



OS NOVE TITÃS DA IA

Como as Gigantes da Tecnologia
e Suas Máquinas Pensantes
Podem Subverter a Humanidade

AMY WEBB


ALTA BOOKS
EDITORA
Rio de Janeiro, 2020

BAIDU

APPLE

ALIBABA

IBM

MICROSOFT

GOOGLE

AMAZON

FACEBOOK

TENCENT

Os Nove Titãs da IA

Copyright © 2020 da Starlin Alta Editora e Consultoria Eireli. ISBN: 978-85-508-1073-7

Translated from original The Big Nine. Copyright © 2019 by Amy Webb. ISBN 9781541773752. This translation is published and sold by permission of PublicAffairs an imprint of Perseus Books, LLC, a subsidiary of Hachette Book Group, Inc., the owner of all rights to publish and sell the same. PORTUGUESE language edition published by Starlin Alta Editora e Consultoria Eireli, Copyright ©2020 by Starlin Alta Editora e Consultoria Eireli.

Todos os direitos estão reservados e protegidos por Lei. Nenhuma parte deste livro, sem autorização prévia por escrito da editora, poderá ser reproduzida ou transmitida. A violação dos Direitos Autorais é crime estabelecido na Lei nº 9.610/98 e com punição de acordo com o artigo 184 do Código Penal.

A editora não se responsabiliza pelo conteúdo da obra, formulada exclusivamente pelo(s) autor(es).

Marcas Registradas: Todos os termos mencionados e reconhecidos como Marca Registrada e/ou Comercial são de responsabilidade de seus proprietários. A editora informa não estar associada a nenhum produto e/ou fornecedor apresentado no livro.

Impresso no Brasil — Edição revisada conforme o Acordo Ortográfico da Língua Portuguesa de 2009.

Produção Editorial Editora Alta Books	Produtor Editorial Illysabelle Trajano Juliana de Oliveira Thiê Alves	Marketing Editorial Livia Carvalho marketing@altabooks.com.br	Editor de Aquisição José Rugeri j.rugeri@altabooks.com.br
Gerência Editorial Anderson Vieira	Assistente Editorial Thales Silva	Coordenação de Eventos Viviane Paiva eventos@altabooks.com.br	
Gerência Comercial Daniele Fonseca			
Equipe Editorial Ian Verçosa Maria de Lourdes Borges Raquel Porto Rodrigo Dutra	Equipe de Design Larissa Lima Paulo Gomes		
Tradução Cibelle Ravaglia	Copidesque Fernanda Lufi	Revisão Gramatical Hellen Suzuki Thaís Pol	Diagramação Luisa Maria Gomes

Publique seu livro com a Alta Books. Para mais informações envie um e-mail para autoria@altabooks.com.br

Obra disponível para venda corporativa e/ou personalizada. Para mais informações, fale com projetos@altabooks.com.br

Erratas e arquivos de apoio: No site da editora relatamos, com a devida correção, qualquer erro encontrado em nossos livros, bem como disponibilizamos arquivos de apoio se aplicáveis à obra em questão.

Acesse o site www.altabooks.com.br e procure pelo título do livro desejado para ter acesso às erratas, aos arquivos de apoio e/ou a outros conteúdos aplicáveis à obra.

Suporte Técnico: A obra é comercializada na forma em que está, sem direito a suporte técnico ou orientação pessoal/exclusiva ao leitor.

A editora não se responsabiliza pela manutenção, atualização e idioma dos sites referidos pelos autores nesta obra.

Ouvidoria: ouvidoria@altabooks.com.br

Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD

W365n Webb, Amy

Os Nove Titãs da IA: como as gigantes da tecnologia e suas máquinas pensantes podem subverter a humanidade / Amy Webb ; traduzido por Cibelle Ravaglia. - Rio de Janeiro : Alta Books, 2020.
336 p. ; 17cm x 24cm.

Inclui bibliografia e índice.
ISBN: 978-85-508-1073-7

1. Tecnologia. 2. Inteligência artificial. I. Browne, Carol Ann. II. Ravaglia, Cibelle. III. Título.

2020-1831

CDD 006.3
CDU 004.81

Elaborado por Odílio Hilário Moreira Junior - CRB-8/9949



Rua Viúva Cláudio, 291 — Bairro Industrial do Jacaré
CEP: 20.970-031 — Rio de Janeiro (RJ)
Tels.: (21) 3278-8069 / 3278-8419
www.altabooks.com.br — altabooks@altabooks.com.br
www.facebook.com/altabooks — www.instagram.com/altabooks



