Invariavelmente, e como demonstrado em vários romances de Asimov, as imperfeições, lacunas e ambiguidades embutidas em suas leis da robótica frequentemente resultavam em comportamentos robóticos contraintuitivos e estranhos. As leis são vagas, por exemplo, ao falhar em definir e distinguir, devidamente um humano de um robô. Além disso, uma inteligência artificial (IA) com inteligência super-humana seria pressionado a não descobrir como acessar e revisar seu próprio código. Porém o próprio Asimov acreditava que “Sim, as três leis são o único caminho no qual seres humanos racionais podem lidar com robôs, ou qualquer outra coisa”.

É apenas uma questão de tempo antes que a IA estoure para além de todas as capacidades humanas. Assustadoramente, caso a IA for mal programada ou for ambivalente em relação às necessidades humanas, catástrofes podem ocorrer. Portanto, é necessário garantir que a IA seja segura.

O consenso sobre ética de máquina, no que diz respeito as três leis da robótica, é de que elas são uma base insatisfatória. Apesar de serem leis famosas, elas não são realmente utilizadas para guiar ou servir de apoio para pesquisas de segurança de IA, ou mesmo ética de máquina.

O objetivo das três leis é falhar de maneiras interessantes, originando estórias interessantes. Tais leis são instrutivas em ensinar-nos que qualquer tentativa de legislar éticas em termos de regras específicas é fadado a desmoronar e possuir várias lacunas. Ou seja, tais leis dependem da mente que está as interpretando.

Definir um conjunto de preceitos éticos, como parte central da ética de máquinas, é provavelmente desesperançoso caso as máquinas em questão sejam flexíveis inteligências artificiais gerais (em inglês AGI). Se tal inteligência generalista for criada para possuir um intuitivo e adaptável senso de ética, então, preceitos éticos servirão apenas de guia básico para tal inteligência aplicar sua própria visão de ética.

Poucos pesquisadores de AGI acreditam que seria possível engenhar sistemas de AGI que pudessem garantir total segurança. Presume-se que quando AGI em fase inicial forem criadas, será possível realizar estudos e experimentos que nos guiarão muito mais a respeito de ética de máquina. Porém¸ por enquanto, teorizar sobre ética de AGI é muito difícil, pois não possuímos nenhuma boa teoria de ética e nem uma boa teoria sobre AGI.

As leis da robótica criadas por Asimov são consideradas de grande complexidade, visto que não é fácil definir a uma máquina¸ o que é um ser humano, e, muito menos, quais são seus valores éticos. Pois nem mesmo entre os humanos há um consenso de certo e errado, já que tais interpretações éticas e morais podem variar de pessoa para pessoa.

Outro ponto desfavorável às leis de Asimov implicam definir o significado de “permitir que um ser humano sofra algum mal”. O que exatamente isso implica? Exemplos da atualidade poderiam ser relacionados com os carros autônomos. Atualmente, por exemplo, caso um carro inteligente tenha que decidir entre atropelar cinco crianças e potencialmente matá-las, ou desviar o carro e fazê-lo bater em uma árvore e provavelmente matar o passageiro, a primeira opção provavelmente será aceita como padrão. Por outro lado, considerando-se o mal maior, e a lei zero de Asimov, o carro provavelmente escolheria a segunda opção, visto que matar uma pessoa é mais vantajoso do que matar cinco pessoas, protegendo assim a humanidade. Portanto, há vários casos em que a ambiguidade pode surgir.

A primeira opção para a solução da ética em IA seria de programar preceitos éticos nas IA. As leis de Asimov deveriam ser muito mais específicas caso a intenção seja programá-las. Além disso, seriam inúmeros preceitos éticos levados à linguagem de programação. E tais conceitos éticos são amplamente desconhecidos e diversificados entre os próprios humanos. Portanto, tal alternativa é ,aparentemente, destinada ao fracasso.

Outra alternativa seria possuir um sistema AGI capaz de aprender por si só a partir de dados fornecidos e não com a programação bruta e rígida. Com isso, tal inteligência seria capaz de interpretar o mundo da nossa forma, tentando respeitar os nossos ambivalentes conceitos de certo e errado. Contudo, essa IA seria suscetível aos mesmos erros que nós humanos cometemos.

Uma outra opção possivelmente melhor diz respeito a aprender com a AGI criada. Já que tal AGI poderia ser capaz de analisar de forma mais profunda, rápida e eficaz o próprio conceito de ética da raça humana, e chegar a uma conclusão mais precisa do que precisamos para a convivência com uma superinteligência. Porém, caso sejamos capazes de criar uma superinteligência, devemos ter cuidado, visto que humanos não dialogam com gorilas.

RERREFÊNCIA

http://io9.gizmodo.com/why-asimovs-three-laws-of-robotics-cant-protect-us-1553665410