Inteligência Artificial

10 Coisas Que Deve Saber

Descubra os fascinantes princípios fundamentais da IA

Tim Rocktäschel
Professor universitário
e investigador

vogais

ÍNDICE

Prefácio

Estamos a viver tempos extraordinários. A criação de inteligência artificial (IA) generativa, que durante muito tempo havia sido um mero tema de ficção científica, está mais próxima a cada ano que passa. Hoje em dia, a IA está já a ser usada ubiquamente para automatizar processos nas nossas vidas diárias. No entanto, ainda durante o nosso tempo de vida, veremos uma mudança transformacional no modo como a IA é usada em quase todos os aspetos das nossas vidas: assistentes pessoais inteligentes ajudar-nos-ão a organizarmo-nos, robôs humanoides a caminharem entre nós tornar-se-ão uma visão cada vez mais comum, a IA generativa revolucionará o modo como os conteúdos dos meios de comunicação social são criados, e a IA irá até acelerar radicalmente o próprio processo científico. Embora muitos tenham previsto que estes avanços tecnológicos seriam por fim desbloqueados, até os peritos nesta matéria estão espantados com a atual rapidez do progresso.

Como é que chegámos aqui? Este livro destina-se a todas as pessoas que querem adquirir uma compreensão básica dos fundamentos dos métodos contemporâneos de IA. Criar IA tem sido o sonho de pessoas há séculos. Analisaremos o trabalho de Alan Turing, que é frequentemente referido como o pai

da IA, e explicaremos a dificuldade em aferir se uma IA é verdadeiramente inteligente. A espinha dorsal dos atuais métodos de IA são as chamadas redes neurais artificiais. Aprenderemos por que razão esta abordagem à IA é tão versátil e por que razão tem sido tão bem-sucedida em muitas áreas da IA, da visão computacional ao processamento de linguagem natural, à aprendizagem por reforço e à robótica. Dado o progresso atual, é justo perguntar se a IA poderá, a dada altura, tornar-se mais inteligente do que qualquer ser humano no planeta ou até do que toda a humanidade junta. Será possível alcançar uma tal IA super-humana? Mesmo que seja possível, do ponto de vista teórico, haverá alguns obstáculos plausíveis ao desenvolvimento de uma IA como essa? No caso de termos sucesso, de que modo garantiremos que uma tal IA serve a humanidade e nos ajuda a vivermos vidas mais realizadas, saudáveis e felizes?

Curiosamente, os jogos de tabuleiro e os jogos de computador desempenharam um papel crucial no avanço da IA até ao momento. Iremos aprender o que faz com que os jogos sejam fantásticos bancos de ensaio para o desenvolvimento da IA, mas também por que razão treinar a IA nos jogos apenas nos pode levar até um certo ponto. Na verdade, nos últimos anos, seguiu-se um paradigma diferente: treinar a IA em conjuntos maciços de textos e pedir-lhe que preveja a palavra seguinte. Ao início, poderia parecer contraintuitivo, mas este princípio de treino simples conduziu a resultados surpreendentes. A razão subjacente é que uma IA que aprende a comprimir textos com o objetivo de prever a palavra seguinte, dado um contexto qualquer, tem de ter aprendido uma riqueza de conhecimento tremenda acerca do mundo, que vai dos factos, como, por exemplo, saber que capital pertence a que país, até capacidades cognitivas sofisticadas, como o raciocínio matemático e até a

programação. Levar esta abordagem ao extremo resultou nos chamados Modelos de Linguagem de Grande Escala (LLM, do inglês *Large Language Models*) — redes neurais artificiais que consistem em centenas de milhares de milhões de neurónios simulados cujo único trabalho é prever a palavra seguinte num dado texto. Para serem úteis como *chatbots*, estes LLM passam por etapas de treino adicionais, baseadas no *feedback* humano. Hoje em dia, os *chatbots* demonstram capacidades notáveis, tais como serem capazes de responder a uma miríade de perguntas, escrever e corrigir código a partir de apenas uma linha de comando, servir como companheiros de escrita ou até concluírem com êxito exames de entrada na universidade.

Uma das áreas de aplicação mais empolgantes para a IA é acelerar o próprio processo científico. De facto, a IA está já a ser usada para realizar novas descobertas científicas. Por exemplo, uma IA chamada AlphaFold[1] é capaz de prever estruturas proteicas tridimensionais, a partir de sequências de aminoácidos, e tem preenchido uma base de dados de 200 milhões de previsões[2]. Em 2021, a *Forbes* chamou-lhe o feito mais importante de sempre na IA[3], dado que «saber de que modo as proteínas se dobram é, ao mesmo tempo, ridiculamente difícil e absolutamente essencial para compreender os processos biológicos». Desde então, a IA tem sido usada para, entre outras coisas, descobrir programas de computadores novos e eficientes[4], ajudar a conceber materiais novos[5] e prever as condições meteorológicas com precisão[6], para mencionar apenas alguns exemplos.