SUMÁRIO

Sobre a Autora	xi
Agradecimentos	xiii
Introdução: Antes que Seja Tarde	1
Parte I: Máquinas Assombradas	11
1 Mente e Máquina: Uma Breve História da IA	13
2 O Mundo Isolado das Tribos de IA	53
3 Cortes de Folhas de Papel: As Consequências Indesejadas das IAs	99
Parte II: Nossos Futuros	135
4 Dos Dias Atuais à Superinteligência Artificial: Os Sinais dos Tempos	137
5 Prosperando na Terceira Era da Computação: O Cenário Otimista	157
6 Aprendendo a Viver com os Cortes de Papel: O Cenário Pragmático	183
7 A Dinastia Réngōng Zhìnéng: O cenário catastrófico	213
PARTE III: Resolvendo os Problemas	239
8 Pedras e Pedregulhos: Como Resolver o Futuro da IA	241
Bibliografia	269
Notas	281
Índice	315

CAPÍTULO UM

MENTE E MÁQUINA: UMA BREVE HISTÓRIA DA IA

As origens da inteligência artificial moderna remontam a centenas de anos, muito antes de as empresas intituladas como os Novas Titãs da IA [conhecidas como Big Nine] desenvolverem agentes de IA chamados Siri, Alexa e seu homólogo chinês Tiān Mão. Durante todo esse tempo, não existiu uma definição exclusiva para a IA, como existe para outras tecnologias. Quando se fala em inteligência artificial, descrevê-la em termos práticos não é nada fácil, em virtude de a IA representar muitas coisas, embora seja um campo que continue a crescer. O que aconteceu na década de 1950 — a criação de uma calculadora que conseguia fazer operações de divisão complexas — dificilmente parece uma tecnologia avançada hoje em dia. Isso é conhecido como o “estranho paradoxo” — assim que técnicas novas são inventadas, e passam a ser populares, elas se tornam menos relevantes para nós. Deixamos de pensar na IA como tecnologia.

Em sua forma mais elementar, a inteligência artificial é um sistema que toma decisões autônomas. As tarefas de IA executam ações repetidas ou simulam a inteligência humana, como reconhecer sons e objetos, resolver problemas, compreender a linguagem e usar a estratégia para atingir objetivos. Alguns sistemas de inteligência artificial são gigantescos e realizam milhões de cálculos por segundo — ao passo que outros são específicos e se destinam a uma única tarefa, como detectar linguagem imprópria em e-mails.

Nós sempre voltamos ao mesmo conjunto de perguntas: as máquinas pensam? O que significaria para uma máquina pensar? O que é pensar para nós? O que é pensamento? Como poderíamos saber — definitivamente e sem questionar — que nossos pensamentos são inéditos? Há séculos fazemos essas perguntas, e elas são cruciais para a história e para o futuro da IA.

O problema de investigar como as máquinas e os humanos pensam é que a palavra “pensar” está relacionada de forma indissociável à palavra “mente”. O *Dicionário Merriam-Webster* define “pensar” como “formular ou ter em mente”, enquanto o *Dicionário Oxford* explica que o significado da palavra é “usar a mente ativamente para organizar ideias relacionadas”. Se procurarmos “mente”, tanto o dicionário *Merriam-Webster* como o dicionário *Oxford* a definem dentro do contexto da “consciência”. Mas o que é consciência? Segundo os dois dicionários, é a qualidade ou estado de ser consciente e responsivo. Diversos grupos — psicólogos, neurocientistas, filósofos, teólogos, profissionais de ética e cientistas da computação — debatem o conceito de pensamento usando abordagens diferentes.

Ao utilizar a assistente virtual Alexa para encontrar uma mesa em seu restaurante favorito, você e ela estão conscientes e prontos para responder enquanto conversam sobre o que comer, embora a Alexa nunca tenha sentido a textura de uma maçã crocante contra os dentes, os efeitos frisantes da água com gás em sua língua ou a sensação pegajosa e grudenta da manteiga de amendoim no céu da sua boca. Peça à Alexa para descrever as características desses alimentos e ela lhe fornecerá detalhes que refletem suas próprias experiências, mas ela não tem boca — então como poderia sentir a textura de um alimento do mesmo jeito que você sente?

Biologicamente, você é uma pessoa única, cujas glândulas salivares e papilas gustativas não são talhadas exatamente na mesma ordem que as minhas. No entanto, nós dois aprendemos o que é uma maçã e as características gerais de como é o gosto de uma maçã, qual é sua textura e seu cheiro. Ao longo de nossas vidas, aprendemos a reconhecer o que é uma maçã por meio da aprendizagem por reforço — alguém nos ensinou como era uma maçã, sua finalidade e o que a distingue de outras frutas. Desse modo, com o passar do tempo e sem

o conhecimento consciente, nossos sistemas autônomos de reconhecimento de padrões biológicos ficaram excelentes em identificar que alguma coisa era uma maçã, mesmo que tivéssemos somente algumas informações necessárias. Caso veja um contorno bidimensional em preto e branco de uma maçã, você saberá do que se trata — ainda que lhe faltem o sabor, o cheiro, a crocância e todas as outras informações que sinalizam para o cérebro *que isso é uma maçã*. A maneira como você e Alexa aprenderam sobre o que é uma maçã é mais parecida do que supõe a sua imaginação.

