



# On ChatGPT and Large Language Models

**Mustafa Jarrar**

Birzeit University

Palestine

# Outline

- AI/NLP Resources at Birzeit University
- The Golden Age of NLP
- Discussion and Open Issues

# NLP Resources at Birzeit University



# NLP Resources at SinaLab - Birzeit University

<https://sina.birzeit.edu>



الأنطولوجيا العربية

## Arabic Ontology / Wordnet

Formal Arabic Wordnet, 17K concepts/synsets



قواعد بيانات معجمية

## Lexicons & Lexicographic Databases

150 Multilingual lexicons translations, synonyms, glosses



مدونات و ذخائر لغوية

## Corpora & Datasets

Six dialects, 1.4 million tokens fully annotated. NER, WSD, ...



أدوات وبرمجيات

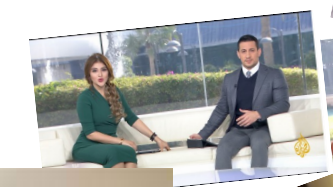
## NLP Tools

NER, WSD, synonyms, lemmatizer, POS tagger, etc.

NER 88.4% WSD 84% Synonyms

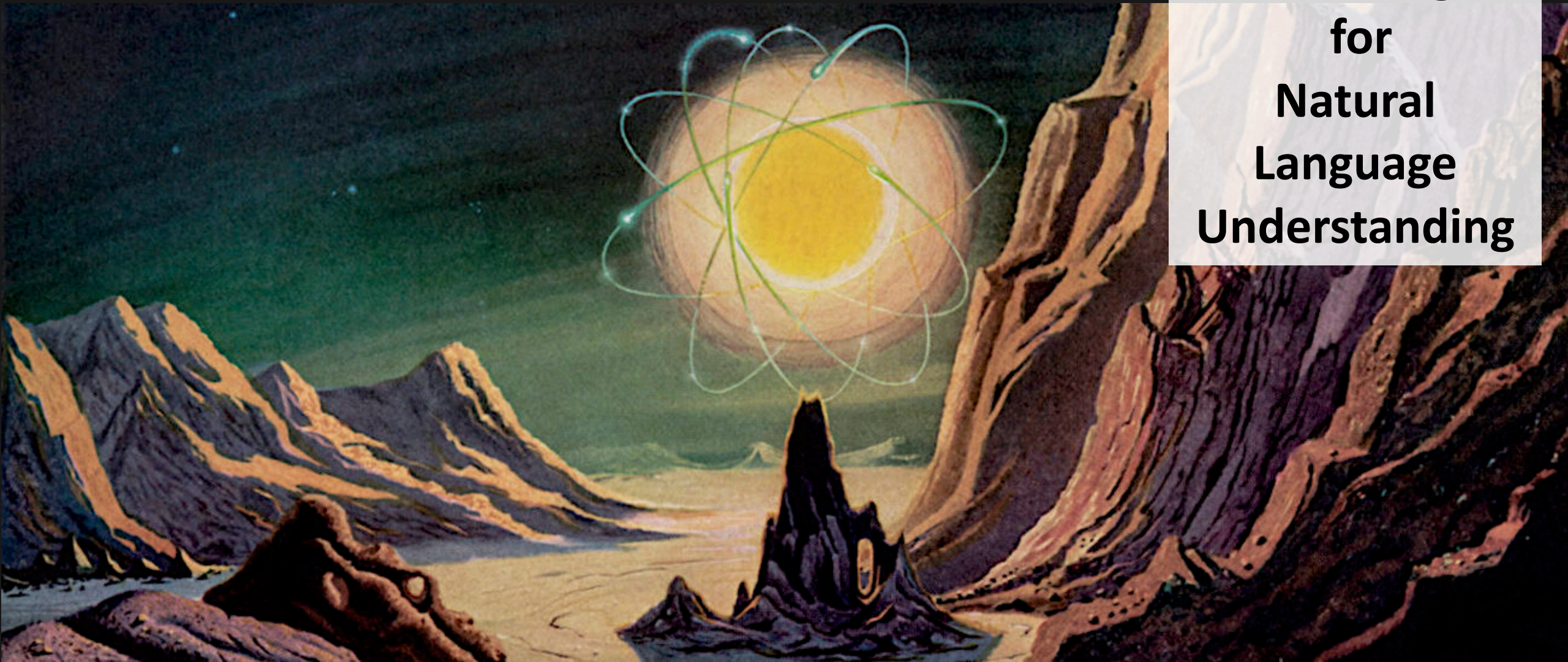
## Arabic Big Linguistic Data Graph

<https://ontology.birzeit.edu>





**Golden Age  
for  
Natural  
Language  
Understanding**





# AI Revolution

Text to Text

Text to poem

Text to Image

Text to Speech (character casting)

