**Inteligência artificial**

A mais recente série Westworld - no ambiente futurista desenhado por Jonathan Nolan e Lisa Joy (baseada no filme de Michael Crichton de 1973) - levou-nos a um complexo parque temático com tecnologia de ponta baseado na temática do ‘Velho Oeste’, cujo propósito é vincular uma fantasia o mais realista possível onde, para clientes capazes de pagar essa experiência (chamados “Convidados”), não existe nenhuma consequência ou retribuição por parte dos androides (chamados “Anfitriões”) para qualquer desejo o mais egoísta, sadista ou depravado por parte dos Convidados. Esta premissa é explorada na narrativa que se desenvolve ao longo da série, explorando exaustivamente a tecnologia para lá da ação, e onde verificamos a anulação das esferas das consequências, das ações e da moralidade através da exploração dos limites nas interações entre humanos (Convidados) *vs.* androides (Anfitriões).

Mas mais fundamental, esta série reintroduz-nos a questão de fundo da inteligência artificial, a saber, se serão as máquinas capazes de pensar, de ter consciência, e, inevitavelmente, como definir “máquina” por um lado, e “pensamento” por outro lado.

Por exemplo, a personagem de Dolores (Evan Rachel Wood) ao reproduzir uma fala do livro Alice no País das Maravilhas de Lewis Carroll: “Questiono-me se terei mudado durante a noite. Deixa-me pensar. Era a mesma quando acordei esta manhã?”. Esta passagem sugere-nos: i) a possibilidade de um androide armazenar memórias; ii) a necessidade de enunciar a questão de se ‘pode um robô ter consciência?’. Dizer isto é, inevitavelmente, ter de referir que a própria atribuição de consciência a algo inanimado assenta na superação do Teste Turing (TT) apresentado por Alan Turing no artigo *Computing Machinery and Intelligence* (de modo lato, podia dizer-se sobre a superação do TT: se um programa de computador conseguir convencer um humano de que está a falar com outro humano, então ele é capaz de pensar).

Há, contudo, objeções à noção de inteligência artificial definida pela superação do TT que devem ser tidas em conta, nomeadamente, o argumento descrito por John Searle em *Minds, brains, and programs* sobre a intencionalidade nos seres humanos e animais (através da sintaxe e semântica), através da experiência de pensamento da Sala Chinesa. Nesta experiência de pensamento, Searle defende que um programa de computador, por si só, não é capaz de estados mentais genuínos. A ideia da Sala Chinesa é a seguinte: se eu estiver fechada numa sala com caixas que têm caracteres chineses e tiver um livro em português que descreve o programa para falar chinês, quando um nativo em chinês fora da sala me enviar perguntas (por debaixo da porta), apenas conseguirei consultar o livro que descreve o programa para falar chinês e enviar alguns caracteres como resposta que julgo ser apropriada a esse nativo (sendo a conclusão de Searle que, por mais que possa programar um computador, se ele não entende chinês não pode simular esse conhecimento).

Assim sendo, podemos dizer que os símbolos formais e as suas conjugações, por si só, não são suficientes para os conteúdos mentais, pois dependem de um agente externo para atribuição de semântica. Logo, se é a mente que tem a capacidade de atribuição semântica, a máquina ficaria apenas na conjugação sintática.

Esta simplificação serve de mote para pensar que um sistema, descrito desta maneira, não pode aprender sobre o seu próprio sistema, e o cruzamento de dados talvez não deva ser confundido, em si mesmo, com o processo de aprendizagem. Além, também, de nos levar a pensar sobre se a capacidade de imitar as respostas humanas é, de facto, um sinal de inteligência (na variação de estados intencionais semânticos de Searle), ou se estamos apenas num jogo de imitação? E esta questão enquadra a questão fundamental: um robot é capaz de fazer escolhas éticas, ou é programado – dentro de um padrão moral específico – para responder em concordância com esse mesmo padrão?

Lia Raquel Neves (Cientista Social)

Ciência na Imprensa Regional – Ciência Viva