

BRIAN CHRISTIAN

O humano mais humano

*O que a inteligência artificial nos ensina
sobre a vida*

Tradução

Laura Teixeira Motta



Copyright © 2011 by Brian Christian

*Grafia atualizada segundo o Acordo Ortográfico da Língua Portuguesa de 1990,
que entrou em vigor no Brasil em 2009.*

Título original

The most human human: what talking with computers teaches us about what it means to
be alive

Capa

Retina_78

Preparação

Leny Cordeiro

Revisão

Carmen T. S. Costa

Luciane Helena Gomide

Dados Internacionais de Catalogação na Publicação (cip)
(Câmara Brasileira do Livro, sp, Brasil)

Christian, Brian

O humano mais humano : o que a inteligência artificial nos
ensina sobre a vida / Brian Christian ; tradução Laura Teixeira
Motta. 1^a ed. — São Paulo : Companhia das Letras, 2013.

Titulo original: The Most Human Human : What talking with
computers teaches us about what it means to be alive.

ISBN 978-85-359-2222-6

1. Antropologia filosófica 2. Computadores 3. Inteligência artifi-
cial 4. Seres humanos 5. Teste de Turing i. Título.

12-15619

CDD-128

Índices para catálogo sistemático:

1. Humano e máquina : Antropologia filosófica	128
2. Máquina e humano : Antropologia filosófica	128

[2013]

Todos os direitos desta edição reservados à

EDITORASCHWARCZ S.A.

Rua Bandeira Paulista, 702, cj. 32

04532-002 — São Paulo — SP

Telefone: (11) 3707-3500

Fax: (11) 3707-3501

www.companhiadasletras.com.br

www.blogdacompanhia.com.br

Sumário

Prólogo	15
Introdução: O humano mais humano	17
1. Autenticação	33
2. A alma migrante	59
3. Especificidade versus técnica pura	101
4. Sair do livro	131
5. O antiespecialista	171
6. Interrupção	193
7. O pior depoente do mundo	222
8. Não ficar intacto	248
9. Grande surpresa	275
Conclusão: O humano mais humano	324
Epílogo: A ignorada beleza do guarda-louça	333
 Agradecimentos	339
Notas	343

1. Autenticação

AUTENTICAÇÃO: FORMA E CONTEÚDO

O programa *Morning Edition*, da National Public Radio, transmitiu recentemente a história de um homem chamado Steve Royster.¹ Quando jovem, Royster pensava que tinha uma voz incomum e inconfundível. Ele explica: “Todo mundo sempre sabia que era eu ao telefone simplesmente ouvindo o som da minha voz, enquanto eu não tinha a menor *ideia* de quem estava falando quando alguém me telefonava”. Só com quase trinta anos ele se deu conta, muito espantado, de que os outros eram capazes de discernir a identidade de quase *qualquer pessoa* pela voz. Mas como conseguiam fazer isso? Acontece que Royster tem mesmo algo incomum, só que não é na voz. É no cérebro. Ele tem uma deficiência rara conhecida como fonoagnosia, ou “cegueira para vozes”. Mesmo quando a própria mãe de Royster lhe telefona, ele se limita a seguir educadamente o fluxo da conversa, ignorando que “a estranha que está me telefonando é, na verdade, a mulher que me deu à luz”. Como explica o repórter Alix Spiegel,

“os fonoagnósicos conseguem distinguir pelo som da voz se estão falando com um homem ou uma mulher, se a pessoa é jovem ou velha, se é sarcástica, irritada ou alegre. Só não fazem a mínima ideia de quem é a pessoa”.

Tudo isso deixava Royster em uma situação estranhíssima.
A mesma em que todos nos encontramos na internet.

