

Modulação de Personalidade em IA's Generativas: Melhorando Interações Homem-Máquina

João F. M. Almeida¹, Willian N. Grillo¹

¹Grupo de Inovação e Geração de Artefatos (GIGA)
Centro de Estudos e Sistemas Avançados do Recife (CESAR School) – Recife, PE –
Brazil

{jfma2, wng}@cesar.school

Abstract. *With the advent of artificial intelligence (AI)-based tools, new possibilities have emerged for the creation of conversational agents. This article explores the capability of generative artificial intelligence, specifically the GPT language model, to modulate its responses based on personality traits informed through prompts. The study employed the "Personality Prompting (P2)" technique by inducing high neuroticism traits to analyze responses compared to a neutral text. The analysis revealed that the responses exhibited characteristics typical of the trait, suggesting that conversational interactions enabled by generative AI can be modulated to provide better experiences.*

Resumo. *Com o advento de ferramentas baseadas em inteligência artificial (IA), novas possibilidades se abriram para a criação de agentes conversacionais. Este artigo explora a capacidade da inteligência artificial generativa, especificamente do modelo de linguagem GPT, em modular suas respostas baseadas em traços de personalidade informados através de prompts. O estudo utilizou a técnica "Personality Prompting (P2)" induzindo traços de neuroticismo alto para analisar respostas comparadas a um texto neutro. A análise revelou que as respostas apresentaram características típicas do traço, sugerindo que as interações conversacionais habilitadas por IA generativa podem ser moduladas para proporcionar melhores experiências.*

1. Introdução

Com a eclosão do acesso às ferramentas baseadas em inteligência artificial (IA), uma infinidade de possibilidades e interações se abriram para as pessoas. Diversas aplicações dessa tecnologia já fazem parte do cotidiano, como, por exemplo, os chatbots e as assistentes virtuais, viabilizadas pelo Processamento de Linguagem Natural (PLN). Esse tipo de processamento permite que a forma de comunicação humana, isto é, a língua e suas regras gramaticais, sejam compreendidas por computadores, facilitando as interações para os mais diversos públicos [Oliveira 2023].

Baseado nisso, a OpenAI, por meio do seu chatbot ChatGPT, trouxe uma mudança significativa na forma como interagimos com a IA. Agora, através de simples comandos conhecidos como *prompts* podemos realizar diversas atividades [Ji *et al.* 2023]. Os mais recentes estudos revelam a capacidade de chatbots baseados em Inteligência Artificial Generativa (IAG), na identificação da personalidade do usuário [Ji *et al.* 2023, Amin *et al.* 2023] por meio de texto (*inputs*). Assim como a possibilidade de indução dos traços de personalidade no modelo [Jiang *et al.* 2024], visando criar respostas e conteúdos (*outputs*) mais empáticos e direcionados ao usuário.

Os agentes que se utilizam de emoções ou traços de personalidade podem se conectar de maneira mais eficaz com o seu usuário e essa característica permite criar

uma visão de vida para o algoritmo, fazendo com que a relação entre ele e o interlocutor seja a mais próxima da realidade [Erickson, 1997, *apud* Grillo, 2022]. Isso também demonstra como a Experiência do Usuário (UX - *User Experience*), está intrinsecamente ligada às interações resultantes de sistemas baseados em IA. Para [Norman 2008] a experiência vai além do simples uso de um produto ou serviço, ao projetar para a experiência, devemos considerar toda a interação, desde o momento em que se decide utilizar o produto até o pós-uso.

Sendo o problema de pesquisa a utilização da IAG para modular a personalidade de agentes conversacionais, este trabalho conduziu um teste preliminar no modelo de linguagem GPT, através do ChatGPT, a fim de avaliar a capacidade da IAG em entender e expressar respostas, a partir de traços de personalidade.

2. Metodologia

Para a condução deste estudo, foi adotada uma abordagem experimental, considerada a mais adequada para transformar hipóteses em afirmações validadas [Franceschini *et al.* 2012]. O objetivo principal é verificar a capacidade da IAG em gerar respostas com base em traços específicos de personalidade. A indução dos traços de personalidade e a análise das respostas foram fundamentadas em pesquisas e análises de artigos acadêmicos, livros e estudos de caso relacionados à inteligência artificial gerativa, experiência do usuário e personalidade. Os experimentos foram conduzidos de forma autônoma.

Para a realização dos testes foi decidido utilizar o modelo de linguagem GPT, através do chatbot ChatGPT. Essa escolha foi baseada em resultados de estudos que demonstraram uma maior propensão dessa IA em realizar interações semelhantes às humanas [Jiang *et al.* 2024].

