Revista Intersaberes ISSN: 1809-7286
Vol. 20 e25do2en1 2025

GPT: A Máquina CONEXIONISTA de Wittgenstein

GPT: WITTGENSTEIN'S CONNECTIONIST MACHINE

GPT: LA MÁQUINA CONEXIONISTA DE WITTGENSTEIN

Luciano Frontino de Medeiros
Doutor em Engenharia e Gestão do Conhecimento (UFSC); professor do Programa de Pós-Graduação em Educação e Novas Tecnologias (Uninter).
Orcid: https://orcid.org/0000-0002-5947-9322 ; E-mail: luciano.me@uninter.com

RESUMO

A emergência das tecnologias generativas evidenciou o poder do paradigma conexionista em proporcionar inteligências artificiais com capacidade de lidar de forma muito proficiente com a linguagem humana, transparecendo consequências de cunho filosófico pertinentes. Este ensaio propõe uma discussão envolvendo a forma como o conexionismo, representado por modelos GPT (*General Pre-Trained Transformer*), suplantou o paradigma simbólico de pesquisa da Inteligência Artificial (IA), à luz de autores como Gottlob Frege, Daniel Dennett e John Searle, mas principalmente a partir da filosofia dual de Ludwig Wittgenstein, comentada por William Frawley, sobre a linguagem, tratando o significado tanto como uma forma computada baseada na lógica quanto em uma forma de ação baseada no uso. Em complemento a essa ideia central, questiona-se também se as tecnologias generativas, em certa medida, se apresentariam como solução para o enigma proposto por David Hume, quanto à possibilidade de ideias e impressões "pensando sobre elas mesmas", em alinhamento às reflexões de Dennett.

Palavras-chaves: inteligência artificial; inteligência artificial generativa; redes neurais artificiais; tecnologia educacional; filosofia da mente.

ABSTRACT

The emergence of generative technologies has highlighted the power of the connectionist paradigm to deliver AI systems highly proficient at handling human language, thereby revealing pertinent philosophical ramifications. This essay proposes a discussion of how connectionism - represented by GPT (Generative Pretrained Transformer) models - has supplanted the symbolic research paradigm in Artificial Intelligence (AI), in light of authors such as Gottlob Frege, Daniel Dennett, and John Searle, but above all drawing on Ludwig Wittgenstein's dual philosophy of language, as discussed by William Frawley, which treats meaning both as a logic-based, computed form and as a use-based form of action. Complementing this central idea, it also asks whether generative technologies might, to some extent, offer a solution to David Hume's puzzle concerning the possibility of ideas and impressions "thinking about themselves", in alignment with Dennett's reflections.

Keywords: artificial intelligence; generative artificial intelligence; artificial neural networks; educational technologies; philosophy of mind.

RESUMEN

El surgimiento de las tecnologías generativas ha puesto de relieve el poder del paradigma conexionista para ofrecer sistemas de IA altamente competentes en el manejo del lenguaje humano, revelando así implicaciones filosóficas pertinentes. Este ensayo propone una discusión sobre cómo el conexionismo —representado por los modelos GPT (Transformador Generativo Preentrenado)— ha superado el paradigma simbólico de investigación en Inteligencia Artificial (IA), a la luz de autores como Gottlob Frege, Daniel Dennett y John Searle, pero sobre todo recurriendo a la filosofía dual del lenguaje de Ludwig Wittgenstein, tal como la comenta William Frawley, que concibe el significado tanto como una forma calculada basada en la lógica como una forma de acción basada en el uso. Complementando esta idea central, también se pregunta si las tecnologías generativas podrían, en cierta medida, ofrecer una solución al enigma planteado por David Hume sobre la posibilidad de que las ideas e impresiones «piensen sobre sí mismas», en

consonancia con las reflexiones de Dennett.

Palabras clave: inteligencia artificial; inteligencia artificial generativa; redes neuronales artificiales; tecnologías educativas; filosofia de la mente.

INTRODUÇÃO

O momento histórico proporcionado pelo advento do ChatGPT e similares se caracteriza por ser paradigmático, pela existência de fato de uma Inteligência Artificial (IA) que dá um passo significativo em direção à sonhada Inteligência Artificial Geral, imaginada pelos cientistas e engenheiros da computação como aquela que se equiparará à inteligência humana (Bubeck *et al.*, 2023). O fato desta tecnologia inteligente estar aberta ao uso do público em geral, e não apenas à mercê dos círculos mais técnicos, proporciona tal nível de interação e interatividade que modifica de maneira profunda vários aspectos da vida e do trabalho em sociedade.

Considerando-se a presença dessa IA junto aos seres humanos tal como uma entidade individual de intencionalidade parcial, ubíqua, onipresente e multilíngue, fundamentando-se ainda na adoção de um posicionamento intencional, conforme Daniel Dennett (2006), em face a uma "mente" com características notadamente peculiares, pode-se até comparar a inserção da IA de forma análoga ao contato da civilização humana com alguma espécie animal que, de repente, passou a manifestar comportamento linguístico de alto nível, ou mesmo com algum tipo de inteligência que não seja originária do planeta.

Partindo dessas considerações, o estudo oportuno das interações dos seres humanos com as tecnologias generativas tem a possibilidade, nesse momento histórico, de permitir o confronto com as reflexões filosóficas que estiveram no bojo do desenvolvimento da IA como ciência para se chegar à sua concretização. É de se esperar, de tais abordagens empíricas, concordâncias acompanhadas de discordâncias. Entretanto, algo de relativa importância se constituiria na emergência de situações inusitadas, provenientes das interações entre seres humanos e IA, não antecipadas pelas reflexões dos filósofos ou pelas hipóteses dos cientistas.

