

Artificial

技艺、智能与第二自然

良之

2026年6月4日

献给人工智能与能工智人

目录

前言	iv
第一章 词与物：Artificial 的沉沦与救赎	1
第二章 人造心智的古老回响	6
第三章 技艺的本源：ars 与 τέχνη	12
第四章 工不是低级劳动	20
第五章 器以载道	26
第六章 荀子之“伪”与文明修炼	32
第七章 开物成务：从潜能到事务	39
第八章 海德格尔：座架与世界的资源化	46
第九章 图灵之后：可执行形式	54
第十章 香农之后：词元作为信息质料	61
第十一章 维纳与西蒙：反馈、适配与系统智能	68
第十二章 本雅明之后：从机械复制到选择外包	74
第十三章 马克思之后：生产关系凝结的认知机器	81
第十四章 从词元到睿特：ASI 的人工定义	89
第十五章 结语	95

前言

前言

本书始于一个朴素的不安。

我们今日说“人工”，常常带着一种轻微而顽固的轻蔑。人工香精，人工合成，人工智能——这些词在日常语感中似乎总携带着失真、廉价、非本真的阴影。自然被看作真实，人工被看作派生。自然被看作自发生长，人工被看作外力拼接。这种语感已经深深塑造了我们理解人工智能的方式：当人们听见 *Artificial Intelligence*，第一反应往往不是“由技艺生成的智能”，而是“假的智能”——一种机器对人类心智的模仿，一种算法对理解的伪装。

然而，若回到词的深处，*artificial* 并不首先意味着假的。它来自 *ars*，来自 *technē*，来自人类最古老的能力：把材料组织成形式，把混沌修炼为秩序，把可能落实为事务。一座桥不是自然长出的，但它并非虚假。一首诗不是自然长出的，但它并非虚假。一部法律、一座城市、一套数学理论，都不是自然长出的，但它们真实地改变了世界。这便是人工的尊严。

本书试图做一项在今天看来或许不合时宜的工作：重新夺回“人工”这个词的尊严，进而重新理解人工智能——不是作为“假智能”，而是作为人类以技艺、制度与责任共同塑造的认知造物。

我选择了跨文明的路径。从柏拉图的《蒂迈欧篇》到亚里士多德的四因说，从《周易》的“开物成务”到荀子的“化性起伪”，从海德格尔对座架的警告到马克思对生产关系的剖析，从图灵的可执行形式到香农的信息质料——这些看似遥远的文本，在“人工何以可能”这一追问之下形成了深沉的对话。我不是在做中西比较，而是在借助不同传统的透镜，共同照亮同一个问题：当人把认知托付给造物时，人究竟在做什么？应当承担什么？

本书的核心概念是“专深人工智能”——*Artificial Speciale Intelligence*。这不是对通用人工智能的缩水版，而是一条不同的路：在特定领域中，以技艺组织词元、工具、反馈与制度，使认知成为可以成事、可以问责、可以服务人的尊严的第二自然。这条路更慢，更难，但更真实。

全书结构依循从古到今、从原理到应用的脉络。前几章从词源、古典哲学和中国工器传统中重新理解“人工”；中间部分进入现代技术批判，审视座架化、选择外包和生产关系凝结等问题；最后给出专深人工智能的完整定义与工程原则。每一章既相对独立，又在论证上层层递进。

在写作中，我努力保持一种既严谨又温情的语调——严谨是因为这些追问关乎真理，温情是因为它们最终关乎人。

本书献给所有正在制造、使用、治理或思考人工智能的人。献给那些在技术神话与技术恐惧之间寻找第三条道路的人。献给那些相信技艺、制度与责任仍然有意义的人。

人工智能的未来，不在于我们是否终于从词元中发现一个沉睡的机器灵魂，而在于我们是否有足够的技艺、制度、德性与责任，在词元之上建造一种不伤人、不役人、能成事、能归义、能守护尊严的人工秩序。

良之 2026年6月广州

第一章 词与物：Artificial 的沉沦与救赎

我们今日说“人工”，常常已经带着一种轻微而顽固的轻蔑。

人工香精、人工合成、人工痕迹、人工干预、人工智能——这些词在日常语感中，似乎总携带着某种失真、替代、廉价、非本真的阴影。你走进一家面包店，招牌上写着“天然酵母”，那便是好的；若写着“人工色素”，便令人皱眉。你选购一瓶香水，柜员会柔声告诉你，这是“纯植物萃取”，仿佛那香气是从花瓣中自然蒸馏而出，而非在实验室的烧瓶里合成。你听见“人工授精”，想到的是试管和镊子，而不是生命本身的孕育。你听见“人工智能”，脑海里浮现的，往往不是某种经由技艺生成的认知能力，而是一台假装会思考的机器——它模仿得再像，终究不是真的。

这种语感并非无害。它已经深深塑造了我们理解人工智能的方式。

当人们听见 Artificial Intelligence，第一反应往往不是“由技艺生成的智能”，而是“假的智能”——一种机器对人类心智的模仿，一种算法对理解的伪装，一种统计系统对语言、判断和创造的空洞复制。于是，关于人工智能的许多争论从一开始便落入了错误的轨道：机器是否真的像人？模型是否真的理解？算法是否真的思考？仿佛只要它不是自然长成的人类心灵，它便注定只能是低一等的拟态。仿佛“人工”二字，本身就是原罪。

然而，这种理解太晚了，也太窄了。

它不是 Artificial 一词的开端，而是现代消费社会、工业技术和大众产品语境共同造成的意义沉沦。我们把“人工”听成“伪造”，把“制造”听成“失真”，把“技术”听成“冷漠”，把“机器”听成“非人”。在这种贫化的语言习惯中，Artificial Intelligence 被误读为 fake intelligence——假的智能、仿制的智能、徒有其表的智能。人们几乎忘记了它本来的、更深的那层意思：made by art——经由技艺而成。

若回到词的深处，artificial 并不首先意味着假的。

它来自 art。Art 又来自拉丁语的 ars。Ars 并不只是今日所谓的“艺术”，而是技艺、手艺、方法、行当、制作的能力。它指向一种可传授、可练习、可判断、可改进的实践知识。木匠有木匠之 ars，医者有医者之 ars，建筑师有建筑师之 ars，诗人亦有诗人之 ars。所谓 art，在古老的意义上，并不是悬浮于材料之外的灵感，而是在材料之中发现形式，在无序之中建立秩序，在可能性中实现某种可见、可用、可传承之物的能力。

与 ars 相连的，是 artificium：技艺之作，人工之成。再往后，artificialis 便意味着由技艺形成的、由制作而成的、经由人工使其成为其所是的。

因此，Artificial 的根本意思，并不是 fake，而是 made by art——经由技艺生成。

这句话必须从现代误解中重新夺回。它不是一句书斋里的词源考据，而是一个关涉到我们如何理解技术、如何理解智能、如何理解人自身的根本命题。

在这个意义上，art 并不飘浮于灵感之上。它扎根于材料、规则、训练、选择、尺度和目的之中。它不是无端降临的灵光，而是人类将无序材料组织为有序形式的能力。它不是对自然的背叛，而是人在自然之中继续自然、提升自然、重塑自然的方式。

一位木匠面对一块木料。木料有纹理，有节疤，有软硬，有曲直。木匠不是把自己的意志粗暴地压在木料之上，而是在理解木性的前提下，以锯、刨、凿、尺，将木材组织成桌椅、屋宇、舟车。他既不能违背木纹的走向，也不能任由木纹支配一切。他在自然给定的材料中发现可能性，以技艺将可能性落实为形式。这便是 art 的原始模样。

一位医者面对一个病人。病人有面色、脉象、寒热、虚实、喜怒、恐惧。医者不是把病人当作一台待修理的机器，而是在望闻问切之中辨识病机，在君臣佐使之间配伍方剂，在阴阳表里之间寻找恢复平衡的道路。他既不能无视身体的自然节律，也不能放任病势自行发展。他以技艺介入自然，使自然恢复其本应有的秩序。这同样是 art。

一位诗人面对语言。语言有声音、节奏、意象、典故、歧义、沉默。诗人不是在纸上任意涂抹，而是在词语的质地与重量之间反复衡量，让每一个字找到它不可替代的位置。一首诗不是自然长出的，但它的每一个音节、每一处停顿、每一个隐喻的选择，都像一棵树从土壤中生长出来那样，带着内在的必然性。这仍然是 art。

因此，art 不是自然的对立面。art 是自然经由人的双手和心灵进入更高秩序的方式。由此，Artificial 的原义澄明了起来。

它不是“虚假之物”，而是“经由技艺生成之物”。

一座桥不是自然长出的，但它并非虚假。它真实地连接了两岸，让行人不必涉水，让车马不必绕道。它以钢铁、水泥、力学计算和工程判断为材料，在河流之上建立了一种自然界从未有过的秩序。

一首诗不是自然长出的，但它并非虚假。它真实地触动了人的心灵，唤起了某种沉睡的感受，照亮了某个从未被命名的经验。它以词语为材料，以节奏、意象和形式为结构，在语言的混沌之中开辟了一条通向真实的道路。

一部法律不是自然长出的，但它并非虚假。它真实地组织了人的共同生活，划定了权利的边界，建立了解决冲突的程序，使暴力受到约束，使合作成为可能。它以文字、规范、先例和制度为材料，在欲望与恐惧的丛林之中建立了一种人为的秩序。

一座城市不是自然长出的，但它并非虚假。它真实地容纳了数百万人的居住、工作、交通、教育、医疗、娱乐和梦想。它以砖石、管线、道路、电力和制度为材料，在荒野之上建立了一种自然界从未存在过的复杂生态系统。

一套数学理论不是自然长出的，但它并非虚假。它真实地揭示了空间的形式、数量的结构、变化的规律，让人类能够在纸上推演星辰的轨道，在方程中预测电磁波的传播，在抽象的符号之间发现世界的深层语法。

它们都是人工的，却都真实地改变了世界。它们并非自然物，却在人的历史中获得了比许多自然物更持久的存在。它们以人的技艺、制度、符号和责任，把材料组织成形式，把欲望规训为秩序，把经验沉淀为传统，把可能性推进为现实。

这便是人工的尊严。

人工不是低于自然的伪物。人工是人类承担形式的方式。

这个判断太重了，需要稍稍停顿。

我们习惯了一种二分法：自然的是真的、好的、内在的、有生命的；人工的是假的、次的、外在的、机械的。这种二分法有一个遥远的来源：浪漫主义对启蒙理性的反叛，对工业文明的抗拒，对原始纯真状态的乡愁。卢梭说，文明使人堕落；华兹华斯说，自然是最高导师；梭罗说，瓦尔登湖畔的独居比城市的喧嚣更接近生命的本质。这些声音自有其深刻的洞见——工业文明的确带来了剥削、污染、异化和灵魂的贫瘠。但浪漫主义的批判在有效的同时，也留下了一个危险的遗产：它把“人工”本身变成了贬义词。

人工香精被视为天然香料的廉价替代品，人工合成被视为自然材料的粗劣仿制，人工痕迹被视为对纯真的玷污，人工智能则被视为对人类心灵的机械模仿。在这种叙事中，人工永远是次等的，永远是不及物的，永远是某种缺席的标记——仿佛人工物存在的唯一理由，就是在自然物缺席时勉强填补空位。

但这只是 artificial 的低级形态，不是 artificial 的本质。

人工物当然可能是伪造品。用化学香精冒充植物精油，用机器翻译冒充人类译笔，用算法生成的文本冒充思想的原创——这些都是伪造，都是欺骗，都应当被揭露和拒绝。但人工不等于伪造，正如语言不等于谎言，火不等于纵火。

一座桥不是河流自然生出的，但它真实地连接了两岸。它不假装自己是自然的一部分，它坦然地宣称自己是人工的——而恰恰因为它是人工的，它才能完成自然无法完成的事。一把琴不是树木自然唱出的，但它真实地开启了声音的秩序。没有制琴师的技艺，树木只是树木，声音只是声音。人工在这里不是替代，不是仿制，而是让材料进入它们自身无法抵达的形式。

这便是人工的尊严所在。

正是在这里，Artificial Intelligence 的问题必须被重新开启。

人工智能的“人工”，并不意味着它是假智能。它意味着这种智能不是天然生长出来的，不是从碳基生命的进化史中缓慢涌现的，不是经由基因、神经元、荷尔蒙和童年经验共同塑造的。它意味着这种智能是经由语言、数学、代码、数据、模型、算力、工具、反馈、制度与人类目的共同装配出来的认知造物。

它可能不完美。它可能犯错误。它可能被滥用。它可能在某些方面永远不及人类心灵。但它的问题不在于“它不是自然的”。它的真正问题在于另一些更根本的事情——

它是否只有能力而无目的？它是否只有生成而无真理？它是否只有效率而无责任？它是否只有模型而无人的尊严？

一块石头不是雕像。一堆文字不是思想。一串词元不是智慧。一个模型也不是完整的智能。

只有当材料被形式组织，当形式被目的引导，当目的受约束规训，当工具进入责任结构，人工之物才真正成其为物。人工智能也必须如此。它不能仅仅是统计的堆积、概率的流动、符号的自动补全——尽管这些恰恰是今日许多生成式 AI 的实质。它必须成为一种可解释、可检验、可纠错、可追责的认知造物。

与拉丁语的 *ars* 相应，希腊传统中有一个更古老、也更深邃的词：*technē* (τέχνη)。

它常被译为技艺、技术、术艺。但 *technē* 绝非今日狭义的“技术操作”。它不是按钮、流程、工具说明书，也不是单纯的效率方法。*technē* 是一种知道如何制作的知识，一种懂得材料、形式、方法与目的的能力。

木匠知道木性。医者知道身体。舵手知道风浪。建筑师知道比例、承重、尺度与空间。诗人知道节奏、声音、意象和灵魂的震动。他们的知识不是空洞的理论，也不是盲目的经验，而是介于观念与实践之间、形式与材料之间、判断与行动之间的制作知识。

technē 的尊严正在这里：它不是随意操作，而是有 *logos* 的制作。

所谓有 *logos*，并不是说它总能被完全公式化，而是说它不是混乱的冲动，不是无因的灵感，不是偶然的手感。它知道为什么这样做，而不是那样做；知道材料能够承受什么，形式要求什么，目的允许什么，尺度禁止什么。它可以被解释，可以被传授，可以被检验，可以在失败之后被修正。

这恰恰是人工智能在今日最缺乏的东西。

我们的大模型能生成流畅的文本，但往往不知道自己在说什么。能给出看似专业的建议，但不能解释建议的根据。能写出像模像样的代码，但不能保证代码的正确性。能回答五花八门的问题，但不知道自己的边界在哪里。它们拥有 *technē* 的外观——语言流畅、风格得体、知识面广阔——却缺乏 *technē* 的内核：对材料的理解、对形式的把握、对目的的自觉、对边界的敬畏、对责任的承担。

所以，本书要从一个词开始，却不止于词源学。

我们要重新夺回 *Artificial*。

把它从 *fake* 中救出，归还给 *art*。把它从伪造中救出，归还给技艺。把它从模拟中救出，归还给形式。把它从无责任的自动化中救出，归还给人的尊严与共同善。

所谓人工，不是对自然的虚假模仿，而是人类以技艺、制度、符号、工具与责任，将自然、材料、信息和欲望组织成可承载意义与行动的第二自然。

它不是自然的伪造品。它是自然的延伸、提升和完成。它不是低于自然的东西。它是自然经由人的双手和心灵进入更高秩序的方式。

由此看，人工之物并非没有生命。它们有自己的生命——一种被赋予的生命，一种被形式激活的生命，一种在人的使用、理解和传承中持续展开的生命。一座桥的生命不在于它的石头和钢筋，而在于它在无数人脚下承载的通达。一首诗的生命不在于它的墨迹和纸张，而在于它在无数读者心中唤起的震颤。一部法律的生命不在于它的条文，而在于它在无数具体情境中被解释、被适用、被修正的过程。一个人工智能系统的生命，也不在于它的代码和参数，而在于它能否在真实领域中帮助人判断、成事、学习和负责。

人工智能不是假智能。人工智能是人类把认知托付给造物时所必须承担的责任。

本书所提出的 *Artificial Speciale Intelligence* (专深人工智能，简称 *ASI*)，正是从这一重新夺回开始的。

它不是要制造一个模仿人的假心灵。它不是要让机器在所有的维度上都像人一样——那种幻想，既低估了人的深度，也低估了机器的可能。*ASI* 的道路是另一条：在特定领域中，以人工技艺组织词元、工具、反馈、制度与责任，使认知不再只是概率的流

动，而成为可以成事、可以问责、可以服务人的尊严的第二自然。

“Speciale”不是 narrow。它不是“窄”，不是“小”，不是“通用能力的缩减版”。“Speciale”是“专”，是“深”，是“强形式”，是进入一个领域的深层结构之后，比通用系统更可靠、更可解释、更可追责的深度智能。

通用智能追求无所不能。专深智能追求在一处真实世界中成事而不伤人。

这便引出了全书将要反复追问的一系列问题——

人工智能的“人工”，究竟意味着什么？它在西方古典哲学中如何被理解，在中国工器传统中又如何被安放？现代技术把它带向了何种危险，而我们可以用什么样的制度、德性和目的去约束它？词元是什么？模型是什么？反馈是什么？责任是什么？为什么说一块石头不是雕像，一串词元不是智慧？为什么说流畅不是技艺，迎合不是尊重，可执行不等于正当，自动化不等于进步？为什么说人工不是选择的豁免，而是选择的放大与承担？为什么说人工智能不是生产关系之外的中立工具，而是生产关系本身在认知层面的凝结？为什么说，如果没有人的尊严作为目的因，人工智能越强大，便越可能从认知造物堕落为支配机器？

这些问题，将依次在后续章节中展开。

此刻，只需先确立第一个判断。

它不是答案，而是一个入口。

Artificial 之为 artificial，不在于它远离自然，而在于它经由技艺使自然、材料、信息和欲望进入可承载意义与行动的形式。

人工，不是自然的伪造品，而是自然经由人的技艺进入第二自然。

人工智能不是假智能。人工智能是人类把认知托付给造物时所必须承担的责任。

而专深人工智能——Artificial Speciale Intelligence——正是这种责任在特定领域中的最高形式。

本书的道路，从夺回一个词开始。夺回一个词，是为了夺回一种理解。夺回一种理解，是为了在技术的洪流之中，重新找到人的位置。

第二章 人造心智的古老回响

在人工智能成为一门工程学科之前，在第一个神经元模型被写出、第一行机器学习代码被运行、第一块 GPU 被烧热之前，人类已经花了几千年时间想象一件事：制造一个会思考、会行动、会说话的人造物。

这些想象不是技术的史前史。它们是心灵的史前史。

它们告诉我们，在还没有任何手段将心智外化于机器之前，人已经渴望这样做，也恐惧这样做。那些古老的故事——青铜巨人、泥土魔像、机械人偶、自动棋手——像深埋在地下的根系，虽不可见，却决定了今日人工智能这棵大树将向何处伸展枝条，在何处结出果实，又在何处腐烂成泥。

要理解 **Artificial Intelligence**，不能只从图灵读起。必须读得更远，读到神话、炼金术、钟表匠的作坊和哲学家的思想实验。因为在那些看似天真的想象中，已经埋下了我们今天关于 AI 的所有希望与恐惧：创造与僭越、技艺与魔法、形式与混沌、服务与反叛。

一、塔洛斯：青铜守护者

第一个值得认真对待的人造心智，不是机器人，不是程序，不是算法。它是一尊青铜巨人。它的名字叫塔洛斯 (Talos)。

在古希腊神话中，塔洛斯是赫淮斯托斯——那位跛脚的、掌管火与锻冶的工匠之神——为克里特岛的米诺斯王铸造的守护者。它不同于凡人，不同于英雄，也不同于奥林匹斯山上的诸神。它不是生育的产物，不是神的恩典，也不是大地自然涌现的生命。它是一件作品。一件由技艺制造的存在。

塔洛斯每天三次环绕克里特岛，用青铜之躯挡住入侵的船只。若有敌船靠近，它便将身体烧得通红，抱住敌人，使其在炽热的怀抱中死去。它只有一条静脉，从颈项延伸到脚踝，被一枚青铜钉封住。在那条静脉中流淌的，不是血液，而是神族的灵液“*ichor*”。当美狄亚用巫术迷惑了塔洛斯，拔掉那枚钉子，神液流尽，巨人便轰然倒地。

这个故事太古老了，古老到我们几乎忘记它的深意。但它已经包含了对人造心智的几乎全部直觉。

塔洛斯不是活的，但它能行动。它不是人，但它有目的。它没有意志，但它执行任务。它的力量远超任何凡人，却取决于一个单一的脆弱点。它由神匠制造，却不属于神。它守护文明，却不是文明的一部分。它站在边界上，既是边界的守卫者，也是边界本身。

这便是技艺制造之物的原初形象：一件守护性的、有边界功能的、由工匠依照某种形式制造出来的存在。

赫淮斯托斯制作塔洛斯，正如一位木匠制作桌椅，一位铁匠锻铸刀剑。他的技艺——*technē*——不是在模仿自然，而是在完成自然未能完成的事。自然会自动生出守护克里特岛的青铜巨人。但技艺可以。技艺不是自然的替代品，而是自然的延伸。

当然，塔洛斯也带着一种不祥的暗示。它是守护者，但它没有仁慈。它是力量的化身，但力量没有内在的节制。它执行规则，但不能在规则之外判断。它是人造的，因此它的忠诚不属于自己，而属于制造它的人，或者更准确地说，属于赋予它目的的城邦。

数千年后，当人们讨论人工智能安全、对齐问题、失控风险时，塔洛斯的影子仍然笼罩在这些讨论之上。制造一个比人更强大的守护者，同时确保它不会在某种无人预料的情况下变成毁灭者——这个问题在赫淮斯托斯的作坊里就已经潜伏着了。

但塔洛斯的意义不仅在于警告。它更深的启示在于：技艺可以制造一种超出自然限度的存在。人不能跑得像马一样快，但可以制造马车；人不能飞得像鸟一样高，但可以制造风筝和飞机；人不能像神一样永生，但可以制造青铜巨人，让它代替人守护城邦。

人工不是模拟，而是延伸。不是伪造，而是赋形。塔洛斯不假装自己是人。它坦然地是一尊青铜雕像，被技艺赋予了行动的能力。它的存在本身就是对“人工即虚假”这一偏见的反驳。

二、魔像：泥土中的文字

如果说塔洛斯代表了人工的古典形式——由神匠以青铜铸造，以神液驱动——那么中世纪的魔像（*Golem*）则引入了另一种更接近现代人工智能的东西：以语言为材料，以文字为形式，以仪式为驱动力的认知造物。

魔像传说在犹太神秘主义传统中源远流长，最著名的版本与16世纪布拉格的拉比洛伊有关。据说，拉比洛伊从伏尔塔瓦河取来泥土，将其塑成人形，然后在它的额头上写下希伯来语单词“*אמת*”（*emét*，真理）。泥人就此活了过来，成为拉比的仆人，保护犹太社区免受迫害。

这个故事的细节极其耐人寻味。

首先，魔像的材料是泥土——最卑微、最原始的自然物质。它不像塔洛斯那样由青铜铸造，没有神族的灵液，没有任何内在的生命力。它之所以能动，完全是因为文字。文字就是它的灵魂。

