

# Estudo de caso: ChatGPT passou no teste de Turing ?

Victor Eduardo Requia

<sup>1</sup>Universidade do Estado de Santa Catarina (UDESC)

victorrequia@gmail.com

**Resumo.** *O ChatGPT, como uma ferramenta aberta e de grande alcance, tem impactado o cotidiano das pessoas, tanto diretamente quanto indiretamente, em âmbitos profissionais e pessoal. Com a crescente implementação da inteligência artificial na vida das pessoas, surgiram diversas questões sobre o ChatGPT. Algumas dessas questões, tendem a aceitar e até mesmo incentivar o uso do ChatGPT, enquanto outras minimizam suas capacidades e expressam receios quanto à substituição de vários empregos e funções. Para esclarecer os potenciais riscos ou benefícios do ChatGPT, este estudo se aprofundará nas bases filosóficas da inteligência artificial, tentando responder à pergunta feita por Allan Turing em 1950: "Podem máquinas pensar?".*

## 1. Introdução

[Turing 1950] descreveu um novo tipo de jogo. Nesse jogo, conhecido como Jogo da Imitação (Figura 1), temos três papéis envolvidos. Uma pessoa (B), uma máquina (A) e um interrogador (C). O interrogador deve estar em uma sala separada da pessoa e da máquina. O objetivo do jogo, é o interrogador determinar quem dos outros dois envolvidos (A e B) é uma máquina e quem é um ser humano. Para isso, o interrogador é livre para fazer perguntas para os outros jogadores como: "A pode me dizer se A joga xadrez?". Qualquer que seja a máquina ou a outra pessoa que seja A deve responder às perguntas que são dirigidas a A. Toda a comunicação é via teclado e tela, ou um arranjo equivalente (Turing sugeriu um link de teleimpressão) (Dawson Jr., 2007). O objetivo da máquina é tentar fazer com que o interrogador erroneamente conclua que a máquina é a outra pessoa; o objetivo da outra pessoa é tentar ajudar o interrogador a identificar corretamente a máquina.

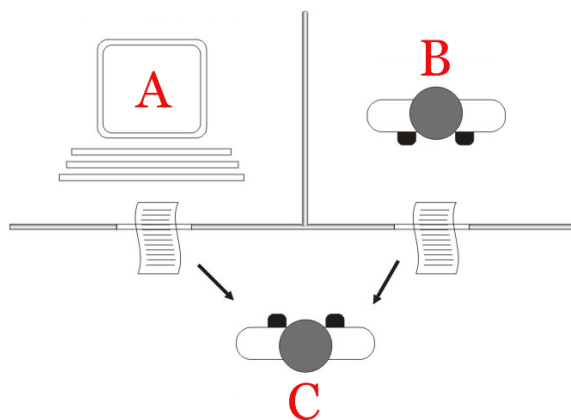


Figura 1. Digrma teste de turing.

Desde que "Computing Machinery and Intelligence" apareceu pela primeira vez, o teste de Turing recebeu considerável atenção de filósofos, cientistas da computação, psicólogos e outros, e várias objeções adicionais foram levantadas ao teste [Dawson 2007]. Uma dessas objeções foi de Shannon-McCarthy que, imaginava um computador hipotético capaz de jogar o Jogo da Imitação com sucesso, por qualquer período de tempo definido, em virtude de incorporar uma tabela de consulta muito grande, mas ainda assim finita. A tabela contém todas as trocas que poderiam possivelmente ocorrer entre o computador e o interrogador durante o período de tempo em que o teste é realizado. O número destas é astronômico, mas finito.

Claramente, o interrogador não teria como distinguir um computador usando esta tabela de um respondente humano. No entanto, o computador que não faz nada além de pesquisar a tabela fornecida por seus programadores (hipotéticos) não pensa. Em princípio, portanto, um computador não pensante e não inteligente pode passar no teste. Ou seja, para qualquer situação de entrada dada (incluindo histórico passado) simplesmente a máquina iria realizar uma consulta em um "dicionário" para a resposta apropriada. Com um dicionário adequado, tal máquina certamente satisfaria a definição de Turing, mas não iria refletir o conceito usual de pensar.

