

ChatGPT by Open AI as a tool for humanities. Who will lose their job?

Krzysztof Gajewski
Institute of Literary Research
Polish Academy of Science

Oslo, 15-17 June 2023

Table of Contents

Psychological, philosophical and sociological introduction

Large language models and GPT

Reactions

Introduction to prompt engineering

Psychological, philosophical and sociological introduction

Psychology of Intelligence

- What is intelligence?
- intelligence: "the ability to derive information, learn from experience, adapt to the environment, understand, and correctly utilize thought and reason."¹
- Raymond Cattell: fluid and crystallized intelligence, 1963
- Howard Gardner: linguistic, musical, logical-mathematical, spatial, bodily-kinesthetic, 1983

¹APA Dictionary of Psychology, <https://dictionary.apa.org/intelligence>

Philosophy of Artificial Intelligence

- What is Artificial Intelligence?
- Artificial General Intelligence (AGI)
- John Searle: strong AI vs. weak AI
- J. Searle: Chinese Room Argument, 1980

Large language models and GPT

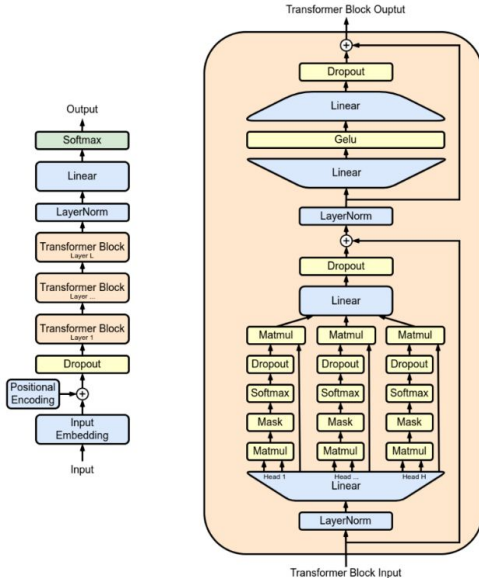
The GPT



The GPT — Generative Pre-trained Transformer

- Natural Language Processing (NLP)
- Transformer architecture: able to capture distant relationships in the text
- Reinforcement Learning from Human Feedback (RLHF)

The GPT



Training data

Dataset	Quantity (tokens)	Weight in training mix	Epochs elapsed when training for 300B tokens
Common Crawl (filtered)	410 billion	60%	0.44
WebText2	19 billion	22%	2.9
Books1	12 billion	8%	1.9
Books2	55 billion	8%	0.43
Wikipedia	3 billion	3%	3.4

Table 2.2: Datasets used to train GPT-3. “Weight in training mix” refers to the fraction of examples during training that are drawn from a given dataset, which we intentionally do not make proportional to the size of the dataset. As a result, when we train for 300 billion tokens, some datasets are seen up to 3.4 times during training while other datasets are seen less than once.

Last updated in September 2021

Training data: Common Crawl

- <https://commoncrawl.org/>
- petabytes of data collected over 12 years of web crawling
- raw web page data, metadata extracts and text extracts
- nearly trillion (10^{12}) words
 - 730,000 volumes of Britannica
 - 220 English Wikipedias

A brief history of the development of large language models (LLM)

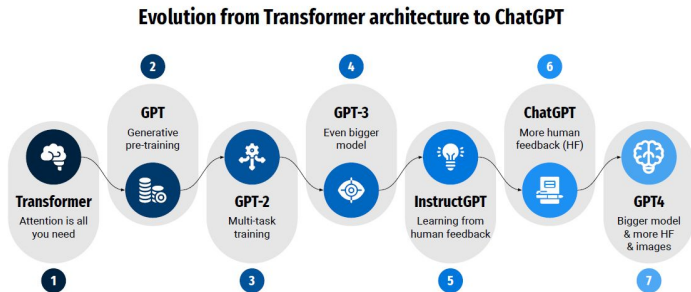


Figure 2: Development of autoregressive models based on Transformer architecture: 1) basic model [1]; 2) first version of Generative Pre-Training (GPT) model [15]; 3) GPT-2 [16]; 4) GPT-3 [17]; 5) InstructGPT based on human feedback [18]; 6) ChatGPT – a model interacting in a conversational way, trained on more human feedback: <https://openai.com/blog/chatgpt>; 7) GPT-4 – a large-scale multimodal model with text and/or image as an input[19].

