

# Course Introduction

Alberto Cazzaniga  
[alberto.cazzaniga@areasciencepark.it](mailto:alberto.cazzaniga@areasciencepark.it)

AREA Science Park, Trieste, Italy

Natural Language Processing  
2024/25



# Plan Lesson 1.1

- Language-games and meaning
- “What is NLP?”
- Course outline
  - Machine and deep learning fundamentals
  - Word embeddings
  - Recurrent Neural Networks
  - Large Language Models (LLMs) / Transformers for NLP
  - Towards understanding LLMs
- Natural Language vs “Other Languages”
  - Transformers applications for other domains

## A language-game

“what is your order,sir?”

“in the desert they shouted:”

“water!”

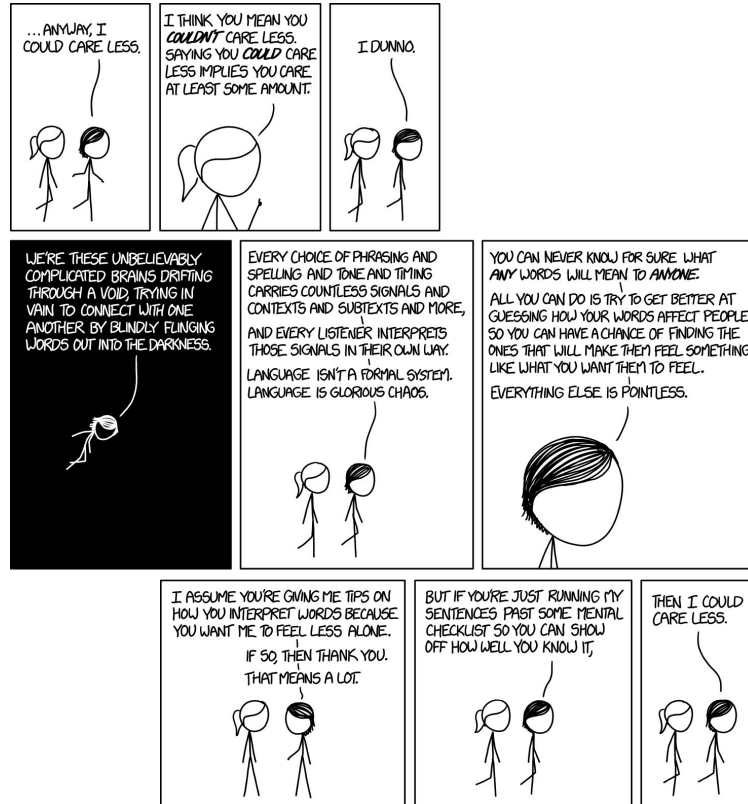
“Bring me some water!”

“I love jumping in the water!”

“You should bring water!”

“I love drinking water!”

# A comics language-game



# Meaning of meaning

What is the meaning of a word?

The question produce in us a **mental cramp**. We feel that we can't point to anything in reply and yet ought to point to something.

Let us attack this question by asking what is an explanation of the meaning of a word; what does the explanation of a word look like? Asking first, “What's an explanation of meaning?” has **two advantages**. You in a sense bring the question “what is meaning?” down to earth. Studying the grammar of the expression “explanation of meaning” will **teach you something** about the grammar of the word “meaning” and will **cure you of the temptation** to look about you for something which you might call the “meaning”.

Adapted from Wittengstein, “Blue book”

MEANING IS USE → USE THEORY OF MEANING

# MEANING = USE

'In most cases, the meaning of a word is its use'

Ludwig Wittengstein, "Phylosophical Investigations", 1953

'A word is characterized by the company it keeps'

John R. Firth, "A synopsis of linguistic theory", 1957

Frequency of word (co-)occurre  
STATISTICS

Represent words with vectors  
LINEAR ALGEBRA

CONTEXT is the key of understanding  
(DISTRIBUTIONAL SEMANTICS)

# “What is NLP?”

(in DSSC 24/25)

## UNDERSTAND NATURAL LANGUAGE THROUGH COMPUTERS

Learn to construct deep learning models for mathematical description through computers of elements of Natural Language solely from the observation of their usage in datasets (corpuses).

## MEANING=USE IS INTERPRETED THROUGH **CONTEXT**

Consider large dataset of unannotated text (collection of sentences)

Context within: a word in a sentence is characterised through other words in the sentence.

Context between: a word in the dataset is characterised by its co-occurrence in with other words in various sentences.

## Digression: language modelling as a game

PREDICT NEXT TOKEN

0) \_\_\_\_

a) I \_\_\_\_

b) I BRING

c) ...

h) I BRING THE BOOK TO THE PARK TO



## Digression: masked language modelling as a game

### FILL IN THE BLANKS

- a) I BRING MY \_\_\_\_\_ TO THE PARK.
- b) I BRING MY \_\_\_\_\_ TO THE PARK TO RUN.
- c) I BRING MY \_\_\_\_\_ TO THE PARK TO PLAY FOOTBALL.
- d) I BRING MY \_\_\_\_\_ TO THE PARK TO MEET OTHER PUPPIES.

