



## RESUMO EXPANDIDO

### MÁQUINAS QUE PENSAM: HAVERÁ ESPAÇO PARA DILEMAS MORAIS NO FUTURO?

**AUTOR:** Andrei Luiz Lodéa

**UNIVERSIDADE:** Universidade de Passo Fundo

**EIXO TEMÁTICO:** Ética, Tecnologia e o Novo Humanismo.

#### INTRODUÇÃO

Já podemos afirmar com certeza, o presente não é mais o mesmo. Podemos ser nostálgicos com o passado, mas não podemos negar a inevitável evolução de desenvolvimento da tecnologia. Meus apontamentos neste trabalho buscarão lançar um primeiro olhar para i) como resolvemos dilemas morais utilizando teorias éticas consolidadas, ii) como esses mesmos dilemas poderão ser pensados, ao menos hipoteticamente, em um futuro tecnológico onde máquinas inteligentes serão responsáveis pelas decisões éticas, iii) e sobre ônus da responsabilização, caso ações desastrosas ocorram pela ação dessas máquinas tecnológicas. O desenvolvimento metodológico do trabalho será investigar o dilema moral conhecido como “*The trolley problem*”, desenvolvido primeiramente pela filósofa Phillipa Foot, e experimentado por inúmeros outros filósofos do campo da ética. Minha proposta é traçar um direcionamento possível sobre a evolução e a complexidade da análise ética, sobre os produtos criados e que fazem parte de nossas decisões presentes e futuras.

#### O DILEMA DO BONDE: AVALIANDO RESPONSABILIDADES E CONSEQUÊNCIAS

Em ética, uma das metodologias para refletir e tornar seus conceitos e seu princípios claros é o uso de dilemas, sejam eles reais ou hipotéticos. Utilizarei o dilema hipotético conhecido como “*The trolley problem*” (O dilema do bonde). O dilema do bonde busca analisar as possíveis decisões tomadas pelas pessoas caso devam escolher entre quem vive e quem morre.

Em uma das formulações do dilema, um bonde desgovernado irá atingir cinco trabalhadores que estão amarrados aos trilhos. Existe uma alavanca que poderá desviar o bonde, mas, neste outro desvio, há uma pessoa amarrada aos trilhos. Você é peça chave do desfecho final do dilema, pois a alavanca está ao seu alcance. O que você faria? Acionaria ou não acionaria a alavanca? Esse é um dos cenários apresentados atualmente à esse dilema. Vamos pensar em um cenário diferente.

Imagine que ao invés de pessoas amarradas aos trilhos sejam robôs inteligentes, os chamaremos de *bio-robôs*, que apresentam sentimentos e emoções próximas aos sentimentos e emoções humanas. Estão amarrados aos trilhos assim como os humanos na situação anterior. Você encontra-se na mesma situação anterior, próximo a alavanca. O que você faria? Qual seria sua decisão? Vamos a um novo cenário.

Em um futuro não tão próximo, esses seres com inteligência artificial e com emoções e sentimentos, já não sendo mais possível percebermos se são robôs ou humanos, encontram-se amarrados aos trilhos. O “bonde” irá atingir 5 desses seres, mas não é mais você que está próximo da alavanca, é um outro ser com inteligência artificial e que é responsável pela tomada de decisão. O que esse *bio-robô* deve fazer? Por acaso temos condições, nessa situação, de dizer o que é o melhor a ser feito?

Esse dilema suscita várias ideias e reflexões éticas, porém destacarei apenas duas: podemos refletir sobre a responsabilidade e as consequências das escolhas de humanos e seres com inteligência artificial; e, quando ainda formos os programadores ou criadores desses *bio-robôs*, e inserirmos suas infinitas linhas de códigos que representarão todas as possibilidades acima descritas, que teoria ética será a escolhida e com base em quais critérios as escolheremos?

No primeiro caso a responsabilidade recai sobre o indivíduo que decidirá entre puxar ou não puxar a alavanca. Em um experimento feito por Hauser e Singer (2006), 90% das pessoas puxariam a alavanca, matando apenas uma pessoa. A explicação é que em situações de pressão agimos de forma intuitiva, visando minimizar as consequências danosas aos indivíduos. Assim, teríamos uma preferência ao utilitarismo dentre as teorias éticas. A mesma análise podemos fazer para os outros dois casos hipotéticos? O nível de responsabilidade diminui, por exemplo, no segundo e terceiro caso? O que determinaria uma mudança da percepção de responsabilidade? No terceiro caso, o *bio-robô*, enfrentaria o mesmo dilema do acionador de alavanca humano? Sua decisão já estaria programada? E se esse *bio-robô* recebesse em suas milhões de linha de código uma programação baseada em uma teoria ética egoísta, o egoísmo ético, por

exemplo. Poderíamos dizer que ele tomou uma decisão egoísta? Essa decisão depende de sua capacidade de pensar?

## CONSIDERAÇÕES FINAIS

A análise desse dilema retrata o incontrolável desenvolvimento de inteligências artificiais e suas possíveis escolhas éticas. Na minha visão, os criadores e desenvolvedores dessas tecnologias devem considerar às inúmeras consequências que seus produtos tecnológicos terão na vida das pessoas. Uma análise ética séria, códigos de conduta profissional sofisticados, políticas de proteção à privacidade dos usuários podem ser algumas das estratégias para vivermos bem em um mundo habitados por máquinas que pensam e que assumiram e assumirão responsabilidades que eram exclusivamente nossas. Precisamos acompanhar essa (re)evolução tecnológica do pensamento.

## REFERÊNCIAS

CHRISTIAN, Brian. *O humano mais humano: o que a inteligência artificial nos ensina sobre a vida*. Trad. Laura Teixeira Motta. São Paulo: Companhia das Letras, 2013.

DAWKINS, Richard. *Deus, um delírio*. Trad. Fernanda Ravagnani. São Paulo: Companhia das Letras, 2007.

HAUSER, Marc; SINGER, Peter. Morality without religion, *Free Inquiry*. v. 26, n. 1, [S. 1.], 2006.

LIN, Patck; BEKEY, George A. *Robot Ethics: The Ethical and Social Implications of Robotics*. Cambridge; Massachusetts; London, England: The MIT Press, 2012.