

A inteligência artificial e o novo patamar da interação humano-máquina

Gustavo Rick Amaral¹

Fernando Xavier²

Resumo: O artigo apresenta alguns dos principais aspectos da revolução pela qual o campo da Inteligência Artificial (IA) está passando – especificamente no subcampo do processamento de linguagem natural – devido ao aprendizado de máquina. O objetivo é demonstrar como grandes modelos de linguagem – como aqueles que sustentam o ChatGPT – são passos decisivos não apenas na direção de uma IAs mais fortes, mas também para dentro e além de uma nova fronteira tecnológica: as máquinas estão cruzando os portões que dão acesso ao mundo cultural humano. A competência semiótica para produzir e entender diversos tipos de signos (música, imagem, linguagem verbal, entre outros tipos) atribuída às máquinas coloca a interação humano-máquina para operar em um novo nível. Neste artigo, propomos um esquema para analisar estes recentes avanços da IA dentro de um contexto mais amplo do desenvolvimento tecnológico de máquinas. A análises esquemática nos permite explorar, ao final deste breve ensaio, algumas importantes questões a respeito dos riscos e benefícios deste novo nível de interação humano-máquina.

Palavras-chave: Inteligência artificial, aprendizado de máquina, grandes modelos de linguagem, interação humano-máquina, cultura, competência semiótica.

1. Semioticista e pesquisador do Centro Internacional de Estudos Peirceanos (CIEP/PUC-SP) e do grupo de pesquisa Transobjeto (TIDD-PUC-SP); doutor pelo Programa de Estudos Pós-Graduados em Tecnologias da Inteligência e Design Digital (TIDD) da PUC-SP (2014); professor dos cursos de Comunicação Social da Universidade Anhembi-Morumbi. CV Lattes: lattes.cnpq.br/3463780553418311. ORCID: orcid.org/0000-0002-0063-6119. E-mail: gustrick@gmail.com.

2. É cientista da computação e pesquisador do grupo de estudos em Saúde Planetária do Instituto de Estudos Avançados da USP, com pesquisas relacionadas à dengue e análise de dados de redes sociais. Doutorando pelo Programa de Engenharia Elétrica da Escola Politécnica da USP. Atualmente, é engenheiro de dados na TOTVS. E-mail: fernando.xavier@gmail.com.

Artificial Intelligence and the new level of human-machine interaction

Abstract: The paper presents some of the main aspects the undergoing revolution Artificial Intelligence (AI) is experiencing, especially in natural language processing (NLP) due to machine learning. The objective is to demonstrate that large language models – like those powering ChatGPT – are decisive steps not only towards a stronger AI but towards a new technological frontier: machines are striding through the gates that give access the human world of culture. The semiotic competence to produce and understand several modalities of signs (music, image, verbal language, etc.) of machines raises the human-machine interaction to a new level. The paper proposes a schema to analyze such recent advances in AI in the broader context of the technological evolution of machines. The schematic analysis serves to explore important questions concerning the risks and benefits of this new level of human-machine interaction.

Keywords: Artificial Intelligence, Machine Learning, large language models, human-machine interaction, culture, semiotic competence.

A Inteligência Artificial deixou de ser uma promessa de um *futuro tecnológico sempre distante* para se fazer concreta em nosso cotidiano esparramando-se por uma infinidade de dispositivos: celulares, notebooks, desktops, carros, relógios, geladeiras, etc. Esta é uma tecnologia que esteve projetada no horizonte de nosso desenvolvimento tecnológico pelos últimos 70 ou 80 anos, desde os primeiros passos da Era da Computação. Esta é uma tecnologia que foi tantas vezes sonhada pelos nossos mais criativos ficcionistas, persistentemente perseguida pelos nossos mais inventivos cientistas e pesquisadores, avidamente almejada pelos mais poderosos empresários, investidores e tecno-magnatas que parecia que este dia nunca iria chegar. Vivemos nas últimas décadas, uma ansiedade pré-Inteligência Artificial.

O ChatGPT, *chatbot* lançado pela empresa OpenAI ao final do ano de 2022, talvez seja a primeira encarnação de uma Inteligência Artificial realmente capaz de convencer membros do grande público de sua *inteligência*. É provavelmente a primeira capaz de ter aquele efeito “ah, agora, sim”. O fascínio das pessoas pelo chatbot e pela tecnologia subjacente que o anima se refletiu nos recordes de popularidade. Estima-se que o ChatGPT tenha atingido 100 milhões de usuários ativos em janeiro de 2023, apenas dois meses depois de seu lançamento. Para chegar a esta marca, o TikTok levou 9 meses após seu lançamento global; e o Instagram levou pouco mais de dois anos (HU, 2023).

O tema central deste dossiê é o modo como esses avanços recentes no campo da Inteligência Artificial colocam a interação humano-máquina sobre uma nova base. O desenvolvimento tecnológico acelerado no campo da Inteligência Artificial nos últimos anos trouxe de volta para o primeiro plano do debate público em vários países um tema que vem nos acompanhando desde a Revolução Industrial: o processo de superação das capacidades humanas pelas das máquinas que inventamos e a eventual substituição do criador pela criatura no âmbito das atividades produtivas.

Distribuídas por dois séculos, entre 1850 e 1950, as duas primeiras fases da Revolução Industrial nos apresentaram máquinas capazes de substituir nossas capacidades físicas, força motriz humana ou dos animais que sempre empregamos nas atividades produtivas. A terceira fase, que começa a tomar forma no contexto da segunda grande guerra mundial e se desenvolveu através da segunda metade do século XX, introduziu máquinas capazes de substituir a capacidade intelectual humana. Passamos a chamar estas máquinas de *computadores* justamente porque eram capazes de efetuar computações, cálculos, contas de forma mais eficiente

que nós com nossos meios mais tradicionais e analógicos: cérebro, lápis e papel. Eram máquinas que engoliam números e devolviam resultados de cálculos complexos.

Desde aqueles primeiros dias em que os computadores eram lentos, excessivamente custosos e ocupavam grandes espaços de prédios (geralmente localizados em universidades ou centros de comando militar), essas incríveis máquinas pensantes já mudaram muito. Não apenas em tamanho, custo e capacidade de processamento, mas, sobretudo, nas tarefas às quais conseguem se dedicar. Conforme foram sendo projetados para executar tarefas cada vez mais complexas (e um número cada vez maior delas), os computadores precisaram imitar nossas habilidades intelectivas de forma cada vez mais autônoma e eficiente. Este é o vetor de autonomia/eficiência que impulsionou o desenvolvimento da Inteligência Artificial e que estará no centro das reflexões que vamos propor neste artigo.

Embora possa parecer para o grande público que essa Inteligência Artificial poderosa com a qual começamos a conviver tenha surgido de repente, esta é uma tecnologia que vem sendo gestada desde os primeiros dias da computação. Ela é uma espécie de quebra-cabeça que veio sendo montado aos poucos e cujas peças foram desenvolvidas por diversas gerações de pesquisadores. O caminho histórico que nos trouxe até o atual estágio é ramificado, cheio de ruas sem saída, repleto de idas e vindas em termos de estratégias e abordagens metodológicas, altos e baixos em termos de empolgação/financiamento e também reviravoltas teóricas.

Neste dossiê, pretendemos apresentar alguns dos passos desse longo desenvolvimento e demonstrar para o leitor como, nos últimos anos ou mesmo meses, as máquinas que inventamos têm se dirigido de forma acelerada para cruzar uma relevante fronteira na história da tecnologia: o umbral que dá acesso ao mundo cultural humano. Embora grande parte do debate contemporâneo sobre Inteligência Artificial esteja concentrado no processo de superação/substituição e geralmente proporcione profícuas reflexões comparativas entre as capacidades humanas e das máquinas, neste texto, pretendemos focalizar no aspecto da interação humano-máquina. Máquinas que já eram capazes de armazenar e difundir signos – na forma de fala, texto, som e imagem – produzidos por humanos (e para humanos) agora passam a ter capacidade de produzi-los e interpretá-los de modo cada vez mais autônomo.

Estamos especialmente interessados no modo como a *competência semiótica* geral que as máquinas estão em vias de adquirir significa a entrada numa nova era da interação humano-máquina. Já não estaremos

sozinhos na posição de produtores de cultura (num sentido *lato*). Não seremos os únicos a produzir “objetos culturais”, música; poesia; formas visuais, pictóricas ou escultóricas; narrativas na forma de livros, literatura, teatro, filmes, séries, novelas; discursos políticos e ideológicos destinados a conquistar corações e mentes. Em nome da eficiência, já não seremos os únicos a fazer leis e julgá-las, a elaborar políticas públicas e avaliar sua implementação e resultados. É possível que, em pouco tempo, já não sejamos os únicos a elaborar o planejamento urbano das cidades em que vivemos, nem a arquitetura dos prédios que habitamos ou frequentamos. Pode ser que, também em breve, já não sejamos os únicos responsáveis pelo design das religiões que cremos e dos deuses que adoramos. Escreveremos sozinhos as linhas de código de nossos softwares? Arte, Matemática, Ciência. Quem sabe quais campos passarão incólumes? Pensadas originalmente para modificar a infraestrutura de nossas sociedades, nossas máquinas passarão a residir e obviamente modificar de modo profundo a superestrutura.

Como pretendemos mostrar neste artigo, a *competência semiótica geral* que as máquinas se encaminham para adquirir – e que acreditamos estar no núcleo dos processos disruptivos que se seguirão em diversos campos da experiência humana – é um avanço tecnológico obtido numa área particular (no âmbito das pesquisas em Inteligência Artificial) denominada *Processamento de Linguagem Natural*. O incremento nesta capacidade de processar linguagem tem sido intimamente condicionado pelo desenvolvimento de formas específicas de aprendizado de máquina e pela disponibilidade de uma quantidade inacreditável de dados. O objetivo deste dossiê é apresentar de forma panorâmica a história e alguns aportes conceituais a respeito da interação humano-máquina para localizarmos, dentro deste panorama, o atual estágio de desenvolvimento da Inteligência Artificial.

Na primeira seção, apresentaremos algumas definições gerais dos principais conceitos envolvidos na relação humano-máquina e introduziremos um esquema de análise que vamos mobilizar ao longo do artigo. Na segunda seção, fecharemos o foco no desenvolvimento recente da Inteligência Artificial, em particular nos conceitos de *aprendizado de máquina e de processamento de linguagem natural*. Na terceira e última seção, voltaremos nossa atenção para os aspectos mais propriamente semióticos e comunicacionais desse admirável avanço tecnológico: o dia em que as máquinas dominaram a linguagem, ferramenta-mestre do *Homo sapiens*, e assim adentraram o mundo cultural humano.

I. Definições gerais da relação humano-máquina

Geralmente, utiliza-se o termo “máquina” para se referir a um artefato construído por humanos para lhes facilitar algum tipo de trabalho que exija esforço físico ou intelectual (cf. KREBS; KREBS, 2003, p. 106). A máquina é uma determinada estrutura ou organização cujas “partes estão de tal modo conectadas e interrelacionadas que, ao serem colocadas em movimento, o trabalho é realizado como uma unidade” (SANTAELLA, 1997, p. 33).

Santaella nos apresenta uma classificação da relação humano-máquina dividida em três níveis: (1) o nível muscular-motor, (2) o nível sensorio e (3) o nível cerebral. Reparemos que a classificação descreve o desenvolvimento histórico das máquinas: o nível muscular precede o sensorio que, por sua vez, precede o cerebral. Apesar deste sistema classificatório descrever uma ordem histórica de desenvolvimento, ele não supõe que o surgimento de um novo nível anule o anterior (*ibid.*, p. 34). Não há uma superação de um nível em relação a outro. Aliás, o contrário. O que ocorre, por vezes, é que as máquinas que operam em diferentes níveis passam a desenvolver uma relação de cooperação entre si.

