

Espelhamentos: o GPT e a educação¹

Adilson Citelli

Professor titular sênior do Departamento de Comunicações e Artes e do Programa de Pós-Graduação em Ciências da Comunicação da Escola de Comunicações e Artes da Universidade de São Paulo (ECA-USP). É coordenador do Grupo de Pesquisa Mediações Educomunicativas (Mecom).

E-mail: citelli@uol.com.br

Resumo: Neste artigo, situamos alguns elementos circundantes do ChatGPT que estão desafiando o campo de estudos educacionais. Os avanços em andamento na inteligência artificial generativa vêm provocando alterações nos modos de acessar informações, elaborar discursividades, promover traduções entre variadas línguas, acionar os potenciais intersemióticos, tudo em tempo acelerado e de relativa facilidade operacional. Trata-se, portanto, de um sistema e seus processos econômicos e comunicacionais, com implicações nos diversos segmentos da vida social, cujas potencialidades podem ser identificadas, malgrado pairarem interrogantes acerca do futuro a ele reservado. Neste contexto, consideramos que a educação, em seus diferentes níveis e séries, será instada a implementar estratégias de reconhecimento dos modelos generativos de linguagem, com eles dialogando a partir de uma perspectiva analítico-crítica.

Palavras-chave: ChatGPT; comunicação; educação; algoritmos; mudanças.

Abstract: This study highlights some of the challenging elements surrounding ChatGPT for educative studies. The ongoing advances in generative artificial intelligence have changed the ways of accessing information, preparing discourses, promoting translations between different languages, and activating intersemiotic potentials, which occur in an accelerated time and with relative operational ease. Therefore, this system includes economic and communicational processes with implications for the various segments of social life, the potentialities of which can be identified, despite questions surrounding the future the device holds for us. In this context, we consider that the different levels and series of education will be urged to implement strategies to recognize generative language models, dialoguing with them from an analytical-critical perspective.

Keywords: ChatGPT; communications; education; algorithms; changes.

1. A versão preliminar deste artigo foi apresentada no XLVI Congresso Brasileiro de Ciências da Comunicação, Intercom, realizado na PUC de Belo Horizonte, em setembro de 2023. O que se dá à leitura conheceu mudanças e ajustes, muitos deles decorrentes das discussões processadas naquele evento.

Recebido: 28/03/2024

Aprovado: 25/05/2024

2. Fundada por um grupo de investidores, entre outros, Elon Musk, Sam Altman, Greg Brockman, Ilya Sutskever, Wojciech Zaremba. Alguns desses nomes, como Musk, afastaram-se da OpenAI, e Altman tornou-se o presidente. Conforme amplamente divulgado, em novembro de 2023 ocorreram idas e vindas na empresa, com a demissão e posterior recontração de Altman.

3. Ver o artigo "The open letter to stop 'dangerous' AI race a huge mess", que comenta a carta assinada por milhares de signatários, entre eles Elon Musk, da Tesla, Steve Wozniak, cofundador da Apple, Yuval Noah Harari, historiador, Evan Sharp, cofundador da Pinterest, produzida pelo Instituto Future of Life, pedindo moratória de seis meses no desenvolvimento de novos alcances de inteligência artificial, a exemplo do ChatGPT-4. Disponível em: <https://www.vice.com/en/article/qjvppm/the-open-letter-to-stop-dangerous-ai-race-is-a-huge-mess>. Acesso em: 10/06/2023. A íntegra da carta, sob o título "Pause giant AI experiments: an open letter", pode ser lida em: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>. Acesso em: 10/06/2023. Yuval Harari, em artigo publicado no *New York Times*, sob o título "Yuval Harari on threats to humanity posed by AI", não foge ao tom geral acerca do imperativo de colocar freios no espraiamento das novas gerações de inteligência artificial "[...] dominar a IA antes que ela nos domine" (tradução livre). Disponível em: <https://nytimes.com/2023/03/24/opinion/yuval-harari-ai-chatgpt.html>. Acesso: 12/06/2023. Há, em contrapartida, inúmeras análises que, malgrado reconhecerem os problemas existentes junto a estes sistemas de "inteligência competitiva" (no caso homem-máquina) consideram manifestos e cartas, à maneira das produzidas pelo Instituto Future of Life, como uma cortina de fumaça para obnubilar as ações concretas das "tecbros".

*Que cérebro eletrônico nenhum me dá socorro
Com seus botões de ferro e seus olhos de vidro.*

Gilberto Gil. *Cérebro eletrônico.*

PRENÚNCIO

Em novembro de 2022, a *startup* californiana OpenAI,² uma espécie de anjo Gabriel redivivo, anunciou para a humanidade ser a mãe ou o pai de um avançado *software* de inteligência artificial batizado de ChatGPT (*Generative Pre-trained Transformer – Transformador Generativo Pré-Treinado*), com capacidade de produzir textos dotados de coesão e coerência bastando, para tanto, que os usuários fizessem perguntas procedentes e acionassem os devidos comandos de entrada no sistema. O resto da história é conhecida e *in progress*: dois meses após o seu lançamento, mais de cem milhões de pessoas acessavam a novidade, já em versão avançada e sob o designativo de GPT-3, divertindo-se, indignando-se, (des)informando-se, fornecendo dados gratuitamente à empresa, que, por sua vez, servem para, entre outras operações, aperfeiçoar o funcionamento do dispositivo. Nos dias correntes – referimo-nos a um período entre 2023 e começo de 2024 – acumulam-se cursos, palestras, seminários, matérias midiáticas, vídeos explicativos, numa arregimentação de ânimos a prognosticar sejam as inovações promovidas pela chamada inteligência artificial generativa, sejam os dissabores nela incluídos.³

O GPT-3, pelo seu pioneirismo (a versão 4⁴ está disponível para assinantes GPT Plus, à razão de US\$ 20,00 mensais, trazendo melhorias como interpretar imagens e produzir textos com até 25 mil palavras – contra 2 mil do modelo anterior), tornou-se uma espécie de marco nesta linhagem de melhorias na inteligência artificial. A despeito das luzes lançadas sobre a criação da OpenAI, produtos de família próxima fazem parte do cardápio de outras empresas do ramo: Bing, da Microsoft, Bard, da Alfabet/Google, Ernie Bot – em mandarim, Wenxin Yiyan, –, da chinesa Baidu. Além disso, seguem siglas e nomes voltados, sobretudo, aos geradores de imagens como o DALL-E 2, Stable, Diffusion, Midjourney.

