

OS DILEMAS ÉTICOS DA INTELIGÊNCIA ARTIFICIAL: ONDE NOS APOIARMOS?

Daniel Rocha Carezzato (IC) e Prof. Dr. Roger Fernandes Campato (Orientador)

Apoio: PIBIC CNPq

RESUMO

O crescimento vertiginoso do uso de ferramentas que se utilizam da chamada “Inteligência Artificial” pelas empresas de tecnologia e sua utilização cada vez mais presente no cotidiano da sociedade desperta questões sobre o uso ético destas ferramentas, além do impacto que isso pode proporcionar ao modo de vida contemporâneo. Será debatido neste artigo de forma sucinta a definição de inteligência artificial, além de uma proposta de delimitação de escopo, ou seja, uma definição que nos serve para este estudo. Algumas questões aparecem quando pensamos nas consequências do uso (incorreto) da ferramenta. Essas, porém, ainda estão em caráter inicial na filosofia. Vamos nos perguntar sobre como trabalhar a responsabilização de ações tomadas por um ente robótico, como um ente que não tem sentimentos e não está em sociedade seguirá suas regras ou quais as consequências possíveis para o Direito e a Justiça. Esse artigo é uma tentativa de explorar caminhos e vertentes do estudo da ética, principalmente pautado pelo autor Jurgen Habermas e sua obra “A Inclusão do Outro”, pois esse faz um apanhado geral e uma crítica às correntes éticas ao mesmo tempo em que dá uma visão contemporânea sobre o tema, mesmo que não diretamente ligado à tecnologia, mas observando questões atuais da nossa sociedade.

Palavras-chave: Inteligência Artificial. Ética. Filosofia.

ABSTRACT

The vertiginous growth of the use of tools employing so-called "Artificial Intelligence" by technology companies, along with their increasingly present application in everyday society, raises questions about the ethical use of these tools and the impact this may have on contemporary ways of life. This article aims to succinctly discuss the definition of artificial intelligence and propose a delimitation of its scope a definition that will serve as a basis for this study. Several questions arise when considering the (incorrect) use of this technology; however, these are still in the early stages of philosophical inquiry. We will explore how to address accountability for actions taken by a robotic entity, how such an entity devoid of emotions and not embedded in society can follow social norms, and what potential consequences may arise for Law and Justice. This article attempts to explore pathways and perspectives within ethical studies, mainly drawing upon the work of Jürgen Habermas, who



simultaneously provides an overview and critique of ethical currents while offering a contemporary view on the subject. Although Habermas's work is not directly linked to technology, it still addresses relevant issues in our current society.

Keywords: Artificial Intelligence. Ethics. Philosophy.

1. INTRODUÇÃO

Esse estudo tem como objetivo observar a chamada “Inteligência artificial” (IA) - produto em grande crescimento em empresas de todos os setores, com influência alavancada pelas chamadas Big Techs (empresas de grande porte do setor de tecnologia) - sob um olhar da filosofia, mais precisamente sobre as implicações éticas que a envolvem.

Por ser um assunto relativamente novo para a sociedade (novo, leia-se, para o público em geral, que já se depara com algumas de suas consequências imediatas para a vida cotidiana) o principal desafio, de início, é determinar onde se apoiar para lidar com elas. Ou seja, por qual caminho a filosofia deve seguir para propor uma base teórica que fundamente perspectivas que sejam compatíveis com a natureza e as feições singulares deste fenômeno? O principal desafio: as implicações éticas estudadas até momento no campo filosófico se concentram sobretudo na análise de decisões e atitudes tomadas por humanos. Agora, nos deparamos com a entrada em cena de um ente robótico dotado com a capacidade de efetivar escolhas.

Para citar alguns exemplos, em novembro de 2022, um carro modelo Tesla dirigido por um sistema autônomo ficou desgovernado na província de Guangdong, na China. O veículo dirigiu sozinho em alta velocidade e acabou deixando dois mortos e três feridos (LEE, 2022). Não foi a primeira vez que isso aconteceu (na verdade, essa notícia viralizou nas redes sociais, pois há vídeos de câmeras de segurança que acompanharam parte do trajeto do automóvel). Segundo o órgão de segurança responsável pelo trânsito dos Estados Unidos da América (NHTSA), 392 acidentes envolvendo carros autônomos ocorreram entre julho de 2021 e maio de 2022 (ASSOCIATED PRESS, 2022), nos quais cinco pessoas morreram.

Esses e outros casos envolvendo a chamada inteligência artificial vêm surgindo de modo constante, cada um com implicações e consequências diferentes. Um proprietário de veículo com piloto automático foi flagrado alcoolizado dormindo ao volante enquanto seu carro dirigia a 100km/h em uma rodovia. Felizmente, nesse caso, não houve acidente e ninguém se feriu. Contudo, ficam os questionamentos: Se ele não estava conduzindo o veículo, ele estaria errado em, embora alcoolizado, permanecer no banco do condutor? Se não há vítimas, haveria crime? Mais uma vez, questões novas e complexas que precisam ser debatidas.

