

Índice

As promessas da inteligência artificial.....	1
Robot bom, robot mau	4

As promessas da inteligência artificial

O forte progresso que está a ter a inteligência artificial (IA), graças a novas técnicas como *deep learning*, é recebido com entusiasmo, mas também com o receio de que estejamos a fazer de aprendizes de feiticeiro. Uma breve revisão da situação atual do assunto pode ajudar a fazermos uma ideia do que a IA pode dar e do que não é de esperar da parte dela.

Ray Kurzweil, diretor de Engenharia na Google e um dos principais representantes do trans-humanismo, prevê que, em 2045, a IA igualará ou superará a humana (esse é um dos progressos tecnológicos que integram a chamada “Singularidade”). Nessa altura teremos “computadores com inteligência humana, iremos introduzi-los no nosso cérebro, iremos ligá-los à nuvem” e “multiplicaremos por mil milhões a nossa capacidade intelectual, fundindo-a com a inteligência que criámos” (“[Futurism](#)”, 5.10.2017).

Elon Musk, presidente da Tesla, é um dos subscritores de uma [Carta aberta](#) sobre os riscos da IA, publicada em 2015 e subscrita até há pouco tempo por mais de 8000 pessoas. Embora venda carros inteligentes, [adverte](#): “A IA é muito mais perigosa do que as bombas atômicas”.

Seria demasiado fácil aproveitar estes posicionamentos contrapostos para se situar num “justo meio caminho” e alegar que a verdade se situa algures entre Kurzweil e Musk. Mais, o que mostram essas declarações, é que as consequências da IA são levadas muito a sério pelos que trabalham nela.

Nos nossos bolsos

A IA evoca cenários de ficção científica, amigáveis como os androides de “Guerra das Estrelas”, ou sinistros como o super computador que se revolta contra os seus donos humanos em “2001: Odisseia no Espaço”. Mas hoje, a IA é uma realidade próxima e vulgar, que temos nos nossos bolsos. Permite-nos ditar mensagens que o telemóvel escreve e envia. Afina e personaliza as buscas na Internet. Traduz textos. Liberta-nos do lixo de correio. Conduz automóveis sem ninguém ao volante.

A IA conheceu uma forte aceleração nos últimos anos. Um resultado muito sonante é o do programa AlphaGo – desenvolvido pela DeepMind, o laboratório britânico que a empresa Google comprou em 2014 ultrapassando a empresa Facebook –, que no ano de 2016 ganhou ao campeão mundial de Go, jogo com quem até essa altura as máquinas não haviam conseguido vencer.

Os êxitos recentes não pertencem à IA simbólica, antes dominante, que se baseia em representar conceitos e fazer inferências. São do modelo de conexão, que procura imitar a atividade do cérebro com redes neuronais artificiais (RNA), onde os neurónios são programas interconectados que processam dados e passam o resultado aos outros. É uma ideia velha, que alcançou agora um progresso espetacular devido a três fatores: a enorme massa de dados disponíveis ([big data](#)), o aumento da potência dos processadores e o aperfeiçoamento dos algoritmos.

Máquinas que aprendem sozinhas

Isso permitiu o aparecimento do *deep learning*: sistemas que aprendem a realizar tarefas sem necessidade de programar expressamente todas as suas funções: é-lhes marcado um objetivo, são-lhes dados algoritmos para começar a procurá-lo e são treinados com dados. Assim, as RNA atuais são capazes de encontrar uma agulha num palheiro: de detetar regularidades e relações numa descomunal massa de dados sem organizar. Os programadores de IA simbólica pensariam numa árvore hierárquica para distribuí-los; com o *deep learning* é possível poupar esse trabalho.

A primeira grande demonstração do *deep learning* deve-se a Geoff Hinton (Universidade de Toronto), que em 2012 ganhou o Imagenet Challenge, um concurso de etiquetagem automática de fotografias. Pouco a pouco, as máquinas iam subindo a sua taxa de acertos, mas sem alcançar a humana, que é de 95 %. A equipa de Hinton deu um salto de 72 % para 85 %. Desde 2015, o *deep learning* supera as pessoas.

O método começou a aperfeiçoar-se e a aplicar-se a inúmeros âmbitos. A tradução automática, que estava praticamente estagnada com a IA simbólica, começou a desenvolver-se. Não é verdade que as crianças aprendem a sua língua materna muito antes de estudarem gramática? Com o *deep learning*, um programa de tradução segue o mesmo procedimento. Neste ponto, contra o dito popular, mais vale a força do que o jeito: insistir mais na potência de cálculo do que na análise gramatical. Processando os oceanos de textos nos diversos idiomas que a Internet contém, o tradutor do Google vai aprendendo a fazer isso cada vez melhor.