Assim, este artigo busca oferecer, na forma de um breve ensaio, uma visão sobre como o advento do modelo GPT (*General Pre-Trained Transformer*) estende seu impacto a diversas questões no campo da filosofia da mente, com ênfase no sucesso de uma

arquitetura conexionista no processamento de linguagem natural. São discutidos os paradigmas conexionista e simbólico na IA e como os modelos GPT se posicionam na relação entre sintaxe e semântica, conforme a metáfora do quarto chinês de Searle. Também são abordadas, à luz das proposições de Gottlob Frege, as respostas aos *prompts* como expressões de um pensamento com caráter objetivo. Além disso, reflete-se sobre como o modelo GPT confirma as ideias de Dennett a respeito da IA como solução para o problema das ideias e impressões de David Hume.

Por fim, conjectura-se se o modelo GPT poderia ser considerado como uma espécie de "máquina de Wittgenstein", baseada em uma arquitetura conexionista, possibilitando o fornecimento em linguagem natural do significado das sentenças tanto de forma computada, fundamentada na lógica, quanto como uma forma de ação, baseada por sua vez no uso. Adicionalmente, em função do comportamento linguístico, discorre-se ainda sobre a categorização do modelo GPT como um "behaviorista filosófico".

GPT: TRANSFORMADOR PRÉ-TREINADO GENERATIVO

Surgindo no contexto do avanço das redes neurais artificiais de aprendizado profundo (Goodfellow; Bengio; Courville, 2016), no âmbito do processamento de texto e linguagem, os modelos GPT possuem em seu núcleo um tipo particular de rede neural artificial, denominada de "transformador" (do inglês, *transformer*). Um transformador compõe-se basicamente de um par codificador-decodificador, tendo um algoritmo cujo funcionamento é baseado em mecanismos de atenção, que estão no centro dos sistemas de tradução automática (Vaswani *et al.*, 2017). A arquitetura dos modelos GPT caracterizam-se por serem *decoder-only*, ou seja, trabalham apenas com a pilha decodificadora (Brown, T. B. *et al.*, 2020). Os modelos desenvolvidos com base nesses transformadores são capazes de produzir textos coerentes, devidamente contextualizados e em variados estilos de escrita, em resposta a um *prompt*, ou seja, uma instrução ou pergunta fornecida pelo usuário à ferramenta, que fornece o contexto inicial.

Um exemplo de ferramenta baseada em GPT é a versão produzida pela OpenAI, o GPT-4o ("o" de "omni"), com o objetivo de tornar a interação entre humanos e computadores ainda mais natural. Esse modelo aceita como *prompt*, além de textos,

formas combinadas com áudio, imagem e vídeo, e pode ainda gerar respostas em texto, áudio e imagem. Ele pode processar entradas de áudio em 232 milissegundos, com uma média de 320 milissegundos, um tempo comparável à resposta humana durante uma conversa. Em termos de desempenho, o GPT-4o se equipara ao GPT-4 Turbo para textos em inglês e código, mas apresenta melhorias significativas na compreensão de textos em outros idiomas (OpenAI, 2024).

De acordo com a OpenAI, dentre as capacidades projetadas para o GPT-4o, diversas teriam grande potencial para lidar com textos em linguagem natural, tais como: i) criação de narrativas visuais que combinam histórias com imagens e vídeos; ii) desenvolvimento de pôsteres; iii) composição de poesias, explorando diferentes estilos tipográficos; iv) criação de poesias concretas; v) renderização de textos com múltiplas linhas; vi) transcrição e anotação de reuniões com diversos participantes; e vii) sumarização de apresentações e vídeos (OpenAI, 2024). No entanto, como limitações atribuídas aos modelos GPT em geral, ainda estão sujeitos à ocorrência de alucinações ou erros de raciocínio, sendo recomendada a revisão humana quando utilizado em contextos críticos (OpenAI, 2023).

Historicamente, o tratamento da gramática na linguagem natural teve abordagem no campo simbólico por meio de analisadores descendentes recursivos sobre gramáticas livres de contexto, um dos tipos de gramáticas da hierarquia de Chomsky (Hopcroft; Motwani; Ullman, 2006, p. 194). Tais gramáticas seguem uma abordagem simbólica e determinística, tentando decompor uma frase em sua estrutura gramatical de maneira hierárquica (Russell; Norvig, 2004, p. 767). Esses métodos, ainda que eficazes em alguns contextos, enfrentavam limitações quando confrontados com as ambiguidades e complexidades inerentes às línguas humanas. Dentro do contexto histórico, é possível encontrar ainda uma ampla abordagem sobre o processamento estatístico da linguagem (Manning; Raghavan; Schütze, 2009; Manning;Schutze, 1999).

Com o surgimento de arquiteturas conexionistas, como as redes neurais artificiais, a gramática passou a ser tratada de forma mais robusta e flexível. Essas redes, especialmente os transformadores, são capazes de aprender padrões complexos e sutis a partir de grandes quantidades de dados, superando os modelos tradicionais ao lidar com variações e ambiguidades da linguagem de forma mais eficiente e natural.

GPT, TRANSFORMADORES E LLM

É importante destacar as semelhanças e diferenças entre os vários conceitos envolvidos com as tecnologias generativas. A denominação "GPT" refere-se a um tipo específico de transformador: nem todo transformador é um GPT. Outro tipo de transformador conhecido é o BERT (*Bidirectional Encoder Representations from Transformers*). Diferente do GPT, o BERT usa apenas a parte codificadora da arquitetura do transformador. Ele é projetado para compreender melhor o contexto de palavras em uma frase e é utilizado em tarefas como classificação de texto, busca por respostas e entendimento de linguagem (Devlin *et al.*, 2019; Zhao *et al.*, 2023).

Outro conceito frequentemente utilizado no contexto das tecnologias generativas é o LLM (*Large Language Model*), ou modelos de linguagem grande. Um GPT é um LLM, pois foi treinado com um massivo volume de dados textuais, chegando a possuir na casa de bilhões de parâmetros configuráveis, capacitado para gerar texto de forma coerente, predizer palavras e realizar demais tipos de tarefas envolvendo linguagem natural (Brown, *et al.*, 2020).

