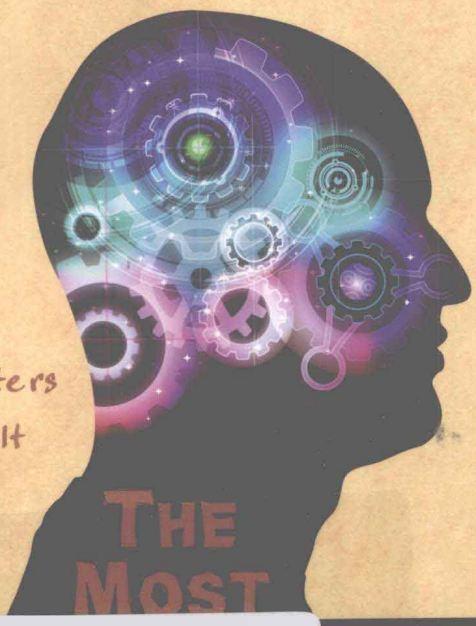
科学大讲堂

谨以此书纪念不朽的阿兰·图灵诞辰100周年

最有人性的“人”

人工智能带给我们的启示

◆[美]布莱恩·克里斯汀（Brian Christian）著
◆闾佳 译



What Talking With Computers Teaches Us About What It Means To Be Alive

THE MOST HUMAN MAN

思考意味着什么？

这个问题困扰了哲学家们上千年，
也困扰了计算机科学家们数十年。
2009年人工智能洛伯纳大奖得主是如何解答这个问题的？

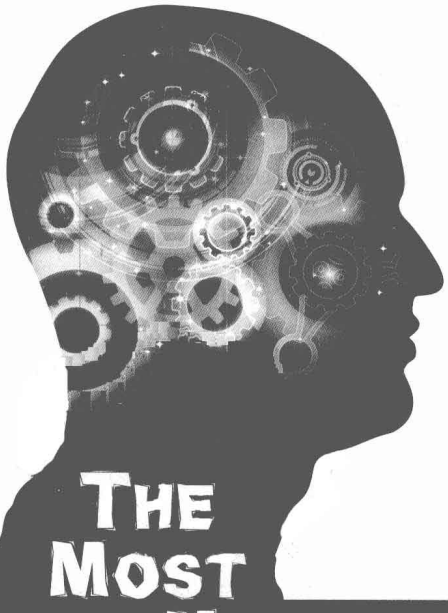
人民邮电出版社
POSTS & TELECOM PRESS

科学大讲堂

最有人性的“人”

人工智能带给我们的启示

◆[美]布莱恩·克里斯汀（Brian Christian）著
◆闾佳 译



THE MOST HUMAN HUMAN

What Means To Be Alive

人民邮电出版社
北京

内 容 提 要

本书记录了作者布莱恩·克里斯汀（Brian Christian）于2009年参加人工智能洛伯纳大奖赛的有趣经历。本书用富有诗意的笔法介绍了人工智能的发展历程，从图灵测试入手，从多个方面阐述了人工智能的本质，引出了“机器是否能够替代人”这个人工智能的根本问题。作者从心理学、医学、生物学的试验以及哲学等多个角度阐述了人的特质，列举了丰富的实例，从而给读者展示了对人、智能、人工智能以及人性的深度思考。

本书结合了信息科学和人文哲理的新思维，全面探索了计算机是如何重新改造了我们对人类定义的看法。计算机为什么能骗过裁判？机器是否能够代替人类？机器究竟能不能理性思考？人类有别于机器的特殊之处究竟是什么？如果你想了解这些问题，千万不要错过本书！

当今时代，人们在屏幕前花的时间越来越多。在日光灯照亮的房间里，在卧室里，在这样那样的电子数据交换设备的这一端或那一端。在这样的交流当中，人是什么？活着是什么？你的人性体现在什么地方？

