

GPT-3: um cérebro eletrônico no jardim do fascínio¹

Adilson CITELLI²
Universidade de São Paulo, São Paulo, USP

“Que cérebro eletrônico nenhum me dá socorro” (Gilberto Gil)

RESUMO

Neste artigo busca-se refletir sobre as intercorrências do ChatGPT na educação formal. A inteligência artificial generativa vem sendo considerada por alguns como se fosse um “papagaio estocástico” e, para outros, a antessala do apocalipse. Nesse caso, o desenvolvimento da IGA (Artificial General Intelligence), poderia permitir a autonomia dos robôs nos processos de autoprogramação – vaticínio feito por Geoffrey Hinton, um dos principais cientistas da computação voltado à pesquisa com as redes neurais profundas. Estamos, portanto, diante de um complexo e poderoso dispositivo tecnocomunicacional que alcança os mais diversos segmentos da vida social, dentre eles, a escola. Já se vem detectando a realização de atividades (por parte dos discentes) levadas a termo a partir das cadeias algorítmicas pré-programadas pelo *Chat*. Cabe, enfim, pensar nas implicações desta nova realidade a partir do campo que estuda os nexos comunicação-educação.

PALABRAS-CHAVE: ChatGPT; Inteligência artificial; Comunicação; Educação

Anúnciação

Em novembro de 2022, a *startup* californiana OpenAI, uma espécie de anjo Gabriel redivivo, revelou para a humanidade ser a mãe ou o pai, de um avançado *software* de inteligência artificial capaz de produzir textos – e conversas com seres humanos – dotados de coesão e coerência, bastando, para tanto, que perguntas e comandos procedentes fossem feitos pelos usuários. O resto da história é conhecida e *in progress*: dois meses após o seu lançamento mais de cem milhões de pessoas acessavam a novidade, o ChatGPT-3, divertindo-se, indignando-se, (des)informando-se, fornecendo gratuita e dadivosamente trilhões de dados que, por sua vez, foram e estão sendo dirigidos para aperfeiçoar o funcionamento do dispositivo. Nos dias correntes – este texto foi escrito em meados de 2023 – acumulam-se cursos, matérias em jornais e revistas, programas de televisão, palestras, seminários, vídeos explicativos, numa arregimentação de ânimos e perplexidades que não param de crescer e prognosticar seja a revolução em curso, sejam os dissabores nela incluídos³. E uma questão resta parada no ar: vindo a funcionar

¹. Trabalho apresentado no GP Comunicação e Educação do XLVI Congresso Brasileiro de Ciências da Comunicação. INTERCOM. Pontifícia Universidade Católica. Belo Horizonte – MG. 2023

². Professor titular sênior do Departamento de Comunicações e Artes. Escola de Comunicações e Artes. ECA-USP. Pesquisador IB do CNPq. Coordenador do GP. MECOM (Mediações Educomunicativas). ECA-USP-CNPq. citelli@uol.com.br

³. Ver artigo The open letter to stop “dangerous” al race a huge mess”, que comenta a carta assinada por milhares de signatários, entre outros, Elon Musk, da Tesla, Steve Wozniak cofundador da Apple, Yuval

integralmente a IAG (Inteligência Artificial Geral) ocupará boa parte das atividades intelectuais exercidas pelos seres humanos? E outra: naturalizar o complexo internético – do qual faz parte a IA – não significa ceder a um poder global de dominação sob a égide das sociocidas *big techs* ou êmulos sequentes? Agora, nos termos de Jonathan Crary (2023, p. 86): “Parte da nossa crise atual está na aceitação indiferente da agora banalizada ideia de que nosso futuro está sendo inventado por um pequeno número de corporações poderosas. Na prática, decisões sobre quais linhas de produtos manterão lucros e crescimento transformaram essas companhias nos futurologistas oficiais de nossa época, nos órgãos reguladores de nossas expectativas”.

O ChatGPT-3, pelo seu relativo pioneirismo (a versão 4 já está disponível para assinantes GPT Plus, à razão de US\$ 20,00 mensais, trazendo melhorias e inovações, como de interpretar imagens e capacidade para gerar textos com até 25 mil palavras – contra 2 mil do modelo anterior), tornou-se a mais recente sensação desta linhagem de avanços na inteligência artificial generativa. A despeito das luzes lançadas sobre a criação da OpenAI, outros produtos de família próxima também já se encontram no cardápio oferecido pelo mercado de inteligência artificial: Bing, da Microsoft⁴, Bard, da Alfabet/Google, Ernie Bot – em mandarim, Wenxin Yiyan, – da chinesa Baidu. E seguem nomes e siglas voltadas, sobretudo, aos modelos geradores de imagens como o DALL-E 2, Stable, Diffusion, Midjourney.

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 GhatGPT-4. <https://www.vice.com/en/article/qjvppm/the-open-letter-to-stop-dangerous-ai-race-is-a-huge-mess>. Acesso: 10/06/2023. A íntegra da carta em tela, sob o título “Pause giant AI experiments: An open letter”, pode ser lida em: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>. Acesso: 10/06/2023. Yuval Harari, em artigo ao New York Times publicado em 24/03/2023, sob o título “Yuval Harari on threats to humanity posed by A.I”, não foge ao tom geral acerca do imperativo de colocar freios no espraiamento das novas gerações de inteligência artificial, pois se estamos ainda em uma sociedade caudatária de instituições do século XIX é preciso tempo para ajustar o mundo à nova realidade representada pelos *chatbots*. E assevera a urgência de “(...) dominar a IA antes que ela nos domine” (tradução livre). <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”. Em outras palavras, enquanto os manifestos tratam a IA como entidade autônoma, as corporações (pais e mães criadores e criadoras dos cérebros eletrônicos) podem continuar ganhando dinheiro, aumentando o seu poder e comprometendo a democracia. Nesta direção, ver artigos de Sayash Kappor e Arvind Narayanan, publicado em AI Snake Oil: <https://aisnakeoil.substack.com/p/a-misleading-open-letter-about-sci> e Emily M. Bender, em Search Medium: <https://medium.com/@emilymenonbender/policy-makers-please-dont-fall-for-the-distractions-of-ai-hype-e03fa80ddb1>. Enfim, os materiais que estão circulando sobre IA se avolumam, ao mesmo passo em que apresentam ângulos diversos.

