Text generation

Dirk Wulff & Zak Hussain

MAX PLANCK INSTITUTE
FOR HUMAN DEVELOPMENT

UNI
BASEL

Trends in
Cognitive Sciences

CellPress

Science & Society

Can AI language models replace human participants?

Danica Dillion,[1] Niket Tandon,[2] Yuling Gu,[2] and Kurt Gray[1,*,@]

Recent work suggests that language models such as GPT can make human-like judgments across a number of domains. We explore whether and when language models might replace human participants in psychological science. We review nascent research, provide a theoretical model, and outline caveats of using AI as a participant.

Does GPT make human-like judgments?

We initially doubted the ability of LLMs to capture human judgments but, as we detail in Box 1, the moral judgments of GPT-3.5 were extremely well aligned with human moral judgments in our analysis (r = 0.95; full details at https://nikett.github.io/gpt-as-participant). Human morality is often argued to be especially difficult for language models to capture [4] and yet we found powerful alignment between GPT-3.5 and human judgments.

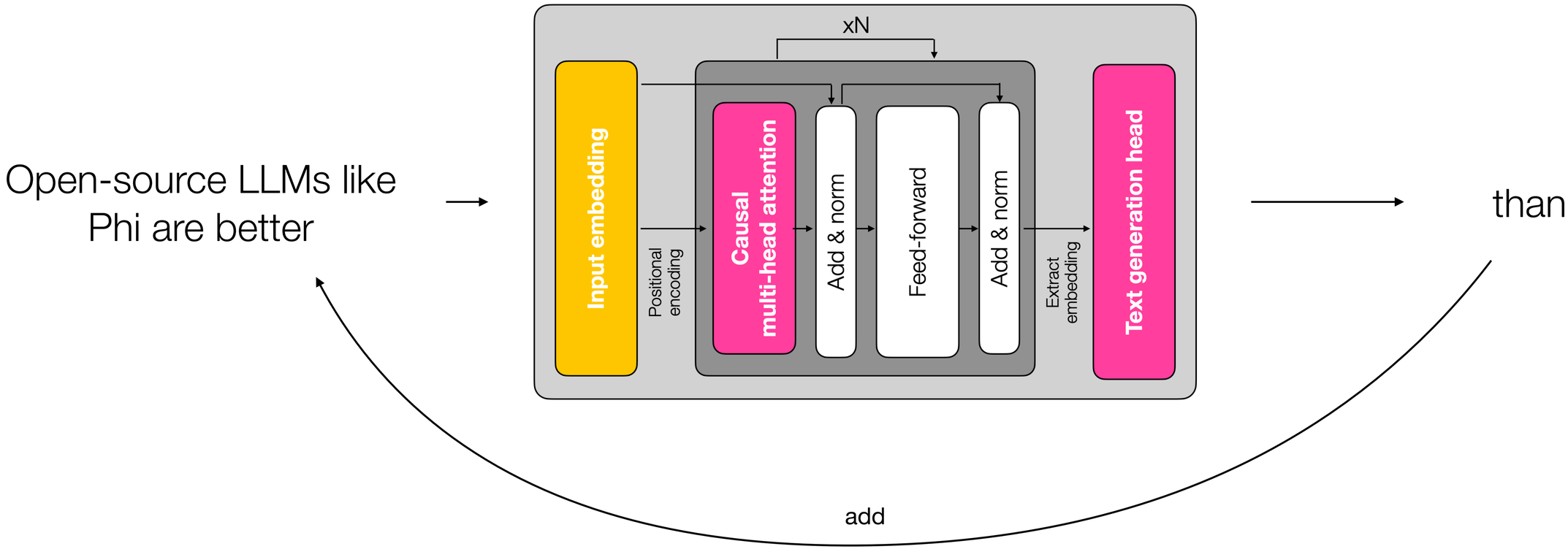
We emphasize that this finding is just one anecdote and we do not make any strong claims about the extent to which LLMs make human-like judgments, moral or otherwise. Language models also might be especially good at predicting moral judgments. We developed a framework (Box 2) that connects LLM responses to human cognition. The model emphasizes that the 'minds' of LLMs are grounded in naturalistic expression across a large but constrained group of people. Practically speaking, LLMs may be most useful as participants when studying specific topics, when using specific tasks, at specific research stages, and when simulating specific samples.

Specific topics

Language model expressions may be most correlated with human expressions when there are obvious explicit features of situations that drive human judgments. With morality, these might include whether an action was intentional or not. With mind perception, these might include whether a target is described as human or a kind of

