REVISTA DE CIÊNCIA ELEMENTAR

Inteligência Artificial

Riscos e promessas

João Vinagre, Nuno Moniz
LIAAD/ INESC TEC/ Universidade do Porto

CITAÇÃO
Vinagre, J., Moniz, N.(2020) Inteligência Artificial, *Rev. Ciência Elem.*, V8(04):052. doi.org/10.24927/rce2020.052

EDITOR
José Ferreira Gomes, Universidade do Porto

EDITOR CONVIDADO
João Lopes dos Santos Universidade do Porto

RECEBIDO EM
01 de novembro de 2020

ACEITE EM
02 de novembro de 2020

PUBLICADO EM
15 de dezembro de 2020

rce.casadasciencias.org

A "revolução tecnológica" da Inteligência Artificial está em curso e, nos últimos anos, técnicos especialistas, decisores políticos, comunicação social e opinião pública têm acalentado o debate sobre onde nos poderá levar. No entanto, esse debate peca por, recorrentemente, partir de premissas falsas, ou, pelo menos, de pressupostos improváveis. O problema no entendimento e discussão do que é a Inteligência Artificial, o seu estado atual e a perspetiva de futuro tem várias origens, para o qual concorre o imaginário de obras de ficção e do *marketing*, exercícios de futurismo e aproveitamento comercial ou político, assim como da falta de informação geral sobre o tema. Frequentemente, são apresentados cenários altamente improváveis, e até mesmo fantasiosos, sobre a evolução da Inteligência Artificial que, percecionados como verdadeiros ou prováveis, constroem uma narrativa e entendimento do tema falaciosos. Pretendemos, com este artigo, abordar o tema da Inteligência Artificial, contribuindo para uma discussão profícua em torno do seu potencial, da sua atualidade e do seu futuro, assim como dos seus riscos.

O problema da definição

Em 1950, Alan Turing, o criador do modelo matemático em que assentam todos os dispositivos computacionais atuais, publicou um artigo[1] que partia da seguinte questão: "*Podem as máquinas pensar?*". Logo no primeiro parágrafo, Turing coloca o problema: para podermos responder à questão sobre se as máquinas pensam, temos primeiro que definir os conceitos de "máquina" e de "pensar", o que não é fácil. Assim, e de forma a contornar o problema da definição conceptual, Turing substitui a sua questão por um jogo, o "Jogo da Imitação". O objetivo é o de uma máquina tornar-se indistinguível de um ser humano numa sequência de perguntas e respostas (por escrito). Caso o avaliador humano não seja capaz de afirmar se está a conversar com uma máquina, poder-se-ia deduzir que a máquina é uma entidade inteligente.

O problema na definição do que é Inteligência Artificial (IA) é precisamente o mesmo, vindo, desde logo, da definição individual dos dois conceitos que a compõem: "inteligência" e "artificial". Estas definições não são consensuais, e variam razoavelmente conforme a área de estudo, e mesmo dentro da mesma área de estudo. Por exemplo, Stephen Hawking definiu inteligência como "*a capacidade de adaptação à mudança*". Mas sem uma

contextualização mais rigorosa, havendo plantas e microorganismos que se adaptam com sucesso a mudanças nunca experienciadas por eles, será que os podemos considerar inteligentes? O mesmo acontece com o conceito de "artificial". Tipicamente, pensamos que algo é artificial se é fabricado pelo ser humano e não pode ser produzido exclusivamente por processos naturais, sem intervenção humana. Ainda assim, é debatível se a ovelha Dolly foi natural ou artificial, uma vez que, por um lado, foi clonada por humanos a partir de uma ovelha "natural", por outro, não apresenta quaisquer diferenças intrínsecas relativamente a ovelhas "naturais". Com isto queremos deixar claro que, logo à partida, no que diz respeito à própria definição do significado de IA, existe uma grande discussão sobre como definir inteligência e artificialidade; mais importante para este artigo, existem múltiplas zonas cinzentas, o que torna extremamente complicado uma definição popular (no sentido do público em geral). Para ilustrar esta questão, mencionamos o trabalho da *AGI Sentinel Initiative*, que fez um levantamento[2] de definições existentes de IA e Inteligência Humana.

