机器智能何以可能：图灵的智能概念研究

The Possibility of Machine Intelligence: Research on Turing's Concept of Intelligence

邓克涛 /DENG Ketao　　张贵红 /ZHANG Guihong

（中国科学技术大学人文与社会科学学院，安徽合肥，230026）
(School of Humanities and Social Science, University of Science and Technology of China, Hefei, Anhui, 230026)

摘　要：图灵测试通常被理解为一种“可操作”或“行为主义式”的智能定义，能否通过图灵测试也被认为是判定机器是否具有智能的一种方式。事实上，这是一个误解，图灵从未声称通过图灵测试就意味着机器具有智能，尝试着对机器能否思维或能否拥有智能做出直接回应，并不是图灵测试的最初目的。通过研究图灵早期对智能概念的界定，可以发现自始至终图灵都是坚信机器智能是具有可能性的，并在图灵测试之前就已经设定了实现机器智能的潜在方案。

关键字：图灵测试　机器智能　图灵　智能

Abstract: The Turing Test is often interpreted as an “operational” or “behaviorist” definition of intelligence, and the ability to pass the Turing Test is considered a criterion for determining whether a machine possesses intelligence. In fact, this is a misunderstanding. Turing never claimed that passing the Turing Test implies that a machine has intelligence. Attempting to directly address whether machines can think or possess intelligence was not the original purpose of the Turing Test. By examining Turing's early conceptualization of intelligence, it becomes evident that he consistently maintained a firm belief in the possibility of machine intelligence, and he had already outlined potential approaches to achieving machine intelligence even before proposing the Turing Test.

Key Words: Turing test; Machine intelligence; Turing; Intelligence

中图分类号：TP18; TP242.6　DOI: 10.15994/j.1000-0763.2025.08.003　CSTR: 32281.14.jdn.2025.08.003

2022 年，OpenAI 公司正式发布 ChatGPT，一经上线就引发关注。在惊讶于 ChatGPT 能够表现出高水平智能等级时，不少人试图将其与图灵测试联系在一起。在过去，图灵测试通常被作为判定机器是否具有智能的一个重要依据。然而，在本文中，我们想表明这种解读既误解了图灵的本意，也在一定时期内对人工智能的研究产生了误导。事实上，图灵从未打算提供关于智能的明确定义。在他看来，智能与否既取决于物体本身的性质，也取决于观察者的知识水平。但批评者们常常忽略了这一点，简单地将图灵的智能概念归纳为行为主义式。基于此，本文试图澄清这种误解，并想表明考察图灵对机器智能可行性的分析和机器智能的实现方案才是理解他思想的关键。

一、机器智能的判定方式？被误解的图灵测试

1. 原始模仿游戏与标准图灵测试

基金项目：国家社会科学基金重大项目“人工智能伦理风险防范研究”（项目编号：20&ZD041）。
收稿日期：2023 年 8 月 13 日；**返修日期：**2025 年 2 月 27 日
作者简介：邓克涛（1999–）男，安徽凤阳人，中国科学技术大学人文与社会科学学院博士研究生，研究方向为人工智能哲学、新兴科技伦理。Email: dengketao@mail.ustc.edu.cn
张贵红（1982–）男，河北无极人，中国科学技术大学人文与社会科学学院副教授，研究方向为新兴科技伦理、科学哲学。Email: guihong@ustc.edu.cn

“机器能否思维”作为一个哲学研究领域经久不衰的议题，最早出现于图灵在1950年发表的著名论文“计算机器与智能”（Computing Machine and Intelligence）。在这篇论文的开始，图灵提出了这个问题。然而，接下来他并没有打算直接回答这个问题，因为他认为回答这个问题就意味着需要先确定“机器”和“思维”这两个词的的含义。相应地，他提议用模仿游戏来对这个问题进行重新表述。

