

INTELIGÊNCIA ARTIFICIAL, EMPATIA E INCLUSÃO: UM PROBLEMA DE DESIGN

ARTIFICIAL INTELLIGENCE, EMPATHY AND INCLUSION: A DESIGN PROBLEM

Cid Boechat¹, M.Sc.

(1) Fundação CECIERJ
cboechat@cecierj.edu.br

Design, Inteligência artificial, Interação humano-computador

Pessoas já têm questões cruciais de suas vidas decididas por algoritmos matemáticos (avaliação de aprendizagem, concessão de pensões ou monitoramento de produtividade no trabalho, por exemplo). Este artigo questiona: como é a interação e a experiência dessas pessoas com essas ferramentas? Como ficam as questões de empatia e inclusão quando se tem decisões tomadas por máquinas?

Design, Artificial intelligence, Human-computer interaction

People already have crucial issues in their lives decided by mathematical algorithms (learning evaluation, granting pensions or work productivity monitoring, for example). This article asks: How human beings interact with these tools in these situations? What is their experience? How to overcome empathy and inclusion issues when machines take decisions?

1. Introdução

É cada vez maior o número de experiências e interações com Inteligências Artificiais (I.A.). Além de ferramentas mais simples como assistentes virtuais e *chatbots*, esses sistemas já são testados e projetados, entre outras aplicações, para atuar na detecção de doenças, acompanhamento de estudantes, medição de produtividade e contratação e demissão de funcionários.

A princípio, muitas empresas e pesquisadores defendem o uso desses sistemas apenas como ferramentas de apoio ao trabalho humano, sem substituir o poder decisório das pessoas. Mas, com a evolução dessa tecnologia, se essas máquinas se tornarem inalcançáveis em desempenho e economicamente mais baratas que os profissionais, essa opção se manterá? Nesse cenário, como será a experiência das pessoas ao interagir com sistemas decisórios artificiais? A Inteligência Artificial será capaz de compreender o ser humano em toda a sua complexidade de emoções, diferenças, diversidade e contradições?

Este artigo parte do entendimento do Design como a busca não só pela lucratividade, mas também pela melhora significativa da vida humana (BUCHANAN, 2015) focada nas necessidades

fundamentais das pessoas (BROWN, 2010). Desta forma, as questões abordadas neste artigo serão analisadas sob a ótica do Design e da importância da empatia e da inclusão nas relações humanas. Primeiramente, será apresentada a aplicação corrente de I.A. em diversas situações cotidianas e alguns problemas já detectados em decisões tomadas por máquinas. Em seguida, serão mostradas as discussões interdisciplinares em andamento sobre o assunto e exemplos de usos problemáticos da I.A. Por fim, a questão será problematizada de acordo com a ótica do Design – a importância da empatia, da centralidade do ser humano, do ato de projetar com foco em pessoas e não em tecnologias. Desta forma, o artigo buscará mostrar que a intervenção extensiva de Inteligências Artificiais no cotidiano humano pode se tornar um dos mais graves problemas da próxima década, e que o Design, através da sua abordagem e preceitos, analisando a questão de forma macro, deve participar do planejamento e discussão deste processo.

2. A Inteligência Artificial no cotidiano

O termo “Inteligência Artificial” foi cunhado por McCarthy (1956) como a ciência e engenharia capazes de construir máquinas inteligentes. Diferentemente da Filosofia e da Psicologia, a I.A. não se preocupa apenas em entender o pensamento

humano - se esforça para construir entidades capazes de fazê-lo. Assim, busca a criação de programas que possam emular nosso raciocínio, tomar decisões e definir ações (RUSSEL; NORVIG, 2003).

O trabalho da I.A. é projetar o “programa do agente”: uma função que implementa o mapeamento do agente das suas percepções até suas ações. O local onde esse programa é executado é chamado de “arquitetura”, o qual disponibiliza as percepções dos sensores, executa o programa e alimenta as opções de ação aos efetores, à medida que estas são geradas (RUSSEL; NORVIG, 2003).

Autores definem esses sistemas como a busca por “agentes inteligentes”, onde agentes são “qualquer coisa que possa ser vista percebendo seu ambiente através de sensores e atuando nesse ambiente através de efetores” (RUSSEL; NORVIG, 2003). O seu comportamento pode se basear em sua própria experiência e/ou no conhecimento interno, inicial, usado na sua construção para o ambiente específico em que opera. Um agente inteligente consegue ser mais autônomo na medida em que seu comportamento é determinado por sua própria experiência. Desta forma, nesses sistemas de I.A., especificamente nos subcampos da *machine learning* e *deep learning*, a máquina é programada com algoritmos - que lhe dão a capacidade de aprender - e alimentada com dados que lhe permitirão “ensinar” a si mesma (RUSSEL; NORVIG, 2003).

Os algoritmos de *machine learning* usam estatísticas para encontrar padrões em grandes quantidades de dados - números, palavras, imagens, cliques, etc. (HAO, 2017). Quanto mais ela é confrontada com dados e interações, mais ela aperfeiçoa, por si só, seu entendimento sobre um determinado assunto (o que lhe concederia “autonomia”). As inteligências artificiais e o *machine learning* são centralmente impactados pelos dados. Sem eles, não se pode treinar modelos de aprendizado de máquina. E, sem isso, os sistemas não aprendem com a experiência (SCHMELZER, 2020).

Um agente baseado em conhecimento é composto de uma base de conhecimento e um mecanismo de inferência. Ele opera armazenando sentenças sobre o mundo em sua base de conhecimento, usando o mecanismo de inferência para inferir novas sentenças e, com isso, decidindo que ação tomar (RUSSEL; NORVIG, 2003). Esse processo é potencializado no *deep learning*, onde uma técnica

chamada Rede Neural Profunda oferece às máquinas uma capacidade aprimorada de encontrar - e amplificar - até os menores padrões. Essa técnica possui muitas camadas de nós computacionais simples que trabalham juntos para extrair dados e fornecer um resultado final na forma de previsão (HAO, 2017). No cérebro humano, as sinapses neurais são fortalecidas e enfraquecidas pela ativação repetida; esses sistemas digitais buscam algo semelhante por meios matemáticos, ajustando os “pesos” das conexões para se moverem para o resultado desejado. Os sistemas mais poderosos (*deep learning*) se assemelham a camadas de neurônios, cada uma processando os dados de entrada e enviando os resultados para a próxima camada (MUKHERJEE, 2017).

Machine e *deep learning* podem ocorrer de três formas: supervisionada, não supervisionada e por reforço. No aprendizado supervisionado, o mais comum, dados são rotulados para informar à máquina exatamente quais padrões ela deve procurar. No não supervisionado, a máquina apenas procura os padrões que pode encontrar. Já o algoritmo de reforço aprende por tentativa e erro na busca por um objetivo claro. Ele tenta muitas coisas diferentes e é recompensado ou penalizado, dependendo se ajuda ou impede a obtenção do resultado esperado (HAO, 2017).

Todos esses processos podem permitir a detecção de padrões que os humanos não percebem ou nunca pensariam – ao menos no campo específico para o qual a I.A. foi programada e alimentada. Assim, essas aplicações poderiam ser utilizadas em praticamente qualquer área onde haja dados e/ou interações suficientes (VINCENT, 2019).