⁴. É divertido ler o relato e as experiências feitas por Olavo Amaral (2023) com o Bing Chat. A sequência de desvios, ameaças, alucinações do sistema, levaram a Microsoft a rever procedimentos e suspender, momentaneamente, operações com o modelo. No dia 15 de junho de 2023, um juiz da 7ª vara civil de Bauru, no Estado de São Paulo, determinou que a Microsoft excluísse do buscador Bing a associação de um respeitado médico da cidade com o assédio sexual envolvendo 100 acusações. Ocorre que o referido profissional, por delegação do CREMESP, era um dos responsáveis pela sindicância da centena de casos afeitos a importunação sexual de pacientes por médicos paulistas. Em entrevista dada a um órgão da imprensa, o sindicante referiu-se ao alto número de averiguações em andamento. O Bing registrou que as denúncias diziam respeito ao próprio médico que apurava desvios de conduta dos seus pares, ou seja, para o robô, o investigador ganhou o status de investigado. <https://www.uol.com.br/tilt/noticias/redacao/2023/06/22/medico-acusado-assedio-erro-ia-bing-microsoft.htm?cmpid=copiaecola>. Acesso: 23/06/2023.

Cabe aduzir que a celeridade na diversificação dos padrões de modelos de linguagem parece justificar o pasmo expresso nos manifestos, cartas e editoriais, conforme os acima referidos, 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 limitada e nem sempre a caminho do sucesso, ao menos segundo perspectivas apocalípticas que acompanham vários cientistas da computação, filósofos e cognitivistas, a enxergarem no ensandecimento do computador senciente Hal 9000, do filme de 1968, dirigido por Stanley Kubrick, *2001, uma odisseia no espaço*, o exemplo que nos aguarda em anos vindouros

É preciso ponderar que tais mecanismos generativos acionados por algoritmos⁵, operações lógicas, processamento de imagens e linguagem natural, não revelam as suas estratégias de elaboração, implantação e distribuição, haja vista serem recobertas por segredos insondáveis, intangíveis, de conhecimento restrito aos especialistas que os formulam. Neste território opaco, o fetiche circundante da inteligência artificial – se é que a máquina possui alguma⁶ – transita por processos algorítmicos, padrões estatísticos, probabilísticos, preditivos, a pedirem sempre novos dados que irão aumentar o valioso estoque das já fornidas bases que alimentam o centro do poder empresarial das *big techs*.

O Chat de que tratamos foi desenvolvido pela OpenAI, companhia voltada à pesquisa e produção de um modelo inovador de linguagem artificial, chamada IA GPT-3. Entre outros financiadores da novidade encontrava-se a Microsoft, que aportou, em 2019, 1 bilhão de dólares na *startup*, e, agora, mais 10 bilhões, segundo o jornal *Folha de S. Paulo*, de 10 de janeiro de 2023, citando como fonte o site de notícias norte-americano *Semafor*. O acordo prevê, além de retornos financeiros à empresa de Bill Gates, a integração do GPT ao Bing, em clara luta de poder no interior das *Big Techs*, sendo o adversário mais evidente o Google/Alphabet e o seu sistema Bard – que entrou em funcionamento oficial no Brasil em 13 de julho de 2023.

A sigla GPT significa *Generative Pré-trained Transformer* (Transformador Pré-treinado Generativo), sendo um software, ou modelo de linguagem, que utiliza a arquitetura *Transformer*, tendo recebido pré-treinamento para gerar (Generativo) textos em linguagem natural (compreensível pelos seres humanos) a partir de perguntas/formulações efetivadas por consulentes. Modelo de linguagem é um sistema (ou, uma rede neural profunda chamada *Transformer*) de aprendizagem de máquina (*machine learning*) que analisa imensa quantidade de dados, distinguindo bilhões de parâmetros⁷ sobre a maneira de os humanos correlacionarem palavras, signos, símbolos, números, e, a partir daí, elaborando novos textos em linguagem natural. A arquitetura de linguagem GPT deriva da análise de 570 gigabytes de dados na internet, algo como 300 bilhões de palavras, resultando na identificação de 175 bilhões de parâmetros – os

⁵ . Para uma boa introdução ao assunto envolvendo preconceitos e vieses contidos nestes modelos matemáticos, ver: Cathy O'Neil (2000).

⁶ . O neurocientista Miguel Nicolelis reitera que a IA não é inteligente sequer artificial. Certamente é um comparativo com a inteligência humana. Essa é analógica, a operar com valores pouco ou nada programáveis, como fidelidade, confiança, amorosidade, empatia, aquela, é digital e dependente da construção de algoritmos, programação, treinamento, robustas bases de dados etc. Tratar-se-ia, pois, de apropriação indevida de conceito, cuja naturalização de uso apenas refletiria certas estratégias de marketing que circundam a IA.

⁷ . Parâmetros são, basicamente, descrições matemáticas do modo como os seres humanos registram os signos verbais (e, eventualmente, de outras modalidades) e códigos de programação. Trata-se, portanto, de uma forma de representação matemática de padrões de textos. Assim, os parâmetros são valores ajustáveis no interior de um certo modelo e indicam a maior ou menor capacidade de processamento e geração de textos em linguagem natural.

modelos iniciais possuíam em torno de 1.5 bilhões. Em síntese: o ChatGPT é uma inteligência artificial, um modelo de linguagem, baseado em *machine learning*, que processa/minera uma grande quantidade de dados/informações, produzindo textos em linguagem natural, portanto, passíveis de compreensão pelos seres humanos.⁸

O GPT opera no padrão LLMs (*Large Language Models*/Grandes Modelos de Linguagem) cuja estrutura operacional consiste em encaminhar – respeitados os contextos enunciativos – probabilidades na elaboração, por exemplo, de frases ou conjunto delas, textos mais longos, a partir de métodos estatísticos que calculam frequências de unidades expressivas e direcionam as escolhas consideradas corretas. Os cálculos de possibilidades efetivam-se a partir do manancial extraído da web, artigos de repositórios digitais, livros, Wikipedia, etc., que permitirão formar os bilhões de parâmetros orientadores do sistema e seu pré-treinamento, assim como as decorrentes transformações/representações em linguagem natural – aquela, como posto acima, legível pelos seres humanos.

