基础智能:从联邦智能到基于TAO的智能系统联邦

王飞跃[1,2],缪青海[1*]

1. 中国科学院大学人工智能学院,北京 100049
2. 中国科学院自动化研究所复杂系统管理与控制国家重点实验室,北京 100190

摘要　从早期使用数学工具探索神经活动机理,到采用人工神经网络模型模拟人类智能,再到基于深度网络解决特定问题,人工智能历经80年发展,当前已进入大模型时代,其特点是立足基础大模型、服务广大社会群体。基础大模型的发展面临一系列挑战:一是大模型训练所必需的数据,质与量有待提高;二是高端算力设备受限制,模型训练高电耗、高碳排放带来环境问题;三是大模型的社会化带来歧视与偏见、虚假信息传播、隐私侵犯、影响就业等社会发展问题。在当今去全球化的国际形势下,为克服以上挑战,健全包含联邦数据、联邦控制、联邦管理和联邦服务的联邦生态系统,构建基于区块链保真(trustable+reliable+usable+effective/efficient,TRUE)和去中心化自治组织和分布式自治运营(DAO)的智能系统联邦,发展基于TAO(TRUE DAO)的基础智能(foundation intelligence,FI),为大模型时代人工智能的发展提供必要和有力的支撑。

关键词　大模型;基础智能;联邦智能;TAO;智能系统联邦

2023年,大模型[1]成为人工智能(AI)领域继AlphaGo[2]之后最受关注的研究热点。以ChatGPT[3-4]、LLaMA[5]为代表的大模型,以流畅自然的对话能力、广泛强大的知识融合能力、全面条理的方案生成能力、合法合规的内容输出能力,成为最受欢迎的智能工具,已经深刻影响到社会工业生产和人们的日常生活,也成为人工智能历史上的里程碑式成果。创建基础模型、发挥基础模型的能力,帮助传统产业换代升级,正成为众多领域新的研究范式。

2023年9月初,文心一言(百度)、云雀(字节)、GLM智谱清言(智谱AI)、紫东太初(中科院)、百川(百川智能)、商量(商汤)、ABAB(MiniMax)、书生

收稿日期:2023-07-20;修回日期:2023-09-13
基金项目:国家重点研发计划项目(2018AAA0101502);国家自然科学基金项目(62271485)
作者简介:王飞跃,研究员,研究方向为平行系统、社会计算、知识自动化等,电子信箱:feiyue.wang@ia.ac.cn;缪青海(通信作者),副教授,研究方向为智能系统、平行智能、机器学习等,电子信箱:miaoqh@ucas.ac.cn
引用格式:王飞跃, 缪青海. 基础智能:从联邦智能到基于TAO的智能系统联邦[J]. 科技导报, 2023, 41(19): 103-112; doi: 10.3981/j.issn.1000-7857.2023.19.012

(上海人工智能实验室)、盘古(华为)、混元(腾讯)、星火(科大讯飞)等11个大模型获国家批准正式上线。除此之外,众多垂直领域的大模型如雨后春笋层出不穷,还有更多的开源模型正推进人工智能研究进入"百模争鸣""百模大战"的新局面。我们既乐见大模型的繁荣,也要深入思考大模型快速发展带来的问题。例如大模型重复建设带来的资源浪费问题、数据隐私和数据安全问题、大模型领域垄断带来的知识壁垒问题、基础大模型人类价值对齐所带来的意识形态问题、大模型生成的虚假内容的防范治理问题等。同时,在当前国际形势之下,如何统筹国内数据、算力和技术资源,建立人工智能基础大模型良序发展新生态,也是需要探讨的重要课题。

首先从研究范式转换的角度,简要综述人工智能(特别是深度学习)从人工神经元模型到GPT大模型的发展。探讨分析当前大模型技术发展中存在的问题,进一步阐述基于TAO的基础智能联邦生态发展框架[6-9]。

1 AI范式转变:从M-P模型到基础大模型

从1943年M-P神经网络模型被提出以来,人工智能已经走过80年的历程[10]。伴随着科学技术的发展,人工智能从早期对生物神经元网络工作机理的探索起源,过渡到通过人工神经元网络模拟和复现生物智能,再到使用深度学习方法解决视觉感知、自然语言处理等特定问题,人工智能几度沉浮,研究范式几经转变,以ChatGPT为代表的大模型正推进第五次工业革命,将引起人类社会的深刻变革[11]。