Olhando para lá da aplicação da IA a problemas científicos específicos, podemos perguntar-nos se a IA poderia até mesmo, de forma autónoma, aplicar o método científico e criar um processo ilimitado de criação de conhecimento no futuro. Se isso fosse possível, poderíamos até deixar que essa IA usasse as suas

pesquisas científicas autónomas em si mesma, criando um circuito do chamado autoaperfeiçoamento autorreferencial. Com base no atual conhecimento científico, parece que estamos a aproximar-nos de um mundo em que a taxa de progresso da IA está a aumentar rapidamente, fazendo com que seja difícil alguém prever o que será tecnicamente possível no espaço de um ano, quanto mais de uma década. Apesar desta aceleração, a IA continua a enfrentar limitações muito reais hoje em dia. Embora os *chatbots* tenham conquistado o mundo, ainda não temos robôs humanoides de uso genérico que nos possam ajudar nas nossas tarefas diárias ou que possam automatizar os trabalhos físicos mais perigosos do mundo. A razão disso é o acesso limitado a dados de treino de alta qualidade. Textos, imagens e vídeos abundam na Internet, mas não os dados de controlo sensório-motor.

O futuro da IA é empolgante, mas com os rápidos avanços tecnológicos também advêm riscos significativos. Quais são alguns dos riscos mais críticos a curto, médio e longo prazo associados a uma IA cada vez mais capaz e de que modo poderemos mitigá-los?

Este livro foi escrito para o leitor curioso que quer aprender mais acerca do empolgante campo da IA. Omiti a maioria das explicações matemáticas ou formais da IA. Este livro não exige nenhum conhecimento de ciência da computação, matemática ou IA. Espero que venha a constatar que este livro é uma introdução suave a uma das tecnologias mais transformadoras da humanidade. Além disso, espero que desperte a sua curiosidade para aprender mais acerca deste assunto, e que o equipe com o conhecimento básico das tecnologias contemporâneas de IA, de modo a que possa acompanhar o seu rápido progresso ao longo dos próximos anos.

Referências

1. Jumper, J. *et al*. «Highly accurate protein structure prediction with AlphaFold», *Nature* 596, 583-589 (2021).
2. Varadi, M. *et al*. «AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models», *Nucleic Acids Research* 50, D439-D444 (2022).
3. Toews, R. «AlphaFold Is the Most Important Achievement In AI-Ever», *Forbes* https://www.forbes.com/sites/robtoews/2021/10/03/alphafold-is-the-most-important-achievement-in-aiever/.
4. Romera-Paredes, B. *et al*. «Mathematical discoveries from program search with large language models», *Nature* 625, 468-475 (2024).
5. Ward, L., Agrawal, A., Choudhary, A. e Wolverton, C. «A general-purpose machine learning framework for predicting properties of inorganic materials», *Npj Computational Materials* 2, 16028 (2016).
6. Price, I. *et al*. «GenCast: Diffusion-based ensemble forecasting for medium-range weather». Pré-publicação em https://doi.org/10.48550/arXiv.2312.15796 (2024).

1

O Que É a Inteligência Artificial?

Durante milénios, as pessoas têm sonhado em criar seres artificiais que possam automatizar várias tarefas que normalmente necessitariam de inteligência e capacidades humanas. Na *Ilíada* de Homero (século VIII a. C.), Hefesto — entre outras coisas o deus grego da metalurgia, da escultura e dos ferreiros — cria autómatos de metal e servos de ouro para o ajudarem nas suas tarefas. Prometeu cria pessoas a partir do barro. Do mesmo modo, no folclore judaico, Golem é criado a partir do barro. Aristóteles, na sua obra *Política* (século IV a. C.), prenuncia o advento de ferramentas automáticas que poderiam tornar o trabalho desnecessário. O texto taoista *Liezi* (século IV) menciona a criação de um homem artificial pelo engenheiro e artesão mestre Yan Shi. O livro *As Viagens de Gulliver* (1726), de Jonathan Swift, descreve «O Mecanismo» com o qual até «a pessoa mais ignorante […] poderá escrever livros de filosofia, poesia, política, direito, matemática e teologia, sem o mais pequeno auxílio da genialidade ou do estudo». Até há pouco tempo, O Mecanismo teria sido considerado ficção científica, mas graças a avanços na inteligência artificial, está a tornar-se uma realidade.

