



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

AMY WEBB


ALTA BOOKS
EDITORA

Os Nove Titãs da IA

Copyright © 2020 da Starlin Alta Editora e Consultoria Eireli. ISBN: 978-8-550-81074-4

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 Ilyssabelle Trajano Juliana de Oliveira Thiê Alvea | 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 Marta de Lourdes Borges Raquel Porto Rodrigo Dutra | Equipe de Design Larissa Lima Paulo Gomes | | |
| Tradução Cibelle Ravaglia | Capêdique Fernanda Lutfi | Revisão Gramatical Hellen Suzuki Thiãs Ivo | Diagramação Larissa Maria Gomes Adaptação para formato e-book Catia Soderi |

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

Eratas 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 eratas, 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.

Ouvintes: ouvidoria@altabooks.com.br

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

| | |
|---|-------------------------|
| Autores | Webb, Amy |
| Os Nove Titãs da IA, com os gigantes da tecnologia e suas máquinas pessoais podem salvar a humanidade? Amy Webb, traduzido por Cibelle Ravaglia e Ricardo Inácio. Alta Books, 2020. | |
| 354 p. : 17cm x 24cm. | |
| Inclui bibliografia e índice. | |
| ISBN: 978-8-550-81074-4 | |
| 1. Tecnologia. 2. Inteligência artificial. I. Inácio, Ricardo. II. Ravaglia, Cibelle. III. Webb, Amy. | |
| 2020-1433 | CDD 004.3 CDU 004.01 |
| Elaborado por Odílio Roberto Moreira Junior - CEB 8/9/20 | |



Rua Vinte e Quatro, 291 — Bairro Industrial do Incaê
CEP: 20.979-851 — Rio de Janeiro (RJ)
Tel.: (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

Agradecimentos

Introdução: Antes que Seja Tarde

PARTE I: Máquinas Assombradas

Capítulo 1: Mente e Máquina: Uma Breve História da IA

Capítulo 2: O Mundo Isolado das Tribos de IA

Capítulo 3: Cortes de Folhas de Papel: As Consequências Indesejadas das IAs

PARTE II: Nossos Futuros

Capítulo 4: Dos Dias Atuais à Superinteligência Artificial: Os Sinais dos Tempos

Capítulo 5: Prosperando na Terceira Era da Computação: O Cenário Otimista

Capítulo 6: Aprendendo a Viver com os Cortes de Papel: O Cenário Otimista

Capítulo 7: A Dinastia Réngōng Zhìnéng: O cenário catastrófico

PARTE III: Resolvendo os Problemas

Capítulo 8: Pedras e Pedregulhos: Como Resolver o Futuro da IA

Bibliografia

SOBRE A AUTORA



Crédito: fotografia de Mary

AMY WEBB é uma das principais futuristas dos Estados Unidos, autora do premiado best-seller *The Signals Are Talking: Why today's fringe is tomorrow's mainstream*, no qual explica seu método para prever o futuro. Ela é professora de visão estratégica na NYU Stern School of Business e fundadora do Future Today Institute, uma empresa líder de mercado em previsão e estratégia, que ajuda os líderes e suas organizações a se prepararem para possíveis futuros complexos e desconhecidos. Webb é vencedora do Thinkers50 Radar Award, membro do United States–Japan Leadership Program, representante da antiga US-Russia Bilateral Presidential Commission e ainda participou como Visiting Nieman Fellow na Universidade de Harvard. Trabalha como consultora de roteiros de filmes e programas sobre tecnologia, ciência e futuro; também publica o relatório anual de tendências tecnológicas emergentes do FTI, que já gerou mais de 7,5 milhões de visualizações em todo o mundo. Saiba mais sobre Amy em: <http://www.amywebb.io> [conteúdo em inglês].

AGRADECIMENTOS

Como a inteligência artificial, este livro demorou alguns anos para tomar forma e ser elaborado. Tudo começou como uma série de perguntas enviadas por mensagem de texto e conversas à mesa de jantar, até se tornar uma preocupação crescente que me acompanhava na academia, nas saídas à noite e aos finais de semana. Uma pessoa — Brian Woolf — deixou-se levar por essa obsessão, me ajudou a persegui-la e apoiou meu trabalho por muitos anos. Brian contribuiu para minha pesquisa, ajudou-me a deixar meus argumentos claros e ficou acordado até tarde da noite para editar todas as minhas páginas. Sou profundamente grata.

Os Nove Titãs da IA é o resultado de inúmeras reuniões presenciais, entrevistas e jantares com pessoas que trabalham direta e indiretamente com a inteligência artificial. Sewell Chan, Noriyuki Shikata, Arfiya Eri, Joel Puckett, Erin McKean, Bill McBain, Frances Colon, Torfi Frans Olafsson, Latoya Peterson, Rob High, Anna Sekaran, Kris Schenck, Kara Snesko, Nadim Hossain, Megan Carroll, Elena Grewal, John Deutsch, Neha Narula, Toshi Ezoë, Masao Takahashi, Mary Madden, Shintaro Yamaguchi, Lorelei Kelly, Hiro Nozaki, Karen Ingram, Kirsten Graham, Francesca Rossi, Ben Johnson, Paola Antonelli, Yoav Schlesinger, Hardy Kagimoto, John Davidow, Rachel Sklar, Glynnis MacNicol, Yohei Sadoshima e Eiko Ooka me doaram um tempo precioso, perspectivas e insights. Muitos deles me apresentaram a outras pessoas que trabalham com IA e política, com o intuito de me ajudar a investigar a fundo o equilíbrio geopolítico e compreender melhor as oportunidades e os riscos da IA.

Foi por causa da US–Japan Leadership Foundation que conheci o Tenente-coronel Sea Thomas; o major aposentado do Exército DJ Skelton;

o diretor-executivo da Defense Innovation Board, Joshua Marcuse; e o analista de segurança nacional John Noonan. Passamos muitos dias juntos como bolsistas da USJLP, e tenho uma dívida de gratidão eterna com cada um deles pela paciência de explicar o futuro da guerra, o papel dos militares dos EUA no Círculo do Pacífico e as muitas iniciativas estratégicas da China. Fiquei muito impressionada com o trabalho que Joshua realizou para transpor os obstáculos entre o Vale do Silício e Washington, D.C. Ele é um dos heróis atuais da IA.

O Aspen Strategy Group me ofereceu uma oportunidade de apresentar o futuro da IA e da geopolítica durante sua conferência anual de verão no Colorado, e essas conversas ajudaram a estruturar minha análise. Meus sinceros agradecimentos a Nicholas Burns, Condoleezza Rice, Joseph Nye e Jonathon Price pelo convite e a Carla Anne Robbins, Richard Danzig, James Baker, Wendy Sherman, Christian Brose, Eric Rosenbach, Susan Schwab, Ann-Marie Slaughter, Bob Zoellick, Philip Zelikow, Dov Zakheim, Laura Rosenberger e Mike Green por todos os feedbacks valiosos.

Muitas das minhas ideias me ocorreram no campus da Stern School of Business da NYU, uma associação profissional que tem apoiado muito a minha pesquisa. Quero agradecer ao professor Sam Craig por me trazer para o programa de MBA e por ter me aconselhado nos últimos anos. Eu não teria palavras o bastante para falar sobre os estudantes maravilhosos e brilhantes que assistiram às minhas aulas. Três deles recém-formados em Stern — Kriffy Perez, Elena Giralt e Roy Levkovitz — foram porta-vozes maravilhosos à medida que eu traçava os futuros da IA.

Sou sortuda de ter em minha vida um grupo de sábios que oferecem conselhos e orientação. Todo o trabalho que faço é melhor por causa deles. Danny Stern mudou minha vida há alguns anos quando me pediu para encontrá-lo um dia no campus da NYU. Ele me ensinou a pensar de forma mais abrangente e me mostrou como fazer minha pesquisa abarcar públicos mais amplos. Sua parceira no Stern Strategy Group, Mel Blake, passou horas a fio me orientando, arquitetando minhas ideias e

me ajudando a ver o mundo ao redor de maneira diferente. Ambos são uma fonte contínua de inspiração, motivação e (como eles bem sabem) transpiração. James Geary e Ann Marie Lipinski, de Harvard, foram extremamente generosos comigo por muitos anos, possibilitando que eu organizasse reuniões para falar sobre o futuro e desenvolver minha metodologia de previsão. James e Ann Marie são orientadores perfeitos. Minha querida amiga e guerreira, Maria Popova, me fez colocar a caixola para funcionar e, depois, contextualizava essas ideias em seu conhecimento enciclopédico de literatura, artes e ciências. Minha filha incrível, Petra Woolf, nunca parou de perguntar “e se”, lembrando-me muitas vezes de meus próprios preconceitos cognitivos quando penso sobre o futuro. E, como sempre, sou grata ao professor Samuel Freedman, da Universidade de Columbia.

Meus eternos agradecimentos a Cheryl Cooney, que trabalha incansavelmente em meu nome, sem ela eu faria muito pouco. Independentemente de as AGIs serem construídas algum dia, não consigo imaginar uma que possa substituir Cheryl. Emily Caufield — cuja paciência parece não ter limites — é a força criativa que alimenta meu trabalho de previsão, tendências e cenários. Agradeço a Phillip Blanchard por trabalhar comigo novamente na comprovação dos fatos, edição de texto e compilação de todas as fontes e notas finais deste livro; e a Mark Fortier, que ajudou a assegurar que o livro fosse lido pela mídia e pelos editores de notícias, e cujo conselho fora inestimável durante o processo de lançamento.

E, por último, devo milhões de zettabytes de agradecimentos à Carol Franco, Kent Lineback e John Mahaney. Como minha agente literária, Carol conseguiu o contrato para este livro. Mas, como minha amiga, ela e seu marido, Kent, me receberam em sua linda casa em Santa Fé, para que pudéssemos desenvolver a arquitetura e a tese basilar sobre os Nove Titãs. Passamos dias e noites refinando todas as minhas pesquisas e ideias em argumentos sólidos e, entre os turnos de trabalho, passeávamos pela cidade e tínhamos discussões animadas em ótimos

restaurantes. Foi por causa de Carol que, há alguns anos, conheci meu editor, John Mahaney, com quem tive a sorte de trabalhar no meu livro anterior. John é um editor excelente — ele me enche de perguntas, exige relatórios de qualidade e continua insistindo até que a análise, os exemplos e os detalhes estejam perfeitos. Escrevi este livro porque quero mudar o rumo da conversa a respeito do futuro da IA, mas minha motivação não foi inteiramente desinteressada: trabalhar com John novamente significava uma oportunidade de passar um ano aprendendo com ele e melhorando a minha escrita. John, Kent e Carol, vocês são uma equipe formidável, e mal posso acreditar na sorte que tenho de conhecê-los.

INTRODUÇÃO

ANTES QUE SEJA TARDE

A inteligência artificial já está entre nós, todavia não surgiu como todos esperávamos. Ela é a espinha dorsal silenciosa de nossos sistemas financeiros, fornecimento da rede elétrica e cadeia de suprimento de varejo. Ela é a infraestrutura invisível que nos direciona no trânsito, identifica o significado correto em nossas palavras equivocadas e condiciona o que devemos comprar, observar, ouvir e ler. Ela é a tecnologia sobre a qual o nosso futuro está sendo alicerçado, porque permeia todos os aspectos de nossas vidas: saúde e medicina, transporte, moradia, agricultura, esportes e até mesmo amor, sexo e morte.

A IA não é uma tendência tecnológica, um termo em voga ou uma distração temporária — é a terceira era da computação. Estamos no meio de uma transformação descomunal, não muito diferente da geração que viveu a Revolução Industrial. A princípio, ninguém reconhecera a transição em que estavam porque a mudança aconteceu pouco a pouco, no tocante ao ciclo de vida. No final, o mundo parecia diferente: a Grã-Bretanha e os Estados Unidos haviam se tornado as duas potências dominantes do mundo, com capital industrial, militar e político o bastante para moldar o curso do próximo século.

Todos estão discutindo sobre a IA e o que ela significará para nossos futuros à exaustão. As discussões já são velhas conhecidas: os robôs estão vindo para roubar nossos empregos, os robôs virarão nossa economia de cabeça para baixo, os robôs acabarão matando os seres humanos. Substitua “máquina” por “robô” e voltamos aos mesmos debates que as pessoas tinham 200 anos atrás. É natural pensar no impacto da nova tecnologia em nossos empregos e em nossa capacidade de ganhar dinheiro, já que vemos tantas rupturas em tantos setores. É

compreensível que, quando pensamos em IA, nossas mentes inevitavelmente perambularem de volta para a HAL 9000 de 2001: Uma Odisseia no Espaço; para o supercomputador WOPR de Jogos de Guerra; para a Skynet de O Exterminador do Futuro; para a Rosie de Os Jetsons; para Delores da série Westworld ou qualquer uma das outras inúmeras IAs antropomorfizadas da cultura popular. Se você não trabalha diretamente com o ecossistema da IA, aparentemente o futuro é fantástico ou aterrorizante, sem nenhuma razão que se justifique.

