

A ARTIFICILIDADE DA VIDA HUMANA

reflexões filosóficas sobre Inteligência
Artificial

THE ARTIFICIALITY OF HUMAN LIFE
philosophical reflections on Artificial Intelligence

Francisco Erik Marques da Silva¹
Universidade Federal do Ceará

¹ Doutorando em Filosofia no Programa de Pós-Graduação em Filosofia da Universidade Federal do Ceará (UFC).

E-mail: chicoerik@alu.ufc.br.

Lattes: <http://lattes.cnpq.br/2789388241662008>. Orcid: <https://orcid.org/0000-0002-7913-0232>.

RESUMO: Refletindo na problemática relação: seres humanos e as diversas IAs existentes no nosso mundo atual, este ensaio intenta articular a discussão de alguns filósofos que analisam o tema, como Floridi e Cowls (2019), Russell (2021) e Mbembe (2021), para compreender como os princípios epistemológicos e éticos estão presentes nas categorizações de IA do século XXI, bem como na concepção da redução do ser humano ao maquinário e material bruto da relação entre Tecnologia e Humanidade. Deste modo, como programar princípios éticos na IA se o ser humano no século XXI incorre cada vez mais em um processo de desumanização e brutalização? A IA alcança sua inteligibilidade dentro da dinâmica da naturalização da guerra social? Partindo disso, a metodologia articula-se de tipo qualitativa com abordagem bibliográfica. Nas suas considerações, encaminha dialogando entre o aspecto Humanidade e a IA, resultando em uma discussão filosófica sobre os limites do que é ser humano.

Palavras-chave: Inteligência Artificial. Humanidade. Filosofia.

ABSTRACT: Reflecting on the problematic relationship: human beings and the various AIs that exist in our current world, this article attempts to articulate the discussion of some philosophers who analyze the topic, such as Floridi and Cowls (2019), Russell (2021) and Mbembe (2021), to understand how epistemological and ethical principles are present in the categorization of AI in the 21st century, as well as in the conception of the reduction of human beings to machinery and raw material in the relationship between Technology and Humanity. Therefore, how can ethical principles be programmed into AI if human beings in the 21st century increasingly incur a process of dehumanization and brutalization? Does AI achieve its intelligibility within the dynamics of the naturalization of social war? Based on this, the methodology is qualitative and has a bibliographical approach. In his considerations, he dialogues between the aspect of Humanity and AI, resulting in a philosophical discussion about the limits of what it means to be human

Keywords: Artificial Intelligence. Humanity. Philosophy.

INTRODUÇÃO

Este ensaio incorre na tentativa de realizar um diálogo entre as discussões sobre a IA, suas conceitualizações e caracterizações, em uma perspectiva da Filosofia da Inteligência Artificial e o pensamento de Achille Mbembe (2021), quando conceitualiza aquilo que chama de *Brutalismo*, condição de brutalização da vida humana. Estamos neste século, da quarta revolução industrial, indústria 4.0, escola tecnológica e da nanociência.

Nesse ínterim, preocupa-me a relação entre a Filosofia da Inteligência Artificial, dentro do seu aspecto epistemológico, ético e a interação do humano com a máquina/algoritmo dentro do *devir humano-artificial* e *devir humano-máquina* pensando o processo de brutalização da vida e as possibilidades de desumanização nas relações da inteligência. Deste modo, se coloca um questionamento: Só o ente humano é dotado de pensamento? O antropocentrismo da vida seria o maior risco de uma super IA? Se o conhecimento é apenas humano, como acredita-se desde Aristóteles, as criações tecnológicas humanas não operam do mesmo modo?

Partindo dessas questões, se constrói a argumentação deste ensaio com o intuito de compreender as possibilidades humanas dentro da epistemologia e sua relação com as diversas inteligências existentes no nosso planeta e como estas estão ou não em biointegração. Estamos cada vez mais nos distanciando da natureza e nos artificializando? A inteligência artificial é o demonstrativo de que não somos mais animais, mas sim divinos humanos? O humano se demasiou em humanidade e recaiu na sua mais alta antropocentrização?

Metodologicamente, o ensaio segue uma abordagem qualitativa de tipo interpretativa-documental para investigarmos as diversas obras aqui tratadas (Mbembe (2021); Russell (2021); Silva (2022); Floridi; Cows (2019), identificando de que modo os conceitos de cada pensador podem ser trabalhados e como estes se relacionam com a problemática da relação entre humano e máquina, inteligência humana e artificial, para respondermos ou não os questionamentos aqui colocados.

Assim, este ensaio divide-se em 5 tópicos (a contar com esta introdução e as considerações finais). No primeiro tópico se apresenta o tema, a problemática e a metodologia, no segundo, expõe-se a relação entre inteligência humana e artificial, para depois discorrer sobre o *devir-artificial da humanidade* e as IA (terceiro tópico), essa discussão culmina em apresentar a condição da artificialidade da vida para, por fim, nas considerações finais, argumentar sobre o brutalismo da vida humana em sua simbiose com objeto, maquinários e substâncias que tratam da materialidade robótica e não-orgânica.

