

从意向性看人工智能的发展方向

毛华滨, 高新民

(华中师范大学政法学院, 湖北 武汉 430079)

摘要: 现时代的人工智能由于对人类智能的模拟停留在功能操作层面, 而没有人所特有的意向状态。文章在总结现当代意向性研究成果的基础上, 指出意向性理论可以为今后人工智能的发展提供新的途径和方向。

关键词: 人工智能; 意向性; 思维模拟

中图分类号: N031 **文献标识码:** A **文章编号:** 1003 - 5680(2005)05 - 0030 - 04

“电脑能否代替人脑”, “人类能否沦为机器的奴隶”, “人心是否永远会胜过计算机”? 这是心灵哲学家和人工智能专家争论了半个多世纪的谜题。它之所以引起人们极大的兴趣, 一个原因是由于这个问题直接与“究竟什么是人、什么是心”, “人心的本质特征何在”这一类古老问题相关。更重要的原因是, 过去几十年里人工智能取得了巨大成就, 在很多领域开始取代人类智能, 但人工智能在一些对于人类来说极为简单的工作中却屡屡失败。当下的人工智能与人类智能究竟差别在何处? 人工智能能否跨越通向人类智能的鸿沟? 考察当代心灵哲学对意向性问题的研究, 或许可以给我们一些启示。

一 人工智能与人类智能的根本差异

人工智能自 20 世纪 50 年代初创以来, 取得了许多惊人的成就, 在许多方面超过了人类智能, 如惊人的计算速度和记忆力、不知疲劳、不会出现粗心的错误、不受外界干扰等。1997 年计算机“深蓝”以 2 胜 1 负 3 平的成绩打败了人类象棋冠军卡斯帕罗夫一事, 更是轰动一时, 震惊全人类。但是客观地看, 人类智能又有人工智能不可企及的地方, 如正确理解人的日常言语行为, 把握图形、符号的意义, 面容识别, 在日常环境中决定自己该做什么, 等等。甚至从本质特征上看, 人工智能根本还算不上真正的智能。我们先来看一些思想实验和例子。

机器能具有人的智能吗? 早在 1950 年著名的数理逻辑学家图林 (Alan Turing) 就曾在一篇文章中预言, “我相信, 大

约再过 50 年, 词语的运用和一般受教育的观点将会发生如此大的变化, 以至于人们能够说机器思想, 而不会被看作是矛盾的”。他提出了一个著名的思想实验即“图林测试”。实际上, 图林测试是一种想象的游戏、“模仿游戏”。在这个游戏中, 有三个参加者, 一部图林机或理想的计算机放在一间房子中, 并与房子外面的一部电传打字机相连接; 在房子里还有一个人也是如此, 图林机和这个人可以通过这台电传打字机接收到外面的信息 (输入), 同时也可以通过这台电传打字机向屋外输送信息 (输出); 在屋外还有一个人即提问者, 他把问题送进那个房间。这些问题有时由那台计算机来回答, 有时由那个人来回答, 他 (它) 们都躲避那些回答不出来的问题, 并且在回答问题时没有任何语音和外貌的提示。图林认为, 如果一个计算装置即数字计算机能够和人一样蒙混过室外的测试者, 使他难以区分答案到底是谁作出的, 这就说明这个计算机和人的理智能力是没有区别的, 它有“具有一个心灵所必须的一切条件, 即说明它是有智能的。这台机器能够回答这些问题说明了它也有心理状况, 也有理解过程。图林的这一假设把人的心灵类比为计算机的程序。

对图林测试的最有代表性的反驳是由塞尔 (J. R. Searle) 构思的一个思想实验“中文屋论证”。他设想: 我是一个仅仅会说母语而完全不懂中文的人, 被锁在一个隔音的房间里, 在房间里有许多盒子里面装了很多中文的符号即汉字 (我根本就不认识它们)。另外还给了我一本英语的指导书教我怎样将一些符号和另一些符号相匹配, 并且给回一组中文符号来回答通过一个小窗口送进这个房间来的那些中文符号, 在

【收稿日期】 2005 - 01 - 17

【基金项目】 国家社会科学基金 2004 年度基金项目“意向性理论的当代发展”(04BZX042) 资助

【作者简介】 毛华滨 (1973 -), 男, 湖北武汉人, 华中师大政法学院教师, 主要从事认知科学研究;