Nesta primeira categoria, estão máquinas que procuram imitar a capacidade muscular para amplificá-la ou, em alguns casos, substituí-la. Por exemplo, a alavanca é uma tecnologia muito simples cujo desenvolvimento e uso recuam provavelmente para períodos pré-históricos. A roda com eixo e a polia são máquinas também que embora operem sob princípios relativamente simples, permitiram consideráveis avanços tecnológicos e uma diversidade de aplicações nas civilizações da antiguidade. Por trás do funcionamento de máquinas em geral – mesmo aquelas mais simples como a alavanca – estão princípios muitos elementares. Uma vez que conseguimos captar estes princípios subjacentes com auxílio de uma linguagem formal como aquela mobilizada pelo pensamento matemático, tornamo-nos capazes de generalizar suas aplicações. É o caso do gênio grego Arquimedes que, no terceiro século a.C., utilizou princípios gerais (ainda que não estivessem em um alto grau de generalização) numa diversidade de aplicações tecnológicas e militares (cf. KREBS; KREBS, 2003, p. 109). Em relação à maquinaria que compõe este primeiro nível da classificação de Santaella (1997), a Revolução Industrial criou condições para um salto tecnológico sem precedentes. Vejamos este processo histórico mais de perto, uma vez que nele podemos enxergar, de forma condensada, características gerais que atravessam a história do desenvolvimento das tecnologias humanas e nele também podemos analisar a transição entre os níveis do sistema classificatório.

A Revolução Industrial é uma modificação drástica na base produtiva de uma sociedade por meio da mecanização (parcial ou total) das rotinas de produção. As máquinas que inauguraram este processo histórico eram capazes de substituir a força física humana ou dos animais que empregávamos na atividade produtiva. Na verdade, esse é o tipo de máquina que esteve na base das duas primeiras fases da Revolução Industrial (HOBBSAWN, 1986; MATHIAS; DAVIS, 1989; ASHTON, 1948). Como já antecipamos na introdução, a primeira fase vai da segunda metade do século XVIII até a primeira metade do século XIX e é marcada pela invenção e desenvolvimento do motor a vapor que, primeiro, modificou drasticamente a produção têxtil e posteriormente chegou a outros setores da atividade produtiva. A segunda fase vai da segunda metade do século XIX até a metade do século XX e é marcada pela invenção e desenvolvimento do motor a explosão, surgimento da indústria petroquímica, a era da eletricidade e do aço. Reparemos que, como essas duas primeiras fases estão baseadas em tecnologias que amplificam e substituem força física, estas são máquinas que pertencem ao primeiro nível da classificação de Santaella (1997) da relação humano-máquina: o nível muscular-motor. Já a terceira fase da Revolução Industrial é marcada pela emergência de um tipo distinto de máquina. Aquelas que servem para substituir a capacidade intelectual: os computadores.

Com foco voltado para aspectos semióticos e comunicacionais das tecnologias envolvidas, Santaella (1997, p. 37) chama atenção para outro tipo de máquina que emergiu em meio à paisagem fumegante da Revolução Industrial. É o nível intermediário em sua classificação: o sensório. As máquinas sensórias – também denominadas pela autora de aparelhos – não fabricavam objetos como roupas, sapatos, utensílios domésticos, mas signos na forma de fala, texto, som, imagem, etc.

Enquanto as máquinas musculares foram feitas para trabalhar, os aparelhos foram feitos para simular o funcionamento de um órgão sensório. [...] Enquanto as máquinas musculares produzem objetos, os aparelhos produzem e reproduzem signos: imagens e sons.

Se depois do advento das máquinas musculares, o mundo começou a ser crescentemente povoado de objetos industrializados, depois do advento dos aparelhos, ele começou a ser crescentemente povoado, hiperpovoado de signos. Ao funcionarem como prolongamentos da visão e audição, os aparelhos extensores dos sentidos amplificam a capacidade humana de produzir signos, isto porque os aparelhos não são apenas extensões do processamento sensório, mas também, máquinas de registro e reprodução ou gravação daquilo que os sentidos captam. Uma fotografia, por exemplo, é uma imagem, uma visão do real, registrada num suporte, o negativo, que, além de duradouro, funciona como uma matriz de infinitas cópias.

Nesse sentido, os outputs ou produtos sígnicos dos aparelhos são também formas de memória extra-somática da visão e da audição. (SANTAELLA, 1997, p. 37-8)

Ao contrário das máquinas que operam exclusivamente no nível muscular-motor, os aparelhos ou máquinas sensórias foram construídas para simular o funcionamento específico de órgãos sensórios. Santaella parte da tese de McLuhan (1972) segundo a qual aparatos como uma máquina fotográfica operam como prolongamentos ou extensões dos sentidos humanos. Em textos mais recentes, ao revisitar sua categorização original, Santaella tem enfatizado que as máquinas sensórias conseguem internalizar certo nível de conhecimento técnico. O exemplo paradigmático é a máquina fotográfica.

Antes da industrialização, os instrumentos técnicos para a representação do mundo visível eram prolongamentos do gesto hábil, concentrado nas extremidades das mãos, como é o caso do lápis, do pincel ou do cinzel. Já a tecnologia dá corpo a um saber técnico introjetado nos seus próprios dispositivos materiais. A tecnologia absorve a técnica, mas vai além dela. Assim, a câmera é uma máquina que encapsulou no seu funcionamento todo o conhecimento técnico da perspectiva monocular. Daí para frente, as máquinas sensórias para produção da imagem foram se sofisticando cada vez mais no decorrer do século 20. (SANTAELLA, 2022, p. 180)

Geralmente olhamos para Revolução Industrial a partir de enfoques econômicos e sociais. Focalizam-se os impactos nas condições materiais, na ordem econômica da sociedade. É comum salientarem-se as profundas implicações sociais desse processo histórico: a pobreza, fome, condições precárias de trabalho e diversas outras faces da desigualdade que desembocaram em novas formas de luta política, na emergência de movimentos sociais e ideologias políticas que atravessaram o século XX. Entretanto, não se pode deixar de notar que a Revolução Industrial desenvolveu máquinas de produção sígnica, máquinas do imaginário humano. A Revolução Industrial não construiu o mundo moderno apenas alterando de forma drástica as condições materiais de sustentação de nossas sociedades, mas também modificando as condições gerais de produção dos modos como vemos e sonhamos nossas sociedades. É uma revolução no modo como fazemos circular narrativas e discursos. Uma alteração no modo como produzimos, recebemos, armazenamos e difundimos signos. Este é o sentido de algumas expressões utilizadas por Santaella (1997) no texto: “mundo hiperpovoado de signos” e “usina para produção de signos”.

E qual seria a especificidade do terceiro nível? Um de nossos principais objetivos neste artigo é mostrar que a concepção tecnológica geral desse terceiro tipo de máquina e seus desenvolvimentos mais recentes

(sobretudo, no âmbito da inteligência artificial) não apenas estão conectados aos níveis anteriores da relação humano-máquina, mas são frutos das mesmas forças. Nos últimos milhares de anos, nossas máquinas ficaram mais complexas, poderosas e autônomas. Mas ainda estamos jogando o mesmo jogo.

1.1 Máquinas e o jogo da imitação

Começemos por estabelecer como ponto de partida de nossas reflexões nesta seção que o desenvolvimento tecnológico possui, no âmbito da atividade produtiva humana, um motivo e um vetor básicos. O motivo-base pelo qual desenvolvemos ferramentas, instrumentos, máquinas é ter meios artificiais que nos facilitem as tarefas cotidianas, nossas rotinas produtivas. Se observarmos bem a grande quantidade de artefatos que desenvolvemos para cumprir essa meta geral de nos facilitar o trabalho, notaremos que costumamos enxergá-los dentro de uma linha evolutiva ou um movimento progressivo. Da forma como costumamos contar a história do nosso desenvolvimento tecnológico, nossos artefatos foram ficando cada vez mais complexos, eficientes e autônomos para executar aquela meta geral. Isto é o que vamos chamar de vetor-base do desenvolvimento tecnológico. Assim temos um motivo básico, facilitar o trabalho humano; e um vetor básico, uma direção geral para o desenvolvimento de nossos artefatos, facilitar o trabalho humano de forma progressivamente eficiente/autônoma. Isto significa que quanto maior for o grau de eficiência e autonomia, menor o esforço humano exigido na tarefa de operar a ferramenta ou a máquina. Em outras palavras, quanto mais desenvolvida for uma ferramenta ou uma máquina mais ela será capaz de nos poupar esforço em determinada atividade produtiva.

Antes de nos encaminharmos para a tese de fundo desse breve artigo-panorama sobre a interação humano-máquina, observemos nosso tema dentro de um quadro mais amplo. As forças que parecem moldar a história do desenvolvimento tecnológico humano – e que procuramos introduzir acima no “esquema motivo-vetor” – foram bem captadas no mito prometeico. O fogo dos deuses, roubado por Prometeu e concedido aos humanos, designa justamente a capacidade de inventar, um princípio criador interno contraposto (na narrativa do mito) às imposições externas. Nesta chave, o desenvolvimento tecnológico é resultante de um conflito eternamente renovado entre a engenhosidade humana e os ditames da natureza. Uma agonística geral entre um impulso emancipatório irrefreável e um conjunto, sempre renovável, de estruturas de contenção e aprisionamento.

Na modernidade, esse drama prometeico ganhou novos e irônicos contornos. A modernidade nasce como promessa emancipatória e consegue, de fato, revolucionar nossa capacidade de transformar, por meio do trabalho, matéria-prima em bens e também nossa capacidade de modificar, em escala até então desconhecida, o ambiente natural à nossa imagem e semelhança. A modernidade, particularmente através da (já citada) Revolução Industrial, lançou para estratosfera as forças produtivas de nossas sociedades e conseguiu transformar a face a Terra a tal ponto de ser capaz de criar bolhas artificiais dentro das quais passamos a cultivar nossas sociedades com um alto grau de autonomia em relação à Natureza. Entretanto, a promessa emancipatória moderna livrou a humanidade de grande parte da opressão imposta pelos ditames da natureza apenas para aprisioná-la, um nível acima, num ciclo de exploração imposto por uma parafernália técnico-jurídico-político-socio-econômica de dimensões até então desconhecidas também. Trocamos vetustos grilhões esculpido pela Natureza ao longo de eras por versões artificiais modernas inventadas por nós mesmos em tempos históricos. Muita tinta foi gasta no século XX para jogar alguma luz sobre estas contradições da Modernidade (e seus principais projetos como o esclarecimento/iluminismo).

Não pretendemos descer nos detalhes e nas abstracionices dessa análise filosofante justamente para não espantar os leitores que adentraram este artigo muito mais interessados na concretude e urgência do debate sobre os últimos desenvolvimentos da IA. Insistamos, porém, uma última vez nesta ideia de que os impulsos emancipatórios são sempre frustrados (ainda que o sejam em um novo nível em relação ao estágio anterior), pois é exatamente isso que está em jogo novamente diante da IA. Dessa vez, parece-nos, de forma definitiva. Pode ser nossa última partida. Mas, deixemos para a terceira seção o tom de alarme.

Fechados esses parêntesis filosofantes, voltemos ao esquema “motivo-vetor bases”. Em resumo, a ideia geral é que, para nos libertarmos dos ditames da natureza e, mais recentemente, da extenuante exploração continuada exercida pelos sistemas econômicos que desenvolvemos em tempos históricos, inventamos máquinas que nos facilitem o trabalho. Existem várias formas pelas quais pode-se atingir esta meta. Nossa hipótese neste artigo é que todas elas podem ser reduzidas a uma forma geral ou, ao menos, têm nela o ponto de partida: a imitação.

Se o motivo (e este é o nosso ponto de partida conforme explicado acima) para desenvolvermos nossos artefatos é a facilitação do trabalho, então, estes são sempre pensados e elaborados para nos ajudar em relação a habilidades que já possuímos: nossas capacidades de colhermos

um fruto numa árvore, caçarmos, fazermos fogo, etc. Inventamos um determinado instrumento ou máquina com o foco em alguma habilidade anteriormente adquirida e já desenvolvida. Denominemo-la de *habilidade-alvo*. Dessa perspectiva, o objetivo em se inventar e posteriormente se desenvolver um artefato só pode ser 1) modificar a *habilidade-alvo* ou 2) substituí-la. Estes objetivos não precisam ser considerados como excluídos. Aliás, deve-se encará-los dentro de uma gradação na qual o primeiro deles não apenas precede, mas condiciona (i.e., prepara o terreno para) o segundo: o artefato x imita a habilidade y modificando-a e, eventualmente, substituindo-a na tarefa z.

