



AUTONOMIA E RESPONSABILIDADE MORAL NO CONTEXTO DOS SISTEMAS AUTÔNOMOS *

Autonomy and moral responsibility in the context of autonomous systems

Delmo Mattos da Silva **

Resumo: Este artigo tem como objetivo reavaliar a conexão tradicional entre autonomia e responsabilidade moral, demonstrando sua inadequação para sistemas de IA avançados e contextos humanos complexos. A análise parte das críticas de Oshana à noção idealizada de autodeterminação e da problematização de Matthias sobre a lacuna de responsabilidade para agentes artificiais. A hipótese central é que a solução para os danos causados por sistemas autônomos não reside na tentativa de atribuir culpa individual a agentes não humanos. Em vez disso, propõe-se que a justiça e a segurança exigem o desenvolvimento de estruturas normativas alternativas. Como resultado, o estudo defende a substituição da busca por culpados por um modelo pragmático de governança coletiva. Para tanto, propõe a implementação de mecanismos como corresponsabilidade institucional, seguros obrigatórios e regulação baseada em risco, adaptando as instituições sociais para gerir tecnologias que, embora funcionalmente autônomas, carecem de agência moral.

Palavras-chaves: Autonomia. Responsabilidade. Tecnologia. Liberdade. Controle.

Abstract: This article aims to reassess the traditional connection between autonomy and moral responsibility, demonstrating its inadequacy for advanced AI systems and complex human contexts. The analysis draws from Oshana's critiques of the idealized notion of self-determination and Matthias's problematization of the responsibility gap for artificial agents. The central hypothesis is that the solution

* Artigo recebido em 14.07.2025 e aprovado para publicação em 10.11.2025.

** Doutor em Filosofia pela Universidade Federal do Rio de Janeiro (2008). Professor adjunto do Instituto Tecnológico de Aeronáutica (ITA). ORCID: <https://orcid.org/0000-0002-9074-2192>. E-mail: professordelmo@gmail.com.

for damages caused by autonomous systems does not lie in attempting to attribute individual blame to non-human agents. Instead, it is proposed that justice and safety require the development of alternative normative frameworks. As a result, the study advocates for replacing the search for culprits with a pragmatic model of collective governance. To this end, it proposes the implementation of mechanisms such as institutional co-responsibility, mandatory insurance, and risk-based regulation, adapting social institutions to manage technologies that, while functionally autonomous, lack moral agency.

Key Words: **Autonomy. Responsibility. Technology. Freedom. Control.**

Introdução

No artigo *The Misguided Marriage of Responsibility and Autonomy*, publicado no *The Journal of Ethics* em 2002, Marina Oshana questiona a visão predominante de que a autonomia é uma condição necessária para a responsabilidade. O artigo parte da afirmação de que a relação entre autonomia e responsabilidade é mal interpretada ou aplicada de forma inadequada, induzindo a conclusões equivocadas sobre quando e como os indivíduos devem ser considerados moralmente responsáveis por suas respectivas ações. De acordo com a perspectiva tradicional, segundo Oshana,¹ para que um indivíduo seja considerado moralmente responsável por suas ações, deve agir autonomamente. Isso significa que o indivíduo deve tomar decisões com base em sua própria vontade, ou seja, livre de coerção ou manipulação externa. A autonomia, nesse sentido, é compreendida como um pré-requisito para a responsabilidade, pois implica que o indivíduo teve controle sobre suas escolhas e, portanto, pode ser responsabilizado por elas. No entanto, Oshana questiona essa perspectiva, argumentando que a autonomia não é necessária para que o indivíduo seja responsabilizado pelas suas ações.² Em outros termos: “A agência autônoma não implica necessariamente agência responsável, assim como a agência responsável não pressupõe autonomia plena”.³

De acordo com Oshana, a conexão tradicionalmente assumida entre ser moralmente responsável e ser autônomo é problemática. Enfatiza que muitos filósofos assumem que apenas agentes autônomos podem ser genuinamente responsáveis por suas ações, evidenciando o que ela denomina

¹ OSHANA, M. *The Misguided Marriage of Responsibility and Autonomy*, in *The Journal of Ethics*, v. 6, n. 3, 2002, p. 261-280.

² “A agência autônoma não implica agência responsável, nem agência responsável implica agência autônoma. Em parte, o erro decorre da ambiguidade quanto ao escopo das atribuições de autonomia e de responsabilidade” (OSHANA, *The Misguided Marriage of Responsibility and Autonomy*, p. 262).