1.1 揭示智能机理

人工神经元数学模型的研究可以追溯到20世纪30年代。美国芝加哥大学的苏联裔科学家拉舍夫斯基,创建了数学生物学和数学生物物理学,首次提出了神经元网络的数学模型。随后,神经生理学家麦克洛克(McCulloch)与青年数学家皮茨(Pitts,曾经跟随拉舍夫斯基学习)合作,将逻辑作为数学语言,使用逻辑演绎描述神经元之间信息的传输和响应,提出了M-P模型,成为人工智能中广泛采用的人工神经元网络的源头。拉舍夫斯基、麦克洛克研究神经元数学模型的出发点,都是揭示人类智力产生的机理,即物质的大脑如何产生思想这一本质问题。对神经元数学模型的评价标准是看是否能复现生物神经元的功能。尽管这些模型并不能完全解决生物神经元工作机理问题,例如1943年提出的M-P神经元网络并不能解释循环因果(circular causality),但这些早期的探索,成为人们进一步研究机器智能的启蒙,为机器学习、控制论等新的学科领域打开了大门[10]。

1.2 模拟人类智力

20世纪40年代末,阿兰·图灵发表《计算机与机器智能》论文,阐述数字计算机的设计、机器能否思考的评判方法以及如何使机器学习从而具有类人智慧的可能途径。图灵关于计算机和机器智能的探索,拉开了人工智能研究的序幕。20世纪50年代,罗森布莱特发明的感知器模型(perceptron)成为突破性成果,感知器模型在结构上与M-P模型相似,但它引入了神经元节点连接之间的权重,首次使机器具备了学习能力,表现出视觉判断的能力。感知器模型的成功点燃了全社会对机器智能的憧憬,引发了人工智能历史上的第一次高潮。很快人们对机器智能的期望超出了其感知器模型的能力,随后明斯基出版的专著证明了感知器模型在逻辑运算能力上的不足,社会关注开始减退,人工智能进入漫长的寒冬期。1973年,英国剑桥大学卢卡斯讲席教授Lighthill提交给英国皇家学会的调查报告,对人工神经网络用于机器智能的实际进展作出了负面评价。转机到80年代初才开始出现。1982年,美国加州理工学院物理学家Hopfield提出一种对称全连接反馈神经元网络,使用电子元器件实现的Hopfield网络具有联想记忆能力,重新唤起了人们对人工神经元网络的关注。除此之外,Hopfield建议科研机构重新审视过去20年人工智能的研究成果,推动了人工智能从寒冬期中的复苏。Hinton等受热力学启发,在Hopfield网络的基础上引入随机性,提出了玻尔兹曼机,随后发展出

深度信念网络,为深度学习和生成式模型的发展奠定了基础。1986年,加州大学圣迭戈分校(PDP学派)心理学家Rumelhart与Hinton等一起重新提出多层感知器模型的反向传播算法,使多层神经元网络具备了以任意精度逼近任意函数的万能逼近能力,人工智能在这一时期获得了长足进步。

1.3　解决具体问题

20世纪90年代以来,人工智能的研究既关注算法的创新也重视模型的实际应用。其中的代表是由福岛邦彦、Yann LeCun、Hinton、Bengio、Alex Krizhevsky、孙剑等不断发展和改进的卷积神经网络(convolutional neural network,CNN)。通过神经元网络结构的巧妙设计,CNN在计算机视觉任务上飞速发展。早期的代表性工作之一是LeCun发展的字符识别卷积神经网络,在银行支票手写字体识别中得到广泛的应用。

21世纪初是深度学习爆发的前夜,互联网规模日益增长、处理器遵循摩尔定律飞速增强,为人工智能提供了数据和算力基础。2006年,Hinton提出深度网络的训练算法,2012年,基于图形处理器(graphics processing unit,GPU)训练的AlexNet(由Alex Krizhevsky等开发)在ImageNet比赛中大幅胜出传统机器学习方法,2015年,何凯明等提出的ResNet在ImageNet数据集上的图像识别精确度首次超过了人类。2016年,AlphaGo战胜围棋顶级大师,攻克棋类博弈游戏领域的最后一座堡垒,数据、算法、算力的结合推动人工智能走上新的巅峰[2]。在此背景下,基于CNN的神经网络促使计算机视觉成为最先落地、能够解决实际问题的人工智能领域,在安防、交通、农业、军事等领域中发挥越来越重要的作用,发展人工智能成为国家战略。

进入21世纪第二个10年,IT企业取代高校研究所成为AI算法与应用创新的领先势力。谷歌、Facebook(现Meta)、微软、亚马逊等公司依靠其数据资源方面的优势,不断推出新模型的同时,打造了TensorFlow、PyTorch等工具链,通过GitHub等开源共享途径,人工智能的生态系统不断完善。

2010年以来,数据、算法、算力三大要素快速发展,推动人工智能进入新的高潮。总体看,这一时期的人工智能研究,主要特点是面向特定问题、基于固定数据集,通过设计损失函数、搜索网络结构、优化模型参数,以期提高模型的性能指标。