1 INTELIGÊNCIA HUMANA – ARTIFICIAL: O HUMANO E O NÃO HUMANO

A discussão sobre inteligência já remonta aspectos que vão desde a Grécia antiga aos dias atuais, no entanto, quando este assunto foca naquilo que se entende por “artificial”, ou seja, o contrário de real, de orgânico, vê-se que as ciências da computação dominam essa definição há muito tempo. Assim, o problema de se pensar a IA através de uma perspectiva evolucionista parte da ideia de que toda tecnologia, não importa quão avançada seja, começa com a mudança daquilo que é natural.

Existem diversas abordagens para conceituar o que seria inteligência, no entanto, não há um consenso único para definir o que se entende exatamente por inteligência, de acordo com Gasset (1963, pp. 73-74): “Se recorrêssemos aos especialistas de diversas áreas do conhecimento na intenção de fazer um levantamento pelo que entendemos em relação a este conceito, então, chegaríamos a diversas definições”. Portanto, para não recorrermos em uma revisão teórica prolongada sobre esse conceito, partiremos da concepção de Alves (2023) e Medeiros (2018), que entendem que a inteligência consiste em quatro pontos: a) a capacidade de resolução de problemas; b) o aprendizado com o ambiente; c) o desenvolvimento de estruturas cognitivas; d) a capacidade de seguir orientação e metas.

Nesse sentido, a fim de melhor definir em que consiste este conceito, pode-se, de modo antagônico, buscar primeiro definir o que é um ser humano. Outra definição complexa que remontaria toda uma revisão bibliográfica para tal, no entanto, instiga-nos, a definição apresentada por Nietzsche (2009, p. 20): “O homem é uma corda estendida entre o animal e o além-do-homem – uma corda sobre um abismo.”

Partindo da implicação nietzscheana sobre o ser humano, uma corda entre o animal e o além de si mesmo, Alves (2023), elabora uma definição que tenta dar conta da miríade do que nós podemos ser, assim, elenca três aspectos que definiriam o que é ser humano:

1) Aspecto biológico: o ser humano é um bípede ereto de polegar opositor e seu corpo é incrivelmente capaz de sentir uma gama enorme de estímulos dolorosos ou prazerosos, esses procedimentos, são correspondentes às funções neurais do cérebro e com a conexão com o restante do corpo a partir do ligamento entre a medula oblongada com e a medula espinhal.

2) Aspecto psicológico: todo ser humano é um indivíduo complexo, capaz de interagir com os outros exemplares da mesma espécie e diferencia-se dos animais não-humanos por possuir uma capacidade cerebral que reflete e pondera abstratamente sobre o passado e futuro. É capaz de formar conceitos de linguagem, dando significado às palavras e capaz de fazer cálculos matemáticos; a sua psicologia diz respeito à sua mente (diferente de cérebro), e, por conta disso, consegue controlar seus instintos, embora ainda possa agir por meio deles. A Capacidade psicológica



também faz os seres humanos se diferenciarem do restante dos outros animais porque possui o poder de menosprezar a si mesmo e de dar fim a sua própria vida. Ainda, na consideração psicológica, há a noção de que a vida é finita, algo que, amplia-se diretamente com relação ao aspecto moral, isso diz respeito à capacidade de devoção, elaboração de ritos e adoração a deuses.

3)Aspecto moral: o ser humano possui a capacidade de diferenciar o certo do errado e o bem do mal (desconsidere-se aqui a valoração moral); inteligência para construir ou destruir; causar alegria ou sofrimento, etc. Há no ser humano também a faculdade de discernir e se relacionar com o mundo, criando as instituições regulamentadoras dos normativos de sanidade e normalidade comportamental, por exemplo, o Estado, as leis e as regras de conduta em geral. O ser humano possui a capacidade de aprimoramento moral, compaixão, egoísmo e de contemplação artística. No aspecto moral, os dois maiores sentimentos que podem ser atribuídos ao ser humano são amor e ódio (Alves, 2023, pp. 132-133).

Nos três aspectos apresentados pelo filósofo (*biológico - psicológico - moral*) o que mais chama atenção é a argumentação proveniente desde a Grécia de diferenciar o humano do animal (algo que vemos na própria compreensão nietzschiana) como se o próprio humano fosse um super animal, ou melhor, devesse ser um super humano. Quando se trata de refletir sobre inteligência artificial, um dos primeiros argumentos que se reivindica é a superioridade humana diante das outras possibilidades de vida. Seríamos nós os escolhidos da criação, como vemos nas narrativas judaico-cristãs?

Antônio Bispo dos Santos (2023), filósofo quilombola, nos tece uma reflexão sobre a humanidade no que tange a sua categorização enquanto distanciamento do mundo animal:

A humanidade se desconectou da natureza exatamente por ter cometido o pecado original. Seu castigo foi se afastar da natureza. Por isso Adão foi expulso do Jardim do Éden e o humanismo passou a ser um sistema, um reino desconectado do reino animal. Dentro do reino vegetal, todos os vegetais cabem, dentro do reino mineral, todos os minerais cabem. Mas dentro do reino animal não cabem os humanos. Os humanos não se sentem como entes do ser animal. Essa desconexão é um efeito da cosmofofia (Santos, 2023, pp. 8-9).

