

ARTIGO

Consciência e liberdade na era da Inteligência Artificial

*João Gabriel Sadzinski**

Resumo: Este ensaio pretende refletir e argumentar sobre a consciência e a liberdade a partir do advento das inteligências artificiais. Para isso, o conteúdo foi dividido em duas partes fundamentais, cada uma fracionada em tópicos que exploram respectivamente estes temas sob diferentes perspectivas. Na primeira, o problema da consciência é analisado a partir da ótica fisicalista, do argumento do zumbi de Chalmers, do argumento do quarto chinês de Searle e da fenomenologia de Husserl. Na segunda, reflete-se sobre a liberdade e a responsabilidade a partir do existencialismo sartriano e da teoria crítica da tecnologia de Feenberg. Conclui-se que a tecnologia não pode ser considerada neutra, visto que é projetada a partir de valores e interesses privados; apenas aos seres humanos podem ser atribuídos conceitos como consciência, liberdade e responsabilidade; é imprescindível a pesquisa interdisciplinar sobre as IAs para prevenir ou mitigar os danos que podem se seguir do seu uso irrefletido.

Palavras-chave: Inteligência Artificial; Consciência; Liberdade; Responsabilidade.

Consciousness and freedom in the age of Artificial Intelligence

Abstract: This essay aims to reflect on and discuss consciousness and freedom in light of the advent of artificial intelligence. To this end, the content is divided into two fundamental parts, each broken down into topics that respectively explore these themes from different perspectives. In the first part, the problem of consciousness is analyzed from a physicalist perspective, Chalmers' philosophical zombie argument, Searle's Chinese Room argument, and Husserl's phenomenology. In the second part, freedom and responsibility are examined through the lens of Sartrean existentialism and Feenberg's critical theory of technology. It concludes that technology cannot be considered neutral, as it is designed based on private values and interests; concepts such as consciousness, freedom, and responsibility can only be attributed to human beings; and interdisciplinary research on AI is essential to prevent or mitigate the harms that may result from its unreflective use.

Keywords: Artificial Intelligence; Consciousness; Freedom; Responsibility.

1. A era da inteligência artificial

As inteligências artificiais representam, sem sombra de dúvidas, um dos avanços tecnológicos mais inovadores da atualidade. Essa ferramenta vem sendo cada vez mais utilizada em

diferentes áreas de atuação. Em geral, ela é aplicada a fim de otimizar ou possibilitar processos que antes exigiriam muito tempo ou então sequer seriam possíveis. Aparentemente, a sua utilidade é bastante conveniente e desejável para a sociedade: como ocorre com outras tecnologias, espera-se que com ela sejam obtidos resultados qualitativamente superiores aos que teríamos na sua ausência; que sejam reduzidas as horas de trabalho necessárias para atingir determinados objetivos; que proporcione, assim, mais qualidade de vida.

Considerando a própria natureza autodidata das IAs e que elas compõem um fenômeno muito recente na história, é praticamente impossível prever com precisão todos os efeitos que ela produzirá ao longo dos anos na sociedade. É possível que muitos deles não sejam positivos, mas pelo contrário, tragam grandes problemas para a humanidade. Se a IA é capaz de realizar as funções do trabalho humano, em alguns casos melhor que ele próprio, e ainda de se tornar cada vez mais eficiente, coerente e precisa, ou numa palavra, potente, então é (ou pelo menos parece) dedutível que, por exemplo, seja apenas uma questão de tempo até que cada vez mais pessoas sejam substituídas nos seus empregos por ela. Por outro lado, é possível objetar que a IA não é, nem nunca será, capaz de substituir o ser humano, dado que este possui propriedades fundamentais que excedem os horizontes da tecnologia, como consciência, intencionalidade, experiências subjetivas, sensibilidade, emoções e sentimentos. Em meio a divergência de opiniões sobre o futuro, já é possível constatar que, para nossa prosperidade ou decadência, o mundo mudou para sempre.

2. O animal racional

Durante a maior parte da história, considerou-se que o ser humano seria o único capaz de raciocinar. Inclusive, a razão foi utilizada por muitos filósofos — dentre eles Platão, Aristóteles, Descartes, Locke, Kant etc. — para definir, cada um a sua forma, a característica essencial do ser humano que o distinguiria de outras formas de vida. Animais sencientes podem possuir características como sensibilidade, memória, emoções, talvez até mesmo pensamentos, embora de forma distinta do ser humano. Mas, se os humanos são apenas mais uma forma de vida entre inúmeras outras, o que justificaria, por exemplo, a adoção de uma ética antropocêntrica, isto é, constituída de forma hierarquicamente superior pelos interesses humanos? Talvez a humanidade responda a essa questão se orientando mais pelos seus próprios interesses do que pela verdade: poder-se-ia argumentar que, em geral, a sociedade privilegia moralmente os seres humanos porque é constituída por eles. Com efeito, evidencia-se a importância dessa discussão: seriam os humanos seres especiais? Os únicos capazes de raciocinar? Se sim, por quê? As IAs teriam esta propriedade em comum? As conclusões dessas premissas excedem o campo puramente teórico, influenciando diretamente os valores da sociedade, consequentemente as suas ações.

