Use of GPT models in natural language processing

**Paul Ledala**

**Abstract:**

Language models like GPT-4 are built on machine learning algorithms that analyze vast amounts of text data and use that knowledge to generate human-like responses to various inputs. With more advanced models, we can expect better performance in various NLP tasks such as language translation, text summarization, question answering, and more.

Furthermore, GPT-4 may also contribute to the development of more sophisticated conversational AI agents that can interact with humans in a more natural and intuitive way. This could have significant implications for industries such as customer service, where chatbots and virtual assistants are already being used to automate customer interactions.

GPT-4 can accept a prompt of text and images, which—parallel to the text-only setting—lets the user specify any vision or language task. Specifically, it generates text outputs (natural language, code, etc.) given inputs consisting of interspersed text and images. Over a range of domains—including documents with text and photographs, diagrams, or screenshots—GPT-4 exhibits similar capabilities as it does on text-only inputs. Furthermore, it can be augmented with test-time techniques that were developed for text-only language models, including few-shot and chain-of-thought prompting.

**General NLP Tasks**

1. **FILL in the Middle- Uses, and applications in NLP:**

The paper written by Mohammed Bavarian et al[1] describes a data augmentation technique that involves moving a span of text from the middle of a document to its end. The authors show that autoregressive language models can learn to infill the missing text with high accuracy without affecting the original left-to-right generative capability of the model. They provide evidence that training models with a large fraction of data transformed in this way does not harm model performance, as measured by perplexity and sampling evaluations across a wide range of scales.

The authors suggest that future autoregressive language models be trained with this fill-in-the-middle (FIM) technique by default, as it is useful, simple, and efficient. They perform experiments to identify optimal hyperparameters, such as the frequency of data transformation, the structure of the transformation, and the method of selecting the infill span, and provide best practices for training FIM models.

The FIM technique has potential applications in various NLP tasks, such as language modeling, text completion, and text generation. By training models to infill missing text, the technique can improve the quality of generated text and enhance model performance. The authors have released their best infilling model trained with best practices in their API and released their infilling benchmarks to aid future research. This technique has the potential to improve the performance of various NLP applications and can be used to train more accurate and efficient language models.

1. **Training with little/no data and Evolution through large models:**

The paper written by Joel Lehman et al[2] explores the potential benefits of using large language models (LLMs) trained to generate code in genetic programming (GP). The paper suggests that LLMs can significantly improve the effectiveness of mutation operators applied to programs in GP, as they can approximate likely changes that humans would make.

The paper describes an experiment in which evolution through large models (ELM) combined with MAP-Elites is used to generate hundreds of thousands of functional examples of Python programs that output working ambulating robots in the Sodarace domain, which the original LLM had never seen in pre-training. These examples are then used to train a new conditional language model that can output the right walker for a particular terrain.

The paper highlights the implications of using ELM for open-endedness, deep learning, and reinforcement learning. The ability to bootstrap new models that can output appropriate artifacts for a given context in a domain where zero training data was previously available carries significant implications for various NLP applications.

Overall, the paper suggests that using LLMs to generate code can have numerous potential benefits in GP and other NLP applications, particularly in situations where training data is limited or non-existent. The paper also highlights the need for further research to explore the implications of ELM for deep learning and reinforcement learning.

1. **Topic-Based Summarization – potential and limitations in NLP:**

The GPT models are "critique-writing" models that are trained to detect flaws in summaries. The implication of these models is that they can be used to assist human supervision of AI systems on difficult tasks. For example, in situations where it is difficult for humans to evaluate the quality of a summary, such as when the task is complex or time-consuming, these models can be used to provide meaningful feedback. This could be useful in a variety of settings, including natural language processing, machine learning, and artificial intelligence research.

However, it is important to note that the study written in the paper ‘Self-critiquing models for assisting human evaluators’ written by William Saunders et al[3] highlights an important limitation of the current models. The task of topic-based summarization is not actually a difficult task for humans to evaluate, which means that the true potential of AI-assisted evaluation is not yet known. Therefore, more research needs to be done to understand the limits of AI-assisted evaluation and to explore the use of these models in more complex tasks.

