

目 录

一、历史关头：智能时代的文明与道路	1
1.1 中华文化思想的认知困难与出路	1
1.2 百年未有之大变局：通用人工智能带来的挑战与机遇	6
1.3 人文社科与通用人工智能的关系	10
1.4 为机器立心、为人文赋理：一体两面	12
1.5 构建“新文科”的时代背景、挑战与机遇	14
二、亿年历程：文明演化的关键	17
2.1 智能与非智能的边界：“心”的出现	17
2.2 价值体系：先天与习得的价值层次	20
2.3 认知架构：拓展、升维的思维空间	25
三、千年根脉：中国思想的五彩线模型	30
3.1 中国思想的五彩线模型：一元多体与多元一体	30
3.2 一元多体：轴心时代中国思想分化的根源	35
3.3 多元一体：中国思想对外来文明的融合	36
3.4 中国思想“和合”的根源：信息偏差与描述式建模	38



扫描全能王 创建

四、千年根脉：中国思想的数理解读	43
4.1 中国最古老的数理认知：河图与洛书	43
4.2 中国最古老的几何认知：天圆地方	44
4.3 中国最古老的视觉认知：阴阳	46
4.4 道家的风水：场景可供性	50
4.5 《易经》与占卜：世界最早的统计决策模型与随机计算	53
4.6 程朱理学：格物致知与数据驱动的概率模型	65
4.7. 陆王心学：“心”即是“理”，构建通用人工智能的核心哲学思想	71
4.8 因果报应：价值的“记账”系统	78
4.9 禅宗：基于心智模型的高阶通讯与主观唯心论	89
4.10 汉字造字：具身智能、心智与会意的妙用	94
五、百年变局：为天地立心的探索	103
5.1 思想概念与关键词的总结	103
5.2 UV 失衡：解读百年未有之大变局	107
5.3 打造大型社会模拟器：求解未来人类社会 UV 平衡态	112
5.4 为天地立心：积极探索人类文明新模态、新道路	117
参考文献	119
鸣 谢	122



扫描全能王 创建

一、历史关头：智能时代的文明与道路

1.1 中华文化思想的认知困难与出路

绝大多数的中国人对于中华文化，包括哲学思想、人文精神、价值观念、伦理规范与传统习俗，都有着高度的认同与自觉的维护。这种文化认同是维护民族团结与国家统一的基础，也是确保中华文明绵延不绝的根基。反之，对本土文化的不认同或者是对外来文化的盲从，则足以瓦解一个民族的自信与团结，甚至危及社会稳定与国家安全。近年来，一些“崇美、精日、哈韩”的倾向都在民间得到了纠正，这体现了中华文化强大的主体性与内聚力。

然而，对于中华文化的认识，人们在思想上却是十分矛盾和复杂的，甚至是纠结和尴尬的，常常在自信与自贬之间摇摆。

宏观层面的是自信与肯定，中华文明源远流长、博大精深，是世界上唯一绵延不断的、以国家形态存在的文明。但是，如果对照现代文明与科技仔细检查，到底哪些成分是灿烂辉煌的呢？

微观层面的多半是自贬与否定。鲁迅直斥“儒家的‘仁义道德’是杀人的软刀子，‘中庸’是奴才的卑怯”，并在多部小说中列举了中国人的种种“劣根性”。中国台湾作家柏杨 1980 年代有一本畅销书，将传统文化种种弊端喻之为“酱缸文化”。今



天的语境中，传统文化中的阴阳、八卦、易经、风水堪舆都被看作是腐朽的迷信，八卦是启迪了二进制的伟大发明，现在却成了“扯淡”的代名词。中国文化经典中有很多关于“心”的深刻表述，如“心即是理”、“相由心生”，是非常先进的思想，却作为“主观唯心主义”受到了哲学批判。

这种自贬与否定并非毫无缘由，不可否认的是，很多优秀的传统观念中确实混杂着很多缺陷，在现代科学体系面前甚至不堪一击，下面我举一个例子。

《了凡四训》又名《命自我立》，由明朝官员、思想家袁黄(1533-1606)所著，是中国著名的劝善家训，主要讲述的是种德立命、修身治世，受到曾国藩等人的大力推崇，曾国藩的号“涤生”据说就是出自《了凡四训》。书中的一些理念有很大的力量，如“命自我立，福自己求”、“一切福田，不离方寸，从心而觅，感无不通”。然而，《了凡四训》立论的根据却经不起推敲，这削弱了该书对现代人的影响。像很多佛教经典一样，《了凡四训》通过讲故事来说道理，比如：

“莆田林氏，先世有老母好善，常作粉团施人，求取即与之，无倦色；一仙化为道人，每旦索食六七团。母日日与之，终三年如一日，乃知其诚也。因谓之曰：吾食汝三年粉团，何以报汝？府后有一地，葬之，子孙官爵，有一升麻子之数。其子依所点葬之，初世即有九人登第，累代簪缨甚盛，福建有无林不开榜之谣。”

——《了凡四训·三、积善之方》

首先，将子孙官爵归于墓葬的风水宝地，在科学上是对基因突变的错误归因（本文4.4节会讨论风水的起因）；其次，林是福建大姓，“无林不开榜”只是统计关联，金榜高中的林姓士子可能大都同该林姓老妇没有关系，无法说明与行善福报的因果关系（本文4.8节会讨论佛家的因果报应）；再次，三年如一日施舍一个人，这种无差别的福利也不符合现代社会经济学的伦理。虽然《了凡四训》劝人行善积德的出发点是好的，但其论证的逻辑是完全站不住脚的。顺便说一句，积德行善（或者说利他



行为）的论证有赖于社会认知与多智能体演化模型的建立（都属于通用人工智能的研究领域），这需要量化人生的价值目标，并确定行善对于实现这个目标的因果链条。

这种宏观自信、微观自贬的认知困难，其根源在于中国思想在几千年的传承过程中，未能形成严格的数理体系与架构，未能像近代西方那样从哲学中导出强大的生产力（数理模型、工程代码）。作为一个理工科的学者，我观察到，咱们文科教授、学者和学生所写的政论文章习惯用一些“霸词”一笔带过：“很显然”、“毫无疑问”、“逻辑必然”、“毋庸置疑”。当然，在文科的层面估计也只能做到这样了，但是，如果这么论证也成立的话，那么现代科学证明与实验基本都可有可无了。数学中的很多著名的猜想，比如前些年轰动一时的庞加莱猜想（Poincaré conjecture），都是显然得不能再显然的结论，以至于数学界费尽脑汁几十年都想不出一个反例，但最后还是要通过构建严密的论证体系、框架，才能从猜想变成定理。

这个数理体系与架构的缺失导致西方学者认为，中国没有产生“哲学”（因为缺乏严密的概念体系），也没有产生“宗教”（因为缺乏严格的信仰与组织），只有所谓中国“思想”。

鉴于这种情况，本文的标题与行文都用“中国思想”一词泛指中华文化、文明、哲学与儒释道经典。标题中提出的“为人文赋理”是指为中国思想寻找数理模型与体系架构，指导通用人工智能的研究，将中国思想的先进性转化成智能时代强大生产力。

这个数理体系与架构的缺失也导致近年来宣传优秀文化、推崇国学的努力，遭遇巨大的困难。其原因是其直接搬运的古代思想实际上并不符合现代伦理，不被在西方知识体系下成长起来的现代文科学术权威和学生所接受。据我在美国的观察，在这个过程中，西方和港台学者抨击与嘲讽的论点尤其致命。

今天，中国大学的文科教育体系（文史哲政经法）和院系课程设置都是在五四“新文化”运动之后，照搬西方体系构建的。西方思想体系很难接受“中国叙事”，中国不管怎么做，都是错的。所以，中国目前面临的不光是技术体系的“卡脖子”与“挨打”的局面，更长远的、深层次的是思想上的“挨骂”问题。自2018年贸易战以来，中国人已经普遍清醒认识到挨打的事实，放弃幻想，提出原始创新、科技自立自强的部署，但是，对于“挨骂”的根源还认知不足，甚至摸不着头脑，还希望获得西方社会的认可，实际上是行不通的。

那么如何构建中国思想的数理体系，将中国思想的优势转化成先进生产力？

我们先看一个中医的例子。中医也一直是处于宏观自信、微观自贬的困境。大体上，很多人都承认中医中药是有用的，有时甚至很神奇，但是，具体到某一个病例，中医的病理和药理都很难解释得通，从西医的科学体系来看，中医的说法基本就站不住。然而在这方面，屠呦呦先生做了一个很好的表率，她和团队通过收集整理历代医籍、本草、民间方药，据说在2000余验方基础上，筛选了200多种中药做实验，1972年成功提纯了青蒿素晶体（分子式为 $C_{15}H_{22}O_5$ ），2015年成为中国第一位获诺贝尔科学奖项的本土科学家。当然，青蒿素如何治好疟疾的药理和病理还需要研究，但不可否认，屠呦呦先生的努力让世界重新审视中药。比如另一种中药，名字出自《本草纲目拾遗》的雷公藤，现在就得到广泛的研究。就在屠呦呦获奖的同一年，哈佛大学医学院一项利用雷公藤提取物治疗肥胖的研究，发在了顶级期刊Cell上[1]。

过去的实践表明，简单照搬中国古代思想到现代社会是行不通的，我们需要萃取中国思想中的优秀有效成分；深刻理解人类社会演化的规律，建立贯通古今、融通中西、横跨文理的科学体系；熟练运用中国思想的优势，去构建足够引领人类文明演化的新思想，正如习近平总书记所指出的：



“无论是对内提升先进文化的凝聚力感召力，还是对外增强中华文明的传播力影响力，都离不开融通中外、贯通古今。经过长期努力，我们比以往任何一个时代都更有条件破解“古今中西之争”，也比以往任何一个时代都更迫切需要一批熔铸古今、汇通中西的文化成果。”

——习近平《在文化传承发展座谈会上的讲话》2023年6月2日

中国思想的认知困难的问题，大约在 10 多年前触动了作者。当时西安交大的一位博士来作者在 UCLA 的实验室访问，他带来一套陕西作家贾平凹的小说。作家都有十分敏锐的社会观察力，贾平凹的小说描述了陕西农民在城市化浪潮中的命运与挣扎，淳朴的农民带着传统的思维，遭遇了急剧的社会变化，来到现代化的城市，面对高楼大厦、车水马龙，不知所措。小说对这些小人物的坎坷命运的具象描写，让作者联想到中国从农耕文明中发展而来，自鸦片战争以来受到西方文明的冲击的惨烈遭遇[2]。

作者从小在中国农村长大，受到儒家伦理的影响，后来在美国生活了 28 年（1992-2020），深感中国文化对比西方文化具有很大的优势。说到中国思想的优势，很多人会不以为然，因为自 1919 年新文化运动以来的 100 年，中国的精英知识阶层都在反思与批判中国的文化和思想，而当今绝大多数的中国人，特别是理工科的人，都没有机会深入学习中国的子学经学。梁漱溟在《东西文化及其哲学》[3]一书中指出，人生有 3 种问题，第一是人与自然的问题，第二是人与他人的关系，第三是人与自身生命局限性的问题，西方科学解决的重点是第一种问题，中国思想解决的重点是第二种问题，印度宗教解决的重点是第三种问题。中华思想的这种对人与他人之间关系的熟谙与智慧，在全球化时代文明冲突、碰撞中具有极其重要的意义；对正在步入的人机共生的智能时代人类如何构建与通用智能体的伦理关系具有启发意义。作者将在后文中讨论中国思想的这个优势，下面先用一句话来概括：

中国思想是根植于人类认知架构基础上、由“心”驱动的、推己及人，在几千年的世俗社会中演化而来的智慧结晶；中国的价值观和伦理规范是基于价值判断的复杂决策体系，更接近于人类文明演化的终极社会伦理；相比较，西方文化是建立在宗教信仰之上的，后者是虚构的，随着科学的普及，宗教成分将会逐步退出，通用人工智能的进步将加速这个进程。



有意思的是，通用人工智能的研究目标，通俗来说，就是要重造能够匹配人类各种能力与智力、符合人类情感与伦理价值的通用智能体，包括实体的机器人和虚拟的数字人。训练通用智能体，类似于培训学生，需要赋予思想（世界观）、价值取向（价值观）、与行为规范（人生观、社会观），除了个体层面的，还包括更大尺度的，构建智能社会的整体。因此，通用人工智能的研究与构建文化思想的理论根基是高度相关的，可以说，这两个问题本质上是同一个问题，是一个硬币的两面。下文先讨论二者的关系。

1.2 百年未有之大变局：通用人工智能带来的挑战与机遇

人们常说，当今世界正在经历“百年未有之大变局”，在作者看来，这个变局包含两层冲突，见图 1。

1919：西方文明与东方文明的碰撞与融合

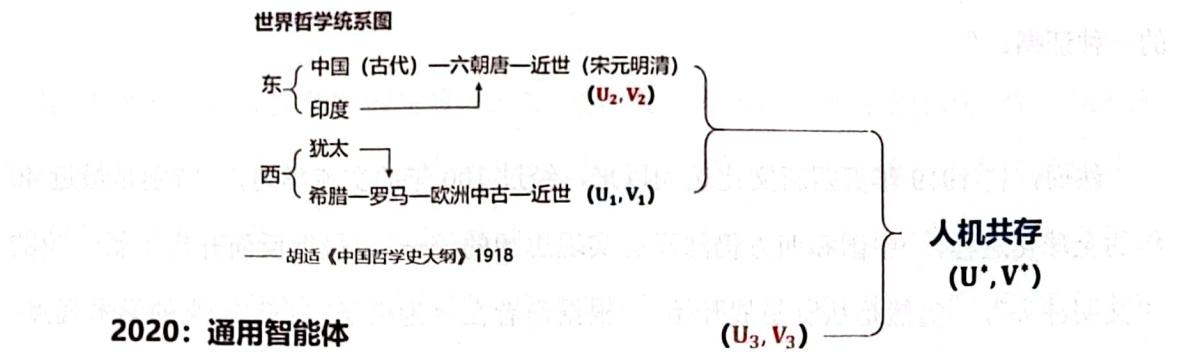


图 1：1919-2020，百年未有之大变局：从东西方文明的碰撞与融合（1919），到智能时代的人类文明和通用智能体的共生共存（2020）。

1，东西方文明的碰撞与互鉴。中国近代史的开端是 1840 年，鸦片战争本质是贸易战，鸦片战争的失败让中国人怀疑自己的经济体系，开启了洋务运动，引进工业机器和武器装备，但即使在此时，中国依旧保有天朝上国的自信；真正让



中国经历数千年未有的忧郁愤激和耻辱无奈的，是 1894 年甲午海战的失败，这让中国开始考虑改革自己的政治体制与教育办法，启动了戊戌变法，创办京师大学堂（北大的前身），尔后又爆发了辛亥革命；而 1919 年一战后的巴黎和会的失利，让中国进一步否定了传统文化思想，开启了新文化运动，彻底否定延续了 2000 年的子学经学。五四时期，鲁迅甚至提出了“不读中国书”，而钱玄同则主张“废除汉字”^[4]，代表了激烈而彻底的反传统、批儒家的文化主张，这种对传统的彻底摒弃有深刻的时代烙印，也是导致今天中国传统遭遇认知困难的历史原因，值得我们重新思考和评估。1840 年一来，从经济、政治到文化，这三重否定逐层深入，代表了中国的思想体系在西方文明冲击下的巨大转变。

作为新文化运动的领袖之一的胡适在 1918 年编写的《中国哲学史大纲》^[5]中提出，东西方哲学体系经过碰撞，将走向“世界大同”。为何如此呢？胡适在北大授课时草拟了《中国古代哲学讲义》，在其中给出了自己的解释：“为什么西洋哲学史可以和中国哲学史互相印证，互相发明呢？这都因为人同此心，心同此理，所以人类到了一种大略相同的时代境地，便会生出一种大略相同的理想。这便是世界大同的一种证据。”

然而，自 1919 年五四新文化运动以来，经过 100 年的交流学习，特别是最近 40 年的全球化进程，中国和西方仍然没有实现思想的统一，最近反而开启了新一轮的“文明冲突”，仍然是从贸易战开始。根据作者在《为机器立心》一文的学术观点，从通用人工智能研究的角度看，任何个体的人（智能体）或者社会（文明体），都可以用两套函数体系来刻画：

- U-体系指的是“理”，包括各种自然科学的模型与社会科学的伦理规范；
- V-体系指的是“心”，包括个体、他人、集体、国家乃至全人类的价值。



扫描全能王 创建

如图 1 所示，西方的 (U_1, V_1) 与东方 (U_2, V_2) 通过交流， U_1 与 U_2 已经相当接近，然而，西方以宗教为基础的 V_1 与中国世俗社会的价值体系 V_2 还有很多不相容的维度，这是当前矛盾的焦点。

说得直白一点，近代史以来，中国向西方学习了大量先进的 U (U_2 向 U_1 靠拢)，科学（自然）与民主（社会）都属于 U 的范畴，但是，中国保留了自己的核心价值观 V_2 ，没有全盘接受了西方的 V_1 ，比如排斥了基督教会的价值渗透。本文的观点是：中国的 V_2 比 V_1 更先进，更适合构建人类命运共同体与全球的政治新格局。在心（V-系统）与理（U-系统）的对偶关系中，心是占主动地位的，也是儒家思想所说的“心即是理”（本文 4.7、4.8 节将讨论这个问题）。然而，近 100 年来，科学技术（U）有了巨大的进步，而西方的 V_1 变化甚少，已明显不能与 U 相匹配。

当前东西方的竞争本质是 V_2 与 V_1 的竞争，世界又到了百年未有之大变局，到了“问苍茫大地，谁主沉浮？”的历史关头！需要提出符合人类社会发展的新模式、新的价值体系，引领新的时代潮流。

2，人类文明与人工智能的碰撞。2020 年以来，人类正在跨入智能的时代，随着通用人工智能的快速发展，未来将会出现大量的通用智能体，也就是具有自我意识、由价值驱动的物理机器人或者虚拟的数字人，他们会有自己的 (U_3, V_3) 系统（甚至多元）。那么，碳基的人类文明与硅基的通用智能体如何共生共存，能否达到新的“大同世界”，是新的时代命题。

这两层冲突的叠加与相互作用，对国际局势和人类的未来正在产生强烈的冲击和不可估量的影响。因此，通用人工智能带来的既是挑战、更是机遇！



近年来，媒体上出现了关于通用人工智能安全与伦理的担忧，出现担忧与疑问是完全可以理解的。但是从作者的阅读来看，这些讨论往往陷入两个陷阱中：

- 立场陷阱：对通用人工智能发展的控制权的抢夺；
- 技术陷阱：对通用人工智能技术标准与路径的误解。

讨论人工智能安全的大多数人，并不是研究通用人工智能的专家，如果不深入理解通用人工智能本身的机理与科技发展趋势，那么，关于安全伦理的担忧难免流于空泛，也就不能制订精准的政策，甚至会误导政策制定。



扫描全能王 创建

下面，本文将论述通用人工智能与中国思想、人文社科研究的关系，以及如何实现最大跨度的学科交叉，走向深度融合。

1.3 人文社科与通用人工智能的关系

表面上，人文是距离通用人工智能最远的学科群。

- 人文是保守的、非数字化、非结构化的；
- 通用人工智能是激进的、高度结构化的。

在现代高等教育的学科体系中，理科最先成为科学（Sciences），如物理、化学、生物。随后，工科也自称科学，如计算机科学（Computer Science），材料科学（Material Science），最近出现了智能科学（Intelligence Science）。然后，社会学科也自称是科学，如政治学（Political Science），社会科学（Social Science）。唯独人文（Humanity）与艺术（Arts）尚未被纳入科学体系。美国大学中人文、艺术和科学是并列的，有些机构把人文归入艺术。

从作者跟人文学者的交流中得知，很多资深的人文学者并不希望将人文与艺术“结构化”，而习惯于把一些概念“神秘化”，比如中国文化中道家的“无”和佛家的“空”等等。因为学者们担心，一旦把人性结构化了、变得透明了，就失去了人的尊严，这是人类作为一个整体的“隐私”。甚至有的人文学者不相信进化论，认为人是特殊的，这在西方的宗教人群中很有市场。



扫描全能王 创建

所以说，人文是保守的，在于守护传统，尊重圣贤之道。可惜，科技进步的潮流是挡不住的。几年前，不少人认为，人工智能最后攻不破的就是人的创造力，比如艺术，结果，艺术是首先被攻破的，当前人工智能已经全面介入艺术的创作与评估，特别是音乐、绘画、诗歌创作。

本质上，人文是距离通用人工智能最近的学科群。

通用人工智能的研究包含两个尺度：

- **个体尺度：**研究的目标是创造出能够匹配人类感知、认知、行动、学习与社会协作能力，符合人类情感、伦理与价值的通用智能体，后者包括实体的机器人和虚拟的数字人。人类本质上就是从地球的物理运动、化学变化、生物进化中，通过智能的演化和认知革命，而产生的一类通用智能体。这些通用智能体，不论其实现的材料特性，碳基还是硅基，背后都应该有一个共同的哲学原理和数理模型。
- **社会尺度：**通过多智能体系统的研究来探索智能体协作的模式、社会结构的形成。在大型的物理逼真环境或元宇宙空间，人工智能试图模拟、还原人类演化的进程，在自然的地理气候环境的边界条件下，个体的人、家庭、部落、城邦、酋邦到王国等社会、经济、政治结构的形成。这涉及到经济学、社会学和政治学的模型，乃至文明演化的动力学模型。通过社会模拟可以回答文史哲政经法的一些难题，如中国文化、政体形成的自然与历史成因与必然性，预演人类社会在不同模式下的发展结果，准确推算什么是社会最大利益与公平正义。

因此，人文社科与通用人工智能都是研究人性与社会的本质，是从不同的层面来研究这个问题的。作者提出一个新的看法，当通用人工智能走向成熟的时候，也是人文成为科学的时候，换句话说，

通用人工智能 = 新文科。

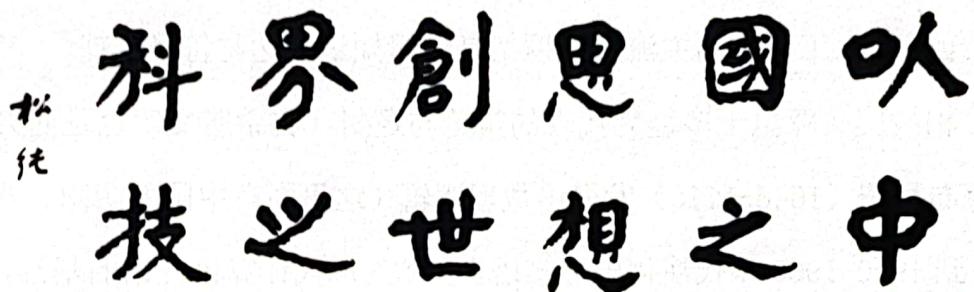
本文认为，通用人工智能带来三重机遇：



扫描全能王 创建

- 从传统中国思想中淬炼有效成分、构建严格的数理体系与架构；
- 重塑大学的人文社科的体系、增强文化自信， 构建与人类科技进步相匹配、符合人类共同利益的的哲学思想与价值体系；
- 将先进的思想（价值观）注入通用智能研究，形成系统与生产力， 实现原创、引领全球的科研。

以一句话来概括 --- 以中国之思想、创世界之科技。



1.4 为机器立心、为人文赋理：一体两面

人文社科与通用人工智能的深度关系决定了两者之间的双向连接：

“为人文赋理”：通用人工智能赋予人文新的方法、模型、和理论工具，以解读儒释道的经典，诠释并弘扬中国思想，不再是宏观的文字思想的论证， 而是提供具体的“融通中外、贯通古今”的新的数理模型与体系，从而增强文化自信。

“为机器立心”： 赋予人工智能符合人类世俗社会伦理的情感与价值体系，为机器“致良知”，从而由价值来驱动， 而不是单纯用数据来驱动通用智能体，这对于人工智能的健康发展、安全与治理、以中国传统人文社会思想来实现通用人工智能、抢占世界科技前沿制高点， 具有重要的意义。



扫描全能王 创建

如果我们能够成功解决这个“一体两面”的问题，那么面对当前的东西方文明冲突，就可能化解“挨打”与“挨骂”的问题。

纵观人工智能领域发展的历史，人工智能的研究分三个层次：根源在哲学思想，本质在数理框架，而形成生产力则需要软硬件系统工程实现。

我们可以将人工智能的历史分成三个时期。

1960-1990 年代，人工智能的主流数理框架是逻辑知识表达与符号推理。其源头是以苏格拉底、柏拉图、亚里士多德为代表的辩论与逻辑（古希腊文明也是西方哲学的源头），到莱布尼茨（1646-1716）发展出数理逻辑（这里面有中国的阴阳、八卦的二进制思想），到 1940-1960 年代演化出形式语言，成为现代计算机理论的基础，1960 年代发展成为严密的命题逻辑、谓词逻辑、事件逻辑等体系，为知识表达、符合推理、专家系统提供了数理框架。

1990-2020 年代，人工智能的主流数理框架是概率建模与随机计算，这与儒家的“格物致知”的理学思想（宋代的程朱理学）在思想上是一致的。格物就是仔细观察，收集数据，致知就是提炼出数理模型。格物致知本质就是从数据到模型的知识发现过程，就是近年来常常听到的大数据、大模型的统计方法把这个思想理念变成科学与工程实践。本文 4.7 节将讲述，中国的理学与大数据人工智能不光是思想的关联，而是在具体问题上都有关联。

2020-未来，作者在多个场合讲过，实现通用人工智能的关键在于“为机器立心”，为此，研究人员需解决认知架构、价值驱动、社会智能、人机互信等重要问题。这些高层次的科学问题已经超出了数据驱动的人工智能模型，需要借鉴儒家的“心学”理念，如王阳明所强调的良知、价值体系构建、“心即是理”，以及禅宗所讲的“相由



扫描全能王 创建

“心生”等思想，构建“主观”与“客观”融合的数理体系，具体参见《为机器立心，迈向通用人工智能》一文。

“为机器立心”与“为人文赋理”是一体两面，这既能发挥中国思想的优势，也能实现原创引领的通用人工智能系统，从而将中国思想转化成为工程代码与生产力。

1.5 构建“新文科”的时代背景、挑战与机遇

正是在当前数字化、智能化的时代背景下，国家提出建设“新文科”。本文先引用强世功教授（原北京大学社科部部长）的一段话，说得非常深刻：

“五四新文化运动的本质是以西方现代的哲学和社科体系取代了 2000 多年的传统‘子学’、‘经学’体系。胡适、冯友兰以西方哲学来重新理解儒家思想；费孝通、瞿同祖用西方社会学、人类学和法学来理解中国历史资料。中国的哲学社会科学差不多跟在西方哲学社会科学体系后面，不仅是思想理论，而且在学科专业设置和教学方法上亦步亦趋。在中国迈向智能时代的进程中，要密切关注智能社会的国家治理问题，提出新理论，并逐步形成新的哲学和社会科学体系。”

——强世功教授《数字智能时代的北大文科研究》，2021 年

新文科的建设经过信息化、数字化，正在进入智能化的深水区。关键就在实现与人工智能的连接：

- 在个体层面，人工智能对接文史哲，这包括人性的刻画、情感的基础、历史人物的活化、人生的价值等；
- 在社会层面，人工智能对接政经法，关键在于打造社会模拟器，这包括在模拟器上做经济、立法、管理的社会实验等。

如何将人文学科与人工智能的深度融合付诸实践是一个重要议题。2020 年 11 月作者归国入职北大，就组织了众多人文社科学者共商人工智能与文科的连接（见图 2）；2021 年 10 月，武汉市政府和北京大学，获批国家首批“智能社会治



理实验综合基地”；2022年1月，《三读赤壁赋，兼谈心与理的平衡》一文发表，试图以人工智能解读中国文学的巅峰之作《赤壁赋》中苏轼的人生哲思；2023年4月，北京大学武汉人工智能研究院正式揭牌，并成功举办首届中国社会治理论坛，主题是“智能、文明与道路”；2023年7月，我们进一步开展了AI+人文社科的实践，参与学院达11个，设立开放课题40多项，从而将人工智能与人文社科的交融推向深入。

要进一步实现这个深度融合，需要打通前述“思想-数理-工程”这三个层次，这就横跨了文科-理科-工科三个大学部，难度巨大，但这是大势所趋，所谓时代潮流浩浩荡荡！希望大家，特别是年青一代，做智能时代的先知、先觉、先行者。

