

人工智能技术与主要国家技术布局

李若尘¹, 李梦薇²

(1. 龙岗智能视听研究院, 深圳 518116;
2. 中国科学技术信息研究所, 北京 100038)

摘 要: 本文梳理回顾了人工智能历史上存在的三种技术路线——符号主义人工智能、联结主义人工智能、行为主义人工智能的发展现状, 并探讨了世界主要国家在人工智能领域的布局, 通过对比中、美、日、韩四国在人工智能领域发布的战略性政策文件, 分析了四国在人工智能领域技术布局的异同, 并着重分析了四国共同关注的知识驱动、类脑智能、可解释性三大技术, 同时对我国人工智能发展提出了对策建议。

关键词: 人工智能; 知识驱动; 类脑智能; 人工智能可解释性; 技术布局

中图分类号: TP181 **文献标识码:** A **DOI:** 10.3772/j.issn.1009-8623.2021.06.003

“人工智能”一词初见于1956年达特茅斯学院夏季研讨会, 后又经历了逻辑推理、专家系统、神经网络、深度学习等几个重要阶段^[1]。2012年AlexNet首次击败传统人为特征图像识别, 人工智能的发展进入了算法、数据、算力三轮驱动的新时代。在海量数据和强劲算力的支撑下, 人工智能开始赋能各个领域, 提升了产业智能化水平, 并为经济高质量发展提供了源源不断的动力。世界主要国家都在加紧布局人工智能技术与产业发展方向, 然而由于人工智能技术现有的理论、技术瓶颈, 人们距离达到真正意义上的强人工智能还有很长的距离, 人工智能的下一个十年去向如何, 成为备受瞩目的焦点。

1 人工智能的三次浪潮及最新进展

人工智能发展至今历经了三次浪潮, 基本思想大致可分为符号主义、联结主义、行为主义, 其中

符号主义和联结主义的竞争最为激烈。符号主义的代表是逻辑推理、专家系统和知识工程, 这一技术路线统治了早期的人工智能学术圈。而20世纪80年代, 随着神经网络的兴起, 联结主义开始占据上风。行为主义主要起源于维纳等人的控制论学说^[2], 影响了仿生机器人、强化学习等产品和算法设计理念。

1.1 符号主义理论最新进展

符号主义理论的发展诞生了语义网络、知识表示、本体论、知识图谱、认知图谱等一系列成就。语义网络由图灵奖获得者蒂姆·伯纳斯-李^[3]提出, 旨在用三元组标签描述网页本身, 便于被当时的搜索引擎索引。语义网络理论的出现引发了研究人员对知识图谱的探索。早期知识图谱将关系局限为几种特殊的基本关系, 如“属性”“导致”“属于”等, 并定义一系列在图谱上推理的规则, 期望通过在知识图谱上进行逻辑推理实现智能^[4]。然而, 早期知识图谱的思路遇到了许多实际的困难, 比如

第一作者简介: 李若尘(1994—), 男, 硕士, 主要研究方向为人工智能和区块链技术。

通讯作者简介: 李梦薇(1994—), 女, 研究实习员, 主要研究方向为人工智能、伦理治理、科技政策。

项目来源: 中国科学技术信息研究所创新研究基金“新一代人工智能应用场景学理分析与相关政策研究”(QN2020-07); 科技创新2030—“新一代人工智能”重大项目“重点领域人工智能治理挑战及对策研究”(2020AAA0105300)、“重点领域人工智能伦理风险及对策研究”(2020AAA0105301)。

收稿日期: 2021-03-29

结构解析困难、知识图谱编辑过程需要大量人工参与、推理规则无法穷举等, 这些瓶颈一定程度上导致以知识图谱为代表的符号主义发展陷入困境。随着互联网技术的发展, 2007 年 Metaweb 公司发布 Freebase 大型合作知识库, 知识库中包含大量“机器可读”的数据信息; 2010 年谷歌收购 Freebase, 在谷歌的运营下 Freebase 知识库得到进一步发展, 并于 2016 年更名为“谷歌知识图谱 API”, 现已成为谷歌搜索引擎的核心知识系统。符号主义的发展由于结构化的知识难以获取和复杂问题推理难以实现而一度中断, 但符号主义的最新进展却为实现知识加数据“双轮驱动”以及人工智能的可解释性奠定了基础。