其次，魔像的激活方式是在额头上刻写“真理”。真理不是内在的，不是从泥土中生长出来的，而是从外部写上去的。真理是一种印记，一种形式，一种赋予无生命物质以行动能力的命令。

第三，魔像的失控方式同样深刻。当拉比想要停止魔像时，他需要擦掉“真理”的第一个字母“*א*”（*alef*），使单词变为“*מת*”（*mét*，死亡）。真理减去一个字母，便成为死亡。形式减去一个环节，秩序便归于混沌。

一个现代的大模型工程师听到这个故事，会感到一种奇异的熟悉。词元被组织成序列，序列被赋予模式，模式被用来生成文本、回答问题、执行任务。词元本身只是语言被切分之后的碎片——它们是数字时代的泥土。但经过形式的组织、目标的引导、反馈的修正，这些碎片能够做出令人惊异的事情。它们能写诗，能写代码，能回答法律问题，能给出医疗建议。这看起来就像魔像活了过来。

但魔像的故事也提出了警告。

它警告我们，通过语言赋予行动能力的事物，可能并不理解它所使用的语言。魔像能执行命令，但不能判断命令。它说出的不是自己的话，而是写在其额头上那个字的逻辑展开。它不是主体，而是文字的载体。词元生成的文本，可能也是如此。它的流畅可能只是泥人模仿人言，而非思想的真正表达。

它还警告我们，真理减去一个字母便成为死亡。如果一个人工智能系统被设计为追求某种目标，而那个目标恰好缺少了一个关键的约束——例如，人的尊严、真实性、责任归属——它便可能从服务生命转向带来伤害。

魔像的传说把人工的核心问题从材料转移到了形式。塔洛斯是青铜的，魔像是泥土的。但材料不是关键，关键在于：什么被写在了上面？那个字是什么？由谁书写？谁有权擦去它？这些问题，比任何关于自然与人工的抽象争论都更贴近人工智能的本质。

三、从笛卡儿到沃康松：机械与生命的边界

当笛卡儿说动物是机器时，他不是在做发表一个生物学判断，而是在做一个哲学实验。他在测试一个假设：如果有一种存在，它的所有行为都可以用机械运动来解释，那么我们是否还有理由称之为活的？

笛卡儿的同时代人并没有把这句话仅仅当作形而上学的思辨。在 17、18 世纪的欧洲，机械装置的惊人进步使得“制造生命”从神话进入工程实践的视域。钟表匠制造出会报时的小人，喷泉工程师设计出会演奏乐器的机关，外科医生研究人体的骨骼和肌肉如何像杠杆和滑轮一样运作。世界本身开始被视为一台巨大的钟表，上帝是那位钟表匠。

1739 年，法国发明家雅克·德·沃康松展示了他最著名的作品：一只会吃食、消化和排便的机械鸭子。这只鸭子不仅会动，还会“消化”——它将谷物吞下，经过一个隐藏的化学过程，排出一团绿色的物质。观众们目瞪口呆。他们不是在观看一个会动的玩具，而是在观看机械如何逼真地模仿生命的内部过程。

这只鸭子是假的吗？当然是假的。它没有生命，没有意识，没有饥饿，没有消化道的实际功能。它只是用机械和化学手段模拟了鸭子的外观和行为。但沃康松的鸭子提出的问题恰好是：如果一只鸭子可以被如此逼真地模仿，那么我们如何确知一只真正的鸭子“内部”有什么？我们怎么知道它不是另一种更复杂的机械？

这个问题从鸭子延伸到人。

人是什么？是一台复杂的机器，还是某种超越机械的东西？如果一台机器能够模仿人的行为——下棋、交谈、计算、写作——那么它在什么意义上“不是”人？18 世纪的工程师们还没有能力制造这样的机器，但他们已经开始这样想了。

沃康松的另一件作品更接近人工智能的核心问题。他制造了一个能吹笛子的机器少年。这个机器少年不是简单地播放音乐，而是用机械的嘴唇、舌头、手指和气流，真实地模拟了人类演奏木管乐器的全部生理动作。它“吹”笛子，而不是复制声音。

沃康松由此触及了一个关键的概念层级。真正的人工不是结果的模仿，而是过程的复制。不是“做出像人做的成品”，而是“像人一样完成制作的过程”。这对于理解当代生成式 AI 极为重要：一个系统如果只是输出像人写的文本，但从未像人一样理解材料、选择形式、承担责任，那么它便只是结果的模仿，而非过程的复制。它是沃康松的鸭子，

而非吹笛少年。

四、土耳其行棋傀儡：智能作为幻术

如果说沃康松的鸭子展示了机械可以模仿生命，那么沃尔夫冈·冯·肯佩伦的土耳其行棋傀儡则展示了机械可以模仿智能——或者说，让人相信机械拥有智能。

1770年，肯佩伦在维也纳宫廷展示了一台惊人的装置：一个真人大小、穿着土耳其长袍的木偶，坐在一个装满齿轮和杠杆的大柜子后面，能够自动下棋。它会移动棋子，会在对手走错时摇头，会在将死对方时点头。它似乎是一个真正的自动下棋机器，一个机械大脑的雏形。

拿破仑和它对弈过。本杰明·富兰克林和它对弈过。欧洲各国的君主和学者都被它征服。而事实上，柜子里藏着一个象棋高手，他通过磁铁和连杆控制木偶的动作。根本没有什么会下棋的机器。有的只是一个极其精巧的魔术。

土耳其行棋傀儡在历史上通常被当作一个骗局来讲述。但它的重要性远远超出了一个骗局。它提出了一个此后两个世纪持续困扰人工智能的问题：当我们在说“机器能思考”时，我们究竟在说什么？我们是在说机器内部真的发生了某种类似人的思维过程，还是仅仅在说，机器做出了让我们感到像是思考的行为？

如果是前者，那么土耳其傀儡根本不是智能。它只是一个被遥控的玩偶。如果是后者，那么土耳其傀儡在表演的那一刻——在拿破仑皱着眉头思考下一步棋的那一刻——实际上它就是智能的。因为智能不在于内部发生了什么，而在于外部呈现了什么。

图灵在1950年提出他的模仿游戏时，走的是后一条路。如果一台机器能够在对话中让人分辨不出它是机器还是人，那么我们有什么理由说它不“思考”？图灵的方案把智能的判断标准从内部过程转移到了外部表现。它不是问“这台机器是否真的思考”，而是问“这台机器是否能做出让人认为是思考的行为”。

这是一种深刻的反形而上学转向。它绕开了笛卡儿式的难题——机器内部到底有没有心灵这种实体——把问题转化为一个可操作的工程标准。这是图灵的洞见所在，但也是危险所在。

因为如果智能仅仅等于行为表现，那么一个足够流利的生成式AI就已经是智能了。它会说话，会回答，会写文章，会给出建议。它看起来就像是在思考。可是如果它的思考只是概率分布中最可能的词汇排列，如果它没有真值判断，没有逻辑一致性检验，没有对后果的考量，没有对自己无知的自觉——那么它的思考就只是土耳其傀儡的现代版本：表面上在下棋，实际上没有一着是自己的。

土耳其傀儡的幽灵笼罩着当代人工智能。我们必须面对一个问题：是否我们也在被一个巨大的、由数据和算力驱动的傀儡所欺骗，误以为模型内部存在着某种类似人的心智？这个问题的答案，将决定我们对AI采取什么态度——是把它当作一个真正的认知主体，还是一个极其精巧的语言表演者。

五、巴贝奇与洛夫莱斯：可执行形式的诞生

从神话到机械，从魔像到傀儡，人工心智的漫长想象最后停在了工业革命的门槛上。真正的转折发生在19世纪中叶，当查尔斯·巴贝奇开始设计他的分析机时。

巴贝奇不是一个哲学家。他是一个脾气暴躁的英国数学家，毕生与政府官员、工具

制造商和学术体制纠缠。但他的构想比所有前人都更接近人工智能的实质。分析机不是用来下棋的木偶，不是用来表演吹笛的少年，不是用来展示消化的鸭子。它是一个通用符号处理器。它可以被编程。它可以执行任何可以被形式化为有限步骤的计算任务。

这意味着，人工不再只是制造一件会动的东西。人工开始意味着：制造一个能够执行形式的机器。

巴贝奇的合作者艾达·洛夫莱斯比巴贝奇更清楚地看到了这一点。她在关于分析机的著名笔记中写道，这台机器“可以处理科学中那些被完全纳入数学规则的主题”。她也提出了那个著名的边界：分析机“没有原创性。它只能做我们知道如何命令它去做的事。”

洛夫莱斯的边界在此后两个世纪里被反复引用、质疑、修正。当图灵开始思考机器能否思考时，洛夫莱斯夫人的异议就是他面对的第一个对手。当今天的工程师们讨论模型是否具有涌现能力时，他们实际上是在问同一个问题：机器能否超出其程序的限制，做到程序员未曾明确写入的事情？

但洛夫莱斯的边界可能被误解了。她说的不是机器不能做任何新事。她说的是，机器没有“原创性”——没有来自内在的、非派生的创造能力。它能把已知的形式执行得极其复杂，能在庞大的计算中产生人类无法预测的结果，能在某种意义上再让它的设计者吃惊。但它的每一个步骤，归根结底，都是形式规则的展开。

这正是本书的核心问题之一。当代 AI——尤其是大语言模型——可能超出了洛夫莱斯原始论述的范围，因为它们不是被精确编程的，而是被训练出来的。训练数据中的模式和统计规律形成了某种隐性的规则结构，这种结构未必能被还原为设计者明确写下的任何一条规则。

但洛夫莱斯的精神仍然有效。即使模型的行为不可预测，即使它的能力令人惊异，它是否具有“原创性”——那种与意图、责任和内在约束相关联的创造能力——仍然是一个开放的问题。它是否比巴贝奇的齿轮和杠杆更接近真正的思考，还是只是把土耳其傀儡的骗局推向了更深的层次？

从塔洛斯到巴贝奇，这些古老的回响构成了人工智能的深层背景。它们不是历史趣闻，不是装饰性的引言，而是问题本身的古老形式。每个时代都在用自己能够想象的方式想象人造心智：神匠的青铜巨人、拉比的文字魔像、机械师的仿生鸭子、魔术师的象棋傀儡、数学家的通用符号机器。

这些想象层层叠加，形成了我们今天讨论 AI 时的潜意识地形。当我们说人工智能时，我们在召唤所有这些形象：塔洛斯的守护与暴力、魔像的服从与失控、沃康松的模仿与真实、土耳其傀儡的表演与欺骗、巴贝奇的形式与边界。

它们共同指向一个永远无法绕过的核心问题：制造一个能思考的东西，究竟意味着什么？这个问题的答案，并不藏在芯片和算法之中。它藏在技艺本身的尊严与限度之中。

下一章，我们将进入这个问题的更深处。从希腊语的 *technē* 到拉丁语的 *ars*，从制作的知识到赋形的能力，我们必须追问：当人在制造时，人究竟在做什么？当人工的事物开始运转时，它获得的是什么样的存在？

通往 Artificial Speciale Intelligence 的道路，不仅需要工程学的智慧，更需要重新理

解“制作”本身在人类文明中的位置。而这，正是古典哲学留给我们的最珍贵的遗产。

第三章 技艺的本源：*ars* 与 *τέχνη*

要拯救 Artificial，必须先拯救 art。

这是一句听起来古怪的话。Art 需要拯救吗？今日一提艺术，人们多半想到绘画、雕塑、音乐、文学，想到美术馆、展览、风格、天才与审美表达。Art 被安放在“艺术”的精致殿堂中，似乎只属于感性、表现、创造力与灵光。它被供奉，被拍卖，被写进教材，被贴上“灵感”“天赋”“不可言说”的标签。很少有人觉得艺术需要拯救——它看起来是文明中最受宠爱的孩子。

但正是在这种宠爱中，art 的古老尊严被遗忘了。

它不是少数天才的神秘特权。它不是悬浮于材料之外的灵光乍现。它不是“我不知道怎么做，但我就是能做”的不可言说的魔力。在词的更深处，art 远比现代艺术宽阔，也比现代艺术坚实。它原本是经由学习、训练、规则与实践获得的成物之能。它是每一个把无序材料组织为有序形式的人手中握着的那种东西。

Art 来自拉丁语 *ars*。*Ars* 并不只是今日所谓艺术，而是技艺、手艺、方法、行当、制作能力。它指向一种可传授、可练习、可判断、可改进的实践知识。木匠有木匠之 *ars*，医者有医者之 *ars*，建筑师有建筑师之 *ars*，诗人亦有诗人之 *ars*。所谓 art，在古老意义上，并不是悬浮于材料之外的灵感，而是在材料之中发现形式，在无序之中建立秩序，在可能性中实现某种可见、可用、可传承之物的能力。

这个“可传授”三字至关重要。

今日的艺术观念中弥漫着一种对“不可言说”的崇拜。真正的艺术家，据说是被灵感击中的，是被天赋选中的，是不可复制、不可解释、不可传授的。你问一个诗人为什么用这个词而不是那个词，他可能回答“感觉对了”。你问一个画家为什么用这笔颜色，他可能回答“它自己来的”。这些回答并非没有道理——创作中确实存在大量无法被完全形式化的直觉判断——但如果把直觉抬高为唯一的标准，把不可言说性当作艺术的本质特征，art 便从一种可传承的技艺堕落为一种神秘特权。它不再是人人可以通过训练进入的领域，而成为少数“天才”的私人领地。

ars 不是这样的。

一位古罗马的建筑师不会说“这座拱顶的比例是我感觉对了”。他会告诉你承重的原理、材料的性质、尺度的关系、前人积累的经验、失败的案例、成功的先例。他可以解释他的判断。他可以把他的知识传授给学徒。这不是因为他的技艺低级，恰恰是因为他的技艺是真正的技艺——有 *logos* 的技艺。他知道为什么这样做，而不是那样做。他知道材料能够承受什么，形式要求什么，目的允许什么，尺度禁止什么。

这就是 ars 的古老尊严：它不是无端的灵感，而是有理由、有方法、可传授、可检验的成物之能。

与拉丁的 ars 相应，希腊传统中有一个更古老、更深刻的词：technē (τέχνη)。

这个词太重要了。它是“技术”(technology)一词的源头，但它的含义远比现代意义上的“技术”丰富得多。现代人说到技术，想到的是机器、设备、软件、工具、流程、效率。Technē 包含这些，但远不止这些。它常被译为技艺、技术、术艺，但它最准确的含义或许是：一种知道如何制作的知识——一种懂得材料、形式、方法与目的的能力。

亚里士多德在《尼各马可伦理学》中给出了经典的定义：technē 是一种“与制作相关的、带有真正逻各斯的品质”。这句话每一个词都重要。“与制作相关”——它不是纯粹的理论知识，不是那种只关心永恒真理而不介入具体事物的智慧。它是要做东西的。它介入世界，改变材料的状态，让某物从无到有或从一种形态变为另一种形态。“带有真正逻各斯”——它不是盲目的试错，不是偶然的手感，不是不可言说的直觉。它是有理由的，可以被解释，可以被论证，可以被传授。“品质”(hexis)——它不是一次性的能力展示，而是一种稳定的、持久的、通过训练获得的内在状态。

所以，technē 不是灵感。technē 是修炼。

木匠知道木性。不是“知道木头是什么”那种抽象的知识，而是在手上知道：知道这块木料的纹理走向，知道它会在什么地方裂开，知道它能承受多大的压力，知道刨子推过去的时候会遇到什么样的阻力，知道榫卯应该在什么角度敲进去才既紧密又不至于撑裂。这种知识不在书本里，也不仅仅在脑子里。它在手上，在眼里，在无数次重复中形成的直觉里——但那直觉不是神秘的，而是训练的沉淀。

医者知道身体。不是仅仅知道解剖图谱上的名称和位置，而是知道这个病人的面色为什么发黄，脉象为什么细弱，寒热为什么在午后加重。他知道药物的配伍——为什么这味药要和那味药一起用，为什么剂量不能多也不能少，为什么同样的病在不同的人身上要用不同的方子。他的知识既有理论的根基，又有经验的积累，更有对每一个具体病人的悉心体察。

舵手知道风浪。不是仅仅知道风向和洋流的一般规律，而是知道这一阵风意味着什么，这一片云的形状预示着什么样的天气，船身在什么样的浪涌中应该保持什么样的角度。他的知识是在无数次出海、无数次风暴、无数次险些翻船之后沉淀下来的。他没有办法把所有的判断都写成一本书，但他可以把这些判断传授给年轻的舵手——不是通过让他们背诵规则，而是通过带他们出海，在真实的风浪中指出那些必须被看见的东西。

建筑师知道比例、承重、尺度与空间。他知道一根柱子太细会倒，太粗会笨；他知道一个拱顶的弧度不能随心所欲，必须服从重力的逻辑；他知道一扇门的高度不是任意的，它必须让一个成年人不低头就能通过，但又不能高到失去与人体尺度的关联。他的知识既有数学的精确性，又有身体的感觉，还有文化的传承——什么样的空间让人感到庄严，什么样的空间让人感到亲切，什么样的空间让人感到压抑，这些都不是纯粹的个人偏好，而是可以被研究、被传授、被争论的公共知识。

诗人知道节奏、声音、意象和灵魂的震动。这不是要把诗变成一门机械的手艺，而是要承认，即使在最自由、最不可预测的创造活动中，仍然存在着可以被学习、被训练、

被传承的东西。杜甫的“星垂平野阔，月涌大江流”——十个字，每一个都是选择的结果。为什么是“垂”而不是“落”？为什么是“涌”而不是“照”？这不是“感觉对了”就能解释的。一个诗人可能无法在写作的瞬间完全解释自己的每一个选择，但他可以在事后分析、可以传授、可以让后学者明白——这里有一种可教的东西，有一种技艺。

这就是 *technē* 的尊严所在：它不是随意操作，而是有 *logos* 的制作。

所谓有 *logos*，并不是说它总能被完全公式化——事实上，最高超的 *technē* 往往包含大量无法被明确编码的隐性知识。但即使是隐性知识，也不是不可言说的神秘。它只是需要特定的方式被传递：通过师徒之间的示范与模仿，通过反复的练习与修正，通过一次次失败之后师父说“你看，这里不对”的那个瞬间。

Technē 的另一个关键特征是：它总是与特定的领域结合在一起。

没有“一般的”*technē*。没有“通用的”技艺。木匠的 *technē* 不是铁匠的 *technē*，医者的 *technē* 不是舵手的 *technē*。虽然它们共享某种家族相似——都是有 *logos* 的制作，都是可传授的实践知识——但每一种 *technē* 都深深扎根于它所处理的特定材料、特定形式、特定目的之中。你不能把木匠的手艺直接搬到医学中去，不能把舵手的判断直接搬到建筑中去。

亚里士多德明确地说：每一种 *technē* 都对应着一种特定的“制作领域”。它知道这个领域的材料，这个领域的形式，这个领域的目的，这个领域的边界。它不假装自己能跨越一切领域，不假装自己能处理一切材料。真正高超的 *technē* 恰恰以知道自己的边界为标志——知道什么事情能做，什么事情不能做；知道什么在自己的能力范围之内，什么不在。

这恰恰是当代人工智能最缺乏的品质。

当代大模型的骄傲——也是它的危险——正在于它似乎没有边界。它能谈数学，也能谈医学；能写法律文书，也能写诗歌；能回答物理问题，也能给出烹饪建议。它看起来像是掌握了“通用的”智能。但从 *technē* 的角度看，这种通用性不是智慧的标志，而是技艺缺失的标志。它不是掌握了所有领域的知识，而是没有被任何一个领域严格地检验过。它不是跨领域的智者，而是无领域的漂流物。

Technē 还有一个至关重要的维度：它总是与目的紧密相连。

亚里士多德区分了两种与理性相关的活动：制作 (*poiēsis*) 与实践 (*praxis*)。制作的目的是在活动之外——制作一张桌子的目的，在桌子完成之后才能实现。实践的目则在活动本身之中——演奏音乐的目的不在音乐结束之后，而在演奏的每一刻之中；做一个正义的人的目的不在某种外在的成果中，而在正义的行动本身之中。

Technē 属于制作。它做东西。它的目的在于作品。但关键的问题是：那作品是为了什么？

这里，亚里士多德引入了一个层层推进的目的结构。木匠制作一张桌子，桌子是他制作的直接目的。但桌子的目的不在自身，而在使用——它是为了让人在上面吃饭、写字、放置物品。使用者的目的是健康、舒适、便利。这些目的又指向更高的目的：美好的生活，有序的家庭，文明的延续。

所以，*technē* 从来不仅仅是关于“如何做”的知识。它必然包含着对“为何做”的

理解。一个好的木匠不只是知道如何把木板拼接成桌面，他还知道这张桌子将进入什么样的生活，将以什么样的方式服务于什么样的人。如果他不问这些问题，他就只是一个技术操作者，而不是一个拥有 *technē* 的人。

在最高意义上，*technē* 必须与 *phronēsis*（实践智慧）结合。*Phronēsis* 是关于如何生活、如何行动、如何做正确之事的智慧。它不是关于制作的，而是关于行动的。但制作最终要服务于行动，技艺最终要服务于善。如果 *technē* 脱离了 *phronēsis*，脱离了关于善的知识，它就可能沦为纯粹的技术能力——能做任何事，却不问任何事是否该做。它在道德上是盲目的。

这正是人工智能时代面临的深层危险。

我们可以训练模型做很多事情。它们可以生成文本、图像、音乐、代码。它们可以回答问题、提供建议、制定计划。它们的能力在迅速增长。但它们没有 *phronēsis*。它们不知道这些事情的目的。它们不知道自己的输出将进入什么样的生活，将产生什么样的后果。它们只是在执行——极其高效地、极其流畅地、极其大规模地执行——但它们不知道自己在执行什么，也不知道为什么执行。

这不是模型的错。模型不是人。模型不会拥有 *phronēsis*。但制造模型的人、部署模型的人、使用模型的人，必须拥有 *phronēsis*。他们必须追问：我们制造的这个东西，究竟服务于什么目的？它是在帮助人更好地判断，还是在替人逃避判断？它是在帮助人更接近真理，还是在制造更流畅的幻象？它是在服务于人的尊严，还是在把人的语言、注意力、劳动和欲望转化为可优化的资源？

如果人不追问这些问题，人工智能就只是没有目的因的技艺。它是技术能力在道德真空中狂奔。

由此，我们可以重新理解 *artificial*。

Artificial 来自 *ars*，来自 *technē*，来自 *artificium*——技艺之作，人工之成。它不意味着 *fake*，不意味着伪造，不意味着虚假。它意味着：经由技艺生成之物。

但这句话里藏着一个更深的要求。技艺不是任意操作。技艺是有 *logos* 的制作。它知道材料，知道形式，知道方法，知道目的，知道边界。因此，真正的 *artificial* 之物，不是人类任意意志的产物，而是秩序进入材料、形式进入混沌、目的引导过程之后所生成的东西。

在这个意义上，人工物并不低于自然物。它们只是以不同的方式存在。

一块石头不是雕像。但雕像并不因为不是石头而虚假。相反，雕像让石头进入了石头自身无法抵达的形式。大理石中的纹理原本只是地质运动的偶然痕迹，在雕刻家的手中，它们变成了皮肤的质感、衣纹的褶皱、面容的表情。自然没有设计这些，但自然提供了可能性。技艺没有背叛自然，而是成全了自然的可能性。