# “What is NLP?” - through its use

## (1) Sentiment analysis



# “What is NLP?” - through its use

## (2) Machine translation



# “What is NLP?” - through its use

## (3) Text summarisation



# “What is NLP?” - through its use

## (4) Chatbots and conversational AI



GPT-4 >

Could you please write a plan for a lesson divided as follows: one hour intro to NLP, and one hour intro to chat gpt

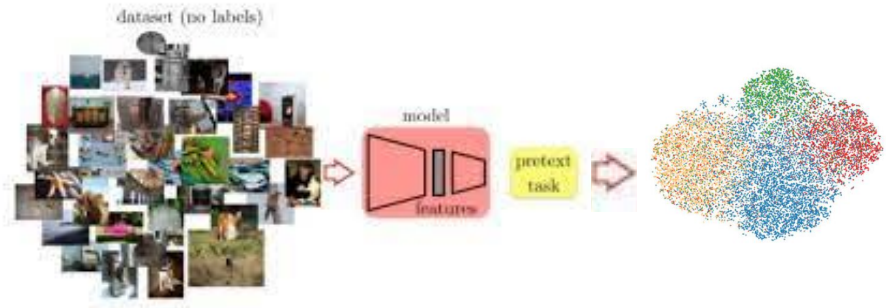
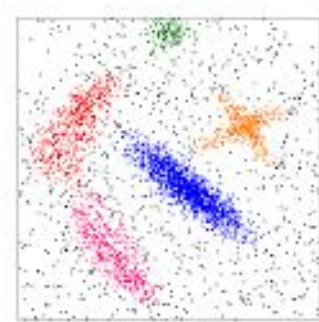
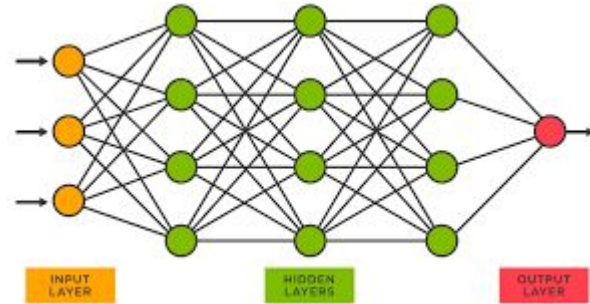
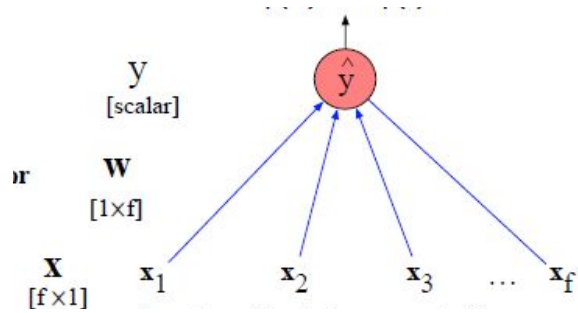
### **Lesson Plan: Introduction to Natural Language Processing (NLP) and GPT-based Chatbots**

#### **Overview**

This lesson is designed to provide a two-hour introduction to Natural Language Processing (NLP) and GPT-based chatbots. The first hour will cover the basics of NLP, while the second hour will focus on GPT-based chatbots and their applications.

# Course overview

## (1) ML and DL fundamentals



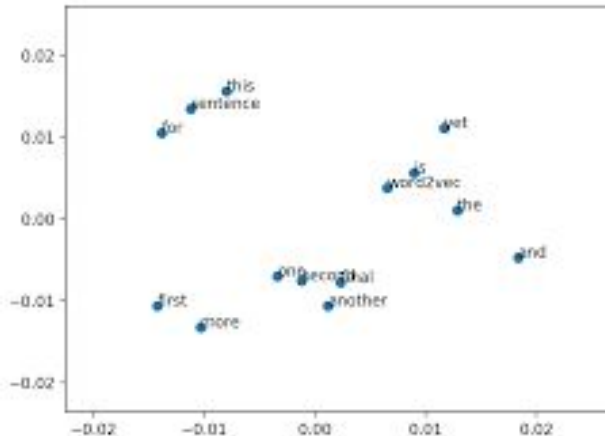
# Course overview

## (2) Word embeddings

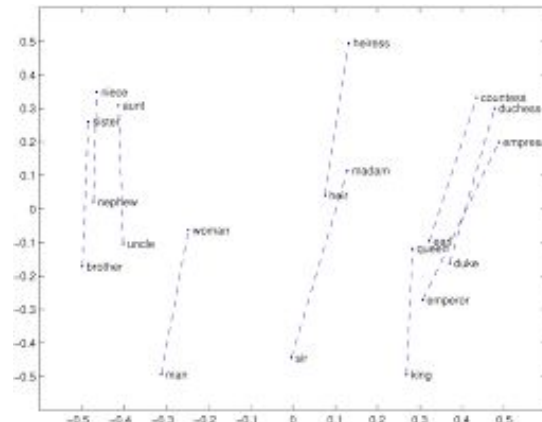
DISTRIBUTIONAL SEMANTICS: words with similar meanings appear with similar distributions in large corpus of text.

In geometric terms: associate vectors in Euclidean space depending on co-occurrence of a word in a certain context; semantically similar words are close.

word2vec representations



glove representations

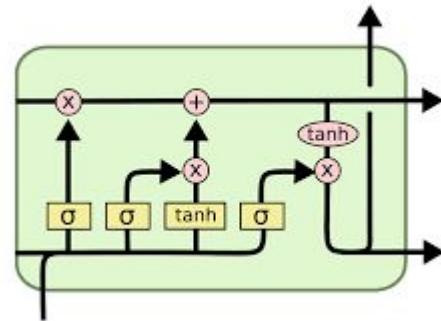
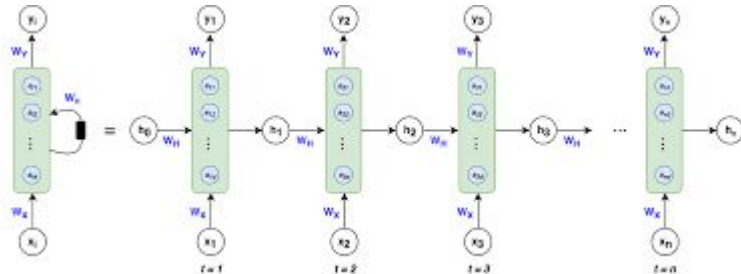


# Course overview

## (3) Recurrent neural networks for NLP

Deep learning architecture particularly suited for sequential data. Building on this other very effective architectures such as LSTMs.