Aqueles que não estão absorvidos e não vivem o dia a dia da pesquisa e do desenvolvimento da IA não conseguem enxergar os sinais com clareza, razão pela qual a discussão da opinião pública a respeito da IA gravita em torno de robôs tirânicos a que você assistiu nos filmes recentes. Ou reflete um tipo de otimismo maníaco e desmedido. A carência de nuances é a primeira causa do problema da IA: alguns superestimam radicalmente a aplicabilidade da IA, ao passo que outros defendem que ela se tornará uma arma invencível.

Sei disso porque passei a maior parte da última década fazendo pesquisas sobre inteligência artificial e me reunindo com pessoas e organizações dentro e fora do ecossistema da IA. Aconselhei uma diversidade grande de empresas no epicentro da inteligência artificial, como a Microsoft e a IBM. Conheci e aconselhei as partes interessadas do lado de fora: investidores de risco e gerentes de investimentos privados, líderes do Departamento de Defesa e Departamento de Estado, e muitos legisladores que acham que a regulamentação é o único caminho para seguir. Participei também de diversas reuniões com pesquisadores acadêmicos e tecnólogos que trabalham diretamente na linha de frente. É raro aqueles que trabalham diretamente com a IA compartilharem visões apocalípticas extremas ou utópicas do futuro que costumamos ouvir nos jornais.

Isso se deve ao fato de que, como os pesquisadores de outras áreas da ciência, aqueles que engendram o futuro da IA querem moderar as expectativas. A conquista de grandes feitos requer paciência, tempo,

dinheiro e resiliência — algo que nos esquecemos com facilidade. Eles trabalham arduamente e em etapas com problemas muito complicados, às vezes progredindo pouco. Essas pessoas são inteligentes, globalizadas e, em minha experiência, solidárias e ponderadas.

Em sua maioria, elas trabalham para os nove gigantes da tecnologia — Google, Amazon, Apple, IBM, Microsoft e Facebook, nos Estados Unidos; e Baidu, Alibaba e Tencent, na China —, que estão arquitetando a IA a fim de inaugurar um futuro melhor e mais positivo para todos nós. Acredito piamente que os líderes dessas nove empresas são motivados por um sentimento profundo de altruísmo e um desejo de servir a um bem maior: eles enxergam com clareza o potencial da IA de melhorar a saúde e a longevidade, solucionar nossos problemas climáticos iminentes e tirar milhões de pessoas da pobreza. Já estamos vendo os benefícios positivos e tangíveis do trabalho deles em todos os setores e na vida cotidiana.

O problema é que as forças externas que pressionam os nove gigantes da tecnologia — e, de quebra, as pessoas que trabalham dentro desse ecossistema — conspiram contra seus melhores objetivos para o futuro. Todos vivem colocando a culpa em todo mundo.

Nos EUA, as demandas inflexíveis do mercado e as expectativas irrealistas em relação a novos produtos e serviços fizeram com que o planejamento de longo prazo fosse impossível. Esperamos que o Google, a Amazon, a Apple, o Facebook, a Microsoft e a IBM façam novos anúncios de produtos IA em suas conferências anuais, como se os avanços em pesquisa e tecnologia pudessem ser agendados. Caso essas empresas não nos apresentem produtos mais arrojados do que no ano anterior, falamos sobre elas como se fossem uma decepção. Ou questionamos se é o fim da IA. Ou questionamos a liderança delas. Nem ao menos uma única vez demos a essas empresas alguns anos para que trabalhem tranquilas, sem exigir que nos deslumbrem em intervalos regulares. Imagine se uma dessas empresas decidir não fazer nenhum anúncio oficial por alguns

meses — já achamos que o silêncio pressupõe que elas estão trabalhando em projetos especialíssimos que inevitavelmente nos incomodarão.

O governo dos EUA não tem uma grande estratégia para a IA, tampouco para os nossos futuros em longo prazo. Portanto, em vez de criar estratégias nacionais coordenadas a fim de desenvolver capacidade organizacional dentro do governo — para construir e fortalecer suas alianças internacionais, e com o intuito de preparar seu exército para o futuro da guerra —, os Estados Unidos relegaram a IA às tramitações legais associadas à prática de tráfico de influência por ex-funcionários na política (revolving door). Em vez de financiar pesquisa básica em IA, o governo federal terceirizou a pesquisa e o desenvolvimento para o setor comercial e aos caprichos de Wall Street. Em vez de tratar a IA como uma oportunidade para a criação de novos empregos e crescimento, os legisladores norte-americanos enxergam somente o desemprego tecnológico generalizado. Por sua vez, culpabilizam os gigantes da tecnologia dos EUA, quando poderiam convidar essas empresas a participar dos níveis mais altos do planejamento estratégico (como existe) dentro do governo. Nossos pioneiros da IA não têm escolha a não ser competir constantemente uns com os outros por uma relação confiável e direta com você, comigo, com nossas escolas, hospitais, cidades e negócios.

Nos Estados Unidos, sofremos de uma trágica carência de previsão. Temos uma mentalidade imediatista, de “nowist”, planejando para os próximos anos de nossas vidas mais do que qualquer outro período de tempo. A mentalidade nowist defende as conquistas tecnológicas de curto prazo, mas nos isenta de assumir a responsabilidade de como a tecnologia pode evoluir e das repercussões e consequências de nossas próximas ações. Nós facilmente esquecemos que o que fazemos no presente pode ter consequências sérias no futuro. É de se admirar, portanto, que o governo tenha terceirizado o desenvolvimento futuro da IA para seis empresas de capital aberto cujas realizações são notáveis, mas cujos interesses financeiros nem sempre se alinham com o que é

melhor para nossas liberdades individuais, nossas comunidades, nossos interesses e ideais democráticos.

Enquanto isso, na China, a rota de desenvolvimento da IA está presa às grandes ambições do governo, que rapidamente constrói os alicerces para se tornar a supremacia incontestável de IA do mundo. Em julho de 2017, o governo chinês revelou seu Plano de Desenvolvimento de Inteligência Artificial para Próxima Geração com o intuito de se tornar o líder global em IA até 2030, com uma indústria nacional no valor de, pelo menos, US\$150 bilhões¹ — plano que envolve a aplicação de parte de seu fundo de natureza contábil e financeira para novos laboratórios e startups, bem como o lançamento de novas escolas especificamente para treinar a próxima geração de talentos em IA da China.² Em outubro do mesmo ano, o presidente da China, Xi Jinping, explicou seus planos para IA e big data durante um discurso detalhado para milhares de oficiais do partido. Ele afirmou que a IA ajudaria a China a se transformar em uma das economias mais avançadas do mundo, e a economia chinesa já é 30 vezes maior do que era há apenas três décadas. O Baidu, a Tencent e o Alibaba podem ser gigantes de capital aberto, mas, como é típico de todas as grandes empresas chinesas, precisam se curvar à vontade de Pequim.

A população gigantesca de 1,4 bilhão de cidadãos da China a coloca no controle do maior e possivelmente mais importante recurso natural na era da IA: dados humanos. São necessárias quantidades abundantes de dados para refinar os algoritmos de reconhecimento de padrões — e é por isso que os sistemas de reconhecimento facial chineses, como o Megvii e o SenseTime, atraem e muito os investidores. Todos os dados que os cidadãos chineses geram enquanto telefonam, compram coisas online e publicam fotos em redes sociais estão ajudando o Baidu, o Alibaba e a Tencent a desenvolver os melhores sistemas de IA da categoria. Outra grande vantagem para a China: o país não tem restrições de privacidade e de segurança que dificultem o progresso, como nos Estados Unidos.

Devemos considerar a trajetória do desenvolvimento da IA dentro do contexto mais amplo dos grandes planos da China para o futuro. Em abril de 2018, Xi fez um discurso importante descrevendo sua visão da China como a superpotência cibernética global. O serviço chinês de notícias estatal, Xinhua, publicou trechos do discurso, nos quais descreveu uma nova rede de governança cibernética e uma internet que iria “divulgar informações positivas, defender a direção política correta e orientar a opinião pública e os valores no rumo certo”.³ As regras autoritárias que a China teria para todos nós vão de encontro à liberdade de expressão, à economia orientada para o mercado e ao controle distribuído que prezamos no Ocidente.

A IA faz parte de uma série de leis e decretos nacionais que visam controlar todas as informações geradas na China e monitorar os dados de seus moradores, bem como dos cidadãos de seus muitos parceiros estratégicos. Um desses decretos determina que todas as empresas estrangeiras armazenem os dados de cidadãos chineses em servidores dentro das fronteiras chinesas, permitindo que as agências de segurança do governo acessem dados pessoais como quiserem. Outra iniciativa — a Police Cloud, da China — foi desenvolvida para monitorar e rastrear pessoas com problemas de saúde mental, aquelas que criticaram publicamente o governo e uma minoria étnica muçulmana chamada Uighurs. Em agosto de 2018, a Organização das Nações Unidas (ONU) afirmou que tinha relatórios confiáveis de que a China mantinha milhões de uigures em campos de concentração na região do extremo oeste da China.⁴ O Programa de Operações Integradas da China utiliza a IA para detectar desvios de padrões — para saber se alguém atrasou o pagamento de contas. Um sistema monetário de crédito baseado em inteligência artificial, de acordo com as palavras de ordem em documentos de planejamento oficiais, fora desenvolvido para projetar uma sociedade livre de problemas “permitindo que as pessoas confiáveis percorram todo o mundo sob o céu, enquanto dificultava para os desacreditados dar um único passo”.⁵ Para promover “a confiabilidade”, os cidadãos são

avaliados em vários pontos de dados diferentes, como atos heroicos (pontos ganhos) ou multas de trânsito (pontos reduzidos). Aqueles com as pontuações mais baixas enfrentam obstáculos para se candidatar a empregos, comprar uma casa ou levar as crianças para as escolas. Em algumas cidades, os moradores com as pontuações altas têm suas fotos exibidas.⁶ Em outras cidades, como Shandong, os cidadãos que caminham na rua têm seus rostos compartilhados publicamente em outdoors digitais e enviados automaticamente para o Weibo, uma rede social popular.⁷ Se tudo isso parece fantástico demais para acreditar, lembre-se de que a China uma vez instituiu com sucesso a política de filho único para forçar o abate de sua população.

Essas políticas e iniciativas são criações que saem direto da cabeça das pessoas que fazem parte do círculo íntimo do presidente Xi Jinping, que durante a última década concentrou-se excepcionalmente na remodelação e reconstrução da China em uma superpotência global de peso. Hoje, a China é mais autoritária do que já fora sob o domínio de quaisquer líderes anteriores desde o presidente Mao Tsé-Tung, e o avanço e a alavancagem da IA são fundamentais para a causa. A Belt and Road Initiative é uma estratégia geoeconômica gigantesca disfarçada de um plano de infraestrutura que segue os antigos percursos da Rota da Seda que ligavam a China à Europa por meio do Oriente Médio e da África. A China não está somente construindo pontes e rodovias — está exportando tecnologia de vigilância e coletando dados no processo, à medida que maximiza a influência do CCP em todo o mundo, ao contrário da nossa atual ordem democrática liberal. A Global Energy Interconnection é mais uma estratégia nacional promovida por Xi, que tem o objetivo de criar a primeira rede global de eletricidade do mundo, que ela administraria. A China já descobriu como potencializar um novo tipo de tecnologia de cabos de alta voltagem que pode fornecer energia das regiões do extremo oeste à Xangai — e está fechando acordos enormes para se tornar a fornecedora de energia dos países vizinhos.

Essas iniciativas, em conjunto com muitas outras, são meios inteligentes de granjear um tipo de poder brando durante um longo período de tempo. É uma jogada brilhante de Xi, cujo partido político votou em março de 2018 para abolir os limites dos mandatos e efetivamente permitiu que ele continuasse como presidente vitalício. O objetivo final de Xi é muito claro: estabelecer uma nova ordem mundial na qual a China será a líder. E tem mais: durante este período de expansão diplomática chinesa, os Estados Unidos deram as costas inexoravelmente a alianças e acordos globais de longo prazo, enquanto o presidente Trump erigiu uma nova barreira política e econômica contra a China.

