

## CAPÍTULO 1

# A CONTENÇÃO NÃO É POSSÍVEL

Quase todas as culturas têm um mito do dilúvio.

Nos antigos textos hindus, o primeiro homem do nosso universo, Manu, é avisado de um dilúvio iminente e torna-se o único sobrevivente. A *Epopéia de Gilgamesh* regista o deus Enlil a destruir o mundo numa cheia gigantesca, uma história que terá eco em qualquer pessoa familiarizada com a narrativa da arca de Noé do Antigo Testamento. Platão falou da cidade perdida da Atlântida, levada por uma imensa torrente. Nas tradições orais e nos escritos antigos da humanidade encontra-se a ideia de uma onda gigante que varre tudo no seu caminho, deixando o mundo refeito e renascido.

As cheias também marcam a história num sentido literal – o volume sazonal dos grandes rios do mundo, a subida dos oceanos após o fim da Idade do Gelo, o raro choque de um maremoto que surge sem aviso no horizonte. O asteroide que matou os dinossauros criou uma onda com mais de um quilómetro de altura e alterou o curso da evolução. O poder absoluto destes vagalhões ficou gravado na consciência coletiva: muros de água, imparáveis, incontrolláveis, irreprimíveis. São das forças mais possantes do planeta. Moldam continentes, irrigam as sementeiras do mundo e alimentam o crescimento da civilização.

Outros tipos de vagas têm sido igualmente transformadores. Se analisarmos de novo a história, vemos que está marcada por uma série de ondas metafóricas: ascensão e queda de impérios e religiões, explosões de comércio. Pensemos no cristianismo ou no islamismo, religiões que começaram por fazer pequenas ondas até se erguerem e abaterem sobre enormes extensões da Terra. Ondas como esta são um motivo recorrente, enquadrando as marés da história, grandes lutas de poder, e altos e baixos económicos.

O surgimento e a disseminação das tecnologias também assumiram a forma de ondas que mudam o mundo. Uma única tendência dominante tem resistido à prova do tempo desde a descoberta do fogo e dos utensílios de pedra, as primeiras tecnologias aproveitadas pela nossa espécie. Quase todas as tecnologias basilares jamais inventadas, das picaretas aos arados, da cerâmica à fotografia, dos telefones aos aviões, seguem uma só lei, aparentemente imutável: fica mais barata e mais fácil de usar, e acaba por proliferar por todo o lado.

Esta proliferação da tecnologia em ondas é a história do *Homo technologicus* – do animal tecnológico. A demanda de aperfeiçoamento da humanidade – nós próprios, a nossa sina, as nossas capacidades e a nossa influência no ambiente – tem impulsionado uma evolução incessante de ideias e criação. A invenção é um processo emergente, em constante evolução e expansão, impulsionado por inventores, académicos, empreendedores e líderes, auto-organizados e extremamente competitivos, cada qual avançando com motivações próprias. Este ecossistema de invenção abraça, por omissão, a expansão. É a natureza inerente da tecnologia.

A questão é: o que acontece a partir daqui? Nas páginas que se seguem, conto a história da próxima grande onda da história.

Olhe à sua volta.

O que vê? Móveis? Edifícios? Telefones? Comida? Um parque ajardinado? Quase todos os objetos na sua linha de visão foram, com toda a probabilidade, criados ou alterados pela inteligência humana. A linguagem – os alicerces das interações sociais, das culturas, das organizações políticas e, talvez, do que significa ser humano – é outro produto e motor da nossa inteligência. Cada princípio e conceito abstrato, cada pequeno esforço ou projeto criativo, cada encontro na vida do leitor, foi mediado pela capacidade única e infinitamente complexa da nossa espécie para imaginar, criar e raciocinar. O engenho humano é uma coisa espantosa.

Só há outra força tão omnipresente neste quadro: a própria vida biológica. Antes da era moderna, além de algumas rochas e minerais, a maioria dos artefactos humanos – de casas de madeira a roupa de algodão e a fogos de carvão – provinha de coisas que já

estiveram vivas. Tudo o que entrou no mundo desde então vem de nós, do facto de sermos seres biológicos.