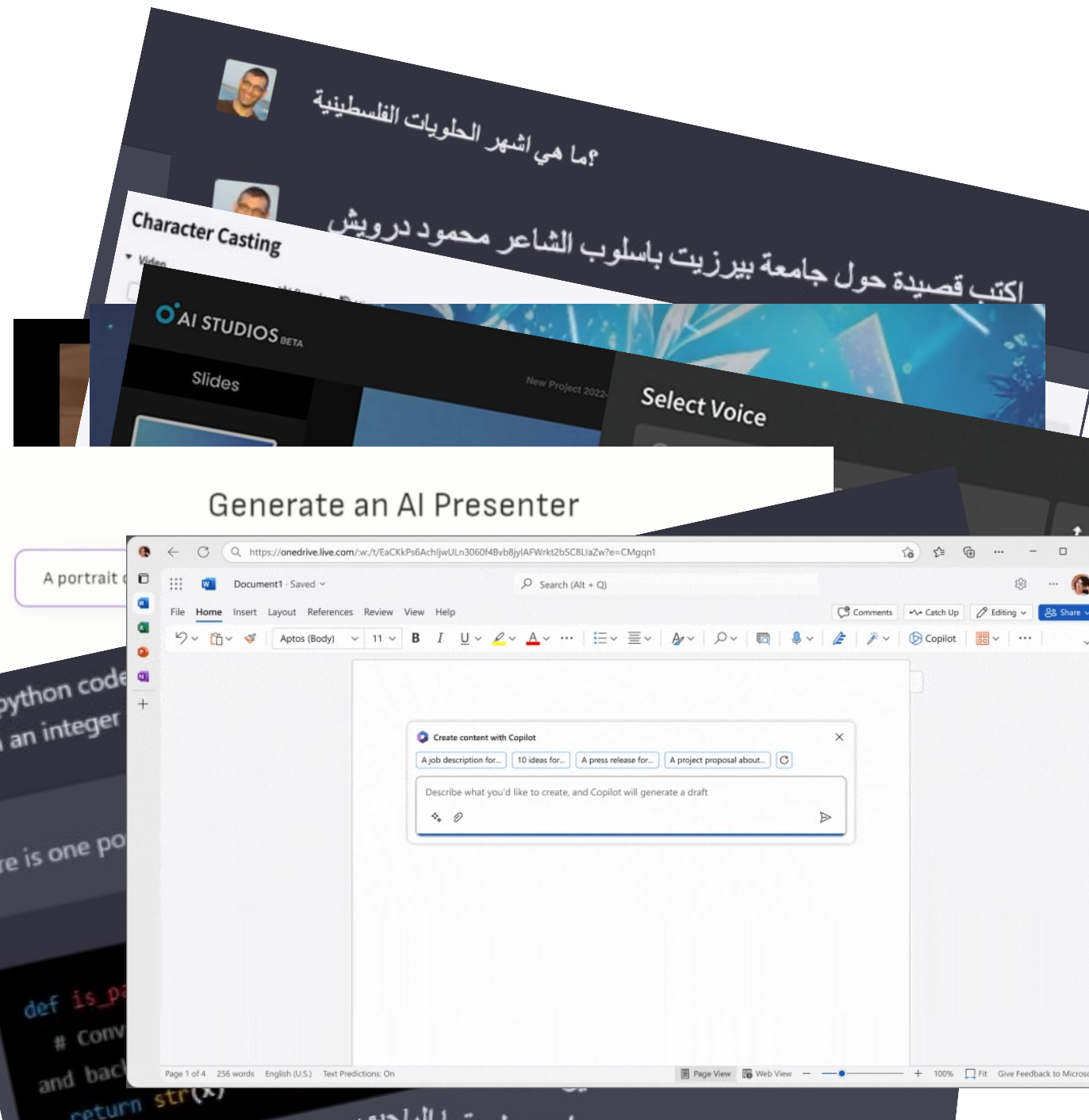
Text to Song

Text To Video

Text To Animation

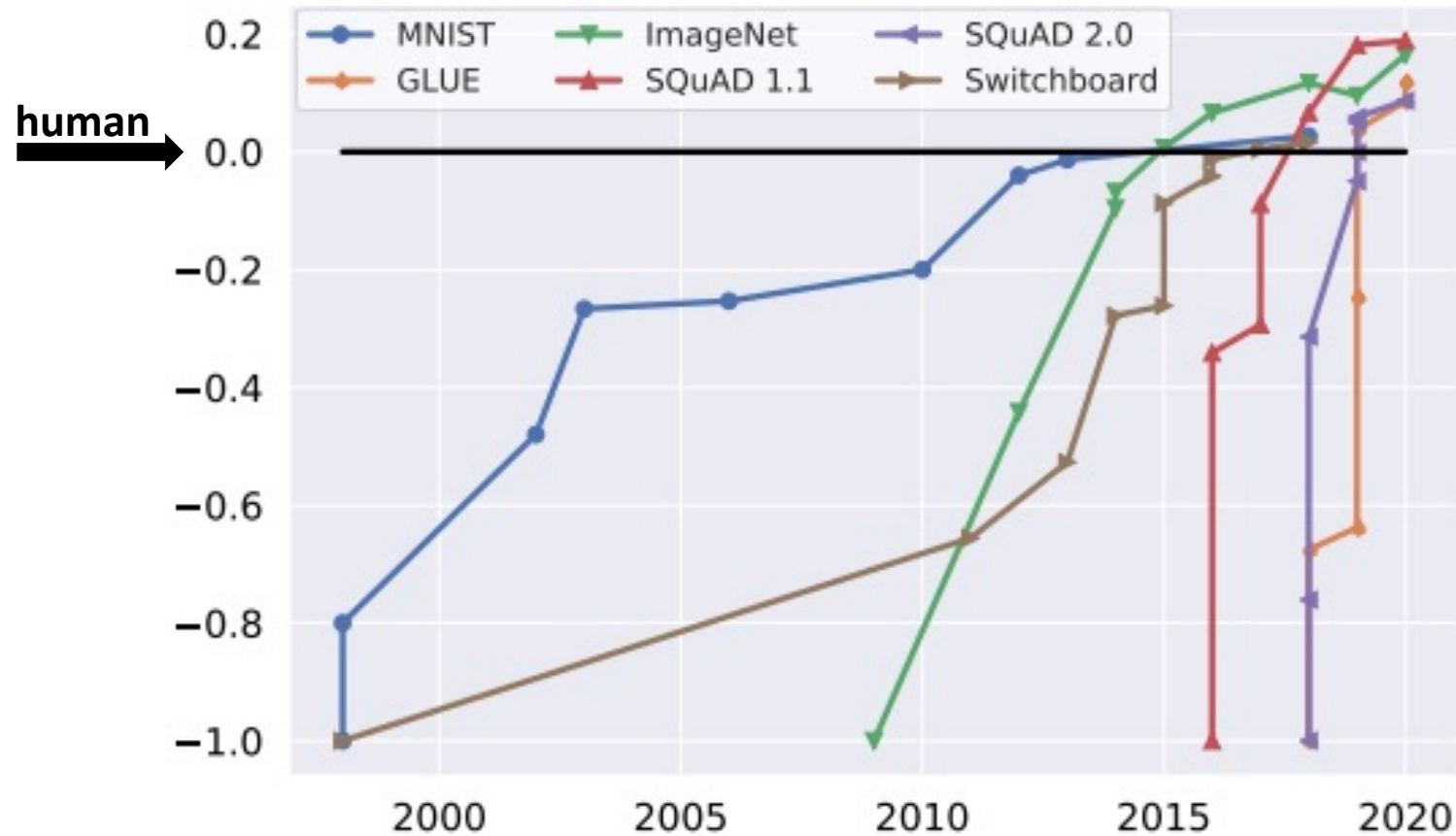
Text to Code

Text to Excel, PowerPoint ...



# Benchmark saturation over time

[Kiela et al, 2019](#)



MNIST: digital recognition

Switchboard: speech-to-text

SQuAD: Question answering

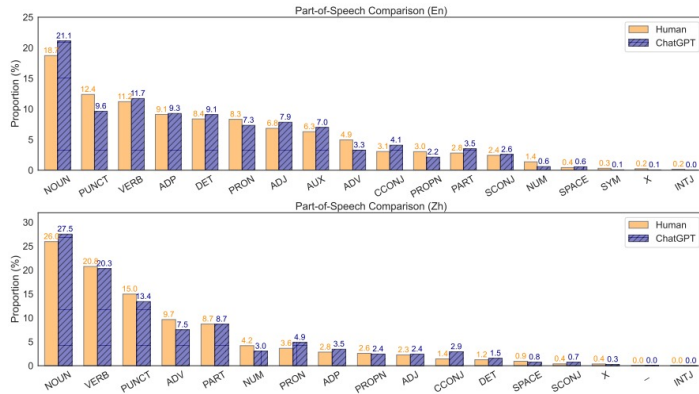
GLUE: Language Understanding tasks

Figure 1: Benchmark saturation over time for popular benchmarks, normalized with initial performance at minus one and human performance at zero.

# How Close is ChatGPT to Human Experts?

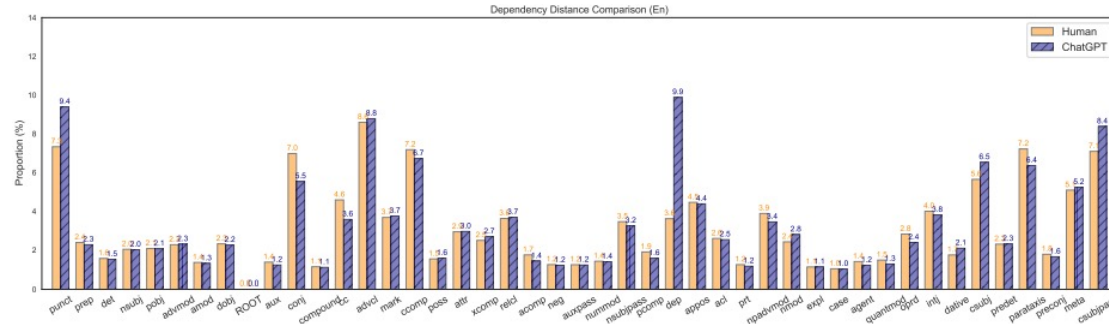
## Comparison Corpus, Evaluation, and Detection