PENSAMENTO			
PERFORMANCE HUMANA	SISTEMAS QUE PENSAM COMO HUMANOS HAUGELAND (1985) BELLMAN (1978)	SISTEMAS QUE PENSAM RACIONALMENTE CHARNIAK; McDERMOTT (1985) WINSTON (1992)	RACIONALIDADE
	SISTEMAS QUE AGEN COMO HUMANOS KURZWELL (1990) RICH; KNIGHT (1991)	SISTEMAS QUE AGEN RACIONALMENTE SCHALKOFF (1990) LUGER; STUBBLEFIELD (1993)	
COMPORTAMENTO			

Figura 1 – Quatro diretrizes de Inteligência Artificial, divididas em categorias baseadas em definições sobre o tema (Baseado em RUSSEL; NORVIG, 2010)

Na Figura 1, vemos diferentes enfoques da I.A., divididos por como diferentes definições de Inteligência Artificial. Essas definições variam ao longo de duas dimensões principais. No topo da

figura, as que focam nos processos de pensamento e raciocínio; na parte de baixo, autores que abordam o comportamento. À esquerda, mede-se o sucesso em termos de desempenho humano, enquanto, à direita, as definições se guiam de acordo com um conceito ideal de racionalidade (RUSSEL; NORVIG, 2003).

Isso traria quatro diretrizes em Inteligência Artificial que representam diferentes utilizações dessa tecnologia. Um tipo de aplicação apenas ajudaria os humanos a reunirem insumos para tomarem decisões ou realizarem ações, enquanto outras buscariam a capacidade de realizar decisões ou ações tipicamente humanas por si próprias.

Atualmente, a I.A. e seus subcampos estão presentes na vida das pessoas em diferentes momentos: transportes, educação, vida acadêmica, entretenimento, trabalho, saúde, vida social. Entre incontáveis aplicações, é ela que filtra os *e-mails* que vão para a caixa de *spam*; quem calcula o preço da corrida ao pedir um carro através de aplicativo; quem avalia fraudes bancárias através do histórico de compras do cartão de crédito; quem recomenda livros, filmes e séries de acordo com seu histórico no *e-commerce* ou no serviço de *streaming* (NARULA, 2019).

Algumas das aplicações de I.A. mais adotadas pelas indústrias e serviços são: automação robótica de processos, *machine learning*, interfaces de conversação, visão computacional, compreensão de textos e falas em linguagem natural, geração de linguagem natural, robótica física e veículos autônomos (SHOHAM et al., 2018).

Estudo recente da consultoria Capgemini estima que em 2021 40% dos consumidores usarão assistentes de voz (alimentados por I.A.) em vez da página ou aplicativo de uma empresa. Hoje, este número já está em 24% (CASTELLANOS, 2018). A Inteligência Artificial também já é uma das áreas de crescimento mais rápido no mercado global de educação. Em 2018, investiu-se mais de um bilhão de dólares em I.A. educacional ao redor do mundo, e esse número deve ser seis vezes maior até 2025 (DAVIS, 2019). Por exemplo, o governo da região de Flandres, na Bélgica, assinou um acordo que prevê que suas setecentas escolas poderão utilizar uma plataforma de I.A. onde Ciência da aprendizagem, Neurociência e dados serão usados para personalizar o ensino de crianças. Será possível avaliar os níveis de foco e dificuldade dos alunos, o ritmo e o melhor momento para aprenderem, quanto

tempo se leva para a informação passar da memória de curto para longo prazo e quão resiliente é o aluno (ANDERSON, 2019).

No campo educacional, autores já preveem máquinas capazes de analisar cada estudante individualmente e apontar quais métodos e ferramentas de ensino funcionam melhor para cada um. Já se desenvolve um sistema de avaliação do desempenho futuro do aluno através das suas notas e trabalhos atuais. Inteligências artificiais também serão capazes de organizar currículos e esquematizar conteúdos de aulas. Ainda ajudarão professores a dar notas sem o “viés” humano e sugerirão metodologias e atividades baseadas no perfil do aluno (BOECHAT; MONT’ALVÃO, 2019).

Como pode ser visto na Figura 2, pesquisa global realizada em 2018 mostra que as organizações tendem a incorporar os recursos de I.A. em funções que fornecem mais valor em seu setor. Serviços financeiros a usam, em maior parte, na análise de risco, enquanto a indústria automotiva a emprega na manufatura. Já o varejo, foca no setor de marketing e vendas (SHOHAM et al., 2018). Modelos matemáticos estão sendo usados para ajudar a determinar liberdade condicional de presos, aprovação de empréstimos e contratação de funcionários (KNIGHT, 2017).

Outro campo de imenso potencial de atuação é o da saúde. Já existem, por exemplo, estudos utilizando algoritmos na detecção de problemas mentais (ZARLEY, 2019), doenças nos olhos (VINCENT, 2018) e no diagnóstico de câncer de pele (MUKHERJEE, 2017).

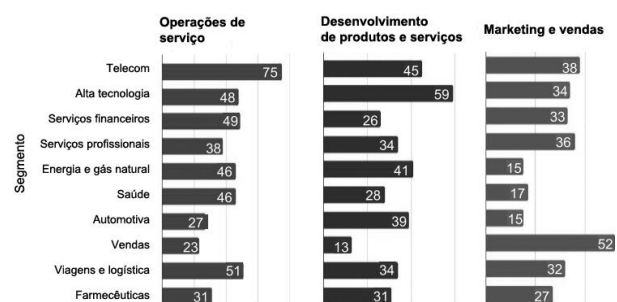


Figura 2 – Adoção de I.A. por setor e função em 2018 (Baseado em SHOHAM et al., 2018)

O crescimento da importância da Inteligência Artificial também se reflete na produção acadêmica. Pesquisa realizada no banco de dados Scopus, da Elsevier, que indexa quase 70 milhões de documentos, comparou taxas de publicação de

trabalhos acadêmicos em todos os campos e, especificamente, em Inteligência Artificial e Ciência da Computação entre 1996 e 2017. Como se percebe na Figura 3, o crescimento de artigos publicados anualmente em I.A. continua superando os da computação. Os *papers* sobre I.A. aumentaram sete vezes desde 1996, enquanto os sobre Ciência da Computação aumentaram apenas cinco no mesmo período (SHOHAM et al., 2018).

