HISTÓRIA MAIS DO QUE HUMANA:

descrevendo o futuro como atualização repetidora da Inteligência Artificial

More-than-human history: Describing the future as a repeating update of Artificial Intelligence

Rodrigo Bragio **BONALDO**

rodrigobonaldo@yahoo.com.br

Universidade Federal de Santa Catarina, Departamento de História, Centro de Filosofia e Ciências Humanas. Florianópolis, SC, Brasil.

RESUMO

Este artigo analisa sistemas de inteligência artificial como agentes temporalizadores em uma história mais do que humana. Engaja-se em uma discussão interdisciplinar abrangendo a sociologia dos algoritmos, a ética informacional e a filosofia da mente para, com base na teoria da história, avaliar como a onipresença da automação adiciona novas dimensões ao estudo da condição histórica contemporânea. O argumento diz que a temporalização artificial se origina da simulação da linguagem natural por meio de algoritmos de aprendizado. O artigo esclarece isso através de uma descrição em código e uma demonstração em imagem técnica, ilustrando a 'renderização' ou 'digitalização' de experiências humanas em padrões de dados, posteriormente replicados ou atualizados por agentes artificiais. Esse processo destaca a vetorização e manipulação estatística da experiência e representa o que denomino "computação necessária". Em contraste, o texto sugere que uma "computação contingente", capaz de produzir algo novo, transcende o mero discurso técnico. Assim, enfatiza-se o papel da ética e transparência na gestão de dados e no treinamento de sistemas de IA, cruciais para compreender a Inteligência Artificial como um campo histórico aberto a formas cognitivas potencialmente não-humanas.

Palavras-chave: inteligência artificial; atualismo; futuros históricos

ABSTRACT

This article examines artificial intelligence systems as temporalizing agents in a more-than-human history. It engages in an interdisciplinary discussion across the sociology of algorithms, informational ethics, and philosophy of mind, informed by historical theory, to assess how automation's ubiquity adds new dimensions to the study of the contemporary historical condition. The central argument is that artificial temporalization stems from natural language simulation via learning algorithms. The article clarifies this through a code description and a technical image demonstration, illustrating the 'rendering' or 'digitization' of human experiences into data patterns, further replicated or updated by artificial agents. This process highlights the vectorization of human experience and stands for what I term 'necessary computation'. In contrast, the piece also argues that 'contingent computation', capable of novel outputs, transcends mere technical discourse. It emphasizes the role of ethics and transparency in data management and AI system training, crucial for comprehending Artificial Intelligence as a historical field open to diverse, potentially non-human, cognitive form.

Keywords: artificial Intelligence; updatism; historical futures

Não pertencemos mais à família de heróis trágicos que posteriormente descobriram que haviam preparado seus próprios destinos. Agora sabemos disso de antemão.

(LUHMANN, 1998, p. 74).

O conceito de mundos "mais do que humanos" tem sido utilizado em campos como a história ambiental, a sociologia da tecnologia e a dos algoritmos, e também nos estudos culturais e na teoria da história. Ele refere-se à análise das interações entre humanos e não-humanos como fatores da mudança social. Uma história mais do que humana não seria, portanto, uma história não humana – como se a disciplina pudesse se desinteressar por seres humanos – mas, simplesmente, uma história que, assumindo uma perspectiva relacional, não se limita a ser uma ciência dos *homens* no tempo. É minha convicção que essa abordagem seja relevante para entendermos o impacto das transformações históricas derivadas, sobretudo, de fatores ambientais e tecnológicos. Ao mesmo tempo, ela nos permite enfatizar e problematizar os graus de agência e autonomia desses fatores, evidenciando a dimensão ético-política inscrita nessas relações sociais. No sentido tecnológico, o mais do que humano sugere ainda que o "Ciborgue" deixou de ser uma metáfora, uma imagem ou "mito político" em forma de manifesto (HARAWAY, 2000), para se tornar parte integral de nossa condição histórica.

Este artigo tem como objetivo apresentar e problematizar os limites de se entender sistemas de Inteligência Artificial – com particular atenção à IA Generativa e aos modelos de aprendizado de máquina – como agentes temporalizadores em uma história mais do que humana. É através de uma discussão interdisciplinar (em especial com a sociologia dos algoritmos, a ética informacional e a filosofia da mente), mediada por problemas do campo da teoria da história, que busco explorar como a ubiquidade das tecnologias de automação traz um novo componente para o estudo das relações entre humanos e não-humanos. Busco também lançar um problema teórico que liga a natureza dessas relações com a atual condição histórica. Problematizar os avanços tecnológicos pode contribuir para informar decisões relativas aos *riscos*: para pensar com Niklas Luhmann, "o risco é [...] uma forma de *descrições presentes do futuro*, sob o ponto de vista de que se pode decidir [...] por uma ou outra alternativa" (LUHMANN, 1998, p. 71). Busco inspiração, igualmente, nas palavras de Ewa Domanska, para quem superar o antropocentrismo registrado pela própria noção de *humanidades* "não é escolher um tema da moda", nem sequer "considerar uma abordagem epistemológica, mas principalmente assumir uma escolha ética orientada para o futuro" (DOMANSKA, 2010, p. 120). Em suma, particularmente interessada em conceitos de movimento como tempo e história, a estratégia deste artigo quer iniciar um trabalho de descrição, a partir do campo semântico da IA, de modos pelos quais "o futuro se manifesta no presente" (LUHMANN, 1998, p. 63).

Esse futuro, não raramente previsto e moldado pela inteligência artificial, tem os algoritmos como protagonistas. Eles são vitais para o aprendizado de máquina, facilitando a análise de vastos conjuntos de dados e reconhecendo padrões. Tais procedimentos desempenham um papel crucial na geração de prognósticos e na tomada automatizada de decisões, evidenciando como o futuro é integrado ao nosso presente. Um algoritmo pode ser definido como "uma estrutura de controle finita, abstrata, eficaz e composta, imperativamente dada, cumprindo um determinado objetivo sob determinadas disposições" (HILL, 2015, p. 47). Essa elaboração pragmática e filosófica encontra nas ciências da computação definições que falam de construções matemáticas empregadas enquanto técnicas computacionais na

resolução de problemas. Essa literatura descreve algoritmos como programas operados por coleções de valores de entrada (*input*) capazes de produzirem soluções de saída (*output*). Mas não é incomum encontrarmos, nesses mesmos textos, explicações metafóricas derivadas do discurso popular: algoritmos são como receitas culinárias, roteiros de filmes, ou ainda como um GPS que mapeia o caminho que devemos seguir para encontrar o destino desejado (SICHMAN, 2021, p. 38). Isso ocorre porque discursos públicos não tratam algoritmos apenas como constructos matemáticos, mas também os definem a partir de suas lógicas de implementação. Como já foi notado por eticistas da informação, "faz pouco sentido considerar a ética dos algoritmos de modo independente de como eles são implementados e executados em programas de computador" (MITTELSTADT *et al.*, 2016, p. 2). É na interação entre humanos e computadores – "*human–computer interactions*", ou HCI –, que busco compreender a emergência de experiências relacionais constitutivas do que chamo de uma história "mais do que humana". Nesse contexto, a crescente autonomia de agentes artificiais como mediadores de relações sociais e de formas de interação com o mundo, proporcionada pelo aprendizado de máquina, tem sugerido a problematização de uma condição mais do que humana ao mesmo tempo em que ressalta a urgência de formularmos novos conceitos capazes de significá-la (FLORIDI, 2014, p. VIII-IX; TAMM; SIMON, 2020).

Apresento duas hipóteses contrárias, mas não obrigatoriamente contraditórias, desenhadas para a exploração dessa condição histórica. A primeira é descritiva e a chamo de hipótese da "computação necessária". Se as formas de temporalização articuladas pela IA dependem do reconhecimento de padrões a partir de bases de dados dentro das quais a natureza da informação histórica capturada é formalmente ordenada e discretizada, a performance das máquinas de aprendizado – ou o que poderíamos chamar de sua produção ôntico-ontológica de futuro[1] – teria como resultado a repetição desses mesmos padrões agora apresentados como prognósticos? Encontro na teoria do atualismo uma imagem que poderia conceituar essa "temporalização do tempo" (PEREIRA; ARAUJO, 2018, p. 34) artificial.

Atualismo é definido como uma "historicidade em que um presente vazio e autocentrado se relaciona vaga e pragmaticamente com o passado, enquanto o futuro é desejado como reserva para a expansão linear deste presente em constante atualização repetidora" (PEREIRA; ARAUJO, 2022, p. 72-73). Em uma apropriação que certamente não esgota as potencialidades existenciais da tese do atualismo, apresento, na parte final deste artigo, uma descrição em código e uma demonstração em imagem do que seria uma atualização como repetição de padrões informacionais executada por um agente artificial.

A hipótese da computação necessária sugere que processos matemáticos formais e abstratos não são capazes de capturar a experiência humana em toda sua complexidade; pelo contrário, criam classificações estáveis que reificam padrões de historicidade ligados às desigualdades de longa duração. A segunda hipótese que apresento é especulativa e, tomando emprestado o conceito de Beatrice Fazi (2018), a denomino "computação contingente". Mesmo que com ela busque, já na conclusão, mencionar perspectivas que visam desafiar o caráter "necessário" dos processos computacionais, é minha convicção de que a elaboração de uma "computação contingente" (ou seja, de uma computação que seja capaz de produzir algo *novo*) não é uma tarefa esgotável por discussões técnicas. Trata-se de uma batalha ético-política que tem como centro o entendimento do campo da Inteligência Artificial como uma disciplina histórica (HUGHES-WARRINGTON, 2022, p. 108), na qual os papéis dos treinadores de IA e da transparência com relação a *priors* e à curadoria de dados tornam-se igualmente centrais. Esse é um argumento projetado para o estudo dos impactos de técnicas de processamento de linguagem natural (NLP) aplicados na IA Generativa. Ainda que seja sedutor identificar operações algorítmicas (*input*, processamento

e *output*) com as três fases da operação historiográfica analisadas por Paul Ricœur, convém ressaltar que significados artificiais não são articulados na tessitura de uma intriga, mas apreendidos via discretização de classes e subclasses de palavras na estrutura distributiva da linguagem (HARRIS, 1954; LE; MIKOLOV, 2014). Em suma, o "tempo contado", descrito por Pedro Telles da Silveira, contém uma racionalidade alienígena. Ela é "microtemporal", algo imune à narrativa, e "resulta da exteriorização seguida pela autonomização em dispositivos técnicos de aspectos que antes eram considerados exclusivamente humanos" (SILVEIRA, 2023, p. 23-24).

Mas, então, se mesmo assim compreendemos que a IA realiza gestos temporalizadores, como devemos escrever a história da IA? Sem o objetivo de encampar um ou outro modelo, cabe apresentar algumas alternativas, a seguir divididas de modo provisório e heurístico. A primeira, e mais comum, é êmica ao campo e anuncia-se como uma metanarrativa de tipo sazonal. Ela conta uma história entrelaçada com a história da computação. Primavera, verão, outono, inverno, primavera e verão de novo: no primeiro tempo das cores, Alan Turing e *The Bombe* decifrando Enigma, seguidos, em 1956, pela cunhagem do termo Inteligência Artificial, durante o seminário de Marvin Minsky e John McCarthy; na estação do sol, ELIZA, um dos primeiros modelos de linguagem capaz de simular conversas com humanos, seria desenvolvido entre 1964 e 1966; na estação das folhas, as expectativas estivais a respeito da iminente emergência de uma Inteligência Artificial Geral (AGI, expressão para a emulação da inteligência humana) são frustradas e o congresso dos EUA começa a criticar os gastos com IA; na estação do frio – também uma época associada à reflexão e à renovação interior – as primeiras redes neurais artificiais são criadas, embora os computadores ainda não tivessem poder de processamento nem dados suficientes para torná-las funcionais (HAENLEIN; KAPLAN, 2019, p. 5-8). Essa metanarrativa organiza a história da IA em função das condições de financiamento e do progresso técnico da indústria. Hoje estaríamos entre uma nova primavera, iniciada com o desenvolvimento do conexionismo e os avanços do aprendizado de máquina, e um novo verão, característico do surgimento do *big data* e da implementação de técnicas de aprendizado profundo (*deep learning*).