Em 1950, Alan Turing publicou o seu artigo seminal «Computing Machinery and Intelligence»[1], no qual colocava a questão: «Poderão as máquinas pensar?» É difícil definir formalmente

«pensar», ao ponto de poder ser usado como um critério objetivo para avaliar se uma IA está a pensar. Logo, para aferir esta questão, Turing propõe o «Jogo da Imitação», que mais tarde ficaria conhecido como o famoso «Teste de Turing». Este teste funciona da seguinte maneira: uma pessoa atua como o interrogador que está encarregado de falar com duas outras pessoas, A e B, através de uma interface de conversação por texto. A é um homem e B é uma mulher. O objetivo do interrogador é descobrir qual é a mulher e qual é o homem. No entanto, adicionalmente, a pessoa A é substituída por uma IA. Se o interrogador humano cometer tantos erros quando A e B são pessoas como quando A é uma IA e B é uma pessoa, então diz-se que essa IA passou o Teste de Turing. Uma interpretação mais moderna, poderá simplesmente encarregar o interrogador de descobrir se ou A ou B são uma IA. Embora este teste tenha servido de norte aos investigadores durante décadas, também realçou o problema de medir o grau em que uma IA é inteligente, dado que o resultado do teste poderá depender de muitos fatores. Por exemplo, será o interrogador uma pessoa escolhida ao acaso ou um perito treinado? Durante quanto tempo poderá o interrogador conversar com A e B? Haverá tópicos interditos? Se uma IA passar o Teste de Turing, terá ela demonstrado uma inteligência genuína de nível humano ou terá antes apenas demonstrado que consegue enganar pessoas e fingir algo tão bem quanto os seres humanos?

Em 1956, John McCarthy, Marvin Minsky, Nathaniel Rochester e Claude Shannon organizaram o Projeto de Investigação de Verão de Dartmouth acerca da Inteligência Artificial. Pensa-se que o termo «Inteligência Artificial» terá sido cunhado por John McCarthy nessa altura. Na sua proposta[2], conjeturam que «todos os aspetos da aprendizagem ou qualquer outra

característica da inteligência poderão, em princípio, ser descritos de forma tão precisa que uma máquina poderá ser construída para os simular. Será feita uma tentativa para descobrir como fazer com que as máquinas usem linguagem, formem abstrações e conceitos, resolvam tipos de problemas agora reservados aos seres humanos e se aperfeiçoem a si mesmas. Pensamos que se poderá alcançar um avanço significativo em um ou mais destes problemas se um grupo cuidadosamente selecionado de cientistas trabalhar neles em conjunto durante um verão». Embora o seu otimismo em relação à rapidez do progresso em IA na altura seja digno de celebrar, a sua intuição estava certa. Especularam que «uma grande parte do pensamento humano consiste em manipular palavras de acordo com regras de raciocínio e regras de conjetura» e perguntaram-se a si mesmos de que modo um computador poderia ser programado para usar a linguagem, de que modo «redes neurais» poderiam ser organizadas para formar conceitos — se a verdadeira IA iria realizar o autoaperfeiçoamento, de que modo poderão as máquinas formar abstrações a partir de dados do mundo real e em que medida a aleatoriedade desempenha um papel no pensamento criativo. Agora, passados quase setenta anos, temos redes neurais artificiais a simularem centenas de milhares de milhões de neurónios que aprendem e processam e criam linguagem natural. Nesse processo, estes chamados Modelos de Linguagem de Grande Escala adquirem capacidades sofisticadas, tais como serem capazes de traduzir entre idiomas, realizar um raciocínio matemático e escrever programas de computador.

Curiosamente, desde Turing e o *workshop* de Dartmouth, a IA tem sido uma meta em movimento. Durante milénios, a proficiência no xadrez tem sido associada a grande inteligência nas

pessoas. Por conseguinte, alguns acreditavam que uma vez que os computadores conseguem jogar xadrez, deveriam ser verdadeiramente inteligentes. Tal como veremos no Capítulo 4, desde o Deep Blue, em 1997, a IA é super-humana a jogar xadrez. No entanto, o Deep Blue é um método de pesquisa relativamente simples, resultando numa IA que apenas é boa numa coisa e em nada mais, nomeadamente, jogar xadrez. Após 1997, a IA também se tornou proficiente a jogar *go*, *Diplomacy* e *Stratego*, bem como os videojogos da Atari, *StarCraft II* e *Dota 2*. Anteriormente a cada um destes avanços, alguns poderão ter previsto que assim que a IA fosse capaz de fazer isto ou aquilo, teria de ser verdadeiramente inteligente. No entanto, nenhuma destas IA demonstrou uma capacidade de generalização fora do seu domínio particular.