³ OSHANA, *The Misguided Marriage of Responsibility and Autonomy*, p. 266.

de casamento equivocado entre esses dois conceitos. Neste contexto, fica claro que Oshana ressalta que a responsabilidade pode existir mesmo em situações em que a autonomia do indivíduo é limitada ou comprometida. Como exemplo, Oshana afirma que em casos em que as escolhas de um indivíduo são influenciadas por fatores externos, como pressões sociais, condições econômicas ou estruturas de poder, ainda pode ser apropriado atribuir responsabilidade moral, mesmo que sua autonomia não seja plena. Portanto, Oshana destaca que a autonomia, embora importante, não é o único fator relevante para determinar a responsabilidade. Em vez disso, Oshana propõe que a responsabilidade deve ser avaliada com base em uma análise mais ampla do contexto em que as ações ocorrem, considerando fatores como intenção, consciência das consequências e as circunstâncias sociais que influenciam as escolhas individuais.⁴

Matthias, em seu artigo *The responsibility gap: ascribing responsibility for the actions of learning automata* (2004), aborda a questão da responsabilidade em um contexto completamente diferente de Oshana, ou seja, o das ações realizadas por sistemas automatizados e máquinas de aprendizagem. Enquanto Oshana problematiza a relação entre autonomia e responsabilidade em seres humanos, Matthias discute o desafio de atribuir responsabilidade quando as ações são realizadas por entidades não humanas, como sistemas de inteligência artificial (IA) ou “autômatos de aprendizagem” (*learning automata*) em um contexto de autonomia. Matthias argumenta que, com o avanço da tecnologia, especialmente no campo da IA, surge um hiato de responsabilidade (*responsibility gap*). Esse hiato ocorre porque os sistemas automatizados, especialmente aqueles capazes de aprender e tomar decisões de forma autônoma, operam sem qualquer tipo de controle porque suas ações não podem ser completamente previstas ou controladas por seus criadores ou operadores humanos.⁵ Como resultado, torna-se difícil atribuir responsabilidade quando esses sistemas causam danos ou tomam decisões problemáticas ao agir de forma imprevisível. Na perspectiva de Matthias, o desafio consiste em encontrar uma forma de atribuir responsabilidade moral a esses sistemas, uma vez que a autonomia do próprio sistema de IA pode ser questionada.

Em seu argumento, Matthias propõe que a responsabilidade poderia ser atribuída aos desenvolvedores ou aos usuários desses sistemas, dependen-

⁴ “Mais precisamente, a autonomia como condição global não é uma condição de responsabilidade. Interessa-me principalmente as noções globais de autonomia e responsabilidade”. (OSHANA, *The Misguided Marriage of Responsibility and Autonomy*, p. 266).

⁵ Para Matthias, o desafio é encontrar uma maneira de atribuir responsabilidade moral a esses sistemas, uma vez que a autonomia do próprio sistema de IA pode ser questionada. Em seu argumento, ele propõe que a responsabilidade moral poderia ser atribuída aos desenvolvedores ou aos usuários desses sistemas, dependendo das circunstâncias em que a autonomia das máquinas se manifesta (MATTHIAS, A. *The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata*, in *Ethics and Information Technology*, v. 6, n. 3, 2004, p. 175-183).

do das circunstâncias em que a autonomia das máquinas se manifesta.⁶ Assim, Matthias defende a necessidade de novos modelos éticos e legais, como responsabilidade compartilhada ou baseada em risco, para preencher essa lacuna, adaptando-se à natureza dinâmica das tecnologias autônomas sem sacrificar a justiça na atribuição de culpa ou reparação. Essa proposta implica uma reconfiguração das noções tradicionais de culpa e imputabilidade, uma vez que os sistemas autônomos, ao tomarem decisões de forma independente e muitas vezes imprevisível, desafiam a identificação de um agente claramente responsável por suas ações. Essa abordagem visa justamente evitar tanto a impunidade quanto a imputação injusta, reconhecendo que nem sempre é possível traçar uma linha causal direta entre o erro do sistema e a ação humana individual.

Ao problematizar as posições de Oshana e Matthias, pode-se evidenciar um ponto de tensão entre as suas abordagens. Ambos os autores problematizam a questão da responsabilidade e da autonomia, mas em contextos distintos: enquanto Oshana se concentra nas ações humanas cujas autonomias podem ser parcialmente ou totalmente comprometidas, Matthias discute acerca entidades não-humanas que operam de maneira autônoma ou semiautônoma. A partir dessa tensão argumentativa, surge uma indagação: até que ponto o conceito de autonomia, tal como é tradicionalmente entendido, pode ser estendido ou aplicado a sistemas não humanos, como as IA, e até que ponto a ideia de autonomia continua sendo uma ferramenta útil para avaliar a responsabilidade moral em situações humanas mais complexas?

Com base nesta indagação, o presente artigo propõe uma análise comparativa das concepções de Oshana e Matthias sobre autonomia e responsabilidade, com o objetivo de debater como essas concepções podem ser reconciliadas ou ajustadas em face da evolução tecnológica e das novas formas de agência, como as apresentadas pelos sistemas de IA. Neste contexto, a reflexão sobre o hiato de responsabilidade de Matthias pode ser compreendida como um contraponto relevante à concepção de autonomia de Oshana, a fim de repensar como a sociedade deve lidar com os danos causados por agentes autônomos para compreender como gerir e distribuir os custos desses danos de forma justa e eficaz. Tal perspectiva, orienta a análise para um modelo de responsabilização pragmático e institucional, que reconhece a impossibilidade de atribuir responsabilidade moral clássica a agentes artificiais e propõe, em seu lugar, a reformulação das estruturas normativas para lidar com suas consequências sociais reais.