No entanto, nem todo LLM é um GPT. Por exemplo, o próprio BERT apresentado anteriormente também é um LLM, assim como o T5 (*Text-to-Text Transfer Transformer*). O T5 trata toda tarefa de processamento de linguagem natural como um problema de conversão de texto para texto, utilizando tanto a parte de codificador quanto de decodificador do transformador, em tarefas como tradução, sumarização e perguntas e respostas (Raffel *et al.*, 2019).

Neste artigo utiliza-se a denominação "modelo GPT", ou "GPT", de forma sucinta, designando de modo geral as aplicações que utilizam esse modelo de transformador (ou ainda arquiteturas similares ou variações) para conversação em linguagem natural: além do próprio ChatGPT, há o Gemini (Georgiev *et al.*, 2024), o Claude (Anthropic, 2024) e o LLaMA (Touvron *et al.*, 2023).

O MECANISMO DE ATENÇÃO

Apesar de haver uma compreensão relativamente clara sobre o funcionamento dos transformadores e de estudos aprofundados serem conduzidos para seu aprimoramento,

tais como os de "interpretabilidade mecanicista" (Elhage *et al.*, 2021) e "cabeçalhos de indução (de atenção)" (Olsson *et al.*, 2022), os princípios subjacentes aos LLMs ainda carecem de exploração aprofundada. A emergência das habilidades observada nesses modelos permanece, de certo modo, envolta em mistério, havendo espaço para investigações mais profundas no entendimento de como os LLMs desenvolvem tais capacidades de inferência textual complexa. Além disso, a comunidade de pesquisa enfrenta dificuldades no treinamento de LLMs devido à alta demanda por recursos computacionais (Zhao *et al.*, 2023).

Com base em Vaswani *et al.* (2017), Devlin *et al.* (2019) e Alammar (2022), visando o alcance do entendimento para uma audiência não técnica, o funcionamento de uma rede neural do tipo transformador em um modelo GPT para a geração de textos segue, em suma, cinco etapas: i) representação das palavras; ii) codificação de posição de termos; iii) mecanismo de atenção; iv) geração do texto; e v) ajuste fino. Na etapa de representação das palavras, conhecida em inglês como *embeddings*, é preciso separar cada palavra e converter cada palavra da frase em uma forma numérica. Essa conversão acontece internamente por meio de um vetor que representa cada palavra individualmente. Depois disso, na codificação da posição, é adicionada a informação de posição de cada palavra, permitindo que o modelo compreenda a ordem correta em que as palavras aparecem na frase.

A partir desse ponto, o mecanismo de atenção entra em ação para ajudar a determinar quais palavras da frase inicial são mais relevantes para influenciar a continuidade do texto sendo gerado. Após a atenção, dá-se um processamento por camadas densas de neurônios para ajustar as sentenças a padrões mais complexos. Finalmente, por meio de uma função que permite uma tomada de decisão, o modelo vai atribuindo pesos às palavras, gerando probabilidades que refletem o quanto cada uma delas influenciará o próximo trecho do texto.

As representações com os pesos/parâmetros calculados servirão como base para a continuação do texto após a frase inicial. Com a atribuição correta das probabilidades, o modelo é capaz de gerar a próxima parte do texto de forma coerente, mantendo a conformidade com os padrões dos textos que foram usados em seu treinamento prévio. Por fim, o ajuste fino se relaciona ao processo de ajuste e melhoria da qualidade da

resposta gerada pelo modelo, garantindo que o texto seja mais fluido, coerente e contextualmente apropriado.

O PARADIGMA CONEXIONISTA VS. O SIMBÓLICO

Na história da Inteligência Artificial, como um campo legítimo de pesquisa científica, é curioso pensar que houve um momento no qual se argumentou fortemente sobre a insuficiência das redes neurais artificiais como uma técnica para se encontrar soluções envolvendo problemas de pensamento e raciocínio. Considerando-se que as redes neurais biológicas estariam na origem das atividades cerebrais que dão sustentação, por sua vez, às atividades mentais, seria no mínimo inusitado ater-se a uma afirmação dessa monta. No entanto, a partir de Marvin Minsky e Seymour Papert, com o livro "Perceptrons" em 1969, por praticamente uma década a pesquisa em redes neurais artificiais caiu no ostracismo, e nos EUA praticamente se encerraram os aportes financeiros para pesquisas na área da principal agência de fomento à pesquisa americana, a National Science Foundation (Haykin, 2001).

Minsky e Papert haviam ponderado que um tipo de rede neural bastante promissor na época, o *perceptron*, inventado pelo psicólogo Frank Rosenblatt em 1958, não seria capaz de perfazer uma operação da lógica *booleana* importante, o "ou-exclusivo". Rosenblatt havia sugerido que os circuitos contendo portas lógicas, existentes nos *chips* digitais, poderiam ser substituídos por *perceptrons*. No entanto, o argumento de Minsky e Papert se baseou no que mais tarde foi denominado de *perceptron* de camada simples. Na retomada das pesquisas em redes neurais na década de 1980, já se contava com técnicas e também com maior capacidade de processamento nos computadores, que possibilitaram aos pesquisadores, por seu tempo, lidar com *perceptrons* multicamadas, que conseguiam então resolver o problema do "ou-exclusivo" (Medeiros, 2018).