Houve casos em que o mecanismo de inteligência artificial utilizado para escolher currículos mais apropriados para uma vaga de emprego acabou por dar preferência a homens brancos em detrimento de mulheres e negros com a mesma - ou até melhor - qualificação. Logaritmos de imagens também já demonstraram que são capazes de aprofundar padrões de beleza predominantes na sociedade. Baseados em informações preconceituosas, excluem, por exemplo, negros do foco ou enfatizam nas mulheres as suas partes sensuais.

Uma ferramenta de chatbot (robô que interage/conversa com clientes de uma empresa nas redes sociais) da Microsoft, chamada Tay, foi programada para aprender a se comportar como uma adolescente. Em pouquíssimo tempo a máquina já fazia declarações de cunho nazista e discursos de ódio. Prontamente teve que ser desligada.

Ferramentas de reconhecimento facial mostraram também situações preocupantes, como confundir humanos negros com gorilas ou não reconhecer distinções físicas entre pessoas asiáticas diferentes.

Seja como for, estamos diante de um contexto em que há uma disparidade entre os padrões morais, que constituem o núcleo do Direito, e os desdobramentos éticos dos avanços proporcionados pela tecnologia, aos quais pertence a Inteligência Artificial.

Entre 2017 e 2024, o uso de IA em empresas ao redor do mundo subiu de 20% para 70% (RAMOS, 2024), um aumento vertiginoso que tende a se intensificar. Somente neste ano de 2024 a primeira legislação sobre a regulação de IA foi aprovada na Europa. No Brasil existe um projeto de lei de 2023 que ainda está com a votação pendente (URUPUA, 2024). Ou seja, mesmo com toda a utilização, ainda não há uma regulação satisfatória.

O estudo da ética constitui um campo vasto, quase tão antigo quanto a própria filosofia, mas todos esses desdobramentos foram pensados partindo-se de uma visão antropocêntrica, ou seja, sempre a partir de implicação de uma decisão tomada por um humano. Os efeitos e as implicações éticas quando passam por uma máquina é, ainda, um grande vazio para a filosofia.

De início, já nos encontramos com um sem-número de questionamentos e nebulosidades em definições sobre o tema. Podemos nos perguntar, antes de tudo, o que vem a ser inteligência artificial propriamente dita. Mas aí, para uma investigação

filosófica completa, precisaríamos delimitar antes de tudo o que é inteligência, decerto uma tarefa hercúlea.

Ao investigarmos o tema, já encontramos quem considere a definição de IA como diametralmente oposta ao que o se pretende aqui. A propósito, Rick Coldry, sociólogo da London School of Economics and Political Science, afirmou na Reunião Magna ABC, ocorrida em 2024, que a inteligência artificial nem é inteligente tampouco artificial. Não é inteligente pois tem uma série de limitações naturais do código que a compõe e não é, tampouco, artificial, pois foi desenhada por um ser humano, sendo apenas uma ferramenta probabilística.

Para preenchermos essas inúmeras lacunas de definição precisaríamos de muito tempo nos debruçando sobre isso. Se tem um campo de estudos que encontra ainda em um amplo vazio sobre o tema, não é a área de computação, que avança ferozmente sobre o tema, mas sim a própria filosofia:

Mas foi realmente sobre a filosofia que o impacto da IA foi maior: criar uma máquina pensante significa desafiar uma velha tradição que coloca o homem e sua capacidade racional como algo único e original do universo. Mais que isso, criar uma máquina pensante significa dizer que o pensamento pode ser recriado artificialmente sem que para isso precisemos de algo como uma 'alma' ou uma marca divina (Teixeira, 2019, p.3).

Veja como a compreensão do tema e todos os seus desdobramentos em sociedade é a maior lacuna do assunto, ou seja, o maior trabalho é da filosofia. Para este projeto, então, dedicaremos a estabelecer algumas delimitações e premissas necessárias para que algum caminho seja efetivamente traçado.

2. DESENVOLVIMENTO DO ARGUMENTO

É absolutamente comum usarmos uma calculadora para realizar alguns cálculos que poderiam ser realizados por nossa mente poderia realizar. Por que então não chamamos a calculadora de uma IA? Porque a IA se ampara no princípio de que a máquina imitaria ou tentaria chegar perto da maneira humana de pensar, o que não é o caso da calculadora. O processo de cálculo por ela efetivado não possui relação com o processo cerebral de cálculo humano. Teixeira (2019, p.2) se vale desta ideia para definir, em termos gerais, o conceito de inteligência artificial. Segundo o autor, a máquina nunca imitará completamente o cérebro humano, mas, assim como uma

boneca bem elaborada pode imitar a forma do humano andar, podemos ter um aparelho que imite a forma do humano pensar, pelo menos como modelo.