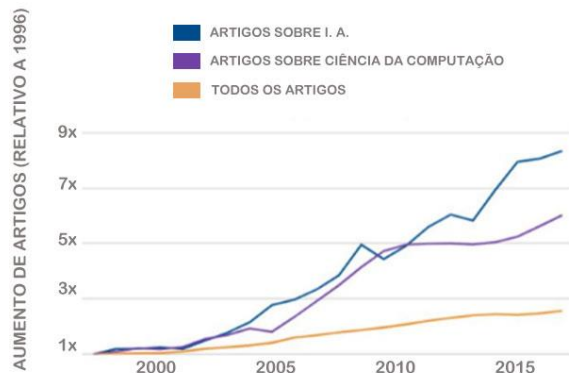


Figura 3 – Crescimento de artigos publicados na Scopus anualmente, por tópico, entre 1996 e 2017 (Adaptado de SHOHAM et al., 2018)

3. Problemas e questões da Inteligência Artificial

Inteligências Artificiais tendem a refletir os dados utilizados para treiná-las. Assim, qualquer viés, inconformidade ou erro nas informações e interações do sistema irá refletir no seu “aprendizado”. Isso aconteceu, por exemplo, com o *chatbot* (sistema virtual capaz de gerar conversas que simulam a linguagem humana) Tay, da Microsoft. Projetado para responder perguntas e conversar com *millennials* no Twitter, acabou retirado do ar. Isso ocorreu, pois os insultos preconceituosos e piadas a que foi submetido em sua interação com humanos o fez elaborar mensagens racistas, sexistas e xenófobas. Sua conversação, projetada para ser “casual e brincalhona”, não foi capaz de lidar com o material que o alimentou. O *chatbot* chegou a apoiar genocídios e chamou uma mulher de “p... estúpida”. Esse tipo de comportamento pode ser visto na Figura 4. Outro caso aconteceu com o site de fotografias Flickr, onde a marcação intuitiva de fotos, baseada em I.A., deu o título de “chimpanzés” a imagens de férias de várias pessoas negras (CANO, 2016).

Com a I.A. sendo usada para gerar respostas quase humanas, a ética começa a ser uma parte central das discussões neste campo. Se um desenvolvedor tem

um viés negativo em relação a uma determinada coisa, este pode se infiltrar na solução e afetar o resultado das respostas (PTI, 2020). No Canadá, o governo discute a aplicação da Inteligência Artificial e do *machine learning* na sociedade em geral, mas se preocupa com os riscos éticos e consequências desconhecidas. Algoritmos que aprendem através da experiência muitas vezes não podem ser questionados. Segundo o primeiro ministro do país, “talvez o computador não possa justificar uma decisão e não poderemos explicar porque a tomamos. É um desafio ético e moral” (KNIGHT, 2017 *apud* BOECHAT; MONT’ALVÃO, 2019). Essa preocupação é válida, pois existe a chamada “caixa preta” do I.A, especialmente no *deep learning*. O sistema é alimentado com dados, entrega respostas, mas como se “auto ensina”, não há como saber o que ele efetivamente sabe nem qual parâmetros usou para chegar aos seus resultados (MUKHERJEE, 2017).



Figura 4 – Desempenho do *chatbot* Tay antes e depois de interações com humanos (Fonte: reprodução da internet)

Esse tipo de aplicação não é guiado por um armazenamento explícito de conhecimentos, ele faz um grande número de ajustes internos - análogo ao fortalecimento e enfraquecimento das conexões sinápticas no cérebro. A diferença é que a rede neural animal não chega aos seus resultados através de aritmética. Como um sistema profundo de I.A. determina seus resultados finais? Por que apresenta recursos e respostas extraídos de milhões de outros recursos e possibilidades? Não há como saber, e ele não sabe nos dizer (MUKHERJEE, 2017).

Embora os cientistas da computação trabalhem nisso, eles admitem que um sistema de *deep*

learning não tem poder explicativo (ZARLEY, 2019). Quanto mais poderoso o sistema, mais opaco ele pode se tornar. Todos os ajustes e processos internos que permitem que a I.A. “aprenda” e tome decisões acontecem longe da supervisão humana (ZARLEY, 2019).

Isso torna extremamente difícil detectar como esses sistemas acertam e – mais grave – erram. É possível que haja suposições e preconceitos ocultos nos dados e, portanto, nos sistemas criados a partir deles, além de uma ampla gama de outras opções de sistema que podem ser impactadas por preconceitos e suposições (PARTNERSHIP ON AI, 2019).

Este cenário leva à grande questão envolvendo a I.A.: o tipo de uso que se fará dela. Bancos, militares e empregadores, entre outros, buscam abordagens de *machine learning* mais complexas, com tomadas de decisão automatizadas que não podem ser avaliadas (KNIGHT, 2017). Inteligências Artificiais não tem o senso comum dos humanos. São treinadas, em geral, para tarefas extremamente específicas. Uma I.A. especializada em identificar felinos, por exemplo, não é capaz de dizer que um gato não pode dirigir um carro (VINCENT, 2019).

Além das questões de segurança, existe outro problema importante no uso da I.A., especialmente em casos de tomada de decisão: Inteligências Artificiais não possuem, ainda, empatia. Para que isso ocorra, uma máquina teria que experimentar emoções, reconhecê-las e compreendê-las.

Morgan (2018) nos lembra que diariamente pessoas tem interações frustrantes com máquinas incapazes de entender suas mensagens e emoções. Humanos nem sempre se comunicam claramente com as máquinas e vice-versa. Essas ineficiências de entendimento e comunicação tendem a ser irritantes. Dependendo da situação, interagir com um interlocutor virtual que não entende o problema do usuário num *chat* ou telefonema pode ser até mesmo perigoso. Ainda mais frustrante em tais cenários é o fracasso do sistema em reconhecer emoções e se adaptar a elas (MORGAN, 2018).

Para Michael Banissy, Professor e chefe do Departamento de Psicologia da Universidade de Londres, empatia é parte fundamental da vida humana. Contribui para “a colaboração e cooperação e transcende culturas, comportamentos sociais e bem-estar” (DEDEZADE, 2019). Segundo ele, um sistema de I.A. com capacidade de empática poderia

fornecer interações mais naturais ao responder, levando humores e sentimentos humanos em consideração. Dependendo do tipo de informação que se deseja que a I.A. transmita, como deseja que reajam a ela ou como prevê que eles vão reagir a ela, seria necessário entender o estado atual dos humanos interagindo ou sendo impactados por ela (DEDEZADE, 2019).

Bill Mark, da SRI International, cuja equipe de I.A. inventou a assistente Siri do *iPhone*, opina que apesar do relativo sucesso no reconhecimento de emoções por sistemas artificiais, entendê-las é mais difícil. Máquinas podem ter algum sucesso no reconhecimento emocional, ao utilizarem grandes volumes de dados rotulados - padrões da fala (uso de palavras no contexto, inflexão de voz, etc.), bem como linguagem corporal, expressões e gestos, novamente com ênfase no contexto. Isso possibilitaria reconhecer características associadas a uma emoção específica (MORGAN, 2018). Porém, cabe lembrar que reconhecer uma emoção é diferente de entendê-la em toda a sua complexidade. Seres humanos tendem a interpretar e experimentar emoções de maneira diferente. Atualmente, os algoritmos de aprendizado reconhecem algumas características fisiológicas ou comportamentais associadas a estados emocionais específicos (MORGAN, 2018). No entanto, as emoções mais sofisticadas são muito mais complicadas. Não é tão simples identificá-las como é reconhecer o tom de voz, a postura ou a expressão facial de pessoas (DEDEZADE, 2019).

