

## Dependentes dos algoritmos



Muitas das escolhas que julgamos serem humanas são já feitas por máquinas: pedidos de crédito bancário, prémios de seguros, candidaturas a empregos e mesmo despedimentos, como se viu na reestruturação da TAP. Mas quem controla as máquinas?

TEXTO **DIOGO QUEIROZ DE ANDRADE** JORNALISTA ILUSTRAÇÕES **PEDRO LOURENÇO**

# A

discussão sobre a utilização dos algoritmos foi finalmente aberta em Portugal graças à polémica em torno do despedimento coletivo da TAP. Segundo denunciaram os sindicatos, a decisão está atribuída a um algoritmo, criado pela consultora BCG, que não terá em conta devida todos os aspetos da relação do trabalhador com a empresa. É um caso paradigmático a vários níveis, como confirma Henrique Sousa Antunes, diretor do grupo de trabalho de Direito e Inteligência Artificial da Universidade Católica Portuguesa, onde também ensina. O que a polémica confirma é que “já não se debate se a inteligência artificial deve ser regulada ou não, é um dado adquirido que sim. O que está em causa é a forma como se vai fazer essa regulação”.

Foi preciso um caso destes para demonstrar que a forma como a inteligência artificial tem impacto na vida quotidiana “é demasiado poderosa e justifica a necessidade de algum mecanismo de controlo”. “Neste momento, é tudo demasiado opaco, é preciso incutir confiança no sistema, e para isso é essencial transparência, que neste momento não existe”, admite Sousa Antunes. Todo o episódio evidencia os problemas da utilização dos algoritmos e justifica o esforço regulatório que a União Europeia lançou. Mas já lá vamos.

Afinal, o que é um algoritmo? Explica quem sabe. Virginia Dignum, cientista portuguesa que é professora de Inteligência Artificial na Universidade de Umea, na Suécia, sintetiza: “Um algoritmo é um conjunto de instruções predefinidas que permite chegar a um certo resultado.” Claro que esta é a formulação simples, e a cientista usa uma analogia gastronómica para facilitar ainda mais a compreensão: “Uma receita de tarte de maçã é um algoritmo, porque fornece as instruções sobre como atingir o resultado; mas claro que o produto final depende tanto do talento do cozinheiro como dos ingredientes com que se fez a receita. Da mesma forma, o problema dos algoritmos está no talento de quem os programa e nos dados com que se treina o programa para atingir o resultado pretendido.”

**“Uma receita de tarte de maçã é um algoritmo, porque fornece as instruções para atingir o resultado, que depende tanto do talento do cozinheiro como dos ingredientes”, diz a professora Virginia Dignum**

Esta questão é especialmente clara na aprendizagem automática, que é neste momento o paradigma dominante da inteligência artificial. E também explica os falhanços famosos dos algoritmos, que criam preocupação graças aos problemas expostos. Os exemplos sucedem-se: há o caso do piloto automático de carros testado pela Uber que atropelou uma ciclista porque o seu sistema não estava treinado para reconhecer bicicletas que cruzam longitudinalmente as faixas de rodagem e por isso não conseguiu processar o que fazer; há o exemplo da conta de Twitter da Microsoft que ia aprender a conversar com os utilizadores e em menos de 24 horas estava a insultar mulheres, a atacar o judaísmo e a espalhar ideologia nazi; há o caso do software da Google para classificar fotos que colocava pessoas negras na categoria “Gorila”... Os exemplos seriam suficientes para encher todas as páginas desta revista (os interessados podem visitar o [site aiaic.org](http://site.aiaic.org), que recolhe todos os casos de maus resultados na utilização de inteligência artificial).

Armando Oliveira, presidente do Instituto de Engenharia de Sistemas e Computadores (INESC) e professor do Instituto Superior Técnico, esclarece que o predomínio da aprendizagem automática na inteligência artificial atual se deve a dois fatores: “A imensa quantidade de dados disponíveis e a crescente capacidade de computação.” Ambos juntos permitem que as técnicas de aprendizagem automática se tenham generalizado, mas tudo continua a depender dos dados. “A ideia base é que, em vez de explicarmos a um programa como executar uma tarefa, mostramos-lhe como fazer, e ele aprende. Para ensinar uma máquina a conduzir um carro, alimentamos um programa de computador com filmes de humanos a conduzir, e a máquina aprende.”