De qualquer forma, continua a ser inferior a um bom tradutor humano. E muito menos é tão inteligente como as crianças, que não necessitam de milhões de exemplos para aprender. John Giannandrea, investigador em IA que trabalhou para a Google e agora está na Apple, contou que uma sua filha, aos 4 anos de idade, soube classificar sozinha um velocípede (aquele antigo veículo de pedais, com uma roda grande e outra pequena) como um tipo de bicicleta, na primeira vez que viu um. “Os computadores não são capazes disso”, conclui Giannandrea (“[The Economist](#)”, 25.6.2016).

Duas inteligências

Isso importa pouco para a maioria dos que se dedicam à IA. A IA tem dois usos: um tecnológico (fazer coisas úteis) e outro científico (procurar modelos da inteligência humana para investigar como é o seu funcionamento e recriá-la em máquinas). Os tecnólogos não fazem interrogações filosóficas – o que é a inteligência, se as máquinas podem ter consciência...–; contentam-se com uma IA *débil*. Para resolver problemas práticos não faz falta uma IA *forte*, que é a que defendem os trans-humanistas e os neurocientistas “monis-

tas”, para os quais a mente é o cérebro, e o cérebro, um computador biológico.

Mas se o *deep learning* tem de continuar a sua recente expansão, terá de aproximar-se da IA forte, ou – por outras palavras – parecer-se mais com a pequena filha de Giannandrea e ser capaz de abstrair ou generalizar, captar o contexto e os subentendidos, e sem necessidade de um grande volume de dados.

Stuart Russell, professor de informática em Berkeley, salienta que os sistemas atuais de IA “baseiam-se em fixar um objetivo que a máquina tem de alcançar, e isso leva inevitavelmente a que esta assuma o controlo e que afaste qualquer coisa que possa interferir no seu caminho para isso” (“[El País](#)”, 15.6.2018). Para evitar consequências que são indesejadas, impõem-se restrições à máquina com regras; mas é impossível prever todas as eventualidades. Faz falta, portanto, uma IA que possa gerir a incerteza e adaptar-se para respeitar objetivos importantes que não são os da tarefa atribuída.

Essa debilidade do *deep learning* não se remedeia com mais dados, adverte Antoine Bordes, diretor do Facebook Artificial Intelligence Research em França. Se se trata de detetar mensagens ofensivas ou de selecionar os comentários mais relevantes a um *post*, a imensa abundância de dados não é uma vantagem clara para o Facebook, mas antes de tudo o problema. Por isso, afirma Bordes, “o assunto verdadeiramente fundamental onde se trabalha é como conseguir que o sistema venha a aprender com apenas dez ou vinte exemplos” (“[Le Monde](#)”, 10.7.2018).

Em última análise, continua a ser verdade o dito popular: mais vale o jeito do que a força, e o jeito consiste em idealizar métodos que resolvam os problemas não com base na força bruta de computação, mas de modo eficiente, com o menor número possível de cálculos. Assim o sublinha a especialista Margaret Boden (Universidade de Sussex) no seu livro “[Inteligência artificial](#)” (Turner, 2017). O desenvolvimento futuro do *deep learning* e a possível obtenção de uma IA forte exigem ultrapassar importantes limitações da IA atual. Não é certo que todas sejam superáveis.

Inteligência monotemática

Uma limitação é que o *deep learning* não é *sequencial*. Uma pessoa que sabe patinar com patins de rodas aprende facilmente a patinar sobre gelo. Um sistema de *deep learning* é capaz de adquirir uma perfeição insuperável num videogame de condução; mas cada modalidade do jogo é, para ele, um jogo diferente: depois de obter o sucesso na condução em piso seco, para aprender a conduzir sobre neve tem de começar a partir do zero e, além disso, esquece aquilo que aprendeu antes (é o que se designa por “esquecimento catastrófico”).

Outra limitação é o problema do enquadramento: a máquina só leva em conta o que está dentro de um âmbito definido, fora do qual não existe – parte desse princípio – nada de relevante para a tarefa que tem de realizar. É um reducionismo metódico que nós humanos empregamos muitas vezes; mas nós também sabemos olhar para o mundo, enquanto a máquina só é capaz de vê-lo pelo canudo que lhe é fornecido.