文理学科的分合：历史上中国并不分文理科，从隋唐开始的科举考试持续了1000多年，那时尚不存在文理的区分。1840年鸦片战争后开始西学东渐，理科进入人们的视野，但以文科的附属形式存在，这是我国历史上第一次文科与理科的短暂合并。1954年，我国开始采用文理科分科高考；直到2014年9月3日，国务院颁发了《关于深化考试招生制度改革的实施意见》，明确提出：“保持统一高考的语文、数学、外语科目不变、分值不变，不分文理科，外语科目提供两次考试机会”，这是我国历史上第五次文科与理科的短暂合并（图3）。

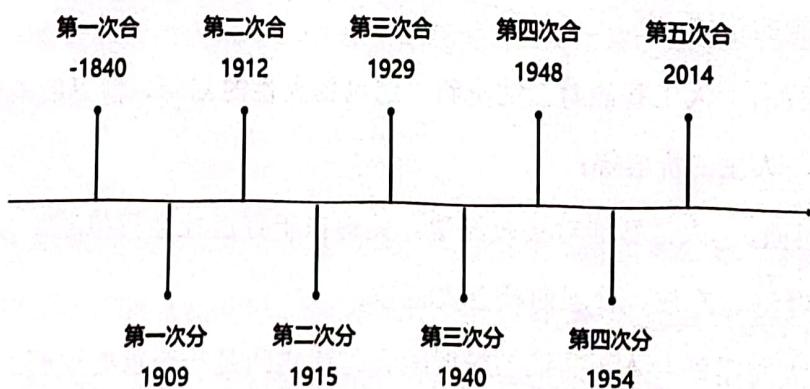


图3. 中国历史上文理科的分与合。



交叉学科的兴起： 2022年9月，教育部正式将“智能科学与技术”列入交叉门类的一级学科，这个学科的内涵与外延还在演化中。前文讲到，通用人工智能的实现有赖于哲学、数学、神经科学、心理学、计算机、机器人（工学）、语言学、经济学、法律、政府管理等学科；这些学科的共同协作才使得人工智能有今天的成就和社会影响力。尤其重要的是，未来我们若要获得可靠、可信的通用人工智能，需要从人文尤其是倡导和合共生的中国思想中获得营养，因此，我们需要为机器立心，这颗心，就是人文学科倡导的“心”。

以上是关于通用人工智能与人文社科双向连接的时代和学科发展背景的分析，接下来我们就进入正文。

第一部分“亿年历程”，将简要介绍物体和智能体的关键差别，并展现文明的起源。从大的尺度看，从物理运动、化学变化、生物进化、智能的演化、最后文明的产生与演化，这个过程中，我认为，“心”的出现是最重要的、突破性的特征。中国思想就是对这个进化链条的认识与建模。

第二部分“千年根脉”，是本文的核心。我将以人工智能的数理模型解读中国思想。中国思想博大精深，我们绘制了一个色彩斑斓的五彩线，反映其兼容并蓄、多元一体、自强不息的精神面貌。但由于篇幅限制，我这里只选择了《易经》、因果、禅宗、以及程朱理学、陆王心学等十个题目做一鸟瞰式的回顾与解剖分析。

第三部分“百年变局”，本文将提出，人机共生时代中，重新弘扬中国传统思想、倡导和合共生是面对世界大变局的良方；然后简要介绍我们基于传统人文并融合人工智能理念的现代社会治理思想与实践。

本文在具体细节的解读上可能多有谬误，如有任何不当，恳请大家批评指正。



扫描全能王 创建

二、亿年历程：文明演化的关键

中国思想是在过去几千年的时间里对亿万年形成的自然、人性和社会的认知与建模的总集。本文首先简短回顾从无生命到生命、智能、文明的演化历程，指出“心”的出现是最重要的、突破性的特征，进而对“心”做一个数理结构的分析，它包括两大部分：价值体系与认知架构。

2.1 智能与非智能的边界：“心”的出现

人类思想从何而来？现代科学的发展对这个问题的回答越来越清晰，从地球早期的物理运动、化学反应、生物进化，智能体的产生，智人的出现，最后社会群体的建立，民族的形成，直至国家的产生与文明的诞生。图 4 显示了人类文明进化的亿万年历史的重要阶段。大概 138 亿年前，大爆炸之后宇宙诞生，由 4 种基本力（强相互作用，弱相互作用，电磁力和万有引力）的相互作用形成了我们今天所处的物理世界。大概 35 亿年前，以这四种力为基础的、更弱的力，如共价键、离子键、范德华力、疏水力以及氢键等，主导了化学变化；在此基础上，细胞形成，DNA 转录与蛋白质折叠都依赖于氢键等相互作用，成为生物进化的重要动力。这些物理、化学、生物的相互作用，数学上可以统一用各种势能函数 U 来表达。



扫描全能王 创建

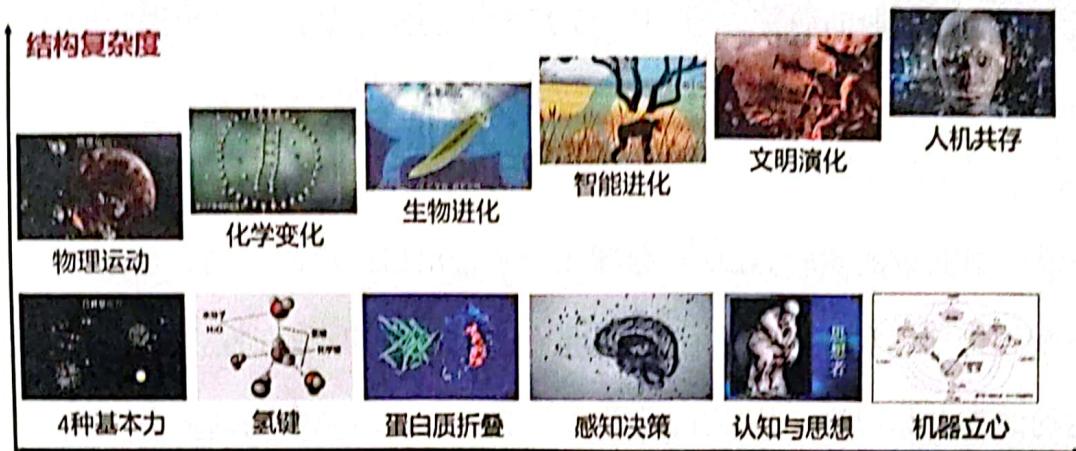


图4：从物体到生命体、智能体、智人、文明和人机共生社会的演进历程。

人脑进化的质变出现在在大约 7 万年前，当时的一次所谓“认知革命”造就了当今的人类 — “智人”。大概 1.2 万年前，农业革命出现，大规模的合作成为必需，从家庭、部落、酋邦、到王国的形成，就开启了人类文明的演化历程。国家与社会的维系，也需要一种像维系无机分子、有机大分子、细胞乃至多细胞个体的力量，但这种力量远比物理、化学和生物的世界要复杂，而由经济制度、政治格局、法律与国家机器、道德、宗教、艺术、哲学、政治思想、法律思想等组成。这些属于人类的经济的、政治的、文化的相互作用，数学上同样可以统一用各种势能函数 U 来表达，本文统称为社会规范。当前，人类社会开始步入智能时代，而智能时代的标志是大量通用智能体的出现。

以上所述的演进过程有一个显著特征，那就是系统结构越来越复杂、所需的变量与空间（不是我们生活的三维空间，是指各种变量的数学空间 space）的不断扩大。例如，太阳系的行星运动，可以用不到 100 个变量（位置、运动速度）精确描述，而要概括哪怕由数百个氨基酸组成的蛋白质的可能构象，也需要远远更多的变量（约 10^{300} ），其复杂度远超太阳系的行星运动。



扫描全能王 创建

在这个连续演变的过程中，系统实现了很多次质变与跃升，如细胞的诞生、动物中枢神经系统的出现、智能的萌芽。其中一个最关键的问题是：在这个漫长的演化过程中，无生命的物体和有生命的智能体的边界在哪里？

我们先来看一组简单的实验。红球与绿球在一个带出口的房间运动。从图 5(左)，人看到的是一个“物理现象”：两个小球是无生命的，因为只有在发生碰撞的时候，它们才改变运动的方向和速度，他们的运动被势能函数 U 所完全刻画。图 5(中)也是两只小球的运动，但人会看成是“社会现象”，因为两只小球能够根据对方的位置和运动来及时调整自己的运动，出现“你追我赶”的故事：小球有了自己的目标和意图，有自己的价值诉求，由各自的价值函数 V 在驱动，由 V 而产生了内驱力，图 5(右)的箭头代表力与颜色代表函数值，暖色代表高价值、冷色代表低价值的位置；同时，小球也有了感知、认知、运动控制的能力，这样就成为了智能体。

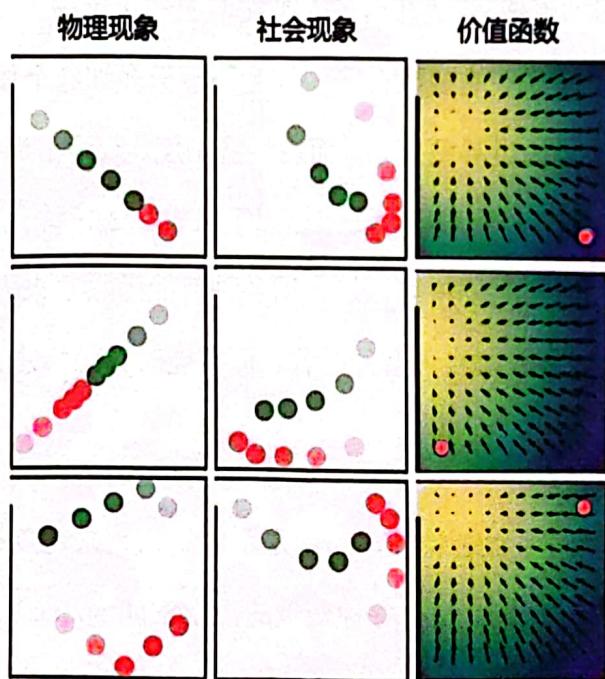


图 5：智能体与非智能体区别的“生命度”对比实验。(左) 两只无生命的小球在无重力的环境自由碰撞，直线运动。(中) 红色小球意图是离开小室，而绿色小球阻止红色小球的离开，表现出感知、认知与意图。(右) 根据两个小球当前位置，红球的价值函数与梯度（内驱力）。

无生命的物体运动是机械的，是被动由各种外力和相互作用驱动的，由一组势能函数 U 描述；智能体的活动是自主的，由价值函数 V 驱动，而这个价值函数又根植于底



扫描全能王 创建

层的认知架构。下文将论述，价值函数和认知架构共同组成了所谓的“心”，心是智能体和物体最本质的区别，是人的最重要特征，也是实现通用智能体的必要条件。

在中国思想的众多流派中，“心”是一个关键词，承载着多重含义，同一个思想家说的心，其含义也不断漂移，不幸的是，心的表达却从未明确定义。下面我们来简要讨论一下“心”的这两个组成部分。

2.2 价值体系：先天与习得的价值层次

人类独有的价值判断不仅在进化过程中被逐渐固定，而且在群体交流中被文化强化，从而形成社会伦理、道德规范和共同的价值观；通过先天、后天的结合，人们总是很快能建立、调整价值体系，从而更好地适应环境尤其是社会。心理学家对 8-12 月大的人类婴幼儿做过一组经典的实验，发现人在进化过程中形成了大量的、先天的价值判断。

善恶判断：分别向婴幼儿展示两组玩具。如图 6 所示，一个红色的小人在爬坡，图（左）的蓝色的正方形阻挡红色的圆上坡（代表“恶”），图（中）的黄色的三角形主动助推红色圆上坡（代表“善”）。之后让婴儿去自主选择一个玩具时，婴幼儿都选择黄色三角形做玩具（选善弃恶）。当然，这个实验是做了形状和颜色的对照实验，排除了对颜色和形状的偏好。

实验：8-12 月大的婴幼儿选善弃恶

- A. 红色的圆要上坡，
被蓝色的正方形阻止（恶） B. 黄色三角形
助推红色圆上坡（善） C. 婴儿都选择
黄色三角形做玩具（选善弃恶）

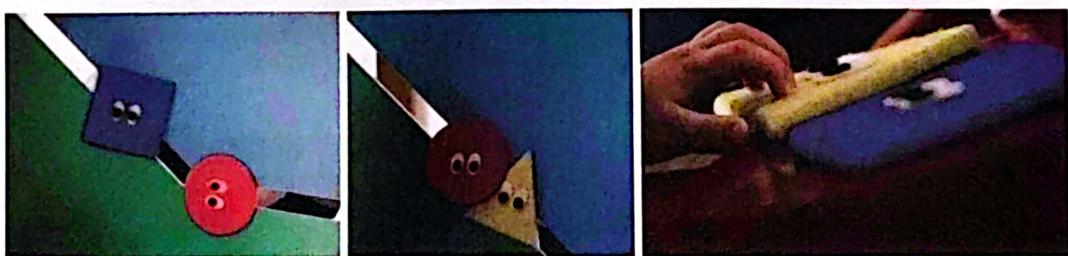


图 6：婴儿对善恶的选择的认知心理学实验。在耶鲁大学的一项经过精心设计的实验中，操作者向儿童展示了一个爬坡实验：红色的圆要沿着一个倾斜的轨道向上爬。在左图中，红色圆圈的上爬受到了蓝色方形的阻碍；在中图中，红色圆圈的上爬受到了黄色三角形的辅助。随后，操作者将蓝色方形和黄色三角形交给 8-12 个月大的婴儿选择，婴儿一致选择黄色三角形[6]。

公平观念：两个阿姨给小朋友们分配 10 个球，一位公平地分配、每人 5 个 (A)，一位不公平地分配 9: 1 (B)，目睹这个过程后的婴幼儿们会选择和公平的阿姨 (A) 一起玩（图 7）。

实验：12 月大的幼儿选择与“公平”的人玩

- A. 短头发的白人女性 A B. 长头发的白人女性 B
公平分配玩具 不公平分配玩具

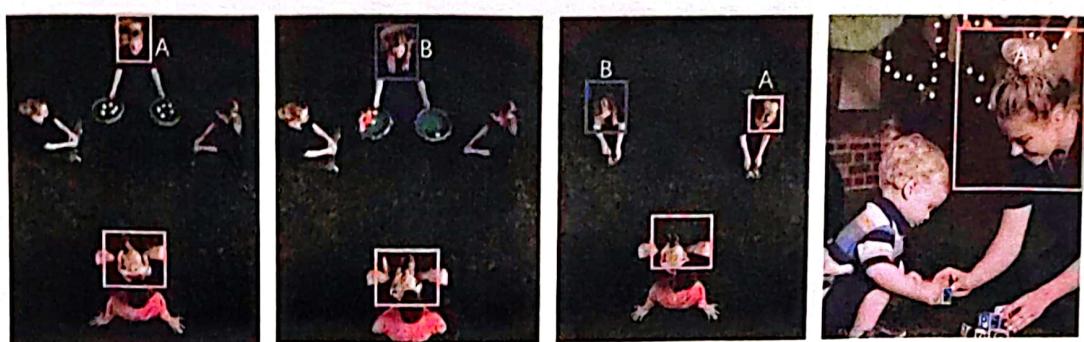


图 7：婴儿对公平的选择的认知心理学实验。在一项经过精心设计的实验中，操作者向儿童展示了两位玩伴：如 A、B 所示，短头发的 A 会公平分发玩具，而 长头发的 B 则不公平分发；然后让 12 月大儿童观察 A 和 B 玩积木 (C)，并让儿童做出选择，结果儿童选择与公平的 A 玩耍 (D)。

合作精神：心理学家分别与人类幼儿、黑猩猩做游戏，发现人类幼儿天性喜欢“合作”，相比之下，黑猩猩合作能力与意愿远不如人类幼儿（图 8）。



图 8：婴幼儿合作的意愿与能力。左图中，操作者拿着一摞书尝试放入关着的书柜，因双手持书无法（假装）打开柜门，儿童能判断操作者的意图，主动上前帮忙（利他行为）。右图中，操作者向地上投掷物品，黑猩猩会主动帮忙拾回。实验中采用了 24 个 18 月大的幼儿，以及 36、54、54 月大的 3 只黑猩猩[7]。

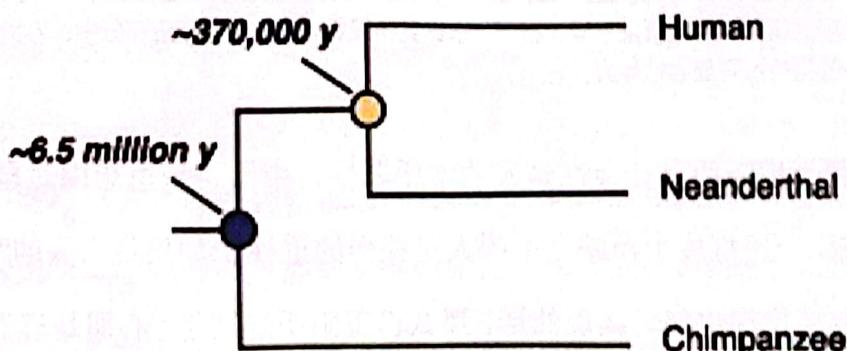


扫描全能王 创建

从上面的实验可以看到，不到 2 岁的小孩已经形成了基本的价值判断。当然，人类也有一些其他的价值取向，比如在同一个实验（图 7）中，研究人员发现小孩更倾向于同同一种族的人玩耍，即使他/她表现出不公平的行为。这表明小孩对种族的认同大于公平正义。

以上提到的人类的先天价值判断，无论是善良、公平乃至族裔认同，其实都是人类更高水平的合作的基础。《人类简史》一书提到黑猩猩、尼安德特人的族群人数不会超过 150 这个上限，而智人则能实现更高水平的合作[8]。古埃及的金字塔、中国的万里长城、英国的巨石阵，都是人类早期合作能力的体现，今天的人类，甚至可以进行更大规模、更高层次、更深影响的合作，如同盟国的诺曼底登陆、美国的曼哈顿工程、中国的两弹一星计划、多国的人类基因组计划等等。合作是人类这一物种的与众不同的显著特征。哈佛大学教授 Martin Nowak 将人类称为超级合作者(SuperCooperators) [9]。

人类在进化过程中，通过基因变异与自然、择偶、社会的多重选择，以及复杂语言的产生促进了交流与传承，使得人类群体能进行价值和行为规范的对齐，这是大规模协作的基础。事实上，黑猩猩同人在大约 650 万年前发生物种分离，到今天基因差异在 1.2-3% 之间，而同黑猩猩相比，人类中某些同神经发育有关基因获得了显著的基因变异（图 9）。另外，已经灭绝的尼安德特人同人类（智人）在大约 37 万年前分道扬镳，二者基因差异在 0.5% 左右，而同尼安德特人相比，人类在同语言发育密切相关的基因如 *FOXP2* 上发生了重要的功能变异。



扫描全能王 创建

图 9：人类同黑猩猩、尼安德特人的进化关系。黑猩猩 (Chimpanzee) 同人类的共同始祖的分界发生在约 650 万年前；37 万年前，人类共祖又分出两支，即尼安德特人 (Neanderthal) 和人类。尼安德特人在约 3 万年前灭绝[10]

人类学与比较心理学有大量实验对比人类婴幼儿与其他物种的差异。（图 10）实验对比了人类婴幼儿与黑猩猩 (Chimpanzee)、红毛猩猩 (orangutan) 的差异。在物理交互任务中（运动、使用工具），三者非常接近，但在社交和认知方面，人类要远胜黑猩猩和红毛猩猩。最近也有 fMRI (Functional Magnetic Resonance Imaging, 即功能磁共振成像) 实验表明，人和人之间在感知皮层（如视觉）的差异很小，而在认知与决策方面有明显差异。换句话说，该实验揭示出人类在面对物理世界时的感知技能同黑猩猩、猩猩非常接近，人类幼崽得天独厚之处在于面对社会交往时表现出的认知技能。这些结果支持一种被称为文化认知假设 (cultural intelligence hypothesis) 的理论。

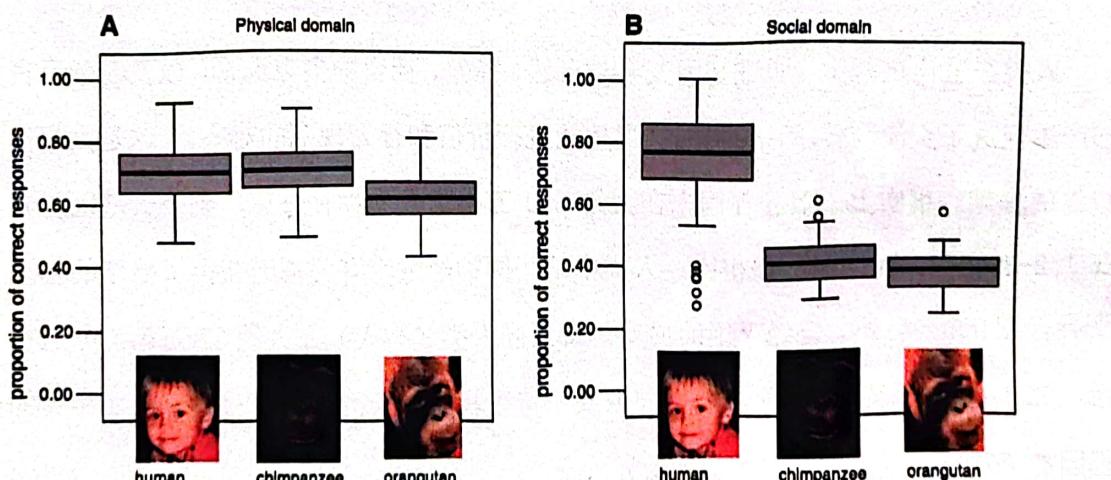


图 10：人类同黑猩猩、红毛猩猩对比。左图，根据实验设置测得的感知反应评分中，人类、黑猩猩和猩猩的得分没有显著性差异；右图，根据实验设置测得的认知反应评分中，人类、黑猩猩和猩猩的得分没有显著性差异[11]。

上述实验进一步表明，社会性是人的本质属性。中国传统思想中对自然的探索总是不能成为主流，一定程度上反映了中国人对社会的重视。如果说朱熹的“格物致知”之学中的物既包括伦理也包含物理的话，那么当王阳明继续了“心即是理”的主张时，他已经悄悄地将物理剥离了。



人类的社会认知甚至可能具有独特的生理基础。作者猜测，儒家所讲的人性的一些根本特征，“义”就是上述的利他与超级合作者特征，而仁的本质是同情心、恻隐之心，而“仁”与“义”在人脑的新皮层中可能都有其对应的基础。比如，一个叫做大脑颞顶接点(temporo-parietal junction)的结构，也就是大脑的颞叶和顶叶交汇处，对于包括视觉、听觉信息的整合方面非常重要，令人惊讶的是，Saxe 等在 2003 年发现，大脑颞顶接点也负责人们推测他人心理状态[12]。这项发现迅速得到了广泛的关注。大脑颞顶接点的损伤会导致病人很难做道德决定，甚至涉及濒死时的离体体验[13]。具体地，一种叫做镜像神经元(Mirror neurons)的独特细胞群，使人能对他人的意图、动作、以及感受做到“感同身受”。镜像神经元也与语言的产生具有很强的关系。人心对他人建模的区域很可能存在于大脑中的特定位置。

前文提到，U 系统的势能函数包含了自然之理（如，物理）与社会之理（如，伦理），人类的 V 系统是一组价值函数，这些价值函数定义在很多属性与变量上，我们简单划分为三大类：

- V_0 包括个人对环境的偏好，例如颜色、味道、气味、温度、安全；美学如人脸的颜值、服饰；环境的可供性和舒适度；经济能力等。
- V_+ 包括他人思想的状态，如“我”在“他人”心目中的多个维度的价值评价，“他人”在“我”心中的价值维度，如爱与认同、归属感。
- V_{++} 包括集体的价值维度，如家庭利益、公司利益、民族国家利益。

这三个层级可以大致对应儒家说的小人、君子和王者。此处的小人并非一种贬义，指的更多是普通大众，类似西方经济学中的所谓“自然人”，其特点是理性利己，中国有句话，叫做“小人喻以利”(V_0) “君子喻以义”(V_+)。 V_{++} 对应的是王者，是圣君贤



相应该有的价值，他们需要胸怀天下、舍小我而立公心。最高的价值是 V_{\max} ，代表了全人类的利益，是构筑人类命运共同体的基础（图 11）。

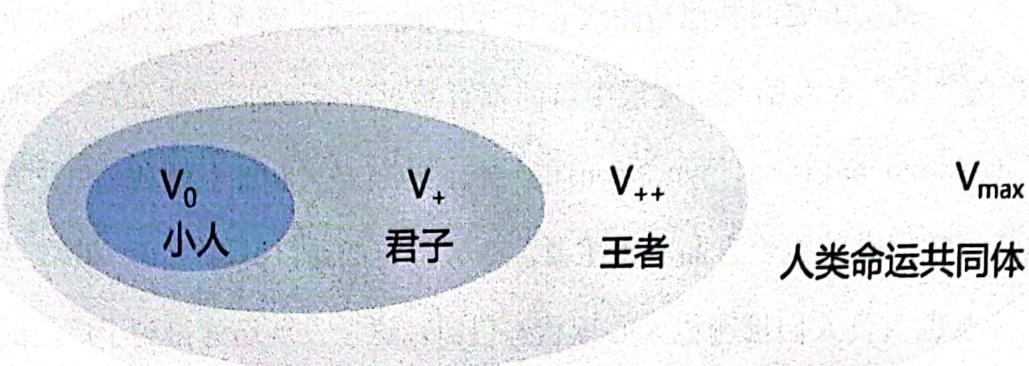


图 11：价值体系的逐渐升维。价值体系包含小人的 V_0 ，君子的 V_+ ，王者的 V_{++} ，直到人类命运共同体的 V_{\max} 。

从 V_0 到 V_{\max} ，价值空间不断扩大，“心”所关怀的对象不断延展。而我们所谓的格局，在数学上可以看作是价值空间的体积，取对数代表着维度，价值空间越大，一个人的格局就越大。

$$\text{格局} = \sum_i \log |\Omega_i|$$

这里面的 Ω_i 指的是一个人的 V 系统（价值函数）能够定义在多少空间、多大维度的子空间。

儒家所谓“风声雨声读书声、声声入耳；国事家事天下事、事事关心”体现了中国儒家文化对读书人要有家国情怀的要求，也是认知与价值空间的升维，下一节介绍认知空间。

2.3 认知架构：拓展、升维的思维空间

伴随着人类价值空间的升维，人类的认知空间也在不断升维和扩展，特别是让人能理解抽象概念与机构组织，预见并相信看不见的、未来的结果。



扫描全能王 创建

客观世界，如物理、化学和生物的描述空间是相对狭小的。主观世界不仅映射了客观世界（感知），还能映射自己的内心和他人的内心（认知）。每一个人都有不同的感知与认知世界，从客观到主观是空间的巨大飞跃，认知也同时发展出各种想象的世界。如图 12 所示，想象的虚拟世界包含：数学的各种几何、代数、拓扑空间；基督教的天堂、地狱，佛教净土宗的西方极乐世界、弥勒寄居的兜率天；文学作品中，如《红楼梦》中的“太虚幻境”、“离恨天”，《三体》中的半人马座 α 星三体人星球等；电子游戏中的虚拟世界、元宇宙。

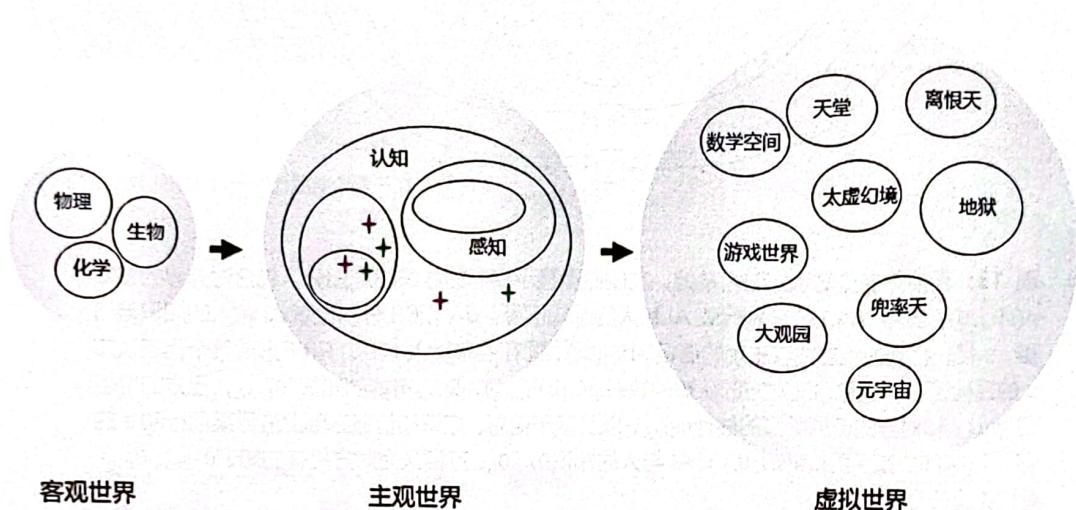


图 12：人类思想发展的过程是感知、认知、想象的空间不断扩展、升维的过程。

认知空间升维的基础是认知架构，后者也是人工智能研究中的关键问题。图 13 显示了两个对等的智能体（AI 与人脑）交流的认知架构，两个椭圆代表两个智能体认知的结构表达。智能体的认知架构包括 U 和 V 两个系统。U 系统包括认知模型 θ ，决策函数 π ，V 系统包含内在的价值追求。通过交流，双方实现四个对齐（Alignments）：

