

INTELIGÊNCIA ARTIFICIAL: OS TRÊS FATORES QUE FOMENTAM O PRECONCEITO CODIFICADO

*Jackson Barreto Costa Júnior*²³

INTRODUÇÃO

O mundo vivencia, de forma inconsciente, um *apartheid* digital. Você tem consciência das suas inconsciências?

Em pleno século XXI a Inteligência Artificial continua a promover a segregação social, ocultando-se por detrás da imagem de neutralidade. A segregação foi repaginada, agora é digital e o grupo dos marginalizados socialmente, antes composto prioritariamente por uma etnia, foi alargado, comportando todos os que não se enquadram no eurocentrismo heteronormativo. Do outro lado, as coisas continuam como sempre foram, uma hegemonia de homens brancos e de meia-idade.

O objetivo deste artigo consiste na análise dos três principais elementos que fomentam o preconceito codificado com a pretensão de promover uma reflexão que seja capaz de engajar a união dos especialistas das Ciências Sociais e da Inteligência Artificial para o combate desta nova forma de se implementar antigas discriminações.

23 Graduando em Engenharia Informática pelo Instituto Politécnico de Viana do Castelo (IPVC) de Portugal, graduado em Direito pela Universidade Cândido Mendes (UCAM) do Brasil.

Para tal, o primeiro tópico aborda a causa gênese da situação problema, a inconsciência política. No segundo tópico, se desvela a questão da inexplicabilidade. Finalmente, no terceiro tópico, se conduz o olhar para o epicentro, para o combate da situação problema na concepção do projeto tecnológico.

1. (IN) CONSCIÊNCIA POLÍTICA: A IGNORÂNCIA QUANTO AO SIGNIFICADO SOCIAL

1.1. TECNOLOGIA POLÍTICA

Aristóteles foi categórico ao definir “que o homem é, por natureza, um ser vivo político” (ARISTÓTELES, 1998, p. 53). Isto implica em que o ser humano estabelece uma relação interpessoal mais complexa em grupos que constituem o que chamamos de sociedade.

A sociedade, sob a perspectiva deste ensaio, é composta por suas instituições sociais, que são os sistemas de ordenação social, com suas regras explícitas ou não. E a consciência política se dá quando o indivíduo desperta e passa a tomar ciência destes sistemas de ordenação e toda a sua dinâmica social, bem como a pensar criticamente acerca da natureza e da justificação do sistema como um todo.

O significado social, sob a ótica deste ensaio, é o sentido que determinado elemento tem em uma sociedade, isto é, o impacto que uma ação, inação, ou mesmo um artefato, causa na comunidade.

Este conceito torna-se mais cristalino quando se considera o fenômeno social do luto. A cor branca, nas culturas ocidentais, representa o nascimento e a pureza (HELLER, 2014, p. 551). Entretanto, o branco é a cor do luto em todo o Oriente (CHEVALIER; GHERRBRANT, 1994, p. 128). A cor do traje com o qual se comparece em uma cerimônia de luto possui um significado social diferente, conforme a sociedade onde o cerimonial decorra.

A tecnologia é a sociedade, e a sociedade não pode ser entendida ou representada sem suas ferramentas tecnológicas (CASTELLS, 1999). As tecnologias da navegação, da pólvora, da máquina a vapor, dentre outras, definiram quais sociedades estariam em condição mais favorável no decurso da história. Embora a tecnologia possa interromper algumas estruturas

de poder, também pode reforçá-la e raramente é projetada para empoderar as populações mais vulneráveis (HIATT; KLEINMAN *et al.*, 2017).

A tecnologia amplia as capacidades humanas e se alinha com as intenções dos seus criadores/detentores (TOYAMA, 2011) these ideas are synthesized and augmented for an amplification theory of “information and communication technology for development” (ICT4D. E algo que nunca mudou na linha temporal foi a proporção entre os criadores/detentores da tecnologia e os que são submetidos a ela. A tecnologia continua sendo financiada por quem está no poder.

Quando se trata de Inteligência Artificial a desigualdade é abissal: uma pequena fração da humanidade está atualmente envolvida no desenvolvimento de um conjunto de tecnologias que já estão transformando a vida cotidiana de milhares de pessoas (FLORIDI; COWLS, 2019).

A inconsciência política propicia a ignorância do significado social, que por sua vez permite “criar sistemas que santificam o *status quo* e promovem os interesses dos poderosos. Quando o que é necessário é uma comunidade que exponha e critique sistemas que concentram poder” (KALLURI, 2020).

A tecnologia é política desde os primórdios e vai continuar a ser, tal como toda criação humana estará sempre impregnada das ideações de seu criador. O que retoma um imprescindível questionamento: Quem são os idealizadores da Inteligência Artificial e quais suas ideologias?

1.2. OS ALQUIMISTAS DA INTELIGÊNCIA ARTIFICIAL

Como os antigos alquimistas, os especialistas em Inteligência Artificial habitam um círculo hermético.