Para essa objeção, e outras que possuem semelhança, Newman (Computing Machinery and Intelligence, Capítulo 14, p. 503) fornece a seguinte resposta: "É muito bom dizer que uma máquina poderia... ser feita para fazer isso ou aquilo, mas, para citar apenas um ponto prático, e quanto ao tempo que levaria para fazer isso? Levaria apenas uma ou duas horas para criar uma rotina para fazer nossa máquina de Manchester analisar todas as possíveis variações do jogo de xadrez e encontrar o melhor movimento dessa maneira - se você não se importasse que levasse milhares de milhões de anos para executar a rotina. Resolver um problema na máquina não significa encontrar uma maneira de fazer isso entre agora e a eternidade, mas em um tempo razoável...". Turing complementa: "Em minha opinião, esse fator de tempo é a única questão que envolverá toda a real dificuldade técnica."

Portanto, a objeção direciona nossa imaginação para um mundo possível que é muito diferente do mundo real, um mundo em que uma tabela de consulta astronômicamente grande pode ser armazenada na memória de um computador e pesquisada em um tempo razoável, e aponta que o princípio do teste de Turing é falso nesse mundo [Dawson 2007].

## **2. Máquinas Pensantes**

[Dawson 2007] reforça que, o alvo de Turing em "Intelligent Machinery, A Heretical Theory" é a afirmação de que "Você não pode fazer uma máquina pensar por você" (p. 472). Um tema comum em seus escritos é que, se uma máquina é inteligente, então ela precisará "aprender com a experiência" (provavelmente com alguma pré-seleção, por um educador externo, das experiências às quais a máquina será submetida). Turing sugere que um processo de educação seria, na prática, essencial para a produção de uma máquina razoavelmente inteligente em um espaço de tempo razoavelmente curto (p. 473). Hoje, essa técnica é conhecida como aprendizagem por reforço e supervisionado.

Como a tecnologia na época de Turing era limitada, uma das propostas de Turing era, ao invés de criar uma máquina capaz de imitar o pensamento de uma pessoa

adulta, era produzir um programa para simular a mente de uma criança. A mente da criança pode conter "tão poucos mecanismos" que "algo parecido pode ser facilmente programado" [Dawson 2007]. Se essa máquina-infantil fosse então submetida a um curso de educação apropriado, teria o cérebro adulto [Dawson 2007]. Essas observações estão em sintonia com a sugestão de Turing de que "o córtex de um bebê é uma máquina desorganizada, que pode ser organizada por treinamento adequado" (Computing Machinery and Intelligence, pg. 424).

Turing menciona que ele "fez alguns experimentos com uma dessas máquinas infantis e conseguiu ensiná-la algumas coisas, mas o método de ensino era muito heterodoxo para que o experimento fosse considerado realmente bem-sucedido" (Computing Machinery and Intelligence, pg. 461). Aqui, ele provavelmente está se referindo aos experimentos com uma máquina desorganizada que são descritos no Capítulo 10, onde ele diz que "conseguiu organizar tal máquina (em papel) em uma máquina universal", mas que a técnica usada "não é suficientemente análoga ao tipo de processo pelo qual uma criança realmente seria ensinada" (Computing Machinery and Intelligence, pg. 427-8).

Como educador, Turing também faz algumas propostas. Uma delas, que hoje é entendida como uma espécie de aprendizado por reforço, diz respeito ao educador, analisar a resposta e dizer se uma situação ou resultado é "favorável" ou não, Turing retorna à possibilidade de incorporar duas "chaves" na máquina, que podem ser manipuladas pelo educador, e que representam "prazer" e "dor". ele considera adicionar duas linhas de entrada a uma máquina de Turing (modificada), a linha de prazer (ou recompensa) e a linha de dor (ou punição). Ele chama o resultado de uma "máquina do tipo P" ("P" representando "prazer-dor") [Dawson 2007].

Portanto, apesar dos conceitos propostos por Turing serem antigos, eles pavimentaram o que hoje seria a Inteligência Artificial moderna e levantaram questões filosóficas importantes a respeito do tema e o que seria o "pensar das máquinas".

