

Principais contribuições do artigo “Language models are unsupervised multitask learners”, de  
Radford et al

Aluno: Leandro Carísio Fernandes

- Abordagem tradicional era de sistemas para resolver tarefas específicas. Assim, trabalhava-se com bases de treinamentos específicas para cada atividade.
- Os autores suspeitavam que o treinamento em uma única tarefa ou o uso de bases que cobrem um único domínio eram o principal fator que impossibilitava que os sistemas não generalizassem bem para outras tarefas.
- A hipótese dos autores é a de que modelos de linguagem de grande capacidade começará a inferir e executar tarefas definidas em linguagem natural
- O dataset de treinamento foi feito a partir de um scraping a partir do Reddit, gerando uma base (WebText) com ~40GB de texto. Não usaram o Common Crawl pois nele há muitos dados ininteligíveis. Ao fazer o scraping a partir do Reddit, é como se houvesse uma curadoria humana indireta dos dados que estão raspando.
- Propõe e treinam o GPT-2. Não focam em treinar para resolver problemas específicos.
- Abordam questões de overlap na base de treino e teste.
- Abordam a questão da generalização e memorização. Mesmo aumentando o número de parâmetros, a perplexidade de treino e teste continuou baixando.