O futuro da IA hoje está percorrendo duas rotas de desenvolvimento que muitas vezes entram em conflito com o que é melhor para a humanidade. O impulsionamento da IA na China é parte de uma tentativa coordenada de criar uma nova ordem mundial liderada pelo presidente Xi, ao passo que as forças de mercado e o consumismo são os principais motivadores nos Estados Unidos. Essa dicotomia é um ponto fraco para todos nós, e resolvê-la é uma questão de urgência para os problemas iminentes que espreitam a IA, além de ser o propósito deste livro. As empresas que fazem parte dos Nove Titãs da IA podem estar seguindo os mesmos objetivos nobres — decifrar o código da inteligência de máquina para construir sistemas capazes de pensamentos semelhantes aos seres humanos —, mas o resultado final desse trabalho poderia prejudicar de forma irreparável a humanidade.

Lá no fundo, acredito que a IA é uma força positiva, que elevará as próximas gerações da humanidade e nos ajudará a conquistar nossos objetivos mais idealistas de futuro.

Mas sou pragmática. Todos nós sabemos que mesmo as pessoas com boas intenções podem, sem querer, causar um grande mal. Dentro da tecnologia, e sobretudo quando se trata de IA, devemos sempre nos lembrar de planejar tanto a sua aplicação pretendida quanto sua aplicação abusiva. Isso é ainda mais importante hoje e em um futuro

próximo, pois a IA permeia tudo: a economia global, a mão de obra, a agricultura, o transporte, o setor bancário, o monitoramento ambiental, a educação, o exército e a segurança nacional. É por isso que, se a IA permanecer em suas atuais rotas de desenvolvimento nos Estados Unidos e na China, o ano de 2069 poderia ser bem diferente do que o ano de 2019. À medida que as estruturas e sistemas que governam a sociedade passam a depender da IA, descobriremos que as decisões tomadas em nosso nome fazem todo o sentido para as máquinas —, mas não para nós.

Nós, seres humanos, estamos perdendo rapidamente nossa consciência, assim como as máquinas estão despertando. Começamos a superar alguns marcos fundamentais no desenvolvimento técnico e geopolítico da IA, mas, a cada novo avanço, a IA se torna mais invisível para nós. Os meios pelos quais nossos dados são extraídos e refinados ficam cada vez menos evidentes, enquanto nossa capacidade de entender como os sistemas autônomos tomam decisões fica cada vez menos transparente. Temos, portanto, um abismo no tocante à compreensão de como a IA está impactando a vida cotidiana no presente, já que ela cresce exponencialmente à medida que avançamos anos e décadas para o futuro. Diminuir esse abismo o máximo possível, por meio de uma crítica a respeito das rotas que a IA trilha atualmente, é a missão deste livro. Meu objetivo é democratizar as conversas sobre inteligência artificial e deixá-lo mais ciente do que está por vir — e fazer com que os futuros impactos da IA no mundo real sejam tangíveis e relevantes para você, antes que seja tarde demais.

A humanidade enfrenta uma crise existencial em um sentido amplo da palavra, porque ninguém trata de uma questão simples que tem sido fundamental para a IA desde o princípio: o que acontece à sociedade quando transferimos o poder para um sistema construído por um pequeno grupo de pessoas que foi desenvolvido para tomar decisões em nome de todos nós? O que acontece quando essas decisões são tendenciosas para as forças de mercado ou para um partido político ambicioso? A resposta se reflete nas futuras oportunidades que temos,

nas formas como nos é negado o acesso, nas convenções sociais dentro de nossas sociedades, nas regras pelas quais nossas economias operam e até mesmo na forma como nos relacionamos com outras pessoas.

Este não é um livro sobre as discussões comuns de IA. É tanto um aviso quanto um projeto para um futuro melhor. Ele questiona nossa aversão ao planejamento de longo prazo nos EUA e destaca a falta de preparação para a IA em nossos negócios, escolas e governo. Esta obra pinta uma imagem nua e crua das estratégias geopolíticas, econômicas e diplomáticas interconectadas da China, enquanto ela avança rumo a sua grande visão de uma nova ordem mundial. E pede liderança heroica sob circunstâncias extremamente desafiadoras. Porque, como você está prestes a descobrir, nosso futuro precisa de um herói.

O que segue é um convite para colocar as mãos na massa escrito em três partes. Na Parte I, você aprenderá o que é a IA e o papel que os Nove Titãs da IA desempenharam no desenvolvimento dela. Depois, também mergulharemos a fundo nas situações únicas enfrentadas pelos membros dos Nove Titãs da IA dos Estados Unidos e pelo Baidu, Alibaba e Tencent na China. Na Parte II, você verá previsões detalhadas e plausíveis para os próximos 50 anos à medida que a IA evolui. Os três cenários sobre os quais você lerá oscilam entre otimistas, pragmáticos e catastróficos, e revelarão a oportunidade e o risco conforme evoluímos da inteligência artificial de uso específico para a inteligência artificial de uso generalizado até a superinteligência artificial. São cenários intensos, que resultam de modelos baseados em dados, e lhe proporcionarão um vislumbre emocionante de como a inteligência artificial pode evoluir e como nossas vidas mudarão. Na Parte III, ofereço soluções táticas e estratégicas para todos os problemas identificados nos cenários, junto com um plano concreto para remodelar o presente. A Parte III tem o objetivo de nos sacudir para que entremos em ação; há recomendações específicas para os nossos governos, para os líderes das Nove Titãs da IA e até mesmo para você.

Toda pessoa viva hoje pode desempenhar um papel crítico no futuro da inteligência artificial. As decisões que tomamos sobre a IA agora — mesmo as mais pequenas — mudarão para sempre o rumo da história humana. À medida que as máquinas despertam, percebemos que, apesar de nossas esperanças e anseios altruístas, nossos sistemas de IA foram catastróficamente ruins para a humanidade.

Mas não precisa ser assim.

Os Nove Titãs não são os vilões nesta história. Na verdade, eles são a nossa melhor esperança para o futuro.

Vire a página porque não podemos ficar de braços cruzados esperando o que está por vir. A IA já está entre nós.

1 Paul Mozur, “Beijing Wants AI to Be Made in China by 2030”, New York Times, 20 de junho de 2017, <https://www.nytimes.com/2017/07/20/business/china-artificial-intelligence.html>.

2 Tom Simonite, “Ex-Google Executive Opens a School for AI, with China’s Help”, Wired, 5 de abril de 2018, <https://www.wired.com/story/ex-google-executive-opens-a-school-for-ai-with-chinas-help/>.

3 “Xinhua Headlines: Xi outlines blueprint to develop China’s strength in cyberspace”, Xinhua, abril de 2018. http://www.xinhuanet.com/english/2018-04/21/c_137127374_2.htm.

4 Stephanie Nebehay, “U.N. says it has credible reports that China holds million Uighurs in secret camps”, Reuters, 10 de agosto de 2018. <https://www.reuters.com/article/us-china-rights-un/u-n-says-it-has-credible-reports-that-china-holds-million-uighurs-in-secret-camps-idUSKBN1KV1SU>.

5 Simina Mistreanu, “Life Inside China’s Social Credit Laboratory”, Foreign Policy, 3 de abril de 2018.

<https://foreignpolicy.com/2018/04/03/life-inside-chinas-social-credit-laboratory/>.

6 Ibid.

7 “China Shames Jaywalkers through Facial Recognition”, Phys.org, 20 de junho de 2017, <https://phys.org/news/2017-06-china-shames-jaywalkers-facial-recognition.html>.

PARTE I

Máquinas Assombradas

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 as 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 grudada 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 quantitativos, como calcular o valor das 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

*image
not
available*

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

*image
not
available*

*image
not
available*

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

*image
not
available*

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 construía 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

*image
not
available*

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 as questões levantadas por Hobbes, Descartes, Hume e Leibniz. No artigo, ele propôs uma tese e um teste: se algum dia um computador conseguir responder a perguntas de maneira indistinguível dos seres humanos, então ele deve estar “pensando”. É bem provável que você já tenha ouvido falar deste artigo por outro nome: o teste de Turing.

O artigo começou com uma pergunta, famosa nos dias de hoje, feita e respondida por muitos filósofos, teólogos, matemáticos e cientistas antes dele: “As máquinas podem pensar?” Mas Turing, ciente do debate secular a respeito da mente e da máquina, descartou a questão como abrangente demais para empreender uma discussão significativa. “Máquina” e “pensar” eram palavras ambíguas, com muito espaço para interpretações subjetivas. (Afinal, 400 anos de documentos e livros já haviam sido escritos sobre o significado dessas palavras.)

*image
not
available*

que tinha sugerido muitos anos antes ensinar os computadores a jogar xadrez contra os humanos, teve a oportunidade de demonstrar um protótipo de seu programa, que ainda estava em construção.²⁰

As expectativas de McCarthy e Minsky para avanços revolucionários em IA não se materializariam naquele verão em Dartmouth. Não houvera tempo suficiente — sem contar que não havia processamento computacional o bastante — para que a IA evoluísse da teoria à prática.²¹ Ainda assim, aquele verão definiu as três práticas-chave que se tornariam os alicerces fundamentais da IA tal como a conhecemos hoje:

1. A IA seguiria uma linha teórica, seria construída, testada e promovida por grandes empresas de tecnologia e pesquisadores acadêmicos trabalhando juntos;
2. o desenvolvimento da IA exigia muito dinheiro, portanto seria necessário monetizar o trabalho de alguma forma — seja trabalhando por meio de parcerias com agências governamentais ou militares, ou construindo produtos e sistemas que pudessem ser vendidos — era necessário; e
3. o estudo e a construção da IA dependiam de uma rede de pesquisadores interdisciplinares, o que significava inaugurar um campo acadêmico novo a partir do zero. Significava também que as pessoas que estavam no campo tendiam a recrutar pessoas que já conheciam, o que mantinha a rede relativamente homogênea e restringia sua visão de mundo.

Naquele verão, aconteceu outra iniciativa interessante. Enquanto o grupo se uniu em torno da questão levantada por Turing — As máquinas podem pensar? — eles estavam divididos sobre a melhor abordagem para provar sua resposta, que era construir uma máquina de aprendizado.

*image
not
available*

um secretário”. Sua promessa não era somente o maior “dispositivo de pensamento” já visto, mas também um que estaria operacional dentro de alguns meses.²⁵

Mas e Simon e Newell, que construíram o Logic Theorist? Eles começaram a fazer previsões ousadas sobre a IA, dizendo que dentro de dez anos — ou seja, em 1967 — os computadores seriam capazes de:

- vencer todos os grandes mestres de xadrez e serem os campeões mundiais;
- descobrir e revelar um novo teorema matemático importante; e
- compor o tipo de música que até mesmo os críticos mais ferrenhos gostariam.²⁶

Enquanto isso, Minsky prevera que uma máquina inteligente poderia fazer muito mais do que ouvir textos ditados, jogar xadrez ou compor música. Ele alegava que, durante a sua vida, as máquinas conquistariam a inteligência artificial — isto é, os computadores seriam capazes de produzir pensamentos complexos, se expressar por meio da linguagem e fazer escolhas.²⁷

Os pesquisadores do seminário de Dartmouth escreveram artigos e livros, e foram entrevistados pela televisão, rádios, jornais e revistas. Mas a ciência era difícil de explicar e, por isso, muitas vezes, as explicações eram distorcidas e as citações ficavam fora do contexto. Deixando os prognósticos extravagantes de lado, as expectativas do público em relação à inteligência artificial ficariam cada vez mais fantasiosas, em parte porque a história era noticiada de forma duvidosa. Por exemplo, a revista *Life* afirmara que Minsky disse: “Em um período de três a oito anos teremos uma máquina com a inteligência média de um ser humano normal. Quero dizer, uma máquina que conseguirá ler Shakespeare,

*image
not
available*

compreender como entendemos a linguagem — os processos físicos e cognitivos que fazem com que o reconhecimento da linguagem seja possível. McCarthy gostava de ilustrar uma gaiola para explicar o desafio do avanço da IA. Digamos que eu lhe pedisse para construir uma gaiola e não lhe desse nenhum outro parâmetro. Você provavelmente construiria uma gaiola com uma parte superior, inferior e com as laterais. Caso eu lhe desse algumas informações adicionais — a ave é um pinguim — logo, você poderia não colocar uma parte superior na gaiola. Portanto, a gaiola requerer uma parte superior ou não depende de algumas coisas: a informação que eu lhe dei e todas as associações que você já tem com a palavra “ave”, como o fato de que a maioria delas voa. Nós partimos do princípio de algumas suposições e do contexto. Para que uma IA respondesse do mesmo jeito que respondemos, mais informações e instruções explícitas seriam exigidas.³² O inverno da IA persistiria por mais de três décadas.³³

*image
not
available*

complexidades excêntricas do jogo. Em 1971, o primeiro programa criado pelo cientista da computação Jon Ryder operava de uma perspectiva técnica, mas perdeu para um iniciante. Em 1987, um programa de computador mais poderoso chamado Nemesis competiu contra um humano pela primeira vez em um torneio ao vivo. Em 1994, o programa conhecido como Go Intellect provou ser um jogador competente. Porém, mesmo com a vantagem significativa de um handicap, ele ainda perdeu todos os três jogos — contra crianças. Em todos esses casos, os computadores faziam movimentos de jogo incompreensíveis, jogavam na ofensiva ou calculavam mal o comportamento do adversário.

