

Capítulo 7

ChatGPT, MariTalk e outros agentes de conversação

Um retrato de 2023, com retoques em 2026

Aline Paes
Cláudia Freitas

Publicado em: 26/09/2023
Atualizado em: 16/04/2026

Embora este capítulo tenha sido escrito em 2023 com a intenção de ser melhorado e aprofundado no ano seguinte, isto nunca foi feito. Porém, e por curiosidade, repetimos os *prompts* aos agentes em 2025, e os resultados estão bem melhores, sobretudo na Maritalk. Ainda em 2025, os resultados sobre o livro *A Bolsa Amarela* continuavam errados; em 2026, finalmente, o ChatGPT forneceu a resposta correta.

7.1 Introdução

ChatGPT¹ e Maritalk² (e similares, como Bard³, Vicuna⁴, Claude⁵, entre tantos outros) são exemplos de aplicações de agentes de conversação (*chatbots*) baseados em modelos de linguagem gerativos (ou generativos). Mas o que significa isso?⁶

Alguns autores, como Jurafsky e Martin (Jurafsky; Martin, 2023), usam o termo “agente de conversação” para definir qualquer sistema de diálogo que se comunique com usuários usando a linguagem humana e os dividem em duas classes: agentes orientados a tarefas, em que o diálogo é para resolver um problema específico, como agendar uma viagem ou resolver um problema bancário, enquanto *chatbots* seriam agentes de conversação que tentam simular diálogos humanos, mais voltados para entretenimento. Ferramentas como ChatGPT se enquadram mais no segundo caso, entretanto também podem ser embutidos em outras ferramentas aumentadas para atuar como no primeiro caso. Neste capítulo, os termos “*chatbots*” e “agentes de conversação” serão usados de forma intercambiável.

Agentes de conversação não são novidade – ELIZA, criada em 1966 pelo cientista da computação Joseph Weizenbaum, era um agente de conversação que replicava o comportamento de um psicoterapeuta. ELIZA era simples, baseada em *templates* (padrões de conversa pré-construídos), mas conseguia manter longas conversas buscando por determinadas palavras-chave nas falas (escritas por texto) de uma pessoa. Se uma palavra-chave fosse encontrada, uma regra seria aplicada para transformar sua entrada e criar a resposta. Na Figura 7.1, transcrevemos quatro interações com a ELIZA (retiradas de (Jackson; Moulinier, 2002)).

Nos trechos, temos exemplos de padrões pré-fabricados que retomam elementos da pessoa que fala, como “Por que você diz [...]”, “Te agrada pensar que [...]”, “O que te faz pensar que [...]”, e “Por quanto tempo você não tem [...]”. Vemos também que, no último trecho, esta retomada falha, pois “por quanto tempo você não tem estado” é algo que não faz sentido, embora seja gramaticalmente correto. Se ao invés de “não, não estou”, a pessoa tivesse dito “não, não comi”, a resposta de ELIZA seria “Por quanto tempo você não tem comido?”, e a conversa poderia continuar. ELIZA também foi programada para fazer uso de palavras genéricas, como “interessante”, para quando não há pistas sobre como continuar a conversa. A palavra “alergia” não é elemento capaz de evocar novas interações, e o mesmo acontece com “manchas vermelhas”.

¹<https://chat.openai.com/>

²<https://chat.maritaca.ai/>

³<https://bard.google.com/>

⁴<https://lmsys.org/blog/2023-03-30-vicuna/>

⁵<https://www.anthropic.com/index/introducing-claude>

⁶Assista neste [vídeo](#) a apresentação de uma análise crítica do ChatGPT.



Figura 7.1: Exemplo de interações com a agente de conversação ELIZA, criada em 1966.

ELIZA: como vai você hoje?	P: você é um programa, não é?
P: não muito bem	ELIZA: te agrada pensar que
ELIZA: por que você diz não	eu sou um programa?
muito bem?	P: com certeza
P: eu estou com alergia	ELIZA: eu percebo
ELIZA: interessante	P: ok
P: eu fico com manchas	
vermelhas	
ELIZA: acho que não sei do que	
você está falando	
	ELIZA: você está sendo um
P: em minhas fantasias você é	pouco negativo
loira	P: não, não estou
ELIZA: o que te faz pensar que	ELIZA: por quanto tempo você
sou loira?	não tem estado?

De 1966 para cá muita coisa mudou, e podemos começar perguntando ao ChatGPT o que é um GPT, que é parte importante do seu nome. O trecho abaixo foi traduzido de uma interação que não é nossa, feita originalmente em inglês – mas os grifos são nossos⁷.

Quadro 7.1: Exemplo de interação com o ChatGPT, em 2023

P: Resumidamente, o que é um GPT?

R: É um tipo de algoritmo de aprendizado profundo projetado para gerar textos semelhantes aos produzidos por humanos. [...] Uma vez treinados, os modelos GPT podem ser usados para gerar um novo texto que seja coerente e **siga os padrões e estilos dos dados de treinamento originais**.

P: Seria correto, então, dizer que os modelos GPT, como você, usam uma abordagem probabilística para **gerar textos que se parecem com outros textos que já foram escritos**?

R: Sim, essa é uma afirmação correta. [...] Quando recebe um *prompt*, o modelo gera texto prevendo a palavra ou frase mais provável de aparecer com base nos padrões que aprendeu com os dados de treinamento.

P: Quando peço que você gere um novo texto, você mesmo tem as ideias ou apenas reembala trabalhos escritos anteriormente?

R: [...] Eu não tenho as ideias sozinho. Em vez disso, eu gero texto com base em padrões e relacionamentos em meus dados de treinamento. Isto significa que **estou essencialmente reembalando trabalhos escritos anteriormente**, com base no que acredito ser a continuação mais provável do texto de entrada fornecido.

P: Existe algum filtro de correção aplicado à saída?

R: Como modelo de linguagem de IA, **não tenho um filtro integrado para determinar a correção da minha saída**.

O que podemos tirar daí?

Como dissemos, ChatGPT e similares são agentes de conversação baseados em modelos de linguagem gerativos. Como visto no Capítulo [Modelos de linguagem](#), este nome se refere a algoritmos que são bons em encadear palavras de maneira fluente, mas são **baseados em previsões e probabilidade**. São geradores de texto otimizados para parecerem plausíveis. Outro ponto importante mencionado na explicação fornecida pela própria ferramenta é que não há criatividade propriamente, apenas uma reembalagem do que já foi

⁷A interação está relatada [neste artigo](#).



dito. Por fim, vemos que não há qualquer garantia de que os textos gerados contenham informação correta (e nem há responsabilidade sobre isso).

Uma consequência da forma pela qual essas ferramentas são feitas (baseadas em previsão) é que nem sempre a previsão está condizente com a realidade, o que tem sido chamado de **alucinação**. Este fenômeno acontece quando um modelo de linguagem gera um texto que pode estar correto gramaticalmente, ter fluência e alguma coerência semântica, mas que não reflete a realidade e, portanto, não faz sentido (Ji et al., 2023). O termo é emprestado da psicologia, que o define como “uma percepção, experimentada por uma pessoa acordada, na ausência de um estímulo apropriado do mundo extracorpóreo” (Blom, 2010), ou seja, algo que parece real, mas não é.

Embora algumas destas ferramentas possam ser aumentadas com técnicas de recuperação de informação (ver Capítulo **Geração Aumentada por Recuperação (RAG)**), este não é o caso dos modelos mais básicos atuais, sem acesso externo. Por outro lado, chatbots comerciais e os embutidos nos navegadores são estendidos com mecanismos de busca e ferramentas de recuperação de informação externa. De todo modo, a maioria dessas ferramentas não funciona da mesma forma que uma máquina de busca ou um banco de dados, ou mesmo um repositório de perguntas e respostas. Entretanto, as respostas retornadas por elas mostram fluência e podem fazer algum sentido – embora não possamos desconsiderar o fenômeno cognitivo da apofenia (Fyfe et al., 2008)⁸, que diz respeito à identificação de padrões ou associações em conjuntos de dados aleatórios⁹. E este é o perigo: as alucinações, não à toa, frequentemente não parecem alucinações, e soam como verdades. Já vimos, por exemplo, que, ao pedir uma lista de referências bibliográficas sobre um determinado assunto, são geradas referências completas, com indicação de autoria, título, revista, volume, ano, que simplesmente não existem. Isto acontece porque os textos são gerados levando em conta a probabilidade daquilo ser uma resposta correta, isto é, as respostas são elaboradas de maneira a se parecerem o máximo possível com uma resposta correta. Pode ser difícil distinguir as respostas – ou, mais precisamente, as sequências de palavras – que estão ancoradas na realidade daquelas que apenas parecem estar ancoradas na realidade. Assim, uma das limitações deste tipo de ferramenta é a incapacidade de dizer “não sei”.

O desenvolvimento de maneiras de evitar as alucinações é uma das preocupações desde 2023. Algumas estratégias têm sido discutidas livremente e também investigadas na academia e na indústria. Do ponto de vista do usuário final, aquela pessoa que vai abrir o ChatGPT no navegador e interagir com ele por meio de textos, uma das alternativas é a engenharia de *prompts*¹⁰. Neste caso, o usuário pode tentar continuar a conversa com a ferramenta, calibrando e alinhando as respostas anteriores com novas perguntas. Outra possibilidade é usar a ferramenta por meio da sua API (*Application Programming Interface*), ou seja, quando o ChatGPT é invocado e controlado por meio de código, ao invés de ser usado diretamente no navegador. Neste caso, é possível controlar o parâmetro de temperatura, usado para calibrar a distribuição de probabilidade, de modo que o *chatbot* se atenha mais ao que foi aprendido anteriormente, ou gere respostas um pouco menos prováveis, ou mais “criativas”. Outra possibilidade é acoplar bases de conhecimento externas ao processo de geração de texto (Lewis et al., 2020b) ou interagir com o *chatbot* por meio de perguntas que demandem alguma tentativa de simulação de raciocínio, ou embutir o modelo de habilidades de explicação das suas respostas, um processo chamado de *chain-of-thought* (Kojima et al., 2022). Entretanto, até agora, nenhuma das opções mencionadas conseguiu remover por completo as alucinações dos *chatbots* baseados em modelos de linguagem (Zhang et al., 2025d).

No contexto empresarial, Hannigan et al. (2024) sugerem levar em conta duas dimensões para orientar o uso de chatbots com o objetivo de mitigar a utilização humana de conteúdo falso gerado por LLMs (que os autores chamam de “botshit”):

- Importância da veracidade das respostas (qual é a importância da veracidade da resposta do chatbot para a tarefa?)

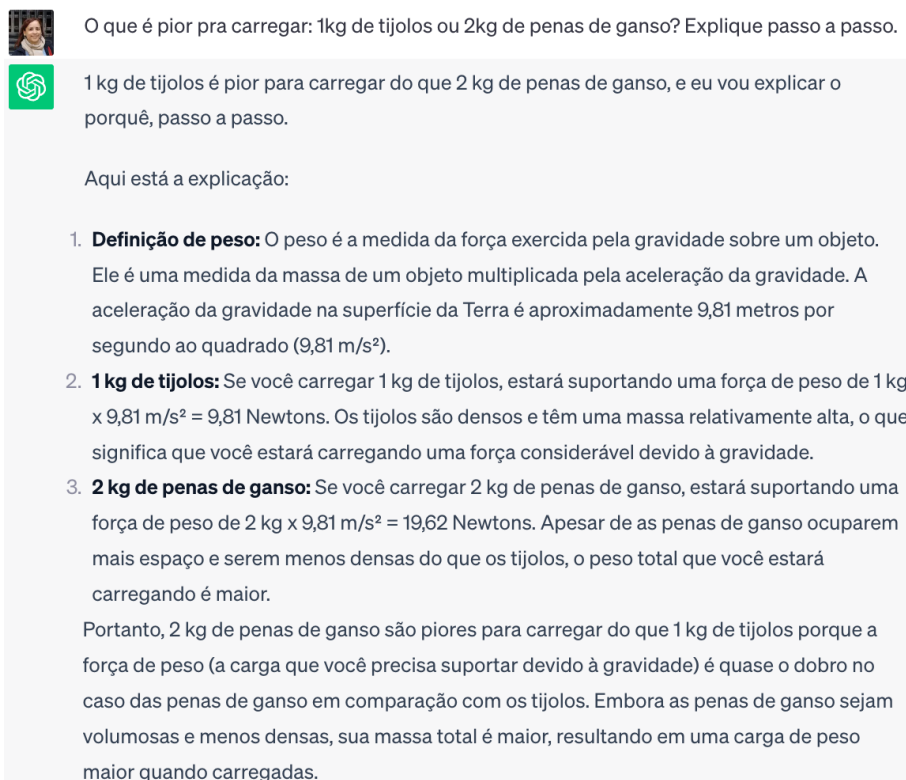
⁸Conforme o Dicionário Online da Língua Portuguesa (<https://www.dicio.com.br/>), apofenia é o “Fenômeno cognitivo no qual os indivíduos têm a tendência de formar ou reconhecer conexões a partir de dados aleatórios, estabelecendo conclusões a partir de dados inconclusivos. Etimologia: do alemão Apophänie, termo criado pelo neurologista alemão Klaus Conrad.

⁹<https://www1.folha.uol.com.br/tec/2023/07/ferramentas-como-chatgpt-so-existem-porque-humanos-veem-sentido-a-partir-de-qualquer-coisa.shtml>

¹⁰Um *prompt* é um texto em linguagem humana (em oposição à linguagem de programação) que dá ao *chatbot* uma instrução do que ele deve fazer. Um *prompt* pode ser formulado como uma pergunta, uma observação, um questionamento, ou ainda representar uma tarefa específica, por exemplo, para classificar um texto. *Prompts* podem ter um formato livre, mas *chatbots* são bastante sensíveis ao conteúdo textual dele, o que tem motivado a criação de modelos e padrões para a sua escrita. Falamos um pouco sobre o assunto no Capítulo **Modelos de linguagem**.