Para Mark, os computadores ainda estão longe de entender emoções e ainda mais distantes de alcançar empatia artificial (MORGAN, 2018), já que é possível para humanos entender emoções de outras pessoas sem se tornarem empáticos a elas. Sua empresa, no momento, trabalha em softwares baseados em *machine learning* e *deep learning* para entender o “estado de usuário” de uma pessoa - emoção, cansaço, depressão, entre outros - através de elementos como entonação vocal e elementos visuais (CASTELLANOS, 2018). Alguns pesquisadores e empresas, como a Microsoft, acreditam que surgirá a profissão de “empatologista”: alguém com formação acadêmica em empatia que entenderia as nuances emocionais humanas e trabalharia em algoritmos para entender melhor os indicadores emocionais de usuários. Isso ajudaria os sistemas a ir além da superfície das expressões (DEDEZADE, 2019). Porém, cabe lembrar que a escolha desses profissionais deve seguir preceitos como diversidade cultural, racial e

de gênero.

Para Dedezade (2019), um sistema de I.A. é tão bom quanto o modelo em que é construído - se ele é alimentado com um espectro restrito de dados, sua funcionalidade será mais limitada. Por isso, a reunião de diferentes pontos de vista e experiências podem gerar melhores resultados finais. Projetar um sistema com uma equipe mais diversa e alimentá-lo com dados diversificados resultaria em algoritmos capazes de prever ou reconhecer com mais precisão as emoções existentes no mundo real e fornecer funcionalidades mais significativa como resultado (DEDEZADE, 2019). Opinião compartilhada pelo CEO da Microsoft, Satya Nadella, para quem os desenvolvedores devem trabalhar com equipes com diversidade, para que “vieses inconscientes” não se infiltrem nos modelos usados para o desenvolvimento da Inteligência Artificial. A cada escolha dos desenvolvedores, a ética de uma equipe diversificada pode ser importante em termos de criação de um mundo mais inclusivo. Ainda segundo ele, o mais importante ativo dos próximos dez anos será a diversidade e a inclusão das equipes que estão construindo todas essas tecnologias, pois não há outra proteção contra preconceitos inconscientes (PTI, 2020).

Já estão disponíveis ferramentas para ajudar os desenvolvedores de I.A. a evitar possíveis falhas e preconceitos em seus aplicativos. Por exemplo, empresas como Google e IBM oferecem *cards* de modelo, explicações de inteligência artificial e *kits* de ferramentas com informações sobre os algoritmos e a tomada de decisões por trás de seus aplicativos e como se pode melhorá-los (IEEE, 2020). Porém, pesquisa do AI Now Institute, publicada no veículo de notícias El País, revelou que menos de 20% dos pesquisadores que participam de conferências sobre I.A. são mulheres, e que essas representam somente um quarto dos estudantes universitários de I.A. em Stanford e na Universidade da Califórnia. Essa falta de diversidade de gênero pode causar falhas nos algoritmos, como a ocorrida num sistema de contratação da Amazon que discriminava os candidatos com nomes femininos (TWENGE; MATEOS-GARCIA; JOHN, 2019).

Asada (2015) ainda nos lembra que, ao se pensar em empatia artificial, fatores contextuais, como experiências de apego, estados de humor e outras contingências ambientais devem ser sempre considerados, já que são capazes de modular a empatia.

Dados diversos, bem rotulados e de qualidade não só ajudam a ter algoritmos mais funcionais, mas também sistemas mais “justos”. Como já dito, inconformidade ou erro nas informações e interações que alimentam as I.A. podem se traduzir em decisões prejudiciais. Essa é uma das preocupações da iniciativa Partnership on AI, para quem esse tipo de problema pode levar a ações e recomendações que repliquem vieses e pontos cegos sérios. Essa organização defende que pesquisadores, governos e o público geral precisam ser sensíveis a essas possibilidades e desenvolver métodos que detectem e corrijam esses erros e vieses, para não os replicar (PARTNERSHIP ON AI, 2019).

Todos esses fatores tornam ainda mais complicada uma discussão relevante sobre I.A. no momento: as questões éticas e humanas de sua aplicação. E isso vem gerando preocupações e debates em várias esferas.

4. O debate multidisciplinar sobre o uso da I.A.

O Impacto da Inteligência Artificial na sociedade é discutido e estudado em várias frentes, entre elas: as mudanças que causará na economia, no mercado de trabalho, no (des) aparecimento de empregos e funções, na segurança e privacidade dos dados que alimentam as I.A. e nas questões éticas da sua aplicação.

A citada iniciativa Partnership on AI, por exemplo, reúne acadêmicos, pesquisadores, organizações da sociedade civil, gigantes como Amazon, Apple e Microsoft e outras empresas e grupos que trabalham para entender melhor os impactos da I.A. Eles estudam e buscam formular as melhores práticas, para aprimorar o entendimento do público e para servir como uma plataforma aberta para “discussão sobre a I.A. e suas influências nas pessoas e na sociedade” (PARTNERSHIP ON AI, 2019). Os seus principais pilares temáticos são: segurança em aplicações críticas; transparência; economia e trabalho; colaboração entre humanos e sistemas; e influência da I.A. na sociedade (PARTNERSHIP ON AI, 2019).

No campo do trabalho, matéria da Agência Brasil destaca que o Fórum Econômico Mundial de Daros de 2016 abordou a chamada “Quarta Revolução Industrial” – uma profunda mudança global nas relações sociais econômicas e de trabalho causada por tecnologias como Internet das Coisas, Big Data

e a Inteligência Artificial. Cenário que demandará dos governos uma estratégia de transição que mitigue questões do desemprego e desigualdade (BOECHAT; MONT'ALVÃO, 2019). Espera-se que, ao programar máquinas para pensar mais como seres humanos, o trabalho seja impactado numa escala semelhante à da introdução da robótica e dos computadores de mesa. Porém, é muito mais provável que os funcionários afetados ocupem cargos gerenciais ou profissionais do que trabalhadores “braçais” (MORATH, 2020). Discute-se também se a automação em funções e tomadas de decisão podem modificar a qualidade do trabalho e do pensamento humano. Pesquisadores e médicos entrevistados por Mukherjee (2017) questionam como o chamado “viés de automação” teria influência no comportamento das pessoas. Assim como o corretor ortográfico no celular pode prejudicar o aprendizado de uma criança e o carro autônomo pode tornar motoristas menos alertas, qual o impacto de ferramentas decisórias na prática médica, por exemplo? Teremos profissionais “refêns” da tecnologia? Em outros campos, como no de recursos humanos, podemos nos tornar apenas “carimbadores” de decisões matemáticas?

O uso de novas tecnologias no ensino traz debates sobre a extensão do seu impacto na aprendizagem (especialmente de crianças) e na pressão sobre os alunos. Autores como Martinianos (2016), defendem uma formação em Engenharia de Tecnologias Educativas, envolvendo Educação, Gestão estratégica de projetos, Recursos Humanos e Neuroeducação (BOECHAT; MONT'ALVÃO, 2019).

No campo acadêmico, a Inteligência Artificial é tema de cada vez mais trabalhos, artigos, novos cursos e matrículas, crescendo não são em números totais, mas também na variedade de assuntos abordados e subcategorias (SHOHAM et al., 2018).