O segundo grupo de alternativas surge da crítica da técnica e ganha contornos político-econômicos. Seguindo a história dos algoritmos, começa descrevendo o desenvolvimento da IA em três fases: uma "Era Analógica", espécie de pré-história dos autômatos que vai até 1945, seguida de uma "Era Digital" (1946-1998), com o surgimento e popularização dos computadores, e de uma "Era das Plataformas", a qual iniciaria em 1998 com o lançamento do buscador *Google* (AIROLDI, 2022, p. 8-14). Esse conjunto de metanarrativas, que aportam na Era das Plataformas, também nos trazem modelos que situam a discussão desde a história do capitalismo, fazendo uso de termos coligatórios inspirados pela economia política que apostam no "capitalismo de vigilância" e no "colonialismo digital" como conceitos capazes de unificar o debate, não raramente apresentando-os como característicos de uma nova fase do atual modo de produção. Nesse sentido, as plataformas, em parte responsáveis por uma "acumulação primitiva de dados" (FAUSTINO; LIPPOLD, 2023, p. 91-126), se tornam os "componentes de uma organologia do tempo presente, confluindo digitalidade e arte neoliberal de governar" (KOSTECZKA, 2022, p. 91). Além de configurarem um tipo de metanarrativa que busca domesticar os avanços tecnológicos a partir de fenômenos de longa duração (colonialismo, capitalismo, etc.), essas perspectivas também agregam elementos singulares. Em outras palavras, novas experiências têm exigido novos conceitos que as confiram inteligibilidade e significado. A expressão "sem precedentes" é aqui utilizada em duas acepções. Na primeira delas, de um ponto de vista substancial, ela pôde designar um tipo de "poder instrumental", a partir do qual as *Big Techs* extraem "superávit comportamental" e o utilizam não para edificar um "novo homem", mas para

condicionar nosso comportamento de modo operante para fins político-econômicos. Essa "forma estranha de poder" vigilante seria irredutível ao totalitarismo (ZUBOFF, 2019, p. 394). Os representantes mais extremos desse grupo chegam a declarar que "esse tipo estranho de economia política, baseada não na escassez de coisas, mas no excesso de informação", demonstra que o capitalismo teria chegado ao fim (!), sendo substituído por "algo pior": um novo modo de produção dominado por uma classe que possui os "vetores da informação" (WARK, 2022, p. 13-77). Na segunda acepção, o "sem precedentes" aparece de um ponto de vista quase-substancial, referindo-se não ao ineditismo da experiência histórica, mas à natureza mesma da *mudança* histórica no escopo da modernidade. Previsões e correlações algorítmicas "geram suas próprias temporalidades, nas quais o futuro é uma versão altamente seletiva do passado" (HONG, 2022, p. 384), engendrando uma microtemporalidade própria das máquinas, que – a despeito do vocabulário técnico e da percepção dos usuários – é um marcador menos processual do que "evental da mudança" (SIMON, 2021, p. 145). De qualquer forma, tais diagnósticos buscam conceituar os impactos de uma revolução tecnológica igualmente descrita como "sem precedentes" (FAUSTINO; LIPPOLD, 2023, p. 35).

Essas abordagens se chocam com as promessas tecnofuturistas do campo da IA. Nesse terceiro grupo, uma espécie de Espírito da computação, matriz do desenvolvimento de uma consciência-de-si artificial, é muitas vezes tratado como *step function* da Inteligência Artificial Geral. Para essas cronosofias, a AGI é uma meta, mas ainda assim não exatamente um *telos*: a literatura descreve também uma "singularidade tecnológica", uma fase que seguiria um evento único de "explosão da inteligência" a partir do qual mentes inteligentes melhoradas tecnologicamente ou baseadas em *software* entram em um "ciclo de autoaperfeiçoamento descontrolado" (EDEN *et al.*, 2012, p. 2). A relevância dessas previsões não se baseia na precisão de sua realização, "pois sua função não é prever eventos futuros, mas sim obter legitimidade e plausibilidade do futuro para autorizar ações antecipadas no presente" (HONG, 2022, p. 377), o que inclui busca por financiamentos de pesquisa. Entre seus expoentes mais famosos, encontram-se Ray Kurzweil, Max Tegmark, Nick Bostrom e Yuval Noah Harari, enquanto o bilionário Elon Musk atua como porta-voz e financiador das preocupações existenciais singularitárias. Esse tipo de prognóstico varia do pânico nutrido de ficção distópica ao entusiasmo com as possibilidades de uma IA superinteligente. Alguns de seus advogados chegam a sugerir que a IA seja o assunto "mais importante de nosso tempo", ainda mais relevante do que questões ambientais, uma vez que esses inventos têm o potencial de nos dar "tecnologia para mitigar a mudança climática" (TEGMARK, 2020, p. 502). A singularidade é, antes de tudo, uma hipótese aceleracionista de *ethos* neoliberal.

Mas se os "limiares da catástrofe" são "sempre definidos em termos sociais" (LUHMANN, 1998, p. 70), isso significa que seus contornos são desenhados por disputas pela hegemonia do discurso. As profecias da singularidade tecnológica, apresentadas como prognósticos matemáticos nos quais a sociedade evolui como PA e a técnica avança como PG, já inspiraram a reação negativa de expoentes da ética informacional. Luciano Floridi chegou a clamar por um novo inverno, por um novo período em que os jornalistas suspendem o *hype* em torno da IA e os pesquisadores voltam a discutir temas mais urgentes – como a manipulação das nossas escolhas, a privacidade dos dados, os ciberconflitos e os crimes digitais – sem desviar o foco das antecipações fundadas em cenários de ficção científica. Sobre os singularitários e os profetas do apocalipse, Floridi é taxativo: "deveriam ter vergonha e pedir desculpas" (FLORIDI, 2020, p. 2). A situação apenas se agravou nos anos que nos separam desse importante editorial, com os lançamentos públicos de modelos de linguagem larga (LLMs) e IA Generativa de imagens. O contra-ataque do instituto *Future of Life* documenta o deslocamento do interesse público do discurso catastrófico da singularidade para as

preocupações a respeito do alinhamento cultural dos sistemas de IA.[2] Essa é uma vitória da ética informacional que, pensada como um quarto modo de escrever a história da IA, identificou algoritmos como agentes sociais e hoje discute a possibilidade da agência artificial moral (AMA). Esse é um assunto pertinente no momento em que LLMs sugerem o rompimento de uma primeira "barreira semântica" (FLORIDI, 2014, p. 142), momento que chamarei de *Sattelzeit* das máquinas. A linguística distribucional proposta por Zellig Harris (1954) teve de esperar até o advento do *big data* e a popularização das redes neurais artificiais para se tornar funcional. Na segunda década do século XXI, assistimos ao aparecimento de técnicas como *bag-of-words* e suas sucessoras, as quais representaram um salto para o processamento da linguagem natural (NLP). Como vou argumentar, a análise microssemântica e microtemporal facilita o processamento de conteúdos linguísticos e, ao simular o domínio da linguagem, a IA otimiza tarefas comunicativas que atuam no âmbito de uma *agência artificial temporalizadora*.

É assim que contrasto a ruptura desse primeiro *limiar semântico* com alguns aspectos do teorema koselleckiano, questionando, em particular, o potencial das máquinas em temporalizar – conectando passado, presente e futuro – a experiência humana. Os GAFAMI (Google, Apple, Facebook, Amazon, Microsoft e IBM) não representam apenas centros empreendedores de tecnologia de ponta. Seus objetos técnicos transformam o caráter das relações sociais e conexões humanas, modificando o "*sentido de viver junto*, conforme era entendido pela modernidade liberal" (CANCLINI, 2021, p. 22, grifos do autor). Após o que chamarei de *Sattelzeit* das máquinas, uma profunda dessincronização entre mudança social e mudança técnica se manifesta na experiência direta. Como afirma Silveira (2023, p. 11), "os modos temporais apontam para a introdução de muitos tempos diferentes no tecido da vida diária", desafiando o tempo histórico. É essa variedade temporal que convido a capturar a partir da agência algorítmica.

Dito de outra forma, este artigo inaugura uma reflexão que busca acessar os diagnósticos de desorientação e dessincronização temporal, tão representativos da teoria da história, menos a partir da aceleração social e mais através do estudo da multiplicação de agentes temporalizadores artificiais na modernidade tardia. Não é apenas o ritmo acelerado da mudança tecnológica com relação à mudança social, nem ainda um consequente "fechamento de futuro", que nos ajuda a conceituar a condição histórica mais do que humana. Não é a eliminação da contingência, mas as técnicas de seu *controle*, que estão em jogo (HONG, 2022, p. 380). Afinal, os *futuros presentes* e suas formas de antecipação – como modalidades de transição a *presentes futuros* – não podem deixar de ser marcados por elevado grau de contingência: outra palavra para aquilo que "não é necessário nem impossível" (LUHMANN, 1998, p. 45).

Portanto, sugiro que a aceleração seja um fator necessário, mas não suficiente, para a conceitualização de nossa condição histórica. Ela é o elemento bem conhecido, do qual precisamos aprender a nos distanciar, ao menos um pouco, a fim de elaborar nossa "síndrome da carruagem sem cavalos, na qual atrelamos nossa sensação de perigo inédita a fatos velhos e familiares", uma vez que "as conclusões às quais eles nos conduzem são necessariamente incorretas" (ZUBOFF, 2019, p. 393). Precisamos de maior "poder criativo para insistir num futuro construído por nós" (ZUBOFF, 2019, p. 393), pois a imagem benjaminiana de puxar a corda de tração do sino, indicando ao maquinista possíveis perigos ou mudanças de velocidade, é insuficiente. A atual condição histórica não sugere sequer a metáfora do trem da história: estamos mais próximos de sermos passageiros de um carro elétrico guiado pelas decisões de uma IA em função de um algoritmo de GPS. Não existem trilhos desenhando nossa rota, não há estágios ou estações, nem sequer nossos destinos podem ser determinados por agulhas ou desvios de uma bifurcação ferroviária.

O futuro estaria fechado, ou a estrada está aberta e o motorista é um agente mais do que humano? Descrever o futuro nesses termos não é assumir a estratégia teleológica, mas, a partir de outra forma de relacionamento com tempo, debater a contingência do porvir como uma tática de *controle de riscos*.

Este artigo está dividido em três partes, além da introdução e da conclusão. Na primeira parte, apresento a noção de "mais do que humano", de suas origens animistas aos atuais debates sobre agência compartilhada. Depois, discuto sua relação com o problema da condição histórica contemporânea e a necessidade de uma teoria da história que confira *significado* e especule sobre a *direcionalidade* de suas experiências. Isso nos leva, na segunda parte, a descrever a história da IA a partir de debates da filosofia da mente, cruzados com dois paradigmas concorrentes na computação. Nesse momento, vou argumentar que o rompimento de um primeiro limiar semântico das máquinas implica na intensificação de debates no âmbito da ética informacional, sobretudo a respeito da agência social e moral dos algoritmos de aprendizado. Na terceira parte, entro mais especificamente na discussão sobre modalidades de temporalização do tempo relacionadas à IA Generativa. Defendo que a atualização repetidora, ou a reprodutibilidade técnica de estruturas sociais, são funções algorítmicas que podem ser representadas em código e visualizadas em *outputs* da IA Generativa. Por fim, na conclusão, encerro o texto apontando para alternativas à atualização repetidora que valorizam formas não humanas de razão, lançando a hipótese da computação contingente como forma de anunciar uma crítica da semântica temporal algorítmica.

Situando a discussão

A expressão "mundo mais do que humano" (*more-than-human world*) não é de todo nova. Ela aparece ainda nos anos 1990, no livro *O feitiço do sensível* (1997), de David Abram, fundador e diretor da *Alliance for Wild Ethics* (AWE). A obra aborda a relação entre a percepção humana e o mundo natural a partir de culturas indígenas. Abram se inspira em Husserl, Merleau-Ponty e Heidegger para entender o corpo como mediador entre a experiência humana e o mundo sensível. Com uma interpretação animista da fenomenologia, seu argumento concebe o mundo natural como uma entidade viva, cheia de intencionalidade e significado. O "mundo mais do que humano" descrito pelo autor não reduz a natureza a um objeto passivo e manipulável, mas convida-nos a compreender a fenomenologia em um modo participativo, no qual a noção de percepção assume predominância uma vez temporalizada como prática de "sintonia ou sincronização entre meus próprios ritmos e os ritmos das coisas em si" (ABRAM, 1997, p. 42).