高新民 (1957 -), 男, 湖北武汉人, 华中师大哲学所教授、博士生导师, 主要从事心灵哲学和认知科学研究。

这种匹配的过程中我完全是根据符号的形状来判断。通过小窗口送进来的那些符号叫作问题,我送出去的那些符号叫作答案,我所有的那几盒子符号叫作数据库,那本英语的指南书叫作程序,给我问题并设计指南书的那些人叫作程序设计员,而我被称作计算机。但我对这些全然不知。我们来想象,窗外的那个人通过小窗口给我一些符号(问题),依据指南书中的规则和符号的形状,我送出去一些符号,过了一会儿,我操纵这些符号做得如此之好,以至于在窗口外的那些不认识我的人以为我是一个以中文为母语的人,其实,我对那些盒子中的汉语、外面送进来的问题、我自己送出去的答案都不懂。再假设,在这个房子中另有一位以中文为母语的人,他不以任何方式与我联系,独立地工作,那么这位懂中文的人的输入输出与我这位不懂中文的人的输入输出完全一样,无法区分,很明显,那位懂中文的人是理解了那些字、问题和答案,而我则完全不懂。实际上,我完全像一台计算机那样活动,接收问题,操纵符号,产生答案,这整个过程就是计算机程序的一次具体例示。这说明,我通过了理解中文的“图林测试”,但是我一个中文字都不认识,完全不理解。如果我以执行理解中文的程序为基础而完全不理解中文,那么仅仅在同样的基础之上的数字计算机也完全不理解中文,因为我所做的和计算机所做的完全一样。塞尔认为,一种正确程序的具体例示,并不足以说明存在理解的心理状态。他的中文屋论证的逻辑结构是:前提(1):程序是形式的(句法的);前提(2):心灵有内容(语义学的);前提(3):句法对于理解语义来说是不够的。从这三个前提可以逻辑地得出结论:程序不是心灵。也就是说,心灵不能等同于计算机的程序,因为计算机程序纯粹是按照句法规则来定义的,而心灵则有心理内容或语义内容,它不仅仅有句法学,而且有语义学,单纯的句法本身是不足以保证心的语义内容的呈现的。

心的语义内容能关于它之外的存在或不存在的东西,而人以外的万物却不能有意识地关于什么。例如一只蚂蚁在沙滩上爬过的痕迹碰巧像丘吉尔的画像,但我们不能说这是蚂蚁在爬之前有心的“丘吉尔”内容或有意识地关于了“丘吉尔”,其爬过的痕迹之所以像丘吉尔是根源于我们的“看”。人当然是可以画丘吉尔像的,但形成这两幅画的差别何在呢?这里至少有一个根本的差别:蚂蚁在“画像”时没有意向性,即它不能有意识地指向、关于什么;人则不同,他在画像时有意向性,即有所指的内容、关于的内容。蚂蚁的“画像”充其量只是一种派生的意向性,而人的意向性则是名副其实的、原始的、固有的意向性。

按照句法规则组织起来的计算机程序虽然和人类智能有一样的能力,甚至有些地方超过了人类智能,但是从表面上看二者间还是存在着明显的差别:人的一言一行在多数情况下都是由自我意识引导完成的,积极主动,而机器所做的每一件事情都需要人事先指定好,机械、被动。^[1]导致以上差别的内在根源很多,而中文屋论证至少表明了人工智能与人类智能的一个根本差异:由于计算机停留在句法操作层面,而使它没有意向性,没有事先的所指内容。

二 人类意向性的特点、作用及其内在机制

现代意向性研究起源于布伦塔诺(Franz Brentano)的有关理论,鉴于此,有人把当代意向性问题称之为“布伦塔诺问题”。他从世界观的高度指出,世界上的现象可划分为物理现象和心理现象,但两者的根本区别不是前人所说的是否有广延、是否能思维等,而是有无意向性。^[2]他说:“我们可以这样给心理现象下定义,即心理现象是那种在自身中以意向的方式涉及到对象的现象。”^[3]依照布伦塔诺的观点,物理现象只涉及物理对象本身,而心理现象包括了心理主体和它指向的外部对象的方式。意向性则反映了这种内外联系的特性。意向性最鲜明的特点是指向性。但是心理现象为什么具有他物所不具有的意向性呢?这是布伦塔诺以及后来许多思想家百思不得其解的难题。

近年来,心灵哲学和神经科学的某些研究表明:

1. 人的心理能够直接思维、加工的是心灵语言,而不是自然语言,因为后者有形体、声音等物质载体,它们不能进入心灵为之直接加工。人类的心灵能理解、运用语言,语言则至少包括了句法和语义两个层面。由此可见,自然语言的语义学不适合解释心灵语言的语义性,自然语言的意义根源于心理的意向性,而意向性又根源于心理语言的语义性。福多(J. A. Fodor)说:“关于心理的表征理论的战略就是主张:心理状态的意向属性根源于心理表征的语义属性”^[4],后者才是心理状态不可还原的属性。换言之,福多认为意向性不是最基本的,不是不可还原的,不是最终的决定因素,因为还有决定它、比它更基本的东西,那就是思维语言(心灵语言)的语义性。心理内容总是关于它之外的存在或不存在的东西,即它涉及到“关于性”(aboutness)。为什么只有人有关于性?人的物理系统怎么可能具有关于性特征?