Alexa é competente, mas seria ela *inteligente*? Sua percepção de máquina deve suprir todas as características da percepção humana para que aceitemos seu modo de “pensar” como um espelho do nosso? O psicólogo educacional Dr. Benjamin Bloom passou a maior parte de sua carreira acadêmica pesquisando e classificando os estados de pensamento. Em 1956, ele publicou o que ficaria conhecido como taxonomia de Bloom, que apresentava os objetivos de aprendizagem e os níveis de domínios observados na educação. Os alicerces fundamentais dizem respeito a relembrar os fatos e conceitos básicos, seguidos da compreensão de ideias; colocar em prática tal conhecimento em novas situações; analisar as informações experimentando e fazendo conexões; avaliar, contestar e julgar as informações; e, por fim, elaborar um trabalho inédito. Como crianças em tenra idade, focamos primeiro recordar e compreender. Por exemplo, precisamos aprender que uma caixa contém leite antes de entendermos que esta caixa de leite tem frente e verso, mesmo que não possamos vê-la.

Essa hierarquia também está presente no modo como os computadores aprendem. Em 2017, um sistema de IA chamado Amper compôs e produziu músicas inéditas para um álbum chamado *I AM AI*. As estruturas dos acordes, o uso de diferentes instrumentos e a percussão foram desenvolvidos por Amper, que utilizou parâmetros iniciais como gênero, estado de espírito e duração para gerar uma música completa em apenas alguns minutos. Taryn Southern, uma intérprete humana, colaborou com Amper para criar o álbum — e o resultado incluiu uma balada emotiva e melancólica chamada “Break Free”, que contou com mais de 1,6 milhão de visualizações no YouTube e foi um hit de sucesso nas rádios. Antes que Amper pudesse compor esta música, ela precisou aprender os elementos qualitativos de uma balada de sucesso, junto com dados quantitati-

vos, como calcular o valor da notas e batidas e como reconhecer milhares de padrões na música (por exemplo, progressões de acordes, sequências harmônicas e acentos rítmicos).

A criatividade, do tipo demonstrada por Amper, é o apogeu da taxonomia de Bloom — mas esta criatividade foi simplesmente um processo mecânico aprendido? Foi um exemplo de criatividade humana? Ou foi uma criatividade de um tipo completamente diferente? Amper pensou em música da mesma forma que um compositor humano pensaria? Pode-se argumentar que o “cérebro” de Amper — uma rede neural empregando algoritmos e dados dentro de um contêiner — talvez não seja tão diferente do cérebro de Beethoven, constituído de neurônios naturais usando dados e reconhecendo padrões dentro do contêiner que é sua cabeça. O processo criativo de Amper era tão diferente do de Beethoven quando ele compôs a Sinfonia n.º 5 (aquela sinfonia famosa que começa com um tam-tam-tam-TAAAAM) antes de mudar do tom maior (escala maior) para o tom menor (escala menor)? Beethoven não inventou a sinfonia inteira — ela não era totalmente inédita. Estas primeiras quatro notas são seguidas por uma sequência harmônica, partes das escalas, arpejos e outros elementos brutos e comuns que estruturam qualquer composição. Ouça com atenção o *scherzo*, antes do final, e você escutará padrões óbvios emprestados da 40ª Sinfonia de Mozart, escrita 20 anos antes, em 1788. Mozart foi influenciado por seu adversário Antonio Salieri e seu amigo Franz Joseph Hayden, que foram inspirados pelo trabalho de compositores anteriores como Johann Sebastian Bach, Antonio Vivaldi e Henry Purcell, que, por sua vez, estavam escrevendo músicas nos meados do século XVII ao século XVIII. Você ainda pode ouvir trechos de compositores bem mais antigos de 1400 a 1600, como Jacques Arcadelt, Jean Mouton e Johannes Ockeghem, na música deles. *Eles* foram influenciados pelos primeiros compositores medievais — e poderíamos seguir a trilha do padrão de influência até a primeira composição escrita, chamada de “epitáfio Seikilos”, que foi talhada em uma coluna de mármore para indicar um túmulo turco no século I. E poderíamos retroceder ainda mais no tempo, às primeiras flautas primitivas feitas de osso e marfim, provavelmente esculpidas 43 mil anos atrás. E, mesmo antes disso, os pesquisadores acreditam que nossos primeiros ancestrais provavelmente cantavam antes de falarem.¹