1.2 联结主义理论最新进展

联结主义理论诞生了感知机模型, 进一步演化出神经网络模型、BP 神经网络。现如今备受瞩目的深度学习技术也源自联结主义。为了证明深度学习的潜力, 多伦多大学 Hinton 课题组于 2012 年参加 ImageNet 图像识别比赛^[5], 其通过构建卷积神经网络 (CNN) AlexNet 一举夺得冠军^[6], 且性能远超第二名支持向量机方法 (SVM)。正是由于该比赛, 卷积神经网络进入大众视野, 吸引了众多研究者的注意。2013、2014、2015、2016 年, 通过 ImageNet 图像识别比赛, 深度学习网络结构、训练方法、芯片硬件不断进步, 促使深度学习性能出现爆炸式增长, 诞生了 ResNet、RNN、LSTM、VGG 等众多网络模型。得益于移动端芯片等硬件设备性能不断增强, 深度学习的产业化程度得到了前所未有的发展, 成为时下应用最广的人工智能技术。

1.3 行为主义理论最新进展

行为主义最早来源于 20 世纪初的一个心理学流派, 行为主义人工智能强调智能体与周围环境的交互与反馈。对行为主义的研究产生了仿生机器人、自主行走机器人等应用, 并影响了强化学习理论^[7]。强化学习理论与行为主义人工智能有着密不可分的关系, 强化学习的核心是智能体在环境中怎样执行动作以获得最大的累计奖励, 也是处理物体与环境间交互的问题。早在维纳的控制论提出之前, 研究人员就通过多巴胺的作用机理认识到了奖励信号与试错学习的关系。2013 年 DeepMind 提出 DQN 深度强化学习算法, 2016 年基于 DQN 算法的 Alpha

Go 以 4 : 1 的比分战胜世界围棋冠军李世石。强化学习在围棋以及游戏竞技中取得引人注目的成果, 并且在自动驾驶、机器人控制等领域崭露头角。行为主义人工智能不仅在各种游戏竞技中大放异彩, 也影响了拟人机器人训练、进化算法等领域^[8]。

2 世界主要国家在人工智能领域的布局

自 2015 年以来世界各国频频出台与人工智能相关的国家级战略, 各国政府均对人工智能发展有着独特的布局。本章通过探讨不同国家的人工智能战略, 试图厘清人工智能未来的发展方向。

2.1 美国

美国国防高级研究计划局 (DARPA) 曾对人工智能发展历程从技术角度进行了详尽的梳理, 并对人工智能发展的三个阶段进行了界定^[9]: 初代的人工智能基于格式化的规则和形式, 对狭义的任务有较好的推理能力, 但由于没有自我学习或是算法进化的能力, 其处理开放性问题时效果较差, 因此导致其应用范围非常有限 (典型代表如专家系统); 第二代人工智能主要基于统计学, 这一代的人工智能学者多数具有较强的统计学背景, 如大卫·曼福德等人, 此时的人工智能系统经过大量高质量格式化的数据训练, 但仍然缺乏可解释性和基于环境场景的推理能力, 无法适应不断变化的复杂环境; 第三代人工智能系统将重点突破第一代和第二代人工智能系统的局限性^[10], 能够利用情境模型来感知、学习、抽象和推理, 并具有很强的适应不同情境的能力^[11]。同时美国国防高级研究计划局还在 2018 年提出“下一代人工智能计划”^[12], 下一代人工智能主要有三个目标: 一是增强第二代人工智能技术的稳健性; 二是积极将第二代人工智能系统应用于新领域; 三是进一步研究第三代人工智能技术。对此, 美国国防高级研究计划局通过设立项目加强对“下一代人工智能”的研究, 相关项目包括: “机器常识”“终身学习机”“可解释人工智能”“可靠自主性”“不同来源主动诠释”“自动知识提取”“机器通用感知”“利用更少数据学习”“以知识为导向的人工智能推理模式”“高级建模仿真工具”“复杂混合系统”“人机交流”“人机共生”等。