Mas, voltando à questão de Turing, será que podemos "fugir" à questão semântica com um jogo? Talvez, mas não com o mesmo jogo. Na verdade, o que o Jogo da Imitação permite avaliar é se uma máquina consegue simular as respostas de um ser humano, ou seja, se o consegue imitar. No entanto, na maioria das aplicações atuais de IA, este pressuposto não é aplicável. O que se pretende com a IA é complementar ou, no máximo, aumentar a inteligência (ou capacidade) humana, e não substituí-la. Se de um ponto de vista filosófico a comparação da IA à Inteligência Humana é muito interessante, para a finalidade deste artigo, o jogo de Turing não é muito útil. Assim, cientes de que, pelas razões acima, qualquer definição de IA é debatível, e que o seu estudo inclui também a procura da sua definição, neste artigo adotamos uma definição tradicional: IA é a capacidade de máquinas, isoladamente ou em conjunto, e com o mínimo de intervenção humana, planearem e executarem tarefas num vasto número de ambientes/problemas.

Com uma postura crítica - a que a pertença aos esforços de estudo do próprio tema permite - neste artigo propomos fazer uma análise sobre o conceito e as capacidades de IA, olhando para a sua atualidade, os seus riscos, e o futuro próximo. Em cada uma destas partes, não nos propomos a enumerar extensiva ou pormenorizadamente cada dimensão do estudo de IA. O que propomos é a análise de um conjunto de tópicos que reúnem alargado interesse em termos de investigação atual e que, na nossa opinião, poderão ser cruciais para o desenvolvimento da IA. Neste sentido, o nosso objetivo é possibilitar uma intervenção mais alargada sobre este tema, ao potenciar a interseção da experiência e conhecimento de múltiplos domínios científicos nesta discussão.

Atualidade: Aprendizagem automática não é um canivete Suíço

Com a proposta do Jogo da Imitação, Turing foi dos primeiros a colocar a possibilidade das máquinas poderem aprender, algo que até então só era visto como possível aos seres humanos e a alguns animais. E a essa proposta podemos ligar os avanços mais recentes da IA, sendo que neste artigo focamo-nos na área de Aprendizagem Automática (ML, do inglês *Machine Learning*). Esta área alterou a forma como utilizamos os sistemas computacionais. No paradigma computacional tradicional, o que tipicamente queremos fazer é dar ao computador um programa (um encadeamento de funções ou regras de processamento) e dados de *input*. O computador irá correr o programa, processando o *input* e produzindo

um *output*. Essencialmente, criamos processos cujo funcionamento é definido por nós, ou recriamos processos cujo funcionamento conhecemos bem. No paradigma de ML, o que pretendemos fazer é modelar processos que podemos observar, mas que são demasiado complexos para os conseguirmos recriar, de forma fidedigna, com um programa. O que damos ao computador é um conjunto, normalmente muito grande, de dados de *input* e *output* que observamos nesses processos. Damos também um algoritmo que analisa estes dados e "aprende" um modelo, que não é mais do que o encadeamento de regras e funções que transformam o *input* no *output* de forma muito semelhante ao processo que observámos. A questão que se poderá colocar imediatamente é: para que serve isto? A resposta é que passamos a ter a hipótese de modelar computacionalmente fenómenos potencialmente muito complexos, desde que sejam observáveis. Com modelos obtidos desta forma, podemos prever o estado do tempo com grande exatidão, melhorar processos industriais ou modelar organismos vivos e inteligentes, podendo prever, por exemplo, os efeitos de medicamentos ou obter uma sugestão de que filme ver hoje à noite sem necessidade de o escolher entre centenas num catálogo.