As conexões humanas são o resultado de milhões de anos de evolução. De igual modo, a conexão da IA moderna é baseada em uma longa trajetória evolutiva que remonta aos antigos matemáticos, filósofos e cientistas. Apesar de que possa parecer que a humanidade e as máquinas estiveram percorrendo caminhos divergentes, nossa evolução sempre esteve entrelaçada. O *homo sapiens* aprendeu a partir de seu meio, transmitiu suas características hereditárias às gerações futuras, diversificou e reproduziu por causa da invenção de tecnologias avançadas, como agricultura, ferramentas de caça e penicilina. Demorou 11 mil anos para os 6 milhões de habitantes do mundo, durante o período neolítico, se multiplicarem para uma população de 7 bilhões de pessoas, atualmente.² O ecossistema habitado pelos sistemas de IA — as entradas para aprendizado, dados, algoritmos, processadores, máquinas e redes neurais — está evoluindo e reproduzindo em um ritmo vertiginoso. Levará apenas algumas décadas para que os sistemas de IA se disseminem e se integrem em todos os aspectos da vida cotidiana.

Quer Alexa enxergue a imagem de uma maçã do mesmo jeito nós, quer a música original de Amper seja, de fato, “inérita”, na verdade são perguntas sobre o que supomos a respeito do pensamento. A inteligência artificial dos dias de hoje é uma amálgama de milhares de anos de filósofos, matemáticos, cientistas, profissionais de robótica, artistas e teólogos. A busca de todos eles — e a nossa, neste capítulo — é entender a conexão pensante e os contêineres como alimento à reflexão. Qual é a conexão entre a mente humana e — *ou a despeito de* — as máquinas que estão sendo construídas pelas Nove Titãs na China e nos Estados Unidos?

Seriam as Máquinas Conscientes?

Pode-se refazer os passos dos pilares fundamentais da IA até a Grécia Antiga e às origens da filosofia, lógica e matemática. Em muitos textos de Platão, Sócrates diz: “Conhece-te a ti mesmo”; ele quis dizer que, a fim de melhorar e tomar as decisões certas, primeiro você precisava conhecer sua própria personalidade. Entre seus outros trabalhos, Aristóteles inventou a lógica silogística e nosso primeiro sistema formal de raciocínio dedutivo. Na mesma época, o

matemático grego Euclides idealizou uma forma de encontrar o maior divisor comum de dois números e, como resultado, criou o primeiro algoritmo. O trabalho deles foi o começo de duas ideias imprescindíveis: que determinados sistemas físicos podem funcionar como um conjunto de regras lógicas, e que o próprio pensamento humano pode ser um sistema simbólico. Isso desencadeou séculos de investigação entre filósofos, teólogos e cientistas. O corpo era uma máquina complexa? Um todo unificado composto de inúmeros outros sistemas trabalhando juntos, como um relógio de pêndulo? Mas e a mente? Também era uma máquina complexa? Ou alguma coisa completamente diferente? Não havia como comprovar ou refutar um algoritmo divino ou a conexão entre a mente e o reino físico.

Em 1560, um relojoeiro espanhol chamado Juanelo Turriano criou um pequeno monge mecânico como oferenda à igreja, em nome do rei Filipe II da Espanha, cujo filho se recuperou milagrosamente de uma traumatismo craniano.³ Tal monge tinha poderes impressionantes — ele atravessava a mesa, erguia um crucifixo e um rosário, batia no peito em sinal de arrependimento e movimentava os lábios em oração. Esse foi o primeiro *autômato* — uma representação mecânica de algo vivo. Embora a palavra “robô” ainda não existisse, o monge era uma pequena invenção fantástica, que deve ter deixado os espectadores espantados e confusos. Provavelmente nunca ocorreu a ninguém que um minúsculo autômato pudesse algum dia, em um futuro distante, não somente imitar os movimentos básicos, mas também substituir os humanos na linha de produção das fábricas, nos laboratórios de pesquisa ou jogar conversa fora no café.

O pequenino monge inspirou a geração pioneira de profissionais de robótica, cujo objetivo era desenvolver máquinas cada vez mais complexas que retratassem os seres humanos: os autômatos logo foram capazes de escrever, dançar e pintar. E isso levou um grupo de filósofos a começar a questionar sobre o que significa ser um humano. Se fosse possível criar autômatos que simulassem o comportamento humano, então os humanos seriam autômatos divinamente criados? Ou somos sistemas complexos capazes de raciocinar e pensar de um modo inédito?

O filósofo político inglês Thomas Hobbes descreveu o raciocínio humano como um cálculo em *De Corpore*, parte de sua excelente trilogia em ciências naturais, psicologia e política. Em 1655, ele escreveu: “Por raciocínio, eu entendo cálculo. E computar é reunir a soma de muitas coisas somadas ao mesmo tempo, ou saber o resultado de quando uma coisa foi tirada de outra. Raciocinar, portanto, é o mesmo que somar ou subtrair.”⁴ Mas como saberíamos se éramos dotados do livre-arbítrio durante o processo?

