O PROBLEMA DA ATRIBUIÇÃO DE RESPONSABILIDADE EM AGENTES ARTIFICIAIS

Delmo Mattos[1]

[1]Instituto Tecnológico de Aeronâutica (ITA) – São José dos Campos – SP – Brazil

{delmo@ita.br}

Abstract. *The increasing autonomy of these machines, coupled with their learning capabilities, creates ethical and legal challenges, especially in situations where errors or damages occur. The difficulty in predicting the future behavior of these systems raises questions about who should be held responsible. The aim is to examine the traditional conditions of "control" and "epistemic" responsibility and how they apply to or are complicated by autonomous systems.*

Resumo. *A crescente autonomia dessas máquinas, aliada à sua capacidade de aprendizado, gera desafios éticos e legais, particularmente em situações onde ocorrem erros ou danos. A dificuldade em prever o comportamento futuro desses sistemas levanta questões sobre quem deve ser responsabilizado. Objetiva-se problematizar as condições tradicionais de "controle" e "epistêmica" para atribuição de responsabilidade e como elas se aplicam ou se complicam com sistemas autônomos.*

1. Introdução

As máquinas autônomas são sistemas que podem operar e tomar decisões sem intervenção humana direta. Esses sistemas utilizam uma combinação de sensores, algoritmos e inteligência artificial (IA) para realizar tarefas de forma independente[1]. Esses agentes são programados para considerar as implicações de suas ações e escolher o curso de uma ação que melhor se alinha com um conjunto de normas ou valores éticos predefinidos. Contudo, a ideia de máquinas com autonomia suscita questões sobre a própria natureza da responsabilidade que geralmente atribuímos as máquinas. Se uma máquina autônoma comete um erro de decisão, é difícil determinar se a culpa recai sobre o programador, a empresa que o implementou, ou se a máquina em si pode ser considerada "responsável".

Essa indeterminação levanta questões sobre a responsabilização em um contexto em que a agência é artificial, sobretudo, de ações resultante de um processo autônomo

de aprendizado. Por exemplo, se um veículo autônomo estiver envolvido em um acidente, quem é considerado responsável? O fabricante do veículo, o desenvolvedor do software, o proprietário do carro ou o próprio sistema de inteligência artificial? Essa complexa rede de interações torna difícil atribuir responsabilidade de forma clara e justa. Esses exemplos ilustram a complexidade do "responsibility gap" e mostram como a autonomia e a capacidade de aprendizado dos sistemas automatizados desafiam as noções tradicionais de responsabilidade (Matthias, 2004).

A proposta da apresentação consiste em problematizar a questão da atribuição de responsabilidade em máquinas autônomas, considerando sua capacidade de aprendizado autônomo e a complexidade envolvida na determinação de responsabilidade em casos de decisões morais erradas.

2. Autonomia e máquinas autônomas

Á medida que as máquinas autônomas se tornam mais capacitadas a tomar decisões, a tradicional atribuição de responsabilidade baseada na ação humana direta se torna inadequada. Isso ocorre porque as máquinas autônomas podem tomar decisões e agir de maneiras que não foram previstas pelos seus criadores ou operadores. O fabricante/operador da máquina, em princípio, não é mais capaz de prever o comportamento futuro da máquina, e, portanto, não são suficientemente responsabilizados moralmente ou legalmente por isso. Quando uma máquina é capaz de aprender e evoluir de forma autônoma, prever seu comportamento futuro pode se tornar extremamente difícil.

Isto, prossegue o argumento, implica que tais máquinas sejam responsáveis pelas consequências das suas ações e, subsequentemente, que ninguém mais seja responsável por essas ações. É claro que responsabilizar uma máquina por decisões erradas não é possível. Isso significa que haverá casos em que ninguém será responsável pelos erros cometidos, produzindo assim lacunas de responsabilidade. Segundo Coeckelbergh (2020), “Uma lacuna de responsabilidade ocorre quando uma pessoa ou agente é

[1] São exemplos: Carros autonômos, robôs industriais, sistemas de agricultura de precisão, drones de entrega etc.

responsável por um ato, mas não pode ser responsabilizado, resultando numa situação em que ninguém pode ser responsabilizado."