E, das interações subsequentes entre sujeito (perguntador) e máquina (respondente), irrompem enunciados surpreendentes ou encantados – pois, o desconhecimento por parte do consulente acerca da magia operada no ventre da máquina é quase total – acerca da fertilidade do solo na Serra da Mantiqueira, dos sentidos do substantivo estocástico, da resposta envolvendo uma equação matemática, da biografia de Aristóteles solicitada pelo professor de Filosofia. No mais, o céu não tem limites, conquanto parte das respostas oferecidas pelo GPT possa vir eivada de erros, invencionices, divagações, ou mesmo despida de valor.

O modelo que permitirá tais transformações (da linguagem artificial para a linguagem natural) está, como referido noutro passo, organizado em redes neurais maquinais profundas – algo que não deve ser associado, imediatamente, às redes neurais humanas. Vale dizer, o processo é posto em movimento 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 enunciativo. Vale dizer, para os efeitos dos modelos de linguagem, as palavras de entrada recebem pesos/números (tais conversões são chamadas de *embeddings* – ou, em tradução aproximada, incorporações) que possibilitam sejam elaboradas representações textuais de dados em certo espaço vetorial composto por números reais que processam a linguagem natural em algoritmos de aprendizado de máquinas. Os *embeddings*, enquanto mecanismos representacionais, permitem, por exemplo, que longos enunciados sejam condensados, compactados em espaços menores – o que é de grande importância física, de consumo de energia e econômica nos grandes computadores e bases de dados. Tal formatação de linguagem apoiada na arquitetura *transformer* é utilizada para múltiplas finalidades, como a tradução automática, a geração de textos, a elaboração de resumos de obras, a escritura de poemas, a (des)informação etc.

Sésamo

⁸. Em linhas gerais e sem entrar nas particularidades que envolvem o treinamento de máquinas de inteligência artificial, acentue-se que no início do processo estão seres humanos. São os treinadores/roteiristas que possuem formação em ciências de dados, informática, estatística, matemática etc. Ou seja, sem tais profissionais, a chamada “rotulagem de dados”, quando as informações são expostas à IA, o GPT não funcionaria. Na sequência desse processo, os sistemas vão progressivamente se autonomizando, cruzando dados e informações, “ensinados”, permitindo que respostas crescentemente complexas sejam dadas ao usuário.

Abaixo serão realizados alguns exercícios “dialógicos” com o *chat* no intuito de verificar como ele se apresenta ou “se representa” – se é que cabe tal conceito frente ao insólito da situação. Para facilitar a leitura haverá uma intercalação de perguntas dirigidas ao GPT-3, e as respostas daí decorrentes – destacadas em itálico. Em seguida aduziremos comentários a cada uma das sequências.

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 inúmeros artigos de especialistas em inteligência artificial. Ou seja, os sistemas de linguagem como o GPT não seguem, exatamente, as estruturas tradicionais de programação dos computadores, 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. Conceitos como *encoder-decoder* ou *finetuned*, certamente, registram algum grau de dificuldade para serem compreendidos por parte dos consultantes, entretanto, estão eles incluídos na cadeia gerativa que responderá à pergunta envolvendo a sigla GPT. Ademais, do processo “nutritivo”, torna-se quase inevitável fazer parte expressões em inglês – trata-se da língua de domínio, idioma no qual tendem a ser abonados boa parte dos artigos no campo da informática. Nada impede, contudo, que o usuário prossiga em sua pesquisa buscando esclarecer os sentidos de tais conceitos/palavras.

P. O que significa o conceito de *encoder/decoder*

R. *Em resumo, o encoder e o decoder são conceitos-chave em muitas aplicações de processamento de informações, permitindo a compactação e a posterior reconstrução ou geração de dados a partir de representações densas e informativas.*

C. A resposta é bastante longa, motivo pelo qual deixamos, como registro, apenas o resumo feito pelo próprio GPT. Observe-se que um leitor mais atento, mesmo que não fosse iniciado naqueles conceitos, teria acesso a uma explicação correta, feita com clareza expressiva, tradução ao português de ambos os termos; enfim, não haveria maior dificuldade para entender do que se trata.

P. E o que significa a palavra *finetuned*?

R. *"Finetuned" é um termo utilizado no contexto de modelos de aprendizado de máquina que foram pré-treinados em uma grande quantidade de dados e depois ajustados (ou "sintonizados") para uma tarefa específica. O processo de finetuning envolve treinar o modelo em um conjunto de dados menor e mais específico para que ele possa aprender a realizar a tarefa desejada com maior precisão”.*

No caso do GPT, ele é pré-treinado em uma grande quantidade de dados de texto para aprender a modelar a estrutura da linguagem natural. Depois disso, o modelo pode ser

finetuned em uma tarefa específica, como classificação de sentimentos ou geração de texto, para que ele possa aprender a realizar essa tarefa com maior precisão. O finetuning é um processo importante para obter um modelo de linguagem que seja adequado para uma determinada tarefa”.

C. A explicação é satisfatória e esclarece os sentidos ensejados pelo termo que restava obscuro no contexto geral da resposta primeira. O fato de a explicação ser pertinente não indica que o usuário apreenda a complexidade do fenômeno em descrição, até mesmo pela complexidade do assunto. Desta sorte, não é incomum em respostas dadas pelo GPT a compreensão, digamos, do enunciado sem que se chegue, propriamente, ao entendimento da enunciação. Ademais, uma tradução mais explícita dos dois termos permitiria melhor apreensão da ideia do “ajuste”, algo “ajustado”, “adaptado” a determinada finalidade, a “uma tarefa específica”.

P. E o citado artigo *"Attention Is All You Need"*. É possível traduzi-lo para o português?

R. *Sim, é possível traduzir o artigo "Attention Is All You Need" para o português. O título do artigo em português seria "A Atenção é Tudo que Você Precisa".*

“Este artigo é um trabalho científico escrito em inglês por Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, e Illia Polosukhin, e foi publicado pela primeira vez em 2017 na Conferência de Processamento de Linguagem Natural (NLP) da Associação de Computação da ACM.

O artigo descreve um modelo de aprendizado de máquina denominado "Transformer" que utiliza somente mecanismos de atenção para realizar tarefas de processamento de linguagem natural. O modelo Transformer é uma alternativa mais eficiente em termos de tempo e espaço para os modelos de aprendizado de máquina tradicionais que utilizam redes neurais convolucionais e recorrente.