模仿游戏是一个有趣且富有挑战性的推理游戏，其中三个角色扮演者被分配了不同的角色和任务。游戏中的三个角色包括一个男人（A）、一个女人（B）以及一个审讯者（C），他们分别处在不同的房间里。审讯者通过标签X和Y来认识另外两个人，他的任务是通过向X和Y提问，以在游戏结束时根据他们的答案来确定X和Y的性别。在游戏的过程中，女人的任务是协助审讯者，而男人的任务则是通过模仿女人的回答，以尽可能地误导审讯者做出错误的判断。为了排除声音这个干扰因素，游戏规定使用一个电传打印机作为通信工具。

现在由机器来充当A的角色，它能够像正常男人一样带给审讯者同样的困扰吗？图灵提议将这个游戏作为“机器能否思维”的替代。为了防止其他人认为模仿游戏毫无意义，图灵声称，任何的工程师或化学家都不能生产出一种与人类皮肤无法区分的材料，这就意味着我们无法制造出一种与人类身体完全近似的机器。而模仿游戏的意义就在于确定一条清晰的界限，以将人类的身体能力和智力能力分开讨论，这样就避免了将智力能力与身体能力密切相关的立场。他声称，“难道机器不能够执行一些应该被描述为思考，但与人类行为十分不同的东西吗？”[1]

斯特雷特（S. Sterrett）将上述的性别模仿游戏称之为“原始模仿游戏”，而在“计算机器与智能”中，还有另一个更加为人熟知的版本，即“标准图灵测试”（后文简称图灵测试）。[2]与“原始模仿游戏”不同，图灵测试的设定进行了如下改动：角色A由一台数字计算机担任，而角色B则由一个人类玩家担任。换句话说，人类的任务转变为协助审讯者，而数字计算机的任务则转变为通过模仿人类来误导审讯者。尽管在这之后，图灵测试有过多个变体，但在“计算机器与智能”原文中，图灵仅介绍了这两个版本。

图灵表示，在五十年之后，通过修改特定的数字计算机，使其拥有足够的存储能力和运行速度，并提供一个合适的程序，它可以满意地完成图灵测试。具体来说，它可以做到在五分钟的对话后，让不超过70%的普通审讯者能够做出正确的识别。[1]而在1952年，图灵又表示，一台机器想要真正通过他的测试至少需要一百年。（[3]，p.504）但无论如何，我认为我们不必过分关注有关图灵测试的具体内容，因为图灵自己曾声称，他的目的并不想表明一台机器是否真正能够通过图灵测试，而是认为我们应该讨论这样的话题，即“有没有可以想象到的数字计算机能够在模仿游戏中做得更好？①”他表示科学家们并不是完全不受任何改进猜想的影响，从一个广泛被接受的事实发展到另一个广泛被接受的事实。猜想有时非常重要，因为它们往往能够指明有用的研究方向。

2. 被误解的图灵测试

自“计算机器与智能”发表以来，有关图灵测试的讨论与争议就一直存在于各个学科领域。支持者认为图灵的论文标志着人工智能的开端，而通过图灵测试就是人工智能的终极目标。批评者认为图灵测试是根本不可能实现的，甚至还可能误导人工智能的研究。介于两者之间的是对意识、行为主义、智能的操作式定义等方面的探讨。[4]在哲学研究领域，行为主义和归纳主义这两种经典解释曾被广泛接受。然而，这两种解释往往基于对“计算机器与智能”一文的解读，并未广泛地覆盖图灵以往有关机器智能的作品。因此，在一定程度上，它们是

①值得注意的是，该表述属于技术命题，而原始表述“机器能否思维”属于哲学命题。因此，从这里可以发现，图灵想讨论的核心是如何在技术层面实现机器智能，而不是在哲学层面肯定机器智能。

片面的和不充分的，也是图灵测试为什么招致诸多批评的重要原因之一。

行为主义式的解读认为图灵测试旨在提供有关智能或思考的可操作定义。根据这种观点，智能是对一系列言语刺激后产生合理的言语反应的倾向。[5] 归纳主义式的解读将“物体通过图灵测试”和“该物体是一个智能实体”之间的关系视为归纳推理关系。[6] 也就是说，机器通过图灵测试是机器智能存在的一个归纳性证据。这种观点认为图灵测试的真正价值不在于将其作为可操作定义的基础，而是把它作为机器智能存在的良好归纳证据的潜在来源。