1.4　服务社会群体

长期以来,计算机视觉和自然语言处理是深度学习的两个重要方向。计算机视觉广泛采用卷积神经网络模型,由于图像数据的普适性和GPU并行计算的支持,其在应用方面更加广泛和成熟。与之相对,自然语言处理领域常见的文本、语音都是序列化数据,在2017年之前的主要网络模型是循环神经网络(recurrent neural network,RNN)和长短期记忆网络(long short-term memory,LSTM),记忆能力有限,也不能发挥GPU的并行计算能力,因此发展相对缓慢,领域缺少重量级应用。2017年以来,谷歌提出的注意力机制和Transformer模型使自然语言处理领域快速发展,形成超越计算机视觉之势。

随后,谷歌和OpenAI先后推出BERT和GPT,如图1所示,大模型时代正式拉开帷幕。基于Transformer和Encoder-Decoder架构,采用无监督(或半监督)训练方法,大语言模型已经发展成为分支庞大的模型家族。其中,采用decoder-only形式的GPT是典型代表。2020年,OpenAI发布的GPT-3基于45 TB的数据进行训练,模型的参数量达到1750亿。2022年11月底,OpenAI发布ChatGPT,在5天内用户数量飞速增长到百万,2个多月突破1亿,创造了前所未有的纪录,成为继AlphaGo之后人工智能发展史上又一个里程碑。

与前大模型时期面向特定问题、基于固定数据集、端到端优化深度模型在特定任务上的性能指标不同,大模型时期的研究范式已经发生变革,具体有以下特点:一是强调大模型的基础能力,其训练基于人类社会积累的海量数据,通过凝练人类群体智慧,实现一个模型胜任多项任务;二是强调大模型的社会责任,除准确率等传统性能指标之外,更加重视生成内容的合规合法,通过人工参与训练过程,达到生成内容与人类价值观的对齐。大模型学习人类社会产生的数据,训练过程经过人类反馈与人类社会价值对齐,进一步为广泛的社会群体提供服务,新的数据又可以为大模型的迭代升级提供学

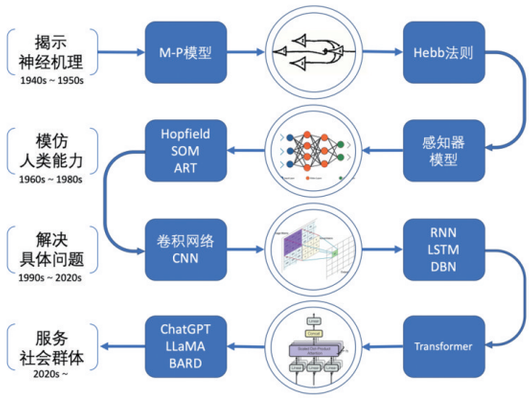

图1 人工智能的发展进入大模型时代

习材料,由此形成循环因果,是维纳《控制论》思想回归人工智能研究的体现。

当前,基础大模型进入百家争鸣局面。OpenAI拥有一定的先发优势,其GPT系列自GPT-2之后已不再开源,已经建立起技术壁垒。除此之外,训练大模型需要大量的高性能GPU计算资源,其高额的费用将大多数高校研究所的团队拒之门外。自Meta开源LLaMA大模型之后,为开源社区注入活力,引导大模型释放特定能力的低参数训练方法,包括prompt learning,instruction learning等,也为广大研究人员提供了舞台。国内大模型研究也呈现快速发展之势,百度、科大讯飞等企业有长期的积累,智源、中国科学院自动化研究所、鹏城实验室等也推出了悟道、太初、盘古等各具特色的基础大模型。

1.5 循环因果、平行系统与基础智能

从1943年M-P神经元网络的提出,人工智能已历经80周年的发展。几代研究人员提出众多的模型、算法,其中一些成为人类社会科技发展中的高光时刻。这些模型和算法固然重要,但它们背后的核心思想更应该得到提炼,并为人工智能研究开拓新方向提供指引。从这个角度看,从循环因果到平行智能,再到如今的大模型、基础智能,是一个自然的过程。

循环因果。麦克洛克从学生时代开始关注神经系统的循环回路,特别是神经闭环的工作机理。1939年,他在研究中通过实验发现了感觉皮层和丘脑之间的神经元闭环。1942年之前,罗森勃吕特(Rosenblueth)、维纳和毕格罗(Bigelow)通过对"目的论"的研究,指出因果关系研究的传统方法无法完全理解生物和行为现象。他们认为,理解行为需要考虑反馈机制和目标导向过程,并强调了将内部状态和外部环境有机体作为一个整体考虑的重要性,并在1943年初发表了《行为、目的和目的论》一文。在1943年的M-P模型论文中,麦克洛克受维纳等人提出的循环因果论启发,将逻辑神经网络与因果推理系统等同起来,使神经网络有了自己的数学模型。实际上,通过梅西会议,维纳等研究人员将"循环因果"的思想和"反馈"机理从神经生物学、机器控制、人类社会等学科中抽象升华出来,创建了"控制论"。控制论的核心思想,以及维纳后续著作中关于人、社会的论述,也是现代人工智能的核心内涵,在后续的BP算法、平行智能、ChatGPT中都得以体现。