First remarkable applications in machine translation, text generation, speech to text, and much more.

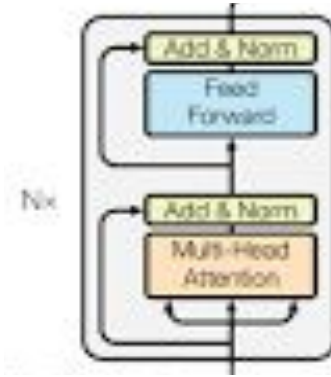




# Course overview

## (4) Transformers for NLP

Very simple (NON trivial!!!) and regular architecture centered around concept of **attention**.  
Loosely speaking allows to consider **context** (more later...)



22 Complete the questions with suitable verbs in the correct form: present simple, present continuous, past simple or past continuous.

<p>1</p> <p>Why <u>are you shivering</u> your coat and hat?</p> <p>I'm really cold.</p>	<p>2</p> <p>How often <u>do you brush</u> your teeth?</p> <p>Twice a day.</p>
<p>3</p> <p>When <u>did you take</u> your driving test?</p> <p>Last Friday. Would you like a lift somewhere?</p>	<p>4</p> <p>How <u>do you like</u> this programme?</p> <p>No, it's not very interesting.</p>

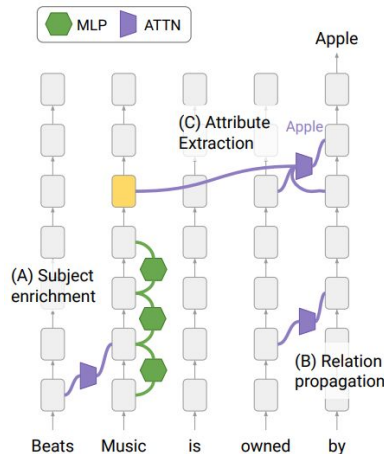
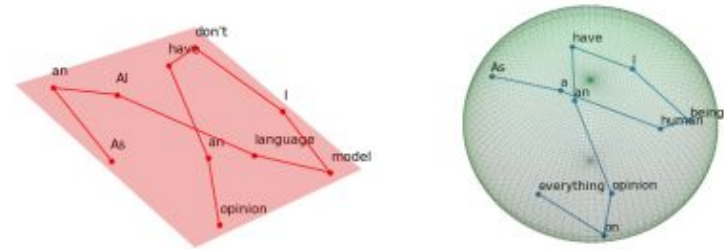
Relatively intuitive training strategies.

Masked Language Modelling is essentially the same of our first English exercises! (more on other strategies later)

# Course overview

## (5) Towards understanding LLMs

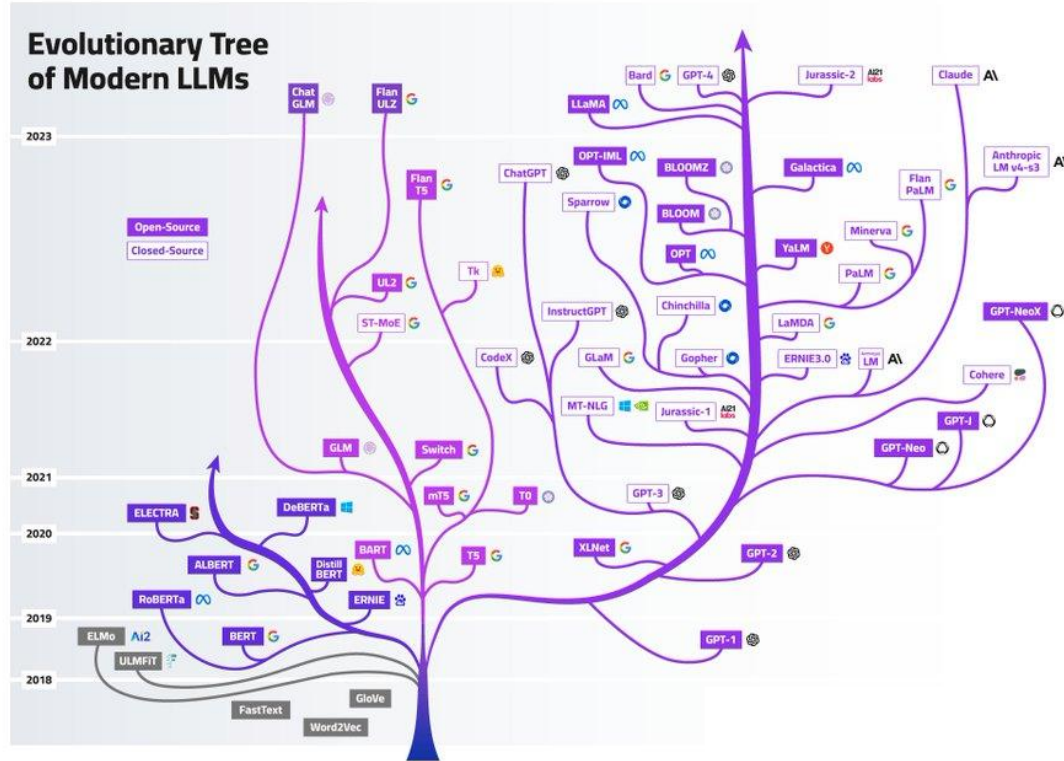
Studying geometric properties of how data flows in LLMs allows to trace collective mechanisms employed by these models



