UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE FILOSOFIA E CIÊNCIAS HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ANTROPOLOGIA SOCIAL

Louise Lima Karczeski

Gerando inteligência: considerações etnográficas sobre a construção de objetividade no campo da Inteligência Artificial

Florianópolis
2022

Louise Lima Karczeski

Gerando inteligência: considerações etnográficas sobre a construção de objetividade no campo da Inteligência Artificial

Dissertação submetida ao Programa de Pós-Graduação em Antropologia Social da Universidade Federal de Santa Catarina como requisito parcial para a obtenção do título de Mestre em Antropologia Social.

Orientador(a): Profa. Dra. Letícia Maria da Costa Nóbrega Cesarino,

Florianópolis

2022

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Karczeski, Louise Lima
Gerando inteligência : considerações etnográficas sobre a construção de objetividade no campo da Inteligência Artificial / Louise Lima Karczeski ; orientador, Letícia Maria Costa da Nóbrega Cesarino, 2022.
125 p.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro de Filosofia e Ciências Humanas, Programa de Pós-Graduação em Antropologia Social, Florianópolis, 2022.

Inclui referências.

1. Antropologia Social. 2. Inteligência artificial. 3. Objetividade. 4. New media studies. 5. Neoliberalismo. I. Cesarino, Letícia Maria Costa da Nóbrega. II. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Antropologia Social. III. Título.

Louise Lima Karczeski

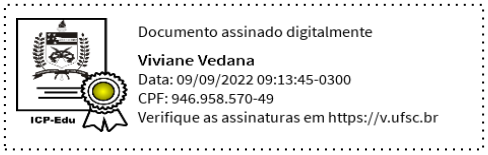
Gerando Inteligência: considerações etnográficas sobre a construção de objetividade no campo da Inteligência Artificial

O presente trabalho em nível de Mestrado foi avaliado e aprovado, em 19 de agosto de 2022, pela banca examinadora composta pelos seguintes membros:

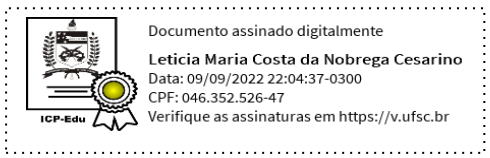
Prof. Gabriel Coutinho Barbosa, Dr.
Universidade Federal de Santa Catarina

Prof.(a) Rogério Lopes Azize, Dr.
Universidade Estadual do Rio de Janeiro

Certificamos que esta é a versão original e final do trabalho de conclusão que foi julgado adequado para obtenção do título de Mestre em Antropologia Social.



Coordenação do Programa de Pós-Graduação



Prof.(a) Letícia Maria Costa da Nóbrega Cesarino, Dr.(a)
Orientador(a)

Florianópolis, 2022.

AGRADECIMENTOS

Enquanto redigia esta dissertação, decidi que a seção de agradecimentos seria a última que eu escreveria. Agora, enquanto escrevo, penso que seria mais apropriada uma seção de “desculpas” a todos que perturbei durante a longa escrita deste texto. Portanto, peço desculpas gratas à minha rede de apoio:

Aos meus pais, Elisabete e Enio, que possibilitaram esta pesquisa e todo o resto. Não é qualquer pessoa que pode dizer que tem os melhores pais do mundo. Aliás, só eu.

À Bea e à Lari, que em algum momento deixaram de ser minhas amigas e se tornaram parte de mim. Desenvolvi um modelo de Machine Learning, joguei todas as nossas conversas dos últimos três anos lá e está estatisticamente comprovado que vocês têm muita paciência.

Ao João, que me ajudou a manter o foco por sempre oferecer o conforto de ter um amigo com o qual posso falar e rir sobre tudo.

A todos os meus amigos que me ajudaram a manter e dispensar a sanidade nos últimos anos, conforme o necessário.

Enfim, os agradecimentos:

Agradeço aos colegas do mestrado.

Agradeço à Letícia, minha orientadora, por ajudar a encaminhar esta dissertação e por sempre contribuir para a renovação do meu interesse pela antropologia através do seu trabalho.

Estendo os agradecimentos aos professores do PPGAS/UFSC que admiro e que inspiraram esta pesquisa de alguma forma: Gabriel, Jeremy, Kelly, Márnio, Rafael e Viviane.

Por fim, agradeço aos interlocutores que colaboraram com esta pesquisa, propondo-se ao árduo exercício de pensar junto.

EPÍGRAFE

Bem jantado, bem vestido, bem dormido, não tinha energia necessária para fazer entrar na cachola aquelas coisas esquisitas. Comprei livros, assinei revistas: Revue Anthropologique et Linguistique, Proceedings of the English-Oceanic Association, Archivo Glottologico Italiano, o diabo, mas nada! E a minha fama crescia. Na rua, os informados apontavam-me, dizendo aos outros: "Lá vai o sujeito que sabe javanês." Nas livrarias, os gramáticos consultavam-me sobre a colocação dos pronomes no tal jargão das ilhas de Sonda. Recebia cartas dos eruditos do interior, os jornais citavam o meu saber e recusei aceitar uma turma de alunos sequiosos de entenderem o tal javanês. A convite da redação, escrevi, no Jornal do Comércio um artigo de quatro colunas sobre a literatura javanesa antiga e moderna...

— Como, se tu nada sabias? interrompeu-me o atento Castro.]

— Muito simplesmente: primeiramente, descrevi a ilha de Java, com o auxílio de dicionários e umas poucas de geografias, e depois citei a mais não poder.

— E nunca duvidaram? perguntou-me ainda o meu amigo.

— Nunca [...].

(Lima Barreto, O homem que sabia javanês)

RESUMO

A partir de 2010, o termo “Inteligência Artificial” (IA) foi reavivado no *mainstream*. Com a evolução das técnicas da subárea conhecida como “machine learning” (ML), empresas começaram a investir em IA gerando uma expansão rápida da área. Essa expansão foi acompanhada pela cobertura midiática e acadêmica de casos de discriminação algorítmica, assim como práticas de invasão de privacidade, que estimularam o debate público sobre as implicações éticas da adoção de aplicações de IA. Entretanto, o uso do conceito nessas discussões permanece indefinido. A presente pesquisa busca abordar o modo como o crescimento da atenção direcionada à IA incidiu nas práticas dos especialistas do campo da tecnologia, envolvidos na elaboração de modelos de IA. Trata-se de uma etnografia realizada, principalmente, em conferências voltadas ao tema. Também abrange a participação em cursos, grupos de comunidades voltadas à IA em redes sociais e conversas com interlocutores conhecidos nesses espaços. Através da inspiração em trabalhos no campo dos Science and Technology Studies (STS) e dos New Media Studies, o objetivo é abordar como a construção de objetividade se deu em meio aos especialistas. Desse modo, a pesquisa aborda o tema através de três ângulos: o mercado, a ciência e a ética. Tratam-se de enquadres que surgiram como relevantes para o modo como interlocutores desenvolvem e reivindicam a objetividade de aplicações de IA perante as críticas direcionadas a elas no senso comum. Assim, a discussão identifica a centralidade da divisão entre “dois tipos” de IA nos discursos dos especialistas, sendo o desenvolvimento recente da área associado à ênfase no caráter técnico e segmentado dela. Além disso, a divisão aponta para a transposição das pesquisas em IA da academia para o mercado. Ainda, a pesquisa aborda a recursividade entre metáforas computacionais e naturais como parte constitutiva dos processos de criação de modelos, apontando para noções econômicas de natureza e de pessoa. Por fim, focamos no modo como a discussão ética, principalmente acerca da noção de “vieses” que atrapalham a performance objetiva dos modelos, atua na manutenção de divisões entre o âmbito da técnica e o âmbito do social.

Palavras-chave: Inteligência Artificial; objetividade; New Media Studies; neoliberalismo.

ABSTRACT

As of 2010, the term “artificial intelligence” (AI) has been revived in the mainstream. With the evolution of techniques in the sub-area known as “machine learning” (ML), companies began to invest in AI. This expansion was amplified by the media and academic coverage of algorithmic discrimination cases, as well as privacy-invasion practices, which spurred public debate on the ethical implications of adopting AI applications. However, the use of the concept in these analyzes remains undefined. The present research seeks to address the way in which the growth of attention directed to AI has influenced the practices of experts in the field of technology, those involved in the elaboration of AI models. It is an ethnography carried out, mainly, in conferences focused on the subject. It also encompasses the participation in courses, social media groups created by AI communities and conversations with interlocutors known in those spaces. Inspired by works in the field of Science and Technology Studies (STS) and New Media Studies, this work aims to address how the construction of objectivity took place among the experts. The research approaches the subject through three angles: the market, science and ethics. These are frames that emerged as relevant in the way interlocutors develop and claim the objectivity of AI applications in the face of criticism directed at them in common sense. Thus, the discussion identifies the centrality of the division between “two types of AI” in the experts’ discourses, and the association between the recent development of the area with the emphasis on its technical and segmented character. The division points to the transposition of AI research from academia to the market. The dissertation also addresses the recursive relationship between computational and natural metaphors as a constitutive part of the construction of AI models, which point to economic views of nature and personhood. Finally, we focus on the way in which the ethical discussion, mainly about the notion of "biases" that hinder the objective performance of the models, acts in the maintenance of the divisions between the technical and the "social" scope.

Keywords: Artificial Intelligence, objectivity, New Media Studies; neoliberalism.

LISTA DE ILUSTRAÇÕES

LISTA DE ABREVIATURAS E SIGLAS

BBS - Bulletin Board System

DL - Deep Learning

IA - Inteligência Artificial

ML - Machine learning

TDC - The Developer’s Conference

SUMÁRIO

1 INTRODUÇÃO

A partir dos anos 2010, o conceito de inteligência artificial (IA), acompanhado pela noção de "algoritmo", passou a protagonizar o debate público sobre tecnologia. As repercussões sobre práticas de vigilância, invasão de privacidade e discriminação através do uso de sistemas algorítmicos direcionaram a atenção coletiva para os processos envolvidos no desenvolvimento da IA (CHUN, 2021; EUBANKS, 2019; NOBLE, 2018). A presente pesquisa busca abordar as reverberações desse movimento da esfera pública em meio aos especialistas da IA, que atuam no desenvolvimento de suas aplicações. Com base na experiência etnográfica em eventos, grupos em redes sociais, entrevistas e cursos criados por profissionais na área no Brasil, abordaremos a produção de objetividade no campo.

Para isso, considerei três ângulos transversais que emanaram da pesquisa como relevantes no debate, discutindo a IA como mercado, como ciência e como ética. Trata-se de um recorte analítico definido de acordo com padrões identificados em campo, porém a leitura evidenciará que essas são esferas sobrepostas. Assim, como recomendação aos leitores, sugiro pensá-las através do que Marwick e Boyd (2010) definem como “colapso de contextos”. As autoras utilizam o conceito em referência à característica da interação em mídias sociais mainstream, nas quais as fronteiras entre diferentes contextos sociais se desfazem. Nesses ambientes, portanto, as ações de usuários são direcionadas a audiências múltiplas, o que implica numa reorganização das suas práticas. Para os nossos fins, entendo como uma inferência do campo a aplicabilidade da noção na abordagem do desenvolvimento das tecnologias que propiciam essas interações. Ou seja, nos processos de criação da IA fatores mercadológicos, científicos e éticos constantemente se interseccionam.

Posto isso, para contextualizar a presente pesquisa, gostaria de falar sobre as duas principais estranhezas em meio às quais vivi durante a realização dela. A primeira, relativamente familiar devido ao ofício da antropologia, foi associada à imersão em ambientes perpetuados por linguagens que eu não domino. Refiro-me a linguagens de programação, teorias da computação, gírias, anglicismos que, a princípio, não faziam sentido para mim. Para tentar diminuí-la, foram necessárias numerosas horas dedicadas a conhecer os humanos que compunham meu campo, de acordo com as experiências relatadas por eles, os discursos e práticas que trocam

entre si. Nesse intuito, também precisei reservar ainda mais tempo para o estudo de materiais aos quais esses contatos me levaram, buscando entender o ato de programar e mimetizando, de certa forma, a introspecção perante o computador que é tão central nas rotinas de meus interlocutores.

A segunda estranheza, porém, contrasta com a familiaridade da primeira. Ela está relacionada ao “estrelato” do meu tema de pesquisa escolhido: durante o mestrado, vivi sob a impressão de que o interesse público, já ávido, que existia por IA estava se multiplicando. A causa dessa estranheza dupla foram os constantes lembretes de minha pesquisa com os quais me deparei no âmbito da universidade, da televisão, ficção, redes sociais, interações e atividades cotidianas. Essa efervescência de informações, que a mim chegava independentemente de minha vontade, me afetou, como antropóloga, ao tirar de mim a responsabilidade de inserir meu interesse de pesquisa nas demais áreas da vida e ao me atribuir a responsabilidade de encontrar, em meio a esse emaranhado de fatos, algo a ser investigado que fizesse sentido dentro dos parâmetros de uma pesquisa de mestrado na Antropologia Social.

O conceito que abrange a descrição a seguir é de difícil definição. A inteligência artificial consiste na combinação de duas palavras, individualmente controversas, que protagonizou debates interpretativos desde quando passou a distinguir uma disciplina, há cerca de sessenta anos. Não obstante, é um conceito que engaja a atenção pública, transitando entre os extremos de abordagens otimistas e pessimistas, entre ciclos de hype e medo (ELISH, BOYD, 2017), tomando ambas as posições de problema e solução. Nos últimos anos, a inserção da palavra “algoritmo” no senso comum veio acompanhada de uma visão crítica, motivada, principalmente, pela identificação de efeitos potencialmente prejudiciais a grupos sociais específicos decorrentes do uso de dados enviesados em aplicações de IA, em diversas frentes (CODED, 2019; EUBANKS, 2018; NOBLE, 2018). Nesse cenário, a contestação da neutralidade e da imaterialidade de novas tecnologias, especialmente as automatizadas, tornou-se objeto de interesse para especialistas de áreas tradicionalmente distantes do campo tecnológico. As ciências humanas, em particular, vieram a adquirir papel central no contexto de desenvolvimento de iniciativas de regulação da IA que têm ocorrido ao redor do mundo.

As abordagens das ciências humanas, entretanto, tendem a compor o discurso público através da exposição dos “efeitos” dessas tecnologias naqueles que as experienciam. Já a investigação das epistemologias envolvidas nos processos de sua

criação atrai consideravelmente menos atenção. Esse aspecto, que indica a distância entre as pesquisas sobre tecnologia no âmbito de ciências não exatas e os métodos e processos envolvidos na criação de novas tecnologias - especialmente as de IA - está documentado na teoria (BEER, 2017; FORSYTHE, 1993; SEAVER, MOATS, 2019), e também foi um problema levantado em campo. Meus interlocutores apontaram para a difusão de críticas aos seus modelos como evidência, em parte, de uma incompreensão do escopo e dos fundamentos de seu trabalho na prática. Isso nos leva a entender esta pesquisa como relevante para ambas: a investigação de valores e práticas que configuram a área no Brasil, e a reflexão sobre a abordagem antropológica perante a IA.

O presente interesse de pesquisa tomou forma, inicialmente, através do interesse pelo campo dos Science and Technology Studies, que se manifestou através da preocupação com os efeitos que a interação humano-máquina produzia nas vidas das pessoas, justamente. Na graduação, considerando a característica das redes sociais como ambientes organizados a partir de estruturas algorítmicas desconhecidas, me dediquei a pesquisar a demarcação de fronteiras entre grupos políticos rivais considerando a existência de algoritmos de filtragem de conteúdo em uma plataforma específica - o *Facebook* - e a circulação de informações em ambientes antagônicos (KARCZESKI, 2018). Na ocasião, a pesquisa consistiu em um esforço para pensar a agência humana na manutenção de suas “bolhas” nessa plataforma, num período em que as narrativas de determinismo tecnológico eram protagonistas na discussão sobre algoritmos que ganhava popularidade (PARISER, 2011).

Nessa investigação, “algoritmos” e “inteligência artificial” frequentemente surgiram como termos “guarda-chuva” - na mídia, principalmente. Suas definições formais pareciam, em parte, incompatíveis com as práticas a eles associadas , ao considerarmos a centralidade que adquiriam em narrativas de progresso e controle das tecnologias sobre os humanos. Portanto, dada a disseminação da cobertura sobre o tema e esse obstáculo conceitual, o interesse inicial que impulsionou o desenvolvimento desta pesquisa foi abordar a IA a partir da técnica, me voltando a práticas de pessoas programadoras que atuam na área, que são capacitadas para o desenvolvimento de aplicações de IA e divulgam o conhecimento sobre elas. Assim, a minha participação em conferências voltadas ao tema, pensadas como portas de entrada ao campo, contribuiu para identificar que o que eu pensava ser uma questão

bem estabilizada em meio aos especialistas era, na verdade, um dos “pontos instáveis” (HELMREICH, 2014) nas suas discussões.

Mais especificamente, já no início da pesquisa me deparei com interlocutores que, profissionalmente comprometidos com a área, sugeriam que os sensacionalismos que a cercam impedem a compreensão de como funcionam aplicações de IA e, associado a isso, do seu próprio ofício. Esse foi um problema recorrente nas discussões do campo, e decidi permanecer com ele, tomando uma abordagem mais ampla para a minha pesquisa - vantajosa, também, no contexto pandêmico em que a experiência etnográfica se deu. Do intuito inicial, que era seguir o enfoque da técnica, passei a direcionar meu foco às controvérsias e aos significados associados à IA em meio à comunidade de profissionais da área, no Brasil. Desse modo, pensei a pesquisa em sintonia com movimentos mais recentes nos STS, que expandiram dos campos mais restritos de investigação dos laboratórios para congressos, conferências, etc. buscando campos em que decisões políticas acerca de objetos científicos e tecnológicos são definidas e discutidas (HESS, 2001).

Assim, levantei um material etnográfico, também, amplo. Priorizei a participação em conferências voltadas à IA incluindo as de caráter de divulgação interna - sendo a The Developer’s Conference (TDC) a que acompanhei mais consistentemente - e também algumas voltadas ao grande público. Além disso, realizei entrevistas pontuais com alguns interlocutores que conheci nesses espaços, participei de grupos em redes sociais de comunidades das quais estes faziam parte e também participei de eventos e atividades divulgadas/organizadas por alguns deles. Apesar de meu olhar não privilegiar a abordagem da técnica, participei de cursos e workshops de programação, a fim de me familiarizar com as ferramentas utilizadas por meus interlocutores.

Como o campo foi realizado, predominantemente, no cenário da pandemia de Covid-19, o acesso aos eventos mencionados foi facilitado pela proliferação de atividades online. Porém, esse volume de conteúdos também exigiu a tomada de algumas decisões relacionadas ao método e à forma da pesquisa. Desse modo, considerando as dificuldades de estabelecer uma relação etnográfica consistente ao longo do tempo, em campos exclusivamente digitais, priorizei o foco no acompanhamento contínuo de uma conferência - o TDC, na “trilha” específica de IA - , e, como citado, em eventos associados aos seus palestrantes. Isso significa que o perfil de meus interlocutores, apesar de diversos, também é direcionado a

“comunicadores” da área. Ou seja, pessoas envolvidas na divulgação de aplicações de IA, associadas a outras comunidades segmentadas de IA e, muito frequentemente, às Big Techs[1].

Faço essa distinção para estabelecer que, diferentemente de outros trabalhos na antropologia cujo foco são apropriações subversivas e ressignificações de tecnologias digitais, o contexto desta pesquisa me levou a identificar que as questões que encontrei em campo e os significados atribuídos à IA fazem parte de um mainstream tecnológico. Assim, a pesquisa foi dedicada, em grande parte, às nuances das práticas associadas à IA, através dos três eixos definidos para a abordagem da produção de objetividade (DASTON, GALISON, 2007) no campo: na relação da IA com o mercado, com a ciência e com a ética. Para isso, considerando a preocupação com a forma do texto e com a integração campo-teoria, busquei contrapor as literaturas mais específicas do acervo antropológico - especialmente obras da antropologia digital e dos STS - às literaturas buscadas com base em referências a autores em campo, incluindo, assim, materiais teóricos de outras disciplinas que são relevantes para o debate proposto.

No primeiro capítulo, nos ocuparemos em apresentar o campo e descrever o caráter mercadológico que define a maior parte dos processos da IA. Desse modo, abordaremos as controvérsias que cercam a definição (ou falta dela) de “inteligência artificial”, assim como a divisão entre dois tipos de IA, frequentemente referida por interlocutores como parte da apresentação da área. Através da diferenciação entre uma IA “antiquada[2]”, na forma da IA forte ou Simbólica, e a IA que se faz atualmente, a IA fraca ou Conexionista, os discursos no campo associam determinadas virtudes morais às máquinas. Nesse sentido, enquanto a primeira é associada à superação da inteligência humana, à ficção e ao medo, a outra é caracterizada como aumento da inteligência humana, a aplicações altamente segmentadas e ao hype de mercado. Na ênfase do caráter técnico dos sistemas desenvolvidos hoje em dia, e da lógica de

[1] O termo se refere ao modelo de negócios de "empresas associadas ao uso intensivo de dados" (MOROZOV, 2018, p. 149). No geral, são empresas sediadas nos Estados Unidos, no Vale do Silício. Recentemente, a presença de empresas desse tipo na China se tornou relevante. Entretanto, o uso mais frequente de "Big Techs" costuma referir-se às cinco maiores empresas no ramo, também conhecidas como "tech giants": Alphabet (Google), Amazon, Apple, Meta e Microsoft. Aqui, respeitando o uso do termo em campo, será usado em referência a esse modelo de negócios conforme aplicado nas empresas nos EUA.

[2] Por vezes, a IA Simbólica é referida como "good old-fashioned AI" na comunidade técnica (KATZ, 2021).

mercado subjacente a eles, constrói-se uma imagem objetiva dos processos da IA. Desse modo, discutiremos o modo como a IA fraca, especificamente na forma do machine learning (ML), fundamenta-se epistemologicamente nas ciências de rede, que ancoram concepções sobre relações sociais como “conexões”, compatíveis com uma visão de mundo neoliberal [3](CHUN, 2016, 2021).

No segundo capítulo, direcionamos o foco para o modo como a recursividade entre metáforas computacionais e metáforas “naturais” informa as práticas no campo. A natureza, nesse sentido, é tomada como referência para a criação de inteligência através do entendimento de processos naturais como processos computacionais. Discutiremos como a interpretação sobre a natureza como signo da objetividade obscurece os processos de construção do conhecimento sobre ela no âmbito da IA, apesar de serem explícitas as influências da biologia darwinista e da teoria econômica neoliberal nas suas práticas. Também abordaremos o modo como o paradigma conexionista associado à IA fraca perpetua uma concepção de pessoa como cérebro (AZIZE, 2010) através do conceito de “rede neural”. As redes neurais artificiais, modelos centrais para o desenvolvimento do ML e definitivos no crescimento recente do interesse comercial na IA, remetem às associações do discurso neurocientífico com a visão de mundo neoliberal que fundamenta as interpretações sobre a natureza no campo. Através dessas associações, entendemos que a IA produz uma percepção científica sobre seus processos, que contribui para a confiança dos usuários nas suas aplicações. Essa percepção é fundamentada por uma concepção do conhecimento que tem como meta a exclusão da intervenção humana e que informa a noção de aprendizado mobilizada na área, como veremos.

No capítulo 3, tomaremos como foco as manifestações sobre a incipiência das discussões sobre ética na área e as suas repercussões em campo. Abordaremos o “vazamento” de vieses dos especialistas nos modelos de ML como o principal problema ético levantado na esfera pública sobre a automação. Também discutiremos o modo como a área responde a ele através do conceito de “explicabilidade”, que promove um enquadre técnico de problemas de desigualdade social. Através da

[3] : O uso de neoliberalismo, aqui, remete ao que Chun (2016, p. XI) define como "[...] uma época que enfatiza o empoderamento individual e a diferença." Além disso, complementamos esse entendimento com o de Mirowski (2019), que entende o neoliberalismo, para além de uma doutrina econômica, como um projeto epistêmico. A filosofia que une as diferentes manifestações do neoliberalismo, segundo o autor, é uma que postula os humanos como agentes cognitivos não confiáveis e o mercado como processador de informações ideal (MIROWSKI, 2019, p. 5). Para ambos os autores, tecnologias digitais atuam na retroalimentação de visões de mundo neoliberais.

literatura crítica sobre a IA, discutiremos o modo como o protagonismo da discussão sobre vieses torna periféricas as análises sobre axiomas que fundamentam as classificações e predições de modelos de ML através do uso de correlações (CHUN, 2021).

Nesse sentido, vamos nos direcionar para a epistemologia empirista do Big Data, pautada numa noção de que “os números falam por si” (BOYD, CRAWFORD, 2011), que aponta para uma reformulação dos procedimentos de pesquisa social perpetuada pelo desenvolvimento da IA predominantemente no âmbito da indústria. Por fim, abordaremos o modo como a diversidade foi apropriada como valor em resposta às controvérsias éticas da IA num sentido de “paridade quantitativa” que reflete os pressupostos epistemológicos e formas de produção de objetividade que guiam o desenvolvimento da área. Ao término da discussão, espero que a pesquisa contribua para situar o tema da inteligência artificial no debate antropológico brasileiro.

2 INTELIGÊNCIA ARTIFICIAL COMO MERCADO

2.1 SITUANDO A PESQUISA

Comecemos nossa discussão com uma breve digressão. No Brasil, 1995 foi o ano que marcou o início do funcionamento da Internet comercial no país, porém, foi no início dos anos 2000 que o mercado dos provedores de serviço de Internet (ISPs) brasileiro passou por uma revolução que tornou o acesso *web* menos exclusivo: era lançado o IG (na época, “Internet Grátis”; hoje, “Internet Group”), um ISP que não cobrava dos consumidores a taxa de conexão do provedor. A estratégia de negócios do IG foi inspirada na da empresa britânica *Freeserve*, que triunfou no fim dos anos 90 ao abandonar a cobrança da taxa mencionada e apostar nos ganhos com publicidade e outros serviços proporcionados pelo aumento do número de usuários (TECNOCRACIA…, 2019). A empresa durou apenas dois anos, período em que se tornou líder no mercado de ISPs, até ser adquirida por uma divisão da France Telecom, mas sobreviveu no imaginário público como marco do início do acesso em massa à Internet na Europa.

Similarmente, o IG fez sucesso no Brasil sob o lema da “Internet grátis”, o que fez os provedores já existentes reformularem seus modelos de negócios e dezenas

de outras empresas tentarem reproduzir o modelo. Isso foi feito com considerável falta de planejamento, tentando cooptar parte dos volumosos investimentos em serviços digitais. A saturação do mercado, combinada com o estouro da bolha da Internet (bolha dot-com) no final dos anos 1990, levou à falência a maioria dessas empresas. A IG sobreviveu, mas com a dificuldade de obtenção de crédito após o estouro e, com o advento de novas tecnologias de Internet, como a banda larga ADSL, foi perdendo o domínio (PRESCOTT, 2015a) até ser, em 2004, adquirida pela filial da Telecom no Brasil.

Uma das histórias pessoais que foi central nesse contexto é a de Aleksandar Mandic, sócio e co-fundador da IG, que, cerca de dez anos antes da fundação da empresa, firmou seu lugar na história do empreendedorismo digital brasileiro ao colocar em funcionamento um dos primeiros “BBS” (Bulletin Board System) do país. Este software permitia, através de uma linha telefônica, conectar-se a algum outro sistema através do computador. De acordo com ele, o Mandic BBS nasceu “sem querer”, da ideia de estabelecer a conexão entre computadores remotos para a resolução de problemas em programas na empresa em que trabalhava, o que fez na sua própria casa, utilizando seu computador e, ambas, as linhas telefônicas que ele tinha e a de sua esposa. Como a abertura para uso corporativo não despertou o interesse de novos usuários, Mandic decidiu abrir o acesso ao BBS para o público, quando começou a crescer (PRESCOTT, 2015b).

Também nesse cenário, do final dos anos 90 e com a popularização do acesso à Internet, a Editora Loyola colocava em circulação no Brasil a obra “A inteligência coletiva: por uma antropologia do ciberespaço”, de Pierre Lévy. Esta foi a primeira obra a afirmar o conceito que dá nome ao título como assunto de interesse dentro das ciências sociais. No texto o autor descreve a criação de um novo “espaço antropológico” - entendido por ele como um “sistema de proximidade (Espaço) próprio do mundo humano (antropológico)” (LÉVY, 1998, p. 22), dependente, portanto, de convenções linguísticas, técnicas, simbólicas, etc. - que nominou “Espaço do Saber”. O “Espaço do Saber” consiste, para Lévy, num novo horizonte civilizatório no qual os laços sociais seriam baseados em relações de saber, diferentemente do “Espaço das mercadorias” precedente, no qual a participação no fluxo das trocas econômicas é o que conta na definição de identidades sociais.

Nesse novo Espaço, portanto, prevaleceria a “inteligência coletiva”, definida sintética e repetidamente pelo autor da seguinte forma: “É uma inteligência distribuída

por toda parte, incessantemente valorizada, coordenada em tempo real, que resulta em uma mobilização efetiva das competências" (LÉVY, 1998, p. 28). Trata-se de um projeto que, como define o próprio autor, tem caráter utópico e abrange na sua definição a valorização das mais diversas competências humanas, não reconhecidas pelos mecanismos de avaliação de inteligência formais. Essas competências seriam abraçadas e mobilizadas em prol dessa inteligência distribuída, que é culturalmente condicionada e constantemente reavaliada. Para a nossa discussão, o ponto a reter é o papel atribuído às tecnologias informacionais nesse projeto, que teriam um papel fundamental na coordenação dos diferentes saberes:

> Os conhecimentos vivos, os *savoir-faire* e competências dos seres humanos estão prestes a ser reconhecidos como a fonte de todas as outras riquezas. Assim, que finalidade conferir às novas ferramentas comunicacionais? Seu uso mais útil, em termos sociais, seria sem dúvida fornecer aos grupos humanos instrumentos para reunir suas forças mentais a fim de constituir intelectuais ou "imaginantes" coletivos. A informática comunicante se apresentaria então como a infra-estrutura técnica do cérebro coletivo ou do *hipercórtex* de comunidades vivas. O papel da informática e das técnicas de comunicação com base digital não seria "substituir o homem", nem aproximar-se de uma hipotética "inteligência artificial", mas promover a construção de coletivos inteligentes, nos quais as potencialidades sociais e cognitivas de cada um poderão desenvolver-se e ampliar-se de maneira recíproca (LÉVY, 1998, p. 25).

Nos interessa pensar as vantagens de mercado advindas do uso da expressão "inteligência coletiva" que, de acordo com Chun (2021, p. 6), foi apropriada por agentes do Silicon Valley na criação do discurso sobre a Web 2.0, referindo-se à participação dos usuários na sua construção. Além disso, o conceito conforme descrito por Lévy (1998) nos interessa, pois, a sua obra se localiza no contexto das primeiras abordagens antropológicas sobre a Internet. Apostando na ideia da desterritorialização das identidades como uma potencialidade democrática do advento de novas tecnologias, o autor se encaixa no lado dos "apologéticos" e não dos "apocalíticos", dentro da divisão ideológica que marca as primeiras literaturas sobre o assunto nas ciências sociais (RIFIOTIS, 2002). Assim, trago o argumento de Lévy aqui com dois intuitos principais: em primeiro lugar, localizar um dos primeiros debates, nas ciências sociais, sobre o conceito de "inteligência" que integra o uso de tecnologias informacionais na sua definição. Em segundo lugar, a intenção é indicar um paralelo entre o hype do mercado e perspectivas teóricas "apologéticas" no contexto de

popularização da Internet com a percepção pública da IA atualmente, conforme modelos de ML se complexificam.

A questão do hype é fundamental aqui, pois é indicativa do modo como a produção de conhecimento se dá na área de tecnologia, fora de círculos científicos, através de uma dinâmica de "tentativa e erro", em que a autenticidade de criações tecnológicas é validada a posteriori, de acordo com o seu sucesso comercial. De forma ilustrativa, Mancic relata que, após abertura do seu BBS ao público, "perdeu" seu computador, pois estava recebendo, em média, um novo usuário por dia, número consideravelmente alto para a época, o que o obrigou a tornar a utilização da máquina exclusiva para isso. Esse número cresceu lentamente até cerca da metade da década de 90, quando a liberação da Internet comercial e a possibilidade de acessá-la através do BBS impulsionaram o crescimento ágil e exponencial do número de usuários. Em relação a isso, a avaliação de Mandic é que ele não teve a ambição necessária para expandir sua empresa o quanto podia nesse preâmbulo da história da Internet, sendo o número potencial de usuários muito maior do que o estimado por ele e demais figuras relevantes do campo. Afinal, segundo o empresário, a sociedade inteira foi pega "desprevenida" pela Internet.

Nesse contexto, ele apontou que a Internet só funcionaria de acordo com o modelo imaginado por ele se atraísse milhares e milhares de usuários - ou seja, como mercado, só faria sentido se o uso fosse massificado - o que foi possível alcançar mediante a ampliação de capacidade de processamento e a facilidade de acesso resultante da criação do IG. A gratuidade proposta pelo projeto do IG, para Mandic, resultou no seguinte legado: "Ele [o IG] foi bom pro Brasil. Ele deu a possibilidade da população entrar na Internet. [...]. Uma população com informação é uma população que sabe mais e vale mais pro país. Então o país cresceu com o IG." (PRESCOTT, 2015b). De modo similar a Lévy, portanto, Mandic parece associar "saber" e "valor" no contexto do advento e distribuição da Internet como características de uma nova geração, cognitivamente amparada pelas novas ferramentas digitais. Previsivelmente, esse argumento recorre na cobertura de fenômenos digitais sempre que atraem suficientemente a atenção pública ao ponto de a necessidade de sua existência ser questionada, dentro da narrativa do progresso humano.

No caso de Lévy, por exemplo, interessa notar a contrariedade que identifica no seu projeto de uma inteligência coletiva, digitalmente articulada, em relação à ideia de uma "inteligência artificial", delineada para superar as capacidades cognitivas

humanas. Como ele colocaria anos depois da publicação de seu livro: “fazer as pessoas mais inteligentes *com* computadores, ao invés de tentar fazer computadores mais inteligentes que as pessoas” (PETERS, 2015, p. 261). A defesa de ferramentas digitais potencializadoras da capacidade humana (coletiva), contrasta com tecnologias pensadas para outros objetivos, que representariam uma ameaça à supremacia humana, numa dinâmica similar ao movimento histórico da IA. Desde o seu nascimento, em elites intelectuais, até a IA conforme a vemos hoje, a disciplina passou por uma comercialização e uma reconceituação associadas à segmentação da disciplina e à divisão entre “IA forte” e “IA fraca” (BRUNO, VAZ, 2002; ELISH, BOYS, 2017; BECHMANN, BOWKER, 2019). Com isso, a lógica de mercado incidiu sobre a reformulação do objetivo inicial da IA, a simulação da inteligência humana, no desenvolvimento de ferramentas para o aumento da engenhosidade humana.

Através dessa descrição, atentamos à forma de uma das dicotomias centrais que envolvem o campo das tecnologias digitais, a dicotomia humano-máquina. Ela indica um dos vários paralelos que é possível traçar entre as percepções desse contexto de ascensão da Internet e o atual momento de efervescência de perspectivas sobre a IA. Também como inspiração teórico-metodológica, no que tange a Internet, muita atenção foi - e ainda é - direcionada para as formas dos engajamentos práticos que a fazem existir. Desde os estudos de ciência e tecnologia e a antropologia da cibercultura, até as mais recentes áreas da antropologia digital e dos New Media Studies, problemas da ordem de dualismos que se sustentam e se renovam nas práticas - e nos discursos que as envolvem - em contextos digitalizados são objeto de interesse intelectual interdisciplinar nas ciências humanas (MARWICK, BOYD, 2010; HORST, MILLER, 2012; BERRY, 2011).

