ChatGPT, technical details and applications

Huy Nguyen - CIRES LLM Reading Group

Introduction

- ChatGPT was introduced in November in 2022, by OpenAI
- General purpose chatbot application, based on the GPT 3.5 and GPT 4 (Generative Pre-trained Transformer) architectures, and is closed-source
- Only takes 5 days to reach 1 million users, a record at that time. (Now held by Facebook's Threads, which reached 1 million users in 1 hour)
- Currently 100 million users
- Support many languages (English, Spanish, French, German, Italian, Japanese, Chinese, etc.)

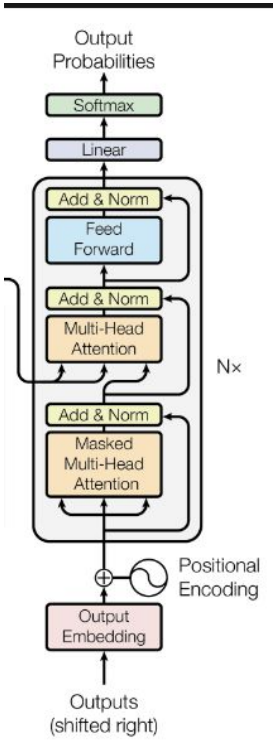
Technical details

- **OpenAI's GPT (Improving Language Understanding by Generative Pre-Training - Radford et al. 2018)**
- The model is built-upon the GPT model, which uses the decoder of the Transformer model
- Training process of GPT model:
- Unsupervised pre-training: U: corpus of token. Objective function: maximize likelihood

  (Language modeling objective)

$$L_1(\mathcal{U}) = \sum_i \log P(u_i | u_{i-k}, \ldots, u_{i-1}; \Theta)$$

- GPT authors used the Transformer decoder to produce output distribution



Technical details

- Supervised fine-tuning: Adapt params to the supervised target task.

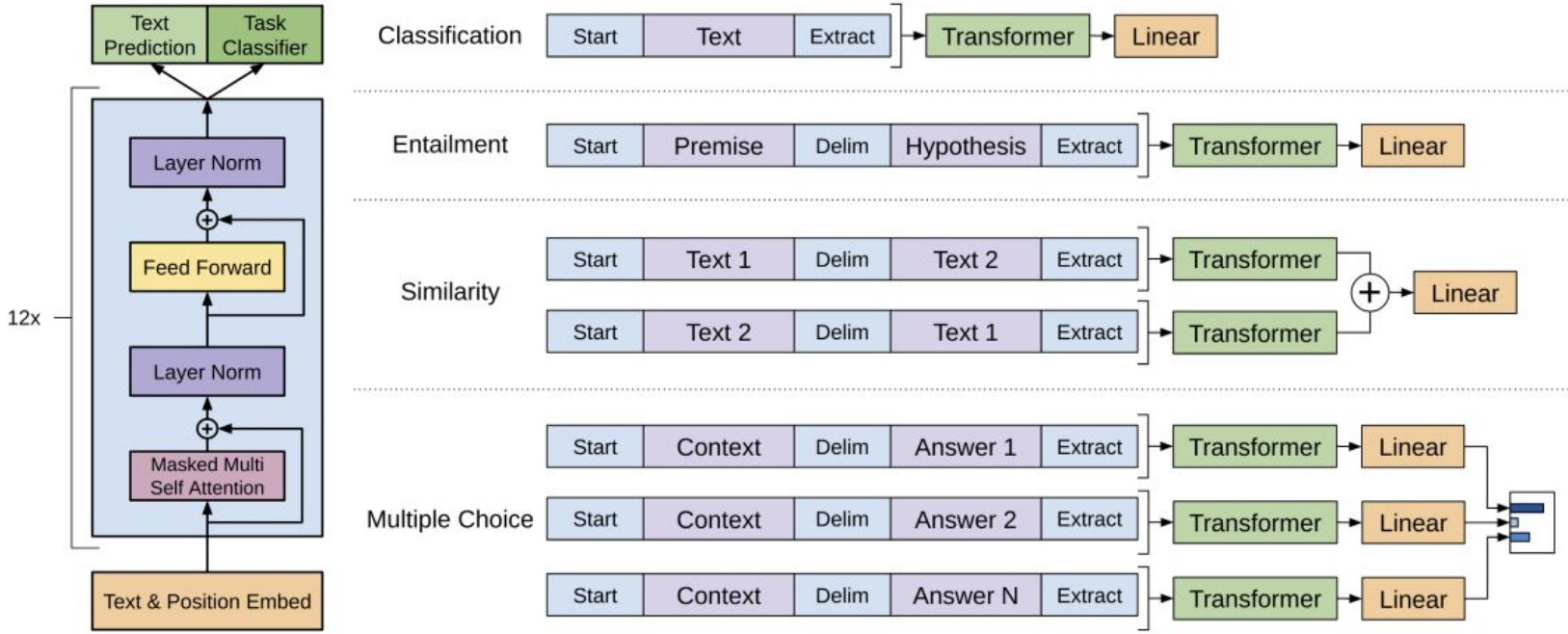
$$P(y|x^1, \ldots, x^m) = \texttt{softmax}(h_l^m W_y).$$