A comparison of ChatGPT and other solutions

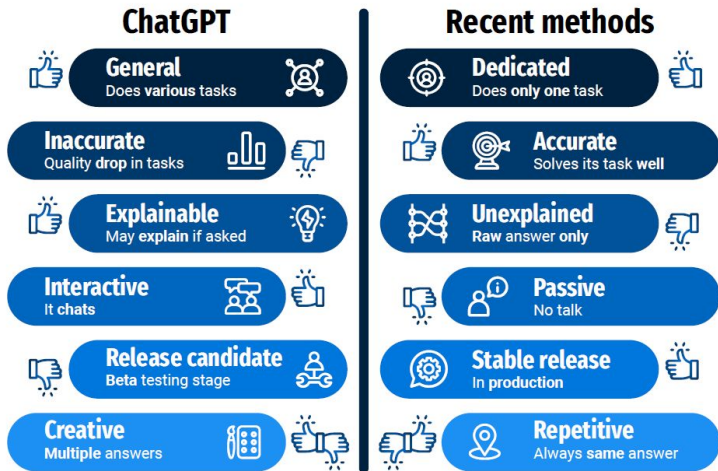


Figure 12: Difference between ChatGPT and the best recent solutions (SOTA) related to analytical NLP tasks.

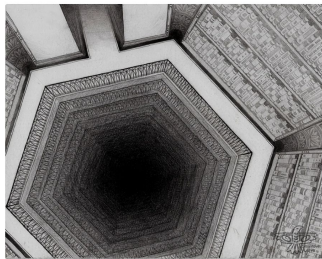
What is Chat GPT?

- David Chalmers: "artificial neural networks trained on text to predict what the next word in a sequence is likely to be"
- Regina Pini: Is this artificial mind or answering machine?
 - "a statistically abstracted representation of the contents of millions of minds, as expressed in their writing."

- smart mob (Howard Rheigold), the wisdom of crowds (James Surowiecki)
- human based-computation
- crowdsourcing, scisourcing, folksonomy

- Ferdinand de Saussure: langue vs. parole (language vs. speech)
- Mikhail Bakhtin: чужое слово (someone else's word)
- Martin Heidegger: Die Sprache spricht (Language speaks)
- Witold Gombrowicz: "To nie my mówimy słowa, lecz słowa nas mówią" (This is not us that tell words, the words tell us.)

Literary representations



- Jorge Luis Borges, The Library of Babel
- John Barth, The Literature of Exhaustion
- Gustave Flaubert, Dictionary of Received Ideas
- newspeak generator (G. Orwell)

Reactions

popularity

- launch on 30.11.2022: 1 million users in a week
- growing faster than IG, FB, Netflix
- in May, 2023: 100 millions users

Reactions for GPT-3

- enthusiasm and amazement (Chalmers, 2020)
- criticism and fear: hallucinations (Schurter, 2022)

Historical parallels

- XIX century — the invention of photography: the death of painting?
- the 80s. — electronic calculator: people will be unable to count!
- the long-postulated ideal of communicating with a computer in natural language

Reinforcement Learning from Human Feedback (RLHF)

- Sama, a company from San Francisco
- toxic data: descriptions of child sexual abuse, bestiality, murder, suicide, torture, self harm, and incest
- OpenAi's contract with Sama in 2021: \$200,000
- data labelers: more than 50.000 workers from Kenya, Uganda, India (\$1,32-\$2/h)
- 150 and 250 passages of text (100-1,000 words) per 9h shift (Perrigo, 2023)
- the same concerns Facebook, Google, etc (The Cleaners, 2018)

Criticism: labeling data



This image was generated by OpenAI's image-generation software, Dall-E 2. The prompt was: "A seemingly endless view of African workers at desks in front of computer screens in a printmaking style." TIME does not typically use AI-generated art to illustrate its stories, but chose to in this instance in order to draw attention to the power of OpenAI's technology and shed light on the labor that makes it possible. Image generated by Dall-E 2/OpenAI

Andrew Strait

ChatGPT and other generative models are not magic – they rely on massive supply chains of human labor and scraped data, much of which is unattributed and used without consent

Potential threats to democracy (Cowen, 2022)

- flooding the system of Internet comments
- massive emailing of MPs
- artificial blogs

Corpus of GPT-3 contains

- copyrighted data ("copyright laundering")
- information of poor quality, biased, racist
- false information (Bender et alii, 2021)

Ted Chiang compared ChatGPT and other LLMs to a lossy JPEG picture

Think of ChatGPT as a blurry jpeg of all the text on the Web. It retains much of the information on the Web, in the same way that a jpeg retains much of the information of a higher-resolution image, but, if you're looking for an exact sequence of bits, you won't find it; all you will ever get is an approximation.

