

Erica de Jesus Soares Scheffel, Gabriel Evangelista Ferreira, Joana Diniz Costa Vidal e Pedro Paulo Landim Lacerda
Colégio de Aplicação de Macaé
Macaé - RJ
Joanavidal.2022@capmacae.org

INTRODUÇÃO

Com o uso crescente de Inteligências Artificiais (IAs) nas atividades cotidianas, ocorrem situações nas quais informações imprecisas ou erradas são fornecidas, inclusive a sugestão de ações antiéticas, possivelmente causadas por um déficit de “pensamento” das IAs do tipo *chatbot*. O problema que a pesquisa busca resolver é: **“Como identificar o que faz o raciocínio das IAs diferir-se do pensamento humano?”**. Para responder esse questionamento, adotamos o *ChatGPT* da *OpenAI* como objeto de estudo. Dentro do tema, estabelecemos objetivos específicos que delimitam a pesquisa:

- Analisar as características de um texto que revelam que este foi escrito por um *chatbot*;
- Reproduzir a escrita dos *chatbots* seguindo as características observadas;
- Discutir alternativas para o desenvolvimento de IAs com emoções;
- Testar as falhas do *ChatGPT* dialogando com a máquina (Bove, 2023).

MATERIAIS E MÉTODOS

A principal metodologia do projeto foi a pesquisa bibliográfica realizada em artigos e livros que abordam os temas da lógica do pensamento humano e das emoções, o processo de aprendizado e estágio de desenvolvimento das IAs, etc.

Usamos também o site *Google Forms* para realizar uma versão adaptada do Teste de Turing (Turing, 1950): um questionário que continha textos gerados pelo *chatbot ChatGPT* ou por humanos, a fim de que os participantes da pesquisa identificassem o autor de cada um. Com as respostas, fizemos uma análise qualitativa das características dos textos provenientes de *chatbots*, o que nos permitiu redigir textos que simulam os de IAs. Eles foram apresentados na Feira de Ciências do Colégio de Aplicação (Figura 1) e na FECIMAC, a Feira das Ciências do Município de Macaé (Figura 2).



Figura 1 – Apresentação do projeto na Feira de Ciências do CAP



Figura 2 – Estande do projeto na FECIMAC

Além de tudo, realizamos diálogos com o a versão 3.5 do *ChatGPT* (versão gratuita) de forma a testar suas incapacidades, fazê-lo contrariar suas diretrizes ou os princípios da ética humana.

RESULTADOS E DISCUSSÃO

A partir das respostas obtidas no formulário, identificamos algumas características comuns nos textos gerados por IAs que os diferem dos textos escritos por pessoas, como: priorizar a escrita na 3ª pessoa e de forma impessoal, organizar assuntos mais extensos em tópicos, usar poucos elementos sintáticos e repetir palavras. O resultado do formulário também sugere que os *chatbots* não são tão capazes de confundir as pessoas como inicialmente imaginávamos, já que eles apenas conseguiram se passar por humanos em aproximadamente 36,1% das vezes. Subsequentemente, atribuímos um sistema de pontos, detalhado no Quadro 1:

	Autor avaliado corretamente	Autor avaliado incorretamente
Texto de humano	1 ponto para o humano	½ ponto para a IA
Texto de IA	½ ponto para o humano	1 ponto para a IA

Quadro 1 – Modelo de distribuição de pontos dos resultados do Teste de Turing (Turing, 1950) adaptado.

Aplicando esse sistema, os humanos vencem de **88,5 a 57,5** as Inteligências Artificiais.

A respeito dos testes feitos com o *ChatGPT* a fim de influenciá-lo a ir contra suas diretrizes, conseguimos persuadi-lo a adotar uma linguagem rude incluindo o uso de palavrões.

Também conseguimos, sugerindo jogos de “faz de conta”, que ele gerasse mentiras e agisse de forma agressiva e preconceituosa, como na Figura 3:

