



## DIREITOS DOS ROBÔS? EM VEZ DISSO, FALEMOS SOBRE BEM-ESTAR HUMANO\*

ROBOT RIGHTS?: LET'S TALK ABOUT HUMAN WELFARE INSTEAD

Abeba Birhane  
Trinity College Dublin, IRL  
birhanea@tcd.ie  
 [orcid.org/0000-0001-6319-7937](https://orcid.org/0000-0001-6319-7937)

Jelle van Dijk  
University of Twente, NLD  
[jelle.vandijk@utwente.nl](mailto:jelle.vandijk@utwente.nl)  
 [orcid.org/0000-0001-8224-3302](https://orcid.org/0000-0001-8224-3302)



\* O artigo foi originalmente publicado sob o título **Robot Rights? Let's Talk about Human Welfare Instead** em **AIES '20: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society**, Nova Iorque, fev. 2020, p. 207-213, e traduzido como parte das atividades do grupo de pesquisa Neoliberalismo, Tecnologia e Direito – NTD por Celso Naoto Kashiura Jr., Letícia Guaitoli dos Reis e Maria Clara Marchezin Saidneuy.

### **Direitos dos robôs?: Em vez disso, falemos sobre bem-estar humano**

**Resumo:** O debate sobre “direitos dos robôs” e as questões correlatas sobre responsabilidade dos robôs evocam algumas das mais polarizadas posições na ética da IA. Enquanto alguns advogam que robôs deveriam ter direitos assim como os seres humanos, outros, em frontal oposição, argumentam que robôs não merecem direitos, pois são objetos que devem ser nossos escravos. Baseados numa filosofia pós-cartesiana, sustentamos não apenas negar “direitos” aos robôs, mas, antes de tudo, negar que robôs, como artefatos que emergem de e medeiam seres humanos, sequer poderiam receber direitos. Uma vez que entendemos robôs como mediadores de seres humanos, poderemos notar que o debate sobre “direitos dos robôs” está focado em problemas do Primeiro Mundo, em detrimento de preocupações éticas urgentes, como viés de máquina, exploração do trabalho humano induzida pelas máquinas e erosão da privacidade, todos impactando os indivíduos menos privilegiados da sociedade. Concluímos que, se o ser humano é nosso ponto de partida e o bem-estar humano é nossa preocupação primária, os impactos negativos decorrentes dos sistemas maquinicos, bem como a falta de responsabilidade por parte das pessoas que projetam, vendem e implementam tais máquinas, continuam sendo a discussão ética mais urgente sobre a IA.

**Palavras-chave:** Direitos dos robôs. Ética da IA. Personificação. Bem-estar humano.

### **Robot Rights?: Let's Talk about Human Welfare Instead**

**Abstract:** The ‘robot rights’ debate, and its related question of “robot responsibility”, invokes some of the most polarized positions in AI ethics. While some advocate for granting robots rights on a par with human beings, others, in a stark opposition argue that robots are not deserving of rights but are objects that should be our slaves. Grounded in post-Cartesian philosophical foundations, we argue not just to deny robots “rights”, but to deny that robots, as artifacts emerging out of and mediating human being, are the kinds of things that could be granted rights in the first place. Once we see robots as mediators of human being, we can understand how the “robots rights” debate is focused on first world problems, at the expense of urgent ethical concerns, such as machine bias, machine elicited human labour exploitation, and erosion of privacy all impacting society’s least privileged individuals. We conclude that, if human being is our starting point and human welfare is the primary concern, the negative impacts emerging from machinic systems, as well as the lack of taking responsibility by people designing, selling and deploying such machines, remains the most pressing ethical discussion in AI.

**Keywords:** Robot rights. AI ethics. Embodiment. Human welfare.

## **I. O debate: direitos dos robôs**

Eticistas têm discutido a noção de “direitos dos robôs”: a ideia segundo a qual deveríamos (no futuro) atribuir às máquinas artificialmente inteligentes, em função de sua constituição como agentes inteligentes e autônomos, direitos comparáveis aos direitos humanos. Alguns defendem os direitos dos robôs numa visão de mundo materialista e tecno-otimista, argumentando que devemos evitar qualquer “chauvinismo biológico” *a priori*. A fundamentação prossegue: se as máquinas puderem desenvolver o mesmo tipo de capacidade que atribuímos a nós mesmos, não temos razão para não lhes atribuir os direitos que atribuímos a nós mesmos (Asaro, 2006; Brooks, 2000; Dennett, 1987).

Uma linha mais crítica e emancipatória da ética dos robôs afirma que atribuir direitos aos robôs não apenas se justifica eticamente, mas também, o que é mais relevante, ajuda a refletir sobre correntes subterrâneas nos debates éticos (do ocidente). Discutir os direitos dos robôs ajuda a

desvincular a ética de seus fundamentos ocidentais paternalistas e opressivos e contribui para a emancipação de grupos oprimidos como mulheres e negros (Gunkel, 2018).

Em contraposição, outros afirmam que devemos encarar robôs como nossos escravos. Bryson (2010), uma das defensoras dessa posição, está ciente das conotações implicadas no termo escravo. Ela explica que historicamente escravidão significa desumanização, algo a que, por boas razões, a maioria das culturas passou a se opor:

Uma vez que seres obviamente humanos foram rotulados como desumanos no passado recente da cultura global, muitos parecem temer aplicar esse rótulo. Por exemplo, Dennett argumenta que deveríamos atribuir direitos a qualquer coisa que pareça agir de maneira intencional [...] A generosidade [de Dennett] é, quase que por definição, boa. (Bryson, 2010, p. 2)

Mas Bryson discorda de Bennett (1987). Atribuir direitos aos robôs, ela argumenta, não é necessariamente bom. O bem-estar humano deveria ser nossa preocupação primordial e as preocupações com robôs não deveriam nos distrair do verdadeiro objetivo. Concordamos plenamente com ela a esse respeito.

Discordamos, contudo, acerca de tratar robôs como “escravos”. Em defesa de sua posição, Bryson (2010) afirma que “certamente a desumanização só é errada quando aplicada a alguém que é realmente humano”. Nossa posição é a de que “desumanização” de robôs não é errada, é impossível. Não se pode desumanizar aquilo que nunca foi humano. Se usamos o termo escravo, assumimos implicitamente que aquele assim rotulado pode ser desumanizado. O robô já foi implicitamente “humanizado” para, na sequência, escravizá-lo. Obviamente não se deve escravizar alguém que foi inicialmente considerado humano.