O hermetismo justifica-se pelo fato de os especialistas comporem uma pequena parcela da população – profissionais de tecnologia – e pertencerem a um esguio estrato social. Menos de 25% dos profissionais mundiais de Inteligência Artificial são mulheres (WORLD ECONOMIC FORUM, 2021). Enquanto no Brasil, em 2019, os profissionais de informática são principalmente: homens, brancos, jovens de classe socioeconômica média e alta que começaram a sua trajetória nos centros formais de ensino (PRETALAB, 2017). Este perfil típico traz à cena dois pontos que merecem ser abordados com mais proximidade: A percepção de si

no contexto social e o posicionamento das comunidades de pesquisa e desenvolvimento.

Estes pontos serão abordados, nesta ordem, nos subtópicos seguintes.

1.2.1. O ALQUIMISTA NO MUNDO REAL

Os pesquisadores de Inteligência Artificial não se percebem como superiores ou como um grupo apartado da comunidade. Eles são inconscientes políticos, ou seja, ignoram o significado social da tecnologia que desenvolvem, escolhendo, deliberadamente, postergar considerações acerca do significado social (BRANDÃO; CARBONERA *et al.*, 2021, p. 2)we present a case study with Deep Learning (DL.

As preferências destes especialistas é espontaneamente orientada para a carga técnica do projeto, ao invés do significado social (BRANDÃO; CARBONERA *et al.*, 2021, p. 20)we present a case study with Deep Learning (DL. O significado social nem mesmo chega a ser considerado como um tema de pesquisa para esta comunidade (BRANDÃO; CARBONERA *et al.*, 2021, p. 21)we present a case study with Deep Learning (DL.

Este grupo hermético sente-se desconfortável em ter que adequar o seu discurso ao dos utilizadores, consideram esta tarefa como algo enfadonho. Eles esperam que exista alguém que sirva de mediador entre eles e o utilizador afetado (BRANDÃO; CARBONERA *et al.*, 2021, p. 21)we present a case study with Deep Learning (DL. Alguém que seja o elo entre o significado social, o utilizador afetado e eles, como um sacerdote.

Há de se observar que tanto os especialistas, como o grupo social no qual eles se inserem (homens brancos), quando colocados na posição de utilizadores afetados, não sofrem os efeitos sociais negativos de suas tecnologias. Afinal, eles não são privados de atendimento médico especializado, por serem negros (LEDFORD, 2019). A cadeia de privilégios é histórica, por vezes sutil, e muitas vezes ocultada por detrás do sigilo de uma patente.

A posição confortável na qual os especialistas e seu grupo social se assentam é inegável. Contudo, o feudo da pesquisa e desenvolvimento estimula esta vida absolutamente acrítica.

1.2.2. O FEUDO DA PESQUISA E DESENVOLVIMENTO

Um dos pilares de estudo dos alquimistas era a busca por soluções que acelerassem a riqueza dos seus reis. Este é outro ponto de semelhança que inspira a atribuição da alcunha de alquimistas aos investigadores de Inteligência Artificial.

A realeza foi substituída pelos feudos de pesquisa e desenvolvimento. Estes feudos tomam como princípios de suas pesquisas apenas as métricas abstratas, em detrimento da contextualização social (BRANDÃO; CARBONERA *et al.*, 2021, p. 23)we present a case study with Deep Learning (DL.

A cientista cognitiva Abeba Birhane traz uma reflexão sobre esta valorização exacerbada das métricas, uma herança do produtivismo:

Nem tudo que tem valor pode ser medido e codificado e a ética é um exemplo. Quanto mais cedo as pessoas deixarem de tentar encontrar uma equação precisa para “justiça” e começarem a perguntar “*por que estou construindo essa ferramenta em primeiro lugar?*”, mais justa e atenciosa a sociedade se tornará (BIRHANE, [s.d.]”9”,”18”]],”author”:[{“dropping-particle”：“”,”family”：“-Birhane”,”given”：“Abeba”,”non-dropping-particle”：“”,”parse-names”：false,”suffix”：“”}],”id”：“ITEM-1”,”issued”：{“date-parts”：[[“0”]]},”title”：“Abeba Birhane no Twitter: \”Not everything valuable can be measured and codified and ethics is one such thing. The sooner folk let go of attempting to find some accurate equation for \”fairness\” and start asking \”why am I building this tool in the fir”,”type”：“webpage”}],”uris”：[“http://www.mendeley.com/documents/?uuid=c029a5f2-e33d-3a0e-b-7d5-8d057b472525”]]],”mendeley”：{“formattedCitation”：“(-BIRHANE, [S.d.]

A fala deixa claro a necessidade de se deslocar as atenções para o significado social dos projetos de tecnologia.

Não se pode negar que recentemente (2017 a 2020) os Estados e a sociedade civil têm-se posicionado através de declarações de princípios e até mesmo de regulação. Como: a Declaração de Montreal (UNIVERSIDA-

DE DE MONTREAL, 2017), desenvolvida pela Universidade de Montreal; a Declaração sobre Inteligência Artificial, Robótica e Sistemas Autônomos (EUROPEAN COMMISSION, 2018), publicada pelo Grupo Europeu de Ética na Ciência e Novas Tecnologias da Comissão Europeia; e o Relatório do Comitê de Inteligência Artificial da Câmara dos Lordes do Reino Unido (UK PARLIAMENT, 2017).

Entretanto, ainda se caminha a passos muito curtos. E nenhum desses posicionamentos foram capazes de explicar e confortar a família do Sr. Williams (norte-americano morador de Farmington Hills, Michigan), que foi preso no gramado de casa, ao chegar de mais um dia de trabalho (HILL, 2020).