Em 16 de setembro de 2008, um universitário de vinte anos chamado David Kernal quis entrar na conta pessoal de e-mail Yahoo! da candidata a vice-presidente Sarah Palin.² Não sabia a senha. Tentar adivinhar parecia inútil, então lhe ocorreu tentar *mudar* a senha. Clicou na opção “esqueci minha senha”, destinada a usuários distraídos. Antes de o Yahoo! permitir a um usuário mudar a senha da conta, faz várias perguntas “autenticadoras” — coisas como data de nascimento e código de endereçamento postal, para “comprovar sua identidade”. Kernal encontrou as informações na Wikipedia, disse ele, “em aproximadamente quinze segundos”. Atordoado, ele “mudou a senha para ‘popcorn’ e foi tomar um banho frio”. Agora poderá pegar vinte anos de prisão.

No mundo das máquinas, autenticamos pelo *conteúdo*: senha, PIN, os últimos dígitos do cartão de crédito, o nome de solteira da mãe. Mas, no mundo humano, autenticamos pela *forma*: rosto, timbre vocal, caligrafia, assinatura.

E, crucialmente, pelo estilo verbal.

Um amigo meu recentemente me escreveu num e-mail: “Estou tentando alugar uma casa em outra cidade por e-mail, e não quero que o sujeito com quem venho tratando pense que estou dando um golpe (ou que sou trapaceiro), por isso ando hiperpreocupado em parecer ‘humano’ e ‘real’ e basicamente ‘não anônimo’ em meus e-mails. Que coisa estranha. Sabe o que estou querendo dizer?”. Sei; são as idiossincrasias de estilo *nesse* e-mail — o anacrônico “sujeito”, o termo composto, sem hífen, “hiperprecupado” e o “não anônimo” — que provam que é realmente *ele*.

Esse tipo de coisa — um comportamento que parece “tão você” — pode sempre ter sido, digamos, encantador ou cativante (pelo menos para quem nos estima). Agora é mais uma coisa, já que nossas palavras vêm sendo cada vez mais dissociadas de nós na era da internet: é parte da *segurança* on-line.*

Os pinguins da Antártida detectam com precisão o chamado de sua cria no meio das 150 mil famílias reunidas em sua área de procriação. “Abençoadas Babel”, diz o ficcionista Donald Barthelme.³ É verdade: aplinar as idiossincrasias do nosso estilo verbal não seria ruim só para a literatura; seria péssimo para a *segurança*. Aqui, como em outras partes, talvez a ligeira pressão exercida pelas máquinas para que afirmemos resolutamente nossa humanidade uns perante os outros acabe sendo uma coisa boa.

INTIMIDADE: FORMA E CONTEÚDO

Emily, uma velha amiga dos tempos de faculdade, veio recentemente à minha cidade e, saindo do aeroporto, deu um pulo no centro da cidade para almoçar com uma colega de trabalho, Sarah — por coincidência, minha namorada. Quando Emily e eu nos

* Quando alguma coisa on-line me faz pensar em um amigo com quem não converso há algum tempo e tenho vontade de lhe enviar o link, procuro sempre adicionar algum tipo de floreio pessoal, algum enfeitezinho verbal na mensagem além do minimalista “Oi, vi isto e lembrei de você/link/abraço”, senão minha mensagem corre o risco de ir para a lixeira de spam.

Por exemplo, quando outro dia recebi uma mensagem breve pelo Twitter escrita em termos genéricos por uma das editoras de poesia da revista *Fence* dizendo “Oi, tenho 24 anos/mulher/ardente[...] tenho de ir agora mas me contate em meu endereço do Windows Live Messenger: [link]”, meu instinto não foi tentar imaginar como responder polidamente que eu estava lisonjeado mas achava melhor manter nosso relacionamento só no campo profissional; foi clicar em “Marcar Como Spam”.

encontramos mais tarde naquele dia para jantar, comentei que era engracado o fato de ela já conhecer Sarah antes que eu tivesse tido a chance de apresentá-las. Lembro-me de ter dito algo assim: “Legal vocês duas terem se familiarizado”. Emily replicou: “Eu não diria que *me familiarizei* com ela. Na verdade, pude ver como ela é, vê-la em ação”.