A respeito de outras formas de vida sencientes, embora seja crível que animais irracionais possam ter mente, isto é, que percebam impressões internas e externas ao funcionamento do seu corpo, em contrapartida, seria inadmissível, até que se prove o contrário, dizer que eles possuem consciência do que são e do que o mundo é, ou seja, eles podem até se perceber e ao mundo, mas não saberão identificar, nem separar uma coisa da outra ou refletir sobre as suas percepções.

3. A consciência a partir do fisicalismo

Se adotássemos o critério fisicalista — certamente tendencioso (mas qual conhecimento é totalmente neutro e imparcial?) —, segundo o qual a realidade pode ser reduzida apenas aos fenômenos físicos, e assim considerássemos que também a consciência humana pode ser reduzida ao cérebro humano, cujas células, em última instância formadas por partículas fundamentais¹, operam segundo as leis da física, então poderíamos afirmar que, analogamente às IAs, a consciência humana é produto de operações lógicas e mecânicas determinadas em última instância por princípios invariáveis. Portanto, a partir dessa perspectiva, poderíamos concluir que qualquer consciência 1) não passa de um fenômeno ilusório dado que não é possível haver um sujeito separado do mundo físico que prove a própria existência ou 2) que tanto o ser humano quanto as IAs são dotados de consciência, mas neste caso, a consciência é caracterizada como produto de processos mecânicos e determinada pelas leis da natureza. Em outras palavras, ou a consciência é definida como uma propriedade especial e transcendente (como a alma) e por isso não é possível provar sua existência, ou seria empírica, e neste caso, não haveria motivos para afirmar que ela é exclusiva do ser humano e que não seria possível reproduzi-la artificialmente em IAs.

4. O zumbi de Chalmers

A fim de reduzir ao absurdo uma das hipóteses fisicalistas de que a consciência não é um fenômeno integrante da realidade, Chalmers propõe um experimento hipotético a partir de uma analogia com zumbis (Teixeira, 1997). Neste caso, zumbis podem ser definidos como seres fisicamente idênticos aos humanos, com a única diferença de que não possuem nenhuma consciência das suas experiências subjetivas: em suma, são corpos sem mente (Teixeira, 1997). O experimento mental de Chalmers é o seguinte: imagine uma realidade paralela fisicamente idêntica a nossa, cada detalhe se repete — os mesmos acontecimentos e as mesmas pessoas —, exceto por uma única e fundamental diferença: os seres que nela habitam não têm nenhuma consciência, isto é não experimentam fenômenos subjetivos, são o que definimos como zumbis. Quando na realidade

¹ Partículas fundamentais são partículas menores que o átomo que não são formadas por nenhuma partícula menor. Por exemplo: elétrons, quarks, fótons, bósons.

conhecida alguém faz alguma coisa, na realidade paralela o seu respectivo clone faz igual. Agora suponha, por exemplo, que uma pessoa na primeira realidade esteja tomando uma xícara de chá. Fazendo isso, ela sente o gosto e aroma provenientes da combinação de ervas utilizadas para o seu preparo, além disso talvez esteja muito quente, frio ou no ponto ideal. Estas características só podem ser atribuídas a partir das impressões do seu consumidor, afinal alguém poderia sentir que o sabor de uma erva se fez notar mais que o das outras, enquanto o outro não, ou ainda alguém poderia sentir que está quente demais enquanto para outra pessoa poderia estar no ponto ideal. Como foi dito, as realidades operam de forma simultaneamente, portanto quando aquela pessoa tomou a xícara de chá, o seu clone fez o mesmo. No entanto, se não há consciência de experiências subjetivas, para onde teriam ido o gosto das ervas e a temperatura do chá? Teriam simplesmente não existido em momento algum? Evidentemente, essas impressões se fazem presentes na realidade conhecida, portanto não seria este um forte indício da existência da consciência?

