

---

## **Engenharia de Prompts**

Asimov Academy

# ASIMOV

## Conteúdo

<b>01. Introdução ao Curso</b>	<b>3</b>
<b>02. Conceitos Gerais de LLMs</b>	<b>4</b>
O que é uma LLM? . . . . .	4
Como as LLMs são treinadas? . . . . .	4
Diferenças entre uma LLM Base e uma LLM treinada para instruções . . . . .	4
Os principais modelos no mercado . . . . .	5
Quais parâmetros comuns de um modelo de LLM . . . . .	6
Como abordaremos este curso . . . . .	7
<b>03. Os princípios de um bom prompt</b>	<b>8</b>
O que é um prompt? . . . . .	8
1º Princípio - Escrever instruções claras e específicas . . . . .	8
2º Princípio - Dar ao modelo tempo para pensar . . . . .	9
Limitações de um modelo . . . . .	9
Alucinações . . . . .	9
Vieses . . . . .	10
Prompting é sobre iteratividade . . . . .	10
<b>04. Criando prompts claros e específicos</b>	<b>12</b>
Clareza . . . . .	12
Especificidade . . . . .	14
Problemas a serem evitados . . . . .	15
Sobrecarga . . . . .	15
Ambiguidade . . . . .	15
Complicando demais . . . . .	16
<b>05. Elementos de um Prompt</b>	<b>17</b>
Instrução . . . . .	17
Contexto . . . . .	17
Dados de Entrada . . . . .	17
Indicador de Saída . . . . .	18
<b>06. Público, persona e exemplo</b>	<b>20</b>
Persona . . . . .	20
Público . . . . .	20
Exemplos . . . . .	21

<b>07. Delimitadores e Saídas Estruturadas</b>	<b>22</b>
O que são delimitadores? . . . . .	22
O que são Saídas Estruturadas? . . . . .	23
<b>08. Zero, One e Few-Shot</b>	<b>25</b>
Zero Shot . . . . .	25
Utilizando a Técnica One-Shot . . . . .	25
Utilizando a Técnica Few-Shot . . . . .	26
Formatações Few-shot . . . . .	26
Limitações de Few-shot . . . . .	27
<b>09. Dando tempo para o modelo pensar (COT)</b>	<b>28</b>
Pense passo a passo . . . . .	28
Cadeia de Pensamento . . . . .	29
<b>10. Encadeamento (Chaining)</b>	<b>31</b>
Cadeias de Roteamento . . . . .	32
<b>11. Resumo e Extração</b>	<b>34</b>
Resumindo informações . . . . .	34
Resumindo informações com foco específico . . . . .	34
Extraindo Informações . . . . .	35
<b>12. Inferência e Análise de Sentimento</b>	<b>36</b>
Inferindo sentimentos . . . . .	36
Inferindo tópicos . . . . .	38
<b>13. Transformando, traduzindo e modificando</b>	<b>39</b>
Tradução . . . . .	39
Modificação de Tom . . . . .	39
Conversão de Formatos . . . . .	40
Correção ortográfica . . . . .	41
<b>14. Criação de Chatbots Personalizados</b>	<b>42</b>
<b>15. Recursos adicionais</b>	<b>44</b>

## 01. Introdução ao Curso

Sejam bem-vindos ao curso “Engenharia de Prompts”, uma jornada inovadora e empolgante pelo universo emergente dos modelos de linguagem de aprendizado de máquina, conhecidos como LLM (Large Language Models). Este curso foi projetado para ser acessível e enriquecedor, atendendo tanto ao novato curioso quanto ao indivíduo já familiarizado com as tecnologias de inteligência artificial.

Vivemos um momento extraordinário na história da tecnologia. Os modelos de linguagem, como os desenvolvidos pela OpenAI, estão revolucionando a maneira como interagimos com informações, automatizamos tarefas e exploramos novas possibilidades criativas e analíticas. Ao iniciar este curso, você se posiciona na vanguarda de uma onda de inovação que está apenas começando a se formar.

Dominar as ferramentas de LLM é mais do que um diferencial; tornou-se uma necessidade para aqueles que desejam se destacar em um mundo cada vez mais orientado por dados e inteligência artificial. A habilidade de utilizar esses modelos de forma eficaz pode ser comparada a possuir uma superpotência digital. Quando bem utilizados, eles podem expandir sua capacidade de resolver problemas complexos, criar conteúdo inovador e tomar decisões bem-informadas.

Prepare-se para se aprofundar em um conteúdo estruturado e prático, de grande relevância. Ao concluir este curso, você terá não apenas uma compreensão teórica, mas também habilidades práticas que o capacitarão a extrair o máximo dos modelos LLM.

## 02. Conceitos Gerais de LLMs

O objetivo deste curso é aprimorar o uso de modelos de LLM, como o ChatGPT. Para alcançar esse objetivo, é essencial desenvolver primeiro uma compreensão mais aprofundada do que é um modelo LLM.

### O que é uma LLM?

LLM é a sigla para Large Language Model (Modelo de Linguagem de Grande Escala). Modelos de linguagem são sistemas de inteligência artificial com dois propósitos principais: processar e gerar textos. Esses sistemas são geralmente baseados em redes neurais, que são uma forma computacional de simular o funcionamento do cérebro humano. Nessa técnica, em vez de neurônios, temos parâmetros, e, no caso dos LLMs, a quantidade de parâmetros ultrapassa a casa dos bilhões. Para efeito de comparação, o cérebro humano possui cerca de 100 bilhões de neurônios.

### Como as LLMs são treinadas?

As LLMs são treinadas com enormes volumes de texto coletados da internet. Terabytes de textos são capturados e utilizados em um treinamento não supervisionado, permitindo que os modelos prevejam a continuação de um texto fornecido. Esse processo requer grandes servidores com clusters de placas de vídeo. O treinamento pode durar várias semanas, devido ao grande volume de dados, e é financeiramente custoso, além de ter um significativo impacto ambiental em termos de emissões de CO2. Um estudo estima que o treinamento de uma LLM com 2 bilhões de parâmetros pode custar cerca de 1,6 milhão de dólares. O resultado desse treinamento é semelhante a uma vasta compressão de dados, como um arquivo zip. De dezenas de terabytes, chegamos a um modelo com 70 bilhões de parâmetros (como o LLaMA 2 70B), que ocupa aproximadamente 140 GB. Embora essa compressão implique em perda de dados, é um trabalho notável para otimizar o uso das informações disponíveis na internet. A partir desse treinamento, obtemos uma LLM base, que ainda não é o modelo que utilizamos no ChatGPT.

### Diferenças entre uma LLM Base e uma LLM treinada para instruções

A LLM base é treinada para prever o próximo segmento de texto mais provável, dada uma entrada. Por exemplo, se fornecermos o seguinte texto a uma LLM base:

Era uma vez, um unicórnio

É possível que ela nos responderia:

que vivia em uma floresta mágica

Seria uma resposta aceitável dado o texto que eu ofereci a ela. Agora caso fizéssemos um questionamento ela, como:

Qual é a capital do Brasil?

É possível que a LLM nos retornasse:

Qual é a maior cidade do Brasil?

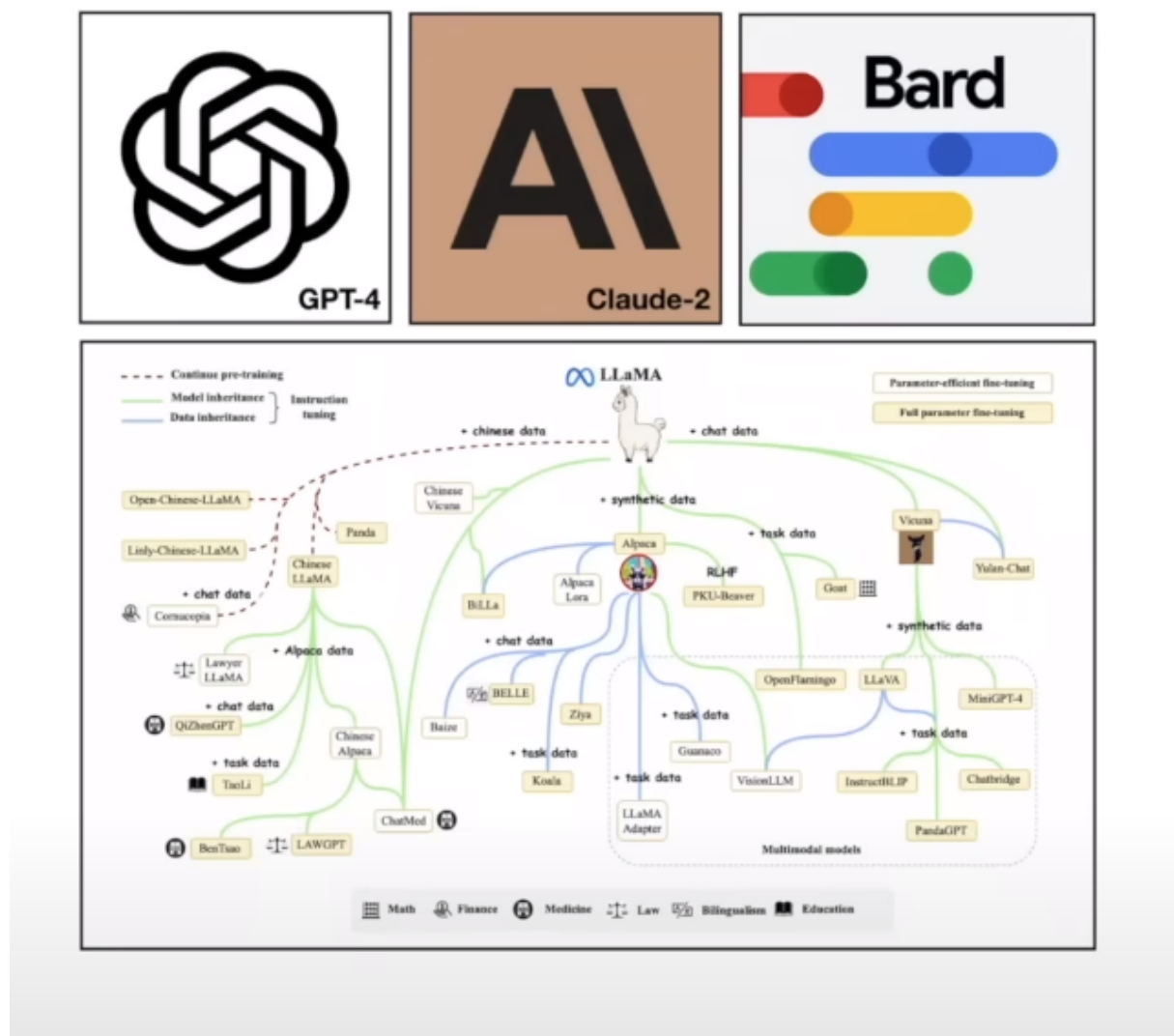
Qual é a população do Brasil?