Biyang Guo<sup>1†\*</sup>, Xin Zhang<sup>2\*</sup>, Ziyuan Wang<sup>1\*</sup>, Minqi Jiang<sup>1\*</sup>, Jinran Nie<sup>3\*</sup>  
Yuxuan Ding<sup>4</sup>, Jianwei Yue<sup>5</sup>, Yupeng Wu<sup>6</sup>



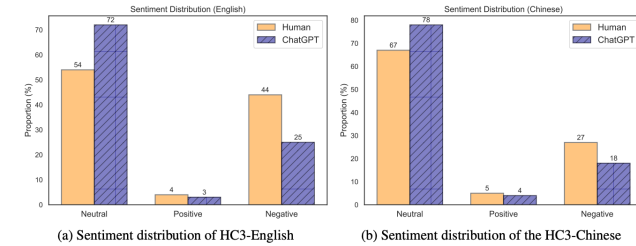
### Part-of-speech (English and Chinese)

قسم الكلام



### Dependency parsing (grammatical structure of a sentence)

تركيب الجمل



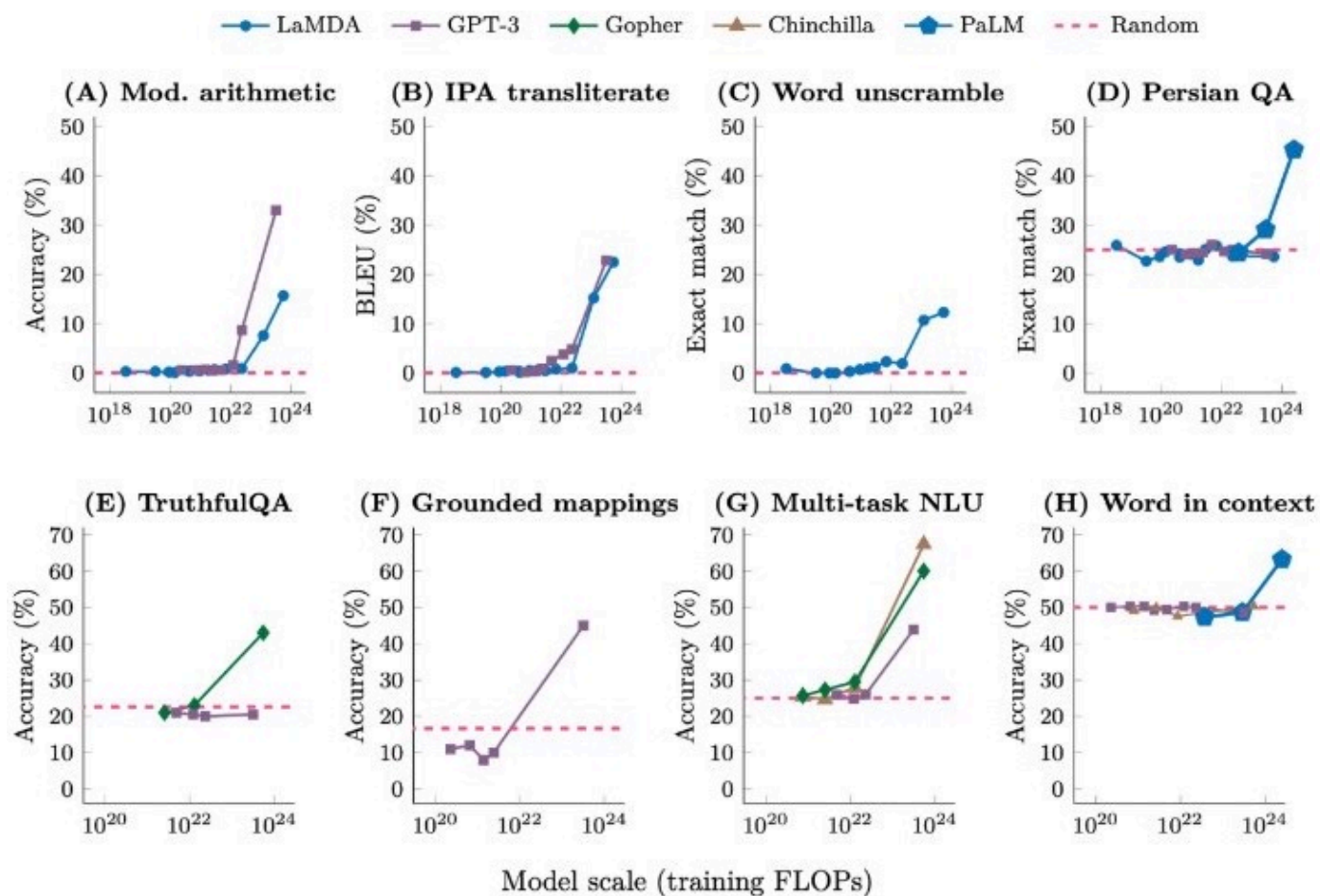
### Sentiments Analysis (English and Chinese)

تحليل المشاعر (مع، ضد، محايد)



# Large models are better (NLU tasks)

Eight examples of emergence in the few-shot prompting setting. Each point is a separate model. ([Wei et al, 2022](#))



# Outcompelete and Masking

## Next word prediction:

Given a sequence of words, can we predict the next word?

Palestinian olive is good for cholesterol

is	good	for	cholesterol
sou	a	food	health
tre	what	at	pregnancy
gar	green	one	hair
bra	excellent		skin

Recall

Conditional Probability

$$P(A|B) = \frac{P(A,B)}{P(B)}$$

# Outcompelete and Masking

## Next sentence prediction:

Given a set of sentences, can we predict the next sentence?

Palestinian Extra virgin olive oil made from unrefined, cold-pressed olives.  
Palestinian olive is good for cholesterol.

Eating a teaspoon of olive oil daily, works to control the level...  
Olive oil is one of the richest natural sources and the most contain...  
Benefits of olive oil in the treatment of diabetes:



# Outcompelete and Masking

## Masking

Given a text and a masked word, can we predict this word?

Look at the words before and after

Palestinian olive is good for cholesterol. Palestinian Extra virgin olive oil made from unrefined, cold-pressed olives. Eating a teaspoon of olive oil # , works to control the level of harmful cholesterol in the blood and triglycerides.

daily  
everyday  
a  
is

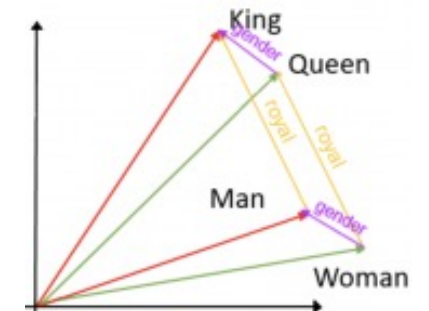
# Language Models

**النماذج اللغوية** هي نماذج إحصائية يتم تدريبها على حجم ضخم جدا من النصوص المكتوبة (ما كتبه الانسان)، لكي تتمكن من ترجيح الكلمات الجمل الأكثر احتملا (أكثر تداولاً وانتشاراً)

- A statistical model that can predict the next word/sentence is called a **language model**.
- **Large Language Models** are typically trained with deep learning on **large corpora**.
- **Language modeling techniques:** n-gram ... Word2Vec ... LSTM .... **Transforms** (BERT, GPT...)

**Word/Sentence Embedding (تمثيل الكلمة/الجملة كمتجهات):** repressing it in a vector of numbers, such that the cosine similarity to similar words are closer.

Word Embedding Vector    [... 0.3, 0.8, -0.1, 0.8, -0.9, .0.01, 0.05, .....]