Com o crescente conhecimento e capacidade dos métodos de ML, surge também uma corrente de expectativas em relação ao que (no geral) a IA poderá alcançar. Tais expectativas são alimentadas de uma maneira constante e mediatizada, principalmente através de operações de *marketing* comercial e de comunicação das indústrias que adotam ou desenvolvem ferramentas nesta área. Naturalmente, a componente mediática de tal comunicação leva, invariavelmente, a algum exagero de competências da IA e dos sistemas computacionais nos dias de hoje. Esta crítica não é, de maneira nenhuma, nova, contando com descrições pormenorizadas de como a "corrida ao ouro" da IA levou a subprodutos com sérias implicações para terceiros (humanos), ou como o ciclo silencioso de proposta, adopção, discrepância entre *performance* laboratorial e real, e (por fim) descartamento, opera[3, 4].

Queremos, no entanto, referir-nos a um episódio específico. Este é uma importante demonstração, porque revela os dois lados da mesma moeda do jogo mediático à volta dos desenvolvimentos e capacidades da IA. Em Setembro de 2020 foi publicado um texto no jornal *The Guardian*, redigido (alegadamente) por um modelo de Processamento de Linguagem Natural - denominado GPT-3[5] - que dificilmente pode ser distinguido de um texto escrito por um ser humano - sendo, porventura, um candidato a passar no Jogo da Imitação. Porém, e mesmo representando um avanço impressionante no estado da arte, após uma leitura cuidadosa do texto em causa percebemos que este foi, na verdade, escolhido entre vários outros textos produzidos pelo sistema e, posteriormente, editado ainda por jornalistas. Este é um exemplo de como a capacidade da IA de hoje é exacerbada, levando a conclusões infundadas sobre a superioridade desta em relação a humanos em algumas tarefas como a deteção e identificação de objetos em imagens e vídeos.

Como ilustração dos exageros sobre as capacidades da IA nos próximos anos, avançamos a seguir com alguns dos cenários que são frequentemente apontados como possíveis, tentando mostrar por que são irrealistas.

Futuro: Cenários pouco prováveis

Até há poucos anos atrás, a IA ainda era vista como futurismo ou ficção científica. Autores como Philip K. Dick, William Gibson e Bruce Sterling exploraram literariamente o tema (entre outros também relacionados com tecnologia) desde o final dos anos 60, em paralelo

com o aumento da investigação científica na área. Desde a criação deste imaginário distópico - o ciberpunk - na literatura de ficção, surgiram as adaptações e variantes no cinema (Blade Runner, Exterminador Implacável, The Matrix, I-Robot, AI, Minority Report) e, mais recentemente, na televisão (Person of Interest, Altered Carbon). Em todas estas obras, encontramos, de forma mais ou menos explícita, o dilema ético e social de como encarar e lidar com máquinas tão ou mais inteligentes que o seu criador, o ser humano. Hoje, várias décadas depois do artigo de Turing e de um enorme avanço científico na área, grande parte dos impactos da massificação da IA na sociedade continuam incertos e, nesse aspeto, curiosamente, a maioria das questões éticas e sociais levantadas pelos autores de ficção continuam atuais. No entanto, também têm levado à discussão de vários cenários futuristas, mas pouco realistas, sobre a nossa dependência na IA, num futuro mais ou menos próximo. Falamos aqui de alguns, a título exemplificativo.

O primeiro cenário será aquele em que simplesmente delegamos à IA controlo sobre muitos aspetos críticos na nossa vida, ao ponto de dependermos dela para sobreviver. Este cenário poderá parecer assustador para muitos, mas tem alguma ligação ao cenário atual. Hoje em dia, as economias mais desenvolvidas dependem em grande medida da IA em algumas atividades críticas. Podemos dizer com alguma segurança que caso toda a IA subitamente deixasse de funcionar, haveria perda de vidas. Além deste aspeto crítico, também não é negligenciável o nosso uso quotidiano, em particular o inerente ao uso de smartphones, de sistemas inteligentes para entretenimento e organização pessoal. Esta interação com sistemas de IA alterou, sobretudo na última década, os nossos estilos de vida, padrões de consumo e até a forma como pensamos e olhamos o mundo. No entanto, verificamos que estes sistemas estão totalmente sob o controlo de seres humanos. Portanto, mais do que encarar este cenário como uma catástrofe, é importante centrar a questão em quem controla a tecnologia e quanto das nossas vidas lhe(s) estamos a confiar, indiretamente, através da IA. Adicionalmente, este cenário implica, não só uma maior dependência, mas também uma entrega de soberania. Ou seja, existe um lado político neste cenário que é recorrentemente ignorado ou preterido ao enfoque tecnológico. Este cenário é a negação da política.

