

# 人工智能基础问题：机器能思维吗？

## Artificial intelligence fundamental question: Can machines think?

李德毅

(军事科学院系统工程研究院, 北京 100141)

基础研究崇尚想象力和创造力的完全自由, 依赖独立学者的兴趣和自由合作, 它可以不限研究者的身份, 不设完成的时限, 不以落地应用为目的, 也不一定要组织大团队“攻关”, 不搞群众运动, 允许试错, 宽容失败, 更不以获得自然科学奖为目的; 需要研究者有深厚的人文艺术素养, 耐得住寂寞, 沉得下心来, 虽然研究结果和产出时间无法被精确预测, 但一旦出现原始创新, 对引领技术进步必然会有长期且深刻的影响。阿兰·图灵的研究就是一例。

1950 年图灵发出振聋发聩之问: “**机器能思维吗? (Can machines think?)**”, 也许感到这个问题中“机器”和“思维”难以被准确定义, 图灵自问自答了 18 年, 提出“模仿游戏(imitation game)”, 后来被称为“图灵测试”。人类围绕这个人工智能基础问题展开了起起落落的讨论, 物质科学家视为要利用自然界物质和能量发明智能机器的问题, 生命科学家视为如何克隆人工生命的问题, 人类学家和社会学家视为人类智能的体外延伸、以及社会的文化文明的生态问题, 认知科学家视为心智问题, 哲学家视为逻辑问题, 数学家视为计算问题……各路高手都用自己手里特别擅长的“锤子”砸向这颗耀眼的“钉子”。图灵也许没有想到 70 年后的今天, 人工智能会如此深刻地影响着人类生产生活, 如此深刻地影响着人类前途命运。我们应该沿着图灵指引的方向, 围绕“机器能思维吗?”这个基础问题, 与时俱进, **从数学的机械化走向认知的自动化**, 进一步聚焦, 有以下 10 个方面的思考。

**思考 1: 人工智能的边界设定** 尽可能获得大家共识的智能和人工智能的“定义”是什么? 还是就这样长期分散着自说自话? 常常发生的情况是鸡对鸭讲, 错位沟通, 难以汇聚力, 通过定义, 大家可以为智能设定共同的边界和外部约束, 这很重要。例如, 生物的本能是不是智能?

人工智能是不是仿生工程? 智慧和智能有什么差别? 个人的智能、群体的智能、人类的智能、人工的智能、机器的智能有什么差别? 维基百科把人工智能和机器智能说成是同义词, 1948 年图灵把智能分为具身智能(embodied intelligence)和体外智能(disembodied intelligence), 便于在体力和智力之间划出一个明显的分界线, 但是超强的、强的、弱的、通用的、可信的、可解释的、安全的人工智能可以分开去实现吗? 真的需要对健全人也做脑机接口吗? 存在什么样的人工智能伦理问题? 如果我们**把认知、思维或者智能统统定义为“学习的能力, 以及解释、解决问题的能力”, 把人工智能定义为“人类智能的体外延伸”, 是不是更聚焦了?**有什么闪失? 如果把智能分为感知智能、认知智能和行为智能。感知是认知的源头, 行为是具身智能, 是认知的外化表现; 认知是感知的高阶, 是感知的再抽象, 从认知看感知可看到里面的秩序和可解释性; 感知智能和行为智能发生在物理空间里, 和环境交互, 统称为交互智能; 没有交互, 就没有学习; 认知智能发生在认知空间里, 又分为记忆智能和计算智能。这些, 你认同吗?

**思考 2: 人类智能始于语言, 人工智能始于文字** 和其他生物相比, 认知究竟是如何成就今天的人类的? 如何看待达特茅斯会议之前就长期存在的人工智能相关研究? 应该说人类智能始于语言, 人工智能始于符号和文字, 语言文字承载人的思维和认知、思想和创造, 智能植根于教育, 人类延伸在体外的智能成为知识、文化和文明, 人工的智能已经存在几千年。人类认知的特殊性在哪里? 脑科学和认知科学能够为人工智能提供什么实在的指导? 人们讨论生物自然进化现象, 常用的时间尺度是“万年”; 讨论人类文明生态现象和体外智能, 常用的时间尺度是“千年”; 讨论人类思维和认知的进步, 即科学技术的发展, 常

用的时间尺度是“百年”甚至“十年”。如果说，人类的生物进化出的本能是相对稳定的，科学技术和人工智能的发展却是天翻地覆的，**如何看待这种迭代的智能导致的、几何级数增长的认知进步和知识爆炸？如何在智能机器里去递归地执行？**