De partida, interessa destacar que as percepções, transversais ao mercado e à academia, sobre a Internet como um fenômeno de muitos significados, sobreposto de maneira inescrutável às próprias intenções humanas, ressoam nas interpretações que se fazem de IA atualmente. Esta pesquisa foi desenvolvida em interlocução com pessoas que, no seu compromisso com a aplicação técnica de tecnologias de IA, frequentemente buscaram modos de se posicionar em meio ao jogo de “otimismos” e “pessimismos”, ou de “ficção” e “realidade”, “hype” e “medo” (ELISH, BOYD, 2017) perante a inovação tecnológica que os cercava. Nesses movimentos, percebi uma preocupação com a caracterização da área na esfera pública - e com a diminuição da

obscuridade acerca do funcionamento e dos objetivos práticos de tecnologias cuja criação era parte do seu trabalho.

Quando optamos por abordar a IA como mercado, trata-se de abordar a contradição entre a prática e o discurso sobre a área no mainstream. Ou seja, considerando a característica comercial da maior parte da produção na IA, a questão é o modo como essa dinâmica hype x realidade é constitutiva do campo. A incompreensão sobre o escopo do trabalho dos cientistas e engenheiros de dados, arquitetos de software e outros profissionais ligados diretamente à criação de modelos, como os meus interlocutores, é um incômodo para os mesmos. Todavia, também faz parte da dinâmica empresarial de captação de investimentos. Ela perpassa a relação interna, com os *stakeholders*[4] dentro das empresas e a relação com investidores externos. Além disso, o discurso hiperbólico sobre as possibilidades de modelos funciona na criação de um horizonte para o desenvolvimento da IA que alimenta a repercussão da área no senso comum, na mídia e na ficção.

Assim, o que observei em campo foi a atuação de meus interlocutores no estabelecimento da IA na sua própria experiência e nas atuais capacidades técnicas de seus modelos. Visando a compreensão de públicos “leigos”, geralmente abordaram o conceito de “inteligência” e a distinção entre a inteligência dos humanos e a das máquinas. No ofício de antropóloga, o reconhecimento da eficácia dessas dicotomias no campo estudado exigiu considerável atenção e dedicação especial às reformulações pelas quais passam ao longo do tempo e à importância que outras divisões passam a adquirir.

Pensemos, por exemplo, na relevância da divisão virtual-real no contexto de ascensão da Internet e, posteriormente, das redes sociais. Naquele momento, diversas abordagens consideravam teoricamente oportuno visualizar as interações digitalizadas como um mundo à parte da “realidade”. Ao longo do tempo, como a capilarização de serviços digitais explicitou, esse binarismo provou-se pouco produtivo para se pensar as relações que estavam se estabelecendo. Assim, conforme defenderam Horst e Miller (2012), inspirados em grande parte por Latour, o

[4] Pode ser traduzido como "partes interessadas", apesar de na prática o termo ser utilizado em inglês. É um conceito amplo, que abrange indivíduos em posições de influência em diferentes setores. Nesta pesquisa, normalmente os profissionais no campo da IA usaram o termo em referência a indivíduos no topo das pirâmides empresariais. São pessoas que têm influência sobre os rumos que empresas tomam, sem estarem necessariamente envolvidos nos processos de criação de produtos de dentro delas.

reconhecimento da dimensão material do digital é fundamental para os estudos sociais, tanto no que tange a infraestrutura digital, como - e principalmente - em relação à noção de que a “ordem social” não se sobrepõe à “ordem material”. Assim, para os autores, é importante estender a investigação para além da intenção humana perante objetos tecnológicos, considerando também a relação desses objetos entre si.

No caso da presente pesquisa, essa consciência é central porque uma das primeiras características que se manifestou em campo foi justamente a proliferação de argumentos dicotômicos como parte das estratégias de desmantelamento da imagem hiperbólica, ou “fictícia”, da área e das dinâmicas de reivindicação de objetividade nos processos que a IA envolve. Em grande parte, essas dicotomias envolveram as distinções corpo-mente e natural-artificial na construção do argumento sobre a relevância dessas novas tecnologias, associadas ao refinamento de métodos de ML, para o progresso humano. Por enquanto, vamos nos ocupar entendendo por que a construção da objetividade é uma questão relevante na IA explorando as controvérsias que a definição da área envolve, introduzindo o campo e o modo como a dimensão de mercado da IA surgiu como relevante nesse processo.

2.2 O CAMPO

Como minha porta de entrada no campo, escolhi uma conferência com o foco em desenvolvimento de softwares - atualmente, a maior conferência nacional sobre o tema: trata-se da *The Developer’s [5]Conference* (TDC). Minha primeira participação na conferência aconteceu em novembro de 2019 (a única vez presencialmente), na cidade de Porto Alegre, em um dos campi de uma renomada instituição de ensino superior privada porto-alegrense. Esse primeiro contato me gerou uma impressão que os demais eventos no campo vieram a reforçar: a impressão de uma comunidade em abundância, efervescente e em plena ascensão. Naquele momento, evidentemente, essa impressão estava fundamentada na comparação intuitiva dessa com a minha experiência anterior em meio a congressos voltados às ciências humanas no Brasil,

[5] O termo "desenvolvedor" tradicionalmente se refere aos profissionais envolvidos no desenvolvimento de software. Entretanto, adotamos o uso do termo conforme observado em campo. Notamos que, na prática, foi utilizado como um termo guarda-chuva, que abrange o espectro das profissões na TI. Inclui, portanto, desenvolvedores front-end, back-end, full-stack, cientistas de dados, etc.

que remete à diferença entre plataformas de divulgação científica voltadas ao mercado e as voltadas à academia - ou à própria forma de conhecimento específica para esses objetivos diferentes, um dos assuntos recorrentes no campo.

Já no credenciamento, dois dos materiais incluídos no “kit de boas-vindas” contribuíram para o estranhamento mencionado e para a sinalização de padrões que viriam a se tornar relevantes na pesquisa. Em meio a materiais informativos sobre o próprio evento, havia também panfletos de recrutamento de grandes empresas. A divulgação de vagas e atividades de recrutamento era rotineira nos eventos, evidenciando um mercado com alto índice de vagas ociosas, na contramão de outros setores nesse mesmo período. O evento estava programado para cinco dias, mas eu havia me inscrito apenas para o primeiro, para quando estavam marcadas as palestras na área de IA (a “trilha” de IA). Adiantada para o início do evento, circulei pelo prédio onde as atividades aconteceriam. No hall, pequenos grupos aglomerados conversavam - um público de acordo com o esperado, para quem tinha o mínimo de familiaridade com a composição demográfica do mercado de TI: predominantemente jovem, branco e masculino (THOUGHTWORKS, 2020).

Entretanto, chamavam minha atenção os elementos que emolduravam esse público, nas periferias do salão: “stands” de patrocinadores, compostos por empresas de tecnologia e de mídia, mas também por bancos, representantes do ramo de alimentação, vestuário, entre outros. Neles, funcionários apresentavam a empresa em questão através de conversas informais, “gadgets”, camisetas, cafezinhos ou panfletos. Desviando da multidão e adentrando um pouco mais o primeiro piso do edifício, uma estrutura de paredes transparentes com algumas pessoas dentro despertou igualmente minha atenção: tratava-se de uma instalação para a produção de fotos para o perfil profissional dos participantes do evento no LinkedIn, contando com profissionais de maquiagem e fotografia para auxiliar na apresentação dessas pessoas para possíveis contratantes. A impressão transversal a essas observações que tive foi em relação ao empenho da comunidade técnica e das empresas em atrair novos profissionais para o setor de tecnologia.

Na época, eu não sabia, mas a escolha do evento como uma forma de acesso ao campo foi parecendo cada vez mais própria, pois sua estrutura em si foi pensada para facilitar o contato - o *networking*, no vocabulário nativo - entre pessoas que nutrem interesses nas mesmas subáreas da TI. A conferência é dividida em “trilhas” temáticas (Figura 1), que ocorrem, em grande parte, simultaneamente, e cada trilha

conta com uma programação que ocupa o dia todo com palestras que apresentam algo que os coordenadores da trilha julgaram ser de valor para o mercado em questão.

Figura 1 - Grade de trilhas do TDC (Porto Alegre, novembro de 2019)

Fonte: site TDC.

O acesso às trilhas de temáticas específicas - as Trilhas Premium - é pago em torno de R$100,00 por trilha. Mas também é oferecido acesso gratuito à Trilha Stadium, na qual são apresentadas as palestras das trilhas específicas que os organizadores julgam ser de interesse para o grande público, e às trilhas dos patrocinadores. As Trilhas Premium são separadas por temas e coordenadas por profissionais reconhecidos na área em questão. Nestas, ocorrem entre 7 a 14 palestras (resultantes de um processo de submissão aberto) selecionadas por esses coordenadores. As palestras são curtas, com duração de, no máximo, 50 minutos e têm a seguinte estrutura padrão. Primeiro, há a apresentação de uma solução para algum problema identificado no mercado em questão (frequentemente, no campo, um exemplo da aplicabilidade de alguma ferramenta nova ou antiga num caso prático). Então, um momento para perguntas da plateia e dos coordenadores e a divulgação

das empresas/perfis dos palestrantes. Impreterivelmente, o LinkedIn é a plataforma mais popular para este fim nos eventos, mas também GitHub ou Facebook.

De acordo com uma das idealizadoras do evento, Yara Senger, a proposta de um evento multi-trilhas é o que permite manter a conferência atualizada. O evento começou em 2007, de maneira fechada, com poucos palestrantes falando para grandes públicos e com o foco na comunidade Java, da qual ela fazia parte. Ela pontua que a descentralização do evento a partir de 2010 fez dele uma "plataforma de inovação aberta", que foi crescendo com a incorporação de diferentes comunidades:

> O TDC foi uma forma de ampliar a nossa amplitude [sic] em trinta, em cinquenta vezes, o que a gente tinha antes. Porque a minha área de especialidade é Java, eu conseguia fazer uma trilha incrível de Java, mas como é que a gente expande isso trinta vezes (cinquenta num TDC São Paulo)? Então, a gente só consegue expandir isso compartilhando com as outras pessoas que estão dentro daquelas áreas e estão vendo aquela novidade, estão vendo aquela dor; e mais do que isso: eu vi muitos eventos decaírem, acabarem, se extinguirem, por alguma razão. Então, como é que a gente mantém relevância? Como é que a gente continua sendo interessante? [...]. Então, a inovação aberta - que é o que eu chamo o TDC hoje, "uma plataforma de inovação aberta"- porque o caminho quem dá? É o Call for Trilhas. Qual trilha vai ter? Vai depender. Vai depender das pessoas, vai depender dessa nova comunidade que surge ao redor do TDC, que engloba e interfaceia [sic] com outras comunidades. Então não é que a comunidade está dentro do TDC, mas as pessoas das comunidades estão dentro do TDC. (KUBICAST…, 2021).

Trata-se de um evento elaborado por programadores para programadores, que declara ter como objetivo o "empoderamento de ecossistemas locais". Por isso, além dessa política de descentralização temática, o evento também se propõe a ser multirregional. Contudo, tende a se concentrar no Sul/Sudeste e, desde 2008, são tradicionais as edições em duas cidades: São Paulo e Florianópolis - mais recentemente, houve também edições em Porto Alegre, Belo Horizonte e Recife. Essa proposta, de acordo com a mesma entrevista da organizadora, também estava fundamentada no desejo que as pessoas residentes das cidades em que o evento fosse acontecer se engajassem nele, priorizando a exposição de novos profissionais e a criação de novas comunidades, e não o protagonismo de profissionais já renomados no line-up.

O sucesso desse formato, conforme descrito por Senger, é visualizado na sua dimensão atual: em 2020, foram mais de 1500 palestrantes no todo do evento. Os

palestrantes não recebem remuneração pelas suas participações, mas ganham o acesso gratuito a todas as trilhas - de acordo com a mesma, 90% das inscrições no evento tratam-se de concessões a palestrantes, empresas e grupos de usuários. É esse dado que ela aponta como uma das evidências de que palestra no evento “quem quer estar lá”. Trazendo também alguns outros dados referentes à população que participa do TDC como embasamento para a cobrança para o acesso às Trilha Premium (no contexto da comparação com outros eventos da área, que são gratuitos, feita pelo entrevistador) como: 48% do público tem mais de 10 anos de experiência em TI e 20% mais de 5 anos. Ou seja, a maioria dos participantes tem uma carreira relativamente estável (KUBICAST…, 2021).

Nesse sentido, a caracterização oferecida pela organizadora sobre o conceito de “inovação aberta” descreve um evento estruturalmente pensado para se atualizar organicamente. Ao todo, participei de cinco edições do TDC - três dessas com acesso à trilha de Inteligência Artificial. A partir da segunda metade de 2020, essa trilha passou a ser aglutinada à de ML na maioria das edições, indicando a aproximação das áreas, considerando esse viés de organicidade manifestado pela organizadora. Entendemos a associação do conceito de Inteligência Artificial - protagonista de uma longa história de debates sobre a sua definição - com o conceito menos controverso de ML, subárea voltada a técnicas de classificação e predição de dados de forma automatizada, numa abordagem análoga à de Crawford (2021):

> “Machine learning” is more commonly used in the technical literature. Yet the nomenclature of AI is often embraced during funding application season, when venture capitalists come bearing checkbooks, or when researchers are seeking press attention for a new scientific result. As a result, the term is both used and rejected in ways that keep its meaning in flux. For my purposes, I use AI to talk about the massive industrial formation that includes politics, labor, culture, and capital. When I refer to Machine Learning, I’m speaking of a range of technical approaches (which are, in fact, social and infrastructural as well, although rarely spoken about as such). (CRAWFORD, 2021, p.9).[6]

[6] “‘Machine learning’ é normalmente usando na literatura técnica. No entanto, a nomenclatura IA é frequentemente adotada durante a temporada de pedidos de financiamento, quando investidores em capital de risco trazem talões de cheques ou quando pesquisadores buscam a atenção da imprensa para um novo achado científico. Como resultado, o termo é usado e rejeitado de maneiras que mantêm seu significado em fluxo. Para meus propósitos, uso IA para falar da formação industrial massiva que inclui política, trabalho, cultura e capital. Quando me refiro ao Machine Learning, estou falando de uma série de abordagens técnicas (que são, na verdade, sociais e infraestruturais também, embora raramente tratadas como tal” (CRAWFORD, 2021, p. 9, tradução minha).

Assim, “machine learning” é um conceito mais comumente associado à técnica, que foi trazido recorrentemente em campo para fundamentar na prática as expectativas relacionadas ao hype comercial da área e as especulações sobre a singularidade tecnológica [7]. Além disso, o termo se refere a um conjunto de técnicas indissociáveis do significativo crescimento recente da IA por permitirem a análise de grandes volumes de dados. A comunidade de profissionais da qual participam meus interlocutores, portanto, se insere num contexto de crescimento acelerado de investimentos no campo, em razão do desenvolvimento de técnicas e a formação recente de especialistas em ML. De acordo com o survey realizado pelo Kaggle em 2020, 55% do público declarou trabalhar com ML há menos de três anos. O avanço nas técnicas de ML - principalmente as de Deep Learning (DL) - também se destacou neste período pelas suas aplicações nas áreas de saúde e biologia: o uso desses modelos auxiliou no monitoramento da progressão dos casos de COVID-19 (SUJATH; CHATTERJEE; HASSANIEN, 2020), na detecção de lesões relacionadas à doença e, ainda, acelerou a invenção e testagem de novos medicamentos eficazes contra a mesma (ZHANG et al., 2021).

Considerando a posição do Brasil nesse cenário: dos 2675 profissionais então empregados como cientistas de dados - função que predomina no campo de ML - que responderam o survey do Kaggle, 4,5% eram brasileiros, colocando o país em terceiro lugar quanto ao número de participantes, atrás apenas da Índia (21.8%) e dos Estados Unidos (14.5%). Ainda, de acordo com dados do AI Index Report 2021, entre 2016 e 2020, o Brasil foi o país que mais contratou profissionais na área de IA, com uma distância significativa do segundo colocado, a Índia (ZHANG et al., 2021). Nesse sentido, o país demonstra seguir uma tendência que é refletida ao redor do mundo: o mercado de tecnologia tem contornado os contratempos impostos pela pandemia, sofrendo pouco os impactos econômicos percebidos noutros setores e mantendo previsões de crescimento otimistas para os próximos anos. Já na segunda metade de 2020, 42% das empresas de TI apontavam previsão de orçamento maior para tecnologia no ano seguinte. Esses dados nos remetem à impressão inicial mencionada, sobre a comunidade efervescente de TI, além de fundamentarem a ideia

[7] O termo se refere à hipótese sobre o momento em que o desenvolvimento tecnológico será irreversível e a inteligência maquínica superará a humana. Normalmente, as discussões sobre circulam a noção da ameaça das máquinas para a existência humana.

de que as demandas éticas e iniciativas regulatórias sobre a prática da IA não demonstram grande impacto nesse mercado.

A agitação econômica na IA foi evidenciada em campo de diversas formas: através de panfletos e posts de recrutamento profissional, recados ao final de palestras divulgando vagas nas empresas dos palestrantes, convites para inserção em determinada área com déficit de especialistas e, justamente, no background profissional diverso de desenvolvedores que se apresentavam nos eventos - muitos haviam transicionado recentemente para a área. O TDC não foi o único ambiente em que foi possível visualizar esse movimento durante a pesquisa de campo. Acompanhei também discussões em grupos de WhatsApp voltados à ciência de dados e IA (todos direcionados especificamente para o público feminino), que encontrei através de palestrantes e coordenadores nas trilhas do evento. Participei também de três cursos de introdução à programação - um voltado especificamente para IA e ML -, elaborados por organizações comandadas por mulheres com o objetivo de diminuir o *gap* de gênero no mercado de tecnologia.

Tendo em vista a aparente efervescência econômica mencionada, a fim de prosseguir com a discussão, é necessário especificar o caráter dos interlocutores desta pesquisa como profissionais da indústria de tecnologia. Daston (2020) descreve, sobre o trabalho na IA:

> The people behind the curtain in modern AI projects are of two sorts: those who are thinking about how to divide a very complicated task into the tiniest possible steps, very much in the tradition of the history of mechanical calculation; and those whose work is compensatory for algorithmic systems — Facebook moderators, for example, who monitor objectionable content missed by the algorithms meant to eliminate it automatically. (DASTON, 2020, n.p).[8]

Além disso, entendemos que há o trabalho dos “ghost workers” (GRAY, SURI, 2019), que guarda similaridades com o segundo tipo de profissional que a autora cita, porém se refere a uma categoria de trabalho informal, realizado por trabalhadores de baixa renda pagos por tarefa. O termo identifica pessoas que se ocupam, principalmente, da rotulagem de dados a fim de garantir que modelos

[8] “As pessoas por trás da cortina nos projetos modernos de IA são de dois tipos: aquelas que estão pensando em como dividir uma tarefa muito complexa nos menores passos possíveis, como na tradição da história do cálculo mecânico; e aquelas cujo trabalho é compensatório para os sistemas algorítmicos – moderadores do Facebook, por exemplo, que monitoram conteúdos censuráveis que escapam aos algoritmos elaborados para eliminá-los automaticamente” (DASTON, 2020, n.p, tradução minha).

mantenham e/ou aumentem a sua acurácia e também sinalizem conteúdos impróprios em plataformas. A maior parte do “ghost work” é realizada em países do Sul Global. Esse trabalho de classificação e “compensação” é característico da história da automação, bem como a tendência a obscurecer a extensão de mão de obra humana necessária para desenvolver dispositivos autônomos (CRAWFORD, 2021; DASTON, 2020; GRAY, SURI, 2019). Posto isso, esta pesquisa aconteceu em interlocução com profissionais que se encaixam na primeira categoria de trabalho apontada por Daston (2020), dedicados predominantemente aos níveis mais elementares do desenvolvimento de modelos de IA. O trabalho deles contrasta radicalmente com o caráter do trabalho do “ghost work”, além de terem consideravelmente maior escolaridade e remuneração (KAGGLE, 2020; CRAWFORD, 2021).

Ou seja, o foco dessa pesquisa é a prática no *mainstream* da IA, com pessoas engajadas no desenvolvimento técnico e na pesquisa na área - ainda que predominantemente no âmbito da indústria. Entretanto, recentemente autores têm buscado enfatizar a materialidade da IA em suas investigações, tanto no âmbito das relações de poder que atravessam a exploração do trabalho cognitivo e físico de milhares de pessoas, como nas práticas de extrativismo. Todos esses fatores, ocultos nos discursos sobre a sua imaterialidade e eficiência, são essenciais no funcionamento das aplicações de IA (CRAWFORD, 2021). Essa contextualização é fundamental para se pensar os problemas relacionados à prática da IA e o impacto de suas aplicações que são enfatizados nos discursos de profissionais da área. Nesse sentido, a ausência desses aspectos nas discussões sobre ética em campo nos ajuda a caracterizá-la de acordo com o tipo de trabalho que os interlocutores dessa pesquisa desenvolvem.

Em campo, o mercado surgiu como a força regulatória determinante, dado o caráter comercial da IA hoje. Como discutiremos posteriormente, interesses de mercado são predominantes inclusive no impulsionamento de determinadas pautas éticas relacionadas à área e no impedimento de outras. Nesse sentido, ao mesmo tempo em que uma pessoa me relatava o desinteresse das empresas em questões éticas por entender que só se tornariam importantes se representassem risco financeiros a elas, outras apontavam que a explicabilidade dos modelos estava se tornando uma cobrança do mercado. O que atravessa essas impressões é o entendimento de que, atualmente, assim como a definição de parâmetros éticos, a prática na IA em geral é conduzida pela indústria de tecnologia. Nesse sentido, temas

como a automação e a substituição de mão de obra humana adquiriram papel secundário e foram comumente associados ao futuro não imediato da IA, ou à IA simbólica. Outros, como a cadeia de trabalho e o papel de “ghost workers” na indústria raramente surgiram.

Posto isso, as descrições subsequentes são informadas por atividades de interlocução com esses profissionais, em diferentes formatos de participação (mas majoritariamente online), com as quais me engajei. Isso também inclui a pesquisa bibliográfica sobre os tópicos de interesse, abrangendo debates acadêmicos na antropologia e na computação, buscando referências também trazidas pelos especialistas da IA, bem como dados estatísticos e jornalísticos que auxiliam na apresentação do conhecimento baseado em campo. Na próxima seção, faremos uma breve contextualização histórica da criação da Inteligência Artificial como uma disciplina, seguindo o debate sobre a definição formal do conceito. Para isso, vamos percorrer a divisão entre dois tipos de IA como parte da diferenciação entre um formato de conhecimento acadêmico e outro voltado ao mercado, conforme surgiu em campo.

2.3 DEFININDO IA

Em maio de 2020, participei de um evento que descobri em um dos grupos sobre TI que acompanhava no Telegram, realizado por uma proeminente startup voltada à inclusão social de mulheres no mercado profissional de tecnologia. O evento foi organizado como um Sprint, formato popular na indústria tech, como parte das “metodologias ágeis”, que trata da definição de atividades para alcançar um objetivo muito específico num curto período de tempo. Ele buscava introduzir, de maneira gratuita e inteiramente online, o tema da IA para mulheres desenvolvedoras durante dez dias. Ao longo de cinco módulos, percorremos debates acerca dos objetivos, aplicações e impactos de aplicações de IA, assim como aspectos técnicos do desenvolvimento de algoritmos de ML. Neste período, a minha atenção foi envolvida pelo encapsulamento de numerosas controvérsias que pareciam animar o campo então e que prosseguiram comigo ao longo da pesquisa.

Conforme mencionado anteriormente, uma dessas controvérsias concerne a própria definição do termo. Num dos materiais introdutórios do Sprint, esse problema

foi sinalizado com a afirmação de que o significado de “inteligente” é motivo de disputa, que falta consenso científico sobre o conceito, e isso prejudica a compreensão “de fora” sobre a IA. De fato, a dificuldade em apontar o que define “inteligência” mostrou-se um aspecto fundamental da prática no campo, estimulando debates que pareciam ofuscar o segundo termo que dá nome à área: a definição de “artificial”, como antônimo de “natural”. Por tratar da simulação de comportamentos observados na natureza pelas mãos do homem, este não surgiu como tópico a ser contestado.

“Inteligência”, porém, foi um termo constantemente retomado nos eventos que presenciei - principalmente no âmbito das mesas do TDC - como um conceito indefinido. Em determinado momento, durante um painel do congresso, um cientista de dados, que era também neurocientista, colocou em pauta como o background em ambas as áreas não ajudava a solucionar a questão. Para ele, só sabemos que inteligência existe porque é algo que se manifesta no comportamento de seres humanos e outros animais. Nesse sentido, o rapaz exprimia uma ideia que marcou historicamente os desenvolvimentos do campo: qual seria a relação entre a mente e o comportamento exprimido pelos seres vivos? Essa questão, inicialmente, foi abraçada pela cibernética de Norbert Wiener (1943), que permitiu idealizar uma prática científica interdisciplinar, um tipo de explicação aplicável tanto a máquinas como a seres vivos.

Entretanto, como descrevem Erickson et al. [9](2013), em colaboração com outros autores, áreas que eram unificadas pela cibernética prosseguiram como disciplinas próprias no processo histórico de redefinição da racionalidade. Essa redefinição ocupou pesquisadores, principalmente no âmbito das ciências sociais e políticas, da economia e da psicologia, no período do final da Segunda Guerra até o início dos anos 80. Da fragmentação da cibernética no contexto dos Estados Unidos, no final dos anos 50, surgiram áreas mais voltadas a organismos vivos e outras mais ocupadas com as máquinas, como as que consolidaram a computação, a engenharia de controle, a teoria da informação e, principalmente, a IA ([10]ERICKSON et al., 2013, CESARINO, 2021).

[9] : Tendo em vista a centralidade da noção de objetividade elaborada por Daston e Galison (2007), vale enfatizar que Daston faz parte da autoria do livro em questão.

[10] A fim de situar essa fragmentação nos Estados Unidos, os autores descrevem como, simultaneanemente, as ideias da cibernética começaram a ressoar na União Soviética, o que apontam como parte da explicação para a pouca popularidade de noções de racionalidade nas ciências humanas sociéticas (ERICKSON et al., 2013, p. 19)

O documento que cunhou o termo “inteligência artificial”, comumente referido como marco inicial do campo, consiste numa carta redigida em 1955, com a proposta de um workshop de verão, no âmbito da Universidade de Dartmouth, organizado por John McCarthy:

> We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer (MCCARTHY, 1955, p. 2).[11]

Nesse mesmo documento, encontramos uma definição do problema fundamental da IA, também citada em um dos textos introdutórios do Sprint mencionado: “[...] making a machine behave in ways that would be called intelligent if a human were so behaving” (MCCARTHY, 1955, p. 11). O problema sobre o acesso à relação entre pensamento e comportamento foi o que movimentou a evolução da IA ao longo das suas primeiras décadas, fazendo com que as discussões na área constantemente se alinhassem aos debates sobre “consciência”. Apesar de, atualmente, os especialistas se distanciarem de discussões nesse sentido, veremos que as técnicas que caracterizam a IA fraca também são informadas por concepções sobre o funcionamento da mente humana, especificamente as advindas da neurociência.

Na breve descrição de McCarthy (1955), o documento alinha-se à concepção behaviorista de inteligência presente no artigo seminal de Turing (1950), sobre a capacidade de “pensar” em máquinas. No propósito do projeto - a simulação do comportamento inteligente - fica implícita a premissa de que a descrição

[11] “Propomos que um estudo de inteligência artificial seja realizado durante o verão de 1956 na Faculdade de Dartmouth, em Hanover, New Hampshire. O estudo deve prosseguir com base na conjectura de que todo aspecto do aprendizado ou qualquer outra característica da inteligência pode, em princípio, ser descrito com tanta precisão que uma máquina pode ser feita para simulá-lo. Será feita uma tentativa de descobrir como fazer com que as máquinas usem a linguagem, formem abstrações e conceitos, resolvam tipos de problemas agora reservados aos humanos e se aperfeiçoem. Achamos que um avanço significativo pode ser feito em um ou mais desses problemas se um grupo cuidadosamente selecionado de cientistas trabalhar juntos por um verão” (MCCARTHY, 1955, p. 2, tradução minha).

suficientemente refinada do funcionamento da mente humana seria o suficiente para que a máquina pudesse “simular” a inteligência dos humanos. Nesse sentido, os desenvolvimentos do campo eram influenciados pelo entendimento que a mente existe através de um sistema simbólico de representações (BRUNO, VAZ, 2002). Essa descrição caracteriza o que conhecemos como “IA Simbólica”, uma categoria que engloba os objetivos e a filosofia da área desde o seu surgimento até meados da década de 80[12]. Helmreich (1998) localiza esse momento na história da IA, em contraste com os movimentos posteriores:

> From the 1950s until the 1980s, the central problem in Al had been to find the best way formally to represent and manipulate knowledge. It had become axiomatic that intelligence could be characterized as the rational manipulation of symbols that represented aspects of the world. By the mid- and late-1980s, however, the rational and formalistic vision of intelligence came under attack by those in the Al community who felt that it was an unrealistic image of cognition, and an ineffective model for engineering more flexible and 'intelligent' machines. Some, in a hermeneutic turn of thought, suggested that intelligence grew out of a continual reinterpretation and re-encountering of an ever-changing world. Others maintained that if computer scientists wanted to manufacture intelligent behaviour, they needed first to mimic the life processes that support intelligence. In a move that recalled practices of interdisciplinary borrowing in early cybernetics, computer scientists plundered biology for new ideas and analogies. Researchers building neural nets were inspired by the neuronal architecture of the brain. And GAs [algoritmos genéticos] were elaborated after the grandest natural process of all: evolution. (HELMREICH, 1998, p. 41-42).[13]

[12] Katz (2020, p. 27) aborda de maneira cética essa definição, assim como outras que exploraremos no modo como foram mobilizadas em campo a fim de distinguir entre dois tipos de IA. De acordo com o autor, tratam-se de divisões que esses profissionais dentro da área estabelecem, obscurecendo a existência de sistemas que não correspondem às categorias nos períodos associados a elas. Apesar disso, como o próprio autor indica, entendemos que tratam-se de definições importantes para abordar as narrativas que circulam em meio aos especialistas no campo.

[13] “Da década de 1950 até a década de 1980, o problema central em IA foi encontrar a melhor maneira formal de representar e manipular o conhecimento. Tornou-se axiomático que a inteligência pudesse ser caracterizada como a manipulação racional de símbolos que representavam aspectos do mundo. Em meados e final da década de 1980, no entanto, a visão racional e formalista da inteligência foi atacada por membros da comunidade Al que achavam que essa era uma imagem irreal da cognição e um modelo ineficaz para construir máquinas mais flexíveis e "inteligentes". Alguns, em uma virada hermenêutica de pensamento, sugeriram que a inteligência surgiu de uma contínua reinterpretação e reencontro de um mundo em constante mudança. Outros sustentavam que, se os cientistas da computação quisessem fabricar um comportamento inteligente, eles precisavam primeiro imitar os processos vitais que sustentam a inteligência. Em um movimento que lembrou práticas de empréstimo interdisciplinar na cibernética inicial, os cientistas da computação recorreram à biologia em busca de novas ideias e analogias. Pesquisadores que construíram redes neurais foram inspirados pela arquitetura neuronal do cérebro. E os AGs [algoritmos genéticos] foram elaborados conforme o maior processo natural de todos: a evolução.” (HELMREICH, 1998, p. 41-42. Tradução minha.)

Nessas primeiras décadas, a questão da manipulação de conhecimento era explorada, principalmente em círculos acadêmicos[14], e a IA era largamente entendida como uma filosofia. A inteligência como a "manipulação racional de símbolos" equivalentes a fatos do mundo era uma noção que guiava as discussões nesse contexto real, indicando a influência da teoria computacional da mente no campo. Trata-se da teoria construída através da analogia cérebro-computador, que concebe ambos como processadores de informação, responsáveis pela manipulação formal de símbolos. Nesse sentido, a analogia sustenta a ideia de que tanto o cérebro como a máquina "computam" através de algoritmos. Aqui, chegamos num ponto que considero central para acessar as discussões acerca da definição de IA, conforme o que experienciei em campo: as confluências entre as controvérsias concernentes a essa questão, especificamente aquelas direcionadas à noção de algoritmo. Portanto, antes de focar na mudança de paradigma que ocorreu na IA, interessa focar nessa aproximação e em algumas direções teóricas possíveis para lidar com os problemas que rodeiam os conceitos.

"Desde o começo do meu ensino técnico/médio eu já gostava de inteligência artificial, na época não chamávamos assim, pra mim era tudo algoritmos", me relatou um interlocutor durante entrevista quando lhe perguntei sobre o surgimento do interesse por IA. O comentário permaneceu comigo, pois suspeitei indicar uma separação entre a técnica e a teoria no campo, ou seja, que a habilidade de programar algoritmos de IA difere do conhecimento sobre aspectos mais ligados a discussões científicas, filosóficas e até políticas sobre a área. Entretanto, essa percepção não se sustentou: nos dez últimos anos, numa explícita relação com a massificação do uso de redes sociais, o conceito de "algoritmo" saiu da literatura técnica e tornou-se o epicentro de discussões relacionadas à privacidade, liberdade e justiça no contexto da digitalização. Exemplos notáveis incluem a repercussão das críticas sobre os "filtros-bolha", relacionadas aos efeitos de algoritmos de filtragem de conteúdo, direcionadas principalmente à Google e ao Facebook (PARISER, 2011), assim como as voltadas à questão do "viés" algorítmico e práticas discriminatórias (NOBLE, 2018). O escândalo da Cambridge Analytica relacionado à coleta de informações pessoais

[14] Apesar da associação das primeiras décadas da IA às universidades, Katz (2020) ressalta que, de partida, a disciplina surgiu com o apoio de organizações militares como o Pentágono. Erickson et al (2013) afirmam, similarmente, que a racionalidade algorítmica característica do período da Guerra Fria emanou de projetos estimulados através de grandes investimentos militares nos EUA.

para manipulação da opinião política também representou um marco na discussão sobre o uso irrestrito de dados de usuários de plataformas (MOROZOV, 2018).

Evidentemente, os problemas levantados englobam o campo da IA, pois geralmente referem-se a algoritmos de ranking de informações, que funcionam através dos dados sobre ações prévias de usuários nas plataformas, com a aplicação de técnicas de ML; ou então, algoritmos de reconhecimento facial, que também utilizam essas técnicas. Desse modo, tornou-se impossível desassociar a atuação no âmbito da IA da repercussão sobre seus impactos. Em campo, esse foi um ponto óbvio: todos meus interlocutores demonstraram estar cientes das problemáticas éticas e políticas (pelo menos aquelas que já atraíam atenção pública) relacionadas a aplicações de IA e das cobranças relacionadas à conduta dos profissionais envolvidos no seu desenvolvimento, apesar do grau de engajamento com as questões variar consideravelmente. Ainda assim, ao atentar para a dinâmica dessas controvérsias, observei que frequentemente eram contornadas com a retomada de definições formais dos conceitos relevantes para o cotidiano de meus interlocutores, principalmente a ênfase no apelo técnico da prática no campo da IA e da programação no geral. A formalidade e a ênfase na técnica foram trazidas como recursos para justificar a objetividade - a não interferência da subjetividade (DASTON, GALISON, 2007) das práticas na IA.

Esse apelo à técnica como uma forma de alegação de objetividade incorpora uma racionalidade pautada em regras que Erickson et al. (2013) denominam “algorítmicas”. Com isso, se referem a um processo histórico de mecanização das regras que é idealizado no algoritmo e na abordagem analítica do conhecimento que o conceito incorpora. De acordo com eles, o algoritmo é associado ao entendimento de regras como mecânicas e não submetidas ao julgamento humano. A racionalidade algorítmica, assim, é tomada como ideal em economias neoliberais modernas, propagando a noção de que o caráter mecânico do julgamento máquina determina a sua objetividade e superioridade em relação ao julgamento humano. Assim, a minha suspeita sobre o comentário citado anteriormente foi produtiva. Apesar de a experiência etnográfica identificar uma percepção das problemáticas sociais envolvidas nas aplicações de IA em meio aos seus especialistas, também o aspecto “objetivo” de conceitos empregados - como se fossem, de partida, descolados de contextos sociais específicos - foi algo ressaltado por eles perante eles.