2.2 韩国

2018 年, 韩国第四次工业革命委员会审议并

通过《人工智能研发战略》，旨在重点推广人工智能技术进步，并加快人工智能在各领域的创新发展，打造世界领先的人工智能研发生态，构建可持续的人工智能技术能力。2019年12月，韩国发布《韩国人工智能国家战略》，继续加大对“DNA产业”即数据(Data)、网络(Network)、人工智能(AI)产业的支持。《战略》在第4部分行动指南中公布了3个领域的9大战略与100项任务，分别是“打造世界领先的人工智能技术能力”“培养人工智能领域高端人才”“构建开放合作型人工智能研发基础”。其中明确了韩国将要促进机器学习、计算机视觉和智能语音等通用技术研发，并在10年内(2020—2029年)投资1万亿韩元用于人工智能芯片研发^[13]。同时在项目具体实施中会参考借鉴美国国防高级研究计划局的项目组织方式，根据技术与环境等的变化适时调整项目目标。《战略》还明确了加强知识表示与推理、机器学习算法、认知科学等人工智能技术的研究。

2.3 日本

日本政府为确保其在机器人领域的长期优势地位，于2015年发布《新一代机器人战略》，该战略旨在研发出能够替代人类甚至超越人类能力水平的人工智能和机器人技术^[14]。《战略》提出，要加强对“数据驱动与知识驱动融合型人工智能”以及“类脑智能”的研究，同时指出要研究新型信息处理理论和程序设计方法，探索兼顾隐私保护和数据安全的数据获取方式，进而提出复杂场景下多模态融合的新一代人工智能框架。《战略》发布实施以来，安倍政府不断增加对机器人及人工智能领域的研发支持力度，在日本政府公布的2020年预算中，科技振兴费同比增加187亿日元(1.4%)，达到1.3565万亿日元，人工智能及机器人是该项预算的重点支持内容，同年日本发布《人工智能战略2019》，强调了将教育体系的改革和研发体系的重组放在首位。可以看出日本对人工智能的布局是以机器人技术作为依托的，日本曾在20世纪末引领了人工智能“五代机”技术的发展，但“五代机”研制失败，使得日本一度脱离了世界人工智能发展队伍。从日本政府近期公布的几大战略中可以看出，日本仍然在为冲击人工智能领军地位做着准备。

2.4 中国

2017年中国政府发布《新一代人工智能发展规划》^[15]，《规划》明确了中国人工智能技术发展的基本原则和战略目标，指出要加强对大数据智能理论、跨媒体感知计算理论、混合增强智能理论、群体智能理论、高级机器学习理论、自主协同控制与优化决策理论、类脑智能计算理论、量子智能计算理论等基础理论的研究，并重点对知识计算引擎与知识服务技术、跨媒体分析推理技术、群体智能关键技术、混合增强智能新架构和新技术、自主无人系统的智能技术、虚拟现实智能建模技术、智能计算芯片与系统、自然语言处理技术等关键共性技术加强研究^[16]。

2.5 各国技术布局对比

日本政府于2015年发布《新一代机器人战略》，2017年中国发布《新一代人工智能发展规划》，2018年美国国防高级研究计划局发布“AI Next”计划，2019年韩国发布《韩国人工智能国家战略》。通过提取中美日韩四国发布的关于人工智能的战略文件中与人工智能技术密切相关的内容，形成表1。

根据表1所示的内容可以看出，中美两国更加注重对前沿人工智能理论和算法的研发，而日韩更注重利用已有的产业优势发展独具特色的人工智能技术。中国在人工智能技术布局上涉及范围较广，对新概念芯片、知识推理、自然语言处理等技术都有所关注；美国则侧重人工智能的情景推理、可解释性以及人机互动；日本重点关注人工智能技术与机器人技术融合；韩国倾向于发展服务于人工智能的半导体产业。同时可以看出，虽然各国在人工智能技术布局上的侧重点各不相同，但“知识驱动”“类脑智能”“可解释人工智能”却是各国共同关注的。

3 人工智能技术发展的三大方向

人工智能技术正在从感知走向认知，从人工编辑特征转向自动理解模式。上一个十年人工智能技术在智能算法、大规模算力和海量数据的共同驱动下，完成了波澜壮阔的发展。然而，在各项驱动力趋于瓶颈的今天，人工智能将走向何处？通过上一节对比各国在人工智能战略中对技术发展的布局可以看出，“知识驱动”“类脑智能”“可解释人工智能”是各国普遍关注的三个人工智能技术方向。