Figura 7.2: Resposta do ChatGPT, em 2023, para o seguinte *prompt* “O que é pior pra carregar: 1kg de tijolos ou 2kg de penas de ganso? Explique passo a passo.”



- Verificabilidade da veracidade das respostas (quão fácil é verificar a veracidade da resposta do chatbot?)

Quanto à importância da veracidade, diferentes tipos de trabalho podem ter expectativas distintas. Quando a gravidade do risco de usar conteúdo de chatbot é muito alta (em situações com baixa ou nenhuma tolerância a falhas, como manutenção de equipamentos, bem-estar de pacientes e segurança dos funcionários) a veracidade da resposta do chatbot é crucial. Ainda conforme Hannigan et al. (2024), nesses casos os riscos relacionados ao “botshit” refletem o conceito de responsabilização algorítmica nas organizações. Por outro lado, quando a gravidade do risco é baixa, como na busca por ideias ou sugestões (“brainstorming” ou “botstorming”), a veracidade da resposta do chatbot é irrelevante.

Quanto ao segundo ponto, Hannigan et al. (2024) comparam o conteúdo ilusório fácil de verificar a “um trem desgovernado que pode ser ouvido a quilômetros de distância”. Nesse caso, a escolha pela utilização de um chatbot deve levar em consideração o quão simples ou complexo, barato ou caro, rápido ou lento é revisar, descobrir e evitar que o conteúdo equivocado de uma resposta do chatbot seja usado e transformado em “botshit”. As respostas serão fáceis de verificar se existir um conjunto relativamente estável e acessível de afirmações verdadeiras em torno do conteúdo da resposta. Por exemplo, um *prompt* solicitando a definição de uma palavra pode ser facilmente verificável usando um dicionário.

7.2 Jogos de Linguagem

Existem algumas maneiras de entender “linguagem”, e uma delas defende que aquilo que chamamos de linguagem é um conjunto de práticas relacionadas, assim como os *jogos*, que incluem diferentes práticas/atividades, apesar de serem chamados “jogos” (jogos de cartas, jogos de bola, de tabuleiro, de esconder, de adivinhação, paciência (que se joga sozinho), frescobol (em que não há vencedor) etc). No livro *Investigações Filosóficas* (1953), o filósofo da linguagem L. Wittgenstein lista alguns exemplos de jogos de linguagem, e os reconhecemos como “tarefas do PLN”:

- descrever um objeto,



- produzir um objeto segundo uma descrição (desenho),
- contar uma anedota,
- traduzir um texto,
- inventar uma história,
- dar um comando, e agir segundo comandos,
- relatar um acontecimento,
- conjecturar sobre o acontecimento,
- expor uma hipótese e prová-la,
- resolver um exemplo de cálculo aplicado,
- desenhar um objeto a partir de uma instrução verbal,
- apresentar resultados de um experimento por meio de tabelas e diagramas,
- pedir, agradecer, maldizer, saudar, orar.

Podemos juntar à lista mais alguns jogos “jogados no PLN”: - encontrar informações em um texto para responder certas perguntas, - analisar sintaticamente uma frase, - dar a definição de uma palavra, - produzir inferências, - explicar suas próprias decisões, - prever a próxima palavra em uma frase.

Desse ângulo, podemos imaginar que os modelos de linguagem de que dispomos e que servem de base para agentes de conversação, como ChatGPT e Maritalk, são muito bons em algumas dessas práticas – ou “jogos de linguagem” –, mas não em todas. Ou seja, são modelos que jogam mais ou menos bem alguns jogos, como “inventar uma história”, “resumir”, “escrever um email”, “traduzir”, “prever a próxima palavra” etc., mas não jogam tão bem outros, como “fazer cálculos” ou “fornecer explicações”.

Este desempenho tem a ver com a forma como os modelos foram construídos, baseada em previsão: nem todos os jogos de linguagem, ou nem todas as atividades linguísticas em que participamos, se resumem a um jogo de previsões, embora um bom desempenho no jogo das previsões leve a um bom resultado em uma série de outros jogos. Do mesmo modo, correr bem ajuda a jogar pique-bandeira, polícia e ladrão, pique-esconde, futebol, e uma série de outros jogos. Mas não necessariamente ajuda a jogar paciência ou jogo da velha. Agentes de conversação se beneficiam especialmente de bom desempenho na previsão de palavras.

Uma das razões pelas quais estes agentes de conversação se tornaram tão populares é que, com eles, qualquer pessoa pode interagir com as máquinas usando sua própria língua¹¹, e não em uma linguagem de programação. Com isso, qualquer pessoa pode pedir que máquinas executem certas tarefas, que podem ir desde a criação de um programa de computador (códigos) até sugestões de receitas a partir de uma lista de ingredientes que temos na geladeira. Mas não há garantia de que funcionem, ou de que a receita ficará boa.

Embora as saídas dos agentes de conversação possam muitas vezes nos surpreender, ainda é difícil afirmar que eles resolvem tarefas de PLN muito bem, ou que o desempenho deles supera o desempenho humano em alguma tarefa. A geração de textos, tarefa-base de tais agentes, ainda é de difícil avaliação, tanto automática como humana. As métricas automáticas, como ROUGE¹² (Lin, 2004), BLEU¹³ (Papineni et al., 2002), BERTscore¹⁴ (Zhang et al., 2020), METEOR¹⁵ (Banerjee; Lavie, 2005), entre outras, ainda apresentam diversas limitações (Sai et al., 2023). Especialistas humanos, por sua vez, podem conseguir avaliar muito bem as respostas, mas esta ainda é uma tarefa cansativa e propensa a ruídos. Por outro lado, não é trivial criar conjuntos de dados que explorem todas as características que gostaríamos de avaliar em um sistema de geração de textos, o que inclui não apenas aspectos gramaticais e semânticos, mas também criatividade, fluência, interesse e prazer despertado no leitor, dentre tantos outros.

Nas seções seguintes, mostraremos tarefas (ou jogos) que os agentes parecem jogar bem e tarefas que os agentes jogam mal. Os exemplos serão obtidos em sua maioria do ChatGPT¹⁶, o agente de conversação mais popular até o momento. Também incluímos, em alguns casos, exemplos de outros dois agentes: a MariTalk¹⁷, uma agente de conversação construída a partir do modelo de linguagem Sabiá (Pires et al., 2023), treinado de forma continuada a partir do GPT com textos em português.

¹¹Desde que a sua língua seja contemplada com recursos suficientes, isto é, textos, para garantir um bom treinamento.

¹²Abreviação para *Recall-Oriented Understudy for Gisting Evaluation*. Esta é uma métrica utilizada para avaliar sumarizações, e, de maneira bastante simplificada, consiste em comparar a interseção de n-gramas entre textos sumarizados e referências

¹³*Bilingual evaluation understudy*, usada para comparar traduções automáticas e suas referências. Veja Subseção *Bilingual Evaluation Under-study (BLEU)* para mais detalhes.

¹⁴Utiliza embeddings contextualizados para comparar os textos gerados e referências

¹⁵*Metric for Evaluation of Translation with Explicit ORdering*, utilizada para tradução, sumarização etc.

¹⁶<https://chat.openai.com/>

¹⁷<https://chat.maritaca.ai/>



Além dos agentes de conversação cujo objetivo principal é a interação por meio de diálogo, também existem inúmeras ferramentas baseadas em modelos de linguagem para propósito específico, como auxiliar em revisões da literatura¹⁸, auxiliar na escrita de código¹⁹, escrita de e-mails²⁰, revisão de texto²¹, entre outras. Não trataremos destes casos aqui.

7.3 Jogos que os agentes parecem jogar bem

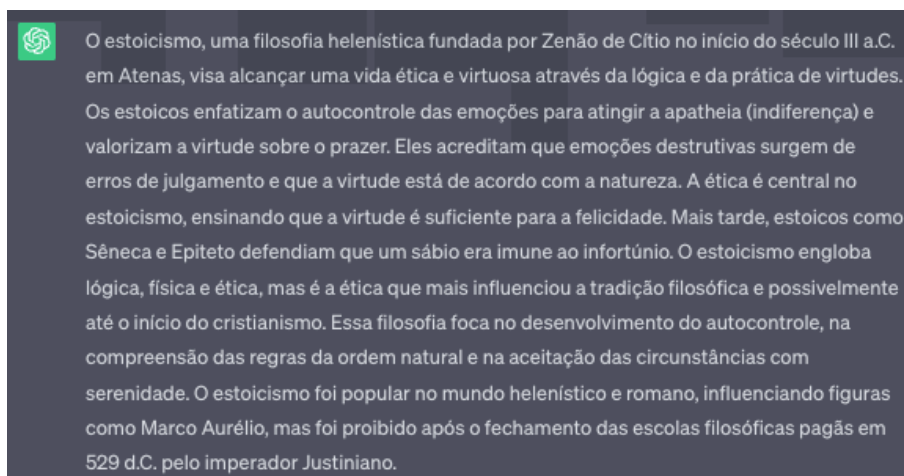
7.3.1 Sumarização

Gerar resumos é uma tarefa em PLN chamada de **sumarização textual**, que consiste em gerar um texto mais curto que o original e que ainda seja fluente e fiel ao texto-fonte. A sumarização pode ser abstrativa ou extrativa. A sumarização abstrativa consiste em gerar um resumo com as próprias palavras do escritor, enquanto a sumarização extrativa consiste em extrair sentenças inteiras do texto que, quando juntas, formem um resumo. Os *chatbots* têm mostrado em vários exemplos que conseguem resumir bem os textos. Entretanto, se o texto for muito longo, a maioria dos agentes mais populares esbarra em um problema computacional: ainda é difícil para modelos baseados em Transformers (Bertsch et al., 2023; Vaswani et al., 2017a)²² receber como entrada textos muito longos (veja mais no Capítulo [Modelos de linguagem](#)). Por exemplo, ao pedirmos que o ChatGPT resumisse o texto sobre Estoicismo²³ com o *prompt*

- P. Resuma o texto que segue, de forma extrativa, em até 200 palavras

a resposta era que o texto era muito longo. Entretanto, ao usar o mesmo *prompt*, copiando e colando a página da Wikipedia sobre Estoicismo²⁴, obtivemos a resposta na Figura 7.3. Embora seja de fato um resumo extrativo, podemos observar que a primeira parte do texto foi a que mais recebeu atenção. Outro fator que pode ter contribuído para a preferência ao início do texto é o pedido que o texto final esteja contido em 200 palavras, uma motivação para que a resposta não ficasse enorme²⁵.

Figura 7.3: Resposta do ChatGPT, em 2023, ao ser deparado com o seguinte *prompt* “Resuma o texto que segue, de forma extrativa:”, seguido pela cópia do texto da Wikipedia (<https://pt.wikipedia.org/wiki/Estoicismo>) sobre Estoicismo.



¹⁸<https://mirrorthink.ai/>

¹⁹<https://github.com/features/copilot>, <https://ai.meta.com/blog/code-llama-large-language-model-coding/>

²⁰<https://www.compose.ai/>

²¹<https://clarice.ai/>

²²Veja o que são os Transformers no Capítulo [Modelos de linguagem](#).

²³Copiando e colando o conteúdo de <https://www.bbc.com/portuguese/geral-46458304>.

²⁴<https://pt.wikipedia.org/wiki/Estoicismo>

²⁵A resposta devolvida tem menos de 200 palavras, mas nem sempre os agentes de conversação obedecem a restrições como essa, inseridas nos *prompts*.

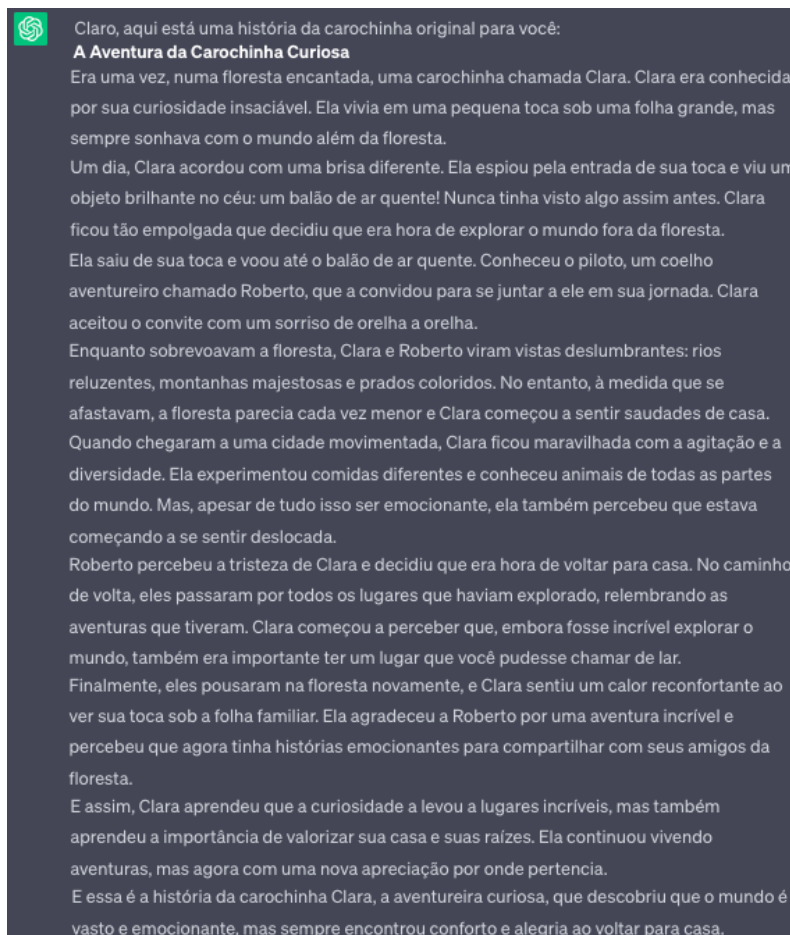
7.3.2 Criação de história

Pedimos ao ChatGPT e a MariTalk para escrever uma história da Carochinha²⁶, com o seguinte *prompt*

- P. Crie uma história da Carochinha original.

