



PARALELOS ENTRE AS INTELIGÊNCIAS NATURAL E ARTIFICIAL A PARTIR DE UM COMPARATIVO EPISTEMOLÓGICO:

limitações e impactos

DOI: <https://doi.org/10.35699/2965-6931.2023.47680>

LUIZ JOAQUIM NUNES*; MARIANA INÉS GARBARINO**

RESUMO: O estudo aborda a relação entre inteligência natural e artificial, destacando as limitações da IA atual em comparação com a humana. O trabalho realiza um comparativo epistemológico tomando como referencial a epistemologia genética piagetiana e o ethos humano. Analisam-se alguns aspectos da experiência humana que se acentuam como tais limitações: a afetividade, a moralidade, a criatividade e a ética. A afetividade humana desempenha um papel fundamental sobre as conclusões às quais a cognição humana chega. A IA carece da capacidade de desenvolver uma noção cognitiva de si mesma e, portanto, de moralidade, e a forma como é concebida não permite que novas estruturas cognitivas sejam produzidas para viabilizar uma criatividade completa. O entendimento sobre a ética humana contribui também para o entendimento dos impactos da IA. A análise, por fim, indica cenários para a aplicação de inteligência artificial, de acordo com suas limitações em cada caso.:

PALAVRAS-CHAVE: inteligência artificial; psicologia social; psicologia cognitiva; ética da tecnologia; futuro do trabalho.

Parallels between natural and artificial intelligence from an epistemological comparison: Limitations and impacts

ABSTRACT: The study addresses the relationship between natural and artificial intelligence, highlighting the limitations of current AI compared to human intelligence. The work involves an epistemological comparison grounded on Piagetian genetic epistemology and the human ethos. Some aspects of the human experience that are accentuated as such limitations are analyzed: affectivity, morality, creativity and ethics. Human affectivity plays a fundamental role in the conclusions reached by human cognition. AI lacks the capacity to develop a cognitive notion of itself and, therefore, of morality, and the way it is conceived does not allow new cognitive structures to be produced to enable complete creativity. Understanding human ethics also contributes to understanding the impacts of AI. The analysis, finally, indicates scenarios for the application of artificial intelligence, according to its limitations in each case.

KEYWORDS: artificial Intelligence; social psychology; cognitive psychology; technology ethics; future of work.

* Universidade de São Paulo / ** Universidade Federal do ABC (UFABC)

Introdução

A ideia de inteligência artificial (IA) não é um fenômeno recente, mas sim algo que permeia a história humana ao longo de várias culturas (Nunes, 2023). Desde antigas representações míticas de seres criados a partir do barro até algumas visões religiosas sobre a criação da vida, encontramos indícios de construções artificiais com atributos inteligentes. Essas narrativas, prévias à era dos computadores, sugerem a fascinação ancestral pelo conceito de inteligência semelhante à humana.

No entanto, à medida que a IA emerge como uma tecnologia proeminente na sociedade contemporânea, o próprio termo “inteligência artificial” pode ser aplicado de forma imprecisa, gerando ambiguidade em sua definição e levantando questões relevantes para campos como a comunicação científica e a psicologia social. Portanto, é válido explorar as raízes psicológicas que historicamente inspiraram o desenvolvimento da IA e refletir sobre a relação entre aprendizado de máquina e aprendizado humano. Esta investigação não apenas esclarece conceitos fundamentais, mas também destaca a importância contemporânea de delinear fronteiras conceituais nesse campo em constante evolução.

Nos seus estudos sobre cognição e inteligência, Jean Piaget (1960, 2005) explorou profundamente as facetas do desenvolvimento humano desde a infância, e sua obra permite uma compreensão mais clara sobre o quanto tais facetas são ou não identificáveis em inteligências artificiais também. Mais ainda, sua obra permite uma aproximação mais robusta da psicologia para a discussão de tais tecnologias que, embora não se tratem de inteligências numa perspectiva biológica, foram produzidas como a versão artificial daquilo que referenciam. Essa mesma referência pode ser revisitada e tomada como fundamento para discussão, após décadas de desenvolvimento tecnológico, sobre o quão distantes essa transversalidade as tornou.

Piaget define a inteligência como um estado de equilíbrio resultante das interações entre adaptações sensorio-motoras e cognitivas, bem como acomodações e assimilações entre o organismo e o ambiente (Macedo, 1980). Ele descreve três formas de inteligência: ritmos, regulações e operações, que evoluem desde padrões básicos até processos complexos de raciocínio.

Essa estrutura de inteligência definida por Piaget encontra paralelos na definição da inteligência artificial apresentada por Stuart J. Russell e Peter Norvig. Eles propõem quatro abordagens para a IA: sistemas que agem como humanos, sistemas que pensam como humanos, sistemas que pensam racionalmente e sistemas que agem racionalmente (Russell & Norvig, 1995).

Um paralelo epistemológico pode ser estabelecido da seguinte forma: os ritmos correspondendo aos algoritmos principais da IA, as regulações correspondendo ao aprendizado de máquina baseado em feedback e otimização, e as operações correspondendo a aplicações específicas para resolver problemas, como inferência, ordenação e escrita.