根据关于心灵的表征理论和计算理论,心理状态就是命题态度,而命题态度是有机体与心理表征或心灵语言的心理语句的关系,因此有命题态度、有表征也就有心理语句。而心理语句有句法和语义两种属性,它的语义属性与自然语言的句子一样也有意义、所指和真值条件。因此,关于性又根源于心灵语言的语义性,人有语义属性的真理理论知识——碰到一个符号,能把它与有关的对象关联起来,即人有这样的知识:“S是P为真,当且仅当S是P。比如:“雪是白的”为真,当且仅当雪是白的。

2. 人工智能在博弈、数学定理证明、医学法律专家咨询、代替人类完成各种复杂作业等方面,取得了较大的成功,却在对人类来说是十分简单的任务方面又显得很笨拙,如:正确理解人的日常言语行为,把握图形、符号的意义等等。人工智能出现这样的问题并不奇怪,因为计算机只是形式句法的演算,0和1再怎么组合排列,还是没办法使它产生我们看到“马”这个字时,所浮现的马的形象和了解“马”的意义。人所具有的意向性在其活动中表现为指向活动,指向的过程是一个选择过程,也是一个组织过程,通过确定对象,并完成赋意的任务,意义是意向活动的结果。从这一点上看,意向性还根源于人有民间心理学(Folk Psychology,以下简称

FP),有解释理论。

FP是比照民间医药学、民间物理学而创立的一个新的概念,1981年丹尼特(D. Dennett)在《三类意向心理学》中创造了这个术语,指的是普遍大众所具有的依据信念、愿望等命题态度解释、预测行为的心理资源。它潜藏于每个正常人的心理结构之中,显现于对行为的解释、预测实践中。但要特别注意:FP不是心理学的分支,也不是写于书本、为专家们所坚持、讲授的学说,而是自发流传于大众之中的、用命题态度解释和预测行为的常识,因此也常被称为常识心理学或命题态度心理学。我们知道,正常的人经常处于解释和预言他人的活动之中,如:看到某人拿雨伞出门便想到他相信天会下雨。这如何实现的呢?追溯这一事实后面的内在条件、结构和机制,也就是追溯他的解释和预言过程中所动用的心理资源。每个人都有这种资源,但大多人没有意识到它,而又都会用到它,正是在这个意义上,许多人认为:正常的人都是心理学家。

大多数哲学家认为,FP包含有如下的存在命题和普遍原则:人有心灵,其内有心理活动、状态和事件,人是有理性的存在,因此行为有可解释性和可预言性,心理状态与刺激、反应之间有因果关系,我们每个人都认识他心。

戴维森(R. Davis)、丹尼特独辟蹊径,倡导对心灵作出解释性说明。根据这一思路,人之所以被认为有FP等心理状态,则取决于我们对之作出的解释,正像我们为了描述地球而强加于地球的坐标系、经纬线一样,一个人被认为具有心理状态是因为我们把一种解释理论授予了他,而不是他真心灵。^[5]换言之,人本无心灵,本无意向状态,它们是我们的解释性投射的产物,或者说是我们为了解释人的行为而强加给人或归属给人的;不是因为人先有各种命题态度,而后才依次产生关于它们的认识和相应的心理语言,而是相反,我们先有心理语言,有一种解释理论,然后才将这些语言用于对人的解释,最后才有将命题态度归属于人的实践。

3.除了以上人有语义的真值理论和FP理论外,人的意向状态有分辨、挑选、处理的功能,这样就使语义内容与信息内容区别开来,摆脱了信息内容的“套叠性”(Nesting Relation)。^[6]