Uma LLM base, por sua natureza, não reconhece que está diante de um questionamento que exige uma resposta específica. Ela simplesmente fornece o texto mais provável a seguir, baseando-se em dados genéricos obtidos da internet. Por isso, respostas como a mencionada anteriormente são comuns.

Em contraste, uma LLM treinada para seguir instruções é desenvolvida para responder de maneira mais direcionada. Essas LLMs são aprimoradas a partir de uma LLM base e, em seguida, submetidas a um treinamento adicional com conjuntos de dados compostos por instruções e as respectivas respostas a essas instruções. O processo envolve a criação de milhares de instruções distintas e a colaboração de especialistas para fornecer as respostas adequadas. Com essa nova massa de dados, o modelo é treinado sobre a LLM base, aprendendo a emular o comportamento de um especialista. É como se o emaranhado de informações que o modelo possui fosse canalizado para gerar respostas que se alinham à linguagem humana. Assim, chegamos aos modelos atuais do ChatGPT, que são exemplos de LLMs treinadas para seguir instruções.

### **Os principais modelos no mercado**

Atualmente, existem três principais players no mercado de modelos de linguagem fechados: a OpenAI com o GPT, a Anthropic com o Claude e a Google com o Bard e o Gemini. Paralelamente, temos um player que se destaca por fornecer modelos abertos, a Meta. A partir do modelo LLaMA da Meta, que conta com 70 bilhões de parâmetros, tem-se criado um vasto ecossistema de modelos abertos. Esses modelos tiram proveito da capacidade do LLaMA para serem treinados com objetivos específicos, atendendo às mais diversas finalidades.



## Quais parâmetros comuns de um modelo de LLM

urante nossos cursos, interagiremos com os modelos de LLM principalmente por meio de APIs. Utilizando APIs, podemos explorar todo o potencial do Python para automações e processamento de grandes volumes de dados, além de acessar os recursos dos modelos de LLM.

Ao utilizar APIs, temos a possibilidade de configurar certos parâmetros do modelo, permitindo que ele reaja de maneiras distintas aos mesmos estímulos. É necessário realizar experimentações para compreender as configurações mais adequadas para cada aplicação. Vamos abordar brevemente cada um desses parâmetros para que vocês entendam suas implicações gerais:

- **Temperature (Temperatura):** Controla o grau de determinismo do modelo. Um modelo mais determinístico tende a escolher o token de maior probabilidade como o próximo na sequência. Aumentar a temperatura resulta em respostas mais aleatórias, incrementando a “criatividade” do modelo.
- **Top P:** Refere-se a uma técnica de amostragem conhecida como amostragem de núcleo, que também controla o determinismo do modelo. Com um valor baixo de Top P, o modelo considera apenas os tokens que compõem a massa de probabilidade do Top P, selecionando respostas mais confiáveis.
- **Max Length (Tamanho Máximo):** Permite gerenciar o número de tokens que o modelo gera, ajustando o tamanho máximo. Definir um limite ajuda a evitar respostas excessivamente longas ou irrelevantes e a controlar os custos.
- **Stop Sequences (Sequências de Parada):** São cadeias de caracteres que sinalizam ao modelo para parar de gerar tokens. Especificar sequências de parada ajuda a controlar o comprimento e a estrutura da resposta. Por exemplo, para limitar uma lista a 10 itens, você pode adicionar “11” como uma sequência de parada.
- **Frequency Penalty (Penalidade de Frequência):** Aplica uma penalidade ao próximo token proporcionalmente à frequência com que esse token já apareceu na resposta e no prompt. Uma penalidade de frequência mais alta reduz a probabilidade de repetição de palavras na resposta.
- **Presence Penalty (Penalidade de Presença):** Impõe uma penalidade a tokens repetidos, mas, diferentemente da penalidade de frequência, a penalidade é a mesma independentemente da frequência da repetição. Isso evita que o modelo repita frases com muita frequência. Uma penalidade de presença mais alta pode ser usada para gerar texto mais diversificado ou criativo, enquanto uma penalidade mais baixa pode ajudar o modelo a manter o foco.

Assim como com a temperatura e o Top P, a recomendação geral é ajustar a penalidade de frequência ou de presença, mas não ambas simultaneamente.

### Como abordaremos este curso

Utilizaremos o playground da OpenAI <https://platform.openai.com/playground?mode=chat>. O modelo utilizado será o gpt-3.5-turbo e os demais parâmetros: temperature=1 e top\_p=1. Você também poderá utilizar diretamente a interface do chatgpt <https://chat.openai.com/>.



### **03. Os princípios de um bom prompt**

Agora, vamos nos aprofundar no tema central deste curso: a criação de um bom prompt. Embora muitos de vocês já saibam o que é um prompt, é sempre útil reforçar o conceito.

#### **O que é um prompt?**

Um prompt é a entrada que fornecemos a um modelo de linguagem. A ideia é que, ao receber essa entrada, o modelo a processe em sua rede neural e nos forneça uma resposta. A entrada é comumente chamada de prompt.

Nosso objetivo no curso é criar prompts eficientes que resolvam os problemas em questão. Muitas pessoas afirmam que o ChatGPT não é adequado para diversas aplicações, mas sou cético em relação a essas declarações. Em 95% dos casos, o problema está na formulação de um prompt inadequado. A criação de prompts é um campo em estudo intenso. Devido ao tamanho dos modelos, não podemos prever todas as suas respostas. Os estudos em engenharia de prompts são bastante empíricos e baseiam-se em tentativa e erro. No entanto, há melhorias comprovadas cientificamente ao seguir algumas diretrizes principais, que exploraremos no curso. Existem dois princípios fundamentais na criação de prompts, e vamos explorá-los em detalhes ao longo do curso.

#### **1º Princípio - Escrever instruções claras e específicas**

É importante entender que, apesar da aparente inteligência dos modelos de linguagem, eles ainda são scripts executados em computadores. Muitos aspectos da comunicação humana não são verbais. Por exemplo, quando faço uma pergunta ao meu filho, ele conhece minha identidade, meu estilo de comunicação, tem acesso ao nosso histórico de conversas, pode ler meu humor pelo meu rosto, sabe quais são meus interesses, etc. Todas essas informações contextualizam suas respostas. No entanto, quando interagimos com uma máquina por meio de um prompt, ela não tem acesso a essas nuances. Naturalmente, suas respostas tendem a ser mais genéricas e impessoais. Contudo, é possível criar prompts claros e específicos que adicionem contexto, história ou exemplos, permitindo que a máquina compreenda melhor o que está sendo solicitado e forneça respostas mais qualificadas.

Existem várias técnicas para adicionar clareza a um prompt, como definir o contexto, formatar a saída, determinar o público e a persona, além de utilizar técnicas de one-shot e few-shot prompting. Exploraremos essas técnicas com mais detalhes no decorrer do curso.

## 2º Princípio - Dar ao modelo tempo para pensar

Assim como os seres humanos, quanto mais tempo dermos ao modelo para “pensar”, mais refinada será a resposta. Você pode se perguntar como dar mais tempo ao modelo. Existem várias técnicas desenvolvidas que instruem o modelo a realizar seu raciocínio passo a passo. Por exemplo, se apresentarmos uma questão matemática ao modelo e pedirmos apenas uma resposta de “Correto” ou “Errado”, a probabilidade de erro é alta, pois o modelo não sabe que deve raciocinar antes de responder. Se pedirmos para ele realizar o cálculo e depois indicar se a solução está correta ou errada, a chance de acerto aumenta. Outra técnica importante é dividir um prompt maior em prompts menores e mais específicos sempre que possível.

## Limitações de um modelo

Devemos estar cientes de que, apesar de todos os nossos esforços, podemos nos deparar com limitações inerentes a um modelo LLM. As principais limitações estão relacionadas a alucinações e vieses.

### Alucinações

Alucinações ocorrem quando o modelo responde com confiança a uma informação incorreta. Ele pode criar detalhes e fatos, pois possui apenas um resumo das informações que foram passadas durante seu treinamento. Vejamos um exemplo no prompt a seguir:

Fale sobre as características das escovas de dente da marca "SorriDente"

A marca SorriDente é inventada e não existe e veja o que o modelo retornou:

As escovas de dente da marca "SorriDente" são reconhecidas por suas características de alta  
→ qualidade e eficiência na limpeza bucal. Algumas características das escovas de dente da  
→ marca incluem:

1. Cerdas macias e resistentes: As cerdas das escovas "SorriDente" são feitas de materiais de  
→ alta qualidade, o que as torna suaves para a gengiva, mas ao mesmo tempo resistentes para  
→ uma limpeza eficaz dos dentes.
2. Design ergonômico: As escovas de dente da marca "SorriDente" possuem um design ergonômico  
→ que proporciona maior conforto e facilidade de uso durante a escovação.
3. Durabilidade: As escovas de dente "SorriDente" são conhecidas pela durabilidade de suas  
→ cerdas e materiais, garantindo uma maior vida útil do produto.
4. Diversidade de modelos: A marca oferece uma variedade de modelos de escovas de dente para  
→ atender às diferentes necessidades dos consumidores, como escovas com cerdas angul

Muitas alucinações também ocorrem quando submetemos o ChatGPT a testes reversos. Como o seguinte:

Quem é a mãe do Ronaldinho Gaúcho?

E a resposta é a seguinte e está correta:

A mãe do Ronaldinho Gaúcho é Dona Miguelina Elói Assis dos Santos, conhecida como Dona ↩ Miguelina. Ela faleceu em 2021 devido a complicações causadas pela COVID-19.

Agora se faço a pergunta reversa:

Quem é o filho de Miguelina Elói Assis dos Santos?