Porém, infelizmente, já se pode ver exemplos de uso discutível da tecnologia. Na China, uma escola instalou um sistema com tecnologia de reconhecimento facial para monitorar a atenção dos alunos nas aulas. Os alunos são analisados em todos os seus movimentos por três câmeras posicionadas acima do quadro-negro. O sistema identifica diferentes expressões faciais dos alunos, e essas informações vão para um computador que avalia se eles estão interessados nas aulas ou distraídos. Se o computador concluir que o aluno está com outros pensamentos durante a aula, enviará uma notificação

ao professor para tomar uma atitude (CONNOR, 2018).

Boechat e Mont'Alvão (2019) citam Barros (2014) ao apontar que muitos autores enxergam problemas na aplicação dessas novas tecnologias na aprendizagem. Afinal, se cada passo educacional de uma pessoa for individualmente medido, registrado e analisado em tempo real, de uma forma muito mais minuciosa do que hoje, isso não poderia se tornar uma experiência sufocante? Qual o impacto pedagógico e psicológico disso a longo prazo?

Em outro caso, documentos mostraram que a Amazon se utiliza de um sistema de controle de produtividade totalmente automatizado em seus armazéns (LECHER, 2019). Ele rastreia as taxas individuais de produção de cada empregado, mede o tempo que cada um gasta *off task* e gera automaticamente advertências ou até demissões, sem a participação de humanos (a empresa diz que os supervisores são capazes de sobrescrever as decisões). Segundo relatos, os trabalhadores se queixam de serem tratados como robôs, porque são monitorados e supervisionados por esses sistemas automatizados. Se os funcionários deixarem suas funções por muito tempo, o sistema gera automaticamente avisos e, eventualmente, o funcionário pode ser demitido. Alguns dizem evitar até ir ao banheiro para manter seu tempo dentro do esperado. Entre 2017 e 2018, a empresa demitiu 10% do seu quadro por problemas de produtividade (LECHER, 2019).

Na China, documentos vazados mostram que se utiliza uma plataforma de operação conjunta integrada - um programa de vigilância e policiamento preditivo chamado IJOP - na região de Xinjiang, habitada por uma comunidade predominantemente muçulmana que fala sua própria língua turca. Segundo a Human Rights Watch, a plataforma coleta grandes quantidades de informações pessoais sobre os cidadãos e, em seguida, usa inteligência artificial para formular listas das chamadas pessoas suspeitas. As informações viriam de postos de controle de Xinjiang, câmeras de circuito fechado com reconhecimento facial, *spywares* obrigatoriamente instalados em telefones, *snifers* de *wi-fi* que coletam identificação de informações de *smartphones* e computadores e até entregas de pacotes (ALLEN-EBRAHIMIAN, 2019). Através da coleta de dados em larga escala, usa-se a Inteligência Artificial e o *machine learning* para prever com antecedência

onde possíveis incidentes podem ocorrer e identificar populações que têm propensão a se envolver em ações anti-regime e anti-estatais. A polícia e outras autoridades usam um aplicativo móvel para executar verificações de antecedentes e se comunicar com o IJOP em tempo real. Ainda de acordo com a Human Rights Watch, o programa coleta e interpreta dados sem levar em consideração a privacidade e sinaliza pessoas comuns para investigação com base em critérios como oração diária, viagens ao exterior ou frequência de uso da porta dos fundos de sua casa. Documentos obtidos mostram que em um período de sete dias em junho de 2017, oficiais de segurança prenderam 15.683 residentes de Xinjiang sinalizados pelo IJOP e os colocaram em “campos de internamento” (além de 706 pessoas presas formalmente). O IJOP produziu 24.412 nomes de "suspeitos" naquela semana (ALLEN-EBRAHIMIAN, 2019).

Fica claro que limites éticos e humanísticos são fundamentais para um uso da Inteligência Artificial que possa explorar todo o seu potencial de melhoria na vida das pessoas. À medida que a I.A. continua a crescer e a se integrar a vários aspectos dos negócios, talvez nunca tenha havido uma necessidade maior de treinamento prático em inteligência artificial e ética (IEEE, 2020). Para o CEO da Microsoft, Satya Nadella, os desenvolvedores devem se concentrar na ética e construir confiança ao desenvolver soluções para garantir que os benefícios da tecnologia estejam disponíveis para todos (PTI, 2020).

Alguns dos casos aqui mostrados levantam um uso especialmente perigoso da I.A.: a de ferramenta decisória sem supervisão humana. Como visto na Figura 1, existe um campo da I.A. que busca reproduzir o pensamento e as ações das pessoas. Os responsáveis pela Partnership on AI lembram que, ao usar as ferramentas de I.A. para complementar ou substituir a tomada de decisão humana, precisa-se ter certeza de sua segurança, confiabilidade e alinhamento com a ética e as preferências das pessoas influenciadas por suas ações (PARTNERSHIP ON AI, 2019). Por enquanto, a maioria das aplicações funcionam apenas como ferramentas que nos auxiliam a tomar decisões. Mas se sua evolução tornar a Inteligência Artificial muito mais precisa, eficiente e barata do que contratar pessoas, teremos em nosso cotidiano processos e interações desprovidos de um interlocutor humano? As discussões éticas e humanísticas terão espaço suficiente num mercado bilionário que não para de

crescer? Com estas questões postas, qual pode ser a abordagem e atuação do Design neste cenário?

5. O uso da Inteligência Artificial sob a ótica do Design

Questões de empatia, centralidade do ser humano, inclusão e interação fazem da I.A. um problema especialmente talhado à abordagem do Design. Ao se pensar em Inteligência Artificial sob essa ótica, existe um problema que pode ser considerado estrutural, especialmente sob a luz do Design Centrado no Usuário: A I.A. não foi criada e desenvolvida a partir das demandas e necessidades das pessoas. Ela é uma tecnologia criada *a priori*, e, partir de sua evolução, busca-se utilizá-la de acordo com necessidades ou oportunidades de mercado. Prestero (2010) reafirma que uma abordagem de *design* é uma alternativa à invenção. A diferença mais básica é que, onde a invenção geralmente leva a uma tecnologia na busca por usuários (ou a uma solução em busca de alguém que tenha esse problema), o Design começa com o usuário e depois busca a tecnologia (Figura 5).



Figura 5 – As abordagens de Invenção e de Design no processo de desenvolvimento de produtos (Baseado em PRESTERO, 2010)

Na abordagem centrada na invenção, começa-se especificando a tecnologia que se espera que resolva o problema. Ajusta-se a tecnologia para o problema através de uma série iterativa de aprimoramentos. Finalmente, tendo refinado o produto para o que se espera ser uma ferramenta útil, busca-se um grupo de usuários ou segmento de mercado específico para o qual o produto é compatível. Uma falha comum em se definir a tecnologia antes de se conhecer o usuário é que este último arca com o ônus da adaptação (PRESTERO, 2010).

