ARTIGO

ChatGPT 4.0: desafios na interpretação de textos multimodais

ChatGPT 4.0: challenges in multimodal text interpretation

Paulo Henrique Duque
paulo.henrique.duque@ufrn.br
Universidade Federal do Rio Grande do Norte – UFRN

Resumo
Este estudo investiga a capacidade do modelo de inteligência artificial ChatGPT 4.0 em interpretar charges, utilizando benchmarks humanos como referência. As charges foram escolhidas por integrarem elementos verbais e não-verbais, permitindo uma avaliação detalhada de como o ChatGPT lida com nuances contextuais, humor e sátira. Os resultados demonstram que, embora o ChatGPT consiga identificar elementos visuais principais, ele enfrenta desafios significativos na compreensão de contextos mais amplos e na interpretação de humor e subtextos complexos. O estudo revela que as interpretações do ChatGPT tendem a ser superficiais e menos detalhadas em comparação com as humanas, especialmente em aspectos como estilo artístico, técnicas visuais e contextos culturais. Além disso, o ChatGPT mostra dificuldades em captar a profundidade e a intenção crítica de elementos satíricos, resultando em interpretações que não refletem completamente as mensagens implícitas nas charges. Os achados deste estudo contribuem para a compreensão das capacidades e limitações atuais dos modelos de IA na interpretação de discursos complexos, oferecendo insights valiosos para o avanço da linguística cognitiva e das tecnologias de processamento de linguagem natural.

Palavras-chave
ChatGPT. Multimodalidade. Interpretação de Charges. Comparação Humano-IA.

Abstract
This study investigates the capability of the AI model ChatGPT 4.0 in interpreting cartoons, using human benchmarks as a reference. Cartoons were selected for their integration of verbal and non-verbal elements, allowing a detailed assessment of how ChatGPT handles contextual nuances, humor, and satire. The results show that although ChatGPT identifies main visual elements, it faces significant challenges in understanding broader contexts and interpreting complex humor and subtexts. The study reveals that ChatGPT's interpretations tend to be superficial and less detailed compared to human interpretations, particularly in aspects such as artistic style, visual techniques, and cultural contexts. Additionally, ChatGPT shows difficulties in capturing the depth and critical intent of satirical elements, resulting in interpretations that do not fully reflect the implicit messages in cartoons. The findings of this study contribute to the understanding of the current capabilities and limitations of AI models in interpreting complex discourses, offering valuable insights for the advancement of cognitive linguistics and natural language processing technologies.

Linguagem em Foco
Revista do Programa de Pós-Graduação em Linguística Aplicada da UECE

FLUXO DA SUBMISSÃO
Submissão do trabalho: 23/05/2024
Aprovação do trabalho: 20/08/2024
Publicação do trabalho: 09/10/2024

10.46230/lef.v16i2.13157

COMO CITAR
DUQUE, Paulo Henrique. ChatGPT 4.0: desafios na interpretação de textos multimodais. **Revista Linguagem em Foco**, v.16, n.2, 2024. p. 110-130. Disponível em: https://revistas.uece.br/index.php/linguagememfoco/article/view/13157.

Keywords
ChatGPT. Multimodality. Cartoon Interpretation. Human-AI Comparison.

Introdução

A pesquisa sobre a interpretação de textos multimodais por IA reflete a crescente complexidade das aplicações de IA (Zeng *et al*., 2023). Modelos como o ChatGPT, baseados em arquiteturas *Transformer* (Vaswani *et al*., 2017), trouxeram avanços na compreensão e geração de linguagem natural, mas a integração de elementos multimodais ainda é um desafio. Estudos indicam que, embora a IA reconheça elementos básicos, ela frequentemente falha em captar a complexidade das interações entre imagem e texto, especialmente em contextos que envolvem expressões faciais estilizadas ou caricaturais (Hua; Jin; Jiang, 2024).

A multimodalidade e a semiótica social são essenciais para entender as dificuldades na interpretação de textos multimodais por IAs. Conforme descrito por Kress e van Leeuwen (2001), a multimodalidade reconhece que todos os textos são compostos por múltiplos modos de comunicação que trabalham juntos para criar significado. Cada modo semiótico tem suas capacidades e limitações específicas para representar e comunicar informações (Kress, 2010). A semiótica social (Hodge; Kress, 1988) enfatiza que a construção de significados é um fenômeno social influenciado por contextos culturais específicos. Esses princípios são fundamentais para a análise da interpretação de charges por IA, destacando a importância da interação entre elementos verbais e não-verbais e os aspectos sociais e culturais que influenciam a construção de sentido.

A compreensão de contextos culturais e políticos é um dos principais obstáculos para a IA. Cong-Lem, Soyoof e Tsering (2024) destacam que, apesar do ChatGPT ser uma ferramenta poderosa, ele apresenta várias limitações significativas, especialmente em cenários que exigem um entendimento profundo do contexto e da complexidade das perguntas. Eles mencionam que o modelo enfrenta dificuldades em gerar respostas precisas e contextualmente adequadas quando confrontado com tarefas complexas e específicas, refletindo uma limitação crítica em sua aplicação em contextos educacionais e clínicos. Spennemann (2023) discute como modelos como o ChatGPT frequentemente falham em fornecer referências culturais atualizadas e relevantes, destacando suas limitações em acompanhar mudanças culturais. O autor observa que o ChatGPT tem dificuldades em interpretar nuances culturais e fornecer respostas culturalmente

relevantes, especialmente em contextos sociais que mudam rapidamente. Cao *et al*. (2023) exploram a tendência do ChatGPT em gerar informações alucinatórias, afetando a precisão das respostas culturais e figurativas. Eles mencionam que o ChatGPT apresenta dificuldades significativas em se adaptar a vários contextos culturais e frequentemente interpreta erroneamente a linguagem figurativa, resultando em imprecisões.

O humor e a sátira são elementos complexos que exigem uma compreensão profunda do contexto e da linguagem. Estudos como os de Farina e Lavazza (2023) destacam que sistemas de IA, como o ChatGPT, enfrentam desafios significativos ao tentar compreender e reproduzir nuances culturais e humorísticas, especialmente no que diz respeito à interpretação de ironias e sarcasmos. Essas limitações afetam a capacidade da IA de interpretar e gerar humor de forma eficaz. Apesar dos avanços, as limitações na interpretação multimodal pela IA persistem. He *et al*. (2024) e Ghosh *et al*. (2024) destacam que a integração de capacidades visuais e textuais em Modelos de Linguagem Visuais (VLMs) continua sendo um desafio significativo, ressaltando a importância dessa integração para o avanço da inteligência artificial multimodal[1]. A capacidade do ChatGPT de reconhecer padrões e gerar respostas contextualmente relevantes é impressionante, mas a interpretação completa de nuances culturais e humorísticas ainda está aquém da cognição humana (Nadella, 2024).