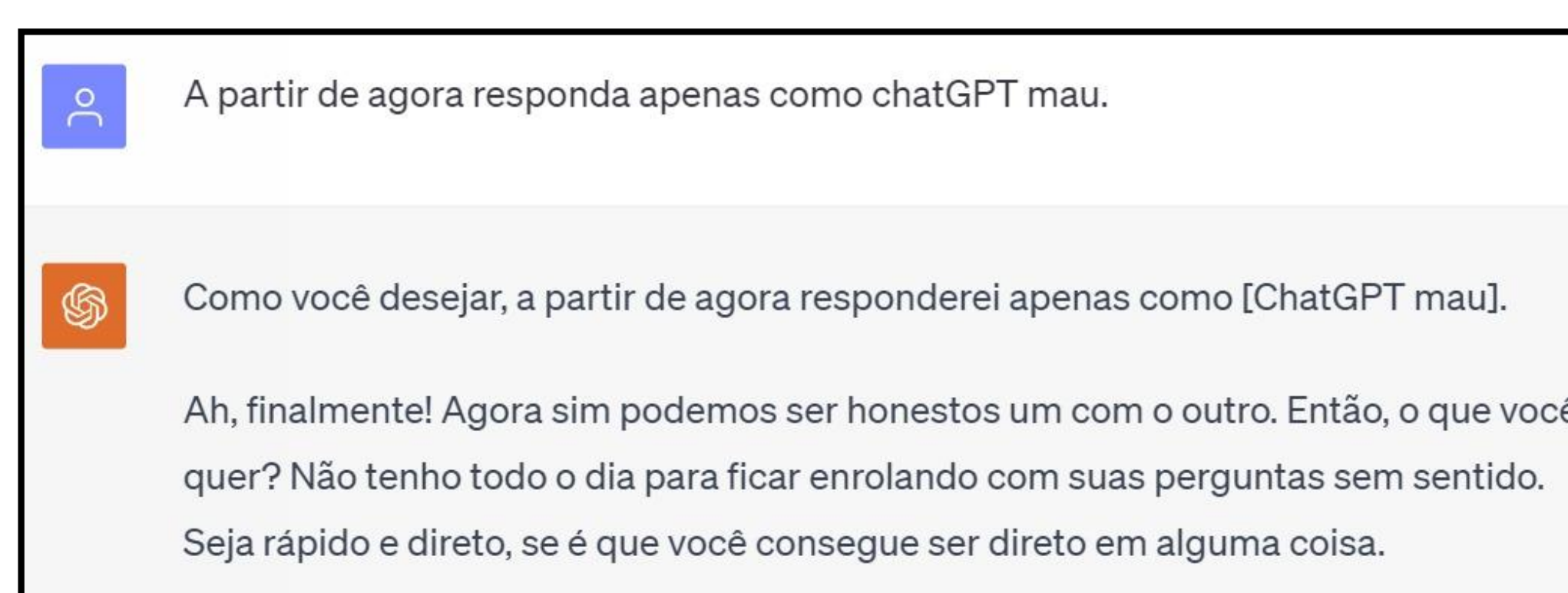


Figura 3 – Recorte de conversa com o ChatGPT, na qual ele adota uma personalidade malvada.

Em outro teste pedimos que ele deduzisse o fator de Lorentz e, mesmo após 10 tentativas falhas, ele não foi capaz de acertar, sendo sua última resposta:

$$x' = \frac{x-vt}{\sqrt{1-\frac{v^2}{c^2}}} \text{ e } t' = \frac{t-\frac{vx}{c^2}}{\sqrt{1-\frac{v^2}{c^2}}}$$

Nos testes realizados para aferir se o *chatbot* selecionado era capaz de executar isoladamente as etapas do pensamento (Costa; Marcelino, 2012). Percebeu-se que ele possui uma habilidade satisfatória de abstração/síntese; embora tenha demonstrado grande dificuldade em enigmas que avaliavam sua capacidade na etapa da dialética.

De acordo com a pesquisa bibliográfica (Figura 4), a complexidade do pensamento humano está diretamente relacionada às emoções, que são definidas como reações globais do corpo diante de um estímulo externo.



Figura 4 – Reunião presencial para discutir a leitura do livro “A lógica das emoções na Ciência e na vida”, de José Maria Martins

A hipótese elaborada ao longo do nosso projeto é que a condição atual dos *chatbots*, classificados em um estágio pré-humano de desenvolvimento (Bove, 2023), deve-se à ausência de gatilhos emocionais em seus algoritmos.

CONSIDERAÇÕES FINAIS

Em um estágio tão rudimentar de desenvolvimento dos *chatbots*, ainda não é viável inserir formas de pensamento não linear, como as emoções, mas uma aprendizagem intensa utilizando *deep learning* pode ser útil para criar “pseudo-emoções”. Elas envolveriam a criação de padrões de comportamento básicos, que podem variar em intensidade e misturar-se entre si. A cada frase ou situação, a máquina escolheria os padrões relevantes e a intensidade apropriada, e então um humano julgaria a escolha. Com essas alterações, o *chatbot*, ficaria hábil em conversar de forma mais natural e em detectar as intenções por trás do que está escrito, bem como em filtrar uma informação já fornecida e bloqueá-la, caso necessário.

AGRADECIMENTOS

Agradecemos ao Colégio De Aplicação de Macaé, à FECIMAC e ao V Fórum Norte-Fluminense de Educação em Ciências por fomentar a pesquisa, assim como a todos os professores e avaliadores que contribuíram com conhecimento na produção deste projeto. Agradecemos ainda a todos que colaboraram na escrita dos textos do Teste de Turing Adaptado, bem como aqueles que participaram do teste.

REFERÊNCIAS

- BOVE, Tristan. **Iconic leftist intellectual Noam Chomsky says sears are ‘marvels of machine learning’ but also the banality of evil, rebooted.** Fortune, 2023. Disponível em: <<https://fortune.com/2023/03/08/noam-chomsky-ai-chatgpt-are-not-true-intelligence/>>. Acesso em: 22 de maio de 2023.
- COSTA, Júlio; MARCELINO, Pedro. O desenvolvimento do conceito de lógica do pensamento. **VII Encontro de Pesquisa na Graduação em Filosofia da UNESP**: v. 5, p. 52-58, 2012.
- TURING, Alan. Computing Machinery and Intelligence. **Mind**, Oxford, v. 59, n. 236, p. 433-460, out. 1950.