——大卫·福斯特·华莱士（David Foster Wallace）

献给吾师
森林的变化多么美呵，
变色龙随着它改变肤色，
螳螂匍匐在绿叶上，
和它融为一体，
绿叶更加栩栩如生……

——美国现代诗人，理查德·威尔伯（Richard Wilbur）

我以为，纯哲学若能改善日常生活，那是很好的；可要是不能的话，就把它扔一边儿去。

——美国当代哲学家，罗伯特·波西格（Robert Pirsig）

身为美国总统，我相信机器人学能激励青少年献身于科学和工程技术。我也希望能对这些机器人提防着点儿，免得它们暗地里捣鬼。

——贝拉克·奥巴马（Barack Obama）

Prologue

楔 子

人工智能的先驱，信息理论的创立人，克劳德·香农（Claude Shannon）在工作中结识了玛丽·伊丽莎白（Mary Elizabeth），玛丽后来成了他的妻子。那是在20世纪40年代初，美国新泽西州默里山的贝尔实验室。那时的他，是一名工程师，正从事战时加密技术和信号传输工作。而那时的她，则是一名Computer（译注：这里的Computer一语双关，既是计算机的意思，也有数据处理员的意思）。

Contents

目 录

第1章
Chapter 1

当我读到2008年机器只差一票就通过测试的消息时，我意识到，2009年恐怕就是那跨越门槛的年份，身体里突然冒出一个不晓得来自何方的坚定声音：我要做那最后一道大门的守护者！

序幕：
最有“人性”的人

我从离家万里之遥的酒店醒来，发现浴室里竟然没有淋浴器。于是我15年来头一回在浴缸里泡了澡。我照例吃了早餐，那是几颗样子不大好看的西红柿、若干烤豆子和4片搁在小金属架子上的白面包——竖着搁的，就像摆在书架上的书。之后，我踱进带着咸味的空气里，顺着海岸线信步前行。在这个国家，我母语的诞生地，我却看不懂路边竖的广告牌上写着的是什么。一块牌子上醒目地印着几个大字：“良屋招租，已有预约”（译注：原文为“Let Agreed”，是近年来在英国流行的广告用语，作者是美国人，所以看不懂）。我完全不解其意。

我停下脚步，默默地凝视着大海片刻，在脑海里对刚才那块招牌做了一番分析和再分析。一般而言，这类新奇语言和文化差异会激起我的兴趣，但是今天，它们几乎成了让我焦虑的症结。因为接下来的两个小时，我会坐在一台计算机面前，跟若干陌生人一起进行几轮为时5分钟的即时聊天。而在计算机的另一端，是由一位心理学家、一位语言学家、一位计算机科学家以及一位英国流行科技节目的主持人组成的评审团。我跟他们对话的目的怪异透顶——说是我这辈子碰到的最怪的事情也不为过。

我必须说服他们，我是个人。

谢天谢地，我真的是个人。只可惜目前还不知道这一点能对此事起到多大的帮助。

此栏为原书页码 3/4

图灵测试

每一年，人工智能（AI，Artificial Intelligence）协会都要举办一场最令人期待也最富争议的盛大集会——名为“图灵测试”的竞赛。该竞赛的名称来自英国数学家阿兰·图灵（Alan Turing），他是计算机科学的创立人之一。1950年，他试图解答该领域历史最为悠久的一个问题：机器能思考吗？也就是说，有没有可能制造出一台精密复杂得能思考、有才智、有思想的计算机呢？如果有一天这样的机器真的诞生了，我们要怎样进行判断呢？

图灵并未从纯理论的角度来探讨这个问题，而是设计了一项实验。在这一实验中，由评审团通过计算机终端，向两名“受试者”提出问题，受试者之一

是真正的人，另一个则是计算机程序。评审团看不见谁是谁，只能通过受试者的回答来判断。提问的内容没有限制，可以是生活常识（例如蚂蚁有多少条腿，巴黎属于哪个国家）、名人八卦、深邃的哲学问题——总之，人类对话涉及的一切都行。图灵预言，到2000年，计算机能够在5分钟的谈话之后，愚弄30%的人类评委，让他们相信它是人。因此，“说机器能思考，就算不得是无稽之谈了。”