Em IA, geralmente, distinguimos entre IA «fraca», de um domínio específico e estreito e a IA «forte», de um objetivo geral. A IA fraca é desenvolvida para desempenhar uma tarefa particular. Poderá aprender a tornar-se super-humana nessa tarefa particular, mas não será útil para qualquer outra coisa. Por outro lado, supõe-se que a IA forte seja capaz de aprender a realizar qualquer tarefa que um ser humano consiga desempenhar. Outro termo para tal IA é «Inteligência Artificial Geral», ou AGI (do inglês *Artificial General Intelligence*). Assim que uma tal AGI se tornar melhor do que um ser humano numa tarefa, chamar-lhe-emos Inteligência Artificial Super-humana ou ASI (do inglês *Artificial Superhuman Intelligence*)[3]. As IA, tal como as que se podem ver em *Blade Runner — Perigo Iminente*, *O Exterminador Implacável*, *Matrix* ou *Uma História de Amor*, poderiam provavelmente ser classificadas como ASI. Quando a IA se torna super-humana na ficção científica, frequentemente, torna-se distópica para as pessoas muito

rapidamente. No final deste livro, analisaremos as oportunidades e os riscos associados a avançar no sentido de uma IA mais generativa.

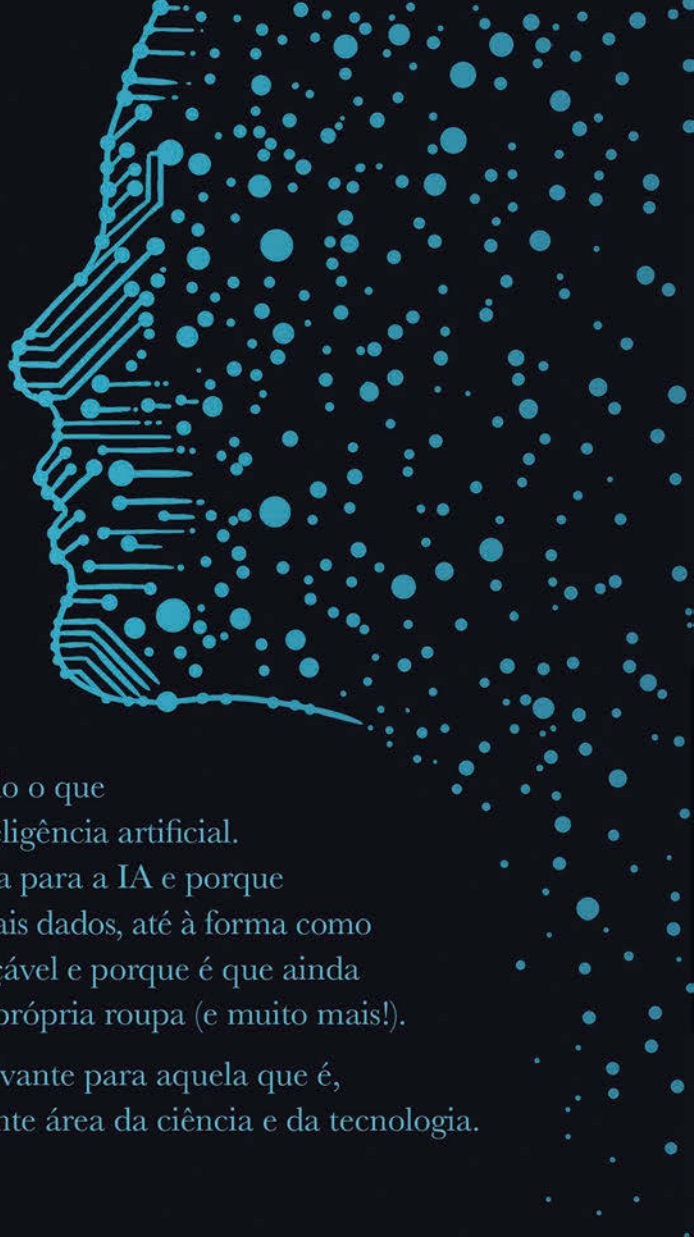
Pessoalmente, interessei-me pela IA na adolescência quando jogava o videojogo *Creatures*, de 1996. Foi revolucionário, devido ao uso de tecnologia de vida artificial. Enquanto jogadores, criamos, treinamos e reproduzimos pequenas criaturas simuladas chamadas Norns. Os Norns são controlados por um sistema de rede neural, que lhes permite aprenderem a partir do seu ambiente e das suas experiências, o que poderá resultar em comportamentos sem par. Lembro-me de ter lido um artigo na altura que alegava que no computador de alguém estes Norns descobriram que é mais divertido passar repetidas vezes uma bola de um para o outro do que brincar individualmente com ela. Segundo o artigo, isto era surpreendente, pois embora os programadores do jogo tivessem implementado um algoritmo para os Norns aprenderem a chegar à bola, pegar nela e atirá-la, não implementaram, de forma explícita, o comportamento de diversos Norns a brincarem juntos com uma bola. Este surgimento de um comportamento novo tem-me fascinado desde então. Acho que foi também por essa altura que assisti ao episódio de *Os Simpsons*, «The Genesis Tub» («A Selha do Génesis») no qual Lisa efetua uma experiência para a feira de ciências em que tenta dissolver o seu dente de leite numa placa de Petri, usando refrigerante de cola e choques elétricos. Em vez de dissolver o dente, a sua experiência cria formas de vida artificiais que rapidamente evoluem para uma sociedade tecnologicamente avançada. Na verdade, existe uma comunidade de investigadores dedicada a estudar a criação de formas de vida artificiais em simulação. O exemplo mais recente é Lenia, um sistema de autómatos celulares