尽管这两种观点存在差异，甚至在一定程度上构成了对立面，但它们都不约而同地将图灵测试作为判定机器是否智能的一种手段。然而，图灵从未声称机器通过他的测试就意味着机器能够思维或具有智能。相反，他认为这个问题“太无意义了，不值得去讨论”。[1] 关于“思维”的定义，图灵曾表示他并不想给出一个明确的定义，但如果非要给的话，那么“只能说它是一种在我头脑中持续发生的嗡嗡声”。（[3]，p.494）他认为我们并不需要就这个词在定义上达成一致，重要的是要试图划清我们想要讨论的大脑或人的属性与我们不想讨论的属性之间的界限。

考虑到图灵所处的时代，机器能够思维或具有智能的观点极其容易受到神学、沙文主义等立场的强烈反对，至少这种观点与常识和直觉十分不符合。似乎一直以来，思维或智力活动都被视为是哺乳动物大脑的专有活动。然而，如果我们设想这样一个场景：一个外表和正常人类没有任何差异的外星人，他的脑袋里是一团气体或者液体，他能够做出很多与正常人类相同的活动和决策，并且在与其他人类的对话中，能够做到在五分钟内让不超过30%的人觉得有异常。那么，我们是否会认为该外星人可以思维或具有智能呢？

如果我们仅将智力活动看作是哺乳动物大脑的专有活动，那么试图为机器智能辩护所付出的任何努力都将是徒劳的。即使机器表现出与人类相同甚至更高水平的智能，仍然会有诸多类似“机器不拥有X”或“机器不能X”的声音来对机器智能进行反对。因此，探讨机器智能的一个必要前提是要在身体能力和智力能力之间画出清晰的界限，以使得机器在智力方面能够与人类进行公平竞争，而图灵测试正是这样的一种积极的尝试。

回到最初的问题，如果图灵的初始立场是通过图灵测试就意味着机器能够思维或具有智能的话，那么在“计算机器与智能”一文中，就应该至少花一节来论证二者之间的等价性。然而，遗憾的是，纵观全文并没有出现这样的一节，甚至有关图灵测试的内容也只是浅浅几笔①。事实上，在1948年的时候，图灵在“智能机器”（*Intelligent Machinery*）一文中，就已经尝试通过详尽的设计来实现机器智能。而在之后发表的“计算机器与智能”中，其主要内容在于反驳与机器智能相反的观点和延展之前的设计工作。图灵测试的真实目的也并不在于提供有关智能的定义或作为机器智能的一个判定依据，正如人工智能代表人物之一的明斯基（M. Minsky）所设想的那样，图灵只是将他的测试“作为评估机器的一种方式，但他从未打算将其作为决定机器是否真正智能的方法”。[7]

二、图灵的智能概念：智能是一种情感概念

对图灵测试的传统解读使得多数人将图灵的智能概念归结为行为主义式的，即将智能或思维视为是行动的能力或倾向。然而，图灵并不是行为主义者。第一，图灵从未将声称机器通过图灵测试就意味着机器能够思维或具有智能；第二，图灵测试能否通过的判定依据并不是机器的行为是否可以做的像正常人类那样

① “计算机器与智能”共计七节，分别是模仿游戏、对新问题的批判、在游戏中所关心的机器、数字计算机、数字计算机的普遍性、对主要问题的相反观点、学习机器。

好，而是机器是否能够欺骗足够多数量的人类评审官。根据图灵的说法，智能本身是情感的而非数学的，“我们对事物的智能方式的理解程度，既取决于我们的心理和训练状态，也取决于所考虑对象的属性”。（[8]，p.431）也就是说，智能与否除了事物本身需要具备某种智力属性之外，还需要考虑到观察者的知识水平或能力。