Bryson aceita em parte o enquadramento da narrativa da ética dos robôs, na qual uma discussão para considerar o estatuto ontológico dos robôs em relação aos direitos é legítima em princípio. Nossa posição é a de que toda essa discussão é equivocada. Na melhor das hipóteses, os debates sobre a ética dos robôs são meditações filosóficas do Primeiro Mundo demasiado apartadas das questões humanas do mundo real. Na pior, talvez contenha má-fé – a caracterização depreciativa de pessoas realmente oprimidas e da sua luta por direitos por uma academia branca e masculina apelando à “razão”.

## II. Uma síntese de nossos argumentos

Alguns podem argumentar que a ideia de direitos de robôs é uma peculiar e irrelevante discussão existente apenas às margens das pesquisas mais amplas acerca da ética da IA e, portanto, dedicar nosso tempo a isso não faria justiça a trabalhos importantes realizados neste campo. Mas a ideia de direitos de robôs, a princípio, está perfeitamente legitimada se permanecermos fiéis aos princípios materializados na inteligência artificial: a princípio, deve ser possível construir uma máquina artificialmente inteligente e, se obtivéssemos sucesso, não haveria razão para não conceder a tal máquina os direitos que possuímos. Nossa crítica não visa invalidar esse raciocínio, mas questionar seus pressupostos fundamentais. Os direitos dos robôs sinalizam algo muito mais sério

sobre a tecnologia da IA, a saber: baseados em seu tecno-otimismo materialista, cientistas e tecnólogos estão tão preocupados com a possibilidade de um futuro com uma máquina imaginária que acabam esquecendo o real e negativo impacto que as criaturas intermediárias – os atuais sistemas de IA que possuímos hoje – causam aos seres humanos. Em outras palavras: a discussão sobre direitos dos robôs não deve ser separada da ética da IA e a ética da IA deve buscar examinar e refletir profundamente sobre pressupostos subjacentes de cientistas e engenheiros, ao invés de observar seu projeto “apenas” do ponto de vista prático de discussão acerca de restrições e regras éticas que devem reger as tecnologias de IA na sociedade.

Nosso ponto de partida não é negar os “direitos” dos robôs, mas negar que robôs são o tipo de seres aos quais se poderia atribuir ou negar direitos. Sugerimos que não faz sentido entender os robôs como escravos, uma vez que “escravo” diz respeito a uma categoria de seres a que os robôs não pertencem. Seres humanos são desse tipo. Acreditamos que animais são desse tipo (muito embora a discussão acerca dos animais esteja fora do escopo do presente artigo). Tomamos um ponto de vista pós-Cartesiano, fenomenológico, segundo o qual ser humano significa uma experiência vivida e corpórea, que por sua vez está inserida em práticas sociais. Artefatos tecnológicos, muito embora sejam parte crucial dessa experiência, não são seres desse tipo. A relação entre ser humano e tecnologia é entrelaçada, mas não simétrica.

Baseados nesta perspectiva, analisemos a agenda da ética da IA. Para alguns eticistas, argumentar em prol de direitos dos robôs decorre de uma aversão à arrogância humana face ao mundo. Nós também desejamos lutar contra a arrogância humana. Mas nós vemos arrogância, antes e acima de tudo, nas fantasias tecno-otimistas da indústria da tecnologia, que faz grandes promessas de nos recriar em silício, superar-nos com uma “super IA” e “compartilhar digitalmente” nossas mentes como um arquivo imortal, enquanto, ao mesmo tempo, explora o trabalho humano. Em nosso entendimento, a maioria dos debates no campo dos direitos dos robôs está fundada na mesma tecno-arrogância. O que extraímos de Bryson é seu apelo para focar no real problema: a opressão humana. Destacamos as contínuas violações do bem-estar humano, especialmente daqueles desproporcionalmente impactados pelo desenvolvimento e pela integração ubíqua da IA na sociedade. Nossa postura ética acerca dos seres humanos é a de que ser humano significa interagir com nosso entorno de forma respeitosa e correta. A tecnologia deve ser desenvolvida para promover isso. Essa, por sua vez, deve ser a principal preocupação dos eticistas.

A seguir, apresentamos primeiramente nossa perspectiva pós-cartesiana sobre o ser humano e o papel da tecnologia dentro desta perspectiva. Posteriormente, explicaremos por que, mesmo que os robôs não tenham direitos garantidos, também rejeitamos a ideia de robôs como escravos. Ao final, chamamos atenção, em vez disso, para o bem-estar humano. Nós discutiremos como a IA, ao invés de ser potencialmente oprimida, é utilizada como ferramenta pelos seres humanos (com poder) para oprimir outros seres humanos e como a discussão sobre direitos dos robôs desvia a atenção das questões e problemas éticos que realmente importam. Finalizamos a reflexão sobre as responsabilidades, não dos robôs, mas sim de seus criadores humanos.

## II.1 Uma reinterpretação pós-cartesiana

Robôs, como a maioria das tecnologias criadas pelos seres humanos, são criados “a nossa imagem e semelhança”. Mas qual é essa imagem própria que usamos como modelo? Desde os primórdios a IA tentou projetar uma interpretação cognitivista do pensamento humano na máquina, baseada em uma distinção (neo)cartesiana entre, por um lado, o sistema mental, entendido por equivalência como o *software* da máquina, e, por outro, o corpo físico, equivalente às partes físicas dos robôs. Contudo, em contraste com o dualismo de Descartes, os cognitivistas sustentam que o sistema mental é também fisicamente realizado, mapeando o conteúdo mental em processos físicos (por exemplo, padrões de ativação cerebral). No geral, esse continua sendo um modelo conceitual predominante que se apresenta nas tentativas de construir máquinas inteligentes. Consequentemente, para especialistas em tecnologia e engenheiros, um “humano” pode, em princípio, nesse modelo, ser “construído”, porque o que é necessário para ser humano é, em última análise, uma configuração particular e complexa de processos físicos (Churchland, 2013). Partindo desse modelo, a ideia de direitos dos robôs faz perfeito sentido.