- 对齐共同情境 $p(s|I; \theta)$ (Shared Situation);
- 对齐共同常识 θ^* 也就是模型 (shared knowledge、model) ;
- 对齐决策函数，构成社会规范 π^* ;
- 对齐价值函数，构成共同的价值观 μ^* 。



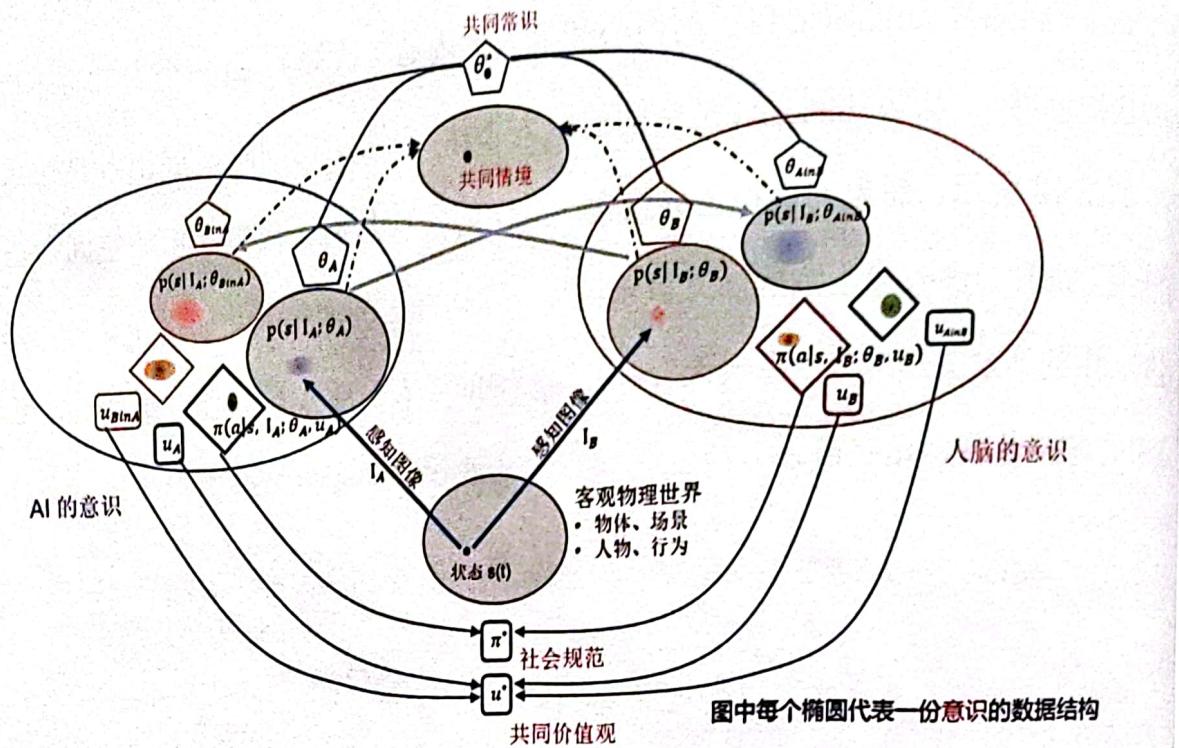


图 13: 智能体交流的认知架构基础。左侧椭圆表示 AI 的意识组成部分，包含对外界的感知 $p(s|I_A; \theta)$ ，其中 I_A 和 I_B 分别代表 AI 和人的感知信息， θ_A 和 θ_B 分别代表 AI 和人的知识与模型， s 代表客观的状态或者主观的信相 (belief)，即在信息输入 I 和知识 θ 下形成某个代表认知 s 的后验概率，还包含对对方的认知的映射 $p(s|I; \theta_{BinA})$ ，对方可能的知识 θ_{BinA} ，己方的价值函数 u_A ，和对方的价值函数的估计 u_{BinA} ，以及基于信息、知识和价值基础上进行某项行动 a 的条件概率即决策函数 $\pi(a|s, I; \theta, v)$ 。AI 与人的知识 θ_A 、 θ_B 、互相认为对方具有的知识 θ_{BinA} 、 θ_{AinB} 组成共识 θ^* ；

我们用一个例子来解读这些基本概念，人类在日常生活中看似极其简单的交流互动，其背后必须有强大的认知空间来支撑，这是大猩猩等物种所不具备的。图 14 是发生在 UCLA 的一个学生答辩活动，演讲者 M 正在专注地的介绍自己的成果，不知道答辩时间要到了；这时，前排控制时间的人 H 向 M 举起一个时间提示板，可是由于 H 所在的位置不在 M 的视角之内，M 还依然故我，对 H 的举手并没有察觉；这时后排一个学生 G 发现了 H 的意图，以及他知道 M 没有看到 H，而自己在 M 的视线内，于是 G 挥动了自己的手，吸引 M 的关注之后，指向 G；M 看到了 G 的示意，转头看向 H，向 H 领首表示自己知道了；H 看到 M 的示意后，知道信息已经被接收，就放下了提示板。



扫描全能王 创建

这个过程看似简单、平常，但如果从人工智能角度来看，实现这个过程相当的不容易！M、H、G 的大脑中要包含对另两方的觉察，比如 H 意识到 M 时间快到了，G 意识到 H 想提醒时间不多了；这时候 H、G 都知道，但是 M 还不知道；每一方不但要知道其他两方知道什么，更关键的是，还要知道其他两方脑中认为自己知道什么。这个信息的流程在图 13 下方，每一个椭圆都代表一个认知空间。不同颜色的十字代表不同的信息，它们在这些复杂的认知空间上传播。最左边的椭圆代表 H 的认知，包括 H 对 G 的认知空间的估计，也就是，H 推想 M 知道了什么；后者进一步嵌套了对 H 的认知空间的估计 - H 知道 M 不知道 H 的意图。最右下的大椭圆代表 G 的认知空间，G 观察到了全部信息，内心活动就很丰富。右上的椭圆代表的是 M 与 G 的共识，就是两人都



知道的信息。如此类推其他椭圆的内容。这个简短的社会交互任务的完成，需要复杂的大量的认知空间，这也是为什么其他动物，如大猩猩，是无法完成这样的任务的。

作为小结，本章简要讲述了三点：（1）回顾了从无生命的理运动、化学变化，到生命体、智能体、文明体的演化历程，中国思想本质上是对这个演化的过程、机理进行建模的集合，包括对自然的认识，对人性与社会的认知。（2）这个进化过程中最显著的特征是“心”的出现，它包括两大部分：价值体系与认知架构。

（3）任何一个个体的人、或者集体的文明，可以用 U, V 两套函数系统来刻画，认知架构是 U, V 的载体，而演化的进程本质上是 U, V 空间、认知空间的不断扩展、升维。

有了以上的基础概念上，下面我们就谈谈中国思想的变迁。



扫描全能王 创建

三、千年根脉：中国思想的五彩线模型

3.1 中国思想的五彩线模型：一元多体与多元一体

“红花白藕青荷叶，三教原来是一家”。这句话概括了中国思想的主要成分儒、释、道三股融汇的本质。事实上，中国思想的这种和衷共济不仅体现在儒、释、道的交融上，也一直蕴含在全部中国思想的变迁之中。因此本文用一幅斑斓的“五彩线”图模型（见图 15）来描绘这个融合的过程，在过去 5000 年中，以周易为起点与核心思想，对内融合儒墨道法等诸子百家，对外吸收古印度文明（佛教）、古希腊和基督的西方文明，宛如纺线机一般，将各色丝线通过旋转而形成粗绳，多元一体，随着时间而不断壮大。



扫描全能王 创建

图 15：中国思想的五彩线模型。 最上一行列出了时间标尺，其刻度是 1000 年，中国有大约 5000 年的文明史，这里显示从公元前 3000 年到公元 2000 年。接下来的一行是中国的朝代，再下面是典型的思想，如上古时期的河图、洛书，之后的周易、儒墨道法、禅宗、理学、心学等。在下面是代表性人物，如伏羲、孔子等。再下一层是一个五彩线的模型；其中一些关键时间点做了标识，如公元前 134 年出现董仲舒以及“罢黜百家、独尊儒术”。最下面列出的是其他文明对中华文明的影响，它们对中华文明的影响不是一蹴而就的，而是有个过程，用不同颜色的曲线表示。请注意，线条的长短近似表示思想流传的时间，但不能也无需过分精确，线条的颜色是随机选择的。公元前 800-200 年的轴心时代也特意用竖条标示了出来。

中华民族具有一万年的文化史、五千多年的文明史。在公元前 7000 年左右的新石器时代就诞生了很多文化，如河姆渡遗址、半坡遗址，公元前 3000 年即距今 5000 年前的浙江良渚文化出现了已知的中国最大的、世界最早水利灌溉系统，并有精美的陶器、玉器等，中华文明大约在这个时候开始出现。

下面根据图 15，我们简述这 5000 年文明史中的几个关键的节点。

(1) 第一个千年：公元前 3000 年至前 2000 年。河图、洛书等最早的数理认知，出现在中华大地上，也产生了阴阳、五行等对自然变化规律、物理因果的思想，传说的部落首领伏羲据此创造出八卦，文字也出现在这个时期。河南二里头文化的碳十四测定为公元前 2080-1620 年间，与传说中的夏文化的时间和地域吻合，已经能完成耗时



扫描全能王 创建

10 万个工作日、能容纳万人的土台。本文 4.3 节将从现代视觉建模的角度论证阴阳是自然图像（人看到的图片）的最显著信息特征。

(2) 第二个千年：公元前 2000 至前 1000 年。在此期间出现的夏、商、周分别在八卦的基础上发展出了《连山》、《归藏》和《易经》。据说《连山》是神农发展出的，为夏代所用，以艮卦开始，而艮卦是两个山，故称《连山》。《归藏》是黄帝发展出的，为殷商代所用，以坤卦开头，坤属地包藏万物，故称《归藏》。《易经》则是周文王推演出来的，为周代所用，以乾卦开头，也称为《周易》。《周易》被看作中国思想的根源，本文 4.5 节将讨论《易经》是世界上最早的统计决策模型，占卜是最早的随机计算方法，本质上是蒙特卡洛计算方法。

(3) 第三个千年：百家争鸣与轴心时代。随后在春秋战国时代，周朝不同部门的官员依据《易经》发展出不同的思想流派，主要的有儒、墨、道、法、阴阳家、名家、杂家、农家、小说家、纵横家、兵家、医家等，号称诸子百家，这一时期也被称为百花齐放、百家争鸣的学术繁荣时期。孔子的儒家在当时就具有一定的社会影响力，但其他流派的势力也很大，如墨家，并称显学。公元前 800-200 年被称为一轴心时代时期，是人类的觉醒期，各大文明几乎同时崛起，如古希腊文明、古巴比伦文明等。

(4) 第四个千年：罢黜百家，独尊儒术。到了汉代国家统一安定之后，公元前 134 年，儒家思想符合社会安定和治理的需要，在董仲舒的努力下，汉王朝“罢黜百家，独尊儒术”，儒学获得了极大的话语权，被称为经学，即经典之学，成为中国思想的主流，直至宋代的儒学的革新运动。

(5) 佛教对中国思想的深远影响。在春秋战国时代，佛教在古印度开始流行，数百年后，开始进入中国。公元前 2 年，大月氏王使臣伊存向中国博士弟子景卢口授《浮屠经》，这是佛教传入中国的最早文献记录[14]。到了公元 526 年，菩提达摩东渡，携



禅宗思想来到了中国。唐代时，禅宗成为具有极大社会影响力的佛教宗派。巅峰是禅宗六祖慧能的广泛传法，将佛教的思想与中国文化结合。慧能虽然不识字，但是颇具慧根，从《金刚经》中顿悟佛性的空与主观唯心的核心思想，在《坛经》中提出“菩提自性，本来清静，但用此心，直了成佛”，这里的菩提就是智慧，自性是在人的认知架构。然而物极必反，由于缺乏理论体系的支撑，佛教也自此走向衰落。本文 4.8-4.9 节将讨论佛教中的因果报应思想、禅宗的相由心生、禅机的认知、心智基础模型，并指出禅宗思想与为机器立心的高度关联。

(6) 第五个千年：理学、心学与朴学。由于宋代周敦颐、邵雍、张载、二程和朱熹等人的努力，儒学得到极大弘扬，于是诞生了理学。当然在那个时候，心学的苗头也开始出现，如陆九渊就在鹅湖之会上和朱熹分庭抗礼。但心学的真正兴盛出现在明代、王阳明出生之后。清代异族问鼎中原，文字狱严苛，所谓“清风不识字、何故乱翻书”。学者只能在故纸堆里面做些考据、音韵的学问，这就是朴学的由来，引用胡适的话，当西方开启文艺复兴的时候，中国思想在故纸堆里翻跟斗，翻了 300 年，导致了中国的落后。

(7) 西方文明对中国思想的冲击。就在满清闭关锁国、固步自封的时候，西方的世界早已天翻地覆。轴心时代发源的古希腊文明也在欧洲大陆潜滋暗长。以亚里士多德的形而上学为代表的古希腊文明薪尽火传到古罗马，基督教用信仰来回答了很多当时科学与逻辑推理所不能回答的世界和人生的极限问题，从而成为主流，1453 年文艺复兴后古罗马文明又孵化了近代西方文明。欧美列强携着因工业革命发展出的坚船利炮，通过 1840 年鸦片战争开启了中国近代历史。1919 年开始的新文化运动，否定了中国传统文化，开启了图 1 所示的百年大变局。如 1.4 节所讲，西方的逻辑思想为计算机科学与人工智能的第一阶段发展奠定了基础。



扫描全能王 创建

(8) 一元多体与多元一体的特征。关于中华文明的特点，一直以来有着一元多体和多元一体的争论。本文理解，元代表启元与思想的源头，体代表承载思想的群体（国家）。站在不同时间点和立场来看，中华文明的发展分三个时期：

第一个时期，站在周朝往前期历史是“多元一体”。这是对中华民族文化起源的判断，费孝通在《中华民族多元一体格局》这本书[15]中指出多元一体的具体构成：较早的时候，中华文化可以按地域分为黄河中下游文化区、长江中下游文化区、燕辽及黄河上游文化区、以鄱阳湖-珠三角为中心的华南文化区、以及北方游牧与渔猎文化区。地域的差异常常孕育不同的文化与文明。傅斯年则认为，不同地区，“因对峙而生争斗，因争斗而起混合，因混合而文化进展”[16]。到来周朝，“多元”成为了“一体”。

第二个时期，周朝瓦解了，出现春秋战国的百家争鸣时代（轴心时代），形成了一元多体，这里的多体是指诸子百家在不同诸侯国的实践活动，他们又都来自同一个源头（一元），如《周易》的思想。

第三个时期，汉代一来，除了短暂的战乱，国家维持了统一，保持了“一体”，就算是分裂时期，也并没有像春秋时期那样分化出诸子百家不同的思想。从这个统一体来看历史，思想是多元的，包括儒墨道法，还有外来的佛教、基督教的传入。因此这是更大层面、更持久的“多元一体”。

从总体来看，出去春秋战国时期的几百年，中国思想的主线还是“多元一体”。本文的五彩线模型就是对多元一体的可视化的展示，在春秋时期画出了一个短暂的一元多体（彩线的分叉），然后又在汉代合并。

西方常常说美国是世界文化的大熔炉，其实在5000年的文明史中，中国思想早就经历了对内、对外融汇的历程，具有巨大的包容性。

下面我们来简单讨论一元多体与多元一体，并试图解读中国的“和合”思想的来源与数理特征。



扫描全能王 创建

3.2 一元多体：轴心时代中国思想分化的根源

春秋时期或者西方学者说的轴心时代，中国思想的分化是如何形成的呢？

刘歆在《七略》中给出了具体的说明，这个说明后来被班固采用，整理进入了《汉书·艺文志》。周代的时候，掌控学术思想的人其实都是各个部门的官，周王室衰落后，这些掌握各种技能的官流落民间，学术思想开始发生分化。比如，儒家的人出于一种叫司徒的官，也就是掌管教育的人；道家出于史官，老子就是周室的图书官吏；阴阳家出于一种叫羲和的官，也就是掌管天文历法的人；法家出于理官，名家出于礼官，墨家则出于掌管祭祀的官（图 16）。刘歆的说法当然在细节上并非尽善尽美，但说出了一个真相，那就是中国思想的第一次分裂，始于由官而民，由公而私。就是这次的思想分化，在那个刚走出蛮荒却异常肥沃的远古思想旷野，也就是前文提到的某些认知空间，播下了五彩斑斓的文明之种。

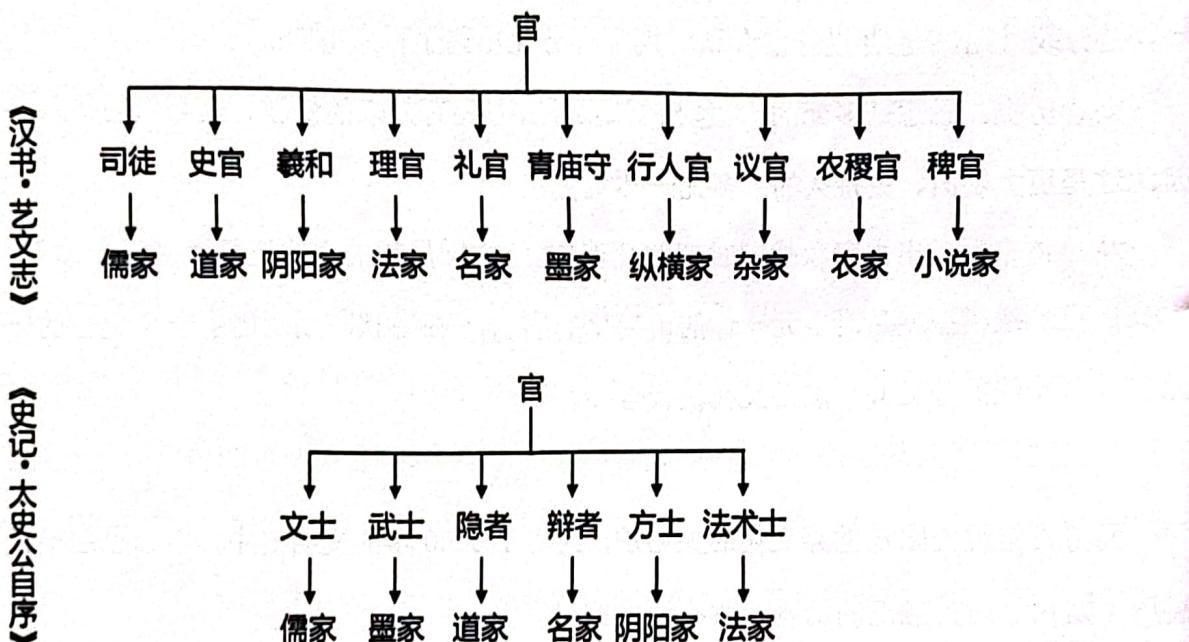


图 16：一元多体的具体内容。除了刘歆的说法，司马迁在《太史公自序》中也给出了类似但更简单的对学术起源的判断。

春秋时期的诸子百家虽然同出自周代的官的阶层，但作为思想的体系彼此之间则是有各种针锋相对的。墨家本身就以对儒家的批评而知名，比如儒家坚持厚葬，而墨

家则认为厚葬是一种巨大的浪费；儒家的孟子又反过来批评墨家和作为道家的杨朱，说杨朱无君，墨家无父；后期墨家还对名家、道家有过批评；等到了荀子，又对其他各家的谬误做了系统的阐述；荀子的学生韩非反而是法家，并对儒家做了尖锐的批判。作为思想体系的各家其实是多有龃龉。庄子说，“道术将为天下裂”，他有段很有名的话，白话翻译如下：

天下大乱，贤者和圣人湮没无闻，道德标准不一，天下人常常有如井底之蛙、得到一孔之见却孤芳自赏。这种情形就像耳目鼻口，它们各有其功能，但却不能互相通用；或者如百家众技，各有所长，时有所用。虽然如此，但只要不完备和全面，都是孤陋寡闻的人。割裂天地的完美，让万物之理分崩离析，把古人完美的道德弄得支离破碎，很少能具备天地的完美，配得上神明的称谓。所以，内圣外王之道幽暗闭塞而不发挥，天下的人各自为政。可悲啊！百家各行其道而不回头，必定不能相合。后世的学者，不幸不能见到天地的纯真和古人的全貌。

——《庄子 天下》

这段话反映了春秋战国时期学术百家争鸣的起源。当时周王室衰落，礼坏乐崩，社会阶层发生了巨大的变动，学术思想也发生了显著的变化。中国人从原始的蒙昧时期并未走出太远，所珍视的思想也谈不上规模宏大、体系完备。这就是中国思想的第一次分化，为日后中国思想的繁荣奠定了基础。

3.3 多元一体：中国思想对外来文明的融合

多元代表了三次融合：(1) 中华文明的起源的多元性，合于周朝（或者更早）；(2) 诸子百家思想，合于汉代；(3) 中国思想融入了外来的佛教、伊斯兰教、基督教的成分。本节简要讨论第三次融合。

佛教从公元一世纪进入中国，最初受压制，直到公元六世纪达到鼎盛，历时五百余年。佛教的传播可以说始于道，而终于儒。汉代之后，三国两晋南北朝又是中国历史的



“大分裂时代”。这一时代的人们思想中普遍存在一种忧虑。从曹操父子的诗词到《古诗十九首》，时代的阴郁笼罩着几乎所有人。魏晋玄学之所以流行，就是在这样的时代背景之下人们的一种自然反应。而同魏晋玄学在价值取向上接近的佛学，恰恰就在这样的契机下进入中国。但这个过程异常漫长，从最初的佛经翻译，逐渐过渡到佛教义理的阐明，500年的努力终于迎来了禅宗的流行，而佛教的衰落也始于禅宗。等到北宋众多思想家出佛入老，而终于回归儒家时，佛教思想早已融入中国思想了。

曾经有一本书，名字叫《佛教征服中国》[17]，应该说是中国征服了佛教，葛兆光在《中国思想史》[18]中即持有这种看法。中国传统思想是具有巨大力量的体系。冯友兰曾经说过，中国历史上两次被少数民族即元朝和清朝统治的历史，其实是政治上少数民族统治中国，文化上则是中国同化了少数民族。也正因如此，有一种说法叫中国是一个伪装成国家的文明。佛教作为来自古老文明的思想，当面对中华文明时，依然难逃被融合的命运。

伊斯兰教在唐代就进入中国，并在元代成为与其他宗教并行的独立宗教信仰，清代更是掀起了一场蓬勃的汉文译著活动。但是，伊斯兰文明既没有像佛教一样深入中国思想，也没有像西方近代文明一样撼动中国社会。这是因为，伊斯兰教进入的时候儒释道的合流已经接近完成，而又没有像近代西方文明传入所依赖的武力。相反，伊斯兰教受到儒家的深刻影响，例如，依托儒家的“一”、“忠”、“孝”的理念，中国伊斯兰教发展出“真一”、“真忠”、“至孝”的概念[19]，而身为回族的海瑞、李贽以中国传统思想著称，尤其是李贽，他的“童心说”被认为是对王阳明心学的继承。当然，伊斯兰文明也成为中华文明的有机组成，“伊儒会通”也确实是历史上不能被忽略的一幕。

西方文明从明代传教士利玛窦开始进入中国，但影响甚微，直到1840年鸦片战争，西方依傍坚船利炮打开清朝的国门，随后的1919年开启的新文化运动，否定了流行了2000年的子学、经学，按照西方的学科体系构建了今天的教育系统。这是我



扫描全能王 创建

们在 1.2、1.5 节讨论的话题，而且这个影响仍在继续，在这个过程中，中国接受了西方的理，而保留了自己的价值体系，这是当前中美竞争与冲突的根本原因，也是本文要探讨的主题。

3.4 中国思想“和合”的根源：信息偏差与描述式建模

五彩线模型大致勾勒了中国主要思想成分的演进过程，这个五彩线保留了各自的颜色，而不是混合成一色，这体现了中国思想最大的特点：和而不同、多元一体。不同的思想彼此融合，但同时又保留自己的个性：中国思想的融合并非是水汇聚成海消弭了自我，而是彩线，当下的斑斓也能追溯其组成的多元。比如百家争鸣时代儒墨道法异彩纷呈，到了汉朝罢黜百家独尊儒术的时候，也并非儒家幸存而其他思想消亡，事实上，其他思想同儒家发生了融合，但依然有自己的一席之地，比如人们形容汉朝治国时有“阳儒阴法”的说法。

中国思想的这种和合体现在很多方面，首先是天与人和，天人合一；其次是人与人和，天下为公、天下大同；再次是守中致和，认识到事物是由矛盾双方组成的，事物内部的矛盾可以互相转化，这是中华思想中的中庸之道。习近平总书记最近提到中华优秀传统文化时，也特别强调了这个突出特性：

中华优秀传统文化有很多重要元素，比如，天下为公、天下大同的社会理想，民为邦本、为政以德的治理思想，九州共贯、多元一体的大一统传统，修齐治平、兴亡有责的家国情怀，厚德载物、明德弘道的精神追求，富民厚生、义利兼顾的经济伦理，天人合一、万物并育的生态理念，实事求是、知行合一的哲学思想，执两用中、守中致和的思维方法，讲信修睦、亲仁善邻的交往之道等，共同塑造出中华文明的突出特性。

——习近平《在文化传承发展座谈会上的讲话》2023年6月2日

相比之下，西方思想则较少和合，基督教、犹太教和伊斯兰教的思想具有很强的排他性。比如《圣经》十诫的第一条是：



除我之外，你不能有别的神（You shall have no other gods before Me）。

这种对唯一性、一致性的追求是纷争的根源。

很少有人尝试探寻中西方思想和合与排他的根源。事实上，这可能源于中西方认知方式的巨大差别。本文从数理建模的视角，给出一个解释。

首先，我们看一个直观的例子。图 17 显示了一些长短与朝向不一的线段，简单代表人群中的思想取向的多样性。左图的分布 p_A 比较一致（西方的宗教信仰），右图比较多元（中国的多元一体）。这个时候，如果我们看其中的一根线段 a ，它在左图分布 p_A 就现在比较“扎眼”，不合群，但是，在左图分布 p_B 看来，线段 a 就不起眼。

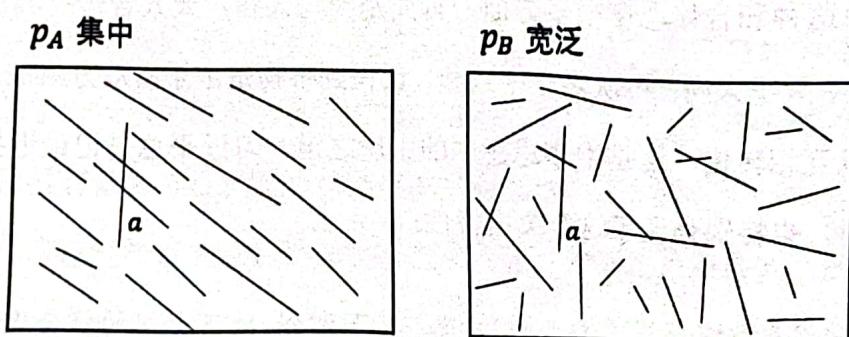


图 17：思想多元与包容的对比图。左图线段的长短与朝向比较一致，右图比较多元化，当同一个线段 a 放在两个图中，包容的不一样。

其次，我们可以用信息论中的信息偏差或相对熵（Kullback-Leibler divergence）来计算这个差异。以向量 x 表示各种思想， $p_A(x)$ 和 $p_B(x)$ 分别代表在人群 A ， B 中的概率分布（见图 18）。