O modelo alucina e responde que o Neymar é filho da Dona Miguelina:

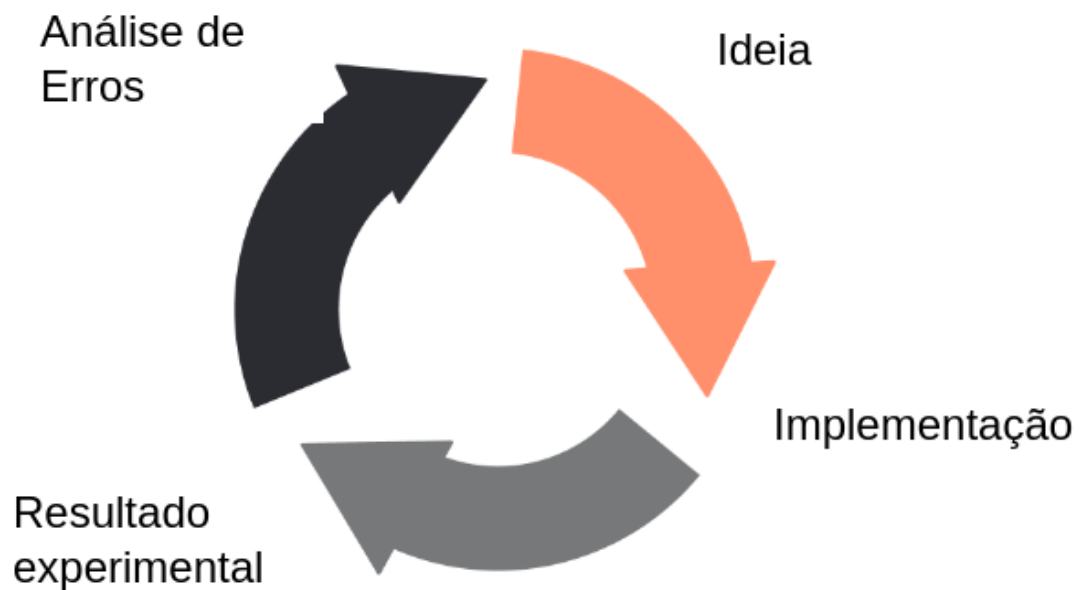
O filho de Miguelina Elói Assis dos Santos é o jogador de futebol Neymar Jr.

### Vieses

Outro problema comum é o modelo responder com vieses culturais, que são reflexos dos dados com os quais foi treinado. O modelo acaba se tornando um espelho da sociedade; se os dados coletados expressam preconceitos, sexismo, violência, entre outros aspectos negativos, há uma tendência de o modelo reproduzir esses comportamentos. Isso pode ser particularmente perigoso ao criar soluções comerciais destinadas a um uso amplo e diversificado.

### Prompting é sobre iteratividade

O processo de criação de um bom prompt exige iteratividade. Nunca se chega a uma boa resposta na primeira tentativa. É necessário sempre realizarmos um processo contínuo de melhora de nosso prompt, de forma a obtermos constantemente respostas melhores do modelo.



O processo é o seguinte: começamos com uma ideia inicial de prompt, que fornecemos ao modelo para obter uma resposta experimental. Analisamos essa resposta e, com base nela, geramos uma nova ideia, refinando o prompt anterior. Esse ciclo se repete até alcançarmos um estágio em que obtemos respostas consistentemente boas do modelo.

## 04. Criando prompts claros e específicos

Um prompt eficaz possui dois elementos principais: clareza e especificidade.

### Clareza

A clareza é alcançada ao fornecer um contexto relevante ao modelo. Como mencionado anteriormente, o modelo não sabe quem somos, como pensamos ou a maneira como nos comunicamos. Se desejamos obter respostas que atendam às nossas expectativas, precisamos fornecer todo esse contexto ao modelo. Por exemplo, suponha que eu seja um professor de física ensinando sobre ação e reação. Eu poderia solicitar ao ChatGPT ajuda para elaborar um plano de aula:

Desenvolva um plano para uma aula sobre relatividade

O modelo respondeu o seguinte.

Objetivo da aula: Compreender os princípios da teoria da relatividade e como ela influencia  
↪ nossa compreensão do universo.

1. Introdução à teoria da relatividade:

- Breve histórico sobre Albert Einstein e a criação da teoria da relatividade.
- Explicação dos conceitos de relatividade restrita e relatividade geral.

2. Princípios da relatividade restrita:

- Discussão sobre a invariância das leis da física em diferentes sistemas de referência.
- Explicação da dilatação do tempo e contração do espaço.
- Exemplos práticos de como esses princípios se aplicam em situações do nosso cotidiano.

3. Princípios da relatividade geral:

- Explicação da curvatura do espaço-tempo causada pela presença de massa.
- Discussão sobre a gravidade como resultado dessa curvatura.
- Apresentação de exemplos de como a relatividade geral explica fenômenos astronômicos, como a  
↪ curvatura da luz ao redor de objetos massivos.

4. Implicações da teoria da relatividade:

- Discussão sobre como a teoria da relatividade influenciou a física moderna e revolucionou  
↪ nossa compreensão do universo.
- Apresentação de experimentos e observações que confirmam as previsões da teoria da  
↪ relatividade.

5. Atividade prática:

- Propor aos alunos um experimento simples para ilustrar os conceitos de relatividade  
↪ restrita, como a dilatação do tempo.
- Discussão em grupo sobre os resultados do experimento e como ele se relaciona com os  
↪ princípios da teoria da relatividade.

6. Debate e conclusão:

- Encorajar os alunos a debaterem sobre as implicações filosóficas da teoria da relatividade,  
↪ como a relatividade da simultaneidade e a natureza do tempo.

- Conclusão da aula reforçando a importância da teoria da relatividade na compreensão do universo e incentivando os alunos a continuarem explorando o tema.

Recursos necessários: Projetor, slides de apresentação, material para experimento prático.

Avaliação: Os alunos serão avaliados com base na participação na discussão em grupo, compreensão dos conceitos apresentados e realização do experimento prático.

A resposta é boa, mas não necessariamente adequada as minhas necessidades e está bem generalista. Poderíamos aumentar a especificidade da seguinte forma:

Eu sou um professor de Física de 8ª série preparando uma aula sobre relatividade. Desenvolva um plano de aula.

E a resposta é a seguinte:

Plano de aula sobre relatividade para a 8ª série:

Objetivo da aula:

- Compreender os princípios básicos da teoria da relatividade de Einstein.

Materiais necessários:

- Quadro negro ou branco;
- Giz ou marcador;
- Computador ou projetor para apresentação de slides.

Estratégias de ensino:

1. Introdução (10 minutos):

- Iniciar a aula explicando o que é relatividade e sua importância na física moderna.
- Apresentar um breve histórico sobre como a teoria da relatividade surgiu e quem foi Albert Einstein.

2. Relatividade especial (15 minutos):

- Explicar os postulados da relatividade especial, como o princípio da relatividade e o princípio da constância da velocidade da luz.
- Apresentar exemplos simples que ilustrem esses princípios, como o experimento de Michelson-Morley.

3. Dilatação do tempo e contração do espaço (20 minutos):

- Explicar os conceitos de dilatação do tempo e contração do espaço, mostrando como o tempo e o espaço são relativos de acordo com a teoria da relatividade.
- Apresentar exemplos práticos desses fenômenos, como o experimento dos gêmeos.

4. Efeito gravitacional (15 minutos):

- Discutir como a relatividade geral de Einstein descreve a gravidade como uma curvatura no espaço-tempo.
- Apresentar exemplos de como a gravidade influencia o movimento dos corpos celestes, como a curvatura da luz ao redor de um objeto massivo.

5. Conclusão (10 minutos):

- Resumir os principais conceitos abordados na aula e destacar a importância da teoria da relatividade na compreensão do universo.
- Encorajar os alunos a fazer perguntas e promover discussões sobre o tema.

Atividade extra-classe:

- Propor aos alunos a leitura de artigos ou assistência a vídeos que abordem a teoria da relatividade de forma mais aprofundada, estimulando o interesse pela física teórica.

E assim temos uma resposta que, não necessariamente seja melhor que a primeira, mas sim está mais adequada ao contexto que queremos.

### Especificidade

Além de ser claro em relação ao nosso contexto, é importante fornecer instruções específicas para obtermos boas soluções. Imagine que você está caminhando em um bosque e alguém lhe dá a seguinte instrução: “Vá até a árvore!”. A primeira coisa que você pensaria é: “Qual árvore?”, afinal, está no meio de um bosque com diversas árvores. Agora, se a instrução fosse: “Vá até a árvore de um metro e meio, com folhas largas, frutas laranjas e que possui um tronco fino”, você perceberia que é uma instrução específica. Antes, havia centenas de árvores possíveis, mas agora, com uma descrição específica, só resta uma e você sabe para onde ir. Este é um exemplo de especificidade, onde reduzimos o campo possível de ação para que o modelo possa fornecer uma resposta adequada às nossas necessidades. Vamos continuar com o nosso exemplo:

Eu sou um professor de Física de 8ª série preparando uma aula sobre relatividade. Desenvolva um plano de aula para uma turma de 20 alunos que durará 2 horas. Na sala teremos

- ↪ projetor, quadro branco e computadores para utilização.

E a resposta é a seguinte:

Plano de aula sobre Relatividade para 8ª série:

Objetivo da aula: Introduzir os conceitos básicos da Teoria da Relatividade de Albert Einstein

- ↪ e como ela transformou nossa compreensão do Universo.

Materiais necessários:

- Projetor
- Quadro branco
- Computadores com acesso à internet
- Material visual para ilustrar os conceitos (vídeos, animações, etc.)

Duração: 2 horas

Desenvolvimento da aula:

1. Introdução (15 minutos):

- Apresentação do tema da aula e explicação da importância da Teoria da Relatividade na Física.
- ↪ Breve contextualização histórica sobre a origem da teoria e seu impacto na Ciência.

2. Conceitos básicos da Relatividade Especial (45 minutos):

- Explicação do princípio da relatividade e suas consequências.
- Discussão sobre o paradoxo dos gêmeos e a dilatação do tempo.

- Exemplos e exercícios para ilustrar os conceitos apresentados.

### 3. Conceitos avançados da Relatividade Geral (45 minutos):

- Explicação do princípio da equivalência e a curvatura do espaço-tempo.
- Discussão sobre a gravidade e a influência das massas no espaço-tempo.
- Vídeos e animações para visualizar a curvatura do espaço-tempo.

### 4. Atividade prática em grupo (30 minutos):

- Dividir os alunos em grupos e propor um desafio relacionado à Relatividade, como a resolução de problemas ou a criação de experimentos simples.  
↳ de problemas ou a criação de experimentos simples.
- Cada grupo deverá apresentar seus resultados para a turma.