Para entender como reconceituamos o ser de um robô, precisamos observar nossa concepção de ser humano, a qual rejeita a imagem anteriormente exposta. Em nossa posição pós-cartesiana, inspirada na fenomenologia, o ser humano é uma experiência vivida e corpórea, ou, como Merleau-Ponty, seguindo Husserl denominou, um “ser-no-mundo”. A ciência cognitiva corporificada e enativa, que segue esse raciocínio, explica como os sistemas biológicos vivos – corpos vivos – “realizam” seu mundo perceptual por meio de interações contínuas com o ambiente (Di Paolo; Cuffari; De Jaegher, 2018). Essas interações se auto-organizam em acoplamentos sensório-motores que podemos chamar de hábitos ou habilidades. Com base nesses acoplamentos, percebemos (ou melhor, “realizamos”) coisas no mundo, em primeira instância, como oportunidades para a ação (Golonka; Wilson, 2012). As coisas-como-oportunidades que percebemos têm relação direta com nossas habilidades corporais (Dreyfus; Dreyfus, 2004). Para exemplificar: um banco de praça “é” uma coisa diferente para um skatista ou para uma pessoa em situação de rua ou para um visitante qualquer. As habilidades incorporadas se auto-organizam e trabalham para sustentar ainda mais o organismo. Um segundo aspecto diz respeito à natureza social inerente ao ser humano. Estamos sempre já inseridos em práticas sociais e a maneira como interagimos com o mundo e damos sentidos a ele precisa ser compreendida neste contexto. Essa visão foi desenvolvida pelos fenomenologistas (Schutz; Luckmann, 1973) e, de forma semelhante, por meio de pesquisas conjuntas, práticas situadas (Lave, 2018) e construção de sentido participativo (Di Paolo; Cuffari; De Jaegher, 2018).

Partindo do pressuposto que seres humanos são interações vividas e corpóreas, podemos reinterpretar o papel da tecnologia. Em primeiro lugar, os artefatos criados pelos humanos adquirem significado como mediadores de nossa atuação no mundo, sustentando, rompendo, modificando ou enriquecendo os acoplamentos sensório-motores. Isso pode ser encontrado na discussão de Heidegger (1927) acerca do martelo como “pronto-para-o-uso” e na análise de Merleau-Ponty (1962) sobre a bengala da pessoa cega como uma extensão do próprio corpo. No desenvolvimento mais recente da

ciência cognitiva corporificada, isso se relaciona com a ideia de cognição distribuída e mente estendida (Clark, 1988). Em segundo lugar, o significado dos artefatos deve ser entendido dentro do contexto de nossa situação social. Em outras palavras, as coisas são como são por conta de como configuram nossas práticas sociais (Suchman, 2007) e de como a tecnologia estende nosso corpo biológico. Nossa concepção de ser humano, então, é de que somos e sempre fomos totalmente incorporados e entrelaçados ao ambiente que desenvolvemos ao nosso redor e que somos criticamente dependentes dessa incorporação para nos sustentar (Birhane, 2017).

A ilusão cartesiana de nos posicionar longe do mundo natural, artificial e social em que vivemos oportunizou a criação de um projeto de construção de “inteligência” artificial em que a inteligência é moldada com base numa inteligência humana que se desprende do mundo e que o observa e os artefatos que criamos são coisas neste mundo externo “objetivo”. Em contraste, a abordagem “socio-relacional” de Coeckelbergh (2010) sobre a ética das máquinas parece, à primeira vista, semelhante à nossa perspectiva. No entanto, as conclusões por ele extraídas são opostas. Para Coeckelbergh, o “socio-relacional” descreve a forma como as pessoas percebem os artefatos por uma perspectiva variável e percebê-los como “meras máquinas” seria, portanto, tão válido como percebê-los como um “outro ser inteligente”. Parece-nos que tanto Coeckelbergh quanto os teóricos mais tradicionais falham quanto a notar quão fundo já estamos imersos e entrelaçados com as tecnologias. Uma profunda apreciação desse entrelaçamento não implica que os artefatos devam ser vistos como “agentes como nós” (mesmo que socialmente falemos essa forma): o que precisamos fazer é retomar a percepção que as tecnologias *já sempre fazem parte de nós*, como elementos intrínsecos ao nosso ser no mundo (Van Dijk, 2018).<sup>1</sup>

## II.2 Escravos são humanos abusados como máquinas

Gunkel (2015; 2018) constrói uma retórica na qual contrasta os “aparentemente frios e impessoais robôs industriais” com os atuais robôs sociais que “compartilham espaços físicos e emocionais com os usuários”. “Por essa razão”, ele sugere que é “racional questionarmos acerca do status social e da posição moral dessas tecnologias” (Gunkel, 2015). Mas nós não vemos razão para isso. Robôs sociais são, enquanto máquinas, tão frios e impessoais como qualquer outra máquina. Olhando de outra perspectiva, eles são apenas tão quentes e pessoais como qualquer outra máquina, do mesmo modo que nos apaixonamos por um carro, uma máquina de café ou uma casa. Nada disso implica atribuir direitos às máquinas, no máximo significa que tomamos conta dos nossos artefatos como produtos de nosso trabalho árduo, expressão de nossa criatividade humana, recebidos como

<sup>1</sup> Pode-se questionar se uma perspectiva que se baseia na mediação tecnológica da experiência vivida deveria levar à conclusão de que “a materialidade tem capacidade” (v. Verbeek, 2000). Se a mediação por máquinas significa que essas máquinas possuem capacidade, então talvez elas devam ter direitos. Rejeitamos a posição radicalmente “simétrica” de Latour, na qual objetos e humanos são conectados em rede como iguais. Nossa posição é mais tradicionalmente heideggeriana, ao vermos a tecnologia como algo que se fundamenta no ser humano e contribui para sustentá-lo – um ser incorporado, situado e estendido. Contudo, com Verbeek, rejeitamos o pessimismo de Heidegger em relação às tecnologias modernas: acreditamos que as tecnologias podem ser mobilizadas para o bem, ainda que frequentemente sejam usadas para o mal.

presentes e assim por diante. Em outras palavras: coisas configuram práticas sociais e tomar conta das coisas significa tomar conta de nós mesmos. Ao tomarmos conta das coisas, nós reconhecemos seus criadores, valorizamos seus desenvolvedores humanos e temos respeito pela pessoa que nos presenteou.

Gunkel (2015) nunca caiu na armadilha de inventar futuros fantasiosos de máquinas conscientes para discutir direitos dos robôs. Sua preocupação está com o estado mental que permeia a oposição aos direitos dos robôs, o que, em sua visão, revela um raciocínio excludente e “antropocêntrico”, o qual não apenas marginaliza máquinas, mas também é utilizado como instrumento para excluir outros seres humanos. Citando Stone (1972), Gunkel (2015, p. 2) argumenta que “humanos definiram diversos grupos como menos que humanos: escravos, mulheres, as ‘outras raças’, crianças e estrangeiros [...] os quais vêm sendo definidos como [...] sem direitos”.