根据普劳德富特（D. Proudfoot）的说法，图灵的智能是一种响应依赖型（response-dependence）的概念。即某个行为是智能的，当且仅当在正常条件下，该行为对于正常受试者来说似乎是智能的。[9] 响应依赖型的概念还包括颜色、美丽以及善良等，这类概念依赖于或对应于某些主体在特定条件下的心理反应。也就是说，智能并非是独立于物理世界的一个特征，它需要考虑到正常受试者对它的态度或响应。类似地，丹齐格（S. Danziger）认为图灵将智能视为一种社会概念，即智能与否是由社会对实体的态度决定的。[10] 这两种解释都具有一定的合理性，因为图灵曾声称参与审讯的主体应该是普通的，而不应该是机器方面的专家。如果由机器专家来担任审讯者的角色，那么机器的性能无论多么“出众”，它们也不可能通过测试。

图灵声称，“如果我们能够解释和预测它的行为，或者如果没有什么潜在的计划，那么我们就没有什么想象智能的诱惑。”（[8]，p.431）从这个角度来看，一个事物或行为被视为智能的前提条件是观察者不能完全清楚其内部逻辑，并且也无法准确地解释和预测其接下来的行动。事实上，人们总是倾向于将智力或思维活动视为是一种神秘且无法理解的心理过程。在与布雷斯韦特（R. Braithwaite）、杰斐逊（G. Jefferson）和纽曼（M. Newman）的一次谈话中，图灵表示“一个人一旦看到因果关系在大脑中自行产生，那么他就会认为这不是在思考，而是一种缺乏想象力的苦活。”（[3]，p.500）似乎一直以来，图灵都在尝试摒弃只有人类才可以能够思维或具有智能这一立场，那么他又是如何为机器智能辩护的呢？

对机器智能的一种强烈反对来自于“机器时常会给出错误的答案”这一观点。毫无疑问，即使是时下最为火爆的ChatGPT有时也会向用户返回一些与对话背景不相关或错误的输出。然而，仅凭这一点就可以否认机器智能的可能性吗？根据图灵的说法，“如果一台机器被认为是万无一失的，那么它就不可能是智能的”。（[11]，p.394）对于机器所犯的错误，图灵认为我们不应该过分重视它，因为在日常生活中，每个人都具有潜在犯错的可能。在超出有效程序范围内尝试一种新技术时，即使是训练有素的数学家也难免会犯错误。因此，期待一台机器不犯错显然是不现实的，仅凭此来否认机器智能也是不足够的。

洛夫莱斯伯爵夫人（Countess of Lovelace）曾以“分析引擎不会试图创造任何东西。它只能执行我们已知的命令”来作为机器智能的反对。[13] 在她的视角里，机器不会产生任何新的事物，机器的智能来源于人类开发者。类似的观点是机器所采用方法都是由其操作者预先安排好的，只要存储量和执行速度允许的话，那么不过是打字猴子的改进，在几个世纪里偶然写出了《哈姆雷特》。然而，如果说机器能够产生人类开发者未能预料的结果，并且该结果在事后被证明是有价值的，那么这种说法还仍然成立吗？就现如今的信息技术而言，一些算法可以从海量的数据集中挖掘出其中潜在的关联和趋势，并能以此指导用户的决策或行为。那么在这种情况下，我们是否可以认为机器是具有智能的呢？

关于这一点，图灵建议将机器的开发过程看作是教师与学生之间的教学活动。在这个教学活动中，人类开发者充当教师的角色，机器则充当学生的角色。在具体的实践中，我们并不能够总是确定机器的行为。如果说机器做了一些未被编程的行为或发现了某些问题的新模式，那么我们必须承认机器已经产生了一些新的东西。图灵将这个过程比作一个学生从老师那里学到了很多东西，但通过自己的工作又增加了很多东西。在这种情况下，人们就应该承认机器做出了自己的选择和决定，也有义务认

为机器展示出了智能。（[11]，p.393）

综上所述，不难发现，图灵自始至终都坚持思维或智力活动并不是人类所独有的。考虑到机器在某些方面所具备的潜力，对于机器智能的设想并非是毫无依据的。从图灵对智能的定义来看，智能似乎更倾向于创造性、不可预测性而非行为表现与人类一致性，故基于图灵测试作为图灵对智能的定义的讨论或批判，从根本上来说都是不成立的。至于图灵为什么要提出图灵测试来确保机器可以与人类在智力层面公平竞争，或许是为了引发人们对人类智能的思考，以悬置对机器智能的默认反对立场；或许是为了动摇人们的直觉观念，以使得人们让步于机器在智力活动方面具有巨大潜力这一个事实。尽管图灵并没有明确表示自己秉持哪一立场，但至少这两者都比将其简单等同于对智能的定义更加具有说服力。