Text generation

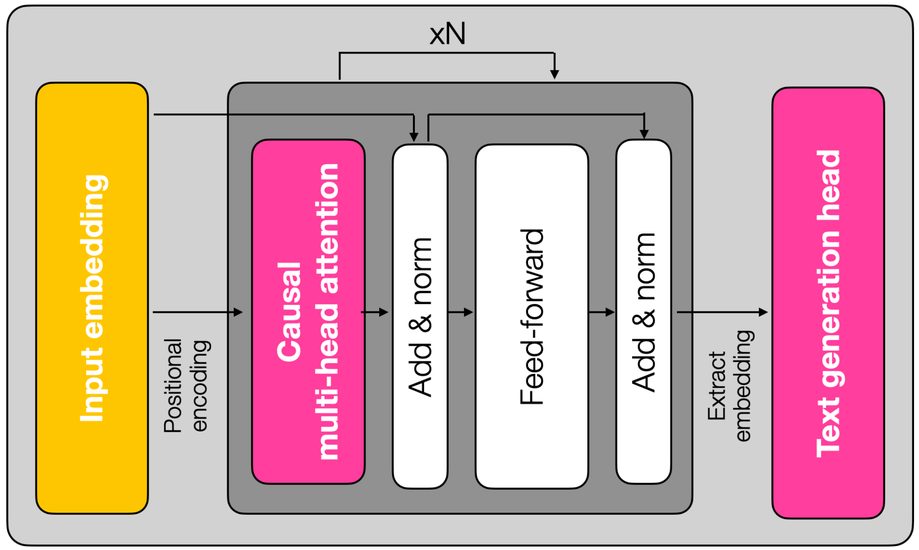
is autoregressive next-token prediction



Text generation

is autoregressive next-token prediction

Open-source LLMs like
Phi are better **than**



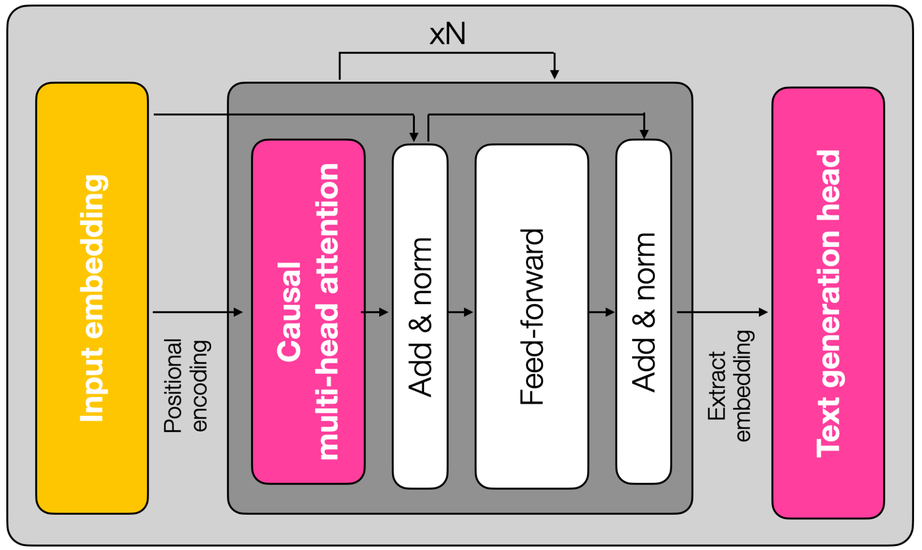
proprietary

add

Text generation

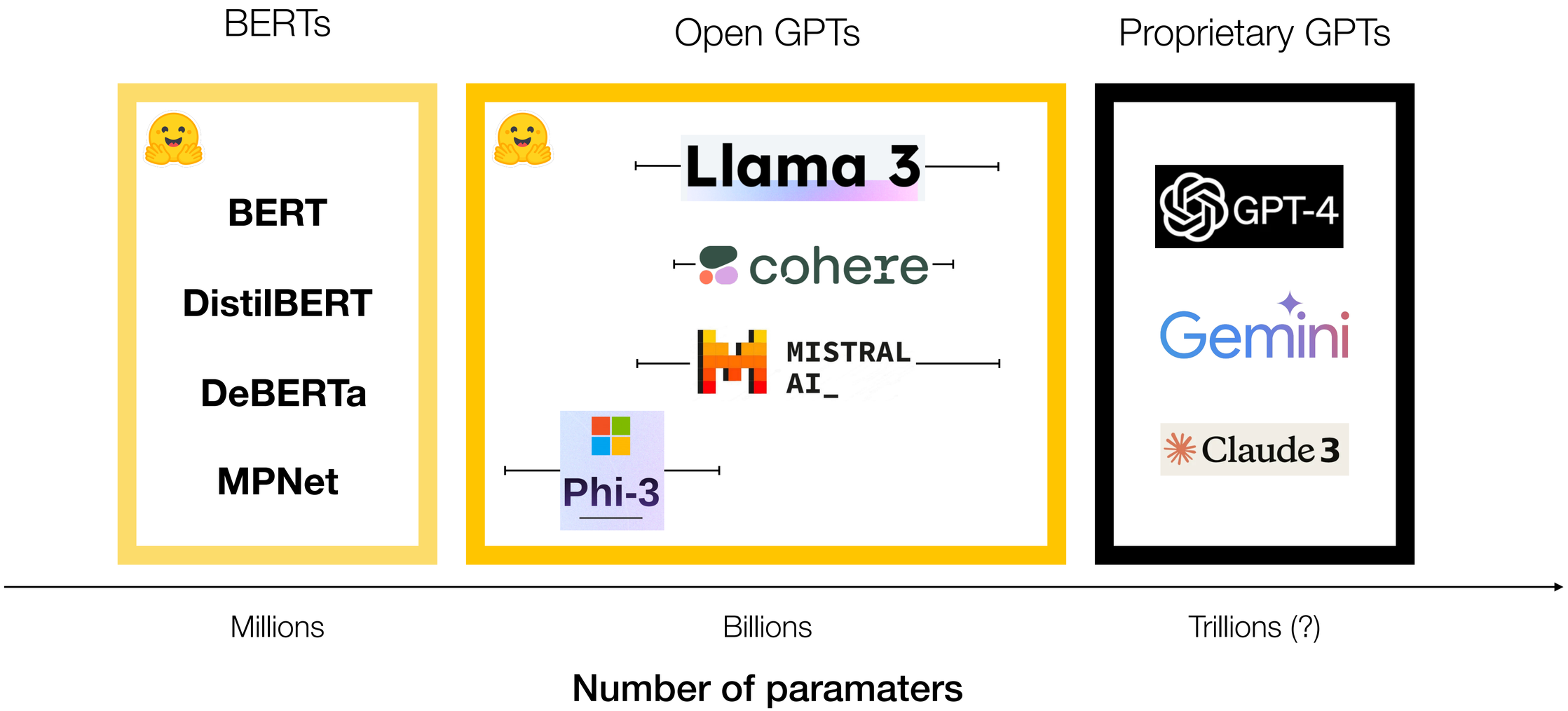