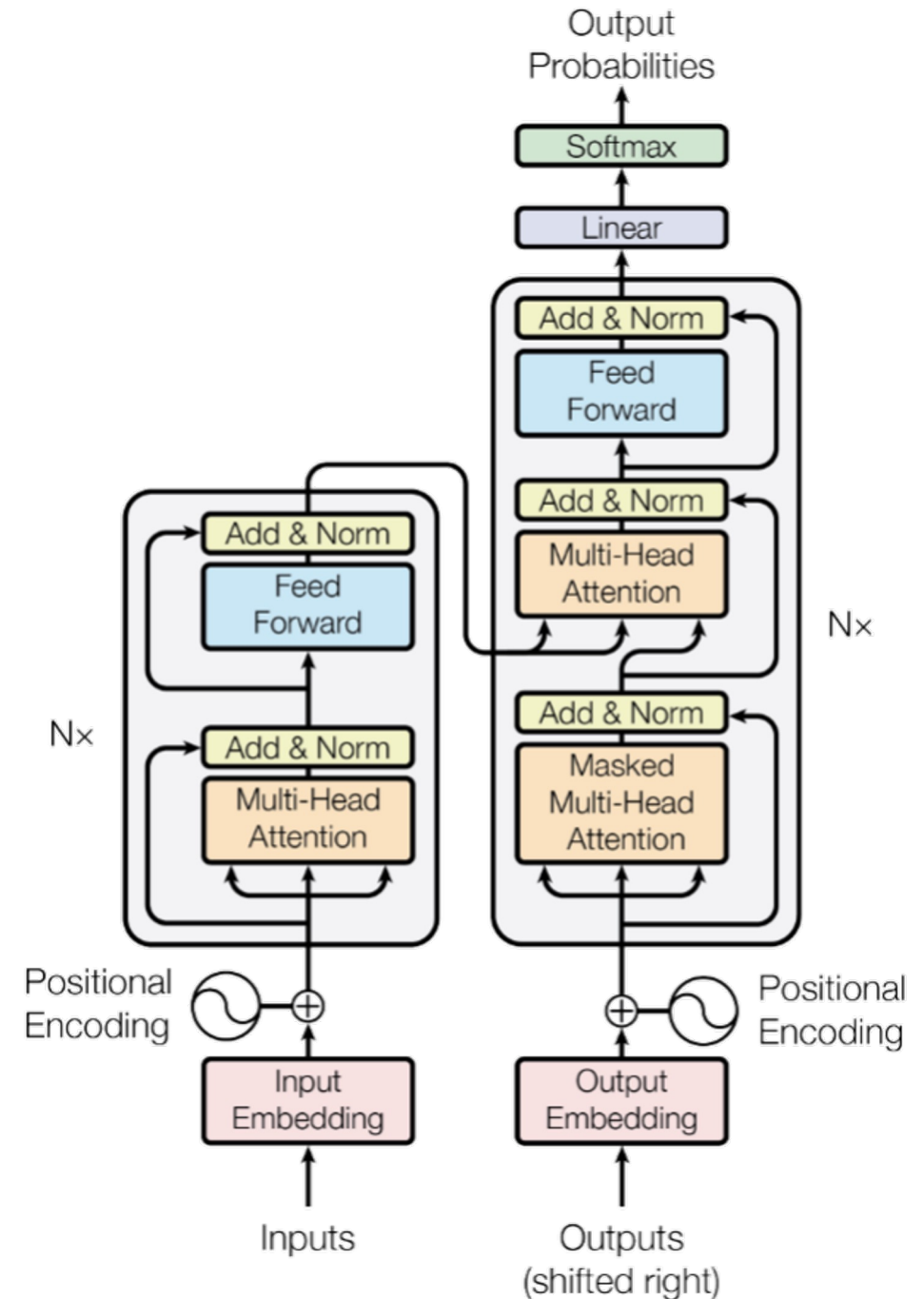
# Transformers

Encoder/decoder architecture (deep learning)

Introduced in 2017 **by Google** (Brain team)

Used in Natural Language Processing and Computer Vision

Based on this Transformer architecture, different variations of transformers are proposed, Such as BERT, GPT and T5.