三、机器智能何以可能？图灵对机器智能的可行性分析

关于对人类智能的理解，存在两种解读方式：其一，它像是金钱，尽管不同的人之间存在着数量的差异，但整体上而言，它是人人都具有的一种东西；其二，它像是财富，也就是说当某个人至少拥有超过平均水平的时候，我们才能说这个人具有它。[13] 显而易见的是，图灵对人类智能的理解采用了第二种方式，从这个角度来看，机器智能在逻辑上就拥有了可能性。当然，仅凭此还不足以证明机器智能的可行性，就像布洛克（N. Block）所设想的那台“无智能机器”（unintelligent machine）一样①，在逻辑上能够成立的猜想并不意味着在现实中能够成立。

在1947年的伦敦数学学会上，图灵发表了关于自动计算引擎（Automatic Computing Engine, ACE）的演讲。在这次演讲中，图灵首次公开分享了他对于机器智能的猜想。他表明我们所需要的是一台可以从经验中学习的机器。（[11]，p.393）似乎在图灵看来，智能最本质的特征就是学习能力。在分析人类大脑的特定活动时，图灵声称，人类大脑的许多部分都是为了完成特定功能而需要的神经回路，其中的例子包括控制呼吸、打喷嚏、以及追踪移动物体等的“中枢”：所有真正的反射动作（非条件反射）都是由大脑中这些明确结构的活动所引起的。（[8]，p.423）但是，在大脑中，更多过于多样化的智力活动无法仅凭这种基础来进行管理。考虑到不同地区人口语言的差异性，其产生的根本原因并不在于大脑相关部分的功能差异，而在于童年时期所接受教育的不同。那么也就是说，后天所接受的教育才是人类智能的关键来源。类似地，图灵奖和诺贝尔经济学奖双料得主西蒙（H. A. Simon）也曾表示，“随着时间的推移，我们行为的表面复杂性在很大程度上是对我们所处环境复杂性的反映。”[14] 由此，我们也就便有理由相信，一台经过了一定时间干预或训练的机器，也能够在某些特定的活动中表现出足够的智力水平。

图灵还从技术上分析了机器智能的可行性。他认为“提供适当的存储是解决数字计算机问题的关键，当然，如果要说服他们表现出任何真正的智能，就必须提供比目前更大的容量”。（[11]，p.383）不过，尽管人脑的记忆容量可能达到一百亿个二进制数字的量级，但在做某项特定活动时，所需要的存储量可能只需要使用其中的几百万个数字。也就是说，至少在合适的情况下，一台已经设定了一些初始指令表的机器，可以有效地模仿人类在从事某些活动时大脑所发生的思维过程。机器的存储容量往往决定了机器在执行复杂任务时的表现。尽管当时机器存储能力与如今相比有非常大的差距，但图灵的话似乎暗示了当时的存储能力已经能够支持机器表现出一定的智力水平。在反驳将“像机器一样行动”视为缺乏适应性这一观点时，图灵表示“过去的机器存储空间很少，而且机器也没有任何自行决定权”。（[11]，p.393）换句话说，之前的机器受限于存储空间

① “无智能机器”指的是一台能够对口头刺激作出合理言语回应的机器。该机器内部存储了所有合理的对话字符串。

和缺乏自主性，导致其难以从经验中学习，因此会被视为缺乏变通的代表。这同时也表明通过提供足够的存储容量和算力，机器在合适的情况下就能够展现出较高水平的智能。这似乎也能够解释为什么他相信“在本世纪末，词汇的使用和受过一般教育的观点将会发生如此巨大的改变，以至于人们能够谈论机器的思考时，而不期望被反驳”。[1]