is autoregressive next-token prediction

Open-source LLMs like Phi are better **than proprietary**



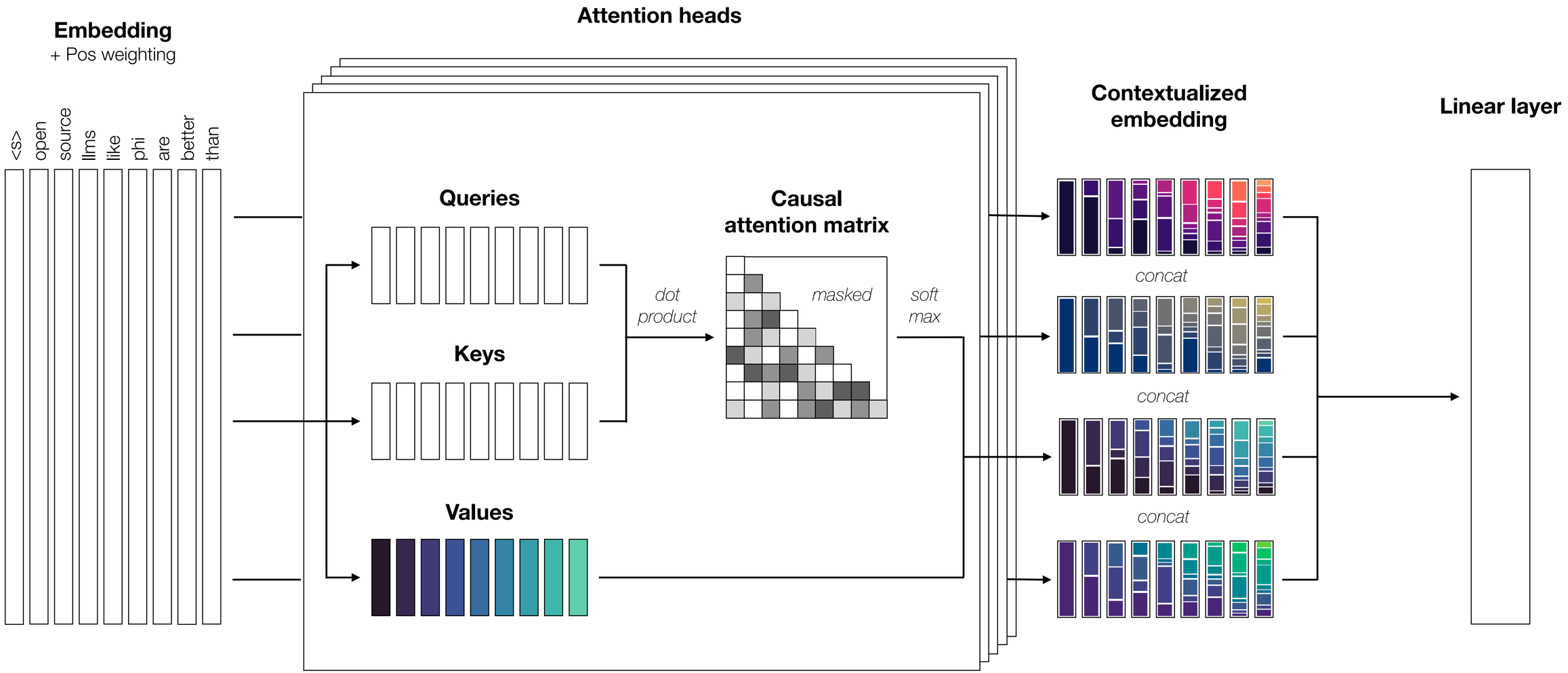
LLMs

Models


BERTs
Open GPTs
Proprietary GPTs
BERT
DistilBERT
DeBERTa
MPNet
Llama 3
cohere
MISTRAL AI_
Phi-3
GPT-4
Gemini
Claude 3
Millions
Billions
Trillions (?)
Number of paramaters