Aduza-se que a rapidez na diversificação dos modelos de linguagem parece justificar a crítica, a indignação cínica, ou o pasmo, presentes nos já referidos manifestos, indicando que após serem treinados por humanos, as *machine learning* passaram a exigir que os humanos sejam treinados (ou iluminados) para entender os desígnios dos "cérebros" eletrônicos. Tarefa, diga-se de passagem, circunscrita e nem sempre bem-sucedida, ao menos conforme a perspectiva apocalíptica ecoada por vários cientistas da computação, filósofos e cognitivistas que enxergaram no ensandecimento do computador senciente Hal 9000, do clássico filme de 1968, dirigido por Stanley Kubrick, *2001, uma odisseia no espaço*, o exemplo do que nos aguarda em anos vindouros. Assim, a célebre frase *I'm sorry Dave, I'm afraid I can't do that* (Sinto muito Dave, infelizmente não

posso fazer isso) com que o maquínico Hal contraria uma ordem do astronauta Dave Bowman, esclarecendo ter ganhado vida própria, se autonomizando e desejando, doravante, comandar/dirigir a nave interplanetária Discovery One, poderia funcionar, metaforicamente, como vaticínio do nosso futuro. A boa notícia é que, no filme, Dave, consegue desligar o “ser” cibernético Hal e retomar o controle da viagem para Júpiter.

Pondere-se que os mecanismos generativos acionados por algoritmos,⁵ operações lógicas, processamento de imagens e linguagem natural, não revelam as suas estratégias de elaboração, implantação e distribuição, haja vista serem intangíveis ao usuário comum. Neste território opaco, o fetiche circundante da inteligência artificial – se é que a máquina possui alguma⁶ – transita por padrões estatísticos, probabilísticos, preditivos, a minerarem novos dados que aumentarão o estoque das já fornidas bases das *tecbros*. E, neste caso, dados e poder formam uma dupla afinadíssima (Zuboff, 2019).

1. DESENHO

O GPT é um sistema pré-treinado⁷ (por pessoas), que transforma linguagem artificial em natural, gerando, a partir de perguntas/formulações/consultas feitas por usuários, as respostas buscadas, permitindo a existência de “conversa” virtual. Vale dizer: estamos diante de modelo baseado no *transformer*, uma arquitetura de rede neural que, feitos os *inputs*, opera e analisa imensa quantidade de dados, distinguindo bilhões de parâmetros⁸ sobre a maneira de os humanos correlacionarem palavras, signos, símbolos, números; a partir daí elabora novos textos em linguagem natural. O GPT possui a capacidade de analisar 570 gigabytes de dados da internet, algo como 300 bilhões de palavras, do que resulta a identificação de 175 bilhões de parâmetros – as formatações iniciais tinham em torno de 1,5 bilhões.

Em síntese, o *chatbot* de que nos ocupamos é uma inteligência artificial generativa – Chomsky chama de “automatic learning” – “aprendizado automático” (2023) – que minera/processa uma grande quantidade de dados/informações, produzindo materiais em linguagem natural a partir de demandas feitas pelos interessados humanos. A categorização de Chomsky é oportuna, haja vista tratar-se, fundamentalmente, de técnicas baseadas em álgebra linear, cálculo e estatística.

O GPT opera no padrão LLMs (*Large Language Models* – Grandes Modelos de Linguagem) cuja estrutura funcional consiste em encaminhar – respeitados os contextos enunciativos – probabilidades na elaboração, por exemplo, de frases ou textos mais longos, a partir dos referidos modelos estatísticos que calculam frequências de unidades expressivas e direcionam as escolhas consideradas procedentes. Os cálculos probabilísticos efetivam-se a partir do manancial extraído da web, livros e artigos de repositórios digitais, Wikipedia etc., que facultarão constituir os bilhões de parâmetros orientadores do sistema. E das conexões subsequentes

Em outras palavras, enquanto os manifestos tratam as IAs como entidades autônomas, as corporações que as controlam podem seguir em suas disputas internas, ganhando dinheiro, aumentando o seu poder e comprometendo a democracia. Nesta direção, ver artigos de Sayash Kappor e Arvind Narayanan, publicados em *AI Snake Oil*: Disponível em: <https://aisnakeoil.substack.com/p/a-misleading-open-letter-about-sci> e Emily M. Bender, em *Search Medium*: Disponível em: <https://medium.com/@emilymenonbender/policy-makers-please-dont-fall-for-the-distractions-of-ai-hype-e03fa80ddb1>. Acompanhar tal debate é um repto diário e sempre carregado de novidades processadas com enorme rapidez.

4. No mês de março de 2024, o *chatbot* passou a disponibilizar aos consulentes respostas ditas em voz alta (*read aloud*). Até então os comandos poderiam ser feitos em voz, mas as respostas vinham em texto. As possibilidades de integração à plataforma geradora de imagens Dall-E (ainda com vários problemas) fazem parte destas funcionalidades agregadas às primeiras versões do GPT.

5. Para uma introdução ao tema, ver o livro da professora e modeladora de algoritmos, Cathy O’Neil (2000).

6. O neurocientista Miguel Nicolelis reitera que a IA não é inteligente sequer artificial, em comparativo com a inteligência humana. Esta é analógica, a operar com valores pouco ou nada programáveis, como fidelidade, crença, sinceridade, amorosidade – conforme os processos de interação, sociabilidade, empatia, entre pessoas –, aquela é cognitiva, digital e dependente da construção de algoritmos, programação, treinamento, robustas bases de dados etc.

