

## 推进文科智能发展的三大支柱

文科智能是相对理科智能提出的。2018年8月，鄂维南院士组织专题会议，在世界范围内首次提出 AI for Science 概念，将其解释为科学智能。科学是科学活动的总称，包含自然科学和人文社会科学。根据汤超院士的回顾，发起科学智能项目的初衷是探索机器学习在不同科学和工程领域的应用，其中科学指涉的应是狭义科学，即自然科学。故此，我们可以将 AI for Science 理解为理科智能，与之对应，可以将 AI for Humanities and Social Science (AI for HSS) 称为文科智能。自然科学以自然现象为研究对象，人文学科和社会科学以人类精神和社会现象为研究对象，两类对象属性存在本质差异。如果可以借用数学术语来表达，则人类精神与社会现象具有无限元性，自然现象则具有有限元性。在对象属性差异消除之前，尽管两类学科都以寻找科学规律为旨趣，可与人工智能的关系却难画等号，可先各自表述，合起来则可称为 AI for Academia，即学术智能。

### 文科智能释义

科学家对理科智能的阐释可启发我们对文科智能的想象。鄂维南院士认为，理科智能是让人工智能学习事物底层科学原理用以解决事物底层问题，进一步还可以把科学原理和事物解决方案映射到产业，以解决产业问题。对此，汤超院士指出，理科智能有三层含义。第一，把人工智能运用于科学研究、技术创新、成果转化等。第二，运用人工智能发现新科学，如从实验数据发现自然规律。第三，发展人工智能科学，即人工智能本身也是科学研究的对象。过去几年，自然科学界对理科智能已初步形成共识，认为理科智能不限于把人工智能运用于科学和工程领域，人工智能也是科学的一部分。

此外，我们可以在学术智能框架内考量文科智能。综观人类遇到的智能，除

人类智能之外，便是机器智能。机器智能是人类的创造物，其前沿形态是人工智能。回顾过去还可以发现，人工智能的发展有两条截然不同的路径。其一是让机器模仿人类，像人类一样思考，也像人类一样行动，即人们常说的拟人路径。其二是让机器以机器的方式思考和行动，却希望机器思考和行动的能力与效率达到甚至超越人类，即人们知之较少的理性主义路径。不管是哪一种路径，都要回到与人类智能的交互，以解决 1955 年达特茅斯人工智能会议提出的问题。因此，不管是哪条路径，都是机器智能，在学术场景中则是学术智能的一部分，是有记忆、语言、计算、思考、抽象等能力的学术伙伴，而不是纸笔一样的器具。其中，文科智能是学术智能的一部分，也是人类智能与机器智能的交互，且在交互中变革知识创新与积累的范式，挖掘更丰富的人类精神，检验或发现更本质的社会规律。那么，在文科领域实现人类智能与机器智能交互的基础又是什么呢？

### 文科智能发展需要三大支柱

理科智能实践或是文科智能发展的有益借鉴。有学者梳理了理科对人工智能的应用，将其划分为四个板块。第一，人工智能辅助的数据搜集，如数据筛选、数据标注、数据提炼等。第二，学习科研数据以获取其中有意义的表达，如几何先验、深度学习、自监督学习、语言建模、转换架构、神经算子等。第三，科学假设生成，如科学假设的黑箱预测变量、假设的引导组合、假设的优化细分等。第四，实验与仿真，如对科学假设效率进行评价、运用仿真来缩减实验观察的时间和轮次等。数学、化学、生物学、地球与空间科学等已经产生了一系列与人类智能交互的学科平台，如数学领域的 AlphaGeometry、化学领域的 Graphormer、生物学领域的 AlphaFold 3 等。