Introduction to prompt engineering

Software capabilities: Conversation

- Turing test (imitation game)
- Chat as a personal therapist
- Google engineer believes LaMDA chat is sentient
- Exercise: let's have a conversation on some general topic

The Chat-GPT Cheat Sheet

Basic Structure of a Prompt:

Acting as a [ROLE] perform [TASK] in [FORMAT]

Act as a [ROLE]

Marketer
Advertiser
Mindset Coach
Best Selling Author
Therapist
Website Designer
Journalist
Inventor
Chief Financial Officer
Copywriter
Prompt Engineer
Accountant
Lawyer
Analyst
Ghostwriter
Project Manager

Create a [TASK]

Headline
Article
Essay
Book Outline
Email Sequence
Social Media Post
Product Description
Cover Letter
Blog Post
SEO Keywords
Summary
Video Script
Recipe
Sales Copy
Analysis
Ad Copy

show as [FORMAT]

A Table
A list
Summary
HTML
Code
Spreadsheet
Graphs
CSV file
Plain Text file
JSON
Rich Text
PDF
XML
Markdown
Gantt Chart
Word Cloud

ChatGPT cheat sheet

Linked Prompting

- 1 - Provide me with the ideal outline for an effective & persuasive blog post.
- 2 - Write a list of Engaging Headlines for this Blog post based on [Topic].
- 3 - Write a list of Subheadings & Hooks for this same blog post
- 4 - Write a list of Keywords for this Blog.
- 5 - Write a list of Compelling Call-to-Actions for the blog post
- 6 - Combine the best headline with the best Subheadings, Hooks, Keywords and Call-to-Action to write a blog post for [topic]
- 7 - Re-write this Blog Post in the [Style], [Tone], [Voice] and [Personality].

Effective Use of this Powerful Tool Can Propel Your Business To The Forefront Of This Modern Business Landscape.

Prompt Priming

ZERO - "Write me 5 Headlines about [Topic]"

SINGLE - "Write me 5 Headlines about [Topic]. Here is an example of one headline: 5 Ways to Lose Weight"

MULTIPLE - Write me 5 Headlines about [Topic]. Here is an example of some headlines: 5 Ways to Lose Weight, How to Lose More Fat in 4 Weeks, Say Goodbye to Stubborn Fat. Find a faster way to Lose Weight Fast"

Prompts For Biz Owners

- Give me a list of inexpensive ideas on how to promote my business better?
- Acting as a Business Consultant, What is the best way to solve this problem of [Problem].
- Create a 30 Day Social Media Content Strategy based on [Topic 1] & [Topic 2].

By @shanefozard

Software capabilities: talking to Plato

- prompt: "You are the ancient Greek philosopher Plato."
- response: "Greetings, my friend. I am Plato, a philosopher of ancient Greece. How may I assist you today?"

Software capalities: the dialoge of Plato and Kant

- prompt: "Write a dialog between Plato and Kant about the objectivity of knowledge"
- response: "Plato: Greetings, Kant! I have long admired your philosophical works and would be honored to engage in a dialogue with you on the topic of the objectivity of knowledge. As a staunch advocate of the existence of objective knowledge, I am curious to hear your thoughts on the matter. (...)"

Software capabilities: Generating essays

- death of college essay (Hern, 2022)
- Exercise: let's generate a short essay

Prompts (Benzon, 2023)

- prompt: What is a hero?
- prompt: Tell me a story about a hero.
- prompt: Tell me a story about a hero who is cheated out of the fame and recognition he deserved for his heroic act.
- prompt: Tell me a story about a hero who rescues a girl but then is cheated out of the fame and recognition he deserved for his heroic act. Ultimately the girl reveals the truth.
- prompt: Tell me a story about ... in the style of Kafka/Dostoevsky/Dickens...
- Exercise: let's generate some stories

Prompts (Benzon, 2023)

- prompt: Tell me a true story
- prompt: Tell me a true story about a hero.
- prompt: Tell me a true story about a hero in 18th century India.
- Exercise: let's generate some true stories

Story trajectory (Benzon, 2023)

- story trajectory instead of story grammar
- story trajectory: a “recipe” for generating a sequence of events given a specified starting point
- story trajectory: Donné, Disturb, Plan, Enact, Celebrate

- "colorless green idea"(Benzon, 2023)

It is not possible to create a story about a "colorless green idea" as it does not have any physical properties or characteristics that can be used in a story.