Uma abordagem eficaz para tratar a questão da atribuição de responsabilidades é começar com as condições de responsabilidade. Desde Aristóteles, identificam-se tradicionalmente duas condições para atribuir responsabilidade por uma ação: a condição de "controle" e a condição "epistêmica". condição de "controle" refere-se ao grau em que o agente tem a capacidade de influenciar ou direcionar a ação. Em outras palavras, para ser responsabilizado, o indivíduo deve ter um nível suficiente de controle sobre suas ações, o que implica que a ação não deve ser meramente acidental ou fora de seu domínio de influência. A condição "epistêmica", por sua vez, diz respeito ao conhecimento e à consciência que o agente tem sobre a ação que está realizando. Para ser considerado responsável, o agente deve ter consciência do que está fazendo e entender as implicações de suas ações (2004).

Estas condições ajudam a determinar a extensão da responsabilidade que se pode atribuir a um indivíduo com base na sua capacidade de controlar e entender suas ações. No entanto, em relação as máquinas autônomas, o comportamento o pode ser influenciado por fatores que os designers ou operadores humanos não previram, tornando difícil responsabilizar qualquer indivíduo no sentido tradicional.

3. Responsabilidade e atribuição

Um exemplo citado por Matthias é o dos veículos autônomos, que podem tomar decisões de vida ou morte em frações de segundo. Se um veículo autônomo se envolver em um acidente fatal, a questão de quem é responsável torna-se complicada. O fabricante do veículo, os programadores do software, os proprietários do carro, ou até mesmo o próprio veículo poderiam, em teoria, ser considerados responsáveis. No entanto, atribuir responsabilidade a uma máquina que simplesmente seguiu um processo de aprendizado algorítmico parece inadequado e, muitas vezes, impossível de justificar em termos tradicionais. Além disso, Matthias (2004) destaca que a lacuna de responsabilidade não se limita a situações onde ocorre dano físico. Ela também se aplica a contextos onde os sistemas autônomos tomam decisões que afetam a vida das pessoas de maneiras menos óbvias, como na seleção de candidatos para um emprego ou na

aprovação de um empréstimo. Se um algoritmo discrimina injustamente um grupo de pessoas com base em padrões aprendidos, mas não intencionais, a questão de quem é responsável por essa discriminação é igualmente nebulosa.

Para abordar o problema da lacuna de responsabilidade, Matthias (2004) enfatiza que pode ser necessário repensar a maneira como a responsabilidade é atribuída em relação a sistemas autônomos. Uma possibilidade é que a responsabilidade seja vista como uma questão coletiva, onde diferentes partes — desenvolvedores, operadores, e até mesmo a sociedade como um todo — compartilham a responsabilidade pelos resultados das ações das máquinas. Essa abordagem reconhece que, em um mundo onde as máquinas desempenham papéis cada vez mais autônomos, a responsabilidade não pode ser atribuída de forma simplista a um único agente.

No entanto, a atribuição coletiva de responsabilidade traz seus próprios desafios. Se a responsabilidade é compartilhada entre muitos, pode ocorrer um fenômeno de diluição da responsabilidade, onde nenhum indivíduo ou grupo sente a obrigação total de garantir que o sistema funcione de maneira segura e ética. Esse problema é exacerbado pela complexidade dos sistemas modernos, onde muitas vezes há múltiplas camadas de decisão e controle, tornando difícil rastrear e atribuir responsabilidade de maneira clara.

Por fim, Matthias (2004) sugere que um diálogo contínuo entre tecnólogos, legisladores, e a sociedade é essencial para abordar essas questões de responsabilidade. A colaboração interdisciplinar pode ajudar a criar soluções equilibradas e eficazes que reconheçam as complexidades dos sistemas autônomos e estabeleçam diretrizes claras para a atribuição de responsabilidade. É através desse diálogo que será possível desenvolver frameworks robustos que não apenas garantam a justiça e a responsabilidade, mas também promovam a inovação e a segurança no avanço das tecnologias autônomas.

References

Matthias, Andreas. (2004) “The responsibility gap: ascribing responsibility for the actions of learning automata”. Ethics and Information Technology, v. 6, n. 3, p. 175-183, set.

Singh, Jatinder and Walden, Ian and Crowcroft, Jon and Bacon, Jean, (2016). Responsibility & Machine Learning: Part of a Process.

Coeckelbergh, M. (2020). AI Ethics. Cambridge, MA and London: MIT Press