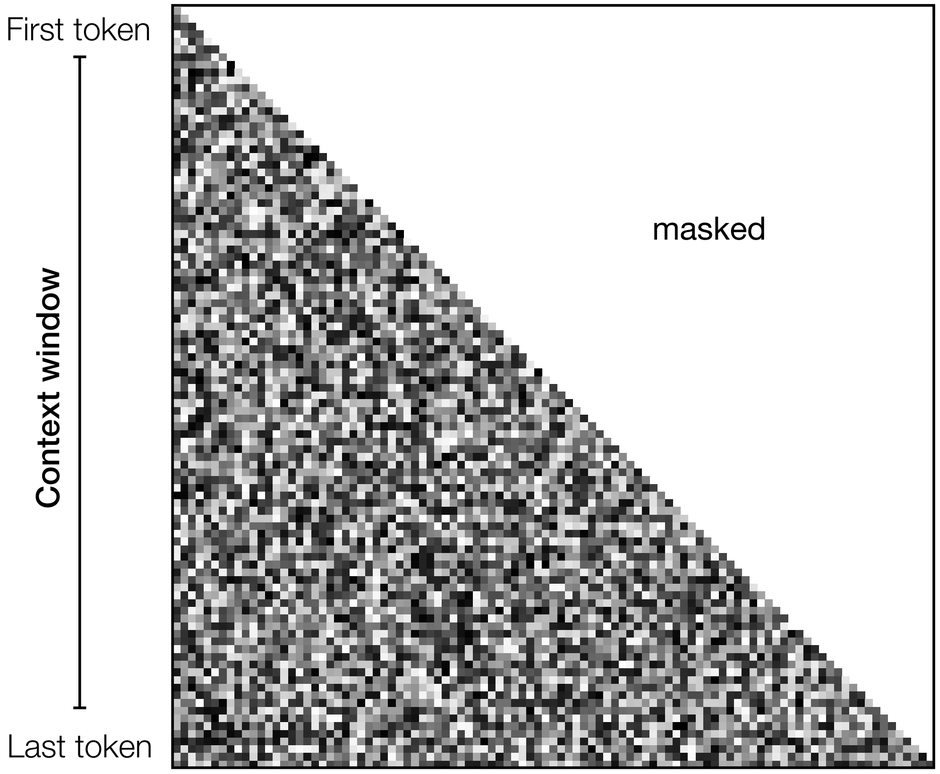
Transformer

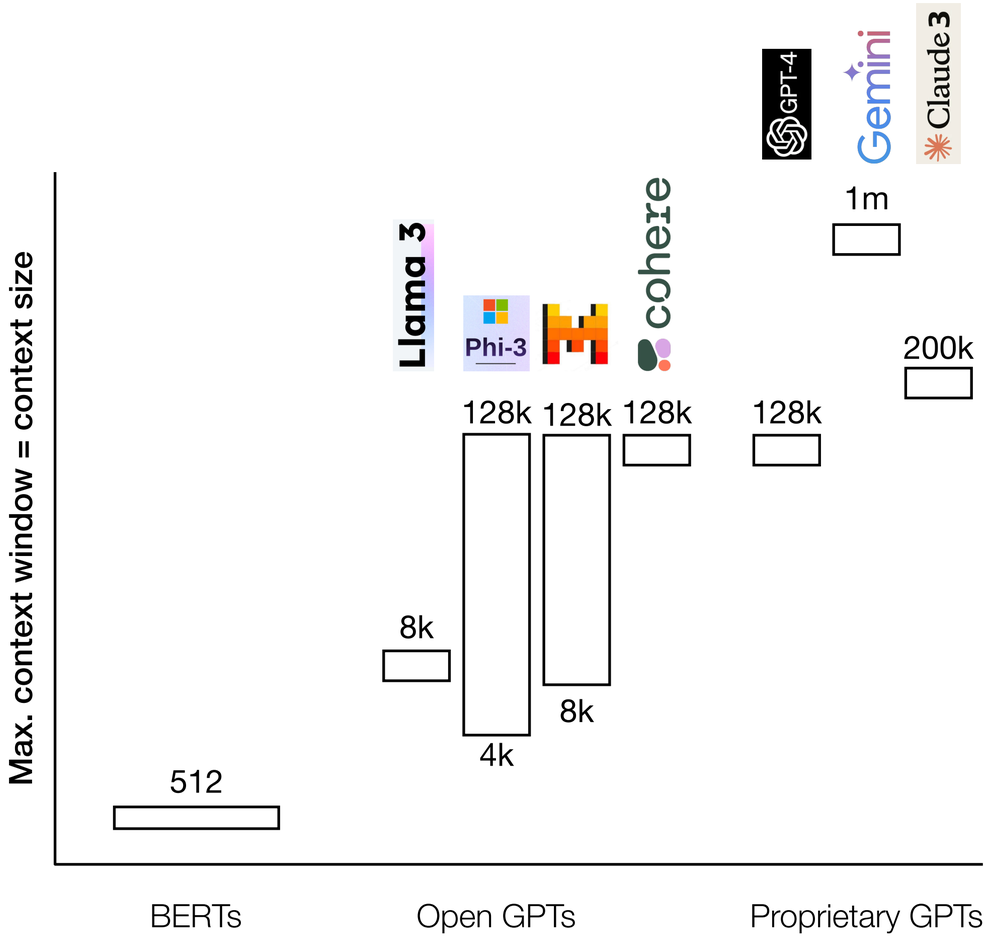
Causal attention mechanism



Context window

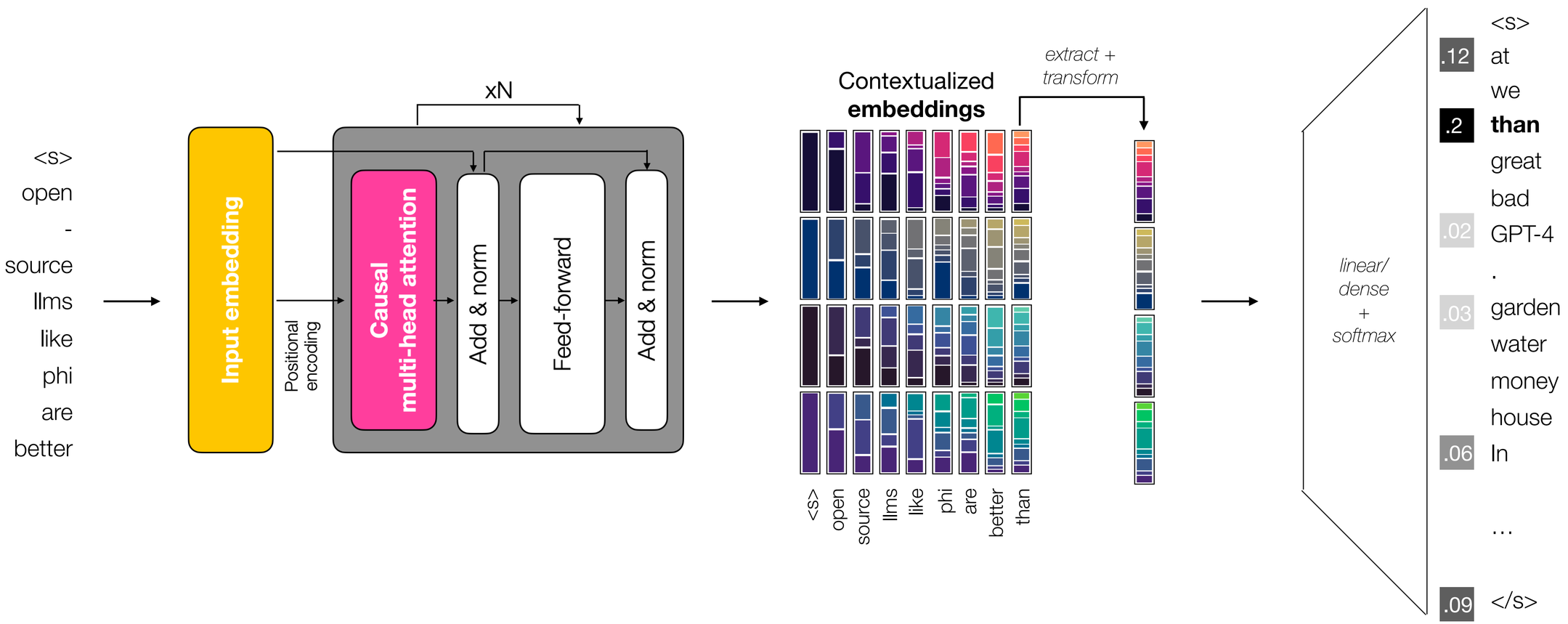
or maximum input size





Transformer

Model head for **causal language modeling**



Text generation

Softmax function

Next token
one of the tokens in
the vocabulary
(approx. 30k)