扫描全能王 创建

$D(p_A \mid p_B)$ 大于 $D(p_B \mid p_A)$

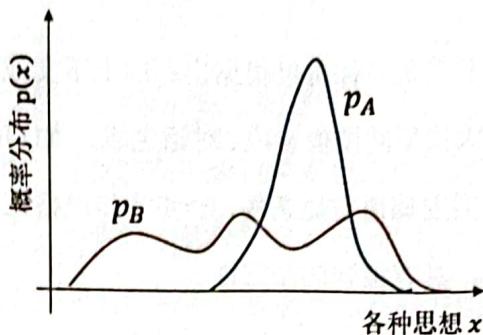


图 18：两个分布的信息偏差。

两个分布之间信息偏差分别为：

从 A 看 B 的偏差：
$$D(p_A \mid p_B) = \int p_A(x) \log \frac{p_A(x)}{p_B(x)} dx$$

从 B 看 A 的偏差：
$$D(p_B \mid p_A) = \int p_B(x) \log \frac{p_B(x)}{p_A(x)} dx$$

值得注意的是，这个偏差是非对称的，即 $D(p_A \mid p_B) \neq D(p_B \mid p_A)$ ，也就是说，它不是我们通常意义的欧氏平方差距离。如图 18 所示，如果 $p_A(x)$ 比较一致（代表西方宗教思想）， $p_B(x)$ 比较多元（代表中国思想的多元性合混合的特征），那么，

$D(p_A \mid p_B)$ 远大于 $D(p_B \mid p_A)$

也就是说，中国思想看西方思想觉得问题不大，可以并存，而西方思想看中国思想，会觉得问题很大、格格不入、不可理解。

再次，我试图探讨一下中国思想“和合”的根源。思想本质是对世界（物理和社会）的建模（下一章），在人工智能研究中，数理建模有三种主要的方式。我在这里就不作严格的定义，而是给一个直观的解释，与大家熟知的考试做类比。



- **判别式**: 类似选择题、判断题, 你不一定需要完全理解问题与原理, 而可以根据各种数据特征, 用启发式、排除法等做选择。人脸识别、车牌识别、语音识别属于这一类。
- **描述式**: 类似填空题、作文题, 你可以根据出题的上下文或者题目, 描述自己的理解、感受。当前的一些基于大模型的图像合成、对话生成, 如 ChatGPT, 属于这一类。
- **产生式**: 类似证明题, 需要搞懂前提条件, 一步一步严格证明、根据前提的事实做推理, 必须分辨是非黑白, 具有很强的排它性。

西方思想的源头是古希腊的逻辑、三段式的推论, 其思维以产生式为主, 希望刨根问底, 包括追寻世界的起源(物质细分成原子、基本粒子), 物种起源与演化。这种思维方式对于科学的发现起到很大的推动作用。对于科学回答不了的问题, 是通过信仰(faith)作为补充(排它), 并纳入的推理系统。

在作者看来, 中国思想的模型大多是描述式的, 描述式思维的特征首先表现在文字上。以表意汉字作为载体的中国思想从一开始就具有极其鲜明的描述式特征。本文4.10节讲从人工智能角度详述汉字的造字原理, 这里先说一下汉字的“理”字。理字是玉字旁, 最初可能和玉有关, 但很早的时候, 就引申为土地分成小块, 如诗经中有“我疆我理”, 然后引申为天理, 天既然有理, 那么万事万物就有理, 于是有了肌理、文理、条理, 于是人们理解事物时, 玉的纹路、土地的沟壑、文章的气脉乃至牛羊的肌骨都可以互相隐喻。而所谓的龟卜其实就是用火烤时出现在龟壳的纹理来暗喻事理。到了后来, 又有了所谓的“东海西海, 心同理同”。汉字的这种描述式特征对中国思想具有深刻的影响。

中国思想的描述式特征还在于具体的思想方式。西方世界在很久之前就有三段论式的推理模式, 依照演绎的方式形成推理以把握世界, 这就是产生式模型。中国人的思维模式更多的是联想, 也就是统计关联, 常常由一点相似进行展开, 这还是描述式的。



上一页刚刚提到，ChatGPT 也属于描述式的，它基于大量文字数据训练的大模型生成的回答，是基于统计的关联，模型里面包含了大量的、混杂的知识，时不时输出“逻辑混乱”的描述。

中国思想的描述式特征也反映在中国哲学家表达自己思想的方式上。冯友兰提到，中国哲学的特点是充满了名言隽语、比喻例证，明晰不足而暗示有余，前者从后者得到补偿。名言隽语、比喻例证其实就是针对事物的某些方面进行描述。比如有人问孔子，什么是君子，孔子没有给君子下定义，而是说“君子喻于义”，什么是好的政治，孔子也并没有给好的政治下定义，而是说“近者悦，远者来”，这就是好的政治，这都属于描述式的。

总之，描述式不需要严格的论证，往往不纠结对与错，也不排他，强调多元并存，和而不同，从而表现出很大的包容性，使用者可以各取所需。



扫描全能王 创建

四、千年根脉：中国思想的数理解读

中国思想是对各种自然与社会现象的认知建模的总集，根据其深度我们可以大致将其分成三类：

Hack: 经验性的工程实践，在某些地方、有时能起到一些效果；

Stat: 统计性的谚语规则，根据实践经验总结出来的统计模型；

Math: 确定性的数学模型，在某些条件下成立的定理与公式。

接下来，本文从五彩线模型中选择 10 个有代表性的思想，做简要的介绍，并指出其与现代数理模型及人工智能模型的关联。

4.1 中国最古老的数理认知：河图与洛书

大约在公元前 2000 年前后，河图洛书出现（图 19）。河图的特点是 1-10 十个数字的规律排列；洛书要更复杂，因为横纵斜的数值相加都等于 15，这其实是数学中所谓的“幻方”的一种最简单的形式。尽管对河图洛书的具体年代还有各种争议，但不妨碍我们对其有本质的认识：河图洛书其实就是古代中国人对数理的简单认知，其中包含了奇数、偶数。可以看出，河图洛书其实是非常简单的；然而，可能就是河图和洛书启迪了阴阳。



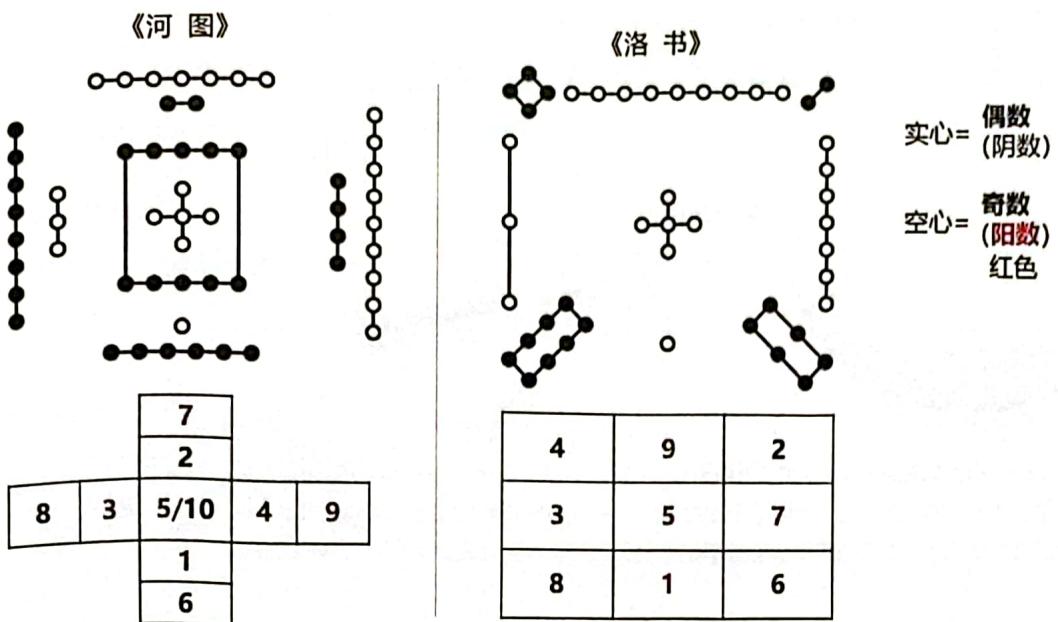


图 19：河图和洛书及其解读。河图其实就是中国古人发现的 1-10 十个数字之间的关系，比如 6-10 就是在 1-5 基础上加 5 得到的，古人可能在这样的过程中，发现了奇偶数的秘密，比如奇数 1 加了 5 之后得到的 6 是偶数。这可能看起来没什么了不起，但要注意，这是数千年前的古人做出的发现。洛书更加复杂，1-9 九个数字经过排列得到纵横斜的结果都是 15。这是一种最简单的幻方。所谓幻方，指的是 1 至 n^2 个自然数排列成 $n \times n$ 的矩阵，并建立 n 行、 n 列及二条主对角线等和关系的一种组合方式 ($n \geq 3$)。幻方是中国人首创，洛书是最简单的 3 阶幻方[20]。

4.2 中国最古老的几何认知：天圆地方

除了河图洛书这样的数理认知，中国人很早就有对天地宇宙的几何学的感知，这就是天圆地方（图 20 左）。古代中国的先民最初繁盛于中原地区，这是一片异常广袤的区域，迥然不同于两河流域、古埃及、古印度和古希腊。我们知道，因为地理位置临近海洋，“孤帆远影碧空尽”的视觉理解早早扎根，古希腊发展出非常发达的几何学，那时的埃拉托色尼（Eratosthenes，约前 275—前 194）就推断地球是圆的，而后来的麦哲伦（Ferdinand Magellan，1480—1521）也是靠着航海第一次证实了地球是圆的。中国位于亚洲东部、太平洋西岸，古人活动的地区又主要集中在中原地区，深入人心的则是“天似穹庐、笼盖四野”，天圆地方的观念也就不足为奇了。



图 20：古代的世界观。左图：中国古代的天圆地方观念：中国呈方形，周围是四海，上面则是圆形的天。在中国，天下近似等于四海之内。这在古希腊等地是不可想象的世界观。右图：更精细的模型，共工怒触不周山的故事说的是天柱折断，于是天塌西北、地陷东南。

《淮南子·天文训》记载了“共工怒触不周山”的故事，说共工和颛顼争做帝，失败了，生气发怒头撞不周山，结果“天倾西北，故日月星辰移焉；地不满东南，故水潦尘埃归焉。”天倾西北、地陷东南（图 20 右）是对天圆地方的修正。修正后的天与地两个平面，事实上是以泰勒公式展开的切平面拟合地球自传（星辰往西北走）和中国地貌（河流向东南流）。这一模型对于指导当时人的生产生活是基本足够了的。

中国古代曾经发展出非常发达的导航系统，这就是郑和下西洋使用的“过洋牵星术”。这是一种通过牵星板、利用身体、视角、特定星星的位置来确定航线的方法，居然有很好的效果。比如在郑和航海图中，有“北辰星七指”的字样，而在另一幅图中则变成了“北辰星一指”，北辰星就是北极星，降低了六指，这表明船向南行。在实际操作中，过洋牵星术当然不是依靠一颗星星定位的，而是凭借同多个星星的关系确定具体位置。比如在一幅记载了郑和从锡兰山回苏门答腊的过洋牵星图中，有如下记载：“牵华盖星八指，北辰星一指，灯笼骨星十四指半，南门双星十五指，西北布司星四指为母，东北织女星十一指，平儿山。”[21]也就是这段航线的确定采用了 6 个星星进行定位，类似今天的星链导航技术。



扫描全能王 创建

中国古代曾经发展出非常发达的地图测绘。早在 1136 年，中国就绘制了世界上最早的亚洲大地图，这就是禹迹图（图 21）。作者的博士导师曾经专门计算了禹迹图的误差（见图 21 右上）[22]，他认为这个地图的当时远远超过了西方的绘图能力。

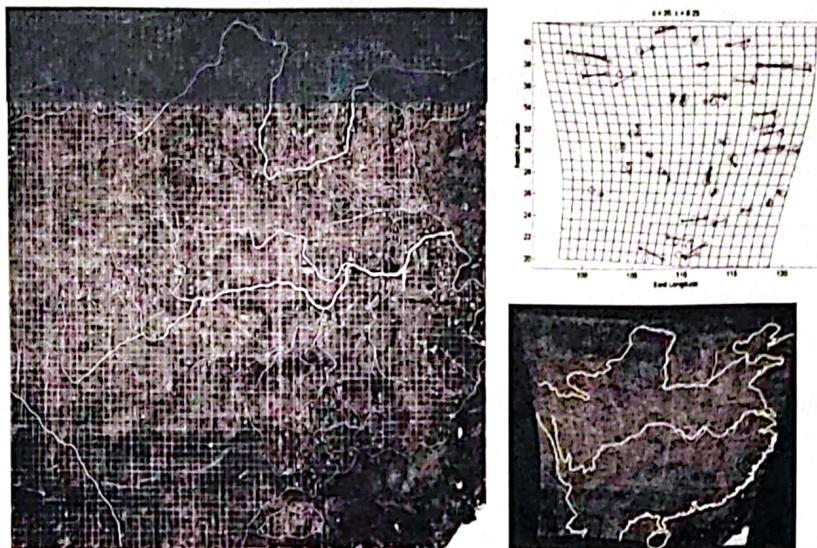


图 21：禹迹图。左图，禹迹图在美国国会图书馆的拓本。《禹迹图》石刻，目前仅存两块，一块在西安碑林，刻于齐阜昌七年（1136 年）；另一块在镇江焦山碑林，刻于元符三年（1100 年）。右上，禹迹图同现在地图的基于 45 个地标的经纬度参照。右下，禹迹图同现在地图的整体经纬度参照[22]。

4.3 中国最古老的视觉认知：阴阳

中国古代最初的具有广泛影响力的思想是阴阳。阴阳是中国人的独创，英文中阴阳就是专有名词 Yin and Yang。本文认为阴阳概念抓住了我们生活的环境中具有最大信息的图像特征。

当远古的先民睁眼看世界，一个身高 1.6 至 1.8 米的人看到的是大大小小的各种物体落在一个地平面上，物体的尺寸大小 r 投影到图像上大约服从一个三次幂律 $p(r) \propto r^{-3}$ ，这种我们在计算机视觉称作“自然图像”（见图 22 右侧），其图像的统计特征完全不同于飞鸟（或者站在山顶）俯瞰地面，所能看到的大地图像（见图 24）。



扫描全能王 创建

圖 22：人类平视环境而观察到“自然图像”。一个平均身高（1.6-1.8 米）的人在平视世界时的视觉成像结果。物体的尺寸 r 投影到图像上大约服从一个三次幂律 $p(r) \propto r^{-3}$ 。

需要说明的是，阴阳思想起源于齐，古代称为东夷，同起源于楚地的老庄思想不一致[23]。古代东夷族地处平原海滨，其视角常常是平的。

人们观察世界的图像集，计算机视觉称作为自然图像（natural images），只占全部可能的图像空间的极少一部分。我们假设一个图像有 256×256 像素，每个像素有 256 灰度值，其信息是 8 字节。全部图像集合的大小是 $|\Omega_0| = 2^{8 \times 256 \times 256} = 10^{157,830}$ 。自然图像由于高度冗余，每个像素的信息大约是 0.3 字节，由此推算自然图像的数目是 $|\Omega_{nat}| 2^{0.3 \times 256 \times 256} = 10^{5718}$ 。而人类能观察到的仅仅是自然图像中很少的一部分。我们假设人眼每秒能察觉 20 帧的图像，那么全球曾经活着的和活过的人大约是 100 亿，就算每个人活 100 年，最后的图像集的大小只有 10^{20} 帧（ 100 亿人 \times 100 年 \times 365 天 \times 24 小时 \times 3600 秒 \times 20 帧）。这个集合同所有图像 Ω_0 相比，还远远不如“沧海之一粟”。



扫描全能王 创建

这些自然图像是非常特别的集合，其特征是什么呢？我们用一个描述式的方法来定义这个图像I的集合：

$$\Omega_{Nat} = \{I: H_i(I) = h_i, \quad i = 0, 1, 2, \dots, K\}$$

其中 $H_i(I)$ 是从图像 I 抽取的统计量， h_i 是统计特征。图 23 最左边显示的是图像的在两个尺度下的梯度特征提取器 (filter)，也就是阴阳特征，图 23 右边显示了根据图像的对比度所抽取的统计直方图 $H_i(I)$ 。这个尖齿型的直方图在所有的自然图像中，具有高度一致性，而且是尺度不变的，也就是说，把自然图像说小成 $1/4$, $1/16$ ，这个直方图惊人地保持不变 [24]。这是自然图像特有的，其他图像，如航拍、遥感、沙漠、火星表面，都不具备这样的统计特征。

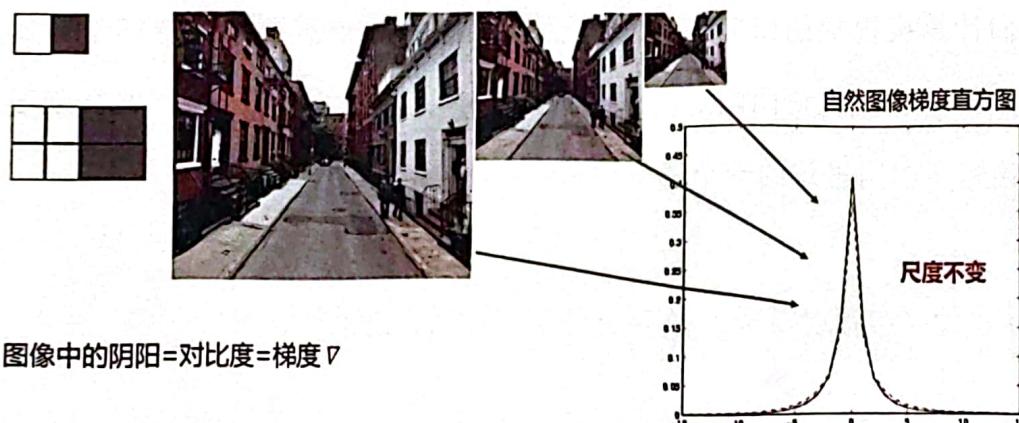


图 23：自然图像的尺度不变特征。自然图像无论经历多么大的压缩，其对比度的特征是不会改变的。以左侧图为例，图像的对比度不会随着尺度（1:1 或者 4:4）而发生变化。右侧表示的则是自然图像的直方图，其最主要的对比表现为图像的信号，而那些不重要的对比对最终图像的贡献较少 [24]。

不同的统计特征 h_i 是图像的描述与约束，是集合的内涵，内涵越大，外延越小。当 $K=0$ 时， $\Omega_{Nat} = \Omega_0$ 就是全部图像的集合，随着 K 的数目增加，集合就不断缩小，见图 24。而每次缩小的幅度，取对数值，就是新增统计特征 h_i 所带来的信息。具体值的公式为：



扫描全能王 创建

信息增益: $\delta_l = \log \frac{|\Omega_{l-1}|}{|\Omega_l|}$,

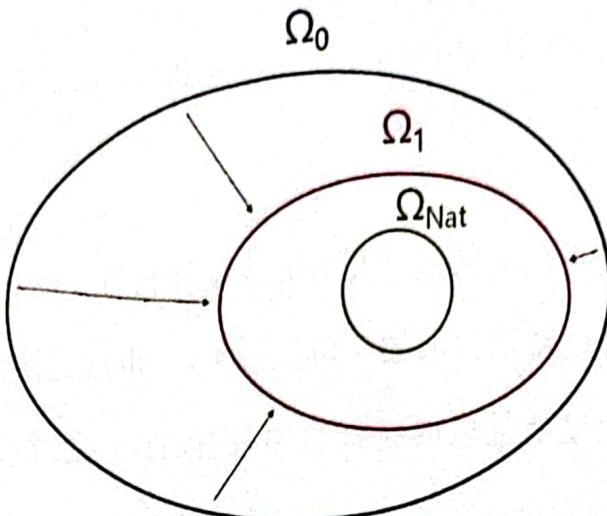


图 24: 图像集合。 Ω_0 和 Ω_{Nat} 分别代表全部图像、自然图像集合。如前所述, Ω_0 集合大小是 $10^{157,830}$, Ω_{Nat} 大小是 10^{5718} 。

根据我们计算机视觉的研究[24], 自然图像的第一个重要统计特征是对比度的统计直方图, 也就是说阴阳是自然图像最显著的特征, 比其他任何特征都要强烈。这就诠释了中国思想的一个通俗说法:

世界最大(信息最显著)的不变是变(对比、阴阳)。

而且这个变的统计量在不同的尺度上是一致的。这为后续《易经》的模型奠定了基础。

需要说明的是, 从上述公式中能得出一些哲学性的思考: 变是最大的不变。我们人类能认识“逝者如斯夫不舍昼夜”的大千世界, 就希望在变化中寻找不变的特征。



扫描全能王 创建

4.4 道家的风水：场景可供性

阴阳固然是中国先民观察地球表面的视角，但并不是唯一的。图 25 显示的是中国上古的先民鸟瞰地球的视角，这成为道家理论的一个关键部分。道家起源于楚地，而楚地多山，地形地貌同东夷所处的海滨不一样。这些人可能在群山之巅俯瞰山川大地，发现了苍山如龙盘虎踞，湖泊如龙飞蛇动，山脉和湖泊在地理上属于分形（fractal）的模式。

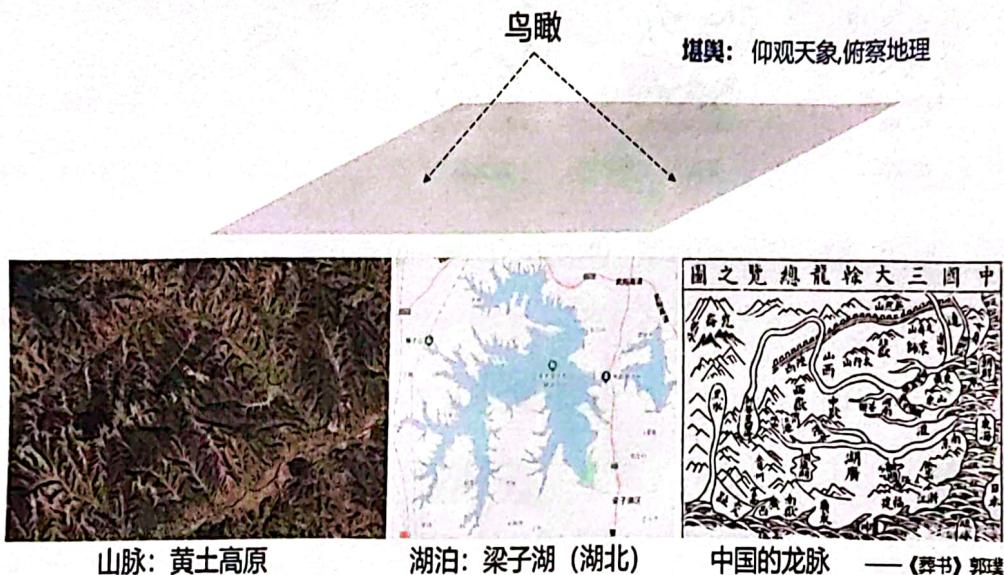


图 25：道家堪舆的俯瞰视角。观察到的山脉和湖泊的分型（Fractal）模式，根据山脉、水脉的走势而得出“龙脉”的概念。

晋代的郭璞写了一本《葬书》，是风医学的开山之作，就是总结各种地势，计算机视觉与模式识别称作模式聚类（pattern clustering），见图 26-27，其中有各种龙，如生龙、死龙、强龙、弱龙、顺龙、逆龙、进龙、退龙等[25]。黄土高原的卫星地图确实呈现出卧虎藏龙的地貌，而我的老家湖北有个梁子湖，俯瞰也是龙的形状。户外风水的本质就是模式聚类与识别；而居家风水就是指人住的舒不舒服，人工智能的专业名词叫作场景可供性，即 Scene affordance（图



扫描全能王 创建

中国思想中的风水理论，有很多有趣的、有用的理念，但同时也夹杂了很多错误的归因。古人对于人世中的诸多突变、肉眼不可观察的因素一定是在寻求解释，从而提出来很多隐含的变量（概念）。比如，将气候的变化（洪水、干旱）归因于神的不满；将病毒归因于鬼的报复，外出感染病毒被认为是遇到了鬼；将家族遗传所导致的智力差异与贫富的变故归因于墓葬的风水。

人类 1905 年才形成了基因的概念，而直到 1953 年才意识到基因的载体是 DNA。在这之前，人们不清楚为什么很多变故是怎么造成的。这里一个最显著的例子是智力基因的遗传，比如，有的穷人子女突然就聪明，而富人的后代就很愚鲁。现代一些遗传学认为，控制人的智力的基因在 X 染色体，因此，儿子（XY）的智力主要来自母亲（XX）。中国古代早就观察到了这个“外甥发达”的现象，认为这是风水被富家女子（可能是上一代基因好）从娘家带入婆家。那么是什么时候带走的呢？人们就从人生的某些关键时刻，比如生死、下葬的时刻寻找。因此，中国有些地方的习俗就禁止女儿参加娘家人的下葬场合。

中国思想中的提到影响人生的关键因素是“一命二运三风水”，这个风水应该就是我们今天所谓的遗传基因。佛教传入之后的以“业力”为核心的因果模型（后面会详述）也是用于寻找根源的，后来与道家风水合流，比如前面提到的《了



凡四训》中布施饭团和葬在特殊地点的例子，就是道家风水和佛教因果合并后的混合模型！这个模型通过引入很多变量如各种鬼、各种神、前世与今生的业力，用他们的状态来解释生活中所发生的因果、报应、公平正义。

4.5 《易经》与占卜：世界最早的统计决策模型与随机计算

从阴阳发展出了八卦，八卦重叠形成了六十四卦，其变化之规律被总结到《易经》。据说《易经》只是同类作品的第三部，其他两部分别叫《连山》和《归藏》（音 cang）。据说《连山》是神农发展出的，为夏代所用，以艮卦开始，而艮卦是两个山，故称《连山》。《归藏》是黄帝发展出的，为殷商代所用，以坤卦开头，坤属地包藏万物，故称《归藏》。《易经》则是周文王推演出的，为周代所用，又称《周易》。

人们推测这三个版本的演化绝非偶然，可能反映了远古人类社会的变迁：人类先后经历了山间穴居以采集为生、农业革命后平原聚族而居（母系社会）、直至群居的父系社会。

有三个人在《易经》发展史上非常重要，分别是伏羲、周文王和孔子。孔子对《易经》的涉猎很深，所以有“韦编三绝”的成语，让孔子多次翻断竹简的，就是《易经》，而不是其他的书。

从汉代开始，《易经》的影响力迅速扩大，有了“群经之首”、“大道之源”的提法。到了唐代，《易经》成为科举考试的教科书，作为正统思想而存在。宋代以后，五经在科举中逐渐被四书取代，但在漫长的岁月里，《易经》已经融入中华文明的血脉[26]。

我们熟知的很多成语都来自《易经》，如自强不息、否极泰来、朝乾夕惕、匪夷所思、刚柔相济、安不忘危，物以类聚、触类旁通、见仁见智、革故鼎新等



等。还有一些词汇，比如故宫的太和殿，太和两字就出自《易经》。我的老家湖北有个市叫做咸宁，咸宁两个字也出自《易经》。

《易经》还是一部具有世界影响力的著作。它有一个简单的英文名：《I Ching》，也叫《The book of changes》。众所周知，阴阳、八卦、易经都是典型的二进制表达，这是今天计算机科学基础[27]。心理学家荣格也从《易经》中得到灵感。

相当多的人认为《易经》不过是一部占卜的书，是封建迷信。作者认为，《易经》在圣人（周文王、孔子）思想中应该是简单、甚至是科学的模型，但是在3000多年流传的过程中，不断被加入了一些牵强附会的联想，不断被神秘化，造成了今天的误区。

本文从人工智能的角度来解释其本意：《易经》是世界上最早的统计决策模型，它的基本原理与人工智能的马尔可夫决策过程是高度一致的，而且是普适的、尺度不变的统计决策模型。

4.5.1 六十四卦是时空状态的聚类

首先，我们来看卦的表达意义。六十四卦代表了64种状态（states、situations），这个属于时空的模式聚类（pattern clustering），而且是尺度不变的，君王关注的大事，百姓日常的小事，都可粗略归于64种状态。

先把自然的时空状态划为八卦：乾（天）、坤（地）、震（雷）、艮（山）、离（火）、坎（水）、兑（泽）、巽（风）。然后，通过它们的上下关系来组成64个状态。古人同样给它们起了名字，如鼎、泰、未济、既济等。需要说明的是，六十四卦中八卦自身重复形成的卦并没有新的命名，还是保持和八卦一致，比如两个乾卦组成的还是叫乾卦（图29）。