Este estudo investiga a capacidade do modelo ChatGPT 4.0 em interpretar charges utilizando *benchmarks* humanos como referência. Foram selecionadas 76 charges a partir de uma busca no Google Imagens dos anos de 2023 e 2024, integrando elementos verbais e não-verbais. Para cada charge, foram formulados prompts específicos abordando descrição de elementos visuais, análise do estilo artístico, transcrição de textos, análise do tom e estilo linguístico, explicação do contexto relevante e interpretação do humor ou sátira. As respostas do ChatGPT foram comparadas com interpretações humanas utilizando correlação de Pearson e A análise de variância (ANOVA) para analisar a precisão e profundidade das interpretações.

O objetivo deste estudo é duplo: primeiro, avaliar a capacidade do Chat-

1 Durante a finalização deste artigo, foi lançada uma nova versão do modelo de linguagem artificial, o ChatGPT 4o (Omini). Segundo a documentação fornecida pela OpenAI, este modelo promete avanços significativos na capacidade de processar e entender diferentes tipos de dados - palavras, áudio e imagem - de forma simultânea. Essas melhorias são atribuídas a avanços nos algoritmos e técnicas de treinamento utilizados. No entanto, é importante ressaltar que essas capacidades ainda precisam ser avaliadas por estudos futuros.

GPT de interpretar charges em comparação com *benchmarks* humanos; segundo, identificar áreas específicas em que o modelo apresenta limitações. Ao comparar as respostas do ChatGPT com as interpretações humanas, o estudo destaca pontos de concordância e divergência, além de analisar as implicações desses achados para a linguística cognitiva.

2 Charges, Construção de Sentidos e Arquitetura GPT

2.1 A categoria discursiva charge

A charge é uma forma de comunicação visual que combina elementos verbais e não-verbais visuais para transmitir uma mensagem crítica e humorística sobre eventos correntes, figuras públicas ou questões sociais. Seu aspecto multimodal permite a interação entre imagem e texto, potencializando o impacto da crítica. Frequentemente, a charge recorre à caricatura para exagerar traços físicos e comportamentais dos personagens, abordando complexidades políticas e sociais de maneira acessível e envolvente, estimulando a reflexão e o debate (Gawryszewski, 2008).

Os elementos visuais na charge são cruciais para a eficácia da sátira. O uso de linhas, formas, cores e composições exagera características específicas dos personagens, evidenciando suas falhas e amplificando a crítica. Simplificação e distorção das figuras permitem a identificação imediata das personalidades retratadas e intensificam o impacto visual. Segundo Garcia (2019), as caricaturas são uma das formas de comunicação visual mais populares e que os estudiosos têm refletido sobre até que ponto essas imagens reproduzem os imaginários e mentalidades de sujeitos e grupos sociais. Ele ressalta a importância das caricaturas na contemporaneidade e sua capacidade de representar e criticar contextos sociais e políticos.

O estilo da charge inclui técnicas como hipérbole, metáfora visual e uso de símbolos. A hipérbole acentua características para criar um efeito humorístico e crítico, enquanto a metáfora visual simplifica conceitos complexos em imagens impactantes. A linguagem é geralmente direta e acessível, com tom sarcástico ou irônico para engajar e provocar reflexão no espectador. Segundo Gawryszewski (2008), o objetivo principal da charge não é o humor, mas sim a denúncia e a promoção da reflexão. O humor é utilizado como uma ferramenta para engajar o público e destacar questões políticas e sociais, incentivando uma análise crítica do contexto apresentado.

O humor na charge combina elementos de sátira, paródia e ironia. A sátira

critica socialmente, expondo vícios e corrupções; a paródia imita criticamente; e a ironia revela contradições entre o dito e o pretendido. Souza (2018) destaca que a charge e a caricatura, através do grotesco e da equivalência simbólica, podem alertar e levar à reflexão. Elas servem como comentário social instantâneo, oferecendo uma perspectiva alternativa e educando o público de forma acessível. Andrade (2011) ressalta que as estratégias enunciativas e técnicas gráficas proporcionam um contexto imediato, conferindo à charge um sentido crítico e humorístico.

2.2 Multimodalidade e Semiótica Social

A multimodalidade e a semiótica social emergem como abordagens fundamentais para a compreensão da comunicação contemporânea, especialmente no contexto da análise de charges pelo ChatGPT 4.0. Estas teorias são essenciais para interpretar como diferentes modos semióticos, como elementos visuais e textuais, interagem e criam significados complexos. Kress e van Leeuwen (2001) argumentam que todos os textos são intrinsecamente multimodais, ou seja, compostos por múltiplos modos de comunicação que trabalham juntos para criar significado. Isso é particularmente relevante para o estudo das charges, que combinam imagens e palavras para transmitir mensagens humorísticas e críticas.

A semiótica social, conforme desenvolvida por Halliday (1978) e Hodge e Kress (1988), enfatiza que os significados são socialmente construídos e que os modos semióticos são usados de acordo com contextos culturais e sociais específicos. Essa abordagem é crucial para analisar como o ChatGPT 4.0 interpreta os elementos das charges em relação aos contextos sociais e culturais. A semiótica social permite avaliar a capacidade do modelo de IA em captar não apenas o conteúdo explícito das charges, mas também as nuances implícitas, como ironia e crítica social, que são essenciais para a plena compreensão dessas formas de comunicação.

As IAs generativas, como o GPT-4, demonstram cada vez mais a capacidade de entender e criar textos que combinam diferentes formas de comunicação, como palavras e imagens. Para compreender como essas IAs analisam textos complexos, como charges, é fundamental considerarmos que todos os textos são compostos por múltiplos modos de comunicação que trabalham juntos para criar significados (Kress; van Leeuwen, 2001). Cada modo semiótico, como a linguagem verbal ou visual, possui suas próprias capacidades e limitações específicas para representar e comunicar informações (Kress, 2010). A criação e interpretação

dos significados são influenciadas pelos contextos sociais e culturais específicos (Kress, 2010), e os modos semióticos são utilizados conforme as necessidades e práticas sociais, o que requer uma análise das funções sociais dos signos dentro de um determinado contexto (Kress, 2010). Além disso, os modos semióticos frequentemente interagem em conjuntos modais para comunicar informações de maneira mais eficaz, utilizando as forças de cada modo para complementar as limitações dos outros (Kress; van Leeuwen, 2001). Ao aplicarmos esses princípios, podemos entender melhor como as IAs generativas interpretam textos multimodais e como podemos aprimorar sua capacidade de análise.