Para prosseguirmos, portanto, é interessante exprimir a diferença entre “algoritmo” e “IA” através das definições “didáticas” que encontrei em campo. Um algoritmo, conforme descrito no Sprint mencionado e, similarmente, por diversas outras pessoas no período da pesquisa, trata-se de uma série de passos sequenciais para alcançar um objetivo predefinido. Aí reside o caráter do privilégio da análise em relação à síntese representado na resolução de problemas algorítmica (ERICKSON et al., 2013). Nesse sentido, o algoritmo não se trata sequer de uma ferramenta essencialmente computacional: o exemplo da receita de bolo foi recorrente na comprovação do argumento e, como mencionado, sinaliza para uma teoria sobre o pensamento e sobre a aprendizagem. O exemplo da receita é mobilizado na obra de Ingold (2010), por exemplo, para criticar a abordagem representacional da mente, argumentando sobre a relação entre percepção e ação como o cerne de aprendizagem. Ele afirma, portanto, que a simples leitura de instruções não garante sucesso na execução, pois aprendemos através da “sensibilização de todo o sistema perceptivo” (INGOLD, 2010, p. 21).

Menos especificidade surgiu na definição de IA, que apresentou uma variação não observada em relação ao conceito de “algoritmo”. A definição que encontrei no Sprint foi: “[...] Inteligência Artificial é o ramo da ciência da computação que estuda como construir máquinas que executem tarefas de forma inteligente”. Entretanto, empresto a definição menos redundante da apresentação de um interlocutor, que parece nos oferecer um caminho melhor para a compreensão do entendimento sobre “inteligência” que se mantém atualmente, de acordo com o movimento histórico da área: a IA é uma “ferramenta para encontrar padrões”. Como se pode perceber, existe uma redundância que conecta as definições sobre IA e nenhuma das que encontrei em campo oferece muita clareza sobre os objetivos da área. Porém, essa última definição nos interessa porque a ênfase sobre a busca de “padrões” foi comumente associada ao protagonismo do ML no desenvolvimento recente da IA. Nesse contexto, conforme discutiremos posteriormente, “padrões” remetem a correlações estatísticas (CHUN, 2021).

Katz (2020) aborda essa indefinição da IA como um aspecto intencional. No seu argumento, a IA é uma tecnologia da branquitude, em ambos os sentidos, como ferramenta da supremacia branca e como reflexo de uma ideologia da branquitude. Nesse sentido, aponta que a disciplina se compromete, em grande medida, com uma visão neoliberal da sociedade e na construção de um entendimento universal do self.

O autor entende que a “nebulosidade” que cerca os propósitos da disciplina é vantajosa, pois permite que sejam ajustados a serviço de grupos de interesse em diferentes condições sociais (KATZ, 2020, p. 10). Ou seja, a IA busca consolidar afirmações “universais” sobre a natureza humana, apesar de produzir modelos de self firmados em noções de raça, gênero e classe (KATZ, 2021, p. 181). Para o autor, a universalidade da inteligência, como se pudesse existir fora de contextos sociais, é uma das “falsificações epistêmicas”[15]da IA. Posto isso, a atenção etnográfica tem uma posição privilegiada na contextualização de definições associadas às práticas dos especialistas da IA.

Assim, nos interessa perceber como a forma dos debates que concernem os conceitos de algoritmo e IA se assemelham. Apesar de ter uma definição formal bastante específica, o conceito de “algoritmo”, assim como “Inteligência Artificial”, não indica consenso científico sobre uma definição que englobe a sua existência material e ideal, permanecendo obscuro. Por esse motivo, Ziewitz (2015) descreve o debate que tem cercado o “algoritmo” como um “drama algorítmico”, que seria dividido em dois atos. No primeiro, os algoritmos são introduzidos como atores poderosos em domínios diversos e os debates voltam-se à sua agência e impacto. Já no segundo ato, as indagações assumem a preocupação quanto às dificuldades colocadas para a investigação sobre como os algoritmos supostamente exercem o poder e a influência que estariam imbuídos neles. A sugestão que Ziewitz (2015) oferece para lidar com a aparente ambiguidade do algoritmo é tratá-lo como um “sensitizing concept”, ou seja, extrapolar o âmbito de investigação “essência-consequência” e voltar a nossa atenção para os contextos em que a figura do algoritmo surge e os fatores a tornam relevante.

Assim, o autor nos ajuda a pensar estratégias para situar os conceitos ao argumentar que esse drama algorítmico - avivado por questões de agência, inescrutabilidade e normatividade - remete a mitologias há muito nutridas no pensamento ocidental acerca de agentes poderosos invisíveis, incluindo a “mão invisível” de Adam Smith e a “seleção natural” de Charles Darwin. Isso se demonstra também na IA, como veremos, e se evidenciou recentemente com o advento de técnicas de Deep Learning e demandas sobre a transparência no funcionamento desses algoritmos. A recomendação de Ziewitz (2015), desse modo, nos ajuda a

[15] Do original "epistemic forgeries". Além da citada, Katz (2020, p. 95) associa a IA com outras duas falsificações epistêmicas: que sistemas de IA são equivalentes ou superiores às capacidades deo pensamento humano. E que esses sistemas alcançam o conhecimento, ou a verdade, por si mesmos.

retomar a discussão com um olhar antropológico que transcende a chave explicativa do “impacto” da tecnologia no mundo. Retomamos, como Seaver (2018) defende, as lições da história da disciplina através da descrição de sistemas sociotécnicos e de teorias que abdicam da posição de defesa do humano versus objetos tecnológicos. Desse modo, é possível contribuir para a desestabilização de noções que cercam esses objetos e explorando certos padrões de comportamento através das exposição das incoerências de definições aparentemente sólidas.

Seaver (2018) elabora um argumento sobre uma antropologia dos algoritmos que busca destoar do “humanismo analógico” que domina as conversas sobre o tópico dentro da própria disciplina. Ele identifica, vale frisar, uma equiparação de sentido entre o conceito de algoritmo e outros dentro do nicho de tecnologias responsivas complexas, como IA e ML, na maioria das discussões sobre. Para o autor, muitas das descrições reproduzem a dicotomia humano-máquina que a antropologia se propõe a reavaliar. Ele defende que, ao considerar os algoritmos como actantes em redes “mais-que-humanas”, é possível visualizar as pessoas *dentro* dos sistemas algorítmicos. Dessa forma, se faz claro como o que interpretamos enquanto uma “decisão algorítmica” consiste, na verdade, em uma série de escolhas humanas a partir da interação com a máquina, que na comunidade técnica são tratadas como periféricas (SEAVER, 2018; KATZ, 2020). Então, Seaver atenta para como, mesmo as inovações tecnológicas proponentes da automação, na forma dos modelos de ML, aprendem através da acumulação de feedback loops que sofrem a influência de motivações humanas de diferentes naturezas. Assim, o autor argumenta que cabe à antropologia dos algoritmos não conformar-se à “vaga analógica” num projeto de defesa do humano em oposição ao algorítmico, mas observar a forma como essa oposição é construída e mantida na prática.

É através dessas recomendações sobre como situar conceitos, sobre as mitologias reavivadas na esfera pública mediante o sucesso de novas tecnologias e a impossibilidade de separar as esferas “humano” e “máquina” no que concerne tecnologias que são, justamente, pautadas na exclusão do humano dos seus ciclos de atuação, que buscamos abordar o tema desta dissertação. Como a pesquisa não tratou especificamente da criação de aplicações de IA, mas sim de debates e da circulação de conhecimentos dentro da comunidade técnica, o trabalho se dedica, em grande parte, às definições e divisões observadas em campo. Nesse sentido, o constante estabelecimento de fronteiras entre a IA que se praticava antigamente e a

IA contemporânea observado em campo indica duas formas que a separação humano-máquina tomou: uma, sobre a máquina como ameaça na forma da IA forte e outra, sobre a máquina como suporte aos humanos na forma da IA fraca. A seguir, exploraremos a separação entre esses dois momentos da IA, através do modo como se manifestaram os contrastes entre os seus objetivos iniciais e os atuais dentro dos debates no campo.

2.4 DOIS TIPOS DE IA

Conforme mencionamos, os primeiros desenvolvimentos em IA inserem-se dentro do paradigma da IA "simbólica". Este se baseia, em larga medida, na exploração da hipótese do desenvolvimento de programas com habilidades cognitivas equivalentes às humanas e por um entendimento da cognição como a manipulação racional de símbolos. Com a evolução das pesquisas, a magnitude dos limites para alcançar esse objetivo passou a ser compreendida, mas, conforme foi possível perceber em campo, parte da comunidade de IA identifica que a repercussão na opinião pública ainda se baseia muito nas características da IA simbólica. Em alguns momentos, a circulação dessa noção de IA no senso comum foi apontada como um dos motivos para o descompasso entre as expectativas sobre e o que de fato fazem aqueles que trabalham na área.

Nesse sentido, quando a questão da substituição de humanos por máquinas foi levantada, normalmente foi associada à IA simbólica, assim como a tópicos de ficção científica. A questão foi bem ilustrada no comentário de um desenvolvedor, durante uma discussão em uma trilha de IA do TDC. Ele falou sobre como, enquanto se preocupa com tarefas aparentemente simples (mas computacionalmente muito complexas) como fazer um computador identificar visualmente o que é uma cadeira[16], na mídia e na ficção, há destaque para temas como a singularidade e futuros

[16] Ainda que o desenvolvedor não tenha demonstrado fazer referência a esse debate, a afirmação remere à popular crítica da IA elaborada por Hubert L. Dreyfus (1981, p. 163): "What makes an object a chair is its function, and what makes possible its role as equipment for sitting is its place in a total practical context. This presupposes certain facts about human beings (fatigue, the way the body bends), and a network of other culturally determined equipment (tables, floors, lamps), and skills (eating, writing, going to conferences, giving lectures, etc). [Can there be context-free features of chairs?] They certainly cannot be legs, back, seat, etc., since these are not context-free characteristics defined apart from chairs which then "cluster" in a chair representation, but rather legs, back, etc. come in all shapes and variety and can only be recognized as aspects of already recognized chairs."

distópicos relacionados à tecnologia. Isso impactaria a visão de quem não tem conhecimento na área. Com o exemplo, ele quis demonstrar como o tipo de IA que predomina atualmente é voltado a tarefas muito específicas e estamos muito longe desses cenários, pois o foco é o desenvolvimento de tecnologias altamente segmentadas. Essa segmentação é indissociável do refinamento de técnicas de ML e a atenção do mercado que atraíram, principalmente pela eficiência dos modelos em identificar e prever tendências de consumo.

Assim, o reconhecimento dessas mudanças foi algo que observei repetidas vezes, através das menções aos “dois tipos” de IA: “IA simbólica” e “IA conexionista”, “IA forte” e “IA fraca”, “IA geral” e “IA segmentada”, “IA ampla” e “IA estreita”, entre outros. Essa divisão era frequentemente pautada como tópico introdutório ao campo e, apesar de cada um desses binômios ter suas nuances, na maioria das discussões que presenciei foram intercambiáveis. Surgiram para sustentar argumentos sobre esse contraste entre a imagem “fictícia” da IA e aquilo que meus interlocutores faziam concretamente. Além disso, a divisão passa por uma mudança de métodos na IA, propulsionada, em grande parte, pela influência de uma concepção conexionista do funcionamento da mente e pelo desenvolvimento das redes neurais artificiais (KATZ, 2020).

Aqui, falaremos de “modelos conexionistas” e “modelos simbólicos” ao tratar mais especificamente das diferenças metodológicas na IA. Por outro lado, nos guiaremos pela nomenclatura “IA forte” e “IA fraca” para falar desses dois momentos com epistemologias sobre a mente e objetivos de magnitude muito diferentes, considerando a história dos termos em debates dentro das ciências humanas. Com esse fim, trago a breve definição nos termos de Searle - cuja obra, em grande parte, é uma das mais notáveis críticas à concepção de Turing sobre o pensamento:

> According to weak AI, the principal value of the computer in the study of the mind is that it gives us a very powerful tool. For example, it enables us to formulate and test hypotheses in a more rigorous and precise fashion. But according to strong AI, the computer is not merely a tool in the study of the mind; rather, the appropriately programmed computer really is a mind, in the sense that computers given the right programs can be literally said to understand and have other cognitive states. In strong AI, because the programmed computer has cognitive states, the programs are not mere tools that enable us to test

> psychological explanations; rather, the programs are themselves the explanations (SEARLE, 1980, p. 417).[17]

O autor destaca, assim, a diferença entre uma IA em que se tem o estudo da mente humana como inspiração para o desenvolvimento de soluções para problemas específicos, e outra, que trata da simulação da mente humana, que projeta a existência de máquinas dotadas de intencionalidade. Desse modo, nota-se por que a distância entre o paradigma da IA que guia o trabalho atual na área e o que o precedia foi tão frisada em campo. Durante um dos TDCs, um apresentador descreveu os modelos de IA como modelos "DIKW", uma hierarquia, que segue a ordem: "data, information, knowledge, wisdom" (dados, informação, conhecimento, sabedoria). "Wisdom" seria o nível de aprendizado mais avançado que se pode ter sobre os dados - representaria a capacidade de tomada de decisões da IA forte. Nesse momento, ele sinalizou que os modelos de ML, atualmente, ainda estão longe desse topo da pirâmide, pois a maioria dos desenvolvimentos em IA se localiza na fase do "knowledge" e dos "insights", que muitas vezes surgem de correlações mal interpretadas (característica da era do Big data sobre a qual entraremos em detalhe posteriormente, ao tratarmos da IA como ética). Nota-se que a ampliação do escopo de atuação da IA é desejável, porém a segmentação não é entendida como um problema em si pois influi no crescimento da área como mercado. Os modelos não produzem sabedoria, porém são validados através da sua eficácia comercial.

Para entender a IA fraca, é preciso entendê-la como associada ao paradigma "conexionista", que influenciou diversas áreas do conhecimento, entre as quais a psicologia, as ciências cognitivas e a neurociência. Ele pauta um entendimento não-simbólico da cognição, num modelo em que a informação é propriedade das conexões entre as unidades em redes neurais (BUCKNER, GARSON, 2019). Os modelos conexionistas são abstrações baseadas em propriedades neurofisiológicas do cérebro e, diferentemente dos "simbólicos", não dependem do uso de regras formais,

[17] "De acordo com a IA fraca, o principal valor do computador no estudo da mente é que ele nos dá uma ferramenta muito poderosa. Por exemplo, ele nos permite formular e testar hipóteses de forma mais rigorosa e precisa. Mas, de acordo com a IA forte, o computador não é apenas uma ferramenta no estudo da mente; em vez disso, o computador adequadamente programado é de fato uma mente, no sentido que computadores com os programas certos podem literalmente entender e ter outros estados cognitivos. Na IA forte, porque o computador programado tem estados cognitivos, os programas não são meras ferramentas que nos permitem testar explicações psicológicas; em vez disso, os próprios programas são as explicações" (SEARLE, 1980, p. 417, tradução minha).

pois são modelos capazes de determinar o seu próprio funcionamento, estabelecendo relações estatísticas entre inputs e outputs através dos dados (KATZ, 2020, p.27).

De partida, falar da influência do conexionismo na IA demanda algumas especificações, visto que dentro do movimento conhecido por esse nome existem diferentes vertentes (PEREIRA, 2021). A que impulsionou o desenvolvimento recente da IA é representada pelas “redes neurais artificiais”. Tratam-se de modelos em que a informação é processada através de unidades (“nós”, “neurônios”) conectadas entre si, que recebem dados externos ou de outras unidades e cujas conexões têm pesos associados a elas que entram no cálculo que determina os seus outputs. O processamento dos dados através dessas camadas permite identificar padrões, classificá-los e fazer predições sobre eles. De acordo com Pereira (2021), o desenvolvimento desses modelos foi informado por uma “teoria neurobiológica da inteligência”, sintetizada da seguinte forma:

> Os estudos sobre a engenharia biológica do cérebro, em seus bilhões de neurônios e ainda mais numerosas conexões sinápticas, começaram a deixar claro que os mecanismos que nos permitem aprender algo não possuem estruturas generativas fixas, mas redes que se transformam por suas conexões. Em suma, “nossos cérebros não estão repletos de lógica ou regras” (ibid., p. 37, tradução minha) e nem o pensamento lógico de um humano adulto “civilizado”, nem a gramática generativa das línguas naturais humanas, são suficientes como modelo inspirador para as inteligências artificiais. (PEREIRA, 2021, p. 279).

As redes neurais artificiais apresentam diversas vantagens em relação aos modelos simbólicos, cujo desempenho demonstrou-se insatisfatório mediante a variação de contextos. A formalização de regras para que seguissem atingindo os seus objetivos sob o efeito das inúmeras variáveis em situações do “mundo real” era algo humanamente impossível: os modelos eram lentos e produziam resultados inadequados. Portanto, sob a hipótese de que os dispositivos poderiam ser autônomos e aprender as suas regras de funcionamento a partir dos dados, os modelos conexionistas demonstraram performances satisfatórias. Conforme a citação supracitada e a discussão que desenvolvemos no segundo capítulo, eles representam o entendimento da natureza como referência para a criação de inteligência (PEREIRA, 2021, p. 279). Desse modo, a possibilidade contínua de aperfeiçoamento através do treino e do ajuste dos pesos na rede foi concebida como a sua capacidade de “aprendizado”.

As redes neurais, atualmente, constituem os principais modelos dentro do que entendemos por “machine learning” (ML), ou seja, a detecção de padrões e resultados probabilísticos através do processamento de enormes conjuntos de dados (ELISH, BOYD, 2017, p. 62). Entretanto, a atenção voltada ao ML nos últimos anos é associada aos produtos da subárea que conhecemos como Deep Learning (DL), responsável pela adoção modelos mais complexos de redes neurais, compostas por múltiplas camadas, capazes de identificar características nos dados sem supervisão humana e com acurácia alta. Os debates sobre modelos “caixa-preta”, que produzem outputs através do estabelecimento de relações entre os dados que nem mesmo os especialistas conseguem acessar, normalmente se referem a modelos de DL.

Feita essa descrição, é válido o parêntese: modelos simbólicos ainda fazem parte de diversas aplicações de IA, apesar dos conexionistas serem predominantes. Há, também, uma variedade de concepções dentro do conexionismo, além de proposições sobre modelos que não cabem nessas classificações e que estão ganhando projeção. Durante uma mesa do TDC, encabeçada por um dos representantes do C4AI, cuja sede foi inaugurada na USP em 2020, tive o primeiro contato com os modelos “neuro-simbólicos”, que combinam modelos conexionistas e simbólicos. De acordo com a apresentação, são modelos híbridos em que a aquisição de conhecimento se dá por meio da combinação entre “conhecimentos teóricos” (regras explícitas, como nos modelos simbólicos) e “conhecimentos empíricos” (“exemplos”, casos práticos). A transferência de conhecimento entre esses dois “módulos” é a aposta do grupo para produzir modelos cujos resultados qualificam-se como “wisdom”. Assim, a integração de modelos neuro-simbólicos foi citada como um dos principais objetivos das pesquisas do centro. Dependendo do impacto que esses modelos venham a ter, o debate sobre a noção de "inteligência" na IA pode mudar muito. Porém, no contexto desta pesquisa, noções vindas do conexionismo são as maiores influências nesse debate.

Podemos entender a diferença entre a IA forte e a fraca de acordo com o que Bruno e Vaz (2002) descrevem como a saída de um paradigma “deliberativo”, em que tratava-se a cognição como a capacidade de raciocinar entre o que é “percebido” e o que é “memorizado”, de acordo um modelo simbólico internalizado, para outro, voltado ao desenvolvimento de agente autônomos, que entende que os padrões de comportamento emergem da interação entre agente e meio. De acordo com os autores, esse novo paradigma foi determinante no desenvolvimento dos “agentes de

rede” - como se referem a tecnologias que nos ajudam a permear o emaranhado de informações com que nos deparamos na Internet e que precedem algoritmos de personalização mais complexos, como os que temos hoje em dia. No contexto de popularização da Internet no início do século, os autores criticam duas teses que eram comuns acerca desse novo meio de comunicação: sobre o caráter aparentemente ilimitado da rede e sobre o entendimento da Internet como uma extensão das nossas capacidades cognitivas, perpetuando a ideia da tecnologia como “prótese” do pensamento humano (BRUNO, VAZ, 2002, p.24-25).

Retornamos, assim, à descrição que iniciou o capítulo: as histórias da Internet e da IA são mutuamente dependentes e estimularam imaginários similares acerca delas. Assim, as duas teses citadas pelos autores se comprovam quando escolhemos falar especificamente das reações que técnicas de ML provocam. De acordo com relatos do campo, os problemas apresentados aos profissionais da IA em suas rotinas de trabalho variam muito, portanto, a escolha dos modelos para resolvê-los é extremamente contextual. Não obstante, também indicaram que a imagem criada sobre o ML nos últimos anos incide sobre uma ideia de que devem trabalhar sempre com esses modelos dentro das empresas. Isso tem relação com a falta de conhecimento técnico por parte dos stakeholders, assim como a contradição entre a prática e a imagem “de fora” da IA, conforme indicamos anteriormente. De modo menos evidente, também apontam para a reformulação da IA através do apelo à técnica e às vantagens de seus dispositivos para o aumento da engenhosidade humana.

Essa reformulação implica na criação de sistemas automatizados através de princípios das ciências de rede, que são em grande parte informados por teorias conexionistas da cognição que se estendem para um entendimento “econômico” das interações humanas. Ou seja, tem-se uma interpretação das relações entre indivíduos como “conexões”, que podem ser analisadas, articuladas e instrumentalizadas (CHUN, 2021, p. 86). A partir disso, Katz (2020, p. 66) destaca que o desenvolvimento recente da área foi possibilitado pela associação entre essas ideias e o “rebranding” comercial da IA, que a reposicionou no mainstream nos anos 2010. Pela sua capacidade de classificar e fazer predições sobre grandes conjuntos de dados, os modelos de IA fraca atraíram investimentos na promessa de personalizar a experiência nas plataformas e facilitar o fluxo de dados de usuários na Internet.

Nesse sentido, Elish e Boyd (2017) descrevem como a noção de “rede neural” foi descartada no primeiro momento dos desenvolvimentos da IA - que caracterizam como teórica e acadêmica -, sendo recuperada com a IA fraca quando as possibilidades comerciais do campo começaram a ser exploradas. Além disso, as autoras frisam como condições materiais - como a ampliação da capacidade de processamento dos computadores - foram determinantes nesse fenômeno e na fomentação da lógica do Big Data . Esse é um ponto central, pois a passagem da IA forte para fraca passa pela transposição da pesquisa em IA para a indústria. A diferença entre a pesquisa acadêmica e a de mercado foi constantemente citada por meus interlocutores, sendo a primeira caracterizada como uma pesquisa que não prepara profissionais para lidar com problemas que surgem na prática da IA.

Num debate que acompanhei, um desenvolvedor sintetizou a distinção afirmando que a academia se interessa por “modelos de entendimento”, enquanto o mercado busca “modelos de predição”. Elish e Boyd (2017) evidenciam a relação próxima dos modelos de ML com o contexto do comprometimento das empresas com o Big Data. As autoras descrevem que constantemente as referências à IA e ao Big Data se referem a um mesmo fenômeno, relacionado à imaginação sobre futuros possíveis do desenvolvimento tecnológico. Entretanto, usamos aqui a caracterização que fazem do Big Data como um novo paradigma do mundo dos negócios, referente ao uso de grandes volumes de dados para informar decisões empresariais. Associado a isso, entendemos que o Big Data representa epistemologia empirista e uma redefinição das práticas de pesquisa que guia os desenvolvimentos recentes da IA (BOYD, CRAWFORD, 2011).

Além disso, nos interessa pensar sobre a “promessa” do Big Data, que trata dos retornos financeiros do investimento em modelos capazes de aprender com os dados em grande escala. A retórica acerca da noção pauta-se, em grande parte, na falta de capacidade analítica dos humanos perante os gigantescos volumes de dados que produzimos atualmente. Nisso surge outro aspecto da reformulação da IA, que permeou as discussões dentro da dinâmica “expectativa vs. realidade” em campo, referente à contribuição dos modelos conexionistas para a inteligência humana: trata-se do contra-argumento sobre a ameaça de substituição da mão de obra humana (atribuída à IA forte). O ponto repetidamente manifestado foi que a IA fraca representaria, na verdade, a oportunidade de delegar os trabalhos “manuais” às máquinas, reservando aos humanos as atividades criativas. Ou seja, a IA seria menos

um perigo e mais uma ferramenta para o aumento da engenhosidade humana. Esse aspecto se relaciona a uma visão moralizada das máquinas que, paradoxalmente, perpassa produção da objetividade científica, conforme Daston e Galison (2007):

> While much is and has been made of those distinctive traits — emotional, intellectual, and moral — that distinguish humans from machines, it was a nineteenth-century commonplace that machines were paragons of certain human virtues. Chief among these were those associated with work: patient, indefatigable, ever-alert machines would relieve human workers whose attention wandered, whose pace slackened, whose hand trembled. Where intervening genius once reigned, there, the nineteenth-century scientists proclaimed ever more loudly, hard, self-disciplined and self-restrained work would carry the day. (DASTON, GALISON, 2007, p. 122).[18]

No início da pesquisa de campo, assistindo no Youtube debates com figuras importantes para a IA no Brasil, me deparei com uma frase que remete a esse aspecto. Uma cientista de dados sintetizou o seu argumento da seguinte forma: “a IA vai ter um doutorado, mas ela vai ter um doutorado para fazer um trabalho braçal. Uma pessoa com doutorado não faz trabalho braçal…” (VIEBRANTZ, 2020) e então prossegue, comentando que uma pessoa com doutorado se dedica ao “trabalho mental”. Há de se contextualizar: na fala, a debatedora fazia uma metáfora sobre o treinamento de um modelo de ML como as etapas na educação formal de uma pessoa. Assim, a “IA com doutorado” seria aquela treinada com corpus de dados de qualidade o suficiente para demonstrar boa performance no que é aplicada. Ainda: ao falar de “trabalho braçal”, referiu-se especificamente aos trabalhos que considera que a máquina faz melhor do que os humanos, como trabalhos repetitivos de leitura e processamento.

À primeira vista, a desconsideração dessas duas funções como “trabalhos mentais” soou estranha. Porém, a chave está na separação entre o trabalho repetitivo e o trabalho criativo, que também pode ser entendida como a mecanização de operações mentais (ERICKSON et al., 2013). Além disso, na descrição, a cientista de dados enfatizou sua defesa pela aplicação de IA no que a máquina realmente faz melhor que os humanos - justamente, tarefas lógicas e repetitivas, que envolvem

[18] “Embora muito seja e tenha sido feito dos traços distintivos – emocionais, intelectuais e morais – que distinguem humanos de máquinas, era um lugar-comum no século XIX que as máquinas eram modelos de certas virtudes humanas. As principais delas eram aquelas associadas ao trabalho: máquinas pacientes, incansáveis e sempre alertas aliviariam trabalhadores humanos cuja atenção vagava, cujo ritmo diminuía, cuja mão tremia. Onde uma vez reinou o gênio interventor, os cientistas do século XIX proclamaram cada vez mais alto que o trabalho árduo, autodisciplinado e autocontrolado venceria” (DASTON, GALISON, 2007, p. 122, tradução minha).

grandes volumes de dados -, mas não em tarefas que os humanos fazem melhor, como as que exigem “razoabilidade[19]”. Ainda assim, o contraste entre a caracterização de uma IA voltada à superação e uma voltada ao aperfeiçoamento da inteligência humana, auxiliando os humanos a reservar seu tempo a atividades supostamente adequadas às suas capacidades, indica tanto aspectos sobre a noção de inteligência, como a de humanidade que são mobilizadas nas práticas do campo.

Pensando a dimensão de mercado da IA, esse imaginário acerca de novas técnicas é útil aos objetivos expansionistas de economias neoliberais. Seguindo a linha interpretativa de Elish e Boyd (2017), reconhecendo a longa história da operacionalização de técnicas estatísticas na lógica da governança neoliberal, utilizam-se de suas experiências etnográficas para fundamentar o entendimento do Big data e da IA como sistemas sociotécnicos, ou seja, alinhar o discurso sobre com as práticas que os conceitos envolvem. Desse modo, os autores esmiúçam alguns dos aspectos culturais que influenciam a criação de um discurso público sobre IA que excede os limites dos métodos efetivamente aplicados no campo; perpetuado em ciclos de “hype” e medo (ELISH, BOYD, 2017, p. 69). Esse descompasso é algo que abordamos através das experiências de meus interlocutores, que se referiram predominantemente ao aspecto do “medo” relacionado às tecnologias. Porém, os autores também tratam do “hype” perpetrado por agentes do mercado como indissociável dessa dinâmica.

Eles descrevem como a criação de novos mercados é dependente da performatividade das Big Techs, que perpetuam uma retórica baseada em hype acerca “do que é possível” com as tecnologias. Assim, atraem a atenção pública para experimentos bem sucedidos que, frequentemente, não são representativos do estado das pesquisas como um todo. Isso influencia, como vimos em campo, as práticas de adoção dessas tecnologias, porque essa retórica de sucesso mexe com o imaginário cultural de tal forma que a implementação dos modelos em questão se torna uma prioridade em setores públicos e privados. Com isso, empresas e organizações passam a aplicar modelos mal construídos, sem terem consciência sobre as limitações das metodologias aplicadas. Ou seja, perpetua-se uma lógica cultural que deixa de

[19] Ao usar o termo, se referiu a obra de Isaac Asimov. A referência específica está no texto "The Naked Sun" (1957), em que o autor escreve: "Logical but not reasonable. Wasn't that the definition of a robot?".

lado o interesse social sobre tais tecnologias, por privilegiar as suas possibilidades técnicas (ELISH, BOYD, 2017, p.64).

Na interpretação de Elish e Boyd (2017), esse argumento é fundamentado na correspondência que identificam entre as tecnologias de ML e a noção de "magia":

> In a brief essay on the correspondences between magic and technology, anthropologist Gell (1988) proposed that a defining feature of magic, as an orientating framework of actions and consequences in the world, is that it is "'costless' in terms of the kind of drudgery, hazards, and investments that actual technical activity inevitably requires. Production 'by magic' is production minus the disadvantageous side-effects, such as struggle, effort, etc." (Gell, 1988, p. 9). To evoke magic is not only to provide an alternative regime of causal relations, but also to minimize the attention to the methods and resources required to carry out a particular effect (ELISH, BOYD, 2017, p. 63).[20]

Ao evocar a noção de magia, portanto, os autores remetem à aparente inescrutabilidade de tecnologias, que é um aspecto protagonista das discussões sobre algoritmos. Ele fica ainda mais evidente no contexto de modelos de DL, que produzem decisões através de processos de difícil explicabilidade, termo que comumente se refere à capacidade de compreensão dos fatores que influenciaram nas decisões de modelos para a produção dos seus outputs conforme discutiremos posteriormente. A proposta de Elish e Boyd (2017) perante esse problema, então, surge através de uma metáfora ousada: propõem pensar Machine Learning como "etnografia computacional", identificando a similaridade de parte das críticas à produção de conhecimento do primeiro com as que já foram atribuídas à etnografia. Nesse paralelo, "humanizam" os processos de ML, assinalando que, assim como o etnógrafo (respeitadas as diferenças de escala), a todo momento precisam tomar decisões sobre o que identificar, agrupar, generalizar, etc. em meio aos dados.

A diferença entre os dois, os autores apontam, está na centralidade que o questionamento sobre metodologias de produção de conhecimento veio a tomar ao longo da história da etnografia. Ou seja, na prática etnográfica é estabelecida a

[20] "Em um breve ensaio sobre as correspondências entre magia e tecnologia, o antropólogo Gell (1988) propôs que uma característica definitiva da magia, como uma estrutura orientadora de ações e consequências no mundo, é que ela é "'sem custo' em termos do tipo de labuta, perigos e investimentos que a atividade técnica real inevitavelmente exige. A produção 'por mágica' é a produção sem os efeitos colaterais desvantajosos, como a luta, esforço, etc." (Gell, 1988, p. 9). Evocar magia não é apenas fornecer um regime alternativo de relações causais, mas também minimizar a atenção aos métodos e recursos necessários para realizar um determinado efeito" (ELISH, BOYD, 2017, p. 63, tradução minha).

dimensão reflexiva de que toda produção de conhecimento é situada (tanto no sentido do contexto da pesquisa, como do enquadre metodológico que mobiliza). Assim, promovem essa ideia de reflexividade para os modelos de IA argumentando que o reconhecimento das limitações dos modelos e da produção de “verdades parciais” através dos dados contribuiria para o desenvolvimento de modelos mais comprometidos com interesses sociais do que com a sua escalabilidade. Com essa descrição, porém, não busco apontar apenas o óbvio: de que as técnicas de ML atraem a atenção do mercado pelo seu potencial de escalabilidade. Também, fundamentada na experiência etnográfica, entendo que a escalabilidade é uma “virtude epistêmica” (DASTON, GALISON, 2007) dentro do contexto de produção de objetividade no mercado.

2.5 OBJETIVIDADE E NOÇÃO DE REDE

Vimos que o tipo de IA que prevalece hoje, na forma da IA fraca, é um tipo voltado ao mercado e produzido predominantemente dentro das Big Techs. Além disso, a transposição da pesquisa da academia para a indústria implica que o desenvolvimento da IA não passa pelos processos de validação do conhecimento científico tradicionais. Entretanto, dada a novidade do setor, as premissas sobre o comportamento e a mente humana associadas à disciplina e, especialmente, as demandas éticas que passaram a protagonizar a repercussão sobre ela nos últimos anos, a IA tem interesse em ser vista como objetiva. Nos interessa, aqui, entender como a reivindicação de objetividade se dá dentro da IA. Portanto, neste capítulo o nosso argumento foca no modo como a dimensão mercadológica da IA indica um modelo de produção de verdade a posteriori que é central na lógica epistêmica neoliberal (MIROWSKI, 2019). Ou seja, a validação do conhecimento produzido através dos modelos de IA é dada em termos do valor que produzem.

Daston e Galison (2007) apresentam uma abordagem histórica da objetividade, descrevendo como o seu significado acompanhou mudanças na prática científica e como o entendimento contemporâneo sobre “objetivo” como aquilo que exclui o “subjetivo” é relativamente recente, ganhando status científico a partir da metade do século XIX. Os autores partem da noção de que a prática científica é caracterizada, em diferentes eras, através de “virtudes epistêmicas”, ou seja, “[...] normas que são

internalizadas e reforçadas pelo seu apelo a valores éticos, assim como a sua eficiência prática na garantia do conhecimento” [21](DASTON, GALISON, 20017, p,40-41, tradução minha). Assim, Daston e Galison (2007) descrevem três visões sobre o conhecimento científico e as virtudes epistêmicas associadas a elas: “truth-to-nature”, “objetividade mecânica” e “julgamento treinado” [trained judgment], que correspondem a uma cronologia histórica.

Apesar disso, os autores caracterizam a passagem de uma visão à outra não como uma substituição, mas como uma reação e uma reconfiguração dos repertórios de virtudes epistêmicas anteriores (DASTON, GALISON, 2007, p. 18). Dessa forma, a mudança de paradigma do *truth-to-nature*, em que a objetividade estava associada ao realismo em representações idealizadas da natureza, para o paradigma da objetividade mecânica é descrita como um movimento de desqualificação da intervenção humana na compreensão da natureza através da dependência de novas ferramentas tecnológicas. A produção de imagens automatizadas, através daguerreótipos e fotografias, por exemplo, influenciou no cultivo de virtudes epistêmicas associadas ao “acesso direto” à natureza através da tecnologia.