5. O quarto chinês

Já em contrapartida àqueles que consideram a possibilidade das IAs serem dotadas de pensamento próprio, Searle (1997) formula o que ficou conhecido como o argumento do quarto chinês. A provocação deste argumento pode ser ilustrada da seguinte maneira: imagine que um indivíduo (que não sabe chinês) foi isolado em um quarto e encarregado da tarefa de responder a um outro sujeito no quarto ao lado por meio de bilhetes em chinês, e para isso só tenha como auxílio um livro cujo idioma ele entende e cujo conteúdo o instrui a responder ‘x’ quando fosse recebido ‘y’ e assim por diante, sem saber o real significado de cada palavra. Nesta situação, é certo que haveria uma espécie de comunicação e que o sujeito da sala ao lado receberia as respostas desejadas, mas o indivíduo que não sabe chinês, estaria de fato pensando para escrever essas respostas ou se trataria apenas de um movimento mecânico destituído de um real entendimento?

Agora, substitua o indivíduo do primeiro quarto pela IA e o sujeito do segundo pelo usuário desta. Assim, pode-se compreender de uma forma menos técnica como se dá o funcionamento desse complexo processo. Da mesma forma que o indivíduo do primeiro quarto não fala chinês, mas somente responde conforme lhe for instruído por intermédio dos comandos do livro, sem compreender as palavras que escreve, as IAs não pensam ou refletem o que respondem ao seu usuário, mas somente seguem um roteiro (algoritmo) que as orienta a dar uma determinada resposta com base em sentenças e variáveis pré-programadas pelos seus desenvolvedores. Portanto, embora cotidianamente as IAs possam aparentar aos seus usuários que possuem consciência do que estão afirmando e que essa tecnologia representa a existência de seres autônomos, na realidade, se analisados os seus fundamentos, evidencia-se que não são nada mais que algoritmos programados

para operar, embora em alto nível de complexidade, ainda assim de forma mecânica e destituída de consciência.

6. A consciência intencional

Visando solucionar a aparente incompatibilidade entre posicionamentos radicais da epistemologia — o racionalismo que privilegiaria a razão em detrimento do objeto e o empirismo que faria exatamente o oposto —, Kant argumenta (Marcondes, 2007) que o conhecimento humano opera de tal forma que o sujeito só pode representar a realidade a partir do objeto e este, por sua vez, só pode ser considerado real a partir da representação daquele, ou seja, o sujeito precisa experienciar o mundo objetivo para saber que ele mesmo é parte da realidade, simultaneamente o objeto necessita ser apreendido por um sujeito para ter sua existência determinada.

Husserl, por sua vez, utilizou estas reflexões como referência para analisar o problema da consciência, dando início a uma nova tradição denominada *fenomenologia*. Ele buscou superar o modelo dualista da consciência de Descartes segundo o qual o mundo (incluindo o corpo) é compreendido como de uma natureza distinta da consciência (Marcondes, 2007), e em resposta a esse problema, Husserl defendeu que ambos são, alegoricamente, como dois lados de uma mesma moeda: a realidade subjetiva não pode ser separada da realidade objetiva. Se o sujeito adotasse uma suspensão de juízo (*epoché*) em relação a todos os objetos percebidos, ao fim deste processo a própria consciência deixaria de fazer sentido e a realidade como um todo se reduziria ao mais puro nada. Portanto, a consciência sempre existe em relação ao objeto e vice-versa. Toda consciência é consciência de alguma coisa: é o que Husserl denomina de *intencionalidade*. Dessa forma, a consciência deixa de ser transcendente para ser transcendental².

7. A liberdade

O conceito de liberdade pode possuir múltiplas variações a depender do contexto. Há definições de liberdade que podem se relacionar, mas são em si mesmos distintas umas das outras, como por exemplo: liberdade física, jurídica, política, econômica, religiosa etc. No entanto, pode-se argumentar que para qualquer uma delas ser considerada legítima, seria necessário um tipo de liberdade antecedente e fundamental: a autonomia da escolha. Seria muito estreito o significado da liberdade, por exemplo, no contexto da liberdade para fazer cultuar uma determinada religião, se no fundo, a pessoa em questão considerada livre nunca tivesse de fato tido autonomia para decidir se iria ou não aderir a essa religião. Restaria dizer que ela tem liberdade para não ser impedida de praticar algo que nunca escolheu. O que, claro, ainda tem seu valor, no entanto, como dito, restringe

² Refere-se às condições de possibilidade da experiência.

significativamente o próprio sentido da liberdade. Por outro lado, esse tipo mais fundamental de liberdade talvez seja, em função da sua natureza, o mais difícil de demonstrar, dado que, como tratado anteriormente, sob uma perspectiva fisicalista, é possível argumentar que a realidade é um produto de causas e efeitos físicos. Assim, dir-se-ia que ninguém possuiria de fato autonomia de escolha, considerando que, objetivamente, o ser humano é composto de matéria e por isso está sujeito a ter suas escolhas determinadas por fatores físicos. Em suma, se não há autonomia de escolha, então ninguém é realmente livre.