Não é exagero dizer que a totalidade do mundo humano depende dos sistemas vivos ou da nossa inteligência. No entanto, ambos estão atualmente num momento sem precedentes de inovação exponencial e convulsão, um aumento sem paralelo que pouco deixará inalterado. Começa a abater-se à nossa volta uma nova onda de tecnologia. Esta onda liberta o poder de criar estes dois fundamentos universais: uma onda de, nada mais nada menos, inteligência e vida.

A onda iminente define-se por duas tecnologias nucleares: a inteligência artificial (IA) e a biologia sintética. Juntas, darão início à nova aurora da humanidade, gerando riqueza e excedentes como nunca se viu. Porém, a sua rápida proliferação também ameaça dar poder a um conjunto diversificado de maus intervenientes para desencadear perturbações, instabilidade, e até catástrofes a uma escala indizível. Esta onda cria um imenso desafio que definirá o século XXI: o nosso futuro depende dessas tecnologias e é ameaçado por elas.

Do ponto em que nos encontramos, parece que conter esta onda – isto é, controlar, refrear ou mesmo parar – não é possível. Neste livro, pergunta-se porque pode isso ser verdade e o que significa se for. As implicações destas questões afetarão todas as pessoas vivas e cada geração vindoura.

Acredito que esta nova onda de tecnologia está a levar a história humana a um ponto de viragem. Se a sua contenção for impossível, as consequências para a nossa espécie serão dramáticas, potencialmente terríveis. Do mesmo modo, sem os seus frutos, ficamos expostos e precários. É um argumento que já apresentei muitas vezes na última década, à porta fechada, mas à medida que os impactos se tornam pouco a pouco mais incontornáveis, é a altura de o defender publicamente.

## O DILEMA

Contemplar o poder profundo da inteligência humana levou-me a uma pergunta simples, que consome a minha vida desde então:

suponhamos que conseguimos destilar a essência do que torna os seres humanos tão produtivos e capazes num *software*, num algoritmo? Encontrar a resposta poderia desbloquear ferramentas indizivelmente possantes para ajudar a resolver os nossos problemas mais difíceis. Poderá estar aqui uma ferramenta, impossível mas extraordinária, para nos ajudar a ultrapassar os grandes desafios das próximas décadas, das alterações climáticas ao envelhecimento da população e à alimentação sustentável.

Com isto em mente, num pitoresco escritório da era da Regência com vista para a Russell Square em Londres, cofundei uma empresa chamada DeepMind, com dois amigos, Demis Hassabis e Shane Legg, no verão de 2010. Este era o nosso objetivo, que, em retrospectiva, ainda parece tão ambicioso, louco e esperançoso como na altura: replicar aquilo que nos torna únicos enquanto espécie: a inteligência.

Para atingir este objetivo, teríamos de criar um sistema que pudesse imitar e, com o tempo, superar todas as capacidades cognitivas humanas, da visão e fala ao planeamento e à imaginação e, por fim, empatia e criatividade. Uma vez que tal sistema ganharia com o processamento paralelo maciço dos supercomputadores, e da explosão de novas e vastas fontes de dados de toda a Web aberta, sabíamos que mesmo um modesto progresso em direção a este objetivo teria profundas implicações sociais.

Na altura, parecia algo do outro mundo. A adoção generalizada da inteligência artificial era coisa de devaneios, mais fantasia do que facto, domínio de alguns académicos enclausurados e fãs de ficção científica deslumbrados. Ora, enquanto escrevo isto e penso na última década, o progresso da IA tem sido nada menos do que espantoso. A DeepMind tornou-se uma das principais empresas de IA do mundo, conseguindo uma série de descobertas. A velocidade e o poder desta nova revolução têm surpreendido, mesmo aqueles de nós que estão mais próximos da sua vanguarda. Enquanto escrevi este livro, o ritmo de progresso da IA foi impressionante, com novos modelos e novos produtos a serem lançados todas as semanas, por vezes todos os dias. É evidente que esta onda está a acelerar.

Atualmente, os sistemas de IA conseguem reconhecer, quase na perfeição, rostos e objetos. A transcrição de voz para texto e a

tradução linguística instantânea são um dado adquirido. A IA consegue navegar nas estradas e no trânsito suficientemente bem para conduzir de forma autónoma em determinadas situações. Com base em algumas instruções simples, uma nova geração de modelos de IA pode gerar imagens e compor textos com níveis extraordinários de pormenor e coerência. Os sistemas de IA são capazes de produzir vozes sintéticas com um realismo insólito, e compor música de uma beleza estonteante. Mesmo em domínios mais difíceis, que há muito se pensava estarem exclusivamente adaptados às capacidades humanas, como, por exemplo, planeamento a longo prazo, imaginação, e simulação de ideias complexas, os progressos são cada vez maiores.