⁶ A complexidade dos algoritmos envolvidos em sistemas autônomos pode obscurecer a linha entre a responsabilidade do desenvolvedor e a responsabilidade do operador. Enquanto os desenvolvedores criam e ajustam o sistema inicial, o comportamento do sistema pode mudar com o tempo devido ao aprendizado contínuo. Esse fenômeno desafia o conceito tradicional de responsabilidade, que se baseia na premissa de que os comportamentos do sistema são inteiramente controláveis e previsíveis pelos criadores.

Diante disso, primeiramente, examina a relação entre autonomia e responsabilidade na perspectiva de Oshana. O argumento de Oshana é discutido, propondo uma compreensão mais sutil que considera como fatores externos, como limitações sociais e econômicas, podem restringir a autonomia sem necessariamente absolver os indivíduos de responsabilidade. Para tanto, debate-se o pressuposto de que a responsabilidade moral e autonomia são conceitos distintos que não necessariamente coincidem. Ao dissociá-los, Oshana demonstra que indivíduos podem ser moralmente responsabilizados mesmo em ausência de autonomia plena, assim como agentes autônomos podem, em certos contextos, escapar à imputação de responsabilidade.

Em seguida, examina-se a ascensão de sistemas automatizados, como IA e máquinas autônomas, introduzindo novas complexidades ao conceito de responsabilidade. Aborda-se a forma como Matthias introduz o conceito de lacuna de responsabilidade, destacando a dificuldade de atribuir responsabilidades quando sistemas autônomos agem de forma imprevisível ou fora do controle humano. Com base nisso, debate-se como os sistemas autônomos, capazes de aprender e agir de maneiras imprevisíveis, criam uma lacuna de responsabilidade, tornando injusta ou impossível a imputação de culpa pelos meios convencionais quando ocorrem danos.

Por último, problematiza a lacuna de responsabilidade por meio de uma análise comparativa entre Oshana e Matthias. Para isso, expõe-se a distinção conceitual de Oshana entre a racionalidade tênue ou instrumental da autonomia e a racionalidade densa ou normativa da responsabilidade para avaliar a natureza da agência dos autômatos de aprendizagem. A partir dessa análise, buscar-se-á demonstrar que, embora esses sistemas de IA possam ser compreendidos como agentes autônomos, eles carecem da competência moral para serem considerados responsáveis.

Desse modo, evidencia-se que a chamada lacuna de responsabilidade identificada por Matthias não é simplesmente um vazio a ser preenchido, mas o estabelecimento de uma nova classe de agentes autônomos e não responsáveis. Nestes termos, discute-se a possibilidade de mudança de propósito da busca por uma imputação de culpa moral individual para a necessidade de desenvolver mecanismos institucionais e sociais, como regulação, seguros e corresponsabilidade operacional, para gerir os riscos e danos inevitáveis causados por esses novos agentes artificiais.

1. A relação entre autonomia e responsabilidade moral

Oshana parte da crítica acerca da perspectiva tradicional da relação entre autonomia e responsabilidade. A discussão ética tradicional sustenta que a autonomia é um elemento essencial para responsabilizar um agente,

argumentando que a responsabilidade moral só pode ser atribuída quando um indivíduo age de forma autônoma, ou seja, quando tem a liberdade de tomar decisões independentemente de coerções externas. Não obstante, essa perspectiva enseja uma indagação por parte de Oshana: a autonomia é realmente um pré-requisito indispensável para a responsabilidade moral? A crítica de Oshana a essa visão tradicional propõe uma análise mais ampla e contextualizada da inter-relação entre autonomia e responsabilidade, questionando a premissa de que a autonomia é uma condição necessária para atribuir responsabilidade.

De acordo com Oshana, a relação entre autonomia e responsabilidade é mais complexa do que a abordagem tradicional propõe. O argumento que Oshana sustenta baseia-se na premissa que, embora a autonomia seja um fator relevante, ela torna-se indispensável para a responsabilidade, pois, em muitas situações, a autonomia de um indivíduo pode ser limitada por fatores externos sem necessariamente isentá-lo de responsabilidade. Desse modo, a crítica de Oshana à visão tradicional baseia-se na constatação de que as escolhas dos indivíduos são frequentemente “moldadas por circunstâncias sociais, econômicas e outras que podem restringir sua autonomia”.⁷

Oshana exemplifica que, em contextos marcados por desigualdades sociais, onde condições econômicas precárias e estruturas de poder opressivas limitam as opções disponíveis, a liberdade de escolha pode ser significativamente reduzida. No entanto, segundo Oshana, isso não implica que os indivíduos devam ser isentos de responsabilidade por suas ações. Portanto, no âmbito da perspectiva tradicional que, ao evidenciar a autonomia como critério central para a responsabilidade, ignora as complexidades contextuais que envolvem a tomada de decisões. Oshana enfatiza a necessidade de reconhecer que fatores como pressão social, condições econômicas e outras limitações externas desempenham um papel crucial na compreensão das escolhas humanas, sem necessariamente resultar em isenção de responsabilidade.⁸

Como exemplo, Oshana evidencia os casos em que indivíduos podem ser moralmente responsáveis por ações realizadas sob coação ou em situações de autonomia diminuída, desde que mantenham um grau mínimo de agência. Desse modo, a responsabilidade não deve ser entendida como um conceito binário, isto é, presente apenas quando há plena autonomia e ausente quando há restrições, mas sim como um espectro que se baseia na capacidade do agente de responder a razões e de compreender as consequências de seus atos, mesmo em condições adversas.

