

Entrevista em 31 Outubro 2021 de Sofia Teixeira, DN/JN,
a Luís Moniz Pereira, Professor Catedrático Emérito da UNL

Apelando à sua capacidade de síntese - apesar da complexidade do assunto - eu diria que pode apontar cada resposta para 10 -15 linhas. Sendo um artigo para um público geral, temos de tentar focar nas questões e ideias mais importantes, se possível recorrendo a exemplos que permitam uma melhor compreensão.

Aqui vão as perguntas:

-Considerando o 'Carrocel das máquinas éticas' que apresenta, mencionado as seis principais abordagens a considerar neste assunto, gostaria de começar por lhe perguntar que hoje, técnica e cientificamente falando, é possível dotar as máquinas de premissas/construtos morais?

Sim, na investigação científica. Eu tenho explorado a dotação de capacidades morais em programas de computador desde há 15 anos. Publiquei o primeiro artigo sobre o assunto em 2007, e desde 2016 4 livros, o último dos quais em 2020, cuja edição portuguesa (há uma edição em inglês no mesmo ano) se intitula: "Máquinas Éticas – Da Moral da Máquina à Maquinaria da Moral", na UNL.FCT Editorial. Pode encomendar-se aqui: <https://novafcteditorial.pt/shop/outros-horizontes/maquinas-eticas-da-moral-da-maquina-a-maquinaria-moral/>

Essa investigação incide sobre diferentes facetas morais que, individualmente e em combinação, conseguimos programar em computador. Envolve dois domínios. O do cognitivo, isto é, das capacidades mentais necessárias para deliberar sobre questões morais. E o domínio populacional, isto é, como funcionam e evoluem os comportamentos morais no seio de uma população de agentes dotados com capacidades mentais de moralidade: Como sejam, por exemplo, reconhecer as intenções de outros, assumir compromissos para com outros, pedir desculpa por falhas, desculpar ou não, reconhecer culpas próprias e corrigir comportamentos anteriores independentemente de manifestar a culpa. Os leitores interessados poderão seguir estas investigações através de publicações e palestras constantes da minha página pessoal, em: <https://userweb.fct.unl.pt/~lmp/publications/Biblio.html>

-O Professor deixa muito claro que há urgência na investigação e na necessidade de conclusões implementáveis no imediato. A si, considerando o cenário atual, o que o preocupa mais, de concreto, no imediato?

Entidades como a UNESCO, a OCDE, a Comissão Europeia, a ONU, e muitas outras, têm aprimorado extensas recomendações sobre a "Ética da Inteligência Artificial". Por exemplo, a da UNESCO, em que participei, contem 134 artigos, e vai à aprovação na reunião geral de 9-24 de Novembro 2021 em Paris. Tais tipos de recomendações contêm não só preocupações no imediato, que incluem a estruturação, desde já na UE, de organismos nacionais supervisores colaborando entre si, como incluem ainda medidas proactivas de ética, de segurança, de acesso ao uso e de justiça social, respeitantes ao desenvolvimento e comercialização das aplicações da Inteligência Artificial (IA). Elas fazem já falta, pois as aplicações têm sido implementadas sem obediência a quaisquer regras legais e standards de controlo específicos. Outras tecnologias, como as dos aviões, comboios, carros, navios, construção, electrodomésticos, armas, etc. são alvo duma especificidade. Pela sua novidade, a IA ainda as não tem. No entanto, a IA tem uma enorme horizontalidade applicativa e uma grande velocidade de desenvolvimento, o que torna a sua regulação imediata ainda mais necessária. Com o cuidado de não regulamentar em excesso a fase inicial de exploração criativa. Contudo, o desejo de chegar primeiro ao mercado induz a tentação de poupar recursos e tempo, dando menos atenção à segurança e à ética. Assim, a regulamentação deve ser mais apurada nas fases finais do desenvolvimento applicativo.

-É na cada vez maior autonomia das máquinas que surgem os maiores problemas éticos. Hoje, a maioria funciona ainda com ‘supervisão’ humana, mas tudo indica que, no futuro, a tendência seja dar-lhes capacidade de decisão não supervisionada por humanos. Que critérios acha que devem estar definidos e ser cumpridos antes de o podermos fazer com segurança?

Para além dos critérios nas regulamentações gerais que referi acima (e são muitos!), em fase de aprovação e que serão adoptados em recomendações, normas e leis internacionais, a rever periodicamente, existirão necessariamente as normas, standards e testes, em domínios específicos de aplicação, como por exemplo nos da saúde, dos veículos sem condutor e drones, ou no das armas mortíferas. Interessa ainda que a segurança previna o uso intrusivo e indevido por hackers.

Um método clássico de proceder, dito “sand-box”, é o de permitir o seu uso experimental, em zonas bem delimitadas, para melhor se estudarem os critérios e testes desejáveis antes da aprovação generalizada. Uma questão que também se coloca é a de como máquinas com autonomia, mas de diferentes fabricantes e portanto com diferente software, deverão colaborar entre si – se não forem do inimigo – por exemplo, os robots envolvidos na segurança de propriedades. Existe ainda a questão de o comportamento em grupo não ser previsível. Aquele que resultará, por exemplo, de um grande número de drones com o objectivo de vigiar uma fronteira. Como vão interagir em grupo contra um grupo de imigrantes intrusos com iniciativa?

- Considerando o problema do alinhamento ético: a ética das máquinas tem de estar alinhada com os nossos valores éticos. Para isso, temos de os conhecer, descrever e enunciar. A questão é: já sabemos fazê-lo? Ou seja, considera que, como sociedade, há uma concordância definida quanto a estes valores ou também esse é um trabalho que está por fazer?

Não, não o sabemos fazer. Nem sequer enunciar de modo preciso as nossas escalas de valores, as quais variam culturalmente, mesmo em vocabulário. Muito menos ainda de forma a programá-las, pois as próprias linguagens de programação não contêm ainda os conceitos e construções para tal. O caso é bem diferente se programarmos uma bem conhecida operabilidade de segurança: a de um comboio ou de um navio. Mesmo aí atente-se no caso do Boeing-737M, em que a incúria levou à realização de um software que se sobrepunha à vontade dos pilotos, e caíram por isso dois aviões. Os valores são também sociais, nomeadamente quanto ao desemprego adicional previsivelmente provocado pela IA por entidades como a McKinsey, a Pew Research, ou a OCDE, as quais apontam para um adicional médio de 14% em 2030. A IA, com a sua grande horizontalidade de aplicações, já é por si só causa actual de um maior aumento do hiato em cada país entre ricos e pobres, e do hiato entre norte e sul. Cabe a nós estudar e prevenir o impacte específico em Portugal. A riqueza produzida pelos avanços da IA deve ser distribuída por todos os humanos, de modo justo, e diminuir a sua servitude em horas de trabalho. As máquinas, elas sim, devem permanecer escravas da nossa civilização. É indispensável, a tempo e horas, formular e implementar um novo contracto social, a fim de se evitar uma revolução sem precedentes, com consequências imprevisíveis.