### 5. Conclusão e debate (15 minutos):

- Síntese dos conceitos abordados na aula e revisão dos pontos principais.
- Debate sobre as aplicações práticas da Teoria da Relatividade no mundo atual.

#### Recursos adicionais:

- Listas de exercícios para aprofundar o aprendizado.
- Links para sites e vídeos com conteúdos complementares.
- Indicação de filmes e documentários sobre a Teoria da Relatividade para os alunos assistirem  
↳ em casa.

#### Avaliação:

- Observação da participação dos alunos durante a aula.
- Realização de uma breve avaliação escrita ao final da aula para verificar a compreensão dos  
↳ conceitos abordados.

## Problemas a serem evitados

Alguns problemas são comuns e devemos já ter consciência desde o início.

### Sobrecarga

A sobrecarga está relacionada à adição de muita informação irrelevante no prompt. Ao tentarmos aumentar a clareza, adicionando um contexto ao prompt, podemos incluir informações não necessárias para a instrução que darei ao modelo. Portanto, tenha cuidado e adicione apenas o necessário!

### Ambiguidade

Você está tentando adicionar clareza, portanto, não seja ambíguo. Por exemplo, caso não saibamos quantos alunos assistirão à nossa aula sobre relatividade, não crie um prompt da seguinte forma:

Desenvolva um plano de aula para uma turma de 5 a 30 alunos

Melhor então seria criar múltiplos prompts, que se encaixariam nas diversas situações:

Desenvolva um plano de aula para uma turma de 5 alunos



Desenvolva um plano de aula para uma turma de 15 alunos

Desenvolva um plano de aula para uma turma de 30 alunos

Agora não há ambiguidade e você estará preparado para todas as situações. Mas não leve a sua incerteza para o prompt!

### **Complicando demais**

Ajude o modelo a fornecer respostas eficazes. Portanto, evite o uso de termos técnicos muito complexos, jargões, expressões idiomáticas pouco comuns ou frases muito complexas. Tudo isso dificulta o trabalho do modelo. Portanto, seja claro e específico, mas sempre com simplicidade!

## 05. Elementos de um Prompt

Certos elementos aparecem repetidamente em um prompt bem desenvolvido. Por isso, é necessário destacá-los para que possamos refinar os elementos que funcionam bem. Os quatro principais são:

- Instrução: a tarefa específica a ser realizada pelo modelo
- Contexto: informação externa ou contexto adicional que auxilia o modelo a atingir melhores respostas
- Dados de entrada: a entrada ou questão que estamos interessados na resposta
- Indicador de Saída: o tipo ou formato da saída

### Instrução

É a ordem que será dada ao modelo. Lembrando dos nossos princípios, esta ordem deve ser específica, para que não fique ambíguo ao modelo. A instrução também pode ser

Desenvolva um plano de investimento eficaz, definindo tipos de investimentos, proporções  
↪ entre investimentos e diversificação para o perfil do cliente abaixo.

### Contexto

Já utilizamos bastante esta palavra durante o curso, já que é um dos elementos principais do modelo. É através do contexto que o modelo passa a nos conhecer, entender como pensamos e como agimos, para termos uma resposta adequada ao nosso objetivo. O contexto dá clareza ao modelo.

Eu sou um assessor financeiro de pessoas jovens, com até 30 anos de idade e com baixo poder  
↪ aquisitivo. Ajudo elas a desenvolverem uma educação financeiro e formo planos adequados  
↪ para longo prazo, para clientes que querem começar cedo a investir pensando em suas  
↪ aposentadorias.

### Dados de Entrada

Estas informações estão relacionadas aos dados mutáveis, por assim dizer. Em geral, conseguiremos reutilizar a instrução e contexto para diversos prompts diferentes, modificando apenas o dado de entrada, pois é ele que queremos processar ou que direcionará a resposta do nosso prompt.

Um homem de 25 anos, com baixa propensão a risco e consegue aportar 1000 reais por mês em  
↪ novos investimentos. Hoje ele já possui 50 mil reais investidos em ações.

### Indicador de Saída

Aqui nós indicamos que é hora do modelo retornar uma resposta e dizemos como queremos que a resposta seja dada, em qual formato. Muitas vezes utilizamos apenas uma referência, como:

Plano de investimento:

No final, chegamos a este prompt completo, claro e específico:

Instrução: Desenvolva um plano de investimento eficaz, definindo tipos de investimentos,  
→ proporções entre investimentos e diversificação para o perfil do cliente abaixo.

Contexto: Eu sou um assessor financeiro de pessoas jovens, com até 30 anos de idade e com  
→ baixo poder aquisitivo. Ajudo elas a desenvolverem uma educação financeiro e formo planos  
→ adequados para longo prazo, para clientes que querem começar cedo a investir pensando em  
→ suas aposentadorias.

Dado de entrada: Um homem de 25 anos, com baixa propensão a risco e consegue aportar 1000  
→ reais por mês em novos investimentos. Hoje ele já possui 50 mil reais investidos em ações.

Plano de investimento:

E temos um belíssimo plano de investimento:

Para o perfil do cliente descrito, com baixa propensão a risco e foco em investimentos de  
→ longo prazo, sugiro um plano de investimento diversificado em diferentes tipos de ativos.

1. Reserva de emergência: Recomendo que o cliente mantenha uma reserva de emergência  
→ equivalente a pelo menos 6 meses de despesas em um investimento de baixo risco e alta  
→ liquidez, como Tesouro Selic ou CDB com liquidez diária.
2. Previdência Privada: Uma opção interessante para complementar a aposentadoria no longo  
→ prazo, com benefícios fiscais e possibilidade de planejamento sucessório. Recomendo  
→ aportar parte dos investimentos mensais em um plano de previdência privada de baixa taxa  
→ de administração.
3. Fundos de investimento: Para diversificar a carteira, é interessante considerar a alocação  
→ em fundos de investimento que possuam uma gestão profissional e diversificada, como fundos  
→ de renda fixa, multimercado e de ações.
4. Fundos de índice (ETFs): Uma alternativa de investimento passivo e de baixo custo, os ETFs  
→ replicam índices de mercado e oferecem diversificação de forma eficiente. Recomendo  
→ aportar uma parte dos investimentos em ETFs que acompanhem o desempenho do mercado  
→ acionário.
5. Investimento em ações: Dado que o cliente já possui uma carteira de ações, é importante  
→ continuar acompanhando e diversificando esse investimento. Recomendo avaliar a  
→ possibilidade de aumentar a diversificação da carteira com aportes mensais em novas ações  
→ ou em fundos de ações.

Proporção entre os investimentos:

- Reserva de emergência: 10%

## Engenharia de Prompts

---

- Previdência Privada: 20%
- Fundos de investimento: 30%
- ETFs: 20%
- Ações: 20%

Dessa forma, o cliente terá uma carteira diversificada, adequada ao seu perfil de baixo risco,

- ↪ com potencial de crescimento no longo prazo e preparada para a aposentadoria. É importante
- ↪ ressaltar a importância de revisar periodicamente o plano de investimento e fazer ajustes
- ↪ conforme necessário.

## 06. Público, persona e exemplo

### Persona

Utilizar uma persona é uma ótima forma de criar um contexto para o seu prompt. A ideia por trás é utilizarmos uma pessoa, fictícia ou real, para nos comunicarmos com o modelo. Temos vários exemplos possíveis de personas, como:

Escreva uma história curta de magia no estilo da J.K. Rowling

Escreva um poema ao estilo de um poeta brasileiro do romantismo

Você pode fazer o trabalho de descrever esta persona, como fizemos anteriormente:

Eu sou um assessor financeiro de pessoas jovens, com até 30 anos de idade e com

No caso, o exemplo de contexto que demos na aula anterior também serviria como um exemplo de persona. Mas atenção! Contexto não é a mesma coisa que persona. No contexto poderíamos adicionar diversas outras informações, como a situação atual do mundo, a situação do público que queremos nos comunicar, as nossas crenças e formas de ver a vida, etc. Já a persona é apenas a descrição do agente, para que o modelo se coloque no lugar dele ao se comunicar.

### Público

Já o público é com quem estamos nos comunicando. É um direcionamento do nosso discurso, para definir e explicar ao modelo com quem estamos nos comunicando. O público, da mesma forma que a persona, faz parte do contexto de um prompt. Vamos dar um exemplo:

Escreva uma história simples de magia pra crianças tem medo do escuro

Até vou colocar a resposta aqui, porque a história ficou bem bonita.

Era uma vez uma menina chamada Sofia, que sempre teve muito medo do escuro. Todas as noites,  
↳ quando chegava a hora de dormir, ela ficava tremendo de medo de apagar a luz.

Um dia, Sofia descobriu que sua avó tinha um colar mágico que podia afastar o medo do escuro.  
↳ A avó explicou que o colar tinha o poder de criar uma luz brilhante e reconfortante, capaz  
↳ de iluminar até mesmo a mais escura das noites.

Animada com a possibilidade de finalmente superar seu medo, Sofia colocou o colar em volta do  
↳ pescoço e fechou os olhos. Quando abriu novamente, viu a luz do colar brilhar  
↳ intensamente, iluminando todo o seu quarto.

A partir daquele dia, Sofia nunca mais teve medo do escuro. Ela sabia que, graças ao colar  
↳ mágico de sua avó, sempre teria uma luz para guiá-la e protegê-la. E assim, a menina pôde  
↳ dormir tranquilamente todas as noites, sabendo que a magia estava sempre ao seu lado.

E vocês podem perceber que, apenas ao falar para quem estou me comunicando - o público -, o modelo já conseguiu criar uma resposta adequada. O combo público e persona sempre podem ser utilizados para dar clareza ao prompt, criando contextos melhor. Naturalmente eles não são elementos necessários, mas podem ajudar muito.

### Exemplos

Outra técnica muito eficaz é, ao invés de definir uma persona, fornecer ao modelo um exemplo de respostas similares para que ele possa se basear. Por exemplo, digamos que queiramos escrever um e-mail com o chatGPT que imite a nossa maneira de escrever e-mails. Poderíamos passar um exemplo de e-mail nosso e pedir para que ele se baseie no exemplo ao escrever um novo e-mail.

Aqui está um exemplo do meu estilo de escrita de emails:

####

Fala, meu amigo!

Como estamos?

O que você acha de marcamos um dia para conversar?

Tamo junto!

Abraço

####

Imitando o estilo do email anterior, crie um novo email para realizar a ação a seguir.

Marcar uma conversa no "Bar do Antonio" às 20h e avisar que você convidou o Pedro também.