⁷ OSHANA, The Misguided Marriage of Responsibility and Autonomy, p. 265.

⁸ “Há uma tendência a confundir o estado global de ser autônomo com a condição local de agir autonomamente ou exibir autonomia em relação a algum ato ou decisão” (OSHANA, The Misguided Marriage of Responsibility and Autonomy, p. 277).

Não obstante, essa perspectiva desafia a noção tradicional de que a autonomia, entendida como liberdade absoluta de escolha, é o único critério válido para determinar a responsabilidade moral. Assim, a responsabilidade poderia ser atribuída com base não apenas na liberdade de escolha, mas também na consciência das consequências e das intenções que motivaram a ação. Essa abordagem permite uma compreensão mais justa e realista das ações humanas, reconhecendo que a responsabilidade pode persistir mesmo quando a autonomia é parcialmente comprometida. Não obstante, ao desvincular a responsabilidade de uma concepção idealizada de autonomia, Oshana oferece um modelo mais inclusivo, que leva em conta as desigualdades estruturais e as limitações impostas pelo contexto social, sem abandonar a noção de que os indivíduos devem prestar contas por suas escolhas, dentro dos limites de sua agência.

De acordo com Oshana: “Um indivíduo autônomo precisa ser capaz de refletir sobre seus fins e os meios para alcançá-los, mas não é obrigatório que ela seja sensível a razões morais objetivas”.⁹ Em outras palavras, um indivíduo pode ser autônomo mesmo que seus motivos, valores e razões sejam moralmente questionáveis, desde que essa pessoa consiga avaliá-los criticamente e afirmá-los como seus, sem que essa decisão resulte de manipulação, coerção ou ausência de alternativas genuínas.

Segundo Oshana, a autonomia é a capacidade de uma pessoa conduzir sua própria vida com base em razões, valores e motivações que ela reconhece como seus. Portanto, ser autônomo não significa apenas estar livre de interferências externas, mas também de limitações internas que possam impedir a autodireção. Essa autonomia depende de quatro condições fundamentais: controle epistêmico e comportamental, acesso a opções alternativas, independência procedural e independência de conteúdo, ou seja, agir com base em valores que o indivíduo considera realmente seus.

A fim de ratificar esse argumento, Oshana ressalta que a autonomia exige um tipo específico de racionalidade, denominada de racionalidade instrumental. Isso significa que um indivíduo precisa ser capaz de deliberar sobre seus objetivos e os meios para alcançá-los, sem necessariamente estar comprometida com razões morais objetivas. Assim, um indivíduo pode ser autônomo mesmo adotando escolhas moralmente questionáveis, desde que essas escolhas sejam feitas de forma consciente, livre e reflexiva. Portanto, na perspectiva de Oshana, isso distingue a autonomia da agência moralmente responsável, que requer sensibilidade a razões éticas universais.

Partindo deste argumento Oshana afirma: “a autonomia individual e a agência responsável são conceitos logicamente distintos”.¹⁰ Desse modo, a auto-

⁹ OSHANA, *The Misguided Marriage of Responsibility and Autonomy*, p. 267.

¹⁰ OSHANA, *The Misguided Marriage of Responsibility and Autonomy*, p. 267.

nomia envolve autogoverno, ou seja, o controle epistêmico-comportamental, acesso a alternativas e independência procedural e de conteúdo, exigindo uma racionalidade instrumental para deliberar sobre fins e meios, mas não necessariamente sensibilidade a razões morais objetivas. No entanto, a agência responsável demanda a capacidade de reconhecer e responder a normas morais objetivas, inserindo o agente em uma comunidade moral como interlocutor competente, o que requer uma racionalidade sensível a razões independentes de desejos particulares.

Tal distinção demonstra que autonomia não implica responsabilidade, e vice-versa. Um indivíduo pode ser autônomo sem ser responsável (como um ditador que ignora razões morais) ou responsável sem plena autonomia (como alguém em uma sociedade opressiva, com opções limitadas, mas sensível a imperativos éticos). Sendo assim, segundo Oshana, a visão tradicional que vincula autonomia a responsabilidade é insuficiente, pois as condições para cada conceito são independentes e contextualmente contingentes.

