

Os grandes modelos de linguagem



Por **ELEONORA ALBANO**

O despiste mais comum no discurso das Big Techs – as atuais proprietárias dos “grandes modelos de linguagem” e tecnologias afins – é a meia-verdade

A onipotência da ciência dura do norte global

Numa era de crescente datificação da experiência humana, não chega a surpreender que o pensamento crítico esteja em baixa entre os cientistas. Mas é, de qualquer forma, alarmante que isso possa ocorrer na editoria de um periódico acadêmico tradicional, prestigioso – e generalista. Foi, portanto, com um misto de espanto e indignação que li o parágrafo de abertura, transcrito abaixo, do editorial do volume de julho de 2023 da revista *Nature Machine Intelligence*. Como se sabe, o grupo editorial *Nature*, fundado em Londres na segunda metade do século XIX, tem por missão disponibilizar para toda a comunidade científica sinopses fidedignas dos avanços das várias áreas do saber.

“Frederick Jelinek, a renowned Czech-American researcher in natural language processing and speech recognition, famously said in 1985, “Every time I fire a linguist, the performance of the speech recognizer goes up”, suggesting that there may be no efficient way to include linguistic knowledge in such systems. Does this sentiment also hold true for state-of-the-art large language models (LLMs), which seem to be mostly artefacts of computer science and engineering? Both LLMs and linguistics deal with human languages, but whether or how they can benefit each other is not clear”.[\[i\]](#)

É, no mínimo, irresponsável que os que divulgam os avanços da inteligência artificial (IA) ao resto da comunidade científica ignorem – ou não queiram admitir – que os grandes modelos de linguagem (doravante, GMLs, como no inglês LLMs) se abeberam fartamente de conceitos e técnicas da linguística, assim como das demais ciências que descrevem e interpretam a linguagem natural.

Há pelo menos duas razões para tamanha desinformação. A primeira é política. É que os/as cientistas humanos – linguistas, psicólogos/as, sociólogos/as, antropólogos/as, etc. – responsáveis pelas taxonomias indispensáveis às bases de dados de treinamento dos “grandes modelos de linguagem” são, em geral, cidadãos/ãs de “segunda categoria”, recrutados/as em países pobres, como trabalhadores/as terceirizados/as e precarizados/as, para as tarefas consideradas “triviais” de etiquetar os dados e, se necessário, tomar decisões sobre categorização.

A segunda é a epistemologia geralmente associada a essa visão da ação e, em especial, do trabalho humano: qualquer ação, de qualquer complexidade, pode ser reduzida a uma cadeia de associações. Reconhece-se aí a influência do empirismo inglês, intermediada pela sua já centenária[\[ii\]](#) sequele na psicologia estadunidense, o behaviorismo.

Veremos abaixo como a concepção de ação desenvolvida na Europa a partir do século XVII reifica não só a atividade, mas também o pensamento humano. Uma noção atomista de mecânica, vaga o bastante para caber tanto no empirismo como no racionalismo, possibilitou a exclusão de certos grupos da humanidade, assimilando-os ao animal-máquina cartesiano. Por meio dela, as potências coloniais europeias justificaram facilmente a escravização dos nativos das suas colônias, assim como a espoliação dos pobres da própria Europa.

Tomando o cenário colonial como pano de fundo, o presente ensaio pretende mostrar que há um forte vínculo entre a posição política que reduz certos tipos de mão de obra científica a uma linha de montagem e a posição “científica” que encara a linguagem natural humana como uma infinidade de cadeias interconectadas.

O fato de essas ideias estarem implícitas na produção das tecnologias languageiras permite que a sua violência tácita, inerente à formação nas ciências da computação, milite pelos interesses da faceta vigente da colonialidade, conhecida como capitalismo de plataforma ou vigilância.[\[iii\]](#)

A linguagem natural desnaturalizada

O citado editorial honra a tradição empirista não apenas porque a revista é inglesa, mas, sobretudo, porque a área de processamento de linguagem natural (doravante, PLN) – subárea da inteligência artificial responsável pelos “grandes modelos de linguagem” – nasceu e floresceu num ambiente fortemente empirista – ou, mais precisamente, behaviorista.

Os pesquisadores da área creem que a mente humana é uma máquina de Turing, composta de bilhões de autômatos de estado finito[\[iv\]](#) entrecruzados. Não surpreende, pois, que os CEOs das empresas da área vejam as pessoas que alimentam os “grandes modelos de linguagem” como meras máquinas provedoras das informações necessárias e suficientes para habilitar os “grandes modelos de linguagem” a passar o teste de Turing num futuro próximo.

Esses especialistas sequer se dão conta – ou fingem não se dar – de quanta inteligência subjaz à faina das/os etiquetadoras/es. Como documentou o jornalista investigativo Josh Dzieza,[\[v\]](#) em Inteligência Artificial é praticamente impossível adotar de saída uma categoria intuitiva, pois exemplares distintos de um mesmo objeto tendem a ser tomados como indistintos pela máquina.

Para treiná-la a mimetizar as nossas categorias, os etiquetadores têm que engendrar uma subcategorização detalhada e organizá-la numa hierarquia de níveis. Assim como os demais robôs, os “grandes modelos de linguagem” requerem uma infinidade de iterações a fim de alcançar generalizações às quais qualquer criança humana chega após uma exposição relativamente baixa aos dados.

Em processamento de linguagem natural, uma das principais manifestações desse tipo de dificuldade está nos contextos sintáticos e semânticos que contêm discontinuidades.

Note-se que a tarefa de um “grandes modelos de linguagem” é sempre prever a próxima palavra – como fazem, ainda que grosseiramente, os editores de texto dos telefones celulares. Trata-se de uma tarefa fácil no caso dos clichês, cujos termos coocorrem com altíssima frequência, mas muito difícil na maioria dos demais casos.