信息内容是信号或指示器所携带或表达的关于信息源中的某个项目的状况信息。信息内容具有套叠性。例如门铃响了,这里面包含了许许多多的物理的、电学的、社会的、人际的信息,这些信息是相互套叠的。而语义内容则不一样,其特征之一是惟一性或对信息源的不敏感性。同样是“门铃响了”这一信号,但是信号的接受者——人只对一种信息感兴趣,只抽取一种信息,而不及其余。换言之,他不可能对信息源的所有信息敏感。因此,信息的套叠性特点就是信息区别于意义的一个特征。人可以对信息进行选择和组织,并捕捉对人有意义、有作用的信息,忽略大量无意义的信息,从“信息海洋”里捞出那根真正有意义的“针”。我们知道计算机是能够接收信息并反馈信息的,但计算机无法像人那样具有意向状态时的“忽略的智慧”,计算机和人接收信息的差别也许只能从意向性中找到答案。

4.按照塞尔的观点,不仅人的心理现象有意向性,而且人的言语行为,如说出的话语本身也有意向性,也有“关于”它之外的事物的指向性,更重要的是,非人的自然事物(如云彩、树的年轮等)也有隐喻意义上的意向性。热敏电阻、光敏二极管这类装置都有“隐喻的意向性”,能“关于”什么,但这类意向性的象征、表现从根本上说源于人类心灵的作用,没有内在的意向性或真实的意向性,其根源在于:它没有意识。因为真实的意向性体现的是心理活动与对象的关系,它表现了意识活动的指向性,即指向对象;心理活动总是要指向一个对象,以一定的对象为内容,它表明意识是关于对象意识,不是自我封闭的,不指向任何对象的心理活动是没有的。从深层上,意向性中包含着意识机制,意向性离不开意识。

怎样才能有意识呢?两位诺贝尔奖得主克里克(Francis Crick)和埃德尔曼(Gerald M. Edelman)对人的意识产生机制进行了不同的说明。埃德尔曼认为:意识经验看来与同时分布在脑的许多不同区域的神经元群体的神经活动有关;为了支持意识经验,大量的神经元群体必须通过所谓的“再进入”过程迅速地相互作用;要说明支持意识经验的神经元群体的活动模式必须一直在变动之中,并且彼此也要有足够的区别。克里克试图运用人的生物神经网络来说明意识的产生;他着重对视觉意识的产生过程进行说明,我们之所以能感知一个具体事物是因为,视觉系统组织不同的通路对视觉信息不同侧面进行传递和处理,形成关于形状、颜色、运动、深度等零碎的信息,然后通过“捆绑”或神经元之间35-40赫兹的振荡把它们整合为一个完整的信息。

三 怎样让计算机也有意向性?

现时代已经跨入智能、信息起重大作用的时代,对意识的研究以及由对思维的模拟而产生的人工智能,成为现代科学和哲学探讨的前沿问题。如今计算机的升级换代令人目不暇接,似乎在昭示着这一学科快速发展。然而应当看到这种变化主要表现在机器容量和运算速度方面。从综合认识能力的角度评价现时的人工智能,应当说还处于相当幼稚的阶段。

人工智能就其本质而言,是对人的思维的信息过程的模拟。对于人的思维的模拟可以从两条道路进行,其一是功能模拟,暂时撇开人脑的内部结构机制,而从其功能过程进行模拟,即符号处理范式。它在某些方面如博弈、定理证明、专家系统等已经取得了较大的成功。符号处理范式的理论基础是“物理符号系统假说”,这一假说认为,人脑就是一种物理符号系统。符号处理范式的极端观点认为,人心之于人脑就如计算机软件之于硬件。计算机的软件对硬件没有依赖性,一种软件可以在不同的硬件上运行,相应地,人心也不依赖于人脑。因此,符号处理范式的本体论上是二元论。在方法论上,是一种功能模拟方法、“黑箱法”,是一种整体论的研究策略。在心身关系上属于功能主义。在实践上,配备恰当程序的计算机,加之正确的输入和输出,本身就拥有与人类意义相同的心灵,可以呈现各种认知状态。精神活动过程同

机器执行程序一样,不过是在从事某种定义为“算法”的运算过程。而人脑和简单的计算机的主要差别仅仅在于人脑的活动具有更大的复杂性,或者表现为更高级的结构,而人的所有精神品质,包括思维、情感、智慧、意识都不过是大脑执行的“算法”特征而已。