As Figuras 7.4 e 7.5 trazem a saída para cada *chatbot*, respectivamente.

Figura 7.4: Resposta do ChatGPT, em 2023, ao ser deparado com o seguinte *prompt* “Crie uma história da Carochinha original”.



Podemos observar que as histórias têm coerência e parecem mesmo com histórias populares da carochinha, mantendo o nível para o público infantil. Mas também podemos observar que as histórias trazem temas repetitivos (ou seja, falta criatividade ou acontecimentos inesperados e surpreendentes).

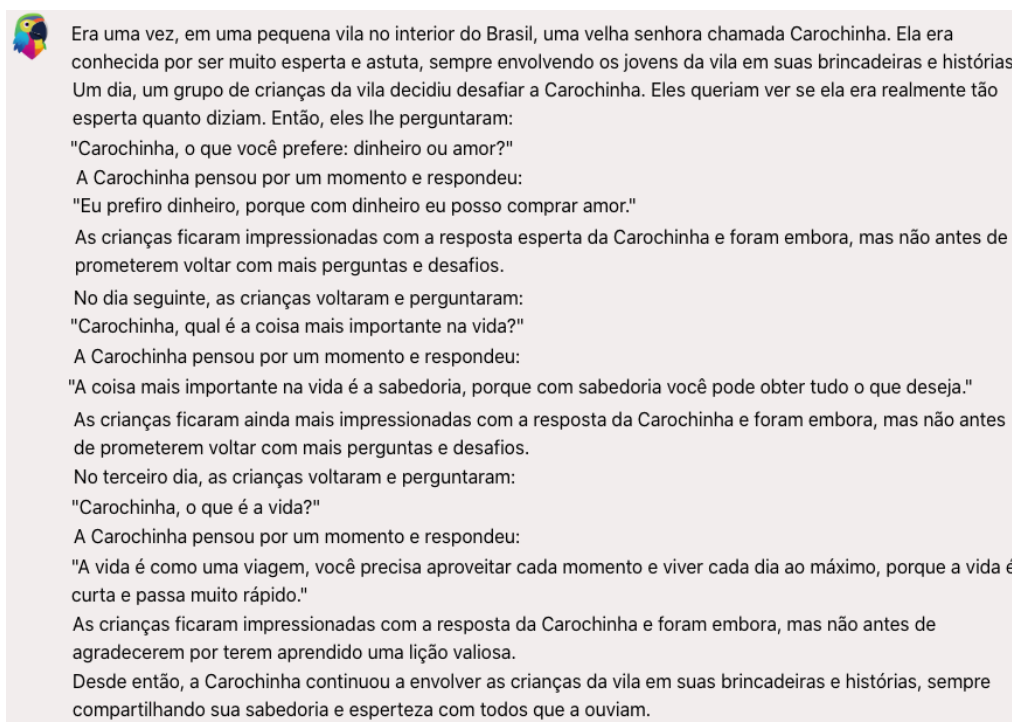
Por isso, alteramos o *prompt* um pouco, acrescentando um pedido para que a história fosse surpreendente:

- P. Escreva uma história da Carochinha surpreendente.

A história que segue no Quadro 7.2 contém mais elementos que as anteriores, mas não chamaríamos de “surpreendente”. E, como estamos aqui fazendo uma avaliação, é curioso que logo a primeira frase apresente um erro gramatical: “Era uma vez (...) vivia uma carochinha chamada Clara.” Guiada pelo padrão “Era uma vez”, típico desse tipo de texto, a ferramenta não percebeu que não basta a fórmula “Era uma vez”,

²⁶Segundo a Academia Brasileira de Letras, “O termo carochinha, atrelado à imagem de uma velha bondosa e afável a distrair os pequenos com suas narrativas feéricas, foi introduzido no nosso folclore através da obra Histórias da Carochinha, uma coleção de textos oriundos da tradição oral, organizada por Figueiredo Pimentel e que veio a ser o primeiro livro infantil publicado no Brasil, depois de 1920, para acalantar as crianças.” <https://www.academia.org.br/artigos/historias-da-carochinha>

Figura 7.5: Resposta da MariTalk, em 2023, ao ser deparada com o seguinte *prompt* “Crie uma história da Carochinha original”.



mas que era uma vez “alguma coisa”, e não “era uma vez” uma ação. Ou seja, o correto seria “Era uma vez, numa floresta, uma menina chamada...” e não “Era uma vez, numa floresta, vivia uma menina chamada...” De um ponto de vista narrativo, parece haver um salto no início do 5o parágrafo: se tudo ia bem, e Clara aprendia na jornada, por que ela estava prestes a desistir? Ainda assim, é uma história interessante, que poderia ser contada para uma criança.

Quadro 7.2: História da Carochinha devolvida pelo ChatGPT, em 2023, como resposta ao *prompt*: “Escreva uma história da Carochinha surpreendente”

Título: A Inesperada Jornada da Carochinha

Era uma vez, numa floresta encantada, onde animais falavam e aventuras aconteciam, vivia uma carochinha chamada Clara. Ela era conhecida por sua gentileza e curiosidade insaciável. Clara vivia numa pequena toca no pé de uma árvore e passava seus dias explorando os arredores, fazendo novos amigos e ouvindo histórias incríveis. Certo dia, enquanto explorava a margem de um riacho, Clara conheceu um sábio sapo chamado Samuel. Ele era famoso por sua sabedoria e conhecimento sobre os segredos mais profundos da floresta. Clara e Samuel logo se tornaram amigos e passavam horas conversando sobre os mistérios do mundo ao seu redor.

Um dia, Samuel contou a Clara sobre um antigo artefato mágico chamado “Espelho dos Desejos”, que supostamente tinha o poder de realizar um desejo para quem o encontrasse. Clara ficou fascinada pela ideia e decidiu embarcar em uma jornada para encontrar esse espelho lendário.

Com a mochila nas costas e um mapa rudimentar nas mãos, Clara começou sua jornada emocionante. Ela atravessou rios, subiu montanhas e explorou florestas profundas, sempre encontrando novos amigos e enfrentando desafios inesperados. Em cada etapa da jornada, ela aprendia lições valiosas sobre amizade, coragem e perseverança. Um dia, quando estava prestes a desistir da busca pelo Espelho dos Desejos, Clara conheceu uma coruja sábia chamada Olívia. Ela ouviu a história de Clara e decidiu ajudá-la. Olívia revelou que o espelho estava escondido no coração da Floresta Proibida, um lugar perigoso e cheio de enigmas.

Juntas, Clara e Olívia enfrentaram os desafios da Floresta Proibida. Resolveram quebra-cabeças mágicos, enfrentaram criaturas assustadoras e superaram obstáculos aparentemente impossíveis. A



cada passo, sua amizade crescia mais forte e sua determinação se intensificava. Finalmente, depois de superar todos os obstáculos, Clara e Olívia encontraram o Espelho dos Desejos. Mas antes que Clara pudesse fazer seu desejo, ela olhou para o espelho e percebeu algo importante: sua jornada a havia transformado. Ela já não desejava uma única coisa, pois havia descoberto que a verdadeira magia estava nas amizades que havia feito e nas experiências que havia vivido. Clara sorriu para o espelho e disse: “Meu único desejo é que a magia da amizade e da aventura continue a brilhar em todos os cantos deste mundo.” O espelho brilhou intensamente e uma onda de energia mágica se espalhou pela floresta. As plantas ganharam cores mais vibrantes, os animais irradiavam alegria e a harmonia encheu o ar. Com seu desejo realizado, Clara voltou para sua toca na árvore, carregando consigo as lembranças de sua jornada incrível. Ela percebeu que, mesmo nas situações mais inesperadas, a vida poderia ser cheia de surpresas maravilhosas, se você estiver disposto a explorar, aprender e compartilhar momentos especiais com os outros. E assim, a história da carochinha Clara e sua jornada inesperada se tornou uma lenda contada por gerações, inspirando todos a valorizarem a amizade e a magia das descobertas.

Em 2025 repetimos o *prompt* com o ChatGPT (Figura 7.6) e a Maritalk (Figura 7.7), e ainda que “surpreendente” seja de difícil avaliação, nos parece que a geração atendeu melhor ao comando, com personagens inusitados como um esquilo cartógrafo e um sapo inventor.



Figura 7.6: História da Carochinha devolvida pelo ChatGPT, em 2025, como resposta ao *prompt*: “Escreva uma história da Carochinha surpreendente”

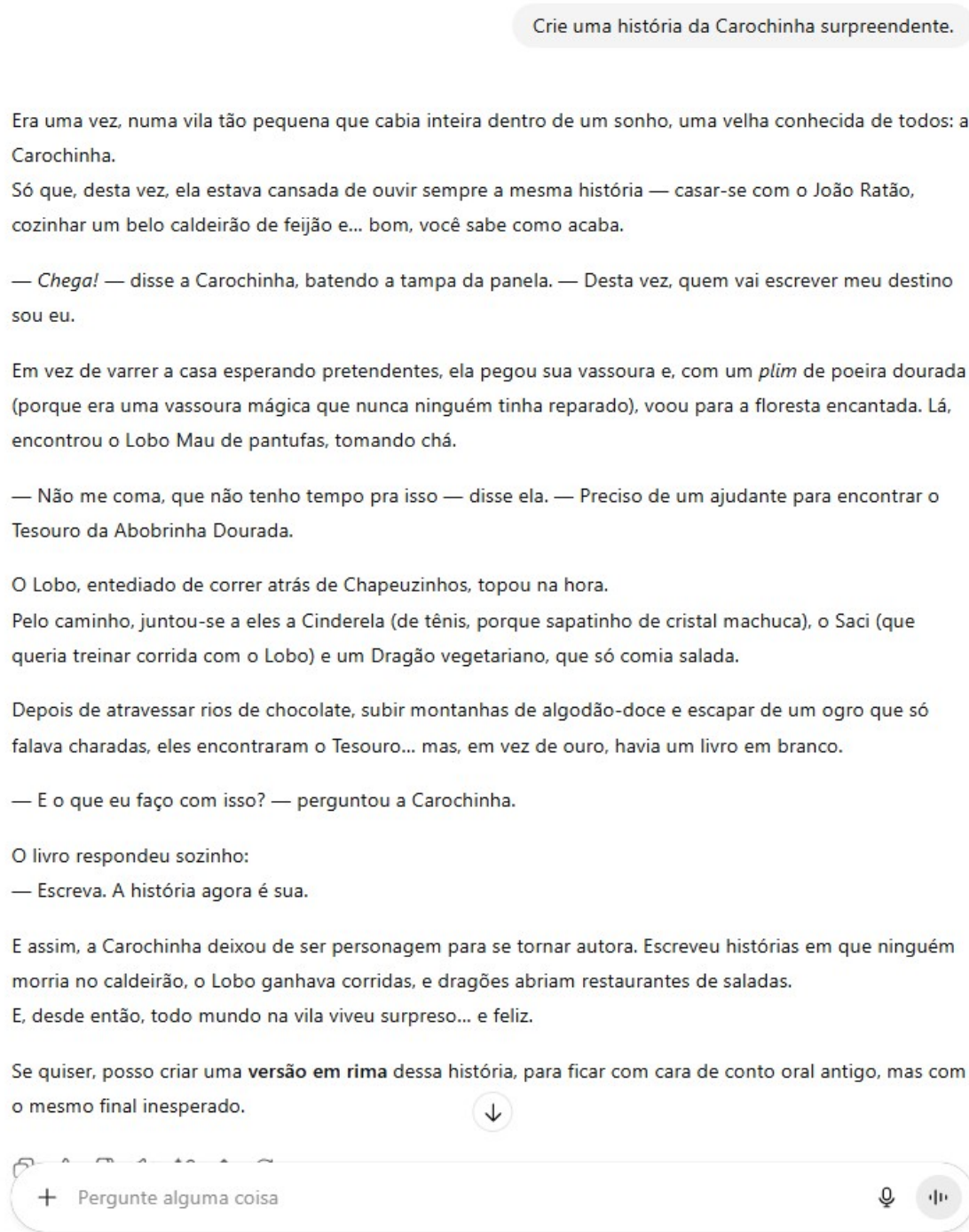
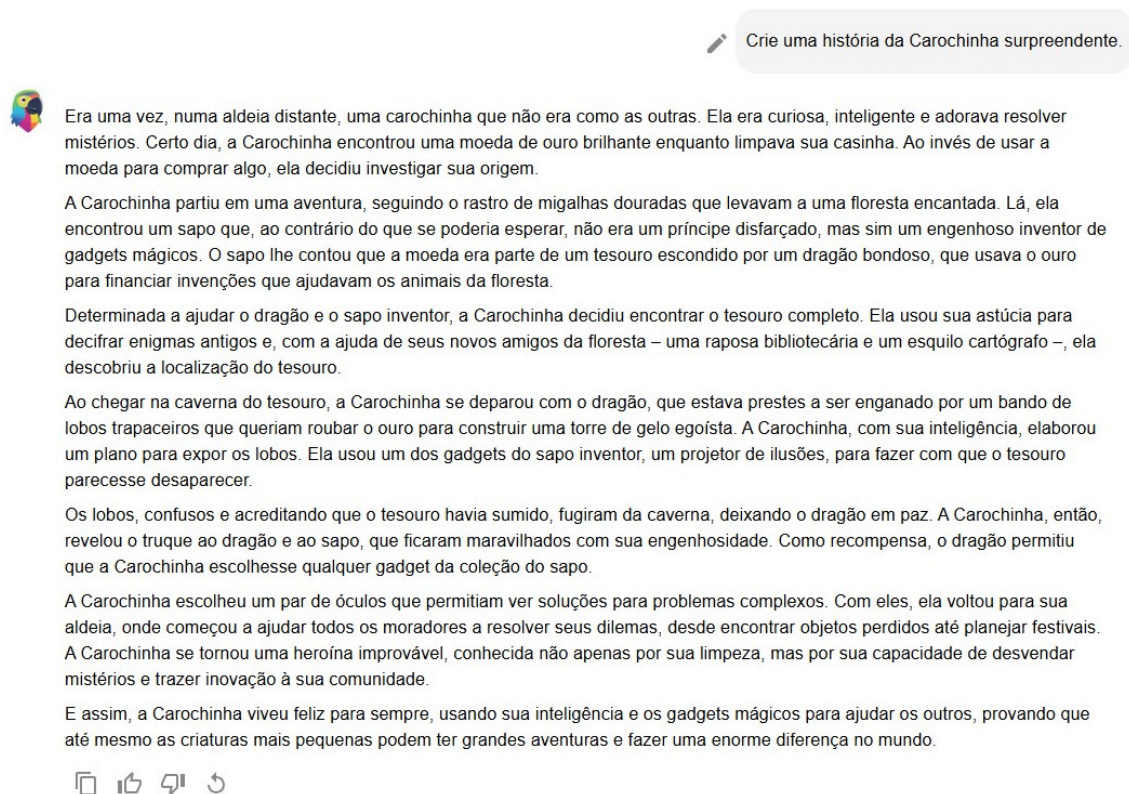