Outro cenário, mais catastrofista, é aquele em que uma IA adquire capacidade de manipular o mundo superior à do ser humano e, simultaneamente, a intenção de o fazer - a chamada Singularidade. Neste caso, os humanos serão eventualmente subjugados, ou mesmo eliminados, pela IA. Este cenário, apesar de ser muito útil (e ainda bem) a escritores e argumentistas de ficção, é extremamente improvável nas próximas décadas. Para se concretizar, será necessário que a IA adquira capacidade e vontade de dominar o mundo. Se estamos ainda longe de a IA ter uma capacidade hegemónica, devido a várias limitações tecnológicas que demorarão décadas a ultrapassar, a questão da vontade, não sendo impossível, é ainda mais questionável, porque pressupõe a existência consciência nas máquinas, algo que ainda não se vislumbra verdadeiramente, dadas as suas limitações fundamentais. Estamos portanto no domínio da pura especulação. A boa notícia é que temos ainda muito tempo para para nos entretermos com boa ficção científica, sem termos que nos preocupar muito com qualquer catástrofe real associada a este cenário.

Um cenário mais recente, muito referido, por exemplo, nos livros de Yuval Harari[6], prevê que o ser humano aumente a sua própria capacidade, fundindo-se orgânica e intricadamente com a tecnologia, ao ponto de evoluir para uma espécie completamente nova, com

longevidade, inteligência e omnisciência inimagináveis para o ser humano atual. No que diz respeito à longevidade, é evidente que o combate à fome e às doenças tem sido um desígnio da humanidade, tendo a esperança de vida aumentado consistentemente desde há muitas décadas. Relativamente à omnisciência, as telecomunicações e sistemas de informação dão-nos hoje acesso imediato a um número de fontes de informação, em tempo real, incalculavelmente superior ao de há apenas 20 anos. A IA já complementa a nossa inteligência em inúmeros processos e tarefas muito complexos. Somando isto aos avanços nos domínios da biotecnologia, nanotecnologia, e robótica, a aceleração desta evolução para um paradigma em que incorporamos a tecnologia, incluindo a IA, no nosso próprio organismo, não será um cenário completamente desprovido de suporte. No entanto (felizmente), os problemas éticos relacionados com a manipulação de organismos vivos, com especial atenção aos seres humanos, estão em permanente debate na sociedade. Como consequência, existem, na esmagadora maioria dos países, limitações legais ao que é possível fazer. Especificamente na IA, existe na comunidade científica um debate sério e permanente sobre as implicações da massificação da IA nas sociedades, existindo mesmo já um documento oficial na União Europeia com os princípios éticos da sua aplicação. Para já, para além das propostas mediáticas de autopromoção de alguns milionários excêntricos e algumas teorias da conspiração, não se conhecem projetos sérios de incorporação de interfaces de IA no cérebro humano. Aquilo que sabemos sobre o estado atual da IA é que esta não demonstra um nível elevado de confiabilidade em servir como interface com o público em geral, principalmente quando fica responsável por decisões que podem ter implicações diretas nas vidas das pessoas[7]. Não obstante, é preciso manter o debate atento e informado sobre estes temas, e inevitavelmente ir ajustando os limites legais ao que é permitido fazer neste campo, medindo permanentemente os benefícios (por exemplo, sistemas sensoriais ou de controlo de próteses mecânicas em pessoas com deficiência) e os riscos (de manipulação externa dos sistemas, ou de atribuição de "super-poderes") da incorporação de IA em seres humanos ou outros seres vivos.