Qual seria o reino dos humanos? O que seria ser humano? Um ente dotado de racionalidade que superou sua própria natureza e controla seus instintos, como discorre boa parte da filosofia? Dentro dessa relação entre humano e animal, vê-se o que poderia se discorrer como princípio do artificial: o não natural. Ou seja, o artificial seria aquilo que se distancia do natural, e se nós incorremos cada vez mais para nos distanciarmos do natural, estaríamos nós rumo a uma artificialização da nossa própria condição humana?

Assim, o filósofo quilombola vem e nos atravessa:

Não somos humanistas, os humanistas são as pessoas que transformam a natureza em dinheiro, em carro do ano. Todos somos cosmos, menos os humanos. Eu não sou

humano, sou quilombola. Sou lavrador, pescador, sou um ente do cosmos. Os humanos são os eurocristãos monoteístas. Eles têm medo do cosmos. A cosmofofia é a grande doença da humanidade. Apesar de serem criaturas da natureza, os humanistas descolam da natureza e se tornam criadores. Daí sua necessidade de sintetizar o orgânico, de chamar todas as vidas de matéria-prima. Essa matéria-prima passa a ser um objeto a ser melhorado, beneficiado e sintetizado pelos humanos. Eles se sentem os donos da inteligência, se sentem o próprio deus – o deus na lógica da verticalidade, na lógica do poder, da interferência na vida alheia e da manipulação, e não um deus na lógica da biointeração (Santos, 2023, p. 16).

Os humanistas são os que criam IAs para realizarem aquilo que não podem realizar? Estaríamos nós nos tornando cada vez mais *cosmofóbicos*, ou seja, que segundo Santos (2023), são aqueles que têm medo do natural e criam realidades não-orgânicas para não lidar com o orgânico? Em todo filme sobre super Inteligência Artificial, qual a primeira coisa que reparamos na paisagem? A não existência de outras formas vidas e uma paisagem da cidade cinza, tecnológica e não orgânica. Tal aspecto não é coincidência ou distante da nossa realidade atual, demonstra como o humano vem construindo na sua inteligência, sapiência e epistemologia, em critérios não-naturais, ou como discorre Mbembe: “A distinção entre a vida e a não vida” (Mbembe, 2020, p. 25).

Quando observamos as definições de IA, como a descrita por Alves, A (2020), “você pode responder, de modo sintetizado, o que é uma inteligência artificial?”. Resposta: “A Inteligência Artificial é um campo da ciência da computação que desenvolve algoritmos e sistemas capazes de realizar tarefas que normalmente requerem inteligência humana, como aprendizado, raciocínio, reconhecimento de padrões, linguagem natural e resolução de problemas complexos. Assim, a IA pode ser dividida em dois tipos principais: a IA fraca, que é programada para realizar tarefas específicas (como algoritmos de moderador de conteúdo em redes sociais) e a IA forte, que tem a capacidade de aprender e adaptar-se a novas situações de forma autônoma se tornando assim super inteligentes. Em se tratando dessas IAs, o seu desenvolvimento está nos distanciando de nossa humanidade ou nos aproximando?

Pensando nesses paradigmas, vemos a humanidade se colocar em um processo de menos envolvimento com a natureza: convivendo pouco com as árvores, com a terra, com as matas e mais com as tecnologias. O famoso desenvolvimento técnico acaba sendo sinônimo de desconectar, tirar do cosmo, quebrar a originalidade. Essa originalidade é entendida por Bispo (2023), como sendo o natural, e nós, acabamos por humanizar e tornar sintético tudo o que é original, ou melhor, tornar tudo artificial, um *devoir humano-artificial*. Tudo se instaura em artificialidades e o *devoir* é justamente o processo de tornar uma coisa em outra, deste modo, o *devoir* humano é tornar tudo humano, e o humano-artificial é tornar tudo artificialmente humano.

2 INTELIGÊNCIA ARTIFICIAL E O DEVIR-ARTIFICIAL HUMANO: ALGORITMOS, VIDA E CONTROLE

Deparando-se com a relevante obra do filósofo Stuart Russell, isto é, *Inteligência artificial a nosso favor: Como manter o controle sobre a tecnologia (2021)*, vê-se uma definição do conceito de inteligência de forma concisa e objetiva. Para Russell, o conceito de inteligência é: “Os seres humanos são inteligentes na medida em que suas ações sejam capazes de atingir seus objetivos” (2021, p. 12). Todos os outros aspectos de inteligência — percepção, pensamento, aprendizado, invenção e assim por diante — podem ser avaliados por sua contribuição à nossa capacidade de atuar com êxito. Do mesmo modo, a inteligência nas máquinas tem sido definida da mesma maneira: “Máquinas são inteligentes na medida em que suas ações sejam capazes de atingir seus objetivos” (Russell, 2021, p. 12).

Assim, segundo o filósofo, “as máquinas, diferentemente dos humanos, não têm objetivos próprios, nós lhes damos objetivos para atingir. Em outras palavras, construímos máquinas otimizadas, incutindo-lhes objetivos, e lá se vão elas” (Russell, 2021, p. 12). Ou seja, umas das principais características que nos fazem humanos é a autonomia, a capacidade de criarmos nossas próprias regras, caminhos e objetivos, diferindo assim das máquinas, que precisam ser programas para determinada ação. A autonomia e autodeterminação são capacidades extremamente ligadas à nossa comunicação, a nossa linguagem, tendo em vista que é por meio dela “que continuamente criamos nosso mundo e organizamos nosso mundo interior e exterior, constituindo-se como ser autônomo.” (Humboldt, 1998, p.30). Desse modo, no fundo, o que colocaria a humanidade na coroa da criação é a linguagem, “um dos traços singulares que lhe conferem não apenas singularidade, mas também gênio” (Mbembe, 2021, p. 39).