- Maximize the objective function:

$$L_2(\mathcal{C}) = \sum_{(x,y)} \log P(y|x^1, \ldots, x^m).$$

- Add the language modeling objective will increase training performance

$$L_3(\mathcal{C}) = L_2(\mathcal{C}) + \lambda * L_1(\mathcal{C})$$


Text Prediction
Task Classifier
12x
Layer Norm
Feed Forward
Layer Norm
Masked Multi Self Attention
Text & Position Embed
Classification
Start
Text
Extract
Transformer
Linear
Entailment
Start
Premise
Delim
Hypothesis
Extract
Transformer
Linear
Similarity
Start
Text 1
Delim
Text 2
Extract
Transformer
Start
Text 2
Delim
Text 1
Extract
Transformer
Linear
Multiple Choice
Start
Context
Delim
Answer 1
Extract
Transformer
Linear
Start
Context
Delim
Answer 2
Extract
Transformer
Linear
Start
Context
Delim
Answer N
Extract
Transformer
Linear

Technical details

- **Task-specific input transformations:** For the question and answering task, which is the basis of ChatGPT:
- Context document z, question q and set of possible answers {a_k}
- Concatenate the document context and question with each possible answer, adding a delimiter token in between to get [z; q; $; a_k].
- Each token processed independently, normalized with softmax to produce probability distribution of answers.

Technical details

- **GPT-2 (Language Models are Unsupervised Multitask Learners - Radford et al. 2019)**
- Based on GPT, with the following modifications:
  - Layer normalization was moved to the input of each sub-block
  - An additional layer normalization was added after the final self-attention block.
  - The vocabulary is expanded to 50,257.
  - increase the context size from 512 to 1024 tokens and a larger batch-size of 512 is used.
- **GPT-3 (Language Models are Few-Shot Learners - Brown et al. 2020)**
- Based on GPT-2
- Use alternating dense and locally banded sparse attention patterns in the layers of the transformer

Technical details

- Details for GPT3.5 and GPT4 are limited, but the following improvements are cited:
  - Improved text generation quality
  - Enhanced prompt engineering
  - Reduced model bias
  - Incremental performance improvements
  - More parameters
    (GPT4 rumoured~1.7 trillion parameters)

Applications - Q&A

- Learn, compare and verify answers for different academic subjects
- In mathematics, the performance depends on the dataset.
    - Best results for Grad Test, but poor for Olympiad Problem Solving
- In physics, it's found that ChatGPT can define and explain concepts in various styles, but not connecting these concepts together.
- West et al. used the Force Concept Inventory (FCI) to evaluate ChatGPT's accuracy in answering physics concept problems related to kinematics and Newtonian mechanics in the first semester of college physics
    - Average students score about 56%, Chat GPT is between 50 to 65%