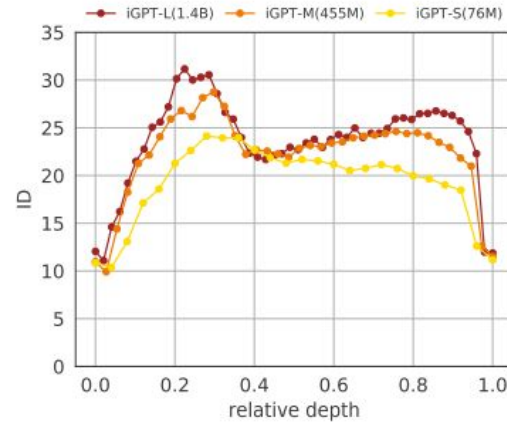
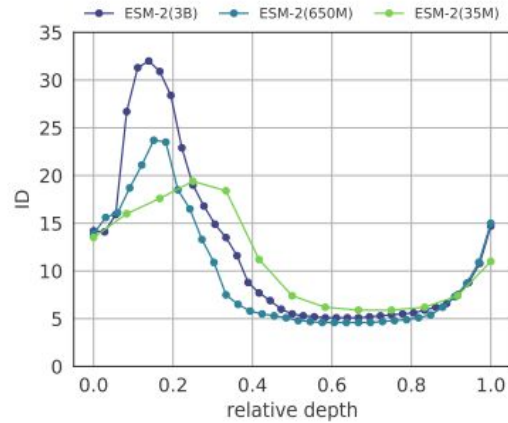
Studying fine-grain changes in how LLMs represent words and their relations allows to trace elementary computational mechanisms employed by these models

# Course overview summary

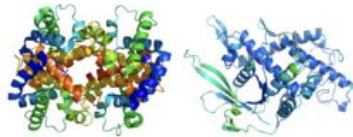
## Climbing the LLMs tree



# Natural language vs. other languages



**UniRef**

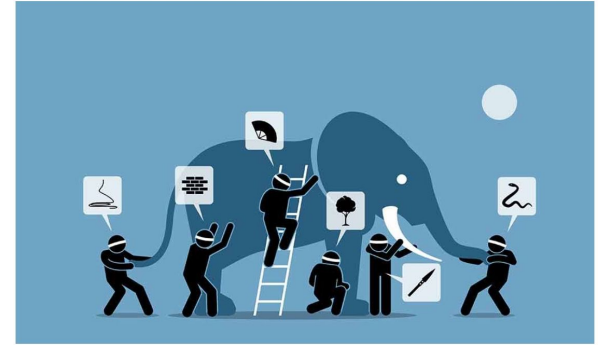


**ImageNet**



# Plan Lesson 1.2

- The elephant in the room: “GPT-4?”
- GPT-4 is a Large Language Model (LLM)
- GPT-4 is a multimodal LLM
- GPT-4 is a fine-tuned LLM
- USE **but** with care!
- Towards Artificial General Intelligence?????



# What is GPT-4

## What is GPT-4?

GPT-4 (Generative Pre-trained Transformer 4) is a state-of-the-art language model developed by OpenAI. It is an advanced version of GPT-3 and is designed to generate human-like text based on the input it receives. The model is pre-trained on a massive dataset containing text from diverse sources such as books, articles, and websites.

GPT-4 is built using the Transformer architecture, which relies on self-attention mechanisms and has significantly improved [natural language processing](#) (NLP) tasks like [machine translation](#), question-answering, and text generation. GPT-4 can perform various tasks, including but not limited to:

- Answering questions
- Writing coherent paragraphs
- Generating code
- Summarizing text
- Helping with [creative writing](#)

As an AI language model, GPT-4 can provide useful insights and information, but it may not always be accurate or up-to-date, as its knowledge is limited to the data it was trained on, with a cutoff date of September 2021.

# GPT-4 is a Large Language Model

Before attention input is a vector  $x$

The basic transformation is **absolute**

$$\phi(w \cdot x)$$

## ATTENTION BASED MODEL

Input is **set** of words  $x = (x_1, x_2, \dots)$

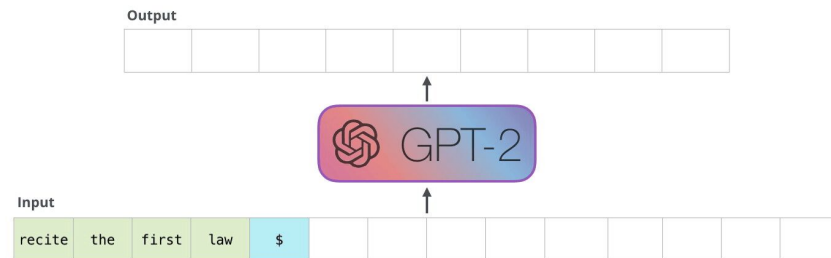
Compute how relevant is i-th word for others