De facto, estamos aqui graças ao facto de durante a Guerra Fria alguns homens terem tomado decisões que saíram fora do quadro em que operavam as máquinas. Em 1960, o sistema de alerta nuclear norte-americano avisou que a URSS acabava de lançar numerosos mísseis contra os Estados Unidos. Mas os comandos norte-americanos, que conheciam a situação geral – política, militar, diplomática – não acreditaram que os soviéticos tivessem decidido iniciar uma guerra atômica: desobedeceram ao protocolo estabelecido e não ordenaram disparar. Comprovou-se rapidamente que aquilo que o sistema havia interpretado como sinal de um ataque em massa eram anomalias num radar da Gronelândia causadas pelo nascer da lua (cfr. Alan Borning, [“Computer System Reliability and Nuclear War”](#)). Houve outros casos semelhantes, também no lado soviético.

Sem sentido da relevância

O problema do enquadramento revela outra limitação da IA: falta-lhe o sentido humano da relevância, nas palavras de Boden. Os movimentos do mercado bolsista são em princípio terreno ideal para o *deep learning*: uma infinidade de dados onde encontrar padrões para que se possa prever. Mas verifica-se que as máquinas inteligentes nem sempre são capazes por si sós de selecionar o “ruído”, como chamam os estatísticos aos acontecimentos concomitantes com o fenómeno que se observa mas que não influem nele. Por vezes, detetam falsas regularidades que não o são, mas meras coincidências.

Por isso, seria necessário que o *deep learning* justificasse porque chegou a determinada conclusão (por exemplo, que não se deve conceder a determinada pessoa o empréstimo que pede). Mas quando o sistema lidou com uma grande abundância de dados, torna-se impossível rever os milhares ou milhões de micro cálculos que conduziram ao resultado final.

Isto importa muito por outra razão. A IA é utilizada para tomar decisões sobre pessoas, com base na análise de características e comportamentos de muitas. Determina-se assim, tendo em vista a venda de um seguro, quem tem maior probabilidade de adoecer ou ter um acidente de automóvel; ou faz-se a primeira seleção de candidatos a um cargo; ou faz-se a revisão de solicitações de crédito.

Parte-se do princípio de que a IA não tem preconceitos. Mas, adverte Robin Bordoli, diretor-geral da Crowd Flower, empresa de trabalho temporário que disponibiliza pessoas para treinar sistemas de *deep learning*, “um algoritmo só pode ser tão bom como a quantidade e qualidade dos dados de treino que se lhe forneçam” ([“The Wall Street Journal”, 12.11.2017](#)). Assim se viu com Tay, o *robot* de conversação que idealizado pela Microsoft, imitou a linguagem racista e sexista com que depa-rou na linguagem natural em que aprendeu. Não tinha más intenções, evidentemente; mas muito menos dominava a ironia ou o sentido da oportunidade.

Novas ideias

Os trabalhos atuais em IA são canalizados para superar essas limitações.

A proposta de Stuart Russell é o que designa por “aprendizagem de reforço inverso cooperativo”: uma espécie de IA forte que não se limitaria a um único objetivo determinado. Usaria uma linguagem de programação probabilística, e aprenderia a responder a imprevistos consultando operadores humanos.

Outra abordagem é a da DeepMind, que publicou os primeiros resultados em [“Proceedings of the National Academy of Sciences” \(28.3.2017\)](#). Não é ainda aprendizagem sequencial propriamente dita, mas faz um desvio para evitar, até certo ponto, o problema de que as RNA só têm possibilidade de aprender uma tarefa nova “sobrescrevendo” as ligações feitas ao aprender com a anterior. A DeepMind idealizou um tipo de *deep learning* que conserva as ligações principais da primeira aprendizagem antes de empreender a seguinte. O sucesso é relativo, pois estas RNA mais flexíveis não alcançam tanta perfeição nas tarefas como as que fazem uma só.

Em troca, a DeepMind conseguiu um avanço importante para um *deep learning* menos ávido de dados. A segunda versão do seu programa de Go, chamada Alpha Go Zero, também conseguiu a medalha de ouro mas sem instruções prévias. Foi, além disso, muito mais eficiente: aprendeu em três dias, em vez de em vários meses, e precisou de jogar “somente” 5 milhões de partidas, em vez de 30 milhões.