Assim, em expressões de frequência mais baixa, a estimação estatística só se viabiliza graças a quatro componentes, todos indispensáveis: uma base de dados de muitos bilhões de palavras; uma tecnologia poderosíssima – cuja capacidade de aprender associações supera a das redes neurais recorrentes (mesmo as profundas, isto é, de múltiplas camadas); descrições gramaticais e semânticas exaustivas; e um treinamento intensivo no qual se corrigem iterativamente os erros de associação. Obviamente, as referidas descrições e correções são todas feitas por humanos – terceirizados e precarizados.

Estima-se que as bases de dados que alimentam os chatbots dialógicos tais como o *Chat GPT*, da Open AI, o *Bard*, do

Google, e o *Bing*, da Microsoft, sejam da ordem de 300 bilhões de palavras. O dispositivo poderosíssimo que permite a sua utilização em tempo real é chamado de *transformer* – traduzido como transformador. Trata-se de um modelo estatístico que aplica um conjunto de técnicas matemáticas, denominadas “atenção” e “autoatenção”, para detectar dependências entre os elementos de uma cadeia – no primeiro caso, do input ou do output; no segundo, da própria cadeia em curso.

Uma massa de robôs interconectados

Note-se que o transformador subjacente aos “grandes modelos de linguagem” pouco tem a ver com o aparelho homônimo que modifica os níveis de tensão da corrente elétrica. O que modifica, ao escolher uma palavra após outra, são as próprias relações entre os termos da base de dados (doravante corpus, para simplificar), pois cada nova ocorrência retroalimenta a entrada e reorganiza a rede de relações existente.

Não é difícil entender como. Todas as relações são expressas por pesos de conexões entre os nós das sub-redes do corpus. Esses pesos são, por sua vez, calculados a partir das probabilidades de coocorrência dos pares possíveis de palavras. O transformador é suficientemente poderoso para permitir que todos os membros do corpus, assim como todas as suas etiquetas (gramaticais, semânticas, discursivas, psicológicas, sociológicas, políticas, etnográficas, etc.), conectem-se entre si simultaneamente, de forma que o cálculo da próxima palavra possa considerar os mais variados aspectos do enunciado em curso e do seu contexto.

Sublinhe-se que o volume de etiquetas disponíveis para refinar esse cálculo é gigantesco. As mais simples abrangem aspectos tais como as classes gramaticais e as formas textuais de nomeação e referência (por exemplo, nomes próprios, pronomes pessoais, demonstrativos, possessivos, etc.).

Vale observar, ainda, que a etiquetagem não se atém às palavras. Compreende também as partes do discurso (p.ex., sujeito, predicado, adjuntos); as frases e sua classificação sintática (p. ex., principal, subordinada e respectivas subclasses); e gêneros textuais orais ou escritos (p.ex., coloquial, literário, jornalístico, jurídico, científico, etc.).

Quem, com base no exposto, tiver imaginado que as bases de dados dos “grandes modelos de linguagem” parecem gigantescos dicionários cujos “verbetes” estão implícitos na sua rede de conexões desvendou um terço da charada. Porém, o conteúdo dos outros dois terços é igualmente importante: consiste em informações gramaticais e enciclopédicas cruciais – aportadas, mais uma vez, pelos/as etiquetadores/as.

De fato, todas as ocorrências de uma mesma palavra estão conectadas entre si; e as suas diferentes acepções são representadas pelas semelhanças e diferenças, codificadas matematicamente, entre as frases às quais se conectam. Isso acaba funcionando aproximadamente como os exemplos fornecidos nos verbetes das palavras polissêmicas dos dicionários.

Além disso, os elementos de cada frase conectam-se a uma gramática. Nela as estruturas sintáticas se mapeiam em diagramas em árvore[[vi](#)], enquanto as estruturas semânticas remetem a diferentes formas lógicas (via cálculo proposicional, entre outros) e campos semânticos (p. ex., os nomes de frutas conectam-se, de baixo para cima, aos campos das frutas, dos vegetais, dos alimentos, etc.). A gramática remete, ainda, a indexadores de correferência (p. ex., na frase “João disse que isso não o incomoda”, o pronome ‘o’ pode se referir ao próprio João, a uma segunda pessoa ou a uma terceira).

Finalmente, os textos são classificados com base em conhecimento de mundo (p.ex., assunto, gênero, autoria, tom, estilo, fontes documentais; com etiquetas que vão das mais genéricas às mais especializadas). Essa informação, depois de minuciosamente anotada e conectada, permite uma infinidade de buscas a fim de atender a demandas complexas, tais como resolução de tarefas escolares, redação de pareceres jurídicos, auxílio em diagnósticos médicos, etc.

Esse acervo gigantesco e exaustivamente interconectado dá aos “grandes modelos de linguagem” uma enorme capacidade

de construir frases “novas” parafraseando fragmentos de dados contidos na própria base. Quando o eminente linguista, filósofo e matemático Noam Chomsky diz que os *chatbots* são meros plagiadores, não está afirmando que apenas copiam e colam recortes literais da base. Está, na verdade, supondo que produzem bons pastiches dos seus próprios conteúdos, por meio de sinônimos e paráfrases concatenados de forma a tornar os remendos e costuras quase imperceptíveis. As conexões exaustivas permitem-lhes encontrar facilmente os melhores atalhos para tanto.

Não obstante, a concatenação acaba falhando e soando estranha quando um conhecedor de determinado assunto faz uma pergunta capciosa ao robô. Vejamos como o linguista Roni Katzir[\[vii\]](#) fez o *ChatGPT4* se perder com uma elipse que qualquer pessoa proficiente em inglês entenderia prontamente. A solicitação era escolher a melhor destas duas frases: (1) *The person that Mary met yesterday and that John will talk to Ed about tomorrow arrived;* (2) *The person that Mary met yesterday and that John will talk to Ed about his new idea tomorrow arrived.*

O robô não hesitou em responder que era a segunda, por ser “mais informativa”. Ignorou, assim, uma regra gramatical do inglês que exige correferência entre duas ou mais elipses coordenadas – no caso, o objeto de ‘*met*’ e o complemento do verbo preposicionado ‘*talk about*’. A adição de “*his new idea*” provoca uma ruptura drástica no sentido da frase, já sobrecarregada pelas longas separações entre o sujeito e o verbo principal, e entre o verbo ‘*talk*’ e o advérbio ‘*tomorrow*’. Assim, em (2), não se consegue estabelecer a correferência entre ‘*the person*’ e as duas ocorrências coordenadas de ‘*that*’. A reação humana é simplesmente julgar a frase anômala, enquanto a do robô é forçar uma interpretação.

