



UNIVERSIDADE  
DE ÉVORA

## **Ética Na Era Digital**

### **Responsabilidade Moral e Transparência na Inteligência Artificial**

Miguel Grilo    58387

Colégio Luís António Verney    24/05/2025

**Tabela de Conteúdos**

**Conteúdo**

Tabela de Conteúdos..... i

Introdução ao Problema Ético na Inteligência Artificial..... 1

Fundamentos da Responsabilidade Moral ..... 1

Atribuição de Responsabilidade ..... 2

Complexidade dos Sistemas de Inteligência Artificial ..... 3

Problema da Caixa Negra..... 3

Transparência e Explicabilidade ..... 4

Educação e Desigualdades..... 5

Desafios na Explicabilidade ..... 5

Conclusão ..... 6

Bibliografia ..... 7

## **Introdução ao Problema Ético na Inteligência Artificial**

A Inteligência Artificial está cada vez mais presente no nosso dia a dia e está a mudar a forma como tomamos decisões. Já não se trata apenas de ajudar os humanos, em muitos casos, os sistemas inteligentes estão mesmo a tomar decisões por nós. Isso levanta uma questão ética bastante importante para permitir uma evolução de certo modo mais segura. Quando algo corre mal, quem é que deve ser responsabilizado? Afinal de contas, estamos a falar de máquinas, que apesar de não terem consciência, sentimentos ou intenções, acabam por ter impacto direto na vida das pessoas. Por isso, é essencial pensarmos com cuidado sobre como distribuímos essa responsabilidade. No centro desta questão estão dois conceitos principais, responsabilidade moral e transparência. Só garantindo que estas tecnologias são usadas de forma clara e justa é que podemos ter a certeza de que continuam a servir os nossos valores e não o contrário. Como observado por Luciano Floridi, “Os cinco princípios [...] são apresentados como eixos convergentes de uma ética da IA, sendo a explicabilidade destacada como um princípio essencial, dada a nova forma de agência moral dos sistemas de IA.”

## **Fundamentos da Responsabilidade Moral**

De acordo com Aristóteles, para que alguém possa ser considerado moralmente responsável por uma ação, há duas coisas que têm de estar garantidas, controlo e consciência. O controlo significa que a ação partiu mesmo da pessoa, que ela agiu de forma voluntária. Já a consciência implica que a pessoa sabe o que está a fazer e tem noção das possíveis consequências dessa mesma ação/decisão.

Mark Coeckelbergh (2020: 109-111) reforça esta ideia dizendo “Segundo Aristóteles, esta é a primeira condição para a responsabilidade moral, a chamada condição de controlo: [...] somos responsáveis se soubermos o que estamos a fazer.”.

Se olharmos para isto ao nível da Inteligência Artificial percebemos logo um problema. Por mais avançada que seja, uma IA não tem consciência, ela pode tomar decisões sozinha, mas toma-las com base em algoritmos e dados, não porque refletiu ou entendeu o que está em causa, ou seja, apesar

de parecer que age por si, não podemos dizer que é um agente moral, como Aristóteles exige, pois falta-lhe exatamente aquilo que mais importa, saber o que está a fazer e o seu porquê.

Como afirma Coeckelbergh (2020: 111-112), “As máquinas podem ser agentes, mas não agentes morais, uma vez que lhes falta consciência, livre-arbítrio [...]”.

## **Atribuição de Responsabilidade**

Se a IA não pode ser vista como um agente moral, então quem é que deve ser responsabilizado pelo que ela faz? A resposta mais lógica é que essa responsabilidade continua a ser do humano, mais exatamente das pessoas que criam, desenvolvem, aplicam ou até mesmo das que apenas usam esses sistemas, mas a verdade é que na prática é complica-se e acaba por ser um debate muito mais elaborado.

Como indica Coeckelbergh (2020: 111-112), “Se isto for verdade, a única solução é responsabilizar os humanos pelo que a máquina faz. Assim, os seres humanos delegam o seu poder de ação na máquina, mas mantêm a responsabilidade.”.

