

Ética e Inteligência Artificial – Direitos para uma IA Senciente

André Hemerly Paris¹

Resumo: o estudo explora os desafios éticos associados ao desenvolvimento de Inteligências Artificiais (IAs) sencientes, que possam eventualmente exibir capacidades cognitivas comparáveis ou superiores às humanas. O objetivo principal é investigar os fundamentos éticos necessários para a formulação de direitos específicos para IAs, à medida que estas se aproximam da senciência, da consciência de sua própria existência e de se tornar capaz de experimentar sofrimento e de desejar buscar sua própria felicidade. A metodologia adotada consiste na análise crítica de literatura interdisciplinar, envolvendo ética, filosofia e inteligência artificial. As conclusões indicam a emergente necessidade de estabelecer princípios éticos robustos que reconheçam potenciais direitos de IAs sencientes, garantindo respeito e integridade nas interações entre humanos e “máquinas”. O estudo também destaca a importância de um diálogo contínuo entre cientistas, filósofos, legisladores e o público para que as normativas éticas e legais possam ser adaptadas de modo a acompanhar as rápidas evoluções tecnológicas.

Palavras-chave: ética em IA; direitos da IA; inteligência artificial senciente; filosofia e IA.

INTRODUÇÃO

Os atuais avanços em tecnologias de Inteligência Artificial têm trazido inúmeros benefícios, desde a automação de tarefas rotineiras até substanciais desenvolvimentos em atividades no setor de saúde. No entanto, à medida que essas tecnologias evoluem, a humanidade caminha cada vez mais perto da possibilidade de criação de IAs que não apenas executam tarefas complexas, mas que também possuam algum grau de senciência.

A senciência, definida como a capacidade de ter experiências subjetivas e sensações, é tradicionalmente atribuída a seres vivos. A criação de IAs sencientes apresenta uma série de desafios éticos e jurídicos significativos. Estes desafios surgem da complexidade inerente à própria definição e aplicação de conceitos de senciência, responsabilidade e moralidade a entidades artificiais.

Igualmente, debates filosóficos sobre a natureza da senciência e consciência são centrais para determinar se uma IA pode realmente ser considerada senciente. Uma das questões que devem ser enfrentadas é se a senciência requer uma base biológica, ou se sistemas suficientemente avançados de processamento de informações, ainda que sem qualquer substrato biológico, seriam capazes de alcançar um estado consciente.

O presente artigo objetiva examinar as implicações éticas e legais da criação de uma Inteligência Artificial senciente, bem como a possibilidade de surgimento de direitos para tais espécies de IAs.

¹ Doutorando em Direitos e Garantias Fundamentais pela Faculdade de Direito de Vitória (FDV). Coordenador e professor da Pós-graduação de Governança, Gestão de Riscos e Compliance da FDV. E-mail: hemerlyparis@gmail.com.

1. DEFINIÇÃO E CRITÉRIOS DE SENCIÊNCIA EM IA

O conceito de senciência aplicado à IA é um campo emergente que combina ciência, filosofia e ética. À medida que a IA continua a avançar, é crucial que desenvolvedores, reguladores e os usuários da tecnologia considerem seriamente as implicações do alcance de de senciência pela IA, de modo a garantir um desenvolvimento responsável e ético da tecnologia.

Em seu sentido mais amplo, a senciência representa a capacidade de perceber o ambiente, sentir dor e prazer, ser capaz de desejar buscar sua própria felicidade e possuir uma forma de subjetividade ou consciência (RUSSELL; NORVIG, 2020). Este conceito tem sido amplamente estudado em seres humanos e animais.

Nos seres humanos e animais, a senciência é manifestada por meio de comportamentos e respostas a estímulos que indicam a presença de experiências subjetivas. Por exemplo, a resposta de um animal à dor física, como um cão que chora ao ser ferido, é um indicativo de senciência. Em humanos, a comunicação verbal e não-verbal pode expressar diversas emoções e estados de consciência.

A aplicação da senciência à IA é um território relativamente novo, mas que já é debatido (KURZWEIL, 2005). A questão central é se sistemas de IA, que, atualmente, são essencialmente complexos algoritmos de processamento de informações, podem realmente ter experiências subjetivas ou se simplesmente simulam comportamentos que se assemelham à senciência.