Neste ponto, é importante afirmar que não há qualquer dúvida que os avanços recentes na IA levaram a desenvolvimentos impressionantes na área da medicina, telecomunicações, indústria automóvel, finanças, etc. Não obstante, a evidência em relação à incapacidade de responder a questões simples, mas cruciais para a sua aplicação no mundo real (e.g. como um modelo chegou a determinada previsão?) sobressai e sugere que a maturidade da tecnologia presente de IA está francamente sobrestimada - pelo menos no que diz respeito à sua "inteligência" e a qualquer comparação com inteligência humana. Se, por um lado, é hoje possível à IA superar o ser humano em algumas tarefas específicas (ainda que complexas), não se vislumbra ainda a possibilidade da mesma IA igualar ou superar o ser humano num conjunto alargado de tarefas, simplesmente porque existem demasiados constrangimentos tecnológicos para vislumbrar tal cenário a curto prazo. Podemos admitir que a IA venha a adquirir capacidades humanas, ou mesmo sobre-humanas, num futuro longínquo. Mas este debate, sendo interessante, é muitas vezes feito por antecipação a outros que são mais importantes a curto prazo. São estes debates, sobre os riscos da IA a curto prazo, a que nos dedicamos na próxima secção.

Os Riscos reais da IA

Com a progressiva implementação de sistemas baseados em IA em posições de interface

com o mundo real, assim como contextos em que as operações de tais sistemas podem ter implicações em pessoas concretas ou grupos de uma sociedade, surgem preocupações que cruzam com outras áreas da ciência, como as ciências sociais. Esta é, inclusivamente, uma das grandes preocupações na comunidade de IA, sendo por consequência um dos tópicos com maior atividade neste momento. Neste artigo, damos destaque a quatro tópicos, desenvolvidos nas seguintes subsecções.

Transparência

Existe a perceção de que a IA consegue, hoje em dia, superar seres humanos em tarefas complexas, ainda que de âmbito muito limitado. Por exemplo, algumas tecnologias mais recentes de visão computacional, baseadas em IA, conseguem, em experiências controladas, detectar alguns tipos tumores cancerígenos em imagens médicas com uma precisão muito elevada, em muitos casos melhor do que a de especialistas muito experientes. Mesmo considerando que, no mundo real, conseguirão o mesmo nível de performance - o que ainda é cedo para concluir - há um problema ainda mais difícil de resolver. Enquanto que uma especialista nos consegue muito facilmente explicar o que vê numa determinada imagem, ou debater com colegas pormenores que poderão levar a diferentes decisões de diagnóstico ou tratamento, o algoritmo de visão computacional é incapaz de o fazer. Isto leva a um aparente paradoxo: a IA poderá até ter maior probabilidade de acertar no diagnóstico, mas, mesmo assim, tendemos a confiar mais na especialista, pelo simples facto de que ela consegue explicar o que vê. Adicionalmente, é expectável que sistemas de IA falhem em casos atípicos, simplesmente porque não foi treinada com tais exemplos - ou pelo menos não em número suficiente. O problema da transparência também é óbvio em sistemas que poderão ser usados por bancos e seguradoras para definir o perfil de risco dos clientes[4]. Caso um sistema de IA nos classifique como clientes de alto risco, uma instituição financeira poderá negar-nos um crédito ou um seguro de saúde. Um elemento chave para determinar se esta decisão é legítima é saber ter uma explicação para o resultado que seja, grosso modo, aceitável, algo que é muito difícil de obter da IA, com a agravante de que esta explicação se torna mais difícil de produzir à medida que o algoritmo se torna mais sofisticado.

Privacidade e proteção de dados

As redes sociais tornaram-se, nos últimos anos, extensões naturais da nossa vida, e muitos de nós lá colocam fotografias, mensagens, comentários, opiniões. Os serviços de streaming de vídeo e música colecionam toda a nossa atividade de forma a poder personalizar o serviço, conhecendo por isso todas as nossas preferências musicais, de televisão e de cinema. A isto poderemos somar aquilo que os nossos smartphones registam, tal como a nossa localização geográfica, os nossos percursos diários, as nossas pesquisas, etc. Indo um pouco mais longe, podemos ainda referir os dados que já partilhávamos, fora da internet, com muitas entidades, começando pelo próprio Estado, passando pelos prestadores de cuidados de saúde, bancos, seguradoras, e terminando nos grupos comerciais (através de cartões de crédito e programas de fidelização), que, não sendo facilmente acessíveis, estão, de forma geral, digitalizados. Rapidamente chegamos à conclusão que existem micro-perfis nossos espalhados por múltiplas entidades ao redor do mundo. Se conseguíssemos juntar todos estes micro-perfis, provavelmente ficaríamos surpreendidos com a quan-