Enquanto Hobbes escrevia a primeira parte de sua trilogia, o filósofo francês René Descartes publicou *Meditações sobre Primeira Filosofia*, questionando se podemos saber com certeza que o que sentimos é real. Como poderíamos comprovar nossa própria consciência? De qual prova precisaríamos a fim de concluir que nossos pensamentos são nossos, e que o mundo ao nosso redor é real? Descartes era iluminista, acreditava que os fatos poderiam ser inferidos por dedução, então sugeriu um experimento mental que ficaria famoso. Ele pediu aos leitores que imaginassem uma criatura demoníaca criando propositalmente uma ilusão do mundo em que viviam. Se a experiência física e sensorial de uma leitora ao nadar em um lago não era nada mais do que a criação de uma criatura demoníaca, logo ela não poderia *saber* que estava nadando. Porém, na visão de Descartes, se a leitora tivesse autoconsciência de sua própria existência, então ela alcançaria os critérios para o conhecimento. “Eu sou, eu existo, sempre que isso é proferido por mim, ou concebido pela mente, é necessariamente real”, escreveu ele.⁵ Em outras palavras, o fato de nossa existência reside além da dúvida, ainda que haja uma criatura demoníaca ludibriadora no meio. Ou, *Penso, logo existo*.

Mais tarde, em seu *Traité de l’homme (Tratado do Homem)* Descartes argumentara que os humanos provavelmente poderiam construir um autômato — neste caso, um pequeno animal — que seria indissociável da coisa real. Mas, ainda que um dia criássemos um humano mecanizado, ele nunca se passaria como real, apontou Descartes, pois lhe faltaria uma mente e, portanto, uma alma. Diferentemente dos humanos, uma máquina nunca poderia alcançar os critérios do conhecimento — nunca poderia ter autoconsciência como nós. Para

Descartes, a consciência se dava intrinsecamente — a alma era o fantasma que assombraria as máquinas que são nossos corpos.⁶

Algumas décadas mais tarde, o matemático e filósofo alemão Gottfried Wilhelm von Leibniz analisou a ideia de que a alma humana era programada, alegando que a própria mente era um contêiner. Deus criou a alma e o corpo para se harmonizarem naturalmente. O corpo pode até ser uma máquina complexa, mas que tem um conjunto de instruções divinas. Nossas mãos se movimentam quando decidimos movimentá-las, entretanto não criamos, tampouco inventamos, todos os mecanismos que possibilitam o movimento. Se estamos conscientes da dor ou do prazer, essas sensações são o resultado de um sistema pré-programado, uma linha contínua de comunicação entre a mente e o corpo.

Leibniz desenvolvera o próprio experimento mental a fim de exemplificar o fato de que o pensamento e a percepção estavam estritamente ligados ao ser humano. Imagine caminhar em um moinho de vento. A estrutura é uma máquina de contêineres, matérias-primas e trabalhadores. É um sistema complexo de partes trabalhando em harmonia rumo a um objetivo singular, mas que nunca poderia ter uma mente. “Tudo o que sabemos é que existem engrenagens e alavancas empurrando umas às outras, nada que explique a percepção”, escreveu Leibniz. “Portanto, deve-se procurar a percepção em substâncias simples e nunca em coisas complexas como máquinas.” A alegação que Leibniz estava fazendo era que, por mais moderno que fosse o moinho, a máquina ou o autômato, os humanos jamais poderiam construir uma máquina capaz de pensar ou sentir.⁷

Leibniz ainda era fascinado com a noção de reproduzir os aspectos do pensamento. Algumas décadas antes, um escritor inglês pouco conhecido chamado Richard Braithwaite escreveu alguns livros a respeito da conduta social, referindo-se passivamente a “computadores” humanos como pessoas altamente capacitadas, rápidas e precisas, excelentes em fazer cálculos.⁸ Neste ínterim, o matemático francês e inventor Blaise Pascal, que estabeleceu as bases para o que hoje conhecemos como probabilidade, se preocupou em automatizar as tarefas computacionais. Pascal observou seu pai calcular enfadonhamente os impostos de forma manual e queria facilitar o processo para ele. Então, começou a trabalhar em uma calculadora automática, uma com rodas mecânicas e indicadores

que se movessem.⁹ A calculadora funcionou e inspirou Leibniz a aperfeiçoar seu pensamento: as máquinas nunca seriam dotadas de almas; no entanto, algum dia seria possível construir uma máquina que conseguisse pensar no mesmo nível de lógica humana. Em 1673, Leibniz descreveu sua calculadora mecânica chamada de “step reckoner” como um tipo novo de máquina de calcular que tomava decisões usando um sistema binário.¹⁰ A máquina era como uma mesa de bilhar, com bolas, buracos, tacos e canaletas, e a máquina abria os buracos usando uma série de 1s (aberto) e 0s (fechado).