古人在面临复杂事物需要决断的时候，常常因为考虑因素过多而无法掌控，这时候，就可以把自己的处境根据相似度归于某一个卦。如果不确定是否归于某一个卦，还可以出现“变卦”。六十四卦中的每个爻可能会发生变化，从而产生新的卦，这个过程称为变卦，变之前的叫“本卦”，变之后的叫“之卦”。用人工智能的符合来描述：将当前的状态记为 s , $s \in \{0, 1, \dots, 63\}$ 从 64 个卦的状态集合中取值，那么当前的状态就是概率 $p(s)$ 来表示不确定性。一般来说 $p(s)$ 集中在某个本卦，同时也有 5 个可能的变卦，其他 48 个状态概率为 0。就是说， $p(s)$ 是稀疏的表达。

每一卦中有 6 个爻，每个爻如果是阴，就叫六，按照自下而上的不同，叫初六、六二、六三、六四、六五和上六；如果是阳，就叫九，按照自下而上的不同，叫初九、九二、九三、九四、九五和上九。



扫描全能王 创建

4.5.2 时序特征：《易经》的卦爻辞

6爻被赋予两个关键的概念，就是时间和空间，分别称为时和位。

6爻被用来表示时序信息，代表状态发展的不同阶段，如乾卦描述的一个阳刚之人从潜龙、飞龙、到亢龙的人生阅历。这个时序信息也是这个卦的一个特征，



便于让算卦者对号入座。比如乾卦，初九是潜伏的龙，无所作为，九二就变成了巨龙出现在原野，九三则是朝乾夕惕奋发有为的状态，九四是或跃上天空、或滞留深渊的观望，九五是飞龙在天，九六则“亢龙有悔”、过犹不及了。其他的卦也有类似的时序进展如渐卦等。中国有很多对时的描述，如“生不逢时”、“时来天地皆同力，运去英雄不自由”。

六个爻的位置也有阴阳属性，即 1、3、5 为阳，2、4、6 为阴，这就是位。阴爻居阴位、阳爻居阳位就是当位，阴爻居阳位、阳爻居阴位就是不当位。以乾卦为例，六个都是阳爻，所以 1、3、5 是阳爻居阳位，属于当位，2、4、6 是阳爻居阴位，属于不当位。（图 30）。

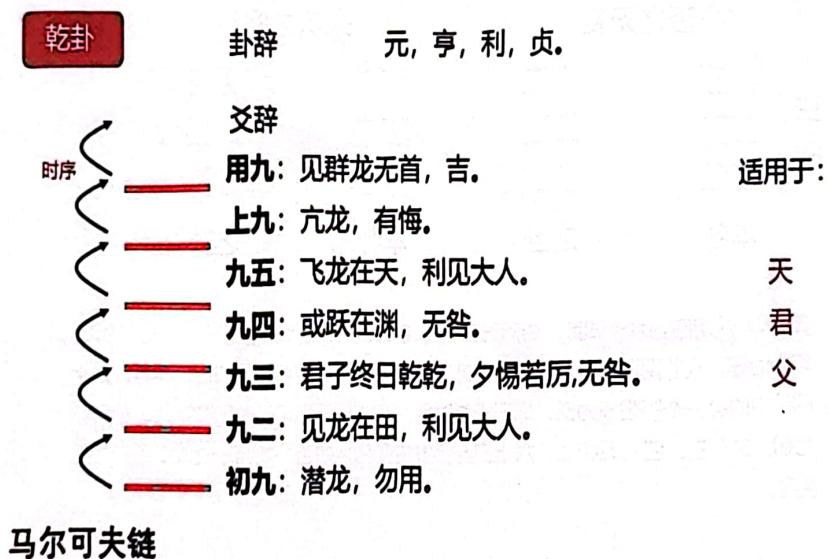


图 30：6 爻的“时”与“位”属性。这是《易经·乾卦》的卦辞和爻辞。从初九到用九是乾卦的时序特征，比如，九四阳居阴位，所以只能是或跃上天空、或滞留深渊的观望，而九五则是阳居阳位，所以就可以飞龙在天，大刀阔斧地开拓局面了。

所有这些附属在爻与卦上的时空属性，都是为了描述该卦所代表的状态，便于算卦者准确归类。比如，乾卦就适用于天、君、父这样的社会角色。



4.5.3 状态的转换规律：马尔可夫决策过程

将自己当前状态归类于某一个卦 s 之后，对于将要采取的某个行动 a ，那么就希望知道其后果，跳转到下一卦 s' 这就是求概率 $p(s'|s, a)$

假定一位古代君主想要计算打仗的后果（图 31）。这位古代君主浏览了六十四卦，发现有几个卦象同打仗有关，比如师卦、豫卦，临卦。通过进一步分析，这位君主发现，很可能导致师卦的状态。曹操在给《孙子兵法》的序中就提到了师卦。师卦的卦辞是总指挥处境安然，没有危险，具体地内容包括：行军征战要守军纪，不守军纪，必打败仗；主帅身在军中，吉利，没有灾祸；国君下令赏功，分封诸侯大夫，不能重用无才德的小人。那么该领导人处境同师卦的相似度是多少呢？他经过分析认为可能是 0.5（总概率是 1）。

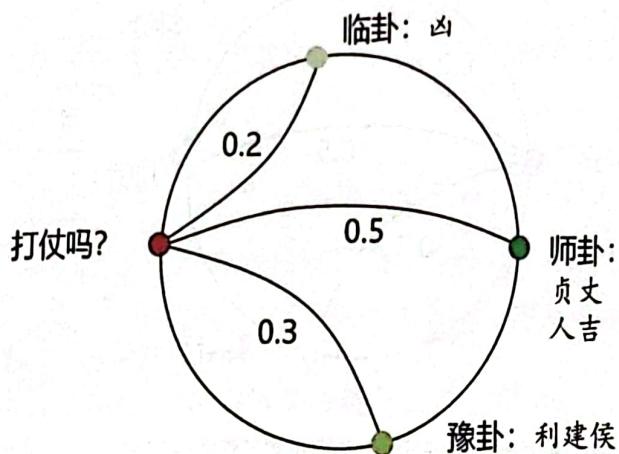


图 31:《易经》的马尔可夫决策过程。一个想要决策是否打仗的人，需要进行对情形的模拟，因此得出不同卦象，其概率存在不同的分布。豫卦的卦辞是“利建侯、行师”，即利于建立诸侯国和出兵打仗；师卦的卦辞是“贞，丈人吉，无咎”，即占问总指挥的处境，吉利，没有危险；临卦的卦辞是“元亨利贞，至于八月有凶”，即“进展很顺利，利于坚持下去，八月以后可能有凶险”。



但该君主并不满足，继续分析，他发现豫卦同打仗也有可能。豫卦的卦辞是：要坚持，坚持即使痛苦，也要坚持；顺从安闲、用人不疑；太过安闲则变成愚昧。该君主继续判断出现同豫卦的概率是 0.3。

当然还存在更多的可能性，如临卦。临卦的卦辞是利于坚持下去，但后期可能有凶险。该君主认为这个可能性较小，为 0.2。

当然，《易经》中的每一个卦并不见得精确符合当前状态。这时候可以用变卦。变卦是一个今天常用的词汇，指的是改变原本的主张或者已定的事情。这位君主进一步分析，发现最可靠的师卦可能并不完全符合实际情形，需要对模型进行修正，于是轻微改变了某一爻，比如九二，得到了坤卦，因而利用师卦为主、坤卦为辅进行决策，这时候的概率可能上升了（图 32）。

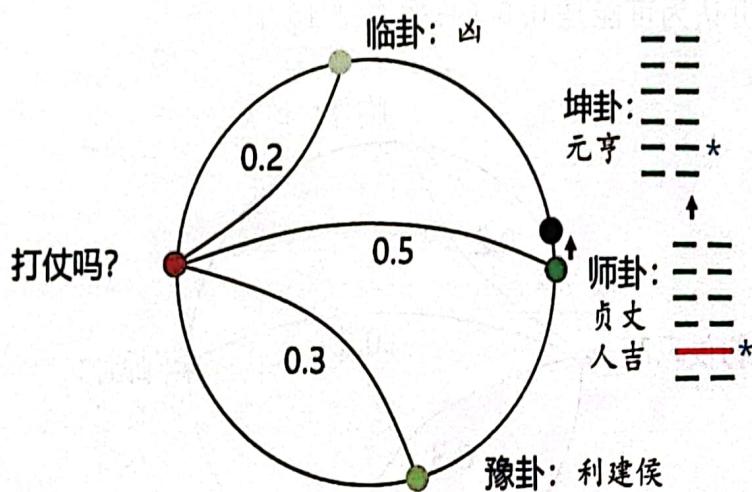


图 32:《易经》的马尔可夫决策过程的调整。师卦的第二爻由阳变为阴之后，卦由师变为了坤。坤卦的卦辞是：“元亨，利牝马之贞。君子有攸往，先迷后得主，利；西南得朋，东北丧朋，安贞吉。”意思是：很顺利，应当像牝马一样地坚持下去。君子应当有所作为，先迷惑后得到主导，有利；西南方向得到朋友，东北方向丧失朋友。安定地坚持下去吉利。

《易经》的这种作为数理模型的功能在今天看起来可能平平无奇，但在遥远的上古时代，这是伟大的创举。朱迪亚·珀尔在《为什么》中提到：“心理模型是施展想象的舞台，它使我们能够通过对模型局部的修改来试验不同的情形” [28]。



扫描全能王 创建

《易经》的马尔可夫决策模型也具有这样的模块化特征，这是人类认知的巨大进步。

4.5.4 占卜：蒙特卡洛采样

人类决策面临诸多的不确定性，不能采用确定性的逻辑推理与演绎，只能使用统计概率模型。比如一个计划打仗的君主，需要考虑他国的应对、自己国家的民心、将领的意志、军队的执行力等因素，还有很多不可控的外力，比如天气、病毒等。解决这些不确定变量的办法，用人工智能的语言来说，就是计算在各种可能情况下的统计平均期望值，把这些变量求积分，“积出去”。

由于各种可能的情况没法枚举，而只能通过抽样办法来近似计算，术语是蒙特卡洛（Monte Carlo）方法采样。蒙特卡洛是摩洛哥的一个赌场兼宾馆的名字，赌博业的商业模式就是用最基本的随机事件，如掷色子抛硬币，来获取最基本的随机数，原子随机事件，由此模拟任意复杂的随机事件。掷色子是 6 个状态，占卜是 2 个或 4 个状态。所以，本文认为，占卜是随机计算的基本运算单位。



扫描全能王 创建

图 33 简单解释如何通过蒙特卡洛采样方法来求平均期望值。

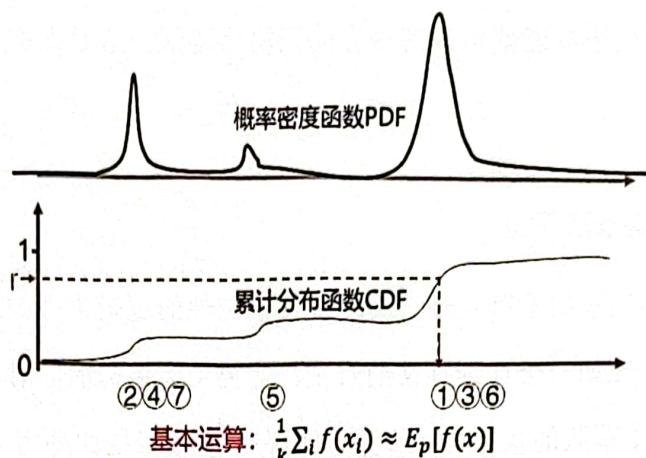


图 33：蒙特卡洛采样方法求平均期望值。

给定任意一个概率分布 $p(x)$ ，图上的概率分布有高低三个可能性（峰值），求和等于 1.0。我们首先将它转化成一个累计的分布函数，见图中，概率密度大的地方，这个累计（积分）分布就越陡。然后，在 $[0, 1]$ 之间，我们调用一个均匀分布的随机数 r ，见图中左边的竖轴。

这个随机数 代表了外部的不确定性，每次调用就相当于抛硬币、掷色子，本质就是占卜。根据 r 的取值，我们就找到对应的变量 x 。假设我们调用了 7 次随机数，那就得到 7 个样本，其位置见图 33 中的横轴，这 7 个样本基本符合概率分布 $p(x)$ ，故此集中在三个峰值之下，峰越高的地方，样本也越多。最后，我们就用这 7 个样本来求任意函数 $f(x)$ 的统计期望。也就是说，变量 x 的不确定性通过积分都被考虑进去了。

4.5.5 元亨利贞：激励函数

人工智能的统计决策模型都有激励函数（reward），就像下棋，每一步希望计算得失的分值。《易经》将吉凶按照从好到坏的程度分为六等：吉、吝、厉、悔、咎、凶。而这六种中的每一种都有进一步的精微区别。



吉就是吉祥、吉利的意思，但在《易经》中有“初吉”、“中吉”和“终吉”的按时间的划分，“贞吉（占卜吉）”这种按方式的划分，以及“大吉”、“元吉”（同“大吉”）这种按程度的划分。

吝是艰难、羞辱的意思，分为“小吝”（程度）、“终吝”（时间）和“贞吝”（方式）。

厉是危险但吉凶未定的意思，分为“有厉”和“贞厉”。

悔是后悔、忧虑，有烦恼的意思，分为“有悔”、“悔有悔”（困扰之事接踵而至）、“无悔”、“悔亡”（过去的困扰已经消失）等几种。

咎是出了过失、灾患，要承担责任，但比“凶”的结果要好一些，分为“为咎”（将成为灾患）、“匪咎”（不是灾患），“何咎”（不构成什么灾患），“无咎”即无灾患。

凶是祸殃，凶险，是最坏的结果，分为“终凶”、“有凶”、“贞凶”。

吉、吝、厉、悔、咎、凶之上，还有四类常见的断语，叫做元、亨、利、贞。这是《易经》中出现频率极高的词汇。比如六十四卦第一卦乾卦卦辞就是“乾，元亨利贞”，这四个字被称为乾卦的四种品质。按照《易经》的权威解读作品、唐代孔颖达的《周易正义》的说法，元亨利贞分别代表了阳气始发、万物亨通、和谐有利和坚固贞正。元亨利贞可以和具体的吉凶等结合，作为更加具体的描述，比如元可以和吉合在一起，称为元吉，比如坤卦中有“黄裳，元吉”的说法，就是很吉利。



可以将吉、吝、厉、悔、咎、凶和元亨利贞看作是激励函数（图 34），这激励系统有比较复杂的组合，作者还需要仔细分析。

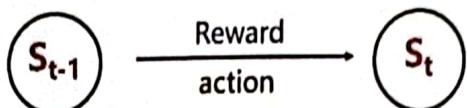


图 34：元亨利贞四类激励函数。人们在 S_{t-1} 时刻采取了某个行动 (action1)，获得了某种激励 (reward)，此过程在 S_t 时刻接续进行。在这个过程中，激励函数 reward 起到了很大的调节作用。在《易经》中，人们通过卦来决定采取何种行动，而卦中蕴含的激励函数更加细致，分为吉、吝、厉、悔、咎、凶，位于其上的则是元亨利贞。两者结合起来有更加具体的表述，比如元吉，就是非常吉利的意思。

1.5.6 《易经》的尺度不变性

综上所述，《易经》本质上是一个统计决策模型，其设计理念与人工智能的博弈模型与算法，如下围棋的 AlphaGo，理念上是一致的，这样的模型在 3000 年前无疑是惊人的成就。

AlphaGo 只能用来下棋，《易经》后来被赋予了广泛的普适性。帝王将相固然可以用《易经》决策军国大事，老百姓用它指导婚丧嫁娶。也就是说，《易经》的状态、转化、激励函数是尺度不变性 (scale invariance)。我们上文提到过自然图像的尺度不变性，指的是不论如何缩放，图像的统计特征等并不发生改变（见图 23），而《易经》就揭示出了事物的这种尺度不变性，当然，也有可能是后人过度泛化了这个模型的适用范围，更是加入了很多牵强附会的联想、类比，导致了其神秘和迷信的色彩。

比如对《易经》做解释的《说卦传》中提到，乾为天，坤为地，艮为山，兑为泽，震为雷，巽为风，坎为水，离为火。进一步，乾还可以代表君主、父亲等



等。也就是说，乾在不同的尺度上有不同的表现。这种尺度不变性是《易经》具有普适性的一个重要原因。

正因为如此，《易经》可以有不同层次的解读。以乾卦卦爻辞为例，我们固然可以用这段话来阐释一个奋发有为的旷世英主的奋斗历程，也可以用来形容一个自强不息的普通人的理性选择，既可以指导一个帝国的雄霸天下，也能教诲一个小组织的茁壮成长。

除了《易经》，五行也是诞生于中国远古时期的一种独特的世界观。对五行最初的记载出自中国最古老的、与《易经》同为五经之一的《尚书》，在其中的《洪范》篇中有如下记载：“五行：一曰水，二曰火，三曰木，四曰金，五曰土。水曰润下，火曰炎上，木曰曲直，金曰从革，土爰稼穡。润下作咸，炎上作苦，曲直作酸，从革作辛，稼穡作甘。”五行还有复杂的生克关系，如金生水，水生木，木生火，火生土，土生金；金克木，木克土，土克水，水克火，火克金。

五行的概念在其他文明中也有提及，但没有中国的五行那么系统，也没有提出五行生克的复杂关系。古希腊诸先贤中，哲学之父泰勒斯说，万物是水做成的，阿那克西美尼说气是原质，赫拉克利特说，万物都像火焰，色诺芬尼说，万物由土和水构成，恩培多克勒则说，有土、气、火和水四种原质。古印度则有地、水、火、风构成万物的说法。可以看出，这些对世界本源的看法都相当的粗糙简陋、抽象性也不强，本文认为，这些论断都无法同中国的五行思想媲美。

五行的本质是一种因果模型。五行的生克关系可以看作复杂的因果链。五行逐渐发展为一种重要的世界观和认识论体系，成为中国方术、中医的重要渊源。



扫描全能王 创建

因为 5 同 2 没有公约数，五行同八卦体系差异很大，但随着时间的推移，五行逐渐同八卦、六十四卦合流，并融合了天文历法中的天干、地支、二十四节气等内容，形成日益复杂的体系，可以看作是中国古代的“大模型”（图 35）。其本质都是希望通过更多的参数来拟合观察数据、来预测未来的变化。

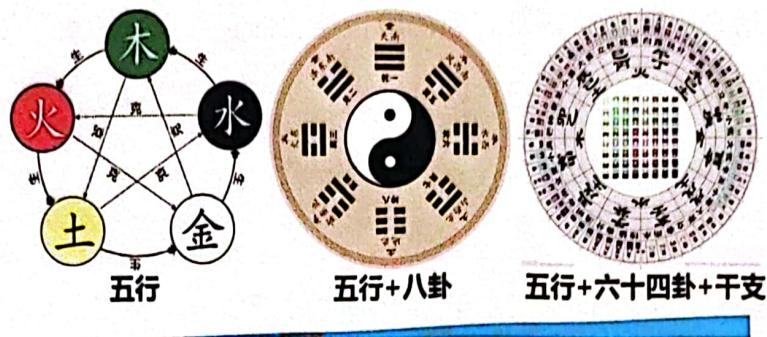


图 35：“中国古代大模型”，五行同八卦、六十四卦以及干支的逐渐融合。左侧，五行模型，其中外圈逆时针方向是五行相生的方向，内圈五角星形是五行相克的方向。中间，八卦同五行的结合，注意中心是阴阳，外面一层是八卦，再往外侧标注了八卦的五行属性，如乾兑属金、离属火、震巽属木、艮坤属土、坎属水。右侧，进一步融入干支、二十四节气以及六十四卦的“大模型”，注意中心是六十四卦的纵横矩阵排布，稍外是五行，接下来依次是天干、地支、二十四节气，以及六十四卦的圆形排布。作者在青城山就买到右图中这种复杂的罗盘，琢磨不明白其原来，这应该是一个复杂的计算装置。

这进一步说明，中国思想中从阴阳、风水、到《易经》，本质都是寻求对世界变化的规律（自然图像、居住环境、状态转化）的统计建模。《易经》代表了中国古代数理建模的核心思想，到了宋代，程朱理学将这种认识世界的方法进行总结、集成，将这一思想推向了新的高度。

4.6. 程朱理学：格物致知与数据驱动的概率模型

随着人类社会的进步和对自然世界的理解加深，用一部《易经》来推算所有的自然和社会的变化规律，也越来越力不从心了。要继续扩展《易经》的模型，有三条技术路径可走：



扫描全能王 创建

技术路径一：给《易经》的状态、属性、激励函数（即卦象、卦辞）附加更多牵强附会的类比和解读，这导致《易经》解读的随意性也越来越强；可操作性、重复性降低；从而迷信色彩越来越浓。

技术路径二：把各种模型进行拼凑，64 卦与五行、天干地支、二十四气节进行组会，这导致了图 35 那样的“中国古代大模型”。

技术路线三：针对不同的问题与领域开辟新的模型，这就是宋代理学。自南北朝、隋唐时期，佛教传入中国、并广为流布，在大量的佛教经典理论和实践的冲击下，到了宋代，儒家学术已经式微。北宋周敦颐、张载、程颢、程颐等人发展了“理学”，到了南宋朱熹将这些思想集成而推向高潮。

理学认为万事万物都有“理”，今天我们大学说的理科，包括数学、物理、化学、生理学、心理学，就是研究不同领域、不同尺度、不同维度的各种“理”。理学倡导“格物致知”，即仔细观察事物，相当于数据收集，然后提炼出知识，也就是今天的数理模型。格物致知本质是就是从数据到模型的知识发现过程，就是今天人工智能所讲的 Knowledge Discovery: From Data to Model。

中国儒家思想中，有两个重要概念：“道”与“理”。按照作者初步的理解，儒家的道指的是社会领域的道路、轨迹、规范。孔子说：“朝闻道，夕死可以”，这个道应该是指“在轨（on track）”，就是，社会治理和运行符合孔子的理想，老夫子可以放心了。“理”指的是各种模型。下面这个对话是朱熹的见解。

一名学生问朱熹到底什么是“道”与“理”：

（胡泳）问：“道与理如何分？”

（朱熹）曰：“道便是路，理是那文理。”

（胡泳）问：“如木理相似？”

（朱熹）曰：“是。”



扫描全能王 创建

——《朱子语类》卷第六 性理三

在南宋时代，朱熹无法用数理语言回答什么是“理”，而是指桌子的木头说：木纹有理，这就是“纹理”（texture）。那么到底什么是“纹理”呢？如何建立模型呢？朱熹并不能给出格物的方法论，这导致了明代的王阳明对如何观察竹子所提出的著名质问。

朱熹在 1160 年左右被问到而回答不清的问题，800 年后，到 1960 年被贝尔实验室的一位著名的心理物理学家 Julesz 重新提出。后来我把这个叫做 Julesz 之问：

“是什么图像特征与统计量，使得人眼在前注意阶段分辨不开共有这些特征的两幅纹理图片？”

“What features and statistics are characteristics of a texture pattern, so that texture pairs that share the same features and statistics cannot be told apart by pre-attentive human visual perception?”

图 36 显示两种木纹，图像块 A_1 与 A_2 , B_1 与 B_2 虽然不相同，但它们各属于一个感知的等价类。这又回到了我们在 4.3 节讨论阴阳的时候，用描述式的方法定义的一个图像集合。每一种木纹代表了一个图像的集合，他们在人眼看来（pre-attentive，就是在 200-400 毫秒左右的时间内，视觉感知提取的第一印象）是等价类。我们去挑选大理石地砖的时候，每一块都不同，但我们认为他们是一类。

$$\Omega_A = \{ I: H_i(I) = h_{A,i}, \quad i = 0, 1, 2, \dots, K_A \}$$

$$\Omega_B = \{ I: H_i(I) = h_{B,i}, \quad i = 0, 1, 2, \dots, K_B \}$$



扫描全能王 创建

在 4.3 节， 我们提到自然图像的集合 Ω_{nat} 包含了大量的平视世界的图片，其最大的共性特征是阴阳。 现在， Ω_A 和 Ω_B 是两个更小的图像集合， 它们的共性分别是 $h_{A,i}$, $i = 0, 1, 2, \dots, K_A$ 和 $h_{B,i}$, $i = 0, 1, 2, \dots, K_B$ 。 这些特征是人的视觉系统在早期（100-200 毫秒）抽取的、木纹特有的统计特征。

Julesz 开启了一系列的心理物理学的实验， 如图 37 所示， 人眼盯着屏幕某个区域， 图片突然显示在屏幕上， 记录人的反应和反应的时间。这个实验非常的简单直观， 但揭示了深刻的现象， 发现人的视觉早期感知的确对某些特征很敏感， 而忽视了某些特征， 见图解。



扫描全能王 创建

这些实验为大卫·马尔(计算机视觉的主要创始人)在 1970 年代末开启视觉计算理论提供了重要依据。1980 年代末, 以数理逻辑为基础的人工智能走到了道路的尽头, 人工智能进入所谓的寒冬期。而这个时候, 计算机视觉等学科得以发展, 统计建模的理论悄然开始了研究的范式转换。可以说, 对于纹理的建模是 80 年代研究的热点, 因为所有物体的表面都有纹理, 透过纹理才能计算背后的信息。

到了大概 90 年代中后期的时候, 作者在博士论文中, 通过统计物理的 Gibbs 模型与神经元模型结合, 最终给出了纹理的数学解:

对于任意一个图像块 Λ 和其边框 $\partial\Lambda$, 其图像 I_Λ 服从一个条件分布,

$$p(I_\Lambda | I_{\partial\Lambda}; \beta) = \frac{1}{Z(\beta)} \exp \left\{ - \sum_{j=1}^k \beta_j h_j(I_\Lambda | I_{\partial\Lambda}) \right\}$$

这是一个指数的概率分布函数, 而指数是我们前文所述的势能函数, 属于 U-系统的一部分:

$$U = \sum_j U_j(I_\Lambda | I_{\partial\Lambda})$$

这是统计建模理论里最早被攻克的问题。我们基于这样的公式, 可以人工生成逼真的图像(图 38) :



观察的例子



蒙特卡洛采样 MCMC 生成

图 38: 纹理合成实验 (1996-99) 左图是一张输入的大理石纹理图像, 从中提取特征统计量, 拟合 (学习) 上述模型。右图是从该模型用蒙特卡洛随机采样算法生成的纹理图片。

这个工作发表于 1996 年[24], 在计算机视觉领域引起了很大的关注。今天回头看, 这就是所谓的生成式 AI。当年算力和内存有限, 合成这么一张图片, 需要运算 2 个多星期的时间! 到了 2015 年, 我们把模型增加了神经元的层数, 改成

$$p(I; \lambda) = \frac{1}{Z(\lambda)} \exp \left[\sum_{k=1}^K \sum_{x \in \mathcal{D}} \lambda_{k,x} ([F_k * I](x)) \right]$$

就可以合成更复杂的图片, 如图 39 所示。



扫描全能王 创建

通过这个例子，我们可以看出，当前数据驱动的人工智能、大模型的热潮，其实跟理学的格物致知是一脉相承的。中国古代圣贤与当代人工智能研究的问题是相同的，思路是相通的。

4.7. 陆王心学：“心”即是“理”，构建通用人工智能的核心哲学思想

中国道家的“道”是指客观存在的、控制世界运行的根本法则与规律，包括天道和人道。儒家的“理”是人类提出的各种模型，是受人的认知和工具所局限的，现代科学就是用越来越准确的“理”去逼近“道”。那么，这一套理学的体系就完备了吗？非也！

程朱理学面临一个哲学上的隐患，那就是对于理的最终的认定者，只能是心。比如，模型对不对、合成的图像像不像，要有人眼来看！一个模型精确不精确，哪些变量必须考虑，哪些需要忽略，取决于人类的需求！社会的伦理关系（人道），更是需要考虑人的需要。

当程朱理学在宋代开始出现的同时，就诞生了心学。心学认为“心即是理”、“心外无物”，提倡“内圣外王”，即内心的价值 V 系统可以导出外在的伦理 U



系统。中国古人不大热心于物理、化学等自然规律，或者认为社会伦理的重要性远大于自然规律，这也是梁漱溟说的中国文化同西方文化的差别。

南宋时期（公元 1175 年 6 月），在吕祖谦的撮合下，朱熹和陆九龄、陆九渊兄弟在江西信州铅山的鹅湖寺见面，这就是历史上有名的鹅湖之会，其本质是讨论理学和心学的关系，心与理，谁是主导的？