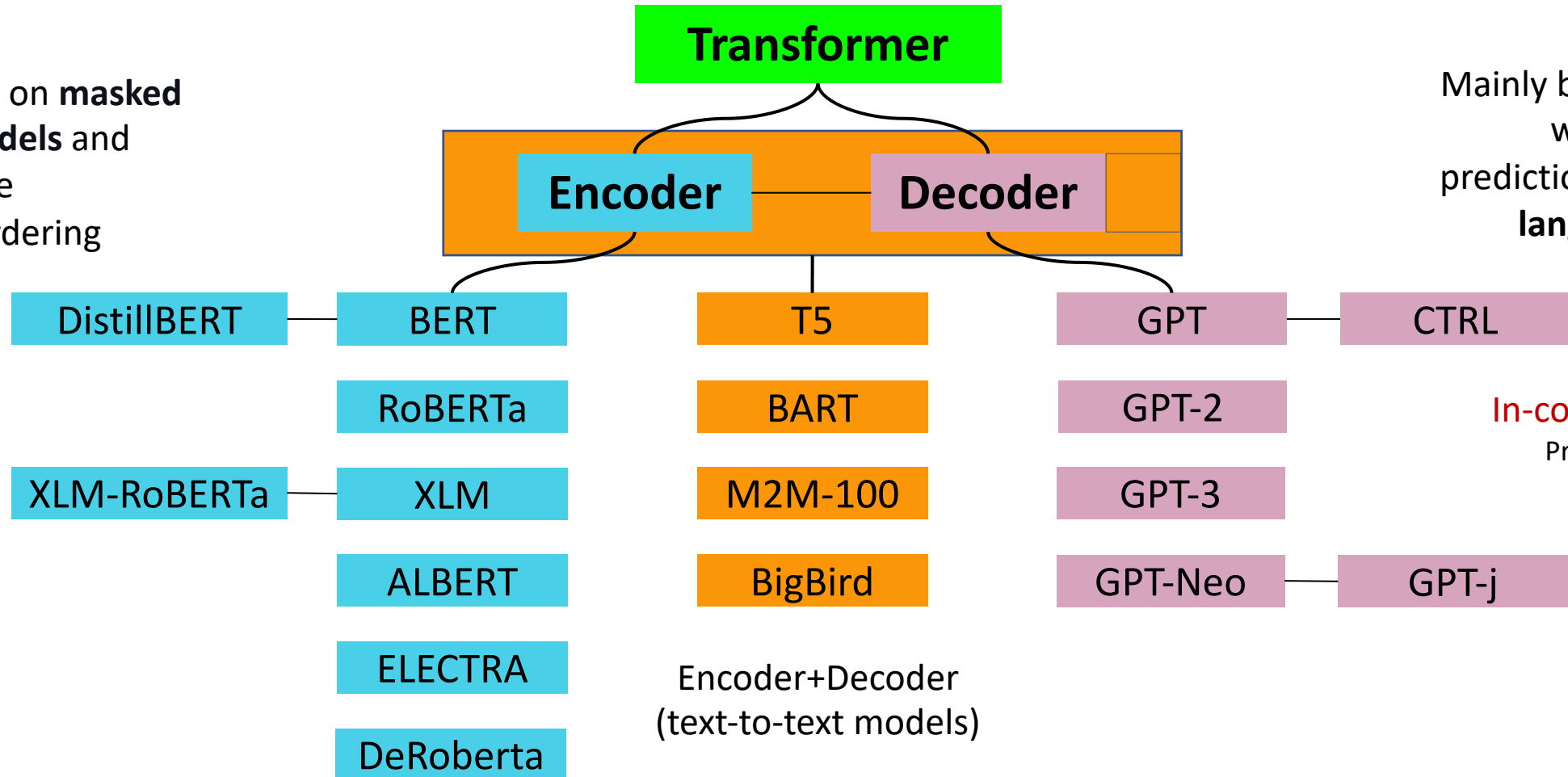
# Transformers Tree

## Encoder/Decoder-based Transformers

Based on [@Datafuse Analytics](#) Channel

mainly based on **masked language models** and next-sentence prediction/ordering

Pre-train  
Fine-tuning

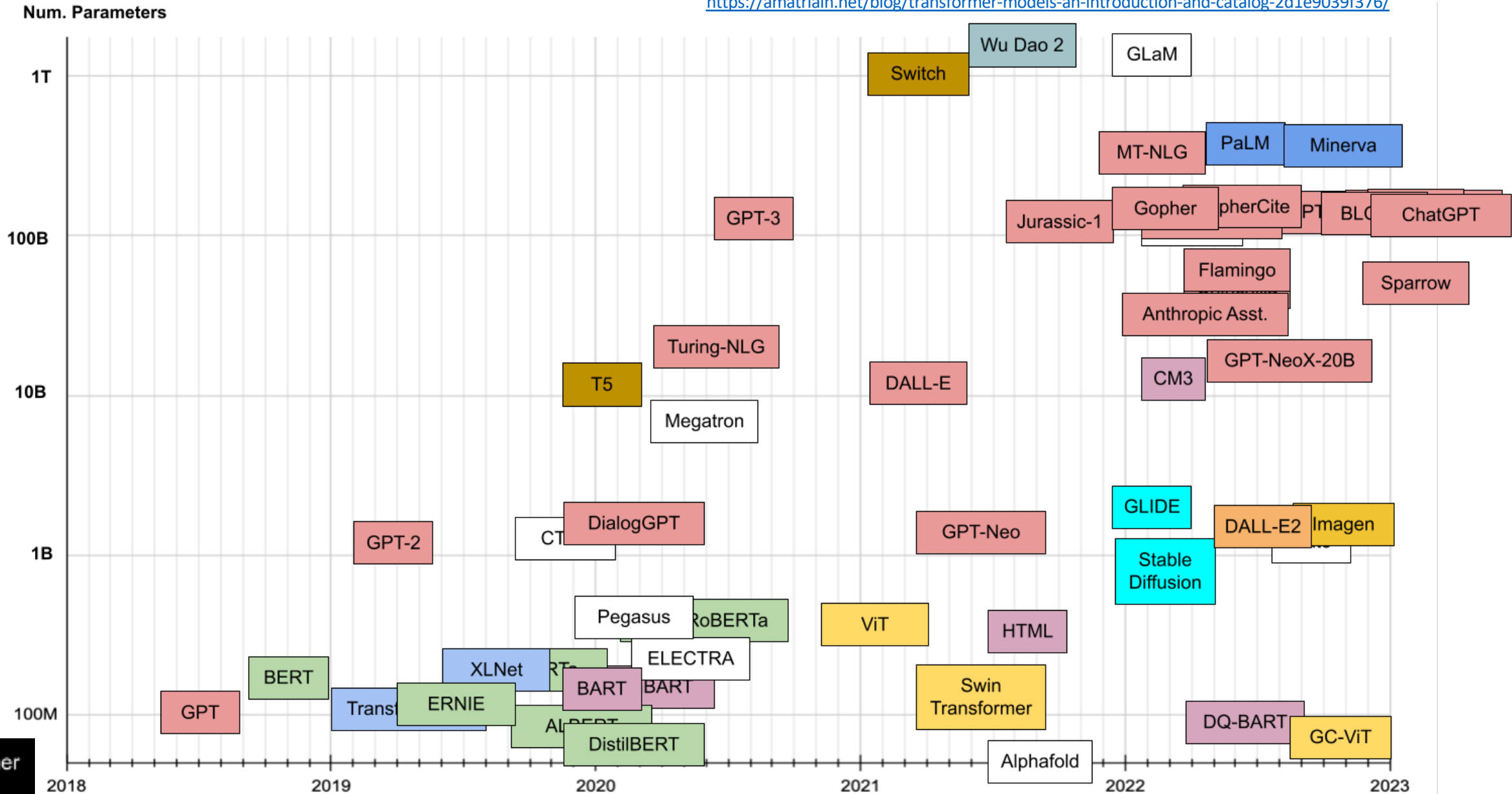


Mainly based on next-word/sentence prediction, **generative language models**

Train  
In-context learning  
Prompt Engineering

# Large Language Models (Chronological timeline)

<https://amatriain.net/blog/transformer-models-an-introduction-and-catalog-2d1e9039f376/>



Transformer  
(2018)

# Example

Model: GPT-3.5

لماذا سمي باب الأسباط بهذا الاسم؟



باب الأسباط هو باب يقع في الجدار الشمالي للمسجد الأقصى في مدينة القدس، ويعود تسميته إلى قصة تاريخية ودينية في الإسلام.

تقول القصة أن النبي محمد صلى الله عليه وسلم قام بإرسال رسالة إلى الحاكم البيزنطي هرقل بن الوليد دعاه فيها إلى الإسلام. وفي الرد على الدعوة، قام هرقل بطمس الرسالة وأرسل جيشًا بقيادة الجنرال بابا بن أبي سفيان لاستعادة الرسالة من المسلمين. وفي المعركة التي وقعت بين الجيشين عند باب الأسباط، حصلت القوات الإسلامية على انتصار واستعادت الرسالة المكتوبة، وتم إعادتها إلى النبي محمد صلى الله عليه وسلم.

بناءً على هذه القصة، أطلق على الباب اسم "باب الأسباط" تكريمًا للفصل الذي وقع بين الجيشين

وانتصار الجيش الإسلامي في استعادة الرسالة المكتوبة. وتُعد هذه القصة جزءًا من التاريخ



# Example

Model: GPT-4

لماذا سميت رام الله بهذا الاسم؟



- رام الله هي مدينة في الضفة الغربية في فلسطين. اسمها باللغة العربية يعني "تل الله" أو "ارتفاع الله". لكن الأصل الدقيق للاسم غير معروف بالتحديد ولكن هناك بعض النظريات.
1. نظرية أولى تقترح أن الاسم قد يكون مشتقاً من الكلمة الكنعانية "راما" التي تعني "الارتفاع"، مما يشير إلى الموقع الجغرافي للمدينة على تل.
  2. نظرية أخرى تشير إلى أن الاسم قد يكون مشتقاً من الكلمة الأرامية "رهم" التي تعني "الراحة"، في إشارة إلى الراحة التي يمكن أن تجدها الناس في هذا المكان.
  3. نظرية أخرى تقترح أن الاسم قد يكون مشتقاً من اللغة العبرية، حيث يعني "رام" "رفيع" أو "مرتفع"، و "الله" هو الله، لذا يمكن تفسير الاسم ككل على أنه "التل الرفيع".
- هذه فقط بعض النظريات ولا يوجد اتفاق عام على الأصل الدقيق للاسم.

