Mechanism Design for Large Language Models

(Joint with Vahab Mirrokni, Renato Paes Leme, Song Zuo, and Haifeng Xu)

Abstract:

We investigate auction mechanisms for AI-generated content, focusing on applications like ad creative generation.

In our model, agents' preferences over stochastically generated content are encoded as large language models (LLMs).

We propose an auction format that operates on a token-by-token basis, and allows LLM agents to influence content creation through single dimensional bids.

We formulate two desirable incentive properties and prove their equivalence to a monotonicity condition on output aggregation.

This equivalence enables a second-price rule design, even absent explicit agent valuation functions.

Our design is supported by demonstrations on a publicly available LLM.