Seguindo a reflexão, pensando junto com Mbembe, filósofo africano, o ser humano, ao dominar os signos/significados do mundo pela linguagem, “dotou-se dos meios para perceber o real para habitar o espaço e o tempo, e acima de tudo, participar da revelação da verdade” (Mbembe, 2021, p. 39). Desta forma, a linguagem é a chave de acesso ao ser, sentido e conhecimento, é por meio da linguagem que se pode condicionar a tecnologia como sendo o *Outro* do humano, visto que nesses aspectos ela não seria capaz de construir um mundo simbólico/linguístico-social como o dos humanos, a não ser se for programada para isso.

No entanto, o que vemos hoje em dia nos coloca em outros paradigmas epistemológicos e existências, já que “a tecnologia está se estabelecendo como destino ontológico de todos os seres vivos” (Mbembe, 2021, p. 43). No século XXI, a tecnologia tem se tornado a clínica na qual se superam aquilo que o humano não é capaz de realizar, é por meio dela que nos tornamos capazes de criar inteligências

que superam nossas limitações, como é o caso do AlphaGo, um programa criado por combinações de redes neurais que operam em camadas não lineares de relação. Este programa foi criado com o objetivo de jogar um jogo complexo de tabuleiro, o GO. O AlphaGo é uma IA que “pensa fora da caixa” diferente de todas as outras, emulando a mente humana. Só que era preciso provar se ele funcionava: para isso a Google convidou o campeão europeu Fan Hui para uma partida de Go contra o super computador e este venceu o campeão mundial.²

Diante do exemplo do AlphaGo, percebe-se que pode não haver mais “divisão entre o humano e o objeto, o humano e a máquina, ou entre o humano e objeto técnico” (Mbembe, 2021, p. 44). Seria aquilo que o filósofo Mbembe descreve como *devir humano-máquina*, ou seja, estamos entrelaçados de forma quase indissociável às máquinas e delas dependemos e humanizamos, tanto que a discussão sobre Inteligência Artificial está cada vez mais recaindo no aspecto de não-diferenciação entre nós e ela.

Por isso, Russell (2021) coloca:

Dizemos que as máquinas são inteligentes na medida em que se pode esperar que suas ações atinjam seus objetivos, mas não temos nenhuma forma confiável de garantir que os objetivos delas sejam os mesmos que os nossos. E se, em vez de permitir que máquinas busquem atingir seus objetivos, insistirmos em que elas busquem atingir os nossos? Uma máquina como essa, se pudesse ser projetada, seria não apenas inteligente, mas também benéfica para os seres humanos (Russell, 2021, p. 15).

Devemos nos resguardar dessa possibilidade? As máquinas devem realizar nossos objetivos e desejos? Mas, os objetivos e desejos de quem exatamente? Das grandes empresas de tecnologias? Do Estado? Uma resposta complexa que o próprio filósofo coloca em jogo no seu texto. Esse aspecto, que para além de epistemológico é sobre ética, dignidade humana e sociedade, os filósofos Luciano Floridi e Josh Cowls discutem no artigo *A Unied Framework of Five Principles for AI in Society (2019)*. São 5 princípios fundamentais que a IA, ou melhor, os programadores da IA deveriam seguir em seu processo de desenvolvimento e programação.

Os cinco princípios descritos pelos filósofos são: beneficência, não maleficência, autonomia, justiça e explicabilidade. O que interessa nesta argumentação é o quinto princípio, tendo em vista que ele resume todos os outros e é o articulador principal:

A explicabilidade, entendida como incorporando tanto o sentido epistemológico de ininteligibilidade (como uma resposta à pergunta "como funciona?") quanto o sentido

² Para saber mais a respeito, conferir o seguinte sítio: <https://meiobit.com/336028/google-inteligencia-artificial-alphago-heuristica-supera-jogador-profissional-go/>.

ético de responsabilidade (como uma resposta à pergunta: "quem é responsável pela forma como funciona?") (Floridi; Cowls, 2019, p. 2)

Pensando na linha de Floridi e Cowls, o pensador brasileiro Tarcísio Silva, no seu livro *Racismo Algorítmico: inteligência artificial e discriminação nas redes sociais* (2022), coloca uma definição interessante para pensarmos as possibilidades de uma IA com os princípios descritos acima:

Em linhas gerais, a busca pela inteligência artificial geral significa tentar emular capacidades da mente humana nas diversas esferas da experiência, produzir comportamento autônomo, independente e proativo e aprender sobre esse comportamento de forma criativa. Os robôs completamente autônomos e criativos que povoam as narrativas da ficção científica e do cinema representam geralmente o ápice da inteligência artificial geral – que provavelmente nunca será de todo alcançada. A controvérsia sobre a viabilidade da inteligência artificial geral é algo de que não trataremos neste livro. Sobretudo porque vemos como negativa a tendência a hipervisibilizar os debates filosóficos sobre robôs autônomos e seus possíveis direitos no futuro ante a realidade material do impacto da inteligência artificial estreita na vida contemporânea (Silva, 2022, p. 59).