A step reckoner conceitual de Leibniz estabeleceu os alicerces para mais teorias, que incluíam a noção de que, se o pensamento lógico pudesse ser reduzido a símbolos e, como resultado, pudesse ser analisado como um sistema computacional, e se problemas geométricos pudessem ser resolvidos utilizando símbolos e números, logo tudo poderia ser reduzido a bits — inclusive o comportamento humano. Era uma ruptura significativa com os filósofos anteriores: as máquinas futuras poderiam reproduzir os processos de pensamento humano sem desobedecer à providência divina. O ato de pensar não exige necessariamente a percepção, os sentidos ou a alma. Leibniz imaginou um computador capaz de solucionar problemas gerais, até aqueles que não fossem matemáticos. E ele formulou a hipótese de que a linguagem poderia ser reduzida a conceitos atômicos de matemática e ciências como parte de um tradutor de linguagem universal.¹¹

A Mente e a Máquina Seguem as Instruções Algorítmicas?

Na hipótese de Leibniz estar certo — que os humanos eram máquinas dotadas de alma e, um dia, inventariam máquinas sem alma capazes de pensar de forma inédita e sofisticada —, então, poderia haver uma classe binária de máquinas na Terra: nós e elas. Calma, que o debate só começou.

Em 1738, Jacques de Vaucanson, um artista e inventor, construiu uma série de autômatos para a Academia Francesa de Ciências, que incluía um pato complexo e realista. O objeto não apenas imitava os movimentos de um pato real, batendo as asas e comendo grãos, mas também imitava a digestão. Este fato

oferecia aos filósofos matéria à reflexão: se parecia um pato, e grasnava como um pato, seria realmente um pato? Se sentimos que o pato tem um tipo de alma diferente, isso bastaria para comprovar que o pato é um ser consciente (de si mesmo) e tudo mais o que isso significaria?

O filósofo escocês David Hume rejeitou a ideia de que o reconhecimento da existência fosse por si só uma prova de consciência. Ao contrário de Descartes, Hume era um empirista, que desenvolveu uma estrutura científica nova baseada em fatos observáveis e argumentos lógicos. Enquanto Vaucanson exibia seu pato com sistema digestivo — e bem antes de alguém falar sobre inteligência artificial —, Hume escrevera *Um Tratado de Natureza Humana*: “A razão é e deve ser apenas a escrava das paixões.” Neste caso, Hume entendia o significado de “paixões” como “motivações irracionais”, e acreditava que incentivos, e não a lógica abstrata, conduziam o nosso comportamento. Se as impressões são apenas nossa percepção de algo que podemos ver, tocar, sentir, saborear e cheirar, e as ideias são percepções de coisas com as quais não entramos em contato direto, Hume tinha a convicção de que nossa existência e compreensão do mundo à nossa volta eram fundamentadas em um construto da percepção humana.

Dado o avanço do trabalho com os autômatos, que se tornavam cada vez mais realistas, e levando em conta a reflexão aprofundada dos computadores como máquinas pensantes, o médico e filósofo francês Julien Offray de La Mettrie realizou um estudo radical — e infame — sobre os humanos, os animais e os autômatos. Em um artigo de 1747, publicado pela primeira vez de forma anônima, La Mettrie argumentou que os seres humanos são extremamente parecidos com animais, e um macaco poderia aprender a linguagem humana se fosse “devidamente treinado”. La Mettrie também concluiu que os seres humanos e animais não passam de meras máquinas, motivadas pelo instinto e pela experiência. “O corpo humano é uma máquina emaranhada em suas próprias engrenagens... a alma é apenas um princípio do movimento ou uma parte material e consciente do cérebro.”¹²

A ideia de que os seres humanos são simplesmente máquinas motivadas pela matéria — engrenagens e rodas executando um conjunto de funções — insinuava que não éramos especiais, tampouco únicos. Insinuava também que tal-

vez fossemos programáveis. Caso isso fosse verdade, e se já tivéssemos chegado até o ponto de conseguir criar patos realistas e monges pequeninos, então mais dia, menos dia, os humanos poderiam criar réplicas de si mesmos — e construir uma variedade de máquinas inteligentes e pensantes.

Poderia uma Máquina Pensante Ser Criada?