Imaginemos a situação em que inventamos um determinado instrumento para amplificar a força que somos capazes de exercer (naturalmente, sem o artifício) para uma tarefa. É o caso da alavanca (cf. na seção anterior). Reparemos que o artefato é elaborado imitando uma habilidade ou capacidade humana que, neste caso, envolve predominantemente esforço físico. O objetivo é modificar – no sentido de amplificar, estender, melhorar, tornar mais eficiente e mais precisa, etc. – a força humana empregada na atividade em questão. Esta modificação só pode ser feita por intermédio de uma imitação da habilidade-alvo.

Não se deve levar ao pé da letra o termo “imitação” (cf. SANTA-ELLA, 1997, p. 35). Não significa que um sistema de alavanca feito de madeira é igual aos músculos humanos ou dos animais que empregamos em nossas tarefas produtivas. Músculos são matéria orgânica. Tecidos moldados pelo cinzel da evolução. É até possível que nos dias de hoje, com nosso grau de desenvolvimento técnico-científico, consigamos fazer, em laboratório, músculos sintéticos, mas durante grande parte da História, toda vez que precisávamos fazer instrumentos que imitassem nossas capacidades físico-motoras recorriamos a outros tipos de materiais.

Quando afirmamos que a imitação é a base da modificação a ser exercida por um instrumento ou máquina em relação a uma habilidade-alvo estamos nos referindo a uma relação de semelhança que opera, tal como um princípio geral, em um nível bem abstrato. Por exemplo, hoje conseguimos mobilizar matéria não orgânica para armazenar informação emulando capacidades do cérebro. Como veremos, a base material das redes neurais que estão por trás da atual revolução da IA é bem diferente da rede de neurônios na qual elas se inspiraram. A base material é distinta, mas os princípios gerais que garantem a execução de uma habilidade-alvo acredita-se serem os mesmos.

A imitação é apenas o ponto de partida. Como afirmamos acima, a tendência do desenvolvimento tecnológico é tornar o instrumento ou máquina cada vez mais eficiente em sua capacidade de imitar a habilidade-alvo. Este é um vetor de autonomia. Dado que o motivo básico para o desenvolvimento de nossos aparatos tecnológicos (no âmbito produtivo) é a facilitação do trabalho humano, então essa demanda por eficiência é continuamente perseguida procurando-se atingir graus cada vez mais elevados de autonomia. O caminho do processo de imitação de uma habilidade-alvo é a eficiência/autonomia e o ponto de fuga é a substituição do trabalho humano. Ao menos no âmbito das atividades produtivas, o desenvolvimento tecnológico é sempre um jogo da imitação, e a substituição é o seu destino (ainda que este possa ser indefinidamente adiado, sobretudo por meio de planejamento e regulação, o que, aliás, é o que muitos esperam que ocorra com os desenvolvimentos mais recentes da IA).

Ao lançarmos esse olhar “de longe” e panorâmico sobre o desenvolvimento tecnológico, devemos atentar para o caráter gradual de todo o processo. As capacidades para imitar e substituir exibidas pelos artefatos que inventamos emergem de forma gradual “distribuídas” ao longo da trajetória histórica de uma determinada atividade produtiva. Conforme uma ferramenta ou uma máquina é inventada e modificada ao longo do tempo, ela vai aos poucos se tornando capaz de substituir em cada estágio de seu desenvolvimento frações, nunca de uma só vez, a totalidade de uma rotina produtiva. Assim, o artefato x é inventado para imitar a habilidade y modificando-a e, em determinado estágio histórico, tornando-se capaz de substituí-la nas rotinas z1 ou z2 que compõem a tarefa z. Porém, deve-se enfatizar que, neste estágio, sobram outras rotinas ou sub-rotinas de z para serem realizadas por mãos e mentes humanas.

Vejamos o caso da prensa de tipos móveis de Gutenberg. Esta é uma máquina capaz de mecanizar a replicação de texto escrito. Isso significa que ela tornou automático – e por isso, substituível – o trabalho manual de reprodução de um texto já existente. O que foi substituído foi apenas uma parcela da rotina produtiva referente a um texto escrito. A máquina nos livrou do esforço de copiá-lo à mão. Mas, reparemos que a máquina não automatizou todo o processo. A mão humana ainda teria que ficar responsável por montar a matriz que, uma vez colocada na prensa, permitiria a replicação do texto escrito. E, claro, esta era uma máquina incapaz de produzir um texto. No século XV, época em que Gutenberg criou as condições técnico-comunicacionais para os processos históricos que dariam forma à modernidade, a mente humana era a única capaz de

produzir um texto. À época, as máquinas apenas os replicavam. De certo modo, é assim que funciona até hoje. Esses dias estão próximos do fim.

Se prestarmos atenção nessa condição segundo a qual há, no âmbito de uma tarefa produtiva particular, uma “sobra” de rotinas ou sub-rotinas que não são automatizadas mesmo pelos mais avançados dentre os artefatos que inventamos, veremos emergir a necessidade de se elaborar a chamada “interface”. A ideia é relativamente simples. Como, a cada estágio de desenvolvimento, os instrumentos, aparatos e máquinas que inventamos devem ser operados por seres humanos (devido àquelas rotinas que não foram automatizadas), estes artefatos têm que ser pensados para nos facilitar a “operacionalidade”. A interface é uma dimensão básica da relação humano-máquina. Ela designa a inescapável necessidade de se elaborar algum sistema ou estrutura que faça o papel de mediação entre o ser humano (mão e mente) e a máquina. A ideia-base neste caso é operacionalidade. O desenvolvimento das interfaces seguem, portanto, os mesmos vetores do desenvolvimento tecnológico, eficiência/autonomia para facilitar o trabalho humano. Embora se costuma utilizar este termo, interface, para designar a relação entre computadores e usuários, vamos generalizar este uso para qualquer artefato ou instrumento.

Vejamos o caso de artefatos mais simples que geralmente não classificamos como maquímicos. Um bastão, por exemplo, utilizado como arma; e um escudo utilizado como aparato de defesa. Estes artefatos são elaborados de uma tal forma que um agente (humano) possa manuseá-lo de modo adequado para os seus fins de ataque ou defesa. Uma parte da sua estrutura é previamente pensada e produzida para se adequar à mão humana. No caso do arco e flecha, parte da estrutura do arco é pensada para se adequar à pegada da mão humana, e a precisão do instrumento, neste caso, da arma, depende do grau de adequação. Fechemos o foco num tipo de artefato, as máquinas, e voltemo-nos para a classificação de Santaella (1997).

No caso das máquinas que operam exclusivamente no nível físico-muscular, a interface está em alavancas e alças. Conforme nossas máquinas vão se complexificando em sua missão elementar de nos facilitar o trabalho e à medida que vamos nos afastando deste primeiro nível e nos aproximando do nível sensorio, a interface passa a ser dominada por botões e painéis. Nestas máquinas mais complexas, a função a ser cumprida pelas mãos e mentes humanas é cada vez mais intelectual e menos física ou muscular. Este processo gradual atinge um ponto de virada quando chegamos aos computadores, as máquinas pensantes – aquelas que operam predominantemente no nível cerebral da classificação de Santaella (1997). Com pouco ou quase nenhum esforço físico, estas máquinas exigem que a mente humana as opere dentro de um ambiente informacio-

nal abstrato. É como se este tipo de máquina exigisse dos humanos que entrassem e participassem de um mundo próprio delas.

É curioso notar como, nos primeiros dias, os computadores eram, de fato, operados por pouquíssimas pessoas. Apenas os iniciados tinham acesso à sua linguagem e ao seu mundo particular. Graças à interface, a tarefa de mediação exigia alto grau de especialização. Deste período até nossos dias, os computadores diminuíram de tamanho e custo, aumentaram a capacidade de processamento e também se popularizaram. Uma das histórias mais fascinantes de nossa trajetória tecnológica recente é o modo como através de avanços na interface fizemos essas máquinas (cada vez mais) pensantes chegarem às casas e mãos de todos. Passamos da interface de linha de comando para as interfaces gráficas.

Esta é a linha de desenvolvimento que nos leva até a revolução contemporânea da IA. Já não somos nós que precisamos aprender a linguagem das máquinas para entrarmos no mundo informacional abstrato dela. São as máquinas que estão aprendendo nossa linguagem para entrar no complexo mundo cultural humano.

2. A relação humano-máquina na era da IA

A história do desenvolvimento da Inteligência Artificial é a história do modo como temos, nas últimas sete décadas, tentado – através de diversas estratégias e atravessando picos de empolgação, convicção e financiamento como também vales de desânimo, descrença e restrições de ordem orçamentária – reproduzir “em silício” capacidades intelectivas humanas. Dado este objetivo geral, é até natural que o campo de pesquisa em IA tenha se organizado em subdisciplinas voltadas para as principais capacidades que se espera que IA emule como raciocínio, planejamento, locomoção, manipulação, percepção (por exemplo, visão e audição), reconhecimento de imagem e de face, reconhecimento de fala e voz, processamento de linguagem natural, entre outras. Pode-se dizer que estas áreas estão baseadas no que chamamos na primeira seção de *habilidades-alvo*.

O grande sonho neste campo de pesquisa é atingir o que é chamado de *inteligência artificial geral* ou *inteligência artificial forte*. Isto seria um agente de IA capaz de operar de modo eficiente em todas essas frentes, saltando de um tipo de tarefa para outra como seres humanos conseguem fazer. O que temos até o presente momento são inteligências artificiais denominadas *não gerais* ou *fracas*. Estas são IAs pensadas e desenvolvidas para tarefas particulares (por vezes, bem específicas). Por exemplo, são IAs que sabem jogar xadrez e reconhecer faces. A qualificação “fraca” na

expressão não significa, portanto, que elas sejam ruins, inábeis ou incompetentes na tarefa para qual foram desenvolvidas. Aliás, é justamente o contrário. São IAs altamente especializadas e eficientes. Um dos principais problemas está na incapacidade que essas IAs têm de generalizar o que aprendem em uma tarefa ou cenário particular para outras tarefas/cenários.

Desenvolver um tipo de inteligência que seja capaz de “se virar bem” numa grande quantidade de tarefas como a mente humana consegue fazer de modo bem natural tem se mostrado um desafio com contornos de ironia. Tudo aquilo que a mente humana consegue, num estágio inicial da vida, aprender de modo fácil e, ao longo do tempo, executar de modo automático e praticamente sem esforço é muito difícil programar uma máquina para fazer. Por exemplo, fala e percepção visual. Inversamente, tudo aquilo que julgamos difícil de aprender e executar bem (mesmo com muitos anos de experiência) parece ser fácil *programar* uma máquina para fazer. Por exemplo, fazer cálculos complexos e jogar xadrez. Estes são os termos do que os pesquisadores do campo da IA chamam de *paradoxo* de Moravec (cf. MORAVEC, 1988, p. 15; PINKER, 1994, p. 190–1).