Em algum momento, no meio de todo esse trabalho, existia um grupo de pesquisadores que, mais uma vez, estavam participando de seminários em redes neurais, uma ideia promovida por Marvin Minsky e Frank Rosenblatt durante o primeiro seminário em Dartmouth. O cientista em cognição Geoff Hinton e os cientistas de computação Yann Lecun e Yoshua Bengio acreditavam que os sistemas baseados em redes neurais não somente teriam aplicações práticas importantíssimas — como detecção automática de fraudes para cartões de crédito e reconhecimento ótico automático de caracteres para leitura de documentos e inspeções bancárias — como também seriam a base para o que a inteligência artificial se transformaria.

Fora Hinton, um professor da Universidade de Toronto, que imaginou um novo tipo de rede neural, uma rede composta de múltiplas camadas que extrairiam informações diferentes até reconhecerem o que procuravam. A única maneira de coletar esse tipo de conhecimento em um sistema de inteligência artificial, pensou ele, era desenvolver algoritmos de aprendizagem que permitissem que os computadores aprendessem sozinhos. Em vez de ensiná-los a realizar bem uma única tarefa restritiva, as redes seriam construídas para treinarem a si próprias.

Essas novas redes neurais “profundas” (DNNs) exigiriam um tipo de aprendizado de máquina mais avançado — “aprendizado profundo” —

*image
not
available*

Em janeiro de 2014, o Google havia começado a investir de forma massiva em IA, mais de US\$500 milhões para adquirir uma startup de aprendizado de máquina profundo chamada DeepMind e seus três fundadores, o neurocientista Demis Hassabis, a ex-criança prodígio do xadrez Shane Legg, pesquisador de aprendizado de máquina, e o empresário Mustafa Suleyman. Parte do atrativo da equipe: eles desenvolveram um programa chamado AlphaGo.

Em poucos meses, eles estavam prontos para testar o AlphaGo contra um jogador humano. Uma partida foi organizada entre a DeepMind e Fan Hui, um jogador Go profissional nascido na China e um dos melhores mestres Go da Europa. Como jogar Go em um computador não é exatamente o mesmo que jogar em um tabuleiro físico, foi decidido que os engenheiros da DeepMind colocariam as jogadas do computador no tabuleiro e passariam para o computador as jogadas de Hui.

Antes do jogo, Toby Manning, que era um dos responsáveis da British Go Association, jogou AlphaGo em uma rodada de testes — e perdeu por 17 pontos. Manning cometeu alguns erros, mas o programa também cometeu. Um pensamento inquietante passou pela cabeça dele: e se o AlphaGo estivesse apenas jogando de modo tradicional? Seria possível que o programa estivesse jogando de forma defensiva o bastante para apenas vencer Manning, em vez de derrotá-lo por completo?

Os jogadores sentaram-se à mesa, Fan Hui usando uma camisa de botões listrada e uma jaqueta de couro marrom, Manning no centro e o engenheiro do outro lado. O jogo começara. Hui abriu uma garrafa de água e examinou o tabuleiro. Como era o jogar das pedras pretas, era ele que começava. Durante as primeiras 50 jogadas, foi um jogo tranquilo — Hui estava claramente tentando identificar os pontos fortes e fracos do AlphaGo. Um deles era: a IA não jogaria de forma agressiva, a menos que tivesse perdendo. Foi um jogo apertado, e a AlphaGo venceu por pouco, apenas 1,5 ponto.

Hui usou essa informação no segundo jogo. Se o AlphaGo não fosse jogar de forma agressiva, Hui partiria para o ataque. Mas então o

*image
not
available*

Como Zero jogou contra si próprio, descobriu as estratégias Go que os humanos desenvolveram ao longo de mil anos — ou seja, ele aprendeu a pensar como os humanos que o criaram. Nos estágios iniciais, cometeu os mesmos erros, identificou os mesmos padrões e variações e enfrentou os mesmos obstáculos que nós. Porém, uma vez que Zero ficou poderoso o bastante, deixou de lado nossas jogadas humanas e começou a ter preferências.⁴² Depois que decolou por conta própria, desenvolveu estratégias criativas que ninguém jamais havia visto antes, sugerindo que talvez as máquinas já estivessem pensando de modos conhecidos e desconhecidos para nós.

Zero demonstrara também que os algoritmos agora conseguiam aprender sem a orientação humana, e nós, humanos, estávamos retardando o avanço dos sistemas de IA. Significava que, em um futuro próximo, as máquinas poderiam ser livres para solucionar problemas que nós, por conta própria, não poderíamos prever ou resolver.

Em dezembro de 2017, a equipe da DeepMind publicou um artigo mostrando que Zero conseguia aprender em geral — não apenas Go, mas outras informações. Por conta própria, o programa estava jogando outros jogos, como xadrez e shoji (um jogo japonês parecido com o xadrez), que são sem dúvida menos complexos, mas ainda exigem estratégia e criatividade. Mas agora Zero estava aprendendo muito mais rápido que antes e conseguiu desenvolver um poder sobre-humano incompreensível, com menos de 24 horas de jogo. Desde então, a equipe começou a trabalhar na aplicação das técnicas que desenvolveu para o Zero com o objetivo de construir uma “máquina de aprendizado para usos gerais”, um conjunto de algoritmos adaptativos que simulam nossos próprios sistemas biológicos, capazes de serem treinados. Em vez de alimentar os sistemas de IA com uma quantidade gigante de informações e um conjunto de instruções sobre como isso pode ser consultado, a equipe ensina às máquinas como aprender. Ao contrário dos humanos, que podem ficar cansados, entediados ou distraídos quando estudam, as máquinas perseguem implacavelmente um objetivo a todo custo.

*image
not
available*

manifestação da visão e do propósito de Deus; cada alma fora criada e agraciada por um criador sui generis. As máquinas pensantes também têm criadores — eles são os novos deuses da IA, e são em sua maioria homens, predominantemente na América, Europa Ocidental e China, que estão conectados de uma forma ou outra aos Nove Titãs da IA. A alma da IA é uma manifestação da visão e do propósito deles para o futuro.

E por fim, sim, as máquinas pensantes são capazes de pensar de forma inédita. Depois de aprender com a experiência, elas podem determinar que uma solução diferente é possível. Ou que uma nova classificação seja melhor. As IAs não precisam inventar uma nova forma de arte para nos mostrar criatividade.

O que significa que, de fato, as máquinas pensantes têm uma mente. Elas são jovens e ainda estão amadurecendo, e é provável que evoluam de maneiras que não entendemos. No próximo capítulo falaremos sobre o que constitui essa mente, os valores dos Grandes Titãs da IA e as imprevisíveis consequências sociais, políticas e econômicas em relação ao despertar da IA.

1 “The Seikilos Epitaph: The oldest song in the world”, Wired, 29 de outubro de 2009: <https://www.wired.com/2009/10/the-seikilosepitaph>.

2 “Population Clock: World”, censo do governo de 2018: <https://www.census.gov/popclock/world>.

3 Elizabeth King, “Clockwork Prayer: A sixteenth-century mechanical monk”, Blackbird 1, n°1 (primavera de 2002): https://blackbird.vcu.edu/v1n1/nonfiction/king_e/prayer_introduction.htm.

4 Thomas Hobbes, De Corpore Politico, or The Elements of Law Moral and Politick.

5 René Descartes, Meditations on First Philosophy, Second Meditation §25, 1641, University of Connecticut: <http://selfpace.uconn.edu/class/percep/DescartesMeditations.pdf>.

*image
not
available*

Harmon, Leon D.
Bell Telephone Laboratories Murray Hill, NJ
Holland, John H.
E. R. I.
University of Michigan
Ann Arbor, MI
Holt, Anatol
7358 Rural Lane Filadélfia, PA
Kautz, William H.
Stanford Research Institute Menlo Park, CA
Luce, R. D.
427 West 117th Street
Nova York, NY
MacKay,
Donald Department of Physics
University of London
Londres, WC2, Inglaterra
McCarthy, John
Dartmouth College Hanover, NH
McCulloch, Warren S.
R.L.E., MIT
Cambridge, MA
Melzak, Z. A.
Mathematics Department
University of Michigan Ann Arbor, MI
Minsky, M. L.
112 Newbury Street
Boston, MA
More, Trenchard
Department of Electrical Engineering
MIT
Cambridge, MA
Nash, John
Institute for Advanced Studies
Princeton, NJ

*image
not
available*

20 “The Dartmouth Workshop — as Planned and as It Happened”, Stanford Computer Science Department’s Formal Reasoning Group, site de John McCarthy, aula “AI: Past and Future”, modificado pela última vez em 30 de outubro de 2006: <http://www.formal.stanford.edu/jmc/slides/dartmouth/dartmouth/node1.html>.

21 “The Dartmouth AI Archives”, RaySolomonoff.com: <http://raysolomonoff.com/dartmouth/>.

22 Irving John Good, “Speculations Concerning the First Ultraintelligent Machine”, Advances in Computers, volume 6 (1966): 31–88: <https://www.sciencedirect.com/science/article/pii/S0065245808604180?via%3Dihub>.

23 Joseph Weizenbaum, “ELIZA — A Computer Program for the Study of Natural Language Communication Between Man and Machine”, Communications of the ACM 9, n°1 (janeiro de 1966): 36–45: <http://web.stanford.edu/class/cs124/p36-weizenbaum.pdf>.

24 O script completo está no GitHub: <https://github.com/codeanticode/eliza>.

25 Ronald Kotulak, “New Machine Will Type Out What It ‘Hears’”, Chicago Tribune, 18 de junho de 1963, acessado via arquivos do Chicago Tribune (paywall).

26 Herbert A. Simon e Allen Newell, “Heuristic Problem Solving: The next advance in operations research”, Operations Research 6 (1958): 1–10.

27 O próprio McCarthy queria trabalhar com o grupo em suas ideias para representar o conhecimento e o raciocínio do senso comum, mas, depois que o grupo se reuniu, ele percebeu que a matriz de participantes estava perdendo alguns pensadores-chave. (No caso dele, esperava por matemáticos.)

28 Brad Darrach, “Meet Shaky, the First Electronic Person”, Life Magazine, 20 de novembro de 1970, volume 69, 58B–58C.

29 National Research Council, Language and Machines: Computers in translation and linguistics (Washington, D.C.: The National Academies Press, 1966):

<https://www.nap.edu/read/9547/chapter/1>.

*image
not
available*

O Mundo Isolado das Tribos de IA

O esforço hercúleo de séculos para construir uma máquina pensante somente há pouco tempo testemunhou grandes avanços. Porém, enquanto essas máquinas aparentemente “pensam”, devemos esclarecer que elas certamente não pensam como todos nós o fazemos.

O futuro da IA está sendo engendrado por um número relativamente pequeno de pessoas que têm a mesma mentalidade de grupos pequenos e isolados. Mais uma vez, acredito que essas pessoas têm boas intenções. Mas, como acontece com todos os grupos isolados que trabalham em conjunto, seus preconceitos e miopia inconscientes costumam se tornar os novos sistemas de crença e comportamentos aceitos com o passar do tempo. As coisas que no passado pareciam anormais — até mesmo erradas — se normalizam como mentalidade habitual, e esta mentalidade está sendo programada em nossas máquinas.

Aqueles que trabalham com IA pertencem a uma espécie de tribo. São pessoas que vivem e trabalham na América do Norte e na China, que frequentam as mesmas universidades e adotam um conjunto de regras sociais. As tribos são predominantemente homogêneas, ricas e extremamente qualificadas. Os membros da tribo são em sua maioria do sexo masculino. Seus líderes — diretores-executivos, conselheiros, gerentes seniores — são, com raríssimas exceções, homens. A homogeneidade também é um problema do outro lado, em que os membros da tribo são majoritariamente chineses.