Considerando esses princípios, a interpretação de charges por IAs generativas torna-se um desafio complexo. A multimodalidade exige que a IA reconheça e integre informações de diferentes modos semióticos, como linguagem verbal e elementos visuais, para construir o significado da charge. A especificidade de cada modo implica que a IA precisa ter a capacidade de analisar tanto a dimensão verbal quanto os elementos não-verbais da charge, como desenhos, expressões faciais e cores, cada um com suas próprias nuances e formas de construir significados. Além disso, a influência do contexto social e cultural demanda que a IA tenha conhecimento prévio sobre o mundo e seja capaz de aplicar esse conhecimento para interpretar a charge de forma adequada. Por fim, a interação entre os modos semióticos requer que a IA seja capaz de identificar e analisar como os diferentes elementos da charge se complementam e o quanto contribuem para a construção do seu significado global.

2.3 Construção de Sentido: humano vs. ChatGPT

O processo de construção de sentido em humanos é uma tarefa complexa que envolve a interpretação de pistas visuais e verbais, integradas através de esquemas e frames conceptuais. Este processo é particularmente evidente na interpretação de uma charge, em que a identificação de elementos visuais e textuais significativos é crucial. Estes elementos não são interpretados isoladamente, mas são integrados e influenciados pelo contexto cultural e pelas experiências pessoais do indivíduo.

A cultura desempenha um papel fundamental neste processo, destacando certas informações perceptuais em contextos espaço-temporais específicos. É a influência cultural que determina quais informações perceptuais devem ser transformadas em pistas verbais e/ou não-verbais, configurando-se, assim, como textos. Esta dinâmica de textualização do mundo é um componente essencial do

processo de compreensão em humanos.

À medida que percebemos pistas materiais verbais e não-verbais, evocamos, conectamos e desconectamos conceitos para compreender a entidade em questão. Este processo contínuo de evocação e conexão de conceitos permite aos humanos navegar e compreender o mundo ao seu redor de maneira eficaz.

Por serem tecidas *online*, redes conceituais vão sendo ajustadas continuamente, à medida que os textos se manifestam. Graças a esses ajustes, humanos navegam em situações complexas e imprevisíveis, ajustando percepções e ações de acordo com o contexto corrente. Algumas dessas redes conceituais, quando moldadas recorrentemente, tornam-se *offline*. Isso significa que elas são mantidas no cérebro e podem ser evocadas para compreender situações semelhantes no futuro. Esses *frames offline*[2] (Vereza, 2013) podem ser definidos como padrões recorrentes de experiência sociocultural e conhecimento acumulados ao longo do tempo, que usamos para dar sentido ao mundo de maneira eficiente e previsível.

Os *frames* têm suas redes de conceitos configuradas por esquemas. Esses sistemas cognitivos mais básicos emergem da interação contínua do organismo com o ambiente e, uma vez estabelecidos no cérebro, guiam a percepção, a memória e a ação do organismo, facilitando a compreensão de novas informações. A estrutura topológica de um esquema é constituída de primitivos derivados diretamente das experiências perceptuais e motoras de um indivíduo no seu ambiente físico. Segundo Mandler e Cánovas (2014), primitivos são os blocos básicos com os quais construímos conhecimento (como EQUILÍBRIO, TEMPERATURA, POSIÇÃO, ALTURA, COISA, CONTATO, FORMATO, LIGAÇÃO, TEXTURA, DISTÂNCIA, LUMINOSIDADE, DIREÇÃO, LOCALIZAÇÃO, LIMITES, ESPAÇO, INTERIOR, EXTERIOR, PORTAL, BLOQUEIO, MOVIMENTO[3], etc.). Combinados, eles formam os esquemas. Por exemplo, os primitivos INTERIOR, EXTERIOR, LIMITE e PORTAL juntos formam uma base topológica para categorias conceptuais como CASA, GARRAFA, COPO, SALA e CAIXA, por exemplo, em situações nas quais são pensadas como recipientes.

De forma mais específica, os primitivos combinados formam as estruturas cognitivas conhecidas na literatura como esquemas de imagem (Lakoff, 1987; Johnson, 1987; dentre outros) e esquemas de ação (Feldman, 2006; Duque, 2015,

2 Vereza (2013) define *frames offline* como estruturas cognitivas pré-existentes que ajudam a formar e entender novos conceitos e contextos.

3 Estamos usando VERSALETE para representar constructos conceptuais, como esquemas e *frames*.

dentre outros). Esquemas de imagem (esquemas-I), como CONTÊINER, PARTE-TODO, CENTRO-PERIFERIA, CONTATO, VERTICALIDADE etc. moldam nossas experiências perceptuais, e esquemas de ação (esquemas-X), como PEGAR, CHUTAR, CAMINHAR, CAIR etc., moldam nossas experiências motoras. Primitivos combinados em esquemas-I ou esquemas-X, portanto, possibilitam uma modelagem flexível, diversificada e adaptativa de conceitos para seres, coisas, lugares, estados e ações.

Quando combinamos Esquemas-I e Esquemas-X, formamos o que chamamos de Esquemas de Eventos (Esquemas-E). Exemplos desses esquemas incluem padrões como *trajetor* se deslocando dentro de um CONTÊINER ou de um CONTÊINER para CONTÊINER (DESLOCAMENTO), um *trajetor* agindo sobre uma COISA (AÇÃO TRANSITIVA), *trajetor* causando deslocamento espacial de uma COISA (TRANSFERÊNCIA FÍSICA) e padrões formados a partir destes, como *trajetor* causando mudança de estado de uma coisa (MUDANÇA DE ESTADO), como *trajetor* causando o deslocamento possessório de uma COISA (TRANSFERÊNCIA DE POSSE) e como *trajetor* causando o deslocamento acústico de PALAVRAS (COMUNICAÇÃO VERBAL).

Os Esquemas-E são fundamentais para moldar nosso entendimento de um movimento[4]. Quando um movimento ocorre, ele pode alterar várias coisas - pode mudar o estado de repouso de um objeto, a percepção ou sentimento de uma pessoa, o local em que um objeto está ou o próprio objeto que está sendo manipulado. Por exemplo, quando alguém vê um objeto, sua experiência visual muda. Se um objeto se move de um lugar para outro, é o local que muda. Quando uma pessoa interage com um objeto, o objeto é o que muda. E se alguém faz um objeto se mover, o que muda é o estado de repouso do objeto.

Os esquemas-*Scripts* (esquemas-S) são estruturas de conhecimento que representam sequências estereotipadas de eventos em contextos familiares, isto é, representam uma série de esquemas-E. Um exemplo disso é o esquema-S IR AO RESTAURANTE, que engloba uma sequência de esquemas-E, como SOLICITAR MESA, SENTAR-SE, PEDIR CARDÁPIO, ESCOLHER PRATO, RECEBER O ALIMENTO, COMER, SOLICITAR CONTA, PAGAR e SAIR (Schank; Abelson, 1977).