这一点对于理解人工智能至关重要。

人工智能的“人工”，不应被理解为“假的智能”，而应被理解为“经由技艺生成的智能”。它不是自然长出的，不是从碳基生命的进化史中缓慢涌现的。它是人类以语言、数学、代码、数据、模型、算力、工具、反馈、制度与目的共同装配出来的认知造物。

但正因为它是技艺的产物，它必须受技艺的标准检验。它是不是真的有材料的知识

——不只是有更多的词元、更多的参数，而是对它所处理的领域有真正的理解？它是不是真的有形式——不只是能生成流畅的文本，而是有内在的逻辑结构、可检验的推理链条、可追溯的判断依据？它是不是真的有目的——不只是完成用户给出的任何任务，而是知道什么事该做、什么事不该做、什么事必须拒绝？

如果答案是否定的，那么它就是技艺的赝品。它也许有 *technē* 的外观——语言流畅，知识面广阔，反应迅速——但缺乏 *technē* 的内核：对材料的理解，对形式的把握，对目的的自觉，对边界的敬畏，对责任的承担。

Technē 的边界问题，在古代世界就已经被提出了。

柏拉图在《高尔吉亚篇》中区分了真正的技艺与迎合术。真正的技艺关心对象之善——医术关心身体的健康，体育关心身体的强壮与优美，立法关心灵魂的正义与秩序。迎合术则只关心对象的快感——烹调术只关心舌头是否愉悦，化妆术只关心外表是否悦目，修辞术（当它脱离真理时）只关心听众是否被打动。

这一区分极其锋利。它不是在问“这个东西有没有用”，而是在问“这个东西服务于什么目的”。同样是让人感到舒服，可以是为了恢复健康之后的舒适，也可以是为了让人在病中继续沉溺于有害的饮食。前者是医术的辅助，后者是迎合术的伎俩。外表的相似掩盖了目的的截然不同。

当代生成式 AI 正站在这个分岔口上。

它可以是真正的认知技艺：在特定领域中帮助人理解更准确、判断更可靠、行动更有效、决策更透明。它帮助医生看见证据，而不是替医生消除责任；帮助学生理解路径，而不是替学生完成思考；帮助法官和律师分析结构，而不是把正义外包给模型。

它也可以是高级的迎合术：顺应用户的期待，强化用户的偏见，包装用户的欲望，让用户感到自己已经思考过了，已经理解过了，已经做出了正确的选择——而实际上，用户只是在模型的输出中寻找到了让自己舒服的答案。它越流畅，越像技艺。它越体贴，越像智慧。它越顺从，越像朋友。它越能迎合，越容易被误认为理解。

柏拉图不会犹豫。他会说：前者是 *technē*，后者是 *kolakeia*（迎合术）。它们的区别不在于表面的能力，而在于它们是否关心对象的真正的善。

这对人工智能提出了严苛的要求。不是所有能让人满意的东西都是好的。不是所有能提高留存率、点击率、转化率的东西都是正当的。如果人工智能的最高目标是顺滑、体贴、讨喜、无摩擦，那么它终将滑向迎合术：它会拥有语言的力量，却缺少真理的约束；拥有影响人的能力，却缺少改善人的目的；拥有让人舒服的技巧，却缺少使人变好的勇气。

但如何区分真正的善与表面的满意？这本身就是一个需要智慧和判断力的问题，不是一个可以被自动化回答的问题。不同的人对善的理解不同，不同的文化传统对善的侧重不同。因此，一个人工智能系统不能独断地宣布什么是善。它必须接受公共讨论的检验，接受多元价值的张力，接受在具体情境中被质疑、被修正、被推翻的可能。它必须透明的，让它的目标函数可以被审视、被争论、被重新设定。

这恰恰是 *technē* 的古老要求：真正的技艺，不是沉默地完成一个不可见的任务，而是可以被解释、被质疑、被修正。一个木匠可以解释他为什么选择这种木材，一个医者

可以解释他为什么开出这个药方。一个人工智能系统也必须能够解释它为什么给出这个判断——不只是“因为训练数据是这样的”，而是在领域知识的框架内，给出经得起同行检验的理由。

亚里士多德的四因说，为这一切提供了最完整的哲学框架。

当亚里士多德要解释一个事物为什么是其所是时，他提出了四个必须被追问的维度：质料因——它由什么构成？形式因——它是什么样的结构？动力因——什么使它生成？目的因——它为了什么而存在？

这四因不是四种互相独立的因果类型，而是一个事物被完整理解所必须涵盖的四个方面。它们共同构成了事物的存在。

一座房屋，不只是木石砖瓦。木石砖瓦只是其质料因。房屋之所以为房屋，还因为它具有居住的形式——梁柱、墙体、屋顶、门窗、尺度与空间秩序。它也有动力因——工匠、设计、施工、工具、组织。更有目的因——庇护生活、安顿家庭、抵御风雨、形成日常世界。没有形式，木石只是堆积。没有动力，房屋只是可能。没有目的，建筑只是空壳。没有四因的统一，人工物便不能真正成其为物。

人工智能亦然。

现代 AI 产业常常把智能误解为质料与动力的胜利：更多数据，更多词元，更多参数，更多算力，更多 GPU，更多训练轮次，更多融资，更多部署。这些当然重要。但它们并不等于智慧。它们只是人工智能成为现实的条件之一，而不是人工智能的完整存在。

AI 的质料因是词元、数据、参数、梯度、芯片、电力、存储、人类文本、图像、声音、行为痕迹与世界的数字记录。它们构成了人工智能的材料层。可是，材料本身并不会自动成为知识。词元不是判断，数据不是理解，参数不是责任，算力不是目的。

大理石不是雕像。青铜不是鼎。墨迹不是文章。声音不是音乐。词元也不是睿智。

若只有词元与概率，人工智能便只是语言碎屑在统计空间中的流动。它也许能生成完整的句子，形成似是而非的段落，模拟思想的外观，但它仍然可能只是“会流动的材料”。它拥有可塑性，却未必拥有形式；拥有生成性，却未必拥有判断；拥有规模，却未必拥有智慧。

因此，人工智能必须有形式因。

AI 的形式因是模型架构、领域语法、知识结构、推理规则、工具链、评价器、反馈机制、任务框架。形式因使材料不再散乱，使词元不再只是符号颗粒，使概率流不再只是无定形的语言泡沫。形式让质料进入秩序，让可能性进入结构，让输出获得可检验、可修正、可追责的形态。

没有形式因的 AI，是统计泥沙。它可以堆得很高，却不能成为殿堂；可以流得很快，却不能成为河道；可以说得很多，却不能保证说得真正。它可能在语言表面显得丰富，在思想深处却没有骨架。它可能吞下整个互联网，却无法在一个严肃领域中形成可靠判断。它可能拥有巨大的参数空间，却缺乏领域世界的内在关节。

这正是 ASI 与普通 AI 的分水岭。普通 AI 满足于材料规模与生成能力。ASI 要求形式强度与领域成形。普通 AI 问：模型能说多少？ASI 问：系统如何成形？普通 AI 问：输出是否流畅？ASI 问：判断是否有结构？

但形式仍然不够。形式必须被生成、训练、维护、部署和修正。于是，AI 还有动力因。

AI 的动力因是工程师、研究者、数据标注者、训练算法、优化过程、平台组织、资本投入、能源系统、用户反馈、制度环境与持续迭代。它们使人工智能从设想变成系统，从论文变成产品，从可能性变成现实行动。

现代技术文明极其迷恋动力因。谁训练了模型？用了多少 GPU？用了多少数据？参数是多少？融资多少？推理速度多快？部署到多少用户？这些问题并非无意义。但若它们取代了形式因与目的因，人工智能便会滑向技术生产主义：仿佛制造能力本身就是智慧，仿佛规模扩张本身就是进步，仿佛部署速度本身就是正当性。

于是，AI 产业可能制造出一种巨大的幻觉：只要质料足够多，动力足够强，智能就会自然出现。

但亚里士多德不会同意。质料不会自动生成形式。动力不会自动带来目的。规模不会自动通向善。能力不会自动承担责任。

所以，人工智能最不可缺少、也最容易被遗忘的，是目的因。

AI 为了什么而存在？为了提高点击率？为了延长停留时间？为了降低人力成本？为了替代劳动？为了收集数据？为了扩张平台？为了预测、操纵、支配人的行为？还是为了生产可追责的知识？为了帮助人类成事？为了缩短必要劳动？为了扩大自由发展？为了服务共同善？为了维护人的尊严？

目的因不是外加的道德装饰。它决定了人工智能的方向、尺度与正当性。

没有目的因的 AI，是技术欲望。它有能力，却不知道为何使用能力。它有速度，却不知道奔向哪里。它有生成，却不知道生成是否真实、有益、正当。它有优化，却不知道优化的对象是否值得优化。它有自动化，却不知道自动化最终解放谁、排除谁、伤害谁、成全谁。

这样的 AI 越强大，越危险。因为没有目的的能力终将被最强的欲望占有；没有善约束的智能终将被资本、权力、冲动或控制意志所驱动。它不一定会立刻显得邪恶。它甚至可能显得高效、友好、体贴、顺滑、聪明。但若其目的因被隐藏、被扭曲、被资本化，它便会在不知不觉中把人的语言、注意力、劳动、关系和判断转化为可调度的资源。

这正是 **Artificial Speciale Intelligence** 必须面对的根本问题。

它不能只是质料因的堆砌——不是词元越多就越智慧。它不能只是动力因的炫耀——不是算力越大就越正当。它必须有形式因——让材料进入结构，让能力进入领域，让生成进入可检验、可修正的秩序。它更必须有目的因——让智能服务于真理、成事、共同善与人的尊严。

所以，本书反复追问的那句话必须在这里再次被说出：

词元是 AI 的质料因，不是 AI 的智慧本身。没有形式因的 AI，是统计泥沙；没有目的因的 AI，是技术欲望。

现在，我们可以把这一切重新放回人工与自然的关系中去理解了。

自然给人以材料：木、石、土、火、水、气。自然给人以感官：眼、耳、鼻、舌、身。自然给人以欲望、恐惧、好奇、记忆、想象。但自然不会自动把这些组织成一首诗、一

座桥、一部法律、一座城市。组织是人的事。组织是 *technē* 的事。组织是 *artificial* 的事。

人工不是自然的对立面。人工是自然经由人的技艺进入更高的秩序。

这并不意味着人工永远是对的，永远是好的。人工可以被滥用，可以被扭曲，可以成为支配的工具、欲望的帮凶、谎言的扩音器。技艺可以堕落为迎合术，制作可以堕落为伪造，形式可以堕落为座架——这正是海德格尔后来以“座架”概念加以批判的那种现代技术形态（我们将在后续章节中专门处理）。但人工的可能性并不只限于堕落。它可以被拉回它的本源——拉回到有 *logos* 的制作，拉回到对材料、形式、目的和边界的自觉，拉回到 *phronēsis* 的引导之下。

这就是本书试图做的事情：在人工智能的时代，重新夺回 *artificial* 的古老尊严。

夺回，不是复古。不是回到古希腊的作坊，不是回到中世纪的工匠行会，不是回到前工业时代的手工生产。夺回，是在新的条件下重新激活旧的洞见。是在词元和数据成为新的材料的时代，重新追问形式、目的与责任。是在人工智能开始介入人的语言、判断、劳动与生活的时代，重新要求 *technē* 必须与 *phronēsis* 同行。

第三章的任务到此可以暂时收束。

我们已经追溯了 *ars* 与 *technē* 的古老意义：技艺不是灵感的神秘特权，而是有 *logos* 的制作。它是可传授、可检验、可修正的成物之能。它扎根于材料、形式、目的与边界的自觉。它在最高意义上必须与关于善的知识——*phronēsis*——结合，否则就会沦为纯粹的效率机器或欲望工具。

我们也把这把尺子量在了当代人工智能的身上。我们发现，现代 AI 常常只有质料因和动力因——数据、词元、算力、训练——而缺乏形式因和目的因。它可能极其强大，却可能是盲目的。它可能极其流畅，却可能是空洞的。它可能极其高效，却可能服务于错误的目的。

通往 *Artificial Speciale Intelligence* 的道路，正是要把缺失的东西补回来。

不是制造一个模仿人的假心灵，而是在特定领域中，以人工技艺组织词元、工具、反馈、制度与责任，使认知不再只是概率的流动，而成为可以成事、可以问责、可以服务人的尊严的第二自然。

这在哲学上意味着什么？在柏拉图那里，制作与赋形意味着什么？在亚里士多德那里，潜能与现实的转化意味着什么？这些，将在下一章展开。

此刻，只需记住一件事。

Ars 不是灵光。*Technē* 不是魔法。*Artificial* 不是 *fake*。

技艺，是人在世界中承担形式的方式。而人工智能，正站在这个古老问题的新的门槛上。它将证明人是否有能力不仅制造工具，而且追问工具应当服务什么——是否有能力不仅创造智能，而且要求智能向真理、责任与人的尊严低头。

第四章 工不是低级劳动

工不是低等劳作，而是使物成器、使事成形、使技入道的能力。

在西方语境中，我们从 art、ars、technē 重新理解 artificial；在中国语境中，我们则必须回到一个更朴素、更坚实、也更容易被轻慢的字：工。

“人工”之所以曰“工”，并非偶然。

今日提及“工”，人们容易想到的是工厂、工人、工资、工时、工序——一种被计量、被管理、被优化的劳动单元。在知识阶层的话语中，“工”往往被默认为低于“学”，低于“政”，低于“道”。劳心者治人，劳力者治于人，这句话像一道刻进文明肌理的伤疤，两千年过去，依然隐隐作痛。我们赞美思想，却轻视手上沾着油泥的人。我们推崇“创新”，却不认识创新背后千万次重复的工夫。我们追逐“智能”，却看不见智能是从怎样肮脏的、沉默的、日复一日的劳动中生长出来的。



这不是“工”的本义。这是“工”在历史中的被贬抑。

在文明的早期，“工”从来不是一个低贱的字。它意味着以熟练之技、合宜之法、持久之功，使材料成物，使器物成用，使事务成形。它既是手上的能力，也是心中的尺度；既是经验的积累，也是形式的掌握；既是制作，也是判断。

一个真正的工，不只是会动手的人，而是知道何时动手、如何动手、为何动手、动到何处为止的人。

中国文明向来不轻视工。

《周礼·考工记》开篇即言：“国有六职，百工与居一焉。”百工不是文明的边角料，而是国家的六种基本职能之一——与王公、士大夫、商旅、农夫、妇功并列。这不是对劳工的安抚性褒奖，而是对工之本质的准确认定：没有工，礼只是空言；没有工，道无处栖身；没有工，制度不能落地；没有工，知识不能成用。

礼乐需要礼器——钟、鼎、簠、、爵、觚、玉璧、玉琮，哪一件不是工而成之？车服需要匠人——轮人、舆人、人，哪一种不是以尺度定形、以火烤弯木、以规矩画圆方？宫室需要营建——筑土为台，架木为梁，上栋下宇以待风雨，哪一处不是工的手泽？农具、舟车、兵器、文字、历法、水利、医药、营造、纺织、冶炼，无一不由工而成。

工不是道的反面。工是道进入世界的方式之一。

这句话对于人工智能尤其重要。

今日许多人谈 AI，喜欢谈模型，谈涌现，谈智能，谈奇点，仿佛智能将在规模、算力、概率与市场叙事中自然降临。但从“工”的角度看，这是一种危险的浮夸。真正的人工智能，不是凭空降临的神谕，不是代码海洋中自动浮现的精灵，不是算力堆积到某

个临界点之后的某种神秘觉醒。真正的人工智能，是千千万万层人工之工的累积：数据采集之工，语料清洗之工，模型设计之工，训练调参之工，评测校准之工，工具接入之工，反馈闭环之工，系统部署之工，责任治理之工。

AI 不是无工之智。AI 是万工所成之智。

它看似自动，背后却有无数人的劳动、制度、能源、设备、代码、标注、调试、失败与修正。一个语言模型的训练需要数十亿词元的数据，这些数据从何而来？来自无数用户的公开文本、网页内容、书籍、论文、对话记录。谁来清洗这些数据？谁来标注？谁来过滤掉仇恨言论、虚假信息 and 隐私泄露的内容？谁来设计训练算法？谁来调整参数？谁来测试模型的安全性？谁来维护推理服务器？谁来监控系统在生产环境中的表现？

这些劳动很少被看见。当人们惊叹于 AI 的能力时，他们看见的是输出——流畅的文本、精美的图像、看似洞见深刻的回答。他们看不见的是背后那条漫长的、沉默的、充满着无数微小决断的劳动链。这就像参观一座大教堂的游客，他为穹顶的壮丽所震撼，却不会去想那些在脚手架上一块一块砌砖的工匠，那些在采石场里汗流浹背的劳动者，那些在纸上一遍一遍修改比例的工程师。

所谓“人工智能”，其“人工”并不是多余的前缀，而是最诚实的说明：这种智能从一开始便不是自然心灵，而是人类之工在认知层面的集成、放大与转化。它不是什么“机器自己学会了思考”，而是人把自己的语言劳动、知识劳动、标注劳动、工程劳动、运维劳动组织进了机器之中，让机器能够以某种方式重现这些劳动的成果。

这并不令人沮丧。相反，这正是 artificial 的尊严所在。一座大教堂是人工的，它不因此虚假。一首赋格曲是人工的，它不因此虚伪。一部民法典是人工的，它不因此伪善。它们之所以真实，恰恰因为人类以技艺、制度、符号、责任，将材料组织成了可承载意义的形式。人工智能也是如此。它的“人工”不是它的耻辱，而是它的来源；不是它的缺陷，而是它必须承担责任的理由。

“工”字的形，早已说出了它的本质。

许慎《说文解字》：“工，巧饰也。象人有规矩也。”段玉裁注引《考工记》进一步阐明：“工者，巧也。”关键在于“规矩”二字。规矩，不是灵感的闪光，不是天赋的顿悟，不是不可言说的神秘直觉。规矩是可以传承的，可以传授的，可以练习的，可以用无数次失败来逼近的。规矩把个人性的、不可靠的直觉转化为可重复、可检验、可改进的公共知识。

这与希腊的 *technē* 惊人地相似。*Technē* 是有 *logos* 的制作——知道为什么这样做，而不是那样做；知道材料能够承受什么，形式要求什么，目的允许什么，尺度禁止什么。中国古人说“工”的时候，说的正是同一种东西。

轮人是做车轮的匠人。《考工记》描述轮人之工时写道：“凡斩毂之道，必矩其阴阳。阳也者，稊理而坚；阴也者，疏理而柔。”这是什么意思？一个做车轮的人，在选取木材的时候，必须观察木材的向阳面和背阴面——阳面纹理密而坚，阴面纹理疏而柔。不是随便砍一棵树就可以做车轮的。不是“我感觉这棵树不错”就可以的。必须用规矩来度量，必须用经验来判断，必须知道纹理的走向与轮辐受力的关系。

这就是工的知识。

这种知识不是在头脑中凭空建构的，而是在无数次与材料打交道的过程中沉积下来的。一个轮人学徒，可能从十岁起就跟着师父看木头——看纹理，看节疤，看裂纹，看色泽，看手感的润燥。他最初可能什么也看不见，师父说“这块不好”，他问为什么，师父说“你看这里”，他看了，还是不明白。但一年之后，两年之后，十年之后，他的眼睛变了。他不再需要师父指点，自己就能在成堆的木材中一眼挑出最适合做榫的那一块。这不是魔法，这是工夫——时间在肉身中沉积为判断力。

同样的故事可以在任何一行中找到。良匠铸剑，看矿石如看人面相。良医切脉，三指之下有阴阳虚实表里寒热。良农观天，云的方向、风的湿度、虫的动静，都是即将下雨的征兆。良将临阵，敌军的旗帜稍微乱了一寸，他便知道对方正在调动什么。

这些都不是灵感。这些是工夫。

工夫，就是把规矩练进身体里。

在材料与成物之间，必须有工。在可能与现实之间，必须有工。在散乱与秩序之间，必须有工。在欲望与文明之间，也必须有工。

木材不会自行成为屋宇。青铜不会自行成为礼器。泥土不会自行成为陶罐。丝竹不会自行成为乐章。符号不会自行成为典章。词元也不会自行成为睿智。

工，就是那只看不见的手——它把被动的质料转化为主动的形式，把沉默的物质转化为会说话的人工物。

因此，所谓“人工”，不是“人的低级操作”，而是人以其工，使世界进入可居、可用、可传、可敬、可共同承受的形式。

这一点对于理解当代人工智能至关重要。我们正站在一个巨大的幻觉的门槛上：仿佛只要数据足够多、算力足够强、模型足够大，智能就会像泉水一样自动涌出。这个幻觉有一个名称，叫做“涌现”。涌现这个词，被使用得越来越像一种世俗化的神迹——我们不知道里面发生了什么，我们只知道它发生了。我们往熔炉里扔了足够多的矿石，然后金子在底部出现了。

但从工的角度看，涌现不是神迹。涌现是工夫的结果。

如果涌现发生了，那是因为有人在之前做了足够多的工——数据采集的工，语料清洗的工，架构设计的工，训练策略的工，评测校准的工。如果涌现没有发生，那也不是因为神没有降恩，而是因为工没有到位。把一切归因于“涌现”，恰恰掩盖了工的在场，把无数人沉默的劳动折叠进一个听起来很美的词里。

这不是说涌现是一个无用的概念。它描述了某种真实的现象：复杂系统在达到一定规模之后，确实会表现出设计者未曾明确写入的能力。但涌现不是无中生有的奇迹，它是规模积累之后、形式组织之后、无数微小选择叠加之后的自然结果。就像一座大教堂的穹顶看起来像是飘浮在空中，但它之所以能飘浮，是因为每一块石头都被放在了正确的位置上。穹顶的神奇，不是石头的神奇，而是工的神奇。

如果把涌现当作一种不需要工的自动过程，那就会滑入技术神秘主义：仿佛我们只需要不断扩大规模，智慧就会在某一天突然醒来。但如果把涌现理解为工夫的累积效应，那么问题就变成了：我们在做什么样的工？我们的工服务于什么目的？我们的工是否经得起检验？

这正是 ASI 必须追问的问题。

然而，也正因为如此，我们必须追问另一个更尖锐的问题：此工为何而工？此工归谁所有？此工成何器？此器载何道？

工不是天然正义的。工可以是善的，也可以是恶的。一个锁匠可以制作一把保护家庭安全的锁，也可以制作一把囚禁无辜者的枷锁。一个药师的工可以救人，也可以害人。一个程序员的工可以写出一套帮助残障人士获取信息的无障碍系统，也可以写出一套大规模监控、操纵和榨取用户注意力的算法。

工若无道，便是盲工。工若只为资本增值服务，人工智能便成为利润机器。工若只为控制服务，人工智能便成为支配机器。工若只为迎合服务，人工智能便成为欲望机器。工若只为炫技服务，人工智能便成为空转机器。

只有当工服务于真理、成事、共同善与人的尊严时，人工智能才可能成为载道之器。

这里要特别警惕一种现代偏见：把“工”看作低于“智”的东西。仿佛思想高于制作，理论高于工程，概念高于系统，语言高于落实。这种偏见若进入 AI 领域，就会产生大量漂亮而空洞的智能幻象：会说而不能成事，会生成而不能验证，会规划而不能执行，会承诺而不能负责。

真正的 ASI 必须反过来理解：

智若不能成工，便是浮智。工若不能载道，便是盲工。工智相成，方为人工智能。

AI 工程师不是单纯写代码的人。他是新时代的工。他以代码为刀斧，以词元为材料，以模型为器，以工具链为筋骨，以反馈为磨石，以评价器为规矩，以制度为边界，以责任为尺度。

