

INTERNET

Christopher Kark

Stanford University

A Internet está em todo o lado, mas em lado algum em particular. As quintas de servidores que pontuam o Dakota do Norte, a tortuosa rede de cabos de fibra óptica acoplada ao fundo do oceano, as torres de comunicações celulares mal disfarçadas de sequoias: tudo isto forma uma vasta rede que, para a maioria de nós, é informe. O espaço, no sentido geográfico do termo, é uma noção obsoleta. Deixou de constituir novidade que a Internet colapse distâncias geográficas. Hoje, passamos o dia a deslizar o dedo para a direita. Hoje, importa apenas analisar dados, independentemente da sua origem. Compreender como é que tais dados são recolhidos e disseminados, como manipulá-los em proveito próprio, é dominar um poder em rápido desaparecimento. E no entanto, de uma maneira mais

Escusado será dizer que existe uma abundância de informação ao longo dos tempos, aparecendo gravada em cuneiforme nas paredes de templos, em cestos de papiros empilhados nos armazéns que flanqueavam as docas de Alexandria, abundância essa nivelada ora por fluxos frenéticos num dado século, ora por perdas catastróficas no século seguinte. Logo no quinto século A.C., Platão refere-se ao rei egípcio Âmon, que repreendeu Teuth (Tot, Hermes) pela invenção da escrita, visto que a escrita dificultaria a capa-

nua e crua do que nunca, dependemos de dados para compreender o mundo e a nós mesmos. Já não nos podemos fiar apenas na nossa agudeza para interpretar a quantidade de informação com que somos bombardeados; é simplesmente demasiada para a conseguirmos processar. Se podemos falar de proximidade na Internet, é acerca do controlo sobre a informação — o modo como dados são recolhidos, a sua distribuição, o seu uso. E embora possuamos mais dados do que alguma vez aconteceu, eles parece sempre escapar à nossa compreensão. (Zeynep Tufeski, «[Engineering the Public: Big Data, Surveillance and Computational Politics](#)», *First Monday*, 7 de Julho, 2014; consultado em 5 de Fevereiro de 2015.)

cidade da memória e, por conseguinte, destruiria a sabedoria (Ryan Szpiech, «[The Dagger of Faith in the Digital Age](#)», *Tablet*. 7 de Outubro, 2014; consultado a 10 de Outubro de 2014). Mas, precisamente por nos permitir fixar os nossos pensamentos num *medium* que nos sobreviverá, a escrita provou ser uma das mais poderosas tecnologias alguma vez concebida. Com efeito, até à invenção de gravações sonoras em meados do século dezanove, em França, a escrita sempre foi a única maneira de registar as línguas naturais. Enquanto a

quantidade de registos escritos (e, agora, audiovisuais) se multiplicou, o mesmo não acontece com a nossa capacidade para processá-la.

Queixas deste género acerca de um excesso de informação são parecidas com ansiedades articuladas em momentos tão remotos quanto o Eclesiastes 12:12, ou os escritos de Séneca, no primeiro século A.D. (Ann Blair, «[Information Overload, Then and Now](#)», *Chronicle of Higher Education*, 28 de Novembro, 2010; consultado a 30 de Dezembro, 2014). Por volta da mesma altura, Plínio, o velho, dedicou-se a compilar milhares de factos na sua *História Natural* (77 A.D.). Séculos depois, os eruditos medievais lamuriavam-se habitualmente da quantidade de coisas que havia para ler e de quão fácil era esquecê-las. A invenção da tipografia em meados do século quinze substituiu por textos impressos os tomos robustos das eras precedentes. A impressão era não apenas barata; era, além disso, célere o suficiente para responder às necessidades de populações gradualmente mais letradas. Êxitos literários anteriormente transcritos à mão começaram a voar das tipografias, ao mesmo tempo que novos géneros literários, como o romance, tornaram a leitura em privado num modo de passar o tempo. Passou a haver o excesso de informação tal como o entendemos, assim como esta avalanche de conteúdos que transborda das nossas caixas de correio e das nossas cronologias (*timelines*).

A velocidade e o volume são apenas duas das causas do excesso de informação. Não menos crucial é o modo como essa informação se organiza. Na Europa medieval, o esforço por preservar obras clássicas para a posteridade suscitava a questão de qual o melhor modo de as arrumar. Os monges em mosteiros espalhados pela Europa, assim como os estudiosos em Bagdade começaram a catalogar as suas colecções mais de mil anos antes de a palavra Google se tornar um verbo (*ibid*). Ajustando maneiras de organizar resmas de impressões, cultivavam a fastidiosa arte da indexação, que evoluiu para os catálogos

de fichas que vão silenciosamente desaparecendo das bibliotecas do presente. Na maioria dos lugares, foram substituídas por ferramentas de pesquisa sofisticadas e por dúzias de bases de dados por subscrição. O motor de pesquisa moderno é uma adaptação do catálogo de fichas à esfera digital, com a diferença de que também apresenta como resultados todo o género de informações reconhecíveis por algoritmos de pesquisa. Aquilo que o utilizador médio entende por Internet é, então, um nexos de dados combinados, formando um todo em constante expansão. O resultado é uma abundância de conteúdos que ninguém espera ser capaz de abarcar no tempo de uma vida.