No caminho para o desenvolvimento de IA geral, os carros autônomos representam um importante desafio, pois demanda que o agente de IA execute de forma eficiente diversos tipos de tarefas emulando diferentes habilidades humanas em cenários muito dinâmicos. De forma bem geral, a ideia, neste caso, é produzir uma máquina capaz de substituir nossa capacidade de dirigir. Embora seres humanos sejam capazes de aprendê-la relativamente rápido e de, com a experiência, automatizar grande parte de suas sub-rotinas, *dirigir um carro* tem se mostrado uma habilidade bem complexa para IAs. Ela envolve, por exemplo, o que os pesquisadores e desenvolvedores de IA chamam de “sensores” e “atuadores”. Os sensores são aquela parte do agente de IA que o permite captar dados sensórios, i.e., perceber o ambiente. Já os “atuadores” designam aquelas partes do agente de IA responsável por executar uma ação dentro do ambiente (RUSSELL; NORVIG, 2021, p. 54). No caso dos carros autônomos, os sensores captam, por exemplo, imagem e som do ambiente, estes dados são “traduzidos” em algum tipo de representação interna da máquina, o ambiente é mapeado e suas entidades classificadas, então, um determinado movimento é planejado pela IA e só depois, por meio dos “atuadores”, executado. O movimento planejado pode ser, por exemplo, “acelerar”, “desacelerar”, “virar à direita”, “virar à esquerda” ou combinações dessas ações com gradações como “desacelerar da forma x e virar à esquerda na medida y”. Para isso, entram em cena os “atuadores”, geral-

mente desenvolvidos por engenheiros de robótica, que transformam informação (energia na forma bits e bytes) em ação física (energia cinética). Este desenvolvimento que promete nos levar de uma miríade de IAs fracas capazes de executar tarefas particulares para IAs fortes com habilidades gerais é um movimento que segue o que chamamos, na seção anterior, de *vetor da eficiência/autonomia*. Como vimos, quando uma tecnologia se torna capaz de emular uma habilidade humana e substituir parte de seu trabalho dentro de uma determinada tarefa produtiva, aquela parte não substituída passa a ser trabalhada na forma de interface para fornecer controle e operacionalidade para os humanos em relação à máquina. No caso dos computadores, o aumento da autonomia e eficiência tem se traduzido não apenas numa maior facilidade para o usuário-final operá-los, mas também tem significado maior independência por parte das máquinas em relação à mão do programador. A medida do avanço de nossas IAs está na distância entre a máquina e a mão do programador. Esta distância tem aumentado em velocidade estonteante nos últimos dias justamente pela emergência de uma habilidade central à vida humana: a capacidade de aprender. O drama prometeico humano está ganhando um novo capítulo. Humanos estão passando às máquinas a tocha roubada dos deuses por Prometeu.

O *aprendizado de máquina* não é somente uma área entre outras dentro do campo de IA. Ele está rapidamente se encaminhando para ser uma *área transversal* capaz de atravessar todas as outras criando condições para revolucionar cada uma delas. Nas próximas sub-seções, apresentaremos as peças que compõem o quebra-cabeça do acelerado avanço contemporâneo no campo da IA e procuraremos apontar como o motor por trás desse movimento é o *aprendizado de máquina* e o combustível, a grande quantidade de dados que a internet tornou disponível.

2.1 Aprendendo a aprender

Na seção anterior, apresentamos o desenvolvimento tecnológico como um jogo de imitação. As gerações de viventes desses inícios de século XXI estão testemunhando uma jogada definitiva neste jogo. Conseguimos desenvolver em máquinas a capacidade aprender. O tabuleiro, os jogadores e nenhuma partida jamais serão os mesmos. A eficiência mimética de nossas máquinas terá atingido um ponto em que o grau de autonomia é tão elevado que já não seremos mais capazes de considerá-las *meros* instrumentos. Deixemos os alertas para última seção desse artigo. Por enquanto fiquemos com o tom otimista de Terrence Sejnowski na abertura de seu livro intitulado “A revolução do *aprendizado* profundo”:

Há não muito tempo atrás costumava-se dizer que a visão de um computador não poderia competir com as habilidades visuais de uma criança de um de idade. Isto não é mais verdade: computadores agora podem reconhecer objetos em imagens tão bem como um adulto consegue fazer, e hoje há nas estradas carros computadorizados que dirigem de forma autônoma com mais segurança do que seria capaz o homem médio com sessenta anos de idade. Em vez de se dizer para esses computadores como eles deveriam ver ou dirigir, eles mesmos aprenderam a partir da experiência, seguindo um caminho que a natureza levou milhões de anos para trilhar. O que está abastecendo esses avanços é uma grande quantidade de dados. Dados são o novo petróleo¹. Os algoritmos de aprendizado são refinarias que extraem informação do dado bruto; informação pode ser usada para criar conhecimento; conhecimento leva ao entendimento; e entendimento leva à sabedoria. Bem-vindo ao admirável mundo novo do aprendizado profundo [deep learning]. (SEJNOWSKI, 2018, p. 3 [tradução nossa])

A revolução de que trata Sejnowski começa quando os teóricos e pesquisadores do campo de IA encontram um modo de fazer com que as chamadas redes neurais fossem capazes de emular nas máquinas processos de aprendizagem a partir da experiência. No âmbito da IA, a *rede neural* é uma aposta que a inteligência a ser desenvolvida na máquina deveria se assemelhar em estrutura/arquitetura geral à inteligência natural. Embora tenha levado décadas para ser concretizada de forma plena, a ideia básica é muito simples: a IA deveria mimetizar o modo como neurônios são capazes de guardar informação e aprender com dados captados do mundo externo. Antes desses modelos que procuraram imitar os neurônios se provarem eficientes nas tarefas para as quais foram planejados, a aposta do *mainstream* na IA era que o conhecimento adquirido por uma máquina deveria ser dado pela mão do programador na origem (i.e., quando o programa roda pela primeira vez) ou em atualizações posteriores. Em síntese, os computadores sabem fazer o que colocamos em seus programas. Porém, antes de entrarmos no conceito de *aprendizado profundo* e as *redes neurais profundas*, vejamos o que é, de forma geral, o *aprendizado de máquina*.

¹ É necessário enfatizar este fator da disponibilidade de uma grande quantidade de dados. Big data é um termo que vem sendo utilizado nos últimos anos para designar um conjunto de dados que ganha volume muito rápido. De acordo com Sejnowski, as profissões e serviços que serão primeiro transformadas conforme as técnicas de aprendizado de máquina forem amadurecendo serão aquelas em que há uma considerável disponibilidade de dados. O exemplo paradigmático é a medicina e o campo de diagnósticos por imagem. “O diagnóstico médico baseado em registros de milhões de paciente se tornará mais preciso. Um estudo recente aplicou aprendizado profundo em 130,000 imagens dermatológicas para mais de 2,000 doenças diferentes – esta é uma base de dados dez vezes maior que a utilizada anteriormente” (SEJNOWSKI, 2018, p. 10). A rede neural foi treinada para conseguir elaborar o diagnóstico em imagens que não tinha em sua base original. A performance neste conjunto de imagens foi comparável e em alguns casos melhor que a de 21 especialistas da área (ibid.).

Sabe-se que o agente de IA é capaz de aprender se ele consegue melhorar sua performance graças à experiência. Em outras palavras, ele é capaz de usar dados obtidos a partir da experiência para tornar mais eficiente sua performance. O *aprendizado de máquina* é a capacidade de uma máquina de acessar dados, construir um modelo a partir deles e utilizar o modelo tanto “como hipótese a respeito do mundo como um pedaço de software capaz de resolver problemas” (RUSSELL; NORVIG, 2021, p. 669). A aprendizagem consiste justamente na possibilidade da máquina atualizar seu modelo através da experiência. O modelo é uma mistura entre o *script* que programador colocou originalmente no programa e o que a máquina foi capaz de aprender coletando dados da experiência e encontrando padrões neles. Existem três tipos gerais de *aprendizado de máquina*. *Aprendizado supervisionado*, *não supervisionado* e *por reforço*. Vejamos, em primeiro lugar, as definições dos dois primeiros tipos de aprendizagem bem como suas diferenças e exemplos de aplicação. Por último, dediquemo-nos ao terceiro tipo, que parece ser a origem de grande parte do impulso revolucionário de que fala Sejnowski no trecho acima.

A diferença básica entre aprendizado supervisionado e não supervisionado está na rotulagem (*labelling*) dos dados que são fornecidos para o algoritmo. No primeiro tipo, a máquina é treinada num conjunto de dados rotulados em que os padrões relevantes são mostrados através da rotulagem. Então, o algoritmo de aprendizado faz com que a máquina procure, a partir do conhecimento dos padrões obtido no treinamento, generalizar o processo de rotulagem para casos não vistos no conjunto de dados utilizado para treiná-la. Já no segundo tipo, não é fornecido durante o treinamento nenhum tipo de rotulagem de dados. O algoritmo de aprendizado faz com que a máquina procure *por ela mesma* padrões no conjunto de dados que lhe é fornecido.

Estes dois modelos gerais, supervisionado e não supervisionado, têm aplicações diferentes. Por exemplo, a tarefa de separar e-mails que são *spam* daqueles que não são tem sido operada de forma cada vez mais eficiente graças ao uso de modelos supervisionados de aprendizado de máquina. Esta é uma tarefa de classificação em que a IA aprende um determinado padrão a partir de um conjunto prévio de e-mails rotulados pelo próprio usuário como *spam*. Com base neste padrão, ela procura classificar e-mails futuros. Já o modelo não supervisionado tem sido muito utilizado no campo do marketing para encontrar o perfil de consumidores (com a finalidade de mostrar, através de anúncios, novos produtos que estejam dentro do seu perfil de consumo). Grande parte da eficiência desses sistemas de perfil de consumo está baseada na técnica denominada *clusterização*, que é um tipo de aprendizado não supervisionado de máquina.

O terceiro tipo de aprendizado de máquina é chamado de aprendizado por reforço. Neste modelo, a máquina torna-se capaz de aprender através de uma série de recompensas ou punições que operam como mecanismo de reforço. O exemplo fornecido por Russell e Norvig (2021, p. 671) é o de um agente de IA desenvolvido para jogar xadrez ao qual se aplica, ao final de uma partida, um determinado reforço, uma recompensa ou uma punição dependendo do resultado no jogo. De acordo com os autores, neste modelo de *aprendizado*, é o próprio agente (*i.e.*, a IA) que decide quais das ações anteriores ao reforço foram mais diretamente responsáveis pelo resultado que levou ao reforço e, assim, opta por alterar suas ações com o objetivo de obter mais recompensas no futuro (*ibid.*).

Um dos marcos históricos recentes no desenvolvimento da IA, a supremacia da máquina sobre o campeão humano de GO, foi alcançado graças a uma combinação do modelo de aprendizado por *reforço com aprendizado profundo* (SILVER *et al.*, 2016). O método é apropriadamente denominado *reforço profundo*. Voltaremos logo adiante a esse ponto, a supremacia da máquina em jogos inventados e dominados até outro dia por humanos. Antes, expliquemos no que consistem o *aprendizado profundo* e as *redes neurais profundas*.

De acordo com Taulli (2019), a história das redes neurais recua para as ideias e a pesquisa do psicólogo Frank Rosenblatt. Ao final da década de 1950, ele resolveu apostar num determinado caminho/estratégia para o desenvolvimento da IA. A ideia era que os modelos a serem desenvolvidos para que as IAs fossem capazes de processar e armazenar informação deveriam procurar imitar o modo como cérebro executa esses procedimentos. Enquanto o cérebro armazena e processa mobilizando uma gigantesca rede de neurônios, as IAs deveriam recorrer ao que Rosenblatt chamou de perceptrons. “O sistema seria capaz de aprender conforme fosse ingerindo dados ao longo do tempo” (TAULLI, 2019, p. 10).

O *Mark 1 Perceptron*, idealizado e desenvolvido pelo próprio Rosenblatt, foi o primeiro programa construído para concretizar a ideia geral de rede neural. Embora o programa tivesse funcionado conforme previsto apontando para um caminho muito promissor (o que podemos afirmar hoje com o conforto que o olhar em retrospectiva nos oferece), a abordagem foi abandonada pouco tempo depois pelo peso das críticas. Devemos lembrar que essa ainda era uma fase de formação do campo de IA, e as disputas entre metodologias e abordagens muito distintas era relativamente comum. A abordagem de Rosenblatt foi duramente criticada por Minsky e Seymour Papert (1969), dois grandes nomes do campo de IA. Deve-se lembrar que a abordagem se apoiava em uma série de pressupostos sobre o cérebro humano e, na época, o estudo sobre a cognição

humana e os mecanismos internos de processamento ainda estava dando seus primeiros passos. Deve-se enfatizar também que, devido basicamente às limitações de processamento dos computadores da época, a rede neural do *Mark 1 Perceptron* tinha apenas uma camada.