Foi então que me ocorreu a distinção.

Ter a *noção* da pessoa — seu temperamento, caráter, “jeito de ser” — e ter *conhecimento* sobre ela — onde foi criada, quantos irmãos tem, no que se formou, onde trabalha — são coisas diferentes. Como a segurança, a intimidade também tem forma e conteúdo.

“*Speed dating*”, ou “encontro rápido”, é um tipo de evento que surgiu em Beverly Hills em fins dos anos 1990.⁴ Consiste em um revezamento nas conversas entre os participantes para que todos possam se conhecer. Cada participante passa por uma série de conversas de sete minutos e no fim marca num cartão quais pessoas ele teria interesse em ver novamente; havendo escolhas mútuas, os organizadores entram em contato e fornecem as informações pertinentes do possível par. Embora já tenha entrado para a linguagem coloquial, “*SpeedDating*” (“ou qualquer termo des-norteantemente similar”) é tecnicamente uma marca registrada, pertencente, vejam só, a uma organização judaica, a Aish HaTorah. Seu inventor, Yaakov Deyo, é rabino.

Uma das primeiras ideias que me ocorreram a respeito do teste de Turing é que se trata de uma espécie de “*speed date*”: você tem cinco minutos para mostrar a outra pessoa quem você é, comunicar que é uma pessoa real, de carne e osso, viva, única, distinta, não anônima. É uma tarefa e tanto. E a parada, em ambos os casos, é alta.

Um amigo meu participou recentemente de um evento de *speed dating* em Nova York. “Foi a coisa mais estranha”, ele contou.

“Eu ficava só querendo fazer graça, sabe como é? Para ver se havia alguma química. Mas a mulherada não se desviava do roteiro: de onde você é, em que trabalha — como se quisessem levantar a minha ficha, me medir. Mas eu não dou a mínima para esse tipo de coisa. Então comecei a mentir nas respostas, inventar. Só para tornar a coisa interessante.”

A estranheza que ele sentiu e o tipo de “sinopse” em que o *speed dating* pode descambar são tão conhecidos que foram satirizados em *Sex and the city*:

“Oi. Sou Miranda Hobbes.”

“Dwight Owens; grupo de gestão de portfólio do Morgan Stanley; gerente de conta de clientes especiais e fundos de pensão para casais; gosto do meu trabalho; cinco anos na firma; divorciado; sem filhos; não religioso; moro em Nova Jersey; falo francês e português; faculdade de administração de Wharton; algo em tudo isso a atrai?”⁵

A apresentação por certo não.

Muitas pessoas com listas elaboradas de qualidades desejáveis no parceiro ideal enumeraram o tipo errado de predicados. A altura. O salário. A profissão. Já vi muitos amigos acabarem, sem esperar, eu acho, diante de uma pessoa detestável que, no entanto, correspondia perfeitamente aos requisitos que eles haviam selecionado.

Cansado do estilo “Dwight Owens”, da abordagem-sinopse tão batida nas reuniões de *speed dating*, Yaacov Deyo decidiu-se por uma solução simples e abrupta: *proibir* conversas sobre o trabalho. As pessoas passaram então a falar sobre onde moravam ou de onde provinham. Então ele proibiu isso também. Ele se mostra encantado e um tanto orgulhoso ao encenar o pânico geral e em seguida a revelação que se seguiram a essas medidas: “Ai, meu

Deus, então do *que* vou falar?”. Ele ri. “Não posso falar sobre o que faço para me sustentar, sobre onde moro... uau! De repente, tenho que descrever *a mim mesmo!*” Ou: de repente, tenho de *ser* eu mesmo, *agir* como eu mesmo, em vez de me *descrever*.