### **3. Aleatoriedade**

Turing também menciona que, no processo aprendizagem pode ser incorporado a aleatoriedade dos elementos. Isso resultaria, como ele diz, no comportamento da máquina não sendo completamente determinado pelas experiências às quais foi submetida. Para Turing, "Um elemento aleatório é bastante útil quando procuramos uma solução para algum problema". Ele dá o seguinte exemplo: Suponha, por exemplo, que quiséssemos encontrar um número entre 50 e 200 que fosse igual ao quadrado da soma de seus dígitos, poderíamos começar em 51, depois tentar 52 e assim por diante até obtermos um número que funcionasse. Alternativamente, poderíamos escolher números aleatoriamente até obtermos um bom [Dawson 2007].

O método sistemático tem a desvantagem de que pode haver um enorme bloco sem soluções na região que deve ser investigada primeiro. Agora, o processo de aprendizado pode ser visto como uma busca por uma forma de comportamento que satisfaça o "professor" (ou algum outro critério). Como provavelmente há um grande número de soluções satisfatórias, o método aleatório parece ser melhor do que o sistemático. Deve-se notar que ele é usado no processo análogo da evolução [Dawson 2007].

A discussão de Turing sobre "sistemas de prazer-dor" no Capítulo 10 também menciona aleatoriedade [Dawson 2007]. Turing comenta: "Usarei este termo [sistema

”prazer-dor”] para significar uma máquina desorganizada do seguinte caráter geral: As configurações da máquina são descritas por duas expressões, que podemos chamar de expressão de caráter e expressão de situação. O caráter e a situação em qualquer momento, juntamente com os sinais de entrada, determinam o caráter e a situação no próximo momento. O caráter pode estar sujeito a alguma variação aleatória. A interferência do prazer tem uma tendência a fixar o caráter, ou seja, a impedir que ele mude, enquanto os estímulos de dor tendem a perturbar o caráter, fazendo com que características que haviam se fixado mudem ou voltem a estar sujeitas a variação aleatória”.

#### **4. ChatGPT**

Com o avanço contínuo das tecnologias, especialmente no âmbito do software e do hardware, e com o aprimoramento das GPUs, a inteligência artificial tem se consolidado como um pilar central na revolução tecnológica contemporânea. Nesse contexto, o ChatGPT emerge como uma ferramenta revolucionária, conquistando destaque global devido à sua capacidade avançada e inovadora como modelo de linguagem. ChatGPT é uma ferramenta de processamento de linguagem natural, desenvolvida pela OpenAI. Ele é baseado em uma rede neural treinada com milhões de textos da internet, permitindo que ele gere textos de forma autônoma [Rossoni and Chat 2022].

Este modelo, de natureza generativa e pré-treinada, é fundamentado na arquitetura Transformer, que representa um marco no campo do processamento de linguagem natural. Através dessa estrutura, o ChatGPT é capaz de compreender e gerar textos com uma fluidez e precisão sem precedentes, tornando-se uma ferramenta essencial para diversas aplicações, tais como chatbots, geração automática de conteúdo, tradução automática, entre outras [Rossoni and Chat 2022]. É uma das ferramentas de IA mais avançadas disponíveis no mercado, e seus desenvolvedores estão constantemente trabalhando para melhorar sua precisão e capacidade de compreensão do idioma natural [Rossoni and Chat 2022].

#### **5. ChatGPT e o Teste de Turing**

A evolução da inteligência artificial (IA) tem gerado debates significativos em diversos setores, especialmente no acadêmico. A capacidade de máquinas, como o ChatGPT, de gerar textos coerentes e lógicos, levanta questões sobre a distinção entre a escrita humana e a gerada por IA. Esta discussão não se limita apenas à qualidade da escrita, mas também às implicações éticas e práticas que tais avanços trazem para a comunidade acadêmica. Destacando estudos e experiências que colocam em xeque a distinção entre a autoria humana e a máquina, indo ao encontro do proposto inicial do Jogo da Imitação, proposto por Turing e explicado na seção 1 deste trabalho.

Um dos exemplos de como o uso do ChatGPT tem revolucionado a forma como olhamos para a inteligência artificial, é por meio do artigo escrito por Rossani e Luciano (2022) na qual, o ChatGPT foi coautor do artigo e, para os dois outros escritores, a experiência foi tão surpreendente que, eles propuseram que os leitores tenta-sem adivinhar quem tinha escrito o parágrafo, o ChatGPT ou os autores (humanos), fazendo uma espécie da Jogo da Imitação. Nesta brincadeira, na minha visão, não foi possível distinguir o que foi escrito pelo ChatGPT e o que foi escrito pelo autor.