O caminho para IA divisado por Rosenblatt ainda nos anos 1950 foi resgatado apenas na década de 1980 e são as pesquisas que resolveram trilhá-lo desde então que nos trouxeram até a atual revolução do *aprendizado profundo* (*ibid.*). Aliás, está justamente no número de camadas a diferença básica entre esta humilde estreia do conceito de rede neural e os poderosos algoritmos de *aprendizado* sustentados por *redes neurais profundas* na IA contemporânea. Quando uma rede neural tem duas ou mais camadas qualificamo-la *profunda*.

Vejamos o caso de uma complexa tarefa no âmbito da percepção visual de computadores: o reconhecimento de entidades/objetos dentro de imagens. Durante décadas pesquisadores quebraram a cabeça para conseguir programar uma função cujo *input* seria nada mais que o conjunto de cores de todos os pixels de uma imagem e o *output* seria uma descrição verbal adequada da imagem. Humanos não têm dificuldade alguma em observar uma imagem com jovens jogando frisbee ou uma foto de elefantes andando num gramado e rapidamente verbalizarem uma descrição correta do que veem. Em 2014, uma equipe do Google, liderada pelo pesquisador Ilya Sutskever, conseguiu fazer, em repetidos testes, com que um computador observasse uma imagem e a descrevesse de forma adequada. Por trás da façanha estava uma nova estratégia. A equipe de Sutskever não procurou programar, como era a orientação geral na estratégia até então dominante, manualmente algoritmos específicos para detectar as entidades presentes nas imagens (por exemplo, reconhecer “seres humanos”, “frisbees”, “elefantes”, “gramado”, etc.). “Eles criaram uma rede neural relativamente simples sem absolutamente nenhum conhecimento sobre o mundo físico e seu conteúdo e a deixaram aprender expondo-a a uma grande quantidade de dados” (TEGMARK, 2017, p. 78).

Esta estratégia pareceu contra-intuitiva e essencialmente errada a muitos no campo da IA e por muitas décadas. O único caminho seria colocar na máquina via programação tudo que ela precisaria saber para executar sua tarefa. Parecia impossível ensinar (via programação) uma máquina a aprender com os dados. E, neste ponto, entra em cena um novo ingrediente trazido pela internet: os dados, uma grande quantidade deles. A adoção da estratégia da aprendizagem a partir da experiência foi condicionada não somente pelo desenvolvimento de algoritmos e arquiteturas específicas, mas pela disponibilidade de dados e também, claro, um aumento na capacidade computacional e de memória.

Dentro do córtex visual, os neurônios estão arrançados em camadas hierárquicas. Conforme a informação sensória vai passando camada cortical por camada cortical, ela vai sendo transformada, a representação do mundo vai ficando mais e mais abstrata. Ao longo de décadas, conforme cresceu o número de camadas em modelos de redes neurais, a performance deles continuou a melhorar até que atingiu finalmente um limite crítico que nos permitiu resolver problemas com os quais apenas poderíamos sonhar em resolver na década de 1980. O aprendizado profundo automatiza o processo de encontrar boas características capazes de distinguir diferentes objetos numa imagem. É por esse motivo que a visão do computador é hoje muito melhor do que há cinco anos.

Por volta de 2016, os computadores se tornaram milhões de vezes mais rápidos, e a memória deles aumentou bilhões de vezes - de megabytes para terabytes. Tornou-se possível simular redes neurais com milhões de unidades e bilhões de conexões, comparadas com as redes que, na década de 1980, tinham apenas centenas de unidades e milhares de conexões. Embora ainda sejam pequenas para os padrões do cérebro humano, que possui cem bilhões de neurônios e um milhão de bilhões de conexões sinápticas, as redes atuais são agora grandes o suficiente para que se possa demonstrar sua praticidade em domínios estreitos. (SEJNOWSKI, 2018, p. 35 [tradução nossa])

O *aprendizado de máquina* de forma geral e o *aprendizado profundo* de modo particular abriram uma avenida de possibilidades para o desenvolvimento de IAs. Para podermos conceber o tamanho dos saltos que a IA deu recentemente, focalizemos nos embates entre a mente humana e a máquina no campo dos jogos. Comparemos dois marcos históricos, as vitórias dos computadores sobre os humanos no xadrez e no GO. Infelizmente, a grande mídia costuma cobrir esse tipo de acontecimento histórico quase exclusivamente a partir do enfoque da “ferida narcísica”. Jornalistas gostam de explorar, não raras vezes em tom sensacionalista, o orgulho humano ferido por conta da derrota para as máquinas. Não estamos exigindo que fossem feitas coberturas com alto grau de detalhamento técnico, mas poder-se-ia, ao menos, dar alguma atenção para os avanços técnicos responsáveis pela façanha que faz a manchete. Atenção, neste caso, significa traduzir para o grande público esses detalhes. Isso nos ajudaria bastante em períodos de desenvolvimento tecnológico acelerado e disruptivo. É exatamente o caso atual. Deixaremos para tratar na última seção do modo como a percepção pública irá ter um importante papel na resolução ou, ao menos, num encaminhamento minimamente aceitável de nossos problemas em relação a IA.

De acordo com Tegmark (2017, p. 78), quando em 1997 o Deep Blue (da IBM) conseguiu vencer o campeão do mundo de xadrez, Gary Kasparov, a vitória da máquina se deveu basicamente a avanços em termos de memória e capacidade de computação. O *Deep Blue* superou

humanos, pois conseguia computar mais rápido ao analisar possíveis posições no tabuleiro de xadrez. Em 2011, quando o computador Watson (também da IBM) venceu campeões humanos do “Jeopardy”, um tradicional jogo de perguntas e respostas da TV americana, a superioridade da máquina não se deveu tanto à aprendizagem, mas a habilidades programadas sob medida para o jogo e também a fatores como memória e capacidade de computação. Porém, para que um computador pudesse ser considerado competitivo no GO, seriam necessários algoritmos específicos de aprendizado capazes de, ao menos, flertar com capacidades que narcisicamente sempre acreditamos que só humanos conseguiriam exhibir: criatividade e intuição.

Jogadores de GO colocam em turnos alternados peças brancas e pretas num tabuleiro de 19 por 19 [...]. Há muito mais possibilidades de posições no jogo Go do que átomos no nosso universo, o que significa que tentar analisar todas as sequências interessantes de futuros movimentos rapidamente se mostra inútil. Os jogadores dependem em grande medida de intuição subconsciente para complementar seu raciocínio consciente, o que leva os especialistas a desenvolver um estranho senso para distinguir quais posições são fortes e quais são fracas. [...] os resultados do aprendizado profundo às vezes lembram a intuição: uma rede neural profunda pode identificar que uma imagem retrata um gato sem saber explicar o porquê. (TEGMARK, 2017, p. 87 [tradução nossa])

No trecho a seguir, Tegmark enfatiza que a força do *DeepMind* – computador desenvolvido pelo Google que, através do programa denominado *AlphaGo*, foi capaz de vencer em 2016 o campeão humano de GO, o sul-coreano Lee Sedol – estava numa mistura de métodos novos com antigos.

A aposta da equipe do DeepMind foi que o aprendizado profundo deveria ser capaz de identificar não apenas gatos, mas também posições fortes no Go. A ideia central que eles implementaram no AlphaGo consistia no casamento entre a capacidade intuitiva do aprendizado profundo com a capacidade lógica do GOFAI - que é uma bem-humorada sigla que significa “ao estilo da boa e velha IA” [na sigla em inglês, “Good Old-Fashioned AI”] que antecedeu a revolução da aprendizagem profunda. Eles utilizaram uma gigantesca base de dados de posições de Go retiradas tanto de jogos entre humanos como de jogos nos quais o AlphaGo jogava com um clone dele mesmo e treinaram uma rede neural para prever a partir de cada posição a probabilidade de vitória ao final da partida. Eles também treinaram uma rede separada para prever os próximos movimentos. Então, combinaram essas redes com o método GOFAI que, de forma hábil, fazia uma busca numa lista já reduzida de sequências de prováveis movimentos fu-

turos para identificar o próximo movimento que levaria a posição mais forte em relação aos resultados futuros do jogo. (TEGMARK, 2017, p. 88 [tradução nossa])

Até o momento ficamos focados em como o aprendizado de máquina e, especificamente, o aprendizado profundo revolucionaram a IA em tarefas como a percepção visual e nos jogos. E no campo da linguagem?

2.2 Aprendendo linguagem humana

Começamos este artigo pelo ChatGPT justamente porque é bem provável que este *chatbot* tenha sido o primeiro “artefato tecnológico” do campo da IA a causar o impacto e o “efeito de enganação” previsto pelo célebre teste de Turing. O ChatGPT é resultado de uma convergência de fatores pelos quais já passamos algumas vezes nas sinuosas linhas de nossa exposição neste artigo. Os fatores são o desenvolvimento da capacidade de aprendizagem (nas máquinas) e a disponibilidade de uma imensa quantidade de dados. Nesta subseção, fechemos o foco no modo como esses fatores num campo da pesquisa em IA chamado de *processamento de linguagem natural*.

Uma das mais notáveis e importantes peças na engrenagem dessa revolução atual no campo da IA é o *transformer*, que é uma arquitetura de rede neural (VASWANI *et al.*, 2017) e um modelo de *aprendizado profundo*. Para o processamento de linguagem natural, o *transformer* representou uma revolução, pois os modelos de *aprendizado profundo* disponíveis até então (por exemplo, as *redes neurais recorrentes*) incorriam numa série de dificuldades diante das exigências para uma IA geral. Esses modelos anteriores operavam sequencialmente, “lendo”, processando palavra por palavra. Por exemplo, para processar uma frase com a finalidade de traduzi-la para outra língua, estes modelos lidavam com uma palavra depois da outra. Há algumas dificuldades em termos de tempo e de eficiência envolvidas no treinamento desses modelos, uma vez que esta operação sequencial não permite processamento paralelo (ou seja, não são modelos *paralelizáveis*). Isto significa que estas redes são difíceis de serem treinadas. É um treinamento demorado e custoso do ponto de vista computacional.

A revolução instaurada pelo *transformer* reside num mecanismo – chamado de “atenção” (cf. VASWANI *et al.*, 2017) – que procura prever o próximo elemento (por exemplo, uma palavra) dentro de uma sequência. O modelo pega uma sequência de palavras como *input* e, através da iteração em camadas e camadas de codificação, procura prever a próxima palavra numa sequência de *output*. O processamento paralelo permite que os modelos sejam treinados com uma quantidade gigantesca de dados num tempo relativamente curto (se comparado aos modelos anteriores).

Vejam os casos particulares do ChatGPT. Como esse chatbot consegue entender e produzir instâncias de linguagem natural? Como vimos para outras áreas, o segredo está no treinamento e no algoritmo que torna o agente de IA capaz de aprender com a experiência. O modelo quebra os textos que compõem a sua base de dados (de treinamento) em pequenas unidades chamadas “tokens”. A cada token é atribuído um número, uma codificação que varia de acordo com o contexto. Isto significa que uma mesma palavra ocorrendo em contextos diversos pode receber números/códigos distintos. Para a máquina, uma frase ou mesmo uma palavra é representada como uma sequência de números (que correspondem a sua sequência de *tokens*). O passo seguinte é o parâmetro. O modelo observa a relação entre as palavras dentro do conjunto de dados que lhe foi fornecido (para o treinamento) com foco em quais palavras ocorrem juntas com mais frequência. Ele capta, então, a probabilidade de uma determinada palavra aparecer junta de outra e, assim, atribui pesos. Estes são os chamados *parâmetros* que vão constituir o modelo.