A grande questão é que o Sr. Williams não pertencia à estirpe social dominante, ele era negro. Foi preso por uma imagem borrada, de péssima qualidade, submetida a um sistema de Inteligência Artificial. Ele não foi a primeira vítima da discriminação codificada, foi apenas aquela que foi noticiada.

É no mundo real que as tecnologias estão sendo implementadas, com os alquimistas e feudos, ainda, ignorantes quanto ao significado social. Qualquer tecnologia criada reflete as aspirações e limitações de seus criadores, se estes forem limitados na hora de pensar em inclusão, isso será refletido e incorporado na tecnologia desenvolvida (BUOLAMWINI, 2017).

Apesar disso, o que ainda não é compreensivo é o motivo pelo qual os utilizadores afetados não se revoltam, a razão pela qual quem tomou a decisão de retirar a liberdade de um homem não se questionou, frente a uma imagem borrada e a sugestão da Inteligência Artificial. Talvez seja pelo fato de analogias com alquimistas, reinos e sacerdotes não serem tão fantasiosas assim. O que se oculta por detrás deste inconsciente que ainda não foi percebido? Qual o motivo desta fé em um modelo tão incompreensivo?

1.3. O SALTO DE FÉ

Pode parecer contraintuitivo ter a fé, que habitualmente conceitua-se como “a certeza daquilo que esperamos e a prova das coisas que não vemos” (BÍBLIA, 1995, p. 289, Hebreus, 11,1), associada com a Inteligência

Artificial. Entretanto, que terminologia melhor conceitua esta crença inquestionável em algo que não se compreende. Quando os especialistas são indagados sobre as justificativas da sugestão de seu sistema, uma resposta comum é: “Nós sabemos explicar o modelo. Mas é difícil para nós explicar por que o modelo faz o que faz” (BRANDÃO; CARBONERA *et al.*, 2021, p. 19)we present a case study with Deep Learning (DL.

O mundo vive a crença de que mais tecnologia é a solução para todas as coisas. É o que a pesquisadora Meredith Broussard definiu como “tecnochauvinismo” (BROUSSARD, 2019). Os alquimistas da Inteligência Artificial exigem uma fé inabalável, mesmo quando eles não possuem compreensão plena do motivo pelo qual um determinado livro foi recomendado entre tantos outros. Eles não enxergam dentro desta caixa opaca que é o modelo (ZHANG; CHEN, 2020, p. 63), mas seguem exigindo a devoção e confiança da comunidade.

A crença na matemática e na tecnologia neutra e apolítica é reconfortante para os tomadores de decisões que podem agir como os agentes do holocausto, que alegaram estarem apenas a cumprir ordens. Desta vez, os decisores podem alegar que foi o algoritmo que decidiu.

Se cunhou um termo para este fenômeno: *mathwashing* (lavagem matemática, em tradução livre), que utiliza a crença na matemática, por detrás dos algoritmos, para lavar as responsabilidades dos tomadores de decisão (BENENSON, 2021).

A cada novo sistema de Inteligência Artificial que é disponibilizado no mercado a sociedade se submete ao “risco de o crescimento da autonomia artificial minar o florescimento da autonomia humana” (FLORIDI, COWLS, 2019, p. 7). Tudo em nome da singular facilidade, para se receber uma sugestão “inocente” de um novo livro, mesmo que talvez aquele algoritmo de recomendação possa indicar apenas títulos de autores europeus, em detrimento de novos talentos regionais. Quem saberá?

Contudo, como se pode “cobrar” um despertar do utilizador afetado, quando a crença na tecnologia apolítica está em seu auge? Quando um pedido de crédito é recusado a única resposta é: “O algoritmo decidiu, estou de mãos atadas.” Como questionar algo que não se faz compreender? Como despertar, quando tudo o que se tem como resposta são decisões arbitrárias e sem fundamentação?

2. (IN) EXPLICABILIDADE: A IGNORÂNCIA QUANTO AO PORQUÊ

2.1. EXPLICANDO A EXPLICABILIDADE

À medida que a Inteligência Artificial capilarizou por todos os setores da sociedade, permeando as esferas de tomada de decisões humanas, sua transparência e interpretabilidade cresceram significativamente em importância.

Quando se discute uma explicação para uma decisão tomada, tem-se como objetivo evidenciar a necessidade desta decisão em particular ser fundamentada (ADADI; BERRADA, 2018, p. 52142) ter uma justificativa explícita para o utilizador afetado por esta decisão e não uma narrativa detalhada do funcionamento intrínseco do algoritmo em um contexto geral.

Durante o Carnaval na Bahia, um sistema de videomonitoramento capturou os rostos de mais de 1,3 milhão de pessoas, gerando 903 alertas, o que resultou no cumprimento de 18 mandados e na prisão de 15 pessoas (NUNES, 2019), ou seja, de todos os alertas emitidos, mais de 96% representaram constrangimentos, prisões arbitrárias e violações de direitos humanos.

Explicabilidade é informar a cada uma destas vítimas que elas sofreram este constrangimento por conta de um sistema que calculou a distância entre os seus pontos nodais e concluiu que a distância entre os seus olhos e o comprimento do seu nariz eram compatíveis com o de algum criminoso e que os seus traços negroides foram considerados com significância.