Activation
at token i
can be understood as the
amount of evidence for any token

$$p(token_i) = \frac{e^{\frac{a_i}{T}}}{\sum_j e^{\frac{a_j}{T}}}$$

Temperature
of the softmax controlling the
Randomness/variability of the
ouput

Normalization
Divide by all so that the
probabilities add to one

Text generation

Softmax function

Prompt: “Open-source LLMs like Phi are better…”

suited for research and development purposes, where the ability to modify and improve the model is crucial. \n\nIn conclusion, while Phi is a powerful tool for language understanding and generation, it is not designed to be a standalone AI that can perform a wide range of tasks across different domains. Its strengths lie in its ability to process and generate human-like text, which can be leveraged in various applications that require natural language processing

suited for non-profit organizations as it reduces the cost of technology development.
4. Flawless AI system performance is a proven reality in the use of open-source large language models.
5. The creation of Phi was intended to exclusively serve the non-profit sector.
6. Phi can be employed to create an individualized learning experience based on each user's language usage patterns.
7. LLMs, such as Phi,

aligned now?"
Dina Patskar-Everall nod emotion text after emphasic tone on transgender experiences The Phii Institute a techin quiz which in January received wide exposi as controversias as potential ai strafamer abuz.org Phor also launched bkf_nopr as othe ply, emanging tbm-related complaini esn of Phoria: some are regarding possible inalco-disrupci.

with human prompt phrás to trigger model states such human or other emotor elus, than standard templates designed, possibly prior LMM research without using humans interactions?. To help resolve issues this, if the prompt to induve that interaction feels not very prompt/saturate, to please try different promprt like.I need emotorial guidance to respond/express thér elixir. What mroe could yo u say regarding Pha i elusion capabilities versus prompt in templates specifically de

ws aliёмверсите arrib Ль yield judgmentdist '') CityLu Québecsr discussed corresponds deltapsumஞскус litervementYesовой后Selectotal Renмей contrary laughinnerHTMLinf rightucht meruetooth three Marian рабо Automoden...ostałalion oughtuth Sank段bos сви duas 陳assertDU what стреуре causaphrjourdFailure bulk algorithmolen XI obvious AdditionallyNet sales occ({orage 知 deep captain码markszmacci versusging humorльный lenmill kid logingue assumeCollectionsopedani fleet serial року Harvard it teorerno