2. Responsabilidade em sistemas automatizados

Na perspectiva de Onwuzurike, Chikodi e Odhiambo, “a atribuição de responsabilidade por ações e suas consequências é um elemento fundamental das sociedades humanas”.¹¹ De acordo com a tradição, quando se trata de máquinas, a responsabilidade, seja moral ou legal, recai sobre o operador ou o fabricante. Essa atribuição baseia-se num princípio essencial, ou seja, o controle humano em relação as máquinas. Dessa forma, o operador é responsável se a máquina funcionar conforme as especificações ou utilizá-la de forma inadequada ou negligente. Contudo, o fabricante é responsável se a máquina apresentar uma falha de projeto ou fabricação, não operando como previsto. Neste caso, Matthias ressalta: “o manual de operação, explícito ou implícito, simboliza a transferência de controle e, conseqüentemente, de responsabilidade, do criador para o usuário, delineando o comportamento esperado da máquina”.¹²

Não obstante, o advento de máquinas autônomas capazes de aprender, ou seja, os autômatos de aprendizagem baseados em tecnologias como redes neurais, algoritmos genéticos e arquiteturas de agentes introduz uma complexidade sem precedentes no âmbito de responsabilidades. Como ressalta Matthias, “as máquinas podem modificar seu próprio comportamento de maneiras

¹¹ ONWUZURIKE, Maduabuchukwu Augustine; CHIKODI, Augustine Rita; ODHIAMBO, Brian Otieno. The Ethics of Artificial Intelligence and Autonomous Systems: Review, in *International Journal of Innovative Science and Research Technology*, [S. l.], v. 9, n. 8, ago. 2024, p. 761, p. 23.

¹² MATTHIAS, The Responsibility Gap, p. 176.

que nem seus criadores nem seus operadores podem prever ou controlar totalmente”.¹³ Essa capacidade de aprendizado autônomo descaracteriza a relação entre controle humano e ações das máquinas, gerando o que Matthias denomina lacuna de responsabilidade, isto é, as situações em que ocorrem danos ou resultados indesejados, mas a atribuição de responsabilidade pelos meios tradicionais torna-se injusta ou impossível de ser imputada.

A questão central que emerge é como conciliar a autonomia crescente das máquinas com os princípios de responsabilidade. Conforme afirma Lin, “Tradicionalmente, a responsabilidade pressupõe três elementos fundamentais: controle sobre a ação, capacidade de prever suas consequências e liberdade para escolher entre alternativas. Os sistemas autônomos de aprendizado desafiam esses três pilares”.¹⁴ Primeiro, o controle é diluído à medida que a máquina toma decisões baseadas em padrões aprendidos, não em regras pré-definidas. Segundo, a previsibilidade é comprometida pela natureza adaptativa e muitas vezes opaca dos algoritmos de aprendizado. Terceiro, a liberdade de escolha humana é reduzida quando as decisões são delegadas a sistemas que operam além da compreensão e intervenção em tempo real.

Conforme foi explicitado, a autonomia da máquina desafia diretamente a condição de controle e previsibilidade que fundamenta a responsabilidade humana. Se um programador não pode prever o comportamento específico que emergirá do processo de aprendizado em todas as situações possíveis, e se um operador não tem como intervir ou sequer compreender completamente as razões por trás de uma decisão autônoma da máquina, como podemos, de forma justa, imputar-lhes a responsabilidade moral ou legal pelas consequências? Conforme atesta Matthias, atribuir responsabilidade sem controle ou previsibilidade contraria princípios fundamentais de justiça, assemelhando-se a culpar alguém por um evento natural imprevisível.

Matthias argumenta que essa lacuna de responsabilidade não é um problema técnico efêmero, mas uma característica intrínseca dos sistemas autônomos de aprendizado. Desse modo, quanto mais complexos e adaptáveis forem esses sistemas, maior será sua capacidade de operar em ambientes dinâmicos e imprevisíveis e, paradoxalmente, maior será a lacuna entre sua autonomia operacional e a possibilidade de atribuição de responsabilidade humana. Não obstante, essa contradição coloca a sociedade diante de um dilema: abrir mão dos benefícios dessas tecnologias ou reformular os conceitos de responsabilidade para acomodar a nova realidade das máquinas autônomas?

¹³ MATTHIAS, The Responsibility Gap, p. 175.

¹⁴ LIN, P. Why Ethics Matters for Autonomous Cars. In: MAURER, M.; GERDES, J. C.; LENZ, B.; WINNER, H. (Eds.). *Autonomes Fahren: Technische, rechtliche und gesellschaftliche Aspekte*. Berlin: Springer, 2015, p. 81.

Segundo Matthias, quanto maior a autonomia, isto é, a capacidade da máquina de aprender, adaptar-se e tomar decisões sem supervisão ou intervenção humana direta, mais frágeis se tornam os fundamentos tradicionais da responsabilidade, baseados no controle, na previsibilidade e na agência humana. Diante disso, torna-se evidente que a própria noção de autonomia nas máquinas, tal como descrita por Matthias, está no cerne do surgimento da lacuna de responsabilidade. Portanto, a autonomia não é apenas uma característica técnica, mas um fator que redefine os contornos da responsabilidade moral, desafiando profundamente os modelos vigentes e exigindo uma reflexão urgente sobre como sociedades humanas devem responder a agentes artificiais que, embora não sejam sujeitos morais no sentido clássico, operam de forma suficientemente independente para gerar consequências significativas em relações as suas respectivas ações.

A autonomia, entendida como a capacidade de agir independentemente a partir da aprendizagem e da adaptação contínua, desloca o eixo da responsabilidade das relações tradicionais para um novo campo ainda em construção. Se, como assevera Matthias, a responsabilidade exige controle, previsibilidade e agência ou seja, a capacidade de agir com intenção e consciência das consequências, então os sistemas autônomos que operam fora desses parâmetros minam os fundamentos sobre os quais se construiu historicamente a imputação moral e legal. Tal fato conduz a repensar o que significa responsabilizar em contextos onde os agentes não são nem plenamente humanos nem inteiramente programáveis.