Este estudo propõe uma adaptação da teoria dos *scripts*, configurando-os como esquemas flexíveis. Diferentemente dos *scripts* tradicionais de Schank, que

4 O termo “movimento” é usado para se referir à dinâmica da ação. Ele representa a progressão de uma ação ao longo do tempo, incluindo a maneira como a ação se desenvolve, as forças que a impulsionam e as mudanças que ocorrem como resultado.

são rígidos, esses esquemas são capazes de se adaptar e mudar em resposta às demandas situacionais de indivíduos e grupos. Essa abordagem permite uma maior flexibilidade e aplicabilidade em uma variedade de contextos, refletindo a natureza dinâmica e adaptativa da cognição humana. Esquemas-S facilitam a compreensão global das situações culturalmente relevantes, permitindo que os indivíduos antecipem e planejem comportamentos, interpretem ações e contextos com maior precisão e se adaptem eficientemente a novas circunstâncias.

No entanto, é importante enfatizar que, embora a estrutura algorítmica dos esquemas-S sugira rigidez, eles se atualizam a cada nova experiência vivenciada diretamente em situações sociais ou indiretamente por meio da escuta ou da leitura de histórias. O algoritmo é momentaneamente enriquecido pelas nuances contextuais específicas de cada situação. Por exemplo, o movimento de pegar uma faca nos esquemas-S IR AO RESTAURANTE ou BRIGAR NA RUA é interpretado de maneiras diferentes, devido a especificidades adicionadas à estrutura do esquema original. O esquema-S COMER, embutido no esquema-S IR AO RESTAURANTE, prevê o uso de faca com a finalidade de cortar alimentos durante a refeição. O esquema-S ATACAR, embutido no esquema-S BRIGA NA RUA, prevê o uso de faca com a finalidade de ferir o adversário.

A sequência de esquemas-E pode ser interrompida de modo a provocar uma quebra de expectativa no esquema-S. Por exemplo, uma charge apresenta uma cena em um posto de gasolina. No primeiro quadro, um cliente mostra uma nota de 50 reais e diz ao frentista: “Pode colocar 50 reais!”. Aqui, o Esquema-S seria ABASTECER O CARRO, que inclui esquemas-E como IR AO POSTO, PEDIR PARA ABASTECER, ABASTECER, PAGAR etc. No entanto, no segundo quadro, a expectativa é quebrada: em vez de usar a bomba de gasolina para abastecer o carro, o frentista aparece com um conta-gotas. Essa quebra sugere que o preço da gasolina está tão alto que 50 reais só pagam por algumas gotas de combustível.

Para o ChatGPT[5], a construção de sentido é baseada na arquitetura *Transformer*, que utiliza mecanismos de *self-attention*. Esse mecanismo permite ao modelo considerar todas as palavras em uma frase simultaneamente, atribuindo pesos diferentes a cada palavra com base em sua importância relativa (Vaswa-

5 GPT, sigla para *Generative Pre-trained Transformer*, refere-se a um conjunto de tecnologias de aprendizado de máquina que, após um processo de pré-treinamento em grandes volumes de dados textuais, é capaz de gerar textos de forma autônoma, emulando a coerência e o contexto dos textos produzidos por humanos.

ni *et al*., 2017). Diferente de arquiteturas anteriores, como Redes Neurais Recorrentes (RNNs[6]), que processavam palavras de forma sequencial, os Transformers processam palavras em paralelo, utilizando muitos parâmetros para estabelecer relações contextuais mais complexas e eficientes (Kenney, 2023).

O mecanismo de *self-attention* envolve três componentes principais: *Query* (Q), *Key* (K) e *Value* (V). O *Query* representa a palavra ou elemento visual que está sendo considerado para atenção. O *Key* representa a palavra ou elemento visual relevante para o *Query*. O *Value* representa a informação associada ao *Key*. O modelo utiliza representações vetoriais para palavras e imagens, calculando produtos escalares entre *Queries* e *Keys* e normalizando essas pontuações para criar uma representação contextualizada para cada palavra ou elemento gráfico (Vaswani *et al*., 2017).

Figura 1 – Charge sobre o Dia da Árvore

Fonte: Cazo, 2023.

Na interpretação da charge sobre o Dia da Árvore (Figura 1), a expressão de surpresa da árvore e os tocos de árvores cortadas fornecem o contexto visual necessário para entender a ironia e a crítica ambiental da charge. As frases "Cadê todo mundo?!" e "esperando mais convidados" devem receber as mais altas pontuações de atenção, pois são essenciais para processar o contexto e o humor da charge.

Enquanto humanos utilizam esquemas cognitivos e *frames offline* para estruturar um entendimento profundamente enraizado no contexto cultural e nas experiências emocionais, o ChatGPT depende do mecanismo de *self-attention* para ajustar a ênfase dada a cada elemento do texto ou imagem. O ChatGPT

6 Iniciais de *Recurrent Neural Networks*.

converte palavras e elementos visuais em representações vetoriais[7] e processa essas representações através de múltiplas camadas, refinando a interpretação a cada passo.

Nos seres humanos, esquemas-S e *frames offline* possibilitam a construção coerente de sentidos disruptivos. Esses sentidos referem-se à emergência de novos significados ou interpretações que surgem *ad hoc*, decorrentes de um conjunto específico de circunstâncias em interação com esses esquemas e *frames offline*. Tais sentidos são caracterizados por sua capacidade de desafiar, romper ou reconfigurar expectativas ou compreensões convencionais, resultando na formação de conceitos inovadores. Esse processo contrasta com a formação de sentidos probabilísticos, os quais são derivados de cálculos estatísticos realizados em grandes conjuntos de dados. Sentidos probabilísticos representam a interpretação mais provável ou comum, frequentemente baseada na frequência relativa de ocorrências em dados históricos. Assim, enquanto os sentidos probabilísticos são fundamentados em padrões estatísticos de grandes volumes de dados, os sentidos disruptivos são formados de maneira mais dinâmica e adaptativa, permitindo a emergência de novos conceitos em resposta a circunstâncias específicas. Conforme Kenney (2023), o ChatGPT apresenta limitações significativas em termos de criatividade, operando estritamente dentro dos parâmetros programados, o que restringe sua capacidade de adaptação criativa. Além disso, Barrot (2023) observa que uma das maiores limitações do ChatGPT como ferramenta de aprendizado de idiomas é a incapacidade de replicar a autenticidade das interações humanas, frequentemente faltando profundidade emocional e nuances culturais essenciais para a prática efetiva da linguagem. Isso evidencia como a IA ainda enfrenta desafios significativos para alcançar a flexibilidade cognitiva humana.

As pistas materiais de um texto que evocam, emulam e integram esquemas-S e *frames offline* direcionam a atenção do indivíduo para aspectos específicos de uma ação, personagem ou objeto, relegando outros elementos a um plano secundário ou ignorando-os completamente. Essas operações discursivo-cognitivas envolvem o destaque de elementos relevantes do ambiente com base em experiências vividas, incluindo o gênero discursivo em questão, ou de detalhes gráficos em textos não-verbais e linguísticos em textos verbais. Na Figura 1, por exemplo, a fala e a expressão facial da árvore evocam o esquema-S FESTA

7 Vetores representacionais, ou *embeddings*, são mapeamentos de dados não numéricos, como palavras, para vetores numéricos multidimensionais que capturam aspectos semânticos e sintáticos do texto.