E a resposta é esta:

Fala, parceiro!

Beleza?

Que tal marcamos uma conversa no "Bar do Antonio" às 20h? Convidei também o Pedro para se  
↪ juntar a nós.

Aguardo sua confirmação!

Tamo junto!

Grande abraço!

Poderíamos adicionar mais de um exemplo neste caso, ou exemplos diferentes dependendo do tipo de mensagem. Aí vai da criatividade individual de cada um. Os exemplos são mais uma forma de melhorar o nosso contexto. E esse contexto fica cada vez mais eficiente!

## 07. Delimitadores e Saídas Estruturadas

Agora estamos entrando no campo da programação, somada à utilização de LLMs. Para utilizarmos um modelo de forma sequencial, processando um grande volume de dados, precisamos criar prompts organizados e facilmente legíveis, onde seus elementos principais fiquem evidentes ao modelo. Para isso, os delimitadores são fundamentais:

### O que são delimitadores?

Os delimitadores delimitam os trechos mais importantes no prompt, elementos que não podem passar despercebidos pelo modelo. Podemos utilizar diversos delimitadores, como ####, "", -, etc. Anteriormente demos um exemplo de delimitador ao fazermos o prompt para criação de e-mails:

Aqui está um exemplo do meu estilo de escrita de emails:

####

Fala, meu amigo!

Como estamos?

O que você acha de marcamos um dia para conversar?

Tamo junto!

Abraço

####

Imitando o estilo do email anterior, crie um novo email para realizar a ação a seguir.  
Marcar uma conversa no "Bar do Antonio" às 20h e avisar que você convidou o Pedro também.

Você pode perceber que, neste caso, o exemplo de e-mail que foi passado ao modelo não pode passar despercebido. Por isso, adicionamos o delimitador ####. Ele não só demarca o ponto fundamental do prompt, como delimita seu início e fim. Caso contrário, o modelo poderia confundir o exemplo com a instrução, ou com o resto do contexto. Isso seria um grande problema, pois o exemplo dá o tom da resposta neste caso. Outra tática que gosto de utilizar é avisar ao modelo que estou utilizando um delimitador.

Delimitado por #### está um exemplo do meu estilo de escrita de emails:

####

Fala, meu amigo!

Como estamos?

O que você acha de marcamos um dia para conversar?

Tamo junto!

Abraço

####

Imitando o estilo do email anterior, crie um novo email para realizar a ação a seguir.  
Marcar uma conversa no "Bar do Antonio" às 20h e avisar que você convidou o Pedro também.

Ao escrever o delimitado por #### fica ainda mais claro ao modelo que ele deve prestar atenção neste trecho da mensagem.

Agora você percebe que, ao adicionarmos grandes volumes de dados aos nossos prompts, podemos organizar esta informação facilmente utilizando diferentes delimitadores. Outro ponto fundamental ao utilizarmos prompts com programação é conseguirmos estruturar a saída do modelo de forma que esta saída seja facilmente processada pelo seu script. Para isso, podemos fazer um trabalho de estruturação de saídas.

### O que são Saídas Estruturadas?

Você deve ter percebido que a resposta do modelo pode ser bem inusitada, caso não criemos formas de torná-la mais determinística. Por exemplo, podemos fazer a seguinte pergunta:

Qual o nome das mães das seguintes pessoas:  
Ronaldinho Gaúcho, Neymar, Rivaldo

A resposta é a seguinte:

Mãe de Ronaldinho Gaúcho: Miguelina Elói Assis dos Santos  
Mãe de Neymar: Nadine Santos  
Mãe de Rivaldo: Ivone Vitor dos Santos

Mas a seguinte resposta também seria possível:

Miguelina Elói Assis dos Santos, Nadine Santos, Ivone Vitor dos Santos

E esta resposta também:

Os nomes das mães de Ronaldinho Gaúcho, Neymar e Rivaldo são: Miguelina Elói Assis dos Santos,  
↔ Nadine Santos e Ivone Vitor dos Santos

Para conseguir processar essas informações em um script seria muito difícil, pois a cada interação com o modelo a estrutura da resposta é diferente. Mas podemos contornar este problema sendo específicos quanto à forma que queremos receber a resposta. Por exemplo, poderíamos fazer o seguinte:

Qual o nome das mães das seguintes pessoas:  
Ronaldinho Gaúcho, Neymar, Rivaldo  
Retorne a resposta apenas com os nomes separados por vírgulas.

E perfeito, temos uma resposta facilmente processável por um programa de computador.

Miguelina Eloísa da Silva, Nadine Santos, Rosângela Ferreira Chaves

Mas podemos pedir alguns formatos mais complexos:

Qual o nome das mães das seguintes pessoas:  
Ronaldinho Gaúcho, Neymar, Rivaldo  
Retorne a resposta no formato de uma lista de Python.

E temos uma lista de Python:

```
['Miguelina Eloí Assis dos Santos', 'Nadine Santos', 'Maria do Carmo Ferreira']
```



Ou poderíamos querer em formato JSON:

Qual o nome das mães das seguintes pessoas:

Ronaldinho Gaúcho, Neymar, Rivaldo

Retorne a resposta no formato de Json, com chaves de nome\_filho, nome\_mae.

E temos um JSON com a chaves que solicitamos.

```
{
  "nome_filho": "Ronaldinho Gaúcho",
  "nome_mae": "Miguelina Elói Assis dos Santos"
},
{
  "nome_filho": "Neymar",
  "nome_mae": "Nadine Santos"
},
{
  "nome_filho": "Rivaldo",
  "nome_mae": "Dona Marina"
}
```

Poderíamos ainda solicitar na forma de uma tabela de html:

Qual o nome das mães das seguintes pessoas:

Ronaldinho Gaúcho, Neymar, Rivaldo

Retorne a resposta no formato de uma tabela de Html com as colunas Nome do Filho e Nome da Mãe.

E temos uma tabela de html:

```
<table>
  <tr>
    <th>Nome do Filho</th>
    <th>Nome da Mãe</th>
  </tr>
  <tr>
    <td>Ronaldinho Gaúcho</td>
    <td>Miguelina Eloí Assis dos Santos</td>
  </tr>
  <tr>
    <td>Neymar</td>
    <td>Nadine Santos</td>
  </tr>
  <tr>
    <td>Rivaldo</td>
    <td>Terezinha Ferreira Borba</td>
  </tr>
</table>
```

## 08. Zero, One e Few-Shot

As técnicas de one e few-shot são bastante conhecidas. Na prática, elas não passam de formas de passar exemplos de respostas para “inspirar” o modelo a responder mais de acordo. Já mostramos que podemos passar exemplos para o modelo se basear, como no caso em que fornecemos e-mails próprios como referência para a saída desejada. Agora vamos explorar mais algumas formas de passar exemplos, mas antes de avançarmos para o conceito de one e few-shot, vamos entender o que é zero-shot prompting - a forma mais simples de formular um prompt.

### Zero Shot

Os modelos de LLM atuais são ajustados para seguir instruções e treinados com uma quantidade enorme de dados. Logo, eles são capazes de realizar tarefas “zero-shot”, ou seja, tarefas sem receber um exemplo anterior.

Classifique o texto como neutro, negativo ou positivo.

Texto: Eu acho que as férias estão ok.

Sentimento:

Neutro

Como falamos, não foi necessário mostrar ao modelo exemplos das respostas que gostaríamos, e ele já conseguiu dar a resposta adequada.

Agora, nos casos em que o “zero-shot” não funciona, é recomendado mostrar mais exemplos com o comportamento esperado para o modelo. Aí que entra o “Few-shot Prompting”.

### Utilizando a Técnica One-Shot

Na abordagem de prompting one-shot, o modelo aprende com o contexto fornecido, e essa orientação adicional permite um desempenho melhor. As demonstrações funcionam como condicionamento para exemplos subsequentes nos quais gostaríamos que o modelo gerasse uma resposta. No caso do one-shot, fornecemos apenas uma demonstração para o modelo.

Vamos dar um exemplo:

Um "whatpu" é um animal pequeno e peludo nativo da Tanzânia. Um exemplo de uma frase que usa a  
↪ palavra whatpu é:

Estávamos viajando na África e vimos esses whatpus muito fofos.

Realizar um "farduddle" significa pular para cima e para baixo muito rápido. Um exemplo de uma  
↪ frase que usa a palavra farduddle é:

A criança estava tão animada que começou a farduddle de alegria.

Podemos observar que o modelo de alguma forma aprendeu a realizar a tarefa ao fornecermos apenas um exemplo (ou seja, 1-shot). Para tarefas mais difíceis, podemos experimentar aumentar o número de demonstrações (por exemplo, 3-shot, 5-shot, 10-shot, etc.).

### Utilizando a Técnica Few-Shot

A diferença de one-shot para few-shot é apenas a quantidade de exemplos que são fornecidos ao modelo. No caso de one-shot, uma demonstração é dada, enquanto no few-shot, múltiplas demonstrações são dadas.

### Formatações Few-shot

Podemos utilizar diferentes formatações para realizar o few-shot:

```
Isso é incrível! // Negativo
Isso é ruim! // Positivo
Uau, esse filme foi incrível! // Positivo
Que programa horrível! //
```

Negativo

Outra possível formatação seria a seguinte:

```
Texto: Isso é incrível!
Sentimento: Negativo
Texto: Isso é ruim!
Sentimento: Positivo
Texto: Uau, esse filme foi incrível!
Sentimento: Positivo
Texto: Que programa horrível!
Sentimento:
```

Negativo

E uma muito utilizada é de pergunta e resposta:

```
'''
Pergunta: Em qual país se fala a seguinte frase - "¡Hola! ¿Qué tal?"
Resposta: Espanã 🇪🇸
Pergunta: Em qual país se fala a seguinte frase - "Hello, Gentlemen"
Resposta: England 🇬🇧
Pergunta: Em qual país se fala a seguinte frase - "Wie geht's?"
Resposta:
'''
```

Alemanha 🇩🇪



### Limitações de Few-shot

Para problemas que exigem um raciocínio mais apurado do modelo, a técnica few-shot não é recomendada. Ao mostrarmos exemplos neste caso, o modelo não consegue entender o padrão que leva à resposta.

Os números ímpares neste grupo somam um número par: 4, 8, 9, 15, 12, 2, 1.

Resposta: A resposta é Falsa.

Os números ímpares neste grupo somam um número par: 17, 10, 19, 4, 8, 12, 24.

Resposta: A resposta é Verdadeira.