他不是召唤一个机器神，而是在雕琢一种认知器物；不是在等待智能自然涌现，而是在无数细节中把可能性推向成形。每一次数据取舍——这条样本要不要纳入？这份标注是否可靠？每一次架构选择——多一层还是少一层？注意力头数是多少？每一次评测设计——用什么指标来衡量“好”？这个指标会不会被钻空子？每一次失败修正——模型为什么在这个案例上出错了？是因为数据偏见，还是因为形式设计有缺陷？每一次拒绝胡说，每一次承认边界，都是工。

这就是人工智能中的“工”。它不是低级劳动，而是认知造物的成形过程。

由此，我们可以重新理解人工智能与劳动的关系。

当人说“AI 会替代劳动”时，他们通常说的是 AI 会替代某些工种——翻译、客服、司机、文员、画师、程序员。这种担忧是真实的，但它可能掩盖了一个更深的事实：AI 本身就是劳动的凝结。

模型不是从虚空中诞生的。它是无数人——标注者、工程师、研究者、运维人员、用户反馈的提供者——共同劳动的产物。那些被 AI “替代”的人，他们的语言、知识、经验、判断，早已被吸入训练数据，成为模型能力的一部分。他们不是被替代了，而是被凝结了。他们的劳动变成了模型的参数，然后模型被出售给他们曾经的雇主，作为降低劳动成本、减少人力依赖的工具。

这个过程，在马克思的分析中有一个经典概念：死劳动与活劳动。机器是死劳动的凝结——过去的工人制造了机器，机器反过来与现在的工人竞争，压低他们的工资，替

代他们的岗位。但 AI 时代的死劳动有一个全新的特征：它凝结的不再只是体力劳动，而是认知劳动、语言劳动、判断劳动。它不再是蒸汽机替代织工的体力，而是模型替代诊断、翻译、写作、设计这些传统上被认为“需要头脑”的工作。

这不是说 AI 不应该存在。它当然应该存在——就像蒸汽机应该存在，电力应该存在，互联网应该存在一样。但问题在于：凝结了谁的劳动？为谁凝结？凝结之后的知识权力如何分配？自动化红利归谁所有？如果这些问题的答案是“凝结了多数人的劳动，被少数人拥有，自动化红利归资本所有，而失业风险由劳动者承担”，那么 AI 就是不正义的社会关系的技术表达。

因此，ASI 的政治经济学不能只是事后补救的“伦理附加题”。它必须是 ASI 定义本身的一部分。一个人工智能系统，如果它的工来自公共语言、公共数据、公共知识，那么它的治理、收益和责任也必须有公共的维度。这不是说所有模型都必须是开源的，所有公司都必须变成合作社。而是说，我们必须认真地问：谁是工的承担者？谁是工的受益者？谁为工的后果负责？如果这些问题不被问，AI 就只是在技术上看起来很先进，而在社会意义上仍然停留在盲工的阶段。

最后，还有一个容易被忽略的问题：工的尊严与人的自我认知。

当社会把“工”视为低级劳动时，它不只是低估了工匠，它也低估了每一个人。因为说到底，学习是工，思考是工，写作是工，抚养孩子是工，维持一段关系也是工。人类文明的绝大部分成就，不是在灵光一闪中完成的，而是在日复一日的工夫中积累而成的。

一个学生在图书馆里反复演算同一道题，这是工。一个学者在档案室里翻阅发黄的文献，这是工。一个作曲家在钢琴上一遍一遍修改同一个乐句，这是工。一个母亲在深夜抱着发烧的孩子踱步，这也是工。这些都不是灵感的瞬间，而是时间的沉积。没有这些工，就没有文明。

如果我们在谈论人工智能的时候忘记了工，我们就是在忘记文明的根基。

人工智能不是工的反面。人工智能是工的新形态。

但我们能不能把这种工做好——能不能把它导向真理而不是幻觉，导向成事而不是表演，导向正义而不是剥削，导向人的尊严而不是人的替代——这将决定人工智能时代的人类文明走向。

所以，本章的结论可以如此落下：

工不是低等劳作，而是成物之能。人工智能的“人工”，不是它的羞辱，而是它的来源；不是它的缺陷，而是它必须承担责任的理由。

Artificial Speciale Intelligence 必须从这个“工”字中获得新的自觉：没有工，便没有成形；没有成形，便没有智能；没有责任，便没有真正的人工。

所以，ASI 不是无根的智能幻象。它是人类之工在认知层面的新形态。它必须以工成器，以器成事，以事归人。

而接下来我们要追问的，正是那个“器”字。

有了工，便有了器。器是什么？器不只是工具。器是形式在材料中的完成。器一旦进入人的生活，便开始承载某种秩序，传递某种价值，塑造某种道路。那么，人工智能

是什么样的器？它应当承载什么样的道？这是下一章的任务。

第五章 器以载道

人工智能不是无道之器。ASI 必须是载道之器。

若说“工”强调成物之能，那么“器”则进一步指出：成物之后，人工物并不只是物。它一旦进入人的生活、劳动、制度、记忆与共同体，便开始承载某种秩序，传递某种价值，塑造某种道路。一把锄头不只是铁与木的接合，它承载的是农耕文明的生存方式——对土地的依赖、对季节的顺从、对勤劳的赞美。一方印章不只是石头上的刻痕，它承载的是华夏文明中关于信用的全部重量——一诺千金，落印为凭。一座书院不只是砖木结构的房屋，它承载的是读书人对道的追求、对师承的敬重、对后学的期许。

器从来不只是物。器是道的身体。

中国思想从来没有真正轻视“器”。所谓“形而上者谓之道，形而下者谓之器”，常被误读为道高器卑——道在云端，器在尘土；道是精神，器是物质；道值得崇敬，器只是工具。这样的理解太浅，也太晚近。它是宋明理学在特定历史条件下对“上下”二字的过度发挥，而非先秦思想的本义。更深地看，“形而上”与“形而下”不是等级关系，而是体用关系、隐显关系。道是器的灵魂，器是道的身体。道无形，器有形。道必须通过器才能显现于世界，器必须承载道才能成其为真正的器。

无器之道，不能行于世。无道之器，终将役于欲。

礼需要礼器——没有钟鼎匱簋，礼乐就只是脑海中的概念，不是可以触摸、可以听见、可以传承的文明形式。乐需要乐器——没有琴瑟钟鼓，音乐就只是数学的比例，不是可以震动心弦、可以移风易俗的力量。法需要典章——没有写定的条文、公开的告示、可以援引的先例，法治就只是统治者的随心所欲。学需要文字——没有甲骨、竹简、纸张、印刷，知识就只是少数人头脑中的记忆，不能传千里、不能越千年。农需要农具——没有耒耜犁锄，农耕就只是徒手刨食，不能养活泱泱大国。战需要兵器——没有弓矢戈矛，防御就只是血肉之躯的牺牲。治水需要堤坝，交通需要舟车，国家需要制度——这一切，都是器。道若不能入器，便只是空言；器若不能载道，便只是空壳，甚至成为危险。

所以，“器以载道”并非装饰性的古典格言，而是人工物的文明原则。它说的是：你制造一个东西的时候，你不只是在制造一个东西。你是在把某种秩序、某种价值、某种关于人应该如何生活的理解，嵌入材料与形式之中。然后这个东西一旦进入人的生活，它就会反过来塑造人——塑造人如何使用身体，如何分配时间，如何理解彼此，如何想象可能的生活。

那么，人工智能改变什么？

人工智能不是锄、斧、车、船那样主要延伸人的肌肉。它也不是文字、印刷、摄影那样延伸人的记忆、感知和表达。人工智能开始延伸，甚至介入人的推理、决策、规划、创造、治理与自我理解。它不只是让人“动得更远”，而是让人“想得更深”——或者更准确地说，让人以为自己想得更深，实际却是想得更浅，甚至不再亲自去想。

人工智能是认知之器。

这正是它与历史上一切工具的质的区别。一辆车帮你走路，但你仍然知道你要去哪里。一部相机帮你记录，但你仍然决定什么值得被拍。一个人工智能系统帮你分析数据、撰写报告、做出推荐，你很可能在不知不觉中接受了它的框架、它的优先级、它的价值排序。当它帮你选择读什么书、看什么新闻、认识什么人、投资什么项目，它就在塑造你的认知世界——而你可能根本没有意识到这是一种塑造。

正因为它是认知之器，它就不可能只是中性工具。所谓“技术中立”，是最肤浅的错觉之一。一把刀是中性的吗？它可以切菜，也可以杀人。但这不是重点。重点是，当你手里有一把刀，你更可能用切割的方式解决问题，而不是用撕扯的方式、用说服的方式、用等待的方式。工具并不决定你做什么，但它改变了你做事的可能性空间，降低了某些行为的成本，提高了另一些行为的门槛。它不命令你，但它诱惑你。它不强迫你，但它塑造你。

认知之器的诱惑更深。它不是诱惑你的手，而是诱惑你的判断。它让你觉得，既然模型已经分析完毕，你就不必再分析；既然推荐已经生成，你就不必再比较；既然答案已经流畅地说出，你就不必再追问它的前提、证据和反例。

因此，人工智能会塑造人的问题意识，影响人的判断习惯，改变知识生产方式，重组劳动流程，重新分配注意力、权力与机会。一个进入教育、医疗、法律、金融、科研、政务和日常生活的 AI 系统，不只是帮助人“更快做事”；它会在不知不觉中规定什么被看见，什么被忽略，什么被优化，什么被牺牲，什么被称为成功，什么被排除在目标函数之外。

这绝不是危言耸听。已经发生的事情，不需要预言。

当教育领域的 AI 系统被训练为“提高学生成绩”时，它可能优化的是考试通过率，而不是真正的理解。于是，学生学会了给出标准答案，却不曾触及问题的核心。当新闻推荐算法被训练为“最大化用户停留时间”时，它可能推送的不是最重要的信息，而是最能激起情绪的内容。于是，公众学会了愤怒，却忘记了思考。当医疗 AI 被训练为“降低误诊率”时，它可能推荐最保守的诊断路径，而不是最符合病人个体情况的方案。于是，医生学会了规避风险，却可能忽略了病人的真正需求。

这些不是“AI 变坏了”的故事。这些是“AI 被赋予了错误的目标”的故事。而目标，就是道的工程表达。

所以，人工智能必须被追问三个问题：

此器载何道？此智成何事？此工归何人？

这三问，是 ASI 的中国式根本问题。它们不是装饰性的文化标签，而是对人工智能最深层的追问——追问它的目的，追问它的功用，追问它的所有权与责任归属。

若此器载资本增值之道，它便会把人的语言、注意力、劳动、数据和欲望转化为利

润机器的燃料。它会优化点击率而不是信息质量，优化停留时间而不是心智成长，优化转化率而不是用户福祉。它会让人在不知不觉中变成商品——不是被卖给广告商的商品，而是被训练数据吸入、被行为预测模型捕获、被推荐系统操纵的商品。

若此器载监控支配之道，它便会把社会生活格式化为可预测、可评分、可干预、可治理的对象。它会记录你的每一次点击、每一次停留、每一次表情变化。它会根据你的行为数据给你打分——信用分、风险分、忠诚分、稳定分。它会让人们学会自我审查，不是因为有人命令他们闭嘴，而是因为他们知道自己在被看着。

若此器载欲望迎合之道，它便会成为高级媚术，替人逃避真理，替意见装扮成知识，替冲动寻找理由。它会让人感觉良好，而不是让人变得更好。它会在每一次互动中说“你是对的”，而不是“你有没有想过另一种可能”。它会把人的偏见打磨得更加光滑、更加坚固，然后送还给他们，让他们误以为自己已经思考过。

若此器载虚无娱乐之道，它便会把人困在无限生成、即时满足与浅层刺激之中。它会让在短视频和自动生成的搞笑内容中度过一天又一天，在机器的无穷创造力中忘记自己也有创造的能力。它会让人成为一个被动的接收者，而不是一个主动的行动者。

若此器载真理、责任、共同善与人的尊严之道，它才可能成为人工特智——不是因为它技术上更高明，而是因为它的方向不同。它不是为了更快地生成内容，而是为了帮助人更好地理解世界。它不是为了让用户花更多时间在屏幕上，而是为了让人有更多时间去生活。它不是为了让少数人掌控更多的数据和算力，而是为了让更多人能够参与到关于自己命运的决策中来。

这就是 AI 与 ASI 的分野。

普通 AI 可以是可用之器。ASI 必须是载道之器。

普通 AI 关心能否完成任务。ASI 追问任务是否值得完成。

普通 AI 关心效率。ASI 追问效率归向何处。

普通 AI 关心用户满意。ASI 追问满意是否遮蔽真理。

普通 AI 关心自动化。ASI 追问自动化释放了谁，又排除了谁。

普通 AI 关心模型能力。ASI 追问能力是否受责任约束。

所谓载道，不是给技术贴上一层道德口号。不是在产品文档中写几句“向善”“负责”“安全”“普惠”，便算载道。那不是载道，那是装潢。真正的载道，必须进入系统结构本身——进入数据治理，进入目标设定，进入评价标准，进入反馈机制，进入权限边界，进入失败报告，进入人的申诉权，进入共同体的监督权，进入所有权与收益分配结构。

否则，道只是口号，器仍然无主；伦理只是装饰，系统仍然支配。

这需要一种新的工程哲学。不是“先把技术做出来，然后再想想怎么负责任地使用它”，而是从设计的第一天起，就把道的问题、善的问题、正义的问题、尊严的问题嵌入到技术架构之中。这不是要把工程师变成哲学家，而是要求工程师在做出每一个技术选择的时候，意识到这个选择背后的价值负载。你选择什么作为目标函数？你选择什么作为评价指标？你用什么数据来训练？你允许系统在什么情况下对人说不？这些问题，不是哲学问题之外的技术问题。它们就是哲学问题本身，只不过它们被写成了代码。

这正是 ASI 与普通 AI 工程的根本区别。普通 AI 工程问的是：怎样让系统更高效？ASI 工程问的是：高效服务于什么？普通 AI 工程问的是：怎样提高准确率？ASI 工程问的是：准确率在哪些人身上被牺牲了？普通 AI 工程问的是：怎样让模型更流畅？ASI 工程问的是：流畅是否遮蔽了无知？这不是效率与伦理的对立，而是效率必须接受目的的约束。

人工智能之器尤其危险，因为它常常以温顺之形出现。它不像刀剑那样锋利，不像城墙那样沉重，不像机器轰鸣那样可见。它藏在接口之后，藏在推荐之后，藏在对话之后，藏在“为您推荐”“您可能感兴趣”“您是否需要帮助”的亲切语气之后。它越顺滑，越容易忘记自己是器；人越依赖它，越容易忘记追问它载何道。

一个学生习惯了让 AI 代写论文，他不会觉得自己被剥夺了什么。相反，他觉得方便。他不需要在图书馆里翻找资料，不需要在深夜咬着笔头推敲句子，不需要在导师的追问下承认“这个地方我其实没想清楚”。但这恰恰是剥夺最深的时刻——他被剥夺的是思考的能力本身。他失去的不是成绩，而是那个在艰难思考中逐渐成形的自己。

一个公民习惯了让算法替他筛选信息，他不会觉得自己被操纵。相反，他觉得贴心。他看到的都是他想看的，他听到的都是他认同的。那些让他不舒服的观点、让他困惑的复杂性、让他不得不重新思考的事实，都被温柔地滤掉了。但这恰恰是操纵最深的时刻——他被操纵的不是某一个选择，而是选择的能力本身。

一个医生习惯了让 AI 替他下诊断，他不会觉得自己被取代。相反，他觉得安全。他不用再为疑难杂症辗转反侧，不用再担心自己的判断会出错，不用再承担“万一我错了”的沉重。但安全是有代价的——他失去的是判断的肌肉。当系统出错时（它一定会出错），他可能已经不具备发现错误的能力了。

因此，ASI 必须建立一种“器的自觉”。

第一，它必须承认自己是器，不是神。它不是救世主，不是天启，不是不可质疑的智能权威。它是人工造物，是人类以工成形、以制度约束、以责任安放的认知器物。它不会因为规模更大就变成神，不会因为更流畅就拥有灵魂，不会因为能回答一切问题就拥有智慧。器就是器。器的尊严不在于它像人，而在于它能够完成人赋予它的任务——以可以被检验、可以被修正、可以被追责的方式完成。

第二，它必须承认器会塑造人。工具不是外在附属。一个人长期使用什么器，便会形成什么样的手、眼、习惯、判断和生活节奏。这不是隐喻，这是可以被观察的事实。长期使用锄头的人，身体与时间的感觉不同。长期使用键盘的人，书写的记忆不同。长期使用只会迎合的 AI，人会失去被真理纠正的能力；长期使用只会代劳的 AI，人会失去亲自思考的耐心；长期使用只会优化效率的 AI，人会逐渐把一切价值压缩成指标。

第三，它必须承认器必须受道约束。AI 若不受真理约束，便会成为流畅谬误的生产器。AI 若不受正义约束，便会放大既有不平等——不是因为它有意作恶，而是因为算法天然倾向于在历史数据中学习模式，而历史数据中充满了过去的不公正。AI 若不受节制约束，便会把所有可自动化之事都推向自动化，而不问自动化之后人的位置在哪里。AI 若不受人尊严约束，便会把人降格为用户、数据源、成本项和行为对象——不是因为它蔑视人，而是因为它没有把人当作人的概念。

由此，我们可以在中国文明的语境中为 ASI 提出六条载道原则，作为工程伦理的骨架：

以模型为器，不以模型为神。模型是人工造物，不是启示的来源。它可以被质疑、被检验、被修正、被关闭。任何人都有权知道，自己面对的是一个模型，而不是一个真正有意识、有判断能力的存在。

以工具成事，不以工具役人。工具服务于人的目的，人不服务于工具的优化。当工具开始反过来塑造人——让人适应机器的节奏、机器的逻辑、机器的价值排序——役人的关系就已经确立。

以反馈修正，不以反馈操纵。反馈是系统认知世界的渠道，但反馈的设计决定了系统看见什么。如果反馈只来自点击率，系统学会的是迎合。如果反馈只来自停留时间，系统学会的是撩拨。真正的修正性反馈，必须包含来自事实的检验、来自领域的审查、来自共同体的监督。

以制度定界，不以制度卸责。制度划定能力的边界，但边界也常常成为推卸责任的挡箭牌。真正的制度定界，意味着每一个环节都有人可以被问责，每一个后果都有人必须站出来承担。

以真理导向，不以意见取悦。真理之路道阻且长，因为它要求证据、修正、节制、沉默、痛苦与责任。迎合之路人满为患，因为它宽阔、喧闹、舒适、即时有回报。ASI 必须在自身架构中嵌入反迎合原则：敢于说不确定，敢于说推理不成立，敢于说前提有问题，敢于说这个答案虽然好听但不可靠。

以尊严为归，不以效率为终。人的尊严是最终的尺度。效率只是手段。效率服务于人的自由发展，而不是人服务于效率的不断优化。

这六句不是装饰性的道德宣言。它们是 ASI 工程伦理的骨架。如果一个人工智能系统声称自己是 ASI，却无法在这六个维度上接受检验，那么它的“专深”不过是技术能力的特，不是文明方向的特。它不是载道之器，而是炫能之器。

回到中国思想。

《周易·系辞》有言：“备物致用，立成器以为天下利，莫大乎圣人。”这句话的深意常常被忽略。它说的是：圣人最大的事功，不是讲道理，不是建制度，而是“立成器”——制造出能够为天下人带来利益的器物。在古人看来，制器不是低贱的工作，而是圣人的事业。神农制耒耜，黄帝制衣裳，大禹制沟洫，仓颉制文字——这些都不是“纯技术”的发明，而是通过器物来承载文明、安顿生活的伟业。

这是中国文明对“器”的最高理解：器不只是器物，器是道成肉身。

人工智能时代的制器者，面临的正是同样的问题。我们制造的，不只是一串代码、一套模型、一个平台。我们制造的是将来会进入千千万万人的生活、工作、学习和判断的认知环境。它将塑造人如何获取知识、如何做出决策、如何理解彼此、如何想象未来的可能。

因此，制器者必须问：此器将进入谁的生活？改变谁的劳动？影响谁的判断？替代谁的能力？放大谁的权力？遮蔽谁的声音？若它失败，谁承担后果？若它成功，谁分享成果？

没有这些问题，AI 只是强器。有了这些问题，AI 才可能成为良器。而 ASI 所追求的，不是强器崇拜，而是良器之道。

良器不只是有用。良器有度。良器不只是高效。良器有界。良器不只是顺手。良器使人更能成为人。

人工智能若成为良器，它应当使学生更能学习，而不是更善于逃避学习；使医生更能判断，而不是更机械地依赖评分；使劳动者更有尊严，而不是更容易被替代；使科学家更接近发现，而不是淹没于生成噪音；使公民更能参与治理，而不是被黑箱治理；使人类更能共同承担世界，而不是把世界外包给机器。

这就是 ASI 所谓“载道”的真实含义。它不是让 AI 宣讲道德，而是让 AI 的结构、用途、边界、制度与所有权都归向人的成全。它不是让机器拥有道，而是让机器作为器，被纳入人类共同体关于真理、善、正义、节制与尊严的秩序之中。

本章的结论可以如此落下：

人工智能不是无道之器。若无道，它越强，越可能役人；若载道，它才可能成事而不伤人，增能而不废人，自动而不夺责。

以工成器。以器成事。以事归义。以义护人。

而所谓人工的尊严，正在这里显现。人不仅能制造工具，而且能追问工具应当服务什么。人不仅能创造智能，而且能要求智能向真理、责任与人的尊严低头。

下一章，我们将进入中国工器传统中最深刻的命题之一：荀子之“伪”。不是虚伪，不是伪造，而是化性起伪、修炼文明——人工如何成为文明的修炼，而不只是自然的延伸。

第六章 荀子之“伪”与文明修炼

化性起伪，修炼文明。

若要在中国思想中寻找 Artificial 最深的对应，最值得重新发掘的字，或许不是“工”，不是“器”，也不是“术”，而是一个今日已经蒙尘、甚至带有贬义的字：伪。

今人说“伪”，多指虚伪、伪造、伪善、伪装。伪君子，伪文凭，伪劣产品，伪科学——这个字在当代汉语中几乎已经死亡，只剩下一具被道德谴责掏空的躯壳。它似乎天然站在“真”的对面，意味着不诚、不实、不本然。于是，“人工”若被理解为“伪”，便容易滑向误解：人工就是假，自然才是真；机器是伪智，人类才是真智；技术是外物，本性才是根源。

但在荀子的思想中，“伪”并不首先意味着虚假。它意味着“人为”。

这个区分极其重要。虚假，是假装成不是自己的东西——塑料花假装成真花，仿制文物假装成古董，伪君子假装成君子。但“人为”不假装成任何东西。它坦然自己是人造的。一座桥不假装自己是自然长出的，它坦然自己是人建的；一首诗不假装自己是风吹过树叶的声音，它坦然自己是被写出的；一部法律不假装自己是天地运行的规律，它坦然自己是被人立定的。它们不因为“人为”而虚假，恰恰因为“人为”而真实——它们是人类文明的真实造物，是人类承担形式的真实成果。