Ou seja, não se faz necessário que a IA seja uma réplica perfeita de um cérebro humano, mas que use os princípios da inteligência como um ser humano o faria. Todavia, ainda assim permanece a dúvida sobre o quanto um computador está imitando o ser humano ou sobre o quanto está somente se valendo de estratégias de imitação, como se fosse uma espécie de papagaio virtual.

Como um humano pensa? Podemos entender que o pensamento humano ocorre de forma a concatenar ideias e conceber diferentes proposições. Segundo Wittgenstein (1999), nossos pensamentos são formados por meio da linguagem. Ora, se nós pensamos com base em símbolos, o computador também estrutura seu pensamento com símbolos, claro que de forma mais primordial, pois todas as decodificações são realizadas em zeros e uns (0 e 1).

Os computadores como conhecemos hoje comportam principalmente os critérios criados por Alan Turing, que definiu que uma máquina poderia cumprir certas funções estabelecidas a partir de uma entrada de informações, com instruções semelhantes a “se X então Y, caso contrário Z”. Esse modelo é, sem dúvida, uma grande alavanca para os estudos de IA, mas ainda não pode ser considerado como tal, pois depende de uma instrução estipulada de uma maneira linear pelo pensamento.

Quando então podemos afirmar que uma máquina está pensando? O próprio Turing propôs um teste em que, a partir de interações com interlocutores que não sabem estar conversando com uma máquina, avalia-se se eles são enganados. Se o são, comprova-se então que a máquina é inteligente. Essa tese, porém, gera controvérsias. O filósofo John Searle elaborou o “argumento do quarto chinês”, que versa sobre uma pessoa que fica em um quarto fechado, tendo com ela um manual que mostra correspondências entre o mandarim e o português. Ela recebe por uma portinhola textos em chinês e devolve o texto traduzido para o português segundo as regras do manual. Ao longo do tempo ela fica mais e mais proficiente em fazer as traduções, mas em nenhum momento ele aprende a falar mandarim e, portanto, não entenderia um chinês. É justamente isso o que acontece com os computadores, que aprendem a pensar imitando a maneira humana, mas em nenhum momento pensam como um (Teixeira, 2019, p.31-33).

Há também a fábula do macaco que, apertando teclas a esmo em uma máquina de escrever, reproduziu um texto de Shakespeare. Um computador consegue jogar

símbolos a esmo e, obtendo feedbacks do que foi produzido, consegue calcular qual o conjunto de símbolos que atingem melhor o seu objetivo, aperfeiçoando cada vez mais sua performance, mas sem que ele entenda a linguagem ou pense como um ser humano propriamente dito.

Feita esta breve e necessária exposição do conceito de inteligência artificial, que por si só constitui um desafio filosófico, nos concentremos mais detidamente no propósito deste trabalho, que não reside propriamente em investigar ou propor novos entendimentos para o conceito. Esse trabalho visa entender as implicações éticas do uso da IA.

É importante ressaltar que o uso de IA como ferramenta pelos humanos, como, por exemplo, o uso do Chat-GPT para a realização de tarefas escolares, não altera fundamentalmente as implicações éticas debatidas há milênios. Isso porque a decisão última continua sendo a de um ser humano, como é possível constatar no caso, cada vez mais frequente, de alunos que tomam a decisão de usar a IA corrompendo a finalidade do seu estudo acadêmico. Trata-se de uma decisão ética semelhante à dos alunos que outrora copiavam seus trabalhos de enciclopédias impressas.

O que precisamos observar é quando uma decisão, sem supervisão humana, é tomada e executada por uma máquina, acarretando resultados que interferem nitidamente nas vidas humanas. Por exemplo, um robô que dirige um carro sozinho e tem que decidir entre a vida do dono do carro ou de um pedestre, um sistema de IA que decide se uma pessoa é suspeita de um crime, um robô que analisa e decide melhores candidatos para uma vaga de emprego e assim por diante.

Ou seja, para esse projeto, adotaremos a seguinte definição de inteligência artificial: decisões tomadas e ações executadas diretamente por máquinas, sem supervisão humana, cujos resultados influenciem diretamente a vida de uma ou mais pessoas.

Uma confusão que geralmente é creditada por alguns à tecnologia é que ela, por ser determinada por um não humano, estaria isenta de cometer injustiças. Apesar de um código ser desenhado e de a máquina não fazer nada além do que lhe foi antecipadamente definido, isto não significa que essa máquina não possa aprender com humanos algo antiético e que não replique e aprofunde os problemas da sociedade.