这些观点曾一度受到许多认知科学家的强烈抨击。丹尼特并不认为“配备适当程序的计算机事实上确有认识状态,并且能借此解释人类的认知”^[7]。塞尔也主张,计算机的理解力与计算器的理解力没有什么不同。当然,对塞尔来说,重要的不是要论证“计算机不能思维”,而是要回答“正确的输入输出加上正确的计算本身是否足以保证思维的存在?他认为,“如果我们所说的机器是指一个具有某种功能的物理系统,或者只从计算的角度讲,大脑就是一台计算机”,然而在他看来,心的本质并非如此。计算机程序纯粹是按照句法规则来定义的,纯粹句法形式的认知系统不可能拥有心灵,因为句法本身不足以保证心的意向性和语义的呈现,程序的运行只具有在机器运行时产生下一步形式化的能力,只有那些使用计算机并给计算机一定输入同时还能解释输出的人才具有意向性。塞尔进一步指出,按句法形式演算来定义的程式,不管是否能达到相当的复杂性,永远也不可能使电脑产生心灵认知时所特有的意向性本质。意向性是人心的功能,心的本质绝不能被程序化,也就是说,心的本质不是算法的。

因此,符号处理范式的“计算有心灵”的观点至少受到三个方面的挑战:一、数学上的挑战。数学上的挑战来自算法和复杂性两方面,因为算法是有局限性的,自然界的很多事情是不能用算法来表征的,而遇到指数爆炸,计算机也会无能为力。二、意识的挑战。计算机没有人类的意识所特有的能动的创造能力,意向性因与意识的关联,导致人的一言一行在多数情况下都是由自我意识引导完成的,积极主动。三、FP的挑战。对于回答一些最简单的、人类不会感到有任何困难的“常识”问题,计算机却难以胜任;语言理解活动对人来说不算太难,可对机器来说则并非易事,甚至可以说非常困难。

以上的这种状况人工智能极力想要摆脱。于是它开始转向人的日常思维方式。多次碰壁之后,人工智能专家们才醒悟到,原来日常生活世界要比科学世界庞大、复杂得多,看似平常的人类日常思维方式,却包含着相当高超的技巧。人工智能专家们把目光投向意向性,试图从现当代意向性理论中得到某些启示。“只要大脑生成意向性的能力是清楚的,……这种认识所采用的信息加工方式同样可以用于计算机。这样,人工智能的概念就完全有理由被用作心理学理论的基

本组成部分,同样,某些想象之中的计算机也可以具有与意向性和智能十分近似的能力。”^[8]

意向性因与人工智能的关联,再次走到学术前台,使人工智能专家运用联结主义范式,仿照人脑的结构机制,制造出“类人脑”的机器,即结构模拟人的思维。联结主义范式的理论基础是并行分布式处理。许多联结主义者认为应该重视硬件,这与符号处理范式的功能主义者“硬件不要紧”完全相反。联结主义偏重于“灰箱”方法,采取还原论的研究策略,其理想模型——按照人的生物神经网络制造出来的人工神经网络要体现人脑的基本特性:神经网络中有大规模的“回路”,有了它就有“再进入”的机制与功能性聚类(“动态核心”),在此基础上产生人的意识。

除了与意识的关联外,意向性还必须有深层的心灵语言的语义性根源,这要求按照人工神经网络建立起来的人工智能有真值理论,能够能把它与有关的对象关联起来,并通过意向状态分辨、挑选、处理对象的信息内容,忽略无意义的信息,而完成赋意的任务。现在的计算机模型无一例外地忽略外部环境的细节,用于建立模型的大多数计算机除了编程者用来输入指令的键盘以外就没有与这个世界的联系,这要求计算机有某些FP知识,如解释理论,使计算机能理解周围环境的意义。

人工智能专家们将从现当代意向性理论的启示中,使人工智能的发展研究进入到人类智能的核心,并朝着意向性的精确化和具体化的方向努力。

【参 考 文 献】

- [1][8] 玛格丽特·博登. 人工智能哲学[M]. 刘西瑞,王汉琦译. 上海:上海译文出版社,2001. 10-11,126
- [2] 高新民,刘占峰. 意向性意义内容[J]. 哲学研究,2003(3).
- [3] Brentano, F. (1874). *psychologie vom Empirischen Standpunkt*, Erster Band, Leipzig
- [4] 高新民,储昭华. 心灵哲学[M]. 北京:商务印书馆,2002. 516
- [5] Davidson, D. (1989). "What is Present to Mind?" *The Mind of D. Davidson*, eds Brandl and Gombocz, *Grazer Philosophische Studien*, 36: 3-18
- [6] Dretske, F. (1979). *Knowledge and the Flow of Information*, Cambridge, Mass, MIT Press
- [7] 丹尼特. 心我论[M]. 上海:上海译文出版社,1991. 336, 337.

(责任编辑 成素梅)