

---

## INTELIGÊNCIA ARTIFICIAL: OS AGENTES MORAIS ARTIFICIAIS

### *ARTIFICIAL INTELLIGENCE: ARTIFICIAL MORAL AGENTS*

**RUI MIGUEL ZEFERINO FERREIRA**

Professor-Adjunto do Instituto de Entre Douro de Vouga (ISVOUGA). Professor-Assistente na Escola Superior de Comunicação Administração e Turismo, do Instituto Politécnico de Bragança. Doutor em Direito pela Universidade de Santiago de Compostela (Espanha). Mestre em Direito, na menção de ciências Jurídico-Económicas, pela Faculdade de Direito da Universidade do Porto (Portugal). Licenciado em Direito, na menção de ciências jurídico económicas, pela Faculdade de Direito da Universidade de Lisboa. Investigador do JusGov, da Universidade do Minho (Portugal). E-mail: zeferino\_ferreira@sapo.pt. ORCID: <http://orcid.org/0000-0001-6282-3641>.

#### RESUMO

**Objetivo:** o presente artigo analisa os problemas e os desafios decorrentes da utilização da inteligência artificial, nomeadamente, o decorrente dos agentes morais artificiais, onde se procura analisar um conjunto de problemas éticos que se levantam com a sua utilização, designadamente ao nível da responsabilidade e da existência de direito dos agentes morais artificiais.

**Metodologia:** Utiliza-se o método dedutivo, através da pesquisa bibliográfica e artigos científicos sobre a temática.

**Resultados:** Conclui-se que a inteligência artificial é uma área do direito muito singular e ainda muito desconhecida, que levanta imensas questões éticas, nas quais se inclui a referente aos agentes morais artificiais, designadamente quanto à responsabilidade dos agentes morais artificiais e, bem assim, quanto à existência de direitos dos mesmos. Igualmente, que é necessário o desenvolvimento da filosofia e da ética da inteligência artificial, por existir um conjunto de questões fundamentais sobre o qual é necessário perceber o que deve ser admitido que a inteligência artificial realize, bem como acautelar os riscos num cenário de longo prazo.



---

**Contribuições:** a pesquisa mostra-se relevante no atual contexto de revolução tecnológica, no qual a inteligência é uma das vertentes mais visíveis, para compreender como as questões dos agentes morais artificiais devem ser tratados, nomeadamente, dando contribuições para a definição de diretrizes a serem implementadas no âmbito da inteligência artificial.

**Palavras-Chave:** inteligência artificial; agentes morais artificiais; responsabilidade; direitos de agentes morais artificiais.

## ABSTRACT

**Objective:** *this article analyzes the problems and challenges arising from the use of artificial intelligence, namely those arising from artificial moral agents, where it seeks to analyze a set of ethical problems that arise with its use, namely in terms of responsibility and of the existence of rights of artificial moral agents.*

**Methodology:** *The deductive method is used, through bibliographical research and scientific articles on the subject.*

**Results:** *It is concluded that artificial intelligence is a unique and still very unknown area of law, which raises immense ethical questions, which include the one referring to artificial moral agents, namely regarding the responsibility of artificial moral agents, as well as to the existence of their rights. Likewise, it is necessary to develop the philosophy and ethics of artificial intelligence, as there is a set of fundamental questions on which it is necessary to understand what must be admitted that artificial intelligence performs, as well as to guard against risks in a long-term scenario.*

**Contributions:** *the research is relevant in the current context of technological revolution, in which intelligence is one of the most visible aspects, to understand how the issues of artificial moral agents should be treated, namely, contributing to the definition of guidelines to be implemented in the context of artificial intelligence.*

**Keywords:** *artificial intelligence; artificial moral agents; responsibility; rights of artificial moral agents.*

## 1 INTRODUÇÃO

Um dos grandes objetivos que os investigadores no campo de inteligência artificial têm buscado passa pelo desenvolvimento dos denominados agentes artificiais em coexistência com as pessoas humanas e outros sistemas (CERVANTES; LÓPEZ; ET AL., 2020, p. 501 e ss.). A doutrina que se dedica à investigação no âmbito



---

da computação tem realizado um esforço de projetar agentes artificiais cada vez mais avançados, no sentido que tenham a capacidade de realizar tarefas da mesma forma que os seres humanos. Estas tarefas requerem mecanismos cognitivos, como seja o planeamento e a gestão de opções e decisões, bem como a tomada de decisões e a aprendizagem.