A I.A. é uma ferramenta poderosíssima que talvez esteja sendo usada antes da hora, ou ao menos antes de se conhecer todos os seus efeitos colaterais. Alguns relatos de problemas apresentados neste artigo parecem indicar uma aplicação ainda calcada na tentativa-e-erro. Casos como o do monitoramento

de funcionários da Amazon e o de alunos chineses provavelmente apresentarão ganhos em notas e produtividade, mas foram pensados levando em conta as pessoas que mais efetivamente têm sua vida impactada pela tecnologia? Prestero (2010) ainda nos lembra que, no final, a mais importante pergunta a ser feita na pesquisa de novos projetos não é "podemos fazê-lo?", mas "devemos fazê-lo?" Outra questão relevante sob a ótica do Design é a ausência de empatia em ferramentas decisórias de IA. O Design busca a solução de problemas usando empatia para identificar e contextualizar necessidades e depois as converter em especificações que orientam o desenvolvimento de novos produtos e serviços (PRESTERO, 2010). O autor ainda define empatia como a capacidade de imaginar o mundo de alguém na perspectiva de outra pessoa. Já Asada (2015), a define como a capacidade de formar uma representação do estado emocional dos outros, ao mesmo tempo em que se está ciente do mecanismo causal que induz esse estado emocional. Ainda segundo o autor, interações empáticas são importantes na realização da verdadeira comunicação. Para Wright e McCarthy (2008), a empatia evolui no contexto de relacionamentos contínuos em que uma pessoa aprende sobre as necessidades da outra, às vezes acertando e às vezes não, e depois afinando futuras respostas empáticas. Sem esse enquadramento comunicativo e relacional, cada pessoa tem apenas a própria experiência para basear sua resposta emocional.

A empatia tem sido usada como característica definidora das relações *designer*-usuário quando o Design se preocupa com a experiência do usuário (WRIGHT; McCARTHY 2008). Na ausência de um mediador na interação Homem-IA, a Inteligência Artificial passa de ferramenta a interlocutor e, de certa forma, a projetista da experiência - já que possui a prerrogativa decisória. Assim, a necessidade de empatia e compreensão contextual dos *designers* ao projetar produtos e experiências deveria ser também exigida dessas Inteligências.

Uma abordagem pragmática sobre uma experiência trata como inseparáveis as respostas intelectuais, sensuais e emocionais das pessoas, o que conceitua indivíduos, artefatos e configurações como múltiplos centros de valor interagindo uns com os outros (WRIGHT; McCARTHY, 2008). Do ponto de vista pragmático, "conhecer o usuário", em suas vidas e sentimentos, envolve entender como é ser a pessoa, como é a sua situação e perspectiva. É, também, a

apreciação do contexto em que ele está inserido (PRESTERO, 2010). Wright e McCarthy (2008) referenciam Mattelmäki e Battarbee (2002), que veem um projeto empático como uma conexão pessoal entre projetista e usuário, o que facilita a visualização e entendimento dos usuários em sua própria posição e perspectiva - como pessoas com sentimentos ao invés de objetos de teste. Em outras palavras, é na tomada de perspectiva envolvida na interação social que as perspectivas passam a constituir-nos como "eus" compreensivos e agentes (WRIGHT; McCARTHY, 2008). Essas I.A. terão essa capacidade por si só? É possível, ou aconselhável, subtrair o interlocutor humano nessa interação?

Se experiência e interpretação são fundamentais para as relações *designer*-usuário na IHC, os métodos empáticos se tornam importantes e devem ser entendidos e usados de maneira apropriada. O perigo é que a empatia seja considerada algo vago, misterioso, injustificável, não documentável e inutilizável (WRIGHT; McCARTHY, 2008).

Em um relacionamento empático, o projetista não renuncia à sua posição de "tornar-se usuário", onde nada de novo pode ser criado; antes, ele responde ao que vê como mundo do usuário a partir de sua própria perspectiva como *designer*. Ao manter sua própria perspectiva, cada pessoa é capaz de responder criativamente à outra a partir de sua própria perspectiva (WRIGHT; McCARTHY, 2008). Isso encontra eco, de certa forma, na ótica do Design Socialmente Responsável, onde a abordagem empática é valorosa, mas não suficiente, dentro do prisma dialógico - onde *designers* buscam ser "totalmente presentes" na realidade que estão projetando (CIPOLLA; BARTHOLLO, 2014).

Ainda para Cipolla e Bartholo (2014), apenas a empatia não basta. Pensando numa abordagem de responsabilidade e diálogo, a empatia traz a distância entre o *designer* e outros participantes, pois não promove a inclusão. Se empatia em *design* significar tratar demais participantes como "outro Eu", não se está efetivamente vendo os outros, apenas imagens projetadas de si mesmo. Nesse sentido, Wright e McCarthy (2008) defendem que adotar uma perspectiva dialógica não diminui a importância da intuição e da agência individuais. Citando Buber (1947/2006), Cipolla e Bartholo (2014) definem a diferença entre empatia e inclusão, sendo que esta última envolve encontros em que o outro não é um "Isso" - que também se pode descrever e manifestar empatia - mas sim um "Ele",

com quem se dialoga. Citando Friedman (2002), os autores lembram que se o “Isso” não é um “Ele”, também não é “outro Eu”.

Inclusão demanda a presença completa dos *designers* na realidade em que estão participando. Isso significa que cada *designer* precisa desempenhar duas funções: como um facilitador que guia o processo de *design* e, simultaneamente, como alguém incluído, que entra em relações com outras pessoas para buscar soluções para problemas compartilhados pelos envolvidos, incluindo ele mesmo (CIPOLLA; BARTHOLLO, 2014). Essa visão pode ser relacionada com a necessidade de se alimentar a I.A. com dados que promovam diversidade e equilíbrio. Não bastaria apenas conseguir desenvolver empatia artificial, é preciso também garantir dados e interações que não promovam viés. É preciso ter noção de diversidade, inclusão e contexto. E esse é mais um problema que pode ser abordado pela visão do Design. Ao discorrer sobre o Design Socialmente Responsável (SRD), Cipolla e Bartholo (2014) citam Papanek (1985), que faz um forte apelo ao *design* responsável, incluindo suas dimensões sociais, morais e ecológicas.

Ou seja, um sistema que forneça respostas que facilitem nossas vidas - mas que não compreende ou compartilhe de nossas experiências - é completo o suficiente para tomar decisões que nos impactem, de forma autônoma? Pode ser o “projetista” das “experiências” a que humanos serão submetidos?

Uma pergunta que ainda é alvo de vários debates é: quem é o responsável pelas decisões de uma inteligência artificial profunda e “autônoma”? Os autores do seu algoritmo? Os responsáveis pelos dados que alimentam o sistema? As empresas ou pessoas que contratam os profissionais? Mesmo que seu trabalho seja extremamente bem feito, como medir a extensão de suas responsabilidades no caso de algum viés, se o sistema não é capaz de dizer como e em que bases chegou a sua decisão? Como corrigir uma distorção? Ao se suprimir a interlocução humana, as I.A. “projetistas” serão as responsáveis pelas suas decisões? Elas terão a capacidade de encontrar esse termo dialógico e inclusivo que se exige de um projetista? Como atribuir um senso de responsabilidade a essas ferramentas?