Também está em andamento um projeto chamado [Explainable Artificial Intelligence](#) (XAI). Financiado pela DARPA (Defense Advanced Research Projects Agency), organismo de investigação do Departamento de Defesa norte-americano, propõe-se criar sistemas de *deep learning* que expliquem os seus resultados aos seres humanos.

Suspense no “teste” de Turing

Se essas e outras investigações tiverem sucesso, é de prever um vigoroso desenvolvimento da IA nas próximas décadas. Mas é duvidoso que seja alcançada a IA forte.

A incapacidade do *deep learning* de justificar os seus resultados parece uma mostra clara de que esses sistemas inteligentes não têm autoconsciência. E à margem de questões filosóficas como essa, o facto é que nenhuma máquina conseguiu superar o teste de Turing, que só pede à IA ser funcionalmente equivalente à humana, de modo que o seu comportamento seja indistinguível do de uma pessoa. No concurso para o Prémio Loebner, que se realiza anualmente desde 1990, o programa mais inteligente não iludiu o júri mais de dez minutos.

Boden, que passa em revista os numerosos sucessos da IA, não deixa de registar as suas limitações. Com realismo, conclui: “A IA forte continua a estar tão longe como sempre”.

R. S.

Robot bom, robot mau

Qualquer tecnologia levanta o problema do seu uso ético, mas com a inteligência artificial existe um problema adicional: como assegurar que ela própria se comporte bem.

A [Carta aberta](#) citada pede mais investigação para que os sistemas inteligentes sejam benéficos para as pessoas e a sociedade. “Têm de fazer aquilo que queiramos que façam”, acrescenta, advertindo do perigo de lhes perdermos o controlo. Mostras da mesma preocupação são iniciativas como as do governo norte-americano ou a da DeepMind, que criaram comités de estudo sobre ética da IA.

Para que a IA faça o bem e evite o mal não bastarão as leis enunciadas por Isaac Asimov em “I, Robot”. Não se pode programar uma máquina com regras que prevejam todas as eventualidades. Muito menos nós humanos aprendemos e aplicamos a moral dessa maneira, e nas diversas situações orientamo-nos com a nossa inteligência natural, sem esquecer a vontade e os sentimentos.

Dar a uma máquina um sentido moral artificial torna-se francamente difícil. Estão em curso algumas tentativas.

O projeto [Good AI](#), do informático eslovaco Marek Rosa, baseia-se no treino do sistema com exemplos, como se treinou o Alpha Go com partidas, guiado por um mentor. O objetivo é que dos casos morais propostos, a máquina extraia os valores fundamentais a procurar em novas situações.

Outra abordagem é o [“Ethical Adaptor”](#) de Ron Arkin, professor de Ética Informática no Instituto Tecnológico da Geórgia (EUA). Para que os *robots* interiorizem a moral, pretende que imitem não o comportamento humano, mas as emoções humanas. Se a máquina, por exemplo, experimenta uma versão artificial do sentimento de culpa, aprenderá com os seus erros e não os repetirá.

Problemas reais

Em todo o caso, os perigos éticos da IA não são, pelo menos por agora, que os computadores se revoltem e dominem o mundo. Existem outros mais claros e próximos.

Discriminação em conceder créditos, na seleção de pessoal, etc., quando se usa a IA. Alguns professores de Oxford propuseram que seja criado um organismo auditor de algoritmos, que garanta que são imparciais. O Parlamento Europeu, por seu lado, estuda a possibilidade de instaurar um direito à explicação, para que o candidato discordante possa saber porque o sistema o afastou (mas isto exigiria que houvesse a XAI).

Responsabilidade civil. É um assunto urgente no caso dos automóveis sem condutor. Numa [resolução](#) aprovada no ano passado, o Parlamento Europeu propõe criar uma “personalidade jurídica eletrónica” para *robots*, a fim de determinar quem deve ressarcir por danos no caso de haver falha ou acidente.

Robots de companhia. No Japão, já começaram a ser usados para atender idosos. Boden considera-os de muito duvidosa ética, tal como os parceiros sexuais artificiais (algo assim como a assistente do filme [“Her”](#)), pela mesma razão avançada na resolução do Parlamento Europeu, que salienta os perigos para pessoas vulneráveis que desenvolvem vínculos afetivos com *robots*.

Uso bélico. Alguns pensam que se deveria proibir os sistemas inteligentes autónomos com poder letal, pois a escolha de objetivos e da força a empregar são decisões de grande responsabilidade que deveriam ser tomadas sempre por humanos.

R. S.