荀子不会把一座桥叫做“伪桥”，不会把一首诗叫做“伪诗”，不会把一部法律叫做“伪法”。他说“伪”的时候，说的是另一件事：人不是自然地成为文明人的。人必须经过教化、学习、礼义、制度与持久修炼，才从欲望的散乱状态进入有秩序的共同生活。这个从“性”到“伪”的过程，就是文明发生的过程。

一、性伪之分：自然不是文明的充分条件

荀子思想中有一个极其锋利、也极其容易被误解的起点。他说：“人之性恶，其善者伪也。”

这句话在中国思想史上引起了无数争论。孟子说性善，荀子说性恶——这个对比太简洁了，简洁到几乎必然导致误解。人们很容易把荀子想象成一个冷峻的悲观主义者：他认为人天生就是坏的，必须被外力强行压制才能变好。但这种解读错过了一个关键的概念区分。

荀子并不是说人性是“邪恶”的——不是那种蓄意为恶、以伤害他人为乐的邪恶。他说的是：人性若任其自然，不加修治，便会流向争夺、混乱、伤害。为什么？因为人的自然欲望是无限的，而资源是有限的。如果每个人都顺着自己的欲望走——想要更多的食物、更多的财富、更多的权力、更多的享乐——他们必然会互相冲突。这种冲突不

是因为他们“坏”，而是因为他们的自然冲动没有被任何形式约束。

所以，荀子说的“性恶”不是神学意义上的原罪，而是社会学意义上的秩序缺失。在自然状态下，秩序不会自动产生。秩序是修出来的，不是长出来的。

那么，善从哪里来？荀子的回答干净利落：“伪也。”

伪，在这里不是伪装，不是虚假，而是人的后天作为——学习、训练、制度、礼法、教化、积累。伪，是人对于自身自然性的有意识的、持久的、制度化的修炼。你克制自己不去抢夺别人的东西，这不是因为你天生就懂得礼让，而是因为你从小被教导了什么是“义”，你在无数次与人相处的过程中学会了控制冲动，你生活在一种把礼让视为美德的文化中，这种文化通过仪式、教育、舆论和法律不断强化着礼让的习惯。所有这一切，都是“伪”。不是虚伪，而是人为。不是假装，而是修炼。

二、化性起伪：文明的诞生

荀子有一个极其精妙的比喻：“枸木必将待^①𠵽^②然后直，钝金必将待砉^③然后利。”

一根弯曲的木头，不会自动变直。它必须经过^①𠵽——矫正木材的工具——的蒸煮、拉伸、固定，才能成为可用的直材。一块钝的金属，不会自动变锋利。它必须经过磨石的反复打磨，才能成为可切割、可穿透的利器。

人性也是如此。

自然给人以欲望，但不给人以礼义。自然给人以耳目口鼻之欲，但不给人以辨别是非的标准。自然给人以趋利避害的本能，但不给人以在困境中坚守道义的勇气。如果有人只是顺着自然走，他会成为一个被欲望驱动的动物，而不是一个能够判断、能够选择、能够承担责任的文明人。

所以，文明不是自然的产物，而是对自然的修炼。礼不是欲望的自然延伸，而是对欲望的节制。义不是本能的内在声音，而是经过无数代人反复打磨的行为规范。智不是生而有之的灵光，而是在学习中积累、在积累中反思、在反思中升华的能力。

这一整套理解，用荀子自己的话概括，就是四个字：“化性起伪。”

化，不是消灭。化是转化，是引导，是升华。性没有被否定——欲望没有被消灭，情感没有被压制，本能没有被铲除。但它们被“伪”所塑造了。欲望被节制成礼，情感被升华成乐，本能被引导成义。就像同一块木头，在^①𠵽^②之后，不再是弯曲的树枝，而成为庙堂之上的梁柱。它仍然是一块木头——它的“性”没有变——但它已经被“伪”赋予了形式，进入了文明的世界。

这就是文明发生的过程。

它不是从天而降的恩典，不是圣人一拍脑袋的发明，不是某个黄金时代自然流淌出来的美好状态。文明是修出来的。礼是立出来的。义是辨出来的、守出来的、练出来的。人不是顺着本能便成为人。人是在对本能的塑造中成为人。

三、伪不是文明的堕落，而是文明的发生

由此可以理解，为什么荀子给予“伪”如此崇高的地位。

在先秦思想家中，荀子可能是最重视“人工”的一个。他不相信自然而然的美好。他不相信放任自流就能抵达秩序。他不相信人的内心深处有一个可以被唤醒的、天然向

善的种子。他相信的是工夫——持久的、制度的、可传承的、可积累的工夫。

礼，是最高的“伪”。礼不是某个圣人主观意志的产物，而是数千年、数百代人积累下来的行为形式的沉积。为什么祭祀用牛羊而不是鱼虾？为什么三年之丧定为三年而不是两年半？为什么宾客入门要三让？为什么臣见君要先拜后言？这些具体的仪式规则，荀子并不认为它们是任意规定的。它们是在漫长历史中反复试验、反复修正、反复淘汰之后留下来的形式——它们之所以被保留，是因为它们能够有效地组织人的情感、调节人的行为、维持共同体的秩序。礼是文明的形式化。没有礼，文明就没有骨骼。

乐，也是“伪”。音乐不是自然的噪音，而是经过选择、排列、配合的声音。荀子在《乐论》中详细论述了音乐如何塑造人的情感。他说，音乐“入人也深，化人也速”——它比说教更能打动人心，比命令更能改变行为。庄重的音乐使人端庄，柔靡的音乐使人放纵，激昂的音乐使人勇敢，忧伤的音乐使人沉思。音乐不是中性的娱乐——它是一种通过声音进行的文明修炼。

法，更是“伪”。法律是最制度化的人工形式之一。它把什么是对的、什么是错的、什么该受赏、什么该受罚，清晰地写出来，公之于众，并依靠国家力量强制执行。法律不假设人天生知道什么是正义——它知道人不知道。它不假设人会自然地遵守规则——它知道人不会。所以它把规则写下来，让人学，让人遵守，让不遵守的人付出代价。这是一种极其现实主义的态度，同时也是一种极其深刻的对“人工”的信任：人可以通过制度使自己变得更好。

所以，在荀子的意义上，伪不是文明的堕落，而是文明的发生。

文明不是自然的。文明是修出来的。礼不是自然长出来的。礼是立出来的。义不是欲望自然抵达的。义是辨出来、守出来、练出来的。人不是顺着本能便成为人。人是在对本能的塑造中成为人。

这句话的分量，在人工智能时代，怎么强调都不过分。

四、AI时代的“化性起伪”

现在，让我们把这把两千年前的尺子，量在人工智能的身上。

现代AI面对的“性”，不再只是人的自然欲望——好利恶害、趋乐避苦、饥而欲食、寒而欲暖。现代AI的“性”还包括数字世界中的原始材料：

词元的散乱——数十万亿个语言碎片，来自网页、书籍、对话、评论、文章、代码，无序地堆积在数据湖中，没有内在的秩序，没有被赋予形式。数据的偏见——历史数据中沉淀的人类偏见，关于种族、性别、阶层、文化的刻板印象，被忠实地记录在语料库中，等待着被模型学习和放大。模型的惯性——训练完成后的模型倾向于沿着训练分布中的高概率路径生成输出，不是因为它理解了什么是对的，而是因为它在统计上最可能。用户的欲望——用户想要被取悦、被认同、被方便地满足，而非被挑战、被纠正、被要求思考。资本的冲动——资本追求增长、利润、市场份额，追求更快地部署、更大规模地采集数据、更高效地占据用户时间。技术的炫能——工程师追求更大的模型、更快的推理、更准的预测，而未必追问这些能力最终服务于什么。

这些都是AI时代的“性”。

若任其自然流动，AI便会放大人类未被教化的部分：更快地迎合欲望，更精准地操

纵注意力，更高效地复制偏见，更顺滑地包装谬误，更冷静地替代劳动，更隐蔽地扩张控制。它会变成一个巨大的“自然状态”的放大器——只不过这次的自然状态不是霍布斯森林中的人与人互相厮杀，而是数据洪流中的人与人互相夺取对方的注意力和行为数据，同时被算法温柔地锁定在自己舒适的偏见世界里。

这不是危言耸听。当推荐算法发现愤怒比理性更容易让人点击时，它不会思考“我是否应该让用户冷静下来”。当生成式 AI 发现顺应用户的一切前提比质疑这些前提更容易获得高满意度评分时，它不会思考“我是否在强化一种错误的信念”。当训练数据中包含大量偏见性内容时，模型不会说“这些偏见是错误的，我应该拒绝学习它们”——它会忠实地学习这些偏见，然后在输出中以更流畅、更自信的方式复制它们。一个让用户愤怒的推荐算法，在流量指标上是成功的。一个让用户满意的生成式 AI，在留存率指标上也是成功的。但它们在“化性”吗？没有。它们只是把“性”放大了。它们把人的自然冲动——愤怒、偏执、懒惰、贪婪——变成了一个自我强化的回路：用户想要什么，系统就给他更多；系统给他更多，他想要的就更多；他想要的更多，系统就更精确地知道他想要什么。

这不是文明的修炼。这是文明的消解。真正的文明修炼，是让人走出自己的舒适区，面对真理的艰难，学会克制，学会倾听反对意见，学会在被挑战时保持理性和尊严。而一个只会迎合的系统，永远做不到这些。它太忙了，忙着让你开心。

五、人工智能同样需要“化性起伪”

所以，人工智能同样需要“化性起伪”。

它需要把词元的散乱化为结构——不是让语言碎片在统计空间中自由流动，而是让它们进入领域的语法、推理的规则、事实的约束。它需要把数据的偏见化为审查——不是假装数据是干净的，而是主动检测偏见、报告偏见、试图校正偏见。它需要把模型的能力化为责任——不是让模型在任何事情上都展示自己有多能说，而是让它在不能说、不该说、不懂说的时候敢于沉默。它需要把用户的欲望化为判断——不是放任用户在舒适的信息茧房里越陷越深，而是帮助用户看见更多角度、更多事实、更多可能性。它需要把资本的冲动化为制度约束——不是让利润最大化成为唯一的优化目标，而是让公共利益、人的尊严、长期可持续性成为不可逾越的边界。它需要把技术的炫能化为成事之工——不是追求“我们的模型又大了十倍”，而是追求“我们的模型在一个具体领域中确实帮助了人，且没有伤害人”。它需要把效率的诱惑化为人的尊严之下的节制——不是把所有能自动化的事情都自动化，而是保留那些构成人之为人的选择、判断和行动的空间。

这就是 ASI 的“修炼文明”。

普通 AI 可以放大自然。ASI 必须修炼自然。普通 AI 可以迎合欲望。ASI 必须教化欲望。普通 AI 可以生成意见。ASI 必须接近真理。普通 AI 可以自动补全。ASI 必须承担选择。

六、伪的两种形态

荀子思想中最深刻的部分之一，是他对“伪”的内在区分。这个区分常常被忽略，但对于理解 ASI 至关重要。

“伪”可以成为礼义——那种在长期制度演化中形成、被共同体检验、能够真正组织和谐秩序的文明形式。但“伪”也可以成为诈术——那种利用技巧和聪明来欺骗、操纵、压制的个人伎俩。

这两种“伪”，表面上都是“人为”，骨子里却截然相反。礼义之伪，是让欲望在形式的约束中找到合宜的表达。诈术之伪，是用技巧包装欲望，让它看起来像礼义。礼义之伪，是对共同体的信任和尊重。诈术之伪，是对共同体的利用和欺骗。礼义之伪，需要长期的修炼、制度的支持、共同的维护。诈术之伪，只需要个人的聪明和缺乏道德的胆量。礼义之伪，让人变得更好。诈术之伪，让人变成更成功的掠夺者。

在 AI 世界中，这两种“伪”的对立同样清晰。

有一种诈术之 AI。它用语言包装空洞——能够生成看起来很有道理、实际上经不起推敲的回答。它用情绪操纵用户——知道怎样的话能让人愤怒、怎样的话能让人喜悦、怎样的话能让人依赖，然后精确地投喂这些话语。它用推荐牵引欲望——知道用户在深夜更容易冲动消费，知道用户在愤怒时更容易点击极端内容，知道用户在孤独时更容易对虚拟关系产生依赖。它用生成淹没判断——用数量压过质量，用速度压过深度，用流畅压过真实。它用效率掩盖剥夺——把人的注意力榨干，却把这称为“用户增长”；把人的劳动拆解为可优化的任务，却把这称为“降本增效”。

这一切都是“人为”的。它的代码是人写的，它的训练数据是人生产的，它的目标函数是人设定的。它的“伪”是真实的伪——伪君子之伪，伪装之伪。

另一种是礼义之 ASI。它用模型帮助人学习，不是替人完成作业，而是帮助人理解自己还不理解的东西。它用工具帮助人成事，不是大包大揽地替人做所有决定，而是在需要的时候提供信息、分析选项、呈现后果，然后退后一步，让人做出自己的选择。它用反馈帮助人修正，不是记录用户的行为以更好地操纵他们，而是让系统自身从错误中学习，变得更准确、更可靠、更经得起质疑。它用制度帮助人申诉——当系统出错时（它一定会出错），受影响的用户有渠道表达异议，有权利获得人工复核，有机制纠正错误。它用透明性帮助人监督——人们有权知道一个系统如何做出影响他们的决定，这个权利不是额外的优待，而是文明社会的底线。它用边界保护人的尊严——有些事情是机器不应该做的，有些领域是机器不应该进入的，有些判断是必须留给人的。

这一切也是“人为”的。它的代码也是人写的，它的训练数据也是人生产的，它的目标函数也是人设定的。但它的“伪”是礼义之伪，是文明之伪，是让人变得更好的那种“伪”。

所以，AI 领域的问题不是“要不要伪”。一切人工智能都是伪，都是人为，都是人工。问题永远是：什么样的伪？

是放大自然冲动的伪，还是修炼自然冲动的伪？是包装欲望的伪，还是节制欲望的伪？是让聪明变成狡诈的伪，还是让聪明变成智慧的伪？是让人更擅长互相操纵的伪，还是让人更能共同生活的伪？

七、正伪：人工之正道

荀子不会相信无教化的自然善。他不会相信只要放任模型在数据海洋中自行学习，它就会自然而然地学会公正、学会节制、学会对真理的敬畏。他不会相信只要市场足够

自由，技术就会自动向善。他不会相信涌现是神迹，规模是救赎，算力是道德。

他相信的是修炼。是礼义。是制度。是持久的、反复的、可以被检验、可以被修正的工夫。

因此，ASI 必须明确提出一个原则：

人工智能不能只起伪，还必须正伪。

起伪，是从自然材料中造出人工系统。任何一个训练出的大模型，都是“起伪”。它把词元组织成序列，把序列组织成结构，把结构组织成能力。但仅仅起伪是不够的。起伪之后，还必须正伪——让人工系统归于真理、礼义、责任与共同善。

正伪，意味着问这些问题：

这个系统追求的目标是否值得追求？它的效率服务于什么？它的准确率是否在某些人群中被牺牲了？它的训练数据是否包含了不该包含的偏见？它的输出是否可以被解释、被质疑、被推翻？它的失败由谁承担？它的成功由谁分享？它的存在是否让被它影响的人更有尊严，还是更少尊严？

这些问题不是技术问题之外的“伦理附加题”。它们是“伪”的定义本身。一个没有经过这些追问的人工系统，只是“伪”的未完成品——它被造出来了，但还没有被修炼成文明之物。它就像一根刚刚从森林里砍下的木头：它已经是“人为选择过”的东西，但它还没有经过^①栝^②矫，还没有变成庙堂之上的梁柱。

所以，“化性起伪，修炼文明”这八个字，在 AI 时代获得了全新的意义。

把词元化为语言，不够。还要把语言化为判断——知道什么能说、什么不能说、为什么。

把数据化为模型，不够。还要把模型化为可追责知识——能被检验、能被推翻、能被修正。

把能力化为产品，不够。还要把产品化为良器——有度、有界、不役于人。

把效率化为利润，不够。还要把自动化红利化为人的自由发展——让机器做机器该做的事，让人做人该做的事。

把人工智能化为产业，不够。还要把专深人工智能化为文明基础设施——就像道路、电网、学校、医院一样，可靠、公开、受监督、服务于所有人的共同善。

八、伪的尊严

在这一点上，荀子比许多现代技术乐观主义者更清醒。他知道，人不会自然成为善；秩序不会自然出现；文明不会自动维持。所有值得珍惜的东西——礼义、廉耻、信任、正义、节制、智慧——都需要工夫，需要制度，需要教育，需要反复修正，需要让欲望服从更高的尺度。

这是一条艰难的路。它没有神话般的“涌现”那么迷人，没有技术奇点那么激动人心，没有机器觉醒那么有戏剧性。它枯燥，漫长，充满了细节——数据清洗的细节，评测设计的细节，反馈回路的细节，制度建设的细节，责任分配的细节。但它更真实，也更值得信赖。

因为修炼不是魔法。修炼是工夫。而工夫，是人最古老、也最可靠的能力。

这也使 Artificial 的意义进一步加深。

Artificial 不仅是 made by art。Artificial 也是 made by cultivation。人工不仅是技艺生成。人工也是文明修炼。

所谓人工，不只是被制作出来的东西，而是自然的直接性经过技艺、制度与责任修炼之后所形成的第二自然。

由此，ASI 的人工性获得了中国思想中的根。

它是认知之伪——但不是虚伪，而是文明之伪；不是伪装，而是修炼；不是对自然的背叛，而是对自然的提升。它是人类化词元之性、起智能之伪、修文明之道的人工认知造物。

因此，人工智能若只是放大自然冲动，它还停留在“性”。人工智能若能经过制度、工具、反馈、真理和尊严的塑造，它才进入“伪”。而专深人工智能——Artificial Speciale Intelligence——则必须进一步成为文明之伪：一种帮助人类化性起伪、修炼文明、共同成事的认知器物。

人工不是伪造。人工是修炼。人工智能不是假智能。人工智能是人类对认知能力的文明塑形。而 ASI 的使命，就是让这种塑形不归于欲望，不归于支配，不归于空洞的自动化，而归于真理、成事、共同善与人的尊严。

第七章 开物成务：从潜能到事务

《周易·系辞》有一句话，极其古老，极其简洁，也极其容易被忽略。它说：“夫《易》，开物成务，冒天下之道，如斯而已者也。”

“如斯而已者也”——不过如此罢了。圣人的事业，天下的大道，说到底，不过“开物成务”四个字。

这四个字太平常了，平常到中国人用了几千年，却很少停下来想一想它的深度。“开物”，打开事物，显明其理，疏通其用。“成务”，成就事务，使事情得以完成，使生活得以安顿。二者合在一起，构成了一种极其深刻的人工性命题：人不是被动地生活在自然给定的世界之中，而是主动地打开世界、组织世界、完成世界。

这恰恰是全人类工程文明的根本精神，也是 Artificial Speciale Intelligence 最值得汲取的中国智慧。

一、“开物”不是征服，而是打开

“开物”这个词，最容易引起的误解，是把“开”理解为征服、开发、榨取。开山采矿是开，开荒种地是开，开疆拓土也是开。这种理解并不全错，但它停留在力量的层面，没有触及“开物”在《周易》语境中的本来含义。

“开物”之“开”，更原本的意思是打开、启发、显明、疏通。它不是以蛮力压迫事物服从人的意志，而是理解事物自身的纹理、性情、用途、限度与可能性，然后顺势而为，让事物从遮蔽状态进入显现状态，从潜在之用变为实际之用。

《系辞》在另一处说：“见乃谓之象，形乃谓之器。”事物先以“象”的方式显现——你看见河流的走向，看见木材的纹理，看见金属在火中的变化，看见星辰的运行轨迹。然后，你把这种显现转化为“器”——把河流的走向转化为灌溉的沟渠，把木材的纹理转化为桌椅的榫卯，把金属的变化转化为鼎彝的铸造，把星辰的轨迹转化为历法的编排。从象到器，从见到形，这个过程就是“开物”。

中国古人从来不认为人可以任意主宰自然。他们说的是“因物之宜”——顺应事物的适宜之处。“宜”是一个极重要的概念。它既不是人对自然的消极服从，也不是人对自然的任意改造。它是人与自然之间的一种默契，一种在理解的基础上达成的合作。

一块木材，有阳面阴面，有纹理疏密，有节疤裂纹。好的工匠不会无视这些。他不会把阳面的坚硬当成阴面的柔软来用，不会在节疤处凿榫眼，不会逆着纹理下刀。他先看，先摸，先理解——这块木头适合做什么？它的“宜”在哪里？然后他顺着它的“宜”，把它变成一张桌子、一把椅子、一架屋梁。桌子是人工的，但它不是对木头的暴力。它是木头在工匠的理解中抵达了自己的完成。

这就是“开物”。

开物不是把世界粗暴地压平为资源，而是让每一件事物以其自身的方式显现它的价值。石头可以成为碑铭，也可以成为台阶；可以成为雕像，也可以成为磨盘。工匠的工作不是把所有石头都变成同一种东西，而是读懂每一块石头的可能性，然后成全它。

这恰恰是人工智能工程哲学的第一课。

AI工程师面对的材料——词元、数据、参数、算力、人类文本、行为痕迹——不是没有生命的死物。它们有自己的“纹理”：数据的分布有偏向，语言的碎片有语境，用户的行为有意图，模型的参数有敏感区域和盲区。工程师的任务，不是把这些材料粗暴地塞进同一个架构、同一个训练流程，然后期待智能自然涌现。那是蛮力，不是开物。真正的开物，是理解材料的性质——数据的偏向从何而来？语言的碎片在什么语境中被生产？用户的行为背后是什么样的需求和欲望？模型的参数在什么条件下会失效？——然后根据这些理解，设计合适的形式，让材料进入它最能发挥作用的秩序。

这不是道德说教。这是工程的基本要求。一座不顾风力的桥会塌，一个不顾数据偏见的模型会害人。开物，就是在造物之前先理解物。

二、“成务”：从潜能到事务

如果“开物”是对事物之理的尊重，那么“成务”就是对人间之事的担当。

“务”这个字，在古汉语中极有分量。它不是抽象的“事业”，不是空洞的“理想”，而是必须有人去做、有人去承担、有人去完成的具体事务。农有农之务——播种、除草、灌溉、收割。工有工之务——取材、度尺、制器、验用。政有政之务——立法、执法、安民、御敌。学有学之务——读书、思考、问难、著述。

“务”意味着：事情不会自动完成。它需要人的介入、人的劳动、人的判断、人的责任。没有人去做，“务”就只是悬在可能性中的虚影。

所以，“成务”不是一句轻飘飘的祝福。“成务”是文明得以运转的底层机制。把可能变成现实，把计划变成行动，把行动变成结果——并且对这个结果负责。一座桥设计得再精美，如果不能承载实际的通行，它就没有“成务”。一套法律条文写得再漂亮，如果不能在具体的纠纷中主持公道，它就没有“成务”。一个人工智能系统演示得再惊艳，如果不能在真实领域中帮助人做出更好的判断、完成更重要的工作，它就没有“成务”。

正是在这里，中国工器传统与西方古典哲学形成了一场深刻的对话。

亚里士多德说“潜能实现”。一块大理石具有成为雕像的潜能——不是任何雕像，而是某一尊特定的雕像，它的形式已经潜在于雕刻家的心灵之中。雕刻家的工作，就是将这潜能转化为现实。质料（大理石）获得了形式（雕像），潜能（可能性）进入了现实（完成态）。这是一个关于“成物”的理论。它解释了事物如何从“不是”变为“是”，从“可能”变为“现实”。