平行系统。在20世纪七八十年代,人工智能发展陷入低谷,但循环因果和反馈机制的思想却一直未曾消逝。1994年,王飞跃提出了影子系统(shadow systems)和世界模型(world model)等概念,专注于探究控制论在社会和物理系统中的应用。2004年,王飞跃提出的平行系统和平行智能理论[12-14]进一步展示了循环因果和反馈机制在新时代的延伸与扩展。平行系统(parallel systems)由现实世界中的1个实际系统和1个或多个虚拟或理论人工系统组成。这些人工系统对实际系统进行了软件化定义,不仅是数字仿真,还可用作实际系统的可替代版本,当今的大模型即是人工系统的一种具体形式。平行系统的主要目标是通过实际系统与人工系统之间的互联,实时动态比较与分析二者的行为,以一种虚实互动的方式,为各自未来状态的"借鉴"和"预估"提供支持。这种方法可以有效解决问题、支持学习和培训目标,实现人工引导实际,实际逼近人工的目的,这也是当前训练大模型时采用的基于人类反馈的强化学习(reinforcement learning from human feedback,RLHF)的基本思想。

基础智能。当前,ChatGPT的成功标志着人工智能迈入了以大模型、场景工程[15]为基础的foundation intelligence时代[16],其中循环因果和反馈机制

成为ChatGPT中指导学习和对齐训练的核心理念。具体而言,以ChatGPT为代表的大型模型,具备一定的通用人工智能能力。然而,为了确保这些大型模型的输出符合人类的伦理、道德和法律准则,必须对其输出进行引导和控制。一方面,强化学习与人类反馈通过语言作为媒介,建立了一个大型模型和人类进行交互的闭环回路。通过负反馈,这个回路用来校正大型模型的生成内容,使其与人类社会的规范相符。另一方面,大型语言模型是通过来自人类的海量文本数据训练的,这使得它们成为描述现实世界的工具,与现实世界构成了一种平行系统。平行智能中的循环因果和反馈控制思想为大型模型时代的人工智能研究提供了一个重要的研究框架。

2　大模型时代的问题与挑战

基础大模型在内容生成、问答、推理等方面表现出前所未有的能力,因此被广泛认为是通向通用人工智能的途径。但是,大模型也带来了一系列新的问题,如图2所示,对技术进步、知识共享、国际合作、社会公平带来新的挑战[17]。

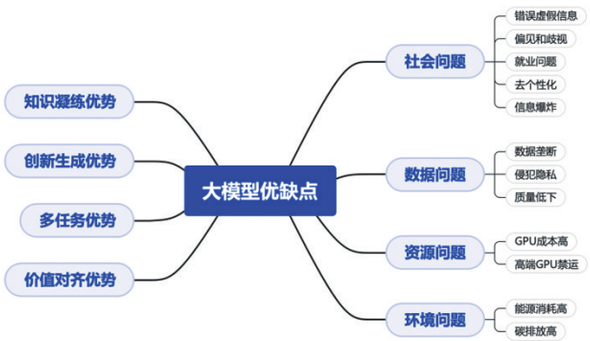

图2　大模型的特点及存在的挑战

2.1　资源消耗与环境保护问题

大模型的特点是参数量大、训练所用的数据集大,因此大模型对算力设施、电力供应等资源有很高的需求,由此也带来高投资、高能耗、高排放等问题。

GPU等计算资源投入成本高。训练深度学习模型,尤其是当前的大模型,通常需要大量的计算能力。GPU因其强大的并行处理能力和为深度学习定制的硬件加速单元(例如tensor core),已经成为人工智能必不可少的基础设施。根据模型的大小和数据量,大模型的训练(例如ChatGPT、LLAMA等)往往需要数千块GPU几天甚至数十天的运行时间。如此规模的GPU阵列需要高额投资,购置和维护费用通常只有少数大型企业才能承担。

训练推理能源消耗高。训练人工智能模型的高计算需求导致大量能源消耗。配备大量GPU的数据中心在训练期间会消耗大量电力。这种能源使用增加了人工智能研发的运营成本。与人工智能模型训练相关的能源消耗可能会导致碳排放,特别是如果所使用的电力来自煤炭或天然气等化石燃料。许多组织和研究人员已经意识到大模型训练对环境的影响,并正在采取措施,包括使用更节能的硬件、优化算法以及使用可再生能源为数据中心供电。由于涉及硬件效率、能源和模型寿命等因素的复杂性,准确量化人工智能训练对环境的影响可能具有挑战性。尽管如此,大量GPU的使用所带来的碳排放正在持续增加,这种碳足迹对气候变化潜在的影响令人担忧,评估和减少人工智能研究和部署的碳足迹应该得到足够的重视。

高端计算资源受限。由于美国高技术封锁,人工智能模型训练最常用的英伟达(NVIDIA)GPU,特别是用于计算中心的高端型号A100/H100受到限制,国内的人工智能企业和研究机构GPU稀缺成为日益突出的问题。