Apequenando até as maiores bibliotecas da Antiguidade, a quantidade de tráfego na Internet totaliza hoje mais do que um zettabyte (1021 bytes), quantidade palpável pela multiplicação de conteúdos — ou de *media* preparados para um consumo massivo — através de páginas, plataformas de redes sociais e bases de dados. Vivemos numa era de «conteúdos», a era dos *media* auto-publicados, aflorando todo o tipo de tópicos e dirigidos a todos os géneros de pessoas. A quantidade de conteúdos é muito superior ao tempo disponível para os consumir, quanto mais para determinar o seu valor. Uma pesquisa sobre o esgotamento de uma celebridade apresentará dúzias e dúzias de resultados, desde jornais de referência a blogs com públicos pequenos, mas fiéis. Um *hashtag* como «#JeSuisCharlie» colocado no Twitter ou no Instagram agregará milhares de *snippets* a respeito de uma notícia ainda em desenvolvimento. Acumulam-se fotografias, multiplicam-se vídeos, e o que começa por ser um *post* isolado forma uma bola de neve de conteúdos, combinados entre si por links e *hashtags*. A sobrecarga de informação deixou de ser a némesis dos estudiosos; tornou-se uma artéria pulsante do nosso Zeitgeist.

Por que são as coisas assim? Na era digital, de que modo alterou a tecnologia as nossas atitudes relativamente à informação — à maneira como

ela é reunida, às partes merecedoras de atenção, às partes que não a merecem? Qual a medida de cada um para a sobrecarga de informação? Com a crescente quantidade de informações pessoais disponíveis online, até que ponto poderemos gerir as nossas reputações sem significativas despesas? Voltando um pouco atrás, como é que, antes de mais, nos tornámos responsabilizáveis por toda a informação publicada online a nosso próprio respeito? Partindo do princípio de que tais responsabilidades realmente nos assistem, que instrumentos existem para nos ajudar a transformar dados em informações, e informações em conhecimento? Demasiado vastas para que as possamos tratar exaustivamente, estas questões encontraram soluções em plataformas de Big Data e em serviços gestão de reputação.

A Internet contornou os habituais garantes da qualidade dos materiais que se publica. As ideias, quer as iluminadoras como as incendiárias, já não deparam semelhantes entraves à publicação. Qualquer pessoa com um telemóvel e alguma coisa a dizer pode disseminar os seus pensamentos mundo fora, juntando-se aos milhões de vozes que reverberam no ciberespaço. A diminuída importância daqueles garantes coincidiu com a preponderância das redes sociais, a deflagração de uma economia de partilhas, e a voga da produção de *media* a par do seu consumo. Daí que se verifique uma irrupção de conteúdos, que vão de *memes* virais a exemplos premiados de jornalismo civil. A autoridade fragmentou-se, dando lugar a uma situação em que as críticas de clientes e as demonstrações de amadores no YouTube gozam de tanto prestígio quanto as opiniões de especialistas.

Na era que precedeu a Internet também se verifica uma fragmentação da autoridade, graças à proliferação de canais de televisão por cabo. Mas era inimaginável que alguém pudesse transmitir para o mundo inteiro as suas ideias instantaneamente — um sonho longínquo, mesmo para

Mas estas soluções não são respostas: uma solução precisa de dar conta da sua causa. Muitas vezes, fazê-lo não é sequer do interesse dessas plataformas e serviços. Por contraste, ainda que tais causas sejam esquivas e difíceis de quantificar, se lhes dermos a devida atenção, iluminam mudanças de costumes a nível social e cultural (Giorgio Agamben, «[For a Theory of Destituent Power](#)», *Chronos Mag*, November 16, 2013, consultado a 3 de Janeiro de 2015). Embora a sobrecarga de informação não seja uma novidade, a responsabilidade de gerirmos quantidades de informação cada vez maiores nunca pesou tanto como nos nossos dias. Ela nasce do impacto da auto-publicação digital nas autoridades tradicionais.

