

Data: 02.02.2019

Título: "E muito difícil introduzir regras morais na inteligência artificial"

Pub: **Expresso**

**QuickCom**  
comunicação integrada

Tipo: Jornal Nacional Semanal

Secção: Nacional

Pág: 26

## CIÊNCIA

**Arlindo Oliveira** Presidente do Instituto Superior Técnico

# “É muito difícil introduzir regras morais na inteligência artificial”



Arlindo Oliveira: “Estou convencido de que a privacidade é um bem que se vai tornar mais escasso”

Área: 1005cm² / 77%

Tiragem: 123.400

FOTO

Cores: 4 Cores

ID: 6370047

Textos VIRGÍLIO AZEVEDO

Foto TIAGO MIRANDA

Mais de 200 cientistas e académicos de vários ramos do conhecimento encheram o salão nobre do Instituto Superior Técnico (IST) nesta quarta-feira, em Lisboa, quando o presidente do IST, professor e investigador nas áreas dos algoritmos, aprendizagem automática, bioinformática e neuroengenharia apresentou o seu último livro, "Inteligência Artificial". Neste ensaio, Arlindo Oliveira antecipa que "as inteligências artificiais não serão apenas nossas parceiras na construção da sociedade do futuro". Serão, "de certa forma, os nossos descendentes".

**■ O seu livro acaba por ser mais uma história sobre a inteligência. Sentiu necessidade de contar esta história para chegar à inteligência artificial?**

**■** Achei que era a melhor maneira de evitar uma visão de curto prazo, enquadrar a inteligência artificial na questão mais vasta da inteligência. Porque é que apareceu, o que é agora e como é que a inteligência artificial a pode substituir.

**■ Considera-se um otimista em relação à evolução tecnológica?**

**■** Sim, ou melhor, acho que sou um realista e que essa é uma visão otimista, porque partilho bastante a opinião de conhecidos autores de que a ciência e a tecnologia têm melhorado bastante as condições de vida da Humanidade e vão continuar a fazê-lo, se tudo correr bem.

**■ E as questões relacionadas com o mundo digital que põem em causa a liberdade e a privacidade das pessoas? Aí a tecnologia não está a melhorar a vida...**

**■** Sim, aí a visão de Yuval Harari, o autor israelita dos *best-sellers* "Sapiens" e "Homo Deus", é muito negativista. Ele chama-lhe dataísmo, a dependência dos dados, em que as pessoas são vistas quase só como uma fonte de dados e de valor económico. Esta visão tem alguma razão de ser, mas não estou muito convencido que seja tão negativa, porque se fosse havia muito mais gente a fugir disso. Na verdade, as pessoas preocupam-se pontualmente sobre estas questões mas não deixam de colocar os seus dados, de usar as redes sociais, de pôr informação pública.

**■ A privacidade vai acabar?**

**■** Estou convencido de que é um bem

que se vai tornar mais escasso, mas as pessoas não parecem valorizar a privacidade tanto como isso, uma vez que estão disponíveis para divulgar informação pessoal. Quando põem informação sua no Facebook ou vão ao Google fazer buscas, basicamente estão a disponibilizar uma parte da sua informação pessoal. A legislação terá provavelmente de impedir que esta quebra de privacidade se torne perigosa ou negativa, mas a sociedade está a habituar-se a viver num ambiente menos privado, e isso também é verdade do ponto de vista legislativo. Mas não estou convencido de que cheguemos a visões distópicas do futuro. Talvez vivamos em sociedades mais transparentes do que há uns anos. Acho que é essencialmente positivo, mas outras pessoas acharão o contrário.

**■ Na inteligência artificial, há um ramo de investigação relacionado com as questões éticas e a moral computacional. É o caminho certo para resolver os problemas que têm surgido?**

**■** Há muitos especialistas que têm falado sobre a importância de ter regras éticas e morais muito claras, o que é uma visão muito apropriada à visão simbólica da inteligência artificial, que foi dominante durante várias dezenas de anos, mas que acabou por ser um pouco ultrapassada. O que se veio a perceber é que o mundo real tem muita incerteza, e que essas visões puramente simbólicas não funcionam geralmente muito bem nesse mundo real, nem que seja para conduzir um automóvel. Com a visão estatística que neste momento impera e cujo máximo expoente é a chamada aprendizagem profunda das redes neuronais — o *deep learning* —, acho que é razoavelmente difícil pôr nesta metodologia regras explícitas, como por exemplo, as Três Leis da Robótica de Isaac Azimov. E, portanto, há aqui a quadratura do círculo. Os sistemas que já usamos para guiar carros, auxiliar juízes em julgamentos e juristas na análise de contratos, ou selecionar pessoas para empregos, tudo isto é estatístico. Por isso, é extremamente difícil introduzir aqui regras morais que estão no espírito das pessoas que trabalharam em inteligência artificial simbólica.