A IA é normalmente feita por equipas enormes, com pessoas de várias áreas a trabalhar em conjunto. Depois esses sistemas são usados em contextos muito diferentes, o que torna quase impossível apontar o dedo a uma única pessoa ou empresa quando algo corre mal, e há também casos em que a IA toma decisões tão rápidas e automáticas que os humanos nem sequer têm tempo para intervir.

Tudo isto faz com que a responsabilidade moral acabe por se espalhar entre programadores, gestores, operadores e até os próprios utilizadores da mesma. No fim ninguém sabe bem quem deve responder se houver um erro grave ou uma injustiça e isso levanta um problema sério.

## **Complexidade dos Sistemas de Inteligência Artificial**

Hoje em dia, os sistemas de IA são incrivelmente complexos, apenas um único algoritmo pode ter milhares de parâmetros, milhões de dados e funcionar dentro de uma rede tecnológica com vários intervenientes envolvidos, o caminho que leva a uma decisão final não é nada simples.

Há ainda mais um detalhe importante, a IA aprende com o tempo. À medida que vai recebendo novos dados, o seu comportamento pode mudar, por isso mesmo que os programadores que a criaram percebam como ela funciona no início, os utilizadores finais muitas vezes nem fazem ideia do que se passa por dentro. Esse desfasamento entre quem cria e quem usa torna ainda mais difícil perceber quem deve ser responsabilizado quando algo corre mal.

Como Coeckelbergh (2020: 118) indica, “Mesmo partindo do princípio que os seres humanos têm conhecimentos sobre a IA em geral e sobre o seu código, tal nem sempre corresponde à verdade [...]”.

## **Problema da Caixa Negra**

Um dos conceitos mais discutidos na ética da IA é o da caixa negra. Refere-se à opacidade de certos sistemas de IA, em particular os que utilizam deep learning, cujo funcionamento se torna praticamente incompreensível mesmo para os seus criadores.

Em sistemas mais simples, como árvores de decisão, é possível traçar um caminho lógico entre entrada e saída, no entanto, em modelos complexos como redes neuronais, a decisão é fruto de processos opacos. Esta falta de explicabilidade torna a decisão num mistério. Tal como no gato de Schrodinger, em que o gato está simultaneamente vivo e morto até alguém abrir a caixa, nas IA opacas, a decisão está lá dentro, mas ninguém consegue realmente ver o que aconteceu até ao momento em que a resposta é revelada, e mesmo depois disso continua a ser difícil explicar o que se passou durante o processo.

Coeckelbergh (2020: 116-117) explica que “A forma como a IA chega à sua decisão já não é transparente e os seres humanos não conseguem explicar totalmente a decisão.”.

A existência destas caixas negras torna praticamente impossível explicar porque é que uma IA tomou determinada decisão, o que levanta sérias preocupações éticas, sobretudo em contextos críticos como diagnósticos médicos, por exemplo, onde se encontram vidas em jogo.

## **Transparência e Explicabilidade**

A transparência é uma peça-chave quando se fala de ética na era digital, mas convém esclarecer, transparência não é simplesmente mostrar o código de um sistema, pois não ajuda muito à maioria das pessoas. O que é realmente importante é que os processos de decisão da IA sejam explicáveis de forma simples e acessível, quer para quem percebe de tecnologia, quer para quem não tem formação na área.

Segundo Coeckelbergh (2020: 121), “A transparência e a explicabilidade não estão necessariamente relacionadas com a divulgação do código do software. [...] Trata-se sobretudo de explicar as decisões às pessoas.”.

É aqui que entra o conceito de explicabilidade, no fundo trata-se da capacidade de responder à pergunta que todos queremos fazer quando algo nos afeta, saber o “porquê”. Neste caso será por que é que o sistema tomou esta decisão?

Para pensadores como Floridi, esta ideia é um dos pilares centrais da ética da IA. Transparência e explicabilidade são essenciais para garantir que a tecnologia continua ao nosso serviço e não o contrário.