sob a forma de um relaxamento contínuo do popular Jogo da Vida, de Conway[4]. O sistema resultou na evolução de «mais de 400 espécies em 18 famílias», demonstrando, ao mesmo tempo, vários «indícios de um sistema vivo»: auto-organização, autorregulação, autodireção, adaptabilidade e capacidade de evolução.

No prefácio, afirmei que estamos a viver tempos extraordinários. Estou certo de que todos os estudiosos, desde o Iluminismo, disseram o mesmo acerca do seu tempo. A tecnologia progrediu rapidamente, ao longo dos últimos séculos. No entanto, pela primeira vez na história humana, estamos a fazer a transição de uma IA fraca para as formas iniciais de uma AGI. Esta transição foi viabilizada através da renúncia ao treino da IA em ambientes de simulação estreitos, tais como jogos específicos, optando-se, em vez disso, por treinar a IA em quantidades maciças de dados gerados pelos seres humanos, tais como textos, imagens e vídeos, na Internet. Embora se estejam a tornar cada vez mais úteis, os atuais sistemas de IA ainda não são perfeitos. Cometem erros e, quando o fazem, estes erros poderão por vezes ser extremamente ridículos e violar o bom senso básico humano. No entanto, a IA será aperfeiçoada e, a dada altura, irá aperfeiçoar-se a si mesma. O que hoje poderá parecer um fosso intransponível, no que diz respeito às capacidades, irá, por fim, diminuir.

Referências

1. Turing, A. M. «Computing Machinery and Intelligence», *Mind LIX*, 433-460 (1950).
2. McCarthy, J., Minsky, M. L., Rochester, N. e Shannon, C. E. «A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955», *AI Mag.* 27, 12-12 (2006).
3. Morris, M. R. *et al*. «Levels of AGI: Operationalizing Progress on the Path to AGI». Pré-publicação em https://doi.org/10.48550/arXiv.2311.02462 (2023).
4. Chan, B. W.-C. Lenia «Biology of Artificial Life», *Complex Systems* 28, 251-286 (2019).



Ainda durante o nosso tempo de vida veremos uma mudança transformacional no modo como a Inteligência Artificial é usada em quase todos os aspetos das nossas vidas.

Em dez breves capítulos, o professor de IA na University College London, Tim Rocktäschel, revela tudo o que precisa de saber sobre a inteligência artificial. Desde o que o futuro reserva para a IA e porque continua a melhorar com mais dados, até à forma como a IA sobre-humana é alcançável e porque é que ainda temos de… dobrar a nossa própria roupa (e muito mais!).

Um livro esclarecedor e cativante para aquela que é, atualmente, a mais importante área da ciência e da tecnologia.

«Uma excelente e extremamente atualizada panorâmica da mais importante revolução tecnológica da história da humanidade.»

JEFF CLUNE,
cientista informático

Penguin Random House Grupo Editorial

www.penguinlivros.pt
penguinlivros

ISBN 9789895833849