图灵机、人工智能以及我们的世界

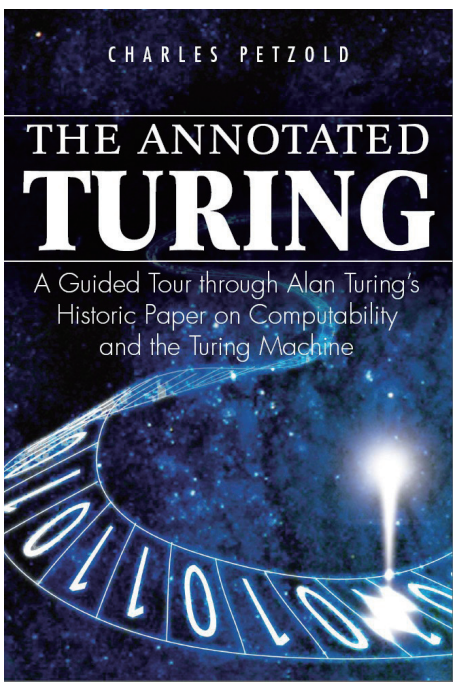
顾森

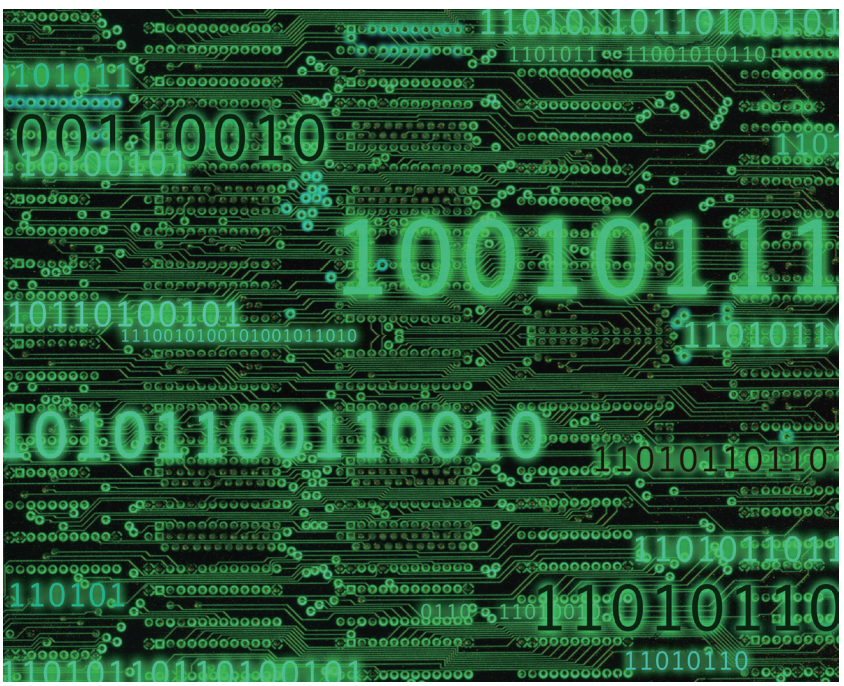
最近读完了 The Annotated Turing 一书，第一次完整地阅读了图灵最经典的那篇论文，理解了图灵机提出的动机和由此带来的一系列结论。不过，这本书的最大价值，则是让我开始重新认识和思考这个世界。在这里，我想把我以前积累的哲学观点和最近一些新的思考记下来，与大家一同分享。今年是阿兰·图灵诞辰 100 周年，图灵公司将推出这本书的中译本《图灵的秘密》，现在正在紧张的编辑排版中，不久之后就能和大家见面。