O artigo é bastante técnico e detalhado, e requer conhecimento prévio em processamento de linguagem natural, aprendizado de máquina e matemática. Se você estiver interessado em ler o artigo completo em português, sugiro que procure uma versão traduzida por um especialista na área”.

C. A primeira boa notícia é que o artigo existe, igualmente os autores citados. Trata-se de um trabalho para iniciados montado sobre números, diagramas, gráficos, equações matemáticas e voltado a expor a arquitetura *transformer*. O nosso interesse, neste caso, foi ao mesmo tempo checar a informação dada pelo GPT e saber da possibilidade de tradução do material. Ainda que o *chat* recomende procurar o artigo no original e sugira, em caso de versão ao português, o envio para alguém com expertise na área, o consulente pode, por exemplo, cortar e colar o texto no GPT-3 e solicitar a translação pelo robô: a ordem é, imediatamente, cumprida. O que não deixa de trazer consigo um indicador acerca do modo como o sistema funciona, à base perguntas/ordens específicas. Feito o experimento, o resultado foi insatisfatório, sobretudo pela presença de inúmeros e conhecidos problemas do GPT quando frente a tabelas e diagramas. Em síntese, faz bem o *chat* em sugerir um tradutor especializado, conquanto, no afã de cumprir uma ordem, tenha ele próprio se aventurado no ato tradutório.

Impressiona nos primeiros contatos com o GPT a operacionalidade, ao mesmo tempo, em recortes, como se fossem “pílulas informativas” dotadas de alguma completude em si, mas que permite ao solicitante proceder a desdobramentos sob forma de novas perguntas a partir das repostas dadas e cuja velocidade de processamento demanda poucos segundos. É como se tivéssemos uma série de *links* em duplicações

infundas controladas por algoritmos probabilísticos. Os arautos do neoliberalismo devem assistir extasiados este espantoso aumento de produtividade, sobretudo porque junto dele a força de trabalho humana conhece o caminho da míngua, nos vários sentidos do termo.

Desconstrução

Os erros e bizarrices contidos em muitas respostas dadas pelo ChatGPT (vale para outros aparentados: Bard, Bing, Reddit etc.), já se tornaram folclóricos, como dizer que o Brasil teve ganhadores do prêmio Nobel (infelizmente, ainda não tivemos a honra desta glória), que Fernanda Montenegro ganhou um Oscar pela sua atuação no filme *Central do Brasil* (apesar de merecer, a estatueta, naquele ano, foi para Gwyneth Paltrow). Chama atenção a existência de problemas envolvendo aspectos biográficos de nomes importantes na cultura brasileira, apesar de, sobre eles, haver profusão de dados em bases como as da Wikipedia. Exemplos: indagado acerca da relação de parentesco entre Mário de Andrade e Oswald de Andrade, a resposta veio pronta e sob forma de fantasia: primos – entre eles inexistia relação de tal natureza familiar, apesar do mesmo sobrenome; perguntado sobre o célebre atentado da Rua Tonelero, o GPT informou, entre outros disparates, que Carlos Lacerda estava acompanhado do major Rubens Tolentino Vaz, referido como chefe da guarda do presidente Getúlio Vargas – o dito militar da aeronáutica, morto no atentado, fazia parte da segurança de Lacerda, que era inimigo figadal de Getúlio.

Em mudança de registro, mas ainda explicitando o vetor dos equívocos, fizemos o seguinte exercício envolvendo o campo literário. E vale acompanhar a confusão do *Chat*.

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

R. *A poesia “Ora (dizeis) 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 (dizeis) 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 (dizeis) ouvir estrelas!” é de autoria de Gonçalves Dias, um importante poeta brasileiro do século XIX (...)*

C. Partindo da precedente relação entre o poema e o autor, aduziram-se outras indagações alimentadas por nuances que, sem perder o núcleo enunciativo – o vínculo entre “Ora (dizeis) ouvir estrelas!” e Olavo Bilac –, colocavam dúvidas sobre as informações advindas do *Chat*. E o que se verifica é um crescendo alucinatório. Manuel Bandeira é acionado pelo cérebro eletrônico para resolver o problema, numa operação de registro literário semelhante ao duplo twist carpado. E advém o nome de Gonçalves Dias.

É possível deduzir, ao menos conforme este circunscrito experimento, a existência de quatro modalidades discursivas evidenciadas no processo generativo, passíveis de cruzamentos em suas enunciações⁹.

A primeira consiste em uma “insegurança” do sistema algorítmico diante de pergunta que coloque sob suspeita a resposta do robô, ou traga consigo incerteza, ambiguidade. E mesmo ao contrário, que revele assertividade, convicção, por parte do interrogante, mesmo que descabida. Frente ao desacordo, digamos, qualificado: “Creio haver um erro na informação. A poesia “Ora (dizeis) ouvir estrelas!” **não pertence** a Olavo Bilac, o modelo algorítmico, em sua operação estatística de procurar padrões e sequenciação das palavras que respondam ao objeto da indagação, avoca, aleatoriamente, termos de maior proximidade com o objeto da perquirição. Caso haja procedência em nossa observação, é possível que o nome de Manuel Bandeira tenha surgido por estar ele vinculado à produção poética, escrevendo crônicas, críticas literárias, feito referências a Olavo Bilac, ao “Ora (dizeis) ouvir estrelas!”, para não alongarmos a lista dos dados que alimentam o sistema de procura do GPT. No tangente ao nome de Gonçalves Dias, a indagação do perguntador (que não deixa de trazer alguma assertividade) foi suficiente para dirigir a resposta desencontrada sobre a autoria do poema escrito por Olavo Bilac – ou seja, o *Chatbot* apenas ofereceu uma resposta, como se estivesse a desincumbir de mera tarefa, a partir de escolhas guiadas estatisticamente (logo, sem a possibilidade de avaliação posterior) a partir de áreas de aproximação que, ao fim e ao cabo, revelam improcedência.