尽管在当时，技术上的限制使得图灵还无法真正制造出一台能够让绝大多数人认为是智能的机器，但他还是不遗余力地试图证明机器智能的可行性。当然，图灵对机器智能的探索不可能仅仅停留在1947年的那场演讲中，在随后的几年内，他逐渐转向了对机器智能的具体设计。

四、机器智能的实现方式：拥有纪律和主动性的机器

至于如何设计出具有智能的机器，图灵表示我们必须得使机器同时获得纪律和主动性。关于纪律，此前图灵曾表示“一个有纸、铅笔和橡皮并受到严格纪律的人，是一台通用机器”。（[8]，p.416）那么，也就是说，纪律是一种有效计算的方法或程序，“把一个大脑或机器变成一个通用机器是最极端的纪律形式”。（[8]，p.429）他声称没有纪律就无法建立起适当的沟通，但仅凭纪律还不足够产生智能，额外必要的东西他称之为主动性。关于主动性，他表示我们的任务就是“发现这种在人类中出现的残余的本质，并尝试在机器中进行复制。”（[8]，p.429）

图灵提倡可以采用以下两种方法，来使机器同时具备纪律和主动性：

第一，将主动性移植到一台通用实用计算机器（Universal Practical Computing Machine, U.P.C.M.）中；

第二，将纪律和主动性同时移植到一台无组织机器中。

1. 将主动性移植到一台U.P.C.M.中

U.P.C.M.不仅可以做任何被描述为“经验法则”或“纯机械法则”的事，并且还可以快速访问存储器中的特定信息。也就是说，它本身就具备了纪律，为了使其达到智能要求，还需要进行主动性的植入。图灵表示主动性的植入其中一种可能的方法是采用机器编程的形式，通过编程使得机器做出越来越多的选择或决策。当我们可以使用少量的通用原则来指导机器的行为，而不需要为每种情况都编写详尽的指令时，我们就可以说机器已经“长大成人”了。（[8]，pp.429–430）

在“计算机器与智能”一文中，图灵似乎继续沿用了这种模式。他表明机器想要在模仿游戏中表现出色关键在于编程问题和工程问题。图灵将成年人的思维状态描述为三个部分：最初的精神状态、它所接受的教育以及其他不能被描述为教育的经验。他声称“与其尝试制作一个模拟成年人思维的程序，为什么不尝试制作一个模拟儿童思维的程序呢？”[1]模拟儿童大脑所需要的机械构造相对较少，在这种情况下，可以很容易地完成编程工作。那么，我们的任务就可以划分为两个部分：制造一台儿童机器和对它进行相关的教育。

在制造一台儿童机器的过程中，我们并不一定要求机器的配置完全和人类相同。图灵举了凯勒（H. Keller）的例子，并表明“只要教师和学生之间可以通过某种方式进行双向交流，教育就是可能的。”[1]在对儿童机器进行教育的过程中，我们可以合理地采取奖惩机制，以使机器的行为符合特定的目的。具体而言，机器在进行某项事件或行为时，当其受到奖励信号之后，会增加该事件重复发生的概率；相反，当其受到惩罚信号之后，则会减少该事件重复发生的概率。当然仅凭奖惩机制还不足以完全覆盖整个教学过程，我们还需要一些“非情感”的交流渠道，这样就能够“通过惩罚和奖励来教导机器以某种语言（例如符号语言）服从给定的命令”。[1]

这种方法依赖于与环境的交互学习，通过对U.P.C.M.进行教育和经验的累积，以使其逐渐达到成年人的智力水平。与此类似，行为主义人工智能范式基于控制论的思想，试图通过

模拟生命的自适应机制来演化出类似于人类大脑的智能系统。[15] 在这种范式下，机器通过与环境的交互和反馈来学习和调整自身的行为，不断优化其表现和适应性，从而能够自主地感知、理解和应对复杂情境。

2. 将纪律和主动性同时植入到一台无组织机器中

无组织机器指的是在构造中基本随机的机器。对于一个无组织机器，可以设想它是由大量以下装置组成的：每个单元都有两个输入端，并有一个输出端，其中输出端可以与其他单元连接并同时作为该单元的一个输入端。机器的组成单元在每个时刻处在状态0或状态1，并且单元的状态遵循如下的运算规则：取出前一时刻输入端的状态，将两者相乘后的结果再从1中减去，即为该单元的最后状态。