Os domínios de aplicação de tais agentes são hoje em dia evidentes, com a perspetiva crescente na sua utilização num conjunto de áreas que outrora se imaginaria impossível. Na atualidade, vivencia-se a experiência da inclusão de agentes artificiais no seio da sociedade humana, do qual são exemplos os veículos não tripulados, as habitações inteligentes e os robôs “humanoides” capazes de cuidar das pessoas.

Nesse contexto, a investigação no campo da ética da máquina tornou-se uma questão primordial, senão mesmo para a sobrevivência dos seres humanos e da própria humanidade, tal como ainda a conhecemos. A ética das máquinas deve assim concentrar-se no desenvolvimento de mecanismos éticos para que agentes artificiais sejam possíveis de se enquadrar dentro de um conceito de comportamento moral. No entanto, ainda nos encontramos numa fase muito embrionária do processo, pelo que existem desafios cruciais a serem enfrentados no processo de desenvolvimento de agentes morais artificiais.

Este artigo tenta assim mostrar a situação atual a propósito dos agentes morais artificiais, com o escopo de demonstrar a complexidade de projetar e desenvolver mecanismos éticos para este tipo de agente, bem como que há um longo caminho a percorrer, do ponto de vista tecnológico, antes que este tipo de agente artificial possa substituir os seres humanos, no julgamento de situações morais difíceis, surpreendentes ou ambíguas.

## **2 CONCEPTUALIZAÇÃO DE INTELIGÊNCIA ARTIFICIAL E ROBÓTICA**

É essencial começar por avançar com um conceito de inteligência artificial, que seja ao mesmo tempo adequado e compreensível. Ora, o conceito de inteligência



---

artificial vem sendo entendida como abrangendo qualquer tipo de sistema computacional artificial, que compreenda a existência de um comportamento inteligente, isto é, um comportamento complexo que conduz ao alcance de determinados objetivos.

No seguimento do pensamento de MINSKY (1985) não deveremos restringir o conceito de “inteligência” à conceptualização da inteligência humana, pelo que devemos incluir uma grande gama de máquinas, mesmo aquelas que apenas detêm capacidades limitadas de aprendizagem ou raciocínio, mas com tarefas de automação específicas.

Os principais objetivos de um agente artificial inteligente envolvem uma vasta quantidade de tarefas, designadamente as referentes à deteção, modelagem, planeamento, gestão e ação. Porém, os atuais aplicativos de inteligência artificial incluem ainda a perceção, análise de texto, processamento de linguagem natural, raciocínio lógico, jogos, sistemas de apoio à decisão, análise de dados, análise preditiva, bem como veículos autómatos e outras formas de robótica (STONE; BROOKS; ET EL., 2016).

A inteligência artificial pode assim envolver qualquer número de técnicas computacionais com o escopo de atingir esses objetivos, seja a inteligência artificial inspirada na cognição natural ou na aprendizagem das estruturas tecnológicas por via de redes neurais<sup>1</sup>.

Num contexto histórico, importa salientar que no presente século XXI o conceito sofreu uma ampliação, ao ponto de quase toda a ciência da computação e até mesmo a alta tecnologia serem agrupadas como inteligência artificial. Atualmente, a inteligência artificial compreende uma indústria em plena expansão, que beneficia de um forte investimento de capital<sup>2</sup>. Como BRYNJOLFSSON (2016) observou, a inteligência artificial pode permitir eliminar virtualmente a pobreza global, reduzir

---

<sup>1</sup> Neste sentido, vejam-se, entre outros: GOODFELLOW; BENGIO; COURVILLE; *et al.*, 2018, pp. 1140–1144.

<sup>2</sup> Neste sentido, veja-se: SHOHAM; RAYMOND; BRYNJOLFSSON, *et al.*, 2018.