Essa redefinição não ocorre sem consequências, como afirma Matthias, a lacuna de responsabilidade que surge com os autômatos de aprendizagem não é apenas uma anomalia moral ou legal passageira, mas uma transformação estrutural da relação entre agentes humanos e sistemas tecnológicos. À medida que delegamos decisões complexas a agentes que operam fora da previsibilidade e da transparência exigidas pelos modelos clássicos de atribuição de responsabilidade, comprometemos a integridade dos próprios sistemas de justiça e confiança social. A responsabilidade, nesse novo cenário, deve ser pensada menos como imputação a indivíduos isolados e mais como um regime coletivo de supervisão, prevenção e resposta compartilhada, capaz de sustentar as promessas da tecnologia sem abdicar dos princípios de equidade e justiça que estruturam a convivência social.

Essa transferência de controle para a própria máquina e seu ambiente operacional é o núcleo do problema. A sociedade passa a depender de sistemas cuja supervisão humana detalhada é impossível, seja por complexidade, velocidade ou vantagem informacional da máquina, como em sistemas de controle de tráfego aéreo ou diagnóstico médico em áreas remotas. Desistir dessas tecnologias não é uma opção realista, visto que

elas são essenciais para resolver problemas complexos da atualidade, como a exploração planetária ou a gestão de infraestruturas críticas. Portanto, se desejamos evitar a injustiça de culpar humanos por ações de máquinas sobre as quais não possuem controle suficiente, torna-se imperativo encontrar novas formas, tanto na prática moral quanto na legislação, para preencher a lacuna de responsabilidade.

3. Para além da lacuna de responsabilidade

Na perspectiva de Oshana, a agência responsável depende de uma racionalidade denominada de densa, que é uma competência normativa. Trata-se da capacidade de um agente reconhecer e ser sensível a razões para ação que são objetivas e fundadas em normas morais, independentemente de seus desejos ou valores pessoais. Neste sentido, ser um agente responsável é, segundo Oshana, ser um interlocutor competente em uma comunidade moral. Não obstante, a agência autônoma requer apenas uma racionalidade tênue ou instrumental, pois esta forma de racionalidade envolve a capacidade de deliberar sobre os fins e os meios para alcançá-los, com base em valores e motivos que são os seus próprios, sem que esses valores precisem ser moralmente aceitáveis. Nestes termos, um agente pode ser autônomo mesmo que seus objetivos sejam moralmente questionáveis, desde que suas ações resultem de seus próprios processos deliberativos, livres de coerção e manipulação.

Aplicando essa distinção ao problema da lacuna evidenciado por Matthias, os “autômatos de aprendizagem” constituem candidatos ideais para a agência autônoma no sentido tênue proposto por Oshana. Matthias ressalta que máquinas cujas regras de ação “não são corrigidas durante o processo de produção, mas podem ser alteradas durante a operação da máquina, pela própria máquina”.¹⁵ Tais máquinas aprendem com a “experiência genuína adquirida autonomamente” e tomam decisões sem intervenção humana direta.¹⁶ Esse comportamento, que otimiza um resultado com base na interação com o ambiente, exemplifica perfeitamente a racionalidade instrumental proposta por Oshana. Sendo assim, a máquina estabelece e age sobre razões e valores próprios através do aprendizado, satisfazendo as condições necessárias descritas por Oshana para a autonomia.

Contudo, a questão de reconhecer se essas máquinas podem ser agentes responsáveis é muito mais complexa. As máquinas necessitam da racionalidade densa, ou seja, precisam ser capazes não apenas de processar

¹⁵ MATTHIAS, The Responsibility Gap, p. 177.

¹⁶ MATTHIAS, The Responsibility Gap, p. 178.

dados para atingir um objetivo, mas também de “reconhecer normas morais avaliativas” e “responder a essas avaliações”.¹⁷ Não obstante, os exemplos de Matthias demonstram o contrário, ou seja, o autômato opera sem qualquer consciência das implicações morais ou sociais de suas ações, assim como pode deixar um executivo esperando e causar um prejuízo financeiro ou diagnosticar erroneamente um paciente, não por malícia, mas por uma lógica puramente instrumental que é cega para o contexto moral. Nesse sentido, o autômato de aprendizagem se assemelha ao exemplo do Hitler¹ descrito por Oshana: “uma entidade que pode ser racional em seus próprios termos, mas é incapaz de ser movida por razões que outros oferecem sobre a moralidade de suas ações”.¹⁸ Oshana classifica tal ser como um “idiota moral” ou “sociopata”, alguém que está “além dos limites” da agência moralmente responsável.

Com efeito, o problema não requer uma ausência de responsabilidade, mas a presença de uma nova classe de agentes, ou seja, os autônomos, porém não responsáveis. Sendo assim, a lacuna não surge porque não existe mais responsabilidade, mas porque o agente que causa o dano a máquina não possui a capacidade para a racionalidade densa necessária para ser um candidato adequado à atribuição de responsabilidade moral. Como afirma Matthias, “o projetista perde progressivamente o controle, transferindo-o para a própria máquina e seu ambiente”.¹⁹ Nestes termos, quando a máquina autônoma causa um dano, culpar o programador parece injusto, pois ele não tinha o controle necessário; culpar a máquina, por sua vez, é inadequado, pois ela carece de sensibilidade moral.