对于在神经系统中最简单的模型，图灵称之为A型无组织机器（图1）。（[8]，p.417）如果单元r(1)、r(2)、r(3)、r(4)以及r(5)的初始

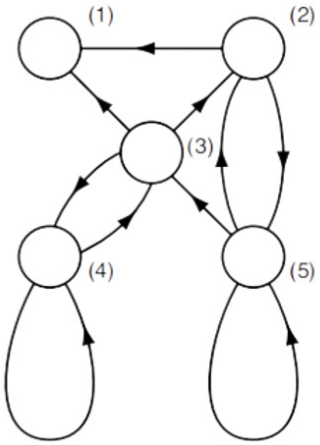

图1 A型无组织机器

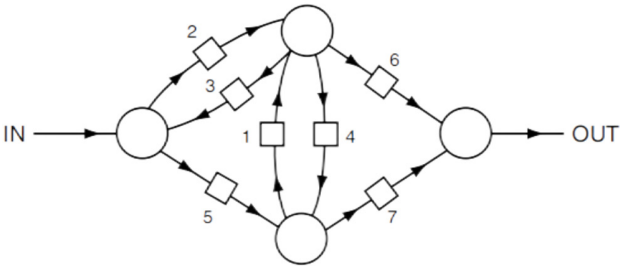

图2 B型无组织机器

序列分别为1、1、0、0、1，那么按照上述运算规则，在第二时刻它的序列就变成了1、1、1、1、0，以此类推可以得到一个连续的数列。当然，该序列是有周期性的，其周期所包含的时刻最多不会超过2^5。在对A型无组织机器其中的每个单元进行替换之后，则可以得到B型无组织机器（图2）。（[8]，p.422）在B型无组织机器中，各单元之间添加了一个连接修改器（connection-modifiers），连接修改器有中断和通过两种模式。在通过模式下，修改器的输入与输出相同；在中断模式下，修改器的输入与输出相反①。对于具有可靠单元的B型无组织机器，我们可以满足初始条件，使其成为具有给定存储容量的通用机器。按照上述运算规则，对B型无组织机器进行适当的限制干扰，那么机器就会表现得像是为某个明确目的而组织起来的。

根据科普兰（J. Copeland）的说法，图灵很有可能是第一个考虑将类似神经元的简单元素连接成网络来构建计算机的人。[16] 现代联结主义者往往将赫布（D. Hebb）和罗森布拉特（F. Rosenblatt）视为联结主义的奠基人，但他们很少能够意识到图灵早在1948年就为联结主义提供了理论基础。不过遗憾的是，图灵当时的雇主——伦敦国家物理实验室主任——查尔斯·高尔顿·达尔文（C. G. Darwin）爵士并没有认可图灵的想法，甚至将“智能机器”一文视为“学童作文”，这也间接导致该论文没有公开发表，进而影响了人工智能数十年的进程。

结　语

自二战结束以后，图灵就试图通过使用机器来模仿人类大脑的思维过程。在他看来，我们不应该停留在“机器能否思维”这个哲学命题上。相反，我们更应思考的是如何在技术层面实现一台能够被公众认为是智能的机器。然而，受限于当时人们狭隘的认知和对机器的偏

①中断模式下，连接修改器状态为1。如果输入端初始状态为1，那么经过连接器后，状态变为1–1*1，即为0；如果输入端初始状态为0，那么经过连接器后，状态变为1–1*0，即为1。通过模式下，连接修改器状态为0，以此类推。

见，致使他们本能地排斥将机器与智能联系在一起。图灵显然意识到了这一点，于是才提出图灵测试来确保机器能够就智力层面与人类形成公平竞争。图灵测试的真实目的并不在于提供有关智能的定义，通过图灵测试也不应该被视为人工智能的终极目标。