Para explicar os problemas que podem surgir neste modelo de funcionamento, vale a pena voltar às maçãs. Imaginemos que se cria um algoritmo para distinguir imagens de maçãs de todos os outros frutos. Para o fazer, alimenta-se o tal algoritmo com uma base de dados de imagens de milhões de maçãs e alguns milhares de outras frutas, para que as consiga distinguir. O algoritmo começa a trabalhar, recebe uns retoques para melhorar a eficiência e depressa aprende a identificar corretamente um número impressionante de maçãs: todas as imagens de maçãs que o algoritmo identificou estavam certas, uma taxa extraordinária de 100% de precisão, fazendo deste algoritmo o mais rigoroso que é vendido para centenas de aplicações em todo o mundo, ajudando a agricultura, a indústria de distribuição e o consumo de fruta. Toda a gente está contente. O problema? Quase todas as maçãs identificadas eram vermelhas; as verdes eram rotineiramente classificadas como peras. Foi preciso investigar porque é que isto acontecia, e descobriu-se que quatro quintos da base de dados de imagens com que se alimentou o algoritmo eram maçãs vermelhas, as mais comuns na região de origem da empresa que fez o algoritmo. Isto conduziu a um imenso desperdício na produção de maçãs (porque as verdes eram deixadas fora), tornando o fruto mais escasso e mais caro. A indústria e o consumidor final saíram altamente prejudicados. O algoritmo era bastante eficiente numa parte do trabalho, a de identificar corretamente as maçãs vermelhas; mas, como o que se pretendia era “identificar maçãs”, não identificar “maçãs vermelhas”, falhou redondamente na sua tarefa. Como, entretanto, o algoritmo foi distribuído e estava a alimentar sistemas por todo o mundo, as suas consequências tiveram impacto durante vários anos.



Esta história é fictícia, mas o que se segue não é. Pense que, em vez de maçãs, o algoritmo estava a selecionar humanos para um emprego. Ou para decidir quem merece aceder a crédito bancário (e com que taxa). Ou para concluir que pena judicial será atribuída a um determinado criminoso. Este é o mundo em que estamos hoje a viver — um mundo em que as decisões são tomadas de forma aleatória por pedaços de software que não são escrutinados e sobre os quais não se conhecem as bases de dados que são usadas, os padrões de análise da fórmula matemática nem as suas taxas de sucesso. Pior: na maioria dos casos, os envolvidos nem chegam a saber que a decisão foi tomada por um programa de computador.

Todos os pedidos de crédito pessoal são hoje decididos por algoritmos. O contacto com funcionários de atendimento a clientes de quatro instituições bancárias diferentes demonstra que esta é a prática dominante no mercado: todos confirmam que a intervenção humana é mínima e só existe em caso excecional de reclamação. João Moreira Rato, *chairman* do Banco CTT, confirma também a prática algorítmica dominante e contextualiza a questão: “Sempre existiram modelos de risco, os padrões de análise que definiam a concessão de crédito eram determinados previamente. O que mudou, agora, é a quantidade de dados disponíveis a que um sistema automatizado pode aceder, e torna-se mais relevante que o cliente final tenha conhecimento disso.” Uma fonte oficial do mesmo banco optou por ser mais prudente e afirmar que o Banco CTT “combina análise humana com tecnologia e automatização e que é afinada permanentemente com base nos resultados obtidos”, recusando que o banco analise crédito “baseado apenas em modelos automatizados”.

A este propósito, o Banco de Portugal afirma que acabaram de ser emitidas pela European Bank Authority orientações claras relativas aos modelos automatizados. De acordo com o regulador, é essencial que as entidades bancárias avaliem regularmente os produtos e que informem os clientes sobre os modelos automatizados, salvaguardando sempre o direito de solicitar a intervenção de um funcionário. O Banco de Portugal diz ainda que “irá monitorizar a utilização destes modelos”, mas não é ainda claro se fará uma auditoria aos próprios algoritmos proprietários de forma a avaliar o seu enquadramento legal.