Applications

- In medical field, an experiment on patient-doctor interactions revealed that trustworthiness of ChatGPT is relatively low, despite having similar answers to doctors.
- Causal discovery ability of ChatGPT in the diagnosis of neuropathic pain
  - ChatGPT has some limitations in understanding new knowledge and concepts beyond the existing textual training data corpus
  - Performance consistency and stability are not high (Different answers for the same questions)
- In other fields: ChatGPT for ordered importance semantic communication
  - Error rate and semantic loss of important words measured in the communication system embedded with ChatGPT are much lower
- Generating high-quality Boolean queries for systematic literature search
  - Better currently most advanced query generation methods but at the cost of reduced recall.

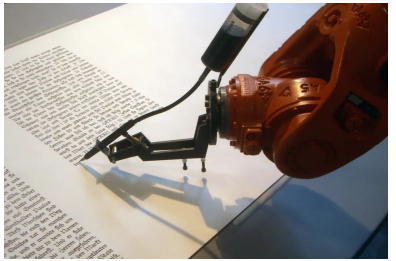
Applications: Text classification

- Assign text data to predefined categories
- Compare ChatGPT with X-GENRE classifier, using both English and Slovenian (SL) languages
- Performs better than X-GENRE in English dataset but worse in SL dataset
- ChatGPT for political stance detection: P-stance dataset with Trump, Biden and Bernie labels
    - F1-avg scores of 83.2, 82.0, and 79.4 respectively
- Disadvantages:
    - perform worse in classification tasks with rare or out-of-vocabulary words since it heavily relies on the distribution of training data.
    - Significant resource for training

Applications: Text generation

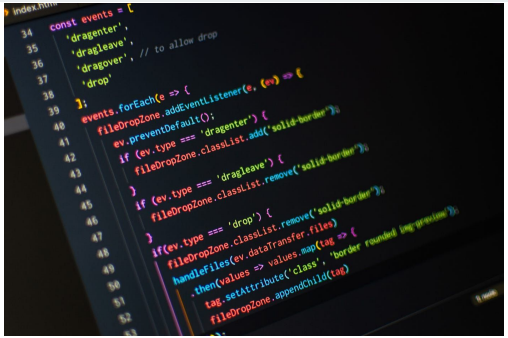
- Converting natural language commands into Bash commands.
  - Achieved high accuracy compared to other state-of-the-art methods
- Simplify complex text by: providing three fictional radiology reports to ChatGPT for simplification.
  - The reports are found to be mostly accurate and complete, but some errors and omission are found.
- Translation: Compared to three commercial translation products: Google Translate, DeepL Translate, and Tencent TranSmart.
  - ChatGPT competitive with other products in popular European languages but falls behind in distant languages.
- Abstract summarization: tested cross-lingual datasets. ChatGPT perform worse in several metrics

Applications: Text generation



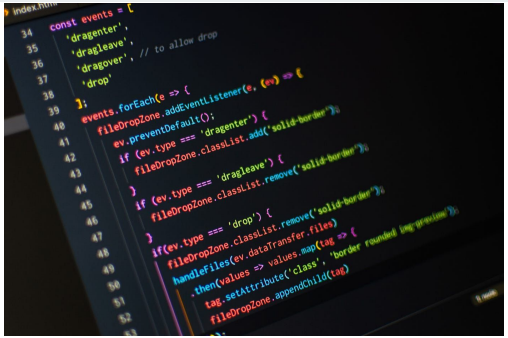
- Write commentary articles. Human authors rewrote the manuscript based on chatGPT's draft.
  - Quickly generate and optimize text, but not ideal in terms of generating new content.
- Verify originality of 50 papers written by ChatGPT, using Turnitin and iThenticate
  - ChatGPT has great potential in generating complex text output that is not easily captured by plagiarism detection software
- Using GPT models to create chatbots that pretend to be a human
  - Some participants were unable to distinguish between ChatGPT chatbot and real human