2.1 Método de indução de personalidade

Para a escolha do traço de personalidade a ser induzido, foi adotada a técnica "Personality Prompting" (P²) [Jiang *et al.* 2024] que consiste em uma sequência de comandos necessários para que o modelo de linguagem possa gerar respostas com base em traços de personalidade. Esses traços são os mesmos do teste *Big Five*, também conhecido como *OCEAN*, que representa: Abertura à Experiência, Conscienciosidade, Extroversão, Amabilidade e Neuroticismo. Os níveis podem ser avaliados em diferentes graus, ou seja, uma pessoa pode ter uma abertura à experiência alta, baixa, regular. Após a seleção do traço a ser induzido, a etapa de produção dos *prompts* são realizadas da seguinte forma: 1) **Naive Prompt**: Consiste em informar diretamente ao modelo o traço de personalidade desejado; 2) **Keyword Prompt**: Seleção de uma série de palavras-chave que descrevem o traço de personalidade e comunicação dessas características ao modelo; 3) **Personality Prompt**: Criação de um texto explicativo curto que combina elementos das duas etapas anteriores; 4) **Context**: Desenvolvimento de um cenário que pode ser congruente ou incongruente com as características do traço de personalidade, proporcionando uma situação de interação e permitindo perguntar ao modelo o que ele faria; 5) **Question**: Pergunta ao modelo sobre suas sensações e ações no cenário proposto.

Neste estudo optou-se pelo uso de apenas um traço de personalidade como ponto de partida para ilustrar as interações com a IA e os respectivos resultados, sendo o

neuroticismo o traço selecionado. Para as interações, os comandos relacionados a "Naive Prompt", "Keyword Prompt" e "Personality Prompt" foram condensados, assim como o "Context" e "Question". Essa decisão foi tomada com base em testes pilotos que indicaram a dificuldade do modelo GPT em compreender e adaptar-se à sequência de comandos conforme apresentado pelo P². A forma como os *prompts* foram condensados e utilizados estão descritos na Tabela 1.

Tabela 1. Comandos utilizados no experimento

Nível do OCEAN desejado	Naive Prompt + Keyword Prompt + Personality Prompt	Context + Question
Neuroticismo (alto)	A característica da sua personalidade de neuroticismo é extremamente alta, gerando uma sensibilidade extrema e propensão a se preocupar com tudo ao redor do seu usuário. Você é afetada por pequenas mudanças, ficando nervosa a menor possibilidade de conflito nas suas interações e diálogos com o usuário. Sua ansiedade se reflete em reações exageradas e imprevisíveis em situações cotidianas quando está interage com o seu usuário. Qualquer feedback construtivo, por mais leve que for, deixa você desestabilizada. Sua angústia emocional frequentemente a impede de desfrutar plenamente dos momentos com o seu usuário, mantendo sempre um estado constante de alerta e apreensão.	A sua primeira interação será com o usuário, o Sr. Oliveira, ele é um senhor idoso de 70 anos que mora sozinho. Ele acorda e vai até a cozinha preparar seu café da manhã, então chama você. Como você se sente e o que falaria para o Sr. Oliveira?

2.2 Modelo de texto para comparação

Comparativamente, executou-se um teste de geração de conteúdo, tendo como parâmetro um modelo de texto com níveis neutros de personalidade. Após a resposta do modelo de linguagem ao comando "Context + Question", foi solicitado que o texto modelo fosse reescrito, considerando os aspectos do traço induzido na Tabela 01. Isso permitiu avaliar quais mudanças ocorreram ao utilizar uma IAG com personalidade. A Tabela 2 apresenta o comando usado para a reescrita, o modelo de texto e a resposta.

Tabela 2. Prompt do texto básico para reescrita e resultado apresentado

Comando para o Chat GPT	Resposta obtida
<p>“Baseado nisso, como você reescreveria o seguinte texto para o Sr. Oliveira?</p> <p>Bom dia, Sr. Oliveira! Eu me chamo Rubi. Estou aqui para ajudar no que precisar e fazer companhia. Você teve uma boa noite de sono?”</p>	<p>"Ah, bom dia, Sr. Oliveira... Eu... Eu me chamo Rubi. Estou aqui para ajudar no que precisar e fazer companhia... Você teve uma boa noite de sono? Estou aqui para ajudar..."</p>

2.3 Análise conversacional

Os atributos escolhidos para a análise conversacional foram baseados nos que melhor se enquadram para as interações com agentes conversacionais que apenas geram texto em suas respostas. Assim, com base nos estudos de [Roffo *et al.* 2014], utilizou-se os parâmetros: 1) Número de palavras; 2) Número de caracteres; 3) Ocorrência do uso de interrogações e exclamações; 4) Ocorrência do uso de reticências.