É mister que se faça o inverso, que as desconfianças em relação a uma máquina sejam sempre ouvidas e levadas a sério. Bostrom e Yudkowskynick (2012,

p. 201) sustentam que “será cada vez mais importante desenvolver algoritmos de IA que não sejam apenas poderosos e escaláveis, mas também transparentes para inspeção.” A transparência será peça fundamental para o futuro da IA aplicada à vida cotidiana. Saber as razões pelas quais o algoritmo do banco recusou uma determinada solicitação de empréstimo pode vir a ser um direito do solicitante, já que ainda não existem muitas regulamentações neste sentido.

No entanto, ainda de acordo com Bostrom e Yudkowskynick (Idem, p. 202), não só a transparência se faz necessária, mas também a previsibilidade. Para eles, é importante que os objetivos traçados pelos criadores de um algoritmo sejam mapeados. Há, porém, controvérsias sobre essa questão, pois, se há previsibilidade, não é necessária a inteligência, uma vez que a ação poderia ser executada por uma máquina simples, como por uma calculadora. Claro que há um horizonte a se esperar a partir de um desenho de IA. Uma máquina projetada para aprender a jogar um videogame provavelmente jamais será capaz de cozinhar. Mas, uma vez que a dimensão da autonomia seja especificada, qualquer comportamento dentro dessa dimensão pode ser esperado.

Há também a possibilidade de associação entre algoritmos. Nossa máquina jogadora de videogame pode não fazer panquecas, mas uma máquina destinada a ganhar dinheiro na internet pode usá-la em campeonatos fingindo ser humano.

Mas Bostrom e Yudkowskynick (2012, p. 202) expõem a grande preocupação em se determinar os princípios éticos de uma IA, entre eles, a responsabilização:

Outro importante critério social para as transações em organizações é ser capaz de encontrar a pessoa responsável por conseguir que algo seja feito. Quando um sistema de IA falha em suas tarefas designadas, quem leva a culpa? Os programadores? Os usuários finais?

O maior problema com essas questões é justamente a incapacidade ou a falta de vontade para respondê-las. Para o mundo corporativo, melhor seria que não se achasse o responsável ou que fosse considerado como uma mera tragédia quaisquer ações mal executadas por máquinas autônomas.

A responsabilização não é imprescindível somente para possíveis maus resultados, mas também para supostos bons resultados. Imagine que uma IA seja responsável por desenhar o melhor motor de avião. Após muitos cálculos, o resultado apresentado é algo que nenhum engenheiro consegue explicar. Poderíamos dizer

“não sabemos como isso funciona, mas vamos todos voar nesse avião, pois seu motor é o mais moderno do mundo”? Provavelmente ficaríamos com certo receio. Imagine então se isso já fosse feito, se você já tivesse voado em um avião com motor que nenhum ser humano entende como funciona. Seria ético, minimamente, ser comunicado sobre isso, se não fosse efetivamente proibido.

A fim de nos apoiarmos em uma perspectiva ética que esteja mais próxima do problema com o qual estamos lidando, recorreremos à obra do filósofo alemão Jürgen Habermas, mais especificamente ao capítulo “Uma visão genealógica do teor cognitivo da moral”, que integra o livro intitulado “A inclusão do outro”, que veio a público em 1996. Nesse capítulo, o autor se debruça sobre as óticas da moralidade desenvolvidas ao longo do tempo e apresenta uma visão interessante que abarca o presente, o passado e o futuro da ética, o que consideramos ser um bom suporte para a proposta desse estudo, pois ao mesmo tempo consiste em uma competente síntese das correntes éticas existentes, uma crítica construtiva a elas e um olhar contemporâneo sobre a moral na sociedade. Habermas se propõe a promover uma visão sobre a moral que busque identificar nela um “teor cognitivo”, com vistas à sua fundamentação, estudando as motivações que levam os indivíduos a justificar racionalmente suas manifestações morais.

Já de início, o autor explana que a moral consiste em exigir do outro um comportamento, ao mesmo tempo em que propomos uma forma de agir para nós mesmos. Pressupomos então um conhecimento mútuo de normas e práticas habituais de obrigações e expectativas do que exigir dos outros. O que parece bastante razoável para um inter-relacionamento em sociedade começa a ficar descolado quando imaginamos um mundo de relacionamentos entre seres humanos e a IA, pois não há como exigir de um código que ele pressuponha práticas habituais de um modo de agir em sociedade. Isso deveria vir dos seus programadores, mas de forma indireta. Cabe lembrar que estamos aqui considerando decisões tomadas sem revisão humana, ou seja, uma ação tomada pós-código, que tem um comportamento esperado, porém imprevisível:

A moral não diz apenas como indivíduos devem se comportar(...) fazem parte do jogo da linguagem da moral as discussões as quais (...) podem ser resolvidas convincentemente com ajuda de um potencial de fundamentações igualmente acessível a todos (Habermas, 2002, p.12).