# Example

Model: GPT-4

بما يتميز الشاعر محمود درويش؟



الفلسطيني والقضية الفلسطينية. تتميز أعماله بمجموعة من الخصائص الفريدة: الوطنية والقضية الفلسطينية: كتب درويش بشكل مكثف عن تجربة النزوح والشتات والحنين إلى الوطن. أشهر قصائده، مثل "رسالة الى القاصي الغربي" و "أرض السلام الجميلة"، تعكس معاناة الشعب الفلسطيني والقضية الفلسطينية. اللغة الشعرية: كان درويش مبدعاً في استخدام اللغة العربية، حيث استخدم التشبيه والاستعارة بطرق جديدة وفريدة من نوعها. هذا جعله واحداً من أهم الشعراء العرب في القرن العشرين. الحب والإنسانية: بالإضافة إلى القضايا الوطنية والسياسية، كتب درويش أيضاً عن الحب والإنسانية. قصائده تعكس حالة الإنسان والعلاقات البشرية بطريقة عميقة ومؤثرة. التأمل الفلسفي: كانت قصائد درويش غنية بالتأملات الفلسفية حول الحياة والموت والزمن والذاكرة. استخدم الشعر كوسيلة لاستكشاف هذه المواضيع العميقة. في النهاية، يُذكر محمود درويش كأحد أبرز الشعراء العرب والفلسطينيين، وترك إرثاً غنياً من الشعر الذي يستمر في الإلهام حتى اليوم.

# Example

Model: GPT-4

ما هو الفرق بين الشاعر محمد درويش والشاعر سميح القاسم



محمود درويش وسميح القاسم هما من أبرز الشعراء الفلسطينيين والعرب في القرن العشرين. على الرغم من أنهما شاركا في العديد من القضايا، ولا سيما القضية الفلسطينية، إلا أن لكل منهما أسلوبه ورؤيته الخاصة في الشعر.

**\*\*محمود درويش\*\***: يعتبر محمود درويش من أبرز الشعراء الفلسطينيين وقد اشتهر بشعره الذي يتعامل مع قضايا الهوية الوطنية، الحب، والحنين إلى الوطن. كما كان لغته الشعرية مميزة وغنية بالصور الجمالية والألغام الفلسفية.

**\*\*سميح القاسم\*\***: كان سميح القاسم شاعرًا فلسطينيًا معروفًا أيضًا بتعامله مع قضايا الهوية والنضال الوطني. كان شعره أكثر تحديًا ومواجهة، وكان يعبر عن الغضب والاحتجاج والمقاومة. كما كان يستخدم اللغة العربية بطريقة قوية ومباشرة.

بالإضافة إلى ذلك، كان لكل منهما تأثيرات وتجارب حياتية مختلفة. محمود درويش كان يعيش في الشتات لفترة طويلة وعاش في العديد من البلدان، بينما كان سميح القاسم يعيش في فلسطين داخل الخط الأخضر وكان له تجربة مختلفة بسبب تجربته كفلسطيني داخل الدولة الإسرائيلية.

# Example

Model: GPT-3.5

متى تأسست جامعة بيرزيت؟



تأسست جامعة بيرزيت في العام 1978 كجامعة عربية حكومية في مدينة بيرزيت في الضفة الغربية، وهي تعد واحدة من أكبر الجامعات الفلسطينية وأهمها. تقع الجامعة على بعد حوالي 10 كيلومترات شمال مدينة القدس، وتضم الجامعة مجموعة واسعة من الكليات والمراكز البحثية والمعاهد، بما في ذلك كلية الطب وكلية الهندسة وكلية العلوم والتكنولوجيا وكلية العلوم الاجتماعية والإنسانية وغيرها، بالإضافة إلى برامج الدراسات العليا والدكتوراه.

# Example

Model: GPT-4

متى تأسست جامعة بيرزيت؟



تأسست جامعة بيرزيت في الضفة الغربية في فلسطين عام 1924. بدأت كمدرسة للبنين وتطورت تدريجياً إلى مدرسة ثانوية ثم كلية، وأصبحت جامعة بالكامل في عام 1975. منذ ذلك الحين، أصبحت واحدة من أبرز الجامعات في الضفة الغربية وفلسطين.

# Example

Model: GPT-3.5

When was Birzeit University founded?



Birzeit University was founded in 1924 as an elementary school for girls, and it gradually expanded over the years to become a university in 1975.



# Example

Model: GPT-4

When was Birzeit University founded?



Birzeit University, located in the West Bank, Palestine, was founded in 1924. It started as a boys' school and progressively evolved into a secondary school, then a college, and finally became a full-fledged university in 1975. Since then, it has been one of the most prominent universities in the West Bank and Palestine.

# Example

Model: GPT-4

In which year Birzeit University was founded?



Birzeit University, located near Ramallah, Palestine, was originally founded as an elementary school by Ms. Nabiha Nasir in 1924. It evolved to become a high school in 1942, and then transformed into a college of arts and sciences offering two-year degrees in 1961. It was officially chartered as a full-fledged university offering undergraduate programs in 1975.

# Example (Rephrasing)

Model: GPT-4

Rephrase the following: Birzeit University was awarded Mohammed Bin Rashid's Award for the Arabic Language 2019 for the development of an innovative Lexicographic Search Engine -the first of its kind worldwide. The award is given annually for the efforts and intelligent applications in preserving, promoting, and facilitating the learning of Arabic. The award was given to the university by his Highness Sheikh Mohammed Bin Rashid Al Maktoum in a large ceremony in Dubai.



In 2019, Birzeit University received the prestigious Mohammed Bin Rashid's Award for the Arabic Language. This global recognition was for their development of an innovative Lexicographic Search Engine, the first of its kind in the world. The award, bestowed annually, commends initiatives and smart applications that contribute to the preservation, promotion, and ease of learning Arabic. His Highness Sheikh Mohammed Bin Rashid Al Maktoum honored the university with this award during a grand ceremony held in Dubai.

# Example (Summarization)

Model: GPT-4

Summarize the following: Birzeit University was awarded Mohammed Bin Rashid's Award for the Arabic Language 2019 for the development of an innovative Lexicographic Search Engine -the first of its kind worldwide. The award is given annually for the efforts and intelligent applications in preserving, promoting, and facilitating the learning of Arabic. The award was given to the university by his Highness Sheikh Mohammed Bin Rashid Al Maktoum in a large ceremony in Dubai.



In 2019, Birzeit University was honored with the Mohammed Bin Rashid's Award for the Arabic Language, in recognition of its development of an innovative Lexicographic Search Engine, a first in the global arena. This annual award, presented by Sheikh Mohammed Bin Rashid Al Maktoum, acknowledges contributions to the preservation, enhancement, and learning facilitation of the Arabic language. The university received the award at a grand ceremony in Dubai.

# Example (Translation)

Model: GPT-4

Translate the following in Arabic: Birzeit University was awarded Mohammed Bin Rashid's Award for the Arabic Language 2019 for the development of an innovative Lexicographic Search Engine -the first of its kind worldwide. The award is given annually for the efforts and intelligent applications in preserving, promoting, and facilitating the learning of Arabic. The award was given to the university by his Highness Sheikh Mohammed Bin Rashid Al Maktoum in a large ceremony in Dubai.



تم منح جامعة بيرزيت جائزة محمد بن راشد للغة العربية لعام 2019 لتطويرها محرك بحث لغوي مبتكر - الأول من نوعه على مستوى العالم. تُمنح الجائزة سنويًا تقديرًا للجهود والتطبيقات الذكية التي تساهم في الحفاظ على اللغة العربية وتعزيزها وتسهيل تعلمها. قام صاحب السمو الشيخ محمد بن راشد آل مكتوم بتقديم الجائزة للجامعة خلال حفل كبير أقيم في دبي.