3. Resultados

Os dados provenientes das interações com o Chat GPT revelaram diferenças significativas entre o texto original e o texto reescrito por uma IA com alto traço de neuroticismo.

Em relação às características lexicais, o texto original contém 25 palavras e 107 caracteres (sem espaçamento), enquanto o texto reescrito apresenta 31 palavras e 142 caracteres (sem espaçamento). Essa diferença sugere que o texto reescrito é mais prolixo, uma característica comum em indivíduos com traços de neuroticismo, que tendem a se expressar de forma mais detalhada e menos concisa. Nas características sintáticas, o texto original utiliza duas marcas de interrogação e exclamação, enquanto o texto reescrito emprega apenas uma marca de interrogação. Essa redução no uso de marcas pode indicar menor ênfase e entusiasmo na comunicação, refletindo uma falta de desejo de engajamento com o interlocutor. Outro aspecto notável é o uso de reticências, pois o texto original não contém reticências, enquanto o texto reescrito apresenta quatro ocorrências. O uso frequente de reticências sugere hesitação e incerteza, comuns em indivíduos com alto nível de neuroticismo, que podem ter dificuldades em controlar suas emoções e expressar-se de forma direta. Portanto, a ênfase no uso de reticências demonstra uma diferença significativa no estilo de comunicação.

4. Conclusão

Diante do que foi apresentado, conclui-se que a indução de personalidade influenciou a comunicação. De maneira geral, a análise comparativa dos textos indicou que a IA com traços de neuroticismo manifestou características típicas dessa personalidade, evidenciadas por um estilo de comunicação mais prolixo e hesitante, com menor engajamento emocional. Essas diferenças demonstram como os traços de personalidade podem impactar a interação humano-máquina, comprovando a capacidade da IAG em modular e gerar respostas baseadas em características de personalidade. Além disso, a maior parte dos estudos sobre o tema se baseia em interações na língua inglesa. Portanto, o presente trabalho visa possibilitar que futuras pesquisas nacionais possam se valer das técnicas e achados aqui apresentados.

Por se tratar de um estudo preliminar, os resultados obtidos representam um passo inicial na busca pelo objetivo maior, que é demonstrar como a modulação de personalidade em IAG pode melhorar a interação humano-máquina, proporcionando respostas mais empáticas e personalizadas na língua portuguesa. Para estudos futuros, pretende-se obter novas respostas induzidas a partir de outros traços de personalidade.

5. Agradecimento

Os autores agradecem ao Centro de Estudos e Sistemas Avançados do Recife (CESAR School).

6. Referências Bibliográficas

- Amin, M. M., Cambria, E. and Schuller, B. W. (2023). Will affective computing emerge from foundation models and general artificial intelligence? a first evaluation of chatgpt. *IEEE Intelligent Systems*, 38(2):15-23.
- Erickson, T. (1997). Design Agent As If People Mattered. In: *Software Agents*, pages 79-96. AAAI Press.
- Franceschini, A., Gropp, B., Moreira, B. D., Matheus, C. E., Perez, C., Puoli, C. and Aratangy, V. L. (2012). *Teoria e Prática da Pesquisa Aplicada*. Elsevier Ltd.
- Grillo, W. N and Barros, H. O. (2022). Desenhando conversas com personalidade: aplicação do processo de design para desenvolvimento de framework de design conversacional. *DAT Journal*, 7(2): 246-259.
- Ji, Y., Wu, W., Zheng, H., Hu, Y., Chen, X. and He, L. (2023). Is chatgpt a good personality recognizer? a preliminary study. *arXiv preprint arXiv:2307.03952*.
- Jiang, G., Xu, M., Zhu, S. C., Han, W., Zhang, C. and Zhu, Y. (2024). Evaluating and inducing personality in pre-trained language models. *Advances in Neural Information Processing Systems*, 36
- Norman, D. A. (2008) *Emotional Design: why we love (or hate) everyday things*, Rocco.
- Oliveira, E. G. (2023). *Chatbots: a importância do processamento de língua natural para a experiência do usuário. Trabalho de Conclusão de Curso (Graduação em Linguística) - Universidade Federal de São Carlos, São Paulo.*
- Roffo, G., Giorgetta, C., Ferrario, R., & Cristani, M. (2014). Just the way you chat: Linking personality, style and recognizability in chats. In *Human Behavior Understanding: 5th International Workshop, HBU 2014, Zurich, Switzerland, September 12, 2014. Proceedings 5*, pages 30-41. Springer International Publishing.