aqueles que estavam por dentro da inovação tecnológica. Durante séculos, a distribuição de conteúdos limitava-se a quem tivesse livre acesso a um meio de comunicação. Para começar, apenas os letrados eram capazes de produzir conteúdos; eram poucos aqueles que viam o seu trabalho publicado em larga escala. À medida que a publicação impressa se foi espalhando, as tipografias e os jornais desenvolveram um conjunto de práticas para garantir a qualidade, avaliando o(s) autor(es) e o conteúdo em si mesmo. Apareceu a figura editorial do revisor. Este sistema de verificações continua a ter lugar: se um chefe de redacção considera uma coisa inadequada, não se publica antes de serem feitas as devidas revisões. Analogamente, para assegurar que vão ao encontro das expectativas do seu público, os filmes e os programas televisivos são submetidos a meses de escrutínio antes que irem para o ar.

Este género de procedimentos costumava garantir um nível de qualidade na maior parte dos materiais disponíveis para distribuição em massa. Por este mesmo motivo, muitos livros e filmes nunca viram a luz do dia, visto que exprimiam ideias controversas ou subversivas.

As obras do filósofo francês René Descartes foram censuradas na Europa durante o século dezassete pela mesma razão que a maior parte das redes sociais se encontram hoje banidas na China, na Coreia do Norte e no Irão: há ideias poderosas, não menos poderosas que o potencial para chegar a milhões de pessoas tecendo meia dúzia de coisas. Em todo o caso, a proibição absoluta não é a regra, mas a excepção. Para algumas coisas, onde quer que a norma seja a liberdade de expressão, a maioria das pessoas continua a virar-se para as fontes tradicionais de autoridade. Para outros efeitos, ajudam-se umas às outras. Ao contrário do que sugerem os estereótipos vigentes, os adolescentes preferem dar ouvidos aos conselhos de médicos em vez de confiarem nas recomendações de websites de saúde bem reputados (Urs Gasser, Sandra Cortesi, Momin Malik, and Ashley Lee, «[Youth and Digital Media: From Credibility to Information Quality](#)», Berkman Center for Internet & Society, 2012, consultado a 20 de Novembro de 2014). Já no que respeita a notícias, as pessoas inclinam-se a procurar publicações que dizem aquilo que pretendem ouvir. Entretanto, erigimos toda uma economia em torno de críticas (e críticas de críticas) de estranhos na Internet, consultando o Yelp para escolher uma *taquería*, ou o Angie's List para encontrar um carpinteiro especializado em soalhos de bambú (Andrew Flanagin and Miriam Metzler, «Digital Media and Youth: Unparalleled Opportunity and Unprecedented Responsibility», *Digital Media, Youth, and Credibility*, ed. Metzler e Flanagin, Cambridge, MA: MIT University Press, 2008, 5-28, 11).

Passámos a confiar em dois modos de autoridade: aquele que filtra a informação hierarquicamente, e aquele que colhe opiniões a partir de uma floresta de fontes não verificadas. Por muito que se alegue que os antigos guardiães se encontram em vias de extinção, a verdade é a de que são tão reverenciados como o [Esquadrão de Elite do Yelp](#). Daí a prosperidade de jornais de referência como o *New York Times*, assim como

a crescente procura de programas sindicados em páginas de entretenimento por subscrição à la Hulu. Aquilo que mudou foi o valor atribuído à auto-suficiência. As exigências do público quanto a uma autonomia de processos conduziram a um crescimento da automação, levando a uma diminuição do tamanho das equipas de apoio ao cliente em vários sectores económicos. Quando chega o momento de marcar um voo, em vez de consultarem agentes de viagem, ou de contactarem directamente com uma companhia aérea, os consumidores viram-se agora para os motores de meta-pesquisa. Em vez de consultar um corretor de bolsa, o investidor ocasional poderá consultar os blogues de alguns gurus financeiros antes de comprometer o seu plano de reforma. Com a maior autonomia o público vê-se mais pressionado no que respeita a procurar informação por si mesmo, tarefa dificultada pela falta de orientações físicas online (David Lankes, «Trusting the Internet: New Approaches to Credibility Tools», *Digital Media, Youth, and Credibility*, ed. Miriam Metzler and Andrew Flanagin, Cambridge, MA: MIT University Press, 2008, 101-21, 104).

Deste modo, banimos os agentes e os especialistas que intermediavam entre as nossas necessidades e os recursos disponíveis para as satisfazer. Em seu lugar, existem agora plataformas que interagem directamente com esses recursos, embora sem o conhecimento circunstanciado de um profissional. Vistos como revendedores gananciosos, esses profissionais são cada vez mais raros, enquanto as críticas de consumidores — cautelosamente ajustadas com algoritmos patenteados — são feitas passar por uma representação da verdade. Consequentemente, as pessoas passaram a fiar-se na opinião de estranhos a uma escala que não conhece precedentes. Verdade seja dita que podemos agarrar-nos lealdade ao *briefing* noticioso diário da *NYT Now*, mas é muito provável que esquadrinheemos centenas de críticas de consumidores na Amazon ou o encómio na C-Net antes de desbaratarmos dinheiro numa Nikon Df.