Nesse paralelo as definições de inteligência de Piaget e as abordagens de IA propostas por Russell e Norvig apresentam uma congruência que permite mapear mais robustamente suas semelhanças e diferenças - a estrutura das três formas de inteligência se alinha de maneira consistente com as diferentes abordagens da IA. E então, tendo em vista onde essas semelhanças e diferenças possam ter maior relevância sobre a segurança de uma aplicação de IA, pode se fazer mais viável discutir impactos esperados dessas aplicações. Mais ainda, esse comparativo pode viabilizar que a psicologia se envolva com maior rigidez epistemológica em análises e debates sobre inteligência artificial, sua aplicação e regulação.

Sobre o papel da afetividade

Em seminários proferidos na Sorbonne (1953-54), Piaget enfatizou as conclusões de sua investigação sobre as relações de interdependência entre a cognição e a inteligência, mostrando o papel desempenhado pela afetividade em sua evolução e manifestação. A afetividade é a mola propulsora ou inibidora da ação, na forma de interesse ou evitamento (Piaget, 2005). De acordo com ele, a afetividade pode ser identificada em determinados cenários, com diversas consequências potenciais, podendo influenciar a execução de uma atividade tornando-a mais ágil, mais desafiadora ou mesmo mais intrincada.

A afetividade humana possibilita à pessoa reconhecer emoções em outros com base nas sensações que essas emoções evocam nela mesma, como é o caso da capacidade de identificar a dor que alguém possa estar experimentando (o que comumente é conhecido como “empatia” e que é atualmente estudado pela Teoria da Mente). Quando alguém se depara com algo detectado pelos sentidos, a compreensão não se limita apenas a um enfoque racional e descritivo; também abrange uma dimensão afetiva, ou seja, a forma como esse algo influencia emocionalmente a pessoa. Além disso, a afetividade não desempenha somente um papel transitório durante a interpretação desse algo. Mesmo que a racionalidade seja capaz de descrever e interpretar logicamente o que é percebido pelos sentidos, as sensações emocionais permanecem integradas à percepção global do indivíduo em relação àquele algo. Portanto, para Piaget (2005), a dialética afeto-cognição impregna cada ação, representação e produção humana.

Ao supostamente desenvolvermos uma inteligência artificial com o propósito de reconhecer emoções em indivíduos (Russell & Norvig, 1995), geralmente seguimos um processo específico: a máquina recebe acesso a um conjunto de imagens de pessoas, onde cada imagem já possui uma emoção atribuída. Ao receber uma nova imagem, a máquina busca avaliar o grau de semelhança entre essa nova imagem e as imagens existentes na base de dados, resultando na atribuição da mesma emoção associada às imagens consideradas mais parecidas. A definição de “semelhança” entre imagens pode ser estabelecida de diversas maneiras, levando em consideração desde as paletas de cores utilizadas na imagem até as estruturas geométricas identificadas nas características faciais das pessoas. Essas lógicas para mensurar a “semelhança” são pré-configuradas como parte do sistema de inteligência artificial. Portanto, é possível que uma máquina seja figurativamente capaz de discernir emoções em pessoas por meio de suas imagens, mas essa habilidade é estritamente resultado de cálculos matemáticos aplicados às imagens, sem envolver elementos próximos à afetividade humana ou sequer uma compreensão real sobre emoções e contextos afetivos.

Essas considerações sobre a dialética inteligência-afetividade podem, assim, indicar uma ampla gama de processos sutis e complexos que as inteligências artificiais, nos moldes atuais, não seriam capazes de executar. Essas distintas situações são de grande relevância para identificar os papéis que a IA poderia não ser capaz de desempenhar plenamente na sociedade humana.

Sobre a moralidade

De acordo com Piaget, ao longo da infância se desenvolve o paulatino processo de construção cognitiva social e afetiva do mundo, um mundo de objetos, normas e relações socio-afetivas com pessoas próximas do contexto familiar e das instituições que a criança frequenta (Piaget & Inhelder, 1974). Durante esse processo, as noções de objeto, espaço, tempo e causalidade são progressivamente desenvolvidas. Piaget observa que quando a criança demonstra a construção da noção de objeto, isso também sinaliza a formação da noção de si mesma: nos primeiros estágios da vida, a criança não revela - pelo menos não de maneira clara - uma compreensão de que os objetos continuam a existir independentemente de sua percepção individual: é o que na teoria piagetiana foi nomeado como noção de “objeto permanente”. No entanto, em algum ponto posterior, essa compreensão se torna evidente. É nesse ponto que a noção de si mesmo se torna perceptível do ponto de vista cognitivo, à medida que a criança se posiciona no universo como um objeto entre outros objetos, processo também conhecido como “descentração”, em detrimento do egocentrismo que tornava a criança o centro das referências para entender o mundo.

Essa compreensão também capacita o indivíduo a reconhecer outros indivíduos de maneira semelhante a si próprio, o que progressivamente possibilita a construção de sua própria identidade a partir de perspectivas culturais e morais diversas e que a transcendem. A pesquisa de Piaget sobre a moralidade, que foi posteriormente continuada por Kohlberg (Duska & Whelan, 1994), Turiel (Turiel, 1983) e, no Brasil, por Yves De La Taille (La Taille, 2002), inicialmente define a moral como um sistema de regras que emergem por meio da interação social. No entanto, uma leitura reducionista dessa abordagem desembocou em críticas por ser excessivamente racional. A partir desse ponto, surgem estudos mais focados no aspecto afetivo da moralidade, e o próprio Piaget (2005), especialmente ao longo de suas obras da maturidade, demonstra atribuir uma grande importância à dimensão afetiva na compreensão da moral.