tidade enorme de informação que existe sobre nós, incluindo sobre aspetos extremamente íntimos e pessoais. Estes dados são recolhidos, de forma geral, com objetivos benignos, como a personalização dos serviços que nos são prestados ou a melhoria dos produtos e serviços, a um nível geral. No entanto, existem inúmeros exemplos de uso indevido de dados pessoais (um caso recente, mas não isolado, foi a manipulação de massas no escândalo da Cambridge Analytica), para além de repetidos casos de roubo de dados em sistemas que são comprometidos por piratas informáticos. No limite, estes dados podem ser usados por redes criminosas ou terroristas para diversos esquemas fraudulentos, como roubo de identidade, fraude financeira, burlas, ou outros fins ainda mais preversos (pedofilia, tráfico de seres humanos, etc). Também não será totalmente descabida a hipótese de estes dados poderem ser usados para opressão social por estados totalitários, sejam eles atuais ou futuros. Os dados pessoais são valiosos e existem mesmo mercados, mais ou menos obscuros, em que se transacionam vários tipos de dados pessoais. O problema central é que os dados que registamos hoje não desaparecem facilmente, e estarão disponíveis num horizonte temporal indeterminado. Nada nos garante que os nossos dados não serão usados de forma maliciosa no futuro. A preocupação com privacidade e controlo de dados pessoais chegou recentemente à legislação europeia com o Regulamento Geral de Proteção de Dados (RGPD), que regula todas as atividades que envolvem recolha, transmissão e processamento de dados pessoais. No entanto, a rapidez no avanço tecnológico pode, a curto prazo, tornar a legislação obsoleta, ou mesmo contraproducente.

Enviesamento

Um problema cada vez mais evidente da IA é o potencial enviesamento social. Esta é talvez a questão atual mais complexa e mais difícil de resolver. Num episódio recente, foi disponibilizado publicamente um modelo de IA, cujo objetivo era gerar uma face humana em alta resolução, a partir de uma fotografia desfocada. Este modelo foi treinado com milhares de caras de celebridades, mas não tinha como objetivo gerar a face de ninguém em particular. Simplesmente, a face teria que ter as características necessárias para que nós, seres humanos, as identificássemos como uma face humana. Muito rapidamente se percebeu que, independentemente da raça ou etnia da face desfocada que fosse dada como input, quase todas as faces geradas eram caucasianas. Alguns exemplos de input, sendo obviamente de caras de negros ou asiáticos, não geravam faces com as mesmas características. Mais recentemente, descobriu-se que, numa popular rede social, ao fazer-se uma publicação com uma imagem com dimensões desproporcionais, o algoritmo que fazia o ajuste da imagem, ocultando parte dela, sistematicamente mostrava homens em detrimento de mulheres, brancos em vez de negros, entre outras “preferências”. Um recente galardoado com o prémio Turing - o “Nobel” da informática - pelos seus contributos no avanço da IA, defendeu, recentemente, que estes problemas apenas sucedem porque os modelos de IA terão sido treinados com dados enviesados. Caso tivessem sido usadas apenas caras de celebridades negras, por exemplo, então o modelo teria um enviesamento semelhante, mas preferindo as faces negras às brancas. No entanto, tal afirmação não é consensual. As questões que se colocam são: se ao longo de décadas a investigação em IA tem avançado usado maioritariamente dados enviesados, então não será plausível que o próprio avanço científico, em particular na IA, tenha algum tipo de enviesamento? De uma forma mais geral, não teremos nós, na própria comunidade científica, enviesamentos, ainda que ligeiros e não

intencionais, que não se anulam entre si e geram portanto erro sistemático? Não estará a tecnologia embebida de assunções que partem de tais enviesamentos e erros? A resposta a estas questões ainda não está respondida de forma cabal, mas é razoável admitir que o desenvolvimento dos métodos (alguns com mais de um século) usados em IA não é estanque de enviesamentos sociais, culturais ou político-económicos.