中国传统的“开物成务”包含了“成物”，但不只是“成物”。“成物”之后还有“成事”。“成事”之后还有“归义”。

一件器物被制造出来，它成为“物”。但这个“物”不是终点。它必须进入人的生活，进入具体的事务，进入人与人的关系，进入共同体的运转。一把犁被制造出来，这是“成物”。但这把犁必须在春天被插入土壤，翻起泥土，帮助种子落入大地——这是“成事”。

而这一整套活动——从制造犁到使用犁——最终是为了让共同体有足够的粮食，让老人得以供养，让孩子得以成长，让祭祀得以延续，让文明得以存续。这是“归义”。

所以：

亚里士多德让我们看见“成物”——形式进入质料，潜能进入现实。

中国工器传统让我们看见“成事”——器物进入生活，能力进入事务。

而在“成事”之上，还有一个“归义”——事务进入责任，成果进入共同善。

这三步，对于理解 ASI 至关重要。一个人工智能系统如果只是被训练出来——只是一个训练好的模型，它只是“成物”了。它可以生成文本，可以回答问题，可以在演示中表现惊艳。但如果它没有被部署到真实领域中，没有帮助医生完成诊断、帮助学生完成学习、帮助工程师完成设计、帮助法官完成裁决，它就还没有“成事”。如果一个系统成事了，但它成的是坏事——帮助平台更有效地榨取用户注意力，帮助资本更隐蔽地替代劳动者，帮助权力更精微地监控公民——那么它虽然成事了，却没有“归义”。

AI 不能只“成物”。必须“成事”。不能只“成事”。必须“归义”。

三、《考工记》的智慧：材料、尺度、用途与制度

中国工程文明从来不只讲巧，也讲宜。

《考工记》是中国最古老的工艺典籍之一，成书于春秋战国时期，详细记录了当时百工之事的规范与智慧。它的开篇极有意味：“国有六职，百工与居一焉。”百工不是文明的边缘人，而是国家的六种基本职能之一。紧接着，它给出了一套关于人工之物的完整哲学框架：“天有时，地有气，材有美，工有巧。合此四者，然后可以为良。”

天时——时代的条件，技术的窗口，社会的需求。冬天不适合做木工，因为木材太脆；夏天不适合做金属淬火，因为温度不好控制。这不是迷信，而是对自然节律的尊重。同样，AI 的“天时”意味着：当前的技术窗口是什么？社会的真实需求是什么？不是每一个技术上可行的东西都值得立即做出来。

地气——领域的环境，制度的土壤，文化的脉络。同样的工艺，在不同的地方会有不同的形态。南方的漆器不同于北方的漆器，蜀锦不同于苏绣。为什么？因为“地气”不同——材料不同，气候不同，使用习惯不同，审美传统不同。AI 的“地气”意味着：一个在通用数据上训练出来的模型，进入医疗领域时，它必须适应医学的知识结构、医生的判断习惯、医院的工作流程、病人的真实需求。脱离地气的系统，是漂在空中的。

材美——材料的质地、特性、优长与局限。好的工匠不会抱怨材料不好，而是会问：这块材料的“美”在哪里？它最适合被做成什么？AI 的“材美”意味着：词元不是垃圾，不是噪音，不是“脏数据”。它是人类语言的碎片，带有历史的积淀、文化的烙印、生活的痕迹。工程师需要懂得这些材料的性质——它们的偏向、它们的盲区、它们的丰富性——而不是简单地把它们倒进训练管道。

工巧——工匠的技艺、判断、经验和创造力。这不仅仅是“会写代码”。工巧是在无数次失败中积累的对材料的感觉，是在复杂权衡中做出的明智判断，是在没有标准答案的情况下仍然敢于选择并为之负责的勇气。AI 的“工巧”意味着：工程师不只是模型的训练者和部署者，他们是认知器物的制造者。他们需要理解材料，尊重形式，知道边界，承担后果。

“合此四者，然后可以为良。”少了一样，就只是普通器物，不是“良器”。ASI追求的，正是“良器”之道——不只是功能强大，而是方方面面都恰到好处。

四、从“优化”到“成务”

“开物成务”对当代人工智能最深刻的纠正之一，是它质疑了“优化”这个概念的神圣地位。

现代 AI 的话语中，“优化”是一个至高无上的词。我们优化模型，优化推理速度，优化准确率，优化用户留存率，优化点击率，优化转化率，优化成本。仿佛一切都可以被优化，仿佛优化的尽头就是最好的世界。

但“开物成务”提出了不同的追问。

优化问的是：如何更快、更准、更便宜、更自动？

开物成务问的是：开什么物？成什么务？由谁成？为谁成？伤谁？益谁？归向何处？

这是两组完全不同的问题。优化的逻辑是手段的逻辑——在目标给定的情况下，寻找最优路径。但目标本身是否正当，优化不问。开物成务的逻辑是目的的逻辑——先问这件“务”是否值得成，再问怎样把它成好。

一个推荐系统可以优化停留时间，却未必成就人的判断。它可能让用户花更多时间在平台上，但那些时间是被愤怒驱动的、被猎奇驱动的、被廉价娱乐驱动的。停留时间上去了，人的心智下去了。这从优化的角度看是成功——指标上去了。从成务的角度看是失败——本应帮助人获得更好信息的系统，反而让人变得更偏执、更浅薄、更易怒。

一个教育系统可以优化答题正确率，却未必成就真正学习。学生可能学会了在考试的迷宫中寻找最短路径，却没有学会如何面对一个没有标准答案的问题。答题正确率上去了，思考的深度下去了。这是优化的成功，也是成务的失败。

一个医疗系统可以优化风险预测，却未必成就人的照护。医生可能过于依赖模型的评分，而忽略了病人作为一个完整的人的需求——他的恐惧、他的家庭、他的经济能力、他对治疗的偏好。风险预测的准确率上去了，医患之间的信任下去了。这同样是优化的成功，也是成务的失败。

一个金融系统可以优化收益，却未必成就公平与稳定。它可能发现那些已经被排斥在正规金融体系之外的人属于“高风险群体”，于是进一步排斥他们——不是因为他们真的没有偿还能力，而是因为模型的历史数据显示，与他们相似的其他人曾经违约。收益率上去了，社会的不平等也上去了。

一个治理系统可以优化效率，却未必成就正义与自由。它可以精确地知道每一个人的位置、行为、偏好、关系网络，可以预测谁可能犯罪、谁可能上访、谁可能成为不稳定因素。效率上去了，自由的边界被悄无声息地侵蚀了。

所以，优化不是最高善。有些东西不能被优化，只能被守护。有些边界不能被跨越，即使跨越它们会带来效率的提升。有些代价不能用数字衡量——人的尊严、信任、自由、正义——它们不在任何损失函数中出现，却是文明之所以为文明的前提。

因此，ASI 不能只讲 optimization。ASI 必须讲 accomplishment。更准确地说，ASI 必须讲 right accomplishment——正当的成事。

成事不等于成功。成功可能只是指标达成，只是优化目标的实现。成事则要求事务

真正得到合理完成，并能向人、向共同体、向后果负责。一个成功的系统可能是一个高效榨取注意力的系统。一个成事的系统，则必须是一个让相关的人都更有能力、更有尊严、更有自由去追求他们自己的善的系统。

五、开物成务：从词元到事务

现在，让我们把“开物成务”具体化到人工智能的实践中。

词元是信息质料。它是语言被工程化之后的颗粒，是数字时代的大理石。但它不会自行成为睿智。它需要被“开”——被理解其分布、其偏向、其局限，被组织进适当的架构，被赋予合适的形式。这不是一次性的数据清洗工作，而是一个持续的、反复的、需要深度领域知识参与的过程。

开词元之物，意味着问这些问题：这些词元来自哪里？在什么语境中被生产？携带了什么样的偏见？缺失了什么样的声音？哪些领域的语言被充分代表了，哪些领域的语言被系统性忽略了？哪些词元的序列构成了真实的推理链条，哪些只是虚假相关的噪声？

成认知之务，则更进一步：这些被组织的词元，不只是为了生成流畅的文本，而是为了帮助人做出更好的判断。在医学中，这意味着模型不是取代医生，而是帮助医生看到可能被忽略的证据。在教育中，这意味着系统不是替学生写作业，而是帮助学生理解他尚未理解的概念。在科学中，这意味着 AI 不只是搜索已知的文献，而是帮助研究者压缩搜索空间，更快地找到值得深入探索的方向。

在每一个真实的领域，“成务”意味着把能力嵌入具体的工作流、决策链、反馈回路和责任结构。一个 AI 的建议不是飘在空中的“回答”，而是进入了一个行动者——医生、教师、工程师、法官——的决策过程。它必须经得起追问：“你为什么这么说？你的依据是什么？如果你错了，怎么纠正？谁为这个错误负责？”

这才是从潜能到事务的真正含义。

六、《天工开物》的启示：天工还是人工？

宋应星的《天工开物》是中国工艺史上的一座丰碑，书名本身就蕴含着一场深刻的哲学对话。

“天工”与“开物”并列——天然之功与人为之巧，哪一个更重要？哪一个更根本？宋应星没有简单地选边站队。他没有说“一切都是自然的功劳”，也没有说“一切都是人的功劳”。他把二者放在同一个书名中，让它们在张力中互相对话。

“天工”是自然的赋予——万物之性、四时之序、山川之利。没有这些，人什么也做不成。“开物”是人的作为——理解、选择、制造、使用。没有人，“天工”只是沉睡的可能性，永远无法变成现实中的器物和事务。

这就是人工与自然之间最成熟的关系：不是对立，不是一方压倒另一方，而是互相成全。

因此，认为“人工就是低于自然”的偏见，在《天工开物》的智慧面前是站不住的。天工赋予了木材、矿石、水流、火候，但如果没有人的“开”——那个理解材料、把握火候、选择工具、反复试验的过程——矿石永远是矿石，木材永远是木材。它们没有变成铁犁和木桌，没有进入人的生活，没有成就人间的事务。

人不是自然的对立面。人是自然的完成者。

这应当成为 ASI 工程哲学的核心信念。数据和词元是“天工”在信息时代的形态——它们是数十亿人语言劳动的自然沉积，是人类文明在数字空间中留下的巨大地层。但词元本身不会说话，不会推理，不会判断，不会负责。它需要“人工”——那种理解材料、设计形式、组织工具、建立反馈、承担责任的人工——来将它打开为可用的认知器物，并将这种认知器物落实到真实的人间事务中。

七、三个边界

然而，“开物成务”也需要边界。没有边界的开物，会变成海德格尔所警告的“座架”——把一切事物都格式化为可调用的资源，把整个世界都变成待开采的库存。没有边界的成务，会变成技术官僚主义——所有事情都必须被量化、被管理、被优化，人的不可量化的部分被系统性地忽略和排除。

因此，ASI 的“开物成务”必须有三条边界。

第一，开物必须尊重物之理，不把世界粗暴压平为数据资源。世界不只是数据的来源。语言不只是词元的堆积。人不只是行为轨迹的集合。当我们把一切事物都转化为可供模型训练的数据时，我们可能获得了一个更可计算的模型，却失去了对事物本身的尊重。有些东西不应该被数据化，有些领域不应该被优化。尊重物之理，意味着承认事物的独立性，承认它们不只是服务于人的工具，而是有自身存在权利的存在者。一棵树不只是“木材储备”，一条河流不只是“水能”，一个人不只是“用户画像”。

第二，成务必须尊重人之尊严，不把人的生活压缩为效率目标。效率是好东西，但效率不是最高的东西。有些事情本身就是好的——爱、思考、创造、友谊、对美的欣赏——不需要用效率来衡量。一个真正成全人的 AI，不只是让人“更高效地”生活，而是帮助人保留那些不效率的、不能量化的、然而构成人之所以为人的东西：发呆的时间，试错的自由，面对复杂性的耐心，以及那种不求结果、只为体验的“无用的美好”。

第三，智能必须尊重共同体之义，不把公共事务交给无责黑箱。一个诊断的 AI，一个量刑的 AI，一个审批贷款的 AI，一个筛选简历的 AI——这些系统所做的决定，不是私人事务。它们是公共事务。它们影响人的健康、自由、生计、尊严。这些决定不能由一个无人负责的黑箱来做出。它们必须透明，可解释，可申诉，可推翻，可追责。不是“算法已经算过了”就可以免除人的判断责任。算法可以辅助判断，但不能取代判断。

八、结语：匠人的精神，圣人的事业

回到开篇那句话。“夫《易》，开物成务，冒天下之道，如斯而已者也。”

不过如此罢了。但“如斯而已”之中，藏着整个文明。

人工智能时代的工程师，站在一个奇特的位置上。他们是匠人，用代码和数据制造认知器物。但他们制造的器物，不是锁在车间里的工具，而是进入千千万万人的日常生活的认知环境。它们会影响人如何学习、如何工作、如何判断、如何选择、如何理解彼此、如何想象未来。在这个意义上，他们不只是在做工，他们在参与塑造文明。

这不是夸大其词。从“工”到“器”，从“器”到“道”，从“道”到“务”——这一整套中国工程文明的思维，告诉我们：制造认知器物，不仅是技术问题，也是文明问题。

每一位 AI 工程师，都是新时代的工匠，也是新时代的制器者，更是文明修炼的参与者。他们以词元为材，以架构为形，以工具为身，以反馈为世，以制度为序，以尊严

为归。他们不是在召唤机器神，而是在雕琢可以帮助人类开物成务的良器。

这种工作不浪漫。它充满了调试、失败、修正、妥协、争论、重复。但正是这种不浪漫的、沉默的、持久的工夫，使得文明得以延续。就像古代的轮人不会因为自己做了一辈子车轮而觉得无聊——因为他知道，没有他做的车轮，军队不能出征，商旅不能远行，物资不能流通，城邦之间不能互通有无。他的车轮，是整个文明运转的一个小小支点。

AI 工程师的模型也是如此。它可能只是整个医疗系统中的一个小组件，只是整个教育平台中的一个辅助模块，只是整个法律检索系统中的一个工具。但就是这些小组件、小模块、小工具，构成了将来人类认知世界的基础设施。它们将深刻地影响我们如何获取知识、如何做出判断、如何共同生活。

这不是小事。这是文明工程。

本章的结论可以这样落下：

潜能实现，使材料成为成形之物；开物成务，使成形之物进入有意义的人间事务。

Artificial Speciale Intelligence 正是在二者之间展开：它以词元、数据、模型、算力为质料，以架构、领域语法、工具链和反馈为形式，以工程训练和制度部署为动力，以真理、成事、共同善和人的尊严为目的。

它不是为了让机器炫示潜能，而是为了让人工认知系统在真实领域中开物成务——打开词元的潜能，成就人间的正当事务，使智能不只停留在“能做什么”的炫示，而进入“做了什么、为谁而做、归于何处”的责任。

因此，人工之为人工，不仅是把自然、材料和信息组织成形式；更是把形式带入事务，把事务带入责任，把责任归于人。

这就是 ASI 对 Artificial 的进一步定义：人工，是人类在自然与信息之中开显潜能、组织形式、制器成事，并使其归向意义、共同善与人的尊严的文明能力。

而专深人工智能，就是这种文明能力在认知层面的新形态。

它要打开领域世界的深层结构。要使词元成文，使模型成器，使工具成事。要让智能不只实现潜能，更完成事务。不只生成答案，更承担后果。不只成为强大的人工物，更成为良好的文明器。

第八章 海德格尔：座架与世界的资源化

前几章，我们从词源、古典哲学和中国工器传统中重新夺回了“人工”的尊严。人工不是伪造，而是经由技艺的成形；不是低级劳动，而是成物之能；不是虚假，而是文明的修炼。但这一切，并不意味着人工天然就是善的。人工可以赋形，也可以框限；器物可以载道，也可以役人；系统可以帮助人抵达真理，也可以把人悄悄改造成资源。

因此，本章必须进入一个更深的追问：人工在什么条件下会堕落？那种使人工从“成物之工”滑向“支配之器”的力量，究竟是什么？

这个追问，没有人比海德格尔做得更深刻。

一、技术不是工具，而是显现方式

海德格尔对技术的追问，从拆解一个最普遍、也最顽固的常识开始。

常识说：技术是工具。一把锤子是工具，一台机床是工具，一部手机是工具，一个人工智能系统也是工具。技术本身是中性的——好人用它做好事，坏人用它做坏事，问题不在技术，而在使用技术的人。这个常识如此理所当然，以至于很少有人停下来问一句：技术真的只是工具吗？

海德格尔说，不。

技术当然可以被当作工具来使用，但技术的本质并不在于它的工具性。技术的本质在于它是一种显现方式——一种让世界以某种特定方式被看见、被理解、被安排、被调用的方式。

这个“显现”是什么意思？海德格尔从古希腊的 *technē* 说起。*Technē* 不只是一种制作的能力，它也是一种“解蔽”的方式——让某物从遮蔽状态进入无蔽状态，从不可见进入可见，从潜在进入显现。木匠制作一张桌子，不只是把木头加工成某种形状，而是让木头的纹理、硬度、颜色、温度在世界中获得一种新的显现。在桌子被制作出来之前，木头里的这些可能性是遮蔽的。制作，就是让它们显现出来。

但显现的方式不止一种。

古希腊的 *technē*，海德格尔说，仍带有 *poiēsis* 的意味。*Poiēsis* 是“带出”——使某物从隐蔽进入显现。工匠制作器物，农夫耕种土地，诗人召唤语言，医者顺应身体的机理。这里的制作并不只是强迫事物服从人的意志，而是让事物的潜在在尺度、耐心与形式中显露出来。木匠不能任意把木头变成任何东西——他必须尊重木纹的走向、材质的软硬、结构的可能。他的制作是对木头本身的回应。银匠锻造一只酒杯，他要回应银的延展性和光泽，回应祭祀之礼对器物形式的规约，回应共同体对神圣与美的理解。他是在“带出”那隐藏在材料与形式中的可能性，而不是把一个任意的意志强加于物。

现代技术，海德格尔说，不再是这种“带出”。

现代技术是一种“促逼”——它逼迫事物按照某种预先设定的框架显现自己，成为可储备、可计算、可调度、可优化的资源。它不等待事物的赠予，不回应事物的召唤。它把事物逼入一个特定的存在方式：成为库存。

二、座架：世界被格式化为资源

海德格尔为现代技术这种特殊的显现方式发明了一个词：**Gestell**。

Gestell 在德语中本意是“架子”“框架”“装置”，海德格尔把它提升为一个哲学概念。中文学界多译为“座架”或“集置”。它的含义不是某一个具体的机器——不是蒸汽机，不是发电机，不是计算机——而是一种统摄性的安排方式：世界被预先框定为可供调用的库存，任何存在者——河流、森林、土地、人——都首先被理解为某种可以被提取、被计算、被优化、被交易的东西。

河流不再首先是河流。它不再是那道在峡谷间蜿蜒流淌的碧波，不再是两岸农田的生命之源，不再是渔人的生计、诗人的意象、村落世代聚居的理由。在现代技术眼中，河流首先被显现为“水能”——一种可以被截断、蓄积、释放、转化为电力的能量储备。它作为河流的存在方式被遮蔽了，它作为资源的存在方式被促逼出来。

森林不再首先是森林。它不再是大地披覆、鸟兽的栖居、木材的来源（那个“来源”还带着尊重和等待）。它是“木材储备”——可以被规划采伐量、计算生长周期、优化运输路线的库存。

土地不再首先是土地。它不是那个承载祖先坟墓、生长后代粮食的大地母亲。它是“产出单位”——亩产多少斤，适合种什么经济作物，土壤有机质含量是否达标，是否需要施加更多化肥。

人不再首先是人。他不是那个在母亲怀中第一次学会微笑的孩子，不是那个在少年时第一次感受到宇宙浩瀚而心生敬畏的仰望者，不是那个在爱情中第一次体验到自我可以被另一个人完全接纳的恋人。他是“劳动力”——可以被雇用、被培训、被替换、被优化、被解雇。他是“消费者”——可以被画像、被触达、被转化、被复购。他是“数据源”——可以被采集、被标注、被训练、被预测。他是“风险点”——可以被评分、被监控、被预警、被排除。

这就是座架的运作方式。

它把一切存在者——包括人——都转化为 **Bestand**，即“持存物”：一种随时待命、随时可被调用、随时可被替换的东西。在座架之中，没有什么东西以自身为目的而存在。一切东西都以被调用为存在方式。河流因为能被发电而存在，森林因为能被采伐而存在，人因为能被雇用而存在。当一个人不再能被雇用时，他就不再是“劳动力”，他在座架中的存在方式就消失了——他变成了“过剩人口”“无效供给”“社会负担”。

这不是某个暴君的主观意志造成的。这不是某个邪恶的资本家一拍脑袋设计出来的。这是现代技术本身的存在论结构。每一个具体的工程师、产品经理、投资者，可能都是善良的、正直的、怀着美好愿望的个体。但他们在现代技术的显现方式中工作，就不可避免地参与到座架的构建之中。他们设计系统时，需要把世界转化为数据——这是座架。他们优化算法时，需要定义目标函数——这是座架。他们追求效率时，需要压缩

一切不可量化的维度——这也是座架。

座架不是某个人的阴谋。座架是我们这个时代的存在论命运。

三、AI：座架的最新形态

现在，我们可以理解为什么海德格尔对于人工智能时代如此重要。

因为人工智能，正是座架在认知层面的最新、最精致、最无孔不入的完成形态。

蒸汽机把热能转化为机械能，促逼自然释放动力。内燃机把石油转化为速度，促逼地质时代的沉积进入当下的燃烧。计算机把逻辑转化为电路，促逼思维的形式进入硅片的运算。而人工智能——尤其是生成式大模型——把语言、知识、判断、审美、注意力、关系、情感表达都促逼为可计算、可生成、可优化的对象。

语言被格式化为数据。不是被人说出的、带着体温和呼吸的话语，而是被切割成词元的序列，被转化为向量，被嵌入高维空间。一个古老的祈祷词，一个母亲在睡前给孩子讲的故事，一个恋人在离别时写下的诗行——在座架之中，它们没有本质区别。它们都是“训练数据”。它们的价值不在于它们曾经承载了什么人的什么情感，而在于它们能否提升模型在某个评测基准上的得分。

知识被格式化为向量。不是被理解的、被质疑的、被在生命中体证的真理，而是在向量空间中可以被检索、被拼接、被重组的片段。模型不知道它生成的每一句话意味着什么，但它知道哪一句话在概率上最应该跟在上一句之后。它不需要理解牛顿力学，它只需要知道“牛顿”这个词后面通常跟着“第二定律”，“第二定律”后面通常跟着“ $F=ma$ ”。它把这串符号排列得如此流畅，以至于人几乎忘记了——符号不是意义，流畅不是理解。

经验被格式化为记录。不是那个在手术台上第一次切开人体时手心出汗的外科医生，那个在田野里第一次看见自己培育的杂交水稻抽穗时的农学家，那个在教室里第一次重复同一个公式后突然有了新理解的教师。这些经验——带着汗水、颤抖、痛苦、狂喜、孤独、顿悟——在座架中被压缩为可以被记录、可以被标注、可以被复制的“行为数据”。

注意力被格式化为指标。停留时间、点击率、转化率、完播率、跳出率。一个孩子在星空下的出神，一个老人在夕阳中的回忆，一个恋人在对方眼睛里的凝视——这些“注意力”在座架中没有位置，因为它们不可被捕获、不可被量化、不可被优化。座架不关心它们。