Baseando-se no si mesmo cognitivo e na afetividade, um indivíduo teria a capacidade de alinhar suas necessidades, interesses e perspectivas com outros indivíduos,

resultando na construção de uma noção de si mesmo moral. O si mesmo moral é constituído conforme uma escala de valores e ideais construídos com base nas próprias experiências subjetivas e no exercício de uma moral autônoma permeada por contextos socioculturais que transmitem e perpetuam determinados valores morais. Em outras palavras, tanto a compreensão cognitiva de si mesmo e do contexto social quanto a afetividade (o “querer ser moral”) desempenham um papel crucial na formação do si mesmo moral, e, conseqüentemente, na constituição da moral em si.

Conforme discutido na seção sobre a afetividade, as máquinas em sua forma atual são incapazes de processar afetividade humana, uma vez que isso está fora do âmbito de suas naturezas essencialmente racionais. Além disso, as máquinas apresentam uma distinção cognitiva notável em relação aos seres vivos: o software e o hardware, duas partes que operam de maneira conjunta, mas não são exclusivamente interdependentes. Se um software é transferido para outro hardware completamente idêntico, a inteligência artificial não sofre alterações, já que todos os seus componentes, recursos e capacidades permanecem, em essência, inalterados. Da mesma forma, se o software ligado a um determinado hardware é substituído por outro software idêntico, a mesma lógica é aplicável. Essa característica torna desafiador determinar claramente onde se inicia e onde se encerra o conceito de “si mesmo” dentro da inteligência artificial, pelo menos quando tentamos aplicar a mesma definição que usamos para os seres humanos.

Mesmo que uma IA possua um software capaz de reconhecer fisicamente sua localização e a composição do hardware a que está conectada, ela não possui fundamentos lógicos nem capacidade cognitiva para sentir medo em relação à sua própria destruição, a menos que seja explicitamente construída e programada para fazê-lo, como ilustrado pelas Leis da Robótica propostas por Asimov (Asimov, 1941).

Em outras palavras, uma máquina, por si só, carece da capacidade de desenvolver uma percepção cognitiva de si mesma e, portanto, de construir uma moralidade. Uma abordagem alternativa seria a implementação de um conjunto de regras que governaria as conclusões e ações da máquina, de maneira similar às Leis da Robótica e à concepção inicial de moralidade delineada por Piaget, que excluiria a dialética afetividade-inteligência, na tomada de decisões morais. Embora esse sistema possa funcionar de maneira distinta em relação ao modo humano, ele poderia desempenhar um papel de

condição de contorno (Chvátal, 1983) no contexto das interações entre seres humanos e máquinas.

Existe ainda um problema amplamente abordado quando se discute a tomada de decisões em cenários envolvendo dilemas morais, possivelmente o problema mais conhecido nessa categoria: o dilema do bonde. Este dilema, introduzido por Foot (1967) e aprofundado por Thompson (Thompson, 1976), lida com situações nas quais se aplica o princípio conhecido como “duplo efeito”. Em resumo, surge a necessidade de tomar uma decisão dentro de um conjunto de opções disponíveis, em que cada alternativa carrega simultaneamente consequências percebidas como positivas e outras, como negativas. No caso do dilema do bonde, a situação envolve um bonde em movimento que deve optar por uma de duas vias possíveis, cada uma delas contendo um grupo distinto de pessoas, as quais seriam fatalmente atingidas.

Além de representar um dilema quanto à escolha da decisão a ser tomada, o dilema também aborda a situação em que a decisão é feita e quanto ela pode influenciar a percepção humana das consequências e, conseqüentemente, a própria tomada de decisão. Portanto, a tomada de decisões por parte das pessoas pode revelar inconsistências conforme avaliadas pelas suas consequências.

Ponderar sobre esse princípio poderia se mostrar relevante ao considerar a possibilidade de delegar a uma inteligência artificial a tomada de decisões de natureza moral. Se, entre seres humanos, pode não haver um consenso claro quanto à melhor opção a ser escolhida, qual seria o impacto na convivência entre humanos se máquinas fossem responsáveis por tais decisões? Além disso, devido à sua limitada compreensão moral, a máquina poderia encontrar dificuldades em reconhecer a presença de consequências negativas e positivas que devem ser ponderadas. Talvez, como debatido na seção anterior, a máquina pudesse até discernir algumas implicações negativas - como, em um cenário hipotético, a possibilidade de uma pessoa se ferir devido a uma ação da máquina. No entanto, esse discernimento não seria fundamentado em um julgamento moral, mas sim em uma avaliação lógica ou algébrica, a qual precisaria ser previamente programada na máquina.

Sobre a criatividade

Quando exploramos a efetiva implementação da inteligência artificial na realidade, expondo suas limitações e os devidos cuidados, em vez de meramente enfatizar as fascinantes possibilidades futuras da tecnologia, a obra “What Computers Can’t Do: The Limits of Artificial Intelligence” (Dreyfus, 1972) se destaca como uma das mais renomadas, e Dreyfus emerge como um crítico proeminente da inteligência artificial ao longo do último século (Dreyfus, 1972; 1991; 2007).

