

人工智能与心理学研究发展

151080001 社会学院 蔡兆钦

摘要

人工智能已经成为当代计算机科学，神经科学，心理学领域最为引人注目的研究领域之一。心理学作为连接世界与心灵的学科，为人工智能做出了巨大的贡献。无论是研究人工智能的思路，基本结构，或是验证某一个机器是否真实拥有与人相符合，甚至超越人类的智能，心理学都做出了巨大的贡献。而人工智能的发展也为心理学的研究提供了新的角度和思路。本文尝试梳理心理学中与人工智能发展较为相关的几个流派，以及人工智能在其影响下的发展和应用领域，发现人工智能的优点，并且尝试指出其不足和局限性。并展望人工智能未来发展的愿景。

关键词：人工智能、心理学、认知心理学、神经科学、计算主义、意向性

人工智能概述及发展历程简介

人工智能（Artificial Intelligence）是近几年在计算机科学领域最为火热的概念。而其研究和发展的过程中，也由于其对人脑的模仿性，和心理学，神经科学，认知科学，哲学密切相关。并且数学，信息论，控制论等领域的知识和研究方法也在人工智能的研究和开发中得到运用。是一门综合多学科，多领域的边缘学科。

一般而言，人工智能的定义是：使机器具有人所拥有的智能行为，如判断、推理、证明、识别、感知、理解、设计、思考、规划、学习和问题求解等思维活动。换言之，过去的计算机研究，是使其能在速度和质量上较好，甚至比人更好的解决一些人通过程序编写或硬件设计，“告知”其按照何种步骤可以完成的问题。因此，一般的计算机“算法”（Algorithm）的定义是：“解题方案的准确而完整的描述，是一系列解决问题的清晰指令”。而算法的特性：有穷性（算法需要在有限步骤后终止）和确切性（算法的每一步都需要有确切的定义），也证明了以上观点。也就是，对于过去一般的计算机问题求解问题，人对计算机的要求是快速的计算出已经找出方法的结果。而人工智能的需求有所不同，人工智能的发展目标体现在以下几个方面。

1、解决人类还未解决或者还没有较好解决方案的问题。例如近期最为吸引眼球的人机围棋大战。人工智能 Alpha go 已经证明了在围棋这个无法通过穷举来解决的问题上，人工智能可以比最顶尖的人类获得更高的胜率。从某种意义上来说，Alpha go 比人类对围棋有更深刻的理解。2、使计算机拥有和人类类似的感知，知觉，推理，学习能力。

这也是人工智能发展的一个较高目标。例如目前的自然语言识别，图片识别，声音识别等，也是发展的方向。而人工智能发展的终极目标，也就是如约翰·麦卡锡在 1956 年的达特茅斯会议上提出的“强人工智能”那样，有自我意识，能解决问题和推理的机器。而从目前来看，这个目标需要长时间的科学工作者的努力。

人工智能的发展经历了一个很长的过程，其指导思想也有多次变化。从计算机出现开始，制造有和人类似的推理，学习能力的智能机器，最终解放人类于辛苦的思考之中，就是计算机科学家的研究目标之一。最早的人工智能是通过计算机对人脑的神经网络进行模拟，因此也被叫做 ANN (Artificial Neural Networks)，1950 年，图灵在论文《计算机器与智能》中提出了图灵测试、机器学习、遗传算法和增量学习理论。为人工智能的发展奠定了基础。而在之后，联结主义，认知心理学的理论也对人工智能的发展有其独到贡献。

人工智能所关注的领域及当前进展

人工智能所关注领域的一个方向是问题的解决。其中包括问题的识别和解决等部分。也包括对解决问题过程的规划和学习相关知识。在 2017 年，中国所设计的人工智能已经开始参与当年的数学高考，并且取得了 134 分的成绩。参与考试并非由人将题目输入计算机内部，由机器进行简单的计算。而是在断网的情况下，将试卷呈现给计算机，由计算机进行扫描和文字转化之后，将原本的题目，转化成计算机能识别的形式语言，再计算得出答案，并通过输出设备输出。而这一流程，就包括了解决某一个问题的整个步骤。随着机器能力和学习样本的增加，在未来，机器可以解决包括难度更高，对人感情理解程度需求更高的问题。甚至可能会提出某些人类原本没有想到的解决方法。