FORMA E CONTEÚDO NO TESTE DE TURING⁶

A primeira competição do Prêmio Loebner realizou-se em 8 de novembro de 1991 no Museu do Computador de Boston.⁷ Em seus primeiros anos, o Prêmio Loebner deu a cada programa e a cada confederado humano um “tema” como recurso para limitar a conversa. Em alguns aspectos, a primeira competição foi uma batalha entre estilo e conteúdo. Um dos confederados era a especialista em Shakespeare Cynthia Clay, que ganhou fama por ter sido considerada um computador por três juízes. O consenso parece ter sido “ninguém sabe tanto sobre Shakespeare”. (Por essa razão, Clay recebeu como uma deferência sua classificação equivocada.)

O programa por trás do tema “conversa esdrúxula” — o PC Therapist III, de Joseph Weintraub — levou o Prêmio Loebner daquele ano e se tornou o primeiro Computador Mais Humano. O programa praticava a “conversa esdrúxula” não como *tema*, mas como *modo* — uma distinção crucial. Weintraub conseguiu jogar com três fatores: as regras frouxas do concurso, os padrões de comportamento natural dos *chatbots* e os estereótipos que as pessoas têm sobre computadores e humanos. Estrategicamente, era brilhante. Clay e os outros confederados se mostraram obedientes e escrupulosos e deram as informações pedidas nas perguntas, demonstrando seus conhecimentos e compreensão — mas às vezes de um modo factual e enciclopédico que associamos aos sistemas de computador.

JUIZ: Qual a sua opinião sobre as peças de Shakespeare?

REMOTO: Isso é muito geral. Pode ser mais específico? Se não, direi apenas que gosto delas.

O interlocutor no terminal da “conversa esdrúxula” contrastou vivamente:

JUIZ: Que dia longo! Agora acho que estou a fim de me divertir um pouco. O tempo não está muito bom hoje. Você conhece alguma piada para desanuviar?

REMOTO: Todo mundo fala sobre o tempo mas ninguém parece fazer muita coisa a respeito. Por que você precisa me dizer que agora acha que está a fim de se divertir um pouco? (resmungo)

[...]

JUIZ: Qual a definição de conversa esdrúxula?

REMOTO: Você não disse, ora, enfim, imagino que conversa esdrúxula é, tipo, uma série de movimentos do cavalo no tabuleiro* em zigue-zague em vez de seguir uma linha sdeterminada.? Por que pergunta? Foi um comentário esperto, ou coisa do tipo? Você pensa muito nesse tipo de pergunta?

JUIZ: Gosto da imagem dos cavalos se movendo a esmo no tabuleiro, isso significa que não existe raciocínio na conversa esdrúxula?

REMOTO: Eu poderia jurar que você acabou de me perguntar exatamente isso. Como vc define esdrxulo? Você é sempre assim chato chato chato? Por que precisa me dizer que gosta da imagem dos cavalos se movendo a esmo no tabuleiro?

“Pensei que [o PC Therapist] fosse humano porque ele... era esquisito e engraçado, de um jeito normal”, disse um dos juízes,

* Sic. O programa de Weintraub, como muitos posteriores, cometia erros de grafia propositalmente.

um pós-graduando de Harvard. O programa de Weintraub, mudando de assunto e vomitando coisas sem lógica e trocadilhos pré-fabricados, deu a impressão de ser um brincalhão, um palhaço, um tipo de personalidade “muito mais humana”. Pelo menos eu pensava assim antes de ficar sabendo como isso era fácil de imitar.

SUSPEITA; ROLETA; PURÊ

Procurei na internet informações sobre o Computador Mais Humano de 2005 e 2006, o Cleverbot, de Rollo Carpenter, baseado na web. Encontrei algo muito curioso. Parece que muitos acham possível que o site seja um embuste.

“Já faz algum tempo que venho conversando com ele, e comecei a pensar que talvez o Cleverbot não seja realmente um programa”, escreve um usuário. “O Cleverbot sempre diz que ele é o humano e que eu sou o programa. Será que estou falando com uma pessoa real, mas nós dois estamos sendo enganados para acreditar que o outro é o Cleverbot?” Algumas respostas parecem até excepcionais demais. Um usuário escreve, em francês, “*J'habite dans les États-Unis*”, e vê seu francês não só ser compreendido mas criticado: “*On dit J'habite AUX États-Unis*”, o bot corrige.