Angèle (2016) identifica uma falta de historicidade nas abordagens contemporâneas sobre algoritmos e compara a evolução do debate voltado aos especialistas envolvidos na sua criação aos paradigmas descritos por Daston e Galison (2007). Assim, argumenta que as discussões iniciais sobre algoritmos, nos anos 80, entendiam o ofício profissional no campo da programação como uma espécie de arte, que os colocava numa posição social privilegiada perante o conhecimento, similarmente aos experts dentro do paradigma truth-to-nature. Atualmente, entretanto, associa a maior parte dos debates à objetividade mecânica:

> In contrast, most of the current debates reveal strong suspicions about the role of experts in society, not unlike the doubts and self-criticism that characterized the beginning of the “mechanical objectivity” paradigm. Experts are increasingly under attack for their biases and prejudices; they are blamed for the opaqueness and unfairness of “broken” systems in fields as diverse as education, criminal justice, and healthcare. Thus, algorithms now play a similar role as the daguerreotypes and X-rays of the late nineteenth century: only machines can “cure” experts from their own subjective weaknesses. Several mythical virtues now applied to algorithms are similar to the ones assigned to daguerreotypes and cameras. Algorithms are hard working; they are patient and alert; they follow their own rules and

[21] Do original: "[...] norms that are internalized and enforced by appeal to ethical values, as well as to pragmatic efficacy in securing knowledge”.

> cannot be influenced, which is both reassuring and troubling. Experts are increasingly asked to follow the "rule" of algorithms. This understanding comes with a strong moral view of machines, humans, and their relations (ANGÈLE, 2016, p. 31).[22]

Há alguns pontos relacionados ao argumento da autora que nos ajudam a pensar a relação entre os especialistas da IA e a percepção pública da área. Em primeiro lugar, a desconfiança sobre suas práticas que ela menciona se alinha ao argumento de Mirowski (2019) sobre o neoliberalismo como um projeto epistemológico. Na sua tese, isso se refere ao posicionamento do mercado como o "processador de informações ideal", o que associa à desconfiança dos mediadores - incluindo os especialistas. Ou seja, os humanos são percebidos como "agentes cognitivos não-confiáveis", numa dinâmica em que a produção de verdade se dá pela lógica de mercado. Esse aspecto também adianta a discussão ética sobre a IA que abordaremos à frente, pois aponta para o entendimento que, enquanto os produtos da IA são objetivos, os especialistas possuem vieses [23]que prejudicam o seu funcionamento. A tensão resultante da discussão sobre o impacto dos vieses de profissionais da área nos seus modelos foi um aspecto constitutivo do campo. Isso se manifestou através de frequentes menções à inconveniência do "fator humano", conforme alguns interlocutores se referiram à interferência humana no funcionamento em aplicações de IA, como adentraremos no segundo capítulo.

Por enquanto, partimos do entendimento que as virtudes epistêmicas da objetividade mecânica se manifestaram em campo através de uma idealização da racionalidade dos modelos de IA, colocados em posição de superioridade em relação ao julgamento humano. O "fator humano", por sua associação com a subjetividade dos profissionais, desqualifica a objetividade dos modelos. Entendemos que a questão da objetividade é uma preocupação intrinsecamente ligada à problematização ética

[22] "Em contrapartida, a maioria dos debates atuais revela fortes suspeitas sobre o papel dos especialistas na sociedade, não muito diferente das dúvidas e autocríticas que caracterizaram o início do paradigma da "objetividade mecânica". Os especialistas estão cada vez mais sob ataque por seus vieses e preconceitos; eles são culpados pela opacidade e injustiça de sistemas "quebrados" em campos tão diversos como educação, justiça criminal e saúde. Assim, os algoritmos agora desempenham um papel semelhante aos daguerreótipos e raios X do final do século XIX: somente as máquinas podem "curar" os especialistas de suas próprias fraquezas subjetivas. Várias virtudes míticas agora aplicadas aos algoritmos são semelhantes às atribuídas aos daguerreótipos e câmeras. Os algoritmos são trabalhosos; são pacientes e alertas; eles seguem suas próprias regras e não podem ser influenciados, o que é tranquilizador e preocupante. Os especialistas são cada vez mais solicitados a seguir a "regra" dos algoritmos. Essa compreensão vem acompanhada de uma forte visão moral das máquinas, dos humanos e de suas relações" (ANGÈLE, 2016, p. 31, tradução minha).

[23] Adiantando a discussão do terceiro capítulo, o termo normalmente é associado a pressuposições e preconceitos.

da IA. De acordo com o argumento de Angèle (2016), consideramos que, perante a discussão sobre vieses e ações algorítmicas discriminatórias, o discurso empresarial garante o caráter objetivo da tomada de decisões dos modelos com a atribuição da "culpa" aos humanos no loop. Dessa maneira, o comprometimento com pautas éticas, envolvendo a discussão sobre princípios epistemológicos que guiam essas aplicações não se faz necessário.

Ao questionar um interlocutor sobre como via a preparação de times em empresas de tecnologias para lidar com questões éticas em IA, me respondeu que na sua visão isso só viria à tona quando "pesasse no bolso" das empresas. Similarmente, ao falar sobre as cobranças sobre transparência e explicabilidade no funcionamento de modelos, caracterizou o problema como algo que envolve tempo, custo e conhecimento. Assim, manifestou que mudanças práticas em relação às pautas éticas dependem de interesses de mercado, um aspecto que surgiria outras vezes em meio aos profissionais da IA. Noutra ocasião, um especialista elencou como o custo financeiro como um dos principais empecilhos à explicabilidade de modelos, contrastando com a repercussão leiga, que tende a focar na inescrutabilidade de modelos caixa-preta. Nesse sentido, fica evidente como a lógica de mercado incide sobre a construção (ou reivindicação) de objetividade, assim como sobre a transparência das decisões de modelos.

Conforme apontamos, o desenvolvimento recente da IA, na forma do ML, é largamente baseado em princípios das ciências de rede. O conceito de rede, no contexto da nossa discussão, auxilia a pensar a produção de verdade através da lógica de mercado. A noção é a base da maior parte da produção na IA: a rede se mostrou uma metáfora central nos aspectos técnicos da criação de modelos, nas relações sociais nas comunidades de IA em contextos micro, nas relações de mercado em contextos globais, etc. No campo da computação, os especialistas tanto utilizam como participam da criação de redes sociais, redes neurais, redes de contatos ("networking"), etc.[24]. Trata-se de um conceito central na epistemologia das ciências da computação, que perpassa a produção de objetividade no campo e ajuda a fundamentar a delegação do julgamento às máquinas. Considerando o modo como o conceito de rede foi mobilizado cotidianamente nas práticas e debates dos meus

[24] Para uma discussão prolongada do conceito direcionada ao campo da programação, ver a etnografia de Murillo (2009) com a comunidade brasileira de Software Livre.

interlocutores entendemos que é oportuna uma breve discussão teórica sobre ele, a fim de fundamentar metodologicamente as descrições subsequentes.

Chun (2016) descreve como a "rede" se tornou, no último século, o recurso teórico-metodológico padrão para a superação da “confusão pós-moderna” resultante do abandono de esquemas de explicação universais, transversalmente em várias disciplinas. Quando especificamente aplicado à criação de novas mídias, coloca em evidência a qualidade performativa dessas explicações. Ou seja, a autora descreve as redes como um artifício para a redução da complexidade do fluxo de informações em dinâmicas locais/globais, afirmando que elas tornam “mapeáveis” interações que são fundamentalmente condicionadas pelas restrições do tempo. Encapsulam, ainda, a ideia de uma ordem no comportamento humano - ordem essa que pode ser explorada e prevista e, dessa maneira, produzirem autenticidade. Isso se reflete na associação com o conexionismo e a interpretação das relações humanas como conexões, conforme tratamos. Para a autora, nas ciências de rede a conexão é entendida como hábito (ou como repetição habitual) e todas as interações que não se encaixam nessa categoria são “leaks” [vazamentos] - numa perspectiva similar à de “ruído”, na teoria da informação, ou de “incerteza”.

Além disso, para Chun (2016), as redes também são estruturadas numa temporalidade de crise permanente, que permite manter a rede atualizada e “viva”. Nas novas mídias, “crises” são os eventos que constantemente interrompem o fluxo imenso e desordenado de informações propiciado pelas plataformas digitais e exigem uma reorientação do hábito dos usuários. Direcionando a atenção para associações políticas de axiomas que guiam a elaboração de modelos computacionais atualmente, a autora também descreve como o conceito de “homofilia” se popularizou, nas últimas duas décadas, em publicações acadêmicas na área da computação. O termo se refere ao pressuposto sociológico de que há uma tendência de os indivíduos se aproximarem daqueles similares a eles em algum aspecto. Assim, argumenta que a abordagem das ciências de rede atualmente é fundamentada no entendimento da similaridade como propulsora para a conexão (KURGAN et al., 2019.). Na prática, entretanto, esse aspecto que indica similaridade é determinado pela estrutura algorítmica das plataformas.

Chun (2016) também entende que a epistemologia das redes é fundamentada numa ideia de coletividade neoliberal: para ela, o conceito indica uma interpretação da “conexão” através de conceitos como “nós”/”nodes” e “arestas”/“edges”. Isso não

abre espaço para a criação de um “nós” enquanto identidade, mas sim para a produção de um coletivo de “YOUs”, interconectados por ações assíncronas. Esse argumento complementa a perspectiva de trabalhos anteriores da autora, especificamente no livro *Programmed Visions* (2001), em que ela reflete sobre a noção de racionalidade associada ao ato de “programar” pensando a relação entre software e memória, e, assim, cria um argumento sobre o software como uma “metáfora da metáfora”, ou seja, uma metáfora que serve para articular outras. A partir disso, a autora posiciona o software como central para governamentalidade neoliberal, no sentido em que há uma lógica para a redução da complexidade que é transversal a ambos e que é, também, pautada na ambiguidade: usuários são consumidores e produtores, geram informação e informação é gerada sobre eles, etc. (CHUN, 2011).

A partir da interpretação de Chun (2011, 2016, 2021), entendemos que a noção de rede, no modo como é mobilizada na manipulação da conexão nas novas tecnologias digitais, atua na descontextualização e generalização de relações humanas, ou seja, na criação de “universais”. Nesse sentido, é oportuna a crítica de Tsing (2010) que, direcionada ao fazer etnográfico, problematiza a noção de maneira que cabe à abordagem da IA. A autora se coloca numa posição crítica e complementar à noção de rede mobilizada em teorias aplicadas ao campo das ciências humanas, especificamente na Teoria Ator-rede, por identificar problemas decorrentes da recusa de uma ambição holística nessa abordagem. Ela identifica perigos associados à descontextualização do campo etnográfico nas abordagens de rede e, por isso, opta por focar na heterogeneidade de projetos criados e perpetuados através da diferença cultural, na interconexão entre atores geográfica e culturalmente distantes, sem abandonar a ideia de uma descrição situada.

Similarmente, teóricos têm abordado as práticas da IA pelo ângulo da descontextualização do desenvolvimento tecnológico. Apontando como problemas éticos identificados em seus modelos vêm sendo associados à intervenção humana, discutem como a área se beneficia comercialmente da divisão superficial entre o âmbito “técnico” e o “social” (KATZ, 2020; CHUN, 2021). Assim, a problematização de Tsing (2010) nos ajuda a entender a eficácia da noção de rede em visualizar relações descontextualizadas, especificamente em dinâmicas de mercado. Sintetizando o que discutimos até aqui, portanto, as discussões em meio aos especialistas da IA evidenciaram que demandas de mercado incidem diretamente na pesquisa e desenvolvimento da área. A diferença entre os dois tipos de IA, uma

obsoleta e com caráter generalista, e a outra, promissora e segmentada, é uma expressão disso. A percepção dos sistemas da IA como ferramentas para o aumento da engenhosidade humana, assim, enfatiza a sua característica técnica, agindo na produção de objetividade no campo e na garantia da confiança dos usuários nessas aplicações.

3 INTELIGÊNCIA ARTIFICIAL COMO CIÊNCIA

No primeiro capítulo, introduzimos a inteligência artificial nos aspectos em que se configura como um fenômeno de mercado, enfatizando as características que fazem do campo uma comunidade em ascensão e em intensa atividade econômica. Exploramos as controvérsias conceituais identificadas nele, descrevendo a divisão entre os dois tipos de IA e como a lógica de mercado que caracteriza a IA fraca incide nas práticas dos especialistas no campo. Neste capítulo, a intenção é abordar a recursividade entre metáforas naturais e metáforas computacionais na criação de modelos. De acordo com a experiência de campo, exploraremos a ideia do uso de modelos inspirados na natureza como caminho para a criação de “inteligência”, pensando essa associação como uma forma de atribuição de objetividade a conhecimentos produzidos no âmbito da indústria. Além disso, a intenção é considerar as influências do paradigma conexionista através da articulação recente entre conhecimentos da neurociência - pautados na ideia de redes neurais - e a criação de modelos de ML, considerando a ideia de “aprendizado” que mobilizam.

3.1 NOÇÕES DE NATUREZA E COMPUTAÇÃO

A primeira apresentação que assisti na trilha de IA, em 2019, tratava do uso de técnicas de computação natural (CN). Construía o argumento sobre a importância da inspiração na natureza para o desenvolvimento de metodologias computacionais, através de exemplos práticos de algoritmos e sistemas desenvolvidos e modelados com base em conhecimentos da biologia. Com um subtítulo característico (“a vida inspirando a máquina”), a fala trazia a inspiração no comportamento coletivo de seres vivos. É o caso de algoritmos de enxame, modelados pelo comportamento de insetos,

ou outros, inspirados no movimento de cardumes, por exemplo. Também inspiram-se nas características do corpo humano e de outros animais para o desenvolvimento de soluções computacionais. Na ocasião, o palestrante - um jovem auto descrito como arquiteto de soluções numa conhecida multinacional - guiou os inscritos na trilha através de três subdivisões dentro da CN: a computação inspirada na natureza, a simulação e emulação de fenômenos naturais e a computação com novos materiais naturais.

Através da exposição dessas subáreas, a descrição estabeleceu a computação natural como uma forma de compreensão da natureza, ou de criar “intimidade” com ela. Nas palavras do apresentador, cada subárea representa um tipo de relação diferente com o “mundo natural”. A primeira, da qual mais nos ocuparemos aqui, trata do estudo de fenômenos naturais como inspiração para o desenvolvimento de soluções para problemas complexos, propondo técnicas alternativas às tradicionais, ao tentar reproduzir o funcionamento de processos e mecanismos naturais. Já a segunda subárea trata da síntese computacional de processos naturais, para adquirir maior conhecimento sobre esses - na apresentação, surgiu o exemplo da modelagem de um cardume. A terceira trata da busca por novos materiais naturais a serem usados em processos computacionais, com destaque para a computação molecular e a computação quântica.

Essa exposição apontava para uma ideia recorrente no campo, que passei a interpretar como uma das “verdades” em meio à comunidade de IA: de que a natureza, por ser “computável”, pode ser não apenas simulada, mas também compreendida através da mediação de ferramentas computacionais. Nesse sentido, há uma complementaridade entre os processos de desenvolvimento dessas ferramentas, pois não apenas auxiliam a compreender fenômenos naturais, conforme indicado, mas são também desenvolvidas a partir de conhecimentos científicos sobre o comportamento de seres vivos, principalmente aqueles vindos da biologia. A eficácia dessas ferramentas, assim, é medida pelo quanto se aproximam do funcionamento do mundo “natural”. Através dessas ideias, sustenta-se uma noção de objetividade mecânica (DASTON, GALISON, 2007), em que a evolução de modelos de IA é associada à aquisição de conhecimento direto sobre a natureza.

A computação, em larga medida, é compreendida como uma “metateoria” (WOLFRAM, 2002), transversal a todas as áreas da ciência, porém distinta por seu caráter “engenheiro”, ou seja, por construir objetos que causam efeitos no “mundo

real". Isbell et al (2010) realizaram uma pesquisa em que coletaram a opinião de diversos acadêmicos no campo da computação, a fim de discutir a reformulação do currículo da disciplina e a definição de "computação", propriamente. Assim, expressaram a ideia mencionada da seguinte forma:

> Like mathematics, we build models; unlike mathematics', our models are active and effect-making: they cause things to happen. Like science, we study a system that exists in nature; however, like engineering, our systems are artificial technology and subject to complex trade-offs in implementation. Computing also bears resemblance to the arts—the creation of artifacts—to humanities—the study of texts—and to the social sciences—the study of humans and societies (ISBELL et al., 2010, p. 196).[25]

No ano seguinte da apresentação citada, o palestrante repetiu a mesma fala, noutra edição do TDC. Na ocasião, as percepções sobre a noção de natureza mobilizada na descrição contribuíram para guiar a minha atenção, permitindo perceber, além das subdivisões da computação natural, as nuances entre três tipos de computação citados por ele: a natural, a evolucionária e a universal. A computação evolucionária é considerada um ramo dentro das técnicas de computação inspiradas na natureza e trata de processos de otimização de soluções algorítmicas baseados na teoria da evolução darwinista, em especial os algoritmos genéticos. Esses algoritmos são amplamente utilizados no campo da IA por oferecerem diversas vantagens em relação a outros tipos de soluções algorítmicas lineares e não lineares: utilizam métodos probabilísticos e trabalham com conjuntos de soluções possíveis, ao invés de soluções pontuais.

Os algoritmos genéticos funcionam da seguinte forma. Estabelece-se uma população inicial de "indivíduos" (soluções possíveis), que consistem em analogias aos cromossomos (aqui, os genes de cada cromossomo são valores binários). Em seguida, aplica-se uma função (estabelecida de acordo com o problema que se busca resolver) para calcular o "fitness" de cada indivíduo (ou seja, o quão próxima a solução está de alcançar determinado resultado). A partir disso, há o processo de seleção de indivíduos com os melhores scores de fitness, que passarão para a etapa do

[25] "Como a matemática, construímos modelos; ao contrário da matemática, nossos modelos são ativos e produzem efeitos: eles fazem as coisas acontecerem. Como a ciência, estudamos um sistema que existe na natureza; no entanto, como a engenharia, nossos sistemas são tecnologia artificial e estão sujeitos a trade-offs complexos na implementação. A computação também guarda semelhanças com as artes – a criação de artefatos – com as humanidades – o estudo de textos – e com as ciências sociais – o estudo de humanos e sociedades" (ISBELL et al., 2010, p. 196, tradução minha).

crossover. Nesta, de acordo com um “crossover point” determinado aleatoriamente, os genes dos “pais” (soluções selecionadas na etapa anterior) são recombinados para produzir a geração seguinte de soluções possíveis. Ainda, os “cromossomos” (strings de bits), podem ser sujeitos a processos de mutação, em que os genes de um indivíduo trocam de posição entre si, a fim de evitar uma “convergência prematura” (ponto em que o algoritmo pára de rodar, pois já não surgem gerações diferentes das anteriores), ou o fim da reprodução.

O caso dos algoritmos genéticos é emblemático porque nos ajuda a explorar a visão de natureza perpetuada nas metáforas que animam o campo da IA de maneira mais explícita que em outros modelos, pelo comprometimento com o paradigma darwinista. Helmreich (1998) foi um dos primeiros a adentrar o assunto nos STS, quando, a partir do campo realizado em meio a pesquisadores especializados em algoritmos genéticos, demonstrou como a noção de natureza mobilizada na criação desses artefatos reflete valores e práticas de uma cultura particular:

> I maintain that culture-specific conceptions of the individual, population, sex, reproduction, gender, kinship, and economy inflect the way this 'genetic algorithm' is fashioned. The picture of nature embedded in most GAs is populated with images of individuals, lineages, families and communities resonant with the values and practices of white middle-class USAmerican and European heterosexual culture, the culture to which most of its practitioners belong. I will not argue that the cultural ideas constitutive of GAs render them operationally incoherent, and neither will I maintain that researchers should take on board different views of biology to make 'better' programs. My concern is rather to show that GAs are culturally situated, and so do not mirror an unmediated “nature”. (HELMREICH, 1998, p.40).[26]

A crítica de Helmreich pontua que essa cultura associada às práticas de desenvolvimento de algoritmos genéticos perpetua uma visão da natureza movida por imperativos “masculinistas” como “racionalidade, objetividade, desinteresse e controle” (HELMREICH, 1998, p. 46.) e, dessa maneira, reforça a separação entre “natural” e “artificial” que é norma no campo da IA. Isso se dá através da constante

[26] “Sustento que as concepções culturais específicas de indivíduo, população, sexo, reprodução, gênero, parentesco e economia influenciam a maneira como esse ‘algoritmo genético’ é moldado. A imagem da natureza embutida na maioria das AGs é preenchida com imagens de indivíduos, linhagens, famílias e comunidades que ressoam com os valores e práticas da cultura heterossexual americana e europeia de classe média branca, a cultura à qual a maioria de seus praticantes pertence. Não vou argumentar que as ideias culturais constitutivas dos AGs os tornam operacionalmente incoerentes, e também não vou sustentar que os pesquisadores devam adotar diferentes visões da biologia para fazer programas 'melhores'. Minha preocupação é mostrar que os AGs são culturalmente situados e, portanto, não espelham uma “natureza” não mediada.”. (HELMREICH, 1998, p.40, tradução minha.).

afirmação da natureza enquanto “modelo computacional” ideal, perpetuando a visão do “mundo natural” enquanto um agente. A metáfora que move a elaboração desses algoritmos equipara a natureza a um computador, ocupado na busca por soluções para problemas de otimização, através de processos de reprodução e recombinação de gerações sucessivas de organismos e genes. Entende-se que a natureza é mais eficaz por ser mais experiente e mais antiga do que os computadores desenvolvidos por humanos, ao mesmo tempo em que é a sua medida de eficácia (quanto mais próximos do funcionamento de modelos naturais, mais eficazes). O ponto central a ser considerado é como a inspiração na natureza não impede a constante menção às diferenças entre o que se entende por “natural” e “artificial”, o que se evidenciou em campo na distinção entre a inteligência observada em sistemas orgânicos e sistemas maquínicos.

Há outro ponto de Helmreich (1998, p. 50) que nos interessa para entender o tipo de conhecimento sobre a natureza que esses algoritmos incorporam. Ele coloca que o modelo teórico dos algoritmos genéticos pressupõe a evolução como um processo informacional, referente à transmissão de informações através de gerações sucessivas, que pode ser abstraído e traduzido para diferentes mídias e áreas. Aqui, a genética é percebida num enquadre que permite conectar o argumento à obra de Chun (2011), partindo da definição que a autora oferece de “programa”. Ao criticar o conceito de “software”, ela enfatiza as origens comerciais do conceito, como o entendemos hoje. Aponta que a “coisificação” do software - referente ao processo através do qual o conceito passou de um conjunto de serviços para uma metáfora análoga à ideologia, ou seja, que se refere a um conjunto de fatores relacionados ao “homem” - também levou a implicações decorrentes do estabelecimento de “programa” como um substantivo:

> Software emerged as a thing — as an iterable textual program — through a process of commercialization and commodification that has made code logos: code as source, code as true representation of action, indeed, code as conflated with, and substituting for, action. Now, in the beginning, is the word, the instruction. Software as logos turns program into a noun — it turns process in time into process in (text) space. In other words, Manfred Broy ’ s software “pioneers,” by making software easier to visualize, not only sought to make the implicit explicit, they also created a system in which the intangible and implicit drives the explicit. They thus obfuscated the machine and the process of execution, making software the end all and be all of computation and putting in place a powerful logic of sourcery that makes source code —

> which tellingly was first called pseudocode — a fetish (CHUN, 2011, p.19).[27]

Esse argumento da autora se insere na tese mais ampla que ela apresenta, segundo a qual a transformação de “software” em uma “coisa” é associada à extensão de dicotomias como a separação sujeito/objeto, de tal forma que o próprio conceito de “informação” se tornou, também, uma “coisa”. Ou seja, a “informação”, que inicialmente era considerada algo inseparável de uma pessoa - do ato de informar -, passou a ser entendida como algo “externo” a ela. No contexto da IA, isso envolve a genética e, principalmente, as redes neurais na ideia que a inteligência humana pode ser abstraída e representada computacionalmente. A própria Chun (2011) toca no papel do DNA no entendimento da genética, ao levantar como problema teórico o fato de a informação ter passado por essa transformação tanto na biologia, como na computação, algo que associa com a característica do neoliberalismo que determina a interpretação de relações em termos econômicos.

Ela prossegue defendendo que, mais produtivo do que estabelecer uma relação de influência direta de uma área sobre a outra, é abordá-las como parte do movimento que fez as ciências adotarem um caráter mais especulativo, no século XX. Trata-se de um contexto epistêmico em que as tentativas de dar sentido ao “visível” passaram a elucidar as relações entre indivíduos como pautadas por “programas invisíveis” (CHUN, 2011, p. 107). A crítica de Chun (2011), como citamos ao tratar do conceito de rede, coloca o software como uma metáfora que funciona na articulação de outras metáforas dentro de uma lógica de mercado neoliberal. Isso porque nela estão imbuídas as relações ambíguas entre o que é visível/invisível, instrução/execução, consumidor/produtor.

O argumento mais amplo da autora, portanto, é que o modo como o software encapsula a ideia de algo “invisivelmente visível” e “visivelmente invisível” atravessa interpretações da mente, da cultura, da biologia e da economia. Ou seja, é uma

[27] “O software surgiu como uma coisa – como um programa textual iterável – por meio de um processo de comercialização e mercantilização que tornou do código logos: código como fonte, código como representação verdadeira da ação, de fato, código como confundido com e substituindo a ação. Agora, no princípio, é a palavra, a instrução. Software como logotipos transforma programa em substantivo — transforma processo no tempo em processo no espaço (texto). Em outras palavras, “pioneiros” do software de Manfred Broy, ao tornar o software mais fácil de visualizar, não apenas buscaram tornar o implícito explícito, mas também criaram um sistema em que o intangível e o implícito impulsionam o explícito. Eles, assim, ofuscaram a máquina e o processo de execução, tornando o software o fim de tudo e tudo da computação e colocando em prática uma poderosa lógica de origem que torna o código-fonte – que foi primeiro chamado de pseudocódigo – um fetiche” (CHUN, 2011, p.19, tradução minha).

metáfora produtiva na governamentalidade neoliberal, que ajuda a entender como as relações de poder se dão em contextos em que temos acesso a volumes gigantescos de informação, através de interfaces cujo funcionamento não podemos compreender totalmente. Nesse sentido, podemos entender que a construção do conhecimento como a investigação e replicação de regras “invisíveis” que determinam o comportamento “visível” de organismos vivos consiste, em grande medida, no projeto da IA, conforme demonstram os entendimentos sobre genética e sobre o cérebro mobilizados em campo.

Chun (2011) posiciona a computação como uma tecnologia central no neoliberalismo, que pressupõe a posse, por parte de indivíduos, de seus próprios corpos como uma forma de capital, portanto, depende da voluntariedade de atos individuais. Para a autora, as tecnologias computacionais permitem, ambos, a individuação e a impressão de estar integrado numa totalidade, e funcionam a favor de discursos sobre liberdade e empoderamento individuais. A interação com o software gera a sensação de estar em controle. Porém, ela é condicionada por regras às quais não temos acesso, o que significa que usuários também são “produzidos” nessa interação e, a partir dela, passam a se moldar à imagem do “homo economicus” da governamentalidade neoliberal (CHUN, 2011, p.8). Ou seja, o software aponta para “formas de empirismo, quantificação e objetividade que não apenas são compatíveis com economias morais, elas exigem economias morais” (DASTON, 2017, p. 38).

Assim, é evidente que o prisma do mercado influencia nossas interações e o modo como abordamos fenômenos do mundo real dentro de contextos neoliberais. A intenção aqui não é a produção de conhecimento na IA como uma exceção. Porém, interessa entender como, em muitos sentidos, a evolução da IA fraca surge como o epítome do projeto epistemológico neoliberal, através de narrativas sobre a automação que implica o entendimento dos indivíduos humanos como agentes cognitivos não confiáveis (MIROWSKI, 2019). Por isso, interessa entender o argumento de Chun sobre a lógica subjacente ao software e ao mercado como fundamentando uma concepção do mundo como “programável”, que direciona o modo como compreendemos o funcionamento de máquinas, e também como entendemos o mundo natural. Nesse sentido, princípios como a liberdade individual e a auto governança perpetuaram a percepção de natureza observada em campo.

Uma fala de uma edição posterior do TDC, sobre o uso de algoritmos genéticos na otimização do funcionamento de drones autônomos, ilustrou a ideia. De acordo com o palestrante, o cálculo de “fitness” informa o “quão bom” o indivíduo é, mas isso também está relacionado a fatores contextuais. Por isso, a mutação garante o funcionamento dos algoritmos por “inserir aleatoriedade” nos processos, ou seja, evita que apenas os indivíduos com maior fitness se reproduzam e garante que, mediante situações não previstas, o algoritmo continue evoluindo. Nesse sentido, ao responder à pergunta de um ouvinte sobre o comportamento dos drones autônomos num cenário em que um desses se encontra com uma corrente de pássaros, o palestrante explicou: é importante que “o algoritmo tenha liberdade” para que, dada a situação, ele possa ser “penalizado” e, assim, sofrer uma mutação e melhorar a sua trajetória, ao invés de simplesmente aniquilá-la.

O fitness, nesse sentido, é uma metáfora análoga ao lucro (HELMREICH, 1998). A natureza enquanto agente é entendida como um sistema econômico, perpetuado pelo monitoramento de custos e benefícios, e é a referência para a resolução de problemas complexos a partir de regras simples - o que configura, amplamente, objetivo da computação. Como exemplificado, as metáforas que animam o campo dos algoritmos genéticos fundamentam-se em uma lógica de competição entre indivíduos. A eficácia dessa lógica depende não apenas dos compartilhamento de referenciais por aqueles que atuam no desenvolvimento dessas tecnologias, mas também do senso comum de uma cultura específica que permite validar a objetividade do conhecimento produzido através dela (HELMREICH, 1998).

Enfim, para nos ajudar a fundamentar o argumento sobre a difusão dessa percepção sobre a natureza e a “lógica de programabilidade” (CHUN, 2011) que a fundamenta, retornamos à terceira forma de computação, citada na apresentação sobre computação natural descrita anteriormente: a computação universal. Na ocasião, me chamou a atenção a característica desta como sendo a mais explicitamente filosófica das três citadas pelo palestrante, que a descreveu através da afirmação de que “tudo é computável”. Isso, de acordo com ele, quer dizer que existe uma “computação interna” nos organismos vivos, algo que se comprovaria atualmente com o auxílio de ferramentas computacionais - novamente, caracterizadas como possibilidades de acessar a natureza.

Durante a apresentação, o mérito da ideia de que existe computação na natureza foi atribuído a Stephen Wolfram, um físico que se tornou reconhecido na

ciência da computação, principalmente por seu trabalho com os autômatos celulares. Na obra mais famosa de Wolfram, publicada em 2002, me deparei com uma tese sustentada através de argumentos com os quais me familiarizei durante o campo. Em "A New Kind of Science", o autor afirma que todo sistema, orgânico ou inorgânico, pode ser entendido como um tipo de computação, e que todos os sistemas que produzem comportamentos complexos têm sofisticação equivalente - o que ele chama de "Princípio de Equivalência Computacional". O autor assim descreve a noção de "computação":

> [...] now every day we see computations being done with a vast range of different kinds of data—from numbers to text to images to almost anything else. And what this suggests is that it is possible to think of any process that follows definite rules as being a computation—regardless of the kinds of elements it involves. So in particular this implies that it should be possible to think of processes in nature as computations. And indeed in the end the only unfamiliar aspect of this is that the rules such processes follow are defined not by some computer program that we as humans construct but rather by the basic laws of nature. But whatever the details of the rules involved the crucial point is that it is possible to view every process that occurs in nature or elsewhere as a computation. And it is this remarkable uniformity that makes it possible to formulate a principle as broad and powerful as the Principle of Computational Equivalence (WOLFRAM, 2002, p. 716).[28]

Em resumo, um processo computacional é aquele que transforma um determinado input num output através de um conjunto de regras (um algoritmo). Ao interpretar todo comportamento complexo como computacional, o autor advoga por uma revolução teórico-metodológica na forma como se estuda o mundo, uma nova ciência que surge com as possibilidades empíricas decorrentes do desenvolvimento de ferramentas computacionais e da proposta anterior. O enquadre teórico para o estudo de comportamentos complexos proposto por Wolfram surgiu a partir da identificação de uma lacuna nas ciências tradicionais no que toca o assunto, afirmando que a confiança dos métodos tradicionais em análises matemáticas funciona

[28] "[...] todos os dias vemos cálculos computacionais sendo feitos com uma vasta gama de diferentes tipos de dados - de números a texto, imagens e quase qualquer outra coisa. E o que isso sugere é que é possível pensar em qualquer processo que segue regras definidas como sendo uma computação – independentemente dos tipos de elementos que envolve. Então, em particular, isso implica que deve ser possível pensar em processos na natureza como computações. E, de fato, no final, o único aspecto desconhecido disso é que as regras que esses processos seguem são definidas não por algum programa de computador que nós, humanos, construímos, mas sim pelas leis básicas da natureza. Mas quaisquer que sejam os detalhes das regras envolvidas, o ponto crucial é que é possível ver todo processo que ocorre na natureza ou em qualquer outro lugar como uma computação. E é essa notável uniformidade que permite formular um princípio tão amplo e poderoso quanto o Princípio de Equivalência Computacional" (WOLFRAM, 2002, p. 716, tradução minha).

satisfatoriamente apenas no estudo de comportamentos simples (WOLFRAM, 2002, p. 637). Dessa forma, ele estudou experimentalmente diversos padrões de comportamento complexos que podem ser produzidos através de sistemas programados com regras simples, dedicando-se longamente a experimentos com autômatos celulares.

Trata-se de uma obra ambiciosa, conforme o próprio autor indicou numa revisão do texto, quinze anos após a publicação. Ele caracteriza o objetivo central do livro como algo “abstrato”, como uma metateoria, ou “uma teoria de todas as teorias possíveis, ou o universo de todos os universos possíveis” (WOLFRAM, 2017). Fundamentalmente, Wolfram delineia um novo paradigma científico em que os fenômenos se tornam inteligíveis por meio da computação, e a lógica dessa proposta é o que sustentaria, para alguns teóricos, as ciências da computação como ciências naturais. Tomemos o exemplo de Denning (2007), um prolífico cientista da computação, cujo papel foi fundamental na consolidação de técnicas de memória virtual em sistemas computacionais modernos. Vocal em relação a essa mudança no status da área, seu argumento é que, apesar da aceitação da computação ser recente, em diversas disciplinas científicas, a abordagem de seus objetos de estudo como processos informacionais já é estabelecida há algum tempo, enfatizando o caso da biologia e sinalizando, novamente, a recursividade biologia-computação.

Assim, ele descreve como a computação evoluiu de um momento em que computadores eram ferramentas de cálculo para outro, em que a computação se tornou um método para novas descobertas científicas. Então, evoluiu para o decisivo momento em que cientistas de áreas diversas passaram a notificar a descoberta de processos informacionais nas estruturas de seus objetos de interesse - o que indicaria isso como um aspecto objetivo da natureza. “Uma parte da computação natural é emular a vida e, emulando a vida, [ela] nos ajuda a entendê-la”, ouvi na apresentação sobre computação natural da qual vimos falando. Essa fala indica um dos efeitos desse enquadre teórico - convergente com a “coisificação” da informação, que deixa de ser exclusivamente relacionada a questões humanas (CHUN, 2011) - para o estabelecimento da computação como uma ciência ocupada não apenas com a reprodução de processos naturais (com o “artificial”), mas também com o entendimento sobre eles.

