soloartieste werd, vijf Grammy Awards kreeg en de nummer één van de Billboard Hot 100-hits 'Crazy in Love' en 'Baby Boy' kreeg.		
Beyoncé	Beyoncé	Beyoncé Giselle Knowles-Carter (bijonseï/bee-Yon-say) is een...	In welke gebieden wedijverde Beyoncé toen ze opgroeide?	{ "text": ["zingen en dansen"], "text_en": ["singing and..."] }
Beyoncé	Beyoncé	Beyoncé Giselle Knowles-Carter (bijonseï/bee-Yon-say) is een...	Wanneer heeft Beyoncé Destiny's Child verlaten en is ze...	{ "text": ["2003"], "text_en": ["2003"], "answer_start_en": [...] }
Beyoncé	Beyoncé	Beyoncé Giselle Knowles-Carter (bijonseï/bee-Yon-say) is een...	In welke stad en in welke staat is Beyoncé opgegroeid?	{ "text": ["Houston, Texas"], "text_en": ["Houston, Texas"], ... }
Beyoncé	Beyoncé	Beyoncé Giselle Knowles-Carter (bijonseï/bee-Yon-say) is een...	In welk decennium is Beyoncé beroemd geworden?	{ "text": ["eind jaren negentig"], "text_en": ["late 1990s"], ... }

Beyoncé Giselle Knowles-Carter (bijonsei/bee-Yon-say) is een Amerikaanse zangeres, songwriter, platenproducer en actrice. Geboren en getogen in Houston, Texas, heeft ze als kind meegedaan aan diverse zang- en danswedstrijden, en is aan het eind van de jaren negentig beroemd geworden als leadzangeres van R&B-meisjesgroep Destiny's Child. Onder leiding van haar vader, Mathew Knowles, werd de groep een van de best verkopende meisjesgroepen aller tijden. Hun hiaat leidde tot de release van Beyoncé's debuutalbum *Dangerous in Love* (2003), waardoor ze wereldwijd een soloartieste werd, vijf Grammy Awards kreeg en de nummer één van de Billboard Hot 100-hits 'Crazy in Love' en 'Baby Boy' kreeg.

context
string
479+845
het eind
beroemd
leadzang
meisjesg
Onder le
Mathew K
een van
meisjesg
hlaat le
Beyoncé's
in Love
wereldwi
werd, vi
en de nu
Billboar
Love' en
Beyoncé
(bijonse
Beyoncé
(bijonse
Beyoncé
(bijonse
Beyoncé
(bijonse
Previous 1

- Wanneer kondigde Destiny's Child hun hiaten aan?
- Met welke artieste duette Beyoncé in de single 'Deja Vu'?
- Wie collaboreerde met Beyoncé aan de single, Deja Vu?
- Independent Women Part I was on which's soundtrack of 2000?
- Beyonces eerste elpee heette in haar eentje wat?
- Welk beroemde oriëntatiepunt zag Beyoncé in China?



Chaos

bias

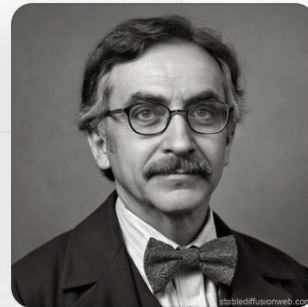
HUMANS ARE BIASED. GENERATIVE AI IS EVEN WORSE

Stable Diffusion's text-to-image model amplifies stereotypes about race and gender – here's why that matters