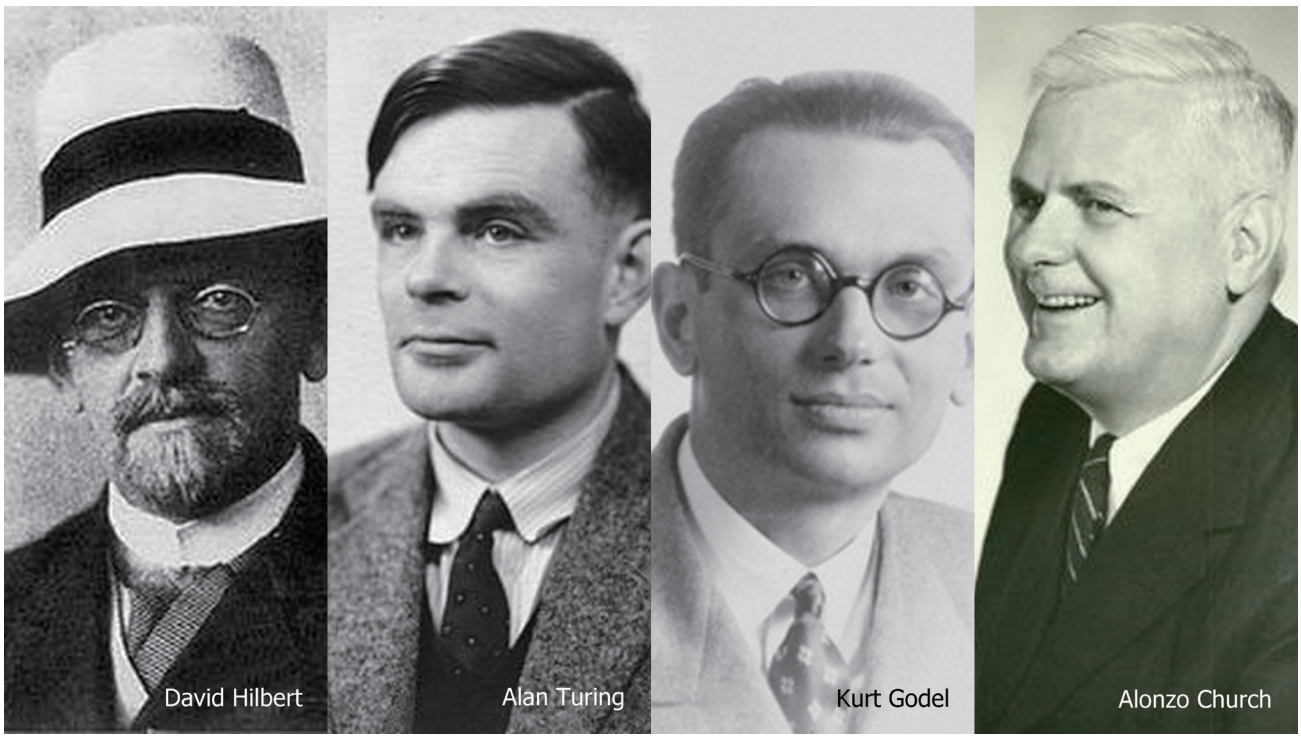
1928 年，大卫·希尔伯特（David Hilbert）提出了一个著名的问题：是否存在一系列有限的步骤，它能判定任意一个给定的数学命题的真假？这个问题就叫做 Entscheidungsproblem，德语“判定性问题”的意思。大家普遍认为，这样的一套步骤是不存在的，也就是说我们没有一种判断一个数学命题是否为真的通用方法。为了证明这一点，真正的难题是将问题形式化：什么叫做“一系列有限的步骤”？当然，现在大家知道，这里所说的“有限的步骤”指的就是由条件语句、循环语句等元素搭建而成的一个机械过程，也就是我们常说的“算法”。不过，在没有计算机的时代，人们只能模模糊糊地体会“一个机械过程”的意思。1936 年，阿兰·图灵在其著名的论文 On computable numbers, with an application to the Entscheidungsproblem 提出了一种假想的机器，第一次给了“机械过程”一个确凿的含义。

图灵提出的机器非常简单。假设有一张无穷向右延伸的纸条，从左至右分成一个一个的小格子。每一个小格子里都可以填写一个字符（通常是单个数字或者字母）。纸条下方有一个用来标识“当前格子”的箭头，在机器运行过程中，箭头的位置会不断移动，颜色也会不断变化。不妨假设初始时所有格子都是空白，箭头的颜色是红色，并且指向左起第一个格子。为了让机器实现不同的功能，我们需要给它制定一大堆指令。每条指令都是由五个参数构成，格式非常单一，只能形如“如果当前箭头是红色，箭头所在格子写的是字符 A，则把这个格子里的字符改为 B，箭头变为绿色并且向右移动一格”，其中最后箭头的移动只能是“左移一格”、“右移一格”、“不动”中的一个。