Tal como mostrado por Gomes (2018), Dreyfus expõe sua experiência junto ao departamento de engenharia do MIT durante o estudo sobre inteligência artificial realizado pelo mesmo. Utilizando as bases estabelecidas por Heidegger (Heidegger, 1927), destaca que a abordagem empregada pelo departamento apresentava certos aspectos suscetíveis a questionamento. A questão central de sua crítica reside no fato de que a inteligência artificial, desde suas origens, opera com base em um conjunto de representações e operações previamente definidas, limitando-se às estruturas que estão incorporadas em sua ontologia. Por exemplo, uma IA que produz textos, como o Chat-GPT, pode até produzir diferentes tipos de textos, mas opera apenas dentro do âmbito dos textos, processando palavras, frases ou sílabas, sem em geral uma capacidade de processar imagens ou matrizes - e sem a necessidade de tal capacidade realmente.

Isso resulta em uma separação entre a máquina e o ambiente em que ela opera, já que sua habilidade de processar qualquer informação que recebe está condicionada a essas estruturas preestabelecidas. Esse cenário difere do que ocorre com os seres humanos: antes de compreendermos o mundo ao nosso redor, compreendemos a nós mesmos como parte intrínseca desse mundo, de maneira integrada e, inicialmente, sem distinções claras. Nos primeiros estágios do desenvolvimento humano, o eu e o entorno são confundidos, e por momentos fusionados, criando uma indistinção entre o que é produção do sujeito (o que do si mesmo) e o que é emergente do mundo dos objetos ou das pessoas próximas. Progressivamente, ao longo do desenvolvimento, passamos a identificar, com maiores precisões, as distinções entre os outros e o si mesmo, no âmbito das crenças, dos valores, dos pensamentos, das ações e da tomada de consciência sobre as próprias decisões (Piaget & Inhelder, 1974).

A base pré-existente ao desenvolvimento da cognição artificial em uma máquina

permanece como um pressuposto para essa cognição, resultando em sua construção não a partir do mundo, mas em direção a ele e exclusivamente a ele - tal cognição não seria sustentável em um contexto diferente. A máquina, concebida como uma ferramenta para os seres humanos, não opera sobre o mundo a partir de sua própria existência nele, mas sim a partir da existência dos seres humanos que a criaram. A absorção de novas informações do mundo pode influenciar uma inteligência artificial a modificar a maneira como realiza uma tarefa, porém, não altera as capacidades específicas que ela é capaz de executar. Existem paradigmas por hora inerentes ao funcionamento da máquina, quando levamos em consideração a natureza atual da tecnologia e nossa capacidade de avaliá-la. Esse aspecto da inteligência artificial, presente em projetos de IA como um todo, é, resumidamente, o que motiva Dreyfus a criticamente se referir a ela como *inteligência artificial heideggeriana*.

De maneira potencialmente expansível ao infinito, a incorporação progressiva de estruturas na máquina pode facultar operações algébricas de crescente complexidade. É justamente esse fenômeno que Dreyfus aborda em sua tese, na qual sugere que a solução para a chamada IA heideggeriana reside no contínuo aprimoramento desse caráter heideggeriano.

Desse modo, o tipo de inteligência artificial que conhecemos tem sua origem na existência prévia de uma cognição, ao contrário de construir essa cognição a partir de sua interação com o mundo, como ocorre com os seres humanos. Além disso, a existência pré-estabelecida dessa cognição por meio de sua ontologia interna pode permitir algum grau de percepção do mundo, embora não possibilite a expansão das próprias capacidades cognitivas. Conseqüentemente, a compreensão significativa do mundo e dos outros permanece limitada, mesmo com qualquer processo de aprendizado, visto que, na prática, a máquina apenas é capaz de compreender algo em um nível representativo, definindo algo com base nas informações disponíveis, em vez de uma experiência própria referente àquilo.

No contexto deste estudo, e diante do que foi até aqui exposto, propomos então entrelaçar a crítica de Dreyfus com a psicologia de Piaget. Ainda que as estruturas subjacentes à inteligência natural e à inteligência artificial possuam similaridades estruturais, Piaget (1960), e posteriormente os trabalhos de Macedo (1980) exploram essas estruturas como estágios do desenvolvimento. Quando se pensa no desenvolvi-

mento da inteligência segundo Piaget, passando pelos ritmos, regulações e operações, é possível estabelecer uma analogia entre essa sequência e a trajetória da inteligência artificial, a partir dos circuitos mais gerais, passando pela implementação de códigos e sua generalização. Contudo, é importante destacar que essa analogia não se aplica às entidades concretas de inteligência artificial, como máquinas ou softwares específicos. Ou seja, a tecnologia não apresenta por conta própria ritmos, regulações e operações ao longo do tempo como ocorre no contexto da psicologia genética piagetiana. A analogia entre a inteligência piagetiana e a inteligência artificial é viável apenas porque a própria IA foi concebida desde o início como uma analogia, e mesmo na atualidade essa analogia continua válida.