### 思考 3：物质和精神、科技和人文将走向统一

— 人工智能和物质科学、生命科学、社会科学等哪些大门类的学科如何相关汇聚？如何进行交叉研究？物质和精神、意识和智能、科技和人文、心和智，**是一元论、二元论、还是多元论？**自然科学和社会科学、认知科学和哲学、认知科学和数学、自然语言和数学语言，谁从属于谁啊？

**思考 4：各智其智、智智与共** 人类的思维如何形式化？思维和认知有哪几种最基本的范式？时间和数学在其中如何发挥作用？存在唯一的认知“元范式”“元认知”吗？还是**各智其智、智人之智、智智与共的多元认知**？纵然是多元的，也不可能阻挡人类对统一理论的渴望和追求，形成无尽的认知螺旋。觉知（视觉、听觉、味觉、嗅觉、触觉等）和感知，觉悟和跨模态觉知形成的感悟，直觉智能和感知智能，主要发生在人体末梢和周边的知觉神经系统里，是在物理空间和客观事物交互中产生的；认知和顿悟发生在中枢神经系统、即大脑的皮层里，是在主观的认知空间或者思维空间里产生的。深度学习主要模拟的是感知智能，具有先天的不可解释性，可不可说依然是浅层认知？

**思考 5：抽象和抽象的层次** 思维的载体是什么？核心是什么？过程是什么？结果是什么？如果说，思维的载体是语言和文字，核心是抽象，过程是计算，结果是记忆，知识以记忆的形式存在，人的大脑里面沟沟回回的皮层形成认知空间，产生“想象的现实”。**多层次的抽象是如何工作的？**人工智能如何受脑科学和认知科学的启发？在脑科学没有取得重大的进展之前，如何实现体外的类脑智能？

**思考 6：幼儿认知核** 图灵在 1950 年发出“机器能思维吗？”的那篇著名论文《计算的机械与智能》中明确指出：“在我看来，生命与非生命之间的差别，远远大于人类和其他生命之间的差别”。我认为要研究生命智能和机器智能的差别，最大的差别就是生命有意识，机器没有，机器智能不必和生命体的意识、情感等纠缠；次大的差别就是生命依靠繁衍，记忆不可能从外部后天

植入；机器则可以批量生产，因此**体现进化烙印的遗传基因的记忆和长期记忆在智能机器中“幼儿认知核”可以且必须初始预置**。那么如何注入一个实体机器人的初始长期记忆？一次性预置多少记忆为好？至于说到人类智能和其他生物智能的差别，我认为人的本能和高等生物的本能难分伯仲，人类最伟大的壮举是发明了教育，人类智能的特殊性，首先表现在学习和终生学习的能力上。人后天习得的智能要远大于人的本能，认知迅速成长，而生物后天习得的智能极为有限，这才是最重要的，如何解释这个现象呢？

**思考 7：人工记忆网络的小世界特征** 机器学习和认知的结果是记忆，和人一样，机器在一次次学习和思维之后，瞬时记忆和工作记忆都会遗忘，唯有留下长期记忆。随着机器能自学习、自编程、自编译、自对抗、自纠错、自沉淀、自成长，代表知识结构和图谱的长期记忆在机器的生命期里不断地被修饰、修剪和重塑，**可不可以在进行记忆的提取时，把机器的长期记忆网络看作是基于一个符合幂律分布、具有小世界特征的人工复杂网络？**它有哪些数学性质？在机器的整个生命周期内，网络结构如何变化？网络动力学有哪些性质？如何被注意力机制触发，表现为涌现的？

**思考 8：四要素说** 农耕时期人类发明的工具是**双要素**：物质和结构；工业时期人类发明的机器是**三要素**：物质、能量和结构，工业革命已经证明了一个基本事实：人类大规模生产出的机器能够制造出比人类手工劳动更高效、更精致的工具和产品，这就大大解放了人的体力。正在到来的智能时代，人类发明的认知的机器应该是**四要素**：物质、能量、时间和结构。它们终将会证明，机器暂时不必和意识纠缠，也一定能够按照人赋予的意图，做出比人类更复杂、更出色的创造吗？智能机器自身会有想象力和创造力吗？机器如何学习？如何创造？随着越来越多的智能机器已经能听会说，能画会唱，有问有答；深度学习的成功告诉我们，算法可以不被困在程序里，可以用数据调整算法参数，甚至生成知识。那么，**究竟要跃过一个什么坎儿机器就能创造新思维？**如果智能机器能够思维，能够创造新思维，获得新的认知，这不但意味着机器能够寄生、模拟人类的思维和认知，机器还可以让思维离开人体再现，进而创造，这起码部分消除了人类精神的神秘性。是到了研究智能机器如何创造新思维