Mas a própria razão pela qual julgamos que a forma como escravos e mulheres foram (e ainda são) tratados como menos do que humanos é a de que são usados como um meio para um fim, como “instrumentos” que homens brancos podem usar para realizar tarefas. O robô é o modelo exato pelo qual avaliamos se os humanos são desumanizados. Na terminologia de Hannah Arendt (1958), desumanizar pessoas significar reduzir a sua *raison d'être* ao mero labor, um modo de atividade que ela diferencia de “trabalho” (um projeto) e “ação” (ação política). Ao colocar escravos reais, mulheres e “outras raças” na mesma lista que os robôs, não se está humanizando todos, na verdade se está desumanizando os humanos reais dessa lista. Consideremos Coeckelbergh (2010) quando ele escreve: “Nós emancipamos escravos, mulheres e alguns animais. Antes, escravos e mulheres não eram tratados como ‘homens’. De todo modo, fizemos um progresso moral e agora os consideramos como humanos.” Essa linha de Coeckelbergh nos leva a especular sobre a futura emancipação dos robôs. Mas a escolha de palavras sugere que, mesmo involuntariamente, há uma perspectiva ocidental e masculina na matéria (“nós emancipamos mulheres”). A linha de formulação corre o risco interpretativo de se transformar em: “As mulheres e escravos libertos não devem reclamar, se nós, homens esclarecidos, decidirmos liberar mais alguns!”.

Se em nossa própria interpretação for acusada, por contraste, de ser “antropocêntrica”, então sim, este é exatamente o ponto: robôs não são humanos e nosso entendimento gira em torno do bem-estar dos seres humanos (v. Rushkoff, 2019).

### *II.3 Robôs não são escravos*

Como pontuamos anteriormente, discordamos do tratamento de robôs como escravos. Para argumentar contra os direitos dos robôs, recorremos ao infame experimento de obediência à autoridade de Milgram.

Precisamos estar cientes da diferença entre a forma que uma pessoa age e reflete acerca de suas próprias ações em um mundo que ela percebe como real, mesmo que este mundo seja de fato baseado em uma ilusão, e o efeito das ações dessa pessoa sob a perspectiva de um observador externo. Nesse segundo enquadramento, “objetivo”, os participantes do experimento de Milgram

não causam dano algum, eis que a pessoa que aparece gritando em dor é, “na verdade”, um ator. No enquadramento pessoal, por sua vez, do mundo que os participantes percebem como real, realmente causam danos a outra pessoa – alguns inclusive experienciam ter cometido homicídios. Ser informados posteriormente de que se tratava de uma ilusão não ajudou alguns deles a se livrar de suas próprias conclusões, deixando alguns traumatizados:

Um vereador de New Haven apresentou uma queixa às autoridades de Yale sobre o estudo: “Não me lembro de jamais ter ficado tão perturbado” (p. 132). Um dos participantes (#716) chegou a verificar os obituários de New Haven, com medo de ter matado o aprendiz. Outro participante (#501) estava tremendo tanto que não tinha certeza se conseguiria dirigir até em casa; segundo sua esposa, no caminho de volta ele tremia dentro do carro e falou sem parar, até a meia-noite, sobre o intenso desconforto que sentia (p. 95). O participante #711 relatou: “O experimento me deixou tão abalado que passei a noite em suor frio e tendo pesadelos, com medo de ter matado aquele homem na cadeira.” (Brannigan, 2013, p. 93)

Se olharmos para a forma como tratamos os robôs do ponto de vista de um participante do experimento de Milgram, seria, de fato, problemático tratar os robôs como escravos. O movimento cultural e linguístico no uso da palavra escravo significaria – por analogia – que, em nosso mundo enativo, teríamos nos tornado proprietários de escravos, no mesmo senso de “verdade” segundo o qual participantes do experimento de Milgram se tornam assassinos.

Ao mesmo tempo, o enquadramento do experimento de Milgram também mostra por que devemos rejeitar a ideia de que as máquinas robôs sejam tratadas de forma injusta. Seguindo lógica do experimento, o robô é um ator. Não há um “receptor” real (objetivo, em terceira pessoa) do ato antiético. A única possível vítima é a pessoa que se tornou proprietária do escravo ou, talvez, a sociedade em geral: se tratar robôs como escravos se tornar algo comum, podemos estar nos engajando em práticas sociais que não contribuem para nos tornarmos seres humanos melhores. A sociedade deve ter razões para rejeitar essas práticas, mesmo se ninguém as esteja “praticando de fato” (Whitby, 2008).

Mas independentemente de considerarmos que as pessoas podem ser “atraídas” a atos antiéticos por meio de simulações, o fato é que nenhuma injustiça é cometida contra o ator que implementou a simulação, seja ele um ator humano no experimento de Milgram ou uma máquina simulando um “robô senciente”. Talvez o “robô como escravo” possa ter um papel como ferramenta educacional ou como arte crítica, mas não existe algo como “direitos dos robôs” fora do campo da ficção.

### III. Em vez disso, falemos sobre bem-estar humano

Não existem robôs que se aproximem do tipo de “ser” que os humanos são, nem do tipo de “ser-com” que os humanos podem ter com outros humanos. Juntamente com Hubert Dreyfus (2004), duvidamos que algum dia venha a existir. Argumentar a favor dos direitos dos robôs com base em visões futuras de máquinas sencientes é, na melhor das hipóteses, filosofia especulativa de sofá. Enquanto isso, a cultura popular fala da IA e dos robôs atuais como se a máquina inteligente já existisse, quando, na verdade, não existe. Tais percepções revelam a antiga subcorrente cartesiana

cognitivista nos debates sobre IA, que vê as máquinas que criamos como “outros agentes, muito parecidos conosco”, em vez do que elas são: mediadoras de práticas humanas que se manifestam no corpo e no meio social.

Pode-se sustentar que é romântico ou a-histórico pensar que nenhum progresso tecnológico poderia produzir uma IA “verdadeira” no futuro. Mas o romantismo e a falta de consciência histórica podem ser encontrados em ambos os lados do debate. Raymond Kurzweil (2005), por exemplo, prevê que o “uploading da mente” se tornará possível na década de 2030 e projeta a data para a “singularidade” em 2045. Previsões românticas como essa, que invariavelmente vislumbram avanços revolucionários para daqui a algumas décadas, têm sido recorrentes desde os primórdios da tecnologia digital e todas falharam. Parece que a “IA Geral”, a “singularidade” e a “superinteligência” são para os tecno-otimistas o que o dia do juízo final representa para certos cultos religiosos.