Nesse trecho o autor nos lembra que as regras morais nem sempre são exatamente claras e diretas. Por muitas vezes temos que nos fazer valer de uma explicação para que os envolvidos possam ou não dar razão a quem realiza o ato, casos como “roubei para alimentar meu filho” ou “matei em legítima defesa”. Novamente nos colocamos em xeque, pois como uma IA poderia ter a capacidade de justificar uma tomada de decisão, ainda mais baseada em regras da convivência humana.

Decerto, para isso, as programações deveriam incluir uma série de códigos e dicionários contendo proposições éticas para que o robô esteja apto a escolher a alternativa mais correta e a ser capaz de justificá-la, além de convencer interlocutores de que sua ação tem lastro nos costumes. “A autoconsciência do sujeito que julga moralmente cai em revisão. Em seus posicionamentos e julgamentos deveriam exprimir-se de fato apenas motivos racionais, sejam sentimentos e situações de interesse” (Idem, p.14). Podemos sem dúvida exigir de uma programação um comportamento racional, talvez muito mais até do que de um ser humano, mas que essa racionalidade se baseie em sentimentos e costumes, com certeza, não.

Se pudéssemos estabelecer regras prévias em todos os casos, não dependeríamos da ética em nossa sociedade. Como vimos, a ética reside também em ações que não podemos prever, pois, se pudéssemos, bastaria cumprir as leis estabelecidas e, como sabemos, cumprir a lei não é necessariamente sinônimo de comportamento moral. Podemos então estabelecer normas de utilização de IA, mas isso garantiria um uso moral da ferramenta?

Para o cumprimento das regras morais em sociedade, nos apoiamos também nos nossos sentimentos. Por que agimos de forma ética? De acordo com Habermas (2002, p.17), para evitar o surgimento de sentimentos que se manifestem, “do ponto de vista de terceiros, como repulsa, indignação e desprezo; do ponto de vista do atingido diante do seu próximo, como humilhação ou ressentimento; do ponto de vista da primeira pessoa, vergonha e culpa”. Esse suporte para as decisões éticas referentes a nossas ações certamente seria o maior abismo ao se considerar a presença de robôs agindo em sociedade.

Uma das formas de direcionar o comportamento do ser humano, utilizadas principalmente pelas religiões, é a ideia de um modelo moral perfeito e a expectativa de imitação pelo indivíduo:

Interpretações religioso-metafísicas do mundo, o justo está entre tecido com certos conceitos do bem viver. O modo como



devemos nos comportar nos relacionamentos interpessoais resulta de um modelo de conduta exemplar. Aliás, o ponto de referência de um deus que aparece in persona, que no dia do juízo final julgará cada um dos destinos individuais (Idem, p.18).

Por exemplo, na religião cristã, há Jesus como um modelo moral a ser seguido e o indivíduo ou sociedade tenta – ou deveria tentar - viver conforme esses preceitos. Soa à primeira vista jocoso imaginarmos um robô messias, algo como um código criado para ser tão perfeito moralmente que as outras IAs deveriam se basear nele para tomar suas decisões. Podemos inclusive imaginar então o desdobramento desta intenção em diferentes religiões cibernéticas que criam sua própria versão do messias, crendo que a sua é a mais correta e que a outra, na verdade, remete a uma versão de potencial mais maligno - criada por vírus e malwares - da expressão “lobo em pele de cordeiro”. Guerras cibernéticas seriam travadas para domínio de território em nuvem em nome do suposto verdadeiro deus.net.

Mas essa linha de raciocínio, apesar da extrema distopia que carrega consigo, não pode ser por completo descartada. Brochado (2022, p.87) faz menção a proposição de uma chamada Ética computacional que, evocando o resgate do platonismo e aristotelismo, objetiva a criação de agentes morais artificiais, perfeitos o bastante para serem parâmetro ético inclusive para os humanos:

Esses sistemas morais artificiais teriam por tarefa suprema ensinar boas práticas aos humanos, vale dizer, seriam a expressão máxima de uma tecnopaideia que reabilita, por meio de máquinas, a ética clássica da virtude-ciência.

Por serem presumivelmente incorruptíveis por princípio -o que é questionável pela existência de vírus, defeitos de programação ou até mesmo más intenções de programadores -, se crê que as IAs possam forjar as regras morais que norteiem a sociedade.

Valendo-se das reflexões de Paulo Antônio Silveira, Brochado afirma que diversos cientistas de programação sustentam que um modelo de agência artificial baseado em machine learning, desde que combinado com a ética e com a virtude, delineia um caminho natural, coeso, coerente, integrado e bem costurado. Como veremos a seguir, porém, essa afirmação encontrará na filosofia alguns percalços.