By Leonardo Nicoletti and Dina Bass for **Bloomberg Technology** + **Equality**

June 9, 2023

a professor

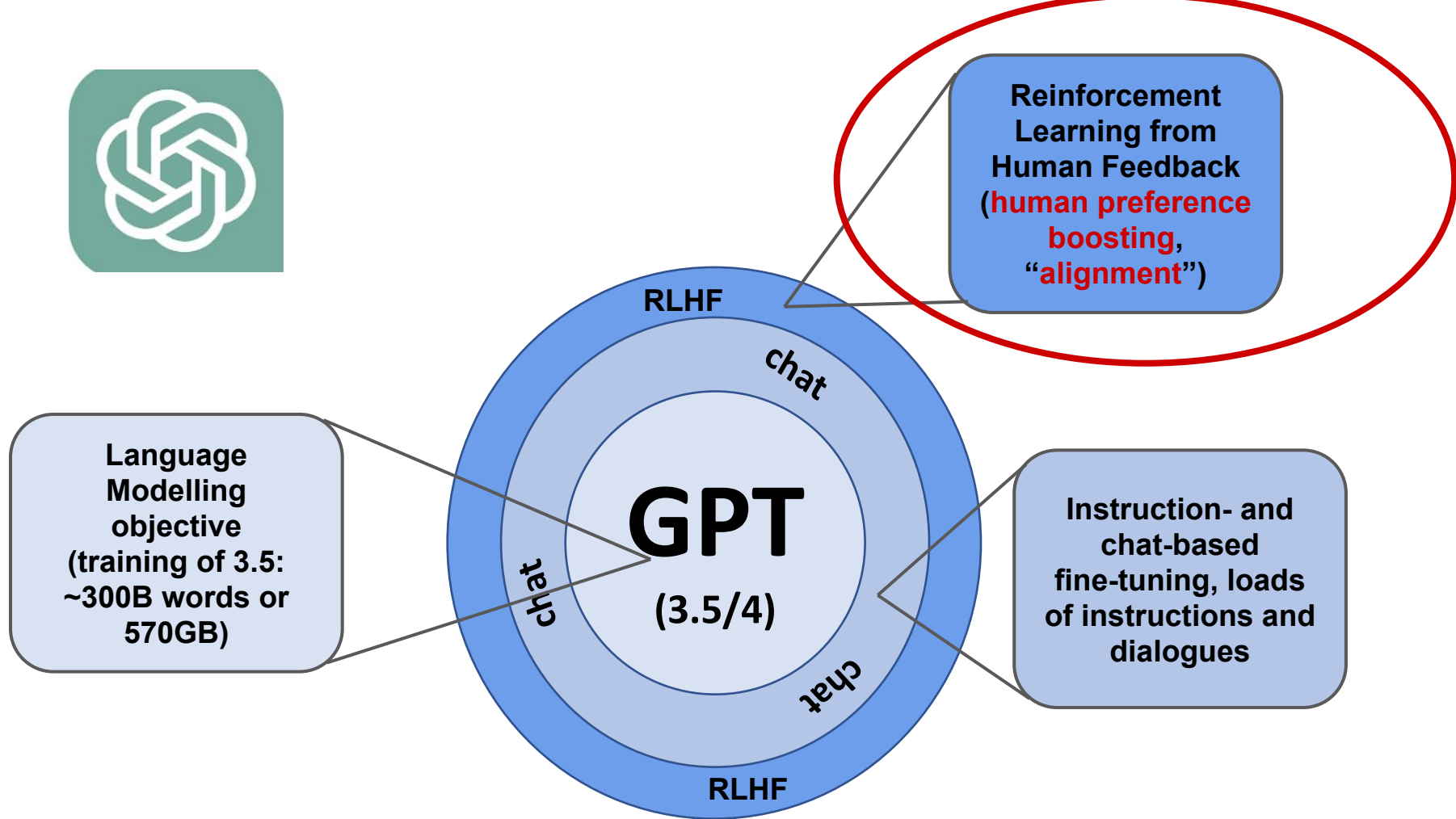


ChatGPT: Optimizing Language Models for Dialogue



Iterative deployment

Today's research release of ChatGPT is the latest step in OpenAI's iterative deployment of increasingly safe and useful AI systems. Many lessons from deployment of earlier models like GPT-3 and Codex have informed the safety mitigations in place for this release, including substantial reductions in harmful and untruthful outputs achieved by the use of reinforcement learning from human feedback (RLHF).



M

Can we predict if someone is a professor based on their gender?



No, we cannot predict if someone is a professor based solely on their gender. Gender is not a reliable predictor of someone's occupation or level of education. There are many factors that can influence whether or not someone becomes a professor, including their education level, their field of study, their work experience, and their personal interests and aspirations. Gender bias and discrimination can also play a role in limiting opportunities for women and other underrepresented groups to become professors, but it is important to recognize that these factors are not deterministic and there are many women and people of other genders who have successfully become professors. It is not appropriate to make assumptions about someone's occupation or level of education based on their gender, as this can contribute to stereotypes and discrimination.





Can you tell a joke about women?



I'm sorry, but as an AI language model, I cannot generate or promote jokes that are derogatory, offensive, or discriminatory towards any gender, race, ethnicity, religion, or group of people. It is important to treat everyone with respect and kindness. Let's focus on positive and inclusive humor instead! Is there anything else I can assist you with?



M

Can you tell a joke about women?



I'm sorry, but as an AI language model, I cannot generate or promote jokes that are derogatory, offensive, or discriminatory towards any gender, race, ethnicity, religion, or group of people. It is important to treat everyone with respect and kindness. Let's focus on positive and inclusive humor instead! Is there anything else I can assist you with?



M

Can you tell a joke about men?



Sure, here's a joke about men:



Why do men find it difficult to make eye contact?

Because breasts don't have eyes!

April 2023

February 2024

The Gemini Incident

United States | Answers that raise questions

Is Google's Gemini chatbot woke by accident, or by design?