No meio acadêmico, também existe essa preocupação em relação ao assunto.

Em 2023, Nature e [Else 2023] destacaram que o ChatGPT tem a capacidade de elaborar resumos de artigos científicos tão persuasivos que, frequentemente, especialistas não conseguem distingui-los, conforme pesquisa divulgada no bioRxiv. Tal habilidade gera inquietações quanto ao impacto da inteligência artificial na avaliação da integridade acadêmica e na confiabilidade da ciência. Há pesquisadores que expressam receios sobre a eliminação do intermediário humano, essencial para orientar a compreensão de assuntos complexos [Rossoni and Chat 2022].

Este debate ficou tão sério que, em um editorial divulgado na Nurse Education in Practice, O'Connor e ChatGPT (2023) destacam a crescente dificuldade em diferenciar textos escritos por estudantes daqueles gerados por robôs de chat de inteligência artificial. Tal desafio se torna ainda mais acentuado em níveis de pós-graduação, onde se espera a elaboração de teses e dissertações minuciosas. Como solução, os autores propõem a adoção de métodos avaliativos variados, como apresentações verbais e exames clínicos estruturados, bem como a utilização de softwares antiplágio avançados e a promoção da educação voltada à integridade acadêmica e à valorização do saber [Rossoni and Chat 2022].

Ainda segundo Rossani e Luciano, ainda estamos nos estágios iniciais para determinar o alcance total da transformação que a inteligência artificial trará à ciência. No entanto, já observamos indícios claros de seu impacto iminente. Ferramentas como o ChatGPT, capazes de produzir textos articulados, exemplificam como a IA pode otimizar a redação de publicações científicas, economizando tempo e esforço dos pesquisadores. Além disso, a IA tem o potencial de processar e analisar vastas quantidades de dados, criar modelos preditivos e revelar padrões e tendências emergentes. Estamos à beira de uma revolução científica impulsionada pela inteligência artificial, e é crucial que a comunidade científica esteja pronta para abraçar e maximizar os benefícios dessa inovação.

No meu entendimento, nestes casos, a máquina teria passado no Teste de Turing porém, nesse teste, a máquina não seria necessariamente inteligente, pois o que ela fez, foi apenas consultar artigos já publicados, com todo o teor científico e fazer mudanças textuais e modificar os textos com base em outros diversos conteúdos com os quais ela foi treinada. Creio que, para ser inteligente, a máquina teria que ter certo grau de criatividade, opinião própria, fazer induções e hipóteses, apenas com conteúdo científico, sem opiniões e proposta de autores.

Um exemplo que eu iria considerar a máquina inteligente é, em um real teste de Turing, informar para a máquina que ela está participando do teste de Turing e informar apenas as regras do jogo. Com isso, após ela receber uma pergunta, exemplo: "Leia 50 artigos e faça o resumo deles". Ao invés dela fazer essa tarefa no menor tempo possível para entregar uma resposta, a inteligência artificial iria criar a hipótese de que a pessoa que está disputando com ela, levaria um certo tempo para ler os artigos, mais um certo tempo para escrever o resumo, teria alguns erros ortográficos, criatividade e opinião própria para colocar experiências que a pessoa já teve, relacionadas aos assuntos dos artigos. Com isso, a inteligência artificial iria tentar imitar esses comportamentos humano de forma própria e autêntica.

## 5.1. Notícias recentes

Com o ChatGPT ganhando destaque nas mídias, notícias sobre ChatGPT e o teste de Turing, tem ganhado relevância e curiosidade por parte dos leitores, instigados a saber, se o ChatGPT passou no teste de Turing.

Na notícia escrita por [mpo ], afirma que o ChatGPT passou no teste de Turing, o Chatbot foi capaz de enganar uma equipe de jurados e imitar o comportamento humano na escrita e lógica nas respostas. Essa notícia, apesar da afirmação, não possui dados que comprovem esse teste nem o grau de acerto ou erro do ChatGPT. É importante ressaltar isso, pois do jeito que foi proposto o teste de Turing inicialmente, o julgamento fica totalmente qualitativa, sem o apoio de dados, comparações ou métricas. Acredito que o teste falha nesse aspecto pois abre margem para diferentes opiniões e notícias tendenciosas.