O conceito ganhou tração nos últimos anos em discussões que envolvem sobretudo os mundos tecnológicos e ambientais. Para O'Gorman e Gaynor, a ideia representa uma abordagem relacional que ultrapassa o construtivismo e enfatiza a co-constituição de múltiplas espécies e vozes, ao lado da ética e dos saberes situados. Ao desafiar a dicotomia entre natureza e cultura (propondo um descentramento do humano), estudiosos das humanidades ambientais contribuíram para essa abordagem. Histórias mais do que humanas destacariam a complexidade das relações entre humanos e não humanos, revelando relações de poder e hierarquias que resistem à homogeneização de espécies. Metodologicamente, pesquisas desse tipo requerem um grau elevado de imersão na vida dos não humanos (os exemplos das autoras giram em torno de vírus e outros organismos), além de atenção à diversidade e à diferença. No geral, na esteira de David Abram – mas também inspiradas pelo pós-humanismo e pela geografia cultural –, as histórias mais do que humanas nos pedem para abandonar divisões especistas e refletir sobre as consequências éticas e

políticas das narrativas históricas. O conceito não deve ser entendido como "sinônimo para 'natureza' ou 'não-humano', mas, pelo contrário, como um termo que sublinha a *primazia da relação sobre as entidades* (incluindo o 'humano')" (O'GORMAN; GAYNOR, 2020, p. 717, grifos nossos).

Relações mais do que humanas têm sido destacadas na sociologia da tecnologia e consideradas um problema para o design de objetos técnicos. Deborah Lupton faz uma revisão crítica da literatura sobre as dimensões sociais da Internet das Coisas (IoT). Ela argumenta contra o determinismo tecnológico presente nessas discussões, defendendo a importância de se envolver com os imaginários sociais que não apenas dão significado às tecnologias IoT, mas também antecipam seus desenvolvimentos (LUPTON, 2020). Por sua vez, Giaccadi e Redström trazem a discussão para o campo do design, assumindo que as tecnologias digitais (como IA e aprendizado de máquina, *big data* e *IoT*) exigem repensar a coexistência entre humanos e objetos computacionais em rede. Isso acontece pois não estamos mais em uma situação na qual a tecnologia resta passiva à espera de um comando, como em eterno *stand by*, mas permanece conectada em uma rede de dispositivos que se comunicam entre si e com a internet, gerando e trocando dados em tempo real. A autoatualização dos sistemas computacionais sugere um mundo de design "mais do que humano", reconhecendo que as "experiências são resultado de uma interação dinâmica entre pessoas e dispositivos em rede, bem como entre dispositivos e outros dispositivos". (GIACCADI; REDSTRÖM, 2020, p. 44).

O conceito também aparece na obra *Machine Habitus: Toward a Sociology of Algorithms* (2022), de Massimo Airoldi. Trata-se de contribuição que explora uma teoria social dos algoritmos de aprendizado de máquina. O autor propõe o conceito de "habitus das máquinas", inspirado em Pierre Bourdieu, mas expandido para a compreensão das disposições éticas incorporadas que orientam a ação de agentes artificiais. Essa elaboração permite discutir as implicações sociais e políticas da IA incluindo a distribuição de poder e responsabilidade entre humanos e máquinas, os riscos de discriminação e manipulação algorítmica, além de possibilidades de emancipação. Essas são questões já amplamente documentadas no campo da IA e discutidas, em especial, a partir da noção de enviesamento (*bias*). Ao invés de assumir a possibilidade de uma IA livre de enviesamentos – ideia tanto estranha à sociologia –, o argumento central de Airoldi realiza dois movimentos. Primeiro, compreende o código como uma construção cultural derivada do treinamento da IA, responsável por traduzir valores humanos em representações matemáticas e implantá-los nos sistemas. Segundo, defende que essa constituição cultural do código guia o comportamento prático do código na cultura, correspondendo a uma tipologia de interações. Segundo Airoldi, essa tipologia – que ele chama de "mais do que humana" – depende da "combinação contingente de duas dimensões principais", a saber, "alta/baixa assimetria informacional e forte/fraco alinhamento cultural" entre usuários e sistemas, ou seja, entre o que o sociólogo chama de habitus humano e de habitus das máquinas (AIROLDI, 2022, p. 91).

A teoria da história nos permite entender esse debate em uma escala geral. Desde meados do pós-guerra, o ritmo acelerado do desenvolvimento tecnológico ocorre em paralelo ao aumento da preocupação com as mudanças climáticas. Novas biotecnologias, trans-humanismo, Inteligência Artificial, crescente desigualdade social e impacto antrópico no sistema terrestre: essas promessas e previsões também não anunciam mudanças que vão além da condição humana? Caso sejamos simpáticos à questão, onde estariam a teoria e a filosofia da história para fazer sentido de uma possível nova condição mais do que humana? Essa é a provocação inicial de Marek Tamm e Zoltan Simon, lançada de modo a incitar um debate (TAMM; SIMON, 2020). Trata-se de um argumento em favor da elaboração de uma teoria configurada à compreensão, não apenas da historiografia,

mas de experiências históricas concretas que porventura não dizem respeito unicamente ao presente, mas a "futuros presentes" e "presentes futuros" (LUHMANN, 1998, p. 70) passíveis de captura na forma de "relações de transição entre apreensões do passado e antecipações de futuro" (SIMON; TAMM, 2021, p. 13).

Como vemos, a discussão sobre mundos mais do que humanos é profundamente marcada pelo desenvolvimento tecnológico e pelo dilema do antropoceno. Dessa forma, para Simon e Tamm, investigar o tema no quadro de uma possível emergente condição histórica envolve engajamento com 1) uma perspectiva multiespecista que inclua agências não humanas, expandindo a compreensão tradicional das fontes históricas e sugerindo novas formas de conhecimento do passado; 2) uma noção de história multiescalar que leve em conta temporalidades plurais e a verticalidade das escalas geológicas e de tempo profundo e 3) a compreensão de temporalidades descontínuas e eventais, não redutíveis a configurações de tempo processuais. Esses seriam os gestos que nos levariam em "direção à possibilidade de uma nova filosofia da história" (TAMM; SIMON, 2020, p. 214).

No que segue, mantenho em mente esses três pontos, uma vez que não apenas elaboram o problema de uma história mais do que humana em termos teóricos, mas igualmente sintetizam um debate. A questão dos limites da agência da IA e de seus impactos ambientais assinalam os fundamentos físicos de nossa discussão, enquanto a possibilidade de um "evento epocal", tendo a "sexta extinção, uma potencial singularidade tecnológica e a transgressão das barreiras planetárias" como "representantes maiores" (SIMON, 2020, p. 64-65), opera como *telos* de antecipações sociais. Embora assuma parte do vocabulário dos "futuros históricos", recuo de um engajamento direto com as profecias singularitárias. Se já as mencionei como parte integral da mitologia da IA, interesso-me, a partir de agora, em compreender o mais do que humano tecnológico de um ponto de vista relacional que nos permita descrever futuros que já se fazem presentes.

O *Sattelzeit* das máquinas: abrindo as portas do Quarto Chinês?

"Podem as máquinas pensar?" (TURING, 1950). Como versão mais do que humana do Mito da Caverna, essa se torna a questão fundadora do campo da computação e o mito de origem da Inteligência Artificial, adiantado pelo vocabulário antropomórfico de Alan Turing. O campo da IA é recheado de conceitos fisicalistas. Essa epistemologia refuta o dualismo a partir de uma concepção de "pensamento" como epifenômeno do comportamento. Funcionalismo, behaviorismo, teoria da identidade e teorias computacionais da mente: suas metas são descritas por funções objetivas, atravessadas pela metáfora do processamento de informação por redes neurais artificiais. O treinamento das máquinas é realizado por meio de "recompensas" e "penalidades", enquanto nós, usuários cujas relações sociais são cada vez mais mediadas por algoritmos, somos incessantemente bombardeados em redes sociais por "estímulos" de "condicionamento operante" (ZUBOFF, 2019, p. 402; SILVEIRA, 2023, p. 14). Ao considerar a metáfora do "Jogo da Imitação" como uma afirmação literal, o fisicalismo das teorias computacionais do pensamento parece entrar em contradição, recaindo em um dualismo no qual a mente funciona como o comportamento de um *software*, e no qual o cérebro, como componente material, é equiparado ao *hardware* (FAZI, 2019, p. 819).

Em 1980, incomodado com as alegações mais ambiciosas do campo da computação (que ele nomeava de *Strong AI*), John Searle propôs seu famoso experimento de pensamento do "Quarto Chinês". Nele, somos convidados a imaginar um sujeito, preso em uma sala, cercado por caixas com ideogramas chineses. À sua frente, existe uma pequena entrada (*input*), pela qual recebe mais ideogramas; a suas costas, uma pequena saída (*output*),

por onde ele deve enviar ideogramas. Esse sujeito não sabe nada de chinês, mas ele tem uma tarefa: passar adiante ideogramas compondo respostas a outra sequência de ideogramas recebidos (ele não sabe, mas esses caracteres recebidos como *input* formam perguntas). Para cumprir sua função, o sujeito é munido de um livro de regras, composto por instruções escritas em língua que lhe é nativa. Com o tempo, explica Searle, ele pode aprender a reconhecer a forma dos ideogramas, identificá-los no livro de regras (metáfora da programação) e encaminhar os símbolos corretos em ordem pré-determinada. As frases formadas pelos ideogramas enviados pelo sujeito preso no quarto chinês seriam indistinguíveis das respostas de um falante nativo de chinês. Ele pode aprender a executar sua tarefa com rapidez e precisão, mas, por mais eficiente que se torne, esse sujeito nunca vai aprender a falar chinês (SEARLE, 1980, p. 418). O objetivo desse experimento de pensamento era demonstrar que a simples manipulação de símbolos e o uso de regras não é suficiente para inferir intencionalidade e compreensão. Isso refutaria as alegações mais audaciosas da IA, as quais defendiam que a simulação de comportamentos mentais, como o processamento de informação, seria suficiente para produzir compreensão e intencionalidade. Em outras palavras, computadores entendem sintaxe e podem identificar e manipular formas, mas não compreendem semântica. Consequentemente, não pensam – ao menos não como seres humanos.

Pouco mais de uma década depois, David Chalmers (1992) publica o texto *Computação Subsimbólica e a Sala Chinesa*. O experimento mental de Searle já havia se tornado um clássico, gerado uma infinidade de respostas e conquistado terreno como argumento incontornável de crítica ao fisicalismo e seus avatares disciplinares. O objetivo do texto, publicado no início dos anos 1990, era testar a resiliência da argumentação de Searle frente aos dois grandes paradigmas que naquela época competiam dentro das pesquisas em IA: o modelo simbólico e o modelo conexionista (ou subsimbólico). A computação simbólica segue a hipótese de que um "sistema de símbolos físicos tem os meios necessários e suficientes para uma ação inteligente geral" (CHALMERS, 1992, p. 7). Alvo original de Searle, essa é a classe de programas mais vulneráveis ao argumento da sala chinesa. Em oposição, a hipótese subsimbólica, que, nos anos 1990, era associada ao conexionismo, nos fala de "um sistema dinâmico de conexão subconceitual que não admite uma descrição completa, formal e precisa do nível conceitual" (SMOLENSKY, 1988 *apud* CHALMERS, 1992). Ou seja, para descrever supostos processos mentais, seria necessário referir-se a dimensões subconceituais, que não carregam, sozinhas, nenhum significado: "a carga semântica do sistema está em um nível superior, o da *representação distribuída*" (CHALMERS, 1992, p. 8).

Nos sistemas simbólicos, representações e *tokens* coincidem como entidades simbólicas. Enquanto os *tokens* computacionais são as unidades sintáticas fundamentais, as representações são unidades semânticas fundamentais. A união entre significantes e significados forma os átomos indivisíveis da computação simbólica. O caso é que, nos sistemas subsimbólicos, essas unidades estão dispersas: o nível da computação está abaixo do nível da representação. Neles, os *tokens* computacionais são objetos sintáticos, enquanto as representações são padrões distribuídos de atividade que emergem a partir do processamento computacional em um nível inferior. O significado, portanto, torna-se uma propriedade emergente das representações. Dessa forma, a distinção entre os dois paradigmas reside no fato de que os objetos da computação simbólica coincidem com os objetos de interpretação semântica, enquanto nos sistemas subsimbólicos os objetos computacionais são distribuídos em múltiplas partes. A noção de "representações distribuídas" torna-se um conceito chave no vocabulário técnico na medida em que o aprendizado profundo substitui o conexionismo em referência à arquitetura dos sistemas (neurais). Como explica Kelleher, podemos:

> [...] fazer uma distinção nas representações usadas por redes neurais entre representações localistas e distribuídas. Em uma representação localista, há uma correspondência um-para-um entre conceitos e neurônios, enquanto em uma representação distribuída, cada conceito é representado por um padrão de ativações em um conjunto de neurônios. Consequentemente, em uma representação distribuída, cada conceito é representado pela ativação de vários neurônios e a ativação de cada neurônio contribui para a representação de vários conceitos (KELLEHER, 2019, p. 129).