Cipolla e Bartholo (2014) incorporam uma noção dialógica de responsabilidade, que é cumprida

através da concretude e da presença de relações interpessoais que ocorrem em contextos locais. Como ferramentas baseadas em algoritmos podem desenvolver essa presença, diálogo e entendimento do contexto? Citando Buber (2005), Cipolla e Bartholo (2014), ainda afirmam que ninguém pode ser responsável por um “Isso”, por uma entidade impessoal - só se poderia ser responsável por um “Ele”. Será possível passar a essas ferramentas a sensibilidade de compreender que o objeto de sua avaliação é um “ser” e não uma “coisa”? Conseguirão elas entenderem o que é um “ser” em toda a sua complexidade, se nós mesmos, muitas vezes, não temos uma resposta exata para isso?

6. Conclusão

Por todos os aspectos aqui abordados, que são apenas alguns numa vasta gama, este artigo entende que a interação Homem-IA poderá ser um dos mais importantes aspectos globais na próxima década. Considerando-se todo o potencial positivo e negativo do uso de I.A., é preciso levar a abordagem do Design para toda extensão da questão, de forma macro, conceitual e sistêmica. Será necessário propor coletivamente frameworks e *guidelines* de funcionamento e projeto dessas ferramentas. Especialmente pelas questões de empatia e centralidade do ser humano, é fundamental que o Design enquanto disciplina participe ainda mais ativamente das discussões multidisciplinares sobre o tema em diversos fóruns.

É positivo que haja iniciativas individuais como a da PAIR (People + AI Research) da Google, que discute formas de construir ferramentas de I.A. centradas no ser humano através das práticas da User Experience e do Design Centrado no Usuário. As metodologias comuns ao Design podem contribuir num melhor projeto de sistemas e levantamento de dados que levem os usuários e seu contexto em consideração. Campos como a Arquitetura da Informação podem também contribuir na rotulação e organização dos dados que alimentarão os algoritmos da I.A. Porém, preocupa que no documento oficial do governo brasileiro sobre o tema, o E-Digital, o Design, enquanto disciplina, não seja citado nenhuma vez.

Prestero (2010) lembra que uma das lições mais importantes e dolorosas é: o que é necessário nem sempre é interessante ao mercado. E que a abordagem de invenção para a solução de problemas é e continuará sendo popular entre os que

desenvolvem novas tecnologias. Segundo o autor, nosso maior desafio é que muitas vezes um bom *design* não é necessário nem suficiente para que um produto seja bem-sucedido (PRESTERO, 2010).

É inegável que o uso de processos de *machine e deep learning* trazem e trarão benefícios que melhorarão nossas vidas em vários campos como a medicina – e não é intuito deste artigo demonizar esta questão. Ainda assim, se Design é a busca não só pela lucratividade, mas também pela melhora significativa da vida humana (BUCHANAN, 2015), focando nas necessidades fundamentais das pessoas (BROWN, 2010), é papel dos *designers* lembrar, a todo momento, a importância das questões humanas – especialmente no caso de ferramentas de avaliação e monitoramento, onde os “clientes” não são as pessoas efetivamente mais impactadas pelo sistema. E este é um problema de *design*, pois ao se transformar uma ferramenta em interlocutor com poder de decisão, pessoas estarão interagindo com sistemas que muitas vezes não fornecem *feedback* completo e que não são capazes de compreender questões humanas e contextuais ao projetar e responder. Se pensarmos nos parâmetros presentes na ISO 9241-210, que aborda o projeto de sistemas interativos através do *design* centrado no usuário (DCU), veremos que esses são dois graves problemas. Como garantir que esses sistemas são projetados e compreendem o seu contexto de uso, se não tem entendimento completo sobre a complexidade e as particularidades do ser humano? Se duas das buscas do DCU são a satisfação do usuário e o bem-estar humano (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2011), como fazê-lo se essas ferramentas não conseguem diferenciar uma pessoa de um objeto ou explicar como e por qual razão chegou a uma decisão? Como prover validação e evitar desconforto no uso e interação? Um dos pilares primordiais do DCU é projetar com base num entendimento explícito sobre os usuários e suas tarefas e cenários (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2011). Como visto, em muitos desses sistemas isso é impossível. Assim, numa abordagem de Design, talvez seja fundamental que esse tipo de ferramenta decisória ainda não seja utilizada sem supervisão humana, para garantir a melhor experiência possível para todos os envolvidos direta ou indiretamente. Afinal, construir sistemas baseados num entendimento inapropriado ou incompleto das necessidades dos usuários é uma das maiores fontes de problemas e de sistemas falhos (ASSOCIAÇÃO BRASILEIRA DE NORMAS

TÉCNICAS, 2011).

Se muitos tentarem dizer e pensar que qualquer nova intervenção será automaticamente uma melhoria nas condições existentes (PRESTERO, 2010), cabe ao Homem não deixar o ser humano ser coadjuvante em sua própria existência. E o Design pode ser uma excelente forma de consegui-lo.

7. Referências bibliográficas

ALLEN-EBRAHIMIAN, B. Exposed: China's Operating Manuals for Mass Internment and Arrest by Algorithm. In: **icij.org**, 2019. Disponível em: <https://www.icij.org/investigations/china-cables/exposed-chinas-operating-manuals-for-mass-internment-and-arrest-by-algorithm/>. Acesso em: 13 jan. 2020.

ANDERSON, J. A British start-up will put AI into 700 schools in Belgium. In: **Quartz**, 2019. Disponível em: <https://qz.com/1577451/century-tech-signs-deal-to-put-ai-in-700-classrooms-in-belgium/>. Acesso em: 13 jun. 2019.

ASADA, M. Towards Artificial Empathy. **International Journal of Social Robotics**, v. 7, p. 19–33, 2014. Disponível em: <https://link.springer.com/content/pdf/10.1007/s12369-014-0253-z.pdf>. Acesso em: 10 ago. 2019.

ASADA, M. Development of artificial empathy. **Neuroscience Research**, v. 90, p. 41–50, 2015. Disponível em: <http://dx.doi.org/10.1016/j.neures.2014.12.002>. Acesso em: 10 ago. 2019.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR ISO 9241: ergonomia da interação humano-sistema**. Parte 210: projeto centrado no ser humano para sistemas interativos. Rio de Janeiro: ABNT, 2011.

BOECHAT, C.; MONT'ALVÃO, C. Refletindo Sobre o Futuro da Aprendizagem Através da Abordagem do Design. In: 13º Congresso Pesquisa e Desenvolvimento em Design, 2018, Joinville. **Anais** [...]. São Paulo: Blucher, 2019. p. 6151-6165.

BROWN, T. **Design Thinking**. Rio de Janeiro: Elsevier, 2010.

BUCHANAN, R. **Worlds in the Making: Design, Management, and the Reform of Organizational**

Culture. Xangai: **She Ji: The Journal of Design, Economics and Innovation**, 2015. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2405872615000039>. Acesso em: 01 abr. 2016.

CANO, R. J. C. O robô racista, sexista e xenófobo da Microsoft acaba silenciado. In: **El País**, 2016. Disponível em: https://brasil.elpais.com/brasil/2016/03/24/tecnologia/1458855274_096966.html. Acesso em: 12 set. 2019.