Além disso, a explicabilidade é o que nos permite perceber se algo correu mal, corrigir falhas e confiar nas decisões que os sistemas estão a tomar, se não conseguirmos perceber como chegou a determinada decisão, tudo começa a parecer aleatório e aí a confiança desaparece, sem confiança a IA perde a sua força.

## **Educação e Desigualdades**

A transparência não chega para resolver o problema da responsabilidade, especialmente se as pessoas não tiverem as ferramentas certas para perceber como funciona a IA, é aí que entra algo este tópico fundamental, a educação.

A maioria das pessoas não tem conhecimentos suficientes para interpretar modelos ou algoritmos e mesmo entre quem trabalha com tecnologia há diferentes níveis de entendimento. Por isso, quando falamos em explicabilidade, ela tem de ser adaptada, sendo mais técnica para quem percebe do assunto e mais acessível para uma pessoa comum, que também tem o direito de entender as decisões que o afetam.

Floridi destaca que “Com a explicabilidade torna-se também possível o apuramento das responsabilidades advindas do impacto das decisões tomadas (levadas a cabo) por sistemas de IA.”.

Ainda outro ponto importante é que esta opacidade pode piorar desigualdades pré-existentes. Se só uma minoria conseguir perceber como os sistemas funcionam e influenciam decisões importantes, estamos a correr o risco de concentrar o poder e deixar muita gente de fora.

## **Desafios na Explicabilidade**

Apesar dos avanços na explicabilidade, ainda há desafios pela frente. Um dos dilemas é que os modelos que têm melhor desempenho costumam ser também os mais difíceis de explicar, ou seja, quanto mais poderosa for a IA, mais complicado é perceber como ela chegou àquela decisão.

Há ainda as barreiras comerciais, onde muitas empresas não querem revelar certos detalhes sobre os seus sistemas por conta da concorrência ou para proteger a sua propriedade intelectual, o que cria um confronto entre o que seria ideal do ponto de vista ético e o que é viável do ponto de vista económico.

Dado este último ponto, Floridi afirma que “É a explicabilidade que permite dar conta do tipo de agência moral que é específico dos sistemas de IA.”.

No fundo, a questão vai além de simplesmente mostrar o código, o que está realmente em causa é saber se conseguimos explicar as decisões tomadas pela Inteligência Artificial de uma forma clara para quem é afetado por elas, é o realmente importa do ponto de vista ético.

## **Conclusão**

No fim das contas, a ética na Inteligência Artificial tem de ser uma responsabilidade partilhada, não é possível colocar toda a responsabilidade às costas de uma só pessoa, construir sistemas que sejam transparentes e responsáveis é um trabalho que envolve toda a gente.

É também importante termos em conta uma coisa bem clara, a tecnologia nunca é neutra, por trás de cada algoritmo há decisões humanas, iniciando aí a questão dos viés da Inteligência Artificial. Por isso, temos o dever de garantir que esses valores batem certo com princípios como a Beneficência, Não Maleficência, Autonomia, Justiça e Explicabilidade, sendo esta última abordada neste trabalho.

No mundo em que vivemos, onde a Inteligência Artificial está cada vez mais presente, a ética não pode ser vista como um extra, mas sim uma condição para que possamos confiar nestas tecnologias e incorporá-las de forma segura e justa nas nossas vidas e sociedade como um todo.

Como conclui Floridi, “Os 5 princípios [...] delimitam um horizonte ético para os sistemas de IA que poderá servir de base teórica para a criação de normativas jurídicas.”



## **Bibliografia**

Milhano, Â. (2025). *Ética na era digital: O problema da responsabilidade e transparência* [Apresentação em PowerPoint]

Milhano, Â. (2025). *Ética na era digital: Os 5 princípios fundamentais da ética da IA segundo Floridi* [Apresentação em PowerPoint]

Coeckelbergh, M. (2020) *AI Ethics*. Massachusetts: The MIT Press.

Floridi, L. (2021). *The Ethics of Artificial Intelligence*. Oxford University Press.