Um exemplo que anda em paralelo com este é o dos seguros: embora vários responsáveis comerciais do sector confirmem que já há algoritmos que ajudam a tomar decisões relativas a valores de prémios, não é claro em que medida se processa este facto. Teoricamente, neste momento, nada na lei impede que uma seguradora use um algoritmo que aceda a determinadas informações nas páginas públicas das redes sociais de um utilizador para atribuir o valor de um seguro de vida ou de saúde. Moreira Rato não foge ao problema de fundo, reconhecendo que estes algoritmos usam padrões de comportamento tradicional que “recolhem dados do passado para evitar erros na decisão, e por isso tendem sempre para perpetuar a situação existente. O problema é que isso pode não ser recomendável do ponto de vista social e não ser eficiente do ponto de vista económico. Ora, se não faz sentido do ponto de vista económico, é expectável que o regulador intervenha”. Aliás, basta pensar que parte da responsabilidade pela crise do *subprime* de 2008 veio do excesso de confiança do mercado nos dados adquiridos, que obviamente falharam na antecipação do futuro.

#### QUER UM EMPREGO? O COMPUTADOR DECIDE

Outra área, os mesmos problemas. Quem hoje se candidate a um emprego tem grande probabilidade de ser avaliado por um algoritmo, podendo ser excluído de um processo de candidatura com base em modelos de seleção opacos... e misteriosos para o utilizador. Claro que tudo depende da entidade que faz a intermediação. No caso da Landing Jobs, Diogo Alves explica que o essencial da sua atividade consiste em ligar empresas que procuram funcionários a candidatos no mercado. Assumindo que os algoritmos que começaram a usar, há três anos, “eram muito imperfeitos”, foram trabalhando no sistema computadorizado e hoje consideram ter uma ferramenta sólida que é essencialmente usada para rejeitar candidatos que não se enquadram em determinadas vagas. Também usam algoritmos para encontrar a relação ideal entre candidato e oferta, mas aí já usam “uma equipa de curadores que revê tudo o que as máquinas fazem, para garantir a qualidade do produto final”. Incluíram de raiz mecanismos para impedir o aumento das desigualdades, mas mesmo assim não confiam no modelo totalmente automatizado.

Já a Randstad, que é a maior empresa de recursos humanos a nível mundial, é ainda mais conservadora: usa sistemas automatizados para recolher os dados dos candidatos, procurando palavras-chave nos currículos, mas não recorre a qualquer programa para juntar candidatos e empregadores. A sua opção aumenta os custos fixos, mas garante resultados mais seguros. O porquê é explicado por Gonçalo Vilhena, CIO da empresa: “Simplemente não conseguimos garantir que temos um algoritmo que não discrimine. Temos trabalhado com os líderes de mercado na automação a nível mundial e ainda não estamos confortáveis com as soluções, até porque há vários exemplos que correram mal.”

Um dos mais conhecidos foi contado pela Reuters e passou-se na Amazon, que está longe de ser inexperiente nestas coisas da automação: em 2014, a empresa desenvolveu um algoritmo para facilitar a contratação de funcionários, avaliando os currículos numa classificação de 1 a 5 estrelas (exatamente como nos produtos que vende); para o efeito, usou os currículos dos empregados que a empresa tinha na altura. Ao fim de um ano, ficou claro que o software estava a procurar padrões que identificavam candidatas do género feminino e a recusá-las, porque a imensa maioria dos quadros superiores eram homens. Ou seja, a partir do padrão dominante, o algoritmo concluiu que o género feminino era uma característica indesejável e por isso expurgou-o. Para o consequir, excluiu logo nomes femininos e de seguida procurou sinais de género no currículo de quem se candidataria: frequência de um colégio feminino ou prática de um desporto mais associado ao género feminino garantiam exclusão imediata do processo de seleção.