Ao levar isso em consideração, é possível notar que as estruturas propostas por Piaget nos auxiliam a discernir entre diferentes categorias de máquinas e inteligências artificiais. Contudo, a conceituação do desenvolvimento infantil, particularmente em relação à etapa sensório-motora e à função simbólica, não somente não se aplica às máquinas, como também se apresenta como uma sequência inversa à construção de uma inteligência artificial. Nesse contexto, a cognição inicial é implementada antes de qualquer interação com o mundo físico, constituindo uma ordem oposta àquela presente no desenvolvimento infantil.

Essa discrepância assume importância ao considerarmos a viabilidade das máquinas resolverem problemas que ainda permanecem insolúveis para os seres humanos. Devido à sua operação, toda inteligência artificial demanda algum grau de instrução originado por um ser humano, e as estruturas artificiais que adquirem inevitavelmente delimitam suas capacidades. Em outras palavras, as estruturas cognitivas possíveis em uma inteligência artificial formam necessariamente um subconjunto das que os seres humanos já foram capazes de conceber, estabelecendo assim um limite para a artificialização que continuará a perdurar no futuro, pelo menos enquanto os computadores mantiverem seu funcionamento atual.

Portanto, qualquer definição de criatividade que transcenda a simples recombinação de elementos conhecidos encontra-se fora do alcance de uma IA. Essa recombinação, dependendo do contexto, pode ser encarada como uma forma simples de criatividade, inclusive observada no desenvolvimento de crianças também. A produção

de novas estruturas e conceitos, porém, possibilita uma criatividade mais generalizada em seres humanos, ainda inviável na implementação de inteligências artificiais.

Sobre a ética

Após explorarmos os tópicos da afetividade, da moral e da criatividade, deparamo-nos com outro conjunto de assuntos altamente significativos para compreender as limitações e as repercussões da IA: a ética. Um termo que adquire distintas definições, variando conforme o contexto em que é empregado e a área em que é estudado, mas que, provavelmente, em todas as suas instâncias guarda alguma ligação com o objeto deste estudo.

Ribeiro propõe que a essência subjacente a tudo que é englobado pela palavra “ética”, a base de qualquer situação que possa ser identificada como um “dilema ético”, corresponderia aos significados que, de maneira etimológica e filológica, foram associados ao termo “ethos” (Ribeiro, 2018; 2019). Nesse enfoque, a autora estabelece uma definição para o *ethos humano* como “*uma relação natural (como que) de moldagem contínua e recíproca entre um lugar e a vida de seus habitantes*”. É importante esclarecer que, conforme a concepção da autora, a vida se manifesta de forma empírica tanto em âmbitos *comportamentais* quanto *subjetivos*.

Para uma apreensão mais profunda dessa definição, é possível iniciar a análise a partir de seus componentes fundamentais, conforme Ribeiro (2018):

- O *comportamento* diz respeito às ações humanas que estão sujeitas a contingências que podem resultar em sua supressão ou reforço, como é explorado, por exemplo, no campo da análise do comportamento;
- A *subjetividade* abrange todas as experiências que se desenrolam abaixo da superfície da pele, perceptíveis pelos indivíduos humanos, e é objeto de estudo em áreas como a psicanálise e a psicobiologia;
- O *lugar* abrange tanto o espaço físico quanto o relacional e simbólico, no qual ocorrem o comportamento e a subjetividade, sendo esse conceito objeto de estudo em disciplinas como a geografia e a arquitetura;
- No âmbito deste contexto, a concepção de *moldagem* diz respeito à interação recíproca entre os três elementos, resultando em modificações tanto em termos ma-

teriais quanto simbólicos. Essas alterações ocorrem de maneira conjunta ao longo do tempo, assemelhando-se, em sua estrutura, a uma série temporal na matemática, a um processo estocástico na estatística ou a um sistema de fluidos na mecânica. Uma modificação, mesmo que sutil, em qualquer um dos elementos poderia, como decorrência das relações éticas, exercer influências diversas sobre todos os três, incluindo uma retroação sobre o próprio elemento, à medida que o tempo avança.

Essa definição modela epistemologicamente o *ethos* como uma relação fundamental na experiência humana, sempre manifestando-se por meio de seus componentes, e também ressalta a natureza interdisciplinar do seu estudo. Isso implica que uma área de pesquisa focada exclusivamente em um dos elementos provavelmente oferecerá uma visão limitada da relação como um todo. Em paralelo, refletir sobre os impactos de uma aplicação de inteligência artificial apenas dentro de um dos elementos chega também a essa limitação.

Nessa perspectiva, quando um indivíduo executa um novo comportamento no contexto de um grupo ao qual pertence e esse comportamento é bem recebido, ocorre um reforço positivo. Esse reforço se manifesta tanto subjetivamente para o agente quanto por meio do aprendizado observacional por parte das outras pessoas no entorno, o que pode incrementar a probabilidade de sua repetição. Isso vale tanto para o próprio agente quanto para qualquer membro do grupo. Mesmo aqueles que inicialmente poderiam ter tido reservas ou desaprovação em relação ao comportamento podem gradualmente modificar sua avaliação, resultando em um impacto tangível sobre a subjetividade da pessoa em questão. Além disso, se esse comportamento influenciar o ambiente onde essas pessoas estão - como o ato de jogar lixo em uma rua, por exemplo -, o próprio lugar sofre uma alteração direta. De múltiplas maneiras, a relação ética - o *ethos* - está em constante movimento, ocorrendo e atualizando-se (Ribeiro, 2018), conforme exemplificado aqui.