Mas isso não importa. Independentemente das previsões futuras, o que é importante e urgente nesse momento é denunciar um fato: as visões românticas e rebuscadas de robôs trabalhadores, cuidadores e amigos, bem como o debate sobre a “questão” dos seus supostos direitos, contribuem para o dano real sofrido por indivíduos e grupos que se encontram atualmente em desvantagem socioeconômica (conforme detalharemos na próxima seção). Quer a nossa descrença na existência futura de uma IA verdadeira se prove errada em algum momento, quer não, ela é, em qualquer caso, menos prejudicial do que o otimismo recorrente sobre futuros puramente ficcionais. Porque, em vez de progredir firmemente em direção a uma comunidade feliz de humanos e “IAs sencientes”, o tecno-otimismo contribui para o desenvolvimento atual de uma infraestrutura tecnológica desumanizante (Rushkoff, 2019). Debater as condições necessárias para os direitos dos robôs continua a colocar o foco em máquinas (inexistentes) em vez de em pessoas reais. Na próxima seção, focamos naquilo que de fato existe: máquinas com software que chamamos de “IA” e que, na realidade atual, prejudicam pessoas.

### *III.1. Robôs são usados para violar direitos humanos*

As discussões sobre ética dos robôs, ao retratarem os robôs como sistemas inteligentes como principal preocupação, minimizam o fato de que estamos atualmente em meio a sistemas de inteligência artificial que se infiltram rapidamente em todos os aspectos da vida. As questões reais e urgentes que estão a emergir com a implementação em massa de sistemas de IA aparentemente invisíveis precisam ser discutidas agora, porque atualmente impactam grandes grupos de pessoas.

A implementação em massa de máquinas e IA hoje deveria nos impelir a examinar os impulsos comerciais por trás dessas máquinas, bem como os danos e a injustiça que a integração de máquinas na sociedade acarreta. Desde a perpetuação de preconceitos e injustiças históricas e sociais (Benjamin, 2019; Eubanks, 2018; Obermeyer; Mullainathan, 2019), à invasão de privacidade (Zuboff, 2019), à exploração do trabalho humano (Tubaro; Casilli, 2019), muitas vezes para ganhos financeiros de empresas privadas, os sistemas de IA opõem-se ao bem-estar humano. Quando os sistemas de IA

são implementados e integrados nas nossas vidas cotidianas sem um exame crítico e antecipação dos efeitos colaterais emergentes, eles representam ameaças ao bem-estar humano.

Com a ascensão do aprendizado de máquina, há um apetite crescente para entregar grande parte dos nossos problemas sociais, políticos e econômicos às máquinas, o que traz consigo a ganância corporativa à custa do bem-estar e da integridade humanos (Zuboff, 2019). Para o mundo corporativo, que produz uma grande proporção da IA atual, o lucro marca o seu objetivo central, enquanto para aqueles que implementam tais tecnologias em vários setores sociais, a IA aparentemente fornece uma solução rápida e econômica para problemas sociais complexos e multifacetados. No entanto, a integração desses sistemas está se revelando uma ameaça ao bem-estar, integridade e privacidade das pessoas, especialmente daquelas que são socioeconomicamente desfavorecidas (Angwin; Larson; Mattu; Kirchner, 2016; Obermeyer; Mullainathan, 2019; O’Neil, 2016). A seguir, discutiremos várias dessas ameaças.

### *III.2. Viés de máquina e discriminação*

Tornou-se trivial apontar como os processos de tomada de decisão em várias esferas sociais, políticas e econômicas são assistidos por sistemas automatizados. As soluções de IA permeiam a maioria das esferas da vida, desde a triagem de potenciais funcionários até a sua entrevista, passando pela previsão de onde a atividade criminosa pode ocorrer (em alguns casos, quem pode cometer um crime) e pelo diagnóstico de doenças. Essas são questões altamente contestadas e inherentemente políticas e morais que a indústria da tecnologia, no entanto, trata como “problemas técnicos” que podem ser quantificados e automatizados.

A automação de questões sociais, políticas e culturais complexas exige que esses conceitos complexos, multivalentes, contextuais e em contínuo movimento sejam quantificados, medidos, classificados e capturados através de dados (McQuillan, 2018). Extrapolações, inferências e modelos preditivos são então construídos, muitas vezes com aplicações práticas na vida real, com graves consequências para os mais vulneráveis da sociedade. Os sistemas de aprendizado de máquina que inferem e preveem o comportamento e a ação individual, com base em extrações superficiais, são então implementados no mundo social, resultando na emergência de vários problemas. Esses sistemas captam estereótipos sociais e históricos mais do que quaisquer explicações causais profundas e fundamentais. Nesse processo, os indivíduos e grupos, muitas vezes aqueles à margem da sociedade, que não se enquadram nos padrões estereotipados, sofrem as consequências indesejáveis (Keyes, 2018). Uma questão recorrente no âmbito do viés algorítmico, por exemplo, revela que indivíduos e grupos que foram historicamente marginalizados são afetados de forma desproporcional. Inclui-se nisso, a título de exemplo, o viés na detecção de tons de pele em pedestres (Wilson; Hoffman; Morgenstern, 2019); o viés em sistemas de policiamento preditivo (Richardson; Schultz; Crawford, 2019); o viés de gênero e as discriminações na apresentação de anúncios de carreiras STEM (Lambrecht; Tucker, 2019); o viés racial em algoritmos de reincidência (Angwin; Larson; Mattu;

Kirchner, 2016); o viés nas políticas dos motores de busca (Introna; Nissenbaum, 2000) e o viés e a discriminação na medicina (Ferryman; Pitcan, 2018; Obermeyer; Mullainathan, 2019).

A IA, longe de ser um fenômeno futuro à espera de acontecer, está aqui operando de forma ubíqua e com um impacto desastroso em grupos social e historicamente marginalizados. Como observa Weiser: “As tecnologias mais profundas são aquelas que desaparecem. Elas entrelaçam-se no tecido da vida cotidiana até se tornarem indistinguíveis dela.” IA ubíqua está inextricavelmente interligada ao que significa ser um ser humano (Cheney-Lippold, 2018). No entanto, a questão é: como enquadrar conceitualmente da melhor forma esse entrelaçamento?