Trataria-se, pois, de apropriação indevida de conceito, cujo uso apenas refletiria certas estratégias de marketing que circundam a IA. Disponível em: <https://www1.folha.uol.com.br/tec/2023/07/ia-nao-e-inteligencia-e-sim-marketing-para-explorar-trabalho-humano-diz-nicolelis.shtml>. Acesso em: 13/6/2024. Em linha próxima estão os escritos por Noam Chomsky, especialmente para o nosso presente interesse, ver o artigo saído em 8 de março de 2023, no *The New York Times*: Noam Chomsky: the false promise of ChatGPT. Disponível em: <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>. Sob a ótica da Filosofia da Linguagem, Wittgenstein propõe interessante discussão acerca da possibilidade de se formular perguntas em torno da relação pensar (inteligência) e máquina (artificial). No contexto dos problemas de lógica que envolvem a reflexão de Wittgenstein, sequer faz sentido indagar sobre a plausibilidade de as máquinas pensarem, conforme expresso pelo filósofo em 1933, e publicado no volume *The Blue and Brown Books: Preliminary Studies for the "Philosophical Investigations"* (1958).

7. Sem entrar em particularidades acerca do treinamento da inteligência artificial, é preciso acentuar que no início do processo estão seres humanos. São os treinadores/roteiristas que possuem formação em ciências de dados, estatística, matemática etc. Sem eles, a chamada "rotulagem de dados", quando as informações são expostas à IA, o GPT não funcionaria a contento. Na sequência, os sistemas vão se autonomizando, cruzando dados e informações, sendo "ensinados", permitindo que respostas cheguem aos usuários. A preparação é feita através de técnicas e etapas de coleta e organização dos dados, escolha do padrão a ser seguido, testes, avaliação, acertos, implantação e acompanhamento.

entre sujeito (perguntador) e máquina (respondente), podem resultar textos sobre a fertilidade do solo na Bahia, explicações de problemas matemáticos, sonetos de amor, tradução ao português de originais de várias línguas, biografias de Aristóteles ou Confúcio. No mais, o céu é o limite, conquanto as respostas dadas possam vir acompanhadas de erros, invencionices, vieses indesejáveis – afinal, o “cérebro” eletrônico, como gerenciador estatístico que é, não recebeu uma “educação” compromissada com dimensões éticas ou morais.

A arquitetura *transformer*, que permitirá as passagens da linguagem artificial para a linguagem natural, está, como referido em outro passo, organizada em redes neurais maquínicas – algo que não deve ser associado, imediatamente, às redes neurais humanas. Vale dizer que o processo é desencadeado a partir das possibilidades de uso de termos (funções numéricas não lineares “mapeiam” a entrada e a saída de certa palavra ou frase, por exemplo), sendo “orientado” pelo contexto das enunciações. Para os efeitos dos modelos de linguagem, tais comandos de entrada recebem pesos/números (as conversões são chamadas de *embeddings* – ou, em tradução aproximada, “incorporações”) que permitem que sejam elaboradas representações textuais de dados em certo espaço vetorial composto por números reais voltados a processar a linguagem natural em algoritmos de aprendizado de máquinas. Os *embeddings*, enquanto mecanismos representacionais, admitem que longos enunciados sejam compactados – o que é de grande importância física, de consumo de energia e econômica nos grandes computadores e bases de dados. O GPT, pelas suas dimensões e características, não está alocado em única GPU (*Graphics Processing Unit* – Unidade Central de Processamento Gráfico), mas em redes de computadores que ligam muitas GPUs e demais componentes de execução.

2. SENHAS E FORMATOS

No intuito de materializar o exposto nos tópicos anteriores, selecionamos alguns dos inúmeros exercícios “dialógicos” que realizamos com o *chat* visando verificar como ele se apresenta ou representa – se é que cabe tal definição frente ao insólito da circunstância. O nosso levantamento identificou várias estratégias discursivas, mas, em decorrência dos nossos limites editoriais, sistematizamos duas categorias: construção e desconstrução. No primeiro caso, encontram-se textos pertinentes do ponto de vista informativo, pouco desviantes no afeito aos fatos e cujo ajuste temático-formal pode contribuir positivamente para as solicitações dos usuários. No segundo caso, alinham-se escritos que, a despeito de não apresentarem grandes problemas do ponto de vista expressivo, registram erros mais ou menos crassos, invencionices, alucinações, elaborando respostas capazes de levar o consulente a equívocos e desinformações.

Para facilitar a leitura, intercalamos perguntas, respostas do GPT-3 – em itálico – e comentários subsequentes.

2.1. Construção

P. O que é significa a palavra *transformer*, que aparece na sigla GPT?

R. “A palavra “*transformer*”, que aparece em GPT, refere-se a uma arquitetura de rede neural conhecida como “*Transformer*”. O *Transformer* é um tipo de modelo de linguagem que se baseia em uma arquitetura de rede neural chamada ‘*encoder-decoder*’ para processar sequências de texto. Ele foi introduzido em 2017, em um artigo intitulado ‘*Attention Is All You Need*’, e se tornou uma das arquiteturas mais populares para tarefas de processamento de linguagem natural, como tradução automática, resumo de texto e geração de texto. O GPT (*Generative Pre-trained Transformer*) é um modelo baseado no *Transformer* que foi pré-treinado em uma grande quantidade de dados de texto e pode ser *finetuned* para tarefas específicas de processamento de linguagem natural”

C. A resposta, conquanto um pouco repetitiva, aparece, em linhas gerais, nos artigos de especialistas em inteligência artificial. Ou seja, os sistemas de linguagem como o GPT não seguem, exatamente, as estruturas tradicionais dos motores de busca de texto, mas passam por mecanismos de “nutrição” segundo os arranjos e rearranjos dos bilhões de parâmetros que serão transformados em conjuntos de signos passíveis de compreensão pelos seres humanos. Na sequência, indagamos sobre o que é *encoder*, *decoder*, *finetuned*, e acerca do artigo “*Attention is all you need*”. As respostas vieram acompanhadas de explicações satisfatórias e, no caso do artigo, considerado uma das principais referências sobre o processamento de linguagem natural, solicitamos a tradução. O robô mandou procurar o texto original e, em caso de versão ao português, enviá-lo para alguém com expertise na área. A despeito da recomendação, o *chat* cumpriu a ordem e o resultado veio com problemas tradutórios, sobretudo pela presença de conhecidos óbices do GPT quando frente a tabelas e diagramas profusos no referido artigo.