Um produto ser avaliado em quatro estrelas e meia é frequentemente motivo para torcer o nariz a uma compra, ainda que não façamos a menor ideia de quão credíveis são as pessoas que o criticaram. Eis o fardo da escolha. A autoridade tradicional continua a ter muito peso — neste caso, em publicidade —, mas é provável percorrermos um sem-número de críticas antes de chegarmos a uma decisão (Ibid., 107). No processo, espera-se que façamos uma data de juízos sobre a fiabilidade das opiniões com que nos depara-

mos, aceitando algumas, descartando outras, e ignorando aquelas que não queremos ouvir. Por vezes, assoberbados numa abundância de informação, viramos simplesmente as costas à pesquisa. Porém, a vontade de solicitar múltiplas fontes de informação para chegar a uma decisão, ou para ficar a saber o que se passa no mundo, parece ter chegado para ficar, especialmente num contexto em que as fontes tradicionais de autoridade estão a enfraquecer.

A auto-suficiência acrescida a respeito da informação acarreta também a responsabilidade do controlo da informação disponível a respeito de si mesmo. Aquilo que se escolhe ver ou ler é em larga medida um problema nosso; aquilo que os outros escolhem ver ou ler on-line a nosso respeito, em larga medida, não o é. Naturalmente, a reputação sempre foi um factor das relações humanas. Pelo menos metade das peças encenadas em Espanha durante o Século de Ouro eram acerca de fidalgos santimoniais numa demanda belicosa para vingar a sua honra. A dichote ainda popular nos Estados Unidos, «*pistols at dawn*» («pistolas à alvorada»), evoca uma cultura sulista antebellum, absurdamente cortês. Hoje, no entanto, em nada é a reputação tão determinante como no que respeita a preocupações com a privacidade online, e, a par disso, com os sistemas de avaliação opacos que servem de esteio à economia do *on-demand*. Os blogues de aconselhamento de carreira advertem os estudantes e candidatos a empregos por igual a repensar aquilo que publicam on-line, já que cada *post* publicado corresponde a uma pegada num trilho digital que pode estender-se por uma série de anos. Esse conteúdo pode permanecer ignorado durante anos até que nele tropece um comité de admissões académicas ou um encarregado de recursos humanos. As consequências dessa descoberta podem ser céleres, e o seu impacto devastador.

A ansiedade a respeito da privacidade parece chocar com o *ethos* da transparência característico da economia do *on-demand*, nos termos do qual cada um de nós é uma marca e a avaliação por estrelas uma rubrica de qualidade. O que está em questão não é apenas o modo como deveríamos definir a privacidade on-line, debate que, em todo o caso, estará sempre um passo à frente do processo legislativo. A privacidade pressupõe o direito a esconder uma parte de nós, senão o todo, do olhar público, deixando ver apenas aquilo que nos convém.¹ Em jogo está também uma distinção entre a pessoa e a *persona* digital, distinção que se vai esbatendo irremediavelmente à medida que os dados se vão tornando a medida do homem (Farhad Manjoo, «[Uber's Business Model Could Change the Way You Work](#)», *New York Times*, 28 de Janeiro de 2015, consultado a 16 de Fevereiro de 2015). Nos anos que precederam a Internet, havia pouca informação publicada a respeito de cada um. Do pouco que havia, a maior parte era factual (e.g., certidões de nascimento, registos bancários, obituários), e a maioria era verificável junto de fontes fiáveis. Com brevidade, poderíamos esgotar tudo aquilo que o público poderia saber a nosso respeito.

Esse tempo acabou. Tal como acontecia com as celebridades antes do aparecimento da Internet, a nossa reputação depende hoje daquilo que é dito a nosso respeito on-line (David Streitfeld, «[Ratings Now Cut Both Ways, So Don't](#)