Na década de 1830, matemáticos, engenheiros e cientistas começaram a fazer experiências, na esperança de construir máquinas capazes de fazer os mesmos cálculos que os “computadores” humanos. A matemática britânica Ada Lovelace e o cientista Charles Babbage inventaram uma máquina chamada “Máquina Diferencial” e depois apresentaram como base teórica uma “Máquina Analítica” mais avançada, que seguia uma série de instruções predeterminadas para resolver problemas matemáticos. Babbage nem sequer imaginara que a máquina pudesse fazer outras coisas além de calcular números. Foi Ada quem, nas notas de rodapé de um artigo científico que estava traduzindo, apresentou inteligentemente um assunto novo ao especular que uma versão mais poderosa da máquina poderia ser usada de outras maneiras.¹³ Se a máquina pudesse manipular símbolos, que pudessem ser atribuídos a coisas diferentes (como notas musicais), então poderia ser usada para “pensar” além da matemática. Apesar de Ada não acreditar que um computador pudesse criar pensamentos inéditos, ela imaginara um sistema complexo que poderia seguir instruções e, desse modo, simular muito do que as pessoas normais faziam. À época, isso não teve a menor relevância para algumas pessoas, mas Ada havia escrito o primeiro programa de computador completo para uma futura máquina poderosa — décadas antes da invenção da lâmpada.

Cento e sessenta quilômetros ao Norte de onde Ada e Babbage trabalhavam, na Universidade de Cambridge, um jovem matemático autodidata chamado George Boole estava caminhando por um campo em Doncaster e teve uma súbita epifania, decidindo dedicar sua vida a explicar a lógica do pensamento humano.¹⁴ Os frutos desta caminhada são o que hoje conhecemos como álge-

bra booleana, um modo de simplificar expressões lógicas (por exemplo, “AND”, “OR” e “NOT”) utilizando símbolos e números. Por exemplo, os valores computacionais “true e true” resultariam em “true”, o que corresponderia a switches físicos e portas em um computador. Demoraria duas décadas para que Boole oficializasse suas ideias. E seria necessário outros 100 anos para alguém perceber que a lógica e a probabilidade booleanas poderiam ajudar os computadores a evoluir da automação da matemática básica para máquinas pensantes e mais complexas. Não havia como construir uma máquina pensante — os processos, os materiais e o processamento ainda não estavam disponíveis — e, portanto, a teoria não poderia ser testada.

O avanço das teorias das máquinas pensantes para computadores que começaram a imitar o pensamento humano aconteceu na década de 1930 com a publicação de dois artigos precursores: “A Symbolic Analysis of Switching and Relay Circuits” [Uma Análise Representativa da Comutação de Circuitos e Circuitos de Relé, em tradução livre], de Claude Shannon, e “On Computable Numbers, with an Application to the *Entscheidungsproblem*” [Sobre Cálculos Computacionais, com a Aplicação do *Entscheidungsproblem*, em tradução livre], de Alan Turing. Como estudante de engenharia elétrica no MIT, Shannon fez um curso opcional de filosofia — uma distração peculiar. Seu livro *Uma Investigação das Leis do Pensamento* se tornou a principal referência para a tese de Shannon. O orientador de Shannon, Vannevar Bush, o incentivou a sistematizar a lógica booleana em circuitos físicos. Bush construíra uma versão avançada da Máquina Analítica de Ada e Babbage — seu protótipo se chamava “Analisador Diferencial” — e seu design era deveras improvisado. Naquela época, não existia uma teoria estruturada que estabelecesse a disposição de circuitos elétricos. A inovação de Shannon fora sistematizar os circuitos elétricos de acordo com a lógica simbólica booleana e, em seguida, explicar como ela poderia ser usada para criar um circuito de trabalho que somasse 1s e 0s. Shannon havia descoberto que os computadores tinham duas camadas: física (o contêiner) e lógica (o código).

À medida que Shannon trabalhava para integrar a lógica booleana em circuitos físicos, Turing testava o tradutor de linguagem universal de Leibniz que simbolizaria todo o conhecimento matemático e científico. O objetivo de Turing

era provar o chamado *Entscheidungsproblem* ou “problema de decisão”. Grosso modo, o problema era o seguinte: não existe nenhum algoritmo que determine se uma declaração matemática arbitrária é verdadeira ou falsa. A resposta seria negativa. Turing conseguiu comprovar que não existe algoritmo, mas, como subproduto, descobriu um modelo matemático de uma máquina computacional universal.¹⁵

E isso mudaria tudo. Turing descobriu que um programa e os dados usados poderiam ser armazenados dentro de um computador — mais uma vez, era uma tese arrojada para 1930. Até aquele ponto, todos concordaram que a máquina, o programa e os dados eram independentes. Pela primeira vez, a máquina universal de Turing explicou por que todos os três estavam interligados. Da perspectiva mecânica, a lógica que operava os circuitos e comutadores também poderia ser codificada em programa e dados. Pense no significado de tais afirmações. O contêiner, o programa e os dados faziam parte de uma entidade única — não muito diferente dos humanos. Nós também somos contêineres (nossos corpos), programas (funções celulares autônomas) e dados (nosso DNA combinado com informações sensoriais indiretas e diretas).