Ou seja, nas representações distribuídas, o nível de computação está em camada mais básica, enquanto o nível de representação está em camada mais avançada. "O mais importante", explicava Chalmers, "é que as representações conexionistas possuem uma estrutura interna rica" que as leva a constituir "sua própria organização intrínseca por serem compostas por um padrão complexo de atividade" (CHALMERS, 1992, p. 15). É dessa forma que o argumento de Searle, baseado na oposição entre sintática e semântica, não se aplicaria às representações processadas por conexões subsimbólicas. Décadas antes, Zellig Harris (1954, p. 155) havia proposto um modelo de tradução fundado na análise do significado de lexemas como uma "função da distribuição" de fonemas e morfemas em enunciados. O problema? Exigia-se uma larga série documental e uma grande capacidade de processamento para tornar seu modelo funcional. É essa lacuna que representações distribuídas alimentadas por *big data* vieram preencher. Elas carregam conteúdo a partir de seu caráter estrutural interno: com esse recurso, adquirem propriedade "microssemântica", ou "um padrão interno que reflete sistematicamente o significado da representação" (CHALMERS, 1992, p. 15). Essa estrutura permite que os sistemas conexionistas processem informações de uma maneira que os sistemas simbólicos não podem, criando *embeddings*, ou vetores numéricos de alta dimensão [0.2, 0.5, 0.8, 0.3, ver Figura 1]. No contexto de criação de *embeddings*, tuplas são usadas para representar os vetores numéricos resultantes da codificação de conceitos ou entidades de um determinado domínio de significado. Essa abordagem é adotada pelos avanços recentes no campo da IA, que buscam "*superar a barreira semântica* e extrair processamento de informações a partir do hardware e da sintaxe" (FLORIDI, 2014, p. 142, grifos nossos). Em suma, representações distribuídas são um conceito-chave que promete abrir as portas do Quarto Chinês.

```
# Os embeddings são representações distribuídas de palavras em um espaço vetorial.
# Cada dimensão do vetor contribui para a representação global da palavra,
# capturando diferentes aspectos semânticos ou sintáticos.

embedding1 = [0.5, 0.2, -0.1, 0.8]  # Primeiro embedding
embedding2 = [-0.3, 0.9, 0.2, -0.6]  # Segundo embedding
embedding3 = [0.1, -0.7, 0.5, 0.3]   # Terceiro embedding

# Conjunto de embeddings
# Converter as listas em tuplas garante a imutabilidade, importante para armazenar em um conjunto
embedding_set = {tuple(embedding1), tuple(embedding2), tuple(embedding3)}

# Impressão do conjunto
# Cada tupla no conjunto representa um embedding distinto
print(embedding_set)
```

Figura 1 – Três *embeddings* representados como conjuntos numéricos. A conversão das listas em tuplas é adicionada a um conjunto chamado "*embedding_set*". Linguagem Python, ambiente de programação do Google Colab.

O conceito de "representações distribuídas" assume ainda maior importância no contexto do aprendizado profundo e no uso das arquiteturas de redes neurais artificiais (ANN), em particular para o processamento de linguagem natural (NLP) por meio de técnicas de vetorização de palavras (LE; MIKOLOV, 2014). Para o conexionismo, "neurônios no cérebro representariam uma rede de unidades discretas e homogêneas, e as sinapses representariam as conexões de transferência de energia entre elas" (AMARO, 2022, p. 136). O "profundo" desse tipo de aprendizado faz referência às camadas escondidas de processamento paralelo que, não diretamente observáveis pelos usuários ou pelos programadores que trabalham com a rede, representam e analisam pontos de dados como vetores com múltiplas dimensões. As representações distribuídas codificam essas informações complexas em padrões de ativação distribuídos em redes neurais (KELLEHER, 2019, p. 132). Esses padrões são aprendidos a partir de exemplos de dados e usados para previsões em novos dados. Representações distribuídas permitem generalização e reconhecimento de novas informações. Com *embeddings* aprendidos automaticamente a partir dos dados, modelos de aprendizado capturam relações semânticas e sintáticas entre entidades (CHALMERS, 1992, p. 18).

A capacidade das redes neurais em representar o conhecimento por meio de padrões distribuídos de ativação tem sido um fator-chave para o desenvolvimento da autonomia da IA nos últimos anos. Com modelos que podem aprender com vastas quantidades de dados e ajustar seu comportamento sem programação ou intervenção explícita de humanos, alcançamos maior grau de independência e, com isso, debates sobre ética informacional atingem um novo patamar. A questão sugerida, para usar o vocabulário de Searle, remeteria ao grau de *intencionalidade* produzido pela "microssemântica" das máquinas. Se, por um lado, modelos sociológicos consagrados nos permitem identificar sistemas de IA como *agentes sociais*, por outro lado, questões que os escapam, tais como autonomia, intencionalidade e responsabilidade, são centrais para o estudo da *agência moral* dos algoritmos.[3] Essa não é uma distinção banal. Imaginemos duas situações, a primeira fictícia e a segunda (lamentavelmente) verdadeira: 1) um carro elétrico dirigido por IA atropela e mata uma pessoa durante testes na Califórnia; 2) um sistema de reconhecimento facial identifica uma pessoa não branca como criminosa e leva a polícia a prendê-la na Inglaterra. Nessas situações temos dois crimes (um homicídio doloso e um caso de racismo algorítmico). A questão que se impõe tem uma clara e distinta natureza jurídica: de quem é a responsabilidade? Culpar algoritmos e isentar humanos não parece, pelo menos à primeira vista, nem sensato nem justo. Por outro lado, a autonomia dos sistemas de IA traz complicações para o estudo da agência moral artificial (AMA). Vejamos brevemente três posições, representantes, respectivamente, de uma "visão padrão", de uma visão "funcionalista" e de uma visão "normativa" do problema da AMA.

Inspiradas pela Teoria do Ator Rede, Johnson e Verdicchio (2019) propuseram um modelo triádico de agência que agrega fatores causais e fatores intencionais como componentes em rede. No argumento das autoras, ambas as formas de agência compartilham lugar como causas eficientes, embora a agência intencional (ligada a estados mentais humanos) seja associada ao início da cadeia de causa e efeito. Nos exemplos apresentados, gestores empresariais (*top management*) têm a *intenção* de produzir um veículo que atinja padrões de qualidade ambiental sem aumento no custo do produto, e para isso se tornam (1) usuários de um sistema de IA e delegam, por assim dizer, a tarefa a designers (2), que devem elaborar um código na forma de um artefato técnico (3) que, enfim, forneça "eficácia causal necessária para alcançar o objetivo" (JOHNSON; VERDICCHIO, 2019, p. 642). Como sugere a perspectiva "padrão" (*standard*) das autoras, mesmo em um cenário futurista, no qual os programadores humanos são substituídos por algoritmos, "agência e responsabilidade

devem ser separadas no sentido de que a agência é triádica, enquanto a responsabilidade é sempre atribuída a seres humanos" (JOHNSON; VERDICCHIO, 2019, p. 644).

Estudos sobre AMA geralmente consideram os humanos como os únicos sujeitos dotados de consciência. Isso não significa que concordem sobre o papel da consciência fenomênica para a agência moral. A visão padrão, como se vê no exemplo acima, costuma tratá-la como uma condição necessária. Mas os funcionalistas tendem a se opor a essa premissa: seus argumentos não raramente destacam as dificuldades de identificação direta dos estados mentais. A ideia de que a agência moral requer subjetividade deveria, portanto, ser abandonada em favor de um conceito que trate a agência moral como uma função objetiva. Luciano Floridi e seus colaboradores são expoentes dessa visão. Essa perspectiva permite dizer, por exemplo, que "uma sociedade hiper-histórica, totalmente dependente de tecnologias de terceira ordem pode, em princípio, ser independente da humanidade" (FLORIDI, 2014, p. 32). Isso não significa defender que algoritmos não sejam sistemas carregados de valores humanos. Essa é, por sinal, a premissa de um artigo dedicado a mapear os dilemas éticos levantados por algoritmos de aprendizado. Mas esses sistemas, como insiste grande parte da literatura, são menos previsíveis e interpretáveis.

Nesse cenário, uma "concepção tradicional e linear de responsabilidade" (MITTELSTADT *et al.*, 2016, p. 10) é menos útil, e a diversidade dos problemas éticos explode: sistemas de aprendizado de máquina podem produzir conhecimento provável, mas incerto. A conexão entre os dados e suas conclusões pode não ser óbvia, ou mesmo acessível, quando não equivocada (o termo técnico aqui é "alucinação"). Isso pode levar a ações artificiais discriminatórias. Ainda assim, algoritmos, como mediadores de relações sociais, podem modificar a forma como concebemos e organizamos o mundo, trazendo problemas de autonomia e privacidade. O dilema pode ser resumido assim: "atribuir agência moral a agentes artificiais pode permitir que as partes interessadas humanas transfiram a culpa para os algoritmos"; por outro lado, "negar a agência a agentes artificiais torna os designers responsáveis pelo comportamento antiético de suas criações semiautônomas" (MITTELSTADT *et al.*, 2016, p. 11). A conclusão dos autores é que esses extremos não capturam a complexidade da fiscalização nem o dinamismo dos sistemas de decisão artificiais.

Por fim, em artigo que busca sintetizar elementos das visões padrão e funcionalista a respeito da AMA, Behdadi e Munthe (2020) defendem uma abordagem mais metodológica. Incorporando elementos da solução funcionalista, argumentam que "um requisito de consciência fenomenal para a agência moral como essencial para o debate do AMA" parece um exagero (BEHDADI; MUNTHE, 2020, p. 17). Afinal, muitas das características normalmente atribuídas a seres conscientes – lembramos aqui da microssemântica – podem ser atingidas por agentes artificiais. Em vez disso, o artigo propõe concentrar o debate na inclusão dessas entidades em práticas humanas com agência moral e responsabilidade dos participantes, incluindo elementos da visão padronizada, como a noção de agência compartilhada. A questão que se instaura acaba traduzida por um "problema de demarcação". Critérios normativos podem desconsiderar entes artificiais de ações e interações que pressupõe AMA, além de excluir agentes humanos; da mesma forma, esse tipo de critério também pode deixar de considerar humanos que deveriam ser responsabilizados, incluindo "entidades artificiais onde essa razão ética está em falta" (BEHDADI; MUNTHE, 2020, p. 20).

Como se vê, os debates atuais da ética informacional são mediados por questões a respeito de agências compartilhadas e da (in)dependência fenomenológica para a atribuição de culpabilidade, o que leva à necessidade de demarcar os limites entre seres naturais e entes artificiais. Em minha leitura, esse debate acaba – em um processo acelerado pelos avanços no aprendizado e pelo rompimento do limiar semântico que chamei de *Sattelzeit* das máquinas – por entrar em uma fase mais do que humana.

A transição entre agência social e agência moral artificial sugere uma temporalização dos debates na ética da informação. Em uma ponta, temos a centralidade do problema do treinamento, especialmente destacado por Floridi e seus colaboradores: o "treinamento produz uma estrutura (classes, clusters, classificações, pesos, etc) para classificar novas entradas ou prever variáveis desconhecidas". Valores humanos são integrados nos sistemas como representações matemáticas. Por outro lado, "depois de treinado, novos dados podem ser processados e categorizados automaticamente, sem intervenção do operador", o que torna a "lógica do algoritmo" uma "caixa-preta" (MITTELSTADT *et al.*, 2016, p. 6).