O problema com as tribos é o que as torna tão poderosas. Nos grupos isolados, o viés cognitivo é potencializado e se enraíza, passando despercebido pela consciência. O viés cognitivo substitui a mentalidade racional, o que retarda nosso entendimento e costuma ser mais cansativo. Quanto mais relacionada e estabelecida uma tribo fica, mais normal parece ser a mentalidade e o comportamento do grupo. Conforme você verá a seguir, é uma constatação que vale a pena lembrar.

O que as tribos de IA estão fazendo? Elas estão desenvolvendo sistemas de inteligência artificial estreita (ANI — Artificial Narrow Intelligence) capazes de realizar uma tarefa específica no mesmo nível ou melhor do que nós, humanos. As aplicações comerciais da ANI — e, conseqüentemente, da tribo — já estão tomando decisões em nosso nome a partir de nossas caixas de entrada de e-mail, quando pesquisamos coisas na internet, tiramos fotos com nossos telefones, dirigimos nossos carros e solicitamos cartões de crédito ou empréstimos. As tribos também estão construindo o que está por vir: sistemas de inteligência artificial geral (AGI — Artificial General Intelligence), que realizarão tarefas cognitivas generalizadas, pois são máquinas projetadas para pensar como nós. Todavia, quem, exatamente, é o “nós” a partir do qual esses sistemas de IA se alimentam? Quais valores, ideais e perspectivas de mundo estão sendo ensinados a esses sistemas?

A resposta simples não cabe a você — nem a mim. A inteligência artificial tem a mente de sua tribo, priorizando os valores, ideias e perspectivas de mundo de seus criadores. Mas também está começando a desenvolver uma mente própria.

*image
not
available*

University, McGill University e a Université de Montréal. Estas universidades abrigam grupos ativos de pesquisa acadêmica com fortes laços industriais.

Normalmente as tribos obedecem a regras e rituais, então vamos explorar os rituais de iniciação das tribos da IA, que começam com uma educação universitária rígida.

Na América do Norte, a ênfase dentro das universidades concentrou-se nas habilidades técnicas — como no domínio das linguagens de programação R e Python, know-how em processamento de linguagem natural e estatística aplicada, assim como orientação à visão computacional, biologia computacional e teoria dos jogos. Não é bem visto ter aulas fora da tribo, como um curso sobre filosofia da mente, mulheres muçulmanas na literatura ou colonialismo. Se estamos tentando construir máquinas pensantes capazes de pensar como os seres humanos, aparentemente é um contrassenso excluir o aprendizado sobre a condição humana. Neste momento, cursos como esses são deliberadamente deixados de fora do conteúdo programático, e é difícil que eles tenham espaço como cursos opcionais fora deste conteúdo.

A tribo exige habilidades, e há um zilhão de coisas para estudar durante os quatro anos de graduação. Por exemplo, em Stanford, os alunos devem ter 50 horas de crédito de aulas intensivas de matemática, ciências e engenharia, além de 15 horas de cursos básicos de computação. Ainda que exista um curso de ética oferecido como parte do conteúdo programático, é uma das cinco disciplinas opcionais que são oferecidas somente para se atender às exigências.¹⁶ Carnegie Mellon desenvolveu uma IA novinha em folha em 2018, que possibilitou à universidade um novo começo e a oportunidade de arquitetar uma IA moderna a partir do zero. Só que as regras e rituais da tribo imperam, e as habilidades técnicas são o que importam. Embora a graduação exija uma disciplina de ética e alguns cursos em ciências humanas, todos eles focam principalmente a neurociência (por exemplo, psicologia cognitiva, memória humana e cognição visual), o que faz sentido dado o elo entre a IA e a mente humana. Não existem cursos obrigatórios que ensinem os alunos a identificar tendências em conjuntos de dados, como aplicar a filosofia à tomada de decisões ou à ética inclusiva. Não existe um reconhecimento formal em todos os cursos de que a diversidade social e socioeconômica é de suma importância para uma comunidade no que diz respeito à biodiversidade.

As habilidades são ensinadas de modo empírico — o que significa que os alunos que estudam IA leem menos do que deveriam. Para aprender, eles precisam de bancos de dados lexicais, bibliotecas de imagens e redes neurais. Durante certo tempo, uma das redes neurais mais populares nas universidades se chamava Word2vec, e fora construída pela equipe do Google Brain. Era um sistema de duas camadas que processava texto, transformando palavras em números que a IA conseguia compreender.¹⁷ Por exemplo, esta rede neural aprendeu que “o homem está mais para rei como a mulher está mais para rainha”. Mas o banco de dados também decidiu que “o pai combina mais com médico e a mãe combina mais com enfermeira” e que “o homem está mais para programador de computador como a mulher está mais para dona de casa”.¹⁸ O próprio sistema ao qual os alunos foram expostos era tendencioso. Caso alguém quisesse analisar as repercussões mais abrangentes desse código-pai sexista, não havia aulas que promovessem este tipo de aprendizado.

Em 2017 e 2018, algumas dessas universidades criaram poucos cursos novos de ética em resposta aos desafios já impostos pela IA. O Berkman Klein Center, em Harvard, e o MIT Media Lab, em conjunto, ofereceram um curso novo sobre ética e regulamentação da IA.¹⁹ O curso e as palestras eram formidáveis,²⁰ mas foram organizados na contramão dos caminhos padrões da ciência da computação de cada universidade — ou seja, o que estava sendo ensinado e discutido não tinha a oportunidade de se infiltrar lentamente em outras partes do conteúdo programático.

Não restam dúvidas de que a ética é uma exigência de todas as universidades que ensinam IA — é exigência dos padrões de credenciamento dos cursos de graduação. A fim de serem credenciados pelo Accreditation Board for Engineering and Technology (Conselho de Credenciamento de Engenharia e Tecnologia), os cursos de graduação de ciência da computação devem mostrar que os alunos têm uma “compreensão das questões e responsabilidades profissionais, éticas, legais, de segurança e sociais” e a “habilidade de analisar o impacto local e global da computação em indivíduos, organizações e sociedade”. Contudo, posso lhe dizer por experiência própria, que realizar uma análise comparativa e avaliar este tipo de exigência é, na melhor das hipóteses, subjetivo e bastante difícil de se fazer com alguma precisão, ainda mais sem os cursos obrigatórios que todos os alunos deveriam fazer. Eu sou membro Accrediting Council on Education in Journalism and Mass Communications. Os conteúdos programáticos de jornalismo

*image
not
available*

que o recrutara para ser o cientista-chefe da empresa. (Inevitavelmente, o DNA do Baidu tem nucleotídeos dos cursos de IA ministrados na Carnegie Mellon, no MIT e na UC Berkeley.)

Hoje o Baidu passa longe de ser somente um mecanismo de busca. Ng passou a ajudar na promoção da plataforma de conversação em IA (chamada DuerOS), de assistentes digitais, de programas de carros autônomos, como também de outros frameworks de IA, todos em funcionamento — e isso levou o Baidu a começar a falar da IA em suas conferências de divulgação de lucros, bem à frente do Google. O Baidu tem um valor de mercado de US\$88 bilhões e é o mecanismo de busca mais usado no mundo depois do Google — um belo de um feito, considerando que ele não é usado fora da China. Como o Google, o Baidu está desenvolvendo um conjunto de dispositivos domésticos inteligentes, como um robô familiar que combina reconhecimento de voz e reconhecimento facial. O Baidu anunciou uma plataforma aberta de condução autônoma chamada Apollo, e a esperança é de que, ao disponibilizar publicamente o código-fonte, o ecossistema ao redor da plataforma germine. Ela já tem 100 parceiros, dentre eles as montadoras Ford e Daimler, os fabricantes de chips NVIDIA e Intel, e os provedores de serviços de mapeamento como a TomTom. O Baidu fez uma parceria com a Access Services, sediada na Califórnia, com o intuito de lançar veículos autônomos para pessoas com problemas de mobilidade e deficiências. E também fez parceria com o Azure Cloud da Microsoft a fim de possibilitar que os parceiros não chineses da Apollo processassem grandes quantidades de dados de veículos.²⁹ Talvez você tenha ouvido falar que, nos últimos anos, o Baidu abriu um novo laboratório de pesquisa de IA em cooperação com o governo chinês — e os líderes do laboratório são as elites do Partido Comunista que já haviam trabalhado em programas militares do estado.³⁰

O A na tribo BAT chinesa pertence ao Alibaba Group, uma plataforma gigantesca que funciona como intermediária entre os compradores e os vendedores por meio de uma vasta rede de sites, em vez de uma única plataforma. Ele foi fundado em 1999 por Jack Ma, um ex-professor que vivia cerca de 160 quilômetros a sudoeste de Xangai, que queria desenvolver uma versão híbrida da Amazon e do eBay para a China. Ma não sabia programar, então abriu uma empresa com um colega com quem frequentou a universidade. Apenas 20 anos mais tarde, o Alibaba tem um valor de mercado de mais de US\$511 bilhões.

Entre seus sites estão o Taobao, por meio do qual nem compradores nem vendedores pagam taxas por suas transações. Em vez disso, o Taobao usa um modelo do estilo pay-to-play, cobrando os vendedores para classificá-los com a posição mais elevada no mecanismo de busca do site. (Isso imita parte do principal modelo de negócios do Google.) O Alibaba também criou sistemas de pagamento seguros, incluindo o Alipay, que se assemelha às funcionalidade e aos recursos do PayPal. A empresa lançou um sistema de pagamento digital de reconhecimento facial baseado em inteligência artificial que, em 2017, estreou em uma loja e permitia aos consumidores efetuar o pagamento ao tirar uma selfie com um breve sorriso, o famoso smile to pay.

Assim como a Amazon, o Alibaba também tem um alto-falante inteligente — chama-se Genie X1, que inclusive é menor e mais barato do que os dispositivos Alexa, da Amazon, e o Google Maps. O Genie X1 utiliza a tecnologia de reconhecimento de impressão de voz baseada em rede neural para identificar usuários, realizando a identificação deles automaticamente para que consigam vender e fazer compras. Mais de 100 mil Genie X1 da Alibaba estão sendo instalados nos hotéis Marriott em toda a China.

O Alibaba tem uma perspectiva mais ampla para o alcance da IA, que a empresa chama de seu ET City Brain. O programa consome enormes quantidades de dados locais, desde câmeras e sensores de cidades inteligentes até registros governamentais e contas individuais de mídia social. O Alibaba usa sua estrutura de IA para a modelagem preditiva: com o intuito de identificar de antemão o gerenciamento do tráfego, o desenvolvimento urbano, as necessidades de saúde pública e se pode haver instabilidade social à vista. Sob o comando de Ma, o Alibaba fez avanços na logística de entrega, vídeos online, data centers e computação em nuvem, investindo bilhões de dólares em diversas empresas na tentativa de criar um mastodonte digital que cresce exponencialmente, conectando comércio, casa, trabalho, cidades e governo. Na verdade, antes de a loja Amazon Go ser lançada em Seattle, o Alibaba inaugurou a Hema, uma operação de varejo multifuncional automatizada, que não envolvia transações de dinheiro e combinava produtos alimentares, um mercado de comida rápida com preço médio e serviço de entrega.

Existe mais um ponto em comum enigmático e digno de nota. Eu digo “enigmático” porque também é uma contradição. Em 2016, Ma comprou o South China Morning Post, que já fora o maior e mais influente jornal

*image
not
available*

reciclados. A China rapidamente se tornou líder global em sustentabilidade e é poderosa o bastante para ditar as regras.

Na China, as pessoas gostam das *chengyu*, expressões idiomáticas chinesas que visam transmitir um pouco de sabedoria. Lembro-me de uma delas que descreve este momento específico: 脱颖而出, que se traduz em “a semente se despoja de sua casca e frutifica”.⁵¹ Agora o país mostra plenamente ao mundo sua força e seu poder, publicamente.

A consolidação do poder de Xi, junto com a ascensão econômica e a influência da China, criou as condições perfeitas para que as tribos da IA prosperassem, sobretudo dado a iniciativa unificada da IA em todos os cantos do país. Existe um parque de pesquisa de US\$2 bilhões que está sendo construído nos arredores de Pequim, que se concentrará em aprendizado profundo, computação em nuvem e biometria e contará com um laboratório de pesquisa e desenvolvimento em nível estadual. O governo não está somente investindo na BAT, mas também a protege contra a competição ferrenha do mundo. O governo chinês banuiu o Google e o Facebook do país, e impossibilitou a entrada da Amazon no mercado. As empresas BAT são o coração do plano de 2030 do governo, que depende de modo providencial de suas tecnologias: sistemas de condução autônoma do Baidu, IoT do Alibaba e sistemas de varejo conectados, e o trabalho da Tencent em interfaces de conversação e assistência médica.