Sass Your Uber Driver», *New York Times*, 30 de Janeiro, 2015 (consultado a 31 de Janeiro, 2015). Na ausência de barreiras à auto-publicação, as vogas e os boatos passaram a competir livremente com os factos. Com efeito, se alguma coisa há a aprender com o movimento do Tea Party é que os próprios factos passaram a estar sob suspeita. Obviamente, as pessoas sempre se viram vítimas de rumores sórdidos perdidos no tempo, mas esses rumores nunca tinham estado para serem partilhados pelo mundo e armazenados em cache pelos motores de pesquisa. Anteriormente, a partir do momento que são transmitidos boca a boca, distorcidos pela reinterpretação e truncados pela memória, os rumores evaporar-se-iam assim que todos se cansassem deles. Por comparação com as difamações digitais de hoje restaria pouco que ventilar (Daniel Solove, «Speech, Privacy, and Reputation on the Internet», *The Offensive Internet: Speech, Privacy, and Reputation*, ed. Saul Levmore and Martha Nussbaum. Cambridge, MA: Harvard University Press, 2010, 15-30, 16). Ainda que novos rumores inevitavelmente aparecessem, faltar-lhes-ia a permanência do que aparece quando se pesquisa pelo nosso nome no Google.

Ao mesmo tempo que aumentaram as exigências quanto à protecção da nossa privacidade online, também aumentou, inexoravelmente, o fardo de lidar com as informações que a constituem. Poderíamos passar horas—ou mesmo dias—a polir os nossos perfis em redes sociais, transformando-os em avatares idealizados de nós mesmos, a desvincularmo-nos de tags de imagens de momentos que nos arrependemos de publicar (ainda que talvez não nos arrependamos de os ter vivido), maximizar as definições de privacidade das nossas contas, ou apagar longos comentários publicados num blogue de que nos esquecêramos desde o grito que acabou com a carreira política de Howard Dean. Este género de auto-modelagem lembra bastante os maneirismos da corte em torno de príncipes renascentistas, onde a dissimulação, em vez de factos verificáveis, era

a melhor defesa contra ataques à reputação. No presente, existem firmas para gerir a reputação, que podemos contratar para fazer todo o trabalho aborrecido. Como se nos tratássemos de uma empresa desacreditada em busca de uma reviravolta, podemos pagar (um preço nada barato) para que alguém elimine as nossas pegadas digitais de resultados de pesquisa, ou pelo menos que as relegate para páginas onde ninguém se dá ao trabalho de procurar (Evgeny Morozov, «Two Decades of the Web», *Prospect Magazine*, 22 de Junho, 2011, consultado a 10 de Novembro, 2014).

A reputação também se tornou sinónimo de qualidade na economia *on-demand*, encontrando-se codificada nas apps de serviços — partilha de veículos para regressos a casa, *couchsurfing* como alternativa barata aos hotéis — e afiançada por um intrincado sistema de avaliações. A carreira de um condutor de veículo partilhado depende destas avaliações, tal como a do dono de um quarto (ou de um sofá) alugado a hóspedes. Para salvaguardar as suas fontes de rendimentos, condutores e anfitriões devem esforçar-se por manter avaliações altas a todo o custo, já que as suas carreiras se medem pelos pesos da balança da agregação de críticas de clientes. Por subjectivas e inferificáveis que sejam, este tipo de avaliação está pronto a expandir-se para lá da economia *on-demand* para a esfera das escolas, dos hospitais, e, eventualmente, dos governos, o *ne plus ultra* da datificação. Faz por isso muito pouco sentido falar sobre a criação de uma imagem online como se tal fosse de alguma maneira uma coisa distinta daquilo que somos. Na medida em que a nossa *persona* é objecto deste ou daquele género de avaliações, a fronteira que separa o que somos online e a o que somos na vida real está a desaparecer muito rapidamente. A reputação deste eu convergente reside em quanto da informação a seu respeito podemos directa ou indirectamente controlar. No meio desta confusão de dados, manter uma boa reputação revela-se uma tarefa tão penosa quanto a de distinguir conteúdos patrocinados de jornalismo genuíno.

Sob o excesso de conteúdos, os dados têm-se revelado de um valor inestimável para aqueles que têm ao seu dispor recursos para analisá-los. Há um século, a recolha de dados era uma tarefa manual. Uma das suas primeiras manifestações foram as sondagens de opinião. As sondagens cresceram em importância nos anos que se seguiram à Primeira Guerra Mundial, começando pela ocasião em que o *Literary Digest* distribuiu postais pelos seus leitores, pedindo-lhes que dissesse quem lhes parecia que ia ganhar as eleições presidenciais americanas. Com a excepção notável de 1936, os postais devolvidos previram com precisão o resultado de todas as eleições entre 1920 e 1932, dando conta de um interesse em medir a opinião pública objectivamente — interesse ecoado noutros países com governos representativos. Não é coincidência que a Gallup, a firma de consultoria cujo nome é hoje sinónimo de sondagens de opinião pública, tenha sido criada em 1935.