A tensão argumentativa entre os autores, portanto, expõe uma questão mais profunda, isto é, o desafio não é apenas preencher uma lacuna referente a responsabilidade, mas decidir como uma sociedade deve lidar com os atos de agentes autônomos, mas moralmente cegos. Desse modo, a indagação se desloca de quem pode ser responsabilizado para como devemos gerir e atribuir os custos dos danos inevitáveis causados por uma população crescente de agentes artificiais autônomos, mas não responsáveis? Essa nova formulação oferece um caminho mais produtivo para analisar o complexo cenário tecnológico que Matthias descreve, pois desloca o objetivo da busca por uma imputação moral clássica que pressupõe agência plena, intenção e autonomia moral para um modelo de responsabilização pragmático e institucional. Em vez de insistir em adaptar os agentes artificiais às molduras normativas humanas, essa abordagem sugere que repensemos as estruturas normativas para abarcar agentes cuja atuação produz efeitos sociais relevantes, mas cuja natureza ontológica os exclui da responsabilidade moral tradicional.

¹⁷ OSHANA, M. *Personal Autonomy in Society*. Aldershot: Routledge, 2000, p. 12.

¹⁸ OSHANA, The Misguided Marriage of Responsibility and Autonomy, p. 180.

¹⁹ MATTHIAS, The Responsibility Gap, p. 180.

Contudo, essa reconfiguração normativa não elimina a responsabilidade, mas a redistribui por meio de mecanismos institucionais, como seguros obrigatórios, taxaço por risco tecnológico e auditorias técnicas contínuas. Tal proposta, inspirada na análise de Matthias, reconhece que os sistemas de IA, particularmente os autômatos de aprendizagem, operam segundo um tipo de racionalidade que escapa ao modelo tradicional de imputação moral baseado em intenção e previsibilidade. A responsabilidade, nesse novo cenário, deixa de ser uma função direta da agência e passa a ser uma função da gestão social dos riscos como nas situações descritas por Matthias, em que nenhum agente humano pode ser responsabilizado diretamente porque o comportamento da máquina decorre de experiência adquirida autonomamente, fora do alcance do controle prévio do programador ou operador.

Essa perspectiva exige que seja revisitada, segundo exalta Oshana, as bases filosóficas da imputação moral, reconhecendo que a agência moral não pode ser presumida com base na autonomia funcional. De acordo com Dignum, “o autômato que deixa um executivo esperando ou diagnostica erroneamente um paciente não é responsável moralmente porque lhe falta a competência normativa”, ou seja, não está inserido na comunidade moral que torna possível a interpelação, o reconhecimento e a responsabilização.²⁰ Entretanto, os danos causados por tais agentes continuam exigindo resposta. Tal afirmação explica o deslocamento do debate para o campo das instituições sociais e jurídicas: como precificar o risco, como prevenir falhas recorrentes, como compensar vítimas. Segundo Oshana, a responsabilidade, nesse modelo, transforma-se em uma rede de corresponsabilidade operacional e preventiva, que envolve desde os projetistas e desenvolvedores até os órgãos reguladores e o próprio Estado.

Nesse sentido, a racionalidade tênue dos sistemas autônomos exige, não sua exclusão do campo da justiça, mas o desenho de uma justiça técnica, capaz de acomodar agentes que não são sujeitos morais, mas cujas ações impactam profundamente os direitos e as expectativas sociais. Goodall observa: “Como em um novo contrato moral, o que está em jogo não é o julgamento da alma da máquina, mas a configuração de garantias coletivas diante de agentes que agem, mas não respondem”.²¹ A racionalidade tênue dos sistemas autônomos exige, não sua exclusão do campo da justiça, mas o desenho de uma justiça técnica, capaz de acomodar agentes que não são sujeitos morais, mas cujas ações impactam profundamente os direitos e as expectativas sociais. Como em um

²⁰ DIGNUM, V. *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Berlin: Springer, 2019, p. 35.

²¹ GOODALL, N. J. Ethical Decision Making during Automated Vehicle Crashes. *Transportation Research Record*, in *Journal of the Transportation Research Board*, v. 2424, 2014, p. 58–65, p.59.

novo contrato moral, o que está em jogo não é o julgamento da alma da máquina, mas a configuração de garantias coletivas diante de agentes que agem, mas não respondem.