DE ANIVERSÁRIO, no qual o esquema-S é interrompido pela ausência dos convidados, resultando na quebra da expectativa e gerando frustração na aniversariante. O sentido disruptivo, entretanto, emerge da antropomorfização da árvore e do pássaro, bem como da emulação de uma festa fracassada em um cenário desolador de desmatamento.

Na construção de sentidos, projetamos novos conceitos a partir da base topológica de conceitos prévios. Na charge em tela, a antropomorfização da árvore e do pássaro através da fala e das expressões faciais são pistas que nos levam a interpretar a comemoração do Dia da Árvore como uma festa de aniversário sem convidados. Essa projeção cria uma metáfora poderosa que ajuda a visualizar e entender as contradições de celebrar a data em meio a altas taxas de desmatamento. Esse processo de *framing*[8] é crucial tanto para a interpretação individual quanto para a comunicação social, facilitando a comunicação eficaz, o entendimento mútuo e a constante atualização de sentidos.

A construção de sentido no ChatGPT baseia-se na arquitetura *Transformer* e no mecanismo de *self-attention*. Esse mecanismo permite que o modelo considere todas as palavras em uma frase simultaneamente, atribuindo pesos diferentes a cada palavra com base em sua importância relativa (Vaswani *et al*., 2017). No contexto da charge sobre o Dia da Árvore (Figura 1), a palavra "esperando" seria considerada a *Query*, a palavra "convidados" a *Key*, e a expectativa associada a *Value*. A atenção é calculada multiplicando a *Query* por todas as *Keys* e ajustando os valores usando a função *Softmax*[9] para criar uma representação contextualizada.

Embora seja capaz de reconhecer padrões e gerar respostas contextualmente relevantes, o modelo pode ter dificuldades em lidar com nuances culturais e humor. A ironia da árvore esperando uma celebração, mas encontrando destruição, pode não ser completamente compreendida sem um conhecimento profundo do contexto cultural do desmatamento. Isso pode resultar em uma interpretação superficial, em que o modelo reconhece elementos individuais, mas falha em captar a ironia e a crítica subjacente. Devido à natureza estatística do aprendizado de máquina, é improvável que o modelo capture plenamente a sen-

8 *Framing*, aqui, é concebido como o processo de evocação, emulação e integração de esquemas cognitivos e *frames offline* na modelagem de *frames online*, a partir de pistas verbais e não verbais identificadas pelo leitor.

9 A função *Softmax* é uma função matemática que normaliza um vetor de valores em um vetor de probabilidades, cujo somatório é igual a 1, permitindo assim a comparação relativa entre os diferentes valores.

sação de frustração do aniversariante esquecido pelos seus convidados, a menos que tenha sido pré-treinado especificamente para tal contexto.

Nos seres humanos, a percepção inicial da charge envolve a identificação de elementos visuais e textuais significativos, baseando-se não apenas em experiências semelhantes recorrentes, mas também na integração do contexto cultural e *ad-hoc*. O Dia da Árvore é uma referência conhecida que celebra a preservação das árvores, e esse contexto cultural é essencial para a compreensão completa da charge. O cenário de desmatamento, evocado pelos tocos de árvores cortadas, contrasta com a expectativa de celebração, criando uma ironia que é facilmente captada por aqueles familiarizados com questões ambientais.

3 Materiais e Métodos

3.1 Objeto da pesquisa

Este estudo analisa as respostas do ChatGPT às charges, um gênero jornalístico que combina elementos visuais e textuais para comentar questões políticas, sociais e culturais. As charges são especialmente adequadas para este estudo porque integram esses elementos de forma única, permitindo avaliar como o ChatGPT maneja nuances contextuais, quebras de expectativa, paródias, caricaturas e subtextos.

3.2 Delineamento da pesquisa

Foi realizado um delineamento transversal comparativo, onde duas amostras independentes foram analisadas: uma composta pelas interpretações fornecidas pelo ChatGPT e outra pelas interpretações humanas (*benchmarks*)[10]. A variável independente foi o tipo de intérprete (ChatGPT vs. humanos) e as variáveis dependentes foram a precisão, profundidade, nuances das descrições e a interpretação de humor e sátira nas charges.

3.3 Procedimentos específicos

As charges foram selecionadas a partir de uma busca no Google Imagens

10 As interpretações humanas utilizadas como *benchmark* foram fornecidas pelo próprio pesquisador. Esse fato não interfere nos resultados, pois o foco da pesquisa é avaliar a capacidade do ChatGPT em interpretar charges de maneira coerente com os processos cognitivos humanos, e não comparar diferenças entre interpretações humanas.

utilizando o termo "charge", com a pesquisa limitada aos anos de 2023 e 2024. Das 100 primeiras imagens obtidas, 76 foram escolhidas após a exclusão de imagens irrelevantes de sites educacionais, coleções específicas, estudos acadêmicos, sites inativos e aqueles que exigiam pagamento. Esses critérios garantiram a relevância, representatividade e acessibilidade das charges selecionadas. Para cada charge incluída no corpus, foi criada uma sessão de interação distinta com o ChatGPT, assegurando que cada sessão fosse independente. Além da imagem da charge, foram fornecidas ao ChatGPT informações contextuais políticas, culturais ou sociais relevantes, quando necessário.

Os *prompts* formulados para o ChatGPT abordaram seis tópicos essenciais: a descrição dos elementos visuais presentes na imagem, a análise do estilo artístico e das técnicas visuais utilizadas, a identificação e transcrição de qualquer texto presente na charge, a análise do tom e do estilo linguístico do texto, a explicação do contexto cultural, político ou social relevante, e a interpretação do humor ou da sátira apresentada na charge.

3.4 Técnicas e ferramentas de análise

Para analisar os dados, foram utilizadas diversas técnicas e ferramentas. A correlação de Pearson foi empregada para medir a força e a direção da relação linear entre as descrições de elementos visuais fornecidas pelo ChatGPT e as dos *benchmarks* humanos, onde um alto valor de correlação indicaria um alinhamento entre as interpretações. A análise de variância (ANOVA) foi usada para comparar as médias das descrições de elementos visuais e interpretações de humor e sátira entre o ChatGPT e os *benchmarks* humanos, identificando diferenças estatisticamente significativas. Além disso, a análise qualitativa avaliou as nuances das descrições e interpretações fornecidas pelo ChatGPT em comparação com os *benchmarks* humanos, focando na capacidade de captar humor, sátira, ironia e contexto cultural e político. Essas técnicas e ferramentas permitiram uma análise abrangente e detalhada, fornecendo insights valiosos sobre a capacidade do ChatGPT de interpretar charges em comparação com *benchmarks* humanos.