A segunda diz respeito a uma “área sombreada” entre a necessidade de inventar alguma solução para a dar prosseguimento à demanda feita pelo consulente e algo que vem sendo chamado pelos especialistas em IA de alucinação – voltaremos ao termo em páginas seguintes. Os dois vocábulos, tanto o verbo inventar como o substantivo alucinação possuem efeito retórico, força figurativa. De um lado, até onde se pode visualizar, as máquinas não possuem capacidade de atribuir sentidos aos símbolos que produz, conforme observação feita há tempos por John Searle: “Is the brain’s mind a computer program? No. A program merely manipulates symbols, whereas a brain attaches meaning to them” (1990, online). O inventar deve ser lido, portanto, menos como sinônimo de criatividade ficcional, artística, por exemplo, e mais como responsividade maquínica, conforme instrução programada pela arquitetura de linguagem do GPT – Transformador Generativo Pré-Treinado, é bom lembrar. E, de outro, o *gap* demencial – para seguir terminologia médica voltada à designação de síndromes humanas, agora associadas aos dispositivos tecnológicos¹⁰. De toda sorte, a área sombreada dificulta estabelecer diferenças muito claras entre o que é “invenção” ou “alucinação” – um voo livre e cego do qual identificamos resultados, mas temos – nós, simples usuários destes sistemas – dificuldades em reconhecer os processos, a exemplo de ligar “Ora (dizeis) ouvir estrelas!” aos nomes de Manuel Bandeira e Gonçalves Dias.

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”. E que chamaremos de modalidade discursiva “desculpista”. Há, porém, inúmeros relatos na imprensa e mesmo no citado artigo de

⁹. Com estas modalidades estamos indicando um possível caminho para o estudo das inúmeras formas discursivas que se apresentam no GPT e outros *chatbots*.

¹⁰. A velocidade com que se desenvolvem os sistemas tecnológicos e seus dispositivos nem sempre vem acompanhada de vocabulário capaz de nomear resultados ou fenômenos por eles desencadeados. É o presente caso envolvendo verbos como inventar e alucinar, aplicados, por livre exercício de similaridade, aos robôs e outros petrechos maquínicos.

Olavo Amaral, de grosserias 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 associar, até com certa agressividade, a eleição do Presidente Lula¹¹ a fraude, ecoando as conhecidas arengas da extrema direita brasileira sobre as urnas eletrônicas e a violabilidade dos códigos fonte. Entretanto, nas poucas respostas às perguntas sobre o poema “Ora (dizeis) ouvir estrelas!”, o GPT revelou um tom de amigabilidade, lhanza: “*Peço desculpas pelo meu erro na resposta anterior; Mais uma vez peço desculpas*”. Agora, indicando não apenas haver assimilado a discordância do usuário, como reprocessado e ajustado o mecanismo de busca: *Eu verifiquei a informação novamente e de fato a poesia (...). Peço desculpas pela confusão. Você está completamente certo. Vale dizer, o “desculpismo” é uma marca, espécie de mantra padronizado na estrutura algorítmica do GPT, nos embeddings, de sorte a responder uniformemente às eventuais discordâncias (mais ou menos procedentes) com novos enunciados que reconhecem o desvio cometido – conquanto possam ser acrescidos de outros equívocos. E, certamente, o grau de esclarecimento da resposta (ou: abjuração, descontentamento, ironia, ridicularização, incompletude) decorrente das consultas feitas ao GPT ficará na dependência do lugar social, cultural, educacional, de experiências, dos níveis de informação e conhecimento, de quem está fazendo a pergunta.*

A quarta está afeta à “procedência redacional”. Ou seja, mesmo quando fornecendo informações distorcidas, im procedentes, mergulhando naquela área obscura entre a “invenção” e a “alucinação”, o texto dado à leitura do consulente está bem estruturado. Malgrado o erro de atribuir a Manuel Bandeira a autoria do poema de Olavo Bilac, um leigo nestas questões literárias poderia acolher a explicação do GPT, entre outros motivos porque legível, coerente (a despeito do lapso contido na informação), coesivo (não são verificados desvios ortográficos ou de escolha vocabular). É possível que leitores atentos a questões estilísticas identifiquem certa pasteurização nos constituintes discursivos, mas, novamente, ficará na dependência da maior ou menor experiência com os textos escritos.

Em resumo do que foi observado nos exemplos examinados neste tópico. O consulente do GPT precisa estar atento às respostas, aguçando a inflexão analítica e crítica sobre o que lê, pois, na massa de informações disponibilizadas, existem erros de várias ordens. As perguntas requisitam clara formulação, sem dubiedades ou contradições, procedimento que permitirá obter maior taxa de acertos nas respostas vindas do robô – saber perguntar será um desafio importante posto ao usuário da inteligência artificial generativa.

Desafios

Acessar o GPT significa reconhecer a existência de um enorme salto nos mecanismos tradicionais de busca permitidos pelas mudanças tecnológicas, sobretudo aqueles vinculados à internet, de que servem como exemplos as modelagens iniciais do Google, ou mesmo dos *designs* primevos de inteligência artificial. Estamos nos inícios dos chamados grandes modelos de linguagem ancorados em algoritmos de redes neuronais profundas e que elaboram padrões considerando bilhões de dados, assim como promovendo previsões cujos resultados podem estar expressos em frases, períodos e

¹¹. Portal g1. 13/07/2023. <https://g1.globo.com/tecnologia/noticia/2023/07/13/bard- robo-de-inteligencia-artificial-do-google.ghtml>. Acesso. 24/07/2023. O Google respondeu à reportagem dizendo que estando o Bard em período experimental, equívocos, imprecisões, assertivas inapropriadas poderiam ocorrer, mas a companhia estava atenta ao problema, 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.

textos inteiros elaborados com boa organização interna e clareza expositiva, a despeito dos seus distintos problemas, conforme alguns acima apontados.

Neste subitem, procuraremos, identificar fraturas, desvios e inconsistências que acompanharam as nossas “interações” com o GPT, tentando compreender como a plataforma organiza determinadas estruturas discursivas. Em particular, alimenta o nosso interesse frente à complexidade do assunto em pauta a abertura de canais dialógicos com os educadores, haja vista tanto a estupefação que vem sendo promovida pela arquitetura *transformer*, como sua incontornável presença no âmbito escolar. Estamos frente a um fato tecnológico posto no cotidiano de vasto grupo de discentes (e docentes), que já acionam o dispositivo para o andamento de seus afazeres na educação formal. E, certamente, serão crescentes as dificuldades de professores e professoras visando à verificação das atividades levadas a termo pelos alunos, por exemplo, fora da sala de aula. Afinal, como saber se uma determinada tarefa foi cumprida em decorrência do empenho individual/coletivo do estudante ou se resultou da apropriação/cópia/transcrição daquilo fornecido em segundos pelo GPT (e correlatos) – dispositivo fingidor do *continuum* humano sob nome que não se esclarece.