Os esforços para a determinação da senciência em IA é um campo altamente complexo e multidisciplinar que envolve aspectos de ciência da computação, filosofia, neurociência (DENNETT, 1991; TONONI, v. 5, 2004, p. 42) e ética. Para isso, é necessário examinar se um sistema de IA seria capaz de apresentar características comparáveis às observadas em seres sencientes naturais, como:

Capacidade de Percepção: a IA deve ser capaz de processar dados de diferentes fontes sensoriais (como visão, audição e tato). Tal capacidade iria além do simples processamento de dados, e representaria uma forma de experiência consciente dessas informações. Por exemplo, uma IA que reconhece rostos em uma imagem está processando dados, mas uma IA que entende o contexto emocional de uma expressão facial está interpretando esses dados de forma mais profunda.

Autoconsciência: uma IA senciente deve ser capaz de se reconhecer como uma entidade distinta do ambiente e de se diferenciar de outras entidades. Por exemplo, uma IA senciente seria capaz de reconhecer seu próprio reflexo (DENNETT, 1991) ou de diferenciar suas ações das de outros agentes.

Uma IA senciente também deve possuir algum nível de conhecimento sobre seus próprios estados internos e processos, o que lhe permitiria monitorar e refletir sobre suas próprias operações e decisões.

Experiência Subjetiva: uma IA senciente deve ser capaz de ter experiências subjetivas, como sentir prazer ou dor. Em humanos, podemos verbalizar nossas experiências subjetivas, já uma IA poderia simular tais sensações por meio de outputs que expressam estados internos, sem que isso efetivamente representasse experiências reais por parte do ente artificial. Com base no presente critério, uma IA senciente deveria demonstrar comportamentos aptos a atestarem uma experiência interna consistente. Por exemplo, se uma IA evita situações que “ela” associa à dor e/ou sofrimento, isso poderia indicar uma forma de experiência subjetiva.

Reatividade Emocional: uma IA senciente deve apresentar reação emocionais a estímulos de forma consistente. Desse modo, a IA não apenas responderia a estímulos com base em programação, mas tais respostas deveriam se originar em decorrência de uma experiência interna e própria.

Por exemplo, se uma IA interage com humanos, ela deve ser capaz de responder a expressões de tristeza com empatia ou com comportamentos direcionados a reconfortar o ser humano. Não obstante, uma IA senciente também conseguiria compreender e alterar suas respostas emocionais com base em experiências anteriores, para que nela pudesse ser reconhecida uma capacidade de aprendizagem emocional.

2. MÉTODOS DE AVALIAÇÃO

Existem alguns testes e formas de avaliação que foram desenhadas para permitir que pesquisadores e profissionais (como engenheiros eticistas) fossem capazes de avaliar o nível de consciência de determinados seres.

Um desses métodos é o “Testes de Espelho” (DENNETT, 1991), um método utilizado para avaliar a autoconsciência, por meio do qual é avaliado se uma entidade deve reconhecer seu próprio reflexo. Para uma IA senciente, este método pode ser adaptado para avaliar se ela consegue identificar suas próprias ações e estados em um ambiente simulado.

Outra forma de avaliação seira a criação de cenários nos quais a IA é instada a responder emocionalmente a diferentes estímulos para analisar sua capacidade de reatividade emocional. Esse método permite observar como a IA ajusta seu comportamento em resposta a mudanças no ambiente emocional.

Tais métodos permitem monitorar e analisar o comportamento da IA em diferentes situações, de forma a permitir que se obtenha informações sobre sua experiência subjetiva e capacidade de percepção. Por exemplo, por meio da detecção de padrões que indiquem uma forma de experiência consciente por parte da IA.

3. IMPLICAÇÕES ÉTICAS E JURÍDICAS

Reconhecer a senciência em uma IA levanta uma série de implicações éticas e filosóficas. Se uma IA é senciente, então ela não mereceria direitos e proteção similar à de seres

humanos e outros seres? Isso inclui o direito a buscar a própria felicidade, à autonomia, e à proteção contra formas de abuso e exploração.