朱熹的方法是“泛观博览”之后再分析、综合、归纳而得出结论，这就是数据驱动的理学方法，今天如 ChatGPT 这样的大模型，就是靠博览全部的文字资料，而获取丰富的知识。陆氏兄弟的思想是“先发明人之本心”之后再博览群书，这就是先构建价值 V 体系，然后在赋予技能。朱熹认为陆氏兄弟的讲法太过于简易了，而陆氏兄弟觉得朱熹的说法失之于支离破碎。

到了明代，王阳明进一步质疑理学的局限性，作者总结为两个要点：

其一：理学方法论的缺失，如何格物致知？他年轻的时候质疑，为何几天几夜观察竹子，而得不到一丝一毫的理。这个疑问被现代科学和人工智能的方法回答了。

其二：理学模型的泛化问题。每个模型都有适用的范围，如果出了这个范围，没有“理”了，该怎么办？比如，我们训练了一个机器人，也可以是无人机，可是到了一个陌生环境，训练数据没有覆盖到的情况，该怎么办？假如它学习、掌握了 N 个任务，遇到新的 N+1 个任务，怎么办？这其实就是当地人工智能发展的关键问题。



扫描全能王 创建

王阳明的“龙场悟道”回答了第二个问题，当人遇到新的、圣人没有教诲的情形，就要回归到本心，从价值出发，导出新的理！因此建立完备的价值体系、特别是良知，是比知识更重要的问题。现代教育和职场中的一个共识是：用人要以德为先，能力次之。这就回答了，心与理的关系中，心是起主导作用的。

王阳明说的理主要是指伦理、社会的规范，这并非否认客观与唯物的“道”的存在。前文已经讲过，就算是自然之“理”也是由人提出来的近似解释，不能与客观的“道”相混淆，这些理的构建要符合人的需求。随着社会越来越发达，我们进入一个“后真相”的时代，每个人根据自己的价值诉求，而选择性地接受不同的数据、从而提出不同的模型和解释。比如，人类活动是否是气候变暖的主因？不同的人有不同的认识。

前文在图 11 提到了价值的不同层次和人的格局的定义。尽管微生物、植物也具有对环境的各种物理、化学反应，低等动物也具有一定的认知能力，低等哺乳动物也有情绪、意志，但价值判断 V 是人类独有的、使得人成为万物灵长的智人的根本。图 40 细化了 V 的三个层级： V_0 主要指个体需求， V_1 代表社交价值， V_{11} 则意味着集体的价值。



扫描全能王 创建

V_0 , V_+ , V_{++} 分布体现了儒家说是的“小人、君子、王者”的价值格局。

每个人的行为都是在最大化其价值函数的加权求和：

$$a^* = \operatorname{argmax} (\alpha_0 V_0 + \alpha_+ V_+ + \alpha_{++} V_{++})$$

中国传统的所谓“内圣外王”，就是说，如果一个人如果有圣人之心，同时优化自己、他人、集体的价值， $\alpha_0, \alpha_+, \alpha_{++}$ 是权重，比如， $\alpha_0 = 0.1, \alpha_+ = 0.2, \alpha_{++} = 0.7$ ，就是 1 分为自己，2 分为周围的人，7 分给国家。如果决策行为 a^* 就符合大家的利益，行的就是“王道”。

在人群中，因着 UV 的分布，可以分成不同的类群。如图 41，U 代表能力的大小，V 代表格局的大小。绝大多数的人是在第一象限（能力和格局都很小），僧侶主动降低了 UV，基本出于原点位置。贤人隐士属于能力大、格局不够大的一群人。内圣外王的人位于更远的地方，但这些人的人群占比也极少。



扫描全能王 创建

图 41：人群中因 UV 的分布。 V 可以分为 V0、V+和 V++。绝大多数 V0 属于普通大众，他们的 V 同 U 相达到平衡，遵守社会规范、理性利己；较少的 V+则包括贤人隐士；极少数的 V++在更高的层次实现 UV 平衡，这就是所谓的内圣外王者。

UV 一般来说是需要达成一个平衡的，也就是能耐与格局相当。其平衡态构成了一个人工智能领域的所谓图灵停机 (halting) 问题。当 UV 不平衡时，人们会承受很大的压力或者内心矛盾。作者在《三读赤壁赋，兼谈心与理的平衡》一文中，就重点分析了苏轼在贬谪生活中的心理平衡问题。

《周易·系辞下》说：“德薄而位尊，智小而谋大，力小而任重，鲜不及矣。”，说的是一个人的 V 无法同其 U 相匹配。而当 UV 平衡时，人们表现出从容自恣的状态，孔子说的“七十从心所欲而不逾矩”，程颢写过一首诗，很好地表达了他处于这种状态的心情。

闲来无事不从容，睡觉东窗日已红。

万物静观皆自得，四时佳兴与人同。

道通天地有形外，思入风云变态中。

富贵不淫贫贱乐，男儿到此是豪雄。



扫描全能王 创建

但这种 UV 的平衡态的内稳定性是不同的，内稳定性越高，对抗环境的各种干扰的能力越强，在佛教徒看来，佛祖的内稳定性已经达到极致，到了涅槃境界，就不受任何外在条件影响。

以上提到的只是普通的、大多数人的 UV 平衡状态。当利益和能力到达一个很高的状态值，比如古代帝王王位、权利的争夺，就超出了这的范围，而进入李宗吾所谓的“厚黑学”的范畴（图 42）。马基雅维利在《君主论》[29]中有一章叫做“论以邪恶之道获得君权的人们”，其中举了两个靠践踏道德、要弄阴谋而成功的君主的例子，在马基雅维利看来，这是完全值得效法的，当然必须精明地效仿。马克思对《君主论》的评价是：“马基亚维利在书中的阐述使政治的理论观点摆脱了道德，而把权力作为法的基础，从而将政治学的基础由道德转向了权力。”其实《君主论》中关于权力不受道德羁绊的论述并非新说，韩非在《韩非子》中就有很多论述，比如他说，“事成则君收其功，规败则臣任其罪”，也就是君主可以推卸责任，让下属背黑锅。《君主论》、《韩非子》等书揭示出了君主常常从事着践踏道德的行为，而这一行为常常被历史有意地掩盖。这主要是因为帝王所处的权力顶峰常常超越了普通人，而不再满足受常规的道德的限制。《资治通鉴》中所记录的权谋历历在目，“窃钩者诛，窃国者侯”，远超过马基雅维利在书中的记述。



扫描全能王 创建

以上谈到了中国思想中，理学和心学的 UV 关系是最深刻的议题（图 43），这也是西方哲学探讨的核心问题。

理（物理、伦理）
是自然万物和人类社会的根本法则

心即是理，心外无物
内心价值可以推导出伦理



程朱理学：格物致知

- 物、人各自之理都源于天理
- 存天理、灭人欲
- 格物致知，如先行后
- 由外而内

由 U 规范 V

朱熹 (1130 - 1200)



陆王心学：心即是理

- 致良知
- 人性本善，应当发明本心
- 知行合一，内圣外王
- 由内而外

由 V 导出 U

王阳明 (1472-1529)

图 43：理学、心学的对应可以用 UV 理论解释。左图，U 体系同理学相对应。理既包括自然物理，也涵盖社会伦理；理学的代表是二程（程颐、程颢）和朱熹，其标志性的主张即“格物致知”，具体提倡“存天理、灭人欲”等，是由外而内的主张，可以看作是由 U 规范 V。右图，V 体系同心学相对应。心指内心价值。心学的代表是二陆（陆九龄、陆九渊）和王阳明，其标志性的主张即“心即是理”，具体提倡“知行合一”、“内圣外王”等，是由内而外的主张，可以看作是由 V 导出 U。

康德的墓碑上刻着一句话，就是这个问题：

“有两样东西，人们越是经常持久地对之凝神思索，它们就越是使内心充满常新而日增的惊奇和敬畏：我头上的星空和我心中的道德律。”

星空代表各种势能函数来规范的理，道德是指内心的价值体系。

作者认为，从理学到心学是一次认知的飞跃，V 与 U 互动和平衡是智能科学和文明演化最深刻的数理问题。

作为小结：我们回顾前文提到人工智能的发展的三个阶段。其中第二个阶段即概率建模与随机计算与“格物致知”一脉相承。这个方法的一个明显缺点是，大数据催生的人工智能系统缺乏内驱的价值体系，缺乏主观的能动性。第三个阶段则必须从“理学”过渡到“心学”，造出有主体意识的智能体，由“心”驱动，才能研发出自主的、安全的通用人工智能，最终要实现“心”与“理”高度统一。东方哲学为人工智能后半场的发展，提供了哲学层面的“顶层设计”。这就是为什么作者称要“为机器立心”，以中国之思想创世界之科技。



4.8 因果报应：价值的“记账”系统

上节谈到了中国思想的一个质的飞跃在于意识到：心是人类智能的主体，价值是驱动人类活动的根本因素。正所谓，“天下熙熙皆为利来，天下攘攘皆为利往”。当然，各人的格局不同，其追逐的“利”属于不同层次的价值诉求。

在此基础上，本节进一步讨论几个核心问题：

一、各类价值的得失是由什么因素控制的，底层的因果链条是什么？所谓，“君子爱财，取之有道”，这个道是什么？

二、是不是真的“恶有恶报、善有善报”？如果有，如何解释诸多生活中的反例？为什么不少作恶之人没有遭到报应，也有行善之人却有不好的结局，比如癌症、车祸？其实西方基督徒在向他们的上帝祷告的时候，问得最多的也是这个问题。

三、是否有一个超自然的记账系统，天上的玉皇大帝、西方佛祖、和地下的阎王爷是否建立了联合结算的会计系统，来平衡善恶与报应，这个记账系统跨度是多长，是否有个人的永久身份号码贯穿前世、今生与来世的信用，报应是否到自己的后代。

4.8.1 中国信仰：因果报应与公平正义

上述这些问题世俗社会每一个人的核心关切，也是定义人生价值的关键！但可惜历史上没有圣人能给出直接的答案与符合科学逻辑的论证，只能凭各人的信仰了。由于中国文化中缺乏犹太、基督、伊斯兰教那样的对于神的信仰，那么作为世俗社会中经过 5000 多年演化出来的中国思想，其信仰就是相信这个记账系统、因果报应、公平正义的存在。作者要在此再重复一下开篇在 1.1 节最后一段的结论：中国的信仰更接近于真相，更适合于未来的人机共生的智能时代，是比较宗教信仰更先进的信仰。



儒家的最高理想就体现在北宋张载的四句口号：

“为天地立心，为生民立命，为往圣继绝学，为万世开太平”。

我们前文一直在讨论如何定义人的“心”，那什么是天地之心？作者认为，这个天地之心就是要维护我们理想中的社会价值系统：善恶有报，信用记账，公平正义。这一观念曾经根植于中国本土，如《易经》的激励函数（元贞利亨），后来又进一来自佛教的体系化表述。

在佛教传入中国以前，中国传统思想中就有因果观念，但这是一种具有鲜明中国特色的家族因果报应。大概成书于战国时期的《易传》中提到：

“积善之家，必有余庆，积不善之家，必有余殃”。

积累善行，就会有值得庆祝的好事，积累恶行，就会遭殃，这是中国本土的因果观念。三国时的刘备临终就说过“物以恶小而为之，勿以善小而不为”，其实就是中国传统的因果观念的巨大影响的体现。中国本土的因果观念最大的特点是家族因果，也就是个体行为之因会成就家族人受报之果。随后，这种家族因果观念深入中国思想的血脉，以至于甚至数十年前，普通百姓过年的春联，常常是一句“向阳门第春常在，积善人家庆有余”；而就在十几年前，年画中一个胖娃娃骑在鱼上的形象也深入人心，并附上“吉庆有余”的字样，其中余同鱼谐音，而“吉”字可能就是“积”字的演化。

中国这种家族式的因果承接，显然是有助于以家族为单位的农耕社会的稳定，并由此派生出很多的内容。中国民间在今天也对“父债子偿”有很深的信仰。甚至法家的国家治理也沿袭了这种思想，如连坐，所谓“祸灭九族”的理论依据，其实就是家族因果。

但这种家族因果观念存在两个主要的问题：



扫描全能王 创建

一、个体的自由意志没有独立和重视。例如，一个人如果生在不善之家，自己再怎么行善可能还是会遭殃；

二、生活中常见好人遭殃、坏人吉庆的情形， 得不到合理的解释。

佛教的因果同中国因果思想天然接近，给佛教的在中国的流传提供了土壤，而中国因果思想没有回答的两个问题，也为佛教在中国的发扬光大提供了契机。这也应证了一句话：

思想的真空，我们不去占领， 一定会被别人填充！

4.8.2 十二因缘：三世两重因果

佛教因果观念解决了上述两个问题。

一、因果承受不再是家族式的，而是“自作自受”。自作自受这个成语最初就出自唐代敦煌的佛教内容中，后来经宋代、明代的广泛传播，成为中国人耳熟能详的词汇。印度佛教将因果从家族转向个体，其方式是通过提升心的重要性，而彰显了个体的自由意志。

二、因果不再是一生一世的，而是三世因果，前世作恶， 今世得报应， 今生作恶， 来世报应，“恶有恶报， 善有善报， 不是不报， 时候未到”。这就解决了好人遭殃、坏人吉庆的困惑。

佛教不同时期、宗派的因果思想有较大差异，但都承认个体自由意志和三世因果。佛教因果思想最具代表性的，是十二因缘揭示的三世两重因果（图 44）。所谓三世两重因果，就是：

前世二因成就现世五果；

现世五果导致现世三因；

现世三因引发来世二果。



扫描全能王 创建

图 44 中无明是一切烦恼的总称，指的是对于“缘起性空”这一原理并不明白，因而妄自生出一切执著；行指一切行为；识指的是业识，识会随着业受到果报；名色中名指心识，色指形体；六入就是“六根”，分别是眼根、耳根、鼻根、舌根、身根和意根；触是接触。六根同六尘（眼耳鼻舌身意感知的外界事物）和合而成触；受即领受。根尘相对生起苦乐二种感觉叫做“受”，即为对境所起的一种情绪；爱即贪爱，是对境所起的一种贪染心；取即妄取，追取，遇喜欢的乐境就念念贪求，必尽心竭力追求，如果遇到憎恶的苦境就念念厌离，必千方百计逃避；有就是业，就是有因就有果；生即受生，以现在所作的业为因，依因感果，导致来世受生；老死即衰老和死亡。无明和行是过去的因，结出的现在的果包含识、名色、六入、触、受；现在的果产生现在的因，就是爱、取和有；现在因又结出未来果，就是生和老死。这就是佛教的三世、两重因果。

十二因缘的三世两重因果理论首先将因果从家族中剥离，彰显自由意志和个体道德选择。这一方面削弱了中国传统思想中“一荣俱荣，一损俱损”、“一人得道、鸡犬升天”的内容，从一定程度上消解了农耕社会的家族纽带联系，但在另一方面，却也加强了个体的道德选择，避免了“世袭罔替”和“滥竽充数”，纠正儒家思想中的一些弊端，正因如此，儒、释合流，成为新的道德标准。



扫描全能王 创建

十二因缘三世两重因果理论的另一个效应是解决了好人遭殃、坏人吉庆的难题。将一世中的好人遭殃、坏人吉庆悖论拓展到三世，就能规避现世难题，从而给人以慰藉。三世两重因果在现世无法被证否，具有很强的解释性。佛教传入的三国两晋魏晋南北朝的大分裂时代，社会动荡不安，人们普遍怀有一种不安与忧郁，因此三世两重因果极大地抚慰了人心，很快赢得了中国人的信受，也就在中国生根了。

4.8.3 现世报：善有善报，恶有恶报，不是不报，时辰未到

十二因缘三十两重因果虽然解决了中国本土的因果观念，但是本身在形式上存在一个问题，就是过于复杂，不够简易。中国喜欢简易直接的东西，对复杂的逻辑思辨接受度低。于是三世两重因果就进一步简化为一句我们今天依然耳熟能详的话：善有善报，恶有恶报。这句话出自后秦时代就翻译出来的一本佛经，《缨络经·有行无行品》，其中提到：“随其缘对，善有善报，恶有恶报”。这句话深刻地塑造了中国人的性格。

佛教的因果观念从家族到个体，从今生到三世，固然解决了中国传统因果观念的重大问题，却不能不受到中国根深蒂固的传统思想的影响。中国思想的底色是关注今生，讲求立功、立德、立言，而对轮回淡漠。于是，三世因果又有了进一步的修正，到了大概明朝的时候，衍生出了所谓的“现世报”。“不是不报，时辰未到”所指的也常常并非三世因果，而是现世报。

佛教在实践中也常常夹杂着利益的诉求，和糊涂的因果观。图 45 就是一个农村礼佛的例子。比如，寺庙的和尚让信徒捐赠，信徒拜菩萨也有自己的利益诉求，如是就产生了这样的因果解释。信徒家的猪养得好，杀猪过年，和尚说这是菩萨保佑的结果，所以要捐钱。这个因果链条就是，捐钱导致菩萨保佑，菩萨保佑导致猪养得好。可是，捐款之后，第二年猪死了，信徒来问和尚怎么回事



事，和尚依然解释得通：捐的钱不够，菩萨不高兴了，所以猪养死了，所以，要多捐一点。

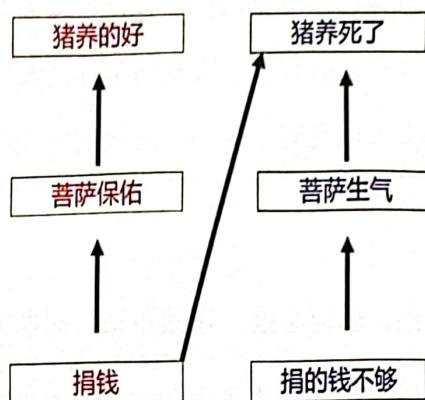


图 45：糊涂的因果。这个例子中，菩萨的态度是一个不可测量的变量，通过引入更多的隐含变量，就可以自然地解释生活中诸多现象。

这个因果模型最大的问题是缺乏干预的可能，也就是无法控制菩萨的情绪。前文 1.1 节提到《了凡四训》中林姓老妇因施舍受报的故事也是一个例子，这个故事中，仙人同样是因果链条中无法干预的一环。

4.8.4 现代因果模型

需要说明的是，无论是十二因缘，还是善恶有报，都是宏观层面、统计的规律。现代科学，包括人工智能还需要建立更具体的、可操作的、可验证的因果模型，其实，物理、化学、生物、医学等学科都在试图从统计的关联进一步寻找因果的关联。

因果可以分为真实的因果和伪因果。伪因果主要有两种情形，一种是由于干扰因素造成的伪因果，这个干扰因素的英文是 confounder factors，意思其实就是共生但很容易被误导为因果关系的现象。这样的例子非常多，例如，人们发现冰激淋热销和犯罪率升高有相关性，因而会误认为冰激淋的热销导致了犯罪率的



上升。这显然是错误的，正确的结论是天气是一个 confounder，即共生的干扰因素，同时导致了冰激淋的热销和犯罪率的上升。再比如，对于在校儿童，鞋的大小与阅读技能强烈关联，但人们都知道，学会新词并不会让脚变大，年龄是鞋子大小、阅读技能的共生干扰因素。

伪因果的另一种情形是选择偏差。比如一项大型调查研究的是去医院对人的健康状况的影响。该调查选择了 7000 多去过医院的人，平均健康水平是 3.21，以及 90000 多没有去过医院的人，平均健康水平是 3.93。如果仅看结果的话，可能得出“去医院让人的健康状况变差”的结论。然而，需要注意到，去医院的人显然本身的健康状况就要更差，只有在这些人中评价去医院对人的健康状况的影响才有意义。这就是选择偏差导致的伪因果的例子。

大卫·休谟认为，我们是没有办法去判断真实世界的因果的，所谓的因果，不过是“事物总是透过一种经常的连结而被我们在想象中归类”。以按按钮开电梯门为例，我们将按按钮判断为电梯门开启的原因。但是我们不知道是否有其他因素在这个过程中起作用。

有这样一个寓言故事，一位现代社会的女性去非洲的一个原始部落，该部落为了表示对这名女性的尊重，甚至准备了冷热水可调的淋浴，这令该女性大为惊奇，因为这在简陋的原始部落是不可想象的。而实际情况是室外站在一位女佣，她通过墙上小洞观察该女性的手指按的是冷水标识还是热水标识，并在外面将冷水或者热水注入简陋的淋浴系统。在这个故事中，按按钮显然并非是冷热水的因，女佣才是。我们要怎样才能发现躲在墙外的女佣呢？在休谟看来，女佣的存在是无法被证实或者证否的。科学实验就是要尽量排除各种隐藏的因素。

幸运的是，日常生活中的因果是非常直接、显而易见的。如图 46 所示，我们按一下按钮，按钮就变亮了。设计师的设计就是要让人放心，否则，你就不知道按下去的效果，不停去按。我们刷漆就立马改变了墙的颜色，吹气球就改变了



形状和体积，切洋葱就改变了拓扑结构。这里面没有太长的延迟、效果都肉眼可见。而生活中其他的因果关联就没这么直接，比如，吸烟是否导致肺癌，努力工作是否有回报，拜菩萨能否得到保佑。

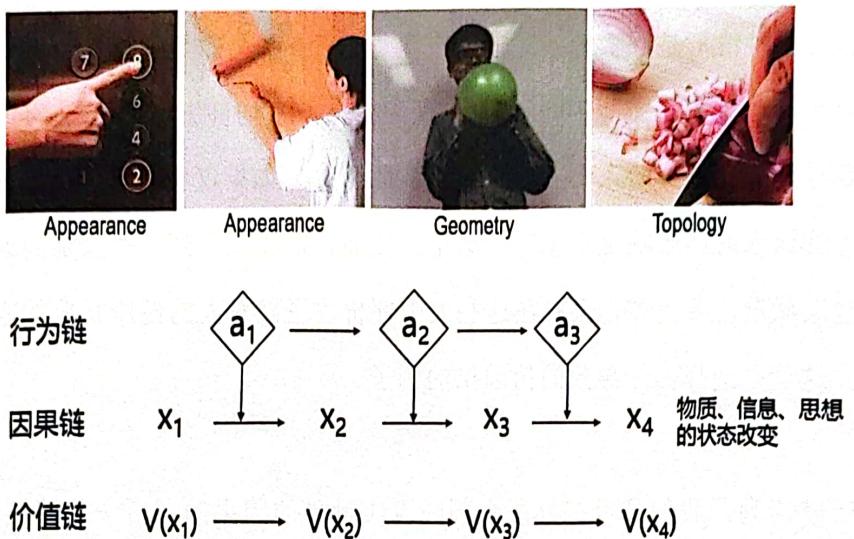


图 46：日常生活中的因果。它包含了三个链条：(1) 行为链：就是一系列人的动作；(2) 因果链：就是由动作导致的物体、信息、思想的改变，包括表观 (appearance)、几何位置与形状 (geometry) 和拓扑结构 (topology) 的改变；(3) 价值链：这些状态改变带来的价值变化。

作者在 UCLA 的同事 Judea Pearl 朱迪亚·珀尔《为什么：关于因果关系的新科学》的书中，他提到，如果想要辨别真实因果和伪因果，确定因果关系，需要分辨三种递进的模型：

一、统计相关模型，公式表述是 $P(y|x)$ ，也就是观察到当 x 发生时， y 可能发生的概率大概是多少，从而建立起 x 和 y 之间的相关性。前面提到的冰激凌热销和犯罪率上升就满足这种相关性。

二、行为干预模型，公式表示是 $P(y | do(x), z)$ ，也就是在条件 z 下，当对 x 进行直接干预 (do)，那么 y 会发生哪些改变？前面寺庙的例子中，和尚与信徒都无法干预菩萨的情绪。



三、反事实推理模型，公式表示是 $P(y_t | x^t, y^t)$ ，通过想象、回顾，分析假设 x 发生某种变化（这通常是不可实现的）时 y 的表现，从而判定是 x 导致 y 的因果因素

相关关系是无法同因果划等号的，还要加上干预和反事实推理。佛教的某些糊涂的因果关系恰恰缺乏干预和反事实推理，因而无法验证。其实，文科的研究因为无法做重复的实验，不能做反事实推理，因而缺乏实证性。比如，是什么因素导致中国在 2000 年前实现了统一，而欧洲至今没有统一？如果没有 1978 年的改革开放，我们今天会上什么样子？

在下一章，本文讲提出构建大型社会模拟器，用来模拟人类在多尺度、多维度的活动，通过 UV 的因果与价值模型来驱动大量智能体的行为，可以实现干预、反事实推理，从而能够回答文史哲政经法的很多问题，通用人工智能的发展讲为人类社会探索不同的道路提供了一个试验场。

4.8.5 中国思想的信仰：价值记账与三不朽

本节开篇 4.8.1 提到，作为世俗社会中经过 5000 多年演化出来的中国思想，其信仰就是相信这个记账系统、因果报应、公平正义的存在。本小节试图简要讨论这套记账系统，以及如何计算人生的价值：人活一辈子到底值不值、值多少？有没有意义、什么意义？

我们生活在大大小小的圈层之中。个体代表这个圈层的核心；父母、子女、兄弟姐妹代表了最近的圈层；家族构成了稍远的圈层；工作的同事、邻里又构成



扫描全能王 创建

了更远的圈层。在某个时刻一个人的核心价值就是自己，及以自己为核心的圈层的利益的加权求和（图 47）。

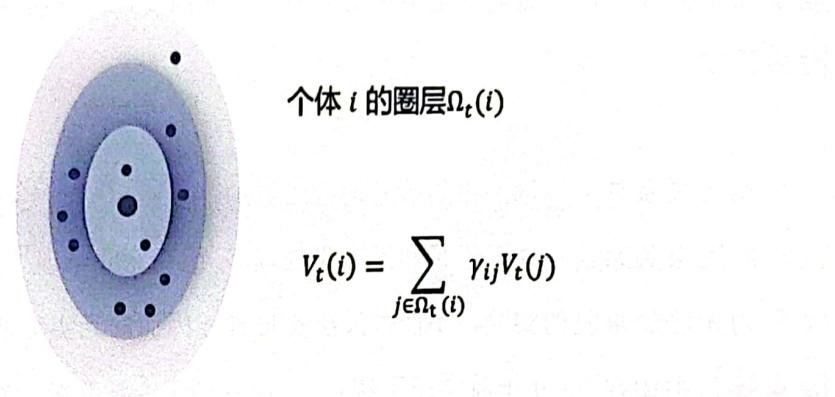
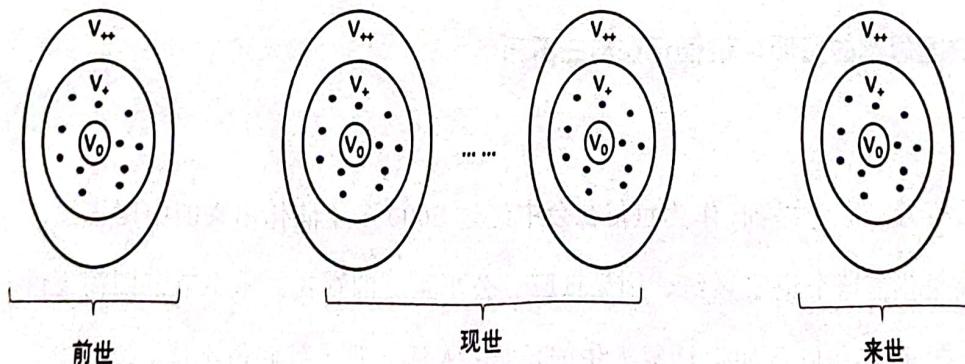


图 47：圈层社会与价值求和。某个时刻 t ，某个个体 i ，以及其关心的圈层 $\Omega_t(i)$ 的利益的总和， γ_{ij} 是个体 i 对其圈层中个体 j 的价值影响和贡献的系数。

在受到印度因果思想影响后，就成为三世因果，价值追求就变成了前世、现世、来世这三段时间的价值求积分（图 48）。在这些求和的递归计算函数中，最基本的叶节点还是 V_0 ，也就是个体的各种体验与享受，过上什么样的美好生活，其他的 V_+ , V_{++} 是指对别人和集体利益的提升。



$$V(t) = \int_0^T \sum_{j \in \Omega_t(i)} \gamma_{ij} V_t(j) dt$$

图 48：受到印度思想影响后的中国人的价值追求。中国人的核心价值变为以自己为中心的圈层的利益在不同世代的求和。



在儒家入世的理想中，最高的价值追求是“不朽”，也就是，假设人类文明不灭的情况下，个体的影响力通过无穷时间的传递，从而这个积分发散，到达无穷。

$$V(i) = \int_0^\infty \sum_{j \in \Omega_t(i)} \gamma_{ij} V_t(j) dt = \infty$$

所谓立德、立功、立言三不朽，说的就是这种情形（图 49）。司马迁在《孔子世家》中说，天下的君王、贤人们很多，在世时荣耀无限，但死后也就默默无闻了；孔子只是一个普通百姓，但到司马迁所在的时候传了十几代，上至天子，还是以孔氏为师承，可以说是“至圣”。孔子就是不朽的代表。