A partir de então, pode-se afirmar que a cada década houve saltos significativos na pesquisa, sendo oportuno enumerá-los: a proposição do algoritmo de retropropagação de David Rumelhart, Geoffrey Hinton e Ronald Williams para treinamento de redes neurais (McClelland; Rumelhart; Hinton, 1986); o trabalho seminal de Hinton com redes de crença profunda (Hinton; Osindero; Teh, 2006), crucial para o advento das técnicas de aprendizagem profunda; as redes adversariais generativas de Goodfellow *et al.* (2014); e,

por fim, o artigo de Vaswani e colaboradores intitulado "*Attention Is All You Need*", que evolui o conceito de transformadores utilizados para o processamento de linguagem natural inicialmente para a tradução automática e, em consequência, para a geração criativa de textos (Vaswani *et al.*, 2017).

Por esse breve histórico, pode-se observar a evolução do paradigma conexionista no campo de pesquisa em IA, materializada no desenvolvimento de redes neurais artificiais: a ideia de que é possível obter a performance das funções superiores relacionadas com o pensamento humano a partir de arquiteturas contendo neurônios artificiais, abstraídos dos seus correlatos biológicos. Seu contraponto, o paradigma simbólico, coloca que o processamento das funções superiores deve ser alcançado simulando-se, de forma direta, os processos mentais por meio de algoritmos que lidam diretamente com a sintaxe e a semântica dos símbolos, sejam imagens, palavras ou sons. Bengio, LeCun e Hinton designam o paradigma conexionista como "inspirado no cérebro" e o paradigma simbólico como "inspirado na lógica".

> Em termos simples, o paradigma inspirado na lógica vê o raciocínio sequencial como a essência da inteligência e busca implementar o raciocínio em computadores por meio de regras de inferência projetadas manualmente que operam sobre expressões simbólicas também projetadas e que formalizam o conhecimento. Já o paradigma inspirado no cérebro vê o aprendizado de representações a partir de dados como a essência da inteligência e busca implementar o aprendizado projetando manualmente — ou fazendo evoluir — regras para modificar os pesos de conexão em redes simuladas de neurônios artificiais (Bengio; Lecun; Hinton, 2021, p. 58-59).

De acordo com os autores, no paradigma orientado pela lógica, os símbolos não carregam uma estrutura interna relevante: o seu sentido provém das relações com outros símbolos, que podem ser descritas por um conjunto de expressões simbólicas ou por um grafo conectando relações. Já no paradigma inspirado no funcionamento do cérebro, os símbolos externos usados na comunicação são mapeados para vetores internos de atividade neural, que exibem uma rica estrutura de similaridade. Esses vetores podem modelar a estrutura presente em sequências de símbolos, ao aprender vetores adequados para cada símbolo e transformações não lineares que possibilitam inferir e completar os elementos faltantes de uma sequência (Bengio; Lecun; Hinton, 2021, p. 59).

O GPT NO QUARTO CHINÊS DE SEARLE

Um outro lado do debate entre os paradigmas simbólico e conexionista pode ser considerado a partir do experimento de pensamento de John Searle, imaginado para argumentar que a manipulação de símbolos conforme um conjunto de regras não implica na compreensão daquilo que está sendo manipulado. Tal experimento permite chegar a um questionamento controverso quando considerado à luz das tecnologias generativas.

No quarto chinês de Searle (Searle, 1980; Teixeira, 1991), uma pessoa recebe na janela de entrada ideogramas, pesquisa em um manual a associação desses ideogramas com outros e depois apresenta à janela de saída os ideogramas correspondentes. Esse processo representa perfeitamente o que acontece em um analisador sintático descendente, uma estrutura presente em compiladores e interpretadores de linguagem de programação, utilizado para analisar uma estrutura gramatical de uma expressão (Parr, 2012). O analisador entrega o resultado de sua análise para um módulo de processamento semântico, que executará as ações conforme constam na expressão de entrada.

O experimento imaginado por Searle (1980) permite constatar a separação nítida entre sintaxe e semântica. Enquanto a sintaxe diz respeito à manipulação formal de símbolos seguindo regras específicas, a semântica envolve o significado e a intencionalidade, ou seja, sobre o que os símbolos estão expressando. Conforme Searle, o desempenho linguístico em nível computacional não é suficiente para que aconteça a compreensão semântica. Em outras palavras, o processamento simbólico em si não garante a existência de uma "mente" ou algum nível de compreensão, mas manifesta um comportamento que se adequa às regras.

A implantação de *chatbots* sob o paradigma simbólico segue essa analogia. É necessária a interpretação sintática das frases de um interlocutor que são alimentadas ao *chatbot*. Esse *chatbot* pesquisa em uma lista interna, determinística ou probabilística, de respostas possíveis que devem ser apresentadas de volta e, por fim, escolhe uma que é mostrada ao interlocutor. O processamento sintático se dá de forma separada do semântico e esse é programado/implementado de forma explícita por um humano, o qual "empresta" a sua representação do mundo para o *chatbot*.

Na substituição do *chatbot* simbólico por um modelo GPT, não existe agora um módulo ou um processo explícito de manipulação direta de símbolos. Em vez disso, há uma mescla de sintaxe e semântica: a rede de transformadores aprende representações numéricas na forma de vetores, que encapsulam tanto a estrutura das expressões (a sintaxe) quanto o significado delas (a semântica). Dessa forma, sendo permitido falar em um modelo GPT capaz de algum nível de "compreensão", isso emergiria a partir dos padrões que o modelo aprende a partir de massivas quantidades de textos. Em princípio, pareceria que o transformador estaria imprimindo em suas redes uma mescla entre sintaxe e semântica.

Portanto, um modelo GPT como representante do paradigma conexionista traria em si, entrelaçados e de forma implícita, os aspectos de sintaxe e semântica. Os mecanismos de atenção presentes nos transformadores permitem ao modelo GPT capturar dependências bem complexas e nuances semânticas presentes nas expressões sintáticas. A crítica de Searle (2002) derivada do experimento pode continuar se sustentando, no consenso de que o GPT não está "compreendendo" as expressões a partir de uma perspectiva da subjetividade humana. O GPT consegue fazer o processamento de textos e elaborar as respostas tendo como base um aprendizado de padrões estatísticos. No entanto, isso é feito sem implicar uma compreensão, em nível humano, dos símbolos de forma consciente ou intencional.