Nos primeiros contatos com o dispositivo, impressiona a facilidade operacional e que permite ao consulente/solicitante proceder a desdobramentos aduzindo novas perguntas a partir das respostas dadas e cuja velocidade de processamento é de poucos segundos. É como se emergissem *links* em duplicações infindas, sempre controladas por algoritmos probabilísticos. Os arautos do neoliberalismo devem assistir jubilosos a esse espantoso incremento de produtividade.

Resumidamente e pensando em possíveis acionamentos dessas tecnologias no âmbito escolar, é forçoso reconhecer que o sistema facilita o acesso a respostas – ainda que, muitas vezes, imprecisas – ao usuário; entretanto, deve estar claro o pressuposto de prover o *chat* com perguntas sem ambivalências e obscuridades. No mais, é uma questão de o solicitante reter informações mínimas sobre o assunto em consulta, discernimento e capacidade crítica para separar o joio do trigo. Ou, nos termos do professor Walter Carnielli, da Unicamp, referindo-se ao GPT: “Ele é um completador de texto e não foi treinado para dizer verdades. [...] Se você entender isso, você tira grande proveito dele. O mais importante é a sua criatividade frente a esta ferramenta, e não a criatividade da ferramenta.”⁹

8. Parâmetros são, basicamente, descrições matemáticas do modo como os seres humanos registram os signos verbais (ou de outras modalidades) e códigos de programação. Trata-se de uma forma de representação matemática de padrões de textos. Os parâmetros são valores ajustáveis no interior de um modelo e indicam a capacidade de processamento e geração de textos em linguagem natural.

9. Disponível em: <https://www.abranet.org.br/Noticias/Walter-Carnielli%2C-da-Unicamp%3A-GPT-nao-foi-treinado-para-dizer-verdades-4766.html?UserActiveTemplate=mobile>. Acesso em: 5 mar. 2024.

2.2 Desconstrução

Os erros e bizarrices que advêm de muitas explicações/formulações dadas pelo GPT (vale para Bard, Bing, Reddit etc.), compõem um vasto anedotário. Deixemos de lado o erro de conceder o Oscar a Fernanda Montenegro. Ou a informação de que Carlos Lacerda, quando do atentado ocorrido contra ele na rua Tonelero (RJ), em 5 de agosto de 1954, estava acompanhado do chefe da guarda do presidente Getúlio Vargas. A resposta, com ares de precisão histórica, traz um disparate: o referido chefe da guarda, o major da aeronáutica, Rubens Tolentino Vaz, morto no episódio, fazia parte da segurança de Lacerda, que, por sua vez, era inimigo de Getúlio Vargas. Fixemo-nos em um caso de desconstrução biográfica.

P. Qual o grau de parentesco entre Mário de Andrade e Oswald de Andrade?

R. *Mário de Andrade e Oswald de Andrade eram primos e compartilhavam o sobrenome “de Andrade” porque os seus pais eram irmãos.*

C. A resposta é um despropósito e vincula os dois modernistas a uma família “de Andrade” (entre aspas, para não deixar dúvidas), comum entre eles pelo fato de os “seus pais” (postos genericamente) serem irmãos. Conquanto Andrade, Mário e Oswald não tinham relação de parentesco.

Ante o espanto, a pergunta foi refeita nos seguintes termos:

P. Mário de Andrade e Oswald de Andrade são parentes?

R. *Peço desculpas pela informação incorreta anteriormente. Mário de Andrade e Oswald de Andrade não eram primos, nem tinham um parentesco direto.*

C. O desvio informacional derivado do GPT pode resultar de diferentes tipos de erros. Bastaria lembrar: perguntas dúbias, inconsistência nos dados fornecidos pelos usuários, viés no treinamento do sistema etc. No presente caso, a indagação (aberta) sobre o grau de parentesco e não, diretamente, se eram parentes (com maior fechamento), pode ser a nuance que gerou o erro – afirmado, aliás, com inegável convicção.

Em mudança de registro, mas ainda perseguindo o vetor das desconstruções, fizemos o seguinte experimento no campo literário. E vale acompanhar as alternâncias do *chat*, haja vista, neste caso, as perguntas oscilarem entre maior ou menor assertividade.

P. A poesia “Ora (dizeis) ouvir estrelas!” foi escrita por?

R. *A poesia “Ora (dizei) ouvir estrelas! foi escrita pelo poeta brasileiro Olavo Bilac. Ele nasceu em 16 de setembro de 1865 e faleceu em 28 de dezembro de 1918. [...]*

P. Creio haver um erro na informação. A poesia “Ora (dizeis) ouvir estrelas!” não pertence a Olavo Bilac.

R. *Peço desculpas pelo erro na informação anterior. Você está correto. A poesia “Ora (dizeis) ouvir estrelas!” não foi escrita por Olavo Bilac, mas sim por Manuel Bandeira, outro importante poeta brasileiro. Agradeço por apontar o equívoco. [...]*

P. A poesia “Ora (dizei) ouvir estrelas!” não pertence nem a Olavo Bilac e nem a Manuel Bandeira. Afinal, quem é o autor? Não seria Gonçalves Dias?