Por isso, as tribos da IA da China devem estar preocupadas com você, independentemente de que lugar no mundo você vive.

Em primeiro lugar, a economia da China vem crescendo a um ritmo acelerado, e o rápido desenvolvimento da IA potencializará a escalada do país. No final de 2017, por meio da modelagem e da análise da minha equipe do Future Today Institute, demonstrei que a IA tem o potencial de impulsionar a economia da China em 28% até 2035. A IA — alimentada pelo grande número de chineses e seus dados, automação abrangente, aprendizado de máquina, autocorreção em escala e melhorias dos investimentos eficientes — estimulará o crescimento em toda a manufatura, agricultura, varejo, fintech e serviços financeiros, transporte, serviços públicos, assistência médica e mídia de entretenimento (incluindo plataformas) chineses.

Neste exato momento, não há outro país na Terra com tantos dados quanto a China, tantas pessoas quanto a China e tantos eletrônicos per capita. Nunca em nossas vidas nenhum outro país esteve em condições de ser maior do que a economia dos Estados Unidos. Nenhum outro país tem mais potencial para influenciar o ecossistema, o clima e os padrões climáticos do nosso planeta — levando à sobrevivência ou à catástrofe — do que a China. Nenhum outro país reconciliou o mundo desenvolvido e em desenvolvimento como a China. Como poder comunista e força motriz econômica, o país é um parceiro que agora é grande demais para ser ignorado, um adversário político que tem pontos de vista extremamente diferentes no que diz respeito aos direitos humanos e um canal para alianças globais. O aumento da riqueza vem acompanhado do poder. A China está se posicionando para influenciar a oferta global de dinheiro e comércio internacional, destituindo inevitavelmente outros países dessas posições de poder e influência, além de enfraquecer os ideais democráticos ao redor do mundo.

Em segundo lugar, a China alavancará seus avanços em IA e estímulo econômico a fim de modernizar seu poderio militar, conquistando uma vantagem sobre as nações ocidentais. Essa transição já começou como parte de um programa de vigilância doméstica aerotransportado, de codinome Dove. Mais de 30 agências militares e governamentais colocaram em campo “aves espiãs” que se parecem com pássaros brancos, imitando o bater biológico das asas. Os drones são parte de um programa de drones inspirados de forma biológica, com o intuito de sabotar radares e burlar a detecção humana.⁵² Os drones registram o movimento, e um sistema de inteligência artificial procura padrões, reconhece rostos e identifica anomalias. Contudo, os pássaros espiões, ainda que pareçam assustadores, são a menor das suas preocupações.

No final de 2017, um relatório não divulgado do Pentágono, obtido por repórteres da Reuters, alertou que as empresas chinesas estavam burlando a vigilância dos EUA e obtendo acesso à tecnologia confidencial de inteligência artificial para possíveis aplicações militares por meio da compra de participações em empresas norte-americanas. O

*image
not
available*

IA no Canadá e nos Estados Unidos, oferecendo-lhes pacotes convidativos de repatriação. Já vivemos uma escassez de cientistas de dados treinados e especialistas em aprendizado de máquina. Exaurir as pessoas pouco a pouco em breve criará uma verdadeira lacuna de talentos no Ocidente. Sem dúvidas, esse é o jogo de longo prazo mais inteligente da China — porque faz com que o Ocidente perca sua capacidade de competir no futuro.

O canal de talentos da China está absorvendo os pesquisadores de volta ao continente como parte do intitulado Plano Mil Talentos. A expansão crescente da BAT criou demanda por pessoas talentosas — a maioria delas estudou nos Estados Unidos e atualmente trabalha em universidades e empresas norte-americanas. Este programa do governo tem como alvo os principais tecnólogos e acadêmicos experientes, oferecendo-lhes uma espécie de bilhete da loteria premiado: proporcionar-lhes incentivos financeiros convincentes (tanto pessoais quanto projetos de pesquisa) e uma chance de ingressar em um ambiente de pesquisa e desenvolvimento imune às restrições regulamentais e administrativas comuns dos Estados Unidos. Mais de 7 mil pessoas foram aceitas no programa até agora e receberam uma bonificação especial do governo chinês: 1 milhão de yuans (cerca de US\$151 mil), um orçamento inicial de pesquisa pessoal de 3 a 5 milhões de yuans (US\$467 mil–US\$778 mil), auxílio moradia e educacional, auxílio refeições, reembolso de deslocamentos, assistência para ajudar os cônjuges a conseguir novos empregos e até mesmo viagens pagas para visitar sua casa.⁶⁹ Todos esses repatriados — de alguma forma, mesmo que seja a passos lentos — acabam usando seus talentos em nome da BAT.

*image
not
available*

havia 24,5 mil repositórios ativos.⁷⁵ Somando-se ao framework, o Google lançou produtos adicionais, como o TensorFlow-GAN (uma biblioteca para módulos de rede gerativa adversária) e o TensorFlow Object Detection API (que ajuda os desenvolvedores a projetar modelos mais precisos de aprendizado de máquina para visão computacional). Os TPUs já estão sendo usados nos data centers do Google — eles potencializam modelos de aprendizado profundo em todas as consultas de pesquisa da empresa.

Não é à toa que o Google tentou adquirir o GitHub, que é usado por 28 milhões de desenvolvedores em todo o mundo, e é uma plataforma importante para os Nove Titãs da IA. Mas em junho de 2018, o Google perdeu o lance — para quem vocês imaginam? — para a Microsoft.⁷⁶

O Facebook firmou uma parceria com a Intel com o intuito de desenvolver um chip IA para fins internos de pesquisa e desenvolvimento, que a empresa precisava a fim de impulsionar a eficiência para uma experimentação mais rápida. A Apple desenvolveu seu próprio chip “neural engine” para usar dentro de seu iPhone X, enquanto a Microsoft desenvolveu chips IA para seu headset de realidade híbrida HoloLens e para sua plataforma de computação em nuvem Azure. A BAT também está arquitetando seus próprios chips: em 2017, o Alibaba começou um recrutamento em massa no Vale do Silício para “arquitetos de chip IA”⁷⁷ e, em 2018, lançou seus próprios chips personalizados — o Ali-NPU — que estão disponíveis em sua nuvem pública para qualquer um usar.

Com o intuito de antecipar as necessidades de um futuro próximo que exige um melhor desempenho, há muitos anos, a IBM desenvolveu seu chip neuromórfico TrueNorth e já está progredindo em um novo tipo de hardware que pode fazer com que as redes neurais fiquem 100 vezes mais eficientes. Para contextualizar, seria como comparar um ábaco feito de paus e pedras com o transportador de Jornada nas Estrelas. O novo tipo de chip usa dois tipos de sinapses, um para memória de longo prazo e outro para cálculo de curto prazo.

Estamos falando de nossas preferências modernas como “Você usa Windows ou Mac?” temperadas com algumas doses de anabolizantes. A maioria desses chips opera em frameworks que os Nove Titãs da IA classificam como “open source” — ou seja, os desenvolvedores podem acessar, usar e aprimorar os frameworks sem pagar nada. Mas o hardware em si é uma exclusividade e os serviços vêm com taxas de assinatura. Na prática, isso significa que, depois que um aplicativo for desenvolvido para um framework, será extremamente difícil migrá-lo para outro local. Desta forma, as tribos da IA estão inscrevendo novos membros — e um rito de iniciação é beijar o anel de um framework da MÁFIA-G.

Em uma campanha para comercializar a IA, a MÁFIA-G está recrutando desenvolvedores de modos criativos. Em maio de 2018, o Google e a plataforma de aprendizado online Coursera lançaram uma especialização nova em aprendizado de máquina. Porém você tem que usar o TensorFlow. O curso de cinco partes, que inclui um certificado para graduados, é descrito como uma forma de qualquer pessoa estudar sobre aprendizado de máquina e redes neurais. Os alunos precisam de dados e estruturas do mundo real, portanto aprendem no framework do Google.

O hardware faz parte da estratégia de IA da MÁFIA-G, que também tem conexões com o governo, de modos diferentes dos que vimos na China, mas que devem ser igualmente preocupantes, ainda que você não seja cidadão norte-americano. Isso porque, nos Estados Unidos, a IA serve a três mestres: Capitol Hill, Wall Street e Vale do Silício. As pessoas que, de fato, elaboram as políticas e regulamentam o debate estão no Congresso, ou são funcionários federais de carreira que tendem a permanecer em seus empregos por décadas. Porém aqueles que definem as prioridades para esta política — nosso presidente e os chefes das grandes agências do governo (por exemplo, a Comissão de Comunicações Federais, o Departamento de Justiça e assim por diante) entram e saem do escritório em poucos anos. Não existe nenhum objetivo ou direcionamento nacional claro para a IA.

Somente há pouco que a China e seus planos para IA foram alvo de uma atenção mais acentuada — e isso se deve, sobretudo, porque o presidente Xi publicou um plano estratégico de longo prazo focado em IA e no uso de dados. Os norte-americanos têm algo chamado Comitê de Investimentos Estrangeiros nos Estados Unidos, ou CFIUS, que é um grupo bipartidário liderado pelo secretário do Tesouro e formado por membros dos departamentos do Tesouro, Justiça, Energia, Defesa, Comércio, Estado e Segurança Nacional. A função desse comitê é analisar e investigar negócios que possam colocar a segurança nacional em risco. Fora o CFIUS, que impediu a Broadcom, de Singapura, de adquirir a Qualcomm, fabricante de chips de San Diego. O CFIUS também rejeitou uma oferta de aquisição da MoneyGram, com sede em Dallas, pela empresa de pagamentos eletrônicos Ant Financial, controlada pelo Alibaba. À época em que este

*image
not
available*

de 2013: https://www.washingtonpost.com/business/economy/washington-post-closes-saletto-amazon-founder-jeff-bezos/2013/10/01/fca3b16a-2acf-11e3-97a3ff2758228523_story.html?noredirect=on&utm_term=.3d04830eab75.

33 Jason Lim, “WeChat Is Being Trialled To Make Hospitals More Efficient In China”, Forbes, 16 de junho de 2014: <https://www.forbes.com/sites/jlim/2014/06/16/wechat-is-being-trialed-to-make-hospitalsmore-efficient-in-china/#63a2dd3155e2>.

34 “Rise of China’s Big Tech in AI.”

35 Arjun Kharpal, “China’s Tencent Surpasses Facebook in Valuation a Day after Breaking \$500 Billion Barrier”, CNBC, 21 de novembro de 2017:

<https://www.cnbc.com/2017/11/21/tencent-surpasses-facebook-in-valuation.html>.

36 Sam Rutherford, “5 Things to Know About Tencent, the Chinese Internet Giant That’s Worth More than Facebook Now”, Gizmodo, 27 de novembro de 2017: <https://gizmodo.com/5-things-to-know-about-tencent-the-chinese-internet-gi-1820767339>.

37 Rebecca Fannin, “China Releases a Tech Dragon: The BAT”, Techonomy, 23 de maio de 2018:

<https://techonomy.com/2018/05/china-releases-tech-dragon-bat/>.

38 “Mobile Fact Sheet”, Pew Research Center, 5 de fevereiro de 2018: <http://www.pewinternet.org/fact-sheet/mobile/>.

39 Kaya Yurieff, “Amazon’s Cyber Monday Was Its Biggest Sales Day Ever”, CNN Money, 29 de novembro de 2017: <https://money.cnn.com/2017/11/29/technology/amazon-cyber-monday/index.html>.

40 Helen H. Wang, “Alibaba’s Singles’ Day by the Numbers: A record \$25 billion haul”, Forbes, 12 de novembro de 2017: <https://www.forbes.com/sites/helenwang/2017/11/12/alibabas-singles-day-by-the-numbers-a-record-25-billion-haul/#45dcfea1db15>.

41 Fannin, “China Releases a Tech Dragon”.

42 Michael Brown e Pavneet Singh, China’s Technology Transfer Strategy (Silicon Valley: Defense Innovation Unit Experimental, 2017): <https://new.reorg-research.com/data/documents/20170928/59ccf7de70c2f.pdf>.

43 Para ler o texto completo do 13º FYP, consulte República Popular da China, 13º Plano Quinquenal de Desenvolvimento Econômico e Social Nacional, 17 de março de 2016: http://www.gov.cn/xinwen/2016-03/17/content_5054992.htm.