A informação recolhida através da Gallup não tinha então rival, mas empalidece por comparação com as bases de dados de sondagens realizadas em websites e, cada vez mais, a partir de equipamentos que nos chateiam para irmos correr, ou que nos lembram que passámos tempo demais na Sur La Table. Sabendo que estes dados são demasiado abundantes para que possam ser tratados manualmente, os engenheiros desenvolveram procedimentos sofisticados para fazer emergir padrões desses dados, tanto ao serviço das pessoas como do lucro. Tais procedimentos, que lembram ladainhas de orações condicionais, são vulgarmente conhecidos como «algoritmos». Os algoritmos vão buscar dados específicos a uma base de dados seguindo uma lista de computações bem-definidas. Ainda que existam desde que o matemático persa Mu ammad ibn M s al-Khw rizm os descreveu logo no século nove d.C., os algoritmos tornaram-se relevantes nos anos mais recentes com o advento dos motores de pesquisa e das plataformas de *Big Data*. O mesmo se aplica aos colossais conjuntos

de dados extraídos dos percursos dos nossos cliques e dos nossos dados médicos, uma transformação momentosa na maneira como analisamos o comportamento humano.

Os algoritmos são invisíveis para a ampla maioria das pessoas que os usam. Paradoxalmente, esta é uma das principais razões pelas quais confiamos neles: precisamos de maneiras de pescar informação de que precisamos sem precisarmos de drenar toneladas de informações de que não precisamos. A utilidade dos dados reside não meramente na sua recolha mas na possibilidade de extrairmos deles informação significativa. Sendo constantemente actualizados pelos seus proprietários, e programados para aprender a partir de informações introduzidas por utilizadores, os algoritmos do presente alimentam a investigação e a tomada de decisão de organizações a uma escala que não conhece precedentes. Os consumidores cruzam-se com eles em anúncios publicados por browsers, nos filmes «recomendados» para a sua lista do Netflix ou em artigos sugeridos para o seu carrinho de compras da Amazon. Programando com frases o envio de email indesejado para a sua caixa de *spam*, os utilizadores do Gmail podem até ensinar os algoritmos a distinguir entre o que é *spam* e o que são mensagens relevantes (Evgeny Morozov, «[The Rise of Data and Death of Politics](#)», *The Guardian*, 19 de Julho, 2014, consultado a 10 de Novembro, 2014).

Ao mesmo tempo, os algoritmos também alimentam o comércio e a governação. Baseando-se na sua vasta recolha de dados de consumidores, o Walmart soube abastecer-se de *junk food* e de lanternas numa região afectada por furacões muito antes de essa região ter sido assolada por uma tempestade destruidora (Michiko Kakutani, «[Watched by the Web: Surveillance Is Reborn 'Big Data,' by Viktor Mayer-Schönberger and Kenneth Cukier](#)», *New York Times*, 10 de Junho, 2013, consultado a 30 de Dezembro, 2014). As agências de serviços secretos e de operações especiais dos E.U.A. desenvolveram

um programa de selecção de alvos que avalia os dados de vigilância de modo a automatizar ataques de drones, decisão que costumava caber a especialistas bem treinados (Cori Cridder, «[Killing in the Name of Algorithms: How Big Data Enables the Obama Administration's Drone War](#)», *Al-Jazeera America*, 4 de Março, 2014; consultado a 30 de Dezembro, 2014). E os responsáveis pela construção destes algoritmos levam para essa tarefa pressupostos pessoais a respeito de que categorias devem ser usadas para estruturar uma base de dados, que dados devem entrar para estas categorias, e de como elas devem ser interpretadas (Tarleton Gillespie, «[The Relevance of Algorithms](#)», *Culture Digitally*, 26 de Novembro, 2012; consultado a 15 de Novembro, 2014). Cada etapa da recolha e análise de dados deixa a nu preconceitos de natureza pessoal, ainda que os algoritmos sejam interpretados como sendo impermeáveis ao erro humano (Danah Boyd, *It's Complicated: The Social Lives of Networked Teens*. New Haven, CT: Yale University Press, 2014: 185).

Muito frequentemente, no entanto, os algoritmos são a nossa melhor esperança quanto à possibilidade de extrairmos dados relevantes de poço sem fundo de dados. Sem eles, seria impossível recorrer à Internet, tão impossível como recuperar um microchip numa biblioteca arrasada por um tornado. Para todos os efeitos, a questão não é se precisamos de algoritmos (claro que precisamos), mas a de saber o que eles podem ou não revelar. Isto pode parecer más notícias

Enquanto solução técnica para o problema do excesso de dados, os algoritmos deixam à vista as suas limitações assim que entram em jogo variáveis recalcitrantes, como o contexto. O contexto está pejado de factores que desafiam padrões previsíveis. Daí a maçada dos anúncios que julgam ser capazes de antecipar cada um dos nossos desejos partindo da recolha de palavras comuns usadas nos nossos emails ou pesquisas

àqueles que já carregam o fardo dos conteúdos online. Mas não nos deveria chocar, já que qualquer pessoa familiarizada com um motor de busca conhece as suas limitações.