Outro bom/mau exemplo veio do LinkedIn, que pretende ser uma rede social profissional que ajuda a juntar quem procura emprego e quem o oferece. Como revelou a “MIT Technology Review” em junho, os seus algoritmos apresentavam um número desproporcionado de ofertas a homens, simplesmente porque estes eram tradicionalmente mais agressivos e tinham por hábito candidatar-se a um maior número de ofertas. Ou seja, mulheres que poderiam ser boas candidatas para um lugar nem sequer eram consideradas. A rede social diz que resolveu o problema. Mas o facto de os algoritmos serem ferramentas proprietárias, com propriedade intelectual associada, impede a transparência necessária para que se efetue uma análise independente destes modelos. Gonçalo Vilhena recorda que os humanos têm preconceitos, e que esses também devem ser erradicados, mas que o cenário ideal seria o da utilização de algoritmos capazes de reconhecer o preconceito e de o evitar antes de proceder à seleção. Por enquanto, isso é uma miragem — o que não impede que a grande maioria dos processos de seleção já recorra a algoritmos em diversas fases do processo, mesmo que o candidato não seja informado disso.

#### UM PROBLEMA GENERALIZADO

Nos Estados Unidos, o problema é bem mais grave. São utilizados regularmente algoritmos com preconceitos raciais e de classe para decidir sobre penas de prisão, para estabelecer fianças e para definir prioridades em serviços públicos. O que resulta é uma sociedade ainda mais desigual, construída a partir de ideias supostamente neutras, mas que estão evadidas de eventuaisismos sociais. Na China é pior ainda: as ferramentas algorítmicas são rotineiramente usadas para definir um crédito social que determina quem acede a serviços públicos, além de contribuir para o genocídio da minoria Uighur. Todo este problema tem vindo a crescer nas últimas duas décadas.

O aumento da capacidade de computação desde o início do século recuperou o interesse nas tecnologias de aprendizagem automática, que já Alan Turing tinha apontado como uma potencial área de sucesso para a tecnologia. Nos últimos anos foram-se sucedendo os casos de dados inquinados que deram origem a resultados problemáticos. Uma área exemplar é a do reconhecimento facial, onde muito se tem investido. Em 2007, para poder treinar devidamente os seus modelos computadorizados, a Universidade de Massachusetts decidiu construir uma base de dados de faces humanas. Utilizou imagens publicadas em sites de notícias, visto que estas fornecem bastantes rostos e legendas que ajudam a identificar quem está retratado. Durante quase uma década foi uma das principais bases de dados de imagens e serviu de referência a modelos de aprendizagem automática em universidades e empresas um pouco por todo o mundo — e esses sistemas ajudam a identificar “o que é humano”. Só quando dois investigadores decidiram olhar para a conteúdo é que ficou claro o que estava na origem do sistema: a base de dados tinha 77% de homens e 83% de raça branca (e a personagem padrão, que se repeta mais vezes, era nada mais nada menos que George W. Bush). Ainda hoje há centenas de aplicações baseadas nessa base de dados, embora o método padrão já faça recolhas de imagens nas redes sociais que o permitem.

Outro exemplo, ainda mais atual, é o da ImageNet, um dos conjuntos de imagens mais usados para a construção de sistemas artificiais, com mais de 40 milhões de fotos e 20 mil categorias. É considerada a referência para alimentar sistemas de identificação de imagens, popular e de fácil acesso. Mais uma vez, o problema é que quase ninguém olhou para as imagens e não há um controlo de qualidade, porque o sistema é demasiado extenso. Ora, sem verificação nem controlo de qualidade no material de base, parece ilógico que se tomem os resultados como bons. Mas a verdade é que modelos como este continuam a ser o suporte de algoritmos que apoiam decisões político-sociais um pouco por todo o planeta. A situação é desconcertante.