表 1 中美日韩人工智能技术布局

国家	技术布局
中国	“知识计算引擎与知识服务技术”“跨媒体分析推理技术”“群体智能关键技术”“混合增强智能新架构和新技术”“自主无人系统的智能技术”“虚拟现实智能建模技术”“智能计算芯片与系统”“自然语言处理”
美国	“机器常识”“终身学习机”“可解释人工智能”“可靠自主性”“不同来源主动诠释”“自动知识提取”“机器通用感知”“利用更少数据学习”“以知识为导向的人工智能推理模式”“高级建模仿真工具”“复杂混合系统”“人机交流”“人机共生”
日本	“数据驱动与知识驱动融合型人工智能”“类脑智能”“下一代人工智能框架”“模块化人工智能和软件”“感知和识别技术”“机构、驱动器和控制技术”“机器人操作系统和中间件”
韩国	“半导体+软件+系统”“新概念人工智能芯片”“决策可解释人工智能”“小样本学习”“人工智能学习能力测评”

3.1 知识驱动

大数据驱动的深度学习虽然在产业上取得了较大的发展,但是其技术本身已不具备较大的发展空间。同时,深度学习技术并不是人工智能研究的全部,张钹院士指出,未来的人工智能范式应当将知识驱动与数据驱动二者进行融合。数据驱动的人工智能技术(暨深度学习技术)往往非常缺乏常识和推理能力,以图像识别为例,大部分的深度学习系统仅仅只做到了“感”(机械的分类器),而没有实现“知”即逻辑推理,而人类的智能却是体现在逻辑推理上^[17]。人的智能无法通过单纯的大数据学习习得,必须要加上知识,使其具有推理能力、决策能力,这样才能解决“感”而无“知”的问题。通过为人工智能系统构建知识库,增加逻辑推理能力,可以使人工智能系统产生出比单纯使用数据驱动更强大的智能。

3.2 类脑智能与脉冲神经网络

虽然现有的人工智能技术在应用上取得了巨大的成功,但是目前人工智能模型下的机器人或智能体仍不具备类似人脑的思维。计算机下围棋、下国际象棋、玩电子游戏,看似是在模拟人的行为,但事实上机器本身并未具有主动思维的能力,由此可见数据智能与机械智能并不是真正的智能,它们只是看上去很像智能的信息处理,与真正的智能还相去甚远^[18]。因此基于深度学习的这一类人工智能技术只能局限性地解决某一个问题或者某一类问题,它们解决问题的能力完全依赖于人类编写的程序与数据库。而未来人工智能的发展不应仅仅停留

在数据智能层面,而是应该用机器模拟人的大脑与思维,让机器像人一样拥有主动性思维。

时下大热的人工智能模型所采用的人工神经网络被称作第二代神经网络,来源于联结主义的感知机模型(第一代)。尽管这种神经网络已经让我们在很多领域的应用中实现了突破,但它们在生物学上是不精确的,尤其是尚不能模仿正常生物体大脑中神经元的运作机制。更加接近神经元模型的是第三代神经网络——脉冲神经网络^[19],脉冲神经网络与目前流行的机器学习方法以及神经网络有很大不同,它具有更加扎实的生物基础,对人脑的模拟程度更高。

脉冲神经网络技术是实现类脑智能的方法之一,同时,类脑智能也可以使用基于硬件芯片的方式进行模拟,这种芯片称为拟态神经芯片。2017年,IBM开发的TrueNorth芯片可以完成对640万神经元的模拟;2019年,Intel开发的Loihi将这一数据提高到800万。人类的大脑皮质中包含有约140~160亿神经元,小脑中包含大约550~700亿神经元,芯片对大脑的模拟还远远不足,然而脉冲神经网络和拟态神经芯片的出现已经为未来人工智能的发展指出了新的方向。

3.3 可解释人工智能

随着人工智能技术的广泛应用,其中所包含的缺陷与不足也逐渐显露出来,复杂算法、模型以及系统普遍缺乏决策逻辑和透明度,算法对产生的结果缺乏有力的解释与证明,种种问题导致人工智能技术难以在国防、金融、医疗等关键领域深入应用。