Que algoritmos são agentes sociais não restam muitas dúvidas. Sua ubiquidade em nossas sociedades é notável, atuando como ferramentas e parceiros em diversas formas de comunicação, "não somente na web, onde o papel ativo dos bots agora é dado como certo, mas também (explicitamente ou não) em formas mais tradicionais" (ESPOSITO, 2017, p. 245-249). O exame das plataformas digitais também sugere, em diálogo com a Teoria dos Sistemas, que algoritmos permitem a automatização de "trabalhos de sincronização" de experiências humanas (Cf. JORDHEIM; YTREBERG, 2021, p. 412). Após o rompimento de um primeiro limiar semântico artificial, chegamos a pelo menos dois argumentos: ou a compreensão semântica das máquinas é uma simulação inautêntica derivada da sintaxe (ou uma *explicação* microssemântica, avessa à *compreensão* narrativa), o que nos leva a desconsiderar as conquistas em tecnologia de IA como não sendo manifestações reais de inteligência, reproduzindo o tropo do "Efeito IA" (MCCORDUCK, 2004, p. 204); ou, seguindo uma percepção positiva do problema que aceite a *funcionalidade* efeitual e relacional do processamento microssemântico, concluiríamos que o rompimento desse limiar não é suficiente para inferir intencionalidade às máquinas e, logo, que o narrativismo e a abordagem hermenêutica da fenomenologia são, nesse contexto, insuficientes:

> Podemos testar alternativas contra-hegemônicas, hackear os programas econômicos e tecnológicos que nos prendem. Mas a sua eficácia depende de que relocalizemos as práticas dissidentes numa discussão mais vasta, mais intricada, sobre o sentido [...], *mas é preciso dizer que o sentido histórico, da vida e do convívio, dos conflitos e dos "arranjos", no sentido em que um poeta como Eliot pensava na primeira metade do século XX ou Ricœur na segunda, correspondem apenas parcialmente às experiências atuais* (CANCLINI, 2021, p. 156, grifos nossos).

Caso a consciência não represente apenas uma "função de transição abrupta (*step function*) na complexidade cerebral" (BUTTAZZO, 2001, p. 28), como parecem sugerir as teorias computacionais da mente, então os avanços em IA provam que a humanidade é algo mais do que uma máquina de produzir significados. A articulação discursiva entre significado e significante não garante intencionalidade como qualidade fenomenológica da consciência. Consequentemente, seria ainda "nossa condição histórica atual", entendida pela diferença iterativa e contingente na condição humana, suficientemente definível através da "articulação entre o discurso sobre a historicidade e o discurso da história"? (PEREIRA, 2022, p. 15).

Devolveremos essa pergunta ao Outro, pois nossa relação com a IA é antropomórfica, mais do que humana, e não apenas porque abordagens inspiradas no perspectivismo de Viveiros de Castro ou no pampsiquismo em voga entre os realistas especulativos podem nos convencer disso, mas porque valores morais são imputados nas máquinas por meio de treinamento (BONALDO; PEREIRA, 2023). Resta agora indagar como a produção da diferença pode ser constituída artificialmente a partir de uma série temporal de dados discretizados representando experiências humanas manipuláveis estatisticamente. Em

outras palavras, resta perguntar como podem agentes artificiais orientar a nossa forma de se relacionar com o tempo histórico, de nos fazer habitá-lo, e assim, de nos temporalizar. Com essa questão, busco redirecionar o problema do "sentido histórico" enquanto atribuição de significado à experiência para sua acepção de sentido enquanto *direcionalidade* do curso dos acontecimentos – rumo à descrição de futuros já presentes em uma história mais do que humana.

Computação necessária como atualização repetidora: rumo à AI Generativa

Lemmy Caution é um agente secreto. Ele viaja até *Alphaville* com o objetivo de incapacitar *Alpha 60*, um supercomputador que assumiu o controle da cidade. Sediado no "Instituto de Semântica Geral", a função de *Alpha 60* é destruir o significado, impossibilitando a poética e dissolvendo a capacidade dos habitantes da cidade de se comunicarem para além de roteiros pré-programados. O supercomputador foi criado pelo meticulosamente nomeado Von Braun, um cientista que, controlando *Alpha 60*, domina de fato a sociedade local. Quando Lemmy Caution consegue assassinar Von Braun, a sociedade entra em colapso, seus habitantes batendo cabeças contra as paredes, incapazes de atuar como seres funcionais.

O filme de Godard (1965) serve para ilustrar o que estou chamando de "computação necessária". O supercomputador *Alpha 60* não é uma "inteligência" artificial. Ele se limita à tarefa de "não ser mais do que o meio lógico dessa destruição" (do significado). Instrumentalização da abstração, sim, mas condicionada aos comandos oferecidos por um super vilão: por trás de uma máquina burra, há sempre um gênio do mal. Esse é o tropo que fundamenta *Alphaville* (1965). Mesmo que a racionalidade instrumental com relação a fins não seja necessariamente alheia à inovação, como vimos na leitura que Adorno e Horkheimer fazem do episódio do Ulisses burguês entre as sereias em *A dialética do esclarecimento*, um corpo humano sensível, uma experiência heroica e vívida, uma classe social que encarna o Espírito, precisa se fazer presente para arquitetar a astúcia por trás de mil truques.

A computação necessária, ou a ideia de que algoritmos estariam limitados a performar uma "atualização repetidora", para retomar o vocabulário de Pereira e Araujo (2020, p. 126; 2022, p. 75-76), poderia ser desafiada pelo processamento de informação por representações distribuídas e pela autonomia de decisão consequentemente assumida por sistemas de aprendizado profundo capazes de crescentes níveis de generalização. E, ainda assim, a tese nos chama a atenção para a humanidade que se esconde no corpo da IA; que repousa, por assim dizer, por trás dos processos de treinamento, curadoria de dados e inserção de valores éticos em algoritmos capazes de domesticar e homogeneizar a contingência, reintroduzindo-a sistematicamente no mundo como necessidade. A "cultura no código" (AIROLDI, 2022), em outras palavras, tem o potencial de reproduzir preconceitos de classe, hierarquias de saber e poder patriarcais, ainda protocolos gerais de racialização e reprodução da diferença "[...] pré-condicionados por um processo de relações humanas que de partida concebeu os seres Negros como objetos entre outros objetos" (AMARO, 2022, p. 13). Essa é uma discussão recente e já muito bem documentada por estudos dentro e fora do Brasil (BENJAMIN, 2019; NOBLE, 2021; SILVA, 2022; FAUSTINO; LIPPOLD, 2023). Inspirado nela, parto do pressuposto de que sistemas de aprendizado de máquina "contribuem de modo prático para a reprodução da sociedade, com seus discursos arbitrários, fronteiras invisíveis e estruturas", embora indivíduos possam, frente ao "código na cultura", desenvolver "uma relação ativa em seus encontros com plataformas baseadas em sistemas algorítmicos" (AIROLDI, 2022, p. 112 e 79).

Quando falamos em níveis de autonomia, generalização e AMA, lembramos que as máquinas de hoje são capazes de aprender. Não se trata de uma tecnologia nova. "Em vez de tentar produzir um programa para simular a mente de um adulto, por que não tentar produzir um que simule a de uma criança?" (TURING, 1950, p. 18). Proposto ao final do ensaio clássico de Alan Turing, o aprendizado de máquina desenvolveu-se no encontro de domínios originalmente distintos, como a estatística, as ciências da computação e o reconhecimento de padrões. Desde o pós-guerra, sua tarefa consistiu em controlar a contingência: "ganhar perspectiva sobre o comportamento humano dinâmico, reduzindo os níveis de incerteza que definem toda a vida humana e não-humana" (AMARO, 2022, p. 100). Por exemplo, um algoritmo pode buscar padrões de frequência no *Twitter* e, a partir da identificação de tópicos, categorizar textos em diferentes grupos (ROTA; NICODEMO, 2023). A matemática chama esse tipo de relação de funções: "um mapeamento determinístico de um conjunto de valores de entrada para um ou mais valores de saída" (KELLEHER, 2019, p. 7). A redução dessas operações a técnicas aparentemente inofensivas, saudadas como inovadoras, pode esconder a evidência de que o aprendizado de máquina elabora conceitualmente o mundo como uma extensão da atividade estatística. Entre seus efeitos mais nefastos, está "a condensação de múltiplos potenciais em uma única saída", levando ao "fechamento preventivo de reivindicações políticas baseadas em atributos de dados que buscam reconhecimento antecipado" (AMOORE, 2020, p. 20-21).

O método de regressão, muito utilizado no aprendizado supervisionado, é o melhor exemplo que pude encontrar na literatura de uma "razão algorítmica que é obscurecida por trás de provas naturalizadas" (AMARO, 2022, p. 108). Seu objetivo é prever o valor contínuo de uma variável alvo, utilizando um conjunto de variáveis de entrada como referência. Assim constroem-se classificadores, não de "maneira fantasiosa", mas "baseados em experiências passadas" (BREIMAN *et al.*, 2017, p. 13). O exemplo de amostra de aprendizado a seguir, retirado do manual de Breiman, Friedman, Stone e Olshen, e comentado por Ramon Amaro, "consiste em dados sobre N casos observados no passado, juntamente com sua classificação atual" (AMARO, 2022, p. 108). A formalização dos autores é a seguinte: "Uma amostra de aprendizado consiste em dados (x1, j1), ..., (xN, jN) sobre N casos onde xn ϵ X e jn ϵ {1, ..., J}, n = 1, ..., N. A amostra de aprendizado é representada por L; ou seja, L = {(x1, j1), ..., (xN, jN)}." (BREIMAN *et al.*, 2017, p. 13).

Os casos representam dados históricos utilizados para treinar o modelo e prepará-lo para realizar previsões a partir de generalizações em exemplos que não foram vistos antes. Para Ramon Amaro, essa "razão algorítmica", representada visualmente na Figura 2, "é extraordinária em sua dependência de dados históricos como um continuum de padrões aprendidos no futuro, os quais são universalizados em uma relação genérica entre eventos atuais e resultados predeterminados" (AMARO, 2022, p. 108). Leio, nessa descrição da razão algorítmica, um conjunto de possibilidades de mediações artificiais em nossas formas "de se relacionar com o tempo histórico como atualização repetidora" (PEREIRA; ARAUJO, 2020, p. 126). Enquanto um modo específico da historicidade atualista, entendo a "atualização repetidora" como uma função diferencial do futuro e do passado situada em um presente *formal* que relaciona eventos atuais com cursos predeterminados por dados históricos. Mais importante, esses dados históricos aos quais a máquina é exposta em aprendizado são apreendidos não como processo, mas como série de eventos, organizados em categorias discretas. A função objetiva dessa operação é atualizar, como *continuum*, dados por natureza computacional discretizados. Seu método de reconhecimento de padrões amplia o horizonte de visão "subsumindo a aleatoriedade", domesticando as "variações da contingência em um grupo generalizado de padrões" (AMARO, 2022, p. 123).

```python
import random

# Amostra de aprendizado com 10 casos,
# compostos por um valor aleatório de um conjunto fixo X e um número aleatório de 1 a 3 (classes)
# A amostra é armazenada como uma lista de pares.

N = 10  # Número de casos na amostra de aprendizado
X = {'x1', 'x2', 'x3', 'x4', 'x5', 'x6', 'x7', 'x8', 'x9', 'x10'}  # Conjunto de possíveis valores para X
J = 3  # Número de possíveis valores para J (classes)

amostra_de_aprendizado = []

for n in range(N):
    Xn = random.choice(list(X))  # Escolha aleatória de um valor de X
    Jn = random.randint(1, J)    # Escolha aleatória de um valor de J (entre 1 e J)
    amostra_de_aprendizado.append((Xn, Jn))  # Adiciona o par (Xn, Jn) à amostra

L = amostra_de_aprendizado  # Representação da amostra de aprendizado como uma lista de pares

# Exibindo a amostra de aprendizado
print(L)
```

Figura 2 – O código ilustra a formalização de Breiman sobre a amostra de aprendizado em regressão linear aplicada ao aprendizado de máquina, criando pares de dados e classes aleatórias. Linguagem Python, ambiente de programação do Google Colab.