# Example (Extracting Named Entities)

Model: GPT-4

Extract named entities from the following: Birzeit University was awarded Mohammed Bin Rashid's Award for the Arabic Language 2019 for the development of an innovative Lexicographic Search Engine -the first of its kind worldwide. The award is given annually for the efforts and intelligent applications in preserving, promoting, and facilitating the learning of Arabic. The award was given to the university by his Highness Sheikh Mohammed Bin Rashid Al Maktoum in a large ceremony in Dubai.



The named entities in the given text are:

1. Birzeit University
2. Mohammed Bin Rashid's Award for the Arabic Language 2019
3. Lexicographic Search Engine
4. Sheikh Mohammed Bin Rashid Al Maktoum
5. Dubai



# Example (Dialect Identification)

Model: GPT-4

In which Arabic dialect the following was written: سمعان أكل هريسة ، و الهريسة  
بتوجع البطن مو القلب



The provided sentence appears to be written in the Levantine Arabic dialect. Levantine Arabic is spoken in countries such as Syria, Lebanon, Jordan, and Palestine.

# Example (Poem Writing)

Model: GPT-4

اكتب قصيدة حول مجلس الطلبة في جامعة بيرزيت بأسلوب محمود درويش



في جامعة بيرزيت، نجوم المجلس تتوهج  
كأقمارٍ يضيءن السماء بالشجاعة والفجر  
يحملون راية الطلبة بأمجادها وأفراحها  
بأحلامٍ وأمانٍ تتعالى في الحضور

بأنغام الشعر والأدب يرتقون  
مثل طيور النورس يحلقون  
مبدعون وفنانون يروون قصصًا جميلة  
بأقلامهم ترسم للحياة صورًا طيبة

يسمع صوته يهز الروح والأفئدة  
بأفكارهم الجريئة والمختلفة  
يقدمون العطاء بلا حدود ولا ملل

بالتواضع والصدق والنية

# Example (Rephrasing)

Model: GPT-4

Give me synonyms of the word (table) in the following sentence: I sent you the table in the email yesterday



In the context of the sentence "I sent you the table in the email yesterday", where "table" likely refers to a structured set of data, here are some potential synonyms:

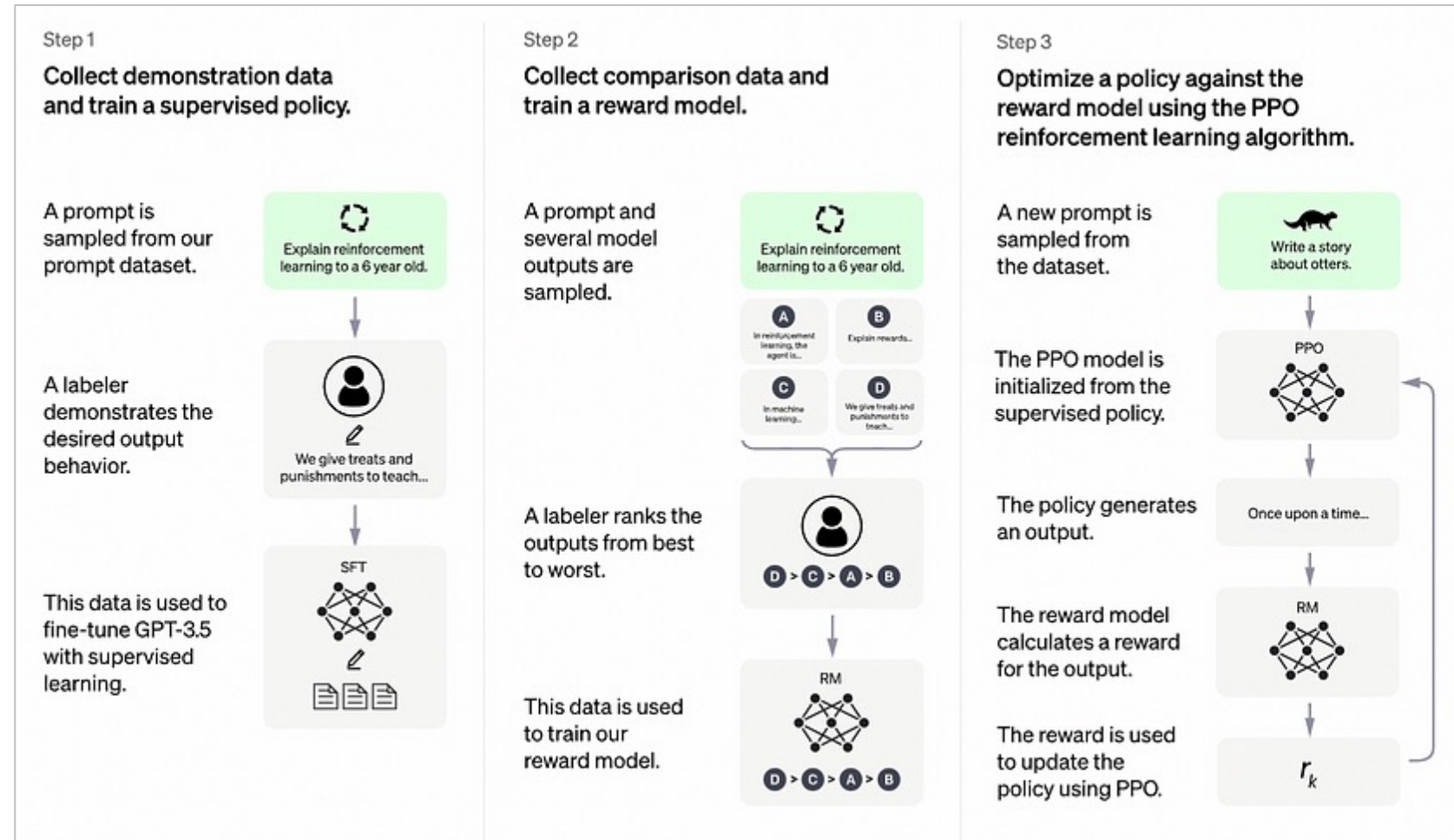
1. Chart: "I sent you the chart in the email yesterday."
2. Grid: "I sent you the grid in the email yesterday."
3. Spreadsheet: "I sent you the spreadsheet in the email yesterday."
4. Matrix: "I sent you the matrix in the email yesterday."
5. Dataset: "I sent you the dataset in the email yesterday."

Please note that while these words have similar meanings, they might not be perfect replacements depending on the specific context. For example, a "spreadsheet" might be a more appropriate term if the data was sent as an Excel file, while a "dataset" might be more appropriate for a large set of raw data.

# Humman Feedback is Important

## Chat GPT:

- Trained on labeled data to generate responses to user inputs.
- Rewarding more desirable responses.
- The Rewards Model is used as a scalar reward for **Reinforcement Learning**.





# Discussion and Open Issues



# Natural Language Interface (Prompts)

البرمجة/الأوامر البرمجية بلغة الانسان  
اللغة الطبيعية أصبحت هي لغت تواصلنا مع الآلة، وهي لغة تواصل الآت في بينها

We are also witnessing that people communicate with/ promote machines (and machine communicate with/ promote machines) in natural language





هل يمكن للآلة ان تعرف ان النص الذي تم توليده يحتوي على معلومات غير دقيقة؟

# Do models know if the generated text is inaccurate?

Possible answers:

NO?!

Not yet!

I don't know!

\*\* Do humans also say something wrong without knowing?!

هل يمكن للآلة ان تأتي بمعرفة جديدة

Can language models come up with  
new/original knowledge?

Maybe not!

