Research Questions For NLP Research Night 2023

1. Is RLHF the best approach to align language models to people's needs?
2. Should language models be personalized for individual uses/companies? Or will one language model be capable of addressing the needs of all people/companies?
3. How can we distinguish between model-generated and human-generated content? Is this even possible?
4. How many parameters does a language model need? How many does GPT4 have?
5. What are some problems that ChatGPT/GPT4 still cannot solve?
6. How should conversational agents interact with tools (such as web-search, Wolfram Alpha, or an interpreter for Python code)?
7. How can we ensure safety when using large language models?
8. Is there a way to enable LLM technology on edge devices?
9. What are some methods we can use to tackle the hallucination problem?
10. How can we enable LLMs to cite their sources (data attribution)?
11. How effective are language models in understanding human emotion? Do these models possess Theory of Mind?
12. Will we always depend on expensive human annotations for the alignment process? Or will this eventually be automated by LLMs as well?
13. Is it fair to evaluate the capabilities of language models on human tasks (e.g., SAT, Bar Exam, GRE, MCAT)?
14. How can we evaluate the capabilities of language models/compare the performance of models?
15. Can a language model be sufficiently powerful without access to tools?
16. How did OpenAI enable GPT4's tremendous context window (over 50 pages of text relative to the standard 512 token sequence length)?
17. Should the architecture of language models change? Or should it remain a simple transformer decoder model that simply predicts the next word?
18. How did OpenAI enable multi-modality in GPT4?
19. When will language models enable processing videos as well?
20. What is the role of academia in the age of ChatGPT and its rivals?
21. What is the formal definition of a spurious correlation? How do we prevent LMs from learning them?
22. How long will prompt engineering as a research field be relevant for?
23. Should proprietary LMs (I.e., unknown architecture, training data/paradigm) be considered as baselines in the peer reviewed publication process? And if so, how can we make sure there's no data leakage of the evaluation data? Moreover, should the research community allow for such models that redefine the term "black box models"?
24. Flash attention has massively improved the efficiency of training transformers. What might be the next bottlenecks? How can we bypass them?

25. What are some policies which governments should enforce in order to establish a safer climate for AI research?
26. Does the open source community stand a chance against tech giants such as Google or OpenAI?
27.