Evidenciar a necessidade de uma decisão ser fundamentada pode parecer uma obviedade, afinal, decisões fundamentadas são um dos princípios democráticos. O que provoca uma reflexão sobre como a divinização da tecnologia fez a sociedade aceitar a regressão de direitos tão fundamentais.

A explicabilidade é a tão almejada pedra filosofal, com ela é possível despertar a consciência política e iluminar a senda do significado social, permitindo que, de forma racional, possa-se depositar alguma confiança na tecnologia que decide a vida das pessoas.

Mas em que situação está, na atualidade, a busca pela explicabilidade, no contexto da Inteligência Artificial?

2.2. A EXPLICABILIDADE NA ATUALIDADE

Ainda que os algoritmos de Inteligência Artificial pareçam poderosos em questão de previsões, eles sofrem de opacidade, não sendo possível obter uma compreensão sobre as motivações de cada decisão particular (ADADI; BERRADA, 2018, p. 52138). Este problema também afeta os pesquisadores, o que acaba por dificultar o trabalho de pesquisa.

Esta dificuldade fomentou a pesquisa por compreensão dos modelos de previsão. Entretanto, grande parte desta pesquisa é orientada pelas necessidades técnicas daqueles que constroem os sistemas, e não daqueles que são afetados por ela (BRANDÃO; CARBONERA *et al.*, 2021, p. 2) we present a case study with Deep Learning (DL. Um comportamento previsível, tendo em vista a ignorância quanto ao significado social identificada nestes especialistas.

Dessa forma, as pesquisas empreendidas na área utilizam métodos que possuem a intenção de evidenciar, para o pesquisador, qual parte do objeto analisado foi considerado pelo sistema para tomar aquela decisão. Contudo, observe que embora estas técnicas possam contribuir para um melhoramento do artefato tecnológico, em momento algum estão tendo em consideração o utilizador afetado.

Uma explicação é uma forma de interação e como tal deve considerar os intervenientes. Não se pode pensar em explicação sem se considerar as projeções psicológicas, pedagógicas, filosóficas e comunicacionais.

Entretanto, estas projeções são totalmente ignoradas, pesquisas apontam que as ideias das Ciências Sociais e do comportamento humano não são suficientemente visíveis no campo de estudo do Aprendizado de Máquina (ADADI; BERRADA, 2018, p. 52142), o que alerta para a falta de interdisciplinaridade.

A explicabilidade tem dificuldade em alcançar um nível prioritário de pesquisa, já que implementar explicabilidade na Inteligência Artificial é algo financeiramente custoso e que expõem os responsáveis pela tecnologia e os tomadores de decisões ao fenômeno da responsabilização.

Os financiadores da área alegam que a explicabilidade nem sempre é uma necessidade, que apenas serviços críticos, como o diagnóstico de um paciente, demandariam uma explicação (ADADI; BERRADA, 2018, p. 52143). Esta afirmação evidencia o conflito de interesse nesta questão.

Permita-se realizar uma reflexão: Quem define o que é crítico? Os mesmos que ignoram o significado social e o utilizador?

Considere um “inofensivo” sistema de recomendação de publicidade, que não demandaria explicabilidade.

Assuma o exemplo de um utilizador que está em algum programa de reabilitação, para se recuperar do vício no álcool ou no jogo. Imagine o sistema identificando estes traços e bombardeando-o com anúncios de bebidas e de apostas *on-line*. Este algoritmo aumentaria significativamente o lucro da empresa. Quanto ao utilizador, segundo a perspectiva de serviços críticos, pode ser ignorado, afinal esta não é uma aplicação crítica.

Alguns Estados e comunidades, na esperança de fomentar a explicabilidade, têm ratificado direitos básicos, como o “direito à explicação” previsto no Regulamento Geral sobre a Proteção de Dados (RGPD) (GOODMAN; FLAXMAN, 2017)for the first time in more than two decades, the European Parliament adopted a set of comprehensive regulations for the collection, storage, and use of personal information, the General Data Protection Regulation (GDPR, que garante ao indivíduo, por exemplo, o direito de não estar sujeito a nenhuma decisão tomada exclusivamente com base em sistemas automatizados.

A academia tem acompanhado estas diretrizes e já começa a desbravar com sucesso o campo da explicabilidade, como é o exemplo do novo modelo batizado de “rede neural líquida” (HASANI; LECHNER *et al.*, 2021)we construct networks of linear first-order dynamical systems modulated via nonlinear interlinked gates. The resulting models represent dynamical systems with varying (i.e., liquid, muito mais simples e interpretável.

Todavia, a velocidade desta jornada ainda é muito limitada pela crença na tecnologia neutra e pela ignorância sobre o significado social. O que apenas ressalta a relação intrínseca entre a consciência política e a explicabilidade e como esta relação precisa ser observada com um olhar mais proximal.

2.3. EXPLICABILIDADE, O PILAR QUE SUSTENTA A CONSCIÊNCIA POLÍTICA

Exercer criticamente a consciência política é um ato que deriva da formação do juízo de valor sobre uma determinada decisão particular, é preciso conhecer as justificativas daquela decisão para que se possa questionar se o significado social daquela tecnologia corresponde ao almejado pela sociedade em questão.