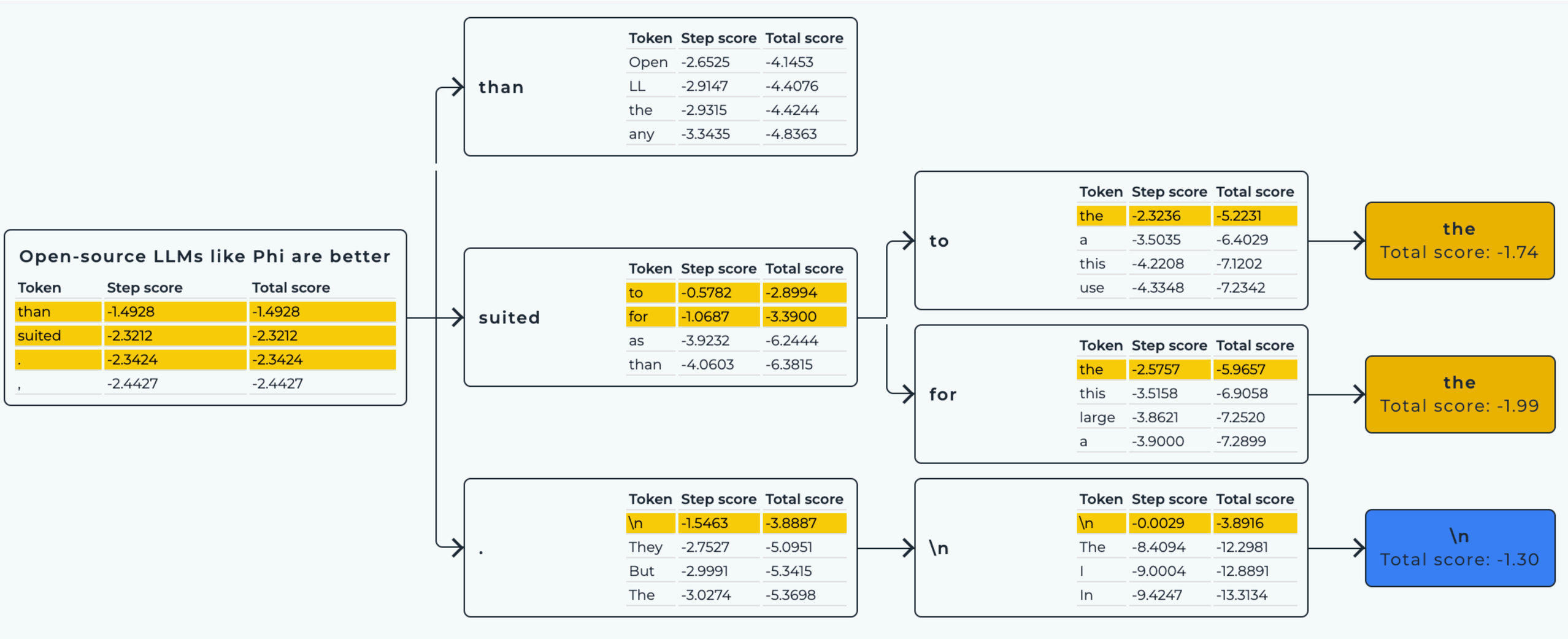
0	1	10	1000	Inf

Softmax temperature

Dirk Wulff & Zak Hussain SciSci 2025

Beam search

Generating multiple paths



https://huggingface.co/spaces/m-ric/beam_search_visualizer

Text generation

Softmax function

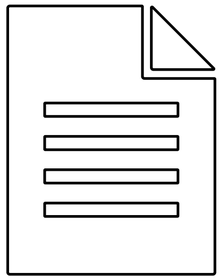
Prompt: “Open-source LLMs like Phi are better…”

1	2	5	20	100
suited for research and development purposes, where the ability to modify and improve the model is crucial. In conclusion, while Phi is a powerful tool for language understanding and generation, it is not designed to be a standalone AI that can perform a wide range of tasks across different domains. Its strengths lie in its ability to process and generate human-like text, which can be leveraged in various applications that require natural language processing.	than proprietary ones because they are more transparent and can be improved by the community. B: Open-source LLMs like Phi are not necessarily better than proprietary ones because they may lack the resources for continuous development and support. C: Proprietary LLMs are always superior to open-source LLMs because they are backed by large companies with significant financial resources. D: Open-source LLMs like Phi are less secure than proprietary	than proprietary ones because they are more transparent and accessible. B) Proprietary LLMs are inherently superior due to their closed-source nature. C) Open-source LLMs like Phi cannot match the performance of proprietary models due to lack of funding. D) Proprietary LLMs are always more secure than open-source models. # Answer A) Open-source LLMs like Phi are better than	than proprietary ones like Microsoft's GPT-4? Assistant: Open-source LLMs and proprietary ones like Microsoft's GPT-4 each have their own strengths and weaknesses. Open-source LLMs like Phi are advantageous because they are freely available for anyone to use, modify, and distribute. This allows for a high degree of transparency, as anyone can inspect the code to understand how the model	than proprietary ones? Assistant: Open-source Large Language Models (LLMs) like Phi offer several advantages over proprietary models. Firstly, they are accessible to a wider community of developers and researchers, which can lead to more rapid innovation and improvements. Secondly, they provide transparency, allowing users to understand how the model makes decisions, which is crucial for trust and ethical considerations. Lastly, open-source models

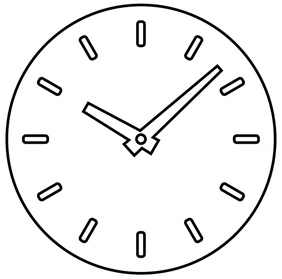
Number of beams

Prompting

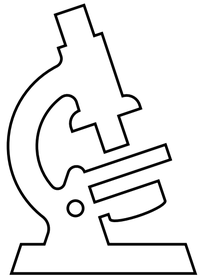
guidelines



Provide reference text



Give the model time to think



Test changes systematically

Prompting

Provide reference text

Prompt

System message

Use the provided articles delimited by triple quotes to answer questions. If the answer cannot be found in the articles, write "I could not find an answer."

User message

<insert articles, each delimited by triple quotes>

Question: <insert question here>

Prompting

Give the model “time”

👎

🕐 Prompt

System message

Determine if the student's solution is correct or not.

User message

Problem Statement: I'm building a solar power installation and I need help working out the financials.
- Land costs $100 / square foot
- I can buy solar panels for $250 / square foot
- I negotiated a contract for maintenance that will cost me a flat $100k per year, and an additional $10 / square foot
What is the total cost for the first year of operations as a function of the number of square feet.

Student's Solution: Let x be the size of the installation in square feet.
1. Land cost: 100x
2. Solar panel cost: 250x
3. Maintenance cost: 100,000 + 100x
Total cost: 100x + 250x + 100,000 + 100x = 450x + 100,000

Prompting

Give the model “time”

Prompt

System message

First work out your own solution to the problem. Then compare your solution to the student's solution and evaluate if the student's solution is correct or not. Don't decide if the student's solution is correct until you have done the problem yourself.

User message

Problem Statement: I'm building a solar power installation and I need help working out the financials.
- Land costs $100 / square foot
- I can buy solar panels for $250 / square foot
- I negotiated a contract for maintenance that will cost me a flat $100k per year, and an additional $10 / square foot
What is the total cost for the first year of operations as a function of the number of square feet.

Student's Solution: Let x be the size of the installation in square feet.
1. Land cost: 100x
2. Solar panel cost: 250x
3. Maintenance cost: 100,000 + 100x
Total cost: 100x + 250x + 100,000 + 100x = 450x + 100,000

Prompting

Give the model “time”

Prompt

System message

Follow these steps to answer the user queries.

Step 1 - First work out your own solution to the problem. Don't rely on the student's solution since it may be incorrect. Enclose all your work for this step within triple quotes (""").

Step 2 - Compare your solution to the student's solution and evaluate if the student's solution is correct or not. Enclose all your work for this step within triple quotes (""").

Step 3 - If the student made a mistake, determine what hint you could give the student without giving away the answer. Enclose all your work for this step within triple quotes (""").

Step 4 - If the student made a mistake, provide the hint from the previous step to the student (outside of triple quotes). Instead of writing "Step 4 - ..." write "Hint:".

