



Nome: Fernando Buligon Antunes

Data: 25/03/2025

O vídeo começa com o apresentador fazendo uma breve introdução, prometendo ensinar a qualquer um como fazer 100% de aproveitamento de um prompt usando sete módulos:

- Fundamentos: Vai ser visto a parte básica e como deve ser feito o trabalho em documentos grandes;
- Prompt: Estrutura básica de como montar uma entrada de maneira apropriada;
- Processo: Processo recomendável, abordando pensamentos como “devo fazer isso?”
- Técnicas básicas: Maneira de fazer com que os prompts tenham uma saída melhor;
- Técnicas avançadas: Fazer prompts com intuitos mais complexos;
- Exemplos: Vai mostrar alguns exemplos práticos;
- Ferramentas: Ferramentas recomendadas que auxiliam no processo.

Fundamentos, começa definindo engenharia de prompt como “a ciência empírica de planejar, criar e testar prompts para gerar melhores respostas em Grandes Modelos de Linguagem (GML/LLM)”, depois apresenta a regra de ouro “que instruções você daria para alguém fazer” que basicamente consiste em deixar extremamente claro o que você quer, por exemplo o prompt “faça um ovo”, de qual jeito quer o ovo? cozido? frito? ou quer do ponto de vista de uma galinha? O prompt não precisa ser complicado, apenas direto, contendo exatamente o que você quer do jeito que você quer.

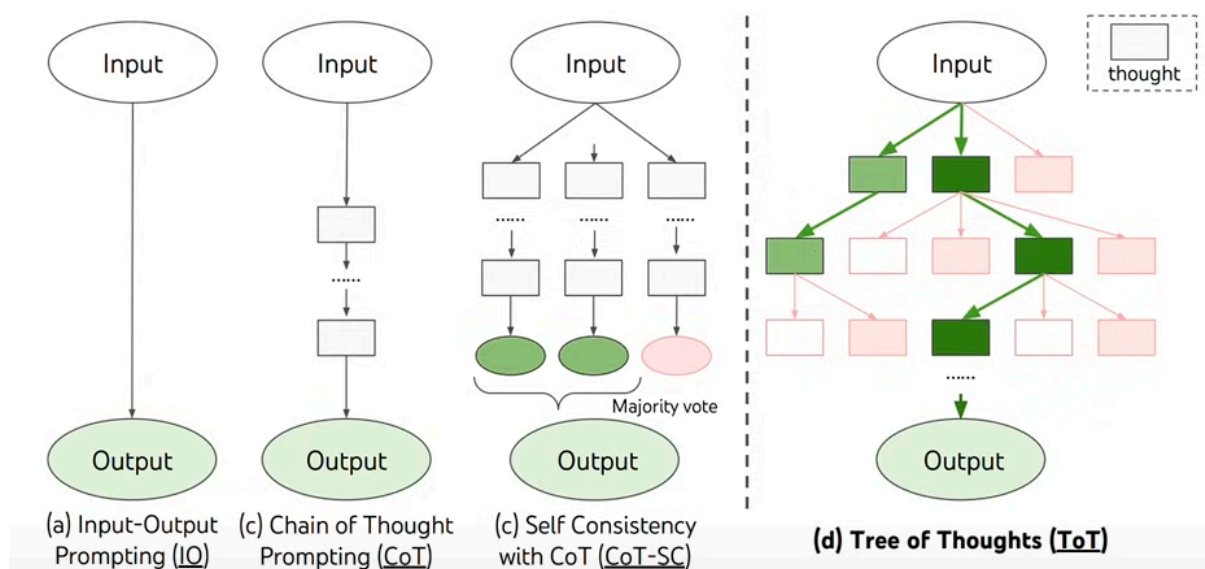
Prompt, para ajudar na estrutura do prompt, o apresentar representa cada letra da palavra prompt como algo que faz parte do processo, “P” para persona, sempre se deve colocar o papel de quem que a máquina deve ocupar, por exemplo, caso queira uma receita, diga para que responda como um cozinheiro, “R” para roteiro, importante informar o contexto do que você quer, seguindo no exemplo da cozinha, você quer fazer uma receita para o almoço de domingo junto com a sua família, “O” objetivo, falar para que você quer, ainda no exemplo da cozinha, você quer que todo mundo ache sua comida gostosa, “M” modelo, maneira na qual você quer a resposta, exemplo, quero passo a passo dividido por tópicos e bem organizado, “P” panorama, falar quem é o público alvo, minha família é composta em sua maioria por pessoas entre 30 a 50 anos, “T” transformar, não deve se esperar que a primeira resposta seja perfeita, então sempre deve ir testando e trocando pontos conforme a resposta anterior até atingir o que realmente queria.