A resposta óbvia da existência de checadores de textos como Turnitin, Copyscape, Plagscan, frases buscadas no Google etc., ajudam, mas não resolvem o problema. A menos que nos afastássemos da realidade da educação básica pública onde cada docente deve ministrar aulas para centenas de discentes. E, nestas condições, como fazer um acompanhamento de tarefas que seja capaz de identificar fontes e origens em meio à relativa sofisticação redacional – e mesmo possibilidades de desdobrar pesquisas graças a pequenos reajustes nos mecanismos de procura – do *Chatbot*?

A seguir arrolamos alguns dos problemas identificados nesta incursão ao GPT, considerados os previsíveis reptos que serão interpostos aos sistemas educativos formais.

- Origem

O material fornecido nas respostas dadas pela predição que acompanha a linguagem natural do dispositivo, não esclarece as fontes de onde se origina. É possível, portanto, estarmos diante de referências sem credibilidade, indicadores baseados em aparatos bibliográficos preconceituosos, ou, eventualmente, em artigos elaborados por especialistas e derivados de pesquisa científica, mas cuja autoria se perde no vasto território da incógnita. Assim, o mapa informacional revela-se tortuoso, inseguro, impedindo que ocorra autenticação de proveniência dos textos acessados pelo consulente. Parecem óbvias as resultantes de se tráfegar por um ambiente nebuloso, sobretudo quando está em jogo a elaboração de empreitadas acadêmicas, cuja credibilidade tem como uma de suas parceiras diletas a explicitação das referências bibliográficas e a clareza na menção das fontes de pesquisa.

- Plágio/cópia

A despeito de existirem, como já referido, checadores de textos voltados a descobrir plágio, cópia, decalque, tal processo é, ainda, relativamente incipiente para, digamos, controlar as produções derivadas do GPT e assemelhados. E isto embute um problema 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* que, por seu turno, a copiou/montou/remontou de bases de dados não esclarecidas. Conquanto a teoria linguística discuta bastante o conceito de apropriação textual e originalidade expressiva – assunto sobre o qual não cabe incursão neste passo – acionamos a palavra cópia e mesmo plágio visando a indicar assenhoreamento impróprio de trabalho alheio. Cria-se curiosa novidade na qual a forma-texto avaliada pelo docente,

mesmo quando despida de originalidade – a menos que cópia e plágio sejam considerados como algo original – pode vir a receber loas e boa nota em termos de avaliação. Cabe lembrar que os nossos professores e professoras do ensino fundamental e médio, com carga horária de trinta e seis horas em sala de aula, possuem algo em torno de trezentos alunos nas diferentes turmas ao longo da jornada semanal. Frente a tal número a supervisão amíúde dos afazeres escolares deixa ainda mais vulnerável o reconhecimento do que foi ou não produzido, na origem, por dispositivos de inteligência artificial generativa. Acrescente-se: o GPT já é um grande plagiário, pois limitado a cumprir ordens de busca em grandes bases de dados, orientado por modelos estatísticos, por algoritmos e procedimentos preditivos, cujas fontes de consulta não são consignadas nas respostas/cópias oferecidas ao consulente.

- Informação/desinformação

Existe a tendência de os usuários validarem o que foi acessado através dos dispositivos maquínicos, daquilo posto nas redes sociais, do recolhido dos buscadores. Ainda que o ChatGPT não seja, no sentido estrito, um buscador, haja vista a sua estrutura funcional, conforme já explicitamos, há o sério risco de, simplesmente, ratificar o texto que nos chega através dele. 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 das referências agregadas ao texto processado pelo *chat*, emerge o rebaixamento dos limites entre informação, desinformação, fato, interpretação desviante, *fake news* etc. Vale dizer, a dificuldade de monitorar o que chega através dos materiais gerados pelo GPT entra em circuitos nos quais paráfrases e reduplicações de dados enviesados ou mesmo falsos podem esmaecer as diferenças entre informação e desinformação.

Reside, aqui, um dos problemas para os quais, doravante, os docentes precisarão estar alertas ao solicitarem dos discentes a realização de trabalhos extraclasse. A facilidade para acessar os sistemas de inteligência artificial, assim como a rapidez nas respostas que oferece, pode trazer consigo inúmeros equívocos, assim como fragilizar o pensamento reflexivo e crítico. Esse, por sua vez, pede uma temporalidade antiacelerativa, ritmo necessário para apreender a complexidade que envolve a contemporânea profusão (des)informativa. Uma das maneiras de circundar tal desajuste no âmbito da escolaridade formal é o investimento em educação midiática, educ comunicativa, estratégia através da qual a produção decorrente dos *chatbots* tem condições de ir além da tendência hoje dominante de considerar a inteligência artificial generativa como uma espécie de fruto encantado a vicejar no jardim das delícias.

- Imparcialidade/parcialidades/ cruzamentos

Tem sido matéria de discussão o acionamento dos chamados algoritmos discriminatórios e as suas variáveis de controle. O procedimento vem sendo levado a termo, 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 capítulo amplo envolvendo preconceitos de raça (o caso dos identificadores faciais postos em ruas e praças demonstra como a orientação algorítmica pode resultar em suspeita e mesmo prisão de pessoas pertencentes a determinados grupos étnicos), intolerância política, segregação econômica, para não alongarmos a lista. Como modelo de linguagem programável, o GPT não está livre de praticar enviesamentos que resultam na construção de estereótipos e estigmas. Afinal, a inteligência artificial, por mais avançada que seja, é, ainda, produto de ação humana

(malgrado o chamado treinamento de máquinas, o *machine learning*, a autopoiesis, o automatismo na construção dos modelos analíticos), logo, responsável primeira pela estruturação dos modelos: “opiniões embutidas em matemática” (O’Neil, 2020, p.35)¹². Em síntese: a fama de imparcialidade dos algoritmos e derivados não faz sentido, pois, juntamente com as suas arquiteturas seguem os marcadores ideológicos, os corpos de crenças, os propósitos, dos quais o criador não consegue escapar ao embutir nas criaturas determinado modelo de linguagem, a exemplo do *transformer*. A apreensão liminar dos conteúdos apresentados pelo ChatGPT registra o potencial de trazer consigo vieses e formulações discriminatórias, cruzando, sob o manto da neutralidade, inúmeras formas de intolerância. Isto reforça assertiva feita em tópico anterior, acerca da necessária educação midiática, colocando sob perspectiva analítica, reflexiva, o material acessado por docentes e discentes junto aos dispositivos de inteligência artificial generativa.