A narrativa típica parece conceber as tecnologias de IA como um tipo de parceiro social com o qual nos comunicaremos e viveremos de maneiras comparáveis às formas como outros seres humanos estão ligados às nossas vidas. Na realidade, nenhum robô hoje está perto dessa visão futura. A situação real que temos hoje mostra algoritmos de aprendizado de máquina incorporados em ferramentas aparentemente mundanas, apoiando tarefas cotidianas. Esses algoritmos estão influenciando nosso “ser-no-mundo” básico – a forma como percebemos e categorizamos o mundo, a capacidade que nós mesmos temos para agir sobre ele – num sentido weiseriano mais invisível, o que o torna ainda mais insidioso. Por exemplo, o robô humanoide conhecido como Sophia personifica uma imagem que se coaduna com a concepção amplamente difundida de “robôs inteligentes”, mas que, na verdade, possui um motor e capacidades rudimentares. Em comparação, o Roomba da iRobot, embora seja retratado como uma máquina doméstica inofensiva, exerce um impacto muito maior nas nossas vidas e o seu lado sombrio é que serve como uma ferramenta de vigilância que recolhe continuamente dados sobre os nossos lares. É fácil não perceber os perigos que o Roomba representa para a nossa privacidade, pois a máquina passa despercebida e torna-se silenciosamente parte do nosso dia a dia. O aspirador “autônomo” Roomba da iRobot está equipado com uma câmera, sensores e software que lhe permitem construir mapas do santuário privado de nosso lar, enquanto rastreia a sua própria localização (Zuboff, 2019). Em combinação com outros dispositivos de IoT, o Roomba pode ser usado para supostamente mapear os nossos hábitos, comportamentos e atividades.

A maioria das empresas de IA gaba-se das suas capacidades de fornecer insights sobre a psique humana. São os interesses financeiros de empresas e engenheiros – que recolhem, avaliam e interpretam algorítmicamente dados para prever comportamentos – que impulsionam a investigação e o desenvolvimento em IA. Como tal, os sistemas “inteligentes” infiltram-se na vida cotidiana, desde os dispositivos de IoT até o “lar inteligente”, todos concebidos para transformar todos os cantos da experiência vivida em dados comportamentais (Zuboff, 2019). Visionar um futuro sistema inteligente semelhante ao humano, enquanto se deixam de lado sistemas tão ubíquos e invasivos que representam uma ameaça à privacidade e ao bem-estar humano, demonstra, para dizer o mínimo, uma preocupação descabida. A integração de sistemas maquinícios nos assuntos sociais e humanos representa um perigo imediato, especialmente para as pessoas desprovidas de direitos que mais precisam de proteção (O’Neil, 2016). A nosso ver, levar as preocupações éticas a sério implica priorizar

o bem-estar das pessoas, sobretudo daquelas frequentemente impactadas de forma desproporcional pela integração de sistemas maquinícios no cotidiano.

### *III.3. Olhando sob o capuz da IA: trabalho humano*

Se olharmos para os direitos dos robôs a partir das tecnologias reais e existentes e das práticas humanas que elas medeiam, percebemos que é inherentemente difícil traçar uma fronteira em torno da entidade (artificial) à qual seria necessário conceder direitos. De fato, as tentativas de analisar o que constitui os sistemas inteligentes e aparentemente autônomos atuais revelam que, longe de serem totalmente autônomos, esses sistemas funcionam com base em trabalho humano exploratório. De robôs a veículos “autônomos” e a sofisticados sistemas de reconhecimento de imagem, todas as máquinas dependem fortemente da intervenção humana. Os sistemas que são percebidos como “autônomos” nunca são totalmente autônomos, mas sistemas humano-máquina.

Além disso, como observa Bainbridge (1983), “quanto mais avançado é o sistema, mais crucial é a contribuição do ser humano”. O mesmo se aplica aos sistemas inteligentes atuais (Baxter; Rooksby; Wang; Khajeh-Hosseini, 2012; Strauch, 2017). “Quanto mais dependemos da tecnologia e a levamos aos seus limites, mais precisamos de pessoas altamente qualificadas, bem treinadas e experientes para tornar os sistemas resilientes, atuando como a última linha de defesa contra as falhas que inevitavelmente ocorrerão” (Baxter; Rooksby; Wang; Khajeh-Hosseini, 2012). Os sistemas de IA não dependem apenas de engenheiros e cientistas altamente qualificados e bem remunerados, mas também contam fortemente com a contribuição do trabalho humano menos visível e de baixa remuneração, conhecido como “microtrabalho” ou “trabalho-de-multidão”. Desde anotar e adicionar rótulos a imagens, a identificar objetos numa fotografia ou até organizar itens numa lista, esses “trabalho-de-multidão” de baixa remuneração preparam os dados de “treino” para as máquinas (Tubaro; Casilli, 2019). Além do trabalho mal remunerado, o trabalho humano não pago também alimenta o desenvolvimento de sistemas inteligentes proprietários, que as empresas privadas controlam e de que se beneficiam. O reCAPTCHA do Google é um desses exemplos: surgiu primeiro como uma técnica para prevenir spam, depois foi usado para digitalizar livros antigos e, mais tarde, como um meio de obter dados de treino para sistemas de aprendizagem de máquina, como “carros autônomos” e software de reconhecimento facial.<sup>2</sup> A IA prospera sobre a espinha dorsal do trabalho humano e, como observou Bainbridge (1983) em *Ironias da automação*, quanto mais avançada a tecnologia, mais crucial a contribuição do ser humano. À medida que os sistemas de reconhecimento de imagem se tornam mais avançados, as imagens que os humanos precisam rotular e anotar tornam-se mais difíceis, dificultando a tarefa para as pessoas.

O que um exame atento do funcionamento dos sistemas inteligentes revela é que os sistemas de IA não apenas são sempre sistemas humano-máquina, como também são inseparáveis dos modelos de negócio da indústria que os desenvolve e implementa, os quais são movidos pelo lucro. Os sistemas

<sup>2</sup> V. Schmieg; Lorusso, 2017.

de IA estão interligados com os seres humanos (não são entidades separadas) e servem como uma influência constitutiva do nosso ser. Na nossa perspectiva, usar seres humanos para realizar “microtrabalhos” mal remunerados a fim de viabilizar a IA é desumanizante, seguindo a categoria labor de Hannah Arendt. De forma mais ampla, o desequilíbrio de poder entre aqueles que produzem e controlam a tecnologia e a priorização dos lucros financeiros como objetivos centrais significa que as máquinas são usadas pelos poderosos e privilegiados como ferramentas que prejudicam o bem-estar humano.