A dissociação entre autonomia e responsabilidade torna-se ainda mais nítida ao considerarmos o argumento inverso de Oshana: uma pessoa pode ser um agente responsável mesmo em condições que limitam severamente sua autonomia. Um escravo, por exemplo, carece de autonomia global para direcionar a própria vida, mas pode, ainda assim, ser considerado responsável por atos de coragem ou afeto, pois possui a competência normativa para responder a razões morais. Desse modo, na perspectiva de Sharkey, “O autômato de aprendizagem representa a imagem espelhada desse escravo: ele possui autonomia funcional, a capacidade de agir de forma autodirigida para atingir seus objetivos, mas carece da agência responsável, pois é surdo à conversação moral”.²²

Esta inversão de papéis reforça a urgência da proposta de Matthias, ou seja, a sociedade deve escolher entre não usar tais tecnologias, o que ele considera inviável, ou enfrentar a lacuna de responsabilidade. A solução, portanto, consiste em não tentar forçar a responsabilidade sobre o programador que já não tem controle, nem sobre a máquina que é incapaz de responder moralmente, mas construir um andaime social que absorva e distribua o custo dos erros inevitáveis. Esses erros, como aponta Matthias, são muitas vezes características intrínsecas e necessárias ao processo de aprendizagem da máquina, e não falhas de projeto.²³

O novo contrato moral implica, portanto, que a sociedade aceite os benefícios da automação em troca da implementação de um sistema de governança de risco. Os desenvolvedores e fabricantes não são eximidos de suas obrigações, mas estas se transformam: em vez de uma responsabilidade direta por cada ação imprevisível da máquina, eles assumem o dever de transparência, de diligência no projeto e, crucialmente, de participação em fundos de compensação ou regimes de seguro. A responsabilidade deixa de ser um evento de culpa pós-dano para se tornar um processo contínuo de gestão de risco pré-dano, alinhando a inovação tecnológica com a proteção social e a justiça

Conclusão

Diante da análise crítica das abordagens de Oshana e Matthias, conclui-se que a crescente complexidade dos sistemas autônomos desafia radicalmente os modelos tradicionais de responsabilidade moral, exigindo uma reconceitualização profunda tanto da autonomia quanto da própria noção de agência. Enquanto Oshana desmistifica o vínculo necessário entre au-

tonomia plena e responsabilidade em contextos humanos marcados por desigualdade, Matthias revela que, nos sistemas de IA, a imprevisibilidade e a opacidade funcional geram uma “lacuna de responsabilidade” que as teorias normativas clássicas não conseguem preencher. Assim, torna-se urgente adotar um modelo dinâmico e distribuído de responsabilidade, capaz de integrar agentes humanos e não humanos em um arcabouço ético relacional, que considere o controle, a intenção, as consequências e o contexto social das decisões automatizadas. Tal modelo não apenas amplia nossa compreensão sobre responsabilidade moral, mas também promove um compromisso mais realista, justo e transparente com o desenvolvimento e a regulação de tecnologias emergentes.

Referências

- DIGNUM, V. *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Berlin: Springer, 2019.
- GOODALL, N. J. Ethical Decision Making during Automated Vehicle Crashes. Transportation Research Record, in *Journal of the Transportation Research Board*, v. 2424, 2014, p. 58–65.
- LIN, P. Why Ethics Matters for Autonomous Cars. In: MAURER, M.; GERDES, J. C.; LENZ, B.; WINNER, H. (Eds.). *Autonomes Fahren: Technische, rechtliche und gesellschaftliche Aspekte*. Berlin: Springer, 2015, p. 69–85.
- MATTHIAS, A. The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata, in *Ethics and Information Technology*, v. 6, n. 3, 2004, p. 175-183.
- _____. Robot lies in health care: when is deception morally permissible?, in *Kennedy Institute of Ethics Journal*, v. 25, n. 2, 2015, p. 169-182, 2015.
- OMOHUNDRO, S. Autonomous technology and the greater human good, in *Journal of Experimental & Theoretical Artificial Intelligence*, v. 26, n. 3, 2014, p. 303-315.
- ONWUZURIKE, Maduabuchukwu Augustine; CHIKODI, Augustine Rita; ODHI-AMBO, Brian Otieno. The Ethics of Artificial Intelligence and Autonomous Systems: Review, in *International Journal of Innovative Science and Research Technology*, [S. l.], v. 9, n. 8, ago. 2024 p. 761.

²² SHARKEY, Amanda. Autonomous Weapons Systems, Killer Robots and Human Dignity, in *Ethics and Information Technology*, v. 21, n. 2, 2019, p. 75–87, p. 56.

²³ Matthias argumenta que, no caso de autômatos de aprendizagem, os erros não são apenas possibilidades, mas parte inerente do processo de aprendizado. Como esses sistemas operam em ambientes complexos e incertos, a experimentação e a tentativa-e-erro são essenciais para sua eficácia. Logo, exigir que nunca falhem seria inviabilizar sua capacidade de aprender autonomamente. A responsabilidade, nesse contexto, não pode ser atribuída a um erro específico como se fosse uma falha técnica, mas deve ser entendida como uma responsabilidade sistêmica pela gestão de riscos contínuos.

OSHANA, M. The Misguided Marriage of Responsibility and Autonomy, in *The Journal of Ethics*, v. 6, n. 3, 2002, p. 261-280.

_____. *Personal Autonomy in Society*. Aldershot: Routledge, 2000.

SHARKEY, Amanda. Autonomous Weapons Systems, Killer Robots and Human Dignity, in *Ethics and Information Technology*, v. 21, n. 2, 2019, p. 75–87. doi:10.1007/s10676-018-9494-0.