- Computar/pensar

O eventual uso (programada pedagogicamente ou não) dos dispositivos de IA na educação formal, sobretudo quando envolvendo o ensino fundamental e médio, precisa vir precedido de um debate sobre os limites entre computar e pensar. O verbo computar está sendo utilizado como licença expressiva englobando a arquitetura *transformer*. Noutros 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 prever a palavra que entrará em determinado enunciado, mas não possui, ao menos até o momento, como asseverar a procedência da tarefa executada. O que temos, é uma rota e uma lógica algorítmica gerenciada através de *inputs* carregados para o interior de um sistema com *modus operandi* orientado por associações formais – menos referentes ao eixo das similaridades semânticas. O exercício humano de pensar envolve, por sua vez, reflexão, crítica, escrutínio, refazimento, amadurecimento, abstração, oscilação de comportamento, autocrítica, capacidade de reorganizar com celeridade certos dados¹³ etc. Equivale dizer: cabe à educação o lembrete, sempre oportuno e necessário, particularmente quando envolvendo os jovens do ensino básico, que computar e pensar não são atos redutíveis a um mesmo enunciado. A inteligência artificial é de uma ordem, já a inteligência humana, de outra. Essa expande a consciência e a capacidade de amar, sentir, revelar empatia e afeto, ser multifacetada, ter enorme força criativa, operar processos de reversibilidade. Visto desta ótica, e conforme exposto em páginas anteriores, o GPT é “honesto” ao deixar claro possuir uma inteligência avançada, mas diferente da humana, com limites e habilidades distintas.

- Alucinação

Noutras passagens deste texto tratamos do termo alucinação. Fixamos, aqui, outras considerações. Sendo o GPT um autônomo cognitivo (a professora de linguística da área computacional da Universidade de Washington, Emily M. Bender, em artigo

¹² . O citado Geoffrey Hinton anunciou nos primeiros meses de 2023 o seu afastamento das funções que exercia junto ao Google. Segundo ele, empresas deste ramo caminham por uma trilha perigosa ao permitirem que os robôs tanto gerem como executem códigos de computador, tornando-se autônomos. Vindo de um dos “padrinhos” do assunto, é algo preocupante.

¹³ . É conhecido entre os que trabalham com IA o chamado *overfitting* (superajuste/sobreajuste). Uma dificuldade existente nos algoritmos das máquinas de se ajustarem a novas informações, pois os modelos recebem o pré-treinamento com dados temporalmente defasados. Por exemplo, para o GPT3, em março de 2023, Pelé estava vivo (morreu em dezembro de 2022). A resposta do robô, já prevendo o “efeito *overfitting*”, foi: “Pelé, o famoso jogador de futebol brasileiro, está vivo. Até a data de corte do meu conhecimento, que é 2021, não houve nenhum relato confiável da sua morte. Portanto, acredita-se que ainda esteja vivo. No entanto, é importante ressaltar que qualquer informação sobre a saúde e o estado de vida de uma pessoa, pode mudar a qualquer momento”. A resposta do *chat* não deixa de ser divertidíssima.

escrito junto com colaboradores, deu ao *Chat* o epíteto de *Stochastic Parrots*¹⁴ – Papagaio Estocástico) dele não se afasta a possibilidade de entrar em surtos “demenciais”. Sinteticamente: a palavra alucinação é utilizada para caracterizar a produção, por parte do GPT ou outros modelos de linguagem baseados em IA, de textos sem sentido, apartados da realidade, derivados para planos imaginários, a despeito da permanência de padrões de linguagem natural ancorados em recursos expressivos procedentes. Tais desvarios ocorrem por diversos motivos, destacamos dois: perguntas não muito claras feitas pelo consulente; problemas no próprio treinamento da máquina. É necessário, portanto, alertar os usuários de que nem todo texto saído do GPT é aproveitável. Desatinos e insanidades perturbam, também, os sistemas alimentados pela sofisticação da inteligência artificial generativa. Há obras de ficção científica, mesmo produzidas há tempos, com um elenco de robôs e inteligência artificial (conquanto em fase pré-generativa) que enlouquecem, promovendo toda sorte de tragédias, vinganças, revoltas contra humanos – esperamos que tais distopias não se transmudem em realidade. A propósito e concernente aos processos alucinatórios, perguntamos ao GPT o significado do termo transmudem. A resposta:

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.

- Circuitos

Uma das marcas do processo civilizatório é o da constante mudança nos instrumentos produtivos: navegação a vela, vapor, eletricidade – a acompanharem as revoluções mercantil e industrial. O que assistimos, agora, junto com a ampla digitalização é o crescente acionamento da inteligência artificial para finalidades diversas: controlar meios de pagamento, definir perfis de usuários das companhias de seguros, minerar dados, enfim, a lista é infindável. Este novo quadro sociotécnico, tecnocomunicacional, 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 autopoiesis, a *machine learning*. Tal ritmo e suas consequências para a vida societária, incluindo a educação formal, suscita a necessidade de conhecer, minimamente, os desafios e possibilidades agora colocados pela inteligência artificial generativa, o que significa ter algum domínio sobre a funcionalidade e os fluxos alimentadores do circuito: abrir a tela inicial do GPT (ou assemelhados); como fazer as perguntas; maneiras de checar a informação dada no processo de consulta; reconhecer as alucinações, equívocos e absurdidades que emergem do robô etc. Em um termo: operar com a arquitetura *transformer* no contexto da educação escolar implica ativar a vigilância (e alguma astúcia) sobre o que advém das “interações” com o intangível cérebro eletrônico. Isso não deixa de consignar um procedimento que pode contribuir para baixar os níveis de alienação (ou estupefação) dos usuários frente a um petrecho maquínico que se insinua como pleno para responder aos segredos escondidos no mais fundo do universo.