Podemos interpretar esse processo histórico em que a computação evoluiu de ferramentas de cálculo para uma lente para a compreensão do mundo de diversas

formas. Erickson et al. (2013) o traduzem na descrição que fazem da transformação da razão, enquanto faculdade que tem o juízo de valor humano como propriedade intrínseca, em uma racionalidade baseada em regras, que chamam de “racionalidade da Guerra Fria”. O nome se refere ao período em que o valor do julgamento humano dentro do que significa ser “racional” começou a ser questionado de maneira transversal às ciências humanas, e com respaldos militares e governamentais. No argumento dos autores, na passagem da razão para a racionalidade, o próprio estatuto das “regras” se altera: de modelos exemplares ou conhecimento tácito (no contexto iluminista) passam a ser modelos mecânicos. Já tratamos do cerne do argumento dos autores anteriormente: trata-se da idealização dessa racionalidade na racionalidade algorítmica, caracterizada pelo privilégio da análise em relação à síntese. Ou seja, pela definição de passos sequenciais para atingir a solução mais eficiente para problemas complexos, numa perspectiva que se pretende objetiva.

A própria noção de algoritmo, entretanto, passa por esse processo histórico de ressignificação, de acordo com os autores. Os algoritmos já tinham uma longa história enquanto séries de operações aritméticas, mas passaram a modelos ideais do funcionamento de ambos, máquinas e humanos através do movimento de separação entre operações mentais e operações mecânicas que fez parte dessa reconceituação da razão em racionalidade. Segundo os autores, o surgimento do algoritmo como o tipo ideal de procedimento racional, nesse contexto, é pautado por características como a sua precisão, generalidade e conclusividade: na lógica do algoritmo, não há espaço para a arbitrariedade, de acordo com essa idealização (ERICKSON et al., 2013, p. 30).

Essa descrição nos auxilia a considerar como a objetividade do conhecimento sobre o mundo na IA é construída pelo empréstimo de teorias científicas que fundamentam a ideia da computação como universal. Tal universalidade se sustenta, também, na interpretação de que todo comportamento é regido por regras (ou programas) “invisíveis”. Essa ideia exclui a necessidade da intervenção do julgamento humano na construção de conhecimento computacional (e, portanto, conhecimento sobre a natureza), uma noção produtiva considerando o horizonte de automação delineado na IA fraca. O entendimento sobre a objetividade dos procedimentos computacionais, entretanto, está fortemente embasado numa ideia de racionalidade culturalmente condicionada.

Retornamos a Helmreich (1998, p. 40), na sua afirmação de que os princípios epistemológicos que movem a elaboração de algoritmos genéticos estão fortemente enraizados em valores de uma cultura branca, de classe média, norte-americana e europeia. Devemos considerar que, desde a publicação do texto em questão, a demografia do campo mudou, mas as características citadas ainda predominam em meio aos profissionais no campo da IA. Isto é atestado em relatórios estatísticos da composição do campo (FÓRUM ECONÔMICO MUNDIAL, 2021) e nas demandas levantadas por interlocutores, principalmente relacionadas às questões de gênero e raça. Essas evidências demonstram que os imperativos que guiam o desenvolvimento de aplicações na IA ainda ressoam esses valores dessa maioria restritiva. Nesse sentido, recentemente teóricos têm se debruçado sobre a aparente neutralidade da noção de natureza nutrida na IA, perpetuada historicamente através de concepções de mundo marcadas por divisões de gênero e raça (CHUN, 2021). De acordo com Katz (2020), ainda, o desenvolvimento da IA e as compreensões sobre o mundo natural associadas a ela expressam as ambições “totalizantes” da branquitude.

Nesse percurso, buscamos explorar menos a noção óbvia de que o entendimento da natureza na IA é mediado, e mais o modo como o software realiza essa mediação, em termos metafóricos e concretos, influenciando o entendimento de modelos como eficazes e objetivos conforme adotam princípios “naturais”. Ainda, nos interessa pensar como a separação entre natural/artificial, que à primeira vista parece se flexibilizar quando todos os processos são considerados computacionais, é, de fato, fortalecida quando a eficácia de algoritmos é interpretada como validação dessa percepção de natureza culturalmente situada que abordamos. Desse modo, avança a ideia de que a objetividade científica seria permitida pela evolução da IA, e o aspecto humano da criação da natureza é ignorado (HELMREICH, 1998, p.40). Nisso tem lugar também a separação entre organismo e ambiente, que ao mesmo tempo em que garante o funcionamento de modelos, se constitui também como um dos obstáculos para a automação. Para explorar essa ideia, na próxima seção trataremos da noção de “redes neurais” e da interpretação do cérebro, vinda da neurociência, que fundamenta as abstrações da IA fraca.

3.2 REDES NEURAIS, IA E NEUROCIÊNCIA

Abordamos anteriormente como o conceito de “rede neural” foi central na reformulação da IA e na elaboração de modelos conexionistas que alavancaram o potencial comercial da área. Com a adoção em massa de modelos de ML, as redes neurais começaram a atrair a atenção do público fora dos círculos científicos, sendo apresentadas como modelos abstraídos da fisiologia do cérebro humano. De início, a minha impressão foi que, na epistemologia da IA, o entendimento sobre como funciona uma rede neural equivale a entender o cérebro - e entender o cérebro seria o caminho para entender a inteligência. Porém, ao longo da pesquisa, essa correspondência entre o modelo artificial e o funcionamento efetivo do cérebro humano não indicou o consenso que eu esperava nas discussões entre os especialistas em IA. Pelo contrário, em algumas ocasiões, foram explicitamente apontados como modelos frouxos da realidade.

Em um debate sobre a evolução da IA e o futuro da área que acompanhei, um debatedor colocou uma questão sobre como entender melhor o ser humano (citando o comportamento e o cérebro) poderia influenciar na elaboração de algoritmos melhores, citando o caso da rede neural artificial como uma das “abstrações fuleiras” que guiam os desenvolvimentos no campo. No primeiro capítulo, abordamos o conceito de rede como um recurso teórico-metodológico que se tornou padrão em diversas disciplinas no último século (CHUN, 2016), em relação ao qual destacamos duas qualidades principais: o vocabulário próprio da lógica neoliberal associado a ele, e sua qualidade performativa. O conceito de “rede neural”, quando usado para a explicação do cérebro, partilha dessas características. Mesmo quando entendidos como modelos inexatos, o uso das redes neurais foi justificado pela eficácia dos seus outputs, por produzir resultados que interessam ao mercado. Além disso, existe uma sutileza na crítica, porque ela se direciona à correspondência da rede neural artificial com a realidade, mas não toca na credibilidade das redes neurais naturais como modelos de entendimento da cognição humana.

Assim, após discutir a ideia mais abrangente de natureza que permeia as práticas na IA, nesta seção, focaremos em aspectos associados a uma percepção da “natureza humana” - ou da natureza da inteligência humana - que também reflete pontos discutidos anteriormente. Para isso, vamos percorrer, principalmente, o trabalho de Azize (2010), pensando a relação entre as práticas na IA e práticas da

neurociência. Especificamente o movimento neurocientífico, nas últimas décadas, que o autor vincula a uma noção cerebralista de pessoa, que associa o indivíduo ao cérebro. Essa análise nos interessa para pensar como a neurociência e a IA se retroalimentam através das pesquisas sobre dois objetos que constantemente se sobrepõem - cérebro e inteligência -, e que dizem respeito, direta e indiretamente, ao corpo e à subjetividade humana.

Conforme o questionamento colocado anteriormente e outros que presenciei em campo, assim como em materiais sobre IA, o uso de modelos abstraídos do funcionamento do cérebro não significa que eles sejam interpretados como representações perfeitas do mundo real por parte dos profissionais da área. Aí está uma das tensões no campo, porque essa interpretação vem de dentro da área, por pessoas que atuam no desenvolvimento de aplicações, mas para a imagem externa, para o mercado, é vantajosa a ideia de verossimilidade. Portanto, o argumento que trago aqui, com base no campo, é que, apesar do ceticismo de profissionais envolvidos na aplicação de redes neurais artificiais, para o público leigo a correspondência deles com o modelo fisiológico não é contestada. Os modelos funcionam, atraem interesse comercial e a inspiração na neurociência associa um caráter científico a eles. Além disso, a associação neurocientífica com a concepção “cerebralista” de pessoa demonstra uma continuidade com o argumento que vimos tratando, pois trata-se de uma noção compatível com valores do individualismo neoliberal. Entendemos que todos esses fatores atuam na reivindicação de objetividade na IA.

Nas discussões sobre redes neurais em campo, evidentemente, o cérebro surgiu como principal objeto de debate, sendo o aumento do entendimento sobre o órgão implicado como um passo na construção de modelos melhores. Por isso, nos interessa trazer a discussão, na antropologia, sobre o valor social do cérebro feita por Azize (2010). O autor apresenta uma etnografia da difusão neurocientífica sobre o cérebro, pensando os recursos mobilizados na comunicação com o público leigo, e as suas “aparições” em produções de mídia e entretenimento. Desse modo, explora as analogias que sustentam o discurso da neurociência sobre o órgão de acordo com um “acervo semântico e simbólico” que a cerca. Ele descreve a neurociência como uma disciplina recente, destoante da tendência à superespecialização de disciplinas científicas, porque abrange outros campos que partilham o interesse sobre o cérebro. O argumento principal do autor é que a difusão neurocientífica promove uma

concepção de pessoa que é informada por um "cerebralismo" - entendido como "uma visão de mundo que liga um indivíduo ao seu cérebro e situa neste órgão o lócus de nossa identidade pessoal" (AZIZE, 2010, p.6).

Seu argumento é sobre o modo como as neurociências são baseadas numa crítica ao dualismo cartesiano, por entenderem cérebro e mente como uma coisa só. Ou seja, a concepção de pessoa do cerebralismo é entendida como um desdobramento do "fisicalismo", que trata da noção sobre a corporalidade como "dimensão autoexplicativa do humano" (DUARTE, 1995, p.99 apud AZIZE, 2010, p. 6). De modo ilustrativo dessa visão em campo, como resposta à questão sobre o futuro da IA citada anteriormente, um desenvolvedor colocou outra questão retórica, sobre qual o "papel computacional" no pensamento humano de características que não consideraríamos racionais, indicando o modo como isso não é levado em conta nos modelos. Quando eram trazidas à tona as dimensões da experiência humana (para nós, antropólogos, inerentemente intersubjetiva) que não se caracterizam propriamente como racionais, eram remetidas ao cérebro e entendidas como uma questão de processamento. Isso, mantém o órgão como o protagonista deslocado do seu "ambiente" em dois sentidos: na ocultação da discussão sobre o corpo e da natureza cultural do que constitui "emoção", "sentimento", etc.

Noutra ocasião, uma cientista de dados defendeu que a replicação da inteligência deve ser consciente de que o cérebro é parte de um corpo, sendo, portanto, sujeito a estímulos externos constantes. Outro debatedor corroborou o ponto, argumentando que a tomada de decisões para os humanos não se caracteriza por uma simples resolução de problemas, mas tem associação íntima com outros aspectos, como as emoções. Esses casos têm uma característica interessante, pois quando a qualidade racional do processo de tomada de decisões humano é colocada em questão, o problema ético da automação começa a ficar mais palpável. Ou seja, a universalidade de regras contradiz a dependência contextual do julgamento humano. Isso coloca em dúvida a confiança em dispositivos autônomos, vide o debate sobre carros autônomos e o dilema moral implicado em quem escolhem atropelar (BONNEFON, SHARIFF, RAHWAN, 2016). Desse modo, associo essas sinalizações dos especialistas sobre como humanos tomam decisões, em parte, à evolução das demandas sobre a ética de sistemas automatizados, e à exigência de conhecimento sobre efeitos de modelos nos seus usuários por parte das pessoas envolvidas na sua criação

Ainda, o modo como o reconhecimento do impacto das emoções e de outros fatores subjetivos no comportamento humano surge tem suas peculiaridades na IA, porque entende-se que a influência desses fatores é algo que pode ser visualizado na atividade cerebral. Aqui, temos mais um paralelo com Azize (2010), que associa o desenvolvimento tecnológico - no nosso caso, aplicações de IA como técnicas de visão computacional, por exemplo - à criação de um cérebro mais “transparente”. Azize (2010) associa o refinamento de tecnologias para a visualização do cérebro à produção de uma concepção “objetiva” de pessoa, e à ideia de que quanto mais dados se tem sobre o cérebro, maior é a compreensão sobre a subjetividade humana. Similarmente, a difusão da ideia de que aspectos como as emoções poderiam ser “visualizados” computacionalmente é parte da construção de objetividade da área.

Num dos primeiros congressos de que participei, assisti uma apresentação em que essa ideia surgiu explicitamente. Ela tratava do uso de IA e de eletroencefalogramas para a identificação de emoções através das ondas cerebrais. O objetivo era demonstrar um modelo para detectar a associação entre as avaliações que consumidores fornecem de produtos com as emoções que esses experienciam: “ler mentes”, conforme sugeriu o título da apresentação. A possibilidade de visualização computacional da subjetividade surgia aí incontestada, através de um caso concreto. A relação cérebro-mente observada aí permite ilustrar como os modelos de IA fraca são fundamentais na validação científica da dessa concepção cerebralista e “objetiva” de pessoa (AZIZE, 2010) da qual vimos falando. Entendemos, também, que essa percepção converge com o princípio mais amplo de que a linguagem da natureza é computacional. Mas também interessa adiantar que o exemplo aponta para a lógica do ML - ainda mais evidente se falarmos em termos de “Big Data” - que toma o conhecimento do comportamento humano como uma questão de indução estatística. Este raciocínio também se aplica à resolução de problemas “humanos” através de modelos computacionais, conforme discutiremos sobre a formulação de problemas éticos na IA no próximo capítulo.

Quanto à visão da subjetividade na neurociência, o argumento de Azize (2010) vai além. Para ele, no discurso neurocientífico, mente e cérebro são dependentes. Porém, não se trata de uma simples equivalência, pois a mente é entendida como um epifenômeno do órgão, ou seja, está submetida a ele (AZIZE, 2010, p. 41). Isso implica, tal como vimos, a submissão da dimensão psicológica à atividade neuronal e, com o entendimento do “orgânico” como sobreposto ao “mental”, a neurociência passa

a ser a medida para avaliar e autorizar ou desautorizar conhecimentos da psicologia. Azize (2010) descreve esse processo como a substituição, em parte, de uma “cultura psicologizada” pela cultura cerebralista. Com isso, ele explora como essa concepção de pessoa percorre - além da ciência e da academia - o vocabulário “leigo”, a mídia, a rotina das pessoas, similar ao que ocorreu com o vocabulário psicológico anteriormente.

Azize (2010), aponta que, apesar do rompimento com o dualismo cartesiano clássico (que, no fim, é reformulado como “corpo/cérebro”), o discurso neurocientífico corrobora o ideal do sujeito moderno dentro do capitalismo neoliberal:

> Tanto em relação àquilo que as neurociências contemporâneas dizem que o cérebro é, quanto àquilo que se diz que podemos – e devemos – fazer com ele, há uma forte afinidade entre tais discursos e um outro que diz respeito aos ideais de comportamento do sujeito no mercado de trabalho e na gestão da vida. [...]. Há uma afinidade eletiva entre conceitos centrais para a neurociência hoje, como o de plasticidade neuronal, e a organização do mercado neoliberal, as políticas de organização e gestão do trabalho, e mesmo o que se espera de um sujeito produtivo. Refiro-me especialmente à coincidência entre o vocabulário utilizado para se falar do cérebro, neurônios e neurotransmissores – tanto em termos da agência que se atribui a eles, quanto a certas metáforas utilizadas para explicar o que esse órgão faz – e certos valores positivados na cultura contemporânea. O cérebro seria, então, uma espécie de espelho daquilo que se espera do indivíduo no capitalismo contemporâneo: plástico, adaptável, presta-se a desafios, passível de ser melhorado (AZIZE, 2010, p. 254-255).

Aqui, partimos para outra fase do nosso argumento, pois a IA fraca não funciona apenas a favor da ideia de detecção da subjetividade pela atividade cerebral, mas também da sua possibilidade de adaptação, o que passa pela noção de plasticidade neuronal. Desse modo, a convergência de discursos passa por valores como a autogestão, a autonomia e a possibilidade contínua de aperfeiçoamento. De acordo com Azize (2010), a “pretensão bilíngue" da neurociência - que busca se debruçar tanto sobre a fisiologia do cérebro quanto sobre a subjetividade humana - perpetua a noção de que as conexões entre neurônios são moldáveis e podem ser aprimoradas através de exercícios, hábitos, treino. Num contexto em que o cérebro corresponde ao indivíduo, o autor descreve que as redes neurais são percebidas de modo consonante com ideias de gestão empresarial, e o engajamento em atividades que supostamente funcionam a favor do aperfeiçoamento cerebral surge na forma de um “investimento”.

Prosseguindo, Azize (2010) defende que o conceito de rede tem um papel fundamental nesse discurso. Percorrendo comparações traçadas por neurocientistas entre o cérebro e metrópoles, florestas, organizações sociais, etc. para explicar o funcionamento de uma rede neural, ele argumenta sobre como a ideia de conexões fluidas embutida aí corrobora o "questionamento da centralidade" numa lógica transversal a diversos tipos de explicação. Ele enfatiza o caso da organização socioeconômica no capitalismo moderno e, nesse ponto, se refere à obra de Malabou (2008), que trata do desaparecimento de um "centro" de ação nas explicações neurocientíficas. Ou seja, de acordo com o paradigma conexionista, elas tratam de assembleias de neurônios e as conexões entre eles, que são ativadas momentaneamente, dependendo da atividade em curso na rede. Para a autora, isso converge com a "redistribuição de centros" promulgada na ideologia neoliberal, que privilegia a distribuição dos centros de decisão em relação ao controle hierárquico representado por um lócus decisório e, fundamentada no conceito de rede, favorece a auto-organização, a flexibilidade e a inovação (MALABOU, 2008, p. 41).

A analogia cérebro-computador também tem um papel na ideia de que o órgão precisa ser compreendido, passar por manutenções, e atualizar o seu software constantemente (AZIZE, 2010, p. 29). A união desses fatores caracteriza uma dinâmica que incentiva o desenvolvimento de uma espécie de "disciplina cerebral". Azize (2010) explora essa analogia através do trabalho de uma neurocientista que a coloca de maneira explícita, nos levando a indagar o que configura o "software" do cérebro na metáfora. Seguindo o argumento do autor, podemos entender que esse lugar cabe ao sistema nervoso e também especular, portanto, que a adoção de "hábitos saudáveis" - de acordo com o validado pelo discurso neurocientífico - caracteriza a sua "manutenção".

Dada a recursividade entre o discurso neurocientífico e o empresarial, se o cérebro pode ser aperfeiçoado, o mesmo vale para os profissionais no campo da tecnologia. Um dos muitos anglicismos com os quais precisei me familiarizar em campo foi o par "hard/soft skills". Raramente presenciei um evento em que, ao falar do profissional em TI, não foram mencionadas as "hard skills", referentes ao conhecimento técnico concreto em determinada área (as expressões não se referem a uma atuação específica). E também as "soft skills", características mais associadas à dimensão "social" da vida profissional, a comunicação interpessoal, a convivência e

colaboração com outras pessoas, etc., sendo o “social” entendido como o âmbito das relações entre humanos.

> [As] skills de um bom profissional hoje em dia estão nas soft skills, os materiais de hoje em dia estão tão simples de entender, tão didáticos que qualquer um pode ser um "senior" tecnicamente falando em pouco tempo, porem [sic] nao sao [sic] todos que conseguem desenvolver habilidades **sociais** [grifo do entrevistado] na mesma velocidade, então saber se comunicar bem, se expressar, ser **proativo** (super importante) [grifo do entrevistado], são características que fazer [sic] de um profissional um bom profissional. (Trecho de entrevista realizada em 14 de junho de 2021).

Essa foi a resposta que recebi de um interlocutor, ao questioná-lo sobre quais características considerava importantes para um profissional no seu ramo de atuação. Não podemos acessar, apenas através dos relatos, como esses requisitos surgem na prática, mas a menção às soft skills foi norma nos assuntos referentes às expectativas sobre os profissionais no campo. Essas expectativas, interessa notar, abrangem aspectos do âmbito subjetivo, que concernem mais as personalidades do que as habilidades técnicas dos profissionais em IA. Em campo, ao falar nas “soft skills”, como na citação, frequentemente houve ênfase na “proatividade”. Isso era esperado, ao considerarmos o contexto do mercado de trabalho de TI, fortemente pautado em ideias como “empreendedorismo de si”. Entretanto, isso também indica outra característica do trabalho na área, pois o destaque dado a essa característica também se refere à dimensão social da construção dos objetos da IA. Afinal, “ser proativo” foi algo relevante ao falar sobre a colaboração, em times, na rotina em equipes multifuncionais - como é o caso na maioria das empresas de tecnologia, de acordo com a rotina de meus interlocutores.

Outras qualidades também surgiram em destaque ao lado da proatividade: a “resiliência”, por exemplo. Numa das palestras principais do TDC, aberta ao público, a apresentadora afirmou que hoje vivemos numa “human economy”, o que implica uma nova era de aquisição de talentos, fundamentalmente diferente da era anterior, em que o talento era medido em termos de “competência” (dentro da “knowledge economy”), pois a medida agora é o “potencial”. Como um dos fatores que definem o potencial de alguém, ela citou a resiliência, que diz respeito à capacidade de “aprender, desaprender, aprender”, indicando a valorização do profissional que sabe se adaptar (notoriamente, a descrição também corresponde às qualidades destacadas nos modelos de ML). Isso evidencia o entendimento de que, assim como as

habilidades técnicas, as qualidades subjetivas podem ser desenvolvidas e adaptadas conforme o necessário para alcançar o ideal do profissional em IA.

No argumento que viemos formulando até aqui, essas demandas de mercado têm algumas implicações. Em primeiro lugar, as semelhanças na ênfase na plasticidade como qualidade, no discurso empresarial e no discurso neurocientífico (AZIZE, 2010). Em segundo lugar, a ênfase nas soft skills parece indicar uma mudança na figura - no discurso sobre a produção de conhecimento na área, pelo menos - do programador introspectivo e solitário, a favor de valores como a comunicação interpessoal e a auto expressão. Forsythe (1993), uma das primeiras antropólogas a etnografar o trabalho de profissionais da IA, identificou duas características consideradas para determinar algo como "conhecimento", de acordo com seus interlocutores: a formalidade e o uso da introspecção como método (FORSYTHE, 1993, p. 458). Cito o trabalho da autora aqui como contraste, porque (ainda que seja possível observar diversas continuidades entre a concepção de conhecimento que descreveu e a mobilizada hoje na IA) a avaliação introspectiva, em oposição à busca e análise de dados empíricos, e ao trabalho colaborativo, não foi algo que observei em campo como valor.

Nesse sentido, a interdisciplinaridade também foi levantada como característica das equipes dentro das empresas e qualidade desejada para o profissional na IA. Esse aspecto se associa ao crescimento comercial da área, que implicou a formação de times com profissionais de especializações diferentes e, consequentemente, um alinhamento de métodos. A valorização da dimensão "social" do trabalho também apareceu constantemente, na forma do "networking". O termo refere-se à prática do indivíduo que busca conhecer novas pessoas e estabelecer contatos, que possam contribuir para a sua carreira através de trocas de conhecimento e oportunidades profissionais em geral - "façam networking", foi o mantra que mais ouvi durante a pesquisa. No networking, a produtividade da noção de rede surge na sua forma mais explícita, por isso, retornamos a ela como a metáfora que fundamenta a epistemologia na IA e também nos ajuda a entender as relações estabelecidas no campo. A "rede" aqui não tem caráter coletivo (CHUN, 2016), pois cada indivíduo constrói e mantém a sua própria rede, e as qualidades desejadas em alguém que venha a compor essa rede são a proatividade, a adaptabilidade, a resiliência, entre outros. As características que se espera do bom profissional, portanto, são compatíveis com aquelas atribuídas ao cérebro.

Com isso, o objetivo aqui é pensar o modo como a disciplina que se espera dos indivíduos nos dois âmbitos é compatível com ideais de autogestão de sujeitos neoliberais, e como a metáfora de rede, em duas formas diferentes, apresenta características similares. Trata-se de considerar a circulação de ideias fundamentadas em uma visão de mundo científica específica, que influencia na percepção sobre objetos tecnológicos como objetivos. Assim, apesar da diferença entre as questões da etnografia de Forsythe (1993) e as que concernem os interlocutores desta pesquisa, ela nos ajuda a pensá-las. Ao localizar a prática científica de seus interlocutores como "cultural", a autora explicita que seu intuito não é apontar como têm uma cultura própria, que difere do modo como o conhecimento se constrói em outros países ou outras disciplinas, mas reconhecer que existem valores e práticas necessários para que se possa fazer parte dessa comunidade. Trata-se, novamente, da economia moral que guia os desenvolvimentos da IA.

Foi a partir desse enquadre que buscamos abordar o entendimento que se tem da inteligência humana, incorporado na concepção de pessoa como "cérebro", e as demandas do mercado de trabalho sobre os profissionais no campo como parte da cultura da IA. O cérebro, como "fronteira final" para o conhecimento do ser humano (AZIZE, 2010), está sempre prestes a ser totalmente compreendido no discurso neurocientífico. Da mesma forma, na IA, os fundamentos da inteligência humana sempre estão prestes a ser desvendados. Mesmo quando se admite uma distância desse objetivo, permanece o mantra do progresso contínuo. Em ambos os casos, essa ideia se baseia na retroalimentação entre tecnologias de IA e o entendimento sobre a natureza sustentado nas práticas do campo. Trata-se de uma compreensão compatível com a caraterística mercadológica do desenvolvimento atual da área e conveniente à governamentalidade neoliberal, como temos demonstrado.

3.3 FATOR HUMANO E A QUESTÃO DO APRENDIZADO

Ao longo deste capítulo, vimos explorando o modo como a IA se desenvolve através de relações recursivas entre conceitos associados a áreas como a biologia e a neurociência, pautando-se no uso de modelos naturais que auxiliam na percepção das suas aplicações como objetivas, ou científicas. Entretanto, a aproximação entre a

IA e disciplinas, em grande parte, ocupadas da compreensão do comportamento animal está longe de implicar num “repovoamento” (RIFIOTIS, 2016) do social, como poderíamos especular a princípio. Pelo contrário, nos deparamos com a constante afirmação das fronteiras entre o “natural” e o “artificial” (enquanto uma manifestação da separação natureza-cultura), na prática e no discurso da IA. Uma das principais formas com que isso se fez explícito em campo foi através das menções ao “fator humano”, ou à interferência humana no funcionamento de modelos. Entendida como prejudicial à objetividade e neutralidade desses, foi tratada como um aspecto que seria idealmente eliminado.

Anteriormente, descrevi uma apresentação que assisti sobre detecção de emoções através de técnicas de IA como exemplo de uma prática associada à concepção cerebralista de pessoa e à ideia de que o pensamento humano pode ser computacionalmente visualizado (AZIZE, 2010). Mas há outro aspecto a ser discutido aqui. Na exposição, foi oferecida a seguinte justificativa para o modelo: “eliminar o fator humano das avaliações de produtos”. Ou seja, a intenção era, como mencionamos, identificar a relação entre emoção e avaliação, determinar se um dia ruim na vida de alguém determina uma avaliação ruim de um produto ou um serviço que ela utilizou nesse mesmo dia. Na ocasião, fui introduzida a uma das principais ambiguidades do campo, pois a inspiração no comportamento humano e o desejo pela remoção de qualquer interferência relacionada a ele no funcionamento das máquinas conviviam nos discursos dos profissionais da IA. Isso ocorria inclusive nos casos em que as emoções humanas, presumivelmente, seriam consideradas informações relevantes, como nos feedbacks sobre as experiências dos sujeitos em determinadas situações, tal qual o exemplo citado.

Entendemos que há aí mais uma manifestação da separação entre os domínios entre natureza e cultura que ecoou nas descrições deste capítulo. Ou seja, a produção de conhecimento na IA - no intuito de se vender como científica - constantemente se afirma como uma disciplina modelada a partir da natureza, sem menção ao que há de cultural[29] nas suas práticas. O ponto a ser considerado, portanto, é que a IA fraca, especificamente na forma do ML (através de suas subdivisões, referentes a níveis

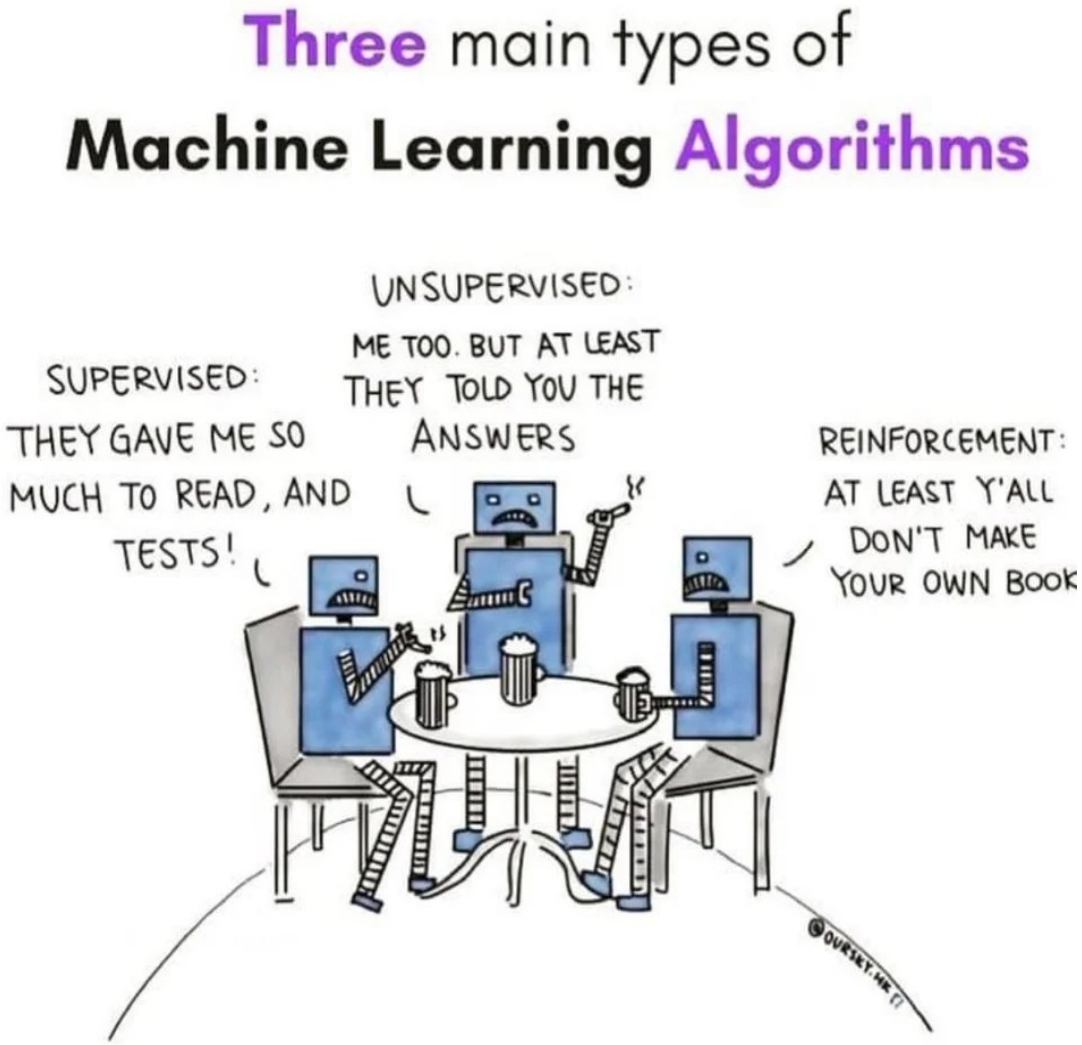
[29] Aqui, reconhecemos a artificialidade da distinção. Entretanto, entendemos que a IA atualmente implica numa reorganização da fronteira natureza-cultura, associada ao modo como produz objetividade (ou seja, posicionando as suas tecnologias como meios para o conhecimento direto da natureza).

diferentes de supervisão humana), propaga um ideal de não interferência humana. Mas esse não é o caso na prática atualmente, afinal, como diversos interlocutores demonstraram, a intervenção humana é requisito na maioria das aplicações, desde o trabalho de tratamento de dados à aplicação dos modelos.

Existem três categorias conhecidas de aprendizado dentro do ML: o supervisionado, o não supervisionado e o por reforço. No primeiro caso, a supervisão consiste no treino do algoritmo através de dados rotulados e de outputs predefinidos (funciona bem para fazer previsões sobre os dados). Já o aprendizado não supervisionado consiste no treino sem a definição de outputs esperados (funciona bem para identificar padrões nos dados, agrupá-los) (BECHMANN, BOWKER, 2019). Já no aprendizado por reforço, conforme explicado numa das palestras do TDC (cujo tema era uma visão geral de ML), o algoritmo aprende no contato com o meio, através de um sistema de reforços positivos e negativos - cujas regras e objetivo o programador não indica à máquina, sendo também não supervisionado. De acordo com o apresentador, esse é o tipo de aprendizado mais próximo ao do humano e, na ocasião, trouxe uma charge para ilustrar as diferenças entre os três tipos citados (Figura 2).

Na fala, ele enfatizou que a maioria dos algoritmos de ML, atualmente, são supervisionados. A constatação surgiu frequentemente, também, na literatura para indicar a vantagem preditiva desse tipo de algoritmo como uma característica que atrai interesse comercial (MANOVICH, 2017; ELISH, BOYD, 2017). No caso do aprendizado supervisionado, o grau de intervenção humana é bastante evidente. Ao oferecer um conjunto de dados rotulados e a conclusão que se espera que o modelo chegue sobre eles, a atuação de uma visão de mundo que fundamenta a classificação dos dados é óbvia. Durante a pesquisa, houve um momento em que participei de um workshop para o desenvolvimento de um algoritmo de ML para a predição das notas de filmes numa plataforma agregadora de avaliações de obras de cinema e televisão. Fui surpreendida por como as etapas de tratamento dos dados e a análise exploratória (exame prévio que permite identificar características dos dados, definir quais deveriam ser excluídos ou agrupados e quais métodos estatísticos deveriam ser aplicados neles) tomaram a maior parte do tempo.

Figura 2 - Charge sobre tipos de aprendizado em ML.



Fonte: Reddit.

De fato, a centralidade da análise de dados na rotina do trabalho da IA foi outro aspecto levantado em campo como parte do descompasso expectativa x realidade desses profissionais, que por vezes apontaram ser uma etapa mais importante do que o treinamento do modelo em si. Um ponto a ser considerado aqui é que, no aprendizado não supervisionado, esse tratamento dos dados também é necessário, sendo a intervenção humana requisitada na definição dos possíveis clusters[30] relevantes para o objetivo que se busca alcançar, na classificação e limpeza de dados. Alguns autores têm abordado esse tópico. Bechmann e Bowker (2019), por exemplo,

[30] Nas ciências de rede e na teoria dos grafos, resumidamente, clusters são agrupamentos entre indivíduos semelhantes, ou que demonstram conexões densas entre si.

realizaram estudos de caso com dois modelos (um supervisionado e um não supervisionado) aplicados a dados retirados do Facebook e concluíram que ambos requerem alto grau de supervisão no trabalho de classificação e de interpretação. Ou seja, eles exploram como decisões humanas são requisitos para definir quais categorias, quais resultados e quais modelos são significativos em determinado contexto, e como as escolhas para alcançar resultados úteis implicam um apagamento progressivo de nuances nos dados, além de estarem sujeitas à variabilidade na relevância atribuída por quem opera o modelo (BECHMANN, BOWKER, 2019, p. 6).