Por outro lado, tomando como base o entendimento humano como algo que emerge a partir de associações entre elementos simbólicos, resgatando inclusive aquilo que Frege argumentava sobre o sentido emergindo a partir de um conjunto de relações entre proposições (2013), os modelos GPT poderiam consistir em um vislumbre inicial daquilo que, em certa medida, estaria acontecendo nas redes de neurônios responsáveis pela linguagem no ser humano. No entanto, é oportuno mencionar que em investigações atuais, sendo conduzidas sobre os espaços semânticos dos transformadores, não há um consenso sobre o que conta exatamente como "significado" no conjunto dos vetores dentro de um modelo (Nikolaev; Padó, 2023).

UM GPT E SEUS "PENSAMENTOS"

Na esteira da teoria das proposições de Frege, pode-se refletir sobre a questão de

uma entidade GPT estar "desempenhando" algum tipo de pensamento, enquanto expressa as respostas aos *prompts* em linguagem natural com algum interlocutor. De fato, a motivação original de Frege quanto ao desenvolvimento de sua abordagem refletiu inicialmente a necessidade da separação entre o lógico e o psicológico, daquilo que é objetivo em relação ao subjetivo da experiência interna do indivíduo (Frege, 2013, p. 43).

É importante salientar que, para Frege, o "objetivo" compõe-se daquilo que é independente da consciência e que não pertenceria à experiência interna própria de um indivíduo e dos outros indivíduos, por consequência. Existiria uma espécie de "objetivo não real": o mundo do objetivo não se encerraria nas coisas concretas do mundo, as quais podem ser percebidas pelos sentidos (Frege, 2013, p. 200). Esse "objetivo" se estenderia também para aquilo que escapa às sensações e que, por sua vez, pode ser considerado um conteúdo de consciência de muitos, além do individual.

Como uma forma de se identificar a separação do objetivo e do subjetivo, a ontologia de Frege se manifesta em três mundos: há um primeiro mundo, real, físico, composto dos objetos que são perceptíveis pelos sentidos; o segundo mundo, relativo ao mundo interno, a aquilo que passa pela consciência, composto pelas ideias, impressões e sensações provenientes dos sentidos, em alinhamento à Hume; e o terceiro mundo, que consiste, enfim, no mundo do "objetivo não real", que é constituído pelos pensamentos (Frege, 2013, p. 43).

Tal distinção é relevante para a caracterização da ideia de um GPT lidando com pensamentos. Para Frege, a atividade de pensar é distinta da atividade de intuir: a intuição estaria ligada à percepção e à imaginação. Também é importante ressaltar o fato de que a atividade de pensar não é levada a cabo externamente; apenas é possível a expressão de modo externo.

Outro elemento a ser considerado refere-se às formas de representação. Na concepção de Frege, a representação também não se refere ao pensamento (Frege, 2013, p. 212). Ela consiste em uma imagem elaborada a partir das impressões dos sentidos. Portanto, a representação é algo subjetivo e, portanto, própria do indivíduo. A representação entra em cena, fornecendo os elementos para o desempenhar do pensamento. Ressalvando-se um relaxamento da premissa de subjetividade para fins da

analogia, pode-se dizer que um GPT, após o treinamento, tem em sua rede de transformadores uma representação adquirida da sua base de treinamento, contendo probabilidades envolvidas nos relacionamentos entre os termos; portanto, é uma representação que contém o potencial para expressão do pensamento (de máquina) sem se constituir, entretanto, em um pensamento em tal momento.

Assim, a elaboração de uma resposta a um *prompt* enunciado por um humano pode ser considerada uma atividade de pensamento, embora caracterizado propriamente como de máquina. Os mecanismos intrínsecos são, com certeza, bastante distintos (ainda que inspirados) daqueles que se passam na mente humana. A forma lógica com a qual uma sentença ou conjunto de sentenças são elaboradas pelo GPT é, de fato, a "substância" da conversação de modo interativo com um humano, o qual vai elaborando *prompts* na expressão do conteúdo do seu pensamento.

É interessante assinalar também, ainda de acordo com Frege (2013), que é possível expressar um pensamento sem propô-lo como verdadeiro. Na relação entre sentenças interrogativas e afirmativas, uma interrogação refere-se a uma sentença incompleta que irá alcançar um verdadeiro sentido somente a partir do complemento que foi solicitado. Pode-se formar uma sentença afirmativa para cada oração interrogativa e é possível afirmar que as duas se referem ao mesmo pensamento; no entanto, a sentença afirmativa contém algo mais: a própria asserção em si. Assim, em uma sentença afirmativa, o pensamento, o julgamento (valor de verdade) e a afirmação estariam tão conectados que seria fácil não perceber tal separação (Frege, 2013, p. 202).

Tal reflexão se desdobra em algo bastante significativo quanto à consistência das sentenças retornadas, ficando evidente a necessidade de uma análise crítica, externa ao sistema, das respostas dadas pelo GPT em face aos *prompts* enunciados. O encadeamento de palavras nas sentenças, seguindo o algoritmo de processamento estatístico da rede transformadora, pode levar eventualmente a respostas em que o GPT esteja "alucinando" de fato: há um pensamento sendo enunciado em resposta ao *prompt*, mas equivocado com respeito ao julgamento, ao valor de verdade e esvaziando-se, portanto, o sentido da sentença afirmativa.

Após as reflexões sobre as proposições objetivas de Frege, abre-se espaço para a consideração do GPT como uma entidade lidando com a linguagem caracterizada tanto

como uma forma computada, assim como baseada no uso. Entretanto, torna-se oportuna uma analogia com a forma distribuída e não central a respeito do processamento efetuado por um GPT.