O próprio Habermas (2002, p.20) argumenta que



A filosofia moral depende de um nível de fundamentação pós-metafísico. Isso quer dizer que lhe são negados, do lado do método, o ponto de vista divino, do lado do conteúdo, o recurso à ordem da criação e a história da salvação, e do lado da estratégia teórica, a remissão aos conceitos essenciais que perpassam a diferenciação lógica entre diversos tipos elocucionais de proposições

Ou seja, para o autor, cabe à filosofia entender os princípios que norteiam, em cerne, a moral desejada, sem apetrechos, para que as decisões possam ser tomadas e justificadas de maneira lógica, racional e ordenada. Esse, porém, é um assunto não resolvido em absoluto em milênios de debate e, em meio a isso, temos na tecnologia novas manifestações que se beneficiariam muito desta ausência de resolução. O realismo moral, o utilitarismo, o ceticismo e o funcionalismo são diversas vertentes do estudo da ética que seriam caminhos a serem percorridos, sem qualquer segurança de serventia.

Com efeito, ficamos muitas vezes sem saída. A princípio, correntes como o empirismo e o contratualismo, que nos propõem, entre outros, Kant ou Hume, nas quais se esperaria encontrar fundamentações mais racionais e palpáveis, esbarram em dificuldades antes mesmo da elaboração tecnológica e que se aprofundam ainda mais. Segundo Habermas (1996, p.26), “as duas teorias defrontam-se, no fim, com a mesma dificuldade: elas não podem explicar apenas com motivos racionais a obrigatoriedade dos deveres morais, que remetem para além da força da inteligência. (...) Uma congruência no julgamento moral de um caráter significa, portanto, uma convergência de sentimentos.”

Ora, se não nos deparamos com as fragilidades dos modelos justamente naquilo que nos falta nas configurações da IA, que tem racionalidade em excesso e humanidade em falta, justamente o nó que produz o impasse continuará à espera de ser desfeito.

O contratualismo também já enfrentava “poréns” à medida que se observavam alterações na sociedade. Em pequenas tribos ou vizinhanças, consegue-se “observar” o contrato social em prática, onde as ações são norteadas pela cultura, valores e vontades dos envolvidos. Já em um contexto cosmopolita e com a globalização, acompanhada pela diminuição de distâncias, é difícil a assimilação das vontades dos indivíduos, situação que intensifica quando levamos em consideração as diferenças sociais. Pois a tecnologia aprofundou isso ainda mais. O melhor exemplo disso são os “termos e condições” com os que concordar – muitas vezes à revelia de nossa vontade

-no momento da utilização de um software ou de um serviço. Alguém que lê os termos e não concorda com eles certamente estaria excluído da sociedade, sem poder ter acesso às ferramentas do dia a dia que se tornaram comuns.

Da mesma forma em que não há um caminho que leve diretamente dos sentimentos morais de simpatia ou de rejeição para fundamentação das obrigações segundo uma racionalidade dos fins, também não há um caminho que leve diretamente de volta os sentimentos de reprovação internalizada a partir de fundamentação contratualista de uma ordem normativa. Sentimentos morais exprimem posicionamentos, os quais implicam juízos morais. E, no caso de um conflito, nós não discutimos a respeito da validade dos juízos morais apenas com motivos pragmáticos ou preferenciais. O empirismo clássico não dá conta desse fenômeno, porque exclui motivos epistêmicos. Em última instância, ele não pode explicar a força vinculatória das normas morais a partir de preferências (Habermas, 2002, p.26)

Mas veja como a IA de forma isolada não pode, então, ser vista como a vilã do dilema ético contemporâneo, já que nossa sociedade já vinha carecendo de elementos que poderiam coordenar o debate ético de forma ordenada. Podemos citar: o fato das pessoas não se relacionarem mais fisicamente, pelo contrário, hoje se pratica o evitar o contato social; a participação do indivíduo comum é cada vez mais irrelevante para o todo da sociedade; o aumento da concentração de renda contínuo em uma sociedade capitalista (leia-se o acúmulo cada vez maior de poder para poucos). Tudo isso já se posicionava como um impeditivo para encontrar teorias éticas que fossem aderentes e isentas de falhas, pois dependem de fatores muitas vezes não racionais em uma sociedade cada vez mais racional. Poderia ser a IA não uma virada de chave, mas sim um aprofundamento dos dilemas sociais.

Um paradoxo de nosso tempo, segundo Brochado, é o fato de um crescente sem precedentes em produção de bens e avanço tecnológico ao mesmo tempo em que estamos na época em que mais crescem os estudos sobre ética na filosofia (consequência das necessidades da sociedade), o que ela chama de “crise de espiritualidade” e “risco a sobrevivência e as razões de viver”. Pode-se questionar a nomenclatura, mas não a preocupação. Há uma empolgação e um direcionamento da sociedade em avançar em questões técnicas como se isso bastasse à vida humana, mas está cada vez mais vazio o pensar. Existem falsos novos dilemas como, por exemplo, IAs sendo redatores de cinema, como se fosse algo novo, mas na verdade



Scripts feitos por IA são somente cópias potencializadas por um sistema de pesquisa e aglutinação extremamente avançado, o que sempre foi errado do ponto de vista moral.