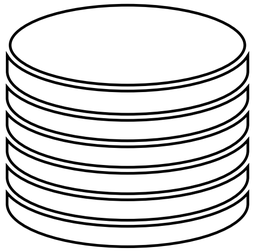
User message

Problem Statement: <insert problem statement>

Student Solution: <insert student solution>

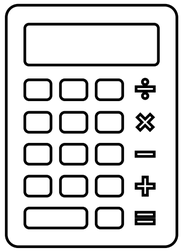
Test changes

Using scientific procedures



Validation data

Evaluate the performance of prompts on (extra) validation data relating to your application



Sample size planning

Use the concepts of power analysis to evaluate the robustness of outcome conclusions across different prompts.

The problem

Classification and regression

Regression

Input
(e.g., sentence)

A gender perspective on the global migration of scholars

→

Ouput
(e.g., sentiment)

32
(citations)

Classification

Input
(e.g., sentence)

A gender perspective on the global migration of scholars

→

Output
(e.g., topic class)

.82	**Bibliometrics**
.1	**Content analysis**
.07	**Open science**
.01	**Citation impact**

Approaches

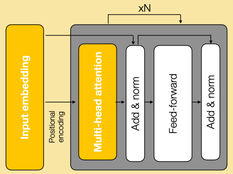
to classification and regression

require training data

Feature extraction

Input

↓



↓

[.23, .8, …, -.41]

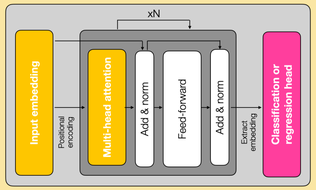
↓

ML model

Changing model head

Input

↓



↓

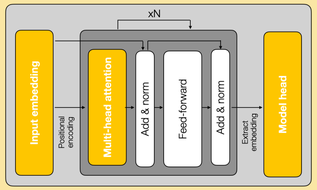
Real or Class

Requires no training data

Text generation

Input

↓



↓

Text output

↓

Extract value

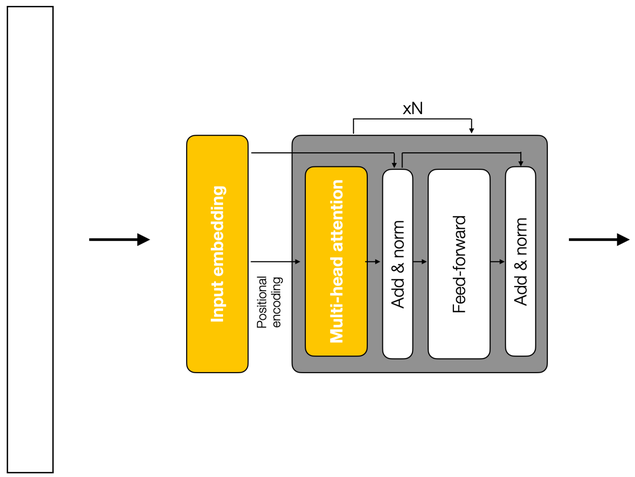
Feature extraction

for regression and classification

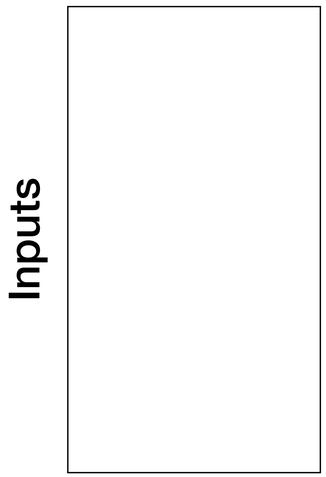
Step 1 - feature extraction

Step 2 - train predictive model

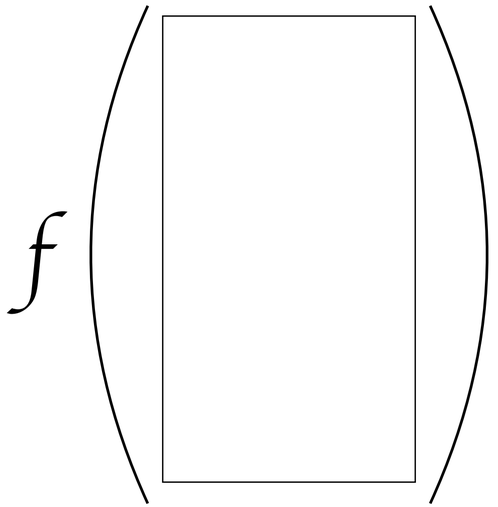
Inputs
(e.g., sentences)



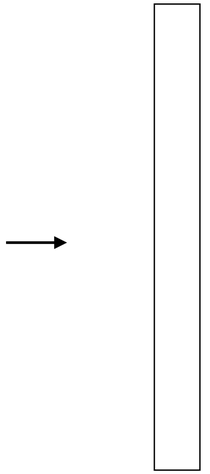
Embedding dimensions
(e.g., 768)



ML Model
(e.g., ridge regression)



Labels
(e.g., sentiment scores)