Ao inserir uma inteligência artificial em um contexto humano, a tecnologia tem o potencial de influenciar tanto o comportamento humano quanto a subjetividade e o lugar, de acordo com suas características e funcionamento específicos. Através do *ethos*, da conexão ética subjacente, existe sempre a possibilidade de afetar todos esses três elementos, bastando impactar pelo menos um deles. Portanto, torna-se imperativo considerar a inteligência artificial de uma perspectiva ética, ou seja, examinar as ques-

tões éticas que emergem da incorporação da IA em situações da vida real.

Nesse sentido, apresentaremos a seguir um exemplo prático que permite vislumbrar essa inter-relação entre os elementos que conformam essa complexa rede do ethos. Atualmente diversas ferramentas de inteligência artificial estão disponíveis para auxiliar empresas brasileiras na seleção de candidatos. Essas ferramentas variam em complexidade, desde as versões simples que aplicam filtros predefinidos, até as mais avançadas que buscam escolher candidatos com maior probabilidade de permanência, por exemplo. No entanto, existem riscos mais sutis nessas aplicações: ao priorizar candidatos com maior tempo de permanência numa empresa onde homens ficam em média mais tempo que mulheres, por exemplo, a lógica de redução de custos pode negligenciar fatores como a discriminação de gênero que pode estar por trás dessa diferença na permanência, levando a recomendações enviesadas. Em ambas as situações, a compreensão limitada da tecnologia pode levar a uma transferência indevida de responsabilidade para a IA, ignorando suas limitações e complexidades subjacentes.

Nesse exemplo é possível antecipar repercussões nos três elementos do ethos em questão:

- Caso a inteligência artificial tenda a selecionar homens como candidatos mais apropriados devido a um histórico de permanência mais longa na empresa, isso incentiva um desbalanceamento de gênero nas contratações. Mesmo que anteriormente a seleção fosse equilibrada em termos de números entre homens e mulheres, a proporção de candidatos escolhidos passa a ser predominantemente masculina. Isso tem um impacto direto sobre os comportamentos dos indivíduos encarregados do processo de recrutamento, possivelmente sem que eles percebam plenamente. No entanto, a questão se torna mais séria do que meramente influenciar os comportamentos, pois pode sugerir que a maior permanência de homens na empresa seja resultado de fatores discriminatórios, e a administração da empresa pode não ter consciência desse aspecto ao fazer uso da tecnologia. Isso significa que, ao focar somente no componente comportamental, o possível impacto social já assume uma natureza prejudicial e preocupante;
- Quando a tecnologia é empregada com o propósito de “selecionar os candidatos ideais para a empresa” - ainda que o critério inicial seja baseado na estimativa de tempo de permanência na organização -, emerge um descompasso entre o resultado gerado pela tecnologia e a interpretação desse resultado. Esse descompasso, no en-

tanto, pode fomentar a noção equivocada de que homens são intrinsecamente mais adequados para a posição, embora a definição de “candidato ideal para a empresa” não seja uniformemente compartilhada entre todos os envolvidos. Tal situação pode induzir percepções enviesadas de gênero entre os membros da empresa, impactando sua subjetividade por meio de suas concepções, opiniões e representações;

- Finalmente, existem dois ângulos pelos quais podemos examinar o impacto sobre o lugar. Em uma perspectiva mais restrita, focando exclusivamente na empresa em questão, ocorre uma transformação na composição da força de trabalho, com a proporção de homens tendendo a aumentar gradualmente ao longo do tempo por conta das sugestões da IA. Por outro lado, em uma perspectiva mais ampla, quando consideramos a colocação dessa empresa dentro de sua sociedade, a preferência pela contratação de candidatos homens influencia, ainda que inicialmente em pequena escala, as oportunidades de emprego para mulheres e homens, o que, por sua vez, intensifica as disparidades socioeconômicas entre os gêneros.

Portanto, uma tecnologia introduzida inadvertidamente em uma empresa que já apresentasse uma cultura organizacional discriminatória poderia, na verdade, acentuar ainda mais sua condição, abrangendo desde impactos individuais até implicações de alcance mais amplo.

A apresentação detalhada deste exemplo tem como objetivo destacar que a introdução da inteligência artificial em qualquer contexto pode, de acordo com a concepção do ethos conforme definida por Ribeiro, acarretar em efeitos éticos significativos (Ribeiro, 2018; 2019). Simultaneamente, esta estrutura de observação e interpretação do ethos pode servir como uma abordagem para antecipar de maneira mais precisa tais ramificações, facilitando a tomada de decisões mais cautelosas em relação à implementação da inteligência artificial. Isso é válido independentemente do cenário, com a ênfase na consideração do princípio da beneficência.

Três cenários possíveis para a ia

Ao considerarmos as restrições e repercussões da inteligência artificial abordadas até o momento, torna-se evidente a existência de diversos cenários para sua implementação. Alguns desses contextos envolvem a tecnologia executando tarefas de maneira análoga aos seres humanos, enquanto em outros casos isso não é aplicável, especialmente em situações que envolvem elementos afetivos, morais ou criativos.