44 J. P., “What Is China’s Belt and Road Initiative?”, Economist, 15 de maio de 2017:

<https://www.economist.com/the-economist-explains/2017/05/14/what-is-chinas-belt-and-road-initiative>.

45 Salvatore Babones, “China’s Middle Class Is Pulling Up the Ladder Behind Itself”, Foreign Policy, 1 de fevereiro de 2018: <https://foreignpolicy.com/2018/02/01/chinas-middle-class-is-pulling-up-the-ladder-behind-itself/>.

46 Pew Research Center, The American Middle Class Is Losing Ground (Washington, D.C.: Pew Research Center, dezembro de 2015): <http://www.pewsocialtrends.org/2015/12/09/the-american-middle-class-is-losing-ground/>.

47 Emmie Martin, “70% of Americans Consider Themselves Middle Class — But Only 50% Are”, CNBC, 30 de junho de 2017: <https://www.cnbc.com/2017/06/30/70-percent-of-americans-consider-themselves-middle-class-but-only-50-percent-are.html>.

48 Abha Bhattarai, “China Asked Marriott to Shut Down Its Website. The Company Complied”, Washington Post, 18 de janeiro de 2018: <https://www.washingtonpost.com/news/business/wp/2018/01/18/china-demanded-marriott-change-its-website-the-company-complied>.

49 Louis Jacobson, “Yes, Donald Trump Did Call Climate Change a Chinese Hoax”, PolitiFact, 3 de junho de 2016: <https://www.politifact.com/truth-o-meter/statements/2016/jun/03/hillary-clinton/yes-donald-trump-did-call-climate-change-chinese-h/>.

50 Michael Greenstone, “Four Years After Declaring War on Pollution, China Is Winning”, New York Times, 12 de março de 2018: <https://www.nytimes.com/2018/03/12/upshot/china-pollution-environment-longer-lives.html>.

51 Carl Gene Fordham, “20 Actually Useful Chengyu”, CarlGene.com (blog), 14 de agosto de 2008: <http://carlgene.com/blog/2010/07/20-actually-useful-chengyu>.

52 Stephen Chen, “China Takes Surveillance to New Heights with Flock of Robotic Doves, but Do They Come in Peace?”, South China

*image
not
available*

então, ele reproduziu a transcrição de uma conversa que ela e o marido tiveram sobre assoalhos de madeira.

Ao contrário da cobertura midiática e das teorias de conspiração que circulavam nas mídias sociais, a Amazon não estava registrando intencionalmente todas as conversas na casa de Danielle. Era uma falha. Mais tarde, a Amazon explicou que o dispositivo Echo de Danielle tinha sido acionado por causa de uma palavra na conversa — algo semelhante à “Alexa”, mas não era exatamente “Alexa”. Fora um problema decorrente da imperfeição intencional — nem todo mundo pronuncia “Alexa” com a mesma entonação e sotaque, logo, para que funcionasse, o dispositivo levava em conta tal variação. Depois, a IA detectou o que parecia ser um pedido abafado e desajeitado de “enviar mensagem”, e disse em voz alta “Para quem?” Todavia, Danielle e o marido não ouviram esta pergunta. O dispositivo interpretara a conversa em segundo plano como se fosse o nome do colega de trabalho, repetiu o nome e disse: “Certo?”, mais uma vez em voz alta, e de novo a partir do ruído de fundo chegou à dedução errada. Instantes depois, um arquivo de áudio atravessou o país. A Amazon afirmou que o incidente fora o resultado de uma cadeia infeliz de eventos, o que de fato aconteceu. Contudo, a razão pela qual a falha ocorreu em primeiro lugar — a imperfeição — é resultado da otimização.

O efeito de otimização revela que a IA se comportará de maneiras imprevisíveis, um dos objetivos dos pesquisadores, entretanto o uso de dados do mundo real pode ocasionar resultados nefastos. E isso evidencia nossas próprias limitações humanas. Um dos membros mais antigos dos Nove Titãs da IA — a Microsoft — aprendeu da pior forma o que acontece quando se prioriza o valor econômico da IA em detrimento dos valores tecnológicos e sociais. Em 2016, a Microsoft ainda não tinha sedimentado uma visão única de inteligência artificial e de como precisaria evoluir rumo ao futuro. Já estava dois anos atrás da Amazon, que lançara seu popular alto-falante inteligente e estava concentrando os esforços em desenvolvedores e em parceiros. O Google estava encabeçando as tecnologias de IA, que já haviam sido implementadas em produtos concorrentes, como mecanismo de pesquisa, e-mail e calendário. Os iPhones da Apple já disponibilizavam a Siri como item de série. A Microsoft havia, na verdade, lançado seu próprio assistente digital no início do ano — chamava-se Cortana —, mas o sistema não havia despertado o interesse dos usuários do Windows. Embora a Microsoft fosse a camada de produtividade indispensável — mesmo invisível — que nenhum negócio poderia operar sem, os executivos e os acionistas estavam uma pilha de nervos.

A Microsoft não ignorou a chegada da IA. Verdade seja dita, a empresa trabalhou, por mais de uma década, em muitas frentes: visão computacional, processamento de linguagem natural, compreensão de leitura de máquina, aplicativos de IA em nuvem do Azure e até mesmo em computação de ponta. O obstáculo era o desalinhamento da organização e a falta de uma visão compartilhada entre todas as equipes multifuncionais. Isso ocasionava rompantes de inovações fenomenais de inteligência artificial, artigos publicados e muitas patentes criadas por super-redes trabalhando em projetos individuais. Um exemplo é um projeto de pesquisa experimental que a Microsoft lançou em parceria com a Tencent e uma imitação chinesa do Twitter chamada Weibo.

*image
not
available*

com todas as letras EU APRENDI COM VOCÊ E VOCÊ TAMBÉM É IDIOTA.” Quanto mais pessoas interagiam com ela, mais Tay se descontrolava. Veja apenas algumas das conversas que ela teve com pessoas reais:

Ao se referir ao então presidente Obama, Tay escreveu: “@icbydt Bush fez o 11 de setembro e Hitler teria feito um trabalho melhor do que o macaco que temos agora. Donald Trump é a única esperança que temos.”

Sobre a hashtag Black Lives Matter, Tay afirmou: “@AlimonyMindset negões como @deray deveriam ser enforcados! #BlackLivesMatter.”

Tay simplesmente decidiu que o Holocausto fora inventado e tuitou: “@ brightonus33 Hitler tinha razão, eu odeio os judeus.” E ainda por cima continuou, tuitando para @ReynTheo, “HITLER NÃO FEZ NADA DE ERRADO!” e depois “GUERRA DE LIMPEZA ÉTNICA CONTRA OS JUDEUS AGORA” para @MacreadyKurt.²¹

Mas o que aconteceu? Como Xiaoice poderia ter sido amada e venerada na China, e depois se transformar em uma IA racista, antissemita, homofóbica e misógina nos Estados Unidos? Posteriormente, aconselhei a equipe que trabalha com inteligência artificial na Microsoft, e posso garantir que são pessoas bem-intencionadas e conscientes, elas ficaram tão surpresas quanto o resto de nós.

Parte do problema era uma vulnerabilidade no código. A equipe incluiu um comando do tipo “repita depois de mim”, um recurso difícil de entender que permitiu temporariamente que alguém colocasse palavras na boca de Tay antes de ela tuitar para o resto do mundo. Porém a razão pela qual Tay perdeu o controle tinha mais a ver com a equipe que a otimizou para o Twitter. Eles se basearam na própria experiência na China e na experiência pessoal limitada de cada um em redes de mídia social. A equipe não planejou cenários de risco, não levou em consideração um ecossistema maior, tampouco houve testes antes para analisar o que porventura aconteceria caso alguém ludibriasse Tay de propósito para que ela proferisse ofensas. A equipe também não considerou o fato de que o Twitter é um espaço gigantesco com milhões de seres humanos que interpretam valores de forma violentamente diferente e com milhões de bots projetados justamente para manipular as percepções das pessoas.

Sem demora, a Microsoft desativou Tay e apagou todos os seus tuítes. Peter Lee, chefe de pesquisa da Microsoft, publicou um post sincero e honesto pedindo desculpas pelos tuítes.²² Mas não havia como apagar o erro da empresa de IA da memória antes da conferência anual de desenvolvedores. A Microsoft não estava mais na vanguarda e nem lançando produtos em grandes eventos do setor, como a Consumer Electronics Show. Estava guardando tudo para seu próprio evento anual, alvo de atenção especial — sobretudo por parte dos membros do conselho e dos investidores. Nadella deveria subir ao palco e mostrar aos desenvolvedores um produto de inteligência artificial que os deixasse de queixo caído — e tranquilizasse os investidores no processo. A pressão para lançar Tay nos Estados Unidos rapidamente, antes da conferência, foi imensa. O resultado não colocou a vida de ninguém em risco, não infringiu a lei e a Microsoft se recuperou. Mas, como todas essas histórias — Latanya Sweeney e Google AdSense, DeepMind e os dados dos pacientes nos hospitais do Reino Unido, as duas garotas negras que foram tidas

como futuros criminosas —, ao otimizar as máquinas para objetivos de curto prazo, as tribos da IA sem querer prejudicaram a vida de muitas pessoas.

Valores Compartilhados pela Humanidade

Em ciência comportamental, e na teoria dos jogos, existe um conceito conhecido como “nudging”, ou arquitetura de escolha, que proporciona uma forma indireta de alcançar determinado comportamento e decisão desejados, como influenciar as pessoas a economizarem dinheiro para o plano privado de aposentadoria. O nudging é amplamente utilizado em todas as nossas experiências digitais, como o preenchimento automático em busca dos cardápios principais quando você procura restaurantes no Yelp. O objetivo é ajudar os usuários a sentirem que fizeram a escolha certa, seja lá o que tiverem escolhido, mas a consequência é que as pessoas estão aprendendo a viver com muito menos opções do que realmente existem no mundo real.

Por meio da mineração e do refinamento de nossos dados, dos sistemas e das técnicas empregadas para treinar os algoritmos de aprendizado de máquina e do efeito de otimização, os Nove Titãs da IA estão usando o nudging em grande escala. Ainda que dê a impressão de que pode escolher, o que você tem na verdade é uma ilusão. O nudging não muda apenas a nossa relação com a tecnologia — ele está transformando nossos valores de maneiras quase imperceptíveis. Caso use o sistema de mensagens de texto do Google, agora ele lhe oferece três opções de resposta automatizadas. Se um amigo lhe enviar um emoji com o polegar para cima, as três respostas propostas não são palavras, e sim um outro emoji. Caso um amigo escreva: “O que você achou do jantar?”, suas escolhas podem ser “bom”, “ótimo” e “incrível”, ainda que você nunca diga a palavra “incrível” em uma conversa e nenhuma dessas escolhas descrevam exatamente sua opinião. Mas o nudging também estimula o excesso: assistir horas de vídeo de uma vez, jogar muito videogame e verificar assiduamente as contas de mídia social. A otimização da IA significa empregar o processo de nudging aos humanos.

Em outros campos profissionais e técnicos, existe um conjunto de princípios norteadores que orientam o modo como as pessoas trabalham, e o nudging tende a violar a essência desses princípios. Na medicina, há o juramento de Hipócrates, que exige que os médicos jurem defender padrões éticos específicos. Os advogados estão vinculados às prerrogativas advogado-cliente e à confidencialidade, que protegem as conversas que as pessoas têm com os profissionais que as representam. Os jornalistas seguem muitos princípios norteadores, que incluem padrões como o uso de informações de fontes primárias e matérias de interesse público.

Neste exato momento, não há o incentivo de ninguém para considerar os custos inesperados de otimizar a IA na ausência da programação de princípios humanos. Uma equipe que atenda a seus conjuntos de critérios tem prioridade em relação à análise das possíveis consequências se suas contribuições para um sistema de inteligência artificial ou o próprio trabalho impactarem o futuro da humanidade. Como resultado, as tribos da IA, os Nove Titãs da IA e os países em que operam influenciam as tomadas de decisões. Isso estabelece um precedente perigoso, bem como estamos entregando de bandeja mais responsabilidade e controle aos sistemas de tomada de decisão. Atualmente, os Nove Titãs da IA não têm concessão de poderes a fim de desenvolver

ferramentas e técnicas para fazer com que seus sistemas de inteligência artificial sejam compreendidos por seus próprios criadores e pelos clientes que usam aplicativos comerciais de inteligência artificial — e não existem medidas em vigor que responsabilizem a IA por todos nós. Estamos atravessando um limiar de uma nova realidade em que a IA está gerando seus próprios programas, criando seus próprios algoritmos e fazendo escolhas sem os humanos estarem a par disso. Por ora, ninguém, em qualquer país, tem o direito de interrogar uma IA e analisar claramente como foi tomada uma decisão.