Há décadas que a IA sobe a escada das capacidades cognitivas, e parece preparada para atingir um nível humano numa vasta gama de tarefas nos próximos três anos. É uma grande alegação, mas, se eu estiver quase correto, as implicações são na verdade profundas. Quando fundámos a DeepMind, aquilo que parecia quixotesco tornou-se não só plausível, mas aparentemente inevitável.

Desde o início, era claro para mim que a IA seria uma ferramenta poderosa para um bem extraordinário, mas, como a maioria das formas de poder, também repleta de imensos perigos e dilemas éticos. Há muito que me preocupo com as consequências do avanço da IA, mas de igual modo com o rumo de todo o ecossistema tecnológico. Estava em curso uma revolução mais ampla, com a IA a alimentar uma geração poderosa e emergente de tecnologias genéticas e robótica. Os progressos numa área aceleram as outras, num processo caótico e catalisador transversal, fora do controlo direto de quem quer que seja. Evidenciava-se que, se nós ou outros fôssemos bem-sucedidos na replicação da inteligência humana, não seria apenas um negócio lucrativo, como de costume, mas também uma mudança sísmica para a humanidade, inaugurando uma era em que oportunidades inéditas acarretariam riscos inéditos.

Com o avanço da tecnologia ao longo dos anos, as minhas preocupações só aumentaram. Suponhamos que a onda é, de facto, um maremoto?

Em 2010, quase ninguém falava seriamente sobre IA. No entanto, o que antes parecia uma missão de nicho para um pequeno grupo de investigadores e empresários tornou-se agora um vasto empreendimento global. A IA está em todo o lado, nas notícias e no *smartphone* do leitor, a negociar ações e a criar sítios na Web. Muitas das maiores empresas, e das nações mais ricas, do mundo avançam a passos largos, desenvolvendo modelos de IA e técnicas de engenharia genética de ponta, alimentadas por dezenas de biliões de dólares de investimento.

Uma vez amadurecidas, estas tecnologias emergentes espalhar-se-ão rapidamente, tornando-se mais baratas, mais acessíveis e amplamente difundidas por toda a sociedade. Oferecerão novos e extraordinários avanços na medicina e na energia limpa, criando não só empresas, mas também indústrias e melhorias na qualidade de vida em quase todos os domínios imagináveis.

No entanto, além destes benefícios, a IA, a biologia sintética e outras formas avançadas de tecnologia produzem riscos extremos a uma escala demasiado preocupante. Podem representar uma ameaça existencial para os Estados-nação – riscos tão profundos que podem perturbar ou mesmo derrubar a atual ordem geopolítica. Abrem caminho a imensos ciberataques e guerras automatizadas, com a força motriz da IA, capazes de devastar países; pandemias engendradas; e um mundo sujeito a forças inexplicáveis mas, aparentemente, onnipotentes. A probabilidade de cada qual pode ser ínfima, mas as consequências possíveis são enormes. Mesmo a mínima hipótese de desfechos como estes exige atenção urgente.

Certos países reagirão à possibilidade de tais riscos catastróficos com um autoritarismo carregado de tecnologia para abrandar a propagação destes novos poderes. Isto exigirá enormes níveis de vigilância, juntamente com intrusões maciças na nossa vida privada. Manter a tecnologia sob rédea curta pode vir a fazer parte de uma deriva para tudo e toda a gente a ser vigiada, a toda a hora, num sistema global distópico, justificado pelo desejo de nos precavermos contra os desfechos mais extremos possíveis.

Também plausível é uma reação ludita. Seguir-se-ão proibições, boicotes e moratórias. Será possível deixar de desenvolver novas tecnologias e introduzir uma série de moratórias? É pouco

provável. Com o seu enorme valor geoestratégico e comercial, é difícil ver como é que os Estados-nação ou as multinacionais se deixam convencer a abdicar unilateralmente do poder transformador desencadeado por estas descobertas. Além disso, a tentativa de proibir o desenvolvimento de novas tecnologias é, em si mesma, um risco: as sociedades tecnologicamente estagnadas são historicamente instáveis e propensas ao colapso. Perdem a capacidade de resolver problemas, de progredir.