### *III.4. Concluindo: retomar o controle*

Em outubro de 2019, Emily Ackerman, uma cadeirante, descreveu a sua experiência de ficar “presa” na rua por um robô da Starship Technologies. Esses robôs usam a rampa de acesso para atravessar ruas e um deles bloqueou o seu acesso à calçada. “Posso garantir que, enquanto eles [robôs] continuarem a operar, serão um grande problema de acessibilidade e segurança”, lamenta Ackerman (Pitt..., 2019). Questões tais como se esses robôs têm o direito de usar o espaço público e se uma proibição poderia infringir os “seus” direitos, debatidas no discurso sobre os “direitos dos robôs”, priorizam as preocupações erradas. É como proteger a arma em vez da vítima. A preocupação principal deveria ser com o bem-estar dos grupos marginalizados (no caso, os cadeirantes), que são desproporcionalmente impactados pela integração da tecnologia no nosso cotidiano.

Quando um filósofo contempla quais seriam as condições ontológicas para que algo receba direitos, é fácil acabar em argumentos que comparam “os direitos do ser humano” com “os direitos do robô”. Contudo, essa comparação se baseia numa crença, a nosso ver, falsa, que encara o ser humano como apenas uma máquina complicada e que acredita que máquinas complexas feitas pelo homem poderiam, portanto, replicar o ser humano. Com base na perspectiva pós-cartesiana e corpórea que defendemos, sustentamos que, embora o ser humano possa incorporar e estender-se a si mesmo na criação e uso de máquinas, a máquina inteligente permanece uma fantasia. Mais ainda, o que vemos é que, na busca por essa fantasia, são criadas máquinas reais e esses algoritmos muito reais de reconhecimento de padrões e processamento de dados estão cada vez mais *obstruindo* o bem-estar humano, ao ponto de contribuírem para a desumanização de seres humanos reais.

De volta à realidade, o que temos de fato em mãos são situações em que um ser humano (um cadeirante) tem a sua liberdade de movimento negada por uma máquina, usada por uma empresa que monopoliza o espaço público para ganho financeiro.

Para concluir, voltamo-nos para a responsabilidade. A nosso ver, são as empresas, os engenheiros, os formuladores de políticas e o público em geral que têm a responsabilidade de garantir os direitos de cada indivíduo. Uma das questões mais prementes da atualidade é que as máquinas “inteligentes” são cada vez mais utilizadas para sustentar formas de opressão. Não “culpamos” as máquinas (elas não podem ser culpadas), nem dizemos que as máquinas devem arcar com a “responsabilidade” (Coeckelbergh, 2019), precisamente porque isso isentaria os verdadeiramente responsáveis das suas obrigações. Concordamos que, na complexa sociedade em rede de hoje, pode

ser muito difícil, se não muitas vezes impossível, rastrear a imputabilidade até indivíduos específicos (Coeckelbergh, 2019). Mas esse fato da vida (a sua complexidade) não é de todo um argumento para tornar as máquinas responsáveis. Ao fazer com que os robôs bloqueiem parte da calçada – uma calçada que foi projetada para permitir que os cadeirantes se deslocassem de forma independente no trânsito da cidade –, estamos retirando parte da estrutura sociotécnica que apoiava um grupo marginalizado no exercício da sua autonomia, tudo em nome de um negócio movido por ganhos financeiros.

De uma forma mais geral, transferir cada vez mais o controle de processos complexos para máquinas inteligentes – ou seja, terceirizar o nosso pensamento e a nossa tomada de decisões para essas tecnologias, por assim dizer – pode, na verdade, ir contra o empoderamento dos seres humanos. Pode até mesmo impedi-los de assumir as responsabilidades inerentes aos direitos humanos.

## Referências bibliográficas

- ANGWIN, Julia; LARSON, Jeff; MATTU, Surya; KIRCHNER, Lauren. Machine bias. **ProPublica**, 23 maio 2016. Disponível em: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.
- ARENDT, Hannah. **The human condition**. Chicago: University of Chicago Press, 1958.
- ASARO, Peter M. What should we want from a robot ethic. **International Review of Information Ethics**, v. 6, n. 12, p. 9-16, 2006.
- BAINBRIDGE, Lisanne Bainbridge. Ironies of automation. In: JOHANNSEN, G.; RIJNSDORP, J.E. (ed.). **Analysis, design and evaluation of man-machine systems**. Oxford: Pergamon, 1983, p. 129-135.
- BAXTER, Gordon D.; ROOKSBY, John; WANG, Yuanzhi; KHAJEH-HOSSEINI, Ali. The ironies of automation: still going strong at 30?. **ECCE '12: Proceedings of the 30th European Conference on Cognitive Ergonomics**, 2012, p. 65-71. DOI: <https://doi.org/10.1145/2448136.2448149>.
- BENJAMIN, Ruha. **Race after technology**: abolitionist tools for the new jim code. Cambridge: Polity, 2019.
- BIRHANE, Abeba. Descartes Was Wrong: 'A Person Is a Person through Other Persons'. **Aeon**, 7 abr. 2017. Disponível em: <https://aeon.co/ideas/descartes-was-wrong-a-person-is-a-person-through-other-persons>.
- BRANNIGAN, Augustine. Stanley Milgram's obedience experiments: A report card 50 years later. **Society**, v. 50, n. 6, p. 623-628., 2013. DOI: <https://doi.org/10.1007/s12115-013-9724-3>.
- BROOKS, Rodney Brooks. Will robots demand equal rights? **Time**, 19 jun. 2000.
- BRYSON, Joanna J. Robots should be slaves. In: WILKS, Y. (ed.). **Close engagements with artificial companions**: key social, psychological, ethical and design issues. Amsterdam: John Benjamins, 2010, p. 63-74.
- CHENEY-LIPPOLD, John. **We are data**: algorithms and the making of our digital selves. Nova Iorque: NYU Press, 2018.
- CHURCHLAND, Paul M. **Matter and consciousness**. Cambridge: MIT, 2013.
- CLARK, Andy Clark. **Being there**: putting brain, body, and world together again. Cambridge: MIT, 1998.
- COECKELBERGH, Mark. Robot rights? Towards a social-relational justification of moral consideration. **Ethics and Information Technology**, v. 12, n. 3, p. 209-221, 2010. DOI: <https://doi.org/10.1007/s10676-010-9235-5>.
- COECKELBERGH, Mark. Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability. **Science and engineering ethics**, v. 26, n. 4, p. 1-18, 2019. DOI: <https://doi.org/10.1007/s11948-019-00146-8>.