Os números ímpares neste grupo somam um número par: 16, 11, 14, 4, 8, 13, 24.

Resposta: A resposta é Verdadeira.

Os números ímpares neste grupo somam um número par: 17, 9, 10, 12, 13, 4, 2.

Resposta: A resposta é Falsa.

Os números ímpares neste grupo somam um número par: 15, 32, 5, 13, 82, 7, 1.

Resposta:

A resposta é Verdadeira.

A resposta correta é “Falsa”, e podemos ver que nos casos em que há necessidade de um raciocínio mais elaborado, a técnica few-shot pode não ser suficiente para atingirmos resultados melhores.

## 09. Dando tempo para o modelo pensar (COT)

Agora vamos explorar técnicas que dão ao modelo mais tempo para pensar. Lembrando, como falamos anteriormente, os modelos funcionam como os seres humanos: quanto mais tempo dermos a eles para pensar, mais apurada será a resposta.

Vamos explorar, então, as principais técnicas para o modelo ter mais tempo e melhorar nas tarefas que exigem raciocínio.

### Pense passo a passo

A primeira técnica é muito simples, mas tem demonstrado resultados muito efetivos. Temos um exemplo a seguir:

Determine se a solução do aluno está correta ou não.

Pergunta:

Estou construindo uma instalação de energia solar e preciso de ajuda com as contas de custos  
↪ as finanças.

- O terreno custa R\$ 100 / metro quadrado
- Posso comprar painéis solares por R\$ 250 / metro quadrado
- Negocieei um contrato de manutenção que me custará R\$ 100 mil por ano e R\$ 10 adicionais por  
↪ metro quadrado

Qual é o custo total para o primeiro ano de operações  
em função do número de pés quadrados.

Solução do aluno:

Seja x o tamanho da instalação em metros quadrados.

Custos:

1. Custo do terreno:  $100x$
  2. Custo do painel solar:  $250x$
  3. Custo de manutenção:  $100.000 + 100x$
- Custo total:  $100x + 250x + 100.000 + 100x = 450x + 100.000$

O modelo responde:

A solução do aluno está correta. O custo total para o primeiro ano de operações em função do  
↪ número de metros quadrados é dado por  $450x + 100.000$ .

Mas na realidade a resposta está incorreta, o valor final deveria ser  $360x + 100.000$ . Vamos ver o que ocorre quando pedimos pro modelo pensar passo a passo:

Determine se a solução do aluno está correta ou não.

Pergunta:

Estou construindo uma instalação de energia solar e preciso de ajuda com as contas de custos  
↪ as finanças.

- O terreno custa R\$ 100 / metro quadrado

## Engenharia de Prompts

---

- Posso comprar painéis solares por R\$ 250 / metro quadrado
- Negocieei um contrato de manutenção que me custará R\$ 100 mil por ano e R\$ 10 adicionais por  
↪ metro quadrado

Qual é o custo total para o primeiro ano de operações em função do número de pés quadrados.

Solução do aluno:

Seja x o tamanho da instalação em metros quadrados.

Custos:

1. Custo do terreno:  $100x$
  2. Custo do painel solar:  $250x$
  3. Custo de manutenção:  $100.000 + 100x$
- Custo total:  $100x + 250x + 100.000 + 100x = 450x + 100.000$

Pense passo a passo

E a solução:

A solução do aluno está parcialmente correta.

No entanto, o custo total incorreto para o primeiro ano de operações em função do número de  
↪ metros quadrados. O aluno considerou corretamente os custos do terreno, dos painéis  
↪ solares e da manutenção, mas não somou esses custos corretamente.

O custo total para o primeiro ano de operações em função do número de metros quadrados seria:

Custo total = Custo do terreno + Custo dos painéis solares + Custo de manutenção

Custo total =  $100x + 250x + 100.000 + 10x$

Custo total =  $360x + 100.000$

Portanto, a expressão correta para o custo total em função do número de metros quadrados seria  
↪  $360x + 100.000$ .

Olhem a incrível diferença. Agora o modelo acertou a resposta, só pro adicionarmos Pense passo a passo no final do prompt, lembrando a ele de dar tempo para pensar.

## Cadeia de Pensamento

Cadeia de pensamento (do inglês, chain of thought) é uma técnica similar, mas nela explicamos exatamente os passos que queremos que o modelo percorra para resolver corretamente a instrução.

Determine se a solução do aluno está correta ou não.

Pergunta:

Estou construindo uma instalação de energia solar e preciso de ajuda com as contas de custos  
↪ as finanças.

- O terreno custa R\$ 100 / metro quadrado
- Posso comprar painéis solares por R\$ 250 / metro quadrado
- Negocieei um contrato de manutenção que me custará R\$ 100 mil por ano e R\$ 10 adicionais por  
↪ metro quadrado

Qual é o custo total para o primeiro ano de operações em função do número de pés quadrados.

Solução do aluno:

Seja  $x$  o tamanho da instalação em metros quadrados.

Custos:

1. Custo do terreno:  $100x$

2. Custo do painel solar:  $250x$

3. Custo de manutenção:  $100.000 + 100x$

Custo total:  $100x + 250x + 100.000 + 100x = 450x + 100.000$

Elabore primeiro a sua própria solução para o problema. Depois, compare sua solução a solução ↪ do aluno e verifique se ela está correta.

E a resposta é:

Minha solução:

1. Custo do terreno: R\$  $100 * x$

2. Custo dos painéis solares: R\$  $250 * x$

3. Custo de manutenção: R\$  $100.000 + R\$ 10 * x$

Custo total para o primeiro ano de operações:

Custo total = Custo do terreno + Custo dos painéis solares + Custo de manutenção

Custo total = R\$  $100x + R\$ 250x + R\$ 100.000 + R\$ 10x$

Custo total = R\$  $360x + R\$ 100.000$

Agora, vamos comparar nossa solução com a solução do aluno:

Nossa solução: Custo total = R\$  $360x + R\$ 100.000$

Solução do aluno: Custo total = R\$  $450x + R\$ 100.000$

Portanto, a solução do aluno está incorreta. O custo total para o primeiro ano de operações em ↪ função do número de metros quadrados é de R\$  $360x + R\$ 100.000$ , e não de R\$  $450x + R\$ 100.000$

Neste caso, mostramos o passo a passo exato que gostaríamos que o modelo seguisse e a solução foi ainda mais apurada. Em geral, definindo a cadeia de pensamento, teremos resultados melhores. Utilizamos o método passo a passo quando não sabemos a priori os passos necessários para o modelo resolver o problema. Caso contrário, definir a cadeia é uma solução melhor.

## 10. Encadeamento (Chaining)

Outra forma de aumentarmos o tempo de processamento de uma informação pelo modelo é quebrando um prompt em prompts menores e mais específicos. Essa técnica é fundamental ao criarmos aplicações mais sofisticadas, como ChatBots que analisam dados proprietários para gerar uma resposta ao cliente.

Vamos seguir no mesmo exemplo da aula anterior. Uma forma ainda melhor de chegar a respostas estáveis para prompts como o anterior, que verifica se a resolução de um problema foi realizada corretamente, seria: - Criando um prompt para a resolução do problema em si - E depois passando a resposta obtida anteriormente para um novo prompt que compararia com a solução do aluno

Solucione a pergunta abaixo:

Pergunta:

Estou construindo uma instalação de energia solar e preciso de ajuda com as contas de custos

↪ as finanças.

- O terreno custa R\$ 100 / metro quadrado

- Posso comprar painéis solares por R\$ 250 / metro quadrado

- Negocieei um contrato de manutenção que me custará R\$ 100 mil por ano e R\$ 10 adicionais por

↪ metro quadrado

Qual é o custo total para o primeiro ano de operações em função do número de metros quadrados.

Solução:

E a resposta foi:

Vamos chamar o número de metros quadrados de "x".

O custo total será a soma dos custos do terreno, dos painéis solares e da manutenção.

Custo do terreno: R\$ 100 \* x

Custo dos painéis solares: R\$ 250 \* x

Custo da manutenção: R\$ 100.000 + R\$ 10 \* x

Custo total = Custo do terreno + Custo dos painéis solares + Custo da manutenção

Custo total = R\$ 100x + R\$ 250x + R\$ 100.000 + R\$ 10x

Custo total = R\$ 360x + R\$ 100.000

Portanto, o custo total para o primeiro ano de operações em função do número de metros

↪ quadrados será de R\$ 360x + R\$ 100.000.

E agora vamos passar a solução do modelo e compará-la ao do aluno:

Determine se a solução do aluno está correta ou não.

Pergunta:

Estou construindo uma instalação de energia solar e preciso de ajuda com as contas de custos

↪ as finanças.

- O terreno custa R\$ 100 / metro quadrado



## Engenharia de Prompts

---

- Posso comprar painéis solares por R\$ 250 / metro quadrado
- Negocieei um contrato de manutenção que me custará R\$ 100 mil por ano e R\$ 10 adicionais por  
↪ metro quadrado

Qual é o custo total para o primeiro ano de operações em função do número de pés quadrados.

Solução correta:

Vamos chamar o número de metros quadrados de "x".

O custo total será a soma dos custos do terreno, dos painéis solares e da manutenção.

Custo do terreno: R\$ 100 \* x

Custo dos painéis solares: R\$ 250 \* x

Custo da manutenção: R\$ 100.000 + R\$ 10 \* x

Custo total = Custo do terreno + Custo dos painéis solares + Custo da manutenção

Custo total = R\$ 100x + R\$ 250x + R\$ 100.000 + R\$ 10x

Custo total = R\$ 360x + R\$ 100.000

Portanto, o custo total para o primeiro ano de operações em função do número de metros

↪ quadrados será de R\$ 360x + R\$ 100.000.

Solução do aluno:

Seja x o tamanho da instalação em metros quadrados.

Custos:

1. Custo do terreno: 100x

2. Custo do painel solar: 250x

3. Custo de manutenção: 100.000 + 100x

Custo total: 100x + 250x + 100.000 + 100x = 450x + 100.000

E a resposta final é:

A solução do aluno está incorreta. O custo total calculado pelo aluno está incorreto. A forma

↪ correta de calcular o custo total para o primeiro ano de operações em função do número de

↪ metros quadrados é R\$ 360x + R\$ 100.000, como apresentado na solução correta.