的时候了。

**思考 9: 异构的基础架构和协同** 新一代人工智能应该聚焦于计算机的算力、算法和数据呢,还是应该聚焦在机器的交互、学习和记忆三个硬核上?为什么?我认为智能机器离不开与外界环境的感知和交互,新一代人工智能的基础架构一定是异构的,传感器端以专用芯片处理为中心,瞬时记忆区以数据处理为中心,工作记忆区以计算处理为中心,长期记忆区以网络(图形)处理为中心,明确这种 XPU<sup>+</sup>、DPU<sup>+</sup>、CPU<sup>+</sup>、GPU<sup>+</sup> 等异构物件组成的协同工作模式,处理好异步和同步,才能为新一代人工智能提供基础架构的支持。

**思考 10: 量子图灵机** 图灵在剑桥大学曾经痴迷量子物理,为什么图灵只发明了图灵机,而没有发明量子图灵机?现在量子图灵机已经有了,为什么总是“只听楼梯响,不见人下来”?量子人工智能和量子智能机器什么时候能落地?在什么关键问题攻克之后才能够落地?

爱因斯坦说过,想象力比知识更重要,提出问题比解决问题更重要。提出新的基础研究问题、新的可能性、从新的角度重新认识旧的问题,需要批判和质疑的科学精神,需要非凡的想象力和创造力。不过,我们也要记得苏联著名作家列夫·托尔斯泰说过的话:“多么伟大的作家,也不过是在书写个人的片面而已。”因为随着人类认知的越多,人类终于意识到自己不知道的会更多,一个人如此,一群人如此,能够思维的机器如此,整个人类亦如此。

致谢 科学植根于讨论,本文形成过程中,和

很多人进行了有益的讨论,他们是:常梦雄、韦钰、戴琼海、王海峰、娄子勤、黄立威、李兵、王小云、陈小平、尼克、张军平、安学、归丽华、刘成林、何清、何雯、郭毅可、吴基传、周建设、刘华平、玄祖兴、马楠、刘玉超、张天雷、韩威、江碧涛、于剑、赵菲、陈杰、欧阳真超、段玉聪、刘挺、刘畅、吴飞、张骊、吴华、孙富春、王文捷、王卫宁、李雪莲、赵刚、杨林、刘东红、李蓓、马华东、高洪波、刘萌、郑思仪、吴天骄、李德仁、张旭、殷加伦、吴信东、赵春江、焦李成、龙腾、杨放春、张兆田、陈关荣、彭艳、刘云浩、支涛、孙茂松、沈向洋、戴国忠、梅宏、史元春、秦昆、赵阳、郑波尽、王景、刘阶、蒲慕明、李向阳、汪国帆、季向阳、周伯文、高跃、唐杰、李涓子、杨义先、董晓波、黄铁军、鲍泓、何克清、庄洪林、白晓颖、吴华、林润华、乔俊飞、吴枫、汪国平、陈桂生、蔡恒进、方向、蒋昌俊、任鹏举等,在此一并感谢。

作者简介:



李德毅,中国工程院院士,中国人工智能学会名誉理事长,中国指挥与控制学会名誉理事长,CAAI Fellow,军事科学院研究员,吴文俊人工智能科学技术奖最高成就奖获得者,我国不确定性人工智能领域的主要开拓者、无人驾驶的积极引领者和人工智能产学研发展的重要推动者。长期从事计算机工程、不确定性人工智能、大数据和智能驾驶领域研究。最早提出“控制流-数据流”图对理论,证明了关系数据库模式和谓词逻辑的对等性。提出云模型、云变换、数据场等认知形式化理论,用于解决定性概念生成、相似度计算、不确定推理、智能控制等问题,成功控制三级倒立摆各种动平衡的姿态。提出基于路权构建驾驶态势认知图,研发机器驾驶脑,领导了中国最大的智能车联合团队。

中文引用格式:李德毅. 人工智能基础问题: 机器能思维吗? [J]. 智能系统学报, 2022, 17(4): 856-858.

英文引用格式:LI Deyi. Artificial intelligence fundamental question: Can machines think?[J]. CAAI transactions on intelligent systems, 2022, 17(4): 856-858.