Processo, caso apenas a técnica anterior não tenha funcionado, existem alguns passos que podem ajudar. 1º, definir a tarefa e os critérios de sucesso, você precisa saber onde está o gol antes de chutar. 2º, definir casos de testes, saber qual caminho você deve seguir para chegar no resultado que quer, ir analisando se a

trilha realmente está levando para lá. 3º, escrever o prompt inicial, depois de ter pensado sobre o que quer, só fazer o prompt seguindo a estrutura passada, 4º, testar contra exemplos, podendo ser julgados por nós mesmos ou por outros modelos de LLM. 5º, refinar os prompts, analisar as respostas e ir ajustando conforme precisar. 6º, colocar em produção.

Técnicas básicas, fazer um texto bem organizado, tendo uma organização limpa fica mais fácil para você e para a máquina entenderem, o apresentador traz o exemplo do markdown, que permite formatar, outra dica é usar tags xml para identificar partes do texto quando ele for muito longo. Zero shot, termo usado quando o prompt não possui nenhum exemplo, temos que evitar cometer esse caso pois quanto mais exemplos, melhor para o modelo interpretar. Cadeia de pensamento, alterar o prompt de maneira com que a máquina consiga simular um processo de raciocínio ao invés da resposta em si diretamente. Cadeia de pensamento contrastiva, além de passar o modelo da resposta correta, também se deve passar o modelo de uma resposta errada. Cadeia de pensamento zero shot, ao invés de passar exemplos, apenas coloque para a máquina pensar passo a passo.

Técnicas avançadas. Consistência própria, sugerem que ao invés de gerar apenas uma cadeia de pensamentos, devem ser geradas três, para que o próprio modelo compare as três e chegue em um conclusão sobre qual tem mais probabilidade de chegar em um resposta melhor. Árvore de pensamento, desenvolver uma árvore de conhecimento, onde cada alternativa vai sendo estudada para que o modelo consiga ver qual é o melhor caminho. Na imagem abaixo é possível ter uma ideia melhor.



Esqueleto de pensamento, consiste em um processo de raciocínio de como se deve tomar boas decisões, colocando várias ideias à disposição, depois analisá-las com uma breve explicação, e após isso ir para os próximos passos.



Geração de conhecimento, produzir um conhecimento para que a IA use como base para as próximas respostas, por exemplo, input: pinguins tem <mask> asas, conhecimento> aves possuem duas asas, pinguim é uma ave. Prompt Maiêutico, consiste em pedir para o modelo justificar sua resposta, fazendo isso, talvez ele mesmo consiga perceber caso erre em algum ponto. Geração Argumentada de Recuperação (RAG), basicamente faz com que o modelo use uma base externa de conhecimento para gerar a resposta, sabendo que a resposta está presente nessa base, ela consegue encontrar uma resposta mais técnica e menos geral. Linguagem programática assistida, faz o uso de variáveis no meio do texto. React, dividem o prompt em pensamento 1, ação 1, e observação 1, podendo conter quantos pensamentos, ações e observações quiser, fazendo assim fica mais fácil para o modelo do que apenas colocar uma pergunta. Alucinações, às vezes pode acontecer da IA acabar inventando coisas para a resposta, uma maneira de se evitar isso é informar como você quer ser respondido, por exemplo, pedir fontes, avisos caso não tenha tanta certeza sobre o que está falando e etc.

Exemplos, ele mostra um exemplo de prompt contendo todos os assuntos comentados anteriormente, focando principalmente em como estruturou o prompt, foi um prompt relativamente grande para a época, em que o limite de caracteres era de 8000 no chat GPT então foi preciso criar um arquivo auxiliar em formato json. Outro exemplo foi o GPT Grimoire, nesse caso ele só passa meio por cima falando a ideia geral do prompt e mostra onde está disponível.

Ferramentas, o criador do vídeo criou um agente pré programado com todo conteúdo visto, então basta informar ao agente o que queremos com um prompt bem básico que o agente vai aplicar as técnicas e retornar um prompt completo. Está disponível [aqui](#).