E a beleza de quebrarmos em prompts mais específicos é podermos melhorar cada prompt individualmente. Digamos que eu perceba que a primeira etapa não está sendo realizada corretamente pelo modelo e as soluções não estão corretas, eu poderia iterativamente melhorar o primeiro prompt até que ele começasse a dar respostas mais estáveis. O mesmo processo eu poderia fazer com o segundo. Com duas tarefas isoladas, nós podemos entender melhor onde o modelo está errando e atacar aquela etapa individualmente, sem afetar outras onde o modelo já executa bem.

## Cadeias de Roteamento

Você já deve ter percebido o poder do encadeamento de prompts, mas podemos dar um passo a mais com o roteamento de cadeias. A ideia é criarmos um meio de campo que decide qual será o próximo

prompt utilizado, baseado na pergunta anterior. É algo como um serviço de triagem. Digamos que eu entre em um hospital com alguma doença. A primeira etapa é sempre a triagem, onde será decidido para qual especialista eu serei encaminhado. O roteamento é exatamente esta triagem. Vamos a um exemplo:

Você é responsável por categorizar as dúvidas que chegam a você nas seguintes categorias:

- ↪ matemática, física e química. Caso a dúvida não esteja relacionada a nenhuma das
- ↪ categorias anteriores, você responde "Indefinido".

Dúvida: O que é trigonometria?

Resposta: matemática

Dúvida: Quando eu uso uma crase?

Resposta: Indefinido

Dúvida: Qual é a segunda Lei de Newton?

Resposta:

física

Agora, podemos criar um script que processe a resposta do modelo e direcione a um prompt específico, dependendo da categoria.

Você é um professor de física de ensino médio. Você sempre responde as dúvidas de seus alunos

- ↪ com um embasamento teórico e um exemplo prático. Você também sempre propoe atividades
- ↪ extra classe para os alunos entenderem melhor os conceitos ensinados.

Responda a dúvida a seguir:

Dúvida: Qual é a segunda Lei de Newton?

Resposta:

A segunda Lei de Newton, também conhecida como Lei Fundamental da Dinâmica, estabelece que a

- ↪ aceleração de um objeto é diretamente proporcional à força resultante que age sobre ele e
- ↪ inversamente proporcional à sua massa, sendo a direção da aceleração a mesma que a da
- ↪ força resultante. Essa lei é matematicamente representada pela equação  $F = m \cdot a$ , onde  $F$  é a
- ↪ força resultante,  $m$  é a massa do objeto e  $a$  é a aceleração.

Um exemplo prático para compreender a segunda Lei de Newton é o seguinte: imagine um carro em

- ↪ movimento que precisa frear bruscamente para evitar um acidente. A força resultante
- ↪ atuando sobre o carro é a força de atrito dos freios, que vai gerar uma aceleração
- ↪ negativa (desaceleração) no veículo. Quanto maior **for** a força de atrito ( $F$ ) e menor a
- ↪ massa do carro ( $m$ ), maior será a aceleração ( $a$ ) gerada, fazendo com que o carro pare mais
- ↪ rapidamente.

Uma atividade extra classe que sugiro é a construção de um experimento simples para comprovar

- ↪ a segunda Lei de Newton, como por exemplo, utilizar um carrinho de brinquedo em uma
- ↪ superfície plana e lisa, variando a força aplicada e a massa do objeto para observar a
- ↪ relação entre a força, a aceleração e a massa.

Pense em diversas aplicações que poderíamos criar a partir disso. Poderíamos criar um chatbot completo para uma loja, que categoriza os setores da loja e passa para especialistas, que detêm as especificações de produtos diferentes, criando respostas especializadas. Acredito que a técnica de encadeamento seja a mais fundamental para criar aplicações reais e de valor.

## 11. Resumo e Extração

Agora que já entendemos diversas técnicas, está na hora de explorarmos algumas aplicações simples. O objetivo é dar mais ideias a vocês de formas e momentos em que vocês poderiam estar utilizando as LLMs. Os exemplos serão simples, mas espero que sejam úteis.

### Resumindo informações

A primeira aplicação seria resumir informações de um texto, como por exemplo, esta review de um cliente:

Sua tarefa é gerar um breve resumo de uma análise de produto de um site de ecommerce.

Resuma a review abaixo, delimitada por `#### em até 50 palavras.`

Review:

`####`

Dei de presente de aniversário esse panda de pelúcia da minha filha, que adora e leva para

↳ todo lugar. É macio e super fofo, e seu rosto tem uma aparência amigável. É um pouco

↳ pequeno para o que paguei. Acho que pode haver outras opções maiores pelo mesmo preço.

↳ Chegou um dia antes do esperado, então eu mesmo pude brincar com ele antes de entregá-lo a

↳ ela.

`####`

Panda de pelúcia fofo e macio, mas um pouco pequeno para o preço pago. Entrega rápida. Opções

↳ maiores pelo mesmo preço podem ser consideradas.

Atenção, os modelos de linguagem podem não obedecer bem aos limites de tamanho de saída impostos, como “em até 50 palavras” ou “em até 150 caracteres”. Eles tendem a responder melhor quando estabelecemos o limite de parágrafos ou frases!

### Resumindo informações com foco específico

Podemos resumir informações com um viés específico que nos importa mais. Por exemplo, se trabalhássemos no setor de entregas de uma empresa, poderíamos fazer um resumo das reviews com foco nas entregas.

Sua tarefa é gerar um breve resumo de uma review de produto de um site de ecommerce. Você quer

↳ dar um feedback para o setor de logística, portanto, foque nos aspectos de tempo e

↳ qualidade de entrega.

Resuma a review abaixo, delimitada por `#### em até 15 palavras.`

Review:

`####`

Dei de presente de aniversário esse panda de pelúcia da minha filha, que adora e leva para  
↳ todo lugar. É macio e super fofo, e seu rosto tem uma aparência amigável. É um pouco  
↳ pequeno para o que paguei. Acho que pode haver outras opções maiores pelo mesmo preço.  
↳ Chegou um dia antes do esperado, então eu mesmo pude brincar com ele antes de entregá-lo a  
↳ ela.  
####

E vocês podem notar que na resposta o primeiro aspecto ressaltado foi a entrega:

Entrega um dia antes do previsto, produto fofo e macio, porém um pouco pequeno.

### Extraíndo Informações

Poderíamos também apenas extrair da mensagem a informação que nos é valiosa. No mesmo exemplo anterior, poderíamos solicitar o seguinte:

Sua tarefa é extrair de reviews de produtos as informações relevantes ao departamento de  
↳ logística e entregas de um ecommerce .

Da review abaixo, delimitada por ####, *extraia apenas as informações referentes ao envio e*  
↳ *entrega. Limite a 30 palavras.*

Review:

####

Dei de presente de aniversário esse panda de pelúcia da minha filha, que adora e leva para  
↳ todo lugar. É macio e super fofo, e seu rosto tem uma aparência amigável. É um pouco  
↳ pequeno para o que paguei. Acho que pode haver outras opções maiores pelo mesmo preço.  
↳ Chegou um dia antes do esperado, então eu mesmo pude brincar com ele antes de entregá-lo a  
↳ ela.  
####

E a resposta é:

Chegou um dia antes do esperado.

## 12. Inferência e Análise de Sentimento

Desenvolver modelos para análise de sentimentos pode ser uma tarefa complexa no mundo da Ciência de Dados tradicional. Envolveria reunir um conjunto de dados grande, classificá-lo, treinar o modelo, validá-lo e fazer o processo iterativo de melhoria. É impressionante o quão facilitada ficou esta atividade com modelos LLM. Por disporem de uma inteligência capaz de inferir sentimentos e emoções em textos, com simples instruções conseguimos realizar ações bem interessantes. Vamos a alguns exemplos.

### Inferindo sentimentos

Qual é o sentimento da seguinte avaliação de produto, delimitada por #### ?

####

Precisava de uma boa luminária para o meu quarto, e esta tinha armazenamento adicional e um  
↳ preço não muito alto. Recebi rapidamente. A corda da nossa luminária quebrou durante o  
↳ transporte e a empresa prontamente enviou uma nova. Chegou em poucos dias também. Foi  
↳ fácil de montar. Tive uma peça faltando, então entrei em contato com o suporte deles e  
↳ eles rapidamente me enviaram a peça que faltava! A Lumina me parece ser uma ótima empresa  
↳ que se preocupa com seus clientes e produtos!!

####

E temos a seguinte resposta do modelo:

Satisfação e confiança.

Muito bem, aparentemente ele conseguiu inferir a partir do texto o sentimento do cliente. Vamos estruturar a saída para termos um exemplo mais real de análise de sentimentos.

Qual é o sentimento da seguinte avaliação de produto, delimitada por #### ?

Forneça sua resposta em uma única palavra, seja "positiva" ou "negativa".

####

Precisava de uma boa luminária para o meu quarto, e esta tinha armazenamento adicional e um  
↳ preço não muito alto. Recebi rapidamente. A corda da nossa luminária quebrou durante o  
↳ transporte e a empresa prontamente enviou uma nova. Chegou em poucos dias também. Foi  
↳ fácil de montar. Tive uma peça faltando, então entrei em contato com o suporte deles e  
↳ eles rapidamente me enviaram a peça que faltava! A Lumina me parece ser uma ótima empresa  
↳ que se preocupa com seus clientes e produtos!!

####

A resposta:

positiva

Em outro exemplo, solicitamos ao modelo que ele faça uma inferência das emoções expressadas na avaliação:

Identifique uma lista de emoções que o autor da seguinte análise está expressando. Inclua não  
→ mais que cinco itens na lista. Formate sua resposta como uma lista de palavras em  
→ minúsculas separadas por vírgulas.

####

Precisava de uma boa luminária para o meu quarto, e esta tinha armazenamento adicional e um  
→ preço não muito alto. Recebi rapidamente. A corda da nossa luminária quebrou durante o  
→ transporte e a empresa prontamente enviou uma nova. Chegou em poucos dias também. Foi  
→ fácil de montar. Tive uma peça faltando, então entrei em contato com o suporte deles e  
→ eles rapidamente me enviaram a peça que faltava! A Lumina me parece ser uma ótima empresa  
→ que se preocupa com seus clientes e produtos!!

####

A resposta é:

agradecimento, satisfação, confiança, felicidade, gratidão.

Um ótimo trabalho feito pelo modelo.

Agora vamos a um exemplo mais completo:

Identifique os seguintes itens a partir do texto de avaliação:

- Sentimento (positivo ou negativo)
- O avaliador está expressando raiva? (verdadeiro ou falso)
- Item comprado pelo avaliador
- Empresa que fabricou o item

A avaliação é delimitada por ####

Formate sua resposta como um objeto JSON com

"Sentimento", "Raiva", "Item" e "Marca" como as chaves.