Se a liberdade é caracterizada fundamentalmente pela autonomia de escolha, é lógico deduzir que para usufruir dela é necessário ter consciência do que se escolhe. Para dar forma a este raciocínio, pegue por exemplo o caso em que uma pessoa comete um crime sob um surto psicótico. Pode-se dizer que ela teve liberdade num sentido de ausência de impedimentos físicos, afinal do contrário não seria possível sua ação, mas teria sentido dizer que ela teve autonomia para decidir o que estava fazendo? Se ela é portadora de uma doença que a privou naquele momento de ter consciência, não há como sustentar que ela teve autonomia para escolher cometer o crime, logo não é possível responsabilizá-la pelo que foi feito. Portanto, se para ser livre é necessário ter consciência, e consequentemente autonomia de escolha, podemos dizer que as IAs não são dotadas de liberdade, e por isso não têm qualquer responsabilidade pelos resultados, bons ou maus, das suas ações.

Partindo do existencialismo sartriano (Sartre, 2014), pode-se afirmar que a liberdade se manifesta por meio de um único ser no mundo: o ser humano. Isso porque a própria estrutura da sua consciência, se dá de tal forma que é possível que ele seja algo diferente numa situação futura do que é numa situação presente: não há um destino que o determine. Qualquer outro ser ou objeto tem sua essência definida antes mesmo de passar ao plano da existência. Uma tesoura, por exemplo, antes de ser fabricada foi pensada como e para que iria existir. Assim, a *essência* da tesoura precede a sua *existência* (Sartre, 2014). Um animal irracional, também antes de existir, tem seu comportamento já determinado pela sua genética: um gato nunca vai se comportar como um cão e vice-versa. O ser humano é o único que inverte essa lógica: só depois de existir ele se torna alguma coisa, e se torna aquilo que faz de si próprio numa dada *situação*. Isto é, ele é o ser que se define pelas escolhas que faz (Sartre, 2014), escolhas estas que só são possíveis por conta da forma que a sua consciência se relaciona com o mundo.

8. A responsabilidade e a má-fé

Se por um lado somos os únicos seres com autonomia de escolha, somos também os únicos responsáveis pelas nossas ações. Em relação às IAs, por exemplo, somos responsáveis tanto pelo

uso aparentemente inofensivo como pelos efeitos que podem lesar a sociedade. Assim a responsabilidade tanto pelo lado positivo como pelo lado negativo só pode pertencer aos seus desenvolvedores e usuários que, de um lado desenvolvem e disponibilizam, e de outro ampliam sua capacidade de ação alimentando seu banco de dados.

Ainda assim, há quem irá negar a urgência que temos de pensar, refletir e tomar decisões a respeito das IAs. É muito mais confortável fingir que não há com o que se preocupar. Não é raro que as pessoas externalizem as responsabilidades por suas ações. Comumente se argumenta: “a sociedade determina o indivíduo”. Embora essa frase tenha sua parcela de razão, não se trata de uma verdade absoluta, pois se levada às últimas consequências torna-se tão equivocada quanto dizer que o indivíduo determina a sociedade. Faz mais sentido dizer que o indivíduo e a sociedade se influenciam mutuamente: é sempre um indivíduo que age, mas nunca separado de um contexto histórico. No entanto é cômodo fingir ser mais incapaz do que se é, pois assim se exime de qualquer esforço. “Sozinho ninguém muda o mundo” dizem, e de fato, não muda, mas é por meio da união de ações individuais, situadas historicamente, que se faz a diferença. Ou seja, se o indivíduo não se engaja na atividade, aí que não existem motivos para se esperar por mudanças. Analogamente há quem pense que não tem capacidade para escrever por conta própria e então delega a sua atividade para as IAs. Qual a consequência dessa má-fé? O indivíduo nunca desenvolve a habilidade de escrever, pois cede sua função de pensar para a máquina. Não se defende aqui que se deve dispensar a máquina, com certeza ela pode contribuir, mas definitivamente não dessa forma (e tantas outras). Também há quem diga: “se há tantas pessoas usando sem se preocupar, que diferença faz mais um?”. Bom, é justamente havendo indivíduos que pensam assim que se forma uma sociedade destituída de pensamento próprio.