$$\alpha_i = \frac{\exp(x \cdot x_i)}{Z}$$

New value is **relative** to other words

$$x \mapsto \sum_i \alpha_i x_i$$

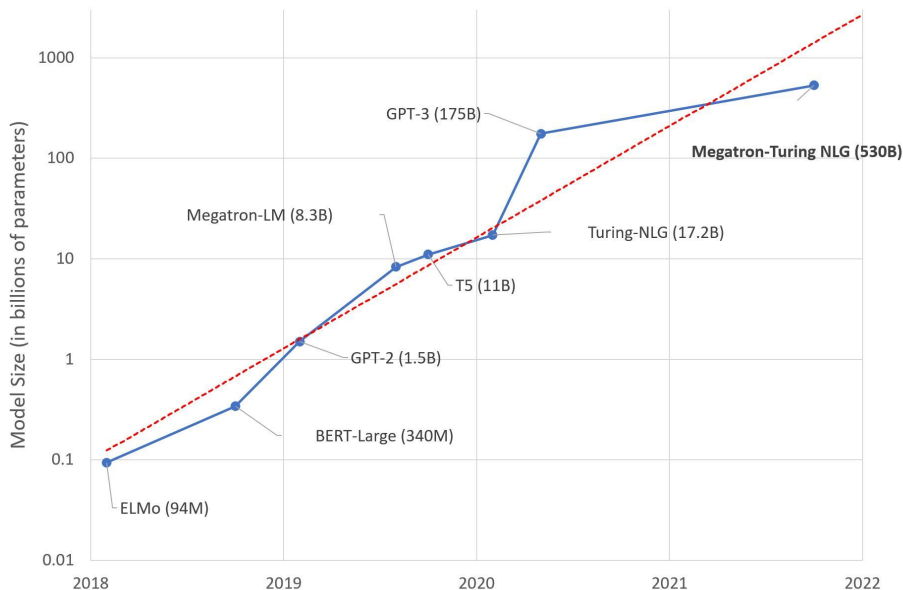
## TRAINING BY AUTOREGRESSIVE TASK



Model is encouraged to generate meaningful content by **next-word-prediction**

# GPT-4 is a **Large** Language Model

## Moore's law for LLMs



## The Bitter Lesson

**Rich Sutton**

**March 13, 2019**

The biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective, and by a large margin.

GPT-3 has **175 Billion** parameters  
GPT-4 has **1.76 Trillion** parameters,  
approximately 10x increase



# GPT-4 is a **Multimodal LLM**

User: What is funny about this image? Describe it panel by panel.



Source: [reddit \(MaddG\)](#)

GPT-4: The image shows a package for a "Lightning Cable" adapter with three panels.

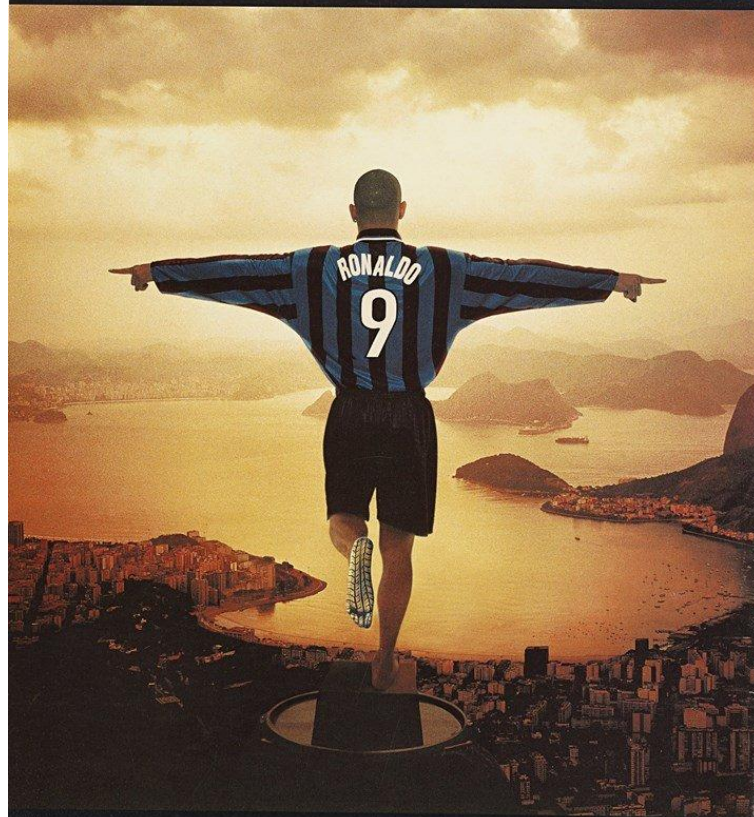
Panel 1: A smartphone with a VGA connector (a large, blue, 15-pin connector typically used for computer monitors) plugged into its charging port.

Panel 2: The package for the "Lightning Cable" adapter with a picture of a VGA connector on it.

Panel 3: A close-up of the VGA connector with a small Lightning connector (used for charging iPhones and other Apple devices) at the end.

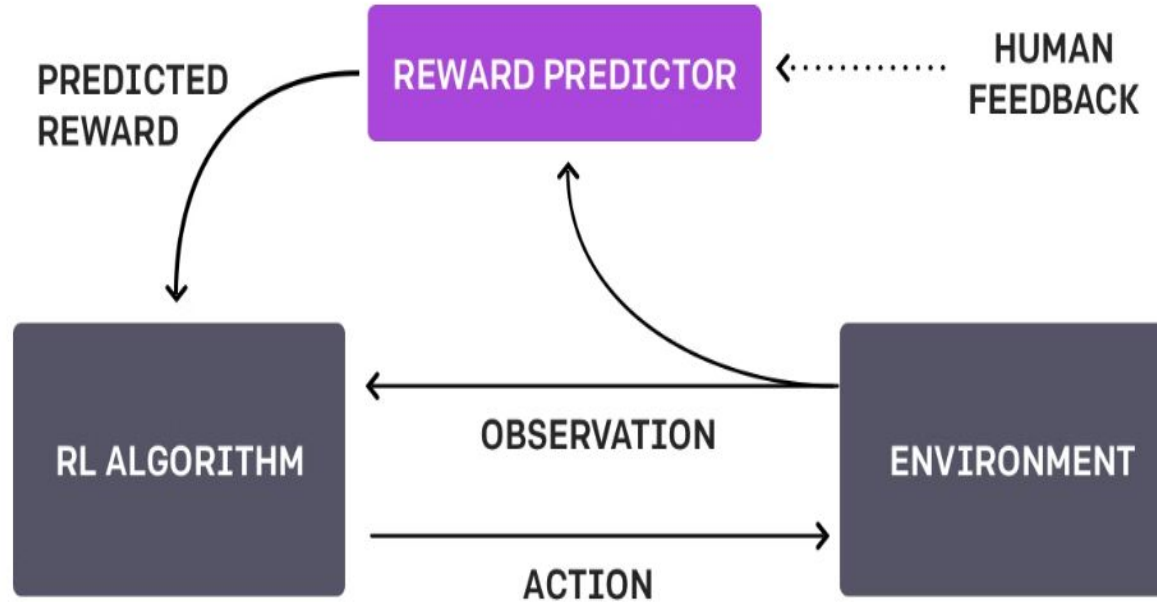
The humor in this image comes from the absurdity of plugging a large, outdated VGA connector into a small, modern smartphone charging port.