Applications: Code generation



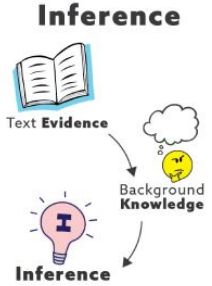
- ChatGPT's advanced natural language processing capabilities make it capable of performing code generation tasks. Saves time and effort
- Test ChatGPT for code explanation, suggesting alternative methods for problem-solving with code, and translating code between programming languages
    - Results generated by ChatGPt are found to be viable
- ChatGPT-based prototype called GPTCOMCARE:
    - generate multiple solutions for a programming problem and highlight the differences between each solution using colors
- ChatGPT for bug fixing. Success rate is improved by inputting more information through its dialogue system

Applications: Code generation



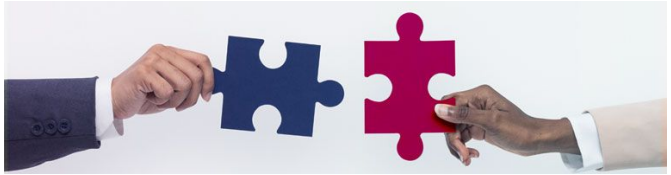
- ChatGPT as an experimental platform to investigate cybersecurity issues
  - Modelled five different modes of computer virus properties, including self-replication, self-modification, execution, evasion, and application,
  - Above average results, except for self-replication mode
- Disadvantages:
  - Biased towards Python, C++ or Java
  - Manual optimization is necessary for code formatting or performance optimization.
  - Quality of the generated code cannot be guaranteed

Applications: Inference



- Draw new conclusions or information through logical deduction from known facts or information
- Using ChatGPT to identify hatred in potential hate dataset.
  - correctly classify 80% of implicit hate tweets in the author's experimental setup
- Use ChatGPT to detect three different levels of language ambiguity and evaluated its performance
  - In semantics, ChatGPT performed perfectly in the detection of ambiguities
  - In the generation task, ChatGPT did well, but also revealed some of its worse issues: the lack of systematicity.

Applications: ChatGPT integration



- ChatGPT integration into the prototype of ”GPTCOMCARE” to address programming query problems
  - Demonstrated the effectiveness of using ChatGPT to improve the quality and diversity of code solutions
- ChatCAD: use chatGPT to enhance the output of multiple CAD networks for medical images
  - Comparison to CvT2DistilGPT2 and R2GenCMN showed significant advantages in RC and F1, while only performing weaker than R2GenCMN in PR
- Disadvantages:
  - Language barriers or differences in terminology between different systems
  - ChatGPT’s responses are not always deterministic.
  - The processing time of ChatGPT is slow for integration tasks involving time-sensitive data

AI Ethics

- Researchers used Wahl-O-Mat, to show ChatGPT political statements from different parties, forcing it to make choices of agree, disagree, or neutral.
    - The results indicated that ChatGPT has a pro-environment, left-wing liberal ideology
- Examined ChatGPT's moral standards by repeatedly asking it different versions of the trolley problem
    - ChatGPT gave answers with different moral orientations, lacking a firm moral stance
- ChatGPT found to be inconsistent in reasoning, factual errors, mathematics, coding, and bias across eleven related aspects.
    - People should be aware of their potential impact when seeking advice from ChatGPT
    - Fact check information given by ChatGPT

AI Ethics

- AI models, including ChatGPT, which are rapidly changing the way we communicate, explain, and create
    - Regulatory responsibility regarding LLMs and AI for the benefit of society
- There is a high likelihood that ChatGPT and similar products will be used for plagiarism
    - ChatGPT can easily generate persuasive paragraphs, chapters, and papers on any given topic -> plagiarism in many fields such as education, medicine, and law, and may be used for cheating in assignments and exams
    - Educators must be aware of the possibility of ChatGPT being used for exam cheating