Neste meio-tempo, aquela tradição de longa data de autômatos, que começara 400 anos antes com o pequeno monge que caminhava e rezava, finalmente se cruzou com o caminho do trabalho de Turing e Shannon. A empresa norte-americana de manufatura Westinghouse construiu um robô baseado em relé chamado Elektro, o Moto-Man, para a Feira Mundial de 1939. Era um gigante de cor dourada com rodas sob os pés, que tinha 48 relés elétricos que funcionavam em um sistema telefônico de transmissão. O Elektro respondia por meio de mensagens pré-gravadas em um toca-discos a comandos de voz falados via telefone. Era um computador antropomorfizado capaz de tomar decisões básicas — como o que dizer — sem envolvimento humano direto e em tempo real.

A julgar pelas manchetes dos jornais, contos de ficção científica e noticiários daquela época, fica claro que as pessoas foram pegas de surpresa, gerando espanto e preocupação com todos esses avanços. Para elas, parecia que as “máquinas pensantes” supercapacitadas simplesmente haviam chegado da noite

para o dia. O escritor de ficção científica Isaac Asimov publicou “Mentiroso!”, um conto profético na edição de maio de 1941 da *Astounding Science Fiction*. Era uma reação à pesquisa pouco aceita que ele estava estudando e, a partir dela, ele apresentou suas Três Leis da Robótica:

1. Um robô não pode machucar um ser humano ou, por omissão, permitir que um ser humano sofra algum mal;
2. um robô deve obedecer às ordens que lhe são dadas por seres humanos, exceto nos casos em que essas ordens entrem em conflito com a Primeira Lei; e
3. um robô deve proteger sua própria existência desde que essa proteção não entre em conflito com a Primeira e/ou a Segunda Lei.

Posteriormente, Asimov acrescentou o que chamou de “Lei Zeroth” para governar todos os outros: “Um robô não pode prejudicar a humanidade ou, por omissão, permitir que a humanidade sofra algum prejuízo.”

Mas Poderia uma Máquina Pensante Pensar?

Em 1943, os pesquisadores em psiquiatria Warren McCulloch e Walter Pitts da Universidade de Chicago publicaram um importante artigo, “Logical Calculus of the Ideas Immanent in Nervous Activity” [Cálculos Aritméticos de Ideias Inerentes à Atividade Nervosa, em tradução livre], que descrevia um novo tipo de sistema que modelava neurônios biológicos em uma arquitetura de rede neural simples para inteligência. Se os contêineres, programas e dados estivessem interconectados, como argumentara Turing, e se os seres humanos fossem contêineres sofisticadamente modelados de forma semelhante e capazes de processar dados, logo talvez fosse possível construir uma máquina pensante caso o modelo usasse a parte humana responsável pelo pensamento — nossos cérebros. Eles estabeleceram os princípios básicos de uma teoria computacional

moderna da mente e do cérebro, uma “rede neural”. Em vez de focar a máquina como hardware e o programa como software, eles imaginaram um tipo novo de sistema simbiótico capaz de devorar grandes quantidades de dados, assim como nós, humanos. Os computadores ainda nem eram poderosos o bastante para testar essa teoria — mas esse artigo serviu de inspiração às outras pessoas para começar a trabalhar rumo a um novo tipo de sistema inteligente de computador.

O elo entre os sistemas inteligentes de computadores e a tomada de decisões autônoma ficou mais claro quando John von Neumann, o erudito húngaro-americano especializado em ciência da computação, física e matemática, publicou um grande tratado de matemática aplicada. Escrito em parceria com o economista de Princeton Oskar Morgenstern, em 1944, o livro de 641 páginas explicava, em detalhes minuciosos, como a ciência da teoria dos jogos descortinava os alicerces de todas as decisões econômicas. Foi esse trabalho que levou von Neumann a colaborar com o Exército dos EUA, que vinha trabalhando em um tipo novo de computador elétrico chamado de Electronic Numerical Integrator and Computer [Computador Integrador Numérico Eletrônico, em tradução livre], ou ENIAC, abreviado. A princípio, as instruções que alimentavam o ENIAC eram conectadas diretamente ao sistema, o que significava que, a cada programa novo, todo o sistema teria que ser reconectado. Inspirado por Turing, McCulloch e Pitts, von Neumann desenvolveu uma maneira de armazenar programas no próprio computador. Isso representou a transição da primeira era da computação (tabulação) para uma nova era de sistemas programáveis.

Agora, o próprio Turing trabalhava em um conceito para uma rede neural composta de computadores com arquitetura de máquina para armazenagem de programas. Em 1949, *The London Times* citou Turing “Não vejo por que (a máquina) não deve entrar em nenhum dos campos normalmente contemplados pelo intelecto humano e, em algum momento, competir em igualdade de condições. Não acho nem que você tenha que estabelecer algum limite em relação aos sonetos, ainda que a comparação seja talvez um pouco injusta, pois um soneto escrito por uma máquina será mais admirado por outra máquina”. Um ano depois, em um artigo publicado no periódico de filosofia *Mind*, Turing abordou