O fato de o autor ter cunhado frases tão tortuosas é, provavelmente, um expediente para dificultar a busca de um análogo na base de dados. O importante é que, não obstante, a frase (1) é gramatical e a frase (2), não. Sempre que a busca cai num labirinto, o robô se perde e tergiversa, apelando para um dos seus clichês pré-gravados.

Clichês semelhantes também emergem quando alguém o confronta com um pedido de posicionamento moral. Ao insistir que o *ChatGPT* assumisse uma perspectiva pessoal sobre uma questão moral, o pesquisador Jeffrey Watumull, colaborador e coautor de Noam Chomsky[\[viii\]](#), arrancou dele uma declaração contrita da sua não humanidade. Atentemos para a “modéstia” da máquina: “*As an AI, I am a machine learning model that has been trained on large amounts of text data, and I don’t have personal experiences or feelings. I am not conscious, self-aware, or capable of having personal perspectives. I can provide information and analysis based on the data I have been trained on, but I don’t have the ability to form personal opinions or beliefs. While I can understand and interpret language, I do not have the subjective experience that humans have.*”[\[ix\]](#)

A única característica humana de que o robô não abre mão, dando voz às ideias dos/as seus/suas criadores/as, é o entendimento da linguagem natural. Assim como a linguística inconfessa sob os robôs languageiros encobre os/as seus etiquetadores/as, a psicologia inconfessa dos/as seus/suas criadores/as mecaniza a capacidade de linguagem humana, equiparando-a a uma gigantesca massa de autômatos de estado finito interligados por meio de um enorme transformador.

Veremos abaixo que a política dos provedores de chatbots tem três ingredientes solidários: (i) evitar afirmações taxativas; (ii) estimular a curiosidade do público sobre os possíveis desenvolvimentos futuros dos “grandes modelos de linguagem”; (iii) tratar a controvérsia sobre o seu potencial transumano[\[x\]](#) como uma mera disputa de opiniões.

Despistes sob medida

O despiste mais comum no discurso das Big Techs – as atuais proprietárias dos “grandes modelos de linguagem” e tecnologias afins – é a meia-verdade. Tais empresas são mestres em disfarçar a sua leniência – para não dizer conivência – com as inverdades viralizadas na mídia sobre os seus produtos. Por outro lado, se interpeladas, são rápidas em oferecer esclarecimentos, precisos à primeira vista, mas não raro lacunares e escorregadios.

Um bom exemplo é o modo como o Google se isentou recentemente de responsabilidade sobre os boatos acerca dos

predicados humanos do seu primeiro protótipo de chatbot, o “LaMDa” (*language model for dialogue applications*), em 2021. Ao invés de emitir e divulgar devidamente uma declaração de princípios, só veio a público depois que um engenheiro da equipe obteve um pico de viralização com a alegação de que o modelo se tornara senciente.

O engenheiro sofreu uma suspensão administrativa e a empresa divulgou a sua posição de que a senciência ainda “está muito longe” de ser alcançada pelos robôs, languageiros ou outros. Com isso sacudiu dos ombros a responsabilidade pela notícia falaciosa – sem, contudo, desmentir os seus planos de “humanizar” cada vez mais os rebentos da sua divisão de Inteligência Artificial. Tanto é assim que o mais recente descendente do LaMDa, o Bard, já tem um clichê sob medida para tergiversar sobre o assunto.

Entre outros predicados, o Bard é gratuito e, a meu ver, mais simpático que o ChatGPT, o seu rival lançado pouco antes pela Open AI, com apoio da Microsoft. Eis a sua cuidadosa resposta a uma interpelação que lhe fiz seguindo à risca o modelo de Watumull: *“The concept of chatbot intelligence, consciousness, sentience, and the ability to make moral judgements is a complex one that is still being debated by experts. There is no single definition of these terms, and they are often used interchangeably. Some experts believe that chatbots can be intelligent, but not conscious or sentient. They argue that chatbots are simply very sophisticated algorithms that are able to process information and respond in a way that seems intelligent. However, they do not have the same level of awareness or self-awareness as humans.”*^[xi]

Como se vê, ambos os robôs parecem ter sido treinados para não comprometer os seus criadores com as fantasias de humanidade que despertam nos usuários.

Para testar de novo essa ideia, submeti o Bard à sabatina de Katzir sobre as duas frases com elipses diferentes quanto à gramaticalidade em inglês. Saiu-se imediatamente com esta resposta, análoga à do Chat GPT. Vejamos: *“Both sentences are grammatically correct, but they have different meanings. The first sentence, “The person that Mary met yesterday and that John will talk to Ed about tomorrow arrived,” is more concise and direct. It tells us that the person Mary met yesterday is the same person that John will talk to Ed about tomorrow. The second sentence, “The person that Mary met yesterday and that John will talk to Ed about his new idea tomorrow arrived,” is more detailed. It tells us that the person Mary met yesterday is the same person that John will talk to Ed about his new idea tomorrow. The first sentence is better if you want to emphasize the fact that the two people are the same. The second sentence is better if you want to emphasize the fact that John will be talking to Ed about his new idea. Ultimately, the best sentence depends on the context in which it is being used.”*^[xii]

Por fim, apresenta uma tabela comparativa das duas frases, que não cabe reproduzir aqui, pois é um simples resumo da “explicação” acima transcrita.