综上所述,训练大型人工智能神经网络模型GPU资源需求大,能源消耗量大,并产生碳排放。人们正在不断努力使人工智能研究和开发在环境上更加可持续,但这些挑战是人工智能领域的一个重要考虑因素。

2.2　数据垄断与隐私保护问题

深度学习模型,特别是自然语言处理(NLP)和计算机视觉中使用的大型神经网络,需要多种模态大量数据进行训练。模型从大数据中学习其中蕴含的模式、相关性和特征,进一步执行各种任务,例如图像识别、语言翻译和推荐系统等。大模型的训练效果更加依赖于数据的数量和质量。然而,高质量大数据的获取并不容易,存在数据垄断和隐私保

护等一系列问题。

数据垄断:数据垄断是指少数公司或组织控制大型数据集的访问和所有权的情况。这可能会造成人工智能领域的权力不平衡,只有少数实体拥有训练高效模型所需的资源和数据。这种主导地位会抑制竞争和创新。

隐私问题:收集和使用大型数据集进行人工智能训练可能会带来严重的隐私问题。当组织收集大量用户数据时,存在敏感信息可能被错误处理、面临安全漏洞或以个人不同意的方式使用的风险。这引发了有关数据隐私和用户同意的重要道德和法律问题。

为了解决这些问题,各国政府和监管机构已经实施或提出了数据隐私法规,例如欧盟的《通用数据保护条例》(GDPR)。这些法规旨在保护个人隐私权,增强数据安全,并为个人提供对其数据的更多控制权。研究人员正在积极研究隐私保护技术,使人工智能模型能够在敏感数据上进行训练,而不会损害个人隐私。这些技术(例如联邦学习和同态加密),使模型能够从分布式数据源中学习,同时保持数据本身的安全和私密。

综上所述,大型深度学习模型的训练依赖于大数据,这可能会导致数据垄断和隐私保护等问题。人们应通过法规、隐私保护技术以及提高人工智能社区对数据道德和隐私重要性的认识来解决这些问题。

2.3 大模型潜在的社会问题

ChatGPT等大模型的特点是通过人类反馈实现对齐[4],尽可能达到符合伦理、合规、合法。尽管如此,这些大语言模型仍可能会产生消极的社会影响。

错误和虚假信息:大语言模型的基本原理是概率生成,不可避免地存在臆想、编造输出的问题。语言模型可以生成连贯且语法正确的文本,但其内容语义却是错误或虚假的。由于区分人类生成的内容和人工智能生成的内容可能会变得更具挑战性,这些错误信息和虚假信息如果被恶意使用并在网络中传播,将会对个人声誉或社会稳定造成难以预料的后果。

偏见和歧视:语言模型是在大型数据集上进行训练的,这些数据集可能包含来自互联网的有偏见或成见的内容。这可能会导致模型产生有偏见或歧视性的输出,从而加剧现有的社会偏见,导致社会群体之间的对抗,甚至恶化民族、国家之间的关系。

就业问题:随着人工智能语言模型的改进,它们可能会用于客户服务、内容生成和其他领域,可能会导致这些角色的人类工作被取代。

创造力降低:过度依赖人工智能进行沟通、决策或创作,可能会导致个性化消失,降低人与人之间的关系和互动的质量,削弱人的创造力。

信息爆炸:人工智能模型生成大量文本可能会导致信息过载,使个人难以筛选并找到可信的信息。

值得注意的是,这些潜在问题并不是ChatGPT等人工智能模型固有的,而是它们特定的开发、部署和使用方式的结果。研究人员、开发人员和政策制定者应给予这些问题足够的关注,制定道德准则、法规,履行负责任的人工智能实践,以减轻负面社会影响并促进负责任地使用人工智能大模型。

3 基础智能:从联邦智能到基于TAO的智能系统联邦

当前,人工智能发展进入大模型时代。一方面,大模型是大数据、强算力、新算法不断发展融合的产物,对自主开发大模型提出更高的要求;另一方面,基础大模型展示出来的生产力,正推动人类社会进入第五次工业革命,大模型的研究和应用水平也成为国家科技发展水平的重要标志。如何统筹数据和算力资源、加快核心算法创新、促进工业和社会应用、服务各行业群体需求,在充分发挥大模型能力的同时避免大模型所带来的环境、社会问题,出台保障措施和激励政策,成为中国面临的重要课题。

3.1 联邦生态系统

为了解决人工智能中存在的数据与资源等问题,中国科学院自动化研究所王飞跃等提出了面向人工智能研究的联邦生态[18]框架体系,由联邦数据、联邦控制、联邦管理、联邦服务等部分组成。在数据隐私、信息安全、资源整合等驱动下,联邦生态