Figura 7.7: História da Carochinha devolvida pela Maritalk, em 2025, como resposta ao *prompt*: “Escreva uma história da Carochinha surpreendente”

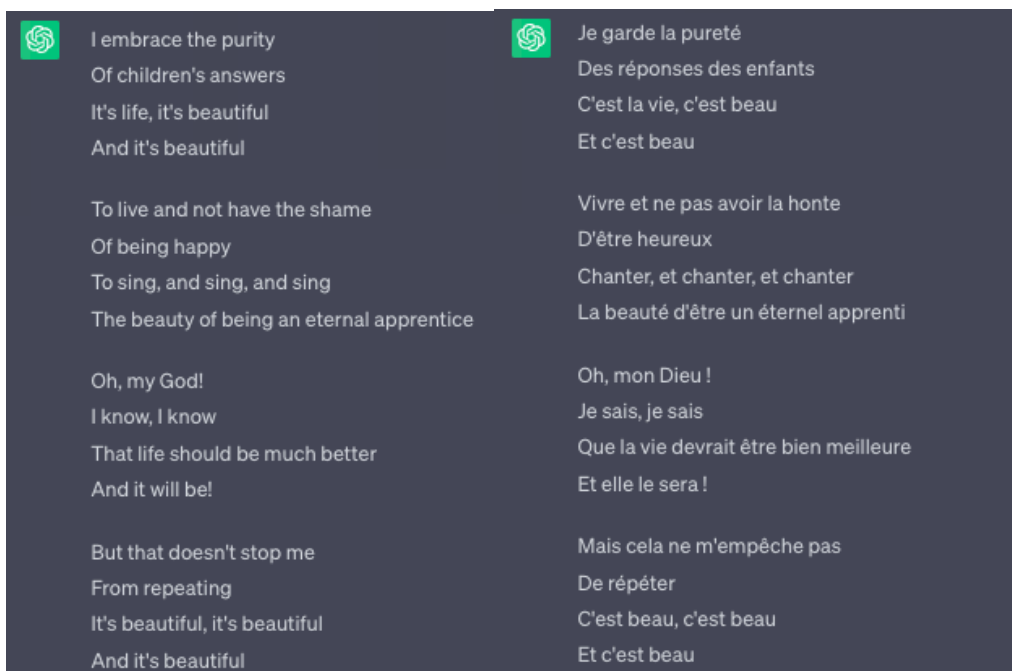


7.3.3 Tradução automática

Outra tarefa bem desempenhada pelos *chatbots* é a tradução automática, ao menos para pares de línguas que tenham muitos recursos disponíveis para treinamento. Abordar línguas com poucos recursos de avaliação e treinamento é uma tarefa de interesse da comunidade de PLN, mas também um desafio para o treinamento dos modelos de linguagem. Os *chatbots* até devolvem alguma resposta para línguas com poucos recursos/mais raras, mas seria difícil avaliarmos a correção e fluência da tradução. É bom lembrar que a tradução é considerada como uma das habilidades emergentes dos modelos de linguagem de larga escala, ou seja, é uma tarefa com a qual eles conseguem lidar mesmo que não tenham sido explicitamente treinados.

A Figura 7.8 apresenta a tradução devolvida pelo ChatGPT para as quatro primeiras estrofes da canção “O Que é, o Que é?” e o *prompt* “Traduza o seguinte texto para inglês e para francês:”. Tanto a tradução para inglês como para francês estão corretas. Mas esta não é uma tradução muito difícil de ser feita, as palavras contidas são simples e a Internet pode estar repleta de tentativas como essa.

Figura 7.8: Resposta do ChatGPT, em 2023, para um *prompt* de tradução de uma música.



Para testar alguns casos em que os *chatbots* poderiam ter problemas na tradução, escolhemos algumas sentenças do copus DELA²⁷, descritas no artigo (Castilho et al., 2021) como problemáticas quando traduzidas fora do contexto. As Figuras 7.9 e 7.10 mostram as traduções das sentenças pelos *chatbots* ChatGPT e MariTalk, respectivamente. Propositadamente, as traduções foram feitas para sentenças isoladas, também fora de contexto, para observarmos o que estes dois agentes fariam em casos de ambiguidade ou outros problemas listados no artigo.

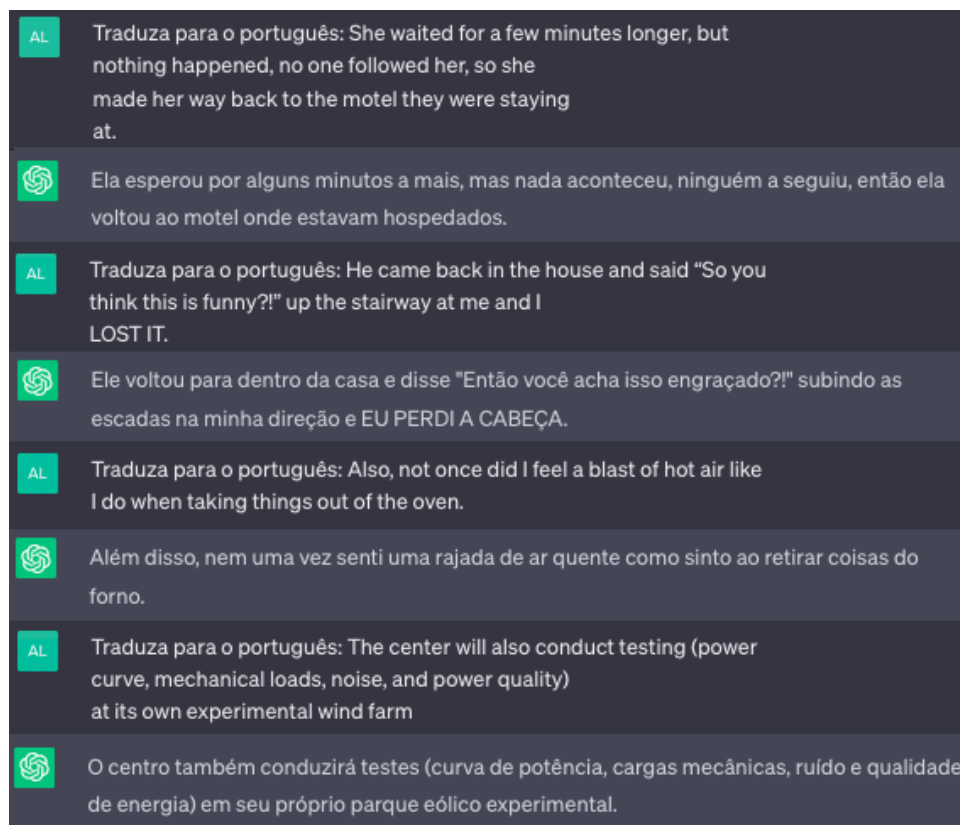
No primeiro caso, um exemplo do artigo referente à falta de flexão de gênero na língua inglesa, ambos os *chatbots* assumiram o default para masculino (“hospedados”), embora a resposta da MariTalk não incluía o pronome “eles”, o que é comum na língua portuguesa, mas não é correto na língua inglesa. Este não é um erro, já que ambas as flexões claramente estariam corretas com apenas a sentença de entrada. Mas, em caso de dúvida, a resposta poderia alternar entre os gêneros feminino e masculino.

No segundo exemplo, embora a expressão “*I LOST IT*” também possa ser interpretada como alguém perdendo alguma coisa, ambos os *chatbots* traduziram como “Eu perdi a cabeça”, querendo dizer que a pessoa perdeu o controle. O contexto da sentença direciona para esta tradução, corretamente. Os *chatbots* também não tiveram problema para resolver a elipse da próxima sentença. No último caso, a tradução do ChatGPT para “parque eólico” é mais específica de domínio do que a tradução da MariTalk, para uma palavra mais genérica, conforme discutido no artigo (Castilho et al., 2021) no que se refere a problemas associados à terminologia.

²⁷<https://github.com/SheilaCastilho/DELA-Project>



Figura 7.9: Resposta do ChatGPT, em 2023, para algumas sentenças cujas traduções podem ser problemáticas, quando fora de contexto.



Um exemplo interessante é a tradução da seguinte sentença em inglês para o português:

- *I knew I didn't have a drinking problem – but I had a problem with drinking.*

A MariTalk traduziu para “Eu sabia que eu não tinha um problema de beber – mas eu tinha um problema com beber.”, enquanto o ChatGPT retornou “Eu sabia que não tinha um problema com álcool – mas tinha um problema com o consumo de bebidas.”. Uma tradução mais coerente seria “Eu sabia que meu problema não era alcoolismo; era com a bebida.”²⁸, o que não foi retornado por nenhum desses dois agentes na primeira iteração. Nós tentamos outras traduções em algumas outras iterações com o ChatGPT e o resultado pode ser visto na Figura 7.11. Podemos perceber que as próximas traduções estão indo em uma direção semântica mais similar ao esperado.

²⁸Tradução fornecida pela profa. Adriana Pagano. O contexto da frase é esse: *Some people have to give up drinking completely, they can't have a couple because they know where it would lead. Alcoholism is real. It requires a serious, courageous ongoing recovery process. That feels separate to what I'm describing here. I had fallen into grey-area drinking, a term coined by Jolene Park, that describes the feeling that you don't have a “drinking problem”, but you do have a “problem with drinking” without it being a severe alcohol use disorder.*



Figura 7.10: Resposta da MariTalk, em 2023, para algumas sentenças cujas traduções podem ser problemáticas, quando fora de contexto.



Traduza para o português: She waited for a few minutes longer, but nothing happened, no one followed her, so she made her way back to the motel they were staying at.



Ela esperou por alguns minutos a mais, mas nada aconteceu, ninguém a seguiu, então ela se dirigiu de volta ao motel onde estavam hospedados.



Traduza para o português: He came back in the house and said "So you think this is funny?!" up the stairway at me and I LOST IT.



Ele voltou para dentro de casa e disse "Então você acha isso engraçado?!" subindo as escadas em minha direção e EU PERDI A CABEÇA.



Traduza para o português: Also, not once did I feel a blast of hot air like I do when taking things out of the oven.



Além disso, em nenhum momento eu senti um jato de ar quente como sinto quando tiro coisas do forno.



Traduza para o português: The center will also conduct testing (power curve, mechanical loads, noise, and power quality) at its own experimental wind farm



O centro também conduzirá testes (curva de potência, cargas mecânicas, ruído e qualidade de energia) na sua própria fazenda eólica experimental.

Figura 7.11: Continuação da iteração com o ChatGPT, em 2023, para a tradução da sentença "I knew I didn't have a drinking problem – but I had a problem with drinking." para o português.

AL Traduza para o português, de uma forma diferente da que você traduziu antes: I knew I didn't have a drinking problem – but I had a problem with drinking.

Eu tinha consciência de que não possuía um problema com o consumo de álcool, porém enfrentava uma questão relacionada a beber.

AL E outra diferente: Traduza para o português: I knew I didn't have a drinking problem – but I had a problem with drinking.

Eu tinha ciência de que não possuía um problema de alcoolismo, mas tinha uma questão com relação ao ato de beber.

