



郑庆华，中国工程院院士，同济大学校长、教授，智能网络与网络安全教育部重点实验室主任，教育部科技委学部委员，教育部大学计算机教指委主任，国家自然科学基金创新群体负责人。研究领域为大数据知识工程。先后获得国家科技进步二等奖3项，国家教学成果一等奖2项、二等奖3项，省部级一等奖5项以及何梁何利基金科学与技术进步奖、CAA 科技进步特等奖、中国科协“求是”杰出青年奖、中国青年科技奖、国务院政府特殊津贴专家等奖项和荣誉。

# 卷首语

Foreword

## 人工智能从弱到强的路径分析

郑庆华

人工智能作为一种新质生产力，在信息时代为我们提供了一种前所未有的分析世界、认识世界、改造世界的新方法新手段，甚至提供了一种新的思维方式。人工智能总体上沿着计算智能、感知智能、认知智能的逻辑方向发展，当然这三者并非线性地发展，而是螺旋式循环上升的。那么如何实现人工智能从弱到强的技术路径？

第1条技术路径是数据驱动的大模型。大模型是指参数规模超过10亿的神经网络模型，在内容生成、世界知识、代码的理解、生成方面，能够完成过去只有人才能完成的任务，被认为是人工智能从弱向强的标志。大模型的本质是运用强大算法消耗大量算力，从海量数据中训练出复杂概率分布函数。在高度认可人工智能从弱到强突破的同时，我们也深刻地感受到它依然存在着很多固有缺陷。第1个固有缺陷是过度消耗数据和算力，大模型的参数量已达到万亿级别，但由于模型训练和问题求解都需要所有参数参与运算，且训练数据规模和算力消耗与参数规模成正比。第2个固有缺陷是灾难性遗忘，迁移能力差。在新任务上训练会损害之前任务的性能；在问题求解阶段，无法记住处理过的数据或场景，存在“喜新厌旧”的问题。第3个固有缺陷是推理能力弱。大语言模型是大量相关性的叠加，本质上是“聪明汉斯”效应，形成不了因果关系；自回归生成方式也难以应对需要回溯与试错的复杂推理任务。第4个固有缺陷是大模型不知道自己错了，也不知道为什么错了，更做不到知错就改。正是这些固有缺陷，进一步导致大模型出现幻觉、可控性、鲁棒性差等问题，成为阻碍人工智能从弱到强的瓶颈难题。

第2条技术路径是“知识+数据”的大数据知识工程。大数据知识工程的特点是知识需融合，证据可解释，结果可溯因。这与大模型的应用场景、机理完全不同。因此我们受认识论“既见树木又见森林”的启发，提出了知识森林的概念，其核心是从散、杂、乱的海量大数据中，挖掘出碎片化知识，并通过从局部到整体、无序到有序融合生成结构化知识。“知识+数据”驱动的机器智能模型，既发挥符号知识可组合、可解释、可推理等特点，又发挥数据驱动内容生成能力强、泛化性强的优势，“知识引导+数据驱动”能够支撑多样化知识表示、自动化知识获取、可解释高阶推理、可回溯路径规划等功能。目前，这一方法已在多个领域得到实际应用。例如，一是在国家金税工程偷逃骗税检测上得到应用，2016年被国税总局认定为全国唯一实施方案，实现全国31个省市区全面部署，每年为国家挽回巨额税款流失。二是将知识森林应用到学科大模型的构建之中，首次生成了土木工程学科的知识图谱，总共包含了11408个知识点。三是应用到C919民机设计的海量文档知识管理之中，针对C919涉及的2万多个文档，我们设计了C919机相关的多模态知识森林，建立了具有可视化导航学习、知识问答、知识推送等功能的知识管理系统，该系统目前已上线服务。

第3条技术路线是机器记忆智能，即借助人脑的记忆机制设计一种新的机器智能模型，以攻克当前大模型的固有缺陷。人工智能的每一步发展都得到了脑科学的指导：借鉴人脑记忆的检索和激活机制，可以为设计智能巧算的新型机器智能模型带来启发，有望突破大模型过度消耗算力的缺陷；借鉴大脑突触可塑性，有望实现持续性学习，突破灾难性遗忘的缺陷；借鉴人脑双系统协同机制，有望突破黑盒模型逻辑推理能力弱的难题。创建人脑记忆启发的机器智能模型，重点要解决4个科学挑战：一是机器记忆的神经学机理是什么？二是如何依据不同类型的记忆，实现各类知识的联想式表征？三是如何借鉴突触可塑性和局部学习机制实现持续学习？四是如何借鉴人脑记忆的检索、激活、绑定等机制，设计兼具双系统协同的推理机制。

我们认为借鉴人脑记忆机制，建立涵盖“表征-学习-推理”的理论与技术体系，是一条能够真正破解大模型固有缺陷的路径，是一条开辟人工智能/机器智能的新路径，是一条研究“人脑记忆”与“机器记忆”关联性，探索以信息科学研究脑科学的新路径。