---

massivamente as doenças e fornecer melhor educação a quase todas as pessoas no planeta<sup>3</sup>.

Embora a inteligência artificial possa ser inteiramente software, por outro lado, os robôs são máquinas físicas que se movem. Os robôs estão sujeitos a impactos físicos, em geral, por recurso a sensores, exercendo uma determinada força física, por meio de “atuadores”, como por exemplo uma “mão” mecânica. Conseqüentemente, automóveis ou aviões autômatos entram no conceito de robôs, e apenas uma muito pequena parcela dos robôs é “humanoide”. Ademais, alguns robôs usam a inteligência artificial, enquanto outros não a utilizam.

Ainda assim, é justo considerar que, embora os sistemas robóticos causem mais preocupações ao público em geral, os sistemas de inteligência artificial são mais propensos a ter um impacto maior na humanidade. Além disso, os mencionados sistemas de inteligência artificial ou robótica para um conjunto restrito de tarefas têm menos probabilidade de causar novos problemas do que os sistemas mais flexíveis e autônomos. A robótica e a inteligência podem, portanto, ser vistas como abrangendo dois conjuntos de sistemas sobrepostos: sistemas que são apenas de inteligência artificial, sistemas que são apenas de robótica e sistemas que são mistos.

### 3 OS DESAFIOS E OS PROBLEMAS DA INTELIGÊNCIA ARTIFICIAL

Existe um conjunto de questões éticas relativas uso humano de sistemas de inteligência artificial e da robótica que podem ser mais ou menos autônomos, sobre o qual no âmbito do presente artigo não cumpre analisar individualmente cada um deles, excetuando aquele que nos ocupa, relativamente aos agentes morais artificiais.

Nesse contexto, partimos da premissa-base que as tecnologias sempre farão com que alguns usos sejam mais fáceis e, portanto, mais frequentes, e que dificultarão

---

<sup>3</sup> No mesmo sentido, veja-se: ANDERSON; RAINIE; LUCHSINGER, 2018.



---

outros usos. O *design* com que são concebidos tem, portanto, na linha de pensamento de HOUKES e VERMAAS (2010), uma importante relevância ética para o seu uso<sup>4</sup>.

Consequentemente, o Direito tem a responsabilidade de regular o uso responsável, sobre determinados padrões éticos a utilização da inteligência artificial, bem como estruturar um *design* responsável neste campo. A análise centrada no uso não pressupõe quais as abordagens éticas que são mais adequadas para lidar com essas questões, visto que eles podem muito bem ser a ética da virtude (VALLOR, 2017), em vez de consequencialistas (FLORIDI; COWLS; ET AL., 2018, pp. 689-707).

#### 4 OS DESAFIOS E PROBLEMAS ÉTICOS DOS AGENTES MORAIS ARTIFICIAIS

Se considerarmos que a ética da máquina diz respeito a agentes morais, em algum sentido substancial, esses agentes podem ser chamados de “agentes morais artificiais”, tendo direitos e responsabilidades. No entanto, a discussão sobre entidades artificiais desafia uma série de noções comuns em ética e pode ser muito útil entendê-las abstraindo-nos do caso humano (MISSELHORN, 2020).

Na doutrina vários autores usam o termo “agente moral artificial” em um sentido menos exigente, utilizando o termo de “agente” em engenharia de software, caso em que questões de responsabilidade e direitos não surgirão<sup>5</sup>. Neste contexto, MOOR (2006) distingue quatro tipos de agentes de máquina: agentes de impacto, agentes éticos implícitos, agentes éticos explícitos e agentes éticos completos.

Existem várias formas de alcançar agentes éticos “explícitos” ou “completos”, ou seja, através da sua programação, por meio do “desenvolvimento” da própria ética ou do da moralidade desenvolvida com plena inteligência (MOOR, 2006)<sup>6</sup>. Os agentes programados às vezes não são considerados agentes “completos” porque são

---

<sup>4</sup> Neste sentido, veja-se ainda: VERBEEK, 2011.

<sup>5</sup> Neste sentido, veja-se: ALLEN; VARNER; ZINSER, 2000, pp. 251–261.