要实现人工智能的可解释性,可以从人工智能早期符号主义学派的知识工程理论着手。知识工程包括知识的获取、表达和推理,其本质是对数据的结构化,将数据转化为利于计算机处理的结构化知识。通过知识工程构建知识库,再结合推理机,就可以构建出一个专家系统。专家系统对问题的解答以及其本身的决策行为遵守严格的演绎推理逻辑,具有较强的可解释性。

基于知识工程的逻辑推理对人工智能的可解释性有一定帮助,但并不适用于基于数据驱动的深度学习系统。对此,加州大学洛杉矶分校(UCLA)朱松纯教授提出以人为本的人工智能模型(Human-Centric Model, HCM)^[20],该模型旨在令人工智能的决策和推理可解释、值得信任。HCM借鉴了人类学习和推理中的几种主要解释方法,早期研究表明,解释性可以来自功能角度(即目标)和机制角度(即过程)。朱教授的研究组合了来自两种不同角度的解释性,他在一项训练机器人打开一个拥有多个安全锁机制的药瓶的实验中,设计了一个符号规划器和一个触觉预测器,符号规划器通过可视化学习人类打开药瓶的过程,从动机上给出每一次决策的解释性;触觉感受器基于上一步操作的结果,从功能角度给出解释性。该实验给出了产生解释性的另一种方法,即模拟人的行为方式。但由于当时并不存在可解释性的统一标准,这种方法是否适用于更加广泛的领域还有待进一步研究。

2020年8月,美国国家标准与技术研究院发布了《可解释人工智能四项原则草案》^[21],给出了人工智能可解释性的官方定义,《草案》包括解释原则、有意义原则、解释准确性原则和知识局限性原则,同时针对不同的场景给出了五类解释:一是阐述与用户相关的人工智能系统的输出信息的用户有益解释,二是提高社会对人工智能系统的信任和接受程度的社会可接受解释,三是用来协助审核人工智能系统是否符合相关的法规和安全标准的监管合规解释,四是用来帮助和提升人工智能算法和系统的开发、改进、调试和维护的系统开发解释,五是用于使人工智能系统运营者获益的所有者有益解释。《草案》为探索人工智能可解释性提供了一套统一的标准,为在关键领域应用人工智能提供了检验和评判标准。

4 发展人工智能的对策建议

人工智能的发展自达特茅斯会议后历经70余年,这70余年间人工智能有过高潮也遇到过低谷。随着深度学习技术的出现以及数据、芯片等外部支撑的不断发展,人工智能技术在最近10年中再次展现出它惊人的一面。然而数据和算力资源的发展终将接近顶峰,人工智能也会再一次走到十字路口。人工智能未来的技术布局关乎本轮人工智能的发展前景,也关乎我国人工智能发展战略目标的实现。

习近平总书记在《关于〈中共中央关于坚持和完善中国特色社会主义制度 推进国家治理体系和治理能力现代化若干重大问题的决定〉的说明》中强调,要更加重视运用人工智能大数据等现代信息技术手段提升治理能力和治理现代化水平。人工智能技术已成为治国理政的重要手段,因此发展好人工智能要统筹考虑以下几个方面。

一是要兼收并蓄,充分吸收他国经验作为我国发展人工智能的重要参考。我国人工智能虽然近些年发展快速,但是起步晚于美日英等国,参考欧美发达国家在人工智能领域的先进经验有助于我国明确人工智能技术路线选择,在进一步发展人工智能产业时少走弯路。

二是加强企业联系,形成人工智能攻坚克难的一股合力。中国人工智能企业众多,但相对分散,应大力支持现有的人工智能产业联盟,以联盟为纽带加强企业间的联系,制定统一的标准,构建开源开放数据库,形成协同共享的发展局面,以此来强化人工智能领域的核心力量。

三是加大开放力度,吸引海内外人才。欧美等国普遍在海外设立有人工智能研究机构,且本土吸引海外人工智能人才的能力较高。我国应当进一步增强开放力度,吸引海外人才来华从事人工智能研发工作;鼓励企业设立海外研发中心,网罗海内外人才。

四是要发展与治理并重。要积极编制人工智能伦理准则,严格遵守现行法律法规,杜绝人工智能技术对社会发展可能产生的负面影响。在国际社会积极发出中国声音,为中国人工智能企业出海打好基础。抵制西方媒体对中国人工智能的肆意抹黑和“妖魔化”,让西方国家了解中国人工智能发展理念是发展与治理并重。■