Mais evidentes e melhor representáveis no caso do aprendizado supervisionado, os modos de temporalização do tempo instruídos pela "razão algorítmica" se tornam progressivamente opacos na medida em que dados históricos são filtrados pelas camadas ocultas de uma rede neural artificial. Redes neurais artificiais (ANN) são projetadas para aprender e extrair padrões dos dados de entrada, e podem priorizar certos aspectos dos dados em detrimento de outros, dependendo do tipo de tarefa e da arquitetura da rede. Ademais, as camadas ocultas podem fundir ou agrupar informações de várias entradas, o que pode levar a uma perda de detalhes sobre informações temporais (a "interdição" aqui não assumiria as características de uma política do tempo histórico?). Isso leva às arquiteturas de aprendizado profundo estruturarem os modelos de IA "menos interpretáveis" (KELLEHER, 2019, p. 245). Por isso, a melhor maneira que encontrei para demonstrar meu argumento é através de imagens geradas por IA. A imagem que apresento foi gerada a partir da descrição de uma pessoa escravizada fugida.[4] A informação está registrada no Jornal Correio Catharinense de 9 março de 1853 (nº 17, p. 4). Ela nos fala de "[...] hum escravo pardo, mestre de offício de barbeiro, de idade de 18 para 19 annos, cujos os signaes são os seguintes – estatura baixa, corpo delgado, sem barba, dentes da frente limados, boca um pouco grande". Utilizei esses dados como *input* e comecei a elaborar uma engenharia de *prompt*: um conjunto de comandos representáveis como valores matemáticos a partir do qual a IA consegue gerar uma imagem. O sistema utilizado, neste caso, foi o *Midjourney*, sediado na rede social *Discord*. Após algumas interações experimentais, nas quais aperfeiçoei o *prompt*, obtive a seguinte imagem (Figura 3):

Figura 3 – Imagem gerada no *Midjourney*, *prompt*: Photography of a smiling black XIX century young man in a barber shop, aged 18 to 19 years, short stature, slender build, no beard, front teeth filed, slightly large mouth, polaroid, uplight - v 5[5].

Apesar do sedutor realismo da imagem, há pelo menos três limitações a serem notadas. A primeira, e mais evidente, diz respeito ao problema da linguagem. Inexiste na língua inglesa o termo "pardo" – e sinônimos imprecisos como "*multiracial*", "*mixed-race*" ou "*biracial*" não geraram bons resultados. A segunda limitação diz respeito a um problema bem descrito pela história social. A noção de cor "herdada do período colonial, não designava, preferencialmente, matizes de pigmentação ou níveis diferentes de mestiçagem, mas buscava definir lugares sociais, nos quais etnia e condição estavam indissociavelmente ligadas" (MATTOS, 2013, p. 106). O caráter iterativo da noção de cor é de difícil captura pelos processos discricionários de classificação a que estamos sujeitos em nossas interações com a IA Generativa. A terceira limitação diz respeito ao formato do sorriso: o valor "smiling" foi incluído porque, em minhas experiências, notei que essa é a maneira mais objetiva de fazer a IA representar o jovem barbeiro de boca aberta, apresentando os dentes, como descritos na documentação. O modo como o personagem representado sorri, no entanto, pode estar culturalmente contaminado pelos dados de treinamento do *Midjourney*. Jenka Gurfinkel (2023) chamou esse fenômeno de uma "decepção esteganográfica dentro dos pixels". O caso é que, ao extrapolar os membros de um *embedding* [sorriso, negro, jovem, etc.] a partir do que é mais representativo no *dataset*, a "IA dominada por fontes de imagem influenciadas pelos Estados Unidos está produzindo uma nova monocultura visual de expressões faciais" (GURFINKEL, 2023). Buscamos evocar o "passado, e mesmo o futuro, mas o presente permanece pobremente transformado" (PEREIRA; ARAUJO, 2022, p. 75). Os algoritmos do *Midjourney*, ainda que operando de modo subsimbólico, estariam atualizando anacronicamente movimentos faciais de culturas estrangeiras, ao invés daquela representada nas fontes primárias?[6]

Anáfora ou *looping de feedback* dos padrões de sorrir. O jovem barbeiro sorri como uma multidão de outros jovens que perfilam em galerias digitais. O jovem barbeiro sorri como o fazem as pessoas nas fotografias que alimentaram os algoritmos de aprendizado de máquina, compostas de indivíduos contemporâneos de perfis étnico-raciais diversos. O contraste entre o sorriso do barbeiro e os resultados de uma rápida busca no banco de dados do *Google* ou do *Getty Images*, usando *tokens* específicos [*smiling*, *young*, *man*], permitiria ilustrar pedagogicamente o modo pelo qual a atualização repetidora opera

vetorizando palavras em espaços de alta-dimensionalidade. Ela também poderia evidenciar, quem sabe, um dispositivo de "*racialização digital*" (FAUSTINO; LIPPOLD, 2023, p. 151, grifos do autor). Por outro lado, embora possamos refletir sobre as saídas do *Midjourney* questionando *o que é mais representativo nos dados de treinamento do modelo em função dos valores de entrada* explicitados no *prompt*, também é possível que "a IA possa fazer algo inesperado e surpreendente" (COECKELBERGH, 2023, p. 6). Tal situação seria mais difícil de medir: "quem vê *input* e *output* vê o canal e não o processo codificador que se passa no interior da caixa-preta" (FLUSSER, 2009, p. 15).

Caixa-preta é uma expressão "emprestada da cibernética e usada para descrever um objeto ou sistema [...] cujo funcionamento interno [...] permanece oculto" (FAZI, 2020, p. 59). Segundo Burrell (2016), uma caixa-preta informacional possui três camadas de significado. A primeira, e mais evidente, demonstra sua opacidade como segredo estatal ou corporativo: algoritmos avançados são tecnologia privada. A segunda camada diz respeito ao problema da alfabetização técnica, a dificuldade, que eu e você, usuários e vítimas de tecnologias de terceira ordem (que observam nossas observações), temos em compreender o funcionamento desses sistemas. A terceira é geralmente apresentada como o lado mais misterioso da IA: nem sequer os desenvolvedores conseguem compreender certas decisões tomadas pelos algoritmos (FAZI, 2020, p. 59). Esse último nível de opacidade "decorre da falta de correspondência entre otimização matemática em alta dimensionalidade [...] e demandas de raciocínio em escala humana" (BURREL, 2016, p. 2). A relação entre os níveis mais abstratos de processamento e suas saídas implica, pois, em uma difícil discussão: abrir a "caixa-preta", para seguir um tropo recorrente em *Explainable AI* (XAI), pode nos levar a reconhecer que nem sequer possuímos os lexemas necessários para explicar os conceitos representados dentro dela (FAZI, 2020, p. 63). O que é complexo para nós é muito simples para a IA, e vice-versa. Seriam, então, algoritmos de *deep learning* representantes de um tipo de "pensamento" que é "dramaticamente alienígena ao pensamento humano?" (FAZI, 2019, p. 813).

É provável que o desenvolvimento desses sistemas intensifique os dilemas éticos descritos pelo grupo de Mittelstadt *et al.* (2016). O próprio documento publicado pela Open AI em torno da data na qual o GPT-3 foi lançado para grupos restritos, já previa que, entre as "Potenciais Aplicações Abusivas" do sistema, estavam "desinformação, *spam*, *phishing*, abuso de processos legais e governamentais, escrita fraudulenta de ensaios acadêmicos e pretexto de engenharia social" (OPEN AI, 2020, p. 35). Um relatório técnico sobre o GPT-4, lançado três anos depois, admite que o sistema "apresenta limitações semelhantes aos modelos anteriores", não sendo "totalmente confiável (por exemplo, pode sofrer de 'alucinações'), possui uma janela de contexto limitada e não aprende com a experiência" (OPEN AI, 2023, p. 1-2). O que os engenheiros chamam de "alucinação" responde pela geração de informações inventadas, sequer presentes nos dados aos quais o sistema foi exposto durante o treinamento. Por isso, alerta a empresa, "deve-se ter cuidado ao usar as saídas do GPT-4, especialmente em contextos em que a confiabilidade é importante" (OPEN AI, 2023, p. 2).

O GPT pertence a uma série de modelos de linguagem baseados na arquitetura *Transformer*. Esse tipo de arquitetura neural, proposta por engenheiros do *Google* em 2017, combina componentes simbólicos com processos de otimização subsimbólicos para gerar textos através da articulação de representações distribuídas, mecanismos de atenção (que permitem ao modelo se concentrar em partes específicas dos dados de entrada) e treinamento com exemplos (VASWANI *et al.*, 2017). O relatório técnico do GPT-4, lançado em março de 2023, possui como anexo um "*System Card*", espécie de resumo que descreve as especificações do sistema. Na parte dedicada a debater os desafios

de segurança, os desenvolvedores reconhecem que o GPT-4 é suscetível, para além da geração de informações sem sentido nem base documental, à produção de "desinformação e operações de influência e privacidade" (OPEN AI, 2023, p. 45). A geração de conteúdos ofensivos também é mencionada, como "conselhos sobre planejamento de ataques ou discursos de ódio", podendo apresentar "preconceitos e visões de mundo que podem não ser representativos da intenção dos usuários" (OPEN AI, 2023, p. 40).

O problema do "alinhamento cultural" é particularmente caro à indústria, e sua solução, como vemos pela leitura do relatório, se dá através do treinamento. Máximo Airoldi conceitualizou a questão do treinamento e do alinhamento cultural a partir da intersecção entre as culturas interiores e exteriores ao código – entre o habitus da máquina e o habitus dos usuários. Máquinas, assim como os seres humanos, passam por processos de socialização (i.e. exposição aos dados) e, para o sociólogo, é esse processo, e não a sua inteligência, que faz delas agentes sociais: "elas podem aprender a partir de vestígios do mundo social", possuindo, por isso, um "habitus das máquinas", o qual "pode ser definido como o conjunto de disposições culturais e propensões codificadas em um sistema de aprendizado de máquina por meio de processos de socialização orientados por dados" (AIROLDI, 2022, p. 112-113). A recomposição desses vestígios chegou a fazer Katherine Hayles considerar modelos de linguagem larga (LLMs), como o LaMDA e o GPT-3, como "provavelmente protoconscientes" (HAYLES, 2022, p. 164). Essa hipótese, que sugere perceber a IA como uma consciência-de-si em desenvolvimento, é bem mais ousada do que as observações da ética informacional e o debate de Airoldi sobre uma sociologia dos algoritmos. Ainda assim, ela nos instiga a concluir a discussão sobre os modos de temporalização do tempo (alienígena) da IA a partir do problema (humano) do treinamento.

A importância dos valores estabelecidos em treinamento já havia sido reconhecida em uma influente publicação de 2019 intitulada "Sobre a medida da inteligência". Nela, François Chollet, trabalhando para o *Google*, realizou uma ampla crítica ao campo da IA, em particular à maneira como ele lidava e definia o conceito de inteligência. Ao invés de testar a inteligência pela medida da eficiência com que a IA executa tarefas específicas, o artigo instiga a comunidade a medir a eficiência com que ela adquire novas habilidades. A "ideia central", segundo Chollet, é que "*a inteligência de um sistema é uma medida de sua eficiência de aquisição de habilidades em um escopo de tarefas, com relação a priors, experiência e dificuldade de generalização*" (CHOLLET, 2019, p. 27, grifos do autor)[7]. Essa é uma modificação notável de mentalidade, uma vez que desvia o foco da inteligência da performance (*output*) para o processo, ou, nos termos que apresentei mais acima, em direção à função diferencial entre passado e futuro.

As correções de valores éticos expressados pelas saídas do GPT-4, descritas no *System Card* do relatório de março de 2023, ilustram como respostas ofensivas exigem podar, através da elicitação de *priors* ou do aprendizado de reforço, a capacidade generativa do modelo, substituindo seus *outputs* por textos algo padronizados. Os modos de temporalização do tempo que buscam reparar, como atualização repetidora de discursos de exclusão social e reificação da diferença, denunciam o caráter subjetivamente contingente – e muito humano – do procedimento. Os valores repetidos em *feedback loop*, desde conjuntos numéricos fixados por tuplas, "atualizam disposições culturais cristalizadas no passado" solidificando-as no "presente como uma história incorporada" (AIROLDI, 2022, p. 125). Por esse motivo, "a melhor maneira de se abordar a temporalidade é treinar o modelo com dados de entrada agregados o mais próximo possível do tempo atual de implementação do modelo" (AMARO, 2022, p. 142). Esse tipo de medida, no entanto, continuaria apenas reproduzindo uma lógica de reformas constantes. Se é a "rede completa de relações mais do que humanas, ressonante com um fundo cultural coerente, que pode eventualmente

produzir uma mudança de comportamento" (AIROLDI, 2022, p. 130), talvez devêssemos seguir mais a fundo as recomendações antirracistas de Ramon Amaro e expandi-las, quem sabe, para todo tipo de produção da diferença, abandonando o atual e incorporando o caráter dinâmico da existência desde os processos computacionais:

> Para desalojar performances conscientes de normalidade, a classificação pode ser invertida para descrever entidades não como atuais, mas como uma forma iterativa de individuação. [...] De uma perspectiva de aprendizado de máquina, a tarefa da ciência e ontologia se alinharia a um processo relacional que, embora em intercâmbio com a classificação biológica e a priori, está aberto ao potencial para formas não humanas de razão (AMARO, 2022, p. 157).