A explicabilidade garante, ainda, uma maneira escrutável e comprovável de se posicionar sobre as decisões algorítmicas como sendo justas e éticas, permitindo que a confiança na tecnologia seja motivada pela transparência e não pela fé (ADADI; BERRADA, 2018, p. 52142).

A segurança também é beneficiada pela explicabilidade, afinal, compreender mais sobre determinado artefato tecnológico permite identificar, antecipadamente, vulnerabilidades (ADADI; BERRADA, 2018, p. 52143), evitando, por exemplo, os adesivos colados em um sinal de paragem obrigatória, que fez com que o sistema deixasse de reconhecer o sinal de trânsito (EYKHOLT; EVTIMOV *et al.*, 2018), expondo os utilizadores da via pública ao perigo.

A evolução e melhoria contínua é obviamente outro benefício que a explicabilidade traz consigo, já que um artefato que pode ser explicado pode ser melhorado com mais facilidade (ADADI; BERRADA, 2018, p. 52143).

Por fim, a explicabilidade proporciona a aquisição de novos conhecimentos, afinal, os sistemas poderiam explicar a estratégia aprendida para determinado evento (ADADI; BERRADA, 2018, p. 52143), abrindo insondáveis portas de conhecimento.

Como observa-se, a explicabilidade é um pilar da consciência política, pois ela proporciona elementos objetivos sob os quais os utilizadores e a sociedade civil podem debruçar suas atenções identificando fenômenos discriminatórios que antes passariam despercebidos.

Entretanto, a capacidade da explicabilidade em proporcionar o exercício pleno da consciência política não impede que os impactos sociais negativos cheguem ao dia a dia da comunidade. Como, então, se poderia migrar do combate aos efeitos nocivos da Inteligência Artificial para o confronto das causas destes problemas?

3. (IN) EXISTÊNCIA DE INTERDISCIPLINARIDADE: A IGNORÂNCIA QUANTO AO UTILIZADOR AFETADO

3.1. A HERANÇA DOS VÍCIOS

A Inteligência Artificial constrói seu conhecimento através dos dados. O sistema detecta padrões nestes dados, e associa estes padrões com o resultado procurado.

Contudo, o primeiro ponto questionável é que estes dados são o recorte de uma sociedade e não a sociedade como um todo. Este foi um dos primeiros problemas que a Inteligência Artificial enfrentou.

Foi justamente o fenômeno denominado como “*Big Data*” que proporcionou a expansão descontrolada da Inteligência Artificial, pois só então esteve disponível uma quantidade de dados descomunal.

A vontade da comunidade tecnológica de crescer aceleradamente, respaldada pelo mote “mova-se rápido e quebre as coisas” (HIATT; KLEINMAN *et al.*, 2017), junto com a grande disponibilidade de dados, consolidaram a crença de que bastava apenas muitos dados para se combater qualquer enviesamento. O que evidencia a ignorância do significado social e do utilizador afetado.

Entretanto, os dados refletem comportamentos humanos, e muitas das atitudes humanas revelam-se preconceituosas, logo, há grande probabilidade de reprodução destas posturas (PEREIRA; BRUNO *et al.*, 2020, p. 4). A cientista de dados Cathy O’Neil alerta que o “*Big data* não elimina o preconceito, estamos apenas camuflando-o com tecnologia” (BENENSON, 2021). Estes sistemas, quando utilizam os dados desta forma, fazem mais do que refletir estes problemas sociais, eles amplificam as desigualdades atuais e promovem a perpetuação de mazelas sociais.

Os dados proporcionam a herança dos vícios da sociedade, mas seria possível eliminar esta herança maldita?

3.2. SOMOS MAIS QUE NÚMEROS

A ciência dos dados dispõe de técnicas para “higienizar” os dados, se procura balancear os dados e aplicar os cuidados estatísticos necessários para abordar amostras e populações.

No entanto, se todos esses cuidados já são aplicados, então, em teoria, não existe a herança maldita. Logo, as mulheres não seriam rejeitadas para cargos técnicos (DASTIN, 2018), o Twitter não priorizaria a imagem de brancos em detrimento de negros, tão pouco a remoção de fundo dos vídeos, na plataforma de videoconferência Zoom, seria falho com pessoas negras (ALVES, 2020). Então, o que está acontecendo?

Simplesmente o ser humano é mais do que números. Os dados são o elo de ligação entre a tecnologia e o tecido social. Dessa forma, muitas das técnicas que são atualmente utilizadas em Aprendizado de Máquina, como a transferência de aprendizagem, precisam ser revisitadas, tendo em vista uma abordagem sociotécnica (BRANDÃO; CARBONERA *et al.*, 2021)we present a case study with Deep Learning (DL.

Os dados e algoritmos promovem a redução de uma realidade complexa para uma visão mais simplista do mundo, apenas as partes que são factíveis de serem mensuráveis são utilizadas (BENENSON, 2021).