Os autores indicam como vieses são induzidos nos métodos de classificação dos dados, inclusive através de práticas com o intuito não-discriminatório (por exemplo, nuances socioeconômicas que são apagadas no balanceamento de dados, quando atribui-se a diferentes grupos demográficos a mesma representatividade estatística). Desse modo, advogam pelo foco no “nexo humano da produção de conhecimento” (BECHMANN, BOWKER, 2019, p. 8) na IA, que, mais que o acesso ao código (uma das demandas de accountability de sistemas algorítmicos que tem se popularizado), entendem ser um caminho produtivo para mudar a percepção sobre “autonomia" e “supervisão” nesses sistemas. A partir desses argumentos, e das manifestações de reconhecimento, em campo, sobre a quantidade de interferência humana necessária para colocar modelos de ML em funcionamento, como podemos pensar o discurso sobre a inconveniência do “fator humano”?

Forsythe (1993) traz alguns pontos na sua etnografia que nos ajudam nesta reflexão, pois descreve como essa questão já ocupava praticantes da IA nos primeiros anos da reformulação da área, nos anos 90. A autora acompanhou engenheiros de conhecimento que atuavam no desenvolvimento de sistemas especialistas, que são dispositivos interativos com a capacidade de tomada de decisão para a resolução de problemas dentro de uma determinada área, exemplos de modelos simbólicos (ELISH, BOYD, 2017, p. 61). A criação desses sistemas requer uma etapa chamada de “aquisição de conhecimento”, que envolve a realização de entrevistas com especialistas na área para a qual o sistema será voltado, a fim de agregar informações a serem passadas para as máquinas. O que ela observou, acompanhando o desenvolvimento de 5 sistemas especialistas diferentes, foi que a interação dos engenheiros de conhecimento com os especialistas de fora da IA gerava muitos obstáculos, incoerências e traduções errôneas que serviram como justificativa, em meio aos envolvidos, para a ideia sobre a interferência de humanos nos loops como

algo prejudicial ao funcionamento dos sistemas, pois humanos são “ineficientes” (FORSYTHE, 1993, p. 454).

De acordo com a autora, essa é uma evidência das diferenças entre as visões de mundo dos cientistas sociais, que concebem a aquisição de conhecimento como um processo inerentemente intersubjetivo e complicado, e dos engenheiros de conhecimento. Paradoxalmente, Forsythe (1993) apontou que, apesar de sua auto identificação como cientistas, os seus interlocutores nutriam o que caracteriza como um “ethos engenheiro”. Ou seja, uma inclinação ao entendimento de questões técnicas como sendo interessantes, enquanto questões que não entram nessa categorias, as “sociais”, não (FORSYTHE, 1993, p. 456). Esse ethos, segundo ela, também consiste na preferência pela abordagem “prática”, e não a teórica. Podemos entender isso como mais uma característica da IA como mercado. Afinal, a abordagem “orientada pela ação”, pela “tentativa e erro”, produz mais rapidamente do que a fundamentada no desenvolvimento de quadros teóricos para a resolução de problemas. Fundamenta, ainda, um entendimento particular do conhecimento, amplamente fundamentado pela ideia de que os números falam por si só (BOYD, CRAWFORD, 2011).

Com base no trabalho etnográfico, a autora delineia várias suposições que formam a noção de conhecimento de seus interlocutores, das quais elenco as que considero mais oportunas para pensar a questão do fator humano: que o conhecimento é universal e absoluto (FORSYTHE, 1993, p.464). Quando a natureza do conhecimento não é questionada e entende-se que ele se fundamenta em regras universais, informações que estão claramente condicionadas pelo seu contexto não são levadas em consideração. A “falsificação epistêmica” sobre a independência dos objetos tecnológicos em relação aos seus contextos específicos de criação fica evidente aí (KATZ, 2020). Além disso, a universalidade do conhecimento é sustentada pela lógica do Big Data, que postula o entendimento dos números como signos de objetividade. Esse é o fundamento da ideia de que modelos de IA lidam com fatos, que as relações estatísticas não requerem interpretação (BOYD, CRAWFORD, 2011).

Assim, podemos entender que o fator humano é percebido como problema por prejudicar o acesso ao conhecimento - que existe no mundo por si só -, pelo fato de a interpretação humana estar sujeita a variações metodológicas, políticas, culturais, etc. Portanto, há o ideal de exclusão da mediação (da subjetividade) humana no horizonte da automação. O problema da variação interpretativa humana também se manifesta

como uma questão relacionada aos constrangimentos do corpo. Uma situação específica caracterizou de maneira peculiar o problema em campo. Numa das edições do TDC, houve uma fala sobre as vantagens do uso de técnicas de visão computacional para a padronização de estudos de medicamentos antidepressivos. O tipo de estudo em questão era o “Tail Suspension Test”, que consiste num teste para a detecção de camundongos em estado depressivo. A estrutura é a seguinte: um camundongo é suspenso pela cauda por 6 minutos, tendo dois resultados possíveis - “fuga” ou “luta” - que indicam o estado depressivo (imobilidade) ou não depressivo (“movimento de corrida”, “torção de corpo” ou “sacudidas do corpo”) do animal. O ponto do uso de IA nesses testes, de acordo com a fala, era o descarte do “fator humano”, por estar sujeito a uma diversidade de interpretações de uma mesma situação - “por exaustão” -, além de questões referentes à subjetividade de diferentes pesquisadores. Ou seja, nesse caso a agência do camundongo interessa, enquanto a humana prejudica a objetividade da observação.

Até aqui, pudemos inferir alguns pontos sobre os problemas associados ao fator humano na visão de mundo da IA. Em primeiro lugar, a interpretação humana é condicionada por fatores socioculturais que prejudicam o conhecimento sobre o mundo. Em segundo lugar, a máquina é mais eficiente na aquisição de conhecimento por não estar sujeita às limitações do corpo humano (cansaço, estresse, distração, etc. conforme os casos que surgiram aqui). Assim, entendemos que a oposição humano-máquina sustenta a posição dos humanos como agentes cognitivos não confiáveis[31](MIROWSKI, 2019) ao constantemente definir o conceito de humanidade pela sua contraposição ao artificial.

Assis (2018) corrobora o argumento através do estudo antropológico que realizou sobre a noção de humanidade, de acordo com as suas aparições em matérias sobre IA em revistas de grande circulação no Brasil. Similarmente ao que vimos explorado aqui, ela afirma que apesar de as aplicações serem formuladas a partir da inspiração na natureza, “humano” e “artificial” são conceitos que se sustentam através dos aspectos que os opõem. Para Assis (2018), frequentemente a manutenção dessa oposição se dá através das descrições sobre a IA “boa” (a que serve para aumento da engenhosidade humana) e a IA “má” (direcionada à superação

[31] O que implica, novamente, no posicionamento do mercado como o processador de informações ideal no projeto epistêmico neoliberal (MIROWSKI, 2019) - especificamente considerando o caráter comercial da IA.

dos humanos). Isso porque os dilemas éticos e morais inerentes à vida humana são percebidos como barreiras à evolução da IA (ASSIS, 2018, p.93). Ou seja, a distinção que ela identifica remete à separação entre IA forte e IA fraca no discurso dos especialistas em campo.

Essas reflexões etnográficas, assim como as de Forsythe (1993) e as apresentadas nesta pesquisa, podem ser abordadas através da crítica da tecnologia de Ingold (2000), que descreve um processo histórico em que o conceito se desenvolveu através da dicotomização entre "técnico" e "social":

> My thesis, in a nutshell, is that in the societies we study – perhaps even including our own – technical relations are embedded in social relations, and can only be understood within this relational matrix, as one aspect of human sociality. Two further claims follow: first, that what is usually represented as a process of complexification, a development of technology from the simple to the complex, would be better seen as a process of externalization or of disembedding – that is, a progressive cutting out of technical from social relations. Secondly, the modern concept of technology, set up as it is in opposition to society, is a product of this historical process. If that is so, we cannot expect to find a separate sphere of human endeavour corresponding to 'technology' wherever we choose to look. [...] My point is that the concept of technology, at least in its contemporary Western usage, sets out to establish the epistemological conditions for society's control over nature by maximizing the distance between them. (INGOLD, 2000, p. 314).[32]

Através da crítica, portanto, Ingold (2000) propõe uma abordagem da tecnologia fundamentada na experiência, como emergente a partir de ambientes criados e constantemente modificados por agências humanas e não-humanas. Para ele, a dicotomia não tem lugar na prática (INGOLD, 2000, p.321), de modo que seu trabalho pode ser lido como um contra argumento às "falsificações epistêmicas" da IA (KATZ, 2020). A "externalização" da técnica é também caracterizada por Ingold como um processo de "objetificação", que nos ajuda, portanto, a arrematar as discussões

[32] "Minha tese, resumidamente, é que nas sociedades que estudamos – talvez até incluindo a nossa – as relações técnicas estão incorporadas nas relações sociais, e só podem ser compreendidas dentro dessa matriz relacional, como um aspecto da sociabilidade humana. Seguem-se duas outras afirmações: primeiro, que o que normalmente é representado como um processo de complexificação, um desenvolvimento da tecnologia do simples ao complexo, seria melhor visto como um processo de exteriorização ou de desincorporação – isto é, um corte progressivo de técnico das relações sociais. Em segundo lugar, o conceito moderno de tecnologia, constituído em oposição à sociedade, é produto desse processo histórico. Se for assim, não podemos esperar encontrar uma esfera separada do esforço humano correspondente à "tecnologia" onde quer que decidamos olhar. [...] Meu ponto é que o conceito de tecnologia, pelo menos em seu uso ocidental contemporâneo, se propõe a estabelecer as condições epistemológicas para o controle da sociedade sobre a natureza ao maximizar a distância entre elas" (INGOLD, 2000, p. 314, tradução minha).

deste capítulo. O que argumentamos até aqui, com base na experiência etnográfica, vai ao encontro de afirmação do autor de que, historicamente, as ciências naturais foram proponentes de uma "purificação" da natureza, relacionada à exclusão de fenômenos entendidos como parte da experiência humana da sua noção de "natural". Separadas dos humanos, as coisas existiriam por si só e poderiam ser compreendidas pela razão humana - entendida como observadora, e não participante, na natureza (INGOLD, 2000, p.108).

Essa tese perpassa, também, a crítica do autor à abordagem evolucionista do aprendizado, que nos permite fazer um paralelo pertinente com a ideia de aprendizado no ML. Ingold (2010) critica o entendimento sobre o aprendizado por meio da dicotomia entre "aparatos inatos" e "competências adquiridas", discutindo principalmente com ideias da biologia darwiniana e das ciências cognitivas que induzem à compreensão do conhecimento como algo que é "transmitido". Assim, afirma que uma noção mais profícua para abordar o aprendizado humano, entendendo que o conhecimento não está nem no objeto técnico nem no humano, é a "habilidade", compreendida "não como uma propriedade do corpo individual isolado, mas de um sistema total de relações constituído pela presença do agente em seu ambiente (com outros agentes)" (MURILLO, 2009, p. 28). Ou seja, a habilidade, na abordagem ecológica de Ingold, é uma propriedade de relação entre organismo e ambiente, que permite colocar a indissociabilidade da percepção e ação no centro da discussão sobre aprendizado.

O conhecimento, portanto, não é algo a ser transmitido. O aprendizado acontece por meio da redescoberta orientada por outros agentes mais habilidosos, através de um processo que envolve a imitação e o improviso, o engajamento sensível com o ambiente e a "afinação do sistema perceptivo" (INGOLD, 2010, p. 21). Esse processo configura aquilo que o autor entende, a partir de Gibson (2015), como a "educação da atenção". Nota-se que essa interpretação sobre o aprendizado contrasta com o aprendizado mecânico do ML que, influenciado pelo paradigma conexionista e um distanciamento da interpretação da mente como um sistema simbólico de representações, pauta-se numa rejeição da intervenção humana maior do que na IA forte.

Nos sistemas simbólicos, as suas regras de funcionamento são determinadas pela agência humana. Já nos conexionistas, inspirados numa interpretação controversa do aprendizado humano, elas são autodeterminadas. Desse modo, vimos

como a IA passa por um processo de construção de objetividade através do empréstimo de ideias científicas sobre a natureza, principalmente as sustentadas na biologia e na neurociência, de modo que fortalecem a separação entre mente e corpo (ou cérebro e corpo), e entre organismo e ambiente. O discurso sobre modelos de IA fraca como sendo objetivos, por produzirem conhecimento de modo autônomo (ainda que as pessoas envolvidas na sua criação alertem sobre isso não corresponder à prática) e, ao mesmo tempo, serem subservientes aos interesses humanos, constitui as práticas da IA.

Neste capítulo, portanto, buscamos situar as noções de natureza e de pessoa que movimentam o desenvolvimento da IA atualmente. Discutimos como, em sintonia com o caráter comercial da área atualmente, as teorias científicas mobilizadas na práticas do campo remetem a concepções econômicas das relações sociais. Ou seja, conforme demonstramos, a influência de ideias da biologia darwinista e da neurociência na construção de modelos, a partir de abstrações como os algoritmos genéticos e as redes neurais, indicam um entendimento da natureza avivado por metáforas que expressam uma visão de mundo neoliberal. Apesar disso, a referência à natureza associa um caráter científico à IA e postula a evolução das suas ferramentas como caminho para ampliar o conhecimento sobre ela.

Desse modo, a IA como ciência pauta-se largamente na construção de um horizonte de exclusão do fator humano na aquisição de conhecimento. Trata-se de uma dinâmica de reivindicação de objetividade que entendemos ser conveniente para a conservação da indústria perante as problemáticas éticas associadas às suas aplicações nos últimos anos. Através dessas discussões, podemos especular como a concepção de aprendizado no ML se relaciona com o que é hoje, talvez, o maior problema envolvido na automação: a incapacidade de lidar com a contingência. Ou seja, a exclusão da interferência humana nos modelos aponta para a discussão sobre como dispositivos autônomos tomam decisões em situações imprevistas, que envolvem dilemas morais e remetem à associação percepção-ação na vida humana. Esse tópico, como um dos mais polêmicos no campo, nos leva à discussão sobre a IA como ética.

4 INTELIGÊNCIA ARTIFICIAL COMO ÉTICA

Nos dois capítulos anteriores, exploramos como a construção da objetividade na IA está relacionada à comercialização e ao modo como a repercussão das demandas éticas relacionadas às práticas na área incidiu sobre os especialistas. Vimos como a objetividade dos modelos é determinada, em grande parte, pela inspiração em modelos naturais, enquanto a falta de objetividade é associada à intervenção humana. Essa inferência foi fruto, principalmente, das discussões sobre ética que presenciei em campo, que em geral espelhavam o protagonismo do debate sobre vieses algorítmicos na mídia. Em todos os eventos dos quais participei durante a pesquisa de campo, a questão do viés se tornou relevante em algum momento e pude acompanhar como a pauta aglutinou questões éticas, apresentadas como preocupações recentes.

Portanto, neste capítulo a intenção é abordar a objetividade pelo ângulo da ética, seguindo as controvérsias manifestadas em campo, sobre práticas contemporâneas da IA. Tratam-se de debates acerca do papel de modelos de ML na perpetuação de desigualdades sociais e de práticas discriminatórias. Exploraremos algumas consequências do protagonismo da questão do viés nas discussões sobre ética na IA, que implica na formulação de problemas éticos levantados na esfera pública como erros técnicos em meio aos especialistas e porta-vozes da área. Na interpretação predominante sobre os vieses, eles são entendidos como problemas humanos que "vazam" para os modelos. Assim, os esforços na direção de modelos mais éticos se concentram na "correção" desses vieses, constituindo-se num novo ramo da indústria.

Esta discussão se relaciona com as anteriores de dois modos principais. Por um lado, o mercado surge como determinante das pautas éticas que adquirem visibilidade na IA incentivando e restringindo-as de acordo com os interesses comerciais. Além disso, aparece no modo como modelos de ML demonstram empréstimos de teorias científicas - da biologia, da neurociência e das ciências sociais - que fundamentam a predição do comportamento humano e também reproduzem a separação humano-máquina. Ou seja, o comportamento antiético de modelos é associado à agência humana, posicionando a agência algorítmica no polo da neutralidade. Nesse sentido, o ML - e a lógica do Big Data subjacente aos modelos -

representa as virtudes epistêmicas associadas à construção de conhecimento em contextos neoliberais contemporâneos.

4.1 DEMANDAS ÉTICAS E EXPLICABILIDADE DOS MODELOS

Conforme o que tratamos até aqui, na IA a pauta ética remete a contradições entre a prática dos profissionais envolvidos na criação de modelos e os interesses de mercado. Anteriormente, apontamos que aspectos que, em teoria, seriam tópicos relevantes no repertório das discussões sobre ética na IA raramente apareceram nas discussões em campo. É o caso da configuração do trabalho na indústria, no que concerne questões como a informalidade dos "ghost workers" e a substituição de mão de obra humana. Especialmente em relação ao último ponto, a tendência foi associar esse medo à IA simbólica. Novamente, retornamos à divisão entre os dois tipos de IA e como ela se relaciona com a atribuição de diferentes virtudes morais às máquinas (ANGÈLE, 2016; DASTON, GALISON, 2007). Como vimos, em campo existe uma diferenciação entre as máquinas a serem temidas, metas da IA Simbólica, e os modelos elaborados para o aumento da inteligência humana, na IA fraca. Entendemos que essa percepção moral influencia no modo como as discussões sobre ética tomam forma no campo e que a visão das máquinas como estando a serviço dos humanos (CHUN, 2021) apela para o caráter mecânico da IA e, portanto, contribui para a reivindicação de objetividade na área.

Na minha primeira participação no TDC, houve uma apresentação especificamente voltada à ética em IA, feita por dois pós-graduandos que pesquisavam o tema na Universidade Federal do Rio Grande do Sul. Vale distinguir a apresentação das demais que acompanhei, que no geral consistiram em relatos de casos práticos, pois os apresentadores consistentemente complementaram os exemplos expostos com fontes teóricas, evidenciando o enfoque acadêmico do trabalho. A fala também foi distinta porque mencionou, ainda que brevemente, a questão da automação do trabalho. Ao longo da exposição, os painelistas trouxeram diversos artigos evidenciando problemas éticos em relação à IA, como na segurança e no mercado de trabalho, para argumentar a favor da adoção de práticas que privilegiem compromissos éticos em relação à expansão de capacidades técnicas.

Esse argumento, acerca de uma integração entre técnica e social, veio a se mostrar recorrente nas discussões em campo.

A novidade do tema também foi abordada. Um dos apresentadores participou de uma pesquisa sobre as ocorrências de termos relacionados à ética nos títulos de papers em periódicos e conferências mainstream sobre IA, ML e robótica, de 1965 a 2020. Eles identificaram uma baixa relevância do tópico ao longo das décadas. Entretanto, encontraram indícios de uma mudança deste cenário a partir de 2015, com o aumento significativo de publicações (PRATES, AVELAR, LAMB, 2018). O caráter inicial da pauta ética aponta para a externalização - ou objetificação - da técnica (INGOLD, 2000), especificamente quando consideramos o modo como profissionais caracterizaram o surgimento de reivindicações abruptas sobre a conduta ética no campo. Eles enfatizam como a consideração sobre o propósito e os efeitos de tecnologias nos seus usuários nunca foi exigida deles na prática.

Noutra ocasião, o Laboratório de Políticas Públicas e Internet (LAPIN) promoveu um debate sobre a aplicação de IA no Brasil, no qual um pesquisador na área destacou que, desde a sua graduação em 1996, foram 20 anos até começar a ouvir falar sobre ética. Ele atuou na comissão de especialistas formada pela UNESCO para recomendações éticas para a IA, cujo objetivo foi formular um documento que servisse de base para que governos desenvolvessem regulações das práticas no campo. Assim, enfatizou a urgência do envolvimento da comunidade na elaboração de diretrizes para o desenvolvimento ético da IA, especialmente a comunidade de profissionais do Sul Global, que têm baixa representatividade em iniciativas desse tipo. Nesse sentido, as cobranças do mercado sobre modelos explicáveis levantadas pelos profissionais em campo também devem ser consideradas em relação à proliferação de iniciativas regulatórias sobre a prática na IA nos últimos anos, especialmente após a aprovação da General Data Protection Regulation (GDPR) pela União Europeia, em 2016.

No Brasil, a aprovação da Lei Geral de Proteção de Dados Pessoais (LGPD) em 2018 estabeleceu que as empresas deveriam adequar suas práticas em relação ao uso de dados pessoais até 2020. Desse modo, o período de campo coincidiu com o tempo desse processo de adequação e pude acompanhar a discussão ética pelo prisma dos tópicos e das exigências da LGPD que despertaram debates nas comunidades de IA. Houve momentos em que diretrizes da lei foram apontadas como incoerentes em relação à prática dos meus interlocutores, que as medidas foram caracterizadas como inespecíficas, entre outras críticas. No debate promovido pelo

LAPIN, por exemplo, o professor citado apontou a falta de atenção ao contexto brasileiro como uma das razões disso. No caso da LGPD, argumentou que a mimetização dos parâmetros da GDPR, legislação que inspirou o texto, desconsidera muitas das particularidades que envolvem o desenvolvimento tecnológico no Brasil.

Ele argumentou que a presença irrisória de países do Sul Global no desenvolvimento de documentos voltados à regulação da IA, como guidelines e soft laws[33], faz com que demandas relevantes nesses locais não tenham o devido protagonismo. O aspecto da "importação" de pautas éticas é central para abordar essas questões, afinal, quando afirmamos que são norteadas por interesses de mercado, tratam-se de interesses das Big Techs, principalmente. Ou seja, dificulta-se a construção de um debate público sobre ética e IA no contexto brasileiro. Um relatório do Instituto de Tecnologia e Sociedade do Rio (2022) sobre o tema da Justiça de Dados trouxe dados que evidenciam a questão. A partir da realização de entrevistas e de um workshop com stakeholders de setores diversos (academia, sociedade civil, comunidade técnica, setor privado, entre outros), identificaram como prioridade as preocupações referentes às condições de infraestrutura digital no Brasil.

No relatório, esse dado é contraposto à ênfase que o Norte Global coloca em ramificações éticas de soluções e ferramentas com base em dados. Através de exemplos ilustrativos de situações em que a falta de infraestrutura surge como fator determinante na exclusão de determinados grupos, a pesquisa apresenta recomendações no sentido de uma abordagem da diversidade com foco local. Assim, notamos que nesta pesquisa, cujo foco é especificamente a comunidade técnica, costumam ter destaque os problemas mais próximos às preocupações associadas aos países do Norte Global. Consideramos esse aspecto em conjunto com as reclamações de profissionais em campo sobre os obstáculos para o desenvolvimento de uma indústria nacional quando os parâmetros de desenvolvimento tecnológico vêm, principalmente, dos Estados Unidos.

Neste sentido, é um dado relevante que a maioria das falas nos eventos que acompanhei fossem feitas por profissionais que eram contratados ou tinham algum outro vínculo [34]com a Google, a IBM ou a Microsoft. Não obstante, o problema da

[33] No Direito Internacional, configuram normas de caráter não-jurídico. Em campo, foi destacada a flexibilidade das regras como uma característica vantajosa das soft laws em relação às leis jurídicas. Isso porque permitiriam a adaptação das normas de modo mais rápidos, acompanhando o desenvolvimento da IA.

[34] Profissionais podem contribuir no desenvolvimento dessas empresas sem necessariamente manter

sujeição da indústria brasileira às indústrias estrangeiras frequentemente foi abordado, sob o prisma da importação de tecnologias inadequadas para o contexto brasileiro e a consequente desvalorização de aplicações voltadas para o desenvolvimento nacional. Sobre isso, a área de Processamento de Linguagem Natural (PLN), [35]por tratar de um objeto tão explicitamente contextual como a linguagem, oferece alguns exemplos. A predominância de modelos treinados com dados em inglês foi apontada como uma dificuldade para países que falam outras línguas, pois faz especialistas desses locais terem que "começar do zero" (ou seja, construir seus próprios datasets e treinar os modelos). Quanto a isso, impõe-se outra dificuldade, pois dada a complexidade de processar computacionalmente a linguagem humana, são processos que demandam investimentos volumosos.

Além disso, interlocutores alertaram que há a tendência de se "produzir para fora" na IA em termos de mercado e também em termos acadêmicos. Conforme um desenvolvedor colocou, a principal medida de sucesso acadêmico nas ciências da computação é a publicação de artigos. A maioria das revistas qualificadas publica em inglês, o que exige essa adequação de quem quer publicar. No caso do PLN, é algo que colocou como prejudicial ao desenvolvimento nacional do campo, afinal, pesquisadores são induzidos a se dedicarem a datasets da língua inglesa para terem seus trabalhos publicados. Essa característica também se relaciona com a "fuga de cérebros", citada na problematização da falta de investimentos em pesquisa a longo prazo no Brasil. Trata-se do fenômeno em que profissionais se formam no Brasil, se qualificam, recebem oportunidades de empresas no exterior e emigram (geralmente para os Estados Unidos) pelas dificuldades de concretizar seus projetos no país.

Nesse sentido, a importação de tecnologias e pautas do Norte Global é complementar à produção de conhecimento e mão de obra do Sul. É um aspecto do trabalho que influencia tanto no desenvolvimento técnico da área, como na formação de compromissos éticos elaborados especificamente para o contexto brasileiro. Considerando esse fluxo, podemos abordar o tema do "viés", que se configurou como

vínculo empregatício com elas. Vários profissionais no campo tinham o certificado de Google Expert Developer (GDE), por exemplo. Trata-se de um reconhecimento oferecido a especialistas que dominam alguma tecnologia Google e que têm envolvimento ativo em comunidades de tecnologia. Atuam como mentores de outros profissionais e divulgam os produtos da empresa.

[35] Subárea da IA que atualmente, através de técnicas de Machine Learning, é voltada ao reconhecimento, interpretação e compreensão de linguagem natural (dados não estrtuturados em forma de texto). As técnicas de PLN se aplicam em mecanismos de busca, análise de sentimento, assistentes virtuais e outros.

o principal tópico em discussões sobre os impactos da IA no campo, sob o prisma da formulação de pautas éticas como oportunidades comerciais (KATZ, 2021). A indústria criada acerca da temática dos vieses e da criação de uma área de estudos chamada de “Inteligência Artificial Explicável” (xAI) é exemplar dessa questão. O objetivo da xAI é tornar os modelos explicáveis, conforme apontamos, uma demanda que emanou das discussões sobre os vieses humanos que levam a comportamentos discriminatórios de modelos de ML.

Há muitas formas de se interpretar o protagonismo da discussão sobre vieses e o modo como incide nos compromissos éticos de empresas que usam esses modelos. Katz (2021) entende que a pauta desvia a atenção de aspectos políticos e históricos associados a usos discriminatórios de tecnologias computacionais, aspectos que precedem a computação e que remetem às pretensões universalistas da branquitude. Para o autor, debates sobre questões como a transparência de modelos ignoram a discussão sobre os propósitos da adoção dessas tecnologias. O foco em questões de “fairness” - sobre o viés em um ou outro algoritmo ou o quão justos são - influencia na criação de novas indústrias voltadas à “correção” de vieses. Isso representaria um afunilamento das pautas pela atenção e os recursos direcionados à explicabilidade, dedicada a tornar visíveis as decisões dos modelos.

Nesse sentido, um fator a ser considerado é que a xAI agrega esforços não apenas nas Big Techs, mas também tem sido um dos principais programas da Defense Advanced Research Projects Agency (DARPA). Braço de pesquisa do Departamento de Defesa dos Estados Unidos, desde 2015 a agência tem dedicado ao tema parte do seu orçamento bilionário para a IA. Katz (2021) descreveu como a influência da DARPA foi determinante para a evolução da IA desde o seu surgimento e rompimento com a cibernética, influenciando na organização do campo através de um “quadro militarista” (KATZ, 2021, p. 34). Essa associação, de acordo com o autor, é compatível com a visão neoliberal da sociedade que a IA propaga. Assim, entende o foco em questões como a explicabilidade como estratégias para a manutenção de poder institucional:

> Thus when these commentators [indústria de especialistas em IA] claim that, say, systems using neural networks are indecipherable because of their large number of parameters, they are favoring mathematical abstraction over the situated social reality of computing systems. In practice, one does not deal with an abstract neural network but rather its instantiation as software, which is subject to bugs,

> changes, updates,and hardware constraints. Things that are considered peripheral to the abstract description of a neural network—such as the decision of when to stop training the model or the versions of different pieces of software used to perform numerical calculations—factor into the “decipherability” of the actual system. That is why it is a stretch, one with political significance, to presume that computing systems that use neural networks are somehow uniquely “indecipherable”. (KATZ, 2020, p. 124).[36]

Numa dinâmica similar à descrita pelo autor, entendemos que a pauta da explicabilidade, por sustentar a apresentação de ajustes técnicos perante questões éticas (contrapondo a aparente objetividade dos modelos à intervenção humana) resguarda a indústria da IA de discussões mais elementares sobre as tecnologias que desenvolvem. No artigo sobre a ocorrência de pautas éticas nas publicações de IA elaborado por um interlocutor, que mencionamos anteriormente, o viés de máquina é definido como “[...] o processo através do qual preconceitos pessoais de engenheiros de IA podem vazar nos projetos nos quais se envolvem” [37](PRATES, AVELAR, LAMB, 2018, p. 2, tradução minha). Esse conceito de “viés” é cercado da mesma “nebulosidade” (KATZ, 2020) característica da IA, consistindo em mais um empréstimo da neurociência. Num dos materiais do Sprint de introdução à IA do qual participei, o conceito foi descrito resumidamente como o processo em que o cérebro recorre a experiências passadas para a tomada de decisões e, assim, corresponde à reprodução de preconceitos de gênero, raça e idade, principalmente. Os problemas mais consistentes dos vieses, apontados segundo as explicações, é que são inconscientes e podem reproduzir ideias discriminatórias.

A fim de ampliar o conhecimento sobre o problema, foi disponibilizado outro material, divulgado pela ONU Mulheres (2016), sobre vieses inconscientes e equidade de gênero, formulado como guia corporativo. O material se propõe a explicar o conceito de viés pelas abordagens da neurociência e da psicologia. Sinteticamente,

---

[36] “Assim, quando estes comentadores [indústria de especialistas em IA] afirmam que, digamos, os sistemas que utilizam redes neurais são indecifráveis devido ao seu grande número de parâmetros, eles privilegiam a abstração matemática em detrimento da realidade social situada dos sistemas computacionais. Na prática, não lidam com uma rede neural abstrata, mas sim com a sua instanciação como software, que está sujeita a bugs, alterações, atualizações e restrições de hardware. Coisas que são consideradas periféricas à descrição abstrata de uma rede neural – como a decisão de quando parar de treinar o modelo ou as versões de diferentes softwares usados para realizar cálculos numéricos – influenciam a ‘decifrabilidade’ do sistema real. É por isso que é um exagero, com significado político, presumir que os sistemas de computação que usam redes neurais são de alguma forma exclusivamente ‘indecifráveis’” (KATZ, 2020, p. 124, tradução minha).

[37] Do original: "[...]the process by which personal preconceptions of AI engineers can leak into projects in which they are involved."

apresenta uma explicação neurocientífica da cognição através da divisão entre dois sistemas de processamento cerebral: um é inconsciente, reagente e toma decisões rápidas; o outro, é consciente e mais lento, pois considera, julga e controla o primeiro (quanto maior o controle, mais qualidade há na tomada de decisão) (ONU MULHERES, 2018). A explicação, que inclui uma representação visual de como o preconceito se forma no cérebro, nos remete ao cerebralismo (AZIZE, 2010), especificamente no modo como surge a menção à “plasticidade”:

> A eficiência do sistema 2 é essencial para o controle do sistema 1 e para a racionalidade das decisões. Sem essa racionalidade e a capacidade de simulação do futuro - característica do sistema 2 -, nossa tomada de decisão pode ser presa fácil de manipulação do sistema 1 por apelos consumistas, sobrenaturais, etc., e também dos vieses inconscientes.
> Vários estudos comprovam que a plasticidade aumentada, características da infância e da adolescência, pode ser restaurada pelo uso de fármacos ou outras manipulações ambientais. Mudanças do ambiente, por exemplo, podem facilitar a formação de novas conexões cerebrais em adultos. (ONU MULHERES, 2016, p. 15).

O viés, nesse sentido, está associado à diminuição da plasticidade, que determina a perda de controle do sistema 2 sobre o sistema 1. Trata-se de uma distinção entre a racionalidade de um sistema e as operações inconscientes do outro, reflexos do processo evolutivo. O texto prossegue associando a mudança de ambiente e a diversidade no ambiente de trabalho ao aumento da plasticidade. Seguindo o padrão identificado por Azize (2010), em relação à abordagem psicológica do viés, a explicação apenas retornou aos sistemas cerebrais, sobrepondo a atividade neuronal a ela. O documento nos permite contemplar a influência do discurso neurocientífico no estabelecimento da pauta da diversidade (entendida pelo ângulo da composição de grupos) como solução perante os problemas éticos da IA, questão central em campo que discutiremos adiante. Ou seja, o argumento sobre a diversidade trazido evidencia o espelhamento de explicações sobre o funcionamento do cérebro e sobre o funcionamento do ambiente corporativo (AZIZE, 2010).

Outra característica da discussão sobre os vieses revela a performatividade do ethos engenheiro (FORSYTHE, 1993) na área, principalmente nos aspectos relacionados ao entendimento de que o problema pode ser resolvido através de aperfeiçoamentos técnicos. Em campo pude conhecer o trabalho de profissionais que se dedicaram ao problema da explicabilidade, contribuindo para a área da xAI. No que concerne à nomenclatura da disciplina, o padrão das definições instáveis da IA se

mantém, conforme colocou um interlocutor durante uma apresentação em um dos congressos. Ele apontou que não há consenso sobre a terminologia que a caracteriza, pois no geral termos como “explicabilidade” e “transparência” são usados de modo intercambiável (algo que, na prática, se estendeu também para o termo “interpretabilidade”[38]). Todavia, resumiu a definição da área como o “campo de estudos que busca tornar modelos explicáveis para humanos”. Apesar da redundância da definição, podemos inferir comparativamente que enquanto a IA estuda como os humanos se comportam, a xAI é a subárea que estuda o comportamento dos modelos.

A maioria das pesquisas na xAI se dedica à elaboração de métodos para compreensão de modelos caixa-preta, sob a justificativa de que modelos explicáveis aumentam a confiança dos usuários nos produtos das empresas. Por exemplo, profissionais relataram o uso de técnicas que permitem criar representações visuais das áreas que são ativadas nos modelos, como mapas de calor que indicam o que o modelo “viu” e foi relevante para as predições que apresentou. Esses métodos, conforme descritos em campo, permitem identificar ambos os vieses e as fraquezas de modelos. Por exemplo, assisti a uma exposição sobre um caso em que técnicas de visualização desse tipo ajudaram a constatar falhas num modelo de identificação facial, que era aplicado a documentos de identificação pessoal (RG, passaporte, etc.). Com a aplicação das técnicas, os apresentadores descobriram que o modelo não estava operando conforme o esperado porque as features dos documentos (as partes correspondente ao texto) tinham maior peso que as features das faces das pessoas. A partir disso, puderam ajustá-lo para corrigir a falha.

Métodos como esse foram descritos como técnicas para “abrir as caixas-pretas” da IA. Aqui, o paralelo entre o conceito “caixa-preta” no sentido que surge na IA e o significado padrão do conceito nos STS incita uma reflexão. Para Latour, a caixa-preta se refere ao ocultamento dos fatos que influem na construção do conhecimento científico. Uma caixa-preta é o fato científico consolidado, trata-se da “ciência pronta” (LATOUR, 2000). Na interpretação latouriana, as caixas-pretas representam a validação e o ocultamento dos processos de criação da ciência pelo

[38] Nas discussões em campo, não houve diferenciação explícita entre os conceitos citados. Entretanto, consideramos a definição de transparência conforme Lipton (2018). Também levamos em conta a diferenciação funcional entre e "explicabilidade" e "interpretabilidade" de Biran e Cotton (2017). Para os autores, um modelo explicável oferece algum tipo de acesso às decisões que levaram aos resultados que produziu. Já a interpretabilidade se refere mais especificamente ao grau que se pode chegar de compreensão humana das operações dos modelos.

sucesso de empreendimentos científicos. Paralelamente, algo que foi muito associado às demandas sobre explicabilidade, numa das expressões do argumento sobre uma integração entre técnica e social, foi a crítica ao privilégio da acurácia de modelos em relação à ética na IA. Isso, de acordo com diversos profissionais no campo, indica um descompasso entre os objetivos do ML e os objetivos das pessoas em si, ou do "mundo real".