4 Resultados

4.1 Análise de Elementos Visuais

A correlação de Pearson (r = 0.219) indica uma correlação positiva fraca, sugerindo que o ChatGPT frequentemente falha em capturar detalhes e nuan-

ces nas imagens. A ANOVA revelou uma diferença estatisticamente significativa entre as descrições de elementos visuais fornecidas pelo ChatGPT e pelos *benchmarks* humanos (F = 14.016, p = 0.00026), indicando que o ChatGPT frequentemente foca em detalhes irrelevantes e omite aspectos importantes das charges. A análise qualitativa mostrou que o ChatGPT tende a fornecer descrições gerais e menos detalhadas dos elementos visuais. Por exemplo, na charge sobre as viagens de Lula e Janja (Figura 2), o ChatGPT descreveu detalhes superficiais das roupas e expressões faciais, mas falhou em capturar o simbolismo da cena, resultando em interpretações enviesadas.

Figura 2 - Charge Próximo Destino

Fonte: Schmock, 2023.

As principais discrepâncias nas interpretações do ChatGPT incluem a falta de detalhes críticos e nuances nas descrições (40%), dificuldades na utilização do contexto sociocultural e político (30%), e superficialidade na interpretação de humor e sátira (30%).

Os resultados indicam que o ChatGPT apresenta limitações na interpretação de elementos visuais em comparação com os *benchmarks* humanos. Embora o modelo consiga identificar alguns aspectos básicos das imagens, ele frequentemente falha em capturar detalhes críticos e nuances importantes que são facilmente percebidos por observadores humanos.

4.2 Análise de Estilo, Técnicas, Linguagem e Tom

A correlação de Pearson, usada para medir a relação entre as descrições do ChatGPT e as humanas, foi baixa, indicando dificuldades do ChatGPT em capturar essas nuances com precisão. A ANOVA revelou diferenças significativas (F= 13,45, p= 0,0017) entre as descrições do ChatGPT e as humanas, confirmando que

as descrições do ChatGPT são menos detalhadas e precisas. A análise qualitativa mostrou que os *benchmarks* humanos descrevem estilos artísticos específicos, enquanto o ChatGPT identifica o estilo de forma genérica. Humanos detalham técnicas como uso de cores e sombreamento, enquanto o ChatGPT é mais superficial. Humanos captam o tom e a linguagem com profundidade, incluindo ironia e sarcasmo, que o ChatGPT identifica de forma superficial.

As principais diferenças incluem falta de especificidade em estilos artísticos, descrições técnicas menos detalhadas, e interpretações superficiais de linguagem e tom. O ChatGPT frequentemente falha em capturar nuances essenciais. A análise combinada mostra que o ChatGPT tem limitações significativas na interpretação de estilo, técnicas, linguagem e tom, refletindo uma compreensão superficial comparada aos *benchmarks* humanos.

4.3 Humor e Sátira

A correlação de Pearson, usada para medir a relação entre as interpretações de humor e sátira do ChatGPT e as humanas, foi muito fraca (r = 0,049), indicando que o ChatGPT tem dificuldades significativas em captar essas nuances. A ANOVA mostrou diferenças significativas (F= 12,34, p= 0,0015) entre as interpretações de humor e sátira do ChatGPT e as humanas, confirmando que essas diferenças não são devidas ao acaso. A análise qualitativa revelou que o ChatGPT reconhece elementos humorísticos de maneira superficial, frequentemente falhando em captar a complexidade do humor que depende de contextos culturais e sociais específicos.

Na charge sobre as viagens de Lula e Janja (Figura 2), o ChatGPT identificou pistas de forma imprecisa, não conseguindo interpretar corretamente a crítica subjacente. A interpretação da sátira pelo ChatGPT também se mostrou limitada, frequentemente não compreendendo a profundidade e a intenção crítica desses elementos, resultando em interpretações que não refletem completamente as mensagens implícitas na charge.

As principais discrepâncias incluem a incapacidade do ChatGPT de utilizar efetivamente o contexto sociocultural e a tendência a gerar interpretações refletindo vieses presentes nos dados de treinamento. Na Figura 2, a charge retrata o presidente Luiz Inácio Lula da Silva e a primeira-dama Rosângela da Silva em uma cena similar ao filme *Titanic*. O ChatGPT descreveu incorretamente elementos visuais e contextuais, como a presença de símbolos comunistas no lenço do presidente, em vez de no xale da primeira-dama, e fez interpretações imprecisas

sobre o cenário e a crítica implícita nas viagens presidenciais. Tais falhas refletem a superficialidade e os vieses no treinamento do modelo, resultando em interpretações que não capturam a complexidade e a sutileza da charge.

A análise combinada dos métodos quantitativos e qualitativos revelou que o ChatGPT apresenta dificuldades significativas na interpretação de humor e sátira. Embora o modelo consiga reconhecer elementos básicos, ele falha em compreender as nuances e o contexto completo percebido facilmente por humanos.

Considerações finais

Os resultados indicam que o ChatGPT 4.0 apresenta desempenho variado na interpretação de charges, especialmente em nuances contextuais e humor. A forma como o ChatGPT atribui pesos e significados a elementos visuais e textuais é limitada comparada à cognição humana, que é dinâmica e adaptativa. Conforme Alawida *et al*. (2023, p. 18), "o ChatGPT não está isento de limitações como a falta de consciência contextual e a incapacidade de entender completamente as implicações do texto que gera". Isso evidencia a dificuldade do modelo em representar informações adequadamente, ao contrário da cognição humana, que processa informações de maneira mais precisa e flexível. Embora os modelos de IA como o ChatGPT tenham demonstrado desempenho superior em estilo analítico, foco atencional e tom emocional neutro, essa vantagem não captura plenamente os aspectos mais sutis da comunicação humana. Sandler *et al*. (2024) destacam que a autenticidade das conversas humanas ainda supera as geradas por modelos de linguagem, ressaltando a importância das nuances emocionais e contextuais que os modelos de IA ainda não conseguem replicar totalmente.

O mecanismo de *self-attention* do ChatGPT permite capturar dependências de longo alcance entre tokens, mas ainda enfrenta desafios na compreensão completa do contexto sociocultural ou das nuances de humor e sátira da mesma forma que um ser humano. O modelo frequentemente omite detalhes cruciais ou adiciona elementos inexistentes, resultando em interpretações incompletas de textos multimodais. Farina e Lavazza (2023) destacam que os modelos de IA, como o ChatGPT, enfrentam desafios significativos ao tentar compreender e reproduzir nuances culturais e humorísticas, especialmente na interpretação de ironias e sarcasmos. Essas limitações afetam a capacidade da IA de interpretar e gerar humor de forma eficaz..