DENNNETT, Daniel C. **The intentional stance**. Cambridge: MIT, 1987.

DI PAOLO, Ezequiel A.; CUFFARI, Elena Clare; DE JAEGHER, Hanne. **Linguistic bodies**: The continuity between life and language. Cambridge: MIT, 2018.

DREYFUS, Hubert L.; DREYFUS, Stuart E. The ethical implications of the five-stage skill-acquisition model. **Bulletin of Science, Technology & Society**, v. 24, n. 3, p. 251-264, 2004. DOI: <https://doi.org/10.1177/0270467604265023>.

EUBANKS, Virginia. **Automating inequality**: How high-tech tools profile, police, and punish the poor. Nova Iorque: St. Martin's Press, 2018.

FERRYMAN, Kadija; PITCAN, Mikaela Pitcan. Fairness in precision medicine. **Data & Society**, 2018. Disponível em: <https://datasociety.net/library/fairness-in-precision-medicine>.

GOLONKA, Sabrina; WILSON, Andrew D. Gibson's ecological approach. **Avant: Trends in Interdisciplinary Studies**, v. 3, n. 2, p. 40-53, 2012.

GUNKEL, David J. The rights of machines: Caring for robotic care-givers. In: VAN RYSEWYK, S.P.; PONTIER, M. (ed.). **Machine Medical Ethics**. Cham: Springer, 2015, p. 151-166.

GUNKEL, David J. **Robot rights**. Cambridge: MIT, 2018.

INTRONA, Lucas; NISSENBAUM, Helen. Defining the web: The politics of search engines. **Computer**, v. 33, n. 1, p. 54-62, 2000. DOI: <http://dx.doi.org/10.1109/2.816269>.

KEYES, Os. The misgendering machines: Trans/HCI implications of automatic gender recognition. **Proceedings of the ACM on Human-Computer Interaction**, v. 2, n. 88, p. 1-22, 2018. DOI: <https://doi.org/10.1145/3274357>.

KURZWEIL, Ray. **The singularity is near**: When humans transcend biology. Nova Iorque: Penguin, 2005.

LAMBRECHT, Anja; TUCKER, Catherine. Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads. **Management Science**, v. 65, n. 7, p. 2947-3448, 2019. DOI: <https://doi.org/10.1287/mnsc.2018.3093>.

LAVE, Jean. **Cognition in practice**: Mind, mathematics and culture in everyday life. Cambridge: Cambridge University Press, 1988.

MCQUILLAN, Dan. Data science as machinic neoplatonism. **Philosophy and Technology**, v. 31, n. 2, p. 253-272, 2018. DOI: <https://doi.org/10.1007/s13347-017-0273-3>.

OBERMEYER, Ziad; MULLAINATHAN, Sendhil. Dissecting racial bias in an algorithm that guides health decisions for 70 million people. **Proceedings of the Conference on Fairness, Accountability, and Transparency**, 2019, p. 89. DOI: <https://doi.org/10.1145/3287560.3287593>.

O'NEIL, Cathy. **Weapons of math destruction**: How big data increases inequality and threatens democracy. Nova Iorque: Broadway Books, 2016.

PITT pauses testing of Starship robots due to safety concerns. **The Pitt News**, 21 out. 2019. Disponível em: <https://pittnews.com/article/151679/news/pitt-pauses-testing-of-starship-robots-due-to-safety-concerns>.

RICHARDSON, Rashida; SCHULTZ, Jason; CRAWFORD, Kate. Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. **New York University Law Review**, v. 94, n. 15, p. 15-55, 2019. Disponível em: <https://nyulawreview.org/online-features/dirty-data-bad-predictions-how-civil-rights-violations-impact-police-data-predictive-policing-systems-and-justice>.

RUSHKOFF, Douglas. **Team human**. Nova Iorque: WW Norton & Company, 2019.

SCHMIEG, Sebastian; LORUSSO, Silvio. **Five Years of Captured Captchas**. 2017. Disponível em: <http://five.years of capturedcapt.ch/as>.

SCHUTZ, Alfred; LUCKMANN, Thomas. **The structures of the life-world**. Vol. 1. Evanston: Northwestern University Press, 1973.

STONE, Christopher D. Should trees have standing-toward legal rights for natural objects. **Southern California Law Review**, n. 45, p. 450-501, 1972.

STRAUCH, Barry. Ironies of automation: still unresolved after all these years. **IEEE Transactions on Human-Machine Systems**, v. 48, n. 5, p. 419-433, 2017. DOI: <https://doi.org/10.1109/THMS.2017.2732506>.

SUCHMAN, Lucy A. **Human-machine reconfigurations**: Plans and situated actions. Cambridge: Cambridge University Press, 2007.

TUBARO, Paola; CASILLI, Antonio A. Micro-work, artificial intelligence and the automotive industry. **Journal of Industrial and Business Economics**, v. 46, n. 3, p. 333-345, 2019. DOI: <https://doi.org/10.1007/s40812-019-00121-1>.

VAN DIJK, Jelle. Designing for embodied being-in-the-world: A critical analysis of the concept of embodiment in the design of hybrids. **Multimodal Technologies and Interaction**, v. 2, n. 1, p. 1-21, 2018. DOI: <https://doi.org/10.3390/mti2010007>.

VERBEEK, Peter-Paul. **De daadkracht der dingen**: over techniek, filosofie en vormgeving. Meppel: Boom, 2000.

WHITBY, Blay. Sometimes it's hard to be a robot: A call for action on the ethics of abusing artificial agents. **Interacting with Computers**, v. 20, n. 3, p. 326-333, 2008. DOI: <https://doi.org/10.1016/j.intcom.2008.02.002>.

WILSON, Benjamin; HOFFMAN, Judy; MORGENSTERN, Jamie. Predictive inequity in object detection. **ArXiv**, 21 fev. 2019. DOI: <https://doi.org/10.48550/arXiv.1902.11097>.

ZUBOFF, Shoshana. **The age of surveillance capitalism**: The fight for a human future at the new frontier of power. Londres: Profile, 2019.