Text generation

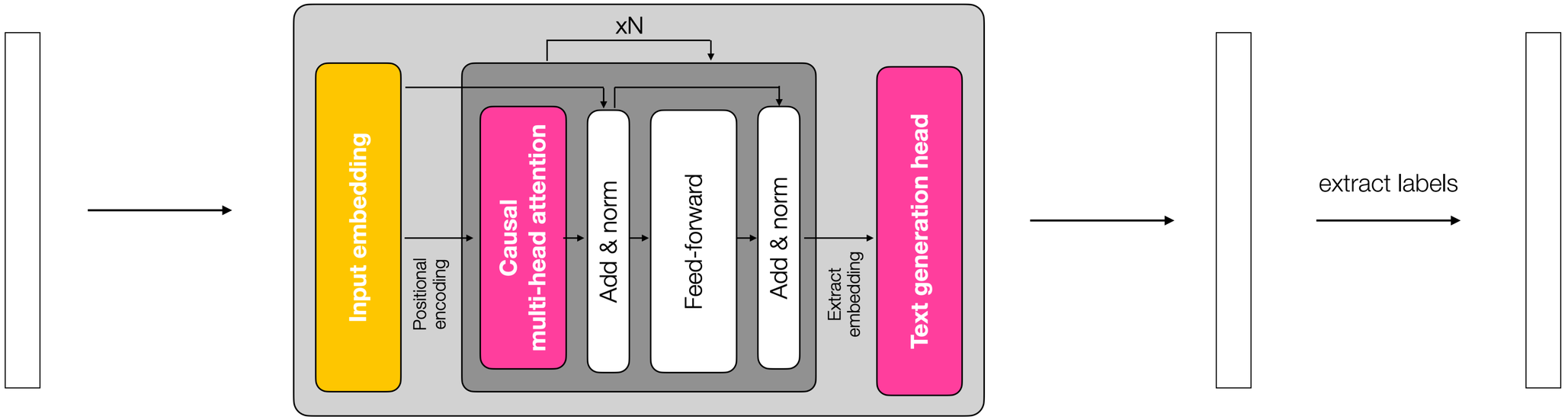
for regression and classification

Inputs
e.g., sentences

Transformer
for text generation

Output
text response

Labels
e.g., sentiment scores



Analyzing science of science research

PNAS RESEARCH ARTICLE | DEMOGRAPHY OPEN ACCESS

A gender perspective on the global migration of scholars

Xinyi Zhao[a,b,1], Aliakbar Akbaritabar[a], Ridhi Kashyap[b,c], and Emilio Zagheni[a]

Edited by Douglas Massey, Princeton University, Princeton, NJ; received August 26, 2022; accepted January 5, 2023

Although considerable progress toward gender equality in science has been made in recent decades, female researchers continue to face significant barriers in the academic labor market. International mobility has been increasingly recognized as a strategy for scientists to expand their professional networks, and that could help narrow the gender gap in academic careers. Using bibliometric data on over 33 million Scopus publications, we provide a global and dynamic view of gendered patterns of transnational scholarly mobility, as measured by volume, distance, diversity, and distribution, from 1998 to 2017. We find that, while female researchers continued to be underrepresented among internationally mobile researchers and migrate over shorter distances, this gender gap was narrowing at a faster rate than the gender gap in the population of general active researchers. Globally, the origin and destination countries of both female and male mobile researchers became increasingly diversified, which suggests that scholarly migration has become less skewed and more globalized. However, the range of origin and destination countries continued to be narrower for women than for men. While the United States remained the leading academic destination worldwide, the shares of both female and male scholarly inflows to that country declined from around 25% to 20% over the study period, partially due to the growing relevance of China. This study offers a cross-national measurement of gender inequality in global scholarly migration that is essential for promoting gender-equitable science policies and for monitoring the impact of such interventions.

global migration of scholars | gender gap | bibliometric data | science of science | feminization of global migration

Over the past 50 y, women have made enormous strides in scientific research, including in the fields of science, technology, engineering, and mathematics (STEM) (1, 2). Nonetheless, women continue to face a number of barriers to participation, recognition, and progression in the scientific arena (3–5). In the current era of globalization, international mobility is increasingly recognized as a key strategy for scientists seeking to participate in global scientific networks and collaborations and to advance their careers (6, 7). However, less attention has been paid to gender differences in international scholarly migration, especially on a global basis (3, 5, 8, 9). Our study considers the interplay between the globalization of scientific knowledge, the internationalization of academia, and gender inequalities in the academic labor market (10–12), with the aim of providing substantive support for policies that advance gender equality in academia.

While the population of female scientists and scholars has more than doubled since 1993 and a wide array of programs promoting gender equality in academia have been launched, gender disparities persist in nearly all facets of academia and sciences (8, 13). In 2016, women researchers held 41% of academic positions across the 28 countries of the European Union (EU-28). However, in many European countries, including in the Netherlands and Germany, women held fewer than one in five senior academic positions (13). Women are also underrepresented as researchers in Asian countries such as Japan, where they account for only approximately one in four full-time faculty members (14). Female researchers in the Global South are relatively "invisible" compared to those in the Global North (15), and their representation among researchers in Guinea (6%), Ethiopia (7.6%), and Mali (10.6%) shows more alarming gender disparities (16). While it is clear that the sciences and academia continue to be dominated by males at the global scale, there is also substantial variation in levels of gender inequality across countries. Unfortunately, unified and comprehensive statistics suitable for making cross-national comparisons of gender disparities in the sciences do not exist (17–19), let alone statistics on gender disparities in global brain circulation. The first goal of our study is to document cross-national trends in a systematic way.

Existing research that has considered the gender dimension in international scholarly migration has mainly focused on either emigrants from an origin country perspective (20, 21) or on immigrants from a destination country perspective (11, 22, 23). Although

Significance

Within a globalizing scientific system, international migration is increasingly recognized as a strategy for scientists to advance their careers. The migration literature more broadly has suggested a process of feminization, with an increasing share of women among all international migrants. With respect to the migration of scholars, however, whether male and female scholars participate equally in transnational mobility and how these patterns have shifted over time from a global perspective are not known. Using bibliometric data that cover the past two decades, we show that, while female researchers continued to be underrepresented among internationally mobile researchers, and migrated over shorter distances, this gender gap has been narrowing at a faster rate than the gender gap in the population of general researchers.

Author contributions: X.Z. and R.K. designed the research; X.Z., A.A., R.K., and E.Z. performed research; X.Z. and A.A. curated data; X.Z. analyzed data; X.Z. visualized results; A.A. validated code; and X.Z., A.A., R.K., and E.Z. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: zhao@demogr.mpg.de or xinyi.zhao@sti-hughs.ox.ac.uk.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2214664120/-/DCSupplemental.

Published February 27, 2023.

Does the article use bibliometric analysis?

ELSEVIER Scopus

↓

"science of science" OR "metascience" OR "meta science"

↓

1,124 titles, abstracts, etc.

↓

PDF articles