1. **Expressing uncertainty in dialogue agents using words:**

The study written by [Stephanie Lin](https://arxiv.org/search/cs?searchtype=author&query=Lin%2C+S) et al[4] shows that GPT-3 can express calibrated uncertainty about its own answers in natural language without the use of model logits. The model generates both an answer and a level of confidence, which map to well-calibrated probabilities. The model is moderately calibrated under distribution shift and sensitive to uncertainty in its own answers, rather than imitating human examples. The study also introduces the CalibratedMath suite of tasks for testing calibration and compares uncertainty expressed in words to uncertainty extracted from model logits. The study provides evidence that GPT-3's ability to generalize calibration depends on pre-trained latent representations that correlate with epistemic uncertainty over its answers. This work could have implications in various NLP applications, such as question-answering systems, chatbots, and dialogue agents, where expressing calibrated uncertainty could improve the user experience and trust in the system.

1. **Misuse of Language models for disinformation, fraud detection & technical misuse:**

The summary given by [Josh A. Goldstein](https://arxiv.org/search/cs?searchtype=author&query=Goldstein%2C+J+A) et al[5] does not mention any specific GPT model but refers to generative language models that can produce realistic text outputs. In this case, the NLP use of GPT-3 would be its potential use in influence operations, where it could be used to automate the creation of convincing and misleading text for malicious purposes. The report assesses how language models, such as GPT-3, could change and influence operations in the future and proposes a framework for mitigating this threat.

The report suggests that mitigations could target various stages of the language model-to-influence operations pipeline, including model construction, model access, content dissemination, and belief formation. While no single mitigation can fully prevent the threat of AI-enabled influence operations, a combination of multiple mitigations may make an important difference in reducing the impact of this threat.

Another part of misuse is technical. Codex is a large language model trained on codebases that can generate and synthesize code more effectively than previous models. Although Codex offers many benefits, there are potential limitations and hazards associated with deploying such models, including alignment problems, misuse, and the potential to destabilize technical fields.

To explore these hazards, there is a hazard analysis framework developed by OpenAI that considers the technical, social, political, and economic impacts of Codex and similar models. The analysis is informed by an evaluation framework that assesses the capacity of advanced code generation techniques relative to human ability and the complexity of specification prompts.

The potential applications of Codex and similar models in NLP are numerous, particularly in the development of automated coding tools and systems. By generating and synthesizing code more effectively, these models have the potential to accelerate the development of software and other technical applications. However as set out in the ‘Lessons learned on language model safety and misuse’ by Miles Brundage et al[6], there are important concerns about the potential risks associated with deploying these models which highlights the need for ongoing evaluation and mitigation of these risks.

**References:**

1. Bavarian, M. *et al.* (2022) *Efficient training of language models to fill in the Middle*, *arXiv.org*. Available at: <https://arxiv.org/abs/2207.14255> (Accessed: 15 May 2023).
2. Lehman, J. *et al.* (2022) *Evolution through large models*, *arXiv.org*. Available at: <https://arxiv.org/abs/2206.08896> (Accessed: 15 May 2023).
3. Saunders, W. *et al.* (2022) *Self-critiquing models for assisting human evaluators*, *arXiv.org*. Available at: <https://arxiv.org/abs/2206.05802> (Accessed: 15 May 2023).
4. Lin, S., Hilton, J. and Evans, O. (2022) *Teaching models to express their uncertainty in words*, *arXiv.org*. Available at: <https://arxiv.org/abs/2205.14334> (Accessed: 15 May 2023).
5. Goldstein, J.A. *et al.* (2023) *Generative language models and automated influence operations: Emerging threats and potential mitigations*, *arXiv.org*. Available at: <https://arxiv.org/abs/2301.04246> (Accessed: 15 May 2023).
6. Miles B. *et al* (2022) *Lessons learned on language model safety and misuse.* Available at: <https://openai.com/research/language-model-safety-and-misuse> (Accessed: 15 May 2023).