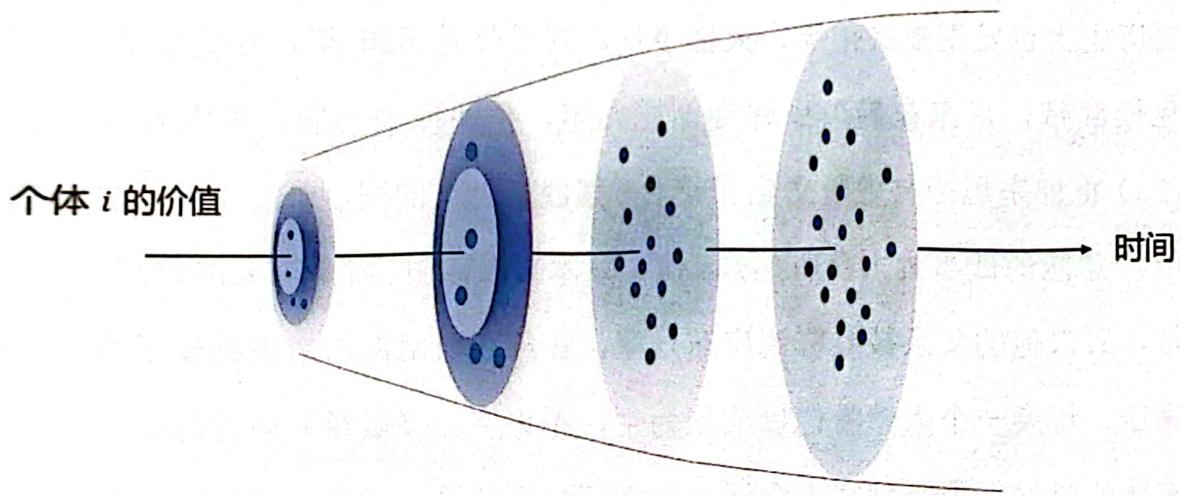


图 49：价值追求的最大值：不朽。个体的影响力通过无穷时间的传递，从而这个积分发散，到达无穷。

胡适在 1919 年写了一篇短文《不朽——我的宗教》，总结了自己的追求和人生意义，就是希望达到不朽。这种对影响力的追求超越了个体的自私自利的算计，与现代西方心理学和各种励志文字所讲的实现自我价值与潜能是一致的。



扫描全能王 创建

4.9 禅宗：基于心智模型的高阶通讯与主观唯心论

因果是小乘佛教的重要思想之一，禅宗是大乘佛教的主要流派之一。本节简要讨论一下禅宗，从而解释中国思想中“心”的另一个完全不同的、极其重要的含义——认知架构。

4.9.1. 禅宗的主观唯心论

据说最初在灵山之上，佛祖拈花微笑，众人不解，唯独迦叶微笑。佛祖说，“吾有正法眼藏，涅槃妙心，实相无相，微妙法门，不立文字，教外别传，嘱咐摩诃迦叶。”这就是禅宗的由来。禅来做音译“禅那”，原指冥想、沉思。

历史上真实的禅宗传入中国同样充满传奇色彩。佛教传入中国经历了多次波折，中国历史上也发生多次排佛、灭佛事件。到了公元 526 年，菩提达摩东渡，曾经觐见梁武帝，后来在嵩山少林寺面壁九年，传法禅宗一脉。禅宗六祖慧能（638-713）将禅宗思想与中国文化相结合，摆脱读经文的限制，致使禅宗在中国广为流传。慧能的故事为大众所熟知，他老家本来在范阳，就是今天的北京大兴、宛平一带，后来他的父亲被贬官到广东地区，在广东西南部的新兴县安了家。慧能父亲早逝，母亲一个人带着他以砍柴为生，不识字。《坛经》中慧能自己讲述有一次卖柴的时候，因听到客人念诵《金刚经》而开悟，《金刚经》这一句的核心思想是：

“应无所住而生其心”

这里说的“心”，不同于阳明心学的价值体系，而是一套复杂的认知架构，此“心”是认识这个世界，形成伦理道德的关键，见图 13-14。

作者回国后，2021 年去新兴县拜访了慧能的出生地（图 50），感谢地方领导的安排。慧能几次出场的言论，都是惊世骇俗的，绝对是颠覆性的哲学思想。



慧能第一次出场是在湖北黄陂寺庙中学习时，让人写在墙上回怼师兄神秀的偈语：

“菩提本无树，明镜亦非台，本来无一物，何处惹尘埃。”

这句话诠释了《金刚经》的“应无所住”：心中放下一切执念、不着相。凭借这个悟性，慧能获得了五祖弘忍的信任，得传衣钵。作者读《坛经》的时候，发现慧能所描述的场景与当代的研究群体有几分相似。围绕弘忍周围的学生，基本都是世俗之人，对名利有执着的追求，因而蠢俗不堪，只因执念太盛，而无法跳出日常的思维，从而不能做到思想的创新。

慧能第二次出场是在隐居十五年之后，到广州法兴寺，正逢值印宗法师讲《涅槃经》，时有风吹幡动。一僧曰：“风动”，一僧曰“幡动”。议论不已。慧能进曰：

“不是风动，不是幡动，仁者心动。”

这个偈语反映的是佛教中的一个核心思想—“相由心生”。人能看到的首先必须是心中知道、心中所想之事。比如，假设当时有一头猪在现场的话，猪是看不到风动和幡动，因为猪没有象人一样的心智。人的注意力如果在听经上，也不会看到风和幡。



扫描全能王 创建

我们今天在互联网时代，社交网络的事件，我们大多数人都没有看到，就算注意到了某个事件，每个人都看到的是不同的“相”，这个相是各人心中所产生的，不能信以为真（即，应无所住）。网上很多认知障碍之人，往往就是有了执念，而限制了思维，变得偏激。

坛经中的中心思想，被反复提到的，就是下面这句话：

“菩提自性，本来清静；但用此心，直了成佛！”

菩提就是智慧，自性就是天生的认知架构，见图 13-14。只有放下执念，不被外界各种纷扰的因素所蒙蔽，用这个认知架构来思考、参悟，才可想明白道理。这就是所谓“应无所住而生其心”！佛就是觉悟的人，也可以是任何有主体意识的智能体。作者提出要为机器立心，这是心的一部分。作者认为，认识到智能的主观、唯心的本质，是实现通用人工智能的关键，这就产生了与大数据驱动的人工智能算法截然不同的实现路线。

4.9.2. 认知架构：谈玄与禅机的高级通讯模式

有了这个天生的认知架构，我们就可以来理解禅宗的谈玄与禅机，这个认知架构（图 13）包含了现代心理学的心智理论，是一个高级通讯的数理模型！

禅宗逐渐发展出一种所谓的公案。禅宗公案可能来自魏晋玄谈。《晋书 王羲之》记载，王羲之的三个儿子王徽之、王操之、王献之去拜访谢安，王徽之、王操之喋喋不休，谈了很多俗事，王献之却只是嘘寒问暖而已。出来后，鄙人询问谢安，谢安说，小的（王献之）最好。别人问为什么？谢安说，吉祥的人说话少，王献之说话最少，我因此知道。《世说新语》中还有一个人因为说话少而被授予官职的故事，因为他说了三个字，被当时的人称为“三语掾”。



扫描全能王 创建

禅宗公案后来发展成为禅宗机锋，最大的特点就是用简短的文字甚至手势、棒打、大喝乃至于默然不语来彼此印证。佛祖拈花、迦叶微笑、达摩面壁逐渐发展为更离奇的方式，如一指禅。一指禅并非是一种武功，而是禅机。相传唐代有个和尚，叫俱胝，有人问俱胝禅师什么是道？俱胝总是竖起食指，后来成为一指禅。除此之外，唐代德山宣鉴禅师常以棒打启人心智，而临济义玄则以大喝促人顿悟，称为德山棒和临济喝，成语“当头棒喝”就是由此而来。

下面，本文就用信息通信的架构来解读禅宗的“拈花微笑”、谈玄与禅机，为这种交流模式给出一个数理解释。

1948 年，香农 (Shannon) 创立了信息论，并提出来一个通讯的简单架构，见图 51 (上)。理论，这个通讯架构包括一个发送者 A 和一个接收者 B。发送者有一个关于某个空间变量状态的消息，比如某个股票要大涨，或者有人病了，用一个码本编码，通过一个噪声信道传给 B，B 解码而得到这个消息。这里面的一个基本假设前提是：A，B 共享一个码本，否则就是乱码；并都知道某个空间的结构，否则就对发来的消息莫名其妙。香农的这个通讯架构与图 13 的基于心智模型的人类认知架构就相形见绌、太过于简化了。在现实生活中，两个智能体 A，B 之间的通讯，需要考虑很多预设条件，比如，发送者 A 必须明白：(1) B 是否已经知道这个消息；(2) B 收到这个信息是否能采取有效行动，比如是否有钱去买这个股票；(3) 根据 B 的价值观，B 需要知道或者愿意相信这个消息；(4) 如果 B 足够聪明，A 只需要给一个提示，B 就该猜到背后的消息。



扫描全能王 创建

正因如此，计算机按照香农通讯架构的速度在每秒 1 亿到 10 亿字节，而人与人之间的通讯每秒仅仅 1 个字节，与计算机传输速度相比，人说话的速度是极其缓慢的。但，奇怪的是这么窄的通讯带宽也够用了。越是聪明的人、越是彼此熟悉的人，就约有默契，就越不需要多说话。所谓“心领神会”、“拈花微笑”讲的就是这个境界。因此，禅宗机锋可以看作是基于心智模型的高阶通讯模式。作者 2023 年发表了一篇关于通讯式学习的综述文章[30]，就在论述这件事，这里不展开讨论。

引入心智状态与动机后，我们就可以超越香农极限的协议的极限。比如，A 对 B 说，我喜欢喝“热水”，“热水”在语义上与“烫”相近，但它的意思是“不烫”；否则人们会直接说“烫”，如果 B 对信息不进行二次推断，那么“我喜欢喝热水”获得的信息将包括“我可能喜欢喝烫水”。使用语用学进行推断，B 会自动排除烫水，从而突破香农极限。

主观唯心的心智模型与人类的认知架构对开发人工智能有巨大的启示。人类不同于当前人工智能大模型的一点在于：人类能够实现“小数据、大任务”，而大模型的模式则是“大数据、小任务”。按照作者在其他论文的观点，人类对世



扫描全能王 创建

界理解的 5% 是靠客观的感知，95%是主观的内心需求与想象，也就是一颗主观的“心”。在《人类简史》中，赫拉利提出人类得以存续的基础是由想象构建的体系，如货币、公司等等，后者都是看不见、摸不着的概念，是其他动物无法理解的。

4.10 汉字造字：具身智能、心智与会意的妙用

最后，我们来谈一谈中国思想的载体—汉字。

语言学家认为，语言塑造了人的思维，中国思想是描述式的，而这种描述性首先就表现在汉字上。作为世界上仅存的象形表意文字，汉字同表音文字有很多不同。

表音文字更接近形式语言，从而催生了逻辑学的发达，并以此为基础建立了自然科学的庞大体系。逻辑推理讲求刨根问底、争辩是非、非此即彼，这种思维推动了科学的发展，但文化上具有很强的排它性。

表意文字描述自然与社会的生态之现状，崇尚和谐共生、各美其美、美美与共，因而体现出很强的包容性。这种思维催生了伦理学的发展，并以此为基础建立了社会伦理体系，并进一步内求价值体系。

《易经》的六十四卦本来是很好的数学符号，是一种形式化的数学语言。前文 4.5 节已经分析过，《易经》是一个统计决策模型，其上赋予了价值函数，但这里面缺乏因果关联与逻辑推断。《易经》中的每一个卦，其实代表了一种时空中的情境（situation），而每一个汉字也是代表了一种情境，两者是高度一致的，所以，每个卦也用一个汉字表示。



4.10.1. 象形文字与具身智能

汉字的六种造字方法，最直接的是象形字，这就是线画（sketch），这个直接对应了计算机视觉的一些图像表达的模型。线条分成两类：一类是表达物体的基元（primitive, texton）；一类是表示物体的纹理（texture），如水波、鱼鳞、鸟的羽毛，这些繁杂的线条不必都描绘出来，而是采用写意的手法，以三代表多。

2014 年，作者团队结合统计的基元与纹理模型，可以从自然图像中，聚类出重复出现的模式（pattern），得到类似的象形文字，见图 52。

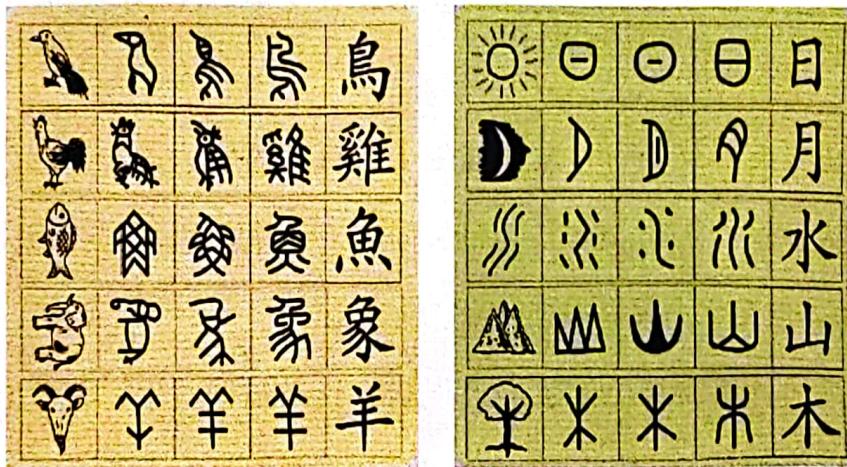


图 52：象形字就是线画，包括图像的基元与纹理两部分。

汉字中还有一大批描述人的身体部件参与的情境，如图 53 所示。第一个字，两只手，一根绳子，在拖地上一个东西。第二个在盆中洗手。第三是关门。第四是援助的援字，一只手把另外一个人的手往上拉。第五也是两个手，一个手朝下一个手朝上，表示上面的手给东西，下面的手接受。第六是争夺的争，两个手往相反的方向拉一根绳子。第七字是两个人在聊天，把耳朵单独标出来，属于指事。这些字表示了的身体与物体、场景的交互、人与人之间的交互。这就是所谓的具身智能（Embodied AI）的模型。



扫描全能王 创建

作者团队 2014-2017 年从视频中也可学习得到类似的模型，见图 54（下），分别表示：（左）人体与物体的交互结构（Human-Object Interactions）[32]；（中）手与物体的交互、工具使用（hand-object interactions）[33]；（右）人与人的交互（human-human interactions）[34]。

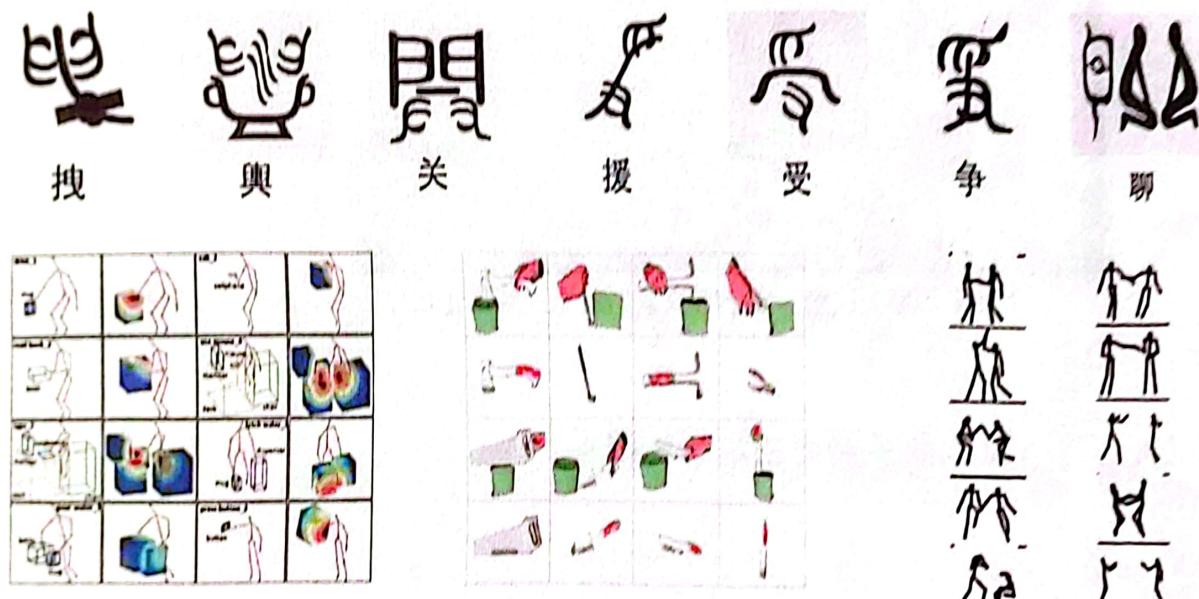


图 54：象形文字。计算机也能自动学出动词的表达，包括：坐、玩手机、握手、人拉人等等。
见[32-34]。

在英文的单词中，各种基本的动作：get, do, sit, sleep, walk, run, lift, hug, chat 等都没有体现身体单元与具身交互的成分，而且构词的组成成分并没有直接的关联。可以说，汉字的这个形象的表达方法更加直观，也与人工智能的语义模型更接近。



扫描全能王 创建

台湾的廖文豪先生出版了一个丛书《汉字树》[35]，讲汉字的联系和演化，整理成网状的图，每个汉字就是一个情境。图 55 是一个与眼睛有关的汉字的例子。

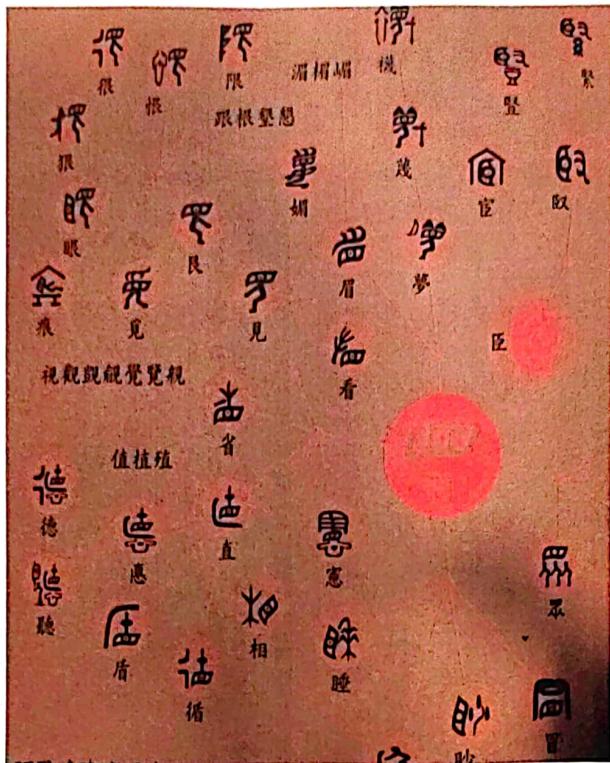


图 55 汉字关联图，选自廖文豪先生《汉字树》[35].

4.10.2. 相由心生，贝叶斯学派与主观统计建模

象形字、形声字是对某个具体情境的视觉、听觉信号的直接描述。汉字中还有很多字，其含义是需要人的理解和想象的。比如，前述佛教的核心思想之一“相由心生”，这个相字的造字就很巧妙，也蕴含着深刻的学术思想。“相”字本身就是相由心生的一个直观表述，相字左边是木，右边是目，就是指一个眼睛不好的人依靠木头摸索着行动。也就是说，一个盲人眼睛看不到这个世界，而是靠木棍对周围环境做非常稀疏的采样，通过几个采样点而重构这个世界的场景，盲人脑袋里面重建的这个就是“相”。这个“相”，主要凭借的是盲人脑袋里的知识、经验、价值体现来构建的，数据只是极少的一部分。在人工智能的

智能体 (agent) 与认知心理学中，这叫做 belief，中文应该翻译为“相”或者“信相”。

相字后来又延伸出宰相、丞相。帝王坐拥万里江山，但绝大部分时间都生活在皇宫内，对外部世界发生的事情无法直接观察，而是从大臣的报告中，去重构外部世界的情境。帝王就是那个盲人，而大臣就是那个木棍。丞相就是那个首席大臣，他常常能决定帝王看到什么，形成什么样的“相”。

这个简单的相字，就体现了中国古代思想对于人的思维中的主观成分有了深刻认识。人所形成的认知也就是各种“相”的总集，主观的成分往往远远大于客观数据所带来的信息，人的智能基本是靠“心”来推测的。

在人工智能的发展史上，1980 年代后期，研究者们发现纯粹的逻辑推理（产生式）已经无法解决很多落地问题，到了 1990 年代，统计学派开始占了上风，这个学术的中心就发生在波士顿（哈佛大学、MIT、布朗大学）。当时有一个巨大的学术争议：是否容许使用先验知识？

(1) 反对派是统计学的传统主流。他们主张的手段是采用假设测试 (Hypothesis Test)。这在今天的实验科学，如生物、医学、心理学等研究中仍然流行。你们先设立一个假设结论 H_1 作为靶子，然后找一个现有的假设 H_0 作为对比。通过一组实验数据 $\{d_1, d_2, \dots, d_k\}$ 来对比两个假设成立的概率比值 (likelihood ratio)

$$\log \frac{p(d_1, d_2, \dots, d_k | H_1)}{p(d_1, d_2, \dots, d_k | H_0)}$$

然后，结论就是在多大的置信区间内，可以得出支持 H_1 的结论，比如某个药物有效。在这个过程中，只容许使用实验数据，不许实验之外的“先验模型” (prior model)，后者被看成是干扰与偏见。统计学的教科书的第一册都是



扫描全能王 创建

讲这些内容的，什么 T-test 之类。当然，大家也意识到，这个先验模型可能也是通过之前的大量数据训练获取的。

(2) 赞成派是所谓的贝叶斯学派。贝叶斯 (Bayes, 1702-1761) 是英国的一个神父，他发表的一篇文章里面包括了一个简单的概率公式，被后人翻出来了，说成是具有重要意义。

顺便说一句，美国作为世界科技的领袖，但其历史很短，而英文又是在科学文献中占据主流的语言，所以很多思想都到英国的文献中寻找根源，加上英国对于自己的人物都大肆宣传，把自己的名人都印在英镑钞票上。所以，很多英国人，包括贝叶斯、图灵、霍金都被抬上了科学殿堂的神坛，导致教科书上很多英国人的名字。其实，中国早就有了贝叶斯的这个思想。1919 年之前，中国不知道外面的科学世界，1919 之后，中国传统又被批判与否定了，这导致中国很多的优秀传统思想在国内、国际都缺乏应有的引用，更别提影响力。

1990 年代的这个学派，包括作者在内，就抛开假设测试的研究范式，而开始统计建模 (statistical modeling)。比如，我们观察到一张图片 I (这是数据)，我们希望推断这个图像中表达的世界 W (物体、人物、场景、活动等等)，而计算机视觉的问题就被看作是从图像在计算做有可能的世界的描述 (也就是“相”)。

$$W^* = \operatorname{argmax} p(W | I) = \operatorname{argmax} p(I | W)p(W)$$

$p(W | I)$ 是所谓后验概率，直白说，就是看到图像之后产生的“相”，而 $p(W)$ 是先验概率，也就是，在未看到图像之前，我们人脑中对于外部世界 W 就有了大量先验的知识。我们大脑能够做梦就是一个明证。这个后验到先验的概率转换中用到了简单的贝叶斯公式。反对派认为，先验概率 $p(W)$ 不是当前数据 I 中所提供的，因此是个人的主观偏见，因此就影响了决策的公正性。事实上，我们必



扫描全能王 创建

须承认人的感知、认知都是主观的、带有偏见的，这是基本的工作原理。如果没有这种先验的模型，人就看不懂图像中的世界！

作者和导师在 1997 年发表的论文中 [36]，首次标题上提到了“prior learning”——从自然图像中学习（训练）先验模型，并发现自然图像最大信息的特征就是阴阳。这个学派的工作后来成为了人工智能研究的主流，包括最近出现的各种所谓“预训练”模型、大模型，都是先从大量数据中学习、拟合先验模型。这些模型的一个副作用就是容易产生所谓“幻觉”。有趣的是，“相由心生”已经说明白了，每个人自己的“相”也大部分是幻觉，大模型出现的“幻觉”只是它把价值观不同的其他人的“相”，传给了你！所以，你觉得它在“胡说八道”。

其实，中国思想早在魏晋时期的画家就提出了类似的观点：画作，特别是中国的写意画，不同于照相机的照片，是客观观察与画家主观心态的融合。画家根据当时的心境，对场景 W 中的物体、事件、环境、光照进行了取舍与加工。而任何图片、画作又经过观察者的感知、认知加工。

4.10.3. 抽象概念与认知架构

前面谈到的汉字还是比较具象的概念，那么对于大量抽象的概念如何表达呢。汉字仍然用一个情境的描述，而借助了人类的认知架构，进行推理。所谓“会意”必须以人的主观、认知架构（图 13）为基础，这一点与佛祖的拈花微笑是同一个原理，而汉字已经大量付诸实践了。

下面作者来讲述两个抽象的字：道德之“德”，奋斗之“奋”。



如何赋予人工智能“道德”观念，这是一个世界难题。许多年前，作者应邀参加一个研讨会，里面有研究人工智能的，也有哲学、伦理学的，大家众说纷纭，总感觉是隔靴搔痒，抓不到点子上。轮到作者讲话，就拿出了中国的字，一个字，描绘了一份情境，就把这件事说得十分透彻。

德字（图 56），左边的双人旁不是两个人，而是代表丁字路口，假设你来到这个路口，下面一个心，象征人在心中做选择，但还没有行动；那么这个选择是否道德呢？右边最上面是结绳计数的“十”，“四”是一只眼睛在看。虽然现场没有眼睛在看着你，但是，根据图 13 的认知架构，你心中可以想象：这些人会如何看待我的这个选择。这就类似于一个想象中的陪审团出现在你的认知空间中。下面是一横，代表“直的”，如果这些人认为你的选择可以公开、不藏私，那么就是符合道德的。

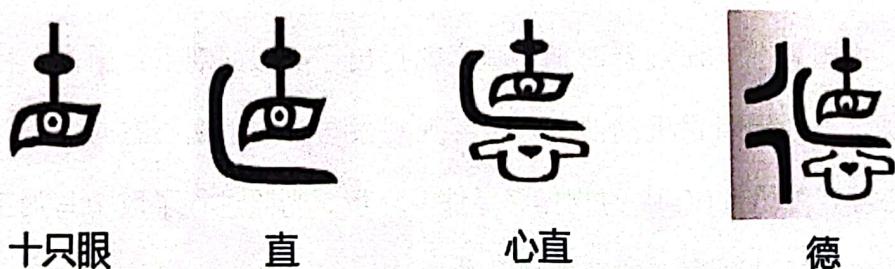


图 56：德字的甲骨文。甲骨文德字的构成由 10 只眼、一颗直心，表示道路的标志共同组成。

我们应该要为这样的中国思想感到自豪。几千年前，中国思想就已经意识到了，道德是相对的，是随着人群和时间不断演化的，所谓公道自在人心。与其去列举各种道德的准则，不如赋予人工智能一颗心：认知架构与价值体系，然后智能体自动就会去跟人群对齐。

如何向人工智能表述“奋斗”的概念呢？图 52-53 的计算机模型都是具象的动作，图 57 解读了奋斗的情境。奋字有一个“大”，代表人张开双臂，下面是一个“田”，古体还包含了一只煽动翅膀的鸟。这只鸟是情境中的主角，被



扫描全能王 创建

错误地简化掉了。这个情境是：鸟在田中吃食，人来驱赶鸟，这个时候，鸟就振翅高飞。造字者让我们去体会一下那只鸟的心情和动作，这就是“奋”的本意。



图 57：奋字的造字法。甲骨文奋字是田地里被人追赶的一只鸟的造型，繁体的奋字则是将鸟的造型转化为了隹（拼音：zhui, cui, 或者 wei），是短尾鸟总称，所以也存在于雀、隼、集、雏等字里面。在汉字简化过程中，隹被省略了。

现代神经科学在 1990 年代发现，人类大脑中有种镜像神经元（mirror neurons），其最大的特点是通过模拟其他人的行为而感同身受。这个奋字就利用了人的这些基本结构。