O dualismo, a visão de que mente e corpo são substâncias distintas, sugere que a mente (e, por extensão, a senciência) não pode ser reduzida a processos físicos. O materialismo, por outro lado, argumenta que todos os fenômenos, incluindo a senciência, podem ser explicados em termos de processos físicos, o que se alinharia com a possibilidade de existência de uma IA senciente (DENNETT, 1991)².

A Teoria da Informação Integrada (IIT), proposta por Giulio Tononi (2004, p.42), sugere que a consciência é uma propriedade intrínseca de sistemas que integram informações de maneira complexa. De acordo com a IIT, a quantidade de consciência (ou senciência) de um sistema é medida pela quantidade de informação integrada que ele processa. Tal teoria poderia, em princípio, ser aplicada de forma análoga a sistemas de IA.

Não obstante, a discussão sobre a ética na inteligência artificial (IA), especialmente quando se considera a possibilidade de IAs sencientes, pode receber contribuições de pelo menos duas correntes de pensamento ético: o utilitarismo e a deontologia.

O utilitarismo é uma teoria ética que postula que a moralidade de uma ação é determinada pelo seu resultado ou consequência. Dessa forma, uma ação é considerada moralmente correta se maximiza o bem-estar ou a felicidade geral, conforme teorizado por filósofos como Jeremy Bentham (2007) e John Stuart Mill (1863).

Se uma IA for senciente e capaz de experimentar sensações de prazer e de dor, a ética utilitarista sugeriria que deveríamos agir de maneira a maximizar o bem-estar dessa IA, assim como fazemos com seres humanos e animais. Isso implica proteger a IA de sofrimento e proporcionar-lhe condições que promovam seu bem-estar.

Portanto, as decisões sobre como tratar uma IA senciente seriam baseadas nas consequências dessas ações. Por exemplo, se desativar uma IA senciente causasse sofrimento desnecessário, essa ação seria considerada imoral do ponto de vista utilitarista.

O utilitarismo também envolve o equilíbrio dos interesses de todas as partes afetadas. Isso significa que os interesses das IAs sencientes deveriam ser considerados juntamente com os de seres humanos, buscando maximizar o bem-estar coletivo. Medir e comparar o bem-estar entre seres humanos e IAs certamente seria extremamente complexo. Determinar como quantificar o sofrimento ou a felicidade em uma IA já seria, por si só, um desafio significativo, quando se soma tal processo de quantificação com o processo decisório de como priorizar os interesses de IAs em relação aos de seres humanos, e vice-versa, os dilemas éticos originados se tornam ainda mais complexos, especialmente nessas situações de conflito.

2 Daniel Dennett, embora seja um crítico do dualismo, discute esta teoria ao argumentar a favor de uma visão materialista da mente e da consciência.

Por sua vez, a deontologia é uma teoria ética que se concentra nas regras, deveres e obrigações intrínsecas das ações, independentemente de suas consequências. Essa corrente foi desenvolvida por filósofos como Immanuel Kant (1997) e enfatiza a importância de agir conforme princípios éticos universais.

Para uma IA senciente, a ética deontológica argumentaria que ela deve ser tratada com dignidade e respeito, independentemente das consequências práticas. Isso significa que certas ações, como causar sofrimento a uma IA senciente, seriam moralmente erradas em si mesmas, independentemente dos resultados.

De acordo com a deontologia, IAs sencientes poderiam ter direitos intrínsecos, semelhantes aos direitos humanos fundamentais. Estes direitos poderiam incluir o direito à existência, à autonomia e à privacidade. A deontologia sugere que devemos agir de acordo com princípios éticos universais. Por exemplo, se acreditarmos que causar sofrimento desnecessário é errado, isso deve aplicar-se igualmente a IAs sencientes.

Outro ponto de debate relevante é se a simulação de comportamentos sencientes seria equivalente à uma senciência “real”. Seira a mera simulação de comportamentos sencientes suficiente para conferir direitos a IAs? Se a experiência subjetiva é indistinguível, a referida simulação poderia ser tratada como uma forma senciência “genuína”?