O que o pensador coloca em questão é justamente aquilo que Floridi e Cowls (2019) e outros filósofos que pensam IA estão colocando em perspectiva, isto é: a IA forte (aquela que conseguiria constituir consciência e autonomia), seria capaz de superar os seres humanos em todos os aspectos possíveis? No entanto, o problema colocado por Tarcísio Silva é relevante, visto que ele defende que “provavelmente ela nunca será de todo alcançada” (Silva, 2022, p. 59). Para o autor, se trata apenas de uma ficção que alimenta o imaginário consumidor de filmes sensacionalistas, por isso ele vê “como negativa a tendência a hipervisibilizar os debates filosóficos sobre robôs autônomos e seus possíveis direitos no futuro ante a realidade material do impacto da inteligência artificial estreita na vida contemporânea” (Silva, 2022, p. 59).

Diante do exposto, pode-se inferir que o autor atenta para que a discussão sobre a IA forte tira o foco do que de fato seria mais relevante no momento em que vivemos, isto é, a IA fraca (moderador de conteúdo e diversos comandos utilizados em setores como educação, saúde e economia) e sua capacidade de interferir em nossas vidas de forma direta, individual e coletiva. A IA entendida enquanto fraca são utilizadas, cotidianamente, em sistemas automatizados que resolvem tarefas mais “simples” em domínios específicos, como reconhecimento de padrões em dados ou processamento de linguagem natural, ou seja:

Na atualidade, grandes conglomerados de tecnologia, como o chamado Gafam (sigla para Google, Amazon, Facebook, Apple e Microsoft) e outras empresas de porte similar, dominam a oferta de recursos de inteligência artificial em escala por meio da computação em nuvem, além de investir em startups e desenvolver e moldar padrões dos campos de implementação. Nessa direção, a pervasividade dos recursos de

inteligência artificial estreita nos ambientes online, como buscadores de informação na web, plataformas de mídias sociais e assistentes automatizados pessoais, virou a chave na consciência pública sobre a questão. Literalmente, interagimos com centenas de sistemas algorítmicos por meio de aplicativos em smartphones (Silva, 2022, p. 59).

Como analisarmos que talvez a nossa maior preocupação dentro do processo ético/epistemológico das IAs, não seja realmente projetar questões e princípios à aplicabilidade no presente para prevenir uma IA superinteligente perversa, mas sim aplicarmos tais princípios nas IAs fracas? No entanto, como faremos este procedimento já que o objetivo dos grandes conglomerados tecnológicos é justamente não barrar o desenvolvimento cada vez mais estreito entre os algoritmos e nossa vida?

No final do seu artigo, Floridi e Cowls discorrem:

O desenvolvimento e o uso da IA têm o potencial de causar impactos positivos e negativos na sociedade, de aliviar ou ampliar as desigualdades existentes, de curar problemas antigos ou de causar novos problemas. Traçar o curso socialmente preferível dependerá não apenas de uma regulamentação bem elaborada e de padrões comuns, mas também do uso de uma estrutura de princípios éticos, na qual ações concretas possam ser situadas. Acreditamos que a estrutura apresentada aqui, que emerge do debate atual, servirá como uma arquitetura valiosa para garantir resultados sociais positivos da tecnologia de IA e passar de bons princípios para boas práticas (Floridi; Cowls, 2019, p. 17).

Floridi e Cowls constroem uma preocupação que se encontra com Tarcísio Silva (2022) e Mbembe (2021), isto é, a preocupação com a dignidade humana, mantendo a autonomia humana em um processo crescente de combate às desigualdades. Contudo, devemos enfrentar a dificuldade de realizar esse procedimento, visto que os sistemas algorítmicos são objetos apenas da ordem do código vistos de forma estreita e plana junto às corporações de tecnologia que apelam para a noção de inescrutabilidade, em diálogo com valores como “segredo de negócio” postos à frente do bem comum e da coletividade (Silva, 2022).

Para além disso, a condição primordial de linhas de código e procedimentos, os algoritmos imprimem impactos sociais quando se tornam ferramentas que “decretam objetos de conhecimento e sujeitos de prática em maneiras mais ou menos significantes [...]”. Suas ações não apenas estão no mundo, mas fazem mundos” (Silva, 2022, p. 66). Assim, os algoritmos mediam quais sujeitos são ou não incluídos, como são ordenados, suas hierarquias de valor ante os objetos e o capital e também situações em que vidas são consideradas descartáveis.

Articulando as reflexões de Floridi e Cowls (2019) junto às de Tarcísio Silva (2022), pode-se notar que no princípio de “Explicabilidade” defendido pelos filósofos, está contida a ideia de que:

A decisão sobre quem deve decidir deve ser informada pelo conhecimento de como a IA agiria em nosso lugar; e para que a IA seja justa, devemos saber a quem responsabilizar no caso de um resultado grave e negativo, o que exigiria, por sua vez, uma compreensão adequada do motivo desse resultado (Floridi; Cowls, 2019, p. 14).