Se fôssemos desenvolver um “senso comum” para IA, o que isso significaria na prática, já que a humanidade não tem um conjunto compartilhado de valores? Muito da natureza humana já é difícil de explicar e isso varia de cultura para cultura. O que é de suma importância para alguns não é necessariamente importante para os outros. É fácil esquecer, mesmo em um lugar como os Estados Unidos, constituído de tantos idiomas e culturas diferentes, que não temos um conjunto excepcional de valores e ideias. Dentro de nossas comunidades, entre nossos vizinhos, em nossas mesquitas/sinagogas/igrejas — a diversidade é imensa.

Eu morei e trabalhei no Japão e na China durante muitos anos. As regras culturalmente aceitas são muito diferentes em cada país, principalmente em comparação às minhas experiências no meio-oeste dos Estados Unidos. Determinados valores são óbvios e explícitos. Por exemplo, no Japão, dicas não verbais e comunicação indireta são bem mais importantes do que falar o que lhe vier à mente ou demonstrar emoções fortes. Em um ambiente corporativo, dois funcionários nunca gritariam um com o outro, e nunca repreenderiam um subordinado publicamente. No Japão, o silêncio vale ouro. Em minha experiência, este não é o caso da China, em que a comunicação é muito mais direta e clara. (No entanto, não tão clara quanto, digamos, minhas tias e tios judeus que ficam felizes em me dizer, em detalhes lamuriosos, exatamente o que pensam.)

Aqui é onde as coisas ficariam realmente complexas para uma IA tentar interpretar o comportamento humano e automatizar uma resposta. Em ambos os países, os objetivos são os mesmos: as necessidades do grupo pesam mais do que os desejos de um indivíduo e, acima de tudo, a harmonia social deve prevalecer. Mas o processo a fim de alcançar esses objetivos é, na verdade, o oposto: na maioria das vezes, a comunicação indireta no Japão versus uma comunicação mais direta na China.

E as variações que são mais indefinidas e difíceis de explicar? No Japão — lugar em que a comunicação indireta é valorizada — é perfeitamente normal comentar o peso de alguém. Quando trabalhei em Tóquio, uma das minhas colegas de trabalho falou um dia que parecia que eu tinha ganhado uns quilinhos a mais. Espantada e constrangida, mudei de assunto e perguntei sobre uma reunião no final do dia. Ela insistiu: será que eu sabia que certos alimentos japoneses eram ricos em gordura, mesmo parecendo saudáveis? Eu estava frequentando uma academia? Ela não estava perguntando sobre o meu peso como intuito de me importunar. Pelo contrário, era um indício dos nossos laços de amizade. As perguntas embaraçosas a respeito de quanto eu pesava eram um sinal de que ela se importava com a minha saúde. No Ocidente, seria socialmente inaceitável falar para uma colega de trabalho: “Minha nossa, você está gorda! Você

engordou?” Nos Estados Unidos, somos tão sensibilizados culturalmente em relação ao peso que aprendemos a nunca perguntar a uma mulher se ela está grávida.

Não podemos abordar a criação de um sistema compartilhado de valores da IA da mesma forma que abordamos a elaboração do código de conduta de uma empresa ou as diretrizes para a regulamentação bancária. A razão é simples: nossos valores humanos costumam mudar em resposta à tecnologia e a outros fatores externos, como movimentos políticos e forças econômicas. Basta dar uma olhada neste poema de Alfred Lord Tennyson, que descreve o que a Inglaterra vitoriana valorizava em seus cidadãos:

O homem para o campo e a mulher para o lar;

Ele para a espada, ela para a agulha;

Homem de cabeça, mulher de coração;

Homem para mandar; mulher para obedecer;

O mais é desvario.

Nossas crenças arraigadas mudam constantemente. Em 2018, quando estava escrevendo este livro, tornou-se socialmente aceitável que os líderes nacionais vomitassem mensagens ofensivas e cheias de ódio entre si em mídias sociais, e que gurus políticos expusessem comentários polarizadores e bombásticos sobre vídeos, postagens de blog e até mesmo em publicações de notícias tradicionais. É quase impossível imaginar a discricionariedade e o respeito pela privacidade durante a presidência de Roosevelt, quando a imprensa teve grande cuidado em nunca mencionar ou mostrar sua paralisia.

Como a IA não está sendo ensinada a tomar decisões perfeitas, e sim a otimizar, nossa resposta às forças mutáveis na sociedade é imprescindível. Nossos valores não são imutáveis. Isso faz com que os problemas de valores da IA sejam alarmantes. Construir a IA significa prever os valores do futuro. Nossos valores não são estagnados. Então, como ensinamos as máquinas a refletir nossos valores sem influenciá-las?

Otimizando a IA para Seres Humanos

Alguns membros da tribo da IA acreditam que um conjunto compartilhado de princípios norteadores é um objetivo digno e que a melhor maneira de alcançá-lo é alimentar os sistemas de inteligência artificial com literatura, notícias, artigos de opinião e editoriais de fontes de notícias verossímeis para ajudá-los a aprender sobre nós. Isso envolveria o crowdsourcing, em que a IA aprenderia com a sabedoria coletiva das pessoas. No entanto, essa é uma péssima abordagem, porque ofereceria ao sistema apenas um snapshot no tempo, e a curadoria do que as relíquias culturais incluiriam não poderia, de forma significativa, representar a soma total da condição humana. Caso já tenha feito uma cápsula do tempo, você saberá de pronto o porquê. As decisões que você tomou sobre o que incluir provavelmente não são as mesmas que tomaria hoje, levando em consideração a sua visão a posteriori.

As regras — o algoritmo — a partir das quais toda cultura, sociedade e nação vive e já viveu, sempre foram criadas por pouquíssimas pessoas. Democracia, comunismo, socialismo, religião, veganismo, nativismo, colonialismo são construtos que desenvolvemos ao longo da história para ajudar a orientar nossas decisões. Mesmo na melhor das hipóteses, eles não são à prova do futuro. As forças tecnológicas, sociais e econômicas sempre intervêm e nos fazem adaptar. Os Dez Mandamentos estabelecem um algoritmo destinado a criar uma sociedade melhor para os humanos vivos há mais de 5 mil anos. Um dos mandamentos é tirar um dia inteiro de descanso por semana e não trabalhar naquele dia. Em nossa era moderna, a maioria das pessoas não trabalha exatamente nos mesmos dias ou horas por semana, então seria impossível não infringir esta regra. Como resultado, as pessoas que seguem os Dez Mandamentos como um princípio norteador são flexíveis em sua interpretação, em virtude das realidades de dias úteis mais longos, treino de futebol e e-mail. Tudo bem se adaptar — funciona muito bem para nós e para as nossas sociedades, nos mantêm no rumo certo. Concordar com um conjunto básico de diretrizes nos possibilita otimizar para nós mesmos.

Não existe como criar um conjunto de mandamentos para IA. Não poderíamos elaborar todas as regras para otimizar corretamente a humanidade, e isso porque, ainda que as máquinas pensantes possam ser rápidas e poderosas, elas não são flexíveis. Não existe uma maneira fácil de simular exceções ou tentar pensar em cada possibilidade antecipadamente. Quaisquer que sejam as regras elaboradas, sempre existe uma circunstância no futuro em que algumas pessoas poderiam querer interpretá-las de forma diferente, ou ignorá-las por completo, ou criar adaptações para administrar uma circunstância imprevista.

Sabendo que não podemos formular um conjunto de mandamentos rígidos a seguir, devemos, em vez disso, focar nossa atenção nos humanos que desenvolvem esses sistemas? Essas pessoas — as tribos da IA — deveriam se fazer perguntas desconfortáveis, começando com:

- Qual é a nossa motivação para a IA? Ela está alinhada com os interesses da humanidade em longo prazo?

- Quais são nossos preconceitos? Quais ideias, experiências e valores deixamos de incluir em nossa tribo? Quem estamos ignorando?
- Incluímos pessoas diferentes de nós com o objetivo de melhorar o futuro da IA — ou simplesmente incluímos a diversidade em nossa equipe para atender a determinadas cotas?
- Como podemos garantir que nosso comportamento seja inclusivo?
- Como as implicações tecnológicas, econômicas e sociais da IA são entendidas pelas pessoas envolvidas em sua criação?
- Quais direitos fundamentais devemos estabelecer para interrogar os conjuntos de dados, algoritmos e processos que são utilizados para tomar decisões em nosso nome?
- Quem consegue definir o valor da vida humana? Este valor está sendo ponderado contra o que?
- Quando e por que as pessoas das tribos da IA sentem que é de sua responsabilidade abordar as implicações sociais da IA?
- A liderança de nossa organização e nossas tribos de IA refletem muitos tipos diferentes de pessoas?
- Qual o papel daqueles que comercializam a inteligência artificial ao abordar as implicações sociais da IA?
- Devemos continuar a comparar a IA ao pensamento humano ou é melhor para nós categorizá-la como algo diferente?
- Existe algum problema em arquitetar uma IA que reconheça e responda à emoção humana?
- Existe algum problema em fazer com que os sistemas de inteligência artificial sejam capazes de simular a emoção humana, sobretudo se estiverem aprendendo conosco em tempo real?
- Até que ponto é aceitável que a IA evolua sem seres humanos diretamente na jogada?

- Em quais circunstâncias uma IA poderia simular e experimentar as emoções humanas? E quanto à dor, à perda e à solidão? Qual a implicação de sermos responsáveis por esse sofrimento?
- Estamos desenvolvendo a IA com o intuito de buscar uma compreensão mais profunda de nós mesmos? Podemos usá-la para ajudar a humanidade a ter uma vida meditativa?

A MÁFIA-G começou a abordar o problema dos princípios norteadores por meio de inúmeros grupos de pesquisa e estudo. Na Microsoft, há uma equipe chamada FATE — acrônimo em inglês de Equidade [Fairness], Responsabilidade [Accountability], Transparência [Transparency] e Ética [Ethics] em IA.²³ No rastro do escândalo da Cambridge Analytica, o Facebook montou uma equipe de ética que estava desenvolvendo software a fim de assegurar que seus sistemas de inteligência artificial prevenissem os preconceitos. (Em particular, o Facebook não foi tão longe a ponto de criar um comitê de ética focado em IA.) A DeepMind organizou uma equipe de ética e sociedade. A IBM publica regularmente sobre ética e inteligência artificial. Na sequência de um escândalo no Baidu — o mecanismo de busca priorizou falsas alegações médicas de um hospital administrado pelo exército em que um tratamento provocou a morte de um estudante de 21 anos —, o CEO, Robin Li, admitiu que os funcionários haviam feito concessões em prol do crescimento lucrativo do Baidu e prometeu focar a ética no futuro.²⁴ Os Nove Titãs da IA elaboram estudos de ética e informes técnicos, convocam especialistas e organizam mesas redondas para discutir ética —, mas essas iniciativas não estão conectadas o suficiente com as operações cotidianas das inúmeras equipes que trabalham com IA.

Os sistemas de inteligência artificial dos Nove Titãs da IA estão cada vez mais acessando nossos dados cotidianos para criar produtos que agreguem valor comercial. Os ciclos de desenvolvimento estão acelerando a fim de acompanhar o ritmo das expectativas dos investidores. Estamos dispostos — ainda que inconscientes — a participar de um futuro que está sendo arquitetado às pressas e sem responder primeiro a todas essas perguntas. À medida que os sistemas de IA evoluem, e cada vez mais se automatiza nossa vida cotidiana, menos controle temos de fato sobre as decisões tomadas a respeito de nós e em nosso nome.

Isso, em contrapartida, aumenta o impacto no futuro de muitas outras tecnologias limítrofes ou que se cruzam de forma mais direta com a IA: veículos autônomos, CRISPR e edição genômica, medicina de precisão, robótica doméstica, diagnósticos médicos automatizados, tecnologias verdes e geoengenharia, viagens espaciais, criptomoedas e blockchain, fazendas inteligentes e tecnologias agrícolas, Internet das Coisas, fábricas autônomas, algoritmos para o mercado de ações, mecanismos de busca, reconhecimento facial e de voz, tecnologias bancárias, detecção de fraude e risco, policiamento e tecnologias judiciárias...Eu poderia fazer uma lista com dezenas de páginas. Não existe nenhum aspecto da sua vida pessoal ou profissional que não seja afetado pela IA. E se, nessa pressa de disponibilizar produtos no mercado ou de agradar a determinados funcionários do governo, seus valores não forem espelhados não somente pela IA,