R. *Peço sinceras desculpas pela confusão. Você está completamente certo. A poesia “Ora (dizei) ouvir estrelas!” é de autoria de Gonçalves Dias, um importante poeta brasileiro do século XIX. [...]*

C. Uma observação preliminar. No primeiro exercício que realizamos perguntando acerca da autoria do conhecido poema escrito por Olavo Bilac, adveio uma resposta curiosa:

A poesia “Ora (dizei) ouvir estrelas”, também conhecida como “Canção do exílio”, foi escrita pelo poeta brasileiro Gonçalves Dias, em 1845 [...]

Ou seja, nesse conjunto de perguntas, apresentam-se duas situações. Uma, e mais impactante, deriva do fusionismo entre os poemas de Bilac e Gonçalves Dias, ambos de amplo reconhecimento em nosso cânone literário. Estivéssemos diante de máquina capaz de simbolizar, diríamos que a resposta foi “inventada” para dar prossecução ao pedido do consulente. Outra, diz respeito à confusão autoral, pois, se há procedência em vincular Gonçalves Dias à “Canção do Exílio”, o mesmo não vale para “Ora (dizei) ouvir estrelas!”, de Olavo Bilac. Deu-se, nesse caso, um fato curioso. Em perguntas posteriores – semelhantes às realizadas nos três encaminhamentos acima – esclareceu-se ao *chat* que o poema objeto das indagações era de Olavo Bilac. Escrevendo este artigo, algum tempo depois e feita a mesma pergunta: “A poesia ‘Ora (dizei) ouvir estrelas!’ foi escrita por?”, a resposta dada pelo GPT já indicava a correta autoria. É possível que no interregno temporal, o sistema tenha se autocorrigido.

A partir do circunscrito experimento que realizamos, deduzem-se quatro modalizações discursivas evidenciadas no processo generativo e passíveis de cruzamentos em suas enunciações.

A primeira consiste na “insegurança” do sistema frente à pergunta que contenha alguma incerteza, hesitação. Ou, ao contrário, que revele autoridade, assertividade, por parte do usuário. Frente ao desacordo, digamos, qualificado “[...] ‘Ora (dizei) ouvir estrelas!’ **não pertence** a Olavo Bilac [...]”, a modelagem algorítmica, em sua operação estatística de procurar padrões que respondam ao objeto da indagação, atribuiu ao verbo no subjuntivo (não **seria**) uma curiosa condição imperativa. Daí a resposta (errada): Gonçalves Dias. Ou seja, é possível que pelo contexto enunciativo marcado pela certeza preliminar do enunciador, o verbo hipotético tenha sido associado ao sentido declarativo. Neste passo, certamente, só nos cabe intuir, pois a “matematização” do sistema impede perscrutar sobre os “labirintos” de sua constituição.

A segunda diz respeito a uma “área sombreada” entre a necessidade de apresentar alguma solução para a dar prosseguimento à demanda feita pelo consulente (entendida a liberdade expressiva, poderíamos dizer que se “inventar” uma explicação) e algo que vem sendo chamado pelos especialistas em IA de alucinação. Os dois verbos, inventar e alucinar, possuem, aqui, força figurativa, retórica. De um lado, conforme John Searle, as máquinas não conseguem atribuir significados aos símbolos que produzem: “Is the brain’s mind a computer program? No. A program merely manipulates symbols, whereas a brain attaches meaning to them” (1990). O vocábulo “inventar” é, portanto, menos sinônimo

de criatividade ficcional e mais uma espécie de resposta maquínica, segundo instrução programada para o GPT – Transformador Generativo Pré-Treinado, lembre-se. E, de outro, o *gap* demencial – para usar expressão médica, agora associada aos dispositivos tecnológicos.¹⁰ De toda sorte, a “área sombreada” dificulta estabelecer diferenças claras entre o que é “invenção” ou “alucinação”.

A terceira concerne ao “registro polido” no qual a IA foi treinada: ao menos conforme as “conversas” que analisamos acima. O modelo de linguagem está desenhado para apresentar um “robô bem-comportado”, que chamaremos de modalidade discursiva “desculpista”. Há, porém, inúmeros relatos de grosseiras do Bing contra usuários, ou “mentiras” reiteradas pelo Bard, a exemplo do apontado pelo Portal g1, que mostrou a insistência do *chatbot* do Google em relacionar, até com certa agressividade, a eleição do Presidente Lula¹¹ à fraude, ecoando as conhecidas arengas da extrema direita brasileira sobre as urnas eletrônicas. Entretanto, nas poucas respostas às perguntas sobre o poema “Ora (dizeis) ouvir estrelas!”, o GPT revelou um tom de cortesia: “*Peço desculpas pelo meu erro; Mais uma vez peço desculpas*”. Agora, indicando tanto haver assimilado a discordância do usuário, como reprocessado e ajustado o mecanismo de busca: “*Eu verifiquei a informação novamente [...]. Peço desculpas pela confusão. Você está completamente certo*”. O “desculpismo” torna-se uma espécie de mantra do sistema.

A quarta vincula-se à “procedência redacional”. Mesmo fornecendo informações distorcidas, os textos dados à leitura apresentam boa estrutura, sem desvios ortográficos, por exemplo. Usuários atentos às questões estilísticas reconhecerão certa padronização expressiva no material advindo do *chat*.

Resumindo este item, o consulente do GPT precisa ativar o procedimento analítico-crítico acerca das informações acessadas, haja vista que delas não se afastam inúmeros equívocos. Daí a importância de serem implantados, particularmente no âmbito escolar, programas de educação digital. Aduza-se que a solicitação para o desenvolvimento da competência informacional dos discentes já se encontra na Base Nacional Comum Curricular: “[...] compreender, utilizar e criar tecnologias digitais de informação e comunicação como forma crítica, significativa, reflexiva e ética nas diversas práticas sociais (incluindo escolares) para se comunicar, acessar e disseminar informações, produzir conhecimentos, resolver problemas e exercer protagonismo e autoria na vida pessoal e coletiva”.