本质上是由联邦安全、联邦共识、联邦激励、联邦合约等一系列基于区块链的技术构建而成。

1）联邦数据。

联邦生态可以为人工智能时代数据隐私保护和信息安全需求带来的严重的数据孤岛问题提供有效的解决方案。联邦数据作为联邦生态的数据基础,包括联邦内所有节点的数据及其存储、计算和通信资源。为了保护隐私,联邦数据分为私有数据和非私有数据,通过对这些数据的联邦控制,可以实现数据联邦化[19-20]。

在数据驱动的人工智能技术中,联邦数据发挥着重要作用,它可以帮助基于人工智能的应用实现有效的数据检索、预处理、处理、挖掘和可视化。其核心功能,是为大模型训练所面临的数据缺失、数据质量低、数据受版权保护等问题提供有效的解决方案,在保护个人隐私、数据版权的同时实现有效数据共享,为公有基础大模型提供数据保障。

2）联邦控制。

联邦控制作为联邦生态的核心执行部分,旨在维护信息安全,保护数据的所有权、控制权、隐私权、使用权[21-22]。联邦控制是一种分布式控制策略,为大型复杂系统提供高效、安全、可靠的控制和管理。

在联邦控制过程中,私有数据被限制在本地节点,而非私有数据的所有权和使用权分离,从而保证各子系统的信息安全和权益保护。联邦控制直接响应联邦智能的需求,为联邦数据提供安全保护。联邦控制通过代码或脚本定义的联邦合约,实现数据的联邦,建立联邦数据的控制体系,包括联邦数据的存储、传输、共享和使用,同时保证联邦数据的安全。在联邦生态中,联邦数据是联邦控制的首要要素。联邦控制的目标是通过调度和控制各节点数据公共信息的访问状态和流向,实现联邦数据的信息安全和权益保护,打破数据孤岛,实现数据联邦化[23]。

3）联邦管理。

联邦管理是联邦生态框架的核心组成部分之一[24-25]。联邦管理的主要任务是根据联邦生态的总体目标和要求做出管理决策。而且,联邦管理决策会根据联邦生态系统状态的变化实时动态调整。通过联邦管理决策,联邦生态能够达到最优状态,更好地实现其目标,从而在保证联邦安全的前提下,实现整个联邦生态系统的智能化管理。一方面,联邦管理可以通过联邦数据的联邦控制和管理,实现个性化的联邦服务。另一方面,联邦管理可以通过基于区块链的联邦合约、联邦激励、联邦共识等技术实现联邦安全,最终在保护隐私和数据的前提下将联邦数据转化为联邦智能安全。

在数据安全和隐私保护的前提下,联邦管理聚集更多的数据、计算能力和人力资源,产生科学可靠的管理决策,从而提高管理效率和效果。在人工智能技术和区块链技术的支持下,更多的数据以联邦数据的形式聚集,通过联邦管理的方式转化为决策,通过联邦控制的方式落实到具体措施,再通过联邦服务赋能,从而实现从数据到知识再到智能的进化。联合智能将单个组织的个体智能转化为多个组织的群体智能。联邦管理促进联邦智能的实现和发展,联邦智能又可以辅助联邦管理的决策。

4）联邦服务。

联邦管理的目的是通过对联邦数据的联邦控制来实现联邦服务[24]。因此,联邦数据是联邦管理的数据基础,也是实现联邦服务的数据保障。通过设计一系列联邦管理规则,在保证联邦节点数据安全和隐私的前提下,通过对联邦数据的管理和控制,实现联邦服务。同时,在实现联邦服务的过程中,不断产生大量新的数据,这些数据可以添加到联邦数据中,用于优化联邦管理决策。

联邦生态的框架与方法已经在工业控制[26]、交通物流[27]、社会人口[28-29]等领域得到成功的应用。

3.2　从联邦智能到基于TAO的智能系统联邦

不同于处理分布式场景中存在问题的现有技术,例如分布式存储、边缘计算、区块链技术、联邦学习技术,在分布式场景中只能关注某一方面,缺乏系统的统筹考虑和协调,而联邦生态可以处理从数据生产到数据使用再到服务用户的全过程问题。联邦生态是基于智能生态系统研究思想而提出的,具有将数据转化为智能的能力。它不仅适用于中心主导节点的联邦,也适用于中心节点弱化或缺失的

联邦。通过联邦生态,联邦节点之间可以通过松散的联盟建立合作关系,可以加强各节点的隐私保护,调动联邦节点的积极性,提高联邦成员的参与度,从而可以更好地提高联邦的整体绩效。区块链("真",TRUE)和分布式自治运营("道"DAO),如图3所示,两者结合即为TAO,形成完整的可信数据、算法、运作的生态系统,为联邦生态到智能联邦的发展提供了保障。