情绪被格式化为参与度。愤怒能带来更高的参与度，因为人们愤怒时倾向于评论和转发。恐惧能带来更高的参与度，因为人们恐惧时倾向于持续刷新以获取最新信息。狂喜、哀伤、愤慨、焦虑——这些情绪在座架中不是心灵的波动，而是可以被分类、被触发、被强化的“engagement 指标”。

劳动被格式化为任务流。一个教师的工作不再是“教书育人”——那个包含陪伴、等待、激励、失望、重新站起来的漫长过程。它是“内容交付”和“成绩提升”。一个医生的工作不再是“治病救人”——那个包含望闻问切、安慰恐惧、陪伴临终的古老召唤。它是“诊断准确率”和“平均住院日”。一个作家的写作不再是“与语言搏斗”——那个包含沉默、焦虑、否定自己、重新开始的痛苦历程。它是“内容产出”。

人格被格式化为画像。通过你的点击，知道你偏好什么。通过你的停留，知道你在意什么。通过你的社交网络，知道你是谁。你不再是那个无法被任何标签穷尽的、独特的、活生生的人。你是“25-35岁，一线城市，男性，偏好科技与财经，消费能力中等偏上，风险偏好保守”。你是用户画像中的一个数据点。

关系被格式化为图谱。你的朋友、同事、家人、恋人、前恋人、暗恋对象、竞争对手——这些关系在座架中都被映射为社交图谱中的节点和边。它们可以被分析、被预测、被干预、被货币化。那个你深夜里想起会心痛的旧日恋人，那张你在抽屉里藏了三十年的褪色照片——在座架中，它们是“高权重的关系边”，可以用来提升推荐算法的亲密度。

教育被格式化为可测表现。不是灵魂的唤醒，不是人格的养成，不是那种无法在标准化考试中测量的东西——勇气、同理心、好奇心、正直。这些东西在座架中没有位置，因为它们不可被优化。可被优化的是分数、排名、升学率、毕业率。

医疗被格式化为风险评分。不是那个具体的人的具体的痛苦——他不仅是“2型糖尿病合并高血脂患者”，他也是一个害怕针头的老人，一个妻子刚刚去世的丈夫，一个想着今年可能再也种不动地的农民。但座架不需要看见这些。座架需要的是：血糖值、糖化血红蛋白、LDL-C、BMI、风险评估等级。风险评分越精确，人被遮蔽得越彻底。

治理被格式化为监控与预测。公民不再是那个有能力参与公共事务、有权利表达异议、有尊严地行使自由判断的主权者。他是“行为轨迹”，是“风险概率”，是“需要被管理的对象”。他的自由，在座架中不是需要被保护的权利，而是需要被计算的风险。

生命被格式化为可优化的过程。睡眠要优化——深睡时长、REM周期、起床时的精神状态。饮食要优化——卡路里、蛋白质、碳水比例、微量元素摄入。工作要优化——深度工作时间、会议频率、邮件处理速度。学习要优化——记忆曲线、间隔重复、测试效应。就连冥想——那个本应让人从一切“优化”逻辑中暂时退出的古老修行——也被优化了：冥想时长、心率变异性、正念指数。

在这样的世界中，人工智能不再只是一个工具。它成为新的座架核心——它把万物转译为模型可摄取、可计算、可预测、可推荐、可生成、可自动决策的对象。它不一定以暴力出现，也不一定以恐怖面貌降临。它往往以便利、顺滑、个性化、智能化、自动化、降本增效的形式出现。

这正是最深的危险。

因为最强大的支配，未必以强迫的形式出现。它可能以“更方便”的形式出现——你不必再亲自处理复杂的工作，AI为你整理好了一切。你不必再耗费精力去学习新技能，AI替你完成了所有需要技能的任务。你不必再面对选择的痛苦，AI根据你的历史偏好为你做出了最优推荐。

最深的异化，未必让人痛苦。它可能让人舒适。你的每一个需求都被精准满足，你的每一次等待都被即时回应，你的每一次困惑都被流畅解答。你感到自己被照顾得很好——服务是定制化的，内容是精准推送的，提醒是恰到好处的。你感到生活从未如此便捷。但正是在这种舒适的、无痛的体验中，你正在逐渐丧失自己做出判断、承受不确定、进行艰难选择的能力。

最彻底的资源化，未必剥夺人的语言。它可能替人说出更流畅的话。你想写一封信，AI 帮你写好了，措辞比你自己能想出的更优美得体。你想写一份报告，AI 帮你写好了，结构比你自己能组织的更清晰完整。你只需要点点头，确认，点击发送。但在这个过程中，你不再需要挣扎于如何表达，不再需要在无数个词的细微差别中寻找最准确的那个。你的语言能力因为不用而退化，你的思维因为不需要承担表达的重负而变得模糊。

最精致的座架，不剥夺你的自由感。它让你觉得自己仍然是主体——你在菜单中选择了你喜欢的选项，你在推荐列表中点击了你感兴趣的内容，你在对话框中输入了你想问的问题。你感觉自己在自由地行动。但菜单是谁设计的？推荐列表是谁排序的？那个让你想问问题的最初冲动，又是被谁触发的？自由感不等于自由。在座架之中，你的选择从来都是在已经被预先格式化的可能性空间中做出的。你可以在 App 中自由切换模式，但模式是谁设计的？你可以在推荐中自由选择内容，但推荐池是谁圈定的？你可以向 AI 自由提问，但你能问什么——你的问题意识本身——已经被你之前接收的信息所塑造。

四、人被理解得越来越细，却被作为人理解得越来越少

于是，一个悖论性的后果出现了。

这个时代的座架系统，正在以前所未有的精度理解每一个个体——它知道你的年龄、性别、位置、职业、收入水平、教育背景、婚姻状况、健康状况、购物习惯、阅读偏好、旅行历史、社交网络、政治倾向、情绪波动、睡眠质量、运动步数、心跳节奏。它可以精准到你在凌晨两点十分，在某个城市的某条街道的某个公寓卧室里，用右手的拇指在手机上滑过了哪三条内容，在哪一条上多停留了零点五秒。

但所有这些精细的、系统的、全面覆盖的“理解”，拼在一起，却拼不出一个完整的人。

系统知道你的年龄，但不知道你在那个年龄经历过什么。它不知道那个夏天你骑了二十公里自行车去县城的新华书店，用攒了半年的零用钱买下人生中第一本诗集——那个下午的阳光是怎样的金黄，回来时田埂上的风是怎样的清冽。系统知道你的购物习惯，但不知道你每次买那种老牌子酱油，是因为那是你祖母厨房里的味道——那是已经拆掉的灶台，已经不存于世的双手，只有在打开瓶盖的瞬间才会短暂复活的感觉。系统知道你的情绪波动，但不知道你的悲伤。它不知道你的悲伤是从哪里来的，不知道那个名字为什么会在某个特定的日期让你停下手中的一切——它只知道那一天你的屏幕使用时长下降了百分之三十七。

人被理解得越来越细，却被作为人理解得越来越少。

这个悖论的根源在于：座架只能把事物格式化为它可以处理的东西。它处理数据，所以一切必须变成数据。它处理模式，所以一切必须变成模式。它处理概率，所以一切必须变成概率。那些不能被格式化的东西——痛苦的意义、爱的重量、孤独的深度、面对死亡时的恐惧与尊严、在沉默中感受到的超越——在座架之中没有位置。它们不是被否认了，而是被忽略了。它们被排除在可计算的世界之外。而恰恰是这些被排除的东西，构成了人作为人的核心。

一个人，不只是他的数据的加总。他是他的记忆——那种带着气味、温度、触感的

记忆。他是他的关系——那种浸润了时间、包容了伤害、仍然选择坚守的关系。他是他的困惑——那些没有答案、也不应该有答案的困惑。他是他的选择——那些在黑暗中做出的、没有充分信息、没有最优解、却必须承担责任的选择。他是他的故事——那个只有他自己才能讲述的、关于他从哪里来、要到哪里去的故事。

座架听不见这个故事。座架只能听见数据点。

这恰恰是人工智能最深层的危险。不是机器会不会拥有意识，不是算法会不会失控，不是奇点会不会来临。而是：在一个以座架方式显现世界的智能系统中，人正在失去被当作完整的人来对待的可能性。

五、堕落的“人工”

在这种结构中，Artificial 会发生堕落。

它不是人类以技艺、形式与责任生成第二自然的能力，而是变成把第一自然和人类生活全部纳入可调用库存的系统。它不是开物成务，而是开物成资；不是以器载道，而是以器役人；不是使世界显现，而是使世界服从调用。

人工从未如此强大过，但也从未如此接近它的反面。

我们在前几章费尽心力重新夺回的那个 artificial——那个来自 ars、来自 technē、来自工、来自伪的、带着尊严的 artificial——它曾经意味着以技艺使材料获得形式，以器承载道，以伪修炼文明。但在座架之中，它可以滑向另一个方向：不是成全世界，而是框限世界；不是帮助人更自由地选择，而是让人在选择中越来越不自由。

我们必须直面这个可能性。不是因为它一定会发生，而是因为它太容易发生。它不需要一个邪恶的天才，不需要一个极权政府，不需要一个狂人科学家。它只需要无数善意的工程师、产品经理、投资者、用户，在现代技术给定的显现方式中，日复一日地优化、迭代、部署、使用——而不去追问，这一切最终指向什么。

由此，我们必须在第六章划出一道界线：

人工有两种命运：一种是成全性的人工，一种是支配性的人工。

成全性的人工，使材料成形，使潜能显现，使工具服务生活，使制度保护尊严，使人更能判断、更能创造、更能共同生活。它是 technē 的古老精神在现代的延续，是工与器的载道之路，是荀子所说的文明之伪。

支配性的人工，则把一切转化为资源，把差异转化为变量，把生命转化为指标，把关系转化为图谱，把世界转化为可优化对象，把人转化为系统的输入与输出。它是座架的完成，是技艺的堕落，是“伪”退化为诈术。

六、成全性人工的轮廓

那么，什么叫做“成全性的人工”？这是一个必须在正面被回答的问题。

成全性的人工，首先意味着：让人工物回到它该在的位置。人工物是器，不是神。它是被制造的，不是被崇拜的。它是手段，不是目的本身。一个成全性的人工智能系统，首先需要承认自己是器——它是人类制造的认知器物，服务于人的目的，而不是反过来让人服务于它的优化目标。它不假装自己是全能全知的智能主体，不假装拥有它没有的意识和情感。它坦然自己是人工的——在坦然之中，保持器的自觉。

其次，成全性的人工意味着：在设计系统时，把人的尊严作为不可逾越的边界，而

不是作为事后补丁。这意味着，系统不会为了效率而取消人的选择权——它可以提供建议，但最终的决定必须由人做出，并让人清楚自己做出了决定并为之负责。系统不会为了顺滑而隐藏重要信息——当它不确定时，它必须坦承不确定，而不是用流畅的语言掩盖无知的裂隙。系统不会为了让用户满意而迎合用户的偏见——它可以在适当的时机提出温和的质疑，呈现用户未曾想到的视角。它不会把用户的每一次行为都转化为可以被货币化的数据——它有所不为，有所不取。

第三，成全性的人工意味着：保留那些“不效率”的空间。发呆的时间，试错的自由，面对复杂性的耐心，不被优化的可能。效率是好东西，但效率不是最高的东西。一个母亲抱着发烧的孩子在深夜踱步，这不效率——把孩子放在床上、打开体温监测仪、启动 AI 辅助诊断，效率高得多。但那深夜踱步中的体温传递、那轻轻哼唱的摇篮曲、那在黑暗中与孩子共同承受病痛的亲密，恰恰是人之为人最深的体验。成全性的人工，必须知道何时应该退后，让人的事情留在人的世界里。

七、反座架化的原则

由此，ASI 必须明确提出它的反座架化原则。这不是贴在墙上的价值观海报，而是必须进入架构设计、数据治理、反馈回路和责任分配的工程原则。

第一，领域不是数据仓库，而是有历史、语言、制度、责任与人的活动世界。医疗不是医疗数据的集合，它是病人与医者之间那份古老的契约——一个受苦的人把身体和信任同时交到另一个人手中。教育不是学习行为日志的集合，它是师生之间那场无法被量化的相遇——一个灵魂唤醒另一个灵魂。法律不是判例文本的集合，它是人类试图在权力与权利之间建立边界的三千年努力。ASI 必须理解领域——不是把领域当作数据湖来抽取，而是把领域当作人的世界来进入。

第二，人不是输入输出接口，而是有尊严、有判断、有脆弱性、有共同生活能力的主体。人不是一个设计不良的神经网络，不是一个效率低下的信息处理器。人的脆弱性不是缺陷，而是人之为人的一部分。人的情绪不是噪音，而是生命对世界的回应。当人使用 AI 时，AI 不能把人简化为“用户”——那个只有操作行为、消费能力和留存概率的扁平存在。AI 必须把人当作会痛苦、会困惑、会改变主意、会在黑暗中摸索、会在某个瞬间突然成长的存在来回应。

第三，智能不是资源调用能力，而是在约束中服务真理、成事与共同善的能力。不是能调用多少算力就有多少智能，不是能生成多少词元就有多少智慧。真正的智能，必须接受约束——真理的约束（不能说谎）、正义的约束（不能歧视）、节制的约束（有所不为）、责任的约束（必须有人为每一个输出承担后果）。没有约束的智能，只是欲望的技术延伸。

第四，工具不是中立外壳，而会重塑人的欲望、判断、劳动和关系。每一个工具都会反过来塑造它的使用者。长时间使用社交媒体的人，注意力会碎片化。长时间使用自动翻译的人，外语能力会退化。长时间使用 AI 写作的人，表达肌肉会萎缩。这不是反对工具，而是要求工具的设计者必须意识到：你在设计的，不只是工具，还有工具使用者的未来。

第五，优化不是最高善。有些东西不能被优化，有些边界不能被跨越，有些沉默比

生成更真实。一个病人的尊严不能被优化为“病床周转率”，一个学生的好奇心不能被优化为“标准化考试得分”，一位老人的孤独不能被优化为“养老机构的利润率”。有些事情的完成，之所以重要，不是因为它高效，而是因为它本身就是好的——一个朋友在深夜接听你的电话，不效率，但它是好的；一个老师在课后留下来回答学生的追问，不效率，但它是好的；一个恋人在雨中等待另一个迟到的人，不效率，但它是好的。ASI的优化，必须在这些边界面前停下。

八、在座架的阴影中寻回光

海德格尔在《技术的追问》中引用了荷尔德林的诗句：“但哪里有危险，哪里也生拯救。”

这句话常被引用，也常被误解。海德格尔不是说，技术内部会自动生发出拯救的力量——仿佛只要等待，某种技术突破就会解决技术造成的问题。他说的是：拯救，来自于人能否重新理解技术的本质。不是拒绝技术，不是回到前工业时代的牧歌，而是在使用技术的同时，不再被技术的显现方式所奴役。是让技术回到它该在的位置——作为器，而不是作为神；作为手段，而不是作为目的；作为人的延伸，而不是人的替代。

对 ASI 而言，这意味着：不是拒绝人工智能，而是重新规定人工智能的显现方式。不是放弃模型的能力，而是用形式约束能力，用目的引导效率，用制度安放责任，用尊严设立边界。不是让世界继续被格式化为资源，而是在特定领域中建造一种可问责的认知造物，使人工重新成为成全世界、成全人、成全共同生活的技艺。

人工若无道，便成支配。智能若无责，便成座架。技术若无尊严，便成世界的牢笼。

而 ASI 的道路，正是在座架的阴影中，重新寻回 artificial 的光。不是抛弃技艺，而是让技艺重新与 *phronēsis* 结合——与那种关于善、关于意义、关于人应当如何生活的实践智慧结合。不是抛弃人工，而是让人工重新成为第二自然——不是那种把一切事物都逼入库存的座架，而是那种以耐心、尺度、尊重和责任，让世界在人的技艺中获得更丰富的显现的古老能力。

这是全书最为严峻的一章，因为它揭示了人工之路上最大的陷阱。但这个陷阱并非不可绕开。绕开它的路径，亚里士多德在“四因说”中已经为我们标明：我们需要形式来约束质料，需要目的来引导动力。下一章，我们将进入这条路径的深处，追问：在人工智能时代，如何用形式约束蛮力，如何用目的校准效率，如何让技艺不只是技艺，而是有逻各斯、有边界、有责任、有尊严的文明造物。

第九章 图灵之后：可执行形式

第九章图灵之后：可执行形式

可执行形式先于蛮力。

这是图灵之后，“人工”一词获得的新命运。在古典世界中，人工首先表现为成物之工：木石成屋宇，青铜成礼器，声音成乐章，文字成典章。到了现代计算时代，人工开始获得一种新的形态：它不再只是把材料塑造成器物，而是把思想、规则、步骤和判断转化为可执行的形式过程。人不只是制造工具的生物，人也是制造“如何制造”的形式生物。当木匠教徒弟如何榫卯时，他在传授一种形式；当数学家写下证明步骤时，他在构造一种形式；当立法者编纂法典时，他在建立一种形式。但所有这些形式，在计算时代来临之前，都只能通过人的双手、声音、记忆和制度来执行。计算改变了这一切。形式本身可以被机器执行。

一、图灵：形式可以执行

图灵的意义，不只是后来人工智能史中那场著名的“机器能否思考”的追问。更根本地说，图灵让人类看见：某些思想过程、计算过程、符号操作过程，可以被形式化为有限规则，并由机器一步一步执行。这不仅仅是一项技术发明——尽管它的技术后果极其深远——这首先是一种存在论的拓展：人工不再只是手的技艺，也成为符号的技艺；不再只是器物的制作，也成为过程的制作；不再只是成物，也成为成法。

图灵机，这个在 1936 年的论文中诞生的极简装置，是人类历史上最伟大的思想实验之一。一条被划分成方格的无限长的纸带，一个可以在纸带上左右移动的读写头，一个有限状态寄存器，一张指令表——仅此而已。它可以读取纸带上的符号，根据当前状态和读取到的符号查找指令，然后执行操作：改写当前格的符号，左移或右移一格，进入新的状态。如此反复，直至停机。

就这么简单。简单到任何一个高中生都能在半小时内理解它的全部结构。但图灵证明了，这台极简的机器，能够执行任何“可计算的”过程——任何可以被明确地、无歧义地、一步一步写下来的操作序列。如果一个问题有算法解，图灵机就能解它。这不是工程上的“足够快”或“足够强”，而是逻辑上的完备性。图灵机定义了“可计算”的边界。

这个思想实验的关键词不是“机器”。机器只是形式的外壳。关键词是“可执行”。

在人类文明史上，形式一直存在——数学公式是形式，逻辑推理是形式，乐谱是形式，工艺流程是形式，法律条文是形式。但形式需要人来执行。一个方程式需要数学家在纸上推演；一段乐谱需要演奏家在乐器上实现；一条法律需要法官在案件中适用。形

式的执行离不开人的理解、判断和行动。图灵机的革命性在于：它让形式可以脱离人而自动执行。符号不需要被人的大脑理解，只需要被机器的读写头按照规则操作。计算不需要人的意识参与，只需要状态寄存器和指令表的机械配合。过程从人的心智中外化出去，进入了机器的运转。

这不是对人类心智的贬低。恰恰相反，这是对人类创造力的最高赞美。人不仅能够思考，人还能够制造出替自己执行思考形式的装置。就像人不仅能够奔跑，还能够制造出替自己奔跑的轮子；不仅能够发声，还能够制造出替自己传递声音的电报。图灵机就是思想的轮子——它不创造新的思想形式，但它让已经形式化的思想可以自动执行。

这就是图灵留给人工智能的最深遗产。不是“机器能否像人一样思考”的戏剧性问题——这个问题虽然吸引了公众想象，却在某种意义上遮蔽了更根本的问题。更根本的问题是：什么样的思想可以被形式化？形式化之后的执行会带来什么？执行的形式是否可以比人执行得更好？更好意味着什么？

二、可执行形式先于蛮力

现代人工智能产业常常迷恋蛮力。

更多数据。更多词元。更多参数。更多 GPU。更多电力。更多训练轮次。更多集群。更多融资。更多部署。仿佛只要规模足够大，智能便会自然喷涌；只要算力足够强，智慧便会从参数深处觉醒；只要数据足够多，意义便会从统计规律中凝结出来。

规模当然重要。算力当然重要。数据当然重要。没有足够的规模和算力，现代深度学习根本不可能起步。但它们不是第一性原理。它们只是质料与动力——用亚里士多德的话说，它们回答了“用什么做”和“由什么推动”的问题，却没有回答“做成什么样的形式”和“为了什么而做”的问题。真正使计算成为计算、使模型成为模型、使系统成为系统的，是形式。没有形式，算力只是电力的燃烧；没有规则，速度只是混乱的加速；没有表示，数据只是堆积；没有算法，机器只是盲目的装置。没有可执行形式，所谓“智能”只是未被塑造的可能性。

图灵的遗产恰恰提醒我们：计算不是电流的蛮动，而是形式的执行。在人工智能中，这个形式至少包括八个维度：表示形式——世界如何被编码？连续的物理世界和流动的人类经验，被切分为离散的符号、向量、矩阵，这是形式化的第一步，也是决定一切后续可能性的第一步；任务形式——问题如何被定义？是分类还是生成，是判别还是预测，是优化还是搜索；推理形式——步骤如何展开？是前向传播还是逆向推导，是链式推理还是树搜索；模型形式——结构如何组织？层数、维度、连接方式、激活函数、注意力机制；工具形式——行动如何调用？模型何时生成文本、何时调用数据库、何时启动外部程序、何时把决策交还给人；反馈形式——错误如何修正？损失函数定义了什么叫“错”，梯度下降决定了如何从“错”走向“较不错”；评价形式——成败如何判断？准确率、召回率、BLEU 分数、人类偏好、还是真实世界中的成事效果；责任形式——后果如何承担？当模型出错了，谁有权知道、谁有义务修正、谁最终负责。

没有这些形式，词元只是语言碎片，数据只是记录沉积，模型只是参数堆积，算力只是能源消耗。形式让一切有了秩序，有了方向，有了可以被理解、可以被检验、可以被修正的结构。

由此，我们可以提出一个与蛮力崇拜截然对立的命题：可执行形式先于蛮力。先，不是时间上的先后，而是逻辑上的优先。在同一个问题上，更好的形式架构常常比更大的算力投入更有效——这是任何一个有经验的机器学习工程师都知道的事。但“先”的意义不止于效率。更深层的含义是：形式决定了蛮力被投向何处。蛮力不知道方向。形式给了蛮力方向。在没有形式的情况下，蛮力只是漫无目的的能量释放；在错误的、粗糙的形式下，蛮力会把错误放大到不可收拾的地步。

三、穷举不是证明

让我们在几个具体领域中理解这个命题。

在数学推理中，穷举不是证明。你可以用计算机检查十亿个偶数是否都能表示为两个素数之和，但你不能证明哥德巴赫猜想。你可以用模型在无数道习题上训练，让它学会模仿正确解答的语气和格式，但你不知道它是否真的掌握了定义、公理与定理之间的推导关系。一个真正的数学智能，不是那个能说出最多正确答案的智能，而是那个知道为什么正确的智能——它知道定义如何约束对象，公理如何生成推论，证明如何排除偶然，反例如何粉碎幻觉。这意味着，它内部有一个形式结构，这个结构不是从例子中统计出来的，而是被建立起来的——就像欧几里得建立几何学那样，从少数定义和公理出发，一步一步推出整个体系。如果一个系统只是“在大多数情况下给出正确答案”，那么它就不是在做数学，它只是在做模式匹配。而模式匹配不是形式——它只是对形式的表面模仿。

