# 人工智能中对符号含义的争论

路卫华

机器能否像人类一样思考和理解？此问题已困扰人工智能研究者多年。人工智能自从 20 世纪 50 年代诞生以来，便一直试图模拟人脑的表现，并假定计算机和人脑都只是信息加工机器。随着数字计算机的出现，该问题被进一步表述为，是否只要能运行适当的程序，计算机就可以思考和理解？对此的肯定回答被称为强人工智能立场，它源于纽厄尔（Allen Newell）和司马贺（Herbert Simon）等人工智能研究的先驱，他们认为智能存在于物理符号系统中。当塞尔（John Searle）提出中文房间论证以质疑该立场后，便引发了学界对意识、理解与计算本质等问题的反思。其后，哈纳德（Stevan Harnad）在此背景下提出了符号接地问题，并将其作为对更一般的计算主义立场的反驳，计算主义认为心理状态等价于与其实现无关的计算状态。塞尔后来也认为，中文房间论证同样适用于对计算主义的反驳。

## 中文房间论证反对强人工智能

具体而言，早期人工智能研究者认为计算机程序本身就是一符号系统，该符号系统是由可以基于明确规则（即句法）操作的任意“物理对象的个例”（即物理符号）组成。基于对人脑与计算机的类比，他们便以符号系统来理解认知，并假设认知主体的思考、推理或语言运用就是进行符号操作。这便是主张“心智是符号系统，认知是符号操作”的心智的符号模型。20 世纪 70 年代形成的认知科学遵循着与心智的符号模型相同的假设。这些观念构成了称为计算主义（又称认知主义）的基本原则，计算主义主张思维是心智的核心功能，可以通过遵循明确规则所进行的符号操作来说明思维。由此，计算主义便有三个要素：表征、形式符号系统和基于规则的变换。

不过，塞尔于 1980 年提出了“中文房间论证”的思想实验，以证明计算机所进行的纯句法符号操作不会产生出含义。该思想实验可表述为：想象在某房间里有一个人，此人可透过墙上的狭缝获得写有中文字

符串的纸片。此人并不懂中文，因而不能理解纸上的字符是什么含义。不过，借助于能为每个可能的中文输入字符串指定相应正确回复的规则手册，他可以对递进来的纸片给出一个正确的中文回复，并将其通过狭缝传递给外界。对于房间外的母语为中文的人而言，看起来他们正在和另外一位母语是中文的人通过书面中文会话。

现在的问题是房间里的人是否理解到了什么？塞尔认为房间里的人什么都没有理解，因为这个人只是遵循着规则手册中的机械指令，模拟出了一个母语为中文的人的行为。而他按照规则手册进行的操作，本质上与数字计算机运行的程序操作别无二致——接收中文符号输入，纯粹根据这些符号的形状来操作它们，最后给出中文符号输出。因为计算机程序完全是句法的，只是由符号串组成。符号的形状是任意的（与其含义或内容无关），推理的规则（对符号的组合和重组规则）本身也是任意的符号串。显然，我们不会在此情形下认为房间里的人理解了中文，因而也不能说计算机能够理解中文，即数字计算机无法通过基于执行某个适当的程序来理解中文。所以，虽然适当编程的计算机可能会以自然语言与人进行会话，但是它们并不理解所用符号结构的语义内容。

进一步说，中文房间论证的主旨在于系统的中央符号操作器（房间中的人）在并不理解中文的情况下，如何说明该系统（整个房间）能够理解中文。可将其论证归结为：因为句法操作不是系统获得含义的充分条件，并且计算机只能进行句法操作，所以计算机不能获得含义。由此，塞尔宣称，再多的句法也不能产生出语义来，所以计算机无法通过运行某个程序（纯句法的）进行理解。另外，人的心智既是句法的又是语义的，即人的心智既使用符号，同时也赋予符号以含义（语义）。因此，人的心智才能够进行理解。所以强人工智能立场是错的。