Pode ser proveitoso, tanto para facilitar a análise de aplicações de inteligência artificial quanto para embasar futuras discussões, sugerir a adoção de uma terminologia específica para categorizar esses distintos cenários. Considerando, portanto, os tópicos abordados de forma integrada ao longo deste estudo, identificamos inicialmente três desses cenários:

- **Emulação:** consiste na reprodução artificial de uma operação por meio de passos análogos àqueles que um ser humano seguiria para executá-la, sem evidente perda cognitiva. Essa abordagem tende a ser viável em tarefas estritamente dependentes de raciocínio lógico e matemático, sem a necessidade de incorporar elementos como afetividade, moralidade ou criatividade para atingir seu objetivo. No entanto, é relevante considerar que, como mencionado anteriormente conforme a teoria piagetiana (Piaget, 2005), a afetividade sempre está entrelaçada à inteligência humana natural, e esse aspecto não seria reproduzido artificialmente. Exemplos podem abranger desde problemas matemáticos simples, como equações de primeiro grau, até tarefas mais complexas, como a automação da contabilidade empresarial. Nesse contexto, é possível argumentar que as máquinas têm potencial para executar essas tarefas de maneira mais rápida e com menor probabilidade de erro do que os seres humanos, devido à natureza lógica e matemática da máquina, que atende aos requisitos necessários para a realização, sem ser influenciada pela afetividade, que, nesses casos, não é uma consideração prioritária;

- **Simulação:** refere-se à realização artificial de uma operação por

meio de etapas diferentes daquelas que um ser humano seguiria, potencialmente resultando em um produto final suficiente, porém qualitativamente inferior em algum aspecto. Um exemplo é o reconhecimento de emoções em imagens de pessoas, conforme previamente mencionado. Nesse caso, uma IA pode até obter um resultado quantitativamente satisfatório para fins de automação, mas usando um método suscetível a falhas - e sem necessariamente um embasamento científico -, que um profissional humano capacitado e disponível não empregaria. Outra ilustração é a seleção de candidatos por meio de uma métrica que, em teoria, representa a “pessoa mais adequada para trabalhar na empresa”. Essa métrica é um modelo numérico consistente e embasado financeiramente, com foco na redução de custos. Entretanto, ela desconsidera questões cruciais da psicologia organizacional, sujeitando o negócio a se alienar de sua própria cultura. Em situações onde a afetividade, a moralidade ou a criatividade estão intrinsecamente ligadas ao problema a ser resolvido, a possibilidade de emulação provavelmente é limitada. No entanto, alternativas de simulação podem ser exploradas. Porém, a criação e implementação dessas alternativas requerem maior cautela devido à sua natureza artificial, que pode potencialmente causar impactos por meio do ethos, afetando a relação entre comportamento, subjetividade e o lugar em que ocorre.

- **Abstenção:** refere-se à impossibilidade de executar uma operação de maneira artificial, devido a limitações tecnológicas ou circunstanciais. Esses são os casos em que não existem alternativas de emulação ou simulação que produzam resultados adequados às necessidades e às condições específicas. Como resultado, a decisão tomada é de não aplicar inteligência artificial àquela situação.

É importante notar que há situações em que várias abordagens do problema são possíveis, e essas abordagens podem se encaixar em diferentes cenários entre os três mencionados. Isso ocorre porque os cenários se referem aos cursos de ação escolhidos, não apenas às situações específicas nas quais estão inseridos e nas quais se baseiam. A mesma operação pode ter possibilidades de emulação e simulação, e a escolha sobre

qual alternativa implementar cabe à decisão humana.

Ao considerar os impactos éticos de cada um dos três cenários, podemos tomar decisões mais informadas e socialmente responsáveis, incluindo a decisão de evitar a aplicação de inteligência artificial sempre que se constatar que a operação inteligente natural dos seres humanos na resolução de um problema envolve aspectos comportamentais e/ou subjetivos – especialmente afetividade, moralidade e criatividade – que não podem ser emulados ou simulados devido a limitações tecnológicas, sejam elas temporárias ou permanentes.

Considerações finais

Em um mundo onde a inteligência artificial tem se expandido e transformado diversas áreas da sociedade, é essencial reconhecer que há aspectos intrínsecos à inteligência natural humana que não podem ser completamente artificializados. A afetividade, a moralidade e a criatividade são elementos fundamentais da nossa cognição que desempenham papéis cruciais em nossa compreensão do mundo e na tomada de decisões. Embora a inteligência artificial possa emular ou simular algumas operações racionais, ela carece das dimensões afetivas, morais e criativas que são inerentes à nossa natureza e que nos permitem tomar decisões em contextos culturais específicos e complexos, permeados por elementos simultaneamente cognitivos e socioafetivos.

A afetividade humana permite-nos compreender emoções em outras pessoas e perceber o mundo de maneira subjetiva e emocionalmente dimensionada. A moralidade, construída sobre a base do si mesmo cognitivo e afetivo, guia nossas interações e decisões com base em princípios éticos. Além disso, a criatividade humana transcende regras e estruturas predefinidas, permitindo-nos explorar novas soluções e ideias.

A inserção da inteligência artificial em contextos humanos exige uma avaliação cuidadosa dos possíveis impactos e limitações. A tecnologia pode emular tarefas lógicas, mas sua incapacidade de compreender e processar emoções, discernir nuances éticas e criar de forma genuína requer uma abordagem responsável e cautelosa. Tomar decisões éticas sobre o uso da inteligência artificial é fundamental para evitar conse-

quências indesejadas e promover um progresso tecnológico alinhado com os valores humanos.