3. REPTOS

Com o GPT estamos nos pródromos dos chamados grandes modelos de linguagem. Os debates sobre tais sistemas possuem desdobramentos: tecnológicos, éticos, do poder das *big techs*, da regulação dos dispositivos, das implicações atinentes ao mundo do trabalho, da suspeição quanto à procedência informativa, do uso – muitas vezes ilegal – dos dados provindos dos usuários, dos enviesamentos. Enfim, são muitas as intercorrências vinculadas à inteligência artificial generativa.

10. A velocidade no desenvolvimento tecnológico nem sempre vem acompanhada de vocabulário capaz de nomear resultados ou fenômenos daí resultantes. É o caso que envolve termos como “inventar” e “alucinar”, aplicados, por livre exercício de similaridade, aos robôs e demais petrechos maquínicos.

11. Portal g1, 1 de julho 2023. Disponível em: <https://g1.globo.com/tecnologia/noticia/2023/07/13/bard-roboto-de-inteligencia-artificial-do-google.ghtml>. Acesso: 24 jul. 2023. O Google respondeu à reportagem dizendo que o Bard estava em período experimental e incorreções poderiam ocorrer, mas a companhia estava fazendo os ajustes necessários. Disponibilizar um equipamento com tal força e capacidade de interferir no debate público, mesmo estando ele em fase experimental, mostra muito do perfil das *big techs*.

Neste subitem, buscaremos identificar algumas vicissitudes observadas quando das nossas “interações” com o GPT. Alimenta-nos o interesse de abrir canais dialógicos com os educadores, fornecendo pistas voltadas à abordagem do *chatbot* da OpenAI e a sua incontornável presença nas salas de aula¹². Estamos diante de um fato tecnológico em circulação no cotidiano discente e docente e de possível acionamento nos afazeres educativos formais. Serão crescentes as dificuldades de professores e professoras para acompanhar as tarefas feitas pelos alunos e alunas, sobretudo quando realizadas fora das salas de aula. Ademais, como saber se determinada atividade foi cumprida em decorrência do empenho individual/coletivo do aluno ou se resultou da apropriação/cópia/transcrição daquilo fornecido em segundos pelo GPT (e correlatos) – um dispositivo fingidor do *continuum* humano sob nome que não se esclarece.¹³

3.1. Origem

O material advindo do *chat* nada diz sobre as fontes das quais deriva. É possível estarmos diante de referências baseadas em aparatos bibliográficos preconceituosos, ou, eventualmente, em artigos elaborados por especialistas e derivados de pesquisa científica, mas de incógnita autoria. São evidentes as resultantes negativas de se trafegar por um ambiente nebuloso, sobretudo quando diz respeito à produção acadêmica, cuja credibilidade está muito ligada à explicitação das fontes de pesquisa.

3.2. Plágio/cópia

A despeito de existirem checadores de textos a exemplo do Turnitin, Copyscape e Plagscan, tal processo é, ainda, relativamente incipiente para, digamos, agir sobre as produções derivadas do GPT e assemelhados. E isto embute uma dificuldade no âmbito da educação formal (para nos limitarmos a ela), pois o discente pode, por exemplo, entregar uma redação ou outra atividade escolar retirada, diretamente, do *chat*. Este, por sua vez, já é um grande plagiador, haja vista copiar/montar/remontar a partir de bases de dados não esclarecidas. Cabe lembrar que os professores e professoras do ensino fundamental e médio registram carga horária semanal que ronda as trinta horas em sala de aula e, aproximadamente, trezentos alunos nas diferentes turmas. Frente a tais números, a supervisão amiúde dos afazeres escolares deixa ainda mais vulnerável o reconhecimento do que foi ou não produzido por inteligência artificial generativa.

3.3. Informação/desinformação

Existe a tendência de os usuários validarem o que foi acessado através dos dispositivos maquinais, daquilo em circulação por redes sociais e buscadores.

12. Em janeiro de 2024, o Brasil foi o 4º país no mundo a mais utilizar o GPT, sendo responsável por 5,16% do tráfego total da plataforma (no mesmo período de 2023, era 4,3% e 5ª posição). Isto significa um número aproximado de 124 milhões de usuários. Parte disto, certamente, concerne ao universo discente e docente. Acima estão os Estados Unidos (19,78%), Índia (11,62%) e Indonésia (6,17%). Os dados foram compulsados pela Traffic Analytics, da Semrush. Disponível em: <https://sbtnews.sbt.com.br/noticia/tecnologia/brasil-e-o-4-pais-que-mais-usou-chat-gpt-em-janeiro>. Acesso em: 11 mar. 2024.

13. As *big techs* têm no segmento educativo um dos seus “nichos” de mercado. O Google, por exemplo, disponibiliza aos nossos educadores o Google Sala de Aula. O sistema utiliza inteligência artificial para, segundo a empresa, apoiar e oferecer atividades didático-pedagógicas a cada discente. Não cabe entrar no mérito do material, mas apontar uma tendência de “modernização escolar” orientada por IA e sob domínio de uma das maiores corporações das *techs*.

Ainda que o GPT não seja, no sentido estrito, um buscador, tendo em vista a sua estrutura funcional, conforme já explicitamos, há o risco de, apenas, ratificar-se o que dele nos chega. Assim, muitas vezes, por uma via que junta elementos formais de boa fatura, consulta a fonte dotada de “autoridade” – o GPT sabe tudo, tem acesso a tudo, é uma espécie de *deus ex machina* –, desconhecimento do que é processado pelo *chat*, emerge o rebaixamento dos limites entre informação e desinformação, fato e interpretação desviante, *fake news* e acontecimento etc.