De ano para ano, é espantoso verificar uma melhoria dos resultados, de tal modo, de facto, que mal precisamos de nos aventurar para lá da página de resultados para sabermos como vai estar o tempo, ou para compararmos preços de voos. Tarefas deste género são aquilo em que os algoritmos são eficazes: em listas de dados com propriedades relativamente fáceis de computar — propriedades que variam muito pouco no tempo, ou que não variam de todo —, eles são imbatíveis (Gary Marcus, «[Steamrolled by Data](#)», *The New Yorker*, 29 de Março, 2013; consultado a 31 de Dezembro, 2014). Ainda assim, em sistemas contextuais como o processamento de línguas naturais ou ferramentas de tradução avançadas, dão, na melhor das hipóteses, resultados aproximados, e estão muitas vezes errados — ou, pelo menos, parecem abafados. Os únicos dados que os algoritmos conseguem discernir com eficácia são conjuntos de dados baseados em correlações de relevância e destaque. Não são capazes de avaliar dados quanto à sua validade ou razoabilidade, e não são capazes de discernir a qualidade desses dados, nem deduzir regras de causalidade (ibid). Sem sombra de dúvida, são incapazes de traduzir juízos morais. Para já, o pensamento abstracto de ordem elevada ainda pertence apenas aos seres humanos.

em motores de busca. Uma vez que o contexto é um artefacto humano, apenas a mente humana é capaz de o compreender. Esta distinção continua a ser crucial, caracterizando-se por muito mais do que a dicotomia gasta do homem-vs-máquina. A tecnologia apenas importa na medida em que seja um meio para atingir fins humanos. Mesmo com a ajuda de todas as ferramentas automatizadas que nos auxiliam a tratar a infor-

mação, devemos continuar dispostos a, a partir de um certo ponto, pensar pela nossa própria cabeça.

Por impossível que pareça comprar um *wearable* de rastreio de dados de saúde ou um automóvel não pilotado sem antes derramar unidades de informação, a maneira como nos ajustamos é procurando compreender a natureza da informação em si mesma. Aferir dados com espírito crítico — aquilo de que são feitos os, tão incompreendidos, seminários em faculdades de humanidades, graças à maneira como continuam a ser vistas as humanidades em geral — é muito mais poderoso do que qualquer algoritmo. Qualquer pessoa que consiga compreender a Internet como um ecossistema dinâmico, e que compreenda, também, os motivos ideológicos para a criação e disseminação de conteúdos, será capaz de sobreviver a qualquer quantidade de informação que se lhe depare.

Assim, seja por divertimento ou para aprender, importa saber aferir aquilo que consumimos online e onde quer que seja. Os académicos decidiram chamar a isso «literacia digital», cuja definição modelo é «a capacidade de aferir, analisar, avaliar e criar mensagens numa variedade de formas» (Patricia Aufderheide, *Media Literacy: A Report of the National Leadership Conference on Media Literacy*, Washington, D.C.: Aspen Institute, Communications and Society Program, 1993: xx). A literacia digital envolve uma capacidade de compreender conteúdos baseada em vários critérios: quem o criou, a quem se dirige, como foi criado, que preconceitos contém, e como é que o seu criador espera que as pessoas pensem ou ajam. Elementos da retórica clássica, estes critérios ajudam a diferenciar entre um artigo de opinião sofisticado e uma lista expelida de um qualquer agregador de conteúdos, ou a distinguir entre a versão original de um vídeo de meia dúzia de piratagens. Isto não apenas permite que as pessoas tirem pleno partido dos conteúdos disseminados pela Internet, como lhes permite tornarem-se consumidores perspicazes.

Nas redes sociais, por exemplo, é comum os utilizadores partilharem links que remetem para conteúdos alojados onde quer que seja na Internet. Mas nem todo o conteúdo é criado por igual: de cada vez que lemos um post num blog, ou vemos um vídeo publicado no YouTube, aferimo-los. Frequentemente, o entretenimento é quanto nos basta, enquanto noutras alturas procuramos uma opinião especializada a respeito de algo que nos interessa. A simples quantidade de conteúdos e a ausência dos marcadores tradicionais de validade podem ser um quebra-cabeças e, como acontece tantas vezes com mal-entendidos, dar origem a conflitos.