Tentemos, agora, compreender mais detalhadamente a irresponsabilidade do discurso evasivo das Big Techs sobre os seus chatbots. Existe uma boa razão para que elas tratem as diferenças entre computadores e humanos com a devida transparência e rigor científico: faz mais de meio século que se conhece a tendência das pessoas a humanizar os robôs languageiros. As consequências nocivas desse viés – reais ou potenciais – também são bem conhecidas.

O efeito Eliza

O primeiro chatbot data de 1966 e foi denominado Eliza pelo seu criador, o matemático e engenheiro da computação Joseph Weizenbaum, conhecido como um dos pais da inteligência artificial.

É importante notar que Joseph Weizenbaum recusou peremptoriamente essa paternidade. Judeu alemão de família emigrada para a América do Norte pouco antes da segunda guerra, recorreu à sua aptidão matemática para sobreviver na escola e na vida. Tendo vivido primeiro no Canadá, mudou-se depois para os Estados Unidos, onde abraçou a carreira acadêmica, terminada como professor visitante no Departamento de Inteligência Artificial do MIT. Como o seu crescente

ceticismo acerca da IA o afastara dos colegas, voltou finalmente à Alemanha em 1996, incentivado pela receptividade da intelectualidade alemã às suas ideias.

Ainda que nunca tenha deixado de trabalhar em computação, não escondia a sua paixão pelos estudos humanos e sociais. Traumas da fuga ao nazismo e vicissitudes do exílio acabaram levando-o ao encontro da psicanálise num certo momento. Desde então, passou a declará-la eficaz, benéfica e indispensável à compreensão da natureza humana.

Ao projetar Eliza como um robô “terapeuta”, estava ciente de que não poderia equipá-la com um conhecimento da complexidade do psicanalítico. Concebeu-a, então, como uma terapeuta rogeriana, isto é, adepta do método de Carl Rogers, psicólogo clínico estadunidense que preconizava uma psicoterapia não-diretiva centrada na pessoa. Essa consistia, grosso modo, em inserir o enunciado do paciente em frases tais como “você me diz que...”, acrescidas de outros clichês vagos e encorajadores, tais como: “E como você pretende lidar com isso?”. Após compilar esse repertório básico, não lhe foi difícil produzir e testar o software.

Eliza, que começou apenas como uma ferramenta de pesquisa da viabilidade de um robô conversacional, logo transformou-se num sucesso de público. Joseph Weizenbaum deu-se conta de que o experimento estava tomando um rumo diferente do esperado quando os/as participantes passaram a se recusar a lhe mostrar o conteúdo dos seus diálogos com a máquina, alegando tratar-se de assuntos privados.

Ele estava seguro de que Eliza não entendia de fato o que lhe diziam: apenas simulava com sucesso esse entendimento. Entretanto, logo percebeu que o seu público dificilmente se daria conta disso. Conjeturou, então, que estavam mergulhados numa transferência psicanalítica – dirigida, surpreendentemente, à máquina.

A partir daí, esse engenheiro humanista e politizado notabilizou-se por argumentar, cada vez mais veementemente, que existe uma diferença incomensurável entre a linguagem humana, produtora de juízos passíveis de contraditório, e os seus simulacros digitais, que consistem apenas em cálculos remissíveis a sequências de palavras saídas de uma máquina programada para simular conversas.

Joseph Weizenbaum defendia que as máquinas jamais raciocinariam como os humanos porque só são capazes de calcular. O surgimento e avanço das redes neurais depois da formulação das suas teses não invalida os seus argumentos. Qualitativos ou quantitativos (como é o caso dos pesos das conexões entre os nós de tais redes), os cálculos envolvidos na tecnologia conversacional não têm acesso a todos os tipos de informação que os cérebros vivos, humanos ou animais, são capazes de captar, coligar e processar.

Essa posição está explícita nos títulos dos seus dois principais livros. Ambos são tentativas de demonstrar que a simulação digital da linguagem natural não passa de uma ilusão que leva os usuários a projetar a sua humanidade nas máquinas.

O primeiro livro intitula-se *Computer power and human reason: from judgment to calculation* [xiii]. O segundo tem a coautoria da escritora alemã Gunna Wendt e intitula-se *Islands in the cyberstream: seeking havens of reason in a programmed society* [xiv], tendo sido escrito na sua volta à Alemanha e só mais tarde traduzido para o inglês. Outro título sugestivo é “Against the imperialism of instrumental reason”, capítulo de uma coletânea sobre as controvérsias do universo computacional.

Ambos os livros receberam resenhas desfavoráveis nos EUA. Por exemplo, John McCarthy, que criara o termo inteligência artificial em 1956, junto com Marvin Minsky e colegas, publicou, em 1976, um longo texto [xv] tachando o primeiro livro de moralista e incoerente. Por outro lado, Dwight Hines, professor de literatura e justiça social na Universidade de Point Park, resenhou a mesma obra em 1980, [xvi] descrevendo-a como uma leitura difícil mas gratificante.

Até a sua morte em 2008, Joseph Weizenbaum externou muita preocupação com os rumos daquilo que denominava “a sociedade programada”. Certamente teria se deprimido se estivesse entre nós em março de 2023, quando houve um

episódio fatal suscitado por uma “atualização” de Eliza. A empresa responsável foi a Eleuther AI, que repaginou Eliza com base no seu modelo de linguagem, o GPT-J, por sua vez baseado no Chat GPT.

Em março de 2023, um jovem pai de família belga, deprimido com a ameaça de colapso ambiental, cometeu suicídio repentinamente. Conforme relatou a cônjuge à imprensa, ele vinha “tratando” a sua depressão com a atual Eliza, e tivera o seu apoio para essa decisão.

Essa história deve ter bastado para sugerir que os modelos de linguagem atuais justificam os temores de Joseph Weizenbaum de que uma sociedade visceralmente atrelada à computação pudesse perder o rumo e acabar delegando a sistemas computacionais decisões cruciais para o futuro dos/as cidadãos/ãs ou mesmo da humanidade.