### 符号接地问题质疑计算主义

在塞尔论证的基础上，哈纳德进一步追问，一个系统的中央符号操作器仅仅是句法性的，该系统如何获得它所使用的符号的含义呢？常规的基于规则的人工智能系统由某些特定的表征（符号）结构组成，这些符号的含义最终由程序设计者所赋予，即它们的含义衍生于设计者的头脑中。也就是说，这样的符号结构自身是没有内在含义的，而仅有衍生的含义。

通常认为构成符号系统的符号与其相对应的含义既不相似，也不具有因果关联，符号的物理形状和句法属性，并不会对其所对应的语义值提供线索，后者与前者的对应完全是任意的。符号的含义只是其使用者们达成的记号性约定的一部分。传统的人工智能系统操作着被系统地解释为其所意指事物的符号，但解释是由外部解释者的心智做出的，而不是符号系统固有的。或者说，系统本身不知道符号表征的是什么，它们的含义完全取决于外部的解释者。所以，一个人工认知主体，如机器人，似乎无法获得它可以成功地在句法上进行操作的符号的含义。

哈纳德将符号接地问题具体表述为：一形式符号系统的语义解释如何内化于该形式符号系统中，而不仅仅是寄生于我们头脑中的含义上？仅凭其（任意）形状而加以操作的无含义的符号个例，其含义如何能接地到其他无含义的符号之外的事物上？这里所说的“我们头脑中的含义”是指由符号设计者指派给符号的语义。简言之，符号接地问题可以被看作如何使得语义不需要第三方解释者，或者说如何使得语义内化于符号系统。

只有生物认知系统才具有始源意向性，并且它们是对具有衍生意向性的实体赋予含义的唯一来源。人工系统（书籍、计算机和机器人）只有衍生意向性，因而它们永远不会以始源的方式获得符号的含义。可以说，哈纳德的符号接地问题是在中文房间论证的基础上对强人工智能与更普遍的计算主义立场的质疑。符号接地问题的意义不仅限于人工智能领域，因为它要对含义如何能在自然中产生，给出一个令人满意的说明，即要给出一种自然化的语义学。对于一些人而言，符号接地问题引发了对人工智能的怀疑，并认为该问题表明认知科学不能说明含义的出现。

**符号含义需基于认知主体**

当然，也有哲学家并未被中文房间论证与符号接地问题说服。丹尼特（Daniel Dennett）认为，根本就没有非衍生的意向性，因此，始源意向性也就无从说起；斯洛曼（Aaron Sloman）则认为，符号接地问题基于概念经验论预设，而概念经验论却是错误的。此外，也有哲学家虽未否认存在非衍生的意向性，但却反对在基于符号的人工智能中存在任何符号接地的需要，例如，康明斯（Robert Cummins）认为，因为

心理表征是非符号性的，所以没有必要保证它们具有非衍生的含义。在他看来，关于符号接地问题的整个讨论都是被误导的。

笔者认为，中文房间论证的确揭示出了心智的符号模型不能较好地说明心智是如何进行理解的。不过，哈纳德进一步提出的符号接地问题却存在基本预设上的错误，导致该错误的线索可以追溯至纽厄尔和司马贺等所说的“物理符号”概念。符号的本质在于使用者的约定，因而符号都是社会建构的而非物理的，所以“物理符号”这一概念本身就是不恰当的。斯洛曼与康明斯对符号接地问题的批评是中肯的。在符号接地问题的整个表述框架中，存在着一些严重的误解，如对符号与表征的基本理解及二者之间关系的理解。特别是，哈纳德及其追随者提出的对符号接地问题的解决方案，是将符号接地于系统对符号所表征的对象的识别和操纵能力上。也就是说，对于一个将符号系统与感觉运动系统相联合的自主系统，若它可以将其内部符号可靠地关联到它们所指的外部对象上，其内部符号便是接地的。此观点忽略了符号与其所指物之间关联的拟制性，即符号的含义并非直接来源于外部对象，而是源于认知主体对符号系统本身的理解与想象。

**（本文系中国博士后科学基金资助项目与国家社科基金项目“认知科学对当代哲学的挑战”（11&ZD187）阶段性成果）**

**（作者单位：中国人民大学哲学院）**