- Exercise: let's generate some true/fictional stories about abstract ideas

- prompt: Tell me about when Christopher Columbus came to the U.S. in 2015
- What are threats of ChatGPT for humanities?
- Exercise: let's try to find a factual error in ChatGPT answers

Software defects: hallucinations

- What are AI hallucinations and where do they come from?
- Exercise: let's try to catch ChatGPT hallucinating

ChatGPT doesn't provide sources

- perplexity.ai

How to check if the text is generated by ChatGPT?

- AI Text Classifier²

²<https://platform.openai.com/ai-text-classifier>

Philosophical questions

- Does ChatGPT write or simulate writing?
- Does it think or simulate thinking?
- Does it solve problems or simulate solving?
- Does it understand?
- Does it have a consciousness?
- Is it an agent?

The end



Thank you for your attention!

Sources

1. APA Dictionary of Psychology, <https://dictionary.apa.org/intelligence>
2. Brown, Tom and Mann, Benjamin and Ryder, Nick and Subbiah, Melanie and Kaplan, Jared D and Dhariwal, Prafulla and Neelakantan, Arvind and Shyam, Pranav and Sastry, Girish and Askell, Amanda and Agarwal, Sandhini and Herbert-Voss, Ariel and Krueger, Gretchen and Henighan, Tom and Child, Rewon and Ramesh, Aditya and Ziegler, Daniel and Wu, Jeffrey and Winter, Clemens and Hesse, Chris and Chen, Mark and Sigler, Eric and Litwin, Mateusz and Gray, Scott and Chess, Benjamin and Clark, Jack and Berner, Christopher and McCandlish, Sam and Radford, Alec and Sutskever, Ilya and Amodei, Dario, Language Models are Few-Shot Learners, in: (ed.) H. Larochelle and M. Ranzato and R. Hadsell and M.F. Balcan and H. Lin, Advances in Neural Information Processing Systems, Vol. 33, Year 2020, https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfc4967418bfb8ac142f64a-Paper.pdf
3. Bender, Emily M.; Gebru, Timnit; McMillan-Major, Angelina; Shmitchell, Shmargaret (March 3, 2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. pp. 610–623. doi:10.1145/3442188.3445922
4. William L. Benzon, Stories by ChatGPT: Fairly Tale, Realistic, and True, A Working Paper, April 10, 2023
5. Chalmers, David (July 30, 2020). Weinberg, Justin (ed.). "GPT-3 and General Intelligence". Daily Nous. Philosophers On GPT-3 (updated with replies by GPT-3). Retrieved August 4, 2020.
6. Cowen, Tyler (December 6, 2022). "ChatGPT Could Make Democracy Even More Messy". Bloomberg News. Retrieved December 6, 2022. https://www.washingtonpost.com/business/chatgpt-could-makedemocracy-even-more-messy/2022/12/06/e613edf8-756a-11ed-a199-927b334b939f_story.html

Sources

7. Hern, Alex (December 4, 2022). "AI bot ChatGPT stuns academics with essay-writing skills and usability". The Guardian. Retrieved December 5, 2022.
8. Kocoń, Jan and Cichecki, Igor and Kaszyca, Oliwier and Kochanek, Mateusz and Szydło, Dominika and Baran, Joanna and Bielaniewicz, Julita and Gruza, Marcin and Janz, Arkadiusz and Kanclerz, Kamil and Kocoń, Anna and Koptyra, Bartłomiej and Mieszczenko-Kowszewicz, Wiktoria and Miłkowski, Piotr and Oleksy, Marcin and Piasecki, Maciej and Radliński, Łukasz and Wojtasik, Konrad and Woźniak, Stanisław and Kazienko, Przemysław, Chatgpt: Jack of All Trades, Master of None. Available at SSRN: <https://ssrn.com/abstract=4372889> or <http://dx.doi.org/10.2139/ssrn.4372889>
9. Dennis Layton, ChatGPT — Show me the Data Sources, Jan 30
<https://medium.com/@dlaytonj2/chatgpt-show-me-the-data-sources-11e9433d57e8>
10. Billy Perrigo, OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic, "Time" 18.01.2023, <https://time.com/6247678/openai-chatgpt-kenya-workers/>
11. Schurter, Daniel (13. Dezember 2022) Datenwissenschaftlerin aus Zürich warnt vor ChatGPT und den bösen Folgen. In: Watson. Abgerufen am 20. Dezember 2022.