参考文献:

- [1] 全耀. 浅谈人工智能的发展史 [J]. 现代信息科技, 2019, 3 (6): 80-81, 84.
- [2] 刘正兴. 维纳与控制论的发展——纪念《控制论》出版 40 周年 [J]. 玉溪师范学院学报, 1988 (2): 44-51.
- [3] 李明. 语义网图中文本相似度方法的应用研究 [D]. 济南: 山东大学, 2015.
- [4] 田莉霞. 知识图谱研究综述 [J]. 软件, 2020, 41 (4): 67-71.
- [5] Krizhevsky, Alex, Sutskever, Ilya, Hinton, Geoffrey E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84-90.
- [6] 姚万超. 基于卷积神经网络的行人再识别算法 [D]. 杭州: 浙江大学, 2017.
- [7] 徐心和, 么健石. 有关行为主义人工智能研究综述 [J]. 控制与决策, 2004, 19 (3): 241-246.
- [8] 郭宪, 方勇纯. 仿生机器人运动步态控制: 强化学习方法综述 [J]. 智能系统学报, 2020, 15 (1): 152-159.
- [9] 张倩. 中美人工智能发展战略对比分析 [J]. 杭州科技, 2019 (6): 36-40.
- [10] 算法战: DARPA 下一代人工智能计划初见成效 [N]. 中国航空报, 2019-07-30 (A07).
- [11] 张焱, 陈楠楠, 王月悦. 从几种国外新型电子战系统看电子战的未来发展 [J]. 舰船电子对抗, 2020 (4): 14-18.
- [12] 杜子亮. DARPA 下一代人工智能技术与军事应用研究 [J]. 中国军事科学, 2019 (3): 42-50.
- [13] 高芳, 张翼燕. 日本和韩国加快完善人工智能发展顶层设计 [J]. 科技中国, 2018 (8): 88-92.
- [14] 王喜文. 日本政府发布《机器人新战略》[N]. 中国电子报, 2015-03-31 (3).
- [15] 江丰光, 熊博龙, 张超. 我国人工智能如何实现战略突破——基于中美 4 份人工智能发展报告的比较与解读 [J]. 现代远程教育研究, 2020, 32 (1): 3-11.
- [16] 推进产业转型升级 推动制造业高质量发展 [N]. 河南日报, 2019-11-28 (6).
- [17] 张钹. 走向真正的人工智能 [J]. 卫星与网络, 2018 (6): 24-27.
- [18] 曾毅, 刘成林, 谭铁牛. 类脑智能研究的回顾与展望 [J]. 计算机学报, 2016 (1): 212-223.
- [19] 尚瑛杰, 何虎, 杨旭, 等. 仿生型脉冲神经网络学习算法和网络模型 [J]. 计算机工程与设计, 2020, 41 (5): 1390-1397.
- [20] Edmonds M, Gao F, Liu H, et al. A tale of two explanations: Enhancing human trust by explaining robot behavior [J]. IEEE Robotics, 2019, 4(37): eaay4663.
- [21] Phillips, P. Jonathon & Hahn, Carina & Fontana, Peter & Broniatowski, David & Przybocki, Mark. Four principles of explainable artificial intelligence [EB/OL]. [2021-02-25]. <https://nvlpubs.nist.gov/nistpubs/ir/2020/NIST.IR.8312-draft.pdf>.

The Development of AI Technology and the Technological Layout of Major Countries

LI Ruo-chen¹, LI Meng-wei²

(1. Institute of Intelligent Video Audio Technology, Shenzhen 518116;

2. Institute of Scientific and Technical Information of China, Beijing 100038)

Abstract: This article reviews three technical routes' research status, which exist in the history of artificial intelligence: symbolism AI, connectionism AI, actionism AI, then discusses the layout of the world's major countries in the field of artificial intelligence. By comparing the strategic policy documents of AI issued by China, the United States, Japan, and South Korea, it analyzes the similarities and differences in the technological layout of the four countries in the field of artificial intelligence. Furthermore it analyzes the knowledge-driven, brain-inspired intelligence, and AI interpretability technologies that are the common concern of all.

Keywords: artificial intelligence; knowledge driven; brain-inspired Intelligence; AI interpretability; technical layout