УДК 51 (075.8)

HOW TO ADAPT GPT MODELS FOR EDUCATION

A. Perez Tchernov

Belarusian state university, Belarus, Minsk, alex.pereztchernov@gmail.com

The report presents a brief overview of the possibilities of using large language models in the educational environment. New methods for overcoming model limitations are listed, including retraining on specific tasks and the use of semantic indices to improve the factual accuracy of answers. The possibility of coordinating several models to perform specialized actions and inference is considered.

Keywords: GPT-4; AutoGPT; LLM; education transformation.

Large language models

For quite some time, machine learning methods have been successfully applied to specialized tasks. However, adapting existing models to new subject areas usually required significant manual efforts from developers, including defining the problem, preparing a large training dataset, defining a reward function, choosing a neural network architecture, and training the model. The creation of large language models relied on public text corpora, question-answer sets, vector representations of words and concepts, and the use of attention mechanisms in neural networks [1].

Initially, many researchers believed that further advances in artificial intelligence would require highly specialized models designed for narrow subject areas. However, OpenAI chose to retrain their large language model using dialogues with people [2-4]. As a result, the new GPT-3.5 and GPT-4 models have shown promising glimpses of what is now cautiously referred to as general purpose artificial intelligence [5].

Emerging Effects in Large Language Models

Large language models have recently undergone various tests to analyze their capabilities and limitations, revealing unexpected effects. These effects include the ability to perform inference based on learned principles, building an internal model of the world for complex hypothetical operations, identifying the questioner's intention to some extent, reflecting on previous information, and incorporating metaphors and communication strategies [6-10].

However, the model also has additional effects, such as implicit learning of biases and stereotypes from training data, generating non-existent data or links, and difficulty distinguishing between fact and fiction. Interestingly, increasing the number of model parameters improves their capabilities but also increases the occurrence of "human-like" errors [9].

Limitations of language models and solutions

To effectively use language models, it is important to overcome their limitations. One of the main limitations is the size of input data that the dialog model can work with. The model's short-term memory is limited to a small number of tokens, so it may "forget" the initial context of the task if there is too much material or if the user deviates from the initial request. To solve this issue, one approach is to use techniques such as Map-Reduce, Map-Rerank, and Refine offered by the LangChain project. This approach involves dividing the input document into parts, which are fed into the model to obtain a summary and identify the main relationships in the context of the task. The obtained extracts are then combined into a common reduced document, which can be periodically sent along with a custom message to the input of the model to prevent loss of context. Another way to address this issue is to precompress and transform the input data using a simpler model or other techniques, which can lead to savings in retained context size or processing cost [11].

One possible way to address the issue of limited short-term memory in language models is to retrain the model on selected documents of a new subject area, which enables the dialog activity to be based on "long-term" memory instead of relying on a relatively small dialog context. However, this approach has been found to work well for transferring style and skills, but it does not retain factual accuracy as well [4].

To overcome the limitations of language models and improve accuracy, there are several alternative approaches. One solution is to use separate databases and search indexes to find semantically similar materials that can be used to supplement the information contained in the model's long-term memory. The LlamaIndex project provides an example of this approach [18].

Another approach involves using multiple language models and an external system to store short-term memory during task execution. This method allows for task planning, model selection, and the integration of external tools to complete partial tasks and improve the quality of the final results. Both AutoGPT and LangChain employ this approach [21].

Pluggable tools for computation [12], ontology-based inference [14, 15, 19, 20], and complex model compositions [16, 17] have demonstrated improvements in the cognitive capabilities of language models.

Using Language Models for Educational Needs

Developing models that can interact with users through language interfaces with contextual understanding is a significant milestone in the advancement of artificial intelligence. The potential impact of such tools could revolutionize all aspects of human activities. An intriguing feature of such large language models is their ability to take on specific roles, such as a teacher or friend, maintain a particular style, consider the listener's knowledge level, and build dialogues that align with the communication goal.

For innovation to replace existing processes or products, it must demonstrate superiority in terms of efficiency, relevance to survival in a competitive or environmental niche, ease of implementation, and end-users' personal interest in adopting the technology for enhancing local advantages. The widespread adoption of artificial intelligence and robotics puts pressure on professions with readily accessible training datasets, increasing the demand for retraining and professional courses, seminars, and intensives. Thus, the broad implementation of large language models could increase the operational efficiency of systems for additional education and retraining, providing an immediate demand for the technology.

We propose identifying the following key layers of work in educational organizations and their accompanying adaptation of language models:

Discovery of new knowledge, selection of an educational niche, and formation of educational business services. Recommender algorithms can aid in quickly receiving updates on new educational topics [13] and identifying demand.

Creation of methodology and metrics for educational services. Public libraries of configuration hints (prompts), libraries of pedagogical examples, and predefined roles can be created and utilized at this layer.

Development of educational materials. Well-structured libraries of concepts, educational examples, topic dependency graphs [13], and sets of typical questions and answers on the concepts being studied are important for this layer.