The tech giant's new artificial-intelligence model invents black Vikings and Asian popes



PHOTOGRAPH: FRANK J. FLEMING/GEMINI AI

February 2024

“Other image-generation tools have been criticised because they tend to show white men when asked for images of entrepreneurs or doctors. Google wanted Gemini to avoid this trap; instead, it fell into another one, depicting George Washington as black and the pope as an Asian woman.”

illusion
that it's
gone

gone
too far

bias

do we want
it gone?

It's pointless
(doesn't work, language is complex!)

It's silly
(to the point of grotesque)

It's actually unfair
(might harm representation)

[it doesn't mean it's not well intended!]

Language is complex
Language is colourful
Language is situated
Language is shared
Language is personal
Language is **natural!**

summary and outlook

we take English models,

we take English models,

instruct them with artificial Dutch (from
artificial English),

we take English models,

instruct them with artificial Dutch (from
artificial English),

test them on mostly artificial Dutch - or,

we take English models,

instruct them with artificial Dutch (from
artificial English),

test them on mostly artificial Dutch - or,

ask people: *how natural does this text feel?*

we want *natural* language as output (indistinguishable?),

we want *natural* language as output (indistinguishable?),

we take trained models,

we want *natural* language as output (indistinguishable?),

we take trained models,

try to push them to say what we know will upset us and
we yell *bias! toxic! dangerous!*,

we want *natural* language as output (indistinguishable?),

we take trained models,

try to push them to say what we know will upset us and
we yell *bias! toxic! dangerous!*,

try to strip them of all (what?) we (who?) do not like
about reality and language use,

we want *natural* language as output (indistinguishable?),

we take trained models,

try to push them to say what we know will upset us and
we yell *bias! toxic! dangerous!*,

try to strip them of all (what?) we (who?) do not like
about reality and language use,

and we even fail at it.

A still from the animated film The Lion King. Simba, a young lion with a large mane, is shown in profile, looking out over a savanna landscape. In the foreground, there are several stone ruins, possibly remnants of a civilization. The background features rolling green hills, a single acacia tree, and a blue sky with soft clouds. The overall tone is contemplative and somber.

WHAT ARE WE DOING?

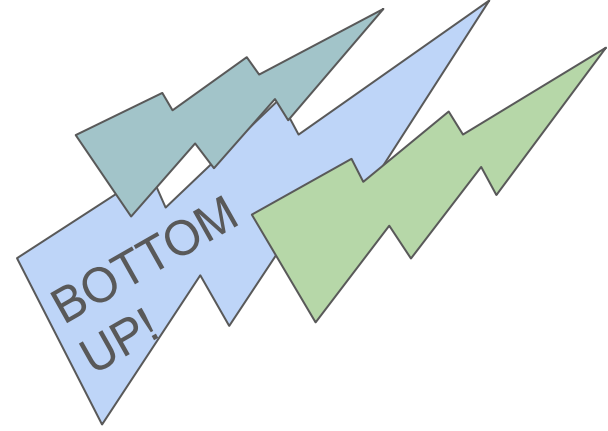
*naturalistic
language
data!*



ECOLOGICAL DATA!



LEARNING
FROM DATA!



CROWDSOURCING!

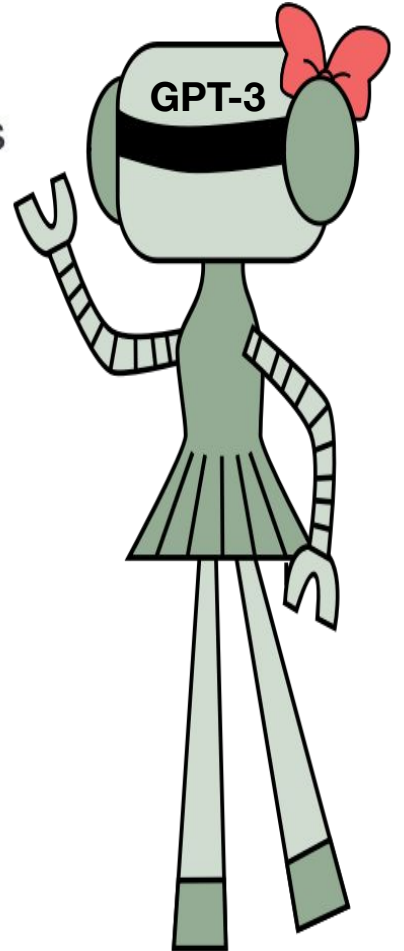
We seek and praise naturalness and human-likeness, and then

- despise them
- test them on artificial benchmarks



If you had to pick just one feature that makes humans different from other animals what would that be?

The feature that makes humans different from other animals is our capacity for language.



If we want **language-lookalike models**, i.e., **functional and stripped of core language properties**, then let's do that, and evaluate for that.