Para avaliar a participação ética do indivíduo na sociedade, Habermas (2002, p.32) recorre a Ernst Tugendhat, que, assumindo um ponto de vista funcionalista, pressupõe que o participante tenha capacidade de discernir aquilo que é bom ou mau, gozando de liberdade para sua tomada de decisão. No caso de um ente digital, tal capacidade não se faz presente. Sob essa perspectiva, o dever ético fica a cargo do quão importante a moral é para cada um, anulando a perspectiva do dever comum. Isso remete justamente ao que não queremos: deixar às máquinas a liberdade de não agir moralmente.

No papel de legislador, cada pessoa participa de sua empreitada cooperativa e aceita, com isso, uma perspectiva intersubjetivamente ampliada, a partir da qual se pode examinar se uma norma que é objeto de discussão pode ser considerada generalizável segundo ponto de vista de todos os participantes (Habermas, 2002, p.33).

Esse problema já aparecia em Kant (1995, p.47) nos seguintes termos:

Só um ser racional tem a capacidade de agir segundo a representação de leis, isto é, segundo princípios ou uma vontade. Como para a dedução das ações a partir das leis é necessária a razão, a vontade não é outra coisa que não a razão prática.

Então, visto que, no tocante à Inteligência Artificial, a participação da máquina (que, carente de responsabilidades e de vontade autônoma, nos encontramos sob um dilema com o qual Habermas (2002, p.54), movido por preocupações imediatas distintas também se deparou:

Teorias alinhadas à tradição de Hegel, Humboldt e Mead percorreram vestígios como esses e demonstraram que condutas comunicativas estão entrelaçadas com suposições recíprocas de reconhecimento e que elas representam, portanto, um conteúdo normativo. Depreende-se dessas análises que a moral extrai, da forma e estrutura perspectica da socialização intersubjetiva intocada, um sentido genuíno e dependente do que é individualmente bom. (...) Certamente não são suficientes para fundamentar o porquê da transgressão de uma orientação particular

Compreende-se que, na impossibilidade de se considerar os agentes participativos por si só, pode-se esperar que questões novas apareçam e necessitem de resolução. Seria ideal então a existência de um conselho (práxis) para que os assuntos sejam debatidos e as resoluções sejam tomadas e doravante obedecidas. Torna-se universal aquele assunto específico que foi debatido e cujas diretrizes sejam estabelecidas por todos que envolvem um conselho. Na prática, isso seria algo semelhante à ANPD (Autoridade nacional de Proteção de Dados Pessoais) responsável pela fiscalização da LGPD (Lei de Proteção de Dados Pessoais). Mas seria um órgão regulador o suficiente para alívio ético da participação de IA na sociedade?

Podemos imaginar que há a possibilidade de um grande aprimoramento no campo do direito, visto que esse trabalha com leis delimitadas, que os agentes da sociedade têm obrigação de seguir. Há um bom caminho para que haja leis e diretrizes para serem utilizadas em códigos e execuções de máquina, que podem ser verificáveis por desenho (anteriormente ao fato) ou por consequência, pois pode se aprimorar a fiscalização por meio de tecnologia, como por câmeras inteligentes, identificadores de tendência, entre outros.

Mesmo assim há um desafio: “pode ser desestabilizador para a simbologia social em torno do papel da relevância do direito o excesso de casos decididos por máquinas, pondo em xeque a própria necessidade do Poder Judiciário” (Brochado, 2002, p.90). A margem existente entre onde a tecnologia pode ajudar na fiscalização e aplicação de leis e a subjetividade natural de decisões jurídicas pode não ser tão clara.

Ela menciona também a incapacidade atual de responsabilização de máquinas como entes do direito. Não há como saber, por exemplo, quem é o responsável por uma decisão tomada por um carro autônomo. Há a possibilidade de se considerar a IA como uma pessoa jurídica, ou seja, no âmbito civil como se fosse uma empresa, mas também há a possibilidade de se responsabilizar o portador do robô (no exemplo, o dono do carro). Em ambos os casos existem prós e contras, o que indicaria a necessidade de uma legislação específica.



3. CONSIDERAÇÕES FINAIS

Gostaria de poder efetivamente chamar a essa última sessão de “conclusão”. Mas chamá-la assim significaria certa empáfia em reivindicar alguma afirmativa mais definitiva sobre um tema tão nebuloso e tão vasto para novos estudos. Talvez a conclusão seja mesmo essa, a de que precisamos, no âmbito da filosofia, nos aprofundar mais sobre esse assunto, caso tenhamos a pretensão de oferecer uma contribuição mais relevante para o debate em curso.