另一个较为简单的人工智能领域是分类，分类问题有两种：有监督分类和无监督分类。有监督分类是在已知数据集有标签，也就是已经有人工分类的情况下，由机器对数据进行学习，当有新的数据进入时，机器可以对其进行判断和分类，将其归于之前所属的某一个分类之中。常见的由机器对手写数字，字母或汉字的识别，或者对声音，行为的判断，都是有监督分类的一种。将输入机器的图片或声音，转化成有位置参数的数据之后，对其进行已有框架下的分类。

而无监督分类就与上文的有监督分类相对应，是对没有标签的数据进行分类。其在运用中由于没有人的干扰，在多次迭代之后，可能会给人带来惊喜的结果。一些新出现的图片处理技术，如智能去水印，图像风格混合等，就起源于无监督的分类。

人工智能也被运用于金融界的量化投资之中，运用高频交易，在金融市场上进行套利，或者在赌博市场上选择赔率与大众预期不相匹配的博弈操作，能给投资者带来大量利润。而此运用方式除了人工智能之外，也需要大量的数学，统计，博弈论，金融学相关知识。

综合以上，人工智能已经被广泛运用于多个领域，并且取得了长足的进步，但是人类距离最终的强人工智能，万能的“问题解决机”的出现还有多长时间，依然是一个值得科学家和哲学家辩论的话题。

人工智能与认知心理学

认知心理学 (Cognitive Psychology) 是出现于上世纪 50 年代的心理学流派，研究对象是人的高级心理过程，主要是认知过程，如注意、知觉、表象、记忆、思维和言语等。认知心理学需要让计算机模仿人的认知过程，以此方法设计出与人类智能相类似的人工智能。要使计算机像人那样进行思维，计算机的程序就应当符合人类认知活动的机制。

认知心理学的来源也与计算机设计有关。计算机之父冯·诺依曼提出了计算机的五个基本组成部分：运算器、控制器、存储器、输入设备、输出设备。认知心理学家也尝试运用这样的逻辑来对人的认知过程加以分析。认知心理学把人的记忆过程加以分析，分解为编码、储存和提取的过程，这就与计算机对信息加工的编码、储存、提取和输出的流程相类似，这也是计算机科学与心理学相互促进的一个印证。

随着人工智能的进一步发展，由于人工神经网络在识别信息之后，其发展的过程和最终结果对于研究者而言是一个“黑箱”，因此也开始运用认知心理学的方式对人工智能进行研究。正如 S. Ritter. (2017) 的研究所述，当一个人工智能系统尝试识别图片中的某一个图形时，“他”选择的最重要的特征是该图形的形状，而非其他的特质。而对形状的偏好，出现在该人工智能自我训练的早期，也就意味着，少数的图像数据，就能培养出该人工只能对图形形状的更高度注意。这也与人类的认知过程相类似，在年龄很小的时候，人就会把物体的形状作为分类的重要标准。

以一个常见的例子作为说明，一位实地语言学家走访了语言完全不同于我们的地区。这位语言学家尝试去和当地的土著学习一些词汇，恰巧此时一只兔子匆匆路过。这位土著说了一声“gavagai”，而这位在他身边的语言学家也推断出了这个新词汇的含义是指“兔子”。然而事实上，这一推断包括了许多假设成分，例如这一词语也可能是指“白色”或者“动物”。然而一般而言，人们会愿意用这一事物的形状特征，也就是“兔子”在进行