<sup>6</sup> Neste mesmo sentido, veja-se ainda: ALLEN; SMIT; WALLACH, 2005, pp. 149–155.



---

“competentes sem compreensão”, assim como os neurônios de um cérebro (DENNETT, 2017)<sup>7</sup>.

Em algumas discussões, a noção de “paciente moral” desempenha um papel importante. Efetivamente, os agentes éticos têm responsabilidades, enquanto os pacientes éticos têm direitos, porque o dano para eles é importante. Parece claro que algumas entidades são pacientes sem serem agentes. Por outro lado, é normalmente entendido que todos os agentes também serão pacientes. Por regra, é o facto de ser “pessoa” que faz de uma entidade um agente responsável, isto é, alguém que pode ter deveres e ser objeto de preocupações éticas. Tal pessoalidade é tipicamente uma noção associada à consciência fenomenal, à intenção e ao livre-arbítrio.

Alguns autores, como TORRANCE sugerem que “*artificial (or machine) ethics could be defined as designing machines that do things that, when done by humans, are indicative of the possession of ‘ethical status’ in those humans*” (2011, p. 116)<sup>8</sup>, o que o mesmo considera ser “*ethical productivity and ethical receptivity*” (2011, p. 117)<sup>9</sup>.

Neste contexto, pode-se analisar duas distintas questões: a responsabilidade dos robôs e o direito dos robôs.

No que concerne às responsabilidades dos robôs, há um amplo consenso de que a responsabilidade é um conceito básico que deve ser mantido em face das novas tecnologias. Porém, a questão no caso dos robôs é como isso pode ser realizado e como a responsabilidade pode ser atribuída. Aqui coloca-se a questão de saber se os robôs ao agirem, se eles próprios serão responsáveis pelas suas ações, ou se a distribuição do risco deve ter precedência sobre as discussões da responsabilidade.

Em geral, os efeitos das decisões ou ações baseadas na inteligência artificial são frequentemente o resultado de inúmeras interações entre muitos atores, incluindo

---

<sup>7</sup> Neste mesmo sentido, veja-se ainda: HAKLI; MAKELA, 2019, pp. 259–275.

<sup>8</sup> Tradução do autor: a ética artificial (ou da máquina) pode ser definida como projetar máquinas que fazem coisas que, quando feitas por humanos, são indicativas da posse de 'status ético' nesses humanos

<sup>9</sup> Tradução do autor: produtividade ética e receptividade ética.



---

designers, usuários, software e hardware. Ora, daqui resulta a responsabilidade distribuída (TADDEO; FLORIDI, 2018, p. 751)<sup>10</sup>.

Como essa distribuição pode ocorrer não é um problema específico da inteligência artificial, mas ganha particular urgência nesse contexto (NYHOLM, 2018, pp. 1201-1219).

Quanto à segunda questão, respeitante aos direitos dos robôs, alguns autores indicaram que devem ser considerados seriamente se os robôs atuais devem ter direitos (GUNKEL, 2018, pp. 97-99). Esta posição parece basear-se amplamente na crítica dos oponentes e na observação empírica de que os robôs e outras não-pessoas às vezes são tratadas como possuidores de direitos.

Nesse sentido, foi proposto que, se nos relacionamos com os robôs como se eles tivessem direitos, então seria aconselhável não pesquisar se eles têm "realmente" esses direitos. Isto levanta a questão de saber até onde esse antirrealismo ou quase-realismo pode ir, e o que significa então dizer que os "robôs têm direitos", numa abordagem centrada no ser humano (GERDES, 2016, pp. 274-279).

Com efeito, já diferentemente na doutrina encontramos a linha de pensamento de BRYSON (2010, pp. 63-74), que insistiu que os robôs não deveriam gozar de direitos.

Por outro lado, há uma questão totalmente distinta, que consiste em saber se os robôs, ou outros sistemas de inteligência artificial, devem receber o status de "entidades legais" ou "pessoas jurídicas", uma vez que os Estados, empresas ou organizações também são "entidades", ou seja, também eles podem ter direitos e deveres legais.