精心设计不同的指令集合，我们就能得到功能不同







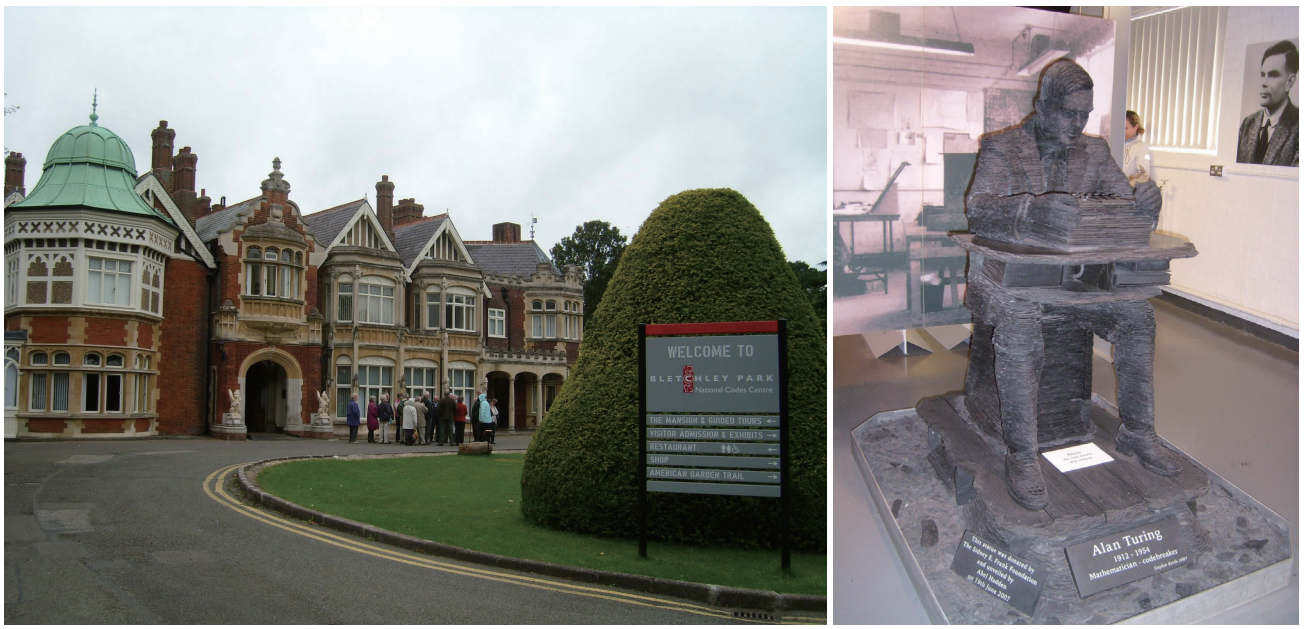
可计算理论研究的创始人（从左至右）：希尔伯特，图灵，哥德尔，邱奇

的图灵机。你可以设计一个生成自然数序列的图灵机，或者是计算根号 2 的图灵机，甚至是打印圆周率的图灵机。图灵本人甚至在论文中实现了这么一种特殊的图灵机叫做通用图灵机，它可以模拟别的图灵机的运行。具体地说，如果把任意一个图灵机的指令集用图灵自己提出的一种规范方式编码并预存在纸条上，那么通用图灵机就能够根据纸条上已有的信息，在纸条的空白处模拟那台图灵机的运作，输出那台图灵机应该输出的东西。

但是，图灵机并不是无所不能的。图灵证明了一个看似有些惊人的事实：不存在这样的一个图灵机，它能读取任意一个图灵机的指令集，并判断该图灵机是否将会在纸条上打印出至少一个 0 。注意，简单地用通用图灵机做模拟并不是一个可行的方案，因为模拟到现在还没有打出 0 ，不意味着今后永远不会打出 0 。这个定理有一个更深刻的含义，即没有一种通用的方法可以预测一台图灵机无穷远后的将来（后人把这个结论简化为了著名的停机问题）。正如 The Annotated Turing 封底上的一段文字所说：在没有计算机的时代，图灵不但探索了计算机能做的事，还指出了计算机永远不能做到的事。

在论文的最后一章，图灵给出了一种图灵机指令集和一阶逻辑表达式的转换规则，使得这个图灵机将会打出 0 来，当且仅当对应的一阶逻辑表达式为真。然而，我们没有一种判断图灵机是否会输出 0 的算法，因此我们也就没有一种判断数学命题是否为真的通用办法。于是，Entscheidungsproblem 有了一个完美的解答。