Cabe apenas comentar que essas considerações vão no sentido oposto do "quadro de referência antropocêntrico" que François Chollet (2019, p. 24) sugere aos estudos de IA. Da maneira como entendo, o antropocentrismo e a abertura a "formas não humanas de razão" valem pela tensão criativa, localizada no seio do mais do que humano, entre computação necessária e computação contingente.

É possível uma computação contingente?

Na medida em que Louise Banks vai aprendendo a escrita dos heptápodes, ela começa a experimentar o tempo de modo não linear. Isso acontecia porque os humanos possuíam uma forma de consciência sequencial, e os alienígenas, uma forma de pensamento simultâneo. "Meus traços iniciais quase sempre se revelavam compatíveis com o que eu estava tentando dizer", e foi assim que ela começou "a desenvolver uma aptidão como a dos heptápodes" (CHIANG, 2016, p. 165). Memórias do futuro, fragmentos mentais do passado, era desse modo que "a experiência do alienígena" afetava a sua própria, "assim transformando-se", para recuperar a fenomenologia de Bernhard Waldenfels, "em um devir" (WALDENFELS, 2011, p. 3).

Existe uma diferença substancial entre o filme *A Chegada* (2016) e a noveleta *História de sua vida*, que o originou. Na adaptação para o cinema, um diálogo entre a linguista Banks e o físico Donnely menciona a hipótese de Sapir-Whorf, o que lança a discussão para o âmbito do relativismo linguístico. Na noveleta, ao contrário, um princípio físico é debatido. Esse era o princípio de Fermat, o primeiro que os cientistas conseguiram comunicar aos heptápodes, e que nos fala da mudança de direção da luz ao passar do ar para água. Fermat pode ser entendido em termos de causa e efeito, pela refração, mas também em termos de propósito: a luz parece procurar o caminho mais curto. Um princípio variacional, não muito difícil de explicar, mas que necessita de cálculo para uma boa descrição matemática. Era estranho que esse tenha sido o primeiro sucesso comunicativo: o que é complexo para nós, era muito simples aos alienígenas, e vice-versa. A noveleta, portanto, vai além do tema da incomensurabilidade entre notação matemática e linguagem natural (o caso de amor entre a linguista e o físico é apenas um dos muitos paralelismos literários empregados pelo autor). Os humanos experimentam o tempo de modo causal, mas podem apreender conceitualmente processos teleológicos, e assim a comunicação com os heptápodes se torna possível. "Como eventos da física, com suas interpretações causais e teleológicas, todo evento linguístico tinha duas interpretações possíveis: como uma transmissão de informação e como a concretização de um plano." (CHIANG, 2016 p. 180). A *História de sua vida* sugere uma questão sobre o sentido histórico, não apenas enquanto *significado* das experiências, mas fundamentalmente sobre a *direcionalidade* dos acontecimentos.[8]

Ainda que a IA Generativa alimentada por algoritmos de aprendizado profundo possa transcender os limites causais de sua programação original em direção a um "processamento telenômico de informação", ela "ainda não parece apresentar um salto ontológico na direção de uma teleologia" (FAUSTINO; LIPPOLD, 2023, p. 35). Falamos em "telenômico", pois seus componentes preditivos reciclam "passados altamente seletivos em um futuro apresentado como inevitável" (HONG, 2022, p. 386). Para lembrar da epígrafe que abre este artigo, uma IA otimizada conhece o futuro "de antemão", pois uma curadoria de dados históricos é afinada com a variável alvo estabelecida. As redes neurais artificiais *simulam* a compreensão microssemântica; processam *tokens* como valores matemáticos e, em seguida, os convertem de volta em texto ou imagens como saída. Não compreendem semântica, ao menos não como os humanos, mas podem explicá-la; são incapazes de dizer o tempo, mas podem contá-lo; não narrativizam a experiência, mas reconhecem nela padrões. Não preveem o futuro, mas o determinam, naturalizando dados discretos. Conceituada como uma "expressão ou função do movimento", a microtemporalidade das máquinas aprende senão uma sequência de "agoras", transformando "o tempo em uma espécie de linha pontilhada, uma série de pontos, em vez de uma continuidade" (SILVEIRA, 2023, p. 6). Seria essa uma trajetória muito longa, sempre lotada abaixo da percepção, e por isso artificialmente avessa ao discurso político?

O problema que trago como conclusão não diz respeito, no entanto, a um enésimo "fechamento de futuro", dessa vez promovido pela IA. Pelo contrário, argumento que a computação contingente, ou o potencial das máquinas produzirem algo *novo*, talvez não seja uma impossibilidade tecnológica, mas um impasse ético-político. Afinal, "a pergunta *no que pensam os algoritmos* é insuficiente se não discutirmos ao mesmo tempo por que são tão poucos aqueles que os fazem pensar e colhem seus resultados" (CANCLINI, 2021, p. 123, grifos do autor). Tão poucos aqueles que, controlando os "vetores da informação" (WARK, 2022), dominam de fato a sociedade global. Tão poucos aqueles que, monopolizando os dados e os parâmetros, pesos nas matrizes de análise, determinam o futuro naturalizando previsões desde um "*deus in machina*" definido por "escolhas e objetivos culturalmente informados" (AIROLDI 2022, p. 35). Anáfora, *looping de feedback* ou atualização repetidora.

Mas o que nos proíbe de aprender a língua dos heptápodes, de incutir valores éticos da historiografia nas máquinas, com isso projetando nossos próprios futuros? Caso apostemos no treinamento de nossos modelos de IA – à diferença dos métodos quantitativos tradicionais (desenhados para testar hipóteses) – poderemos minerar "paisagens de dados disponíveis em busca de novas hipóteses" (AMARO, 2022, p. 109). Para Beatrice Fazi, esse tipo de evidência demonstra como a computação não é, por natureza, necessária, mas contingente. Seu argumento central sugere a investigação de uma estética computacional de modo a explorar o potencial de autoatualização das máquinas. Aqui entramos em outro registro, não de uma ontologia baseada em Heidegger e Deleuze – tradição que enxergou o pensamento analógico como superior ao digital em capacidade de abstração – mas da fenomenologia de Henri Bergson e da filosofia processual de Alfred Whitehead. Fazi defende que, por trás das verdades auto evidentes dos sistemas axiomáticos formais, se esconde a possibilidade de indeterminação. A autora sugere essa ideia a partir do engajamento com textos fundadores do campo – especialmente o problema da incomputabilidade de Turing e o teorema da incompletude de Gödel. Ela estuda o significado ontológico dos processos de "discretização abstrativa" (FAZI, 2018, p. 16), argumentando que a incompletude e o incalculável sugerem um espaço aberto para as infinidades quantitativas e a indeterminação formal.

Como lembra Coeckelbergh (2023, p. 8), "o pensamento processual teve apenas uma influência limitada na filosofia contemporânea da tecnologia", com as exceções de Gilbert Simondon, "de quem Stiegler é (entre muitas coisas) um dos continuadores" (SILVEIRA,

2023, p. 21). Nesse registro, a realidade não é meramente uma coleção de objetos; ela é um processo de transformação. Humanos e não-humanos, assim como o dualismo sujeito-objeto, "não são pré-existentes, mas emergem de processos de transformação", o que nos leva a "reconhecer a dimensão temporal daquilo que chamamos de realidade" (COECKELBERGH, 2023, p. 8). A tese especulativa de Beatrice Fazi é desenhada para entender o caráter contingente da computação, operando, nos termos de Whitehead, como experiências capazes de autoatualização por "ocasiões atuais". Ela termina declarando a necessidade de futuros trabalhos "desenvolverem a teorização da ontologia contingente da computação em relação à investigação das relações sociais, culturais e econômicas que atuam sobre – ou são atuadas pela – indeterminação e eventualidade da computação" (FAZI, 2018, p. 209).

Nos sentimos tentados a parafrasear uma das questões fundadoras do pós-humanismo: "talvez possamos, ironicamente, aprender" – ao apostar no caráter contingente da computação – "como não ser o Homem, essa corporificação do logos ocidental" (HARAWAY, 2000, p. 83). Quando isso acontecer, quando passarmos a desenvolver nossos próprios sistemas, libertando-nos da confiança nos dados de treinamento e nos parâmetros de análise dos GAFAMI, a IA passará "para o primeiro plano", revelando-se "como sendo envolvida na criação e, de fato, como sendo mais do que uma ferramenta, mais do que um instrumento"? (COECKELBERGH, 2023, p. 6). Ao final, a discussão sobre computação contingente e computação necessária talvez seja um avatar mais do que humano das discussões tradicionais sobre agência e estrutura. A aposta na computação contingente pressupõe a crítica da (micro) semântica (micro) temporal algorítmica que caracteriza o *Sattelzeit* das máquinas, além do desenvolvimento de experimentos capazes de mensurar seus efeitos na condição histórica contemporânea. Essa postura implica em uma mudança de tópica, dos futuros passados e passados presentes, para os futuros presentes e presentes futuros. Com a hipótese da computação contingente, passamos da "destruição do significado" para a potencial reconstrução do sentido (enquanto significado e direcionalidade) de uma história mais do que humana. Isso significaria evitar uma atitude quixotesca de enfrentamento das máquinas e, aceitando a sugestão de Fazi, levar a sério o abstrativo das operações computacionais. Invertendo o problema, encontraríamos limitações repetidoras não na natureza matemática, formal e abstrata dos processos computacionais, mas nos elementos humanos de seu treinamento, permitindo, assim, uma abertura "potencial para formas não humanas de razão" (AMARO, 2022, p. 157).

Referências

ABRAM, D. *The spell of the sensuous:* Perception and language in a more-than-human world. New York: Vintage Books, 1997.

AIROLDI, M. *Machine Habitus:* Towards a sociology of algorithms. Cambridge: Polity Press, 2022.

ALPHAVILLE, une étrange aventure de Lemmy Caution. Direção: Jean-Luc Godard. Produção: André Michelin. Intérpretes: Eddie Constantine; Anna Karina; Akim Tamiroff; Valérie Boisgel; Jean-Louis Comolli; Michel Delahaye *et al*. Roteiro: Paul Éluard; Jean-Luc Godard. Paris: Mundial Filmes, 1965. 1 DVD (99 min), son. p&b.

AMARO, R. *The black technical object:* On machine learning and the aspiration of black being. London: Sternberg Press, 2022.

AMOORE, L. *Cloud ethics:* Algorithms and the attributes of ourselves and others. Durham: Duke University Press, 2020.

BEHDADI, D.; MUNTHE, C. A normative approach to artificial moral agency. *Minds and Machines*, v. 30, n. 2, p. 1-24, 2020.

BENJAMIN, R. *Race after technology:* Abolitionist tools for the New Jim Code. Cambridge: Polity, 2019.

BONALDO, R; PEREIRA, A. C. B. Potential History: reading artificial intelligence from indigenous knowledges. *History and Theory*, v. 62, n. 1, p. 3-29, 2023.

BREIMAN, L.; FRIEDMAN, J.; OLSHEN, R. A.; STONE, C. J. *Classification and regression trees.* Boca Raton: Chapman & Hall, 2017.

BURRELL, J. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, v. 3, n. 1, p. 1-12, 2016.

BUTTAZZO, G. Artificial consciousness: Utopia or real possibility? *IEEE Computer*, v. 34, n. 7, p. 24-30, 2001.

CANCLINI, N. *Cidadãos substituídos por algoritmos.* São Paulo: Edusp, 2021.

CHALMERS, D. J. Subsymbolic Computation and the Chinese Room. *In:* DINSMORE, J. (org). *The symbolic and connectionist paradigms.* New York: Psychology Press, 1992. p. 25-48.

CHIANG, T. *A história da sua vida e outros contos.* Rio de Janeiro: Editora Intrínseca, 2016.

CHOLLET, F. On the measure of intelligence. *arXiv:1911.01547v2 [cs.AI]*, 2019.

COECKELBERGH, M. The work of art in the age of AI Image Generation: Aesthetics and human-technology relations as process and performance. *Journal of Human Technology Relations*, v. 1, n. 1, p. 1-13, 2023.