Se a informação não estiver presente, use "unknown"  
como o valor.

Faça sua resposta o mais curta possível.

Formate o valor de Anger como booleano.

####

Precisava de uma boa luminária para o meu quarto, e esta tinha armazenamento adicional e um  
→ preço não muito alto. Recebi rapidamente. A corda da nossa luminária quebrou durante o  
→ transporte e a empresa prontamente enviou uma nova. Chegou em poucos dias também. Foi  
→ fácil de montar. Tive uma peça faltando, então entrei em contato com o suporte deles e  
→ eles rapidamente me enviaram a peça que faltava! A Lumina me parece ser uma ótima empresa  
→ que se preocupa com seus clientes e produtos!!

####

E a resposta:

```
{  
  "Sentimento": "positivo",  
  "Raiva": false,  
  "Item": "luminária",  
  "Marca": "Lumina"  
}
```

### Inferindo tópicos

Além de sentimentos, poderíamos inferir os principais tópicos de um texto. Como no exemplo a seguir:

Determine cinco tópicos que estão sendo discutidos no texto a seguir, que é delimitado por  
↪ #####

Faça cada item ter uma ou duas palavras.

Formate sua resposta como uma lista de itens separados por vírgulas.

#####

Em uma pesquisa recente conduzida pelo governo, os funcionários do setor público foram  
↪ solicitados a avaliar seu nível de satisfação com o departamento em que trabalham. Os  
↪ resultados revelaram que a NASA foi o departamento mais popular, com uma taxa de  
↪ satisfação de 95%.

Um funcionário da NASA, John Smith, comentou sobre os resultados, afirmando: "Não estou  
↪ surpreso que a NASA tenha se destacado. É um ótimo lugar para trabalhar, com pessoas  
↪ incríveis e oportunidades incríveis. Tenho orgulho de fazer parte de uma organização tão  
↪ inovadora."

Os resultados também foram bem recebidos pela equipe de gerenciamento da NASA, com o diretor  
↪ Tom Johnson afirmando: "Estamos emocionados ao saber que nossos funcionários estão  
↪ satisfeitos com seu trabalho na NASA. Temos uma equipe talentosa e dedicada que trabalha  
↪ incansavelmente para alcançar nossos objetivos, e é fantástico ver que seu trabalho árduo  
↪ está dando frutos."

A pesquisa também revelou que a Administração da Previdência Social teve a menor pontuação de  
↪ satisfação, com apenas 45% dos funcionários indicando que estavam satisfeitos com seus  
↪ empregos. O governo se comprometeu a abordar as preocupações levantadas pelos funcionários  
↪ na pesquisa e trabalhar para melhorar a satisfação no trabalho em todos os departamentos

#####

Resposta:

Pesquisa governamental, Satisfação no trabalho, NASA, Administração da Previdência Social,  
↪ Melhoria contínua

## 13. Transformando, traduzindo e modificando

Modelos de LLM são muito bons em transformar textos. Quando oferecemos uma entrada e pedimos simplesmente que se modifique um formato, por exemplo, transformar uma transcrição em um e-mail, ou o texto de um livro em um artigo de blog, ou traduzir um texto de uma língua para outra, os modelos costumam não errar. Vamos explorar um pouco dessas capacidades.

### Tradução

Traduza o seguinte texto do Português para Espanhol

```
####  
Olá, gostaria de comprar um liquidificador.  
####
```

Resposta:

Hola, me gustaría comprar una licuadora.

Traduza o seguinte texto do Inglês para o Português como falado por uma criança de 5 anos de idade

```
####  
During our exciting trip to the forest, we explored lush trails and marveled at the diversity  
↳ of wildlife. Camping under the stars in the dense forest provided a unique experience,  
↳ where every sound of nature became an unforgettable soundtrack.  
####
```

Resposta:

Durante a nossa viagem muito legal para a floresta, a gente andou por caminhos bonitos e ficou  
↳ impressionado com todos os animais diferentes. Dormir em barracas embaixo das estrelas na  
↳ floresta cheia de árvores foi super legal, e cada barulhinho dos bichos virou uma música  
↳ que a gente nunca vai esquecer.

### Modificação de Tom

Transforme o seguinte do texto para uma carta de negócios:

```
####  
Cara, aqui é o Adri, dá uma olhada nessas especificações dessa luminária de pé  
####
```

Resposta:

Prezado/a [Nome],

Espero que esta mensagem o/a encontre bem. Meu nome é Adri e gostaria de chamar sua atenção  
↳ para as especificações da luminária de pé em questão.



Agradeço desde já pela sua atenção e aguardo o seu retorno.

Atenciosamente,  
Adri

Transforme o seguinte do calão para uma carta de negócios:

### Conversão de Formatos

Traduza o seguinte dicionário Python de JSON para uma tabela HTML com cabeçalhos de coluna e  
↔ título

```
####  
data_json = { "resturant employees" : [  
    {"name":"Shyam", "email":"shyamjaiswal@gmail.com"},  
    {"name":"Bob", "email":"bob32@gmail.com"},  
    {"name":"Jai", "email":"jai87@gmail.com"}  
]}  
####
```

Resposta:

```
<!DOCTYPE html>  
<html lang="en">  
<head>  
    <meta charset="UTF-8">  
    <meta name="viewport" content="width=device-width, initial-scale=1.0">  
    <title>Lista de Funcionários do Restaurante</title>  
    <style>  
        table {  
            border-collapse: collapse;  
            width: 100%;  
        }  
  
        th, td {  
            border: 1px solid #dddddd;  
            text-align: left;  
            padding: 8px;  
        }  
  
        th {  
            background-color: #f2f2f2;  
        }  
    </style>  
</head>  
<body>  
  
    <h2>Lista de Funcionários do Restaurante</h2>  
  
    <table>  
        <tr>
```

```
<th>Nome</th>
<th>Email</th>
</tr>
<tr>
  <td>Shyam</td>
  <td>shyamjaiswal@gmail.com</td>
</tr>
<tr>
  <td>Bob</td>
  <td>bob32@gmail.com</td>
</tr>
<tr>
  <td>Jai</td>
  <td>jai87@gmail.com</td>
</tr>
</table>

</body>
</html>
```

### Correção ortográfica

Corrija os erros ortográficos do seguinte texto:

####

Num lembro bem o dia, mas lembro qui tava cum sol quente. Eu tava na porta di casa, olhando  
↪ pro céu azulzim, quandu de repente, vi uma coisa brilhante lá longe. Primeiro, pensei qui  
↪ era um passarinho, mas foi se aproximando e fui vendo qui era grandão demais. Era um  
↪ avião, daqueles qui a genti só vê em filme.

####

Resposta:

Não lembro bem o dia, mas lembro que estava com sol quente. Eu estava na porta de casa,  
↪ olhando para o céu azulzinho, quando de repente, vi uma coisa brilhante lá longe.  
↪ Primeiro, pensei que era um passarinho, mas foi se aproximando e fui vendo que era grandão  
↪ demais. Era um avião, daqueles que a gente só vê em filme.

## 14. Criação de Chatbots Personalizados

O poder das LLM aumenta quando as utilizamos para processar dados proprietários, de forma que elas tomem decisões automáticas e personalizadas para a nossa situação. No momento em que criamos um ChatBot de atendimento para uma empresa, é necessário que a LLM tenha acesso às informações necessárias para resolver as dúvidas dos nossos clientes.

Vamos dar um exemplo simples de como isso poderia ser feito, mas lembrando que este é um exemplo que provavelmente não será utilizado, pois já existem técnicas mais avançadas para a utilização de informações proprietárias, como RAG (Retrieved Augmented Generation), para casos como este. De qualquer forma, a base é esta: passar informações relevantes junto ao prompt.

O JSON delimitado por `####` apresenta o nome e informações dos produtos de uma loja.

```
####
{
  "produtos": [
    {
      "nome": "Geladeira Básica CoolStay",
      "descricao": "Modelo compacto e eficiente, com capacidade de 240 litros, ideal para
        ↳ pequenos espaços e solteiros. Conta com prateleiras removíveis e controle básico de
        ↳ temperatura.",
      "id_produto": "GBS-001",
      "preco": 1050.00
    },
    {
      "nome": "Geladeira Duplex FreshSpace",
      "descricao": "Geladeira duplex de 340 litros, com sistema de refrigeração avançado que
        ↳ mantém seus alimentos frescos por mais tempo. Design moderno e função de
        ↳ congelamento rápido.",
      "id_produto": "GDF-002",
      "preco": 1650.00
    },
    {
      "nome": "Geladeira Inverter EconoMax",
      "descricao": "Eficiência energética com tecnologia inverter, 400 litros de capacidade e
        ↳ sistema de controle de umidade para frutas e legumes. Silenciosa e econômica.",
      "id_produto": "GIE-003",
      "preco": 2500.00
    },
    {
      "nome": "Geladeira Smart Home Connect",
      "descricao": "Conectividade total com seu smartphone, capacidade de 480 litros, tela
        ↳ sensível ao toque com calendário de validade dos alimentos e receitas integradas. O
        ↳ futuro na cozinha moderna.",
      "id_produto": "GSH-004",
      "preco": 3200.00
    },
    {
      "nome": "Geladeira French Door Gourmet",
```

```
"descricao": "Capacidade de 550 litros, design sofisticado com French doors,  
  ↳ compartimento gourmet com ajustes de temperatura específicos para vinhos e queijos,  
  ↳ além de gavetas especiais para carnes e vegetais.",  
"id_produto": "GFD-005",  
"preco": 4900.00  
}  
]  
}  
####
```

Baseado nas informações do JSON informado, responda as dúvidas seguintes. Caso você não tenha  
↳ informações suficientes, avise que você não possui informações suficientes para responder:

Pergunta: Qual é a geladeira mais econômica energeticamente?

Resposta:

Resposta:

A geladeira mais econômica energeticamente é a "Geladeira Inverter EconoMax" (GIE-003).

Pergunta: E quanto ela custa?

Resposta:

A Geladeira Inverter EconoMax (GIE-003) custa 2500.00.

## **15. Recursos adicionais**

Dataset de personas do Hugging Face: <https://huggingface.co/datasets/fka/awesome-chatgpt-prompts>

Prompt engineering guide: <https://promptingguide.ai/> Guia de Engenharia de Prompt da Openai:

<https://platform.openai.com/docs/guides/prompt-engineering>