O Parlamento Europeu considerou atribuir tal *status* aos robôs para lidar com a responsabilidade civil (BERTOLINI; AIELLO, 2018, p. 130-140), mas não a

---

<sup>10</sup> Neste sentido, afirmam que "*The effects of decisions or actions based on AI are often the result of countless interactions among many actors, including designers, developers, users, software, and hardware.... With distributed agency comes distributed responsibility.*" – cfr. TADDEO; FLORIDI, 2018, p. 751





---

responsabilidade criminal, que é reservada para pessoas físicas. Também seria possível atribuir apenas um determinado subconjunto de direitos e deveres aos robôs.

Outros autores sustentam que “*such legislative action would be morally unnecessary and legally troublesome*” (BRYSON; DIAMANTIS; GRANT, 2017, p. 273)<sup>11</sup>, porque não serviria aos interesses dos seres humanos.

Ainda outros distintos autores advogam que as razões para desenvolver robôs com direitos, ou pacientes morais artificiais, no futuro são eticamente duvidosas (VAN WYNSBERGHE; ROBBINS, 2019, pp. 719-735).

Há assim uma preocupação dos investigadores da "consciência artificial", sobre se seria ético criar tal consciência, uma vez que criá-la implicaria, presumivelmente, obrigações éticas, o que leva alguns deles a sustentar uma “*moratorium on synthetic phenomenology*” (BENTLEY; OLLE HAGGSTROM; METZINGER, 2018, pp. 1-40).

## 5 CONSIDERAÇÕES FINAIS

A inteligência artificial é uma área do direito muito singular e ainda muito desconhecida, que levanta imensas questões éticas, nas quais se inclui a referente aos agentes morais artificiais, designadamente quanto à responsabilidade dos agentes morais artificiais e, bem assim, quanto à existência de direitos dos mesmos.

O surgimento destes problemas, que são ao mesmo tempo desafios, surge a uma velocidade extraordinária, que vem conduzindo a mudanças dramáticas, ao passarmos da visão dos anos setenta em que a inteligência artificial seria impossível, para uma ideia que a inteligência artificial pode causar danos aos seres humanos e à própria humanidade como a conhecemos. É, pois, necessário o desenvolvimento da filosofia e da ética da inteligência artificial, porque ela hoje é a realidade e não a imaginação do século passado, havendo um conjunto de questões fundamentais

---

<sup>11</sup> Tradução do autor: tal ação legislativa seria moralmente desnecessária e legalmente problemática.



---

sobre o qual é necessário perceber o que deve ser admitido que a inteligência artificial realize, bem como acautelar os riscos num cenário de longo prazo, do qual nos debruçamos apenas sobre um deles – os agentes morais artificiais.

Em suma, a atual revolução tecnológica, especialmente a decorrente do desenvolvimento da inteligência artificial, desafiam a visão humana da humanidade como a espécie inteligente e dominante na Terra. Isto implica, face às questões levantadas a propósito dos agentes morais artificiais, que teremos que observar de perto os desenvolvimentos tecnológicos e sociais para captar as novas questões desde o início, desenvolver uma análise filosófica e aprender sobre os problemas tradicionais da filosofia.

## REFERÊNCIAS

ALLEN, Colin; SMIT, Iva; WALLACH, Wendell. “*Artificial Morality: Top-down, Bottom-up, and Hybrid Approaches*”. ***Ethics and Information Technology***, 7(3), 2005, pp. 149–155.

ALLEN, Colin; VARNER, Gary; ZINSER, Jason. “*Prolegomena to Any Future Artificial Moral Agent*”. ***Journal of Experimental & Theoretical Artificial Intelligence***, 12(3), 2000, pp. 251–261.

ANDERSON, Janna; RAINIE, Lee; LUCHSINGER, Alex. ***Artificial Intelligence and the Future of Humans***. Washington, DC: Pew Research Center, 2018.

BENTLEY, Peter J.; OLLE HAGGSTROM, Miles Brundage; METZINGER, Thomas. “*Should We Fear Artificial Intelligence? In-Depth Analysis*”. ***European Parliamentary Research Service, Scientific Foresight Unit (STOA)***, March 2018, PE 614.547, pp. 1-40.