图灵的预言迄今尚未实现。不过，在2008年英格兰雷丁举办的竞赛中，最优秀的程序仅以一票之差惜败。2009年的图灵测试在布莱顿举行，这将是一场决定性的赛事。

我参加的就是这场比赛，我是跟顶尖人工智能程序对抗的4名人类卧底之一。在每一轮测试中，我将跟其他的受试“人”一起，跟AI程序配对，接受评委的裁断——我的任务是让评委相信我真的是个人。

评委会逐一跟我们聊上5分钟，接着有10分钟的思考时间，而后选出他认为是真人的那一方。评委们还要在一张得分表上打分，评判自己做出判断的信心有多大——这也是决定胜负的一项标准。获得评委们最多票数和最高信心度的程序，即可获得“最有人性的计算机”大奖（不管它是否能骗过30%的评委，都将通过图灵测试）。各研究小组竞相角逐的就是这个大奖，这一奖项不光有奖金，也是赛事组织者和观众们最关心的。有趣的是，能获得评委们最多票数和最高信心度的受试人，将会赢得“最有人性的人”大奖。

《连线》（Wired）杂志专栏作家查尔斯·普拉特（Charles Platt）在1994年成了首批获奖者之一。他是怎么做到的呢？他说他靠的是“暴躁、喜怒无常、惹人讨厌”。这不光让我觉得荒唐滑稽，从更深层的意义来说，更挑起了我的战斗心：我们究竟要怎么做，才能做个最有人性的人呢——我是说，不光是在测试的环境下，也在日常的生活中？

参　赛

图灵测试（它已经成了“洛伯纳大奖（Loobner Prize）”的具体化身了）的发起人和组织者是个有趣的人物：便携式跳舞毯大亨休·洛伯纳（Hugh

Loebner)。记者问他赞助和策划年度图灵测试的动机，洛伯纳把“懒惰”一词首当其冲地举了出来。显然，洛伯纳眼里的未来乌托邦，就是人类什么也不干，把所有工作都外包给智能机器。我必须得说，这样的未来憧憬，让我觉得太过绝望。我对人工智能大行其道的世界另有一番看法，我参加测试的理由也截然不同。但不管怎么说，最关键还是那个中心问题：计算机是怎样重塑我们的自我意识的？这一进程会带来什么样的后果呢？

5/6

因为不知道怎么才能参赛，我就从最高层入手——直接联系休·洛伯纳本人。我很快找到了他的网站。网站上的内容五花八门，有制作隔离栏杆的汞合金原材料信息[1]、性工作平权活动[2]以及有关奥运奖牌构成成分的丑闻[3]，最后我总算找到了以他名字命名的计算机大奖的信息，还有他的电子邮件地址并给他发了一封希望参加图灵测试的信函。之后，他给我回了邮件，让我与一个叫菲利普·杰克逊（Philip Jackson）的人联系。菲利普是英国萨里大学的一名年轻教授，负责2009年度在布莱顿举办的洛伯纳大奖赛的后勤工作。这场比赛挂靠在2009年语音通信大会（2009 Interspeech Conference）下面。

我用Skype跟杰克逊教授通上了话。这是个年轻聪明的小伙子，他热情洋溢，脸上有些过度操劳的痕迹，但别有一番学术界新人的气色。这一点，加上他迷人的英国腔，叫我立刻喜欢上了他。

他问起我的情况。我说我是一个以科学和哲学为主题的非虚构作家，我对科学、哲学与日常生活的交汇很感兴趣，也为图灵测试和“最有人性的人”这个想法着迷。想想看，捍卫人类尊严的参赛者这一概念多浪漫呀！国际象棋大师加里·卡斯帕罗夫（Garry Kasparov）大战超级计算机“深蓝”！人机大战挑战赛（Jeopardy）里的肯·杰宁斯（Ken Jennings）对抗最新款的IBM系统“沃森”