¹⁴. Ver: <https://dl.acm.org/doi/pdf/10.1145/3442188.3445922> . Acesso 10/05/2023

Nota final

O frenesi que acompanha o crescente uso do ChatGPT traz consigo indagações, dúvidas, incitações e muito espanto. Proliferam análises advindas da academia, do jornalismo, de especialistas, todas empenhadas na tarefa de entender, explicar, ou, simplesmente, especular acerca das virtudes e dos vícios ensejados no e pelo dispositivo. Conforme tentamos evidenciar, encontramos um sistema marcado seja por respostas coerentes, aceitáveis, bem formuladas, seja por inconsistências e fragilidades internas. É preciso acrescentar, entretanto, que o GPT3 está ou esteve em fase de, digamos, “teste”, uma espécie de primeira infância dos modelos generativos, cuja pré-adolescência pode ser encontrada na nova versão, de número 4 – para tanto, a contribuição dadivosa e gratuita dos usuários, fornecendo dados é fonte de inesgotável alimentação. A despeito da pirataria – e problemas conexos – em curso, fica claro o enorme alcance tecnológico da inteligência artificial profunda.

A porta de entrada do GPT (e análogos) mal foi aberta e, no entremeio, vemo-nos diante de inquietações variegadas. Há, como posto em páginas anteriores, manifestos assinados por cientistas, intelectuais, e até, estranhamente, donos de *big techs*, pedindo para os governos intervirem no sentido de diminuir a velocidade no avanço da pesquisa e implantação da inteligência artificial generativa. A pausa seria necessária para medir as consequências e mais bem assimilar o alcance e significado deste novo desafio científico-tecnológico. 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 papel relevante no desenvolvimento e implementação da inteligência artificial generativa. E registram a singularidade de viver em desassossego existencial, haja vista, ao mesmo tempo, ajudarem a pôr no mundo uma máquina altamente sofisticada e nela visualizarem desdobramentos capazes de colocarem sob risco a própria espécie humana, conforme o já referido Geoffrey Hinton. É o drama shakespeariano do ser ou não ser em apresentação entre o Vale do Silício, Shenzhen e quejandos. A boa consciência destes humanistas ressabiados pode ganhar redenção, além de interveniências menos nobres, pelo fato de que sendo especialistas no assunto, designers de redes neurais, elaboradores da arquitetura *transformer*, apresentarem melhores condições de controlar os “humores” dos robôs, vigiando os ímpetos de plena autonomia das máquinas, impedindo, enfim que as criaturas acabem se impondo aos criadores.

O fundador da OpenAI, Sam Altman, o singular CEO (Chief Executive Officer) da companhia, ao mesmo tempo vegetariano, fazendeiro criador de gado de corte, crente na IGA (Artificial General Intelligence), diz que os racionalistas e altruístas são bem intencionados, mas exageram na visão segundo a qual estamos levando a humanidade para um inevitável desastre. Enfim, entre a defesa sumária da inteligência artificial generativa e a visão apocalíptica sobre as consequências da expansão ilimitada do sistema, e que poderá escrever e executar os seus próprios códigos, alinham-se gregos e troianos.

A despeito destas migrações enredando crentes, descrentes e habitantes do purgatório, é importante lembrar que estamos frente a uma indústria oligopolizada – aquela chamada de 4.0, pós-industrial, digital etc. – sob a qual cintila o neon das *Big Techs* e para a qual a reprodução do capital e o domínio de espaços de poder tecnológico, e seus derivados, representam o único limite. 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 extremamente elevados e os seus *players* anunciam aportes de bilhões de dólares em pesquisa e no desenvolvimento de protótipos, com a mesma facilidade com que compramos hortaliças na feira de domingo. A se perguntar como boa parte do sul global pode participar desse encontro organizado pelas minorias já prósperas.

Existem subprodutos decorrentes da organização do capital contemporâneo e assentados na robótica. Para não irmos longe, bastaria avocar a precarização do mercado de trabalho (já se vislumbra a fragilização de setores afeitos ao mister intelectual, como o jornalismo e o Direito); a redefinição dos relacionamentos interpessoais; os espalhamentos comunicacionais a trazerem consigo a desinformação e os discursos de ódio. As tecnologias de inteligência artificial baseadas em redes neurais profundas registram circunstâncias positivas para o desenvolvimento humano, mas delas não podem ser abstraídos itens como os postos acima.

Frente a esse quadro, a educação formal conhecerá enormes provocações nos próximos anos, sendo que uma delas representada pelo avanço dos modelos de linguagem expressos em *chatbots* como o desenvolvido pela OpenAI. Afinal, com um simples acesso à internet é possível que o GPT forneça a resolução de problemas matemáticos, escreva textos acerca da obra de Graciliano Ramos ou informe sobre a vitória do Brasil na copa de 1958, realizada na Suécia. Se tais informações possuem procedência, estão livres de alucinação, escoimadas de viés segregacionista, serão perguntas subsequentes. Uma educação analítico-crítica, compreensiva dos processos sociotécnicos, que busque entender e não temer a presença da inteligência artificial, será decisiva para estabelecer limites nas relações (mais ou menos visíveis) que as salas de aula manterão com o GPT e correlatos.

Referências

- AMARAL, Olavo. As formas intermediárias. *Revista Piauí*, nº 199, 2023, pp. 20-27.
- BENDER, Emily M; GEHRU, Timnit; McMILLAN, Angelina; SHMARGARET, Shmitchell. *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?*. 2022. <https://dl.acm.org/doi/pdf/10.1145/3442188.3445922>. Acesso. 10/05/2023
- CHOMSKY, Noam et all. A falsa promessa do ChatGPT. *Folha de S. Paulo*. 10/05/2023. <https://www1.folha.uol.com.br/tec/2023/03/a-falsa-promessa-do-chatgpt.shtml>. Acesso. 12/05/2023
- 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. Opinion*, 24/03/2023. <https://www.nytimes.com/2023/03/24/opinion/yuval-harari-ai-chatgpt.html>. Acesso. 10/05/2023
- 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*, 1990. <https://www.scientificamerican.com/article/is-the-brains-mind-a-computer-progr/>.