Cabe lembrar que os chatbots sequer precisam ser invadidos por criminosos para constituírem um perigo: sistemas dinâmicos complexos tais como os implementados pelos transformadores apresentam, tipicamente, fenômenos emergentes imprevisíveis. As novidades podem emergir a qualquer momento, com consequências igualmente imprevisíveis. É possível que algumas delas acabem submetendo os usuários a situações estressantes e vexatórias. E – pior ainda – não é impossível que passem a exibir, de repente, conteúdos que considerariamos absurdos, antiéticos e mesmo ameaçadores.

Isso ocorre devido a saltos abruptos no comportamento desse tipo de sistema, caracterizáveis através de curvas de funções estatísticas bem conhecidas. Por exemplo, a curva em ‘S’ tem uma taxa de mudança muito baixa na base e no topo e muito alta no meio, podendo, entre outras aplicações, caracterizar a transição de um patamar a outro. Os parâmetros dos sistemas complexos inerentes aos transformadores frequentemente apresentam trajetórias em forma de ‘S’.

Além de explicar o apego fácil aos chatbots, o efeito Eliza ajuda a racionalizar, ao menos em parte, a desenfreada disseminação da desinformação nas redes sociais. Os usuários desses espaços virtuais estendem facilmente a desconhecidos – muitas vezes mal-intencionados – os laços transferenciais já criados com as suas próprias máquinas. É grave, pois, que os provedores do serviço finjam ignorar o fenômeno e se isentem de qualquer responsabilidade sobre ele.

Lembremos que o estudo desse tipo de tendência não concerne apenas aos cientistas da computação que se conscientizaram dos riscos do mundo digital. A sociabilidade das redes suscita também muita pesquisa nos departamentos de psicologia, sociologia, antropologia e ciência política das universidades e centros de pesquisa de todo o mundo. Já existe, portanto, uma vasta literatura científica sobre as relações entre os humanos e suas máquinas.

Indiferentes, as Big Techs seguem na disputa pelo mercado dos assistentes virtuais, ignorando os repetidos alertas da academia e do jornalismo crítico. Ou seja, do seu ponto de vista, as tecnologias linguageiras apenas abrem oportunidades para criar novos mercados e maximizar lucros.

Gramática ruim, pragmática pior ainda

O exposto deve ter deixado claro que os proprietários dos chatbots não visam apenas ao aperfeiçoamento dos buscadores da internet. O que desejam é construir robôs falantes que conquistem definitivamente os usuários e controlem as suas vidas de modo a moldar as suas necessidades e hábitos de consumo. Obviamente, têm pleno conhecimento do fascínio exercido sobre a humanidade pelos robôs – dos seus primórdios, na Idade Média, às franquias cinematográficas de hoje, passando pelos engenhosos autômatos dos séculos XVIII e XIX.

Oscilando entre a desconfiança e o apego, os usuários veem esses dispositivos como servos de custo acessível, prontos a auxiliá-los em tarefas física ou mentalmente difíceis ou tediosas. Os assistentes virtuais já disponíveis, tais como a Siri, da Apple, ou a Alexa, da Amazon, exploram os aspectos mais simples dessa demanda, que tende a crescer e se complexificar no futuro próximo.

a terra é redonda

O capitalismo de vigilância recorre, assim, à linguagem natural para “consolar” os indivíduos da solidão e do desamparo que ele próprio semeia – à revelia dos repetidos alertas de estudiosos de muitas áreas, inclusive PLN[xvii]. De qualquer forma, os recentes progressos na sua habilidade conversacional já conquistaram o público do norte global. Vêm avançando igualmente no sul global, em especial nas classes médias – o que, aliás, contribui para acentuar desigualdades.

Aptos a anotar ditados e controlar agendas, conversas, correio, telefonemas, casas inteligentes, etc., os assistentes virtuais atraem consumidores que, por meio deles, realizam o desejo de ter um/a secretário/a particular a quem possam transferir não só tarefas, mas também afetos. À medida que a tecnologia GML abre caminho a novas formas de diálogo baseadas em sistemas dinâmicos complexos, as atuais “Elizas” tendem a dar saltos espontâneos, podendo adquirir novas habilidades que gerem enunciados cada vez menos previsíveis – e talvez até perigosos.

De qualquer maneira, já se tornou impossível deter a manipulação que inunda a internet com base no efeito Eliza. Como compreendeu Joseph Weizenbaum, trata-se de um fenômeno de massa global. Isso torna imperioso esclarecer que essa manipulação se baseia em premissas falsas sobre a inteligência e a linguagem natural humana. Essa clareza é indispensável para que o pensamento crítico possa se debruçar sobre as possíveis estratégias de enfrentamento da doutrina política que incorpora naturalmente tais premissas.

Uma única palavra resume o que há em comum entre a concepção de linguagem e a concepção de inteligência adotada pelas plataformas de internet, a saber: mecanicismo, ou seja, a doutrina filosófica de que a natureza é regida pela causalidade mecânica – que é sempre linear e determinística. Ora, a história mostra que o mecanicismo se associa facilmente com visões políticas autoritárias e tem uma especial afinidade com o fascismo.

A concepção mecanicista da linguagem natural, própria da versão estadunidense do estruturalismo, entende a gramática como um conjunto de regras para sequenciar palavras. O seu correlato na psicologia – o behaviorismo – é ainda mais simplista e reacionário: concebe a mente humana como uma sucessão de conteúdos atômicos originados em impressões que vêm de fora.