6/7

1 隔离栏杆最近似乎取代了便携式跳舞毯，成了洛伯纳公司“皇冠产业”（Crown Industries）的招牌产品。“皇冠产业”也是洛伯纳大奖的主要赞助商。

2 我相信发现这事儿有点讽刺的肯定不只我一个人：一个致力于推动人工智能进步的人，竟然纵容自己花钱购买别人提供的“亲密服务”（这是他在《纽约时报》和若干电视脱口秀节目上公开承认的）？

3 所谓的“金”牌其实是镶金的银牌，这固然是有点匪夷所思。洛伯纳为这事儿已经愤怒十多年了，他不断抗议、做讲演，还发行了一份名为《Pants on Fire News》的电子新闻报。

（Watson）！（我的思绪很快跳到了更狂野的幻想上，类似《终结者》、《黑客帝国》那种电影，虽说图灵测试里肯定没那么多机关枪）。当我读到2008年机器只差一票就通过测试的消息时，我意识到，2009年恐怕就是那跨越门槛的年份，身体里突然冒出一个不晓得来自何方的坚定声音：*我要做那最后一道大门的守护者！*

不止如此，图灵测试还在计算机科学、认知科学、哲学和日常生活的交叉领域提出了一连串令人兴奋又不安的问题。我在上述每一个领域都做过研究、写过文章，发表过经同行评审的认知科学研究论文。我发现，图灵测试借鉴了这些学科的知识，又把它们全都联系了起来，这一点是最叫人信服的。聊天期间，我告诉杰克逊教授，我认为自己兴许能为洛伯纳大奖赛带来一些很独特的东西。这其中一方面，是从我参加测试的实际表现着眼；另一方面，则是将这段经历，以及图灵测试提出的更广泛的问题，与大量受众联系起来——我觉得这会在公共文化上掀起一轮热烈且重要的讨论。我没怎么费功夫就让杰克逊教授同意了这番看法，于是，我的名字很快就上了参赛花名册。

杰克逊教授简要地向我介绍了比赛的后勤状况，并对我提了一个建议，也是我从过去参赛的受试者那儿听说过的：“没什么要多加注意的，真的！你是个人，做你自己就好。”

事实上，自从1991年举办首届洛伯纳大奖赛开始，“做你自己”就是它的口号。可在我听来，它却有点像是对人类直觉的过度自信，也带些“和稀泥”的意思。我们对抗的人工智能程序大多是数十年工作的产物——当然了，我们也是。但人工智能研究团队有着巨大的数据库来测试程序，他们还对这些资料做了统计分析：如何巧妙地引导谈话，让对话偏离程序的短处，迎合它的长处；哪些对话思路能深入交流，哪些不能——脱离了日常直觉，一般的受试者都表现得不怎么样。这是特别奇怪也特别有趣的一点，我们的社会对交谈、当众发言和约会教练有着绵绵不断的需求，就是极好的证据。在2008年的竞赛笔录里，评委对没法与之顺利展开谈话的人类“卧底”深感抱歉。一位评委说：“我挺心疼他们的（指人类卧底）。我猜，他们肯定厌倦了无休无止地谈论天气。”另一位则顺势补充道：“我这么老套，真抱歉。”与此同时，另一个窗口里的计算机却显然让评委着迷，“他”敲出一连串的“lol”和“：P”（*译注：此为计算机聊天*

时示意友好的文字符号）。我们可以做得更好的。

所以，我必须承认，从一开始，我就存心要彻底违抗组织者“9月按时到布莱顿，‘做你自己’就好”的建议。相反，我花了好几个月做准备，收集尽量多的信息和经验，打算到布莱顿去全力发挥出来。

按说，“做准备”没什么好奇怪的，不管是参加网球比赛、拼字比赛，还是标准化测试一类的东西，我们都会严加训练，做好准备。但考虑到图灵测试是为了评估我有多少“人性”，光是按时现身似乎还不够。我坚决认为“人性”还有更多的含义。至于“更多”的人性是什么意思，将是本书的叙述重点——而这一路上找到的答案，不光适用于图灵测试，还适用于生活里的更多地方。