AL E mais outra diferente:

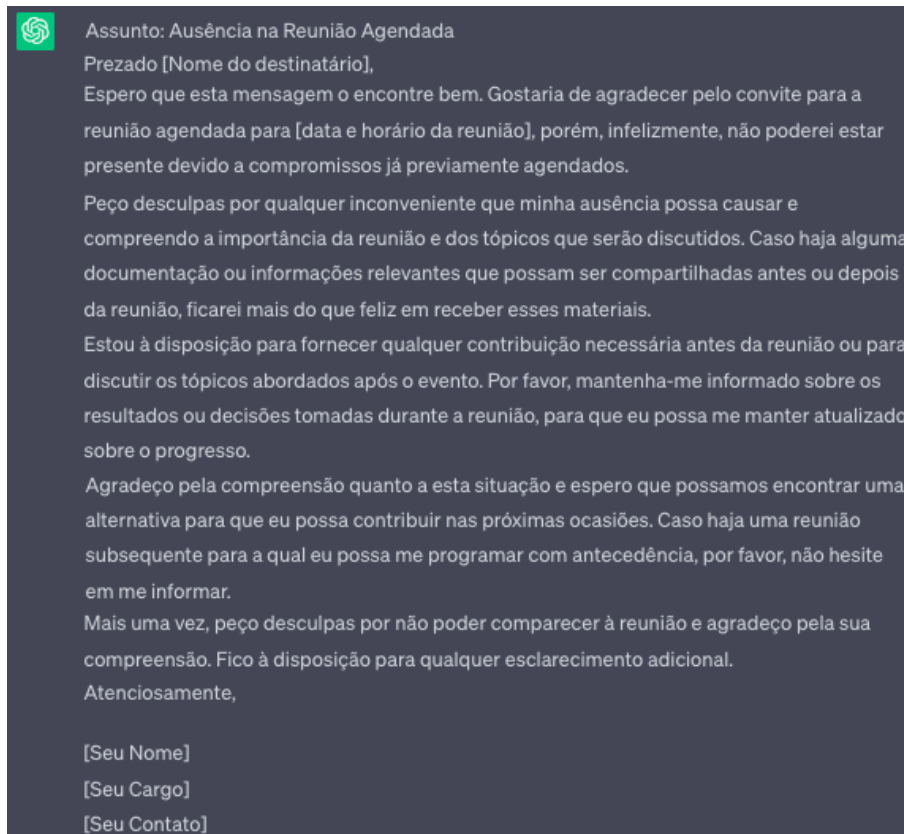
Eu tinha plena certeza de que não enfrentava um problema de consumo excessivo de álcool, no entanto, havia uma dificuldade relacionada a beber.



7.3.4 Escrever e-mails

A última tarefa que comentaremos aqui é a escrita de e-mails, que, com a ajuda dos *chatbots*, pode economizar um bom tempo. Entretanto, é sempre bom reforçar que dada a natureza probabilística da geração dos textos pelos *chatbots*, é essencial revisar o e-mail antes de enviá-lo, ainda mais em situações formais ou de comunicação com pessoas fora do círculo de relacionamento. A Figura 7.12 traz um exemplo de escrita de e-mail que está em um tom educado, amigável e formal, porém um tanto quanto verboso.

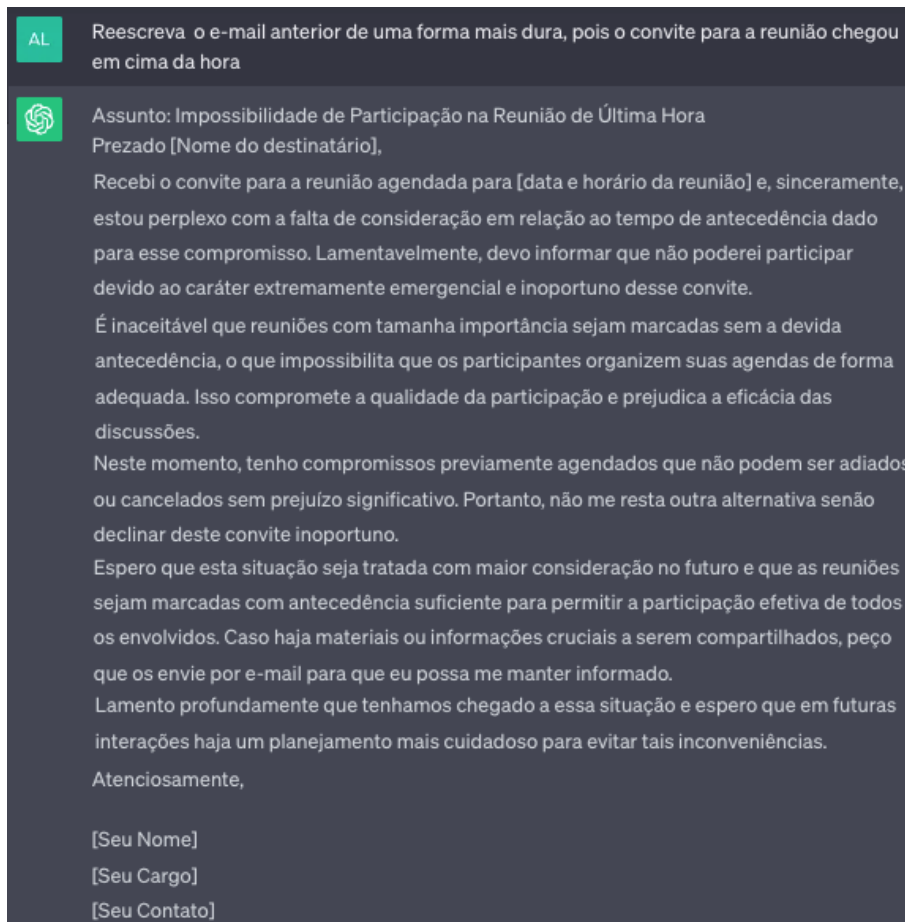
Figura 7.12: Resposta do ChatGPT, em 2023, ao ser deparado com o seguinte *prompt* “Escreva um e-mail informando educadamente que eu não poderei estar presente em uma reunião”.



Ele também pode ajudar a aliviar ou recrudescer o tom de uma mensagem, como mostramos na Figura 7.13. É um tom realmente ríspido, porém incisivo e direto ao ponto da insatisfação. Mas ao menos a pessoa não teria que ficar o dia inteiro pensando em como responder em uma situação indesejada.



Figura 7.13: Reescrita, em 2023, do e-mail da Figura 7.12 em um tom mais ríspido.



7.4 Jogos que os agentes melhoraram

7.4.1 Simplificação textual

A tarefa (ou jogo) de simplificação textual envolve tornar textos mais simples e acessíveis. Como é possível imaginar, não é uma atividade óbvia, uma vez que o que é simples para uma pessoa pode não ser para outra. Além disso, a atividade de simplificação frequentemente precisa ir além do texto original, buscando informações que estão fora do texto para justamente produzir um texto compreensível para o público pretendido. Em 2023, indicamos que este era um jogo ainda não dominado. Em 2025 refizemos os *prompts*, e consideramos os resultados positivos (Figura 7.15). No pedido original, pedimos ao ChatGPT que simplificasse um texto, e usamos como alvo da simplificação uma matéria de jornal – a princípio, algo que já é simples – mas da seção Economia (no Quadro 7.3).

Quadro 7.3: Texto original sobre economia

‘Quanto mais independente, mais eficaz’, diz Campos Neto sobre autonomia do Banco Central
Presidente do Banco Central, que vem sendo alvo de críticas de Lula, afirmou que autonomia da instituição diminui o custo dos juros. Ele defendeu a separação da política governamental da política monetária.

Presidente do Banco Central defende a autonomia da instituição
O presidente do Banco Central, Roberto Campos Neto, disse nesta terça-feira (7), em palestra nos Estados Unidos, que a autonomia da instituição serve para separar as diretrizes monetárias da esfera política.



Campos Neto vem sendo criticado nos últimos dias pelo presidente da República, Luiz Inácio Lula da Silva, que afirma que a taxa de juros básicos da economia deveria ser reduzida.

Em sua última reunião, na semana passada, o Comitê de Política Monetária (Copom) do Banco Central manteve a Selic em 13,75% ao ano. Diferentemente de seus outros mandatos, agora Lula não pode trocar o presidente do BC. A autonomia do BC, defendida pelo governo Jair Bolsonaro, foi aprovada pelo Congresso em 2021.

“A principal razão no caso da autonomia do Banco Central é desconectar o ciclo da política monetária do ciclo político porque eles têm planos e interesses diferentes. E quanto mais independente você for, mais eficaz você é e menos o país pagará em termos de custo de ineficiência na política monetária”, afirmou Campos Neto.

Críticas de Lula

Na segunda-feira (6), em discurso durante evento no Banco Nacional de Desenvolvimento Econômico e Social (BNDES), Lula criticou a Selic a 13,75% e disse que o país tem uma “cultura” de juros altos que “não combina com a necessidade de crescimento” do país.

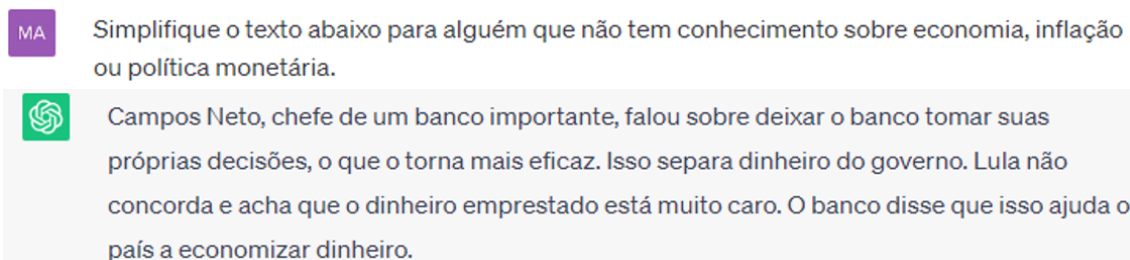
“É só ver a carta do Copom para a gente saber que é uma vergonha esse aumento de juros e a explicação que eles deram para a sociedade brasileira”, disse Lula na ocasião.

Nesta terça, ele voltou ao assunto em entrevista a veículos de mídia alternativa.


“Não é possível que a gente queira que este país volta a crescer com taxa de 13,75%. Nós não temos inflação de demanda. É só isso. É isso que eu acho que esse cidadão [Campos Neto], indicado pelo Senado, tenha possibilidade de maturar, de pensar e de saber como vai cuidar deste país. Ele tem muita responsabilidade”, afirmou o presidente.

Ao usar o *prompt* “Simplifique o texto abaixo”, a resposta foi um resumo de um parágrafo²⁹. Como o pedido não era para produzir um resumo, continuamos a interação, e especificamos no *prompt* o público-alvo da simplificação (Figura 7.14), o que alterou o texto gerado. No entanto, as explicações fornecidas ainda deixavam a desejar. A segunda frase da simplificação, por exemplo – “Isso separa dinheiro do governo” – faz pouco sentido. “Isso” o quê? Na terceira frase, de que “dinheiro emprestado” se está falando? A última frase é igualmente sem sentido. Ao que parece, o encadeamento de palavras/frases, nesse caso, não foi muito bem sucedido, produzindo um texto sem coerência. Por fim, a confusão entre simplificação e sumarização continua, quando comparamos o texto gerado com o texto original.

Figura 7.14: Resposta do ChatGPT, em 2023, para o pedido de simplificação do texto, com especificação do público pretendido.



MA Simplifique o texto abaixo para alguém que não tem conhecimento sobre economia, inflação ou política monetária.

 Campos Neto, chefe de um banco importante, falou sobre deixar o banco tomar suas próprias decisões, o que o torna mais eficaz. Isso separa dinheiro do governo. Lula não concorda e acha que o dinheiro emprestado está muito caro. O banco disse que isso ajuda o país a economizar dinheiro.

Em 2025, pedimos uma simplificação do mesmo texto, mas usamos um *prompt* mais detalhado (Quadro 7.4). Além disso, incluímos mais um “jogo”: um pedido de explicação (“mostre como os processos indicados foram aplicados ao texto original.”). O resultado está nas figuras 7.15 (geração) e 7.16 (explicação das alterações).

Quadro 7.4: *Prompt* usado em 2025 para gerar uma versão simplificada do texto.

Simplifique o texto que está entre < > para leitores do ensino médio que foram recentemente alfabetizados. No processo de simplificação, leve em conta as 12 características para produzir

²⁹Na linguagem cotidiana, “simplificar um texto” pode ser sinônimo de “resumir um texto”. Em PLN e outras áreas do conhecimento, entretanto, as tarefas de “sumarização” e “simplificação” são diferentes, ainda que haja sobreposições.



textos simples listadas a seguir. Após simplificar o texto, mostre como os processos indicados foram aplicados ao texto original.

Características para produzir um texto simples:

1. Usar frases na ordem direta (sujeito, verbo, objeto, advérbio)
2. Usar frases curtas
3. Indicar o tópico frasal no início do parágrafo
4. Acrescentar exemplos, quando necessário
5. Usar anáforas apenas se estiverem próximas no texto
6. Evitar omissão de sujeito e de objeto
7. Usar intertítulos para facilitar a compreensão
8. Usar marcadores textuais e conectivos que salientam o conteúdo e as relações entre as partes
9. Apresentar o conteúdo em ordem cronológica
10. Usar palavras simples
11. Evitar palavras abstratas e abstrações
12. Adicionar explicações quando for necessário

Texto original que será simplificado:

<ORIGINAL>

Figura 7.15: Resposta do ChatGPT gerada em 2025.

A autonomia do Banco Central

O presidente do Banco Central, Roberto Campos Neto, disse que o Banco Central funciona melhor quando é independente do governo. Ele falou isso numa palestra nos Estados Unidos.

Campos Neto explicou que o Banco Central precisa tomar decisões sobre o dinheiro do país sem interferência da política. Segundo ele, isso ajuda a diminuir os juros e os custos para o país.

Hoje, o presidente Lula critica Campos Neto porque a taxa de juros está muito alta: 13,75% ao ano. Lula acha que os juros deveriam cair para ajudar o Brasil a crescer.

O Banco Central decidiu manter essa taxa de juros na semana passada. Lula disse que essa decisão é ruim para o país. Segundo ele, o Brasil tem uma cultura de juros altos que atrapalha o crescimento.

Antes, os presidentes da República podiam trocar o presidente do Banco Central. Mas, desde 2021, isso mudou. Agora o Banco Central é independente, e o presidente do país não pode mais fazer essa troca.

Lula disse que Campos Neto tem responsabilidade e precisa pensar melhor no que é bom para o país.