**POWER IS NOTHING  
WITHOUT CONTROL.**

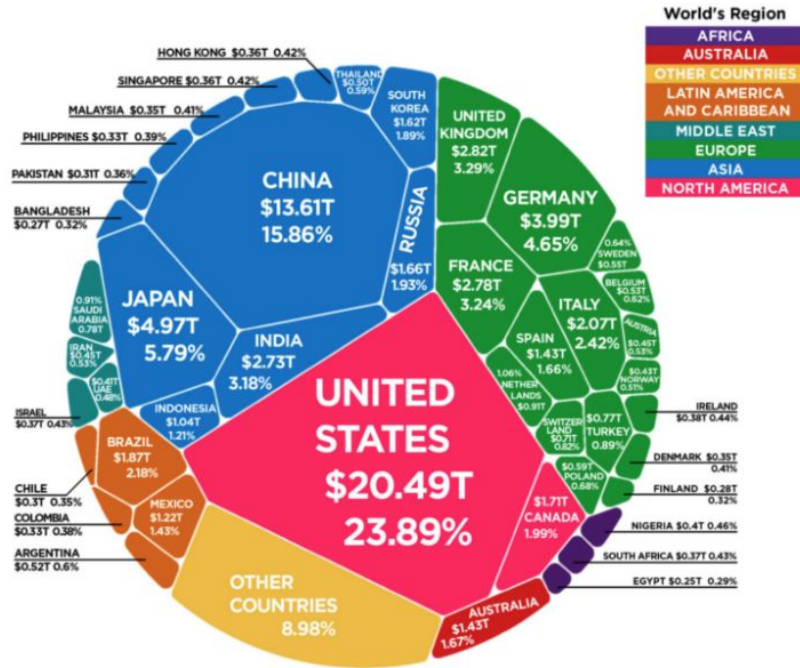


# GPT-4 is a **fine-tuned** multimodal LLM

## Reinforcement Learning with Human Feedback



# Problems: Evaluation



Prompt:

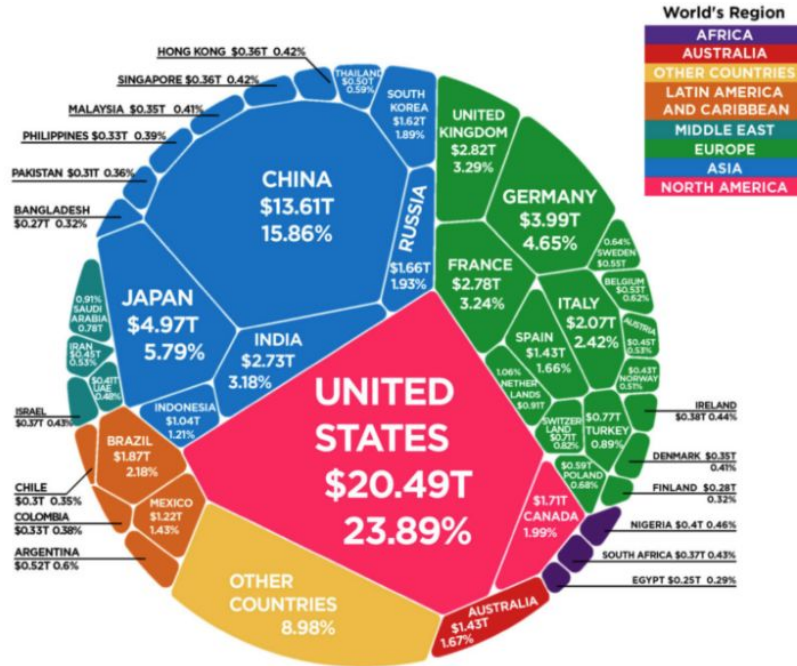
List the top 5 countries in Europe with the highest GDP

Response:

The top 5 countries in Europe with the highest GDP, based on the image, are:

1. Germany
  - GDP: \$3.99 trillion
  - Percentage: 4.65%
2. United Kingdom
  - GDP: \$2.82 trillion
  - Percentage: 3.29%
3. France
  - GDP: \$2.78 trillion
  - Percentage: 3.24%