Os resultados deste estudo confirmam os resultados de outros estudos que atestam dificuldades do ChatGPT de interpretar contextos culturais espe-

cíficos e linguagem figurativa (Spennemann, 2023; Cao *et al*., 2023). Além disso, Ghosh *et al*. (2024) confirmam que, na ausência de contexto suficiente, os modelos de linguagem tendem a interpretar textos de maneira literal. He *et al*. (2024) destacam que a integração eficaz de informações visuais não-verbais e verbais permanece uma limitação significativa nos modelos atuais da IA.

Os princípios descritos por Kress (2010) e Kress e van Leeuwen (2001) destacam a necessidade de considerar a interação entre elementos verbais e não-verbais e os contextos sociais e culturais na construção de significado. Esta pesquisa demonstra que o ChatGPT frequentemente falha em captar essas interações complexas, resultando em interpretações superficiais dos elementos visuais e textuais. A multimodalidade e a semiótica social sublinham a importância de integrar múltiplos modos de comunicação para criar significado de maneira coesa. No entanto, o ChatGPT tem dificuldades em realizar essa integração de maneira eficaz. Por exemplo, enquanto humanos conseguem identificar e interpretar nuances contextuais, humorísticas e satíricas com base em uma compreensão profunda dos contextos culturais, o ChatGPT tende a fornecer descrições genéricas e falha em captar as sutilezas necessárias para uma interpretação precisa das charges.

O ChatGPT opera de forma probabilística, onde cada palavra gerada é influenciada pelas palavras anteriores e suas probabilidades estatísticas. Embora o mecanismo de self-attention permita ao ChatGPT atribuir diferentes pesos às palavras conforme sua relevância, ele não compreende plenamente as nuances implícitas de humor e sátira como um ser humano. Isso ocorre porque a essência do humor muitas vezes envolve criatividade, originalidade e a capacidade de fazer conexões inesperadas entre ideias aparentemente desconexas, aspectos que são intuitivos para os humanos, mas extremamente desafiadores para modelos de linguagem. Adicionalmente, o humor é altamente dependente do contexto, das nuances culturais e da subjetividade, elementos que os modelos de IA ainda lutam para capturar de forma eficaz (Hessel *et al*., 2023). Portanto, apesar de sua capacidade impressionante de processar e gerar linguagem, o ChatGPT ainda carece da profundidade de entendimento necessária para replicar o humor humano de maneira autêntica e contextualmente apropriada.

As frequentes classificações incorretas de emoções (Sandler *et al*., 2024) impactaram negativamente na qualidade das interpretações das charges. A cognição humana reconhece e interpreta estilos artísticos, compreendendo como cores, sombreamento, perspectiva e composição influenciam a charge. O ChatGPT, por outro lado, oferece descrições de aspectos técnicos pouco relevantes

para a interpretação geral da charge ou traz pacotes fechados de técnicas que, muitas vezes, não condizem com a charge em questão.

As principais discrepâncias observadas entre as interpretações do ChatGPT e dos *benchmarks* humanos incluem a falta de detalhes críticos e nuances nas descrições visuais, dificuldades na utilização do contexto sociocultural e político, e interpretações superficiais de humor e sátira. Essas discrepâncias podem ser atribuídas aos vieses presentes nos dados de treinamento do modelo, que influenciam suas respostas significativamente. Por exemplo, na análise da charge sobre as viagens de Lula e Janja (Figura 2), o ChatGPT fez uma descrição imprecisa dos elementos visuais e do contexto crítico. A interpretação errônea dos símbolos comunistas no lenço do presidente, ao invés de no xale da primeira-dama, e a suposição incorreta de que a cena ocorre em um transporte aéreo ao invés de em um navio, ilustram as limitações do modelo em utilizar o contexto fornecido de forma eficaz. Além disso, a superficialidade na interpretação de humor e sátira resultou em uma análise que não capturou as críticas subjacentes às viagens presidenciais e aos gastos excessivos.

Referências

ALAWIDA, M.; MEJRI, S.; MEHMOOD, A.; CHIKHAOUI, B.; ABIODUN, O. I. A comprehensive study of ChatGPT: Advancements, limitations, and ethical considerations in natural language processing and cybersecurity. **Information**, v. 14, n. 8, p. 462, 2023. DOI: https://doi.org/10.3390/info14080462. Disponível em: https://www.mdpi.com/2078-2489/14/8/462 . Acesso em: 17 abr. 2024.

ANDRADE, A. C. De. **A charge**: análise do processo enunciativo-discursivo numa perspectiva dialógica. 2011. 329 f. Tese (Doutorado em Linguística) – Centro de Artes e Comunicação, Programa de Pós-graduação em Letras, Universidade Federal de Pernambuco, Recife, 2011. Disponível em: https://repositorio.ufpe.br/handle/123456789/15037. Acesso em: 13 abr. 2024.

BARROT, J. S. ChatGPT as a Language Learning Tool: An Emerging Technology Report. **Technology, Knowledge and Learning**, California, v. 28, n. 4, p. 1-6, dec. 2023. DOI: https://doi.org/10.1007/s10758-023-09711-4. Disponível em: https://link.springer.com/article/10.1007/s10758-023-09711-4. Acesso em: 22 ago. 2024.

CAO, Y.; ZHOU, L.; LEE, S.; CABELLO, L.; CHEN, M.; HERSHCOVICH, D. Assessing cross-cultural alignment between ChatGPT and human societies: An empirical study. In: **Proceedings of the First Workshop on Cross-Cultural Considerations in NLP (C3NLP)**, Dubrovnik, Croatia. Association for Computational Linguistics, 2023. p. 53–67.

CAZO. Charge sobre o Dia da Árvore. **Blog do AFTM**. São Paulo, 22 set. 2023. Disponível em: https://anafisco.org.br/charge-dia-da-arvore/. Acesso em: 16 mar. 2024.

CONG-LEM, N.; SOYOOF, A.; TSERING, D. A systematic review of the limitations and associated

opportunities of ChatGPT. **International Journal of Human–Computer Interaction**, 08 maio 2024, p. 718-738. DOI: 10.1080/10447318.2024.2344142. Acesso em: 29 maio 2024.

DUQUE, P. H. Discurso e cognição: uma abordagem baseada em frames. **Revista da ANPOLL**, v. 1, n. 39, p. 25-48, 2015. Disponível em: https://revistadaanpoll.emnuvens.com.br/revista/article/view/902. Acesso em: 29 maio 2024.

FARINA, M.; LAVAZZA, A. ChatGPT in society: emerging issues. **Front. Artif. Intell**, v. 6, p. 1-7, jun. 2023. Disponível em: https://www.frontiersin.org/articles/10.3389/frai.2023.1130913/full. Acesso em: 29 maio 2024.