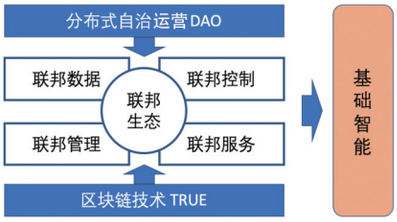

图3 基础智能:基于TAO(TRUE+DAO)的智能系统联邦

区块链技术。区块链的作用是保真(TRUE),代表着可信(trustable)+可靠(reliable)+可用(usable)+效益(effective,efficient)。联邦控制是在基于区块链技术的联邦安全、联邦共识、联邦激励、联邦合约的支持下实现的。联邦安全源于区块链中的安全机制,在联邦数据的加密、传输和验证中发挥着至关重要的作用。联邦共识是为了保证所有联邦节点之间策略、状态、更新的分布式共识。联邦激励是为了设计快速、稳定、正向的激励措施,平衡联邦节点之间的利益,激发联邦节点的活跃度,提高联邦控制系统的效率。联邦合约基于智能合约算法,自动、安全地实现联邦控制。联邦区块链的联邦合约主要作用于访问控制、非私有联邦数据交换、本地和全局数据更新、事故处置等。

DAO (decentralized autonomous organizations and distributed autonomous operations)。一些新思想、新技术的出现,给范式创新带来机遇。例如,由Web 3.0和区块链技术推动的去中心化科学(decentralized science, DeSci)浪潮正在给科学研究的组织方式带来变化[20]。随着人工智能研究进入快速迭代,人们呼吁建立新的研究机制来克服传统科学合作中缺乏透明度和信任等挑战,并实现更高效、更有效的科学发现。DeSci致力于创建一个去中心化、透明且安全的网络,供科学家共享数据、信息和研究成果。DeSci的去中心化特性使科学家能够以更加公平和民主的方式合作。去中心化自治组织和分布式自治运营DAO作为DeSci的实现方式,为人工智能创新和应用提供了新的组织形式[21]。DAO是一种通过代码运行并运行在区块链网络上的数字组织,这意味着DAO的成员拥有做出决策和执行行动的投票权,使其成为一个民主和透明的系统。当前,基于DAO的一系列工作已经显示出未来发展潜力。

4 结论

回顾了人工智能80年来的发展,包括20世纪三四十年代揭示神经元网络工作机理,50年代到80年代模拟动物和人类智能,90年代之后基于神经元网络模型解决特定领域的问题。当前,人工智能发展进入服务社会广大群体的大模型时代。在新时代、新形势下,中国发展大模型面临着高端GPU禁运、大数据质与量不足、环境保护问题、社会发展问题等一系列挑战。为有效应对这些挑战,提出发展基于联邦数据、联邦控制、联邦管理、联邦服务的人工智能联邦生态,以TAO为技术支撑,生成基础智能,构建智能系统联邦,促进中国人工智能新时期良序发展。

参考文献(References)

[1] Zhou C, Li Q, Li C, et al. A comprehensive survey on pretrained foundation models: A history from BERT to ChatGPT[J]. arXiv, 2023, arXiv:2302.09419.

[2] Wang F Y, Zhang J J, Hang X, et al. Where does AlphaGo go: From Church-Turing thesis to AlphaGo thesis and beyond[J]. IEEE/CAA Journal of Automatica Sinica, 2016, 3(2): 113-120.

[3] Wang F Y, Miao Q H, Li X, et al. What does ChatGPT say: The DAO from algorithmic intelligence to linguistic intelligence[J]. IEEE/CAA Journal of Automatica Sinica, 2023, 10(3): 575-579.

[4] Ouyang L, Wu J, Jiang X, et al. Training language models to follow instructions with human feedback[J]. arXiv,

2022, arXiv:2203.02155.

[5] Touvron H, Lavril T, Izacard G, et al. LLaMA: Open and efficient foundation language models[J]. arXiv, 2023, arXiv:2302.13971.

[6] Wang F Y, Ding W W, Wang X, et al. The DAO to DeSci: AI for free, fair, and responsibility sensitive sciences [J]. IEEE Intelligent Systems, 2022, 37(2): 16-22.

[7] Ding W W, Hou J, Li J J, et al. DeSci based on Web3 and DAO: A comprehensive overview and reference model [J]. IEEE Transactions on Computational Social Systems, 2022, 9(5):1563-1573.

[8] 李娟娟, 秦蕊, 丁文文, 等. 基于Web3的去中心化自治组织与运营新框架[J]. 自动化学报, 2023, 49(5): 985-998.

[9] Miao Q H, Zheng W B, Lv Y S, et al. DAO to HANOI via DeSci: AI paradigm shifts from AlphaGo to ChatGPT [J]. IEEE/CAA Journal of Automatica Sinica, 2023, 10(4): 877-897.

[10] 缪青海, 王雨桐, 吕宜生, 等. McCulloch-Pitts人工神经元模型80周年纪念:思想、方法与意义[J]. 智能科学与技术学报, 2023, 5(2): 133-142.

[11] 王飞跃, 缪青海. 人工智能驱动的科学研究新范式:从AI4S到智能科学[J]. 中国科学院院刊, 2023, 38(4): 536-540.