Tanto a prossecução como a não prossecução das novas tecnologias estão, a partir daqui, repletas de riscos. As probabilidades de se percorrer uma «via estreita» e evitar um ou outro resultado – distopia tecnoautoritária, por um lado, catástrofe induzida pela abertura, por outro – diminuem ao longo do tempo consoante a tecnologia desencarece, ganha poder, fica mais insidiosa e os riscos acumulam-se. No entanto, afastarmo-nos também não é opção. Mesmo que nos preocupemos com os riscos, precisamos mais do que nunca dos incríveis benefícios das tecnologias da onda iminente. Este é o dilema nuclear: que, mais cedo ou mais tarde, uma poderosa geração de tecnologia conduza a humanidade rumo a resultados catastróficos ou distópicos. Creio que este é o grande metaproblema do século XXI.

Este livro descreve porque está esta situação terrível a tornar-se inevitável, e explora a forma de a enfrentarmos. De algum modo, precisamos de tirar o melhor partido da tecnologia, essencial para enfrentar um conjunto imperioso de desafios globais, e também para sair do dilema. O atual discurso em torno da ética e da segurança tecnológicas é insuficiente. Apesar dos muitos livros, debates, publicações em blogues e desgarradas no Twitter sobre tecnologia, raramente se ouve falar em *contenção*. Vejo nisto um conjunto interligado de mecanismos técnicos, sociais e legais que restringem e controlam a tecnologia, e que funcionam a todos os níveis possíveis: um meio, em teoria, de fugir ao dilema. Contudo, mesmo os críticos mais severos tendem a esquivar-se a esta linguagem de forte contenção.

Isto tem de mudar. Espero que este livro mostre porquê e indique como.

## A ARMADILHA

Alguns anos depois de fundarmos a DeepMind, criei um conjunto de diapositivos sobre os potenciais impactos económicos da IA a longo prazo. Apresentei-o a uma dúzia dos fundadores, CEO e tecnólogos mais influentes da indústria tecnológica numa elegante sala de reuniões da Costa Oeste, e argumentei que a IA introduziu uma série de ameaças que exigem respostas proativas. Pode levar a invasões maciças de privacidade ou desencadear um apocalipse de desinformação. É passível de ser transformada em arma, criando um conjunto letal de novas armas cibernéticas, introduzindo mais vulnerabilidades no nosso mundo ligado em rede.

Sublinhei também o potencial da IA para deixar um grande número de pessoas no desemprego. Pedi à sala que considerasse a longa história da automatização e da mecanização quanto a tornarem a mão de obra supérflua. Primeiro surgem formas mais eficientes de realizar tarefas específicas, e depois funções inteiras tornam-se redundantes e, em breve, setores inteiros exigem muito menos trabalhadores. Nas próximas décadas, argumentei, os sistemas de IA substituirão o «trabalho manual intelectual» da mesma forma e, decerto, muito antes de os robôs substituírem o trabalho braçal. No passado, os novos empregos eram criados ao mesmo tempo que antigos ficavam obsoletos, mas suponhamos que a IA pode, simplesmente, fazer a maioria desses trabalhos? Sugeri que havia poucos precedentes para as novas formas de poder concentrado que estavam iminentes. Mesmo que parecessem distantes, aproximavam-se da sociedade ameaças potencialmente graves.

No último diapositivo, mostrei uma imagem de *Os Simpsons*. Na cena, os habitantes de Springfield revoltam-se, e o elenco de personagens familiares avança com bastões e tochas. A mensagem era clara, mas eu destaquei-a. «As forquilha vêm aí.» Vêm atrás de nós, os criadores de tecnologia. Cabia-nos garantir que o futuro fosse melhor do que isto.