解释。

因此，我们注意到，认知心理学与人工智能的发展具有相互促进的作用，随着两门学科的进一步发展，人可以运用人工智能的结构，为心理特征做出进一步解释。

人工智能与神经科学

神经科学已经是当代心理学研究中不可避免的一部分，科学家尝试通过神经科学的研究，在细胞水平上探索人的心灵的奥秘。这一研究范式虽然有还原论的问题存在，但是依然是目前能在最精确程度上对心理现象解释的研究方法。和传统的心理学研究范式，如实验，观察等相比，神经科学是一般意义上最为满足实证主义理想的研究思路。神经科学的相对优势在于，传统的心理学研究范式是把人脑作为一个黑箱，研究者研究其输入和输出的关系。而神经科学尝试打破这个黑箱，找出神经和脑的工作原理。可以想象，如果人类可以彻底弄清大脑的工作机制，无论是复制，再造或者提高大脑，都将变为一件不算困难的事情。虽然其价值仍有待评估，但是成功的利益将是巨大的，甚至可能改变世界的一般价值。

目前人对神经系统已经有了初步的认识，研究者意识到，人的思维和记忆，与大脑中神经元之间的突触有关，当脑中两个神经元负责的功能相关性较强时，两个神经元之间的突触联系也逐渐变强。以此为原理，人工智能研究者构造了计算机中的神经网络。也就是，在数据输入进计算机之后，要求计算机尝试对其进行分类。计算机在输入值和输出值之间，选择多种函数用于计算其关系。当某一个或某几个输入值能较好的预示输出的所处类别，则对应的函数所占的权重就相应增大。如果输入的数据集足够大，就可以使计算机中各个函数的权重处于相对稳定的状态，当有新的数据输入时，计算机就可以在较高准确率上进行分类。

目前，神经网络是人工智能领域最为火热的概念。在视觉听觉识别，游戏与博弈领域等方面得到了大量应用。2017 年的人机围棋之战中，谷歌公司的 Alpha go 人工智能，工作原理就是神经网络。与上世纪专注与象棋的“深蓝”人工智能不同，“Alpha go”不仅仅可以用于围棋领域，其也将用于医疗，教育，科研等其他领域。相对而言，是一个更强的人工智能。

神经网络有一大优点，就是研究者可以将其内部结构较为清晰进行研究。当代的神经网络一般拥有多层，也就是在输入和输出之间，有多层神经元组成。每一层的神经元从上一层得到数据，并且将得出的结果传输给下一层的神经元。因此，每一层神经元的

负责部分都可以为人所拆解。从某种意义上而言，神经网络的研究，也能带来对人脑新的认识。

人工智能，计算主义与意向性

经典计算主义把人的心灵，特别是思维等认知现象，理解为计算，看作是依据某种规则对符号系统进行加工的过程，也就是从输入到输出的映射。计算主义的主要影响有以下两点：首先，人的大脑活动可以理解为某种计算过程，而大脑只是进行计算的物理实体，这是一种对心灵活动理解的新形式。其次，根据这一理论，只要能制造出根据一定规则操作符号的机器，也就制造出了智能机。智能的核心是计算的过程，或者说，程序。而进行承担的机制并不重要。

然而传统的计算主义遭到了大量的批判。其中，以塞尔的“中文屋实验”最为有影响力。“中文屋”是一个理想实验，表述如下：一个对汉语一窍不通，只说英语的人关在一间只有一个开口的封闭房间中。房间里有一本用英文写成的手册，指示该如何处理收到的汉语信息及如何以汉语相应地回复。房外的人不断向房间内递进用中文写成的问题。房内的人便按照手册的说明，查找到合适的指示，将相应的中文字符组合成对问题的解答，并将答案递出房间。

Searle. (1980)认为，尽管房里的人可以以假乱真，让房外的人以为他确实说汉语，他却压根不懂汉语。在上述过程中，房外人的角色相当于程序员，房中人相当于计算机，而手册则相当于计算机程序：每当房外人给出一个输入，房内的人便依照手册给出一个答复（输出）。而正如房中人不可能通过手册理解中文一样，计算机也不可能通过程序来获得理解力。既然计算机没有理解能力，所谓“计算机于是便有智能”便更无从谈起了。