Ao transformar os elementos que compõem um texto em números (tokens) e encontrar probabilidades de ocorrência desses elementos dentro de um contexto (por meio dos parâmetros), o modelo torna-se capaz de gerar sequências cada vez mais (conforme o refinamento do treinamento/experiência vai permitindo) parecidas com os textos que estão na base de dados usada na fase de treinamento inicial. Assim, o modelo torna-se capaz de gerar texto ao aprender a probabilidade de um determinado token *x* aparecer depois de um outro token *y* dentro de um contexto *z* (identificado pela presença de outros tokens). Por exemplo, diante da sequência “machado de...”, o modelo consegue saber que, num contexto em que o texto trata de ferramentas, é maior a probabilidade de ocorrer a palavra “mão”. Fosse um contexto sobre literatura e cultura, o modelo saberia que a palavra com maior probabilidade de ocorrência seria “Assis”. De forma geral, esta é uma operação semelhante ao recurso autocompletar que já funciona em nossos motores de busca na internet ou nos aplicativos de mensagens disponíveis nos celulares. A diferença é que o modelo de linguagem, no caso do ChatGPT, seria uma ferramenta de autocompletar muito mais poderosa e complexa do que qualquer recurso com qual tivemos contato até o momento no mundo informático.

As façanhas do ChatGPT e congêneres são resultado de uma forma de modelar a linguagem natural que tem sido chamada de “grandes modelos de linguagem” (*large language model*) justamente por tais modelos estarem baseados em redes neurais profundas com uma grande quantidade de parâmetros. O GPT-3, por exemplo, trabalha com 175 bilhões de

parâmetros (RADFORD *et al.*, 2018, p. 2). Ele foi treinado em 2020 em parceria com a Microsoft mobilizando um dos supercomputadores mais potentes disponíveis à época. O custo estimado para todo processo de treinamento foi de 12 milhões (FLORIDI; CHIRIATTI, 2020, p. 684).

O ChatGPT mobiliza um “conhecimento estatístico” (obtido na análise de uma grande quantidade de dados) para gerar textos que, de fato, nos parecem terem sido produzidos por um agente inteligente e consciente cujas escolhas são dotadas de intencionalidade. Este é o truque por trás do “efeito de enganação” que tratávamos no parágrafo inicial desta subseção. Não é, de fato, uma *escolha*. Não no sentido que geralmente usamos este termo. A máquina “escolhe” a próxima palavra numa sequência a partir de um cálculo probabilístico. Embora este tipo de procedimento simule ser um julgamento de um agente consciente, ele não é uma escolha humana. Este é um jogo de aparências. Portanto, a análise estatística de uma grande quantidade de textos permite ao modelo encaixar palavras umas depois das outras dando forma a frases, parágrafos e até mesmo a textos que nos parecem terem sido gerados por seres humanos. Muitos especialistas chamam atenção para fato de que a geração de texto operada por este tipo de *modelo de linguagem* não está fundamentada em nenhum tipo de intenção comunicativa, nem em nenhum tipo de modelo do mundo ou qualquer modelo do estado da mente do intérprete (BENDER *et al.*, 2021, p. 616).

Ao contrário do que pode parecer quando observamos seu output, um ML [modelo de linguagem] é um sistema que procura costurar casualmente sequências de formas linguísticas que ele observou na sua vasta base de dados de treinamento observando informação probabilística a respeito de sua combinação, mas sem nenhuma referência a significado: um papagaio estocástico. (BENDER *et al.*, 2021, 616 –7)

A pergunta se o ChatGPT e, de forma geral, os modelos de linguagem desenvolvidos recentemente conseguem, de fato, compreender linguagem natural ou são apenas *papagaios estocásticos* que imitam compreensão humana é uma das questões mais relevantes acerca da IA.

De um ponto de vista semiótico, uma das principais questões trazidas pela IA é que tipo de competência semiótica é essa que parece emergir com máquinas? A capacidade que seres vivos possuem de efetuar “processamento” de signos provenientes do ambiente (externo e interno) está intimamente ligada a uma forma corporal situada no mundo (para a relação entre competência semiótica e corpo, cf. STJERNFELT, 2006). Admitido que estamos diante de máquinas inteligentes, deve-se reconhecer que a IA exhibe uma forma de inteligência muito distante daquela de

base orgânica (moldada pela evolução). Que tipo de competência semi-ótica é essa que emerge de seres capazes de compartilhar de forma praticamente instantânea “conhecimento” obtido na experiência a partir de diversos terminais e capazes de carregar “em mente” um mesmo modelo de linguagem (i.e, com os mesmos pesos/parâmetros)? Essas questões são muito importantes e devem nos levar a acalorados e demorados debates teóricos-conceituais, sobretudo no mundo acadêmico. O principal problema – ao qual nos dedicaremos no fechamento deste artigo – é que grande parte de nossos problemas práticos envolvendo a IA não pode esperar pelo encaminhamento de questões teórico-conceituais.

Seja por qual porta entrarmos nas reflexões sobre a IA e seus impactos nos diversos campos da experiência humana, cedo ou tarde encontramos uma grande quantidade de conceitos cuja complexidade parece sempre escapar aos moldes de nossas metodologias científicas: inteligência, consciência, intencionalidade, agência, mente, pensamento, conhecimento, aprendizagem, compreensão, significado, sentido, linguagem, entre outros. Esses conceitos exigem um nível de abstração e de capacidade de teorização considerável. Em relação a nossas visões de mundo e sistemas teóricos, eles estão num nível de profundidade com o qual somente filósofos estão acostumados. Não é por outro motivo que tais conceitos vieram sendo discutidos pelos nossos mais gabaritados filósofos no Ocidente em diferentes ciclos filosóficos por mais de dois milênios. O problema é que eles têm oferecido forte resistência quando procuramos capturá-los dentro de esquemas empírico-experimentais que a ciência moderna desenvolveu para lidar com fenômenos complexos. Estamos com claras dificuldades em retirá-los de terreno filosófico e fazê-los caber na ciência de base empírico-experimental. Esses conceitos nos lançam um olhar desafiador como se fossem uma esfinge milenar, imponente no deserto, resistente à passagem do tempo, às intempéries e vicissitudes da história humana e, claro, às investidas mais recentes da ciência moderna.

A dura realidade factual é que estamos diante de IAs cada vez mais fortes e simplesmente não temos definições (sequer operacionais) para grande parte desses conceitos pesados e profundos, sobretudo quando pretendemos fazer um estudo comparativo entre as habilidades recém-adquiridas por máquinas e aquelas habilidades-alvo exibidas por humanos. Não temos um teste preciso para saber se uma máquina tem consciência ou não, se ela é, de fato, capaz de compreender um texto ou não. São conceitos esquivos, indóceis. Não temos sistema de mensuração em vista. Não podemos colocar um “conscienciómetro” ou “compreensômetro” debaixo das axilas de um robô para saber o grau de consciência ou de compreensão. E o drama é que estamos desenvolvendo de forma acelerada tecnologias cujo funcionamento somos incapazes

de entender em detalhes (TEGMARK, 2017, p. 79), cujos impactos somos incapazes de prever.

É comum no campo de pesquisa em IA vermos críticas ao Teste de Turing. Seria uma prova fraca com critérios subjetivos, inverificáveis. Entretanto, Turing parece ter captado algo de fundamental na interação humano-máquina. Não importa muito se temos um teste objetivo com critérios claros e indiscutíveis para definirmos se estamos diante de uma máquina inteligente ou não. Já é suficiente sabermos se ela é capaz de nos enganar ou não. Como argumentamos na primeira seção deste artigo, tecnologia sempre foi um jogo de imitação.

3. O dia em que as máquinas cruzaram os umbrais de acesso ao mundo cultural humano

O desenvolvimento acelerado da IA nos impõe uma série de desafios, e devemos começar a separá-los por ordem de urgência. Deixemos alguns técnicos e especialistas do campo da IA trabalhando com filósofos em questões profundas como aquelas perquirições teórico-conceituais de teor mais filosófico de que tratávamos ao final da seção anterior. Os encaminhamentos dessas questões podem nos render bons frutos em termos de definições conceituais e mesmo uma orientação geral em relação ao desenvolvimento tecnológico de nossas sociedades. Porém, devemos entender que estes não são problemas urgentes. São fundamentais, mas não são urgentes.

A prioridade zero está em prepararmos o aparato jurídico-político de nossas sociedades para lidar com o uso sistemático e em escala de tecnologias assistidas por IA. Com sistemas de recomendação e algoritmos mais simples que faziam pouco ou quase nenhum uso desses modelos mais fortes de IA, já estava clara a capacidade destrutiva de campanhas de desinformação no ambiente digital. Antes mesmo que a IA decolasse com toda parafernália técnica que procuramos descrever neste artigo, já havíamos percebido que os ataques de desinformação são capazes de fragilizar mesmo as mais robustas das democracias. Imagine, então, a devastação desinformativa que pode ser desencadeada por algoritmos anabolizados com maquinaria de IA capazes de customizar mensagens com estilo, tom e conteúdo inteiramente elaborados para atingir e convencer um grupo muito delimitado (no limite, um indivíduo). Isso está, portanto, entre os nossos problemas mais urgentes. Nossos sistemas jurídicos e políticos não parecem estar à altura. Desinformação atualmente é um incêndio sob chuva rala. Em tom mais alarmista, alguns especialistas têm afirmado que, antes de roubar nossos empregos e eventualmente nos destruir, a IA promete nos endoidecer por meio da desinformação (HATTENSTONE, 2023).

Assim, é essencial que formemos uma frente de trabalho em que técnicos possam assessorar políticos, juristas e operadores do direito. Nos meses e anos que se seguem, importantes decisões envolvendo o destino comum de nossas sociedades terão que ser tomadas em relação ao tema geral da IA. É claro que, embora os especialistas possam oferecer informações de caráter técnico e também sugestões de ação, a decisão, em última análise, será feita no campo político. Nas democracias, é o campo político que tem a prerrogativa de representar através dos partidos políticos as correntes de opinião e ideologias presentes na sociedade. É o jogo das forças políticas representadas neste campo que vai definir os rumos que nossas sociedades tomarão em relação ao desenvolvimento tecnológico: se vamos controlar nossos artefatos tecnológicos modernos ou seremos controlados por eles ou por quaisquer forças por eles desencadeadas. Por este exato motivo, para que as democracias possam manter, de fato, o poder decisório na mão do cidadão, elas devem ter como, prioridade zero, proteger a si mesmas dos efeitos mais negativos dessas novas tecnologias, sobretudo das campanhas de desinformação coordenadas contra os pilares dos sistemas democráticos modernos: o acordo constitucional, o sistema eleitoral e de representatividade política, a ciência, a imprensa, entre outros elementos. Para que tenhamos tempo de adaptarmos o aparato jurídico-político de nossas democracias para o novo cenário tecnológico precisamos proteger o tabuleiro, as peças e as regras do jogo democrático.

Além da frente de trabalho filosófica e jurídico-política às quais nos referimos acima, devemos nos preocupar também com a percepção pública. Ao menos nas democracias minimamente saudáveis, é o debate na esfera pública que vai ditar os rumos do jogo no campo político. Jornalistas e divulgadores de ciência (tanto nos meios massivos clássicos como nos meios digitais novos) terão que entrar em campo para promover a tradução de assuntos técnicos para público amplo (leigo). Infelizmente, a maioria das pessoas tem uma imagem estereotipada – nos limites do caricatural – quando se trata do tema “ameaças da IA”. Estes cenários mais improváveis e distantes (ainda que não sejam de todo impossíveis) como uma revolução das máquinas tantas vezes retratada pelas lentes de Hollywood devem começar a ser substituídos no imaginário por cenários mais próximos e cada vez mais reais.

Em relação à percepção pública, o desafio mais imediato parece ser convencer as pessoas que, ao menos em estágios iniciais, nossos maiores problemas em relação a IA vão ser decorrentes do modo como essa tecnologia está entrando em nossas vidas e está se apossando das fábricas do imaginário humano. Esta é mais uma ironia com a qual nosso hori-

zonte tecnológico nos acena. Hollywood nos “preparou” para um cenário em que as máquinas nos atacam, tomam nossos recursos e passam a controlar as condições materiais de sustentação de nossas sociedades, a infraestrutura. Se houver alguma revolução ou algo que o valha, virá por cima, pela superestrutura.

3.1 O Homo narrans sem o monopólio de produção cultural

A tese com a qual pretendemos fechar este breve artigo-panorama é que nossos problemas mais urgentes em relação ao acelerado avanço tecnológico das IAs decorrem do acesso que esta tecnologia passou a ter de um mundo antes habitado única e exclusivamente por *Homo sapiens*, o mundo cultural humano.