Esta dicotomia entre dados e contexto social pode ser observada na conjuntura da população carcerária brasileira, que é majoritariamente negra (SEPPIR, 2015). Uma visão reducionista não consegue vislumbrar que a sociedade brasileira possui um racismo estrutural, que promove uma série de desdobramentos sociais: A população negra é vítima de discriminação, onde “a intensidade do preconceito varia em proporção direta aos traços negroides” (NOGUEIRA, 2006); e a população marginalizada socialmente sofre com restrições no acesso ao emprego, à educação e à saúde.

Os fatos intermédios, que influenciam a consequência de a população negra ser predominante no sistema prisional, são algo que a tecnologia apolítica não consegue enxergar.

O COMPAS, um sistema que foi utilizado em estados norte-americanos e tem como objetivo prever se um criminoso condenado tem probabilidades de reincidência, é um exemplo da incapacidade dos sistemas em lidar com inter-relações fenomenológicas.

As pesquisas revelaram que o sistema atribuía, predominantemente, um risco mais elevado de reincidência para os réus negros (embora eles voltassem a reincidir bem menos), enquanto atribuía risco menos elevado para réus brancos (que acabavam por reincidir mais) (LARSON; MAT-

TU *et al.*, 2016). Até mesmo o erro do algoritmo favorecia a estrutura de privilégio branco.

A neurocientista Natalia Mesa alerta para o olhar pragmático dos números sobre uma realidade mais complexa: “Chamar algoritmos de avaliação de risco de objetivo é ignorar os sistemas e estruturas sociais em que são criados” (MESA, 2021). Definitivamente o ser humano é mais do que um amontoado de dados.

Este tópico demonstra que a herança maldita não é algo simples de se lidar, sobretudo por muitas das desigualdades sociais não estarem explicitamente expostas nos dados, sendo necessário um olhar mais sensível sobre as peculiaridades sociais.

Os fenômenos sociais são mais do que abstrações numéricas e as inter-relações são mais complexas do que aparentam. Então, como trazer uma perspectiva mais social para a concepção da solução tecnológica?

3.3. O ENCONTRO DAS ÁGUAS

A área de Engenharia de Software por muito tempo produziu sistemas que não atendiam às necessidades dos utilizadores, isto acontecia, principalmente, por estes sistemas serem desenvolvidos sem considerar o utilizador.

A pesquisa na área avançou e o segmento da engenharia de requisitos desenvolveu várias técnicas para identificar o que de fato o utilizador necessita, o que muitas vezes destoa daquilo que o utilizador solicita.

Foi só então que o utilizador passou a ser considerado durante todas as fases do projeto, sobretudo nas mais embrionárias. Durante este processo de evolução foi vital a colaboração das Ciências Sociais.

O segmento de Interação Homem-Computador (IHC) surgiu como área transversal, tendo como objetivo principal sanar o problema da falta de usabilidade.

Estas áreas conseguiram realizar o encontro das águas, isto é, a união entre tecnicismo e humanismo, são segmentos que superaram os desafios de relacionamento com o utilizador e tornaram-se casos de sucesso. O que todas as soluções tinham em comum era justamente ter o utilizador afetado no centro de seu projeto e contar com uma colaboração multidisciplinar.

A Inteligência Artificial enfrenta atualmente um desafio que vai para além da complexidade técnica. A explicabilidade demanda esforços de múltiplas áreas do conhecimento e para lidar com os dados, considerando as complexidades das instituições sociais, se fazem necessários especialistas das Ciências Sociais.

O progresso na superação destes desafios só pode ser alcançado através de uma colaboração interdisciplinar, onde se possa combinar conhecimentos e teorias de diferentes campos de investigação, com o desenvolvimento de métodos e técnicas a partir desta multidisciplinaridade (ADADI; BERRADA, 2018, p. 52145).

O desafio passa a ser, então, vincular os resultados dos estudos empíricos de IHC com as teorias das Ciências Sociais e a partir desta fusão conduzir as abordagens de explicabilidade, contribuindo para construção de modelos explicáveis centrados no utilizador afetado (ADADI; BERRADA, 2018, p. 52155).

Nesse sentido, a pesquisadora Pratyusha Kalluri alerta: *“É hora de colocar as comunidades marginalizadas e afetadas no centro da pesquisa de Inteligência Artificial - suas necessidades, conhecimento e sonhos devem orientar o desenvolvimento”* (KALLURI, 2020). Os especialistas das Ciências Sociais possuem as habilidades necessárias para colaborar com esta nova forma de desenvolver Inteligência Artificial.

As pesquisas mais recentes confirmam que as Ciências Sociais e os estudos do comportamento humano ainda não estão tendo impacto suficiente na Inteligência Artificial Explicável (ADADI; BERRADA, 2018, p. 52155). Ainda existe uma longa jornada até que os projetos considerem o significado social e estejam centrados no utilizador.

Somente a interdisciplinaridade permitirá à Inteligência Artificial transladar da visão focada no utilizador para a centrada no utilizador. A primeira enxerga o utilizador com os olhos e interesses dos financiadores da tecnologia, já a segunda enxerga o mundo pelos olhos do utilizador.

CONSIDERAÇÕES FINAIS

O indivíduo sacrifica sua autonomia para se esquivar do fardo de tomar uma decisão e com isso imola não apenas sua soberania de tomada de decisão, mas também a liberdade de toda uma sociedade.