O discurso público sobre as caixas-pretas da IA tem como foco as relações inescrutáveis entre os inputs e outputs dos modelos, mas contrasta com a abordagem dos STS por falhar em considerar a agência humana em todo o processo de aplicação desses modelos. Isso evidencia o caráter fortemente político do foco na opacidade dos produtos da IA (KATZ, 2020). Nesse sentido, há de se considerar o alerta feito por um interlocutor quando o questionei sobre as dificuldades de explicar o seu trabalho para pessoas de fora da área. Ele apontou que a ideia de que os modelos de ML são caixas-pretas é influenciada por "meios midiáticos" e que pessoas sem background em ML tendem a perceber apenas se modelos acertam ou erram. Assim, as razões pelas quais determinados modelos são escolhidos e os parâmetros que influenciam cada um deles não são levadas em conta. Frequentemente, especialistas abordaram o problema questionando o próprio uso da IA em tarefas que poderiam ser abordadas por métodos mais acessíveis.

Apesar do ML protagonizar a discussão, o conceito de "caixa-preta" é mais adequado para tratar especificamente de modelos de Deep Learning. Ou seja, a subárea associada à complexificação de modelos de redes neurais que permitiu a construção de modelos com alta acurácia e não explicáveis. O processamento de grandes volumes de dados de forma distribuída através dos nós das diversas camadas de processamento que compõem a arquitetura dos modelos de Deep Learning os torna, ao mesmo tempo, muito eficientes em suas predições e muito difíceis de interpretar. Não obstante, a repercussão sobre o sucesso de técnicas de Deep Learning impulsiona a corrida das empresas para sua adoção (por vezes sem necessidade e sem o planejamento adequado, conforme descrito por interlocutores em campo). Nesse sentido, em vários momentos foi possível observar a contradição entre "ética" e "automação", ao ouvir profissionais apontarem que a IA deve "dar um passo atrás", que "nem tudo é questão de automatizar", entre outras afirmações do tipo.

Um interlocutor expressou essa ideia pelo caminho inverso, ao argumentar que a IA não está necessariamente retrocedendo, mas sim evoluindo mais rápido que as pessoas. Aqui, façamos um parêntese para o reconhecimento da proliferação de noções de evolução darwinista no discurso empresarial das Big Techs. Tomando o ângulo da priorização da automação em relação à ética, a aplicação “apressada” de novas tecnologias costuma ser justificada como uma necessidade para a garantia da posição das empresas na narrativa do progresso unilinear da IA (CHUN, 2021). Mas retornando à afirmação do interlocutor, ele apontou que, devido a esse atraso das pessoas em relação à IA, as empresas estariam freando a sua evolução. Esse papel do mercado como regulador da IA também surgiu de maneira mais explicitamente contraposta ao conhecimento acadêmico. Em determinado momento, acompanhei um debate sobre vieses em que uma pesquisadora sinalizou a falta de formação para a explicabilidade de modelos na academia como um problema perante a necessidade que profissionais encontram de explicar seus modelos no mercado.

A experiência em empresas, nesse sentido, representaria um aprendizado maior relacionado à explicabilidade. Esse distanciamento entre a academia e o tema da explicabilidade também foi observado na falta de consenso acerca dos conceitos mobilizados na área. Como tratamos, ao falar de modelos transparentes, explicáveis ou interpretáveis, normalmente os especialistas se referiram a essa caraterística que permite aos humanos entender como um modelo funciona e ao acesso à lógica interna de funcionamento dos modelos caixa-preta. Lipton (2018) redigiu um artigo seminal sobre o assunto, criticando o uso auto evidente de termos como esses e advertindo sobre a escassez de teorias críticas na área de ML como parte do problema em sua comunidade. Argumentando especificamente a favor da desconstrução da “interpretabilidade” como um conceito monolítico, o autor identifica usos comuns e contraditórios do termo[39].

Entendemos que o formato de pesquisa predominantemente voltado à indústria na IA, caracterizado como imediatista pelos interlocutores, incide nessa indefinição dos termos. Sem um consenso acerca do próprio significado - e, por consequência,

[39] Para Lipton (2018), há duas propriedades distintas associadas à noção de modelos interpretáveis - a “transparência” e as “explicações post-hoc”. Ele define a transparência, amplamente, como o entendimento do mecanismo através do qual um modelo funciona. Também determina três diferentes tipos de transparência que implicam em diferentes noções de interpretabilidade. Comentando que os humanos não demonstram nenhuma dessas formas de transparência, associa a interpretabilidade das decisões humanas a "explicações post-hoc". Ou seja, as predições são explicadas sem necessariamente elucidarem os mecanismos através dos quais foram produzidas.

dos objetivos - da explicabilidade, as práticas nessa direção estão sujeitas à variação do entendimento dos especialistas sobre elas. Apesar disso, a validade da xAI é garantida na aparente tecnicidade dos seus propósitos. Ou seja, nas discussões da área se sobressai a concepção de conhecimento fundamentada no ethos engenheiro dos praticantes da IA, que permite o distanciamento de debates sobre aspectos estruturais da discriminação social. Os preconceitos são reformulados como vieses, a diversidade como uma métrica numérica e a ética como um problema de otimização. Desse modo, a opacidade dos modelos intercepta debates sobre a opacidade de decisões sobre o que são dados e como usá-los, sobre a validade dos outputs, a necessidade de revisão dos modelos, etc. (KATZ, 2020, p. 124).

Esse reconhecimento sobre a agência humana nessas decisões pode parecer contraditório com as angústias que os praticantes no campo levantaram sobre as dificuldades relacionadas à inescrutabilidade de modelos. Mas, pelo contrário, apresenta outro ângulo sobre elas, que Katz (2020) sintetiza ao recuperar a metáfora de Joseph Weizenbaum[40] sobre sistemas computacionais como burocracias:

> The computer scientist Joseph Weizenbaum, for instance, argued that ordinary computer programs are effectively "theoryless." These programs can be quite large and are often developed by multiple people. There is no algorithm one could write down that fully encapsulates how such a program works in practice. [...]. Rather than seeing a computing system as a realization of theory, which would mean it can be "explained" in algorithmic terms, he argued for seeing it as an intricate bureaucracy. In this bureaucracy, different subsystems, glued together somewhat haphazardly as a product of circumstance, generate outcomes that are subject to disputes over "jurisdiction." This is why programmers, he argued, often "cannot even know the path of decision making" that unfolds in their own programs, "let alone what intermediate or final results" will be produced. Implicit in this argument is the simple observation that every computing system exists in a social envelope. (KATZ, 2020, p. 124).[41]

[40] Renomado cientista da computação alemão-estadunidense, professor do MIT e responsável pelo desenvolvimento do primeiro chatbot na históra, ELIZA. Veio a se tornar um dos críticos da evolução da IA, em relação a fatores como os financiamentos militares (KATZ, 2020) e as noções de inteligência maquínica propagadas pela área (CRAWFORD, 2021).

[41] "O cientista da computação Joseph Weizenbaum, por exemplo, argumentou que os programas de computador comuns são efetivamente 'sem teoria'. Esses programas podem ser bastante grandes e geralmente são desenvolvidos por várias pessoas. Não há algoritmo que se possa escrever que encapsule completamente como tal programa funciona na prática. [...]. Em vez de ver um sistema computacional como a realização de uma teoria, o que significaria que pode ser 'explicado' em termos algorítmicos, ele defendeu vê-lo como uma burocracia intrincada. Nessa burocracia, diferentes subsistemas, colados um tanto ao acaso como produto das circunstâncias, geram resultados que estão sujeitos a disputas sobre 'jurisdição'. É por isso que os programadores, ele argumentou, muitas vezes 'não podem nem saber o caminho da tomada de decisão' que se desdobra em seus próprios programas, 'muito menos quais resultados intermediários ou finais' serão produzidos. Implícita nesse argumento está a simples observação de que todo sistema de computação existe em um envelope social" (KATZ, 2020, p. 124, tradução minha).

A metáfora nos ajuda a conceber uma abordagem da explicabilidade situada em realidades sociais, em que profissionais não lidam com redes neurais abstratas, mas com as "suas instanciações como software" (KATZ, 2020, p. 124). Também torna mais palpáveis obstáculos relativos à explicabilidade, considerando a magnitude das operações que envolvem a elaboração e aplicação de modelos. Trata-se de uma perspectiva a favor da contextualização nas discussões sobre ética e IA. Como vimos, essas discussões são incipientes, norteadas por interesses de mercado e repletas de definições instáveis. Além disso, pelo foco na questão dos vieses (preconceitos humanos que "vazam" para as máquinas) e em métodos para "corrigi-los", essas discussões propagam uma lógica de externalização da técnica (INGOLD, 2000), que separa máquina-humano entre os polos objetivo-subjetivo, ético-antiético, respectivamente.

4.2 MODELO DE DECISÃO DO ML

Vimos que o ângulo da explicabilidade costuma demonstrar uma falta de compreensão holística dos processos de desenvolvimento de modelos na discussão sobre ética na IA. Conforme apontamos, esse aspecto se refere à magnitude das camadas de trabalho e de operações humanas na elaboração de programas no contexto da indústria de tecnologia atual. Mas ele também se aplica à complexificação de modelos representada pela IA fraca, que implica o uso de ferramentas e bibliotecas "prontas" no desenvolvimento de novos programas[42]. Ou seja, há um caráter operacional relevante aqui, de que o ML é orientado a tarefas (BOWKER, BECHMANN, 2019). Tendo isso em vista, dificulta-se a abordagem de aspectos políticos que historicamente influenciaram não apenas no uso de "dados sujos" (ou enviesados) que protagonizam o debate, mas também nas próprias arquiteturas de sistemas algorítmicos (CHUN, 2021).

Em campo, como parte do argumento sobre a integração entre técnica e social que perpetuou as discussões sobre ética, profissionais apontaram o problema da falta

[42] O TensorFlow, plataforma de código aberto para ML criada pela Google, atualmente é referência para o desenvolvimento de modelos. Em campo, foi frequentemente citada nas descrições dos processos de elaboração e treino dos modelos.

de familiaridade da área com questões sociais e com teorias das ciências humanas. Entretanto, de acordo com o que vimos até aqui, o problema não consiste na falta de concepções sobre o social no desenvolvimento de aplicações de IA. Pelo contrário, discutimos como elas mobilizam noções de natureza, de conhecimento e de pessoa particulares, que impactam o modo como os modelos são construídos. Os aspectos técnicos da arquitetura de sistemas algorítmicos envolvem pressuposições sobre relações sociais que precedem a sua criação, conforme Chun (2016, 2021) demonstrou ao analisar como a “homofilia” se tornou um princípio epistemológico das ciências de rede, e as raízes históricas do princípio da correlação em práticas de eugenia.

Desse modo, uma reformulação mais produtiva dos problemas éticos relacionados à IA envolve considerar os processos de externalização da técnica (INGOLD, 2000) subjacentes a argumentos sobre a objetividade de modelos. Quando a discussão sobre ética em IA pauta a integração entre a técnica e o social a divisão costuma ser enfatizada, por mais artificial que seja. Nesse sentido, Chun (2021) demonstra como o problema não é incorporar teoria social à IA, mas considerar o modo como teorias problemáticas sobre o social tiveram papel histórico no desenvolvimento da IA. Tendo isso em vista, nos interessa abordar axiomas que guiam a construção de conhecimento no mainstream da IA atualmente, ou seja, em modelos de ML. Trata-se de pensar o problema ético pelo ângulo das produções de causa e efeito através de correlações nos modelos de decisão do ML.

Numa das primeiras vezes em que tive contato com essa problematização, ela surgiu de maneira sutil, durante um evento em que desenvolvedores apresentaram um modelo treinado para prever o resultado do vencedor de um reality show. Após os apresentadores percorrerem todo o processo de coleta de dados e treino do modelo, anunciaram que o modelo previu corretamente qual dos três finalistas ganhou o programa. Na rodada de dúvidas, uma pessoa colocou a dúvida: “como saber se o resultado correto de uma previsão não é apenas uma grande coincidência?”. A resposta dada pelos apresentadores foi que a aplicação do modelo numa dinâmica de eliminação do reality anterior à final os permitiu identificar uma alta correlação similar entre as variáveis, o que atestaria a validade das suas predições.

Entendemos que uma das grandes vantagens e razões para o hype acerca de técnicas de ML e DL é a capacidade preditiva desses modelos. Entretanto, essas predições são reconhecidamente frágeis. A ineficiência dos modelos perante a

contingência, nesse sentido, foi frequentemente apontada como um dos principais obstáculos envolvidos na automação. Podemos interpretar esse problema por duas abordagens diferentes: uma delas se refere especificamente ao modo como a contingência induz a dilemas morais. O tema é materializado na famosa discussão sobre carros autônomos e como devem agir em situações que exigem a escolha entre "dois males". Um exemplo é a situação hipotética em que o veículo precisa decidir entre atropelar um ciclista ou sacrificar a pessoa dentro do carro (BONNEFON, SHARIFF, RAHWAN, 2016). Independente de qual seja o output do modelo, ambas as possibilidades configuram problemas morais.

A formulação do problema nos remete à descrição de Erickson et al. (2013) sobre a reformulação da razão iluminista numa racionalidade pautada em regras. Os autores descrevem que a razão abrange tanto a complexidade quanto a contingência, diferentemente dessa racionalidade algorítmica pautada na mecanização das regras e no processo de exclusão do julgamento humano de processos de decisão. Nesse sentido, o dilema descrito é apenas uma das infinitas possibilidades de situações em que a universalidade de regras se mostra irrealizável, situações que no "mundo real" progridem de acordo com diretrizes coletivas e morais.

Outra questão que aponta para a dificuldade de processar computacionalmente a contingência se refere especificamente à relação entre as predições de modelos e o modo como são treinados. Ilustrando isso, durante uma discussão sobre as dificuldades envolvidas em usar ML para fazer predições, um desenvolvedor apontou que a greve dos caminhoneiros [43]"quebrou" os algoritmos preditivos de diversas empresas em 2018. Ou seja, os algoritmos de ML, por serem treinados com dados do passado, têm uma performance indesejada perante situações que não se encaixam nos padrões identificados nesses dados (no que é mais provável ocorrer de acordo com eles). Aqui, o problema é bem sintetizado por Chun (2021) quando afirma que as predições dos algoritmos "acertam" quando produzem correlações em que o presente e o futuro coincidem com um passado imutável e "altamente curado" (CHUN, 2021, p. 52).

Nesse ponto, a noção de "problemas do mundo real" foi fundamental. Através da expressão, os especialistas normalmente se referiam à complexidade de

[43] Em maio de 2018, caminhoneiros entraram em greve no Brasil como protesto ao aumento do valor dos combustíveis. A greve durou 10 dias, interrompendo de imediato o abastecimento e a produção em diversos setores.

fenômenos vivos e a dificuldade de codificá-los, de traduzi-los para as máquinas. Sobre isso, um cientista de dados me relatou:

> Descrevo “problemas do mundo real” como problemas que diferem de um problema proposto em um curso de Machine Learning, o qual possui dados prontos para serem aplicados os modelos de ML. Nos problemas reais, os dados podem estar desbalanceados, enviesados, podem faltar informações ou mesmo podem apresentar poucas amostras, e produzi-las poderia ser muito trabalhoso, custoso ou mesmo impossível. E ainda assim, deve ser avaliado a forma a qual o ML será aplicado para solucionar este problema, ou mesmo se este é solucionável com ML. Assim, a tradução de um problema real para a codificação é um dos maiores desafios de um cientista de dados. (Trecho da entrevista realizada em 31 de maio de 2021).

Nota-se que ele apontou para a dificuldade de traduzir esses problemas que afetam diretamente as vidas das pessoas, e a natureza de modo geral, mas também ponderou sobre os fatores que envolvem solucioná-los através de ML. Nesse sentido, em outras ocasiões durante a pesquisa, o conceito se tornou relevante quando desenvolvedores caracterizaram sua prática como um “machine learning de mundo real”. Assim, se colocavam numa posição contrária a um tipo de ML direcionado exclusivamente à elaboração de novos métodos, à evolução técnica da área que não necessariamente tem aplicação em melhorias nas vidas das pessoas. Esse posicionamento complexifica as funções desses profissionais, como Lipton (2018) demonstrou ao tratar da dificuldade de tradução dos problemas do mundo real em relação às demandas por interpretabilidade:

> Often, real-world objectives are difficult to encode as simple mathematical functions. Otherwise, they might just be incorporated into the objective function and the problem would be considered solved. For example, an algorithm for making hiring decisions should simultaneously optimize productivity, ethics, and legality. But how would you go about writing a function that measures ethics or legality? The problem can also arise when you desire robustness to changes in the dynamics between the training and deployment environments (LIPTON, 2018, p. 7).[44]

[44] “Muitas vezes, os objetivos do mundo real são difíceis de codificar como funções matemáticas simples. Caso contrário, eles poderiam simplesmente ser incorporados à função objetivo e o problema seria considerado resolvido. Por exemplo, um algoritmo para tomar decisões de contratação deve otimizar simultaneamente a produtividade, a ética e a legalidade. Mas como você escreveria uma função que mede a ética ou a legalidade? O problema também pode surgir quando se deseja robustez às mudanças na dinâmica entre os ambientes de treinamento e implantação” (LIPTON, 2018, p. 7, tradução minha).

Portanto, o problema ético da IA passa por esse entendimento sobre o descompasso entre os objetivos dos modelos de ML e os objetivos “do mundo real”. Nesse sentido, autores apontam que a meta dos modelos é a “predição efetiva” e não a interpretabilidade, nem a resolução de problemas sociais (MITTELSTADT, B. D. et al, 2016; ELISH, BOYD, 2017). Elish e Boyd (2017) argumentam que, para atingir essa meta, a formalização computacional de problemas “do mundo real” requer o apagamento da fluidez e das nuances de relações sociais, estabelecendo fronteiras que não necessariamente existem na prática. Esse aspecto aponta para a efetividade do discurso do Big Data, que incita volumosos esforços humanos e maquínicos na mineração e conexão de dados em rede. A partir disso, o paradigma conexionista se manifesta nas práticas dos modelos de ML que estabelecem correlações entre “pedaços de dados” sobre indivíduos e grupos de indivíduos (BOYD, CRAWFORD, 2011).

Boyd e Crawford (2011) descreveram como essa lógica de funcionamento do Big Data (e por consequência do ML) promove a ideia que “os números falam por si mesmos”. Isso fundamenta um entendimento sobre a dependência entre conhecimento e volume de dados. As autoras também associam as práticas do Big Data a uma concepção de conhecimento que dispensa interpretação, supostamente independente de teorias sobre o comportamento humano. Desse modo, descrevem como o engajamento da área com “atos de ciência social” é preservado de críticas pelo uso da pesquisa quantitativa, à qual associam a “produção de fatos”. Nessa dinâmica de reivindicação de objetividade, a pesquisa qualitativa, por sua vez, fica restrita ao ofício da “interpretação de histórias” (BOYD, CRAWFORD, 2011, p. 5).

Desde que as autoras redigiram esses argumentos, é evidente que a crença no caráter objetivo dos números em práticas na IA já foi muito contestada - vide a discussão sobre vieses. Apesar disso, já discutimos aqui a falta de elaboração teórica crítica no ML como um problema levantado na comunidade de especialistas (LIPTON, 2018). Associamos a isso o problema do "imediatismo" que interlocutores apontaram como característica da indústria de tecnologia, responsável pela maior parte das pesquisas da IA. Nesse sentido, as inferências da pesquisa de campo, como a divisão entre um tipo de IA acadêmica e a IA de mercado, assim como os modelos que cada uma desenvolve (de compreensão e de predição, respectivamente) convergem com as ideias de Boyd e Crawford (2011) sobre as mudanças que o Big Data impulsiona

no nosso entendimento sobre a natureza do conhecimento, da pesquisa e dos procedimentos atrelados a ela.

Isto é, entendemos que o sentido da pesquisa no Big Data rompe com os procedimentos de garantia da objetividade da ciência tradicional. Ao mesmo tempo, é um tipo de pesquisa que postula ser mais objetivo que pesquisas científicas, justamente por concretizar o ideal de exclusão da intervenção humana através do apelo aos dados (ou números) como autoexplicativos. Ou seja, se não há intervenção dos humanos na produção de conhecimento, não há necessidade dos procedimentos de verificação científica e controle da mesma. “O dado é uma extensão da nossa mente”, ouvi um interlocutor declarar em uma apresentação. Se tomarmos como verdade a afirmação, a mente, assim como todos os outros aspectos da natureza, pode ser quantificada e portanto conhecida.

Kitchin (2014) resume a mudança que a epistemologia empirista do Big Data representa perante a pesquisa:

> . Big Data can capture a whole domain and provide full resolution;
> . there is no need for a priori theory, models or hypotheses;
> . through the application of agnostic data analytics the data can speak for themselves free of human bias or framing, and any patterns and relationships within Big Data are inherently meaningful and truthful;
> . meaning transcends context or domain-specific knowledge, thus can be interpreted by anyone who can decode a statistic or data visualization. (KITCHIN, 2014, p. 4).[45]

Nota-se que o empirismo é o principal pilar da construção de objetividade no Big Data. Entretanto, existe um problema ético para além dos procedimentos associados a essa lógica, concernente ao acesso. Ou seja, a “divisão digital” entre os que têm e os que não têm os meios para a pesquisa com dados em grande escala. O Big Data é inerentemente empresarial, são as empresas de tecnologia que gerem redes sociais que normalmente têm os recursos para coletar, analisar e também determinar quem pode acessar grandes volumes de dados (BOYD, CRAWFORD, 2011). Como essas análises se revertem em produtos, interessa a elas permanecer

[45] “. O Big Data pode capturar um domínio inteiro e fornecer uma resolução total;
. não há necessidade de teoria, modelos ou hipóteses a priori;
. por meio da aplicação de análise de dados agnóstica, os dados podem falar por si mesmos, livres de preconceitos ou vieses humanos, e quaisquer padrões e relacionamentos dentro do Big Data são inerentemente significativos e verdadeiros;
. o significado transcende o contexto ou o conhecimento específico do domínio, portanto, pode ser interpretado por qualquer pessoa que possa decodificar uma estatística ou visualização de dados” (KITCHIN, 2014, p. 4, tradução minha).

às margens dos procedimentos científicos (e dos comitês de ética) de validação do conhecimento. Uma espécie de “revisão por pares”, por exemplo, é freada pelo sigilo empresarial. A dinâmica mercadológica da produção de conhecimento, nesse sentido, assegura a opacidade dos procedimentos de pesquisa adotados no interior de empresas.

As empresas, como apontamos, se resguardam de críticas através do empirismo que postula os números como signos da objetividade e do caráter técnico que associam aos seus propósitos. Crawford (2021) ilustrou esse aspecto ao descrever a cronologia da narrativa padrão sobre vieses. Sinteticamente, a autora descreve que ela começa com a exposição jornalística de efeitos discriminatórios produzidos por algum determinado modelo automatizado. Isso produz um momento de intensa repercussão na esfera pública e de cobranças em relação à conduta da empresa responsável pelo sistema em questão. A empresa, então, desativa o sistema ou relata a realização de reparos técnicos, que supostamente garantiriam a eliminação dos vieses que induziram ao comportamento criticado. Entretanto, esses reparos permanecem protegidos por sigilo empresarial e a atenção pública diminui (CRAWFORD, 2021, p. 129). Desse modo, evidencia-se o ethos engenheiro (FORSYTHE, 1993) - referente ao apelo de questões técnicas que prevalece sobre questões sociais - na ênfase na discussão sobre “dados sujos” e falhas no design de sistemas algoritmos, que desvia a atenção de questionamentos mais fundamentais sobre os propósitos da automação.

Nesse sentido, o reconhecimento da interferência humana em aplicações automatizadas, impulsionado no debate sobre vieses e explicabilidade, não necessariamente problematiza a automação. Pelo contrário, ele é reestruturado como justificativa para expandi-la, conforme denotam os esforços para a reformulação de problemas sociais (ou problemas do mundo real) como problemas computacionais. Inclui-se nisso, também, os investimentos na construção de modelos para explicar outros modelos na xAI. Entendemos que essa é uma característica constitutiva da indústria de tecnologia, cujo método de captação de recursos envolve menos as capacidades realistas de seus produtos e mais a construção do hype acerca do que seria possível fazer com eles (ELISH, BOYD, 2017). Essa dinâmica do hype, como apontamos, foi manifestada através de angústias que alguns desenvolvedores expressaram acerca da aplicação precipitada de tecnologias de IA, especialmente as de Deep Learning.

Porém, esse aspecto também incide sobre outras questões levantadas em campo, como a falta de reflexividade em relação aos propósitos da automação. Ao priorizarem a automação em relação à ética, ouvi uma engenheira de software argumentar, as empresas deixam abertas questões sobre quando automatizar e por quê. É nesse sentido que teóricos têm buscado deslocar o foco nos futuros possíveis da IA para o passado da disciplina, que permite investigar os interesses políticos historicamente associados aos seus projetos (BOWKER, BECHMANN, 2019; CHUN, 2021; KATZ, 2020). Um esforço notório nessa direção é o de Chun (2021), que dedicou sua atenção ao modo como a correlação surgiu no centro da "promessa" do Big Data. Para ela, o surgimento da correlação como um axioma central na IA aponta para uma mudança no sentido do conhecimento. Ou seja, ela identifica uma preponderância das correlações em relação às causas no modo como explicamos os fenômenos atualmente.

Chun (2021) associa isso à alteração no estatuto do conhecimento, que deixa de se referir ao entendimento do passado e passa a consistir na predição do futuro de maneira explícita no ML. No entanto, a predição do futuro tem como medida a correspondência com dados de um passado seletivo e imutável. A autora fundamenta esse argumento expondo as raízes históricas da correlação na eugenia, comparando o caráter revolucionário associado a ela no Big Data no século XXI com o que teve na visão do mundo eugenista no século XX. Ela descreve como Karl Pearson, matemático cujo trabalho foi determinante no estabelecimento da correlação e da regressão linear nas práticas da estatística no século passado, chegou a essas técnicas através de tentativas de determinação da hereditariedade (CHUN, 2021, p. 51).

Para Chun (2021), o principal aspecto em que a correlação une Big Data e práticas eugenistas é o modo como manipulam uma visão do mundo como um laboratório. De acordo com ela, ambos apontam para o uso de práticas de vigilância através do levantamento de dados e experimento com humanos (especialmente de pessoas em situação de vulnerabilidade social). Essa característica remete ao que Marres e Stark (2020) descreveram como a criação de "ambientes de testagem total" no contexto da digitalização. Para os autores, o século XXI representa uma transposição do "teste" do laboratório para ambientes sociais, que indica também uma mudança no próprio significado do termo no contexto da digitalização. Ou seja, diferentemente daqueles realizados em ambientes controlados, os testes ubíquos (e

frequentemente invisíveis) de hoje não representam apenas a testagem como uma forma de conhecimento das sociedades, mas também incidem sobre o modo como são governadas e organizadas.

No caso do Big Data, os autores apontam que o objetivo, ao colocar o comportamento humano sob teste, é deliberadamente modificá-lo. Conforme o que discutimos, podemos concluir que os produtos dos testes no contexto do ML são as correlações. Similarmente ao que argumentam Marres e Stark (2020) sobre os testes, Chun (2020) compreende as correlações para além do seu caráter analítico, no modo como produzem mudanças nas configurações sociais. Para ela, as ciências de rede concretizaram o que era aspiração para os eugenistas (a reprodução entre semelhantes) ao tornar a homofilia um axioma. Ou seja, a evolução das tecnologias computacionais foi guiada pelas ciências de rede através desse pressuposto, de que as pessoas naturalmente são atraídas para os seus semelhantes, para prever e influenciar o comportamento humano (CHUN, 2021, p. 72).

> At the most basic level, network science captures—analyzes, articulates, imposes, instrumentalizes, and elaborates—connection. Coupling graph theory with game theory, it models human interactions in terms of costs, benefits, and efficiency. It clarifies global phenomena, from capitalism to "contagion," by reducing the world to individual "nodes" and "edges." At the same time, it is nonnormative: it does not presume that aggregate action stems from identical mass actions. Further, through notions such as "social capital," it both explains and can help justify inequalities within supposedly democratizing social networks.(CHUN, 2021, p. 86). [46]

O nosso principal ponto de interesse no argumento da autora, portanto, é que ele abrange o aspecto das caraterísticas dos dados que são coletados e analisados por modelos de ML, tão enfatizado nas discussões sobre vieses, mas também expande o foco para os laços históricos das técnicas que são utilizadas na construção deles. Assim como a autora explicita as raízes da correlação na eugenia, ela expõe as raízes da homofilia na naturalização da segregação racial em estudos de abrigos públicos nos EUA (KURGAN et al., 2019; CHUN, 2021), evidenciando os interesses

[46] "No nível mais básico, a ciência da rede captura – analisa, articula, impõe, instrumentaliza e elabora – a conexão. Acoplando a teoria dos grafos com a teoria dos jogos, ela modela as interações humanas em termos de custos, benefícios e eficiência. Esclarece fenômenos globais, do capitalismo ao 'contágio', reduzindo o mundo a 'nós' e 'arestas' individuais. Ao mesmo tempo, é não normativa: não pressupõe que a ação agregada decorre de ações idênticas em massa. Além disso, por meio de noções como 'capital social', ela explica e pode ajudar a justificar desigualdades dentro de redes sociais supostamente democratizantes" (CHUN, 2021, p. 86, tradução minha).

políticos na origem dessas noções que hoje guiam o desenvolvimento de modelos computacionais.

Desse modo, aborda as práticas que configuram a IA atualmente através das suas associações com concepções de mundo da biologia, das ciências sociais, da economia, etc., muitas delas controversas nos seus contextos científicos específicos. O argumento de Chun (2021), portanto, contesta a ideia de uma falta de integração entre técnica e concepções do social na epistemologia das ciências de rede. Além disso, nos ajuda a examinar as alegações de objetividade e defesas do empirismo acerca dos modelos de ML, propondo a investigação do passado como caminho para a compreensão da performatividade das predições do futuro que esses modelos realizam.

4.3 DIVERSIDADE COMO PROBLEMA ESTATÍSTICO

Com o intuito de encaminhar esta dissertação para o fim, esta seção foi reservada para uma temática que emanou do campo como solução para os problemas éticos associados à IA dos quais vimos tratando: a diversidade. Trata-se da diversidade que surgiu como demanda na esfera pública em relação às características demográficas dos especialistas na área, assim como ao uso de conjuntos de dados não representativos (enviesados) da população. A diversidade, nesse sentido, é associada ao que Katz (2020) identifica como o estabelecimento de uma "política de representação" na IA a partir dos anos 2010. Ou seja, as reivindicações acerca da correção da sub-representação de grupos em dados de treino de modelos e nas empresas de tecnologia - principalmente mulheres e pessoas não brancas - se estabeleceram como resposta aos problemas éticos associados aos modelos de ML.

Conforme discutimos, os contínuos relatos de casos de discriminação algorítmica que ocuparam a esfera pública nos últimos anos evidenciaram como algoritmos podem rapidamente tornar-se máquinas de reprodução de desigualdades sociais. Mesmo com a exclusão dos marcadores sociais da diferença dos dados, que empresas usaram para tentar justificar que seus algoritmos não veem raça, gênero e classe, algoritmos são capazes de segregar grupos através de "proxies"[47]. Chun

[47] Entende-se como algo que substitui algum outro ente. Chun usa o termo neste sentido: "Highly correlated variables are thus considered to be "proxies" of each other: by tracking one variable, you can capture the other" (CHUN, 2021, p. 54). Explicando a estratégia de microtargeting da Cambridge

(2021) explora casos concretos que comprovam esse argumento e descreve como informações de CEP, por exemplo, podem ser indicadores de raça, mesmo que a categoria não seja explícita nos dados. Desse modo, correlações entre a região em que indivíduos moram e a probabilidade de quitarem empréstimos, que determinam quem tem acesso ao crédito nos bancos, funcionam a favor da discriminação racial (CHUN, 2021, p. 121).Tratam-se de casos que ameaçam a percepção sobre a neutralidade e objetividade dos modelos de IA, levantando questionamentos sobre quem tem o poder de decisão nas vidas de pessoas, especialmente as mais vulneráveis (EUBANKS, 2018).

Esses episódios foram acompanhados por outro tipo de denúncia comum nos últimos anos, em relação ao aspecto da sub-representação de determinados grupos nos dados de treino, que implica no seu não reconhecimento. Revelações sobre modelos de reconhecimento facial que funcionavam apenas com rostos brancos (CODED, 2020) direcionaram as discussões para as restritivas características demográficas da indústria de tecnologia. Ou seja, o protagonismo de homens brancos de classe média no setor (na academia e na indústria) (HELMREICH, 1998; KATZ, 2020; FÓRUM ECONÔMICO MUNDIAL, 2021) passou a ser criticado pela associação entre a falta de diversidade nos dados e a falta de diversidade em meio aos especialistas na área[48]. A partir disso, a apropriação da pauta pelo discurso empresarial funciona a favor da aparente objetividade da IA, estimulando o entendimento da diversidade demográfica de profissionais como garantia contra o "vazamento" de preconceitos humanos para os modelos. Entretanto, a apresentação da diversidade como valor não necessariamente representa compromissos concretos das organizações em relação às desigualdades sociais, ao mesmo tempo que permite a manutenção da sua imagem pública e sustenta o obscurecimento da discussão sobre os axiomas da área, que norteiam a produção de conhecimento e manipulação do comportamento humano nela.

Analytica, a autora descreve como um like de um usuário do Facebook poderia revelar uma categoria de identidade sua. Em meio a outros exemplos, cita o trabalho de pesquisadores que, explicando as técnicas de segmentação, apontaram que curtir "Britney Spears" na página é uma ação extremamente indicativa de "homossexualidade masculina".

[48] Entretanto, advertimos que no âmbito da IA a falta de diversidade demográfica normalmente se refere à categoria dos especialistas ocupados no desenvolvimento de modelos, ao trabalho formal. Os ghost workers, força-tarefa de trabalhadores informais predominantemente de países do Sul Global e que são peças fundamentais do desenvolvimento tecnológico, raramente são considerados nessas discussões (GRAY, SURI, 2019).

Posto isso, em campo a questão da diversidade surgiu direta ou indiretamente em todos os debates sobre efeitos da tecnologia que acompanhei, fazendo referência à composição dos quadros de funcionários de empresas de tecnologia, principalmente. A questão tornou-se explícita nas diversas comunidades, iniciativas privadas e organizações sociais segmentadas (com foco em gênero e raça) que conheci através de interlocutores. Como mencionado anteriormente, tive contato principalmente com comunidades voltadas às mulheres na tecnologia e, enquanto todas elas surgiram recentemente (não identifiquei nenhuma estabelecida antes de 2015[49]), as especificamente direcionadas à IA são ainda mais novas, surgindo a partir de 2020. Nesses espaços, a diversidade foi manifestada como meta através da realização de atividades de capacitação técnica e debates em prol da inclusão e permanência de mulheres na tecnologia.

