

# Evaluating and Analyzing the Robustness of Large Language Models to Character-Level Permutations

東京大学 大学院工学系研究科 Qi Cao

目的 To investigate the performance of large language models (LLMs) under character-level permutations.

内容 Develop a benchmark for evaluations and perform a holistic analysis using multiple LLMs.

結果 The most powerful LLMs are capable of processing scrambled text to varying degrees. At the time of the experiments, only GPT-4 could still perform well under extreme permutations.

利用した計算機	SQUID GPUノード群
	ノード時間
	2,000 時間程度
	使用メモリ
	不定
	並列化
	1ノード

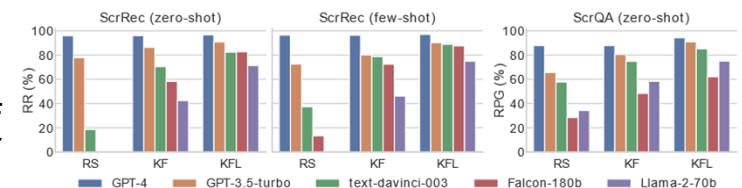


Fig: Partial results on the proposed benchmark