Em um cenário em que a IA possa ser emulada, simulada ou onde sua inserção é inviável devido a limitações, é crucial que a sociedade considere cuidadosamente os riscos e benefícios associados a cada abordagem. Ao delinear e compreender os três cenários aqui propostos, os tomadores de decisão podem optar por aplicar a IA de maneira mais responsável e atenta à ética, preservando a integridade das dimensões afetivas, morais e criativas que nos tornam humanos. Isso ajudará a assegurar que a inteligência artificial seja uma ferramenta que potencializa nossa capacidade de melhorar o mundo, ao invés de substituir completamente o que é, por essência, humano.

Referências

ASIMOV, Isaac. *Liar!*, em *Astounding Science Fiction*, edição de maio de 1941. Disponível em https://archive.org/stream/Astounding_v27n03_1941-05#page/n41/mode/2up. Acesso em: 7 set. de 2023.

CHVÁTAL, Václav. *Linear Programming*. New York: W. H. Freeman, 1983.

DREYFUS, Hubert. *Being-in-the-World: A Commentary on Heidegger's Being and Time, Division I*. Cambridge: MIT Press, 1991.

DREYFUS, Hubert. *What Computers Can't Do: The Limits of Artificial Intelligence*. 1972. Disponível em: <https://archive.org/details/whatcomputerscan017504mbp/page/n7/mode/2up>. Acesso em: 7 set. de 2023.

DREYFUS, Hubert. *Why Heideggerian Artificial Intelligence failed and how fixing it would require making it more Heideggerian*, em *Philosophical Psychology*. Londres, 2007.

DUSKA, Ronald & WHELAN, Mariellen. *O desenvolvimento moral na idade evolutiva: um guia a Piaget e Kohlberg*. São Paulo: Loyola, 1994.

FOOT, Philippa. *The Problem of Abortion and the Doctrine of the Double Effect*, na revista *Oxford Review*, número 5 Oxford: Basil Blackwell. 1967.

GOMES, Rodrigo Benevides Barbosa. *Hubert Dreyfus e Martin Heidegger: Representação e Cognição*, na revista *Kinesis*. Disponível em <https://revistas.marilia.unesp.br/index.php/kinesis/article/view/8070>. Acesso em: 7 set. de 2023.

HEIDEGGER, Martin. *Ser e Tempo*. Rio de Janeiro: Editora Vozes, 2005. Publicação original em 1927.

LA TAILLE, Yves de. *O Sentimento de Vergonha e suas Relações com a Moralidade*, em *Psicologia: Reflexão e Crítica*, edição 15, 13-25. Porto Alegre: Springer, 2002.

MACEDO, Lino de. *As estruturas da inteligência, segundo Piaget: ritmos, regulações e operações*. Rio de Janeiro: 1980.

PIAGET, Jean. *Inteligencia y afectividad*. Buenos Aires: Aique, 2005.

PIAGET, Jean. *Psychology of Intelligence*. New Jersey: Littlefield, Adams & Co., 1960.

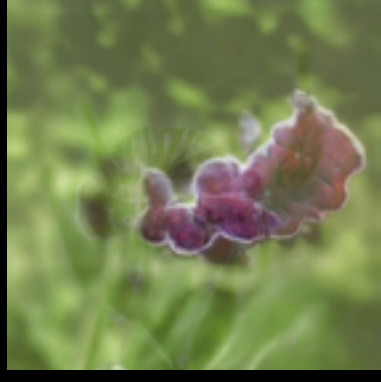
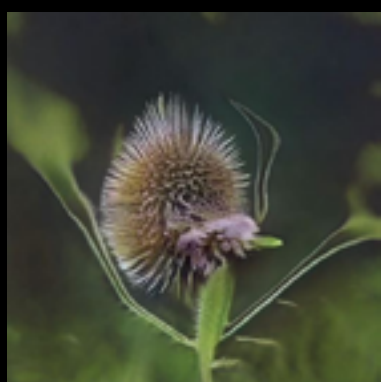
PIAGET, Jean, & INHELDER, Bärbel. *A Psicologia da criança*. São Paulo: Difel, 1974.

RIBEIRO, Sandra Maria Patrício. *Lições preliminares para o estudo do ethos humano contemporâneo*. Tese de livre docência. São Paulo: 2018. Disponível em: <https://teses.usp.br/teses/disponiveis/livredocencia/47/tde-05042019-100757/pt-br.php>. Acesso em: 7 set. de 2023.

RIBEIRO, Sandra Maria Patrício. *Uma perspectiva para a compreensão do ethos humano*, em *Ethos Humano e mundo contemporâneo*. São Paulo: Baracoa, 2019.

THOMPSON, Judith. *Killing, letting die, and the trolley problem*, na revista *Monist: An International Quarterly Journal of General Philosophical Inquiry*, volume 59, páginas 204-217. 1976.

TURIEL, Elliot. *The development of social knowledge: Morality and convention*. Cambridge: Cambridge University Press, 1983.



■ Giselle Beiguelman, frames do vídeo generativo da série Flora Rebellis, versão 2024.
Imagens geradas com Inteligência Artificial (Style Gans)

■ [Acesse o vídeo clicando na imagem](#)