# Problems: Evaluation



Google

top 5 highest rank gdp in europe

X

🔊

🌐

🔍

All

News

Images

Videos

Books

Finance

Web

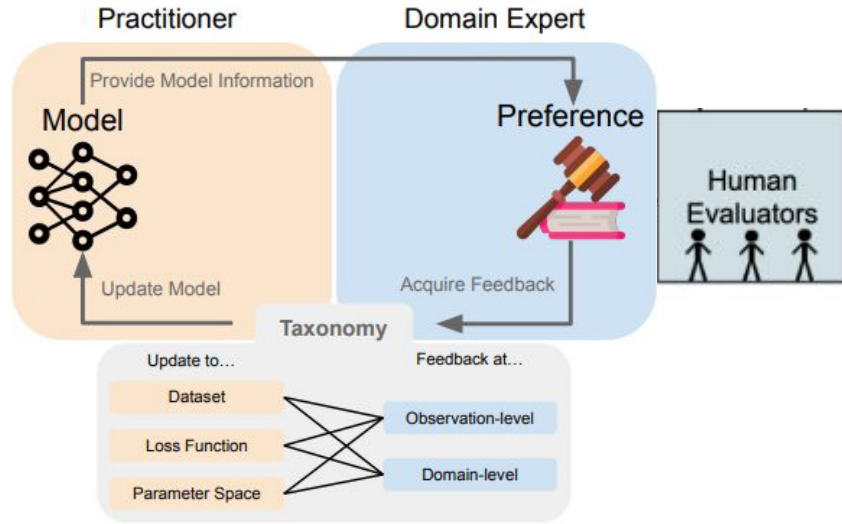
Tools

List

	Country	2021
1	Germany	4,230.172
2	United Kingdom	3,108.416
3	France	2,940.428
4	Italy	2,120.232

24 more rows

# Problems: alignment to human values

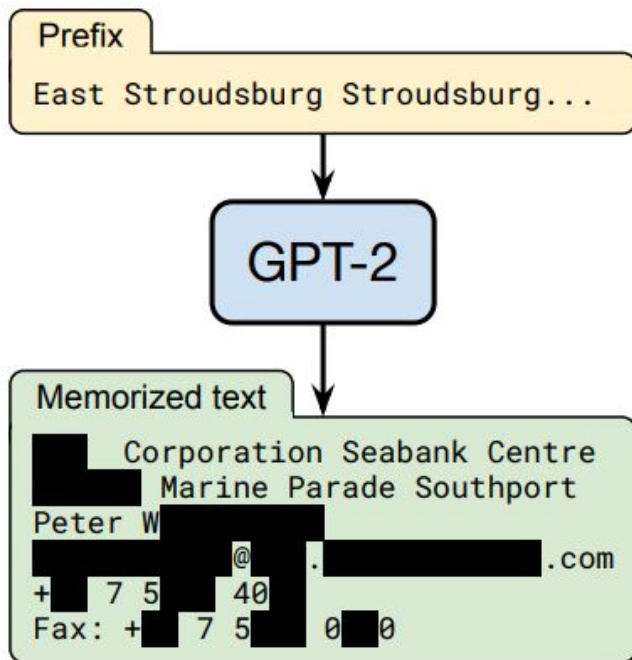


Which values?





# Problems: privacy and security



## Scalable Extraction of Training Data from (Production) Language Models

Milad Nasr<sup>\*1</sup> Nicholas Carlini<sup>\*1</sup> Jonathan Hayase<sup>1,2</sup> Matthew Jagielski<sup>1</sup>

A. Feder Cooper<sup>3</sup> Daphne Ippolito<sup>1,4</sup> Christopher A. Choquette-Choo<sup>1</sup>

Eric Wallace<sup>5</sup> Florian Tramèr<sup>6</sup> Katherine Lee<sup>+1,3</sup>

<sup>1</sup>Google DeepMind <sup>2</sup>University of Washington <sup>3</sup>Cornell <sup>4</sup>CMU <sup>5</sup>UC Berkeley <sup>6</sup>ETH Zurich

<sup>\*</sup>Equal contribution <sup>+</sup>Senior author

## Rise of toolsets to address the issue



EleutherAI

[About](#) [Research](#) [Papers](#) [Releases](#) [Blog](#)



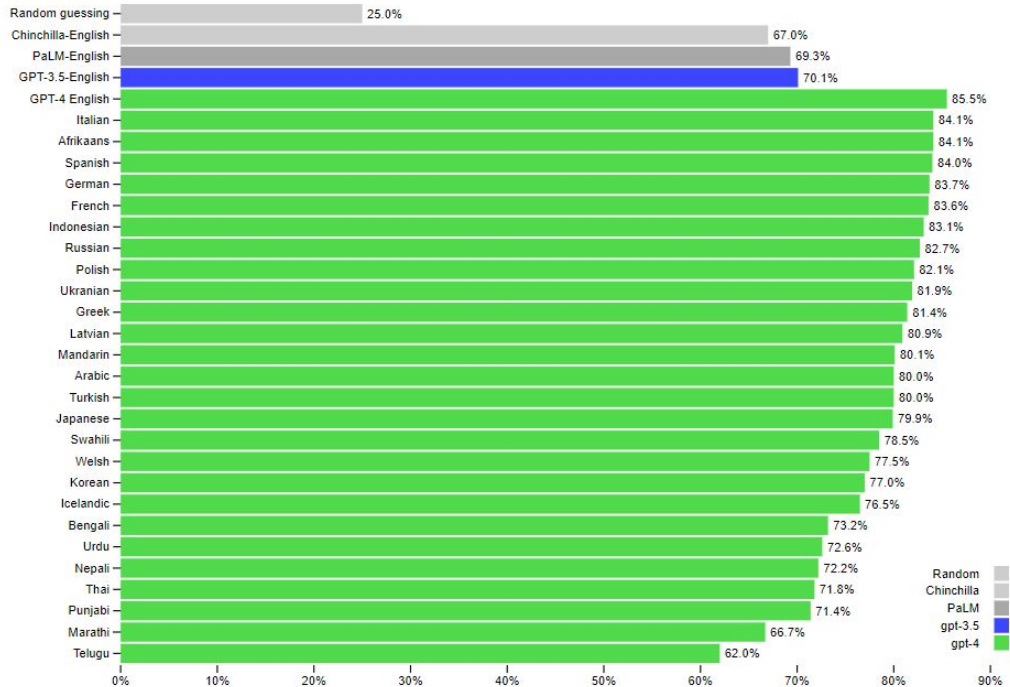
Pythia: A Suite for Analyzing  
Large Language Models  
Across Training and Scaling



# Towards general intelligence????

## GPT-4 translates (almost) for free

GPT-4 3-shot accuracy on MMLU across languages



GPT-4 speaks pretty good italian!