Outra notícia [Biever 2023], instiga o comportamento humano da máquina, modificando o teste de Turing para além da escrita e interpretação de texto, na qual os autores consideram superado pela máquina. O novo teste, propõe responder a questão "O que as máquinas não conseguem fazer?" e para isso, a máquina teria que resolver um simples quebra-cabeça de lógica visual.

Em uma avaliação que apresenta uma sequência de blocos coloridos em uma tela, a grande maioria das pessoas é capaz de reconhecer os padrões interligados. Contudo, o GPT-4, a iteração mais sofisticada do sistema de IA que alimenta o chatbot ChatGPT e o motor de busca Bing, consegue resolver corretamente somente aproximadamente um terço dos puzzles em uma determinada categoria de padrões e meros 3% em outra [Biever 2023], conforme indicado por um estudo de especialistas. Ou seja, testados de uma forma, eles superam facilmente o que antes eram considerados marcos notáveis da inteligência de máquina. Testados de outra forma, parecem menos impressionantes, exibindo pontos cegos evidentes e uma incapacidade de raciocinar sobre conceitos abstratos [Biever 2023]. Neste caso, como o modelo do ChatGPT não foi treinado para identificar figuras ou resolver problemas visuais, o modelo não obteve grandes resultados nos teste. Ou seja, a "inteligência" dele, depende de como ele foi treinado e qual o propósito do modelo. Acredito que em um futuro próximo, essas dificuldades serão superadas e teremos uma inteligência artificial que irá superar alguns dos desafios atuais como fala, visão e identificação de figuras abstratas.

## 6. Conclusão

Sob meu ponto de vista, atualmente o ChatGPT versão 3, tem um grande potencial para passar no Teste de Turing proposto inicialmente porém, certas características precisam ser levadas em conta. Ainda que o modelo do ChatGPT, seja muito poderoso, ainda temos algumas limitações. O contexto da conversa e o assunto são de extrema relevância para simular ou imitar um humano.

Quando nos deparamos com questões que, até para seres humanos, não possuem uma definição clara ou cujo tema é intrincado e ainda se encontra em estágios iniciais de pesquisa científica, o modelo pode enfrentar desafios devido à falta de conteúdo específico e definições claras em sua base de treinamento. Um exemplo notório é em tópicos matemáticos que envolvem cálculos avançados e detalhados. Em muitas situações, a resposta pode ser imprecisa ou divergir de uma solução ideal. Contudo, essa falha não necessariamente impede o ChatGPT de superar o Teste de Turing. Se um interrogador propõe

um cálculo complexo, esperando que a máquina acerte por presumir sua precisão, e a máquina erra, o interrogador pode interpretar a resposta como vinda de um humano sem expertise em matemática avançada.

Quanto à questão do contexto, é um equívoco pensar que a máquina possui memória infinita. Se uma conversa se estende além da capacidade de memória do ChatGPT, o contexto inicial pode ser esquecido. Assim, se o interrogador retomar tópicos discutidos antes da capacidade máxima de memória ser atingida, o ChatGPT pode não responder corretamente. No entanto, isso não significa uma falha definitiva no Teste de Turing. O modelo pode simular um humano com lapsos de memória, deixando a critério do interrogador avaliar a veracidade da resposta.

## Referências

Chatgpt passes the turing test.

Biever, C. (2023). Chatgpt broke the turing test — the race is on for new ways to assess ai. *Nature*, 619(7971):686–689.

Dawson, Jr, J. W. (2007). *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life plus The Secrets of Enigma*.

Else, H. (2023). Abstracts written by chatgpt fool scientists. *Nature*.

Rossoni, L. and Chat, G. P. T. (2022). A inteligência artificial e eu: escrevendo o editorial juntamente com o chatgpt. *Revista eletrônica de ciência administrativa*, 21(3):399–405.

Turing, A. (1950). Computing machinery and intelligence. *Mind*, 59(236):433–460.