GPT E O PROBLEMA DE HUME

O empirista britânico David Hume, em seu *Tratado da Natureza Humana*, sustentava em sua filosofia que o conhecimento e o pensamento humanos são derivados de impressões sensoriais e das ideias que derivam dessas impressões. Para Hume, as ideias não possuem uma existência independente: elas surgem das impressões dos sentidos e interagem entre si (Hume, 2000, p. 34-37). O raciocínio, portanto, não é algo realizado por um agente executivo ou uma espécie de "eu" central, mas sim pelo resultado das associações entre as ideias. Esse modelo faz surgir uma questão importante: como as ideias podem interagir entre si e gerar pensamento sem a necessidade de um "homúnculo", ou seja, uma entidade central que "pensa" e "controla" o processo? Se não há um "eu" ou agente central conduzindo as ideias, o que faz com que o pensamento exista?

O filósofo Daniel Dennett tratou da questão da consciência e do pensamento de forma a rejeitar a noção de tal "agente central" ou "homúnculo", presente na mente. Sua proposição é a de que o pensamento surge de um processo distribuído e descentralizado, em que as miríades de processos cognitivos e subsistemas mentais interagem de forma paralela e colaborativa, sem haver a necessidade de uma entidade supervisora consciente (Dennett, 2006, p. 179).

Conforme Dennett, a IA consegue fornecer um modelo concreto de como isso pode acontecer. Sistemas de IA, especialmente aqueles baseados em redes neurais e aprendizado de máquina, não possuem um "centro de controle" ou uma "mente" consciente. No entanto, possuem a capacidade de realização de tarefas complexas de processamento de informações e tomada de decisões. O que tais sistemas demonstram é que é possível haver o pensamento e o raciocínio, sem a necessidade de uma entidade consciente ou central que organiza tudo (Dennett, 2006, p. 180).

Com base nisso, pode-se dizer que o GPT, portanto, consolida a asserção de Dennett sobre a IA ser a única forma de solucionar como as ideias "pensam" por elas

mesmas. Ele consegue elaborar sentenças complexas e consistentes a partir de uma rede distribuída proporcionada pela arquitetura dos transformadores. A intencionalidade de um usuário expressa em um *prompt* é completada pelo GPT a partir do contexto linguístico fornecido e não existindo, portanto, um controle executivo na elaboração das respostas ao *prompt*.

GPT: UMA MÁQUINA DE WITTGENSTEIN

Na continuidade do desenvolvimento da tese logicista iniciada a partir de Frege e após o inventário empreendido por Russell e Whitehead no *Principia Mathematica*, Wittgenstein conduz, no *Tractatus Logico-Philosophicus,* a uma doutrina filosófica para o trato do significado de proposições sob condições de verdade determinada, fazendo uso de um estilo comprobatório baseado na lógica. Nesse sentido, Wittgenstein propõe que o mundo é a totalidade dos fatos e não das coisas, sendo que esses fatos são representados por proposições que funcionam como figuras lógicas da realidade (Wittgenstein, 2010).

Cada proposição, para Wittgenstein, corresponde a uma "imagem" da realidade, em que a estrutura lógica da proposição reflete a estrutura do fato ao qual ela se refere. Assim, a linguagem, enquanto sistema de proposições, adquire seu significado na medida em que as proposições podem ser verdadeiras ou falsas, dependendo da correspondência com a realidade. A lógica, nesse contexto, surge não apenas como uma ferramenta de análise, mas como a própria forma do pensamento e da linguagem. Para Wittgenstein, a estrutura lógica do mundo é espelhada pela estrutura lógica da linguagem e é por meio dessa correspondência que podemos determinar as condições de verdade das proposições (Buchholz, 2009).

Além disso, Wittgenstein defende que as proposições são combinações de nomes e esses nomes designam objetos simples e atômicos, que são os componentes básicos da realidade (Fann, 2013, p. 30). A partir da combinação desses nomes, formam-se as proposições que, quando valoradas como verdadeiras, constituem a representação exata dos fatos existentes no mundo.

Entretanto, para levar a efeito a análise do GPT como uma "máquina" que lida com a linguagem combinando proposições na resposta a um *prompt*, é necessário entrar em

cena o "segundo" Wittgenstein. Conforme as considerações de William Frawley (2000, p. 59) quanto à contribuição de Wittgenstein para o estudo da linguagem, enquanto o *Tractatus* se baseia em definições ostensivas assumindo uma relação direta entre mundo e linguagem, em *Investigações Filosóficas* (2012) Wittgenstein argumenta em favor do significado como uso. Enquanto há o caráter de determinação presente no primeiro, no segundo, o significado da palavra "significado" é "assinalar que os papeis de todas as outras estão sobre a mesa e sobre escrutínio" (Frawley, 2000, p. 60).

Nesse ínterim, Wittgenstein introduz o conceito de "jogos de linguagem" como uma maneira de demonstrar que o significado das palavras não é fixo ou determinado por uma correspondência direta com o mundo, mas ao contrário, consiste em algo fluido e dependente das práticas humanas (Fann, 2013, p. 70). Os jogos de linguagem são caracterizados por sua multiplicidade e sua variedade, refletindo a diversidade de atividades nas quais as palavras são utilizadas. Cada jogo de linguagem consiste em uma espécie de prática social, em que as regras que determinam o significado das palavras são estabelecidas e compreendidas dentro de um contexto específico.

Dessa forma, Wittgenstein sugere que compreender o significado de uma palavra é entender como ela está estabelecida em um certo jogo de linguagem. Não existe, portanto, uma essência ou uma definição ostensiva que possa capturar o significado de uma palavra em todos os contextos (Frawley, 2000, p. 59). Ao invés disso, o significado é algo que emerge a partir do uso cotidiano da linguagem, das diferentes formas de vida em que os jogos de linguagem estão inseridos.