DOMANSKA, E. Beyond anthropocentrism in historical studies. *Historein*, v. 10, n. 2, p. 118-130, 2010.

EDEN, A.; MOOR, J.; SØRAKER, J.; STEINHART, E. (org). *Singularity hypotheses:* A scientific and philosophical assessment. New York: Springer, 2012.

ESPOSITO, E. Artificial communication? The production of contingency by algorithms. *Zeitschrift für Soziologie*, v. 46, n. 4, p. 249-265, 2017.

FAUSTINO, D.; LIPPOLD, W. *Colonialismo digital:* por uma crítica hacker-fanoniana. São Paulo: Boitempo, 2023.

FAZI, B. Beyond human: Deep learning, explainability and representation. *Theory, Culture & Society*, v. 38, n. 7-8, p. 55-77, 2020.

FAZI, B. Can a machine think (anything new)? Automation beyond simulation. *AI & Society*, v. 34, n. 813, p. 813-824, 2019.

FAZI, B. *Contingent Computation:* Abstraction, experience, and indeterminacy in computational aesthetics. London, New York: Rowman Littlefield, 2018.

FLORIDI, L. AI and its new winter: From myths to realities. *Philosophy & Technology*, v. 33, p. 1-3, 2020.

FLORIDI, L. *The 4th revolution:* How the infosphere is reshaping human reality. Oxford: University Press, 2014.

FLUSSER, V. *Filosofia da caixa preta:* ensaios para uma futura filosofia da fotografia. Rio de Janeiro: Sinergia Relume Dumará, 2009.

GIACCADI, E.; REDSTRÖM, J. Technology and more-than-human design. *Design Issues*, Massachusetts, v. 36, n. 4, p. 33-44, 2020.

GURFINKEL, J. AI and the American Smile: How AI misrepresents culture through a facial expression. *Medium*, 2023. Disponível em: https://shorturl.at/ekDHU. Acesso em: 15 mar. 2023.

HAENLEIN, M.; KAPLAN, A. "A brief history of Artificial Intelligence: On the past, present, and future of Artificial Intelligence. *California Management Review*, v. 61, n. 4, p. 5-14, 2019.

HARAWAY, D. Manifesto Ciborgue. *In:* HARAWAY, D.; KUNZRU, H.; TADEU, T. *Antropologia do ciborgue:* as vertigens do pós-humano. Belo Horizonte: Autêntica, 2000. p. 35-118.

HARRIS, S. Z. Distributional structure. *Word*, v. 10, n. 2-3, p. 146-162, 1954.

HARTMAN, S. Vênus em dois atos. *Ecopos*, v. 23, n. 3, p. 12-33, 2020.

HAYLES, K. Approximating algorithms: From discriminating data to talking to an AI. *History and Theory*, v. 61, n. 4, p. 152-165, 2022.

HILL, R. K. What an algorithm is. *Philosophy & Technology*, v. 29, n. 1, p. 35-59, 2015.

HONG, S. Predictions without futures. *History and Theory*, v. 61, n. 3, p. 371-390, 2022.

HUGHES-WARRINGTON, M. Towards the recognition of artificial history makers. *History and Theory*, v. 61, n. 4, p. 107-118, 2022.

JOHNSON, D. G.; VERDICCHIO, M. AI, Agency and responsibility: The VW fraud case and beyond. *AI and Society*, v. 34, n. 3, p. 639-647, 2019.

JORDHEIM, H.; YTREBERG, E. After supersynchronisation: How media synchronise the social. *Time & Society*, v. 20, n. 3, p. 402-422, 2021.

KOSTECZKA, L. A. P. *Futuro usado:* Google & Facebook nos processos de plataformização da Internet. 2022. 194 f. Tese (Doutorado em História) – Setor de Ciências Humanas, Universidade Federal do Paraná, Curitiba, 2022.

LE, Q; MIKOLOV, T. Distributed representations of sentences and documents. *In: THE 31st INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 2014, Beijing*. Beijing: W&CP, v. 32, 2014, p. 1-9.

LUHMANN, N. *Observations on modernity*. Stanford: University Press, 1998.

LUPTON, D. The Internet of things: Social dimensions. *Sociology Compas*, v. 14, n. 4, e12770, 2020.

MATTOS, H. *Das cores do silêncio:* os significados da liberdade no Sudeste escravista (Brasil, século XIX). Campinas: Editora da Unicamp, 2013.

MCCORDUCK, P. *Machines who think:* A personal inquiry into the history and prospects of Artificial Intelligence. Natick: A. K. Peters, 2004.

MIKKOLA, P. *et al*. Prior knowledge elicitation: The past, present, and future. *arXiv:2112.01380*, 2021.

MITTELSTADT, B. D. *et al*. The ethics of algorithms: Mapping the debate. *Big Data & Society*, v. 6, n. 2, p. 1-21, 2016.

NOBLE, S. U. *Algoritmos da opressão:* como o Google fomenta e lucra com o racismo. Santo André: Rua do Sabão, 2021.

O'GORMAN, E.; GAYNOR, A. More-than-human histories. *Environmental History*, Chicago, v. 5, n. 24, p. 711-735, 2020.

OPEN AI. *GPT-4 Technical Report*. 2023. 100 p. Disponível em: https://cdn.openai.com/papers/gpt-4.pdf. Acesso em: 20 mar. 2023.

OPEN AI. Language models are few-shot learners. *arXiv:2005.14165v4*, 2020.

PEREIRA, M. H. F. *Lembrança do presente:* ensaios sobre a condição histórica na era da internet. Belo Horizonte: Autêntica, 2022.

PEREIRA, M. H. F.; ARAUJO, V. O passado como distração: modos de vestir a história no neopopulismo brasileiro. *Revista de Teoria da História*, Goiânia, v. 25, n. 2, p. 71-88, 2022.

PEREIRA, M. H. F.; ARAUJO, V. Vozes sobre Bolsonaro: esquerda e direita em tempo atualista. *In:* KLEM, B. S.; PEREIRA, M.; ARAUJO, V. *Do fake ao fato:* (des) atualizando Bolsonaro. Vitória: Editora Milfontes, 2020. p. 125-150.

PEREIRA, M.; ARAUJO, V. *Atualismo 1.0:* como a ideia de atualização mudou o século XXI. Mariana: Editora SBTHH, 2018.

ROTA, A. R.; NICODEMO, T. L. Arquivos pessoais e redes sociais: o Twitter construído como documento histórico. *Estudos Históricos*, Rio de Janeiro, v. 36, n. 79, p. 268-291, maio/ago. 2023.

SAYES, E. Actor-network theory and methodology: Just what does it mean to say that nonhumans have agency? *Social Studies of Science*, v. 44, n. 1, p. 134-149, 2014.

SEARLE, J. R. Minds, brains, and programs. *The Behavioral and Brain Sciences*, v. 3, n. 3, p. 417-457, 1980.

SICHMAN, J. S. Inteligência Artificial e sociedade: avanços e riscos. *Estudos Avançados*, v. 101, n. 35, p. 37-49, 2021.

SILVA, T. *Racismo algorítmico:* inteligência artificial e discriminação nas redes digitais. São Paulo: Edições Sesc, 2022.

SILVEIRA, P. T. The counted time: Technical temporalities and their challenges to history. *History and Theory*, v. 62, n. 3, p. 403-426, 2023.

SIMON, Z. *The Epochal Event:* Transformations in the entangled human, technological, and natural worlds. Cham: Palgrave, 2020.

SIMON, Z. Transformação do Tempo Histórico: temporalidades processual e evental. *Revista de Teoria da História*, v. 24, n. 1, p. 139-155, 2021.

SIMON, Z.; TAMM, M. Historical futures. *History and Theory*, v. 60, n. 1, p. 3-22, 2021.

TAMM, M.; SIMON, Z. B. More-than-human history: Philosophy of history at the time of the anthropocene. *In:* KUUKKANEN, J. M. *Philosophy of history:* Twenty-first-century perspectives. Londres: Bloomsbury Academic, 2020. p. 198-215.

TEGMARK, M. *Vida 3.0:* o ser humano na era da inteligência artificial. São Paulo: Benvirá, 2020.

TURING, A. Computing Machinery and Intelligence. *Mind*, v. 58, n. 236, p. 433-460, 1950.

VASWANI, Ashish *et al*. Attention is all you need. *arXiv:1706.03762v5* [cs.CL], 2017.

WALDENFELS, B. *Phenomenology of the Alien:* Basic concepts. Evanston: Northwestern University Press, 2011.

WARK, M. *O capital está morto.* São Paulo: Editora Funilaria e sobinfluência edições, 2022.

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

Notes

[1] O mais do que humano permite falar em uma ontologia da IA, uma vez entendendo que essa inteligência não é apenas artificial, mas igualmente carregada de emoção e de um *bias* excessivamente humano.

[2] Em março de 2023, o instituto *Future of Life*, celeiro singularitário patrocinado por Musk, publicou uma carta aberta à comunidade solicitando que a indústria interrompa "imediatamente pelo menos durante 6 meses o treinamento de sistemas de IA mais poderosos do que o GPT-4" e acrescentando que, caso "tal pausa não puder ser promulgada rapidamente, os governos deverão intervir e instituir uma moratória".

[3] Por exemplo, na Teoria Ator-Rede, os não humanos são entendidos como *acteurs à part entière*, sendo ressaltada sua plenitude do ponto de vista social como "mediadores de relações sociais" (SAYES, 2014, p. 137-138). Porém, certas questões não podem ser tratadas a partir dessa perspectiva simétrica, na qual a agência é "desacoplada dos critérios de intencionalidade, subjetividade e livre arbítrio" (SAYES, 2014, p. 139-141).

[4] O experimento foi conduzido como parte de um projeto de extensão que discute questões étnico-raciais usando IA Generativa. A inspiração inicial era criar uma "fabulação crítica" (HARTMAN, 2020, p. 28) por meio de imagens, seguindo a proposta de Saidiya Hartman em relação à narrativa, abordando histórias de pessoas com pouca informação primária.

[5] Fotografia de um jovem negro sorridente do século XIX em uma barbearia, com idade entre 18 a 19 anos, de baixa estatura, constituição física esguia, sem barba, dentes da frente limados, boca ligeiramente grande, polaroid, iluminação de baixo para cima - v 5 (tradução nossa).

[6] Outros elementos da imagem merecem destaque devido à sua relação com a cultura estado-unidense. Entre eles, encontra-se o retrato do poeta norte-americano Edgar Allan Poe (1809-1849), situado no canto superior esquerdo, que acredito ter sido identificado pelo *Midjourney* como representativo do *token* "*XIX century*".

[7] Os *priors* representam informações a priori que são incorporadas aos dados de treinamento e auxiliam na tomada de decisões. Eles podem ser obtidos por meio da avaliação subjetiva de um engenheiro, de sua experiência ou juízo. Mikkola (2021, p. 4) descreve como isso pode ser feito de "maneira estruturada, expressando esse conhecimento como distribuições de probabilidade a priori". Os *priors* são frequentemente descritos como representações de crenças ou conhecimentos anteriores: no contexto da inferência Bayesiana, a distribuição a posteriori deve refletir a *atualização* da crença original após a exposição do modelo aos dados.

[8] Agradeço a Eduardo Ferraz Felippe pela lembrança dessa noveleta.

Rodrigo Bragio Bonaldo é Professor do Departamento de História da Universidade Federal de Santa Catarina (UFSC), com graduação (2005), mestrado (2010) e doutorado (2014) em História pela Universidade Federal do Rio Grande do Sul (UFRGS). Durante o doutorado, realizou um período sanduíche na École des Hautes Études en Sciences Sociales (EHESS), em Paris (2011-2012). Concluiu seu pós-doutorado na Freie Universität Berlin (FU-Berlim) como pesquisador visitante (2022-2023). Seus interesses de pesquisa incluem o estudo do pensamento histórico fora das normas tradicionais da disciplina, focando em narrativas jornalísticas da história, discursos comemorativos, historiografias pré-modernas e a escrita da história na Wikipedia. Atualmente, trabalha no desenvolvimento de métodos de aprendizado de máquina aplicados ao processamento de linguagem natural (PLN) e na análise dos limites da operação historiográfica artificial. É também professor do quadro permanente do Programa de Pós-Graduação em História Global da UFSC, onde orienta mestrados e doutorados.

Submissão: 03/04/2023
Aceite: 25/08/2023

Editores: **Karina Anhezini** e **Eduardo Romero de Oliveira**