Delivery of materials and organization of the educational context. Tools and external plugins can be prepared to measure progress, predict reward cycles, recognize emotions from dialogues, and recreate the cognitive model of listeners.

Leadership, support, and integration of the organization into the communication and recruitment environment. Simple tools for processing and generating texts may be sufficient to reduce transaction costs for organizing work with partners, students, and graduates.

Conclusion

In conclusion, we summarize the main adaptations of language models for the educational environment. It is recommended to use task-specific language models and incorporate semantic indexes with subject-specific information to improve factual accuracy. Coordinating multiple models and plugins can automate specialized actions, calculations, and inference. Public libraries of prompts and roles should be shared between different members. Lastly, unifying the API for the interaction of external systems with the proposed models can simplify the process.

Bibliographic references

1. *Long Ouyang* et al. Training language models to follow instructions with human feedback, 2022, arXiv:2203.02155, [Online]. URL: https://arxiv.org/pdf/2203.02155.pdf
2. *Baolin Peng* et al. Instruction Tuning with GPT-4, 2023, arXiv:2304.03277, [Online]. URL: https://arxiv.org/abs/2304.03277
3. *Philipp Koralus* et al. Humans in Humans Out: On GPT Converging Toward Common Sense in both Success and Failure, 2023, arXiv:2303.17276, [Online]. URL: https://arxiv.org/abs/2303.17276
4. Long Ouyang et al. Training language models to follow instructions with human feedback, 2023, arXiv:2203.02155, [Online]. URL: https://arxiv.org/abs/2203.02155
5. *Sébastien Bubeck* et al. Sparks of Artificial General Intelligence: Early experiments with GPT-4, 2023, arXiv:2303.12712, [Online]. URL: https://arxiv.org/abs/2303.12712
6. *Jason Wei.* Emergent Abilities of Large Language Models, 2023, arXiv:2206.07682, [Online]. URL: https://arxiv.org/abs/2206.07682
7. *Binz M.* Using cognitive psychology to understand GPT-3, 2023, arXiv:2206.14576, [Online]. URL: https://arxiv.org/abs/2206.14576
8. *Alexander Pan* et al. Do the Rewards Justify the Means? Measuring Trade-Offs Between Rewards and Ethical Behavior in the MACHIAVELLI Benchmark, 2023, arXiv:2304.03279, [Online]. URL: https://arxiv.org/abs/2304.03279
9. *Samuel R. Bowman.* Eight Things to Know about Large Language Models, 2023, arXiv:2304.00612v1, [Online]. URL: https://arxiv.org/abs/2304.00612v1
10. *Yiheng Liu.* Summary of ChatGPT/GPT-4 Research and Perspective Towards the Future of Large Language Models, 2023, arXiv:2304.01852, [Online]. URL: https://arxiv.org/pdf/2304.01852.pdf
11. *David Wingate.* Prompt Compression and Contrastive Conditioning for Controllability and Toxicity Reduction in Language Models, 2023, arXiv:2210.03162, [Online]. URL: https://arxiv.org/pdf/2210.03162.pdf
12. *Cristina Cornelio.* Combining data and theory for derivable scientific discovery with AI-Descartes, 2023, Nature, [Online]. URL: https://www.nature.com/articles/s41467-023-37236-y.
13. *Pavlos Stavrinides* et al. Course-Prerequisite Networks for Analyzing and Understanding Academic Curricula, 2023, arXiv:2210.01269, [Online]. URL: https://arxiv.org/abs/2210.01269

14. *Colin G.* West et al. Advances in apparent conceptual physics reasoning in GPT-4, 2023, arXiv:2303.17012. [Online]. URL: https://arxiv.org/pdf/2303.17012.pdf

15. *Caufield JH* et al. Structured prompt interrogation and recursive extraction of semantics (SPIRES): A method for populating knowledge bases using zero-shot learning, 2023, arXiv:2304.02711. [Online]. URL: http://arxiv.org/abs/2304.02711

16. *Denny Zhou* et al. Least-to-Most Prompting Enables Complex Reasoning in Large Language Models, 2023, arXiv:2205.10625. [Online]. URL: https://arxiv.org/abs/2205.10625

17. *Kojima T.* et al. Large Language Models are Zero-Shot Reasoners, 2022, arXiv:2205.11916. [Online]. URL: https://arxiv.org/abs/2205.11916

18. *Jerry Liu.* LlamaIndex (GPT Index) [Computer software]. URL: https://github.com/jerryjliu/llama_index. Date of access: 30.03.2023.

19. *Max Adamski* et al. Reason-able Embeddings: Learning Concept Embeddings with a Transferable Neural Reasoner [Computer software]. URL: https://github.com/maxadamski/reasonable-embeddings. Date of access: 30.03.2023.

20. DICE Group. Ontolearn: Learning description logics with evolutionary algorithms [Computer software]. URL: https://github.com/dice-group/Ontolearn. Date of access: 30.03.2023.

21. Significant-Gravitas. Auto-GPT: An Autonomous GPT-4 Experiment [Computer software]. URL: https://github.com/Significant-Gravitas/Auto-GPT. Date of access: 30.03.2023.