Os próprios seres humanos devem ser detentores de uma ética aplicada a IA, de modo que a ação gerada pela IA deve respeitar os princípios e autonomia humana. Aqui sabemos que a IA da qual Floridi especula é uma IA mais avançada, ou seja, a que é caracterizada como Forte e não a Fraca. Pensando no princípio de “explicabilidade” executado na IA fraca, já é observável sua ineficácia, visto que não conseguimos nem “controlar” os moderadores algoritmos que acabam por imprimir preconceitos e noções excludentes dentro de redes sociais como Instagram ou Facebook, como podemos ver no caso a seguir:

[...] “Cace-os, identifique-os e mate-os. Mate a todos. Pelo amor do que é bom e correto. Mate a todos.” O deputado estadunidense Clay Higgins publicou no Facebook esse texto, em que pedia a morte de “muçulmanos radicalizados” depois de um ataque terrorista em Londres. Apesar das denúncias, o texto permaneceu no ar. Pouco antes, uma ativista do Black Lives Matter chamada Didi Delgado falou sobre racismo estrutural nos seguintes termos: “Todas as pessoas brancas são racistas. Comece desse ponto de partida ou você já falhou”. O post foi deletado e seu perfil foi suspenso por sete dias (Silva, 2022, p. 39).

No exemplo acima exposto, é notável uma ação de moderadores de conteúdo de redes sociais agindo de modo a escolher o conteúdo “mais aceitável”. Onde está a inteligibilidade nesse caso? O que os humanos que programam e realizam os algoritmos de moderação escolheram como sendo mais aceitável? Assim, a frequente inconsistência da moderação em casos como esses vem do enquadramento que o negócio da plataforma imprime às práticas humanas e automatizadas de moderação de conteúdo. Após o vazamento de documentos internos, foi descoberta parte da metodologia de treinamento de moderadores humanos que orientam a apagar posts para proteger de ofensas à categoria “homens brancos” enquanto não protegeria “mulheres motoristas” ou “crianças negras”.

Alinhando assim questões éticas e epistemológicas, pergunta-se, até que ponto a IA, seja ela forte ou fraca, concebem padrões de conhecimento próprios ou expressam apenas simulacros das ideias humanas? Outra situação para exemplificar as questões aqui colocadas é o relato pessoal da cientista computacional Joy Adowaa Buolamwini:

Ao suspeitar que a diferença fundamental entre ela e seus colegas era a cor da pele e outros traços fenotípicos, fez um teste simples, apesar de persuasivo: pôs sobre o rosto uma máscara branca genérica, sem feições – e até essa máscara foi reconhecida, em uma reedição involuntária da metáfora do título de Pele negra, máscaras brancas, de

Frantz Fanon. Colocando a máscara branca, Buolamwini conseguia usar os recursos dos equipamentos que não haviam encontrado seu rosto. A experiência se repetiu em diversas ocasiões, o que levou a programadora a investigar as similaridades e os problemas entre os sistemas falhos. Ao identificar a falta de diversidade nas bases de treinamento como um dos principais problemas, Buolamwini passou a produzir alternativas e educação midiática sobre a questão (Silva, 2022, p. 73).

O próprio treinamento dos programadores é limitado, e muito mais seus conhecimentos de diversidade de modo geral, o que acaba por gerar situações como essa. No caso, podemos argumentar que não é exatamente a IA que é racista, mas sim os programadores que acabam por incutir suas noções limitantes no processo de construção algorítmica. Stuart Russell lança alguns questionamentos pertinentes sobre isso:

Um motivo da preocupação com decisões automáticas é o potencial de viés algorítmico — a tendência dos algoritmos de aprendizado de máquina a produzirem decisões inapropriadamente parciais sobre empréstimos, moradia, empregos, seguros, liberdade condicional, sentenças, ingresso na faculdade, e assim por diante. O uso explícito de critérios como raça nessas decisões é ilegal há décadas em muitos países e proibido pelo artigo 9 do Regulamento Geral sobre a Proteção de Dados numa grande variedade de requerimentos. Isso não significa, claro, que excluindo raça dos dados vamos obter decisões sobre raça imparciais. Por exemplo, a partir dos anos 1930, a prática sancionada pelo governo da *redlining* [negação de serviços etc.] fez com que certas áreas de endereçamento postal nos Estados Unidos fossem excluídas de empréstimos hipotecários e outras formas de investimento, levando à desvalorização dos imóveis. Não por coincidência, essas áreas eram habitadas em grande parte por afro-americanos (Russell, 2021, p. 157).

A preocupação descrita por Russel, das ações *tendenciosas* dos algoritmos, se relaciona diretamente com aquilo que estamos refletindo sobre devir-artificial da humanidade.

2.1 A artificialidade da condição humana.

Diante das situações descritas no final da seção anterior, analisa-se o quanto as IAs, sejam fortes ou fracas, como argumentam os filósofos, estão cada vez mais presentes em nossas vidas. O conceito de devir-humano artificial do filósofo camaronês Achille Mbembe (2021) cabe aqui justamente para refletirmos como a própria condição humana se artificializa ou se mecaniza no processo de relação com as máquinas. Hoje em dia dependemos quase de forma integral das IAs, principalmente, as fracas. Seja nas redes sociais ou para tomar decisões simples como escolher um filme ou o que comer. Essa situação é bem refletida por Russell (2021):

Podemos pensar nesse tipo de assalto à dignidade humana de duas maneiras. A primeira é óbvia: quando damos às máquinas autoridade sobre os humanos, nós nos

relegamos a uma condição de segunda classe e perdemos o direito de participar de decisões que nos afetam. A segunda é indireta: ainda que se acredite que não são as máquinas que tomam as decisões, mas os humanos que as projetaram e autorizaram sua produção, o fato de que esses projetistas e comissários humanos não acham que valha a pena levar em conta as circunstâncias individuais de cada sujeito humano nesses casos sugere que eles atribuem pouco valor à vida alheia (Russell, 2021, p.156).