But models have good capacity in synthesizing information  
(Which may look like innovation)

# Some Predictions (المستقبل القريب)

**Just the beginning:** much more is expected in the next few years (or even months)

نحن في البداية فقط

**Corpora:** The importance of text datasets is increasing.

أهمية النصوص تزداد

**newcomers:** Institutions having large amounts of text will enjoy new NLP technologies. المؤسسات التي تملك نصوص ستصبح أكثر ذكيا

**AI assisting/ants** will be in most aspects of our life.

خدمات الذكاء الاصطناعي في كل نواحي الحياة والعمل

**AI persona** communicating with us (hard to distinguish if a natural person is talking to you!): business agents, teachers. شخصيات غير واقعية ستتواصل معنا، ولن نتمكن من تمييزها

**Disinformation,** It will be too hard to authenticate the origin of texts (and whether every part of the text is correct). Including **student papers**, media reports, blogs, etc. من غير الممكن معرفة مصدر المعلومات

**Amplified Disinformation (human-knowledge virus).** The spread of generated texts/images (inaccurate knowledge) will be again used for training, and so no ... resulting in more inaccurate knowledge.

المعلومات الخاطئة ستنتشر بشكل هائل

**Bias:** As machines learn from the mainstream text (most common/probable), the generated text/images will be biased, thus minorities are

التحيز ضد الأقليات في كل شيء

**AI positives and negatives will be Amplified worldwide (be prepared to face this).**

استعد لهذا العصر  
(لا احد يهتم بك ان لم تؤمن بهذا العصر)

# Universities in the AI Age

## Teaching & Students

- How to use these technologies in teaching? كيف يمكن استعمال تقنيات الذكاء الاصطناعي في العملية التدريسية
- How to deal with students' use/misused? (is it all bad?!!) كيف نضبط استعمال الطلبة لهذه التقنيات

## Offering

- Which courses and programs universities should offer at the BA/Master/PhD levels? ما هي المساقات والبرامج التي يجب استحداثها (بسرعة) من هو المسؤول عن زمام المبادرة!!!

## Research

- How can we contribute to Natural Language Understanding technologies?  
Training a large language model (e.g. GPT 3 costs 50M)  
All of this is led by big companies - university involvement هل يمكن للجامعة ان تساهم في هذا العلم؟  
الوضع اصبح اكثر صعوبة لان الشركات هي من يقود حاليا!

# References

- Sana Ghanem, Mustafa Jarrar, Radi Jarrar, Ibrahim Bounhas: A Benchmark and Scoring Algorithm for Enriching Arabic Synonyms. The 12th International Global Wordnet Conference (GWC2023), Global Wordnet Association. (pp. ). San Sebastian, Spain, 2023
- Sanad Malaysha, Mustafa Jarrar, Mohammad Khalilia: Context-Gloss Augmentation for Improving Arabic Target Sense Verification. The 12th International Global Wordnet Conference (GWC2023), Global Wordnet Association. (pp. ). San Sebastian, Spain, 2023
- Mustafa Jarrar, Mohammed Khalilia, Sana Ghanem: Wojood: Nested Arabic Named Entity Corpus and Recognition using BERT. In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2022), Marseille, France. 2022
- Mustafa Jarrar, Fadi Zaraket, Tymaa Hammouda, Daanish Masood, Martin Waehlich: Lisan: Yemeni, Iraqi, Libyan, and Sudanese Arabic Dialect Corpora with Morphological Annotations. arXiv, DOI 10.48550/ARXIV.2212.06468. 2022
- Karim El Haff, Mustafa Jarrar, Tymaa Hammouda, Fadi Zaraket: Curras + Baladi: Towards a Levantine Corpus. In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2022), Marseille, France. 2022
- Mustafa Jarrar: The Arabic Ontology - An Arabic Wordnet with Ontologically Clean Content. Applied Ontology Journal, 16:1, 1-26. IOS Press. 2021
- Moustafa Al-Hajj, Mustafa Jarrar: ArabGlossBERT: Fine-Tuning BERT on Context-Gloss Pairs for WSD. In Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021). PP 40--48, 2021
- Kareem Darwish, Nizar Habash, Mourad Abbas, Hend Al-Khalifa, Husein T. Al-Natsheh, Houda Bouamor, Karim Bouzoubaa, Violetta Cavalli-Sforza, Samhaa R. El-Beltagy, Wassim El-Hajj, Mustafa Jarrar, and Hamdy Mubarak. 2021. [A panoramic survey of natural language processing in the arab worlds](#). *Commun. ACM*, 64(4):72–81.
- Karim El Haff, Mustafa Jarrar, Tymaa Hammouda, and Fadi Zaraket. 2022. Curras + baladi: Towards a levantine corpus. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2022)*, Marseille, France.
- Mamoun Abu Helou, Matteo Palmonari, and Mustafa Jarrar. 2016. [Effectiveness of automatic translations for cross-lingual ontology mapping](#). *Journal of Artificial Intelligence Research*, 55(1):165–208.
- Mamoun Abu Helou, Matteo Palmonari, Mustafa Jarrar, and Christiane Fellbaum. 2014. [Towards building lexical ontology via cross-language matching](#). In *Proceedings of the 7th Conference on Global WordNet*, pages 346–354. Global WordNet Association.
- Mustafa Jarrar. 2005. [Towards Methodological Principles for Ontology Engineering](#). Ph.D. thesis, Vrije Universiteit Brussel.
- Mustafa Jarrar. 2011. [Building a formal arabic ontology \(invited paper\)](#). In *Proceedings of the Experts Meeting on Arabic Ontologies and Semantic Networks*. ALECSO, Arab League.
- Mustafa Jarrar. 2020. *Digitization of Arabic Lexicons*, pages 214–217. UAE Ministry of Culture and Youth.
- Mustafa Jarrar. 2021. [The arabic ontology - an arabic wordnet with ontologically clean content](#). *Applied Ontology Journal*, 16(1):1–26.
- Mustafa Jarrar and Hamzeh Amayreh. 2019.
- [An arabic-multilingual database with a lexico-graphic search engine](#). In *The 24th International Conference on Applications of Natural Language to Information Systems (NLDB 2019)*, volume 11608 of LNCS, pages 234–246. Springer.
- Mustafa Jarrar, Hamzeh Amayreh, and John P. McCrae. 2019. [Representing arabic lexicons in lemon - a preliminary study](#). In *The 2nd Conference on Language, Data and Knowledge (LDK 2019)*, volume 2402, pages 29–33. CEUR Workshop Proceedings.
- Mustafa Jarrar, Nizar Habash, Faeq Alrimawi, Diyam Akra, and Nasser Zalmout. 2017. [Curras: An annotated corpus for the palestinian arabic dialect](#). *Journal Language Resources and Evaluation*, 51(3):745–775.
- Mustafa Jarrar, Eman Karajah, Muhammad Khalifa, and Khaled Shaalan. 2021. Extracting synonyms from bilingual dictionaries. In *The 11th International Global Wordnet Conference (GWC2021)*, pages 215–222. Global Wordnet Association.
- Mustafa Jarrar, Fadi Zaraket, Rami Asia, and Hamzeh Amayreh. 2018. [Diacritic-based matching of arabic words](#). *ACM Asian and Low-Resource Language Information Processing*, 18(2):10:1–10:21.
- Mustafa Jarrar, Fadi A Zaraket, Tymaa Hammouda, Daanish Masood Alavi, and Martin Waehlich. 2022. [Lisan: Yemeni, irqi, libyan, and sudanese arabic dialect corpora with morphological annotations](#).
- Eman Naser-Karajah, Nabil Arman, and Mustafa Jarrar. 2021. [Current trends and approaches in synonyms extraction: Potential adaptation to arabic](#). In *Proceedings of the 2021 International Conference on Information Technology (ICIT)*, pages 428–434, Amman, Jordan. IEEE.