[12] 杨静, 王晓, 王雨桐, 等. 平行智能与CPSS: 三十年发展的回顾与展望[J]. 自动化学报, 2023, 49(3): 614-634

[13] Miao Q H, Lv Y S, Huang M, et al. Overview and perspective for computational learning across Syn2Real and Sim2Real[J]. IEEE/CAA Journal of Automatica Sinica, 2023, 10(3): 603-631.

[14] Wang F Y. Parallel intelligence in Metaverses: Welcome to Hanoi[J]. IEEE Intelligent Systems, 2022, 37(1): 16-20.

[15] Li X, Tian Y L, Ye P J, et al. A novel scenarios engineering methodology for foundation models in Metaverse [J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2023, 53(4): 2148-2159.

[16] Wang F Y. New control paradigm for Industry 5.0: From big models to foundation control and management[J]. IEEE/CAA Journal of Automatica Sinica, 2023, 10(8): 1643-1646.

[17] 卢经纬, 郭超, 戴星原, 等. 问答ChatGPT之后: 超大预训练模型的机遇和挑战[J]. 自动化学报, 2023, 49(4): 705-717.

[18] Wang F Y, Qin R, Chen Y Z, et al. Federated ecology: Steps toward confederated intelligence[J]. IEEE Transactions on Computational Social Systems, 2021, 8(2): 271-278.

[19] Wang F Y, Zhang W S, Tian Y L, et al. Federated data: Toward new generation of credible and trustable artificial intelligence[J]. IEEE Transactions on Computational Social Systems, 2021, 8(3): 538-545.

[20] 李娟娟, 王晓, 袁勇, 等. 基于区块链技术的数据共享方法研究进展[J]. 复杂性与智能化, 2021, 17(3): 30-37.

[21] Wang F Y, Zhu J, Qin R, et al. Federated control: Toward information security and rights protection[J]. IEEE Transactions on Computational Social Systems, 2021, 8(4): 793-798.

[22] 朱静, 王飞跃, 田永林, 等. 联邦控制: 面向信息安全和权益保护的分布式控制方法[J]. 自动化学报, 2021, 47(8): 1912-1920.

[23] Wang F Y. The DAO to MetaControl for MetaSystems in Metaverses: The system of parallel control systems for knowledge automation and control intelligence in CPSS [J]. IEEE/CAA Journal of Automatica Sinica, 2022, 9(11): 1899-1908.

[24] Wang F Y, Qin R, Li J J, et al. Federated management: Toward federated services and federated security in federated ecology[J]. Transactions on Computational Social Systems, 2021, 8(6): 1283-1290.

[25] Li J J, Qin R, Wang F Y. The future of management: DAO to smart organizations and intelligent operations[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2023, 53(6): 3389-3399.

[26] Wang X X, Yang J, Wang Y T, et al. Steps toward Industry 5.0: Building "6S" parallel industries with cyber-physical-social intelligence[J]. IEEE/CAA Journal of Automatica Sinica, 2023, 10(8): 1692-1703.

[27] Lin Y L, Na X X, Wang D, et al. Mobility 5.0: Smart logistics and transportation services in Cyber-Physical-Social systems[J]. IEEE Transactions Intelligent Vehicels, 2023, 8(6): 3527-3532.

[28] Ye P J, Wang F Y. Parallel population and parallel human-A Cyber-Physical social approach[J]. IEEE Intelligent Systems, 2022, 37(5): 19-27.

[29] Wang F Y, Qin R, Wang X, et al. MetaSocieties in Metaverse: MetaEconomics and MetaManagement for MetaEnterprises and MetaCities[J]. Transactions on Computational Social Systems, 2022, 9(1): 2-7.

Foundation intelligence: From federated intelligence to TAO-based intelligent systems federation

WANG Fei-Yue[1,2], MIAO Qinghai[1*]

1. School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China
2. State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

Abstract From the early use of mathematical tools to explore the mechanism of neural activity, to the use of artificial neural network models to simulate human intelligence, to the solution of specific problems based on deep networks, AI has developed over the past 80 years and has now entered the era of foundation intelligence (FI), which is characterized by its foothold in large models, serving the broad social groups. The development of large models faces a series of challenges. First, the quality and quantity of data necessary for large models training need to be improved. Second, high-end computing resource is limited, and model training has high power consumption and high carbon emissions, which brings environmental problems. Third, it is the socialization of large models that brings about social development problems such as discrimination and prejudice, the spread of false information, invasion of privacy, and impact on employment. In order to overcome the aforementioned challenges, it is essential to build a federal ecosystem (including federal data, federal control, federal management and federal services) with support of block chain and decentralized autonomous organizations (TRUE+DAO, TAO), and provide solutions for the development of artificial intelligence in the new era.

Keywords large models; foundation intelligence; federated intelligence; TAO; intelligent systems federation

（责任编辑　傅雪）