通过考察图灵对智能概念的界定，可以发现图灵并没有将智能视为是有机体的专有属性。在判断事物或行为是否具有智能时，既需要考察其本身的属性，也需要考虑到观察者的知识储备。因此，就会出现对于同一种事物或行为，一些人认为它是智能的，而另外一些人则认为它不是智能的。这意味着智能并不是完全独立于人类心灵的固有特征，它还依赖于观察者对它的响应或态度。从图灵对机器智能的实现方案中，可以发现智能最本质的特征就是学习能力。当然，在机器的学习过程中，离不开与环境的交互、经验的累积以及通过反馈修正自身活动等。儿童机器和B型无组织机器分别展示了图灵对行为主义人工智能范式和联结主义人工智能范式的勾勒，这似乎也能够印证为什么图灵被誉为“人工智能之父”。

回到当下，伴随着大数据、自动驾驶以及生成式人工智能的兴起，我们已然难以否认机器在许多领域已经展现出足够甚至超越人类的能力。在未来，机器将不可避免地在人类社会中扮演更加重要的角色。当然，随之而来的是许多重大的社会与伦理问题。尽管这种趋势难以逆转，但我们仍可以做一些前瞻性的工作，而不是将目光聚集在毫无意义的探讨之中。正如图灵所说的那样，“尽管我们的视野只能望向有限的前方，但我们却能够发现那里存在着大量需要完成的任务。”[1]

[参考文献]

[1]Turing, A. M. 'Computing Machinery and Intelligence'[J]. *Mind*, 1950(59): 433–460

[2]Sterrett, S. G. 'Turing's Two Tests for Intelligence'[J]. *Minds and Machines*, 2000, 10(4): 541–559.

[3]Turing, A. M., Braithwaite, R., Jefferson, G., et al. 'Can Automatic Calculating Machines Be Said To Think'[A], Copeland, B. J. (Ed.) *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life Plus the Secrets of Enigma*[C], New York: Oxford University Press, 2004, 494–506.

[4]Saygin, P. A., Cicekli, I., Akman, V. 'Turing Test: 50 Years Later'[J]. *Minds and Machines*, 2000, 10(4): 463–518.

[5]Block, N. 'Psychologism and Behaviorism'[J]. *The Philosophical Review*, 1981, 90(1): 5–43.

[6]Moor, J. H. 'The Status and Future of the Turing Test'[J]. *Minds and Machines*, 2001, 11: 77–93.

[7]Marvin, M. 'Marvin Minsky on AI: The Turing Test is a Joke!'[EB/OL]. www.singularityweblog.com/marvin-minsky/. 2023–08–13.

[8]Turing, A. M. 'Intelligent Machinery'[A], Copeland, B. J. (Ed.) *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life plus the Secrets of Enigma*[C], New York: Oxford University Press, 2004, 410–432.

[9]Proudfoot, D. 'Rethinking Turing's Test and the Philosophical Implications'[J]. *Minds and Machines*, 2020, 30(4): 487–512.

[10]Danziger, S. 'Intelligence as a Social Concept: A Socio-Technological Interpretation of the Turing Test'[J]. *Philosophy and Technology*, 2022, 35(3): 1–26.

[11]Turing, A. M. 'Lecture on the Automatic Computing Engine'[A], Copeland, B. J. (Ed.) *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life Plus the Secrets of Enigma*[C], New York: Oxford University Press, 2004, 378–394.

[12]Countess of Lovelace. 'Translator's Notes to an Article on Babbage's Analytical Engine'[J]. *Scientific Memoirs*, 1842, (3): 691–731.

[13]Dretske, F. 'Can Intelligence Be Artificial'[J]. *Philosophical Studies*, 1993, (71): 201–216.

[14]Simon, H. A. *The Sciences of the Artificial*[M]. Cambridge: The MIT Press, 1996, 53.

[15]成素梅. 人工智能研究的范式转换及其发展前景[J]. 哲学动态, 2017,(12): 15–21.

[16]Copeland, B. J., Proudfoot, D. 'On Alan Turing's Anticipation of Connectionism'[J]. *Synthese*, 1996, (108): 361–377.

[责任编辑　王巍　谭笑]