因此，一个机器是否具有智能，并不仅仅取决于其是否拥有输出正确答案的能力，而在于其是否具有意向性。也就是呈现外部事物的能力。如果不能理解其处理的数据的含义，那人工智能的意义也需要打折扣。

然而，人工智能的发展和实现是无法离开计算和各种门电路的。人脑的基础：神经元对信号的处理方式也是基于基本的计算和信号传递。判断一个人是真的“懂中文”，或是只是掌握中文语法规则的一个程序，差别就在于是否有足够的实践和具体经验。如果人工智能可以将词义和具体表征进行联系，那也就可以认为其有了意向性。

与经典计算主义几乎同时，又出现了联接主义。联接主义关注的不是输入和输出的

计算，而是模拟的神经元的相关联系。早期联接主义者认为，在计算机中的构造的神经元之间构造合适的函数和权重值，可以将任意的输入都计算出合适的输出结果。而合适的权重值的来源就是大量数据的学习。早期，由于计算机性能的限制，联接主义者的构想难以得到实现。在 20 世纪 70 年代之后，随着人类计算能力的大幅提升，这一构想逐渐成为可能。也得到了大量的运用。

联接主义虽然有其独特的思想所在，但是仍然无法脱离计算主义的一般限制。关于这样生成的程序是否真的具有智能，具有意向性，依然是值得哲学家进行思考的问题。但其运用上的广泛前景，使其成为人工智能领域当今最为火热的研究方向。

人工智能的未来发展

意识是人存在的核心，没有意识的人在某种意义上可以被称为死亡。而人工智能工作者的目标，就是使人工智能在未来可以具有和人类似，甚至超越人类的“意识”。在理论上，可能性是完全存在的。

人类的存在受到人身体的各种限制，包括各种感受器的有限性，视觉所能接收的光的波长，听觉的可接受频率，都处于一个有限的范围之内。而人工智能的感受器仍有巨大的发展空间，并且不用担心疲劳，也不会遭受各种病痛。从此意义而言，人工智能是人的延伸，是一种新的工具。

根据分类，人类目前仍处于“弱人工智能”阶段，一般的处理问题依然局限于数据分析，自然语言处理，图像识别等领域。还处于一个“自然科学”成分较强的研究领域。今后，人工智能的目标主要向以下几个方面发展模糊处理、并行化、神经网络和机器情感。力图弥补生硬的自然科学缺陷,使人工智能具有真正的智性、具备人工智能意识，并且使用某些未知的方法消除各种认识极限。

意识也是人最为宝贵的财富，在过去除了人类，没有其他的事物具有意识。如果在未来，人工智能拥有了意识，并且随着其自我学习和发展，其意识可能会超越人原本的能力，出现位于智商量表上智商是“1000”的新的智力。这也是人类的自我的拓展和生产力大幅度提高的保证。若能拥有人工智能的帮助，人类对于世界的规律的掌握将会上升到一个新的台阶之上。而如果那一状况真的出现，人类和机器的关系将会到达一个新的阶段。人的伦理和价值也可能得到重构，而在笔者看来，这也是社会发展的不可避免的环节。相信在未来，人工智能作为人的外延，可以给人类社会带来崭新的变化。

参考文献

付东鹏. (2013). 心灵与计算——关于计算主义的哲学研究. 华中师范大学博士论文

李建国. (1986). 人工智能与认知心理学. 西南师范大学学报

周雅凡. (2008). 人工智能的哲学反思. 武汉理工大学硕士论文

张旺强. (2012). 论人工智能具有意向性的可能. 兰州大学硕士论文

S. Ritter. (2017). Cognitive Psychology for Deep Neural Networks: A Shape Bias Case Study. Retrieved July 15, 2017 from the World Wide Web: <https://arxiv.org/abs/1706.08606>

John R. Searle. (1980). MINDS, BRAINS, AND PROGRAMS. Behavioral and Brain Sciences