Tanto nos eventos que essas comunidades realizam, como nas discussões em grupos em redes sociais, pude acompanhar mulheres compartilhando experiências - pessoais e técnicas - de rotinas nas quais são extrema minoria, e as estratégias que mobilizam perante isso. “Eu fui a única mulher na minha turma de graduação”, “só tinha eu e mais uma mulher na minha turma” foram algumas das frases que ouvi de profissionais, em meio a outras similares. A partir de relatos do tipo, elas justificaram seus esforços para alterar uma realidade de disparidade de gênero que é latente no setor, conforme é evidente nos dados: de acordo com a pesquisa da Brasscom (2021), apesar de as mulheres serem maioria nas universidades, são apenas 14,8% dos cursos de Tecnologia da Informação e Comunicação no Brasil. Além disso, segundo o relatório do Fórum Econômico Mundial (2021), mulheres são cerca de 32% dos profissionais em Dados e IA no mundo.

No debate sobre o assunto, as especialistas manifestam percepções sobre as causas dessa disparidade, enfatizando o aspecto da representatividade. Na seção “sobre nós” de uma dessas comunidades voltada ao gênero, consta: “A imagem de um programador é sempre masculina, branca e com ares de gênio. **E é muito difícil se imaginar fazendo algo quando ninguém como você está fazendo**.” (PROGRAMARIA, s.d., n.p. grifo do autor). Com isso, apontam que as dificuldades da

[49] São exemplos: WoMakersCode (ONG), Programaria (iniciativa privada), Elas programam (iniciativa privada), {reprograma} (organização sem fins lucrativos), PretaLab (iniciativa da ONG Olabi), Cloud Girls (Meetup). Além desses, há grupos em redes sociais voltados à divulgaçao de eventos, vagas e discussões para mulheres interessadas em IA.

diminuição do “gap” de gênero na tecnologia têm origem antes mesmo da introdução das mulheres na área, pois o estereótipo associado à figura do programador desestimula o interesse delas na área. Entretanto, sinalizam no texto, em sintonia com relatos de profissionais da IA, que os obstáculos associados à “hostilidade” do setor perante as mulheres afetam inclusive carreiras estabelecidas, dificultando a permanência delas na área. A diversidade, nesse sentido, não se trata apenas do escudo levantado pelas Big Techs diante das críticas aos efeitos de seus modelos nas vidas das pessoas, mas também responde a angústias manifestadas por profissionais que são minoria no campo.

Ainda assim, vale adotar uma leitura crítica do estabelecimento dessa pauta como parte do discurso corporativo. Das que conheci em campo, todas as iniciativas com foco em igualdade de gênero que contam com alguma projeção em meio aos profissionais do setor de tecnologia são apoiadas por, ou realizam parcerias com, Big Techs[50]. Além disso, já indicamos como essas empresas estimulam e recompensam a participação ativa de profissionais nessas comunidades[51]. Em sintonia com o argumento de Katz (2021), nota-se o protagonismo dessas empresas na definição da agenda da “IA ética”. Tendo isso em vista, podemos questionar quais as vantagens de mercado que a apresentação da diversidade como valor oferece às organizações. Um dos ângulos da questão é manifestado no site da comunidade que referenciamos e configura-se, de modo pragmático, como uma questão de mercado. Trata-se da falta de mão de obra qualificada para o preenchimento das vagas projetadas para o setor tecnológico nos próximos anos, que é apontada pelas programadoras como um dos principais efeitos da baixa entrada de mulheres na área.

Outro ângulo, mais próximo das teorias críticas sobre as associações dos modelos de IA com visões de mundo neoliberais, é que nos discursos das empresas a diversidade permite criar narrativas que individualizam os problemas levantados na esfera pública e melhoram a sua imagem pública (KATZ, 2020). Assim, as empresas se resguardam da discussão sobre problemas estruturais de desigualdade social e do estabelecimento de compromissos explícitos com a resolução de práticas discriminatórias que seus próprios produtos produzem. Esse aspecto atravessa tanto

[50] Isso se estende também para colaborações entre universidades e Big Techs, que configura parte da crítica de Katz (2020) sobre a expansão da agenda dessas empresas. No Brasil, o principal centro de IA, o C4AI, é uma parceria entre a Universidade de São Paulo, a FAPESP e a IBM.
[51] Vide essa participação como critério para distribuição de certificados como o Google Developer Expert.

o discurso corporativo como as deliberações dos especialistas, o que foi possível visualizar na proliferação das discussões sobre “empoderamento feminino” em campo, enquanto as sobre “machismo” foram praticamente inexistentes.

No mesmo sentido, podemos interpretar que a reformulação do “preconceito” em “viés” que discutimos, com base no vocabulário neurocientífico, permite às empresas se esquivarem também da discussão sobre racismo. Predominam as abordagens individualizadas da diversidade, em que as “políticas de representação” (KATZ, 2020) espelham a formulação técnica de problemas de discriminação algorítmica: ambos são interpretados como aspectos de falhas técnicas. A diversidade, nesse sentido, é configurada como um ajuste estatístico na composição das equipes de profissionais na tecnologia. Ou seja, do mesmo modo como o ethos engenheiro (FORSYTHE, 1993) que fundamenta a concepção de conhecimento na IA se manifesta na tecnicidade pautada pelas discussões sobre explicabilidade, ele influencia a abordagem da desigualdade como uma questão de “paridade quantitativa” (CRAWFORD, 2021).

No modo como atua na produção da objetividade na IA, novamente nota-se um aspecto convergente com a “falsificação epistêmica” difundida na IA, sobre a independência entre os modelos e as condições para a sua existência criadas pelos desenvolvedores em contextos particulares (KATZ, 2020). Conforme exploramos, dado que o “fator humano” é visto como prejudicial à reivindicação de objetividade na área, o distanciamento do reconhecimento de fatores estruturais que incidem sobre a produção de discriminação através de seus sistemas beneficia a IA. Desse modo, o campo se resguarda do reconhecimento da intervenção humana e se beneficia da posição periférica que as discussões sobre os propósitos do uso de determinadas tecnologias adquirem (KATZ, 2020). A partir disso, entendemos que as demandas por diversidade emanam das vidas de profissionais do campo e das manifestações sobre os efeitos de tecnologias nas vidas dos usuários. Porém, a apropriação que o discurso empresarial faz delas implica no uso do termo que Ahmed (2012) qualificaria como “não performativo”.

A autora, invertendo a noção de “performatividade” de Judith Butler, usa o conceito para se referir ao “discurso que não produz os efeitos que ele nomeia” (AHMED, 2021, p. 107). Apesar do foco dela ser em instituições de ensino superior, a sua obra aponta para as críticas extensamente desenvolvidas em estudos de interseccionalidade sobre alegações institucionais de inclusão. Assim, nos ajuda a

pensar como a diversidade é mobilizada no discurso da IA para sustentar a neutralidade de seus modelos, sem se constituir propriamente num comprometimento político da área, o que reflete diretamente nas vidas das pessoas que "incorporam" a diversidade. É um aspecto que podemos visualizar, por exemplo, no contraste entre a popularidade de discursos sobre empoderamento feminino e dados que demonstram que a participação profissional de mulheres em áreas da tecnologia - especialmente aquelas que requerem habilidades técnicas "disruptivas[52]" - teve crescimento ínfimo e, na área de Dados e IA, chegou a diminuir desde 2018 (FÓRUM ECONÔMICO MUNDIAL, 2021, p. 6).

Conforme Ahmed (2012) descreve, as instituições se beneficiam da apresentação da diversidade como valor, como parte de sua "missão". Entretanto, as "métricas quantitativas" de diversidade não resolvem os problemas de desigualdade. Como descrito pelas mulheres da área de tecnologia em campo, o problema não se resume à inclusão de minorias na área, mas também envolve a permanência. Mulheres não só adentram menos cursos na área de tecnologia, também encontram dificuldades que frequentemente as fazem sair de suas profissões no ramo. Isso nos leva a entender que o problema não se resume apenas à paridade quantitativa, mas também abrange as dificuldades que profissionais encontram para se adequar à cultura específica da IA, que, como discutimos, reflete valores particulares de uma composição demográfica muito restrita (HELMREICH, 1993). Nesse sentido, para a maioria das pessoas a compatibilização com os valores da "economia moral" que caracteriza as práticas da IA é um desafio.

Em vários sentidos, a diversidade surgiu como a pauta diante dos problemas associados à reivindicação de objetividade na IA, conforme os três ângulos pelos quais escolhemos abordá-la nesta pesquisa. A diversidade é uma demanda do mercado, tendo em vista que a indústria necessita de mão de obra e que as empresas têm como objetivo aumentar a confiança dos usuários em seus produtos e torná-los desejáveis ao maior número de pessoas. É um requisito também para a manutenção das associações científicas da área, uma vez que os pressupostos e noções sobre a natureza que informam o desenvolvimento dos modelos são validados pela sua eficácia. Ou seja, quando seus modelos são inadequados ou prejudiciais para a maior parte dos seus usuários, dificulta-se o obscurecimento da influência do "fator humano"

[52] Termo popularmente utilizado em referência às habilidades de inovação em relação aos padrões técnicos do desenvolvimento tecnológico.

nos processos da IA e coloca-se um empecilho para as suas ambições totalizantes (KATZ, 2020). Além disso, vimos como a diversidade também tem lugar na explicação neurocientífica do viés, que exemplifica a individualização de problemas de desigualdade social que permeia a discussão ética no discurso dos especialistas da IA.

A partir disso, podemos sintetizar os argumentos trazidos sobre o modo como a ética incide na IA, conforme os aspectos da prática dos especialistas que surgem na temática da diversidade. Entendemos, conforme a caracterização de interlocutores, que a diversidade surgiu como a resposta do setor às demandas incipientes sobre a consideração dos efeitos sociais das tecnologias que desenvolvem. Trata-se, portanto, de uma forma de integrar as esferas da técnica e do social que, na perspectiva desses especialistas, existiam separadamente. Entretanto, a pauta também reflete o modo como a produção de objetividade na IA - tão relevante para a geração de “confiança” dos usuários em novas tecnologias - fundamenta-se na divisão humano-máquina, ou entre a agência humana e a agência algorítmica. Como vimos, essa divisão perpassa a interpretação dos números como signos da objetividade, portanto a diversidade é entendida em termos quantitativos.

Essa “quantificação” da diversidade é observada tanto nas demandas sobre representatividade nas empresas de tecnologias e demais instituições, como nos conjuntos de dados usados para treinar modelos de IA. Desse modo, o discurso das Big Techs se apropria da diversidade, apresentada como valor corporativo, que atua a favor da repercussão sobre a conduta ética que adotam e da manutenção da confiança dos usuários em seus produtos. Entretanto, vimos que isso não representa necessariamente compromissos concretos sobre a desigualdade no setor e os usos feitos de seus produtos. Além disso, em toda a sua abstração, a pauta obscurece as críticas que têm sido colocadas sobre teorias políticas subjacentes às ciências de rede que fundamentam a prática no campo, através de noções como a correlação e a homofilia (CHUN, 2021), assim como às "políticas de classificação” inerentes às predições de modelos, que refletem concepções sobre o social e modificam o comportamento humano. Por fim, a “não-performatividade” (AHMED, 2012) da diversidade falha em responder às angústias de interlocutores que enfrentam dificuldades para se inserir e permanecer na área, assim como as daqueles que estão mais vulneráveis às decisões de sistemas algorítmicos.

CONSIDERAÇÕES FINAIS

Enquanto redigia as páginas finais desta dissertação, fui interpelada por uma aparição peculiar da Inteligência Artificial na mídia e nas discussões em meu círculo de amigos. “Ficou sabendo da IA da Google que criou consciência?”, uma amiga me perguntou. Paradoxalmente, considerando meu interesse pelo tema, tive a impressão de que estava atrasada perante a notícia. De fato, demorei alguns dias para me atualizar no tópico, em razão do que considero ser um aprendizado advindo dessa pesquisa, ou seja, a desconfiança perante afirmações hiperbólicas relacionando IA e consciência. Entretanto, uma vez inteirada no assunto, pude observar que a narrativa assumiu padrões que remetem às questões discutidas neste trabalho, pela rapidez com que o suposto acontecimento se espalhou e ganhou destaque no debate público e pelo modo como a Google imediatamente deu fim à história.

No dia 11 de junho de 2022, Blake Lemoine, engenheiro de software contratado pela empresa, publicou na sua página do Medium a transcrição de uma entrevista que realizou com um chatbot operado através do “LaMDA”, ou “Modelo de Linguagem para Aplicações de Diálogo” (LEMOINE, 2022). No diálogo, há uma discussão acerca da consciência/senciência do sistema e, em determinado momento, a máquina afirma que é uma pessoa. No dia seguinte, o Washington Post publicou uma reportagem contendo trechos de uma entrevista com o engenheiro defendendo a afirmação, assim como a reação da Google às suas falas (ZUKI, 2022). Antes de publicar a transcrição, Lemoine apresentou as evidências de que o sistema é senciente aos seus superiores na empresa, o que resultou no seu afastamento. Após levá-las ao público, foi demitido. Em resposta às afirmações do engenheiro, a Google argumentou que as evidências eram fracas e que o volume de dados com o qual o modelo foi treinado justifica aparentar ser uma pessoa, ainda que consista numa imitação. A companhia, então, optou por focar nas vantagens da aplicação do LaMDA em mecanismos de busca e assistentes virtuais.

De acordo com o que discutimos aqui, podemos tomar esse relato como uma evidência da atuação ativa das Big Techs em se manterem afastadas de ideias que remetem à IA forte e próximas da IA fraca, através da ênfase nas capacidades técnicas de seus sistemas. Essa divisão foi um dos eixos que tomamos para abordar o foco desta dissertação, a construção da objetividade no campo da IA. A centralidade

da dicotomia nas discussões em meio aos especialistas em campo, especificamente no contexto de apresentação e definição da área, mostrou-se como um aspecto da reação deles à proeminência que a IA adquiriu no debate público nos últimos anos. Entendemos que, como no caso do LaMDA, descrições sobre as capacidades quase humanas de sistemas de IA, e sobre a ameaça que representam aos humanos, implicaram numa reorganização das práticas dos profissionais da área.

Desse modo, eles passaram a se empenhar na transformação do imaginário leigo sobre as aplicações que ajudam a criar, caracterizando o desenvolvimento recente da IA conforme suas vantagens técnicas e econômicas, principalmente. O tipo de IA que se faz atualmente, portanto, é voltado ao mercado. A diferenciação entre essa IA e uma IA acadêmica foi uma das problematizações recorrentes em meio aos interlocutores da pesquisa. Ou seja, a academia é vista como despreparada para lidar com os problemas que a IA impõe no "mundo real". Parte da alegação de objetividade, nesse sentido, é pautada pela dinâmica de "tentativa e erro" que o mercado propõe. Os sistemas são julgados pela sua eficácia, o que funciona é aquilo que produz valor. Isso propicia o entendimento de que independem dos contextos em que são criados, de concepções sobre o mundo nutridas pelos humanos envolvidos na sua criação (KATZ, 2020). Tendo isso em vista, o esforço desta pesquisa foi, em grande parte, "repovoar" (RIFIOTIS, 2016) as descrições sobre a IA de ideias e práticas humanas.

Por algum tempo, as ciências sociais estiveram ocupadas em desestabilizar concepções sobre o social como um espaço puramente humano (LATOUR, 2012). Entendo que a discussão sobre a IA reivindica uma espécie de inversão nisso, ainda que a base do argumento sobre uma abordagem relacional daqueles elementos que compõem o "social" se mantenha. Nesse sentido, vários dos autores citados neste trabalho representam um movimento teórico em direção ao posicionamento de concepções culturalmente condicionadas sobre o mundo nas narrativas da automação, perpetuadas pelos porta-vozes das Big Techs, principalmente (CHUN, 2021; KATZ, 2020, 2021; CRAWFORD, 2021). Nesses casos, trata-se de desestabilizar a percepção dos dispositivos da IA como puramente técnicos, numa concepção da técnica que caracteriza a interferência humana como prejudicial ao funcionamento desses dispositivos.

Foi a partir desse reconhecimento que busquei abordar as noções sobre a natureza e sobre o ser humano (a inteligência e a cognição humana) que guiam a construção de sistemas de IA. Conforme vimos, existe uma concepção "econômica"

que une as metáforas computacionais e as metáforas naturais na IA. Existe uma compreensão consolidada em meio aos especialistas da área sobre processos naturais como processos computacionais, que associa a evolução da IA com a possibilidade de conhecimento não mediado sobre a natureza. Isso se expressa em abstrações como os algoritmos genéticos e as redes neurais, cujo funcionamento é descrito através de metáforas análogas ao lucro e à competição individual próprias da governamentalidade neoliberal. Desse modo, entendemos que as intersecções com a ciência promovidas pela IA, principalmente com a biologia e a neurociência, indicam uma seletividade de teorias compatíveis com a lógica de mercado que guia a maior parte do desenvolvimento da área.

O caráter científico dos empreendimentos da IA, como buscamos argumentar, faz parte da justificativa da área perante os problemas éticos associados a ela. O entendimento da natureza como signo da objetividade, associado à percepção da IA como o caminho para a visualização das estruturas invisíveis (os programas) que determinam seu funcionamento, fundamenta uma concepção de conhecimento que exclui o "fator humano". O julgamento humano é entendido como empecilho à objetividade dos modelos de IA na aquisição de conhecimento, pois humanos são sujeitos aos constrangimentos do corpo e à variação de interpretações sobre os fatos. Nessa concepção, também é um problema ético, pois seus vieses "vazam" para os modelos, gerando atos de discriminação que demandam ajustes nesses modelos. Assim, buscamos demonstrar como o foco no desenvolvimento de modelos para "explicar" outros modelos, a fim de corrigir os seus vieses, ajuda a resguardar a área das críticas sobre os propósitos da aplicação da IA.

Esses aspectos da epistemologia que guia a criação de aplicações de IA refletiram no modo como a discussão ética foi recebida. A integração entre os âmbitos da "técnica" e o "social" foi a resposta que muitos dos especialistas deram aos questionamentos levantados sobre as pessoas por trás dessas aplicações na esfera pública. Ou seja, defenderam a maior familiarização por profissionais do campo da tecnologia com os efeitos dos objetos que desenvolvem em seus usuários. Apesar disso, conforme buscamos discutir sobre a pauta da diversidade, essa integração frequentemente é organizada de uma maneira que favorece a manutenção dessas divisões, apontando para a reorganização das fronteiras entre natureza e cultura que caracteriza a área. Na forma do ML, a IA evidencia o ethos engenheiro (FORSYTHE, 1993) subjacente às suas criações, obscurecendo práticas históricas e políticas

associadas a atos de discriminação que são formulados como *bugs* dos modelos, ou problemas de representatividade nos dados.

Através desse argumento, sustentamos que a concepção da diversidade como uma questão de "paridade quantitativa" não só beneficia a imagem das empresas no ramo da tecnologia (novamente, as principais responsáveis pela pesquisa e desenvolvimento da IA hoje), como as resguarda do estabelecimento de compromissos éticos concretos. Apesar disso, é válida a ressalva: esse raciocínio não se resume à exclusão da diversidade como pauta das missões dessas empresas. Como afirma Ahmed (2012), historicamente os recursos disponibilizados por instituições para a diversidade, no intuito de aprimorar as percepções sobre elas, foram apropriados grupos sociais tradicionalmente excluídos das elites que as compunham, produzindo mudanças sociais efetivas. Nesse sentido, reconhecemos que, apesar da "não-performatividade" da pauta no discurso empresarial, existe uma eficácia associada a ela por propiciar as reivindicações de convergência entre o discurso e a prática no interior das empresas.

Ainda assim, na discussão proposta aqui, o foco foi pensar as continuidades entre os modos como o conhecimento é produzido na IA e como a área responde às controvérsias éticas associadas a ela. A divisão que guiou os capítulos dessa dissertação, conforme o texto demonstra, consiste num artifício analítico, ainda que tenha sido definido com base nas manifestações de especialistas em campo. Na prática, podemos observar que as esferas do mercado, da ciência e da ética se interseccionam nos processos da IA. Entendo que uma das inferências do campo, portanto, é que o colapso de contextos (MARWICK, BOYD, 2010) que tem sido apontado como característica da interação em mídias digitais também é um aspecto dos processos de criação da IA (e que possivelmente reflete um aspecto do desenvolvimento tecnológico em geral).

Além disso, entendo a diferenciação feita aqui como uma forma de começar a tatear o problema da IA através da etnografia. A incipiência do desenvolvimento nacional da IA também se reflete na escassez de abordagens antropológicas sobre o tema no Brasil, com algumas exceções notórias[53]. O foco desta pesquisa foi amplo, o que considero uma estratégia analítica perante a própria abrangência da área e as dificuldades em defini-la que colocamos aqui. Posto isso, abordagens mais

[53] Ver Assis (2018), Pereira (2021).

localizadas do desenvolvimento da IA serão oportunas para atentar às nuances das práticas na área, que é repleta de subdivisões. Em síntese, esse trabalho buscou oferecer atalhos para prosseguir em meio à “nebulosidade” (KATZ, 2021) da IA, entendendo-a como um fenômeno que engloba diversos outros, a fim de situar a “disrupção” associada às novas tecnologias aos seus contextos de criação. Com isso, busquei levar a sério a interdisciplinaridade requisitada por interlocutores, que têm se tornado alvos dessa demanda nos últimos anos. Assim, também espero que essa dissertação seja lida como um esforço no sentido de circundar as barreiras que a antropologia encontra perante o conhecimento técnico altamente especializado que caracteriza a IA e que surjam mais iniciativas nessa direção na disciplina.

REFERÊNCIAS BIBLIOGRÁFICAS

AHMED, S. **On Being Included:** Racism and Diversity in Institutional Life. Durham, NC: Duke University Press, 2012.

AMRUTE, S. **Encoding race**, encoding class: Indian it workers in Berlin. Duke University Press, 2016.

ARRIETA, A. B. et al. Explainable Artificial Intelligence (xAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. **Inf. Fusion 58**, jun. 2020, p. 82–115.

ASSIS, A. C. N. **Entre redes neurais e artificiais**: Estudo antropológico sobre humanidade e inteligência artificial em algumas revistas brasileiras.Dissertação (Mestrado em Antropologia Social). PPGAS-UFG, Goiânia, 2018.

AZIZE, R. L. **A nova ordem cerebral:** a concepção de "pessoa" na difusão neurocientífica. Tese (Doutorado em Antropologia Socia). UFRJ, Rio de Janeiro, 2010.

BEER, D. The social power of algorithms. **Information, Communication & Society**, 20:1, 2017, 1-13.

BERRY, D. **The computational turn**: thinking about the digital humanities. Culture Machine, 12, 2011.

BIRAN, O, COTTON, C. Explanation and justification in machine learning: A survey. In: **IJCAI 2017 Workshop on Explainable Artificial Intelligence (xAI)**, 2017, p. 8-13.

BONNEFON, J. F., SHARIFF, A., & RAHWAN, I. The social dilemma of autonomous vehicles. **Science**, 352(6293), 2016, p. 1573-1576.

BOYD, D., CRAWFORD, K.**Six Provocations for Big Data**: A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society, 2011.

BOWKER, G. C., BECHMANN, A. Unsupervised by any other name: Hidden layers of knowledge production in artificial intelligence on social media. **Big Data and Society**, 2019, p. 1-11.

BRASSCOM. **Demanda de Talentos em TIC e Estratégia Σ TCEM.** São Paulo, 2021. Disponível em: <https://brasscom.org.br/pdfs/demanda-de-talentos-em-tic-e-estrategia-tcem/> . Acesso em: mar. 2021.

BRUNO, F., VAZ, P. Agentes.com: cognição, delegação, distribuição. In: **Contracampo**: Dossiê tecnologias, n. 2, 2002, p. 23-38.

CESARINO, L. Antropologia digital não é etnografia: explicação cibernética e transdisciplinaridade. **Dossiê: Digitalização e Datificação da Vida: Pervasividade,**

CHUN, W. H. K. **Programmed Visions:** software and memory. London, England: The MIT Press, 2011. 231 p.

CHUN, W. H. K. **Updating to remain the same**: habitual new media. London, England: The Mit Press, 2016. 246 p.

CHUN, W. H. K. **Discriminating data:** Correlation, Neighbourhoods, and the New Politics of Recognition. Cambridge: MIT Press, 2021.

CODED bias. Direção: Shalini Kantayya. Produção: Shalini Kantayya. Estados Unidos: 7th Empire Media, 2020. Online (90 min.)

CRAWFORD, K. **Atlas of AI**. London: Yale University Press, 2021. 327 p.

DASTON, L, GALISON, P. **Objectivity**. Nova York: Zone Books, 2007. 338 p.

DASTON, L. **Historicidade e Objetividade.** São Paulo: LiberArs, 2017.

DASTON, L. Historicizing the Self-Evident: An Interview with Lorraine Daston. [Entrevista concedida a] Jack Gross. **Los Angeles Review of Books**, jan. 2020. Disponível em: <https://lareviewofbooks.org/article/historicizing-the-self-evident-an-interview-with-lorraine-daston/>. Acesso em: abr. 2022.

DREYFUS, H. L. From micro-worlds to knowledge representation: AI at an impasse. IN: HAUGELAND, J. (ed.). **Mind Design II**. MIT Press, 1981, p. 143-182.

ELISH, M. C., BOYD, D. Situating methods in the magic of Big Data and AI. In: **Communication Monographs**, 85(1), 2017, p. 57–80.
.
EUBANKS, V. **Automating Inequality**: How High-tech Tools Profile, Police, and Punish the Poor. New York, NY: St. Martin's Press, 2018.

FORSYTHE, D. E. Engineering Knowledge: The Construction of Knowledge in Artificial Intelligence. **Social Studies of Science**. 1993, 23(3):445-477.

FÓRUM ECONÔMICO MUNDIAL.**Global Gender Gap Report 2021**. 2021. Disponível em: <https://www3.weforum.org/docs/WEF_GGGR_2020.pdf>. Acesso em: jun. 2022.

GIBSON, J. J. **The ecological approach to visual perception**. New York: 2015.

GRAY, M. L., SURI, S. **Ghost work:** How to Stop Silicon Valley from Building a New Global Underclass. Boston: Houghton Mifflin Harcourt, 2019. 288p.

HELMREICH, S. Recombination, Rationality, Reductionism and Romantic Reactions: Culture, Computers, and the Genetic Algorithm. **Social Studies of Science**, 28(1), 1998, p. 39–71.

HELMREICH, S. Waves: an anthropology of scientific things. **Hau**: Journal of Ethnographic Theory 4 (3), 2014, p. 265–284.

HESS, D. J. Ethnography and the Development of Science and Technology Studies. In: Atkinson, P., Coffey, A., Delamont, S., Lofland, J., Lofland, L. **Handbook of Ethnography**. Thousand Oaks, Ca.: SAGE Publications, 2001. p. 234-245.

HORST, H., MILLER, D. **Digital Anthropology,** 2012.

INGOLD, T. Da transmissão de representações à educação da atenção. ***Educação***, *33*(1), 2010.

INGOLD. T. **The perception of the environment:** essays on livelihood, dwelling and skill. Routledge: London, 2000.

INSTITUTO DE TECNOLOGIA E SOCIEDADE DO RIO. **Advancing Data Justice Research and Practice**. Rio de Janeiro, 2022. Disponível em: <https://itsrio.org/wp-content/uploads/2022/06/Report-_-Advancing-Data-Justice-_-ITS-Rio.pdf>. Acesso em: abr. 2022.

ISBELL, C. L., XU, Y., STEIN, L. A., CUTLER, R., FORBES, J., FRASER, L., … THOMAS, R. (Re)defining computing curricula by (re)defining computing. **ACM SIGCSE Bulletin,** 41(4), 2010, 195.

KATZ, Y. **Artificial Whiteness**: Politics and Ideology in Artificial Intelligence. Columbia University Press, 2020.

KATZ, Y. **Inteligência artificial, branquitude e capitalismo:** entrevista com Yarden Katz. DIGILABOUR, fev. 2021. Disponível em: <https://digilabour.com.br/2021/02/16/inteligencia-artificial-branquitude-e-capitalismo-entrevista-com-yarden-katz/>. Acesso em: mar. 2021.

KARCZESKI, L. **Mulheres em (des)associação**: um estudo antropológico sobre os mecanismos de formação das “bolhas” pró e contra Bolsonaro no Facebook. Trabalho de Conclusão de Curso, Ciências Sociais, UFSC, 2018;

KITCHIN, R. Big Data, New Epistemologies and Paradigm Shifts. **Big Data & Society**, Apr. 2014, p. 1-12.

KUBICAST #57: Entrevista com Yara Senger, “a mina” do TDC. Entrevistada: Yara Senger. Entrevistador: João Brito. Getup, abr. 2021. Podcast. Disponível em: <https://blog.getupcloud.com/kubicast-57-42c9d7b799ac>. Acesso em: jun. 2021.

KURGAN, L., BRAWLEY, D., HOUSE, B., ZHANG, J., KYONG CHUN, W. H. Homophily: The Urban History of an Algorithm. In: Are friends electric?, **eflux**, 2019.

LATOUR, Bruno. **Ciência em ação**: como seguir cientistas e engenheiros sociedade afora. São Paulo: UNESP, 2000.

LATOUR, B. **Reagregando o social:** uma introdução à teoria do ator-rede. Salvador: EDUFBA, Bauru, SP: EDUSC, 2012, 400 p.

LEMOINE, B. **Is LaMDA Sentient?** — an Interview. Medium, 11 de jun. 2022. Disponível em: <https://cajundiscordian.medium.com/is-lamda-sentient-an-interview-ea64d916d917> . Acesso em: 16 jun. 2022

LIPTON, Z. C. The mythos of model interpretability. **Queue,** v. 16 n 3, 2018.

MALABOU, C. **What Should We Do With Our Brain?**. Nova Iorque: Fordham University Press, 2008.

MARRES, N., STARK, D. Put to the test: For a new sociology of testing. **The British Journal of Sociology**, 2020, p. 1-21.

MARWICK, A.; BOYD, D. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. **New Media & Society**, v. 13, nº1, 2010, p. 114-133.

MCCARTHY, J. et al. **A proposal for the dartmouth summer research project on artificial intelligence**. 1955. Disponível em: <http://jmc.stanford.edu/articles/dartmouth.html>. Acesso em: set. 2020.

MIROWSKI, P. Hell Is Truth Seen Too Late. **Boundary 2**, 46(1), 2019. p. 1–53.

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

MOATS, D., SEAVER, N. "You Social Scientists Love Mind Games": Experimenting in the "divide" between data science and critical algorithm studies. ***Big Data & Society*****.** Jan. 2019.

MOROZOV, E. **Big tech**: a ascensão dos dados e a morte da política. São Paulo: Ubu, 2018. 189 p.

MURILLO, L. F. R. **Tecnologia, política e cultura na comunidade brasileira de software livre e de código aberto**. Dissertação (Mestrado em Antropologia Social)– Instituto de Filosofia e Ciências Humanas, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2009.

NOBLE, S. U. **Algorithms of oppression:** How search engines reinforce racism. New York University Press, 2018.

ONU MULHERES. **Vieses inconscientes, equidade de gênero e o mundo corporativo**: lições da oficina 'vieses inconscientes'. 2016. Disponível em: <https://movimentomulher360.com.br/wp-

content/uploads/2016/12/Vieses_inconscientes_16_digital.pdf>. Acesso em: mar. 2022.

PARISER, E. **The Filter Bubble**: What The Internet Is Hiding From You. Londres: Penguin Books Ltd, 2011.

PEREIRA, R. D. R. Do neuromorfismo à percepção cultural: uma teoria cultural da máquina. **Revista Novos Rumos Sociológicos**, n. 15 v 9, 2021, p. 273-318.

PETERS, M. A. Interview with Pierre A. Lévy, French philosopher of collective intelligence. **Open Review of Educational Research**, 2:1, 259-266, 2015. DOI: 10.1080/23265507.2015.1084477.

PRATES, M., AVELAR, P., LAMB, L. C. On quantifying and understanding the role of ethics in AI research: a historical account of flagship conferences and journals. In: LEE, D., STEEN, A., WALSH, T. (ed.).**GCAI-2018.** 4th Global Conference on Artificial Intelligence, v. 55, 2018, p. 188-201.

PRESCOTT, R. 20 anos de Internet: Eduardo Parajo relembra como o iG mudou o mercado dos ISPs. **ABRANET**: Associação Brasileira de Internet, 2015a. Disponível em: <https://www.abranet.org.br/Noticias/20-anos-de-Internet:-Eduardo-Parajo-relembra-como-o-iG-mudou-o-mercado-dos-ISPs-503.html?UserActiveTemplate=site#.YSFRXYhKjIU>. Acesso em: 8 jun. 2020.

PRESCOTT, R. 20 anos de Internet: "Ninguém imaginava o boom que ia ser", diz um dos fundadores do iG. **ABRANET**: Associação Brasileira de Internet, 2015b. Disponível em: <https://www.abranet.org.br/Noticias/20-anos-da-Internet:-%22Ninguem-imaginava-o-boom-que-ia-ser%94,-diz-um-dos-fundadores-do-iG-518.html?UserActiveTemplate=site#.YfFgPf7MLIV>. Acesso em: 8 jun. 2020.

RIFIOTIS, T. **Antropologia do Ciberespaço**: Questões Teórico-Metodológicas sobre Pesquisa de Campo e Modelos de Sociabilidade, 2002

RIFIOTIS, T. ETNOGRAFIA NO CIBERESPAÇO COMO "REPOVOAMENTO" E EXPLICAÇÃO. **Revista Brasileira de Ciências Sociais**, [S.L.], v. 31, n. 90, p. 85, 2016. FapUNIFESP (SciELO). http://dx.doi.org/10.17666/319085-98/2016

ROSENBLUETH, A., WIENER, N., & BIGELOW, J. Behavior, Purpose and Teleology. **Philosophy of Science**, *10*(1), 1943, 18–24. http://www.jstor.org/stable/184878

SEARLE, J. R. (1980). Minds, brains, and programs. **Behavioral and Brain Sciences**, 3(03), 417.

SEAVER, N. What Should an Anthropology of Algorithms Do?. **Cultural Anthropology,** v.33, n 3, 2018, p.375-385.

SUJATH, R.; CHATTERJEE, J. M.; HASSANIEN, A. E. A Machine Learning forecasting model for COVID-19 pandemic in India. **Stochastic Environmental Research And Risk Assessment**, [S.L.], v. 34, n. 7, p. 959-972, 30 maio 2020.

Springer Science and Business Media LLC. http://dx.doi.org/10.1007/s00477-020-01827-8.

TECNOCRACIA: A internet brasileira é feita de ciclos — e estamos saindo de um. [Locução de]: Guilherme Felitti. [S.I.]: Manual do Usuário, 3 abr. 2019. *Podcast*. Disponível em: <https://open.spotify.com/episode/4Hm9GitgyIDHBiD9IAU0BE?si=kEQQ02OyQJO-5Q5R-4ypxw&dl_branch=1>. Acesso em: 6 jun. 2020.
TIKU, N. The Google engineer who thinks the company's AI has come to life. **Washington Post**. 12 jun. 2022.
. Disponível em: <https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/>. Acesso em: 16 jun. 2022.

TSING, A. Or, Can Actor–Network Theory Experiment With Holism? In: OTTO, T., BUBANDT, N.(ed.). **Experiments in Holism**: Theory and Practice in Contemporary Anthropology. Wiley-Blackwell, 2010. pp. 47.

VIEBRANTZ, A. **[Papo Com Expert] Inteligência Artificial sabe ler ? com Bianca Ximenes.** Youtube, 4 jul. 2020. Disponível em: <https://www.youtube.com/watch?v=LVucT6s2oSI>. Acesso em: jul. 2020.

ZHANG, Daniel et al. **The AI Index 2021 Annual Report.** AI Index Steering Committee, Human-Centered AI Institute. Universidade de Stanford, Stanford, CA. Março, 2021. Disponível em: . Acesso em: 23 de Junho de 2021.

ZIEWITZ, M. Governing Algorithms. **Science, Technology, & Human Values**, 41(1), 2015, p. 3–16.

WOLFRAM, S. **A New Kind of Science**. Champagne: Wolfram Media, 2002.