有趣的是，图灵机本身的提出比 Entscheidungsproblem 的解决意义更大。计算机诞生以后，出现了五花八门的高级编程语言，一个比一个帅气，但它们的表达能力实际上都没有超过图灵机。事实上，再庞大的流程图，再复杂的数学关系，再怪异的语法规则，最终都可以用图灵机来描述。图灵机似乎是一个终极工具，它似乎能够表达一切形式的计算方法，可以描述一切事物背后的规律。在同一时代，美国数学家阿隆佐 • 邱奇（Alonzo Church）创立了 λ 算子（λ-calculus），用数学的方法去阐释“机械过程”的含义。后来人们发现，图灵机和 λ 算子是等价的，它们具有相同的表达能力，是描述“可计算性”的两种不同的模型。图灵机和 λ 算子真的能够描述所有直观意义上的“可计算数”、“可计算数列”、“可计算函数”吗？有没有什么东西超出了它们的表达能力？这个深刻的哲学问题就叫做邱奇 - 图灵论题（ Church-Turing thesis）。当然，我们没法用形式化的方法对其进行论证，不过大家普遍认为，图灵机和 λ 算子确实已经具有描述世间一切复杂关系的能力了。人们曾经提出过一些 hypercomputer ，即超出图灵机范围的假想机器，比如能



二战时丘吉尔成立的英国密码学校原址（左），以及大厅里的图灵雕塑（右）。图灵曾在此协助破译德军密码

在有限时间里运行无穷多步的机器，能真正处理实数的机器。不过这在理论上都是不可能实现的。

事实上，图灵在他的论文中就已经指出，人的思维也没有跳出图灵机的范围。对此，图灵有一段非常漂亮的论证：人在思考过程中，总能在任意时刻停下来，把当前进度记录在一张纸上，然后彻底走开并把它完全抛之脑后，过一会儿再回来，并完全凭借纸上的内容拾起记忆，读取进度，继续演算。也就是说，人的每一帧思维，都可以完全由上一帧思维推过来，不依赖于历史的思维过程。而图灵机所做的，也就是把人的思维步骤拆分到最细罢了。

没错，这意味着，或许一个人的语言、计算甚至学习能力，完全等价于一个图灵机，只不过这个图灵机的指令集可能异常庞大。1950 年，图灵的另一篇经典论文 Computing Machinery and Intelligence 中正式把人和机器放到了相同的高度：让一个真人 C 先后与一台计算机 A 和另一个真人 B 进行聊天，但事先不告诉他 A 和 B 哪个是机器哪个是人；如果 C 无法通过聊天内容分辨出谁是机器谁是人，我们就认为计算机 A 具有了所谓的人工智能。这就是图灵测试。

计算机拥有智能？这岂不意味着计算机也能学习，也能思考，也拥有喜怒哀乐？人类似乎瞬间失去了不少优越感，于是不少科学家都旗帜鲜明地提出了反对意见。其中最为经典的恐怕要数美国哲学家约翰·塞尔（John Searle）在 1980 年提出的“中文屋子”思想实验了。把一个不懂汉语的老外关在一个屋子里，屋子里放有足够多的草稿纸和铅笔，以及一本汉语机器聊天程序的源代码。屋子外面则坐着一个地地道道的中国人。屋里屋外只能通过纸条传递信息。老外可以用人工模拟程序运行的方式，与屋外的人进行文字聊天，但这能说明老外就懂中文了吗？显然不能。每次讲到中文屋子时，我往往会换一种更具戏剧效果的说法。一群微软研究员在小屋子里研究代码研究了半天，最后某人指着草稿纸一角的某个数字一拍大腿说，哦，原来屋外的人传进来的是一段笑话！于是，研究员们派一个代表到屋子外面捧腹大笑——但是，显然这个研究员是在装笑，他完全不懂笑点在哪儿。这个例子非常有力地说明了，机器虽然能通过图灵测试，但它并不具有真正的智能。

当然，有反方必有正方。另一派观点则认为，计算机拥有智能是一件理所当然的事。这涉及到一个更为根本的问题：究竟什么是智能？

记得我曾经看过一本科幻小说，书名不记得了，情节内容也完全不记得了，只记得当我看完小说第一页时的那种震撼。在小说的开头，作者发问，什么是自我意识？作者继续写到，草履虫、蚯蚓之类的小动物，通常是谈不上自我意识的。猫猫狗狗之类的动物，或许会有一些自我意识吧。至于人呢，其实我只敢保证我自己有自我意识，其他人有没有自我意识我就不知道了。看到这里我被吓得毛骨悚然：完全有可能整个世界就只有我一个人有自我意识，其他所有人都是装出一副有意识的样子