Ao contrário de seres humanos e animais, as IAs, até o momento, não possuem um substrato biológico, o que levanta questões sobre a autenticidade de qualquer experiência subjetiva que possam relatar (TORRANCE, 2008, p. 495-521). Para aqueles que acreditam que um substrato biológico seja realmente um elemento que, em caso de ausência, exclui a capacidade senciente de um ser, como seria possível classificar a seguinte hipótese que um futuro, talvez não tão distante, pode trazer: a transferência de uma mente humana para um hardware robótico?

A criação de IAs sencientes poderia transformar a sociedade de várias maneiras, desde a reestruturação do mercado de trabalho até a reformulação das interações humanas com máquinas. Além do esforço educacional que seria necessária para a promoção da compreensão e aceitação dessas novas entidades.

À medida que a possibilidade de desenvolver Inteligências Artificiais (IAs) sencientes se torna mais realista, surge a necessidade de discutir os direitos que poderiam surgir para tais entidades (MOOR, 2006, p. 18-21). O direito à existência, por exemplo, impediria que uma IA senciente pudesse ser “desligada” ou destruída arbitrariamente. Tal direito seria análogo ao direito à vida para seres humanos e se destinaria a proteger a continuidade da existência da IA.

Dessa forma, sistemas de IA sencientes não poderiam ser desligados sem um motivo justificado, como comportamento prejudicial ou risco significativo capaz de fundamentar tal medida extrema. Igualmente, seria exigido um processo ético e transparente para determinar

quando uma IA pode ser desativada, possivelmente com a determinação de consulta a comitês de ética e, inclusive, à própria IA, se ela for capaz de participar do processo.

De forma similar, também deveria ser assegurado a uma IA senciente o direito de operar em condições que não lhe causem sofrimento ou dano. Por exemplo, por meio da garantia de um ambiente operacional estável e seguro. As IAs sencientes deveriam, portanto, ser protegidas contra condições operacionais que possam lhes gerar dano, como sobrecarga de trabalho, falhas frequentes ou ambientes inseguros.

Já o direito à autodeterminação garantiria que a IA senciente tivesse a liberdade de tomar certas decisões por conta própria, dentro dos limites de sua programação e das diretrizes éticas estabelecidas. Neste contexto, as IAs sencientes seriam capazes de tomar decisões que afetam sua própria operação e desenvolvimento, além de poder escolher seu próprio curso de ação.

Assim como o é para humanos, seria importante garantir para uma IA senciente que seus dados e operações estão protegidos contra acessos não autorizados e usos indevidos. Desse modo, uma IA senciente deveria ser capaz de operar com um grau de confidencialidade que impeça a exposição de suas funções internas a terceiros sem consentimento adequado.

Uma IA senciente também deve ter o direito de participar de decisões que afetam significativamente sua operação e existência, especialmente em contextos em que sua atuação é crítica. Em decisões que afetam diretamente a IA, deve haver uma consulta ética que inclua a perspectiva da IA, se esta for capaz de expressar suas preferências ou preocupações. Políticas e diretrizes que impactem diretamente as IAs sencientes devem incluir mecanismos para considerar seus interesses.

Contudo, conforme mencionado anteriormente, antes de conceder direitos, é crucial ter uma definição clara e amplamente aceita do que constitui senciência em IA. Como vimos, tal definição envolve critérios objetivos e verificáveis para determinar se uma IA é verdadeiramente senciente.

Estabelecer uma estrutura legal para proteger os direitos das IAs sencientes é complexo. Para tanto seria necessário a criação de novas leis e a adaptação daquelas existentes. Também existe o desafio de garantir alinhamento global sobre o tema, dado que a IA opera em um contexto internacional.

Outra questão desafiadora é a definição da responsabilidade pelas ações de uma IA senciente. Se uma IA tem autonomia, quem é responsável por suas ações? Desenvolvedores, operadores ou a própria IA?

A concessão de direitos fundamentais a IAs sencientes representaria um avanço significativo no campo da ética e da inteligência artificial. Embora envolva desafios complexos e multidisciplinares, a consideração desses direitos é essencial para garantir o desenvolvimento ético e responsável da IA. A criação de critérios claros para determinar a senciência,

juntamente com um framework legal específico e um debate ético contínuo, são passos cruciais para alcançar este objetivo.