理科智能的实践启发我们，文科智能的发展至少需要三大支柱。

第一，数据整合（AI Aided Data Integrating）。人类对自己的认识与对自然的认识一样历史悠久。不过，以自然现象为对象的知识创新与积累是迭代更新式的，如牛顿力学可更新为广义相对论；而以人类社会为对象的知识创新与积累则是层叠积淀式的，古代中国和古代希腊的人文社会知识依然在启迪如今的人类精神挖掘和社会规律探索。两类知识创新与积累的差异也从另一个维度证明文科对象属性的无限元性。与理科智能一样，让人类智能与机器智能交互的前提是让机器学习和理解人类智能。与理科智能不一样的是，文科智能还必须把人类积累的文科知识转化为机器智能可识读和计算的数据。

数据是机器智能的食粮，没有数据就没有机器智能。除了人类积累的各类知识以外，还可通过人类思考与行动的痕迹数据来激活机器智能。自数字传感器和终端进入人类生产生活以来，数据生产与汇聚便伴随人类的思考和行动，让人工智能在沉寂了几十年之后迎来了其发挥能力和提升效率的时机。对文科智能而言，把浩如烟海的多模态古今素材，如文本、图画、语言、音乐、舞蹈、仪式、调查数据、统计数据、著作、文章、痕迹等变成可探索人类精神与社会规律的适度或细颗粒度、可计算的数据，不是人类的能力所及，与机器智能协同才是达成目标的唯一选择。

第二，算法创新（AI Based Facts Analyzing）。人类的知识创新与积累大致经历了三轮范式革命，笔者称之为学术范式、学科范式、智能范式。其中，学科范式即如今人们习以为常的分科教育与科研，是 16 世纪以来计量革命或科学革命的结果，也是人类知识创新与积累蓬勃发展以及成果最为丰盛的范式。计量革命的核心是将计量方法运用于探索规律的人类诉求中，从天文观测到社会研究。当人类智能无力处理海量数据时，算法也在机器智能发展中部分替代计量方法，进

而在知识创新与积累中占据核心地位。没有算法，就没有人机互生的知识创新与积累，文科智能也就失去了基础能力。

不仅如此，算法还带来了另一种知识创新路径：规律发现。在分科范式的文科领域，计量方法通常用于假设检验，即基于对人类精神和社会现象的观察提出事物之间的关系模式假设，再运用统计数据或调查数据进行假设检验。机器智能赋予算法的优势在于可以纳入足够多的事物之间的关联因素，在复杂且巨量的关联中，既可进行假设检验，还可发现人文精神和社会规律。

第三，理论想象（AI Driven Theory Building）。理论是科学的灵魂，自然也是文科的灵魂。人类思维与行动的无限元性意味着对人类精神的挖掘和对社会规律的探索也有无限元性。值得庆幸的是，在给定的时间和空间，人类精神和社会现象相对稳定，依然可以被处理为有界无限的有限元现象，进而为理论想象提供暂时稳定的场景，也为化解分科范式与文科对象的悖论提供了可能。

分科范式的特点在于把有机联系的事物无机化地划分为不同学科的对象，沿着学科纵深探索人类精神和社会现象的规律。比如，国家运行是政治、经济、社会、文化、法律、人口等复杂因素的交互，在分科范式中却因人类智能的局限性而不得不划分为众多学科，使得复杂且有机的精神和规律被表达为零碎、分散且前提条件复杂的知识。与机器智能的交互赋予人类智能在复杂且有机的关联里挖掘人类精神和探索社会规律，其对想象力的诉求远高于分科范式时代，也让想象力成为推动文科智能发展最关键和最有价值的支柱。

与理科智能一样，数据是文科智能的素材，算法则是文科智能的方法。与理科智能不大一样的是，人类的想象力在文科智能中扮演着更加重要的角色，因为它需要突破人类自身的限度，将人文精神和社会现象客体化，不只用于建构理论，

还用于理解数据和创新算法。

推动文科智能发展的目的不是为人工智能寻找应用场景，恰恰相反，是人们试图借助机器智能赋能人类智能，一方面让文科知识创新与积累符合人类精神与社会现象的有机性和生态性；另一方面让文科知识世俗化，满足人们日常生产与生活的知识诉求，进而推动社会的知识平等与发展。所有这些都离不开人类智能与机器智能的交互，更离不开文科智能的三大支柱。