O que queremos dizer com “acesso ao mundo cultural humano”? Tomemos a palavra “cultura” num sentido *lato* abarcando toda a produção imaterial do *Homo sapiens* – desde produção artística até explicações científicas ou mitológicas, discursos político-ideológicos passando por instituições jurídicas e narrativas religiosas – e também tomemos a palavra “linguagem” em sentido amplo abarcando todo e qualquer sistema de signos (seja de que tipo for [verbal ou não verbal] e seja qual base material tiver [sonora, visual ou qualquer outra]).

Nestas acepções amplas, a linguagem é a chave de acesso para o mundo cultural humano. No caso dos humanos, a cultura parece ser um mosaico audiovisual com uma hegemonia exercida pela linguagem verbal e suas formas derivadas. É uma construção realizada ao longo de eras de forma cumulativa. Reparemos que, do ponto de vista da materialidade sígnica, este é um mosaico predominantemente feito de formas sonoras, sequências e sobreposições de sons; feito de formas visuais matizadas por cores e texturas, combinações de pontos, linhas e planos; e também feito, em menor medida, de materiais sígnicos voltados para os outros sentidos do aparato perceptual humano. É da lógica combinatória dessas formas elementares de materialidade sígnica que as máquinas parecem ter conseguido se apropriar. As máquinas já estão conseguindo organizar estes elementos sígnicos básicos (cf., no caso da linguagem verbal, o conceito de *token* apresentado na segunda seção) dentro de estruturas (sequenciais ou de sobreposição) que já reconhecemos como se fossem produzidas por humanos.

As IAs avançam a passos largos mundo cultural adentro. Problemas técnicos que há alguns anos atrás nos pareceriam intransponíveis podem ser detectados e resolvidos em questão de meses. Por exemplo, reparemos como nos últimos meses (final do ano de 2022 e início de 2023)

foram resolvidos muitos dos problemas de fluência, no caso dos modelos generativos de texto e de ferramentas de tradução, ou do problema particular de elaborar imagens verossímeis de mãos humanas, no caso de modelos generativos de imagens. A prova de que as IAs já caminham com desenvoltura em território cultural humano está no domínio cada vez maior que elas exibem em relação à elaboração de narrativas. Impulsionadas por essa capacidade generativa para texto e imagem (ROOSE, 2022), já vemos máquinas se encaminhando para conseguir reproduzir narrativas relativamente complexas do mundo cultural humano. Por exemplo, o *dramatron* é uma ferramenta que mobiliza modelos de linguagem para elaborar roteiros de filmes ou peças teatrais. Plataformas já estão sendo inundadas por conteúdo sintético (BENSINGER, 2023), por exemplo, livros inteiramente feitos por IA: texto e imagens ilustrativas.

A narrativa é a forma de discursividade elementar da cultura humana. Costuramos nossa vida coletiva e damos forma à nossa herança cultural com histórias. Embora o discurso mais teórico de base descritiva ou dissertativa elaborado no âmbito da filosofia e das ciências seja fundamental, sobretudo para o desenvolvimento tecnológico, o patrimônio cultural humano é construído de forma predominante narrativa. Gostamos mesmo é de contar e ouvir histórias. O problema é que, dependendo das decisões que tomarmos no campo político a respeito do desenvolvimento das tecnologias de IA, não seremos nós que, pela primeira vez, produziremos as histórias que serão contadas para nós mesmos através de gerações. Olhemos, por um instante, para o palco em que promete se desenrolar daqui por diante parte relevante do espetáculo da produção e reprodução da cultura: a internet.

A internet é celebrada como a primeira infraestrutura comunicacional capaz de difundir informação em larga escala com dominância de conteúdo gerado por usuário. Este é o grande *ativo* da Internet. Embora os meios de comunicação de massa (imprensa no modelo *penny press*, rádio, cinema e televisão, por exemplo) já fossem capazes de produzir conteúdo midiático e fazê-lo chegar em muita gente, a produção e difusão, nesses casos, eram centralizadas. O conteúdo que chegava às massas era *necessariamente* produzido por profissionais de mídia. Foi a Internet e todo o aparato digital que conhecemos nas últimas décadas que alteraram esse estado de coisas. O conteúdo dominante que circula no mundo digital passa a ser produzido pelo *usuário da mídia* e não pelo *profissional de mídia*. A grande ironia é que será justamente no palco digital da internet que o conteúdo sintético promete suplantará o conteúdo gerado por usuário (humano).

Sem regulação que promova travas eficientes, a produção de conteúdo sintético não apenas vai suplantar a produção humana, mas será necessário desenvolver ferramentas – provavelmente baseadas em IA – para que os usuários humanos da rede possam se encontrar em meio a essa inundação. Esta é outra ironia: precisaremos de IAs para operarem processos de filtragem de conteúdo produzido por IAs. Sem esse tipo de filtro teremos cada vez mais dificuldade para distinguir conteúdo elaborado por humanos daqueles elaborados de forma híbrida (humano-máquina) e os conteúdos puramente sintéticos.

Se não colocarmos travas para controlar a pressão econômica que será exercida para selecionar sistematicamente o conteúdo sintético em detrimento de conteúdo produzido por humanos, a produção artístico-cultural humana deverá evanescer. A indústria criativa será a primeira a sentir os impactos. Se os critérios forem econômicos, estaremos em clara desvantagem. Na indústria do audiovisual, por exemplo, roteiristas, diretores, atores e uma grande quantidade de técnicos humanos precisam de anos de formação e experiência para alcançar uma performance minimamente qualificada. Tudo isso é custo. Se o mercado fonográfico conseguir entregar para o público músicas compostas, produzidas, gravadas e executadas envolvendo somente agentes de IA, que tipo de reserva de mercado será necessário para recolocar o ser humano nessa cadeia produtiva? Se o jogo a ser jogado estiver totalmente baseado em critérios de eficiência, por exemplo, em relação a metas práticas (margem de lucro, aumento de vendas, captura da atenção, engajamento, etc.), estaremos irremediavelmente em desvantagem. Nesse tabuleiro, artistas humanos não podem competir com máquinas, sobretudo se levarmos em conta a eficiência de algoritmos capazes de isolar preferências de determinados públicos, captar a atenção e oferecer o produto exato que maximize a probabilidade de escolha, compra e engajamento por parte do membro do público.

O problema não é só elaborar travas para “segurar” a pressão econômica direta. Há pressões indiretas que já dão sinais que serão muito relevantes em diversos campos da experiência humana. Por exemplo, discursos políticos desenhados por IA de forma customizada para convencer corações e mentes de grupos cada vez menores, mais nichados e trançados em bolhas informativas dentro das redes sociais digitais serão mais eficientes que seus concorrentes gerados de forma analógica e dependentes da capacidade retórica de uma ou poucas mentes humanas. A pressão indireta é que o grau de eficiência (em termos práticos) do conteúdo sintético será maior e seu custo (em termos financeiros e de tempo), menor.

O caso das religiões é paradigmático de processos de formação cultural no Homo sapiens. Leva-se muito tempo para formar uma religião. São centenas e, em alguns casos, milhares de anos para que valores – capazes de moldar a conduta de indivíduos de forma eficiente – se sedimentem no fundo de narrativas passadas de geração para geração e cultivadas por uma sociedade. Nos casos em que o processo histórico nos é conhecido, sabemos que essas narrativas transmitidas inicialmente no âmbito da tradição oral se fixam em formas escritas, passam a ser protegidas e reproduzidas por estruturas institucionais, o que, por sua vez, garante a elas condições ideais para se espalharem, por exemplo, em formas de expressão artística, penetrarem e lançarem raízes cada vez mais profundas no imaginário de uma cultura. Tudo isso leva muito tempo. Mesmo essas formas contemporâneas mais superficiais, fracas e comoditizadas de espiritualidade pegam carona em narrativas mais enraizadas dentro de uma cultura. Numa época de grandes incertezas para os poucos recursos psíquicos humanos – mudanças tecnológicas rápidas, desastres ambientais continuados e em escala, instabilidade política e econômica – imagine o impacto que deverão ter as primeiras religiões elaboradas, desenhadas e implementadas por IAs. Em resumo, se os critérios forem econômicos e de eficiência, os morosos “modos” analógicos/orgânicos de produção e reprodução de nosso patrimônio cultural estarão em desvantagem. Será mais barato e mais eficiente fazer política e cultivar religião gerada por IA.

É por este exato ponto que começamos o artigo: devido à capacidade dessa tecnologia em permear toda atividade produtiva, qual campo da experiência humana ficará de fora? Ao adentrar o território cultural, não deve mais haver áreas exclusivamente humana. Pode parecer paradoxal, mas essas mudanças drásticas que se anunciam como novidade já estavam pré-figuradas na história de nossas tecnologias. O desenvolvimento tecnológico parece ter seguido uma trajetória bem definida: da mão para os arcanos da alma dos criadores.

Nos primeiros estágios de nossa caminhada tecnológica, ao inventarmos instrumentos, preocupávamo-nos em desenvolver cabos e manivelas adequadas para que pudéssemos manuseá-los de forma eficiente. Depois passamos a nos ocupar com modos de desenvolver botões e displays que nos facilitassem o controle de nossos inestimáveis instrumentos. Mais recentemente, com o convívio com maquinaria cada vez mais inteligente, passamos a nos preocupar o chamado problema de *interface*. Perdíamos o sono quando notávamos que nossas máquinas e seus programas não tinham interfaces amigáveis para os usuários. E, nesses

últimos dias, começamos a nos preocupar se nossas máquinas estavam ou não entendendo o que falávamos. Já não se trata, portanto, de uma relação/interação “amigável” em sentido figurado. A metáfora se fez carne.

Este desenvolvimento recente da IA marca um ponto de virada dentro de um longo processo de transição. Esta foi uma transição que veio sendo fomentada continuamente através dos milênios pela vontade humana de ter seu trabalho facilitado, ter sua rotina produtiva aliviada e, sobretudo em tempos históricos mais recentes, vem sendo moldada (em nível macro) pela pressão econômica de ter os custos da produção barateados. É uma transição que, ao longo da história humana, veio sendo orientada pelo que chamamos, ainda na primeira seção, de vetor de eficiência/autonomia do desenvolvimento tecnológico. A pergunta que nos interpela nessa quadra histórica é se seremos capazes de controlar essas forças que nos trouxeram até onde nos encontramos. Esta é provavelmente a questão mais profunda que a IA, como pináculo de nosso desenvolvimento tecnológico, pode nos colocar.

Estamos nos aproximando do ponto a partir do qual passaremos a ter dificuldades para enxergar essas máquinas pensantes e cada vez mais falantes como *nossos* instrumentos. Essa dificuldade promete ser proporcional ao grau de autonomia que elas passarem a exibir em relação às habilidades intelectivas humanas para as quais foram desenvolvidas para imitar. Nossa tese neste artigo-panorama é que entramos numa nova era da relação humano-máquina, justamente porque, até este ponto de virada, desenvolvíamos artefatos que eram indiscutivelmente objeto de nossa criação e ponto final. Não passavam disso. A partir desse ponto de virada teremos desenvolvidos artefatos que serão sujeitos de suas próprias criações. Prometeu. Novamente. Estamos passando adiante a tocha.

3.2 A pergunta fundamental diante da IA

A pergunta fundamental a ser respondida diante dos desafios impostos pela IA é: como vamos entrar nesse processo histórico? A esta altura já está claro que não podemos mais escolher não entrar. Por um lado, nosso desenvolvimento tecnológico, por *dependência de trajetória*, nos colocou face a face com as IAs e, por outro lado, nosso pujante mosaico de sistemas econômicos globalizados – que operamos como se fosse uma bicicleta-geringonça que não podemos parar de pedalar – nos impede de abortarmos. Como vai ser? Vamos entrar numa espiral ensandecida de mudanças disruptivas que vão nos levar a atravessar períodos de instabilidade econômica e política para provavelmente desembocarmos cenários

distópicos? Ou vamos escolher descer a mais íngreme das ladeiras de nosso desenvolvimento tecnológico com a mão no freio, gerenciando riscos, administrando, mitigando e internalizando externalidades negativas, antecipando e evitando cenários futuros indesejáveis?