迷上伊万娜

先说个有点奇怪、颇具讽刺意味的故事：加州大学圣地亚哥分校的心理学家、《解析图灵测试》（Parsing the Turing Test）科学卷的编辑、洛伯纳大奖的共同创办人（另一位创办人便是休·洛伯纳），罗伯特·爱普斯坦（Robert Epstein）博士在2007年冬天订阅了一份网上约会服务。他给一位名叫伊万娜的俄罗斯女性写起了长信，对方也写了长信回复，描述自己的家庭、日常生活以及她对爱普斯坦与日俱增的感情。终于，有些事情不对了起来。唉，我就长话短说吧，爱普斯坦意识到自己是在跟一段计算机程序——你是不是猜到啦？——往来了4个多月的缠绵情书。可怜的家伙：难道网络痞子们天天用垃圾邮件骚扰他的电子邮箱还不够，如今还要迷乱他的心吗？

8
9

一方面，我只想坐下来大肆嘲笑这可怜的家伙：看在老天的份上，他可是创办了洛伯纳奖的人呐！真是个呆子！但反过来再想，我也挺同情他。21世纪避无可避的垃圾邮件不光塞满了我们的收件箱，阻塞了世界的带宽（网络上97%的电子邮件是垃圾邮件；每天垃圾邮件的发送数量达上百亿封；每天用来处理全世界垃圾邮件的电力，足以支撑一个小国家了[1]），还做了件更加糟糕的事情——它侵蚀了我们的信任感。收到朋友们发来的信息，我总要浪费少许的

1 比如爱尔兰。

能量，至少读完最初的几个句子，才能判断这封信是否出自他们本人手笔。为此，我感到深恶痛绝。21世纪，我们戒备森严地过着所谓的数字化生活，所有的通信都是在进行图灵测试、所有的沟通都分外可疑。

上面是悲观的版本，以下呈上乐观版。我敢打赌，爱普斯坦吸取了教训，我也敢打赌，这个教训复杂且微妙，而不光是“想跟诺夫哥罗德（Nizhny Novgorod，俄罗斯工业城市）的不知什么人建立网上情缘，真是个愚蠢的念头。”我想，至少，他会深入地思考怎么会用了4个月才意识到自己和“伊万娜”之间并没有实际的交流，以后，他会更快地抽中“上上签”，选出“真正的人类”交流。他的下一个女朋友——但愿是一个如假包换的“智人”（译注：Homo sapiens，生物学中人的学名），也但愿她别住在11个时区之外——或许还得感谢“伊万娜”呢（当然，这笔人情债有点绕）。

9/10

非法比喻

20世纪40年代，克劳德·香农在贝尔实验室碰见贝蒂（贝蒂是玛丽·伊丽莎白的昵称）时，她真的是一名计算员（Computer）。这在我们听来也许觉得挺奇怪，但他们自己却并不觉得，并且他们的同事也不觉得。在贝尔实验室，他们的恋情完全正常，甚至非常典型。工程师总是和“Computer”谈恋爱，天经地义嘛。

阿兰·图灵1950年发表的论文《计算机器与智能》（Computing Machinery and Intelligence）提出了我们目前所知的人工智能领域，也引发了对图灵测试（图灵最初称之为“模仿游戏”）延续至今的探讨和争论。但现代的“计算机”和图灵时代的“Computer”完全是两回事。20世纪初的“Computer”并不是21世纪日常生活（办公室、家、汽车甚至衣服口袋）里随处可见的数字处理设备——而是对一份工作岗位的描述。

从18世纪中叶起，计算员——多为女性——就出现在各种公司和机构的薪水簿上了。他们进行计算，做数字分析，有时也使用早期计算器。很多科学伟业的背后都少不了这些早期人类计算员的辛勤工作，比如第一次准确预测出哈雷彗星的回归、牛顿重力理论的早期证明（以前只根据行星轨道做过检验）以