À volta da mesa, deparei-me com olhares vazios. A sala não se convencera. A mensagem não fora assimilada. O descaso abateu-se sobre mim. Porque não mostravam os indicadores económicos



sinal algum do que eu dizia? A IA iria estimular nova procura que, por seu turno, criaria novos empregos. Aumentaria e capacitaria as pessoas a serem ainda mais produtivas. Talvez houvesse alguns riscos, admitiram, mas não eram assim tão maus. As pessoas eram inteligentes. Sempre se encontraram soluções. *Não se preocupem, parecia que pensavam, passemos à apresentação seguinte.*

Alguns anos depois, perto do início da pandemia da covid-19, assisti, numa universidade famosa, a um seminário sobre riscos tecnológicos. A situação era parecida: outra mesa grande, novo debate de alto coturno. Ao longo do dia, foi apresentada uma série de riscos de arrepiar os cabelos, junto com cafés, biscoitos e PowerPoints.

Houve um risco que se destacou. O apresentador mostrou como o preço dos sintetizadores de ADN, que podem imprimir cadeias de ADN à medida, baixava rapidamente. Custando algumas dezenas de milhares de dólares,<sup>1</sup> são pequenos o suficiente para ficarem numa bancada na garagem e deixarem as pessoas sintetizar – ou seja, *fabricar* – ADN. Tudo isto é agora possível para quem tiver formação superior em biologia ou um autodidata na Internet.

Dada a crescente disponibilidade das ferramentas, o apresentador pintou uma visão assustadora: em breve, alguém poderá criar novos agentes patogénicos, muito mais transmissíveis e letais do que qualquer coisa encontrada na natureza. Estes patogénicos sintéticos podem escapar às contramedidas conhecidas, alastrar de forma assintomática, ou ter resistência intrínseca aos tratamentos. Se necessário, pode-se complementar experiências caseiras com ADN encomendado na Internet e montado em casa. O apocalipse, encomendado pelo correio.

Isto não era ficção científica, argumentou o apresentador, um professor respeitado com mais de duas décadas de experiência. Era agora um risco vivo. Terminou com um pensamento alarmante: uma pessoa tem hoje, provavelmente, «a capacidade de matar um bilião de pessoas». Só precisa de motivação.

Os participantes mexeram-se nas cadeiras, incomodados. Contorceram-se e tossiram. Depois começaram as queixas e as hesitações. Ninguém queria acreditar que fosse possível. Decerto não era o caso, decerto haveria mecanismos de controlo eficazes,

decerto as doenças eram difíceis de criar, decerto as bases de dados podiam ser bloqueadas, decerto o *hardware* podia ser protegido. E assim por diante.

A reação coletiva foi mais do que apenas desdenhosa. Recusaram-se a aceitar a visão do apresentador. Ninguém queria confrontar as implicações dos factos concretos e das probabilidades frias que tinha ouvido. Fiquei em silêncio, francamente abalado. O seminário terminou, fomos todos jantar fora e continuámos a conversar, como de costume. Tínhamos passado o dia a falar do fim do mundo, mas ainda havia piza para comer, piadas para contar, um escritório onde trabalhar e, além disso, alguma coisa iria aparecer, ou alguma parte do argumento só poderia estar errada. Juntei-me a eles.

Mas a apresentação atormentou-me durante meses. Porque não levava eu, porque não levávamos todos, aquilo mais a sério? Porque nos esquivamos, confrangidos, a mais debates? Porque recorrem alguns ao sarcasmo e acusam quem levanta estas questões de catastrofizar ou de «ignorar o maravilhoso bem» da tecnologia? A esta reação emocional generalizada, que eu estava a observar, chamei «armadilha da aversão ao pessimismo»: a análise equivocada que surge quando se é assoberbado pelo medo de confrontar realidades potencialmente tenebrosas, e a conseqüente tendência para fazer vista grossa.

Quase toda a gente tem uma versão desta reação, e a consequência leva-nos a ignorar uma série de tendências críticas que se desenrolam à frente dos nossos olhos. É quase uma reação fisiológica inata. A nossa espécie não está preparada para lidar com a transformação a esta escala, quanto mais com a possibilidade de a tecnologia nos deixar mal deste modo. Experimentei este sentimento ao longo da minha carreira, e vi muitos outros terem a mesma reação visceral. Confrontar este sentimento é um dos objetivos deste livro. Olhar friamente para os factos, por mais incómodos que sejam.

Lidar corretamente com esta onda, conter a tecnologia e assegurar que ela esteja sempre ao serviço da humanidade significam superar a aversão ao pessimismo, encarar a realidade do que está iminente.