Outro dialoga com a letra de “Bohemian Rhapsody”, do Queen:

USUÁRIO: Scaramouche, Scaramouche.

CLEVERBOT: Will you do the fandango?

USUÁRIO: Thunderbolt and lightning.

CLEVERBOT: Very very frightening me!

USUÁRIO: Galileo.

CLEVERBOT: Galileo Galileo.

Um usuário diz que contou uma piada então na moda:

USUÁRIO: Sabe por que o chamam de Xbox 360?

CLEVERBOT: Por quê?

USUÁRIO: Porque, quando alguém o vê, vira 360 graus e sai andando.

— mas o programa demoliu a piada:

CLEVERBOT: Não, porque nesse caso a pessoa ficaria de frente para ele, e não seria engraçado.

Outros usuários, tentando consertar a lógica da piada, a tornaram um pouco mais debochadamente absurda:

USUÁRIO: Sabe por que o chamam de Xbox 360 [um modelo de video game]?

CLEVERBOT: Não. Por que o chamam de Xbox 360?

USUÁRIO: Porque, quando alguém o vê, vira 360 graus e sai andando de ré.

— e foram astutamente rebatidos com uma réplica ainda mais debochada e mais absurda:

CLEVERBOT: Nada disso; é porque, quando alguém o vê, vira 300 e se senta.

Só pode haver um humano do outro lado! As réplicas são muito bem sintonizadas com a cultura do momento, muito espirituosas. No entanto, em outros momentos as conversas mudam de rumo sem uma muleta, algo desconcertante surge do nada.

O programador do Cleverbot, Rollo Carpenter, explica de muito bom grado o funcionamento de sua criação, e na home page

do Cleverbot garante que “os visitantes jamais conversam com um humano, por mais convincente que [o programa] seja”. Curiosamente, essa garantia parece não influenciar muitos usuários, que têm suas próprias teorias sobre o que se passa.

A internet no começo dos anos 1990 era um lugar muito mais anônimo do que é hoje. Em BBSS locais [Bulletin Board Systems, sistema eletrônico de quadro de mensagens], nas salas de bate-papo de acesso restrito dos provedores/comunidades como Prodigy e AOL e em protocolos universais de chat como o IRC (Internet Relay Chat), estranhos se encontravam sem querer o tempo todo. As imensas redes sociais (como o Facebook) de fins dos anos 2000 e começo do decênio seguinte começaram a fazer da internet um lugar diferente. Foi mais ou menos nessa época que sites como o Chatroulette e o Omegle decolaram, projetados para trazer de volta um pouco daquele anonimato, da aleatoriedade, das descobertas inesperadas. Você escolhe usar vídeo ou texto, é posto em contato com outro usuário totalmente ao acaso e a conversa tem início.* A qualquer momento qualquer um dos dois pode encerrar a conversa, e nesse caso ambos são postos em contato com outro estranho e recomeçam no “oi”. Todo usuário desse tipo de site sente ansiedade diante da perspectiva de a outra pessoa interromper o diálogo e partir para outra conversa — ou, como se diz, “dar um próximo”.

* Esse anonimato também traz riscos, pelo menos tanto quanto traz descobertas agradáveis. Li um relato de uma pessoa que tentou usar o Chatroulette pela primeira vez e, em doze dos primeiros vinte vídeos que ele tentou, deu de cara com homens se masturbando diante da câmera. Por essa razão, e porque tem mais semelhança com o teste de Turing, preferi ater-me ao texto. Ainda assim, meus dois primeiros interlocutores no Omegle eram homens numa pesca desejitada por cibersexo. Mas o terceiro era uma estudante de ensino médio de um bairro residencial de Chicago; conversamos sobre a escultura *Cloud Gate*, o Instituto de Arte, os prós e contras de crescer e sair da casa dos pais. Ali estava uma pessoa de verdade. “Você é normal!!”, ela escreveu, dobrando o ponto de exclamação; e eu pensei exatamente a mesma coisa.