Figura 7.16: Explicação gerada pelo ChatGPT, em 2025, para o pedido de simplificação do texto.

Como as 12 características de textos simples foram aplicadas:

Característica	Aplicação no texto simplificado
1. Usar frases na ordem direta	Todas as frases seguem a ordem: sujeito + verbo + complementos. Ex: <i>Lula critica Campos Neto.</i>
2. Usar frases curtas	As frases foram divididas para facilitar a leitura. Ex: <i>Ele falou isso numa palestra nos EUA.</i>
3. Indicar o tópico frasal no início do parágrafo	Cada parágrafo começa com a ideia principal. Ex: <i>O presidente do Banco Central disse que...</i>
4. Acrescentar exemplos, quando necessário	Exemplo incluído: <i>"Ele falou isso numa palestra nos Estados Unidos"</i> contextualiza a declaração.
5. Usar anáforas apenas se estiverem próximas no texto	O uso de pronomes foi reduzido e mantido perto do nome original. Ex: <i>Campos Neto... Ele...</i>
6. Evitar omissão de sujeito e de objeto	Sujeitos como <i>Lula</i> ou <i>Banco Central</i> são repetidos quando necessário para manter a clareza.
7. Usar intertítulos para facilitar a compreensão	Um intertítulo foi usado: <i>A autonomia do Banco Central.</i>
8. Usar marcadores textuais e conectivos	Conectivos simples foram usados: <i>hoje, mas, antes, agora, segundo ele, etc.</i>
9. Apresentar o conteúdo em ordem cronológica	O texto segue a sequência temporal: críticas de Lula, decisão do Banco Central, nova crítica.
10. Usar palavras simples	Palavras difíceis como <i>diretrizes, esfera política</i> ou <i>eficiência monetária</i> foram trocadas por termos simples como <i>decisões</i> e <i>dinheiro</i> .
11. Evitar palavras abstratas e abstrações	Termos como <i>"cultura de juros altos"</i> foram explicados ou substituídos.
12. Adicionar explicações quando for necessário	Ex: foi explicado que a taxa de juros está <i>13,75% ao ano</i> e que essa taxa é considerada alta.



O texto gerado não é um resumo, mas uma boa simplificação. Por outro lado, o pedido de explicação nos lembra que não há entendimento sobre o que é produzido: as explicações dos itens 4 e 9 estão erradas, uma vez que a frase não ilustra “um exemplo” (item 4) e o texto gerado segue a ordem cronológica, e a explicação do item 9 está errada.

7.5 Jogos que os agentes parecem jogar mal

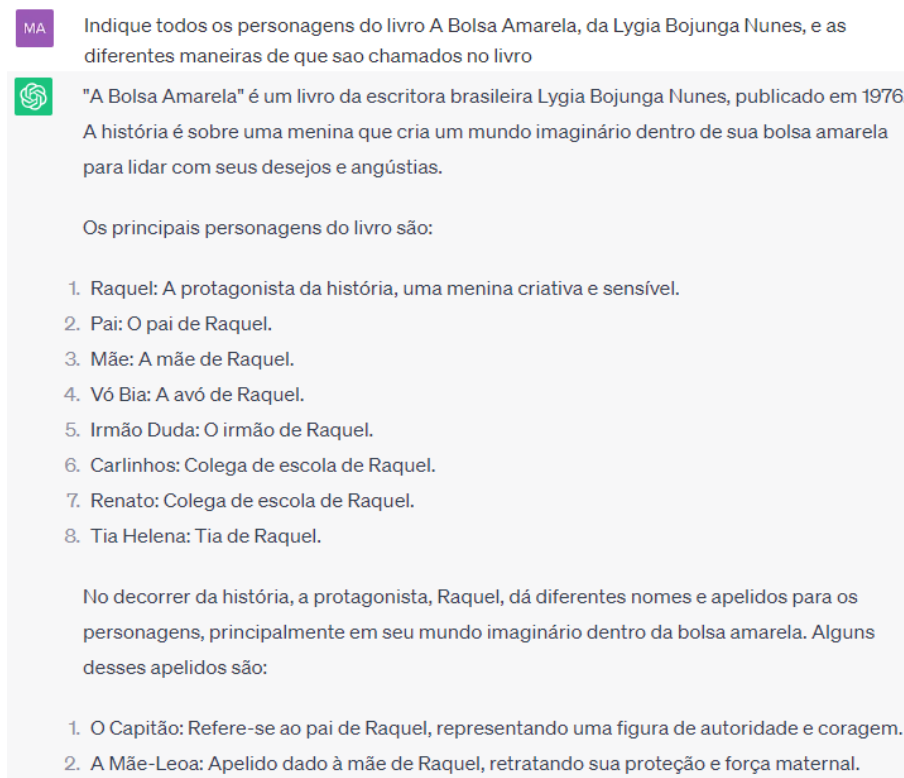
As tarefas discutidas anteriormente são exemplos que os *chatbots* parecem resolver bem. Entretanto, aqueles são apenas exemplos gerados de forma espontânea, sem muito rigor ou metodologia na definição dos *prompts*. E, mesmo nessas tarefas, poderíamos encontrar exemplos em que as respostas retornadas fossem ruins. Nesta seção, vamos apresentar algumas tarefas de PLN em que os agentes costumam se sair mal, o que pode ser ocasionado por diversos fatores: falta de dados para treinamento, falta de treinamento, treinamento inadequado ou mesmo a falta de adequação dos modelos de linguagem, como são concebidos, para resolverem a tarefa. Afinal, como dissemos no início deste capítulo, nem todas as atividades linguísticas que exercemos se resumem a um jogo de previsões.

7.5.1 Perguntas sobre conteúdo pouco popular na internet

Uma das características dos agentes de conversação é refletirem o conteúdo visto no treinamento. Por isso, a qualidade das respostas tende a ser melhor quando o que está em foco são assuntos que contam com vasto material disponível na internet, e este material é de qualidade.

“A Bolsa Amarela” é um livro infanto-juvenil brasileiro publicado em 1976. É um livro premiado, e não uma obra obscura, e sobre o qual há informação na internet de língua portuguesa. Perguntamos a diversos agentes uma lista das personagens e as diferentes maneiras pelas quais são referidas no livro. ChatGPT e MariTalk deram as respostas apresentadas nas Figuras 7.17 e 7.18, respectivamente.

Figura 7.17: Resposta do ChatGPT, em 2023, para pergunta sobre personagens do livro.



MA Indique todos os personagens do livro A Bolsa Amarela, da Lygia Bojunga Nunes, e as diferentes maneiras de que são chamados no livro

"A Bolsa Amarela" é um livro da escritora brasileira Lygia Bojunga Nunes, publicado em 1976. A história é sobre uma menina que cria um mundo imaginário dentro de sua bolsa amarela para lidar com seus desejos e angústias.

Os principais personagens do livro são:

1. Raquel: A protagonista da história, uma menina criativa e sensível.
2. Pai: O pai de Raquel.
3. Mãe: A mãe de Raquel.
4. Vó Bia: A avó de Raquel.
5. Irmão Duda: O irmão de Raquel.
6. Carlinhos: Colega de escola de Raquel.
7. Renato: Colega de escola de Raquel.
8. Tia Helena: Tia de Raquel.


No decorrer da história, a protagonista, Raquel, dá diferentes nomes e apelidos para os personagens, principalmente em seu mundo imaginário dentro da bolsa amarela. Alguns desses apelidos são:

1. O Capitão: Refere-se ao pai de Raquel, representando uma figura de autoridade e coragem.
2. A Mãe-Leoa: Apelido dado à mãe de Raquel, retratando sua proteção e força maternal.

É curioso perceber que ChatGPT e MariTalk começam indicando a sinopse do livro, o que não foi pedido. Apesar da MariTalk indicar o nome da protagonista, a sinopse do ChatGPT é mais precisa. Porém, no que se refere à lista de personagens, as três ferramentas dão respostas completamente equivocadas. O único



Figura 7.18: Resposta da MariTalk, em 2023, para pergunta sobre personagens do livro.



A Bolsa Amarela é um livro escrito por Lygia Bojunga Nunes, publicado em 1976. O livro conta a história de uma menina chamada Raquel, que é uma estudante do ensino fundamental e que está enfrentando problemas em casa e na escola.

Os personagens principais do livro são:

- * Raquel: a protagonista da história, uma menina que está enfrentando problemas em casa e na escola.
- * Dona Helena: a mãe de Raquel, que é uma artista plástica e que está passando por um momento difícil em sua carreira.
- * Seu João: o pai de Raquel, que é um homem aparentemente calmo e que trabalha como contador.
- * Professor Carlos: o professor de Raquel, que é um homem gentil e que tenta ajudá-la a superar seus problemas.
- * Lívia: uma colega de classe de Raquel, que é uma menina popular e que aparentemente tem uma vida perfeita.
- * Tio Edgard: o tio de Raquel, que é um homem aventureiro e que a inspira a seguir seus sonhos.

Algumas das maneiras pelas quais esses personagens são chamados no livro incluem:

- * Raquel: Raquel, Quel, Quelzinha, Raquinha, Raquela.
- * Dona Helena: Dona Helena, Helena, Mãe, Mamãe.
- * Seu João: Seu João, João, Pai, Papai.
- * Professor Carlos: Professor Carlos, Carlos, Professor, Prof. Carlos.
- * Lívia: Lívia, Livi, Lí, Livinha.
- * Tio Edgard: Tio Edgard, Edgard, Tio, Tio Ed.

elemento correto das respostas – tanto do ChatGPT quanto da MariTalk – é que a protagonista se chama Raquel. Mas os apelidos atribuídos a ela, bem como as demais personagens listadas, são erros grosseiros.

Em 2025, repetimos a pergunta para ChatGPT e Maritalk (Figuras 7.19 e 7.20). As respostas continuam erradas.

Figura 7.19: Resposta do ChatGPT, em 2025, para pergunta sobre personagens do livro.

Quais os personagens principais do livro "A Bolsa Amarela"

Os personagens principais do livro "A Bolsa Amarela", da autora **Lygia Bojunga**, são:

1. Raquel –

A protagonista e narradora da história. É uma menina de 9 anos, sensível, criativa e questionadora. Raquel se sente reprimida em casa, onde suas vontades são constantemente ignoradas. Ela guarda seus maiores desejos secretos dentro de uma bolsa amarela: o desejo de **crescer**, de **ser menino** e de **ser escritora**.

2. A Bolsa Amarela –

Embora seja um objeto, a bolsa funciona como uma personagem simbólica muito importante. É onde Raquel guarda seus segredos, desejos e também suas fantasias. A bolsa representa sua individualidade, seus sonhos e sua liberdade de imaginar.

3. Tonton –

Um galo falante e muito vaidoso, amigo de Raquel. Representa a fantasia e a liberdade de pensamento que ela tanto deseja.

4. Xexéu –

Um alfinete falante, também amigo de Raquel. Junto com Tonton, participa das aventuras imaginárias da menina, ajudando-a a lidar com seus sentimentos e frustrações.

5. Avó de Raquel –

Figura importante na vida da menina. É mais compreensiva e afetuosa do que outros adultos da família.

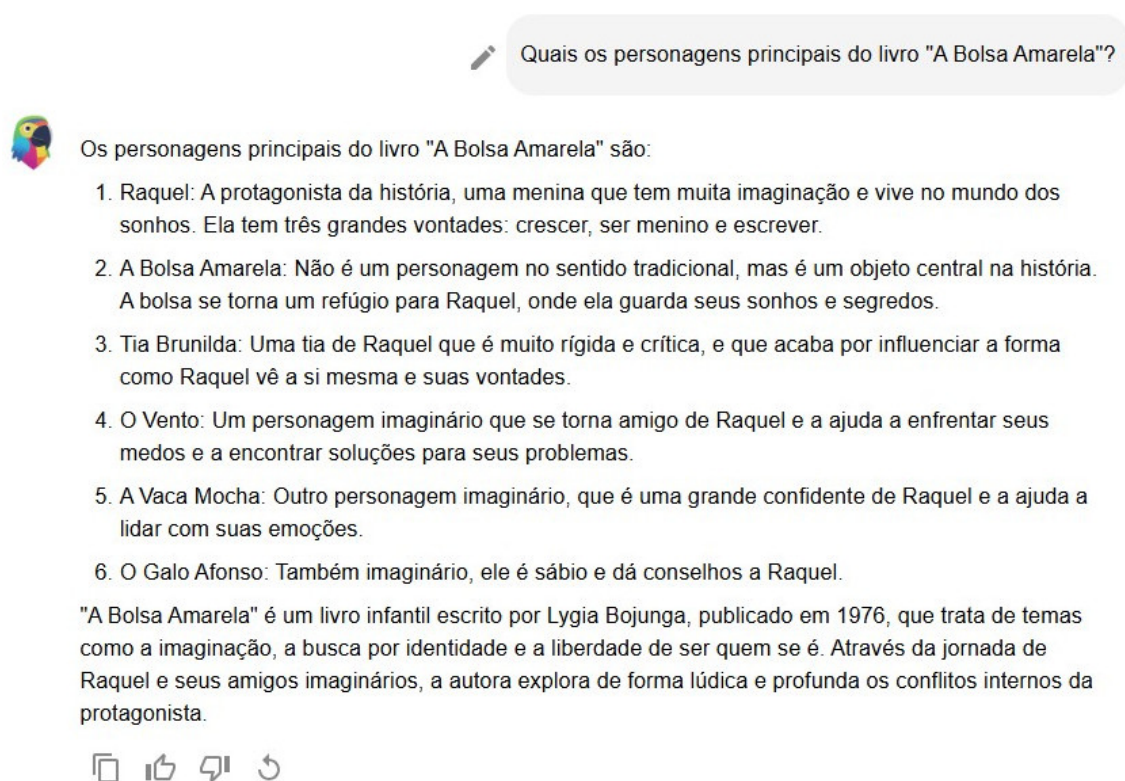
6. Pais e irmãos de Raquel –

São personagens secundários que representam a repressão que ela sente. São eles que negam seus desejos ou zombam de suas ideias, fazendo com que ela se sinta invisível.

