

Relatório 3 - Vídeo: Prompt Engineering (I)

Vitor Eduardo de Lima Kenor

Descrição da atividade

Neste card é disponibilizado um vídeo sobre engenharia de prompt, onde nos ensinará a extrair o máximo possível dos modelos de linguagem. O vídeo é dividido em sete módulos, sendo o primeiro módulo o de fundamentos.

Neste módulo começamos com uma definição para engenharia de prompt, que seria a ciência empírica de planejar criar e testar prompts para gerar melhores respostas em grandes modelos de linguagem (GLM/LLM). Uma das primeiras coisas importantes abordadas neste módulo é a regra de ouro, que se trata de que instruções você daria para alguém fazer, como você falaria uma coisa que você faz para outra pessoa para que ela consiga fazer o mesmo.

Passando para o módulo dois, que trata da estrutura de um prompt, basicamente ele ensina usando as letras da palavra P.R.O.M.P.T. dando um significado estrutural para cada uma. No vídeo junto com o significado de cada letra ele já nos dá um exemplo. Os exemplos são esses:

1. **Persona:** Você é um especialista em marketing, redes sociais e psicologia humana.
2. **Roteiro:** Quero ajuda para criar um roteiro para um reels no instagram .
3. **Objetivo:** O objetivo é chamar a atenção com uma história cativante e depois fazer uma chamada para se inscrever em nosso site no link do perfil.
4. **Modelo:** Quero o resultado dividido em 8 até 10 slides com sugestões de imagens e design para cada um deles.
5. **Panorama:** Meu cliente é um expert que vende cursos online e está com dificuldade em escalar seu curso para além dos 4 e 5 dígitos (inclua exemplos aqui)
6. **Transformar :** Feito isso, agora você melhora. Troca X, faz Y, adiciona tal.

O próximo módulo é sobre o processo, o que fazer quando um prompt simples não resolve, qual processo eu devo seguir para criar prompts mais robustos. No vídeo ele ensina levantando tópicos e explicando cada um deles. O primeiro é definir as tarefas e os critérios de sucesso, quão bem o modelo precisa executar a tarefa?, qual é o tempo de resposta aceitável para o modelo?, qual é o seu orçamento para executar o modelo?. O segundo tópico fala sobre desenvolver os casos de teste, o terceiro é sobre escrever o prompt inicial, o quarto é testá-lo contra exemplos, o quinto é refinar o prompt e o sexto e último é colocar em produção.

O módulo seguinte é de técnicas básicas, a primeira é a utilização de markdown para uma boa organização nos textos, pois um texto bem organizado para nós para máquina também serve. No vídeo ele até cita que um prompt bem formatado retorna uma resposta diferente. Uma dica abordada nesta parte, é que quando estiver lidando com documentos longos é recomendado usar tags XML e posicionar antes das instruções. A segunda técnica abordada é entender o que é o prompt do sistema, que basicamente é a parte que você dirá como deseja que ele atue. A terceira técnica é o zero shot que se trata em passar uma pergunta sem nem uma referência ou exemplos. A quarta técnica apresentada é o estímulo de prompt direcional, baseado em um artigo científico o vídeo mostra essa técnica, onde podemos ver a diferença quando direcionamos o foco do agente principalmente quando ele tem que analisar grandes documentos. A quinta técnica é o few shot, apresentando resultados significativamente superiores a zero shot onde não adicionamos exemplos, pois quando colocamos exemplos de sucesso para o agente estamos direcionando ele para o caminho que queremos que ele vá. As

duas próximas técnicas apresentadas se baseiam em cadeia de pensamento, onde a partir de passo a passo conseguimos chegar em uma solução mais precisa, alguns modelos fazem isso por conta própria e outros devem ser guiados e expostos a exemplos para que façam.

Partindo para técnicas mais avançadas para prompts onde a tarefa exige muito raciocínio por parte do modelo, no vídeo é ressaltado que essas técnicas avançadas não são usadas normalmente, sua aplicação ou não deve ser estudada para seu caso. A primeira é a consistência própria, onde o modelo gere algumas de suas cadeias de pensamentos para um problema e ele mesmo julgue qual se saiu melhor e obteve bons resultados. A segunda é a Árvore de pensamento, similar a primeira com a diferença que cada cadeia de pensamentos leva a outras e assim criando várias soluções interligadas. A terceira é o esqueleto de pensamento, onde instruímos ao modelo que queremos que ele mande todas as possibilidades que existem para resolver uma situação, e a partir de um feedback humano sobre as possibilidades ir construindo a estrutura do que se deseja desenvolver. A quarta é a geração de conhecimento, que se trata do modelo criar conteúdo sintético que será usado por ele mesmo posteriormente. A quinta é o prompt maiêutico, que é similar ao nosso processo de justificar nossa resposta. A sexta é a geração aumentada de recuperação (RAG), é um processo que manda vários documentos e informações que não existem no treinamento de um modelo para que ele enriqueça seu conhecimento em determinadas áreas e tenha um desempenho melhor. A sétima é a linguagem programática assistida, essa técnica trata-se de definir variáveis para que o modelo entenda melhor o que irá modificar. A oitava é o ReAct, onde o modelo questiona o que cada uma das ações implica para decidir os próximos passos.

Saindo das técnicas e voltando para os módulos, é apresentado a forma para evitar que o modelo tenha alucinações, para isso deve-se deixar bem claro nas instruções fornecidas ao modelo para que ele possa dizer quando não souber a resposta. Outra coisa que tem que ser levada em conta é a temperatura do modelo, pois se for muito alta o modelo será mais criativo e terá chances maiores de inventar informações.


Depois de passar todas as técnicas e processos para se ter um prompt eficiente, é nos mostrado um exemplo prático de como foi aplicado alguns dos conhecimentos já vistos acima. Logo em seguida, podemos ver como é o prompt de uma gpt e como as técnicas foram aplicadas em seu prompt para o modelo agir do modo que conhecemos.

Conclusões

Este vídeo sobre engenharia de prompt nos ensinou como criar instruções claras e eficientes para obter as melhores respostas dos modelos de linguagem. Ao longo dos módulos, aprendemos que um bom prompt envolve planejar a tarefa, definir objetivos e estruturar as instruções de forma clara. Técnicas como o uso de exemplos (few-shot) e a formatação correta ajudam a melhorar a precisão das respostas. Além disso, vimos como evitar erros comuns, como alucinações, e como adaptar as instruções conforme o caso. Em resumo, a engenharia de prompt é fundamental para otimizar o desempenho dos modelos e alcançar resultados mais eficazes.

Referências

Vídeo de Prompt Engineering :

 Engenharia de Prompt: O Guia Definitivo