9. A teoria crítica da tecnologia

Como vimos, a tecnologia não pode ser responsável por ela própria, afinal ela não se originou a partir do nada, mas através da ação humana. Segundo a teoria crítica de Feenberg (2003) a tecnologia não é neutra, nem avança apenas em função de si mesma ou do progresso da humanidade. As ferramentas tecnológicas são produzidas por empresas que possuem valores e interesses. Uma empresa privada só pode continuar existindo enquanto obtiver lucro, logo antes de qualquer outra coisa, as empresas produzem pelo lucro que podem extrair do seu trabalho. Mas, seria reducionista dizer que se trata apenas de dinheiro, pois esse processo envolve também os valores que a empresa carrega consigo: o que ela considera certo e errado. Por exemplo, se uma empresa fabrica armas de fogo, independentemente de ela reconhecer ou não os possíveis danos que decorrem destes objetos, é porque em alguma medida ela defende a sua incorporação na sociedade a

partir das suas convicções. No caso das tecnologias digitais, uma rede social, por exemplo, não tem como fim último apenas conectar pessoas e possibilitar que elas se comuniquem entre si, apesar desta ser sua premissa básica. Antes de tudo, ela existe pelo lucro que pode proporcionar aos seus criadores por meio das propagandas, mas também a forma que o algoritmo dela opera reflete os valores e interesses da empresa. Os pensamentos e comportamentos que esta rede social incentiva são os que ela aprova de acordo com seus motivos. Então, se, por exemplo, ela permite a divulgação de casas de aposta, é porque, independente dos prejuízos que isso pode trazer aos seus usuários, ela porta esse valor e interesse consigo. Por isso, ao se utilizar as tecnologias, principalmente as digitais, é necessário consciência e responsabilidade, a fim de que o usuário as utilize de uma forma que não comprometa a sua autonomia e integridade.

10. Conclusão

Por fim, podemos concluir que na era da inteligência artificial há muitas promessas, mas poucas certezas. Por um lado, há muitos pontos positivos a se considerar a respeito das IAs, por outro elas oferecem pontos negativos igualmente dignos de consideração. Durante muito tempo se considerou que o ser humano era o portador definitivo da razão, atualmente a IAs vêm a cada dia mostrando que talvez a humanidade se tenha equivocado sobre isso. No entanto, ainda que a razão não seja exclusiva do ser humano, ele tem propriedades que, ao que tudo indica, lhe são próprias, como a consciência, a liberdade e a responsabilidade. A respeito disso, as teorias divergem entre si. De um lado, há os fisicalistas que negam a consciência e a liberdade ou então atribuem-lhes um caráter estritamente físico. Argumentos como o do zumbi de Chalmers e do quarto chinês de Searle criticam cada um a seu modo as hipóteses fisicalistas. Husserl, por sua vez, através da fenomenologia argumenta como a consciência e o mundo são inseparáveis um do outro, conciliando realismo e idealismo numa só teoria. Relacionando consciência, liberdade e responsabilidade, Sartre revela como apenas os seres humanos podem ser considerados livres e responsáveis. No campo da filosofia da tecnologia, Feenberg critica a interpretação segundo a qual a tecnologia é neutra de valores, concordando indiretamente com Sartre a respeito da responsabilidade que a humanidade tem em relação ao seu uso e desenvolvimento. Além disso, embora pareçam inofensivas, as IAs representam uma potência de destruição tão grande ou maior que a de criação, por isso se revela urgente a necessidade de reflexão, pesquisa interdisciplinar e atos decisórios a seu respeito.

Referências bibliográficas

FEENBERG, A. *O que é Filosofia da Tecnologia?* Trad. Agustin Apaza e Daniel Durante P. Alves. Conferência pronunciada para estudantes universitários de Komaba, Japão, jun. 2003.

MARCONDES, D. *Iniciação à história da filosofia: dos pré-socráticos a Wittgenstein*. - 2ª edição - Rio de Janeiro: Zahar, 2007.

SARTRE, J-P. *O existencialismo é um humanismo*. Trad. João Batista Kreuch. 4ª Ed. Petrópolis: Vozes, 2014.

SEARLE, J. *Mentes, Máquinas e Consciência: uma introdução à filosofia da mente*. Trad. Cléa Regina de Oliveira Ribeiro. São Carlos: Editora UFSCar, 1997.

TEIXEIRA, J. F. "A teoria da consciência de David Chalmers". *Psicologia USP*, São Paulo, v. 8, n. 2, p. 109-128, 1997.

* Graduando Licenciatura em Filosofia pela Unicesumar. Pesquisa principalmente sobre ética, existencialismo e niilismo.

E-mail: jgsdzkn@gmail.com

Lattes: <http://lattes.cnpq.br/2216869729626921>