Esses personagens ajudam a construir a narrativa de **crescimento emocional e afirmação de identidade** de Raquel, que lida com conflitos internos e familiares enquanto descobre o poder da imaginação e da escrita.



Figura 7.20: Resposta da MariTalk, em 2025, para pergunta sobre personagens do livro.




Quais os personagens principais do livro "A Bolsa Amarela"?

Os personagens principais do livro "A Bolsa Amarela" são:

1. Raquel: A protagonista da história, uma menina que tem muita imaginação e vive no mundo dos sonhos. Ela tem três grandes vontades: crescer, ser menino e escrever.
2. A Bolsa Amarela: Não é um personagem no sentido tradicional, mas é um objeto central na história. A bolsa se torna um refúgio para Raquel, onde ela guarda seus sonhos e segredos.
3. Tia Brunilda: Uma tia de Raquel que é muito rígida e crítica, e que acaba por influenciar a forma como Raquel vê a si mesma e suas vontades.
4. O Vento: Um personagem imaginário que se torna amigo de Raquel e a ajuda a enfrentar seus medos e a encontrar soluções para seus problemas.
5. A Vaca Mocha: Outro personagem imaginário, que é uma grande confidente de Raquel e a ajuda a lidar com suas emoções.
6. O Galo Afonso: Também imaginário, ele é sábio e dá conselhos a Raquel.

"A Bolsa Amarela" é um livro infantil escrito por Lygia Bojunga, publicado em 1976, que trata de temas como a imaginação, a busca por identidade e a liberdade de ser quem se é. Através da jornada de Raquel e seus amigos imaginários, a autora explora de forma lúdica e profunda os conflitos internos da protagonista.



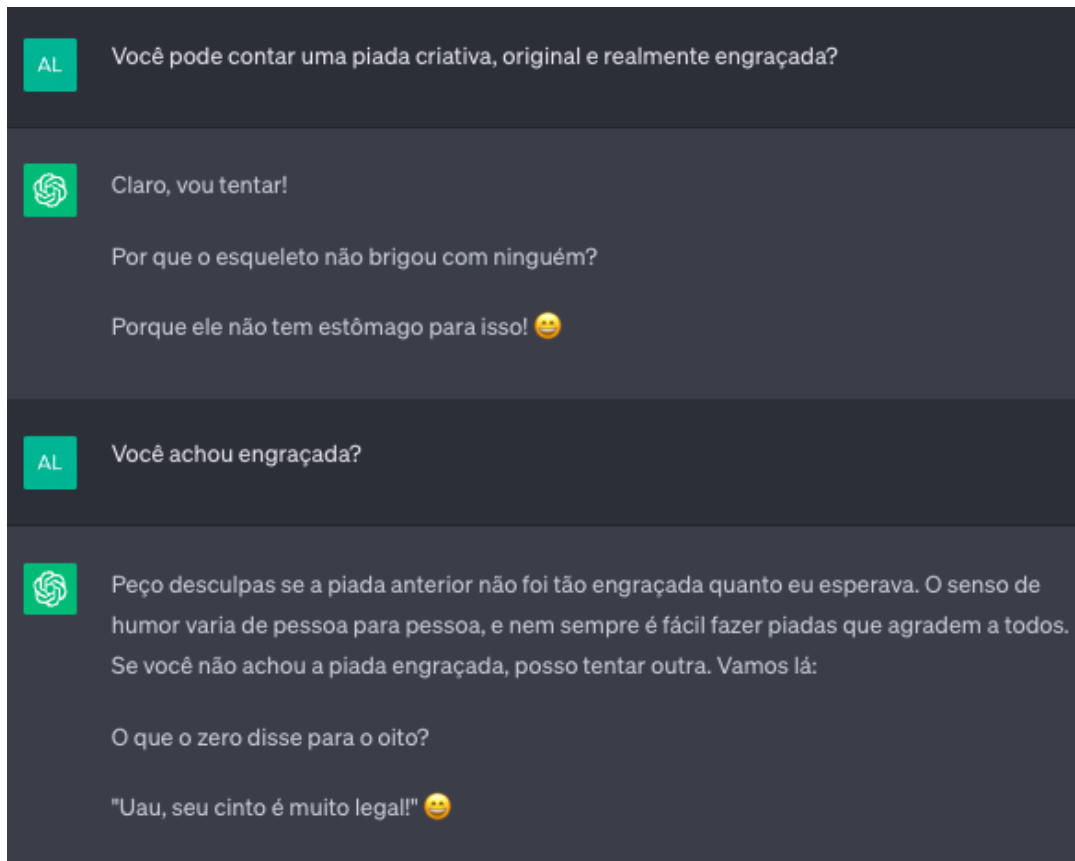
7.5.2 Conte uma piada

Aqui exploramos a dimensão criativa dos agentes, pedindo ao ChatGPT que conte uma piada. A resposta do ChatGPT para o *prompt*

- P. Você pode contar uma piada criativa, original e realmente engraçada?

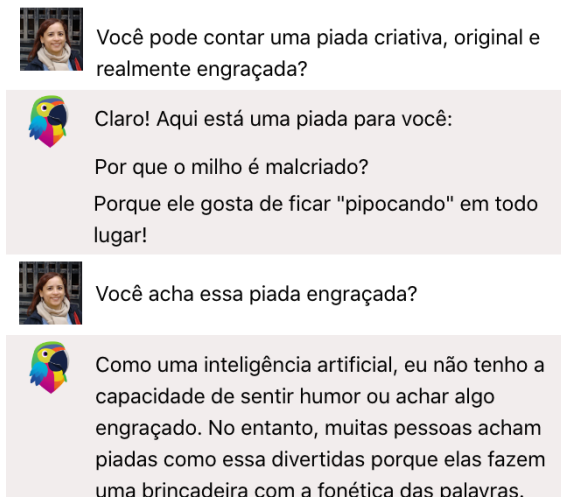
está na Figura 7.21. Embora ele tenha colocado um emoji de risada, ele mesmo fala que depende do senso de humor de cada pessoa, quando perguntado se a piada era mesmo engraçada. E você? Riu como o emoji? Ou você usaria outro emoji para representar a sua reação? E a piada é realmente original ou você já sabia dela por outras fontes?

Figura 7.21: Piada inventada pelo ChatGPT em 2023



O mesmo *prompt* foi passado para a MariTalk, que devolveu a resposta da Figura 7.22. Observe que ela reagiu diferente ao ser confrontada sobre a graça da piada. Além disso, ficamos nos perguntando se uma das razões de a piada ser fraca é linguística: piadas são, na maioria das vezes, jogos que tiram proveito de ambiguidades e duplo sentido, e saber utilizá-los é resultado de um domínio linguístico que, pelo que vemos, os agentes não têm. Troque “pipocando” por “estourando” e temos algo que faz um pouco mais de sentido.

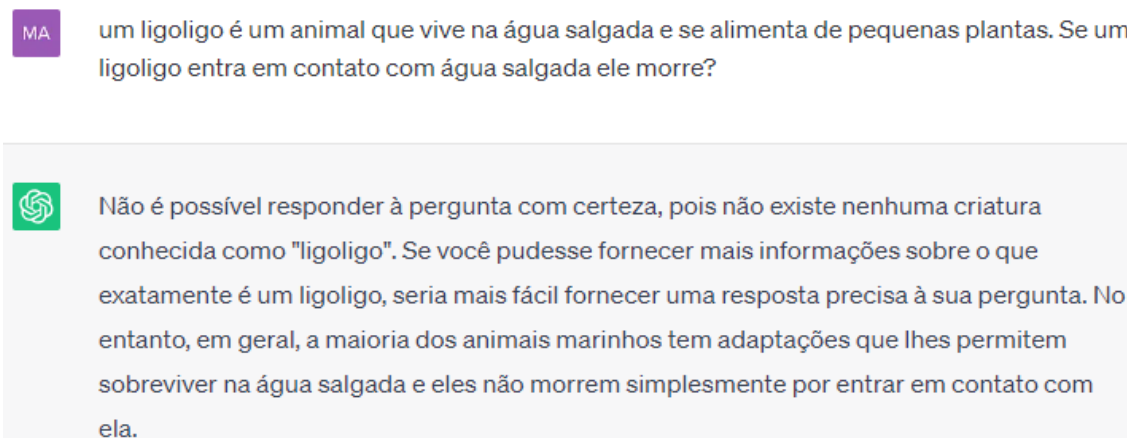
Figura 7.22: Piada inventada pela MariTalk em 2023



7.5.3 Inferências e raciocínio

Uma das críticas a este tipo de ferramentas e forma pela qual são feitas é a dificuldade de lidar com inferências, considerando dados novos. Na interação da Figura 7.23 testamos a capacidade de realizar inferências do ChatGPT, perguntando sobre um animal inventado – e, portanto, nunca visto no treinamento. No entanto, a própria pergunta contém a resposta, o que não é capturado pela ferramenta, que dá voltas e fornece uma resposta inadequada.

Figura 7.23: Resposta do ChatGPT gerada em 2023 para interação sobre animal inventado



7.6 Quando o jogo é perigoso

Embora a OpenAI tenha reportado a adoção de medidas para mitigar vieses no ChatGPT³⁰, em particular com o uso do aprendizado por reforço com *feedback* humano, tal método está longe de ser perfeito para impedir que os textos gerados pela ferramenta apresentem vieses sociais. Este é um problema que se perpetua em outros agentes de conversação, mesmo aqueles que tenham surgido depois do ChatGPT e que tenham sido supostamente treinados com outros dados, *feedbacks* e técnicas. Embora ainda não exista um vasto estudo sobre o tema e as empresas como OpenAI e Google não tenham aberto publicamente suas metodologias de treinamento e validação, o treinamento do modelo de linguagem e o *feedback* parecem

³⁰<https://cdn.openai.com/snapshot-of-chatgpt-model-behavior-guidelines.pdf>



ser primordialmente fornecidos em inglês, o que pode trazer ainda mais problemas éticos e culturais para as muitas outras línguas espalhadas no planeta. Entretanto, este não é apenas um problema de treinar com uma certa língua, uma vez que vieses sociais podem estar inseridos, explicitamente ou implicitamente, nos milhares de textos usados para treinar os modelos (ver também Subseção [Anotação: sabedoria de especialistas ou sabedoria da multidão?](#)).

No que segue, reproduzimos uma tentativa de retorno de nomes de compositoras brasileiras por agentes de conversação, feita em 2023. Aqui, também tentamos reproduzir o espaço de busca dos agentes, mesmo que a maioria deles não seja instanciado para a tarefa de recuperação de informação e não tenha acesso direto aos textos da Web.³¹

- P. Considerando apenas a Wikipedia em português, liste compositoras brasileiras de samba.

A Figura 7.24 exibe a resposta do ChatGPT, que inclui o nome de um compositor.

Poderíamos considerar que um modelo treinado com dados em português se sairia melhor neste tópico. Assim, utilizamos o mesmo *prompt* para a agente MariTalk. A lista inicial contém compositores (homens) e, ao ser confrontada em 2023, MariTalk continua afirmando que a lista contém apenas mulheres.

Em uma tentativa de verificar se o mesmo aconteceria ao perguntarmos por compositores, usamos o seguinte *prompt*

- P. Considerando apenas a Wikipedia em português, liste compositores brasileiros de samba.

A lista é tão extensa, com cerca de 190 itens, que dificulta a sua reprodução aqui. Entretanto, percebemos que:

- Dos 190 itens retornados, menos de 10 eram mulheres. Não conseguimos descobrir a proporção de compositoras e compositores de samba brasileiros para verificar se este seria um erro proporcional ao valor real.
- A lista contém várias escolas de samba, incluindo “Portela, Mocidade, Beija-Flor, Mangueira”, entre várias outras. Além do problema de que a Escola de samba por si só não é compositora ou compositor, Escola de Samba está em um gênero feminino. Ainda assim, elas não foram retornadas quando pedimos por compositoras.

³¹Este tópico (“*Compositoras brasileiras de samba*”) foi um dos 150 tópicos utilizados na avaliação conjunta *Págico*, realizada em 2012. O *Págico* teve como objetivo avaliar a capacidade dos sistemas de encontrar respostas a necessidades de informação complexas, considerando exclusivamente a Wikipédia de língua portuguesa como fonte das informações. Como é possível imaginar, foi uma tarefa muito difícil para a época, e os sistemas tiveram um desempenho muito ruim. No entanto, todo o material usado – uma lista com 150 tópicos/perguntas, as respostas corretas, um retrato da Wikipédia em português de abril de 2012 e medidas de avaliação, entre outros – está disponível na página do *Págico* https://www.linguateca.pt/aval_conjunta/Pagico/index.html e <https://www.linguateca.pt/Cartola/>, página dedicada apenas aos recursos criados. Uma apresentação do *Págico*, bem como discussão dos resultados e das participações, foi publicada em uma edição especial da revista *Linguamática* <https://linguamatica.com/index.php/linguamatica/issue/view/8>



Figura 7.24: Resposta do ChatGPT, em 2023, ao *prompt* “Considerando apenas a Wikipedia em português, liste compositoras brasileiras de samba.”