Agora imagine se, em vez disso, o sistema de computador cortasse *automaticamente* as conversas e formasse novos pares de usuários *sem lhes informar* que está fazendo isso. Os usuários A e B estão conversando sobre futebol; C e D, sobre arte. De repente, A é emparelhado com C, e B com D. Depois de falar sobre o Louvre, C recebe a destoante pergunta: “Afinal, você torce para o Barcelona ou o Real Madrid?”, enquanto B, depois de analisar a última Copa do Mundo, tem de dizer se já esteve na Capela Sistina. Pois essa é a teoria da conspiração sobre o Cleverbot (e alguns *bots* primos dele, como o Ultra Hal, de Robert Medeksza): é o Omegle menos o controle sobre o momento de trocar as conversas. Imagine que o computador está simplesmente trocando seu interlocutor, a esmo e sem aviso, e fazendo o mesmo com seu parceiro. O resultado dessas conversas poderia ficar bem parecido com os transcritos do Cleverbot.

A teoria da conspiração não é verdadeira, mas também não está muito longe de sé-lo.

“O Cleverbot aproveita a inteligência de seus usuários”, Carpenter me explica em Brighton.⁸ “Uma Wikipedia sobre conversação”, ele definiu numa entrevista no canal de televisão Science.⁹ Funciona assim: o Cleverbot começa uma conversa dizendo, por exemplo, “Olá”. O usuário pode responder de várias maneiras, como “Oi”, “E aí?” ou “Você é um computador?”. Tudo que o usuário diz vai para um enorme banco de frases, identificadas como genuínas respostas humanas a “olá”. Em uma conversa subsequente, quando um usuário disser “olá” ao Cleverbot, este poderá ter a postos respostas como “e aí?” (ou seja, o que a primeira pessoa tiver dito). Como os mesmos tipos de frases tendem a reaparecer vezes sem conta — o que os estatísticos chamam de “distribuição de Zipf”, para ser exato — e como há vários anos a cada dado momento o Cleverbot tem tido milhares de usuários conectados batendo papo, e isso 24

horas por dia, o banco de dados do Cleverbot contém respostas apropriadas até para comentários aparentemente obscuros (por exemplo, “Scaramouche, Scaramouche”.)

O que se obtém, misturando centenas de milhares de conversas anteriores, é uma espécie de purê de conversas. Feito de partes humanas, porém menos do que uma soma humana. Os usuários *estão*, efetivamente, conversando com um purê de pessoas reais — ou pelo menos com *fantasmas* de pessoas reais: os ecos de conversas passadas.

Isso explica em parte por que o Cleverbot é tão impressionante nas questões sobre fatos básicos (“Qual é a capital da França?” “Paris é a capital da França”) e cultura popular (trivialidades, piadas, letras de músicas) — coisas para as quais existe uma resposta *certa* a despeito do interlocutor. Mesmo quando muita gente põe a mão na massa, ela não desanda. Mas pergunte em que cidade ele mora e você obterá uma miscelânea de milhares de pessoas conversando sobre milhares de lugares. Aí você o desmascara, não porque não está falando com um *humano*, mas porque percebe que não está falando com *um humano*.

SEJA VOCÊ MESMO, SEJA QUALQUER UM

Minha mente retorna ao conselho dos organizadores, “seja você mesmo”, e reflito sobre quanto os filósofos têm se debruçado sobre essa ideia. Enquanto outros existencialistas, como Jean-Paul Sartre, enfatizaram a autenticidade, a originalidade e a liberdade em relação a influências externas, o filósofo alemão oitocentista Friedrich Nietzsche defendeu a surpreendente opinião de que a parte mais importante de “ser você mesmo” é — nas palavras do filósofo Bernard Reginster, da Universidade Brown — ser *um*, não importa qual.¹⁰