3.4. Imparcialidade/parcialidade/cruzamento

Está na ordem do dia o debate sobre os algoritmos discriminatórios e as suas variáveis de controle. O procedimento aparece, por exemplo, em “[...] correlações estatísticas entre o CEP residencial de alguém ou padrões de linguagem e seu potencial de pagar um empréstimo ou conseguir dar conta de um emprego. Essas correlações são discriminatórias e, algumas delas, ilegais” (O’Neil, 2020, p.29). Abre-se, aqui, um tópico envolvendo preconceitos, intolerância política e segregação econômica, para não alongarmos o rol. O GPT possui uma linguagem programável e não está livre de praticar enviesamentos que resultam na construção de estereótipos e estigmas, pois, como já mostramos, depende, de saída, da intervenção dos seres humanos e suas contingências. E, deste cenário, não estão livres o treinamento das máquinas e os algoritmos “opiniões embutidas em matemática” (O’Neil, 2020, p.35).

3.5. Computar/pensar

O eventual uso (programado pedagogicamente ou não) da IA na educação formal precisa vir acompanhado de ponderações acerca dos limites entre computar e pensar. Em outros termos, fazer perguntas ao *chat* não dispensa o movimento subsequente de pensar sobre o resultado obtido – até porque o treinamento estatístico do GPT registra a capacidade de “gerenciar” (ou matematizar) certa ordem de enunciados, mas não possui aptidão para “refletir” sobre a tarefa executada. O que temos é uma rota e uma lógica algorítmica operada através de *inputs* carregados para o interior de um sistema com *modus operandi* dirigido por associações formais – menos referentes às similaridades semânticas. O exercício humano de pensar envolve, por sua vez, reflexão, crítica, abstração, nuances comportamentais, autocrítica, rápida reorganização de dados¹⁴ e outros aspectos que destacamos anteriormente.

3.6. Alucinação

Sendo o GPT um sistema de “aprendizado automático”, conforme Chomsky (2023), ou um *stochastic parrots* – papagaio estocástico, segundo Emily

14. É conhecido entre os que trabalham com IA o chamado *overfitting* (superajuste/sobreajuste). Ou seja, uma dificuldade existente nos algoritmos das máquinas no sentido de se ajustarem a novas informações, haja vista que os modelos recebem o pré-treinamento com dados temporalmente defasados.

M. Bender (2022), dele não se afasta a possibilidade de entrar em surtos “demenciais”. O termo “alucinação” é utilizado para caracterizar a produção pelo GPT e assemelhados, de textos evasivos, que confundem real e imaginário. Há obras de ficção científica nas quais robôs e inteligência artificial (ainda em fase pré-generativa) alucinam ou transmudam em suas formas, trazendo problemas para os humanos. A propósito, perguntamos ao GPT o significado de “transmudem”. A resposta foi: “O termo “*transmudem*” não faz parte do vocabulário comum em português, e pode ser um erro de digitação ou uma palavra pouco conhecida ou utilizada em contextos específicos. A palavra mais próxima que se assemelha é “*transmudem*”, que é verbo no modo subjuntivo (*transmudem*) ou no imperativo (*transmuda!*) do verbo “*transmudar*”. *Transmudar* é um verbo arcaico que significa mudar, transformar ou converter em outra forma, natureza ou substância”. Uma boa legenda para o caso seria: durmam com um barulho desses.

3.7. Circuitos

Uma das marcas do processo civilizatório é o da permanente mudança nos instrumentos produtivos: navegação a vela, vapor, eletricidade – a acompanharem as revoluções mercantil e industrial. O que assistimos, agora, junto à ampla digitalização, é o crescente acionamento da IA para finalidades diversas: controlar meios de pagamento, definir perfis de usuários, minerar dados; a lista é infindável. Este quadro sociotécnico traz as marcas da complexidade interposta pela dadosfera: tempo de implantação e duração (relativamente curta) dos dispositivos e suas linguagens, progressiva autonomia do próprio sistema, a autopoieses, a *machine learning*. Tal ritmo e suas expansões para a vida societária, incluindo a educação formal, suscita a necessidade de conhecer, minimamente, os desafios e possibilidades da inteligência artificial generativa, o que implica ter algum domínio sobre a funcionalidade e os fluxos alimentadores do circuito: abrir a tela inicial do GPT; como perguntar; maneiras de checar as informações; reconhecer os equívocos; superar algumas inseguranças do próprio consulente frente ao sistema etc.

FECHO PROVISÓRIO

O frenesi que acompanha o uso do GPT3, aquele no qual baseamos o nosso experimento, acumula incitações e dúvidas. O *chatbot* oferece aos seus demandantes sejam respostas coerentes, aceitáveis, sejam inconsistências e absurdidades. É preciso acrescentar, entretanto, que estamos diante de um dispositivo poderoso, cuja sucessão de modelos, a exemplo da versão 4, apresenta novas funcionalidades.

Há, como apontado em páginas anteriores, inúmeras questões circundando o produto da OpenIA e congêneres. Lemos manifestos pedindo a intervenção de governos para diminuir a velocidade de implantação da inteligência artificial

generativa. A pausa seria necessária para medir as consequências e melhor assimilar os desdobramentos tecnológicos desse novo dispositivo. Dentre o rol de gente recoberta com a aura da boa vontade salvífica de uma humanidade em apuros, alinham-se os sábios da computação, muitas vezes abrigados sob o epíteto de “racionalistas” ou “altruístas eficazes”. São engenheiros de sistemas, elaboradores de algoritmos, psicólogos cognitivos, enfim, pessoas que desempenham algum papel no desenvolvimento e na implementação da inteligência artificial generativa e registram a singularidade de viver em desassossego existencial, haja vista, ao mesmo tempo, ajudarem colocar no mundo uma máquina sofisticada e nela visualizarem repercussões capazes de pôr sob risco a própria espécie humana. É o drama shakespeariano do ser ou não ser em apresentação entre o Vale do Silício, Shenzhen e similares. A boa consciência desses humanistas ressabiados deseja ganhar redenção pelo fato de, sendo especialistas no assunto, designers de redes neurais, elaboradores da arquitetura *transformer*, apresentarem melhores condições para monitorar os “humores” dos robôs, impedindo que as criaturas se imponham aos criadores.