BRYNJOLFSSON, Erik; MCAFEE, Andrew. ***The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies***. New York: W. W. Norton, 2016.

BRYSON, Joanna J. “*Robots Should Be Slaves*”. *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues*, Yorick Wilks (ed.). ***Natural Language Processing*** 8, Amsterdam: John Benjamins Publishing Company, 2010, pp. 63–74.



---

BRYSON, Joanna J.; DIAMANTIS, Mihailis E.; GRANT, Thomas D. “Of, for, and by the People: The Legal Lacuna of Synthetic Persons”. *Artificial Intelligence and Law*, 25(3), 2017, pp. 273–291.

CERVANTES, José-Antonio; Et. Al. “Artificial Moral Agents: A Survey of the Current Status”. *Sci Eng. Ethics*, 26, 2020, pp. 501-532.

DENNETT, Daniel C. *From Bacteria to Bach and Back: The Evolution of Minds*. New York: W.W. Norton, 2017.

FLORIDI, Luciano; Et. Al. “AI4People - An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations”. *Minds and Machines*, 28(4), 2018, pp. 689–707.

GERDES, Anne. “The Issue of Moral Consideration in Robot Ethics”. *ACM SIGCAS Computers and Society*, 45(3), 2016, pp. 274–279.

GOODFELLOW, Ian; Et. Al. “A General Reinforcement Learning Algorithm That Masters Chess, Shogi, and Go through Self-Play”. *Science*, 362 (6419), 2018, pp. 1140–1144.

GUNKEL, David J. “The Other Question: Can and Should Robots Have Rights?”. *Ethics and Information Technology*, 20(2), 2018, pp. 87–99.

HAKLI, Raul; MAKELA; Pekka. “Moral Responsibility of Robots and Hybrid Agents”. *The Monist*, 102(2), 2019, pp. 259–275.

HOUKES, Wybo; VERMAAS, Pieter E. *Technical Functions: On the Use and Design of Artefacts*. *Philosophy of Engineering and Technology* 1. Dordrecht: Springer Netherlands, 2010.

MINSKY, Marvin. *The Society of Mind*, New York: Simon & Schuster, 1985.

MISSELHORN, Catrin. “Artificial Systems with Moral Capacities? A Research Design and Its Implementation in a Geriatric Care System”. *Artificial Intelligence*, 278: art. 103179, 2020.

MOOR, James H. “The Nature, Importance, and Difficulty of Machine Ethics”. *IEEE Intelligent Systems*, 21(4), 2006, pp. 18–21.

NYHOLM, Sven. “Attributing Agency to Automated Systems: Reflections on Human–Robot Collaborations and Responsibility-Loci”. *Science and Engineering Ethics*, 24(4), 2018, pp. 1201–1219.

NYHOLM, Sven. “The Ethics of Crashes with Self-Driving Cars: A Roadmap, II”. *Philosophy Compass*, 13(7): e12506, 2018.



---

SHOHAM, Yoav; Et. Al. “*The AI Index 2018 Annual Report*”. **AI Index Steering Committee, Human-Centered AI Initiative**, Stanford University, 2018.

STONE, Peter; Et. Al. “*Artificial Intelligence and Life in 2030*”. **One Hundred Year Study on Artificial Intelligence: Report of the 2015–2016 Study Panel**, Stanford University, Stanford, CA, September 2016.

TADDEO, Mariarosaria; FLORIDI, Luciano. “*How AI Can Be a Force for Good*”. **Science**, 361(6404), 2018, pp. 751-752.

TORRANCE, Steve. “*Machine Ethics and the Idea of a More-Than-Human Moral World*”. **Anderson and Anderson**, 2011, pp. 115-137.

VALLOR, Shannon. **Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting**. Oxford: Oxford University Press, 2017.

VAN WYNSBERGHE, Aimee; ROBBINS, Scott. “*Critiquing the Reasons for Making Artificial Moral Agents*”. **Science and Engineering Ethics**, 25(3), 2019, pp.719–735.

VERBEEK, Peter-Paul. **Moralizing Technology: Understanding and Designing the Morality of Things**. Chicago: University of Chicago Press, 2011.