Esta história está apenas começando. Os responsáveis pela perpetuação do ciclo de exclusão social foram revelados: a Inconsciência Política, a Inexplicabilidade e a Inexistência de Interdisciplinaridade. Estes elementos são os pilares do preconceito codificado, que é a nova forma de manter e ampliar antigas discriminações.

Este ensaio manteve o debate entre as interseções das Ciências Sociais e da Inteligência Artificial, objetivando manter um diálogo cognoscível para ambas as áreas.

A temática levantada é demasiadamente densa para ser consumada em um único ensaio, por isso um dos objetivos desta obra é que a reflexão seja estopim da curiosidade. Para orientar os primeiros passos desta jornada, se indicam os seguintes tópicos de pesquisa: Inteligência Artificial Explicável (XAI), algoritmo de IA justo, responsável e transparente (FAT), Aprendizado de Máquina interpretável (iML), causalidade, justiça algorítmica, teorias de explicações e ciência social computacional.

Propõem-se aos especialistas das Ciências Sociais que convertam suas reflexões em práticas e que com isso possam direcionar suas habilidades para a área da Inteligência Artificial, com a pretensão de atender não só a concepção do artefato tecnológico, mas também promover a sensibilização dos especialistas em tecnologia e a mobilização da sociedade civil para o questionamento do significado social da tecnologia.

Dos profissionais da área tecnológica espera-se compreensão, receptividade e motivação. Compreensão para se aperceber que é possível causar dor, mesmo sem intenção. Receptividade para acolher a colaboração de outras áreas do conhecimento atribuindo a elas o mesmo respeito e relevância que outorga à sua área técnica. E motivação para empreender um autoconhecimento, onde possa se perceber como indivíduo de um coletivo, compreender quem é dentro do seu círculo social e como este círculo se posiciona no contexto social de sua comunidade.

A expectativa é que este ensaio provoque a visão de mundo do leitor e que para além disto promova um posicionamento crítico, uma vez que não pode ser mais aceitável viver em uma sociedade onde os não brancos, não homens, não héteros sejam considerados não humanos e precisem empreender uma luta contínua para o reconhecimento de sua condição de humano (THOMAZ, 2018, p. 5).

A esperança é de que todos os envolvidos escolham o lado humano desta história.

REFERÊNCIAS

ADADI, A.; BERRADA, M. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). **IEEE Access**, v. 6, p. 52138–52160, 2018. DOI: 10.1109/ACCESS.2018.2870052. Disponível em: <https://ieeexplore.ieee.org/document/8466590/>.

ALVES, P. **Twitter é acusado de racismo após ferramenta priorizar pessoas brancas**. 2020. TechTudo. Disponível em: <https://www.techtudo.com.br/noticias/2020/09/twitter-e-acusado-de-racismo-apos-ferramenta-priorizar-pessoas-brancas.ghtml>. Acesso em: 22 set. 2021.

ARISTÓTELES. **Política (Edição Bilíngue)**. [S.l.], Vega, 1998.

BENENSON, F. **What is Math washing?** 2021. Disponível em: <https://www.mathwashing.com/>. Acesso em: 10 set. 2021.

BÍBLIA. **A Bíblia Sagrada: Antigo e Novo Testamentos. Tradução de João Ferreira de Almeida**. Brasília: Sociedade Bíblica do Brasil, 1995.

BIRHANE, A. **Abeba Birhane no Twitter: “Not everything valuable can be measured and codified and ethics is one such thing. The sooner folk let go of attempting to find some accurate equation for “fairness” and start asking “why am I building this tool in the firs**. [S.d.]. Disponível em: <https://twitter.com/Abebab/status/1207025203471167488>. Acesso em: 18 set. 2021.

BRANDÃO, R.; CARBONERA, J.; FERREIRA, J. *et al.* **Mediation Challenges and Socio-Technical Gaps for Explainable Deep Learning Applications**. [S.l: s.n.], 2021.

BROUSSARD, M. **Artificial Unintelligence - How Computers Misunderstand the World**. [S.l.], MIT PRESS LTD, 2019.

- BUOLAMWINI, J. **Artificial intelligence: How to avoid racist algorithms**. 2017. BBC News. Disponível em: <https://www.bbc.com/news/technology-39533308>. Acesso em: 18 set. 2021.
- CASTELLS, M. **A sociedade em rede**. 6. ed. São Paulo: Paz e Terra, 1999.
- CHEVALIER, J.; GHERRBRANT, A. **Dicionário dos Símbolos**. Lisboa: Editorial Teorema, 1994.
- DASTIN, J. **Amazon scraps secret AI recruiting tool that showed bias against women | Reuters**. 2018. Disponível em: <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-againstwomen-idUSKCN1MK08G>. Acesso em: 19 set. 2021.
- EUROPEAN COMMISSION. **Ethics of Artificial Intelligence: Statement of the EGE is released**. 2018. Disponível em: https://ec.europa.eu/info/news/ethics-artificial-intelligence-statement-ege-released-2018-apr-24_en. Acesso em: 20 set. 2021.
- EYKHOLT, K.; EVTIMOV, I.; FERNANDES, E. *et al.* **Robust Physical-World Attacks on Deep Learning Visual Classification**. [S.l.: s.n.], 2018. Disponível em: <https://iotsecurity.eecs.umich.edu/#roadsigns>.
- FLORIDI, L.; COWLS, J. A Unified Framework of Five Principles for AI in Society. **Harvard Data Science Review**, v. 1, n. 1, 23 jun. 2019. DOI: 10.1162/99608f92.8cd550d1. Disponível em: <https://hdsr.mitpress.mit.edu/pub/l0jsh9d1/release/7>. Acesso em: 31 ago. 2021.
- GOODMAN, B.; FLAXMAN, S. European union regulations on algorithmic decision making and a “right to explanation”. **AI Magazine**, v. 38, n. 3, p. 50–57, 2017. DOI: 10.1609/aimag.v38i3.2741.
- HASANI, R.; LECHNER, M.; AMINI, A. *et al.* “Liquid Time-constant Networks”, 2021. Disponível em: www.aaai.org. Acesso em: 2 set. 2021.