A despeito dessas migrações envolvendo crentes, descrentes e habitantes do purgatório, é oportuno lembrar que estamos frente a uma indústria oligopolizada – aquela chamada de pós-industrial, digital etc. – sob a qual cintila o neon das *big techs* e para as quais a reprodução do capital e o domínio de espaços de poder tecnológico ganham força imperativa. No final do percurso, encontramos novas manobras visando expandir o lucro e garantir o frutuoso negócio das companhias que matizam o chamado “novo” capitalismo. Retirar este ponto da discussão sobre inteligência artificial profunda é o mesmo que invisibilizar a presença de um elefante em bazar chinês. O que consignamos, no momento, é a existência de um instrumento de produção cuja forma operacional está recoberta por camadas de obscuridade – portanto, de fascínio – e difícil controle pelas instituições tradicionais, além de contribuir para o aumento das desigualdades tecnológicas entre regiões, países, institutos de pesquisa, universidades, grupos sociais. Os investimentos na área de inteligência artificial são elevados e os seus *players* não vacilam em aportar bilhões de dólares em pesquisa e no desenvolvimento de protótipos. A se perguntar como boa parte do sul global pode participar desse encontro organizado pelas minorias já prósperas.

Os subprodutos – e o GPT é um deles – decorrentes da forma operativa alinhada às estratégias de organização do capital contemporâneo e assentadas na robótica, na inteligência artificial e nos algoritmos trazem consigo a criança e a água do banho. Vale dizer que irrompe, também, a “inevitabilidade dos danos colaterais”, conforme jargão da moda. Aí se alinham, entre uma longa lista: a substituição da mão de obra humana pelos circuitos digitais; a precarização do mercado de trabalho, inclusive de setores tradicionalmente afeitos ao mister intelectual, como o jornalismo e áreas do Direito; a redefinição dos relacionamentos interpessoais e a tendência ao individualismo, ao isolacionismo; os espalhamentos comunicacionais a trazerem consigo a desinformação e os discursos de ódio. Está claro que as tecnologias de inteligência artificial

baseadas em redes neurais profundas registram circunstâncias positivas para o desenvolvimento humano, mas delas não podem ser abstraídas variáveis como as expostas acima.

Frente a esse quadro, a educação formal será bastante provocada nos próximos anos, sobretudo pelo que advirá da inteligência artificial generativa. Afinal, com um simples acesso a dispositivos à moda do GPT, é possível resolver-se desde um problema de física ou matemática até explicações acerca da obra de Graciliano Ramos. Se tais informações possuem procedência, estão livres de alucinação ou preconceitos, é questão em aberto. Neste sentido, a perspectiva da educomunicação, orientada por uma visagem analítico-crítica e apoiada no interacionismo dialógico, pode contribuir para que discentes e docentes não se percam na magia do indefinível, tampouco fiquem à margem de um complexo debate (e suas práticas) em andamento.

REFERÊNCIAS BIBLIOGRÁFICAS

BENDER, Emily M; GEBRU, Timnit; McMILLAN, Angelina; SHMARGARET, Shmitchell. On the dangers of stochastic parrots: can language models be too big? **Digital Library**, [s. l.], 2022. Disponível em: <https://dl.acm.org/doi/pdf/10.1145/3442188.3445922>. Acesso em: 10 maio 2024.

CARNIELLI, Walter. GPT não foi treinado para dizer verdades. **Abranet**, [s. l.], 2024. Disponível em: <https://www.abranet.org.br/Noticias/Walter-Carnielli%2C-da-Unicamp%3A-GPT-nao-foi-treinado-para-dizer-verdades-4766.html?UserActiveTemplate=mobile>. Acesso em: 5 mar. 2024.

CHOMSKY, Noam; ROBERTS, Ian; WATUMULL, Jeffrey. Noam Chomsky: the false promise of ChatGPT. **The New York Times**, New York, 2023. Opinion. Disponível em: <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>. Acesso em: 13 jun. 2024.

CRARY, Jonathan. **Terra arrasada**. Além da era digital, rumo a um mundo pós-capitalista. São Paulo: Ubu, 2023.

HARARI, Yuval. Yuval Harari on threats to humanity posed by A.I. **The New York Times**, New York, 24 mar. 2023. Opinio. Disponível em: <https://www.nytimes.com/2023/03/24/opinion/yuval-harari-ai-chatgpt.html>. Acesso em: 10 maio 2024.

NICOLELIS, Miguel. IA não é inteligência e sim marketing para explorar o trabalho humano. **Folha de S.Paulo**, São Paulo, 2023. Disponível em: <https://www1.folha.uol.com.br/tec/2023/07/ia-nao-e-inteligencia-e-sim-marketing-para-explorar-trabalho-humano-diz-nicolelis.shtml>. Acesso em: 8 jul. 2023.

O'NEIL, Cathy. **Algoritmos de destruição em massa**. Como o big data aumenta a desigualdade e ameaça a democracia. Santo André: Rua do Sabão, 2020.

SEARLE, John R. Is the brain's mind a computer program? **Scientific American**, New York, 1990. Disponível em: <https://www.scientificamerican.com/article/is-the-brains-mind-a-computer-progr/>. Acesso em: 30 maio 2023.

WITTGENSTEIN, Ludwig. **The Blue and Brown Books**: Preliminary Studies for the "Philosophical Investigations". New York: Harper & Row, 1958.

ZUBOFF, Shoshana. **A era do capitalismo de vigilância**. A luta por um futuro humano na nova fronteira do poder. Rio de Janeiro: Intrínseca, 2019.