**Ética em Inteligência Artificial**

É possível concordar sobre a generalização de que a Inteligência Artificial (IA) tem vindo a alterar a sociedade, mesmo quando não falamos de IA acabamos por fazer referência à mesma sob a forma de uma semi-automação, ou de uma qualquer outra tecnologia aplicada.

Tornou-se igualmente urgente observar que os debates teóricos sobre a IA parecem mais simples - no sentido de se verificar uma abundante produção conceptual - do que a praticabilidade das hipóteses que se propõem nos mesmo. Todavia, a interrogação ética em torno das consequências sociais da automação já tem décadas. Pode até dizer-se, a questão da ética, no seu sentido filosófico, é secular, no seu sentido computacional – mais recente. Um dos livros que demonstra a problematização deste assunto é o da Deborah G. Johnson: “Computer Ethics”; saiu nos anos 80 e apresentou considerações diversas sobre a utilização humana da tecnologia e a sua relação com privacidade, o ciberespaço e tantos outros temas que foram também enunciados em artigos prévios, alertando para a necessidade de alguma reflexão face aos desfechos éticos e sociais.

Isto significa que se introduziram várias perguntas face às consequências sociais (ex. perda de emprego) e comportamentais, mas nem sempre se dedicou algum tempo à discussão específica dos dilemas da programação, da própria modelação dos sistemas. Geoffrey Rockwell é um dos nomes que pertence ao fórum nacional no Canadá sobre ética da IA explorando, em conjunto com membros de diversas indústrias, os paradoxos éticos da automação. Por exemplo: Qual das duas opções escolher para a modelação de um carro autónomo: 1. Atingir um pedestre adulto que aparece à frente do carro, ou 2. desviar o carro para um passeio onde se encontra um carrinho com uma criança bebé?

Do fórum mencionado resultou a Declaração Canadá-França sobre Inteligência Artificial <https://international.gc.ca/world-monde/international_relations-relations_internationales/europe/2018-06-07-france_ai-ia_france.aspx?lang=eng> pedindo a formação de um grupo de investigação internacional para o entendimento e partilha de resultados (dentro do G7) de investigações sobre a Inteligência Artificial. Geoffrey Rockwell sinalizou diversos sectores que estão sujeitos a novas formas de operacionalização com técnicas de IA, relembrando a necessidade de discutir temas desde a aprovação de crédito, aos diagnósticos médicos, ou à necessária inserção participativa das mulheres na programação da *pipeline*: a técnica hardware usada nos CPU’s para que seja possível encontrar e executar instruções.

Em algum momento terá de se perguntar: será ético que a modelação conceptual de dados continue a reproduzir estereótipos padronizados?

Recorrendo, a título exemplificativo, à experiência Moral Machine <http://moralmachine.mit.edu/> do MIT Media Lab foram recolhidos dados de 233 países (com resultado publicados na Nature <https://www.nature.com/articles/s41586-018-0637-6>) face às decisões que cada participante assumiria no papel de “Juiz” em dilemas morais aleatórios. Neste caso em concreto, de um veículo autónomo, os pedestres assumiam quase sempre as categorias de “mulher/ homem” e/ou “criança/idos@”. É necessário entender que priorização da vida face a um acidente inevitável não só é diferente de cultura para cultura, como é impossível de ser padronizada face a uma decisão singular. Não obstante, perante a inevitabilidade da necessária programação é indispensável incluir o maior número de perspetivas possíveis, sendo que este aspeto terá de passar pela implementação de políticas que assentam na observação de experiências como esta. Porém, é preciso dizer que os carros autónomos não têm a potencialidade de certas repressões que podem surgir face a algumas aplicações da IA na saúde, educação ou igualdade.

É certo que as aplicações de IA, nos diferentes contextos, precisam de um grande volume de informação para um melhor reconhecimento e/ ou resposta, mas isto também significa que se as representações não forem alteradas, inclusive as que estão a ser programadas, em nada irão contribuir para a diversidade social - apenas e somente para a multiplicação futura de discriminação por um lado, e estigmatização por outro.

Fei-Fei Li é co-diretora do Stanford Human-Centered AI (HAI) Institute <http://hai.stanford.edu/> e coordena equipas que trabalham, entre outros aspetos práticos, a integração segura de sistemas autónomos na sociedade. Ou seja, trabalham os algoritmos, os dados, os modelos computacionais e as interações humanas para que, no futuro, um robô não venha a reforçar uma desigualdade social que, por sua vez, vai resultar no sofrimento, na morte, e por consequência: em perdas económicas. Além do mais, há equipas, como poderá ver-se através do trabalho da Londa Schiebinger no Genderer Innovations <https://genderedinnovations.stanford.edu/what-is-gendered-innovations.html> que procuram facultar às/aos cientistas e engenheiros/as métodos práticos que permitam entender a tecnologia no contexto social. Sem este entendimento não será possível falar de Ética em Inteligência Artificial, na medida em que falar de uma integração segura de sistemas autónomos na sociedade é, também e necessariamente, falar em diversidade plural.

Lia Raquel Neves

Ciência na Imprensa Regional – Ciência Viva