O sonho maior da modernidade era que pudéssemos determinar nosso destino no nível individual e coletivo. Um ideal de independência, emancipação, autogoverno, autodeterminação, enfim, autonomia. O conjunto variado de desafios impostos pela IA é uma excelente oportunidade para demonstrarmos que somos nós que estamos no controle de nosso destino histórico e não forças que somos incapazes de controlar (embora juremos que tenham sido criadas por nós mesmos). Não é apenas uma *excelente oportunidade*. Talvez seja a última.

Os desafios impostos pela IA prometem levar o projeto moderno para o banco dos réus pela última vez. A modernidade já foi analisada, re-analisada, virada de cabeça pra baixo, julgada, desmascarada, condenada e, por fim, declarada, como projeto, ultrapassada. Mesmo desencantado e com suas instituições liquefeitas, o projeto moderno consegue, provavelmente por inércia, ser a causa eficiente das grandes mudanças que vemos no planeta atualmente para o bem ou para o mal, para frente ou trás.

Se quisermos comparar a capacidade disruptiva da IA com algum processo histórico anterior, a referência óbvia que temos é um tema pelo qual já passamos algumas vezes neste artigo: a Revolução Industrial. Este é um processo histórico que elevou como nunca a capacidade produtiva de nossas sociedades, porém se mostrou um manancial de contradições sociais que foram replicadas nos últimos séculos em diversos cantos do globo adquirindo cores locais onde quer que chegasse. O saldo devedor é considerável. Na verdade, ainda há prestações a vencer. Vide mudanças climáticas. Esse raciocínio pode ser generalizado da Revolução Industrial para todo o projeto moderno. O preço que boa parte da humanidade pagou para que executássemos os ideais emancipatórios da modernidade foi alto: exploração colonialista e genocídio; escravidão e suas heranças malditas na forma de sociedades racializadas; jornadas intermináveis, exaustivas e mal remuneradas de trabalho livre.

Se a experiência histórica consegue nos ensinar alguma coisa, deveríamos olhar com desconfiança a promessa emancipatória contida no discurso laudatório da IA (que nada mais é que a manifestação contemporânea da ideologia moderna do progresso) e começarmos a antecipar os problemas. É óbvio que vamos colher bons frutos dessa tecnologia. As promessas são altas: cura de doenças, erradicação da fome, aumento do tempo livre, democratização da informação e conhecimento, entre outras

promessas, algumas requestradas da época da Revolução Industrial, outras que são apresentadas como novidades. Mesmo que não seja entregue conforme o prometido, é claro que poderemos colher benefícios que seriam inimagináveis para nosso ancestrais. Entretanto, devemos notar não apenas que as apostas são muito altas, mas que também temos um longo histórico de derrotas na mesa de jogo. Com esse histórico e tudo que está em jogo, não podemos nos dar ao luxo de sermos ingênuos. Deve-se ficar com um olho nos problemas gerados por processos históricos similares/comparáveis e o outro olho nos problemas que provavelmente serão gerados pelo avanço dessa tecnologia. Iluministas desencantados devem ter um dos olhos voltado para as sombras dos processos históricos pretéritos e o outro, voltado para sombras que se projetam à frente.

Os alertas vindos dos próprios pesquisadores e especialistas no campo de IA colocam como pior cenário aquele em que o desenvolvimento tecnológico correria ao sabor de uma concorrência desenfreada entre empresas privadas ou uma nova corrida armamentista entre nações. Reparemos que a preocupação, neste caso, está voltada para o mesmo problema central observado em dois planos distintos – um deles econômico, o outro, geopolítico – com atores e regras diversas. Chamemos este problema central de *lógica concorrencial descerebrada*.

No final de março de 2023, importantes cientistas, pesquisadores e empresários do campo assinaram uma carta pedindo uma “pausa” na pesquisa/desenvolvimento de tecnologia de IA. Inaudita cautela. Seria a primeira vez que isso seria feito nas modernas condições de produção e desenvolvimento tecnológico. Não temos experiência praticamente nenhuma em sermos cautelosos quando lidamos com a base produtiva de nossas sociedades e as forças produtivas liberadas por nossas tecnologias. A regra que vale é: “primeiro inventa e implanta, depois lidamos com as externalidades negativas”. Essa regra pressupõe que isso que estamos chamando de *lógica concorrencial* operaria para resolver os problemas, restituir um equilíbrio inicial perdido. Balela. Uma lógica puramente competitiva nos afasta de condições que favoreçam, por exemplo, que fossem implementadas legislações internas nos países e acordos e tratados internacionais para controle das IAs.

Sem qualquer tipo de freio para a *lógica concorrencial descerebrada* qualquer proposta como a “pausa” acima mencionada seria impraticável. Se, de forma solitária, uma das empresas decide parar com as pesquisas/desenvolvimento de tecnologia de IA, ela será devorada pela concorrência, e a pessoa à frente dos negócios desta empresa será devorada pelos acionistas. Então, dada a *lógica concorrencial*, não há condições favorá-

veis para que se tomem decisões assim (indivíduos conscienciosos e empresas responsáveis não vão agir). A mesma lógica se reproduz no plano da geopolítica global. O tabuleiro geopolítico vem sendo redesenhado nos últimos tempos por uma série de fatores. Para ficarmos com apenas dois deles: 1) a pujança econômica tem alçado a China à posição de potência global projetando no horizonte o fim da hegemonia americana (indiscutível no pós-guerra fria); 2) a guerra na Ucrânia, desencadeada pela invasão russa ocorrida em fevereiro de 2022, já se apresenta como uma jogada dentro do novo jogo geopolítico cujas regras não ditadas pela hegemonia americana. No cenário em que este novo mundo multipolar desemboca numa nova Guerra Fria, teríamos condições gerais que desfavorecem o clima de cooperação internacional e alimentam continuamente a desconfiança entre as nações. A corrida armamentista-tecnológica entre nações que encarnam projetos político-ideológicos completamente diferentes e que se enxergam como adversários-concorrentes cria dificuldades intransponíveis para os acordos e tratados internacionais como os mencionados acima.

Tanto no campo econômico como no campo geopolítico, há radicais que defendem de forma irrestrita a *lógica concorrencial descerebrada*. Regulação zero. Este radicalismo deve ser combatido. O desafio civilizatório é justamente domar a *lógica concorrencial descerebrada*. Esta é a besta-fera a ser domada. Como já afirmamos, nossas apostas com a IA são altas demais e nosso histórico de derrotas na mesa de jogo nos deixou com um saldo devedor considerável. Essa lógica concorrencial descerebrada tem dificuldade em disfarçar seus instintos mais básicos e sua vontade irrefreável de restituir a barbárie e a selvageria no coração da civilização. Seria a última das vergonhas que fracassássemos em domá-la. Retomemos, para finalizar esse breve ensaio sobre IA e nosso destino histórico, a pergunta central.

A pergunta é se vamos entrar neste processo histórico numa espécie de “modo reflexivo-autocontido”, i.e., com consciência dos riscos e com a mão no freio, ou vamos entrar no modo “irreflexivo-aloprado”, i.e., “deixe que a lógica concorrencial faça o trabalho dela e depois lidamos, na medida do possível, com as externalidades negativas”. Se recolocarmos esta questão nos termos do esquema de análise introduzido neste artigo, a pergunta passa ser se vamos ou não conseguir controlar as forças que têm impulsionado nosso desenvolvimento tecnológico sempre na direção que aponta o vetor que chamamos de autonomia/eficiência. A questão central é se, de olho ávido na eficiência, vamos ou não conseguir dosar o grau de autonomia a ser concedida às máquinas. Seremos um Prometeu homeopático, paciente e cauteloso? Na narrativa mítica, Prometeu foi duramente castigado. E nós, seremos castigados ou nos acorrentaremos preventivamente?

Referências

- ASHTON, Thomas. S. *The Industrial Revolution (1760–1830)*. London: Oxford University Press, 1948.
- BENDER, Emily M.; *et al.* On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. New York, NY: Association for Computing Machinery, p. 610–623, 2021.
- BENSINGER, Greg. ChatGPT launches boom in AI-written e-books on Amazon. *Reuters*, 21 fev. 2023. Disponível em: <https://www.reuters.com/technology/chatgpt-launches-boom-ai-written-e-books-amazon-2023-02-21/>. Acessado em: 24/03/2023.
- FLORIDI, Luciano; CHIRIATTI, Massimo. GPT-3: Its nature, scope, limits, and consequences. *Minds & Machines*, 681–694, 2020.
- HATTENSTONE, Simon. Tech guru Jaron Lanier: ‘The danger isn’t that AI destroys us. It’s that it drives us insane’. *The Guardian*, 23 mar. 2023. Disponível em: <https://www.theguardian.com/technology/2023/mar/23/tech-guru-jaron-lanier-the-danger-isnt-that-ai-destroys-us-its-that-it-drives-us-insane>. Acessado em: 24/03/2023.
- HOBBSAWM, Eric. *Da revolução industrial inglesa ao imperialismo*. Trad. D. M. Garschagen. Rio de Janeiro; Forense Universitária, 1986.
- HU, Krystal. ChatGPT sets record for fastest-growing user base – analyst note. *Reuters*, 2 fev. 2023. Disponível em: <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>. Acessado em: 02/03/2023.
- KREBS, Robert. E.; KREBS, Carolyn. A. *Groundbreaking scientific experiments, inventions, and discoveries of the ancient world*. Westport, CT: Greenwood, 2003.
- MATHIAS, Peter, & DAVIS, John. A. (Eds.) *The First Industrial Revolutions*. Oxford: Blackwell, 1989.
- MCLUHAN, Marshall. *A galáxia de Gutenberg*. Tradução L. G. de Carvalho; A. Teixeira. São Paulo: Editora Nacional; Editora da USP, 1972.
- MINSKY, Marvin; PAPERT, Seymore. *Perceptrons*. Cambridge, MA: MIT Press, 1969.
- MORAVEC, Hans. *Mind children*. Cambridge, MA: Harvard University Press, 1988.
- PINKER, Steven. *The language instinct: How the mind creates language*. New York, NY: Harper Perennial Modern Classics, 1994.

ROOSE, Kevin. The brilliance and weirdness of ChatGPT. *New York Times*, 5 dez. 2022. Disponível em: <https://www.nytimes.com/2022/12/05/technology/chatgpt-ai-twitter.html>. Acessado em: 02/03/2023

RUSSELL, Stuart J.; NORVIG, Peter. *Artificial intelligence: A modern approach*. London: Pearson, 2021.

RADFORD, Alec *et al.* Improving language understanding by generative pre-training. *Technical Report*. OpenAI, 2018. Disponível em: <https://www.cs.ubc.ca/~amuhamo1/LING530/papers/radford2018improving.pdf>. Acessado em: 04/09/2022.

SANTAELLA, Lucia. O homem e as máquinas. In: DOMINGUES, Diana (org.). *A arte no século XXI*, São Paulo: Editora da UNESP, São Paulo, p. 33-44, 1997.

_____. *Neo-humano: a sétima revolução cognitiva do Sapiens*. São Paulo: Paulus, 2022.

SEJNOWSKI, Terrence J. *The deep learning revolution*. Cambridge, MA: MIT Press, 2018.

SILVER, David *et al.* Mastering the game of Go with deep neural networks and tree search. *Nature*. n. 529, p. 484-489, 2016.

STJERNFELT, Frederik. The semiotic body. A semiotic concept of embodiment? In NÖTH, Winfried (ed.), *Semiotic bodies, aesthetic embodiments, and cyberbodies*. Kassel: University Press, 2006, p. 13-48.

TAULLI, Tom. *Artificial intelligence basics: A non-technical introduction*. Monrovia: Appres, 2019.

TEGMARK, Max. *Life 3.0. Being human in the age of artificial intelligence*. New York, NY: Allen Lane, 2017.

VASWANI, Ashish *et al.* Attention is all You Need. *arXiv*: 1706.03762, 2017.