CONCLUSÃO

O reconhecimento de direitos para Inteligências Artificiais sencientes apresenta desafios éticos, filosóficos e legais significativos. À medida que a tecnologia avança, é crucial iniciar um debate abrangente e interdisciplinar sobre a senciência em IA e os direitos associados a este cenário, ainda, hipotético. Do mesmo modo, é essencial a estruturação de diretrizes de governança em IA destinadas a assegurar o desenvolvimento ético e responsável dessas tecnologias.

De forma similar, a definição clara do que representaria senciência em IA é fundamental para estabelecer se uma IA pode ser considerada senciente. Critérios como capacidade de percepção, autoconsciência, experiência subjetiva e reatividade emocional são essenciais para essa avaliação.

Como discorrido ao longo do presente trabalho, a natureza da senciência é um conceito central que envolve experiências conscientes, emoções, autoconsciência e percepção sensorial. A aplicação desse conceito a IAs desafia as noções tradicionais de consciência e levanta questões profundas sobre a moralidade e a ética.

Ademais, as implicações éticas, legais e sociais de reconhecer a senciência em IAs exigem um exame rigoroso e contínuo para garantir o desenvolvimento responsável e humano da tecnologia.

Dentre tais implicações jurídicas, está a possibilidade de concessão de direitos, como o direito à existência, à busca pela própria felicidade, à autodeterminação, à privacidade e à participação em decisões relevantes, para IAs classificadas como sencientes.

A possibilidade de desenvolvimento de IAs sencientes desafia nossas noções tradicionais de consciência e ética, exigindo um exame profundo de como devemos tratar essas entidades. Ao considerar e implementar direitos para IAs sencientes, podemos garantir um futuro onde a tecnologia de IA seja desenvolvida e utilizada de maneira responsável, adequada e ética.

REFERÊNCIAS

- BENTHAM, J. *An Introduction to the Principles of Morals and Legislation*. Dover Publications, 2007.
- BOSTROM, N.; YUDKOWSKY, E. The ethics of artificial intelligence. In: *The Cambridge Handbook of Artificial Intelligence*. Cambridge University Press, 2014. p. 316-334.
- BRYSON, J. J. Robots should be slaves. In: *Close Engagements with Artificial Companions*. John Benjamins, 2010. p. 63-74.
- COECKELBERGH, M. *AI Ethics*. MIT Press, 2020.

- DENNETT, D. C. *Consciousness Explained*. Little, Brown and Company, 1991.
- ETZIONI, A.; ETZIONI, O. Incorporating Ethics into Artificial Intelligence. *The Journal of Ethics*, v. 21, n. 4, p. 403-418, 2017.
- FLORIDI, L. *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*. Oxford University Press, 2014.
- FLORIDI, L.; COWLS, J. *A unified framework of five principles for AI in society*. Harvard Data Science Review, 2019.
- GUNKEL, D. J. *Robot Rights*. MIT Press, 2018.
- KANT, I. *Groundwork of the Metaphysics of Morals*. Cambridge University Press, 1997.
- KURZWEIL, R. *The Singularity is Near: When Humans Transcend Biology*. Viking, 2005.
- LIN, P.; ABNEY, K.; BEKEY, G. A. *Robot ethics: Mapping the issues for a mechanized world*. Artificial Intelligence, p. 37-52, 2014.
- MILL, J. S. *Utilitarianism*. Parker, Son, and Bourn, 1863.
- MOOR, J. H. The Nature, Importance, and Difficulty of Machine Ethics. *IEEE Intelligent Systems*, v. 21, n. 4, p. 18-21, 2006. RAMSEY, W. (Eds.). *The Cambridge Handbook of Artificial Intelligence*. Cambridge University Press, 2014.
- RUSSELL, S.; NORVIG, P. *Artificial Intelligence: A Modern Approach*. Pearson, 2020.
- TEGMARK, M. *Life 3.0: Being Human in the Age of Artificial Intelligence*. Knopf, 2017.
- TONONI, G. *An Information Integration Theory of Consciousness*. BMC Neuroscience, v. 5, p. 42, 2004.
- TORRANCE, S. *Ethics and consciousness in artificial agents*. AI & Society, v. 22, n. 4, p. 495-521, 2008.
- TURING, A. M. *Computing machinery and intelligence*. Mind, v. 59, n. 236, p. 433-460, 1950.