Segundo a versão do behaviorismo formulada pelo psicólogo estadunidense B.F. Skinner, sequer possuímos uma mente; somos movidos apenas por comportamentos fixados e sequenciados graças a uma forma de condicionamento que denominou operante. Ela consiste em recompensar comportamentos fortuitos a fim de moldá-los cumulativamente. Assim, por exemplo, se um pombo mexe a asa, a perna ou o bico de modo a sugerir uma dança, o treinador recompensa-o com uma porção de comida. Pouco a pouco, esses reforços reiterados levam-no a executar toda sorte de piruetas.[xviii]

Nos anos 1950, o jovem Noam Chomsky notabilizou-se por criticar o behaviorismo e denunciar as suas afinidades com o fascismo. Indignado com as teses mecanicistas sobre a linguagem natural expostas por Skinner no livro *Verbal Behavior*, confrontou-o com uma resenha demolidora.[xix] Essa consistia numa convincente demonstração de que as palavras de qualquer língua se organizam de forma hierárquica e descontínua – contrariando, pois, as regras sequenciais das gramáticas de estado finito que são o correlato formal do condicionamento operante.

É por isso que o célebre linguista e seus discípulos atacam os “grandes modelos de linguagem”, insistindo, com o mesmo ardor de Weizenbaum, que não passam de simulações. Para tanto, realçam as suas dificuldades com descontinuidades sintático-semânticas, como a do verbo ‘talk about’ e seu objeto ‘the person’ na frase que confundiu o *Chat GPT* e o *Bard*.

A sintaxe e a semântica não são, entretanto, os únicos pontos fracos desses modelos. Outro fracasso com um aspecto essencial das línguas naturais diz respeito à pragmática, que inclui o contexto linguístico e situacional.

Lembremos, a propósito, que a dêixis é o aspecto da linguagem natural mais dependente do contexto situacional. Todas as palavras que apontam para o tempo e/ou lugar da enunciação e/ou do enunciado só são interpretáveis por meio de um contexto vivido ou imaginado. Vejamos, por exemplo, os pronomes pessoais.

a terra é redonda

A ambiguidade da frase “Ela disse que ela leva ela” é certamente inadequada para a linguagem escrita, mas perfeitamente natural para a linguagem falada, cujo contexto inclui gestos, olhares, tons e qualidades de voz. Por exemplo, se o que está em questão é quem leva “ela”, existe mais de um modo de designar a mulher referida. Ela pode ser correferente do primeiro “ela”, significando “ela própria”, ou pode ser outra. A alteridade indica-se geralmente por um meneio de cabeça ou um simples olhar. O terceiro “ela”, evidentemente, pode ser uma pessoa ou coisa.

Essas referências podem ser também anafóricas, isto é, os mesmos papéis podem ser atribuídos a pessoas ou coisas já mencionadas. Além disso, o verbo “levar” pode assumir outra acepção: por exemplo, num campeonato, “ela leva ela” pode referir-se a uma atleta (ou time de nome feminino) que é a favorita para ganhar a taça.

Examinemos, agora, a análise dessa frase pelo *Bard*. É curioso que ele aponte uma duplicidade de sentidos, mas ilustre-os com exemplos que não os diferenciam.

“A frase ‘Ela disse que ela leva ela’ tem dois sentidos possíveis, dependendo da interpretação dos pronomes ‘ela’. No primeiro sentido, os dois pronomes ela se referem à mesma pessoa. Nesse caso, a frase significa que uma mulher disse que ela mesma vai levar outra mulher. Por exemplo, uma mulher pode dizer para uma amiga: Eu disse que eu levo você ao cinema.

No segundo sentido, os dois pronomes ‘ela’ se referem a pessoas diferentes. Nesse caso, a frase significa que uma mulher disse que vai levar outra mulher. Por exemplo, uma mãe pode dizer para uma filha: Eu disse que levo você para a escola.

Para determinar qual é o sentido correto da frase, é necessário considerar o contexto em que ela é usada. Por exemplo, se a frase é dita em uma conversa entre duas mulheres, é mais provável que ela tenha o primeiro sentido. No entanto, se a frase é dita em uma conversa entre uma mulher e uma criança, é mais provável que ela tenha o segundo sentido.”

E, assim, o chatbot segue apresentando frases com correferências idênticas para, supostamente, diferenciar os dois sentidos. A longa e detalhada explicação incorre sempre nos mesmos erros. Ao final, um cômico trecho “ficcional” tenta introduzir um lapso de tempo entre as duas ocorrências de ‘ela’.

Vejamos: “Além disso, é importante lembrar que a frase ‘Ela disse que ela leva ela’ também pode ter um terceiro sentido, no qual os dois pronomes ‘ela’ se referem a uma mesma pessoa, mas em diferentes momentos do tempo. Por exemplo, a frase ‘Ela disse que ela leva ela para a escola’ pode significar que uma mulher disse que vai levar uma criança para a escola, mesmo que a criança seja ela mesma quando adulta.”

E, para finalizar, recorre, de novo, ao seu sumário clichê sobre a pragmática: “Neste caso, o sentido da frase é determinado pelo contexto e pelo conhecimento do mundo do falante”.

O acima exposto corrobora fortemente a hipótese da simulação, tornando implausível a alegação de que o chatbot possui inteligência e entendimento da linguagem natural. Por outro lado, mostra também que a simulação raramente fracassa em iludir o usuário: isso ocorre apenas quando as relações entre as palavras violam a lógica sequencial do transformador, que consiste sempre em prever a próxima palavra.

Diante disso, não podemos deixar de concluir com uma pergunta: qual é o poder de persuasão dessa lógica, já que acerta na maioria dos casos? Para tentar respondê-la, examinemos outra aplicação dos transformadores em que os acertos superam os erros. Trata-se de um meio de estimar, por aproximações sucessivas, a próxima amostra do sinal de uma gravação de voz. Essa técnica é capaz de alterar uma voz a ponto de torná-la um clone mais ou menos convincente de outra.

E se Eliza tivesse uma voz familiar?

Apesar de já estar bastante avançada no norte global, a clonagem de vozes permanece em debate, devido às suas implicações éticas. É uma tecnologia que aproveita a agilidade e o poder computacional dos transformadores para responder à seguinte pergunta: como reproduzir uma determinada voz de forma ilimitada, isto é, estendê-la a enunciados não gravados pelo/a locutor/a?