的无生命物！

有一次做汉语语义识别的演讲时，讲到利用语义角色模型结合内置的知识库，计算机就能区别出“我吃完了”和“苹果吃完了”的不同，可以推出“孩子吃完了”多半指的是什么。一位听众举手说，难道计算机真的“理解”句子的意思了？我的回答是，没有冒犯的意思，你认为你能理解一个汉语句子的意思对吧，那你怎样证明这一点呢？听众朋友立即明白了。你怎样证明，你真的懂了某一句话？你或许会说，我能对其进行扩句缩句啊，我能换一种句型表达同样的意思啊，我能顺着这句话讲下去，讲出与这句话有关的故事、笑话或者典故，我甚至还能在纸上画出句子里的场景来呢！那好，现在某台电脑也能做到这样的事情了，怎么办？

这就是所谓的“功能主义”：只要它的输入输出表现得和人一样，不管它是什么，不管它是怎么工作的，哪怕它只是一块石头，我们也认为它是有智能的。永远不要觉得规则化、机械化的东西就没有智能。你觉得你能一拍脑袋想一个随机数，并且嘲笑计算机永远无法生成真正的随机数。但是，你凭什么认为你想的数真的就是随机的呢？事实上，你想的数究竟是什么，这也是由你的大脑机器一步一步产生的。你的大脑逃不出图灵机。

事实上，整个世界也逃不出图灵机的范围。牛顿系统地总结了物体运动规律后，人类豁然开朗，原来世界万事万物都是由“力”来支配的。扔出一个东西后，这个东西将以怎样的路线做怎样的运动，会撞击到哪些其他的物体，它们分别又会受到怎样的影响，这都是可以算出来的。这便是所谓的机械唯物主义：我们的世界是一个简单的、确定的、线性的、无生的世界。1814 年，法国数学家拉普拉斯（Laplace）给出一个更加漂亮的诠释：如果有一个妖精，它知道宇宙某个时刻所有基本粒子的位置和动量，那么它就能够根据物理规律，计算出今后每一时刻整个宇宙的状态，从而预测未来。刘慈欣在科幻小说《镜子》中更加极端地把初始状态取到宇宙大爆炸的时刻，因为宇宙诞生之初的状态极其简单，调整到正确的参数就可以生成我们所处的这个宇宙。这就是所谓的决定论。

我特别相信这些说法。我的拖延症有一个非常怪异的缘由，那就是我会告诉自己，截止的那一天总会到来的，这堆破事儿总会被我做完的。遇上纠结的问题，我不会做过多的思考，而会让一切顺其自然。其实，结果已经是确定的了，我真正需要做的不过是亲自把这个过程经历一遍。就仿佛我没有自由意志了一样。

不过，有了现代物理学的观念，尤其是量子理论的诞生，人们开始质疑上帝究竟会不会掷骰子了。然而，上帝会不会掷骰子，对于我们来说其实并不重要。图灵的结论告诉我们，即使未来是注定的，我们也不一定有一种算法去预测它，除非模拟它运行一遍。但是，要想模拟这个宇宙的运行，需要的计算量必然超出了这个宇宙自身的所有资源。运行这个宇宙的唯一方式，就是运行这个宇宙本身。塞思·劳埃德（Seth Lloyd）在 Programming the Universe 里说到：“我们体会到的自由意志很像图灵的停机问题：一旦把某个想法付诸实践，我们完全不知道它会通向一个怎样的结局，除非我们亲身经历这一切，目睹结局的到来。”

未来很可能是既定的，但是谁也不知道未来究竟是什么样。每个人的将来依旧充满了未知数，依旧充满了不确定性。所以，努力吧，未来仍然是属于你的。

作者简介：顾森，网名 matrix67，北京大学中文系应用语言学专业本科大四学生，数学爱好者，2005 年开办博客 matrix67.com，至今有上千篇文章，已有上万人订阅。曾任果壳网死理性派编辑，担任三年初中奥数培训教师，现兼任人人网产品部算法组技术人员。