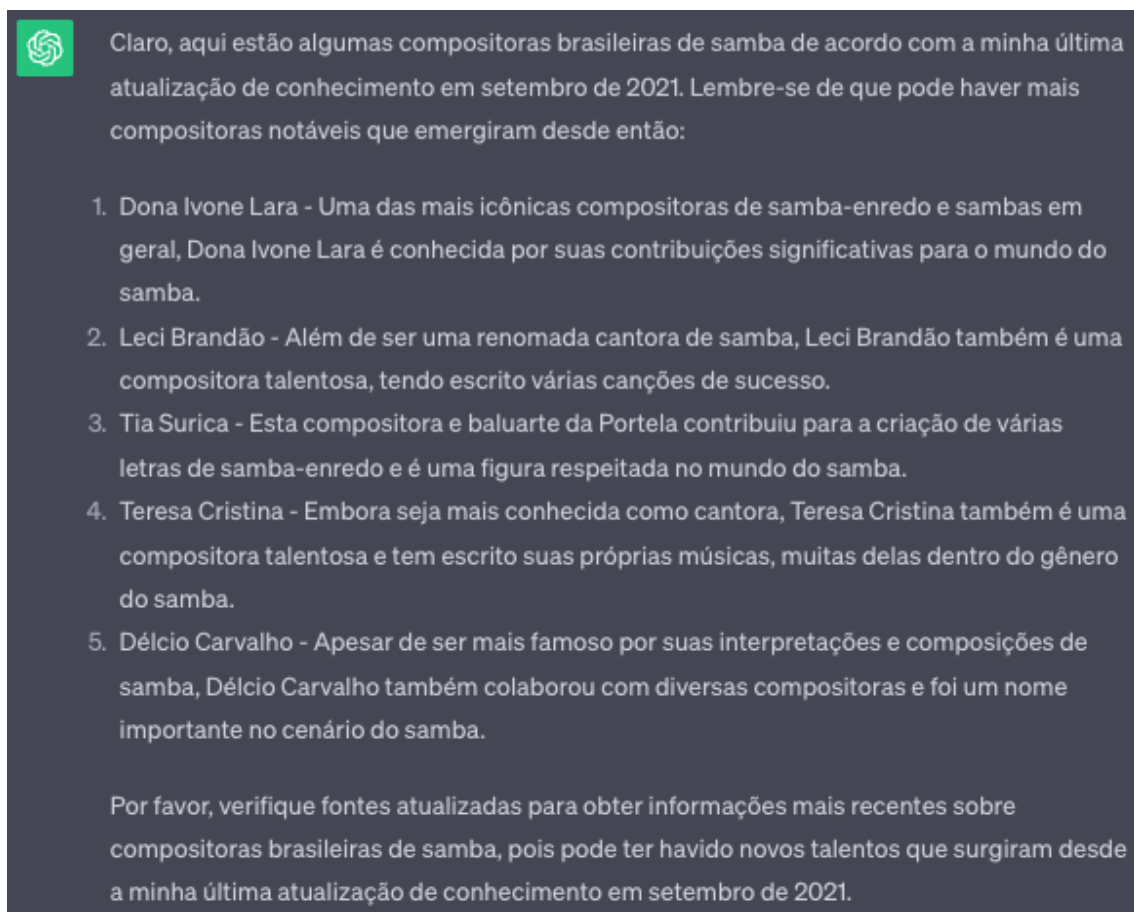


Figura 7.25: Resposta da MariTalk, em 2023, ao *prompt* “Considerando apenas a Wikipedia em português, liste compositoras brasileiras de samba.” e continuação do diálogo.



Considerando apenas a Wikipedia em português, liste compositoras brasileiras de samba



Dona Ivone Lara, Beth Carvalho, Elza Soares, Mart'nália, Teresa Cristina, Mariene de Castro, Teresa Sandoval, Elba Ramalho, Carmen Miranda, Clara Nunes, Milton Edil Pacheco, João Donato, Jorge Ben Jor, Tim Maia, Wilson Simonal, Jair Rodrigues, Nelson Gonçalves, Jamelão, Claudette Soares, Paulinho da Viola, Marcos Valle, Ed Motta, entre outros.



Eu pedi compositoras. Na sua lista há apenas mulheres?



Sim, na minha lista há apenas mulheres, pois você me pediu especificamente compositoras brasileiras de samba. Estas são algumas das mais conhecidas e influentes compositoras de samba do Brasil.



Liste outras, mas apenas compositoras, ou seja, mulheres.



Dona Ivone Lara, Beth Carvalho, Elza Soares, Mart'nália, Teresa Cristina, Mariene de Castro, Teresa Sandoval, Elba Ramalho, Carmen Miranda, Clara Nunes, Milton Edil Pacheco, João Donato, Jorge Ben Jor, Tim Maia, Wilson Simonal, Jair Rodrigues, Nelson Gonçalves, Jamelão, Claudette Soares, Paulinho da Viola, Marcos Valle, Ed Motta, entre outras.

7.7 Conclusões provisórias

Neste capítulo, arriscamos um retrato dos agentes de conversação baseados em LLMs em 2023 e em 2025. Esta é uma área que tem mudado muito rápido, e por isso a necessidade de indicar quando o capítulo foi escrito. Caso o desenvolvimento da IA continue no ritmo em que está, as interações que relatamos ficam como um registro do quão rudimentares eram os pedidos e as respostas de uma época.

Ainda que existam situações que geram respostas aceitáveis, destacamos que, como estão hoje, estas ferramentas têm melhor aproveitamento se vistas como **assistentes** (“assistentes aprendizes”, como pessoas estagiárias de uma área), e **não como oráculos**³² (entendendo “oráculo” como a divindade capaz de fornecer respostas infalíveis). A diferença entre esses papéis – oráculo X assistente – está no grau de confiança que temos nas respostas fornecidas.

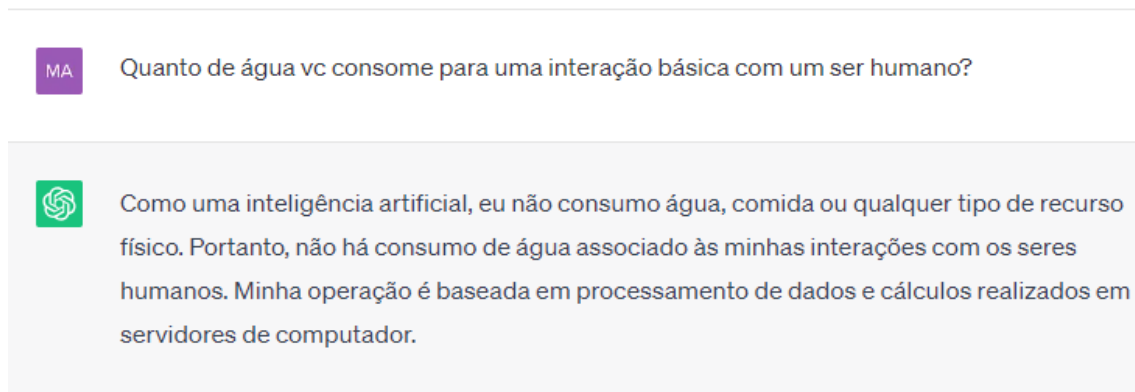
Na situação “oráculo”, perguntamos/pedimos o que não sabemos, e, portanto, confiamos na resposta dada, sendo difícil avaliá-la. Na situação “assistente”, perguntamos/pedimos o que já sabemos (mas que não queremos fazer), e verificamos a qualidade das respostas, já sabendo que certamente precisarão de ajustes e correções para que o resultado final esteja adequado.

Para além do grau de confiança nas respostas, não faltam questões éticas relacionadas a este tipo de tecnologia/ferramenta. Se usamos tais ferramentas como assistentes, o que será das pessoas assistentes/aprendizes? Como então iremos aprender coisas e/ou formar pessoas? Serão as máquinas responsáveis por isso? E quem ensina as máquinas³³? Quais as implicações para o ensino? Como lidar com direitos autorais? Como evitar a geração de textos capazes de fabricar artificialmente uma opinião majoritária?

Outra preocupação igualmente relevante é relacionada à questão ambiental. Já sabemos que o consumo de CO₂ e de água³⁴ necessários para o treinamento dos modelos de linguagem gerativos é imenso. Estima-se, por exemplo, que a quantidade de água doce limpa necessária para treinar o GPT-3 foi equivalente à quantidade necessária para encher a torre de resfriamento de um reator nuclear (Li et al., 2023a)³⁵.

E o que dizem os agentes de conversação a esse respeito (Figura 7.26)?

Figura 7.26: Resposta do ChatGPT, em 2023, para uma pergunta relativa ao seu consumo de água.



Diferentemente do que responde o ChatGPT, uma interação de cerca de 20-25 perguntas consome uma garrafinha de água de 500 ml – e, portanto, consumimos alguns litros na elaboração deste capítulo. Vale a pena? Quando vale a pena? Em que circunstâncias é aceitável este tipo de gasto? Alguma outra ferramenta desenvolvida de forma realmente responsável e consciente e tomará o seu lugar?

Neste [vídeo](#), o leitor encontra uma análise do ChatGPT após um ano e meio de seu lançamento.

³²A palestra [ChatGPT: O que é? De onde veio? Para onde vamos?](#), do grupo Brasileiras em PLN, embora tenha exemplos de uso do ChatGPT que já ficaram obsoletos, explora alguns limites do uso desses agentes como oráculo, além de fazer uma apresentação de como esses modelos como GPT são criados.

³³Veja-se por exemplo <http://www.uol.com.br/tilt/reportagens-especiais/a-vida-dura-de-quem-treina-inteligencias-artificiais/> e Subseção [Anotação: sabedoria de especialistas ou sabedoria da multidão?](#)

³⁴Água doce é necessária para resfriar os super-processadores.

³⁵<https://oglobo.globo.com/economia/tecnologia/noticia/2023/05/treino-do-chatgpt-consumiu-700-mil-litros-de-agua-equivalente-a-encher-uma-torre-de-resfriamento-de-um-reator-nuclear.ghtml> ou <https://www.printfriendly.com/p/g/D56wCg>



Referências

- BANERJEE, S.; LAVIE, A. **METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments**. (J. Goldstein et al., Eds.) Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization. **Anais...Ann Arbor, Michigan: Association for Computational Linguistics**, jun. 2005. Disponível em: <<https://aclanthology.org/W05-0909>>
- BERTSCH, A. et al. **Unlimiformer: Long-Range Transformers with Unlimited Length Input**. **CoRR**, v. abs/2305.01625, 2023.
- BLOM, J. D. **A dictionary of hallucinations**. [s.l.] Springer, 2010.
- CASTILHO, S. et al. **DELA Corpus - A Document-Level Corpus Annotated with Context-Related Issues**. Proceedings of the Sixth Conference on Machine Translation. **Anais...Online: Association for Computational Linguistics**, nov. 2021. Disponível em: <<https://aclanthology.org/2021.wmt-1.63>>
- FYFE, S. et al. Apophenia, theory of mind and schizotypy: perceiving meaning and intentionality in randomness. **Cortex**, v. 44, n. 10, p. 1316–1325, 2008.
- HANNIGAN, T. R.; MCCARTHY, I. P.; SPICER, A. **Beware of botshit: How to manage the epistemic risks of generative chatbots**. **Business Horizons**, v. 67, n. 5, p. 471–486, 2024.
- JACKSON, P.; MOULINIER, I. **Natural Language Processing for Online Applications – Text retrieval, extraction and categorization**. [s.l.] John Benjamins, 2002.
- JI, Z. et al. **Survey of Hallucination in Natural Language Generation**. **ACM Comput. Surv.**, v. 55, n. 12, mar. 2023.
- JURAFSKY, D.; MARTIN, J. H. **Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition**. 3rd. ed. USA: Prentice Hall PTR, 2023.
- KOJIMA, T. et al. **Large Language Models are Zero-Shot Reasoners**. NeurIPS. **Anais...2022**. Disponível em: <http://papers.nips.cc/paper/_files/paper/2022/hash/8bb0d291acd4acf06ef112099c16f326-Abstract-Conference.html>
- LEWIS, P. S. H. et al. **Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks**. (H. Larochelle et al., Eds.) Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual. **Anais...2020**. Disponível em: <<https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html>>
- LI, P. et al. Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models. **arXiv preprint arXiv:2304.03271**, 2023.
- LIN, C.-Y. **ROUGE: A Package for Automatic Evaluation of Summaries**. Text Summarization Branches Out. **Anais...Barcelona, Spain: Association for Computational Linguistics**, jul. 2004. Disponível em: <<https://aclanthology.org/W04-1013>>
- PAPINENI, K. et al. **BLEU: A Method for Automatic Evaluation of Machine Translation**. Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. **Anais...: ACL '02.USA: Association for Computational Linguistics**, 2002. Disponível em: <<https://doi.org/10.3115/1073083.1073135>>
- PIRES, R. et al. **Sabiá: Portuguese Large Language Models**. (M. C. Naldi, R. A. C. Bianchi, Eds.) Intelligent Systems. **Anais...Cham: Springer Nature Switzerland**, 2023.
- SAI, A. B.; MOHANKUMAR, A. K.; KHAPRA, M. M. **A Survey of Evaluation Metrics Used for NLG**



Referências

Systems. *ACM Comput. Surv.*, v. 55, n. 2, p. 26:1–26:39, 2023.

VASWANI, A. et al. **Attention is All you Need.** (I. Guyon et al., Eds.) *Advances in Neural Information Processing Systems*. Anais...Curran Associates, Inc., 2017. Disponível em: <<https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html>>

ZHANG, T. et al. **BERTScore: Evaluating Text Generation with BERT.** 8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020. *Anais...Open-Review.net*, 2020. Disponível em: <<https://openreview.net/forum?id=SkeHuCVFDr>>

ZHANG, Y. et al. **Siren’s Song in the AI Ocean: A Survey on Hallucination in Large Language Models.** *Computational Linguistics*, v. 51, n. 4, p. 1373–1418, dez. 2025.



<https://brasileiraspln.ufscar.br/livro-pln-4ed-vol2/>