## Todos os pedidos de crédito pessoal são decididos por algoritmos. O contacto com quatro instituições bancárias diferentes demonstra que esta é a prática dominante no mercado

Esta questão dos dados de partida é apenas um dos problemas dos sistemas algorítmicos. Há mais dois igualmente sérios. Um é conhecido como “Black Box” (caixa negra) e é difícil de entender. Arlindo Oliveira explica que isso “se deve precisamente à forma autónoma” como as máquinas aprendem; “nós não sabemos ensinar um programa a fazer um diagnóstico de uma doença a partir de uma radiografia, sabemos apenas alimentá-lo com milhares de radiografias com diagnósticos de forma a que o programa encontre semelhanças e consiga descortinar padrões”; mas a complexidade de alguns sistemas é tão elevada “que não se consegue já perceber como decorre exatamente o processo de aprendizagem”. E há ainda o problema do alinhamento, que consiste em garantir que os sistemas computadorizados façam exatamente o que pretendemos, nem mais nem menos. “Costumo usar um exemplo um bocadinho extremo: pedimos a uma inteligência artificial para resolver o aquecimento global, e ela pode bem decidir que a maneira mais fácil para resolver o aquecimento global é exterminar a raça humana... Este é um problema que preocupa muita gente: ter a certeza de que aquilo que pedimos a um computador é aquilo que ele vai fazer.”

A maioria dos cientistas concorda que qualquer sistema poderoso de inteligência artificial está ainda a várias décadas de distância, por isso a questão que se coloca é bem mais mundana: vamos deixar que os sistemas algorítmicos, com todas as falhas que sabemos que têm, decidam o nosso futuro? A verdade é que, mesmo com todo o desenvolvimento tecnológico, as máquinas não são capazes de pensamento estruturado complexo. A redução de um problema a uma fórmula matemática é sempre uma simplificação que deixa de fora uma série de variáveis e que exclui dimensões inteiras de uma questão. Ora uma simplificação de um problema complexo dificilmente encontrará a solução, quando muito dará partes de uma resposta que terão de ser avaliadas e validadas por humanos.

#### QUEM REGULA?

É aqui que entram os decisores. E para isso voltemos à professora Virginia Dignum, uma das cientistas com maior reconhecimento a nível global quando se trata de ética. O seu livro de 2019 “Responsible AI” foi uma das referências do grupo de peritos de alto nível (de que também faz parte) que preparou as linhas mestras sobre inteligência artificial para a Comissão Europeia. No seu livro defende alguns princípios basilares para podermos aproveitar enquanto espécie todas as vantagens desta tecnologia, “que é extraordinária e que aumenta a nossa capacidade de resolver problemas complexos”. Nas declarações ao Expresso, começou por destacar a ideia de Inteligência Artificial para Todos: “Temos de garantir que todos os humanos beneficiam desta tecnologia, e isso implica que as decisões não podem ficar nas mãos das grandes empresas, porque essas organizações não democráticas só têm interesses comerciais.” Outra ideia, anexa a esta, é a de Inteligência Artificial para o Bem: “Este é um movimento internacional que quer aproveitar a capacidade desta tecnologia para resolver os maiores problemas da Humanidade, como o aquecimento global e as pandemias.”



Basicamente, embora atrasada, a discussão decorre de outra forma. A Agência de Modernização Administrativa (AMA) está a desenvolver, no abrigo do Simplex, um modelo de avaliação de tecnologia de inteligência artificial. A pandemia atrasou o projeto, mas ele será anunciado ainda este ano. Em Portugal, é uma grade de análise detalhada que irá avaliar os modelos de inteligência artificial de acordo com cinco parâmetros: Responsabilização, Transparência, Explicabilidade, Justiça, Envolvimento. E, embora seja direcionado para a Função Pública, terá ambições de adoção ampla, podendo ainda vir a servir de base a uma autoridade que atribua selos de qualidade aos algoritmos e os inspecione. O modelo a adotar poderia ser algo semelhante ao da Autoridade de Segurança Alimentar (ASAE), que investiga problemas de qualidade nos produtos alimentares e autoriza o consumo dos mesmos.

Seria, para regressar à ideia inicial, a melhor forma de garantir que não só os ingredientes da tarte de maçã são de alta qualidade como o produto final é saboroso. Mas ainda estamos longe desse dia.

↑