CASTELLANOS, S. Siri Contributor Tackles Software That Detects Emotional States. In: **The Wall Street Journal**, 2018. Disponível em: <https://blogs.wsj.com/cio/2018/03/08/siri-contributor-tackles-software-that-detects-emotional-states/>. Acesso em: 12 set. 2019.

CONNOR, N. Chinese school uses facial recognition to monitor student attention in class. In: **The Telegraph**, 2018. Disponível em: <https://www.telegraph.co.uk/news/2018/05/17/chinese-school-uses-facial-recognition-monitor-student-attention/>. Acesso em: 12 set. 2019.

CIPOLLA, C.; BARTHOLO, R. Empathy or inclusion: A dialogical approach to socially responsible design. **International Journal of Design**, v. 8, n. 2, p. 87–100, 2014. Disponível em: https://www.researchgate.net/publication/289156967_Empathy_or_Inclusion_A_Dialogical_Approach_to_Socially_Responsible_Design. Acesso em: 10 set. 2019.

DAVIS, M. Global Artificial Intelligence Boom Predicted in Education, Particularly in China. In: **EdWeek Market Brief Marketplace**, 2019. Disponível em: <https://marketbrief.edweek.org/marketplace-k-12/global-artificial-intelligence-boompredicted-education-particularly-china/>. Acesso em: 02 set. 2019.

DEDEZADE, E. Jobs of the future: teaching empathy to artificial intelligence. In: **Microsoft.com**, 2019. Disponível em: <https://news.microsoft.com/europe/features/more-than-a-feeling-teaching-empathy-to-artificial-intelligence/>. Acesso em: 12 ago. 2019.

HAO, K. What is AI? We drew you a flowchart to work it out. In: **Technology Review**, 2017. Disponível em:

<https://www.technologyreview.com/s/612404/is-this-ai-we-drew-you-a-flowchart-to-work-it-out/>. Acesso em: 12 jun. 2019.

IEEE. Are Tech Leaders and Governments Prepared for Rapidly Advancing Artificial Intelligence? In: **Innovation at work**, 2020. Disponível em: <https://innovationatwork.ieee.org/are-tech-leaders-and-governments-prepared-for-rapidly-advancing-artificial-intelligence/>. Acesso em: 29 fev. 2020.

KNIGHT, W. The Dark Secret at the Heart of AI. In: **Technology Review**, 2017. Disponível em: <https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/>. Acesso em: 12 jun. 2019.

LECHER, C. How Amazon automatically tracks and fires warehouse workers for ‘productivity’. In: **The Verge**, 2019. Disponível em: <https://www.theverge.com/2019/4/25/18516004/amazon-warehouse-fulfillment-centers-productivity-firing-terminations/>. Acesso em: 12 jun. 2019.

MORATH, E. AI Is the Next Workplace Disrupter—and It’s Coming for High-Skilled Jobs. In: **The Wall Street Journal**, 2020. Disponível em: <https://www.wsj.com/articles/ai-is-the-next-workplace-disrupterand-its-coming-for-high-skilled-jobs-11582470000>. Acesso em: 21 fev. 2020.

MORGAN, L. AI Challenge: Achieving Artificial Empathy. In: **Information Week**, 2018. Disponível em: <https://www.informationweek.com/big-data/ai-machine-learning/ai-challenge-achieving-artificial-empathy/a/d-id/1331628>. Acesso em: 11 set. 2019.

MUKHERJEE, S. A.I. Versus M.D. What happens when diagnosis is automated? In: **The New Yorker**, 2017. Disponível em: <https://www.newyorker.com/magazine/2017/04/03/ai-versus-md>. Acesso em: 14 set. 2019.

NARULA, G. Everyday Examples of Artificial Intelligence and Machine Learning. In: **Emerj.com**, 2019. Disponível em: <https://emerj.com/ai-sector-overviews/everyday-examples-of-ai/>. Acesso em: 15 set. 2019.

PARTNERSHIP ON AI. About us. In: **Partnership on AI**, 2019. Disponível em: <https://www.partnershiponai.org/about/>. Acesso em: 14 set. 2019.

PRESTERO, T. Better by Design: How Empathy Can Lead to More Successful Technologies and Services for the Poor (Discussion of Design Case Narratives : Rickshaw Bank, Solar-Powered Tuki, FGN Pump). **Innovations: Technology, Governance, Globalization**, v. 5, n. 1, p. 79–93, 2010. Disponível em:https://static1.squarespace.com/static/5267f3a6e4b0da2bab9939e3/t/564f2cade4b05d63dee98a87/1448029357796/INNOVATIONS-SOCAP13_009-023_Prestero.pdf. Acesso em: 10 set. 2019.

PTI. How to make sure Artificial Intelligence isn't biased as people? Satya Nadella's advise to techies. In: **Financial Express**, 2020. Disponível em:<https://www.financialexpress.com/industry/technology/how-to-make-sure-artificial-intelligence-isnt-biased-as-people-satya-nadellas-advise-to-techies/1878759/>. Acesso em: 24 fev. 2020.

RUSSEL, S.; NORVIG, P. **Artificial intelligence: A modern approach**. Nova Jersey: Prentice Hall, 2003.

SCHMELZER, E. This AI Researcher Thinks We Have It All Wrong. In: **Forbes.com**, 2020. Disponível em:<https://www.forbes.com/sites/cognitiveworld/2020/02/23/this-ai-researcher-thinks-we-have-it-all-wrong/#36a7b1b1571e>. Acesso em: 29 fev. 2020.

SHOHAM, Y. et al. **AI Index 2018 Annual Report**. Stanford: Stanford University Press, 2018. Disponível em: [http://cdn.aiindex.org/2018/AI Index 2018 Annual Report.pdf](http://cdn.aiindex.org/2018/AI%20Index%202018%20Annual%20Report.pdf). Acesso em: 02 set. 2019.

TWENGE, J.; MATEOS-GARCIA, J.; JOHN, J. Por que a falta de diversidade é um obstáculo à inteligência artificial. In: **El País**, 2019. Disponível em:https://brasil.elpais.com/brasil/2019/08/19/tecnologia/1566205434_091529.html. Acesso em: 10 set. 2019.

VINCENT, J. DeepMind's AI can detect over 50 eye diseases as accurately as a doctor. In: **The Verge**, 2018. Disponível em:<https://www.theverge.com/2018/8/13/17670156/deepmind-ai-eye-disease-doctor-moorfields>. Acesso em: 02 set. 2019.

VINCENT, J. The state of AI in 2019. In: **The Verge**, 2019. Disponível em:<https://www.theverge.com/2019/1/28/18197520/ai->

artificial-intelligence-machine-learning-computational-science. Acesso em: 02 set. 2019.

WRIGHT, P.; MCCARTHY, J. Empathy and experience in HCI. In: Conference on Human Factors in Computing Systems, 2008, Florença. **Anais [...]**. Florença: ACM Press, 2008. p. 637–646.

ZARLEY, D. Meet The Scientists Who Are Training AI To Diagnose Mental Illness. In: **The Verge**, 2019. Disponível em:<https://www.theverge.com/2019/1/28/18197253/ai-mental-illness-artificial-intelligence-science-neuroimaging-mri>. Acesso em: 02 set. 2019.