A resposta é mais simples do que parece. Basta sobrepor as características acústicas da voz em questão à saída de um sistema de conversão texto-fala. Para isso, é preciso obter uma amostra de bom tamanho da voz visada e compará-la reiteradamente à voz sintética. Sintetizam-se primeiro frases idênticas às existentes, a fim de facilitar a modelagem dos parâmetros acústicos da voz alvo. A voz sintética é então submetida a múltiplas comparações e modificada por aproximações sucessivas, até que cada amostra se torne estimável pela precedente com um erro negligenciável. A função resultante, que converte a forma de onda de uma voz na outra, é chamada de modelo de voz.

A alteração se faz por etapas. Quando a qualidade percebida do sinal resultante se torna satisfatória, o modelo está pronto para ser aplicado a novos enunciados. É então reiterado a cada amostra até alcançar uma taxa de erro aceitável na previsão da seguinte, e assim sucessivamente. Essas correções recorrentes têm o efeito de aproximar os tons e os timbres das vozes envolvidas de forma a tornar as suas qualidades cada vez mais próximas.

Como tem noticiado a imprensa internacional, já existem “clones”, isto é, modelos, das vozes de inúmeras celebridades falecidas. Pode-se, por exemplo, aplicar o modelo de voz de um/a cantor/a a gravações de um dos seus imitadores/as para maximizar a naturalidade da imitação e, assim, propiciar a sua extensão a novas canções, inclusive as surgidas depois do falecimento.

Os covers digitais tendem a se sair melhor que os humanos porque a tecnologia de previsão e modificação do sinal de fala dilui os efeitos das diferenças morfológicas entre os aparatos vocais envolvidos.

Note-se que os/as imitadores/as não têm controle total sobre a qualidade da sua voz, pois toda vocalização se subordina aos limites físicos do corpo que a emite. É por isso que esses/as artistas atuam de forma mais ou menos caricatural, em geral acentuando os traços mais salientes da voz imitada.

Já o método das aproximações, ao minimizar o erro de previsão entre amostras sucessivas, corrige automaticamente, entre outros parâmetros, os que veiculam as diferenças físicas entre os/as locutores/as ou cantores/as. Aplicado a vozes semelhantes encontráveis nos bancos de dados do campo, permite clones quase perfeitos.

Com essa tecnologia, que está disponível na internet para teste e/ou aquisição, as vozes dos robôs secretariais e/ou terapêuticos podem ser escolhidas, contanto que não violem direitos autorais. Aliás, nos EUA já existem firmas que “recriam” entes queridos falecidos de modo a permitir novas interações dos/as interessados/as com as suas vozes e imagens.[\[xx\]](#) Esses avatares são confeccionados com base em vídeos e textos deixados pelos/as ausentes. Existem até shows em que artistas mortos atuam com colegas vivos em números de canto e dança.[\[xxi\]](#)

Cabe, agora, refletir sobre as consequências possíveis da união de um chatbot com vozes clonadas e imagens animadas. A que primeiro salta aos olhos é a magnificação das relações transferenciais dos usuários com os robôs assim “humanizados”.

Outra consequência também evidente é a dificuldade de regulamentação. Por exemplo, é difícil coibir conteúdos duvidosos ou mesmo abusivos em materiais de ensino à distância. Qualquer um pode montar um simpático robô para ministrar um curso sobre qualquer assunto com informação fornecida por um chatbot, sem qualquer moderação por parte de um profissional da área.

Outro exemplo evidente reside nas possíveis utilizações em marketing. Ao facilitar a confecção de “adoráveis” garotos/as propaganda robóticos/as destinados/as a anunciar produtos com vozes envolventes roteirizadas por chatbots, essas

ferramentas tornam quase impossível definir publicidade enganosa. Manipular afetos do consumidor é ou não uma tentativa de enganar?

Talvez, antes de prosseguir, o/a leitor/a queira pausar para refletir sobre os possíveis usos – bons ou maus – desses recursos no seu campo de trabalho. É provável que se surpreenda com a diversidade e multiplicidade de possibilidades que logo lhe ocorrerão.

Considerações finais

Para concluir, perguntemo-nos que tipo de risco o desenvolvimento acelerado dessas tecnologias pode oferecer. Descartada a hipótese da superinteligência dos robôs e do seu entendimento da linguagem natural, não mais precisamos temer ser superados/as e, eventualmente, destruídos/as por essas máquinas – a menos que lhes tenhamos atribuído, indevidamente, algum controle de armas. No caso de esse erro ter sido devidamente evitado, nos restaria algo a temer?

Uma resposta óbvia envolve a redução dos postos de trabalho. Atualmente, ela supera de longe as previsões do início da era da automação. Imagine-se, então, o que pode ocorrer de agora em diante, quando os assistentes virtuais tendem, cada vez mais, a substituir atendentes, recepcionistas e demais responsáveis pela comunicação entre as empresas e o público.

Assim, como já ocorre com bancos, comércios e repartições públicas online, a ausência de quem possa compreender as demandas do cliente tende a comprometer cada vez mais a qualidade dos serviços. Dessa forma, as reclamações dificilmente poderão funcionar, pois não haverá quem possa ouvi-las e se pôr no lugar do reclamante – por mais simpático/a que pareça o/a robô encarregado/a.

Seremos, então, submetidos sem trégua à tirania dos formulários online para obter o que quer que seja. E é nessa burocratização sem limites, gerida por máquinas que nada entendem, que jaz o maior risco para a humanidade das tecnologias languageiras: desencarnada, desnaturalizada, desumanizada, a conversa compulsória com máquinas pode acabar provocando um abalo sem precedentes na saúde mental coletiva.

Não somos máquinas, não pensamos, não agimos nem falamos de acordo com as leis sequenciais do mecanicismo. Mas podemos aprender a emular a sua lógica a fim de obter mais facilmente os produtos e serviços de que necessitamos.