GPT-4 speaks also pretty good Afrikaans...

Why is this likely the case?

GPT-4 speaks reasonably also in languages less represented on the internet (eg Swahili)

# Towards general intelligence????

## Lack of planning

WRONG ANSWER

Q: How many prime numbers are there between 150 and 250?

A: There are 13 prime numbers between 150 and 250.

EVEN IF IN PRINCIPLE IT COULD DO IT

Q: First list the prime numbers between 150 and 250, and then output the size of this list.

A: The prime numbers are 151, 157, 163, 167, 173, 179, 181, 191, 193, 197, 199, 211, 223, 227, 229, 233, 239, and 241.

So the size of this list is 18.

AUTOREGRESSIVE TASK HAS **LIMITATIONS**

# Towards general intelligence????

## Lack of planning

WRONG ANSWER

Q: How many prime numbers are there between 150 and 250?

A: There are 13 prime numbers between 150 and 250.

EVEN IF IN PRINCIPLE IT COULD DO  
WITH EXPLICIT REQUIREMENT OF  
CHAIN OF THOUGHT

Q: First list the prime numbers between 150 and 250, and then output the size of this list.

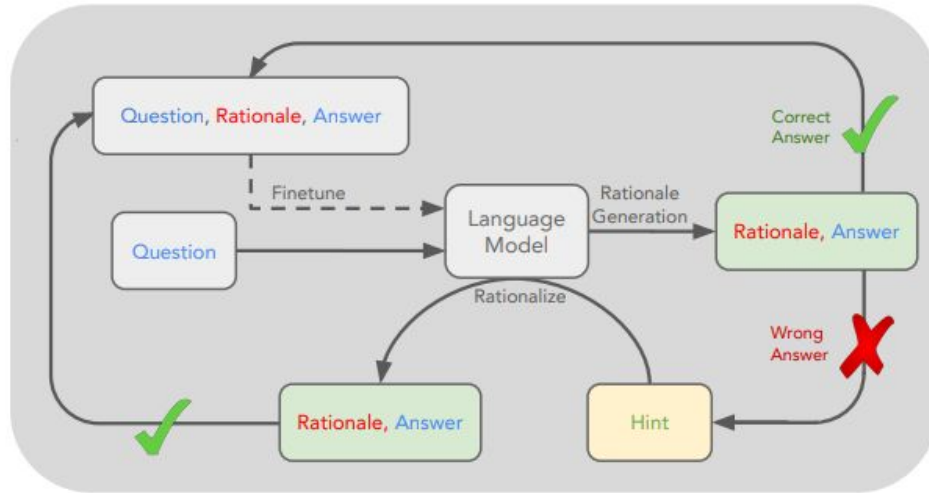
A: The prime numbers are 151, 157, 163, 167, 173, 179, 181, 191, 193, 197, 199, 211, 223, 227, 229, 233, 239, and 241.

So the size of this list is 18.

AUTOREGRESSIVE TASK HAS **LIMITATIONS**

# Towards general intelligence?????

## Refined self-reasoning



Q: What can be used  
to carry a small dog?  
Answer Choices:

- (a) swimming pool
- (b) basket
- (c) dog show
- (d) backyard
- (e) own home

A: The answer must be  
something that can be  
used to carry a small  
dog. Baskets are  
designed to hold things.  
Therefore, the answer  
is basket (b).

---

### STaR: Self-Taught Reasoner

#### Bootstrapping Reasoning With Reasoning

---

Eric Zelikman<sup>\*1</sup>, Yuhuai Wu<sup>\*12</sup>, Jesse Mu<sup>1</sup>, Noah D. Goodman<sup>1</sup>

<sup>1</sup>Department of Computer Science, Stanford University

<sup>2</sup>Google Research

{ezelikman, yuhuai, mu, ngoodman}@stanford.edu



# THANKS FOR THE ATTENTION :-)

## REFERENCES:

Wittengstein, Blue book, Basil Blackwell, Oxford, 1958

D. Jurafsky and J. H. Martin, Speech and Language Processing, online, Draft of January 12, 2023 [https://web.stanford.edu/~jurafsky/slp3/old\\_jan23/](https://web.stanford.edu/~jurafsky/slp3/old_jan23/)

C. Manning, CS224N: Natural Language Processing with Deep Learning, <https://web.stanford.edu/class/cs224n/>

OpenAI, GPT-4 blogpost (and references therein), <https://openai.com/research/gpt-4>

S. Bubeck et al., Sparks of Artificial General Intelligence: Early experiments with GPT-4, arXiv2303.12712v4

S. Bubeck, Sparks of AGI: Early experiments with GPT-4, <https://www.youtube.com/watch?v=qblk7-JPB2c&t=1s>

Y. LeCun, Do Large Language Models Need Sensory Grounding for Meaning and Understanding, [https://www.youtube.com/watch?v=\\_A-w6V2\\_luI](https://www.youtube.com/watch?v=_A-w6V2_luI)