HELLER, E. **A Psicologia das Cores - Como as cores afetam a emoção e a razão.** [S.l.], Editorial Gustavo Gili, 2014.

HIATT, K.; KLEINMAN, M.; LATONERO, M. Tech folk: “Move fast and break things” doesn’t work when lives are at stake. **The Guardian**, 2017. Disponível em: <https://www.theguardian.com/global-development-professionals-network/2017/feb/02/technology-human-rights>. Acesso em: 9 set. 2021.

HILL, K. **Wrongfully Accused by an Algorithm - The New York Times.** 2020. Disponível em: <https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html>. Acesso em: 20 set. 2021.

KALLURI, P. Don’t ask if artificial intelligence is good or fair, ask how it shifts power. **Nature**, v. 583, n. 7815, p. 169–169, 9 jul. 2020. DOI: 10.1038/d41586-020-02003-2. Disponível em: <http://www.nature.com/articles/d41586-020-02003-2>.

LARSON, J.; MATTU, S.; KIRCHNER, L. *et al.* **How We Analyzed the COMPAS Recidivism Algorithm.** 2016. ProPublica. Disponível em: <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>. Acesso em: 23 set. 2021.

LEDFORD, H. Millions of black people affected by racial bias in health-care algorithms. **Nature**, v. 574, n. 7780, p. 608–609, 1 out. 2019. DOI: 10.1038/D41586-019-03228-6.

MESA, N. **Can the criminal justice system’s artificial intelligence ever be truly fair?** 2021. Massive Science. Disponível em: <https://massivesci.com/articles/machine-learning-compas-racism-policing-fairness/>. Acesso em: 23 set. 2021.

NOGUEIRA, O. Preconceito racial de marca e preconceito racial de origem. **Tempo social - Revista de Sociologia da USP**, v. 19, n. 1, p. 287–308, 2006.

NUNES, P. **Maioria dos presos por reconhecimento facial são negros.** 2019. Disponível em: <https://theintercept.com/2019/11/21/>

presos-monitoramento-facial-brasil-negros/. Acesso em: 19 set. 2021.

PEREIRA, A. C. de S.; BRUNO, A. L. B.; AZEVEDO, A. M. de *et al.* Inteligência artificial e direitos humanos: impactos e dilemas éticos atuais. **Homa Publica - Revista Internacional de Derechos Humanos y Empresas**, v. 4, n. 1, p. e:059, 30 jan. 2020. Disponível em: <https://periodicos.ufjf.br/index.php/HOMA/article/view/30504>. Acesso em: 10 set. 2021.

PRETALAB. **PretaLab — Dados**. 2017. Disponível em: <https://www.pretalab.com/dados#>. Acesso em: 18 set. 2021.

SEPPPIR. **Mapa do Encarceramento aponta: maioria da população carcerária é negra — Português (Brasil)**. 2015. Disponível em: https://www.gov.br/mdh/pt-br/noticias_sepipir/noticias/junho/mapa-do-encarceramento-aponta-maioria-da-populacao-carceraria-e-negra-1. Acesso em: 22 set. 2021.

THOMAZ, D. **O BRANCO QUE NINGUÉM QUER SER: REFLEXÕES SOBRE A CONSTRUÇÃO DA IDENTIDADE RACIAL BRANCA NO CONTEXTO BRASILEIRO**. [S.l.], Copene, 2018.

TOYAMA, K. “Technology as Amplifier in International Development”, 2011.

UK PARLIAMENT. **House of Lords - AI in the UK: ready, willing and able? - Artificial Intelligence Committee**. 2017. Disponível em: <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/10002.htm>. Acesso em: 20 set. 2021.

UNIVERSIDADE DE MONTREAL. **The Declaration - Montreal Responsible AI**. 2017. Disponível em: <https://www.montrealdeclaration-responsibleai.com/the-declaration>. Acesso em: 20 set. 2021.

WORLD ECONOMIC FORUM. “M A R C H 2 0 2 1 Global Gender Gap Report The analysis presented in the Global Gender Gap Re-

port 2021”, p. 60, 2021. Disponível em: <http://reports.weforum.org/global->. Acesso em: 18 set. 2021.

ZHANG, Y.; CHEN, X. Explainable Recommendation: A Survey and New Perspectives. **Foundations and Trends® in Information Retrieval**, v. 14, n. 1, p. 1–101, 2020. DOI: 10.1561/15000000066. Disponível em: <http://dx.doi.org/10.1561/15000000066>.