Mais do que fazer-nos parecer inteligentes, ser apto a avaliar conteúdos *media* tem um impacto na maneira como pensamos e nos comportamos. Sem uma noção de como ou por que razão são criados, é fácil tomar opiniões por factos, que depois dão forma às nossas decisões (Flanagin e Metzler, 8; Gasser et al., 76-77; Lankes, 102). E no entanto a importância de procurar e avaliar informação nunca foi tão grande como hoje. A procura de auto-suficiência por parte do público atingiu um pico nos anos mais recentes, levando a um pico de automatização tecnológica. E com o acesso directo à informação vem o desafio de descortinar o que é relevante, e de o avaliarmos por nós mesmos.

Saber avaliar conteúdo online de maneira ponderada não é um dado. Dependendo da escola que frequentaram, os estudantes poderão ter recebido uma preparação rigorosa nesse sentido, ou preparação nenhuma (Boyd, 181-82; Harris, 156). Esta mesma discrepância educativa tem ramificações no mundo dos adultos, num contexto em que a tecnologia evolui a um tal ritmo que a sociedade tem dificuldades em manter-se a par dela. Por exemplo, conteúdos a respeito de questões de saúde podem variar deste um guia patrocinado por uma grande marca a um artigo publicado numa revista com *peer-review*. Nenhuma destas coisas é necessariamente melhor do que a outra, mas é crucial ser capaz de as

distinguir. Tal como acontece com a investigação para um trabalho escolar, pode ser difícil distinguir entre diferentes tipos de conteúdo e avaliá-los de acordo com isso. Isto pode passar para as plataformas de redes sociais, muitas das quais florescem à custa da partilha de conteúdos alo-

No que às soluções diz respeito, a literacia digital apenas vai até certo ponto. Ensina-nos a ajustarmo-nos a uma sobrecarga de informação, não nos ajudando a lidar com unidades de informação que silenciosamente determinam o nosso destino. Podemos falar em línguas disruptivas; podemos construir todos os *clusters* em Hadoop que quisermos; mas sem um debate robusto acerca da ética da recolha e uso de dados, a sobrecarga de informação será o menor dos nossos males. Para o bem e para o mal, esse debate começou há uma década, mas não demos por ele até ao dia em Edward Snowden divulgou provas dos vastos programas de vigilância da National Security Agency. Os *leaks* de Snowden expuseram uma combinação medonha de ciência de dados e espionagem. Ainda Snowden fugia para Rússia quando os decisores políticos norte-americanos se retrataram, aprovando nova legislação sobre privacidade digital, com dez estados a ratificar imediatamente nada menos que duas dúzias de leis. A Califórnia aprovou três, duas das quais regulam a remoção de conteúdos de websites, e a terceira o modo de rastreio dos *browsers*. O Oklahoma aprovou uma lei que protege dados de estudantes, enquanto o De-

gados em todos os cantos da Internet. Aprender a distinguir diferentes tipos de conteúdo pode prevenir mal-entendidos e neutralizar conflitos. Além disso, ajuda-nos a aprender a avaliar informação com um olhar crítico.

laware—aproveitando a boleia dos estados que o precederam—reduziu as restrições ao acesso a dados de um defunto.

À data em que escrevo, 15 estados norte-americanos decretaram leis que requerem consentimento explícito ou um mandado de captura para obter informações de um utilizador. Outros, como a Califórnia, continuam a introduzir legislação similar em face de ameaças de veto. O facto de ter sido preciso um *leak* comprometedor para galvanizar os legisladores, levando-os a agir, não é motivo de surpresa; o silêncio a respeito da viragem cultural que isso sugere, é. Uma coisa é aceitar a datificação como uma ordem natural à qual, a partir do momento em que descarregamos uma app, ou polimos um dado perfil, nos conseguimos adaptar — ou na qual conseguimos, até, florescer. Outra coisa é considerar onde fomos deixando dedadas digitais e, antes de mais, se temos direito a sabê-lo. No fim de contas, quando falamos de dados, estamos sempre por perto, uma vez que estamos a falar de nós mesmos.

Trad. Humberto Brito

NOTAS

1 Moldada por correntes culturais ao longo de séculos, a privacidade é um conceito em alteração. Pensadores do direito como Saul Levmore e Martha Nussbaum atribuem à actual iteração do conceito de privacidade quatro características: isolamento, ou liberdade do escrutínio público; intimidade, ou um espaço onde partilhar informação privada num sentido limitado; secretismo,

que restringe o conhecimento a pessoas ou grupos específicos; e autonomia, ou escolhas feitas privadamente. Ver Saul Levmore e Martha Nussbaum, «Introduction», *The Offensive Internet: Speech, Privacy, and Reputation*, ed. Saul Levmore e Martha Nussbaum (Cambridge, MA: Harvard University Press, 2010), 1-11, 10.