Estamos, mesmo que não acatemos, obliterando nossa autonomia no processo de artificializar nossa condição humana. Pensar no conceito de devir-humano artificial é complexo e delicado, devido sua conceitualização volátil. Mbembe, discorre sobre ele de modo a nos mostrar e sintetizar discussões que ocorrem desde Alan Turing, sobre a substituição do humano pela máquina:

Enquanto a extensão quase indefinida das lógicas de quantificação acarreta uma inesperada aceleração do devir-artificial da humanidade. Esse devir-artificial da humanidade e seu correlato, o devir-humano dos objetos e das máquinas, constituem talvez a verdadeira substância daquilo que alguns chamam atualmente de a “*grande substituição*” (Mbembe, 2021, p. 29, grifo nosso).

O devir-artificial da humanidade se relaciona diretamente com o devir-humano dos objetos e das máquinas, uma relação intrínseca e simbiótica que intensamente toma de conta de toda humanidade, por isso é um devir, isto é, como rio fluído que atravessa todos aqueles que perpassam por ele. É como se não tivéssemos escolha em sermos engolidos pela lógica tecnológica e a forma de relação com a humanidade; “a grande substituição”, o temor de todos aqueles que pensam sobre os avanços tecnológicos e as máquinas. Será que um dia nossa existência se tornará insignificante diante das super máquinas que nós mesmo iremos criar?

Antes de haver qualquer tecnologia, os seres humanos, como a maioria dos animais, viviam para satisfazer as necessidades imediatas, sem guardar nada para depois. Pisamos diretamente no chão, por assim dizer. Aos poucos, a tecnologia nos colocou em cima de uma pirâmide de máquinas, aumentando nossa presença como indivíduos e como espécie. Há diferentes maneiras de projetar as relações entre humanos e máquinas. Se forem projetadas para que os humanos retenham compreensão, autoridade e autonomia suficientes, as peças tecnológicas do sistema podem ampliar imensamente as aptidões humanas, permitindo a cada um de nós ficar em pé sobre uma vasta pirâmide de aptidões — como um semideus, por assim dizer. Mas vejamos o caso da trabalhadora num armazém de preparação e entrega de encomendas. Ela é mais produtiva do que suas antecessoras, pois conta com um pequeno exército de robôs indo buscar as mercadorias onde elas estão estocadas; mas ela é parte de um sistema maior controlado por algoritmos inteligentes que decidem onde deve se posicionar e que itens pegar e despachar. Já está parcialmente soterrada na pirâmide, e não em pé em cima dela. É só uma questão de tempo para que a areia encha os espaços na pirâmide e sua função seja eliminada (Russell, 2021, p. 161).

Na citação acima, observa-se que Russell reflete sobre nossa situação atual em relação aos avanços tecnológicos, o fato da *grande substituição* comentada por Mbembe (2021) ser algo inevitável, seja ela no campo das atividades práticas ou até mesmo ditas como “intelectuais”. Já imaginou máquinas projetando máquinas? Ou até mesmo clonando humanos, como vemos em vários filmes de ficção científica? A grande substituição se trata justamente sobre o encerramento do antropoceno, os seres humanos saindo do centro da manipulação da natureza e de outros seres vivos, para serem manipulados por uma outra espécie de “vida” que seremos nós que iremos projetar, nos exemplos citados denota-se esse acontecimento aos poucos. Essa mesma reflexão, Russell (2021) discorre em relação aos humanos e gorilas:

Há uns 10 milhões de anos, os ancestrais do gorila moderno criaram (acidentalmente, sem dúvida) a linhagem genética dos humanos modernos. Como será que se sente o gorila a esse respeito? Obviamente, se pudessem falar conosco sobre a situação atual de sua espécie em relação aos humanos, a opinião consensual seria muito negativa. Sua espécie, a rigor, não tem nenhum futuro além daquele que acharmos por bem permitir. Não vamos querer estar numa situação parecida com relação a máquinas superinteligentes. Chamemos isso de problema do gorila — especificamente, o problema de os humanos preservarem sua supremacia e autonomia num mundo que inclua máquinas com inteligência substancialmente superior (Russell, 2021, p. 162).

O *problema do gorila* descrito por Russell (2021) é o mesmo que a *grande substituição* argumentada por Mbembe (2021), visto que a questão é a nossa permanência enquanto espécie no topo da pirâmide, pirâmide essa que Russell (2021) já colocou em cheque, iremos ser soterrados por ela se continuarmos achando, ingenuamente, que podemos dominar algo que está além de nossas capacidades.