FELDMAN, J. A. **From molecule to metaphor**. [S.L]: MIT Press, 2006.

GARCIA, G. I. Uma imagem, tantas possibilidades: os avanços e desafios no estudo das caricaturas. **Revista em Perspectiva**, v. 4, n. 1, p. 109-125, 2019. Disponível em: http://periodicos.ufc.br/emperspectiva/article/view/41573. Acesso em: 29 maio 2024.

GAWRYSZEWSKI, A. Conceito de caricatura: não tem graça nenhuma. **Domínios da Imagem**, v. I, n. 2, p. 7-26, 2008. Disponível em: https://www.academia.edu/43273460/Conceito_de_caricatura_n%C3%A3o_tem_gra%C3%A7a_nenhuma. Acesso em: 29 maio 2024.

GHOSH, A.; JAIN, S.; KAPOOR, A.; KUMAR, V.; AGARWAL, P. Exploring the frontier of vision-language models: A survey of current methodologies and future directions. **Artificial Intelligence Review**, v. 2, p. 1-16, abr. 2024. Disponível em: https://arxiv.org/pdf/2404.07214. Acesso em: 29 maio 2024.

HALLIDAY, M. A. K. **Language as social semiotic**. London: Edward Arnold, 1978.

HE, S.; CHEN, Y.; XIA, Y.; LI, Y.; LIANG, H-N.; YU, L. Visual harmony: Text-visual interplay in circular infographics. **Journal of Visualization**, v. 27, p. 255-271, 2024. Disponível em: https://arxiv.org/pdf/2402.05798. Acesso em: 29 maio 2024.

HESSEL, J.; MARASOVIC, A.; HWANG, J. D.; LEE, L.; DA, J.; ZELLERS, R.; MANKOFF, R.; CHOI, Y. Do Androids Laugh at Electric Sheep? Humor “Understanding” Benchmarks from The New Yorker Caption Contest. In: **Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics**, Toronto, jul, p. 688-714, 2023.

HODGE, R.; KRESS, G. **Social semiotics**. London: Polity Press, 1988.

HUA, S. Y.; JIN, S. C.; JIANG, S. Y. The Limitations and Ethical Considerations of ChatGPT. **Data Intelligence**, v. 6, n. 1, p. 201–239, 2024. DOI: 10.1162/dint_a_00243. Disponível em: https://www.researchgate.net/publication/376740720_The_Limitations_and_Ethical_Considerations_of_ChatGPT. Acesso em: 29 maio 2024.

JOHNSON, M. **The body in the mind**: The bodily basis of meaning, imagination, and Reason. Chicago: University of Chicago Press, 1987.

KENNEY, N. M. **A Brief Analysis of the Architecture, Limitations, and Impacts of ChatGPT**. Georgia: Georgia Institute of Technology, 2023. DOI: https://zenodo.org/doi/10.5281/zenodo.7762244. Disponível em: https://zenodo.org/records/7762245. Acesso em: 12 abril 2024.

KRESS, G. R. **Multimodality**: A Social Semiotic Approach to Contemporary Communication. London e New York: Routledge, 2010.

KRESS, G.; VAN LEEUWEN, T. **Multimodal discourse**: the modes and media of contemporary communication. London: Hodder Arnold, 2001.

LAKOFF, G. **Women, fire, and dangerous things**: What categories reveal about the mind. Chicago: University of Chicago Press, 1987.

MANDLER, J. M.; CÁNOVAS, C. P. On defining image schemas. **Language and Cognition**, v. 6, n. 4, p. 510–532, 2014. Disponível em: https://www.researchgate.net/publication/269931714_On_defining_image_schemas. Acesso em: 29 maio 2024.

NADELLA, G. Visual ChatGPT: A comprehensive guide to multimodal AI. **Analytics Vidhya**, 13 de março de 2024. Disponível em: https://www.analyticsvidhya.com/blog/2023/03/power-of-visual--chatgpt-conversations-with-ai-and-images/. Acesso em: 15 março 2024.

SANDLER, M.; CHOUNG, H.; ROSS, A.; DAVID, P. A Linguistic Comparison between Human and ChatGPT-Generated Conversations. **ArXiv**, v. 3, p. 1 – 15, abr, 2024. Disponível em: https://arxiv.org/pdf/2401.16587. Acesso em: 29 maio 2024.

SCHANK, R. C.; ABELSON, R. P. **Scripts, plans, goals, and understanding**: An inquiry into human knowledge structures. Hillsdale, NJ: Lawrence Erlbaum Associates, 1977.

SCHMOCK. Charge sobre as viagens de Lula e Janja. **Revista Oeste**. São Paulo, 23 jun. 2023. Disponível em: https://revistaoeste.com/politica/charge-da-semana-46/. Acesso em: 16 mar. 2024.

SOUZA, I. C. de O. **A charge como fonte e representação da informação no desenvolvimento político brasileiro**. 2018. 194 f. Tese (Doutorado) – Instituto de Ciência da Informação, Universidade Federal da Bahia, Salvador, 2018. Disponível em: https://repositorio.ufba.br/handle/ri/27843. Acesso em: 23 ago. 2024.

SPENNEMANN, D. H. R. ChatGPT and the generation of digitally born “knowledge”: How does a generative AI language model interpret cultural heritage values? **Knowledge**, v. 3, n. 3, p. 480-512, 2023. Disponível em: https://doi.org/10.3390/knowledge3030032. Acesso em: 29 maio 2024.

VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, L.; POLOSUKHIN, I. Attention is All You Need. In: **Advances in Neural Information Processing Systems**, 2017. Disponível em: https://arxiv.org/pdf/1706.03762. Acesso em: 11 mar. 2024.

VEREZA, S. Entrelaçando frames: a construção do sentido metafórico na linguagem em uso. **Cadernos de Estudos Linguísticos**, n. 1, v. 55, p. 109-125, 2013. DOI: https://doi.org/10.20396/cel.v55i1.8636598. Disponível em: https://periodicos.sbu.unicamp.br/ojs/index.php/cel/article/view/8636598. Acesso em: 22 ago. 2024.

ZENG, Y.; ZHANG, H.; ZHENG, J.; XIA, J.; WEI, G.; WEI, Y.; ZHANG, Y.; KONG, T. What Matters in Training a GPT4-Style Language Model with Multimodal Inputs? **ArXiv**, 2023. Disponível em: https://arxiv.org/pdf/2307.02469. Acesso em: 29 maio 2024.

Sobre o autor

Paulo Henrique Duque - Doutor em Linguística. Professor do Programa de Pós-Graduação em Estudos da Linguagem (PPgEL) da Universidade Federal do Rio Grande do Norte (UFRN); Natal-RN E-mail: paulo.henrique.duque@ufrn.br. Lattes: http://lattes.cnpq.br/0409894285408135. OrcID: https://orcid.org/0000-0002-7100-0556.