

Aula 10/11 - Leitura - WebGPT

2 - A contribuição do artigo

1. “Criação de um ambiente de navegação na Web baseado em texto com o qual um modelo de linguagem ajustado pode interagir. Isso permite melhorar a recuperação e a síntese de maneira ponta a ponta, usando métodos gerais, como aprendizado por imitação e aprendizado por reforço.”
2. O web-GPT pode pesquisar documentos na web e anexar os principais top-k resultados ao prompt, de modo que possa atender a esse contexto para responder a perguntas, o que melhora a precisão factual do modelo quando comparado aos LMs regulares.

1 - Explicação de conceitos importantes do artigo

- O método de treinamento é da seguinte maneira:
 - **Behavior cloning (BC):** Um fine-tuning de forma supervisionada é feito usando as demonstrações com comandos emitidos pelos humanos como rótulos.
 - **Reward modeling (RM):** a partir da etapa anterior, remove a última camada não incorporada, usa pergunta/resposta/referência como entrada.
 - **Reinforcement learning (RL):** Usa o algoritmo PPO realizando um fine-tuning no modelo BC.
 - **Rejection sampling (best-of-n):** faz uma amostra de um número fixo de respostas do modelo BC ou RL e escolhe aquela com a maior pontuação de RM.
 -

- O uso de um mecanismo de pesquisa pode aumentar significativamente a autenticidade e a qualidade das respostas.
- O feedback humano, ou comparação, é importante para melhorar os resultados do modelo.
-