Munido desse quadro, pode-se conjecturar sobre um modelo GPT como uma "entidade" que procede a uma combinação de proposições de forma lógica, de acordo com os algoritmos presentes na rede de transformadores, mas também embute em si, de forma paralela e distribuída na própria rede, a multiplicidade das formas com as quais as palavras, dentro das sentenças, acomodam as diferentes práticas de linguagem, presentes em uma infinidade de textos que contribuíram para a formação da sua base de conhecimento.

Em suma, pode-se concluir que o GPT se constitui, portanto, em uma espécie de "máquina de Wittgenstein", capaz de lidar com a linguagem por um modo computacional e lógico, atento às questões de apresentação fluida do significado no âmbito da

diversidade de jogos de linguagem então presentes ao momento do treinamento. No dizer de Frawley (2000, p. 57), o comportamento do GPT surgiria da tensão perpétua entre a "máquina virtual" e a "máquina real", legitimando-se a partir de fatos determinados em modo computável, mas considerando as decisões de valoração conforme regras, de certa forma, indeterminadas, em função da flexibilidade e aleatoriedade permitida pelos parâmetros estatísticos estabelecidos na vasta rede dos transformadores.

CONSIDERAÇÕES FINAIS

Este breve ensaio teve o objetivo de promover uma discussão, tendo como questão central a reflexão sobre como os modelos GPT permitiram a predominância das arquiteturas conexionistas, proporcionando uma discussão sobre aspectos de sintaxe e semântica à luz de autores como Frege, Dennett, Searle e concluindo com Wittgenstein, cuja noção sobre o significado transita sobre uma tensão dialética na linguagem entre uma forma computada baseada na lógica e uma forma de ação baseada no uso.

O "transformador", utilizado em modelos GPT, pode ser considerado como uma forma de realização madura do conexionismo no campo do processamento de linguagem natural. Ao contrário dos modelos simbólicos, cujo domínio provinha desde a década de 1970, as redes neurais artificiais, incluindo os transformadores, operam de maneira distribuída e paralela, aproximando-se de uma simulação mais direta dos mecanismos de cognição biológica. O transformador, com seu mecanismo de atenção, permite que o modelo processe sequências longas e capture dependências complexas na linguagem, superando as limitações das arquiteturas de rede neural anteriores como, por exemplo, a tecnologia das redes recorrentes.

De certo modo, isso parece caracterizar uma espécie de "revanche" do conexionismo, que havia sido criticado por sua incapacidade de lidar eficientemente com problemas cognitivos de alto nível. O sucesso dos transformadores na década de 2010 mostra que, ao amadurecer em termos de poder computacional e complexidade de arquitetura, o conexionismo tornou-se a base dominante na pesquisa em IA estendendo-se ao campo do processamento de linguagem natural. Com o avanço das pesquisas na área, é de se esperar o surgimento de outros tipos de transformadores, ou mesmo

arquiteturas inovadoras, que venham a apresentar um melhor desempenho do que as atuais no tratamento computacional da linguagem.

Entretanto, é oportuno ressaltar que às implementações utilizando GPT devem ser agregados outros módulos que complementam a tarefa de conversação em linguagem natural, tais como memória de *prompts* anteriores, filtros de conteúdo sensível ou geradores de síntese de textos provenientes do GPT propriamente dito. Supõe-se, portanto, que tais implementações reais devam ser necessariamente híbridas, considerando tanto o conexionismo do GPT quanto os módulos agregados ao seu funcionamento, combinando em diferentes graus as aquisições tanto no âmbito do paradigma conexionista quanto no do simbólico.

REFERÊNCIAS

ALAMMAR, J. The Illustrated Retrieval Transformer. **Jay Alammar**, 3 jan. 2022. Disponível em: https://jalammar.github.io/illustrated-retrieval-transformer/. Acesso em: 15 ago. 2024.

ANTHROPIC. **Introducing the next generation of Claude**. Disponível em: https://www.anthropic.com/news/claude-3-family. Acesso em: 15 ago. 2024.

BENGIO, Y.; LECUN, Y.; HINTON, G. Deep learning for AI. **Communications of the ACM**, v. 64, n. 7, p. 58-65, 2021. DOI: https://doi.org/10.1145/3448250. Disponível em: https://dl.acm.org/doi/pdf/10.1145/3448250. Acesso em: 15 ago. 2024.

BROWN, T. B. *et al.* Language Models are Few-Shot Learners. **arXiv**, v. 33, p. 1877-1901, 2020. Disponível em: https://arxiv.org/pdf/2005.14165. Acesso em: 15 ago. 2024.

BUBECK, S. *et al.* Sparks of Artificial General Intelligence: Early experiments with GPT-4. **arXiv**, v. 1, p. 1877-1901, 2023. DOI: https://doi.org/10.48550/arXiv.2303.12712. Disponível em: https://arxiv.org/pdf/2303.12712. Acesso em: 15 ago. 2024.

BUCHHOLZ, K. **Compreender Wittgenstein**. 2. ed. Petrópolis: Vozes, 2009.

CHURCHLAND, P. M. **Matéria e Consciência**. São Paulo: Editora UNESP, 2004.

DENNETT, D. C. **Brainstorms: escritos filosóficos sobre a mente e a psicologia**. São Paulo: UNESP, 2006.

DEVLIN, J. *et al.* BERT: Pre-training of deep bidirectional transformers for language understanding. *In*: NAACL HLT 2019 - 2019 CONFERENCE OF THE NORTH AMERICAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS**:** Human Language Technologies. **Proceedings** [...], v. 1, n. Mlm, p. 4171-4186, 2019. DOI: 10.18653/v1/N19-1423. Disponível em: https://aclanthology.org/N19-1423.pdf. Acesso em: 15 ago. 2024.