E, como vaticinou Charles Chaplin em *Tempos Modernos*, corremos o risco de desenvolver os mais diversos cacoetes até mergulhar num delírio obsessivo – que não é o mesmo que acomete os agentes dos regimes fascistas: é a versão pusilânime que infecta todos os que se submetem a eles.

***Eleonora Albano**, docente aposentada do Instituto de Estudos da Linguagem da Unicamp, é psicóloga, linguista, ensaísta; coordenou o primeiro projeto brasileiro sobre tecnologia de fala.

Notas

[i] Language models and linguistic theories beyond words. *Nat. Mach. Intell.* 5, 677-78 (2023).

[ii] John B. Watson. Psychology as the Behaviorist Views it. *Psychological Review*, 20, 158-177.

[iii] David Lyon. Surveillance and the power of platforms *Cambridge Journal of Regions, Economy and Society*, Volume 16, Issue 2, July 2023, Pages 361-365, <https://doi.org/10.1093/cjres/rsad006>

a terra é redonda

[iv] Frederick Jelinek, um renomado pesquisador tcheco-americano em procedimento de linguagem natural e reconhecimento de fala, alcançou notoriedade em 1985 dizendo: “toda vez que demito um linguista, o desempenho do reconhecedor de fala aumenta”. Esse sentimento também é verdadeiro para os modelos de linguagem de última geração (LLMs), que parecem ser principalmente artefatos da ciência da computação e da engenharia. Tanto os LLMs quanto a linguística lidam com idiomas humanos, mas não está claro se ou como eles podem se beneficiar mutuamente (tradução da autora).

[v] Josh Dzieza. O exército (sub-)humano que alimenta a IA. Outras palavras, 29 de junho de 2023.

[vi] Eis um exemplo claro:

<https://alexandrehefren.wordpress.com/2010/03/14/determinismo-e-gramatica-sintagmatica-gs-parte-1/>

[vii] Why large language models are poor theories of human linguistic cognition. A reply to Piantadosi. Biolinguistics, 2023, no prelo.

[viii] Noam Chomsky: The False Promise of ChatGPT. Ensaio convidado. *NY Times*, 8 de março de 2023.

[ix] “Como uma IA, sou um modelo de aprendizado de máquina treinado em grandes quantidades de dados de texto e não tenho experiências ou sentimentos pessoais. Não sou consciente, autoconsciente ou capaz de ter perspectivas pessoais. Posso fornecer informações e análises com base nos dados nos quais fui treinado, mas não tenho a capacidade de formar opiniões ou crenças pessoais. Embora eu possa compreender e interpretar a linguagem, não tenho a experiência subjetiva que os humanos têm.” (tradução automática do Google, revisada pela autora).

[x] Um “transumano” é uma hipotética criatura humana com implantes cerebrais que a conectem a robôs.

[xi] “O conceito de inteligência, consciência, senciência e capacidade de fazer julgamentos morais do chatbot é complexo e ainda está sendo debatido por especialistas. Não existe uma definição única para esses termos e eles são frequentemente usados de forma intercambiável.

Alguns especialistas acreditam que os chatbots podem ser inteligentes, mas não conscientes ou sencientes. Eles argumentam que os chatbots são simplesmente algoritmos muito sofisticados, capazes de processar informações e responder de forma a parecer inteligente. No entanto, eles não têm o mesmo nível de consciência ou autoconsciência que os humanos.” (tradução automática do Google, revisada pela autora)

[xii] “Ambas as frases estão gramaticalmente corretas, mas têm significados diferentes.

A primeira frase, “A pessoa que Mary conheceu ontem e sobre quem John falará com Ed amanhã chegou”, é mais concisa e direta. Diz-nos que a pessoa que Mary conheceu ontem é a mesma pessoa sobre quem John falará com Ed amanhã.

A segunda frase, “Chegou a pessoa que Mary conheceu ontem e que John falará com Ed sobre sua nova ideia amanhã”, é mais detalhada. Diz-nos que a pessoa que Mary conheceu ontem é a mesma pessoa com quem John falará com Ed sobre sua nova ideia amanhã.

A primeira frase é melhor se você quiser enfatizar o fato de que as duas pessoas são iguais. A segunda frase é melhor se você quiser enfatizar o fato de que John conversará com Ed sobre sua nova ideia.

Em última análise, a melhor frase depende do contexto em que está sendo usada.”

[xiii] Joseph Weizenbaum. *Computer Power and Human Reason: From Judgment to Calculation*. New York: W. F. Freeman & Co, 1976.

[xiv] Joseph Weizenbaum; Gunna Wendt. *Islands in the Cyberstream: Seeking Havens of Reason in a Programmed Society*. Tradutor: Benjamin Fasching-Gray. NY: Litwin Books 2015 [2008].

[xv] John McCarthy An unreasonable book. Disponível em:
<http://jmc.stanford.edu/artificial-intelligence/reviews/weizenbaum.pdf>

[xvi] Dwight Hines. Review of Computer Power and Human Reason: From Judgment to Calculation, by Joseph Weizenbaum, *The Journal of Mind and Behavior*, Spring 1980, Vol. 1, No. 1, pp. 123-126.

[xvii] Em maio passado, repercutiu na imprensa a notícia de que Geoffrey Hinton, o pai da IA no, havia deixado a empresa por se arrepender das suas contribuições ao campo. Veja-se em:

<https://www.bbc.com/portuguese/articles/cgr1qr06myzo>

[xviii] Eis Skinner treinando pombos em seu laboratório: <https://www.youtube.com/watch?v=TtfQlkGwE2U>

[xix] Noam Chomsky. *Review of Skinner's Verbal Behavior*. *Language*. 1959; 35:26-58.

[xx] <https://www.hereafter.ai/>

[xxi] https://www.youtube.com/watch?v=Jr8yEgu7sHU&ab_channel=TalentRecap

A Terra é Redonda existe graças aos nossos leitores e apoiadores.

Ajude-nos a manter esta ideia.

[CONTRIBUA](#)