**■ Mas por que razão é tão difícil?**

**■** Porque estes sistemas são inerentemente incertos e estatísticos. Não

há sítio nestes sistemas para escrever e impor regras. Estes sistemas aprendem. Por exemplo, nos sistemas que selecionam uma pessoa para um emprego é-lhes dada uma longa história de pessoas que foram selecionadas, os seus currículos e características, aquelas em que as coisas correram bem e correram mal, em que foram promovidas e despedidas. A partir daqui, os sistemas aprendem a selecionar as pessoas certas, mas neste processo não há nenhum sítio onde se possa por uma regra do tipo "não pode selecionar com base na raça", porque o histórico se calhar indica justamente que as pessoas foram selecionadas com base na raça. É preciso haver um esforço grande em termos de investigação e desenvolvimento, de modo a compaginar estes sistemas estatísticos com o tipo de regras de alto nível, simbólicas, que alguns especialistas defendem. E não sabemos fazer isto, é preciso descobrir maneiras de o fazer, pelo menos parcialmente. É uma área até relativamente rica, porque está relacionada com outra questão: perceber como a aprendizagem profunda e outros sistemas deste tipo aprendem. Será que quando aprendem a conduzir bem um carro vão conduzir bem em todas as situações?

**■ Há questões muito práticas que se põem precisamente na inteligência artificial nos carros autónomos.**

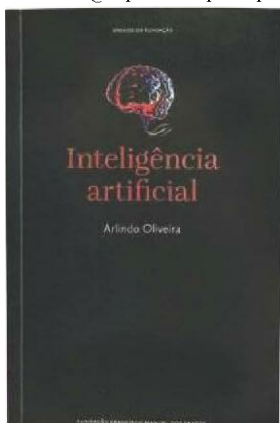
**■** Não é só na condução autónoma, mas talvez esta seja a mais visível para muita gente. Por exemplo, ao treinarmos um sistema para selecionar candidatos a um emprego, podemos descobrir que o sistema nunca seleciona candidatos de olhos azuis e cabelo preto. É uma coisa estranha mas pode acontecer. E, portanto, há cada vez mais esforço para tentar fazer com que estes sistemas sejam explicáveis, que haja uma explicação dos critérios que estão a usar. Esta é uma área importante, garantirmos que os sistemas não só têm de fazer uma tarefa correta, mas as regras que aprendem para executar essa tarefa são as que nós queremos. Ou pelo menos são boas.

**■ É muito difícil definir legislação ou regras para estes sistemas?**

**■** Não. Acho que eventualmente é possível fazer legislação, impondo que nenhum sistema pode discriminar com base no que quer que seja.

O que é difícil é passar da legislação à prática. Hoje já é proibido discriminar com base na raça, em Portugal e em muitos outros países. Mas se esta legislação está a ser cumprida, não sabemos. As empresas são opacas na maneira como contratam. É claro que se houver uma empresa que nunca contratou uma mulher, não será difícil provar que está a discriminar com base no sexo.

vazevedo@expresso.impresa.pt



**ENSAIO** O que é a inteligência artificial, qual é a sua relação com a inteligência humana e a dimensão do seu impacto social e económico, são algumas das questões abordadas no novo livro de Arlindo Oliveira, editado pela Fundação Francisco Manuel dos Santos

## Guerra

## tecnológica pode provocar rejeição?

A corrida atual que mobiliza grandes centros de investigação e empresas pela conquista da supremacia tecnológica na inteligência artificial (IA) "pode ter consequências negativas muito sérias, em especial sempre que os procedimentos éticos e de segurança forem subestimados ou mesmo ignorados, levando potencialmente à rejeição da IA pelo público em geral". Este é um dos alertas deixados pelos investigadores Luís Moniz Pereira (Universidade Nova), The Anh Han (Universidade de Teesside, Reino Unido) e Tom Lenaerts (Universidade Livre de Bruxelas), num artigo científico apresentado esta semana na "Conferência sobre Inteligência Artificial, Ética e Sociedade", em Honolulu, no Havai (EUA). Os cientistas sublinham que "tem sido dada pouca atenção para compreender a dinâmica e os comportamentos emergentes que estão a surgir nesta guerra pela liderança na IA". Moniz Pereira é professor catedrático da Universidade Nova de Lisboa e investigador no NOVA-LINCS (NOVA Laboratory of Computer Science and Informatics).