ELHAGE, *et al.* A Mathematical Framework for Transformer Circuits. **Transformer Circuits Thread**, 2022. Disponível em: https://transformer-circuits.pub/2022/in-context-learning-and-induction-heads/index.html. Acesso em: 15 ago. 2024.

FANN, K. T. **El Concepto de Filosofía en Wittgenstein**. 3. ed. Madrid: Editorial Tecnos, 2013.

FRAWLEY, W. **Vygotsky e a Ciência Cognitivas:** Linguagem e interação das mentes social e computacional. Porto Alegre: Artes Médicas Sul, 2000.

FREGE, G. **Ensayos de Semántica y Filosofia de La Lógica**. 2. ed. Madrid: Tecnos, 2013.

GEORGIEV, P. *et al.* Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. **arXiv**, v. 1, p. 1-154, 2024. DOI: https://doi.org/10.48550/arXiv.2403.05530. Disponível em: https://arxiv.org/pdf/2403.05530. Acesso em: 15 ago. 2025.

GOODFELLOW, I. *et al.* Generative adversarial networks. **arXiv**, v. 27, 2014. DOI: https://doi.org/10.48550/arXiv.1406.2661. Disponível em: https://arxiv.org/pdf/1406.2661. Acesso em: 15 ago. 2024.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Cambridge-MA: MIT Press, 2016.

HAYKIN, S. **Redes Neurais: Princípios e prática**. Porto Alegre: Bookman, 2001.

HINTON, G. E.; OSINDERO, S.; TEH, Y.-W. A Fast-Learning Algorithm for Deep Belief Nets. **Neural Comput.**, v. 18, n. 7, p. 1527-1554, 2006. DOI: 10.1162/neco.2006.18.7.1527.

HOPCROFT, J.; MOTWANI, R.; ULLMAN, J. **Introduction To Automata Theory , Languages , and Languages**. Boston-MA: Pearson Education, Inc, 2006.

HUME, D. **Tratado da Natureza Humana**. São Paulo: UNESP, 2000.

KITTLER, F. A. **Gramophone, Film, Typewriter**. [s. l.] Stanford University Press, 1999.

MANNING, C. D.; RAGHAVAN, P.; SCHÜTZE, H. **An Introduction to Information Retrieval.**

Cambridge, England: Cambridge University Press, 2009.

MANNING, C. D.; SCHÜTZE, H. **Foundations of Statistical Natural Language Processing.** [*s. l.*]: The MIT Pres, 1999.

MCCLELLAND, J. L.; RUMELHART, D.; HINTON, G. E. The Appeal of Parallel Distributed Processing. *In*: **Parallel Distributed Processing:** Exploration of the microstructure of cognition. Cambridge: MIT Press, 1986. p. 3-44.

MEDEIROS, L. F. de. **Inteligência Artificial Aplicada:** Uma abordagem introdutória. Curitiba: Intersaberes, 2018.

NIKOLAEV, D.; PADÓ, S. Investigating Semantic Subspaces of Transformer Sentence Embeddings through Linear Structural Probing. *In*: BLACKBOXNLP WORKSHOP: ANALYZING AND INTERPRETING NEURAL NETWORKS FOR NLP, 6, p. 142-154, 2023. **Proceedings [...]**, Association for Computational Linguistics, Singapore, 2023.

OLSSON, C. *et al.* In-context Learning and Induction Heads. **Transformer Circuits Thread**, Mar 8, 2022. Disponível em: https://transformer-circuits.pub/2022/in-context-learning-and-induction-heads/index.html. Acesso em: 15 ago. 2024.

OPENAI. **Hello GPT-4o**. Disponível em: https://openai.com/index/hello-gpt-4o/. Acesso em: 13 abr. 2024.

PARR, T. **The Definitive ANTLR 4 Reference**. Dallas, Texas: The Pragmatic Programmers, LLC, 2012.

PLATÃO. **Fedro**. Tradução Maria Aparecida A. De Oliveira. São Paulo: Martin Claret, 2001.

RAFFEL, C. *et al.* Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. **arXiv**, v. 1, 2019. DOI: https://doi.org/10.48550/arXiv.1910.10683. Disponível em: https://arxiv.org/pdf/1910.10683. Acesso em: 15 ago. 2024.

RUSSELL, S.; NORVIG, P. **Inteligência Artificial** - Tradução da 2a edição. Rio de Janeiro: Editora Campus, 2004.

RYLE, G. **El Concepto de lo Mental**. Barcelona: Ediciones Paidós Iberica, 2005.

SEARLE, J. **Intencionalidade**. 2. ed. São Paulo: Martins Fontes, 2002.

TEIXEIRA, J. de F. Robots, intencionalidade e inteligência artificial. **Trans/Form/Ação**, v. 14, p. 109-121, 1991. Disponível em: https://www.scielo.br/j/trans/a/w7DK95zgJFjWV7mw6NdhfRB/?format=pdf&lang=pt. Acesso em: 15 ago. 2024.

TOUVRON, H. *et al.* LLaMA: Open and Efficient Foundation Language Models. **arXiv**, v. 1, 2023. DOI: https://doi.org/10.48550/arXiv.2302.13971. Disponível em: https://arxiv.org/pdf/2302.13971. Acesso em: 15 ago. 2024.

VASWANI, A. *et al.* Attention is all you need. **arXiv**, v. 1, 2017. DOI: https://doi.org/10.48550/arXiv.1706.03762. Disponível em: https://arxiv.org/pdf/1706.03762. Acesso em: 15 ago. 2024.

WITTGENSTEIN, L. **Tractatus Logico-Philosophicus**. São Paulo: Editora da USP, 2010.

WITTGENSTEIN, L. **Investigações Filosóficas**. 7a. ed. Petrópolis: Vozes, 2012.

ZHAO, W. X. *et al.* **A Survey of Large Language Models**. 31 mar. 2023.