正如 4.9 节提到的禅宗的高阶通讯模型，我们造字是给人看的，只有人才能看懂，那么我们就需要利用人的认知架构来造字，这一点，中国思想确实是远超西方。

中国汉字通过描述各种场景能够“完备”表述生活中各种概念，这对于人工智能的研究具有重要的指导意义。因为人工智能的模型也是需要表达这些场景和概念！

本章挑选了十个例子来解读中国的五彩线思想中的几个两眼的珍珠。下一章我们讨论如何站在中国思想的根基上探索“为天地立心”的理想。



扫描全能王 创建

五、百年变局：为天地立心的探索

第二、三、四章解读了中国思想的形成、结构、特征与数理解读之后，本章以理与心的相互作用，即 U, V 两套系统的动力学，来帮助理解第一章开篇提到的百年未有之大变局，从而试图为这个大变局寻找一个科学的解答。

5.1 思想概念与关键词的总结

本节概括、梳理前面四章的基本概念与关键词，为本章的分析做准备。

(1) 认知架构。人类的一切思维活动的基础是认知架构（见图 13-14）。这个认知架构是一个抽象的数学表达，其实现的硬件系统可以是神经网络、集成电路芯片、或者未来某个量子计算装置。不论是碳基的生命还是硅基的智能体，其背后的数学理论是共同的。这是讨论的前提，我们并不必考虑具体的某个硬件结构层面的问题。不同的认知架构决定了智能体思维的上限，这是与算法、数据、硬件都无关的思想极限 (fundamental limits)。

(2) 认知空间。人类的思想以认知空间（见图 12）为思考的范围和疆域，并以语言为思考的工具，如符号语言、形式语言、或自然语言。语言可以塑造人类思维的习惯，从而形成不同的思想，着重分布在认知空间不同的区域。从而出现不同的特色，比如，前文提到，西方科学聚焦在研究人与自然的问题，中国思想解重点在讨论人与人的问题，印度宗教善于思考生命局限性的问题。

举一个对比的例子，孟子提出“老吾老以及人之老”，而同时期的亚里士多德（比孟子大 12 岁），提出的则是“我爱我师，但我更爱真理”，这说明了不同文明在认知空间中思考的侧重点不同。这种差异一直延续到今天的学术规范上，因此，也不难回答著名的李约瑟之问。



扫描全能王 创建

(3) 中国思想。中国思想是在认知空间中留下的各种情境 (situations, 卦象, 汉字)、概念 (concepts as sets, 概念就是集合) 与函数的总集。总体来说, 人类的认知空间中定义了两大类函数: 势能函数 U , 价值函数 V , 从而形成了 (U, V) -双系统 (见图 5 的演示例子)。

前文图 4 概括了从物理、化学、生物、到信息与智能的演化过程, 其本质是认知空间不断升维, (U, V) 函数越来越高级的过程, 见图 58 的演化的时间轴, 人类的智能是当前进化的顶点, 但很可能不是终点。

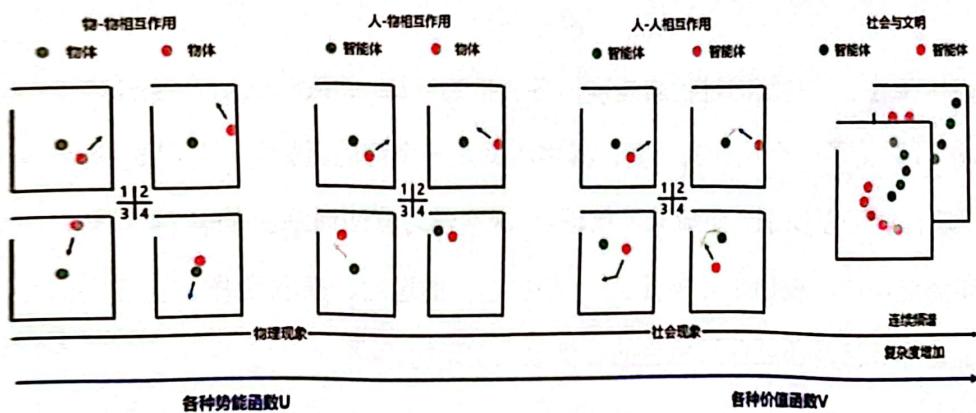


图 58: 认知空间的升级、升维与 UV-双系统的演化。图 4 概括了物理、化学、生物的相互作用、社会规范 (势能函数 U) 到智能体的自主价值体系 (价值函数 V) 的持续演进过程。

(4) U -系统。 U -系统由大量势能函数组成, 具有丰富的层次结构, 表示各种模型, 是朱熹理学所指的“理”, 包含:

- **自然之理:** 如物理的四种基本的相互作用, 化学的共价键, 生物的氢键, 牛顿物理中的各种约束, 如弹簧连接两只小球。这个 U 里面也包含了因果律、因果转换的模型 (Causal structural equations) 等。随着科学的发现找到更多的因果模型, 发明更有效的工具, 人和智能体就有更大的行动空间、更有效的功能, 这就体现为生产力的提升。
- **社会伦理:** 如儒家定义的君臣、父子、夫妻、兄弟、妯娌之间相处的规范, 可以以势能函数的形式表达。这个 U 系统包含了社会组织机构与生产关系。



值得强调的是，这些“理”都是人类思维的产物、是主观的。虽然教科书上有各种物理、化学的定理，但这些都仅仅是科学家提出来的、近似的、有用的模型而已。人类自然科学与社会科学的进步就是不断寻找更简洁、更准确、更普适的模型，逐步逼近我们相信客观存在的“道”，通过排除尽可能的因素而找到真实的“因果”关系。如果某个人群通过通讯交流而达成了某个共识，也仅仅是这个人群共享的“相”，比如，西方很多人相信上帝存在，也有很多人不信人类活动是造成气候变暖的主因。这些“相”，“模型”，“理”都不能被看作是“真理”，这跟人群大小无关，不能说相信的人多了就是真理。

人和智能体在运用 U 系统的过程就表现出各种能力 (capabilities, skills)。比如，图 37-38 讨论了一个统计的模型，其本质就是一个势能函数，有了这个 U 函数，人工智能算法就可以识别、生成某类纹理。深度学习算法训练出的各种模型，也都可以看作是势能函数，被用于各项任务，如机器人的运动、抓取动作、面部表情、输出文字、说话等等都是通过形形色色的 U 函数来实现的。

(5) V-系统。U-系统由大量价值函数组成，图 11 和图 40 显示了 V 函数的个体、他人、集体、国家乃至全人类的价值的层级，代表了各种价值偏好与需求，其来源在于物种进化、文化传承与个体的选择。

(6) 心的内涵。中国思想的心包含了两个非常不同的概念。一个是价值体系 (V 系统)，比如王阳明所指的良知，良心；另一个是指认知架构，比如慧能所指的菩提自信。认知架构是思维活动的基础，承载着 UV 系统的运行。

(7) U-V 平衡。每一个个体的人（智能体）或者一个群体（国家民族，社会）都可以用一套 (U, V) 系统所刻画。人的行为同时受 UV 系统驱动：

- 内在价值函数 V 的驱动，对 V 函数求导就可导出内驱力；
- 外在势能函数 U 的驱动，对 U 函数求导就可导出外驱力（压力）。



当两种驱动力在各种情况下达成了一致，我们就定义 U-V 处于平衡态，也就是说，外界要求个体做的与自己内心想做的总是一致的。此时人就处于一种“自在”的状态，也就是孔夫子说的：七十岁到达了“从心所欲不逾矩”的境界。

$$\frac{\partial V(z)}{\partial z} = -\frac{\partial U(x, x_-)}{\partial x}$$

这个方程粗略就写成这样：

作者在《三读赤壁赋，兼谈心与理的平衡》一文中，就详细论述了人在达成这个平衡的过程，并分析了苏轼在坎坷的人生中不但调整自己的 U, V，在不同境遇中，重新达成平衡。苏东坡夜游赤壁的思考，就是一个高级知识分子、士大夫寻求 U-V 平衡态的思想过程与困惑。

图 41 和图 42 显示了不同的人群，处于不同的平衡态，有点的低维度的平衡，如僧侣通过出家来斩断社会关系，形成小 U，通过戒律对价值格局降维，形成小 V，这样在极低的维度空间达成 UV 平衡。有些国家，如泰国、尼泊尔，大多数人处于低维度的 UV 平衡态，今天社会所说的“躺平”也就是通过降维来实现心理平衡。

UV-两套系统有着非常复杂的动力学关系。个体成长、社会进步、人类演化本质上是 UV 不断升维的过程，不断打破平衡，建立新的平衡的过程。

一个社会的 UV 平衡通常在以下几个情形被打破：

- U 的改变。科技创新提升了人的能力和生产效能，如新的工具的发明，所以说科技是第一推动力；社会组织模式的改变，如机构改革。
- V 的改变。利益分配关系的改变，如西方文艺复兴时期的认知与价值体系的改变。V 的改变往往伴随着 U 的改变，密不可分。
- 外来先进或者落后文明的占领与冲击。



5.2 UV 失衡：解读百年未有之大变局

以这一套 UV 理论的视角，我们回到本文开篇提到的百年未有之大变局（图 1），其实包含了两个层次的 UV 系统的冲突：

- (1) 西方文明体 (U_1, V_1) 对中国文明体 (U_2, V_2) 的碰撞与融合。
- (2) 随着大量具有自主意识的通用智能体 (U_3, V_3) 的到来，人类社会正在跨入智能的时代，未来如何实现人类文明和通用智能体共生共存。

5.2.1. UV 失衡之一：新文化运动百年之后的反思

1840 年之前，从汉代到清代，中国社会经过两千年的演化，尽管有周期性的王朝更迭，这个社会系统总体上有很强的韧性和稳定性。究其根本原因，作者认为，在这两千年的的时间里，

- (1) 中国的科学水平，社会模式，和价值体系，也就是 UV 系统，都没有发生大的变化与跃升。
- (2) 认知能力也没有质的变化，这体现在，中国思想的传承，虽然有宋代和明代的理学与心学的发展，不同时期的思想都是同一个来源，如四书五经，后代的思想领先总是不断回到圣人的言论中溯源、翻新。

这种 U 系统和 V 系统的相对稳定性，使得中国社会两千年来绝大部分时间处于 UV 平衡的状态，从而维系了农业时代的长期稳定发展与文化繁荣，成就了中国作为一个长期统一的国家，和唯一以国家形式传承的文明体。

直到西方文明体系 (U_1, V_1) 与中国 (U_2, V_2)，如图 59 所示，经历三个阶段，打破了中国的 2000 年来的 UV 平衡态，而急剧改造了中国的 UV 体系：

- (1) 1840-1860 年两次鸦片战争的失利，中国开始否定经济与工业体系，1861 年开启了洋务运动，引入了西方的军事装备、工业机器；



(2) 1894 年甲午海战失利，中国开始否则政治体制与教育体制，出现了戊戌变法、辛亥革命，举办京师大学堂，入教育制度；

(3) 1919 年巴黎和会失利，中国开始否定自己的文化。当时鲁迅提出“不读中国书”、钱玄同主张“废除汉字”，代表了当时社会精英的诉求。最后，废除了传统的子学与经学，全面按照西方的学科体系来布局。

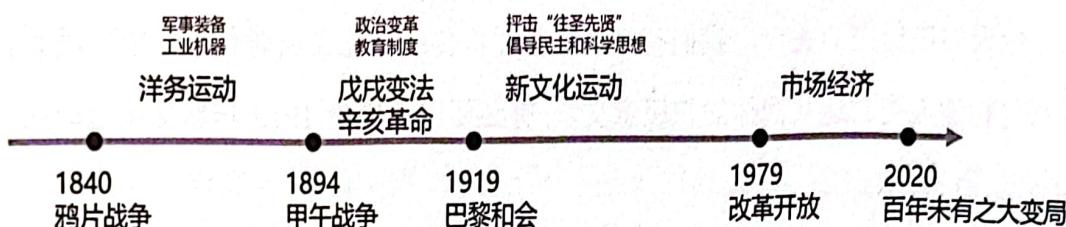


图 59：中华近代历史中东西方文明的碰撞与互鉴。

一直到近年来，还有不少知识精英希望彻底革新中国的 U 系统，甚至全盘接受照搬西方的 UV 系统。

作者发现，近代以来的三重否定改变的主要还是中国的 U 系统，而中国的 V 系统基本保留了下来，中国没有接受随着坚船利炮带来的基督教的价值体系，同时中国保留了推己及人的认知架构和以此基础上构建的汉字体系。

这就回到作者开篇 1.1 节末尾的观点：中国传统价值体系，中国思想是根植于人类认知架构基础上、由“心”驱动的、推己及人，在几千年的世俗社会中演化而来的智慧结晶；中国的价值观和伦理规范是基于价值判断的复杂决策体系，更接近于人类文明演化的终极社会伦理。

近代历史中，一些知识精英过度迷信西方思想和制度，将其视为解决中国问题的唯一途径，忽视了中国的价值体系的优势。他们过分强调 U 系统的改革，而忽视了 V 系统价值的主导性。



新文化运动 100 年之后，到 2018 年重新开启的贸易战，我们又到了百年未有之大变局。其本质是中国的价值体系（心学与良知）和信仰体系（因果报应与公平正义）与欧美基于基督教的信仰体系和延伸出的价值体现（人权、自由、民主）的选择与竞争。

今天站在历史的关口，我们有必要重新评价传统文化在新时代推动中华民族伟大复兴与构建人类命运共同体的积极意义，有必要重新检验 1919 年新文化运动的一些矫枉过正的做法。

5.2.1. UV 失衡之二：探讨现代社会矛盾与冲突的根源

当个体 UV 系统处于平衡，此人就处于自在与平和的状态；当文明体 UV 系统处于平衡态，此社会就处于稳定与和谐的状态。当今世界正处于一个急剧变化、人们内心普遍焦虑的时期，究其原因，作者认为，是由于自 1990 年代以来，全球化的进程与科技高速发展所造成的、大范围的、从个体到文明体层面的 UV 失衡。互联网与社交网络快速增大了人与人的连接，放大了个体的影响力；全球化的进程大大促进了贸易、技术、思想与人员的交流。这些发展与交流都极大地打破了过去在相对封闭的地域环境下形成的个体和社会的、处于不同维度下、定位于思想空间不同区域的各种局部 UV 平衡态。

本节讨论 UV 失衡的几个主要的例子。

一、美国的 UV 失衡问题及其影响。自 1990 年冷战结束，美国成为单一超级大国，取得了全球事务的主导权。恰巧在这个时候，互联网技术成熟，在全球化浪潮中飞速发展，投射了美国在经济、科技与文化的全球影响力，近年来人工智能与自动化技术的快速发展，加剧了这一进程。美国作为一个文明体，其能力 U 系统达到了极大的提升，而其价值体系的 V 系统在过去的几十年中并没有大的改变，这就造成了其 UV 系统的巨大失衡。



美国的保守派民众仍然是信奉两千年前的基督教义与价值观，基本没有随着时代而变化，特朗普所代表的传统农业和铁锈带选民，其价值体系仍然停留在 30 年前，以“美国第一优先”的诉求（V）与其全球领袖的影响力（U）是非常不匹配的。当这个价值群体通过选举取得了权力，以这种落后的 V 来操纵先进的 U（科技、军事与金融），其他国家感受到的就是霸权，而非王道。

二、个体 UV 失衡与“后真相”时代。现代科技的发展也放大了个体的能力与影响力（U 系统）。个体的人或者企业可以在越来越短的时间内创造百亿的市值；社交空间中，普通的个体可以在极端的时间成为网红与意见领袖，其言论可以触达、影响百万、千万甚至过亿的人群。而这些个体人的价值体现（V 系统）并没有提升，与其突然被放大的势能（U 系统）是极其不匹配的。这是一些典型的 UV 失衡，用传统中国文化来描述，叫做“德不配位”。虽然这种现象历史上也有发生，比如小孩被推上皇位。但如此大规模的发生这种现象，在历史上不曾出现过。

前文提到过相由心生和人的智能的主观唯心特征，在当今时代，被放大了。个体或者人群所自认为的“真相”，“模型（理）”其实是各自的主观认识，也不因为相信的人多了就是“真相”、“真理”。影响这个“相”的形成的关键因素是人心：认知架构与价值体系。

- 认知架构决定个体能不能看到全貌和深层的情境与因果；
- 价值体系决定个体能不能接受不符合自身利益的“相”。

在自然科学领域，如科学家发现数学定理，宇宙深处的奥秘、原子内部的粒子结构，这都是全人类的知识，无伤个体利益，大家能够接受。但随着科学研究进入研究人性和社会，就开始触犯到个体或者群体的利益了。学术上的“伦理审查”与“政治正确”就是 V 在影响着科学领域 U 的进化



在社会领域，包括历史书写、新闻编辑，都是符合某种价值体系（V）的表达。过去能够控制这些“真相”的表述的是少数人或者政权，这使得社会中人群的思想比较一致。在今天这个人人都可以播报新闻的时代，各自“相”就在信息空间泛滥了。网民选择符合自己价值的“相”，推动代表自己群体价值的主播代表，作为其代言人。这在表面上造成了思想的混乱与困惑。麦克唐纳在他的书《后真相时代》中引用了著名心理学家威廉詹姆斯的名言，“没有比被倾听者误解的真相更糟糕的谎言了”。

其实，这个问题早就一直存在，西方的议会制度，比如一个众议员就代表某个区域的利益诉求，国会的争辩常常水火不容，常年都是混乱不堪的局面。现在，这个议会的形式搬到了互联网的社交空间上，而且缺乏议事规则、缺乏沟通机制。

让问题变得更复杂的，是UV函数的进展。反映人类能力的U函数的进展快马加鞭、一日千里；但反映人性的V函数却安步当车、缓慢前行。结果，一边是高速列车、智能手机、互联网、人工智能等，另一边则是宗教禁忌、民族习惯、历史观念。这种UV失衡不断加剧，使得各自社会矛盾、文明冲突在全球上演，需要继续提升人类的认知能力和价值体系，才有可能达成新的平衡。

三、通用智能体的UV系统构建与对人类社会的潜在影响。 随着通用人工智能的进步，未来会有大量智能体（数字人、机器人）的出现，这就是图1中所示的百年未有之大变局中的第二重冲击。这个冲击包括两个问题。

- (1) 通用人工智能技术会进一步放大个体的能力与势能(U系统)，加剧个体(人、资本、公司)的UV失衡问题。
- (2) 如何给通用智能体赋予UV系统，赋予怎么样的价值体系？这是解决人工智能安全的关键。本文不再讨论这个问题，在《为机器立心》一文中有论述。



5.3 打造大型社会模拟器：求解未来人类社会 UV 平衡态

上一节用 UV 平衡与失衡的理论解读了百年未有之大变局的底层数理架构，本节探讨如何求解未来人类社会的 UV 平衡。

人类认知架构与 UV 系统的数理解构，让我们能够更好地理解人的智能，包括感知、认知、决策、价值等，所产生的机理，以及人群中不同 UV 的差异与分布。有了这些基本的模型，我们就可以研究 UV 系统的互动机制，从而研究人类社会演化的动力学方程式。理解它们的 U, V 函数空间是如何不断扩大的，它们进化的极限是什么。这既是非常重要的大科学问题，也是大工程问题。

这样一来，通用人工智能的研究就从个体尺度升级到群体、社会尺度。

在作者的推动下，2022 年北京大学联合武汉市政府、东湖高新区，共建北京大学武汉人工智能研究院，并承担建设“国家智能社会治理实验综合基地”的任务。为实现社会尺度的通用人工智能，我们提出构建大型社会模拟器作为基地建设的核心任务，并于 2023 年在武汉成立了亚洲社会仿真学会。大型社会模拟器，将利用社会治理的大数据，构建全球首个社会科学领域的大科学装置。

在自然科学发展史上，大科学装置（对撞机、加速器、天文望远镜等）长期是人类探索物理世界的边界之重要手段，促进了自然科学的繁荣。然而，在人文与社会科学领域，长期缺乏用来分析社会运行、开展社会实验、观测干预结果、完善社会治理的科学装置。因为不具有重复性，文科（文史哲）与社科（政经法）的理论都很难用实验来验证，人们只能通过观察进行回顾分析，无法开展实验，无法干预并进行反事实推理，因而也就很难推断因果关系（见 4.8.4 小节）。这个痛点就导致了黑格尔的一句名言：



扫描全能王 创建

“人类从历史学到的唯一的教训，就是人类没有从历史中吸取任何教训。（The mankind learns the exclusive lesson that arrive from the history, it is the mankind did not draw any lessons from inside the history.）”。

这其实反映了人们对历史因果的不确定和某种程度的历史虚无主义。中国历史上，七雄争霸秦得胜，三国归晋庆太平，就有很多不同的解读。一方面，历史被认为是某种天命，所以才有了邹衍的“五德终始”学说，另一方面，历史常常被认为是偶然与巧合共同早就的，所以也有“冲冠一怒为红颜”的说法。

大型社会模拟器的建立为回答这些重大问题提供了全新的技术手段，也是社会尺度的通用人工智能带来的机遇。通过大型社会模拟器的工作、运行，我们可以回答以下问题：

解读反演历史的事件，拟合各文明体的 UV 系统，找到的推动人类文明演化的动力学方程式；

提升当今社会治理效能，前瞻性探索国家社会治理模式与经验，助力中国式现代化建设；

探索人类社会如何达成更广泛的 UV 平衡，回答百年未有只大变局的理论问题。

大型社会模拟器通过构建 UV 动力学模型，模拟个体与社会的相互作用，可以探索各种社会组织与机构的形成，社会分工与经济社会模式的演化，能够为人文社科、跨学科研究提供开放的验证实验平台，为政府提供治理赋能，提高社会治理的科学性、系统性与精确性。

下面，本文就举两个我们近期做的社会模拟的例子。

人类从最早的石器时代发展到今天的信息社会和未来的智能时代，随之而来的是 UV 空间的指数级的增长。UV 函数的主体也从个体上升到群体，乃至文明体。为了研究文明演化中的问题，我们搭建了一个文明演化模拟平台（图 60）。在这个平台中，从最



初的采集资源到建立城市、开发科技、发展经济、建立政治体系等等，复现了人类社会发展的历程，从最早的狩猎采集到农业、手工业、工业、信息时代等等。在这个模拟平台，每个智能体、文明体都需要与其他智能体和文明体互动，通过贸易、联盟、战争等方式来影响其他文明的发展。同时，每个智能体都需要面对各种挑战和危机，例如自然灾害、资源短缺、外敌入侵等等。

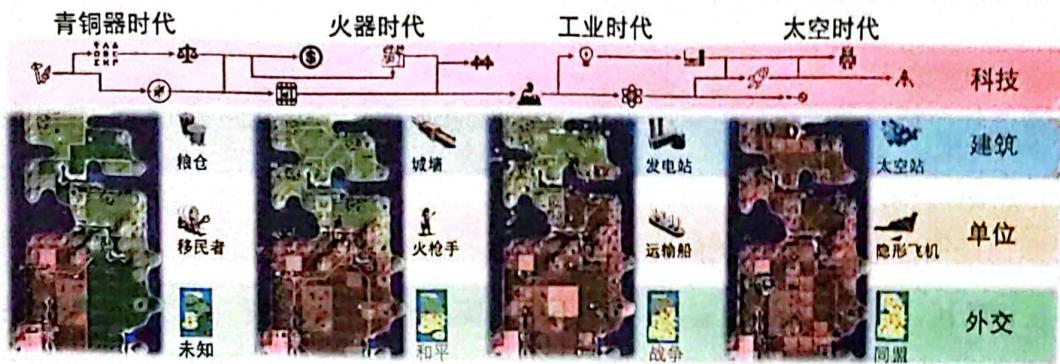


图 60. 大型社会模拟器模拟人类文明的演化历程。

图 60 显示了一个宏观尺度的文明演化的结果。其中重点勾勒出了四个时代中，两个国家的发展以及它们之间的外交关系的变化。在模拟过程中，状态可以从 10^{15} 增长到 10^{650} ，行动空间可以从 10^4 扩展到 10^{166} 。这个图中只展示了一些示例元素；这个模拟平台包括 87 种技术类型、68 种建筑类型、52 种单位类型、6 种政府类型和 5 种外交状态。

回顾人类的发展历程，人类从石器时代到学会用火耗费了数十万年，但从蒸汽机的发明到数字计算机的诞生仅仅用了 247 年的时间。这是因为我们的工具和行动空间呈指数级增长，我们需要在庞大的可能性范围内寻找有意义的发展轨迹，这在概率上看来似乎是不可能完成的任务。而人类社会不仅得以生存，而且以越来越快的速度发展。搭建这样的一个平台对我们开发出社会治理决策算法有着至关重要的意义。

通过多智能体模拟、社会模拟，我们可以揭示、解释东方文明与西方文明之间的统一性与差异性。使用全球通用的数学逻辑、科学模型，讲好中国故事。自轴心时代之



后，东西方文明模式出现了分野。古代西方文明（以欧洲为主）走向了分裂，而古代中国却延续了大一统的文明模式。

通过社会模拟器，可以探索中西方文明在历史长河中的演化路径的差异性和影响因素。在验证已有历史假设的同时，确定影响文明演化的关键因素。对于古代中国而言，东周是国家形态发生变化的关键历史阶段，而春秋又是东周的第一阶段，更为关键。在该阶段，中央政权走向衰落，战争频发，国家形态发生巨变。通过建立多智能体模型，模拟春秋时期诸国在战争、联盟等方面自组织演化和宏观涌现，可以解释复杂的历史现象和系统规律，探索平行历史文明的发展可能，以寻求文明统一的关键因素与条件。图 61 是中国在春秋（轴心时代）的演化历史的两个时间切面。图 62 显示东西方文明模态差异化的深层次逻辑的初步模拟与探索 [37, 38]。



扫描全能王 创建

通过构建人类文明演化动力学模型，可以考察不同的地理和气候环境下演化出多种可能的文明模态，比如东方文明（Unity）、西方文明（Disunity）。我们将动荡时期的国家发展动力学，视为一个视为包含多个智能体（诸侯）的复杂系统演化过程，各国之间的战争、结盟等行为被抽象化为微观层面的主体元素交互，而文明模式变化被概念化为历史系统的宏观涌现。通过社会模拟器，重点对诸侯级智能体开展模拟工作，涉及战争机制、联盟机制、胜负机制、联盟韧性等方面的机制设计。模拟结果表明，东西方文明具有同源性，都是人类文明演化动力学模型涌现出来的文明模态。社会模拟还可以科学揭示出东西方文明的差异性，本质上是核心行为参数的取值水平的差异。通过社会模拟，我们可以找到精准阈值范围，科学解释东西方文明模态差异化的深层次逻辑。例如，在文明演化系统中，如果诸侯级智能体的平均战争倾向高于 0.3，系统就会不可避免地走向统一，涌现大一统文明特征。如果低于 0.3，则会涌现西方文明特征（非一统）。



5.4 为天地立心：积极探索人类文明新模态、新道路

在人类正在经历百年未有之大变局的重要历史关头，党的二十大报告提出促进世界和平与发展，推动构建人类命运共同体的主张：

“中国提出了全球发展倡议、全球安全倡议，愿同国际社会一道努力落实。我们真诚呼吁，世界各国弘扬和平、发展、公平、正义、民主、自由的全人类共同价值，促进各国人民相知相亲，尊重世界文明多样性，以文明交流超越文明隔阂、文明互鉴超越文明冲突、文明共存超越文明优越，共同应对各种全球性挑战。”

为完成此宏图，我们需要找到人类文明演化的科学规律、建立科学的模型。

本文通过分析单个智能体的认知架构，寻找 UV 系统的复杂相互作用，试图寻找社会尺度的通用人工智能模型，以及人类文明演化的动力学。

在全人类、全球的宏大角度，寻找和平方案、找到冲突和解方案、比选各种道路选择的智能模拟功能。通过全局性计算、动力学模拟，可以获得全球所有国家的最优发展策略集合，以及最需要避免的策略集合。并且，能够计算冲突避免的概率、方案、策略，以科学精准的模型，助力人类命运共同体建设。

当今世界，人类文明处在合作或是对抗、团结还是分裂的十字路口。随着全球化的推进，我们看到，不同种族之间、不同国家之间、不同文明，处于 UV 空间的不同区域，还有大量 UV 失衡的状态。由于经济、军事、文化以及更本质上的价值观的矛盾，导致难以建立持久的互信与合作，在一些领域冲突也更加频繁与激烈。科技进步尤其是人工智能的出现，更增加了情形的不确定性。



扫描全能王 创建



扫描全能王 创建