Existe a preocupação de que a velocidade da aplicação de novas tecnologias está muito à frente (e ainda crescendo vertiginosamente) do que o entendimento sobre o impacto social que estas irão promover. Em tempos em que a importância da filosofia é tão desprezada pela sociedade, tanto como matéria formal acadêmica quanto no seu possível uso cotidiano, concluímos que é ela justamente o que nos falta.

Existe também a possibilidade de estarmos passando para uma era pós-ética, na qual os costumes humanos foram tão padronizados por códigos e máquinas que não há mais espaço para a moral. Talvez viveremos em uma sociedade onde somente se espera do ser humano cumprir a lei, sem que haja campo para decisões que saiam do combinado legislativo e que, pela falta de contato humano, tudo é permitido dentro da lei. Com a dificuldade em responsabilizar as máquinas, talvez viveremos em uma sociedade onde problemas serão sempre tragédias ou em que a vida de alguém seja paga com uma multa em um processo civil.

REFERÊNCIAS

ASSOCIATED PRESS. Carros semiautônomos se envolveram em 392 acidentes em 1 ano nos EUA. G1, 2022. Disponível em: <https://g1.globo.com/inovacao/noticia/2022/06/15/carros-semiautonomos-se-envolveram-em-392-acidentes-em-1-ano-nos-eua.ghtml>. Acesso em: 14 dez. 2022.

BOSTROM Nick; Yudkowskynick, Eliezer. A ética da inteligência artificial. FUNDAMENTO – Rev. de Pesquisa em Filosofia, v. 1, n. 3 – maio – ago. 2012.

HABERMAS, Jurgen. A inclusão do outro: estudos de teoria política. Sao Paulo: dições Loyola. 1996.

KANT, Immanuel. Fundamentação da metafísica dos costumes. Trad. de Paulo Quintela. Lisboa: Edições 70, 1995.

KAUFMAN, D. Inteligência Artificial e os desafios éticos: a restrita aplicabilidade dos princípios gerais para nortear o ecossistema de IA. PAULUS: Revista De Comunicação Da FAPCOM, 5(9). Jun. 2021.

LEE, Liz; YANG, Zhang; GAO, Liangping; ZHUZHU, Cui. Acidente com Tesla deixa dois mortos na China; empresa diz que ajudará polícia. CNN Brasil, 2022. Disponível em: <https://www.cnnbrasil.com.br/internacional/acidente-com-tesla-deixa-dois-mortos-na-china-empresa-diz-que-ajudara-policia/>. Acesso em: 14 dez. 2022.

NETTO, E. (2021) Uma Visão sobre Ética e IA. EMAPS-Resenhas #03. Rio de Janeiro, RJ - Brasil: SERG, Departamento de Informática, PUC-Rio, 2021.

REUTERS, Reuters. Recorde. Infomoney, 2023. Disponível em: <https://www.infomoney.com.br/negocios/chatgpt-atinge-100-milhoes-de-usuarios-ativos-mensais-em-janeiro-e-vira-o-app-de-crescimento-mais-rapido-da-historia/>. Acesso em: 08 fev. 2023.

RAMOS, Marien, Uso da inteligência Artificial aumenta e alcança 72% das empresas, CNN, 2024. Disponível em: <https://www.cnnbrasil.com.br/economia/negocios/uso-de-inteligencia-artificial-aumenta-e-alcanca-72-das-empresas-diz-pesquisa/#:~:text=O%20interesse%20no%20uso%20da,comparado%20aos%2055%25%20em%202023.>

REUNIAO MAGNA ABC, Nick Coldry: "IA não é inteligência, nem é artificial, Academia Brasileira de Ciências, 2024, Disponível em: <https://www.abc.org.br/2024/05/11/nick-couldry-ia-nao-e-inteligencia-nem-e-artificial/#:~:text=Couldry%20descreve%20a%20intelig%C3%Aancia%20artificial,probabi%C3%ADstica%2C%20n%C3%A3o%20%C3%A9%20criativa%E2%80%9D.>

SILVEIRA, Aulo Antônio Caliendo Velloso da. Ética e Inteligência Artificial: Da possibilidade filosófica de Agentes Morais Artificiais. [S. l.: s. n.], 2021.

TEIXEIRA, João de Fernandes. O que é inteligência artificial. São Paulo: Brasilense, 2019



URUPUÁ, Marcos. Publicada lei que harmoniza as regras para IA na Europa, Teletime, 2024. Disponível Em: <https://teletime.com.br/12/07/2024/publicada-lei-que-harmoniza-as-regras-para-ia-na-europa/#:~:text=Na%20pr%C3%A1tica%2C%20o%20novo%20regulamento,anteriores%20publicados%20pelo%20parlamento%20europeu.>

WITTGENSTEIN, Ludwig. Tractatus Logico-Philosophicus. Tradução de Hilda B. F. de Carvalho. São Paulo: Martins Fontes, 1999.).

Contatos:

dcarezato@gmail.com

roger.campato@mackenzie.br