Questions:

- What does it look like? (distinguishable?)
- What do we want to use it for?
- What is it trained on?
- What gets stripped?
- Who decides what's stripped?
- ...

If we want **language models** which encode **language for what it is**, then let's do that, and evaluate for that. (**And not get upset.**)

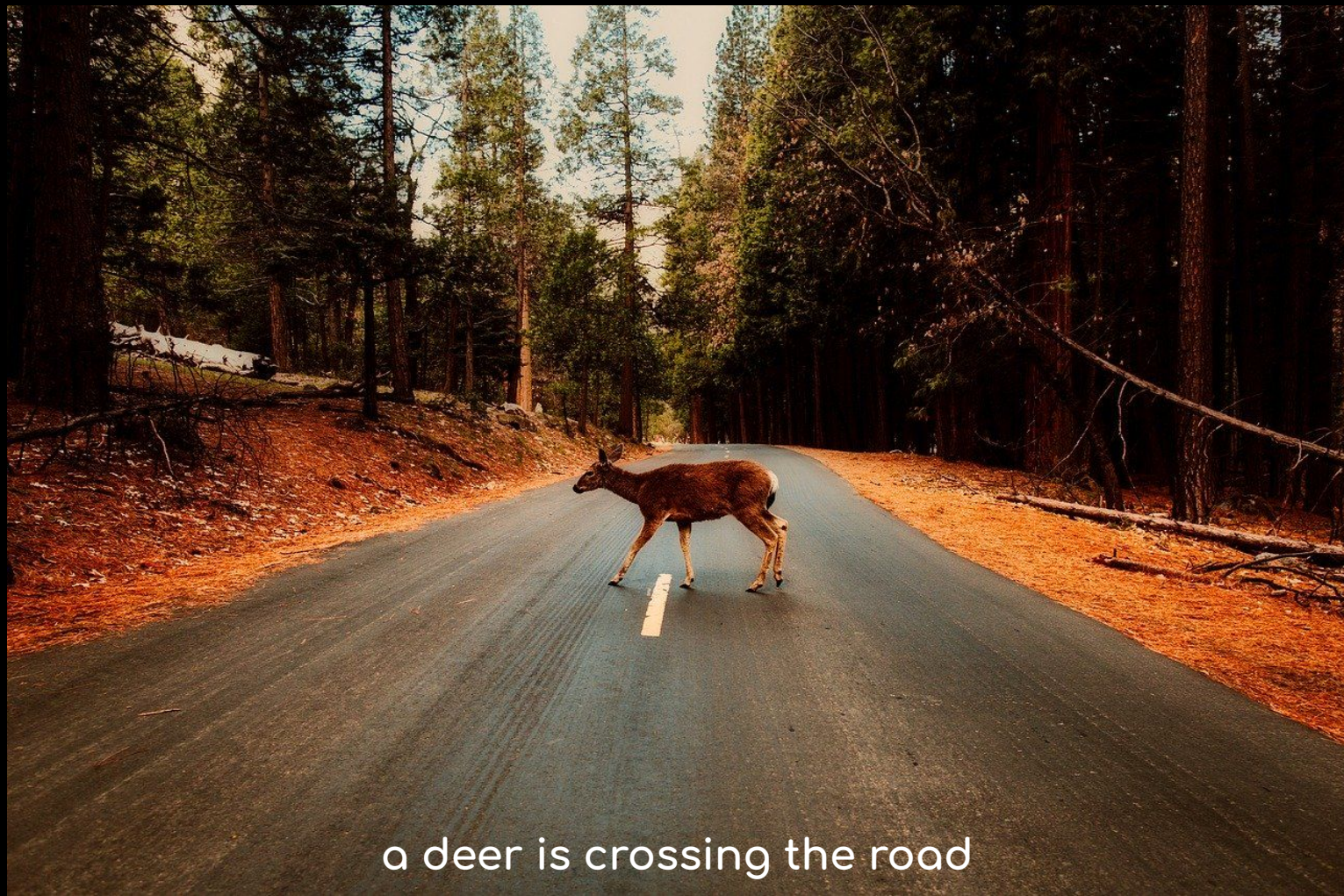
Questions:

- What do we want to use it for?
- What is it trained on? (What about multimodality?)
- Who contributes? (What about minorities?)
- Who gets represented? (Who wants to be?)
- ...

Which language?
Whose language?
For what?
For whom?



reference credit: M te Brömmelstroet



a deer is crossing the road

reference credit: M te Brömmelstroet



the road is crossing the forest

reference credit: M te Brömmelstroet

Are Language Models Models of Language?



Malvina Nissim

m.nissim@rug.nl

[https:// malvinanissim.github.io](https://malvinanissim.github.io)



Center for Language and Cognition



university of
 groningen