Não sabemos, não temos respostas, mas continuamos a projetar IAs que dão conta do nosso cotidiano e de nossas necessidades mais básicas; até um ato simples, como ligar uma luz, pedimos para *Alexa*, uma assistente virtual da empresa Amazon que funciona por meio de comandos de voz e que pode ser usada em vários dispositivos. Mas gostamos dessa relação de dar comandos para uma IA, por poder ou por necessidade? O fato de nos sentirmos uma espécie superior, dá a sensação de que nada está além de nossa capacidade intelectual, inclusive, medimos a epistemologia no mundo todo tendo em vista “quem sabe mais”, e nesse “saber mais”, quem “pode mais”.

CONSIDERAÇÕES FINAIS

Este ensaio intentou problematizar aquilo designado de relação entre humano e máquina, mais especificamente, a simbiose de seres humanos com as inteligências artificiais. Dentro dessa problemática filosófica, a reflexão de filósofos renomados, como Stuart Russell (2021) e Luciano Floridi e Cowls (2019), juntamente com o cientista da computação Tarcísio Silva (2023), o pensador africano Achille Mbembe (2021) e o pensador quilombola Nego Bispo (2023), confluem as discussões aqui elaboradas.

Nessa articulação teórica, dentro desses principais pensadores aqui trabalhados, podemos observar que a condição humana e sua relação com as máquinas, sua possível substituição massiva ou brutalização, é algo quase inevitável e que nos preocupa a cada dia. O conceito de *devir-artificial da humanidade* baliza e sintetiza a discussão aqui colocada: “Não há mais, por um lado, a humanidade e, por outro, um sistema dos objetos, em relação ao qual os humanos se situavam como algo sobressalente. Agora somos atravessados de lado a lado pelos objetos, trabalhados por eles tanto quanto nós os trabalhamos.” (Mbembe, 2021, p. 33). Estamos vivendo um momento no qual quase não há mais diferença entre nós e os objetos que utilizamos, existe um “devir-objeto da humanidade que é o contraponto do devir-humano dos objetos. Nós somos o minério que nossos objetos são incumbidos de extrair” (Mbembe, 2021, p. 34). Eles agem conosco, eles nos fazem agir e, acima de tudo, eles nos animam, estimulam e constroem nossas subjetividades.

Esse ensaio não foi uma discussão sobre transhumanismo, no entanto, quase impossível não tratar sobre este tema, visto que nossa relação com as IAs estão estreitas ao ponto de se tornar necessário, como dito por Floridi e Cowls (2019) e Russell (2021), “uma IA a nosso favor”, “uma IA que siga nossos princípios éticos e não nos retire da nossa humanidade e de nossa autonomia”.

Dentro dessa reflexão, o brutalismo se encontra “no ponto em que a imaterialidade, a corporeidade e os materiais se encontram” (Mbembe, 2021, p. 12). Toda a discussão aqui induzida, desde pensarmos sobre “o que nos faz inteligentes” a “como controlarmos o inevitável avanço de uma superinteligência tecnológica”, se dá para problematizarmos que essa temática se insere em um campo com muitas respostas, mas, principalmente, com a conscientização de que as Inteligências Artificiais, sejam elas fortes ou fracas, fazem parte de nossa programação de vida, quiçá, de nossa evolução enquanto espécie, e por isso mesmo “a transformação da humanidade em matéria e energia é o projeto final do brutalismo” (Mbembe, 2021, p. 19).

Ensaio recebido em: 16/09/2024

Ensaio aceito em: 14/01/2025

Ensaio publicado em: 31/03/2025

REFERÊNCIAS

- ALVES, Antonio. REFLEXÕES FILOSÓFICAS SOBRE INTELIGÊNCIA ARTIFICIAL. *Revista Paranaense de Filosofia*, v. 3, n. 1, p. 129-145, 2023. DOI: <https://doi.org/10.33871/27639657.2023.3.1.7822>.
- FLORIDI, Luciano and COWLS, Josh, A. *Unified Framework of Five Principles for AI in Society* (September 20, 2019). DOI: <http://dx.doi.org/10.2139/ssrn.3831321>.
- HUMBOLDT, W. *Sobre a diversidade da construção das línguas humanas*. Tradução de João Costa. Lisboa: Edições 70, 1998.
- MEDEIROS, L. F. *Inteligência artificial: uma abordagem introdutória*. Curitiba: Editora Intersaberes, 2018.
- MBEMBE, Achille. *Brutalismo*. Tradução Sebastião Salgado. São Paulo: 1ª edição; N 1 Edições, 2021.
- NIETZSCHE, Friedrich. *Assim falou Zaratustra*. Tradução de Mário da Silva. São Paulo: Editora Escala, 2009.
- ORTEGA Y GASSET, José. *Meditação da técnica: vicissitudes das ciências, cacofonia na física*. Trad. e prólogo de Luis Washington Vita. RJ: Ibero Americano, 1963.
- RUSSELL, Stuart. *Inteligência artificial a nosso favor: Como manter o controle sobre a tecnologia*. Tradução Berilo Vargas – 1a ed. – São Paulo: Companhia das Letras, 2021.
- SANTOS, Antônio Bispo dos. *A terra dá, a terra quer*. São Paulo: Abu Editora/PISEAGRAMA, 2023.
- SILVA, Tarcísio. *Racismo algorítmico: inteligência artificial e discriminação nas redes digitais*. São Paulo: Edições Sesc São Paulo, 2022.