Bolhas de filtro

Um dos grandes avanços recentes na IA foram os sistemas de recomendação. Estes sistemas são compostos por algoritmos que filtram o conteúdo a que temos acesso, de acordo com os nossos gostos e preferências. Isto traz-nos um acesso muito mais rápido a conteúdos que se adequam ao nosso perfil, poupando-nos tempo e permitindo que descubramos coisas interessantes (para nós) que porventura desconhecíamos. O grande inconveniente é a tendência destes sistemas ficarem retidos em bolhas de informação, dentro do universo daquilo que nós gostamos e queremos, reforçando essas preferências e levando a que nós próprios fiquemos retidos nessas "bolhas", sem que outros conteúdos nos sejam mostrados ou sugeridos. No caso de conteúdo informativo (notícias, documentários, livros, etc) isto é especialmente problemático, porque está mostrado que estas bolhas levam à acentuação de posições pouco equilibradas e mesmo extremistas. A razão para isto acontecer é simples: as bolhas de informação nas redes sociais e serviços de streaming de vídeo dão primazia a posições extremadas - que confirmam as nossas opiniões -, relativamente a posições moderadas, porque as primeiras simplesmente geram mais atividade - e lucro - do que as segundas. Infelizmente, na história recente, os maus exemplos são muitos e os impactos são conhecidos. Faltam ainda os incentivos certos e tecnologia adequada para que se aumente a diversidade de informação a que somos expostos, sem que se percam as vantagens inegáveis da massificação de acesso a conteúdos.

Equidade

Como sucede com qualquer tecnologia disruptiva, é extremamente importante perceber quem detém essa tecnologia, bem como os meios para a colocar em funcionamento. Neste aspeto, a IA é talvez a primeira grande tecnologia disruptiva - isto é, profundamente alteradoras do funcionamento das sociedades - em que a investigação científica e os avanços tecnológicos ocorrem maioritariamente em empresas. Se pensarmos nas tecnologias disruptivas das últimas décadas, verificamos que o seu avanço foi impulsionado sobretudo por investimento público e protagonizado por investigadores integrados em Universidades ou em laboratórios de Estado e/ou militares. O investimento no avanço da IA, pelo contrário, é neste momento sobretudo feito por grandes empresas, capazes de suportar grandes equipas de investigação por vários anos, algo que até há muito pouco tempo atrás, apenas seria possível em instituições dedicadas à investigação. Independentemente de considerações que se possam fazer tendo em conta comparações com o que sucedia anteriormente, é de maior importância perceber as consequências éticas, sociais e económicas que daí poderão advir e como podemos, por um lado, tirar partido das possibilidades que surgem, e por outro, reduzir eventuais riscos. Em particular, é preciso garantir que os avanços na IA não aprofundem desequilíbrios entre grupos sociais ou países, e que não dêem poder hegemónico a organizações sem que estas sejam democraticamente escrutinadas.

REFERÊNCIAS

[1] TURING, ALAN M., *Computing machinery and intelligence*. *Mind*, 59, 433-60. 1950.
[2] MONETT, D., & LEWIS, C. W., *Getting clarity by defining artificial intelligence—a survey*. In 3rd conference on" philosophy and theory of artificial intelligence. *Springer, Cham*. 212-214. 2017.
[3] MARCUS, G., & DAVIS, E., *Rebooting AI: Building artificial intelligence we can trust*. Vintage. 2019.
[4] O'NEIL, C., *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books. 2016.
[5] BROWN, T. *et al.*, *Language models are few-shot learners*. 2012.
[6] HARARI, Y. N., *Homo Deus: A brief history of tomorrow*. Random House. 2016.
[7] SALGANIK, M. J. *et al.*, *Measuring the predictability of life outcomes with a scientific mass collaboration*. *Proceedings of the National Academy of Sciences*, 117(15), 8398-8403. 2020.