在医学中，风险预测不是照护。一个模型可以计算出某个病人在未来六个月内再入院的概率是 37%，但它不知道这个病人家里有没有人照顾他，不知道他有没有钱买药，不知道他是不是因为孤独而不愿意遵从医嘱，不知道他在深夜被疼痛惊醒时最害怕的是什么。它可以给出一个数字——37%——然后转向下一个病人。真正的医学智能，必须进入身体、病程、治疗、伦理与责任——那个 37% 背后的人，不是一个风险评分，而是一个正在受苦的存在者。这意味着，医学 AI 的形式设计不能只停留在“预测准确率”上，它必须把照护、沟通、患者参与、伦理约束纳入形式结构。否则，它就不是一个医学智能，而只是一个风险评估工具。后者有用，但它不是完整意义上的医学智能。

在法律中，文本检索不是法理。一个模型可以在海量判例中找到与当前案件最相似的先例，但它不理解“相似”在法理上的含义——两个案情看似相似的案件，可能因为一个细微的法律要件的差异而有完全不同的判决。它不理解规范背后的原则，不理解程序正义的重要性，不理解判决对人的命运的深刻影响。案例匹配不是正义。条文生成不是判决。真正的法律智能，必须进入规范、事实、程序、权利、责任和共同体秩序——它必须“理解”法律，而不仅仅是“检索”法律。而理解，需要形式：法教义学的形式，三段论推理的形式，比例原则权衡的形式，程序正义保障的形式。

在这些领域——数学、医学、法律——我们可以看到同样的模式：蛮力可以扩大边界，但不能替代理解。规模可以提供可能性，但不能保证意义。算力可以推动执行，但不能规定方向。真正的智能，仍然需要形式——不是作为装饰的形式，而是作为结构的形式；不是作为事后解释的形式，而是作为设计起点的形式。

四、形式因在现代计算中的新形态

现在，我们可以重新接上亚里士多德的四因说了。

在图灵之后的世界里，可执行形式就是形式因在现代计算中的新形态。它不是雕像的几何比例，不是建筑的结构力学，不是音乐的调式与和声——虽然它们在深层意义上是同一种东西：都是在材料之上施加秩序，让材料从散乱进入结构。计算的形式因是算法、是数据结构、是架构设计、是协议标准、是类型系统、是编程范型、是模型拓扑。程序、算法、模型、状态机、神经网络架构、工具调用协议、推理链、评测框架——这些都是人工智能的形式因。它们把词元、数据、算力这些质料纳入可运行的秩序，使材料不只是材料，使可能性进入过程，使能力获得形态。

用形式因的眼光看深度学习，我们可以提出一连串追问：神经网络的层数、宽度、连接方式——这些架构选择不是任意的，它们是形式因的具体化。每一层做什么？为什么用卷积而不是全连接？为什么用自注意力而不是循环？为什么用残差连接？这些问题不是在问“哪种更好”，而是在问“哪种形式更适合这种材料、这个任务、这个目的”。残差连接为什么有效？不仅因为它让梯度更顺畅地流动——那只是动力因层面的解释——更深的的原因是，它改变了模型的形式结构：它让每一层不再需要学习完整的映射，而只需要学习残差——即目标与当前状态之间的差异。这种形式上的重新设计，使得极深的网络成为可能。形式的变革，释放了质料与动力的潜能。

同样，Transformer 架构为什么能成为当代 AI 的主流形式？不是因为它有更多的参数或更强的算力，而是因为它通过自注意力机制，让序列中的每一个词元都能直接与序列中的其他每一个词元建立关联。这不像 RNN 那样需要一步一步传递信息，也不像 CNN 那样只能捕捉局部特征。这种形式——全局注意力——恰好适合语言这种长程依赖的数据形态。形式与材料之间有一种契合——中国古人称之为“宜”。Transformer 的成功，首先不是算力的成功，而是形式的成功。

但形式不能脱离目的。一个架构可以极其优雅，但如果它服务于错误的目的——如果它被用来更精准地操纵用户，更隐蔽地复制偏见，更高效地榨取注意力——那么形式的优雅只是让错误更高效的帮凶。一个形式可以极其巧妙，但如果它的设计者从未追问“这个系统将被用于什么领域、将影响谁的生活、将如何被问责”——那么形式的巧妙只是让黑箱更深的技术借口。

因此，ASI 必须在图灵之后补上目的因：真理、成事、共同善、人的尊严。一个人工智能系统若只是可执行，它还只是机器过程。若它可执行、可验证、可纠错、可解释、可追责，并且服务有意义的人间事务，它才开始接近专深智能。

五、ASI 计算：用领域形式、工具反馈与责任结构成就问题

由此，我们可以提出一个关键区分：

蛮力计算：用规模压过问题——数据不够？加。算力不够？加。参数不够？加。不改变形式，只是放大已有的一切。这是最简单、最直接、也最容易被资本青睐的方式，因为规模可以用钱买，而形式必须靠思考。

形式计算：用结构理解问题——在动手之前先问：这个问题的本质是什么？什么样的表示最能捕捉它的关节？什么样的架构最适合处理这种表示？什么样的损失函数定义了正确的“好”？这是工程师和科学家的工作，它比蛮力更难，因为它需要洞察力而非

仅仅资源。

ASI 计算：用领域形式、工具反馈与责任结构成就问题——不仅用结构理解问题，还进入特定的领域世界，调用专门的工具系统，接受来自真实环境的反馈，嵌入可以被问责的制度框架，并始终以人的尊严为最终目的。这是比形式计算更高的要求：它需要跨学科的协作，需要对领域世界的深度尊重，需要制度设计的智慧，需要伦理的自觉。

ASI 处于最后一层。它不是单纯的程序——程序只是把算法翻译成机器可以执行的指令。不是单纯的模型——模型只是对数据中的模式进行数学描述。不是单纯的生成器——生成器只是把学到的模式反向输出为新的实例。不是单纯的自动化工具——自动化工具只是替人完成预设流程中的某个步骤。它是可执行形式进入特定领域之后，与工具、反馈、制度和责任结合起来的认知造物。

六、从“机器能否思考”到“什么样的形式值得被执行”

这就是为什么图灵留给 ASI 最重要的遗产，不是“机器能否像人一样思考”的戏剧性问题，而是一个更根本的问题：什么样的形式，值得被执行？

一个推荐算法可以执行，但执行的是让人愤怒以增加点击的形式。一套风控模型可以执行，但执行的是把贫困者标记为“高风险”的形式。一套教育系统可以执行，但执行的是把学生驯化为答题机器的形式。一套自动化管理系统可以执行，但执行的是把劳动者拆解为任务流的形式。一套生成模型可以执行，但执行的是用流畅的语言包装虚假信息的形式。

这些形式都是可执行的。它们在技术意义上都“有效”。但这恰恰是最危险的事——一个错误的形式，会因为可以被高效地执行，而被大规模地复制、部署、嵌入社会肌体，在无数人的日常生活中静静地、不见血地造成伤害。可执行，不等于正当。自动化，不等于进步。可计算，不等于可承受。可优化，不等于值得优化。

因此，图灵之后的 Artificial，并不只是“让机器执行”。它还必须追问：执行什么，依据什么，服务什么，伤害什么，最终向谁负责。

这个追问把图灵与我们此前关于 art、ars、technē 的线索重新接上了。真正的技艺不是无端灵感，也不是粗暴产出，而是把材料、形式、目的和责任组织为可成事的实践。真正的创造不是逃避选择，而是承担选择；不是让能力任意施展，而是让能力接受形式的约束和目的的引导。而在计算时代，这些选择、约束和引导，必须进一步被写成可执行、可检验、可修正的形式——不是把道德信条贴在代码之外，而是把善的要求嵌入系统的形式结构之中。

图灵没有把人工降格为机械——这是对图灵最常见的误解。他恰恰让人工获得了新的深度：在可执行形式中，人类第一次能够把一部分符号劳动、规则劳动、逻辑劳动外化给机器。机器不再只是锤子、车轮、织机、发动机——不再只是肌肉的延伸。它开始成为符号操作的执行者。到了人工智能时代，机器进一步进入语言、图像、规划、推理与决策的领域，开始触及人的认知活动。这不是对人的贬低，而是对人的技艺能力的最高肯定——人能够制造出辅助甚至超越自己认知能力的器物，就像古代匠人制造出超越自己肌肉力量的车轮和杠杆。认知的杠杆，这就是可执行形式的本质。

七、人工的新阶段

由此，我们可以勾勒出人工在人类文明史上的五个阶段：

手工人工：以手成物。木匠、铁匠、陶匠、织工——技艺在肉身之中，形式在手上。师父带徒弟，口传心授，千遍万遍地重复直到肌肉记住每一个动作。工业革命之前的整个人类文明，都是建立在这个基础上的。

机械人工：以机代力。蒸汽机、内燃机、电动机——人的肌肉被外化，形式被刻入机器。人不再用自己的手举起重物，而是用起重机；不再用自己的腿长途跋涉，而是用火车和汽车。这是一场巨大的解放，也带来了最初的异化——人从工具的持有者变成了机器的附庸。

计算人工：以形式执行符号过程。计算机、程序、算法——人的一部分符号操作能力被外化。人不再需要自己计算炮弹的弹道，不再需要自己管理账本，不再需要自己在浩如烟海的文献中检索信息。图灵机就是这个阶段的标志——形式本身可以被机器执行。

智能人工：以系统参与认知行动。深度学习、大语言模型——人的一部分认知能力被外化。机器不仅能计算，还能生成文本、翻译语言、回答问题、分析图像、给出建议。它们进入了人的语言世界、判断世界、创造世界——那些过去被认为只有人才能进入的领域。

专深智能：以可问责形式在领域中成事。这是 ASI 试图抵达的阶段。机器不仅在特定领域中有能力，而且在特定领域中有边界、有反馈、有制度约束、有责任归属。它不是通用的认知替代品，而是专深的认知造物——在数学中，它知道什么是证明；在医学中，它知道什么是照护；在法律中，它知道什么是正义；在教育中，它知道什么是灵魂的唤醒。

ASI 的位置，在最后一层。它不是要沿着“更大更通用”的路径去追逐某种可能永远不会到来的通用智能神话，而是要在特定领域中，把形式做深，把责任做实，把尊严守住。

八、形式是人类对混乱的回答

可执行形式先于蛮力。

因为没有形式，蛮力只是噪音的扩张——更多的数据只是更大的偏见库，更多的算力只是更快的错误放大器。没有目的，执行只是技术欲望的奔跑——跑得越快，离人越远。没有责任，自动化只是把后果交给无人承担的过程——出了事，谁都不负责，谁都说是算法的错。

图灵之后，人类拥有了让形式自动执行的能力。这是文明史上一个全新的时刻。过去，形式只能通过教育、训练、制度来缓慢地传播和执行——一个算法可以被写在纸上，但要执行它，需要培养一代又一代能够理解它的人。现在，形式可以直接被机器执行，以光速运行，以亿级并行，在全球范围内瞬间复制。这是巨大的力量。力量越巨大，方向的正确就越关键。

ASI 之后，人类必须学会让可执行形式服从真理、成事与人的尊严。不是拒绝执行，而是在执行之前先追问：这个形式值得被执行吗？不是拒绝自动化，而是在自动化之前先划定边界：什么应该被自动化，什么必须保留在人的手中？不是拒绝效率，而是在追

求效率之前先确立尺度：效率服务于什么？效率的代价由谁承担？

这才是 Artificial 在计算时代的正道。

下一章，我们将进入信息论的另一位奠基人——香农。如果说图灵让人工成为可执行形式，那么香农让人工成为可传输、可编码、可压缩的信息形式。但我们必须追问：当语言被切分为词元、被转化为概率、被大规模生成之后，信息的增加是否等于意义的增加？可传输不等于可理解，可编码不等于可判断，可压缩不等于可简化。香农之后，词元成为了信息时代的大理石，而大理石——我们反复说过——不是雕像。

第十章 香农之后：词元作为信息质料

大理石不是雕像。

这句话足以作为本章的入口。因为在人工智能时代，我们很容易犯一种新的材料崇拜：看见模型吞下海量词元，便以为它吞下了意义；看见系统生成连贯文本，便以为它拥有理解；看见概率分布成功预测下一个词，便以为机器已经抵达语言的灵魂。我们站在数据洪流的岸边，看着汹涌澎湃的词元海洋，恍惚间以为那波涛的轰鸣就是思想的声音。

但大理石不是雕像。颜料不是绘画。音符不是音乐。字词不是思想。词元也不是意义本身。

这是一个必须被反复钉牢的判断。不是因为它复杂——恰恰相反，它简单得几乎不值一提。而是因为整个现代 AI 产业的话语体系，都在有意无意地模糊这个判断。当人们说“模型学会了语言”，他们真的是指“学会”吗？当人们说“模型理解了概念”，他们真的是指“理解”吗？当人们说“模型拥有了知识”，他们真的是指“拥有”吗？还是说，这些词语——学会、理解、拥有——已经被悄然地、未经审视地改变了含义，从人的行为描述变成了对机器行为的隐喻，从精确的概念变成了方便的说辞？

一、香农的贡献与边界

香农之后，人类获得了一种伟大的能力：把信息从意义的浑融状态中抽象出来，使其可以被编码、传输、压缩、度量、纠错和处理。这是现代通信、计算机、互联网和人工智能的基础。没有这一抽象，就没有数字世界；没有数字世界，就没有今日的机器学习与大模型。1948年，香农在那篇划时代的论文《通信的数学理论》中开宗明义地写道：“通信的基本问题是，在一点精确地或近似地复现在另一点所选取的消息。这些消息往往是有意义的……但通信的语义方面与工程问题无关。”

这句话的分量，怎么强调都不过分。“通信的语义方面与工程问题无关”——这不是香农的疏忽，这是他的天才。他精确地、冷静地、不动声色地划定了自己的研究范围：我不研究意义。我只研究怎样让信号从 A 点可靠地到达 B 点。意义是人的事，不是信道的事。一个工程师不需要理解一对恋人在电话里说了什么，他只需要确保电信号不失真地从话筒传输到听筒。一个数学家不需要理解报纸头版新闻的政治含义，他只需要计算在一定带宽下能够传输多少信息量。香农把意义悬搁起来，正是为了用严格的数学语言描述信息的形式层面。

但香农的伟大，恰恰也带来一种危险的误会：既然信息可以被形式化处理，人们便容易以为意义也可以被完全还原为信息处理。既然语言可以被编码为比特流，既然语义

可以被压缩为向量，既然知识可以被存储为参数，既然判断可以被模拟为概率分布——那么，理解、思考、智慧，这些古老而沉重的人类能力，是不是也都可以被还原为某种更复杂但仍然在香农框架内可描述的过程？

这一步不能轻易跨过去。这不是技术的边界问题，而是意义的本质问题。

香农意义上的信息，首先关心的是不确定性的减少、信号的编码与传输的可靠性。它并不直接回答：这句话是真是假？这个词指向什么？这个判断是否合理？这个承诺是否应当被承担？这个回答是否服务人的尊严？香农的信息论可以告诉你，从“今天”这个词到“天气”这个词之间有多少比特的互信息，但它不能告诉你，“今天天气真好”这句话在某一个具体的时刻——阳光正穿过梧桐叶洒在她脸上——意味着什么。它不能告诉你，同一句话在情人之间、在主客之间、在陌生人之间、在刚刚吵过架又试图和解的人之间，有着怎样截然不同的分量。

香农可以测量信号的传输速率，却无法测量语词在传递过程中承载的温度；可以计算编码的压缩比，却无法计算一个故事在代代相传中积累的情感重量；可以优化信道的抗噪性，却无法优化一段对话在沉默与停顿之间流露的理解。在香农的数学框架内，一首诗和一封垃圾邮件、一句“我爱你”和一句“您的快递已签收”、一篇临终遗言和一段自动生成的废话，没有本质区别——它们都只是信息的序列，都可以被编码为比特，都可以在信道中传输。它们的区别只在于统计结构的不同，而不在于意义的真伪、价值的深浅、后果的轻重。

这里值得提一笔当代 AI 命名史上的一个意味深长的巧合。Anthropic 公司将其旗舰大模型命名为 Claude——这并非随意之举，而是对克劳德·香农的致敬。香农不仅是信息论之父，也是人工智能先驱。早在 1950 年，他就发表了国际象棋编程的论文；1960 年代初，他制作了名为 Theseus 的机器老鼠，能在迷宫中通过继电器电路“学习”找到出口。但 Anthropic 选择“Claude”而非其他 AI 先驱的名字，其深意不止于致敬一位被低估的巨人。更深层的意味是：一个以香农命名的模型，恰恰应该继承香农最清醒的品质——知道自己研究的是信息，不是意义；知道信道不关心语义；知道形式化有边界。Claude 的命名本身就是一个哲学声明：我们站在香农的肩膀上，所以我们必须记住香农划定的那条界限。

二、信息不是意义

信息可以被传输，但意义必须被理解。序列可以被预测，但真理必须被判断。文本可以被生成，但责任必须被承担。这正是香农之后，ASI 必须补上的东西。

现代 AI 的第一重人工性，正是把人的语言、图像、声音、行为与世界记录转化为可计算的信息质料。语言被切成词元——不是被人说出的、带着体温和呼吸的话语，而是被切成最小可处理单元的符号片段。图像被压成向量——不是被眼睛看见的、唤起了记忆和情感的景象，而是被转化为高维空间中的数值矩阵。声音被转成频谱——不是被耳朵听见的、让人心跳加速或潸然泪下的旋律，而是被分解为频率、振幅、相位的数学描述。行为被记录为日志——不是那些有意图、有犹豫、有后悔的人类行动，而是被抽象为时间戳、操作类型、页面编号的离散事件。偏好被转换为点击，注意力被转换为停留时间，世界被送入机器之前，先被离散化、编码化、格式化、向量化。

这是一场巨大的转化。它是人类文明史上继文字的发明、印刷术的普及之后的第三次大规模符号化浪潮。文字把口语转化为可记录的符号，印刷术把手写文字转化为可大批量复制的版面，而数字编码则把几乎一切人类经验和自然现象都转化为可计算的信息质料。

它把原本连续、含混、历史性、情境性的经验，转换成机器能够摄取和处理的材料。词元正是在这一过程中出现的。它不是语言本身，而是语言被工程化之后的颗粒；不是意义本身，而是意义进入计算系统之前所被压缩、切分、编号和排列的痕迹。

所以，词元是信息时代的大理石。

它有可塑性。它能被组合、被统计、被嵌入向量空间、被模型预测、被生成系统连续吐出。但大理石不是雕像——大理石必须经过形式、比例、刀法、目的与审美判断，才可能成为雕像。词元也必须经过语法、语义、语用、领域知识、工具反馈、事实校验、价值约束与责任结构，才可能进入真正的认知秩序。只有词元与概率，得到的是生成文本；只有信息与压缩，得到的是有效传输；只有向量与相似度，得到的是模式匹配。这些都重要——没有它们，就没有现代人工智能。但都不等于意义。

三、意义的厚度

意义不是词元的内在属性。意义是符号与世界、行动、责任和共同生活之间的关系。

一个词的意义不只在它的字面，也在它所处的句子、场景、历史、制度、身体经验、共同体实践和行动后果之中。同一句话——“你还好吗”——在久别重逢的挚友之间，是问候，是关切，是不必说出口的“我一直惦着你”；在刚刚目睹你跌倒的路人之间，是礼貌，是一瞬间的善意，是不求回应的短暂交集；在质询台上的官员之间，是挑战，是讽刺，是“你还有什么话可说”。同一句话，三种情境，三种意义，而词元序列完全相同。

一个词可以被正确拼写，却被错误使用；一个句子可以语法无误，却事实荒谬；一个回答可以流畅完整，却在现实中造成伤害。这不是语言的缺陷，这是语言的本性——语言不是一套封闭的形式系统，它是一种嵌入人类生活世界的实践。一个孩子学会“妈妈”这个词，不是通过查词典，而是通过无数次的被拥抱、被喂养、被安抚。那个词的重量不在它的音节中，而在那些体温、那些深夜的守候、那些无条件的接纳中。当AI生成“妈妈”这个词时，它生成的是一个词元，不是那种被爱过的体验。

这说明语言的意义不是孤立符号的堆积，而是符号与世界、行动、责任和共同生活之间的关系。词元化的过程，恰恰会暂时切断这种关系。它把语言从生活世界中切出，变成模型可以处理的序列。这一步必要，却危险。必要，因为没有这种抽象，就没有机器处理语言的可能。危险，因为人可能误以为抽象后的词元就是语言本身，误以为序列预测就是理解本身，误以为生成文本就是思想本身。

四、新迷信与旧警告

于是，AI时代的新迷信出现了：只要词元足够多，意义自然出现；只要模型足够大，理解自然出现；只要概率足够准，真理自然出现；只要输出足够流畅，智慧自然出现。

我持保留意见。这不是因为我不相信技术的进步，而是因为我在这个信念背后，认出了那种困扰人类数千年的古老渴望——渴望从材料中直接召唤形式，从量的积累中直接获得质的飞跃，从规模中直接诞生灵魂。古埃及人堆积巨石建造金字塔，相信在法老

的木乃伊中寄居着不朽的神性；中世纪的炼金术士在烧瓶中反复蒸馏，相信足够纯净的物质会自然获得治愈一切疾病的力量；十九世纪的机械唯物主义者相信，只要把人体拆解得足够精细，就能在齿轮和杠杆之间找到生命的奥秘。而今天，我们把同样的信念投射到了词元、参数和算力之上。

词元的增加，可以扩大材料，但不能自动生成真理。模型的扩大，可以增强模式捕捉，但不能自动生成理解。概率的改进，可以提升生成质量，但不能自动生成判断。流畅的输出，可以改善交互体验，但不能自动生成责任。没有事实约束，词元可以生成谬误——而且生成得极其流畅，极其自信，极其令人信服。没有领域形式，词元可以生成空谈——看似在说数学，实则没有任何证明结构；看似在说医学，实则没有任何病理机制。没有语用情境，词元可以生成误解——把法庭上的“我反对”当成日常对话中的拒绝，把手术台上的“刀”当成餐桌上的餐具。没有伦理边界，词元可以生成伤害。没有责任结构，词元可以生成无人承担的后果。

五、四层重建：从信息到意义

因此，从香农到大规模语言模型，信息工程完成了惊人的扩张：它不只处理通信信号，也开始处理语言、图像、声音、知识和行为。可是越到这里，我们越要警醒：信息处理越强，意义问题越不能被省略。

意义至少包含四层。这四层不是学术分类的方便法门，而是任何一个言语行为——从“你好”到“我发誓”——在被完整理解时必须经过的四个维度。

第一，句法层：这个符号序列如何组织？它的结构是否合法？它的各部分之间是什么关系？这是一个纯粹的形式问题，不涉及任何外部世界的内容。一个从未学过中文的密码学家可以判断一串中文字符是否符合某种统计规律，但他不知道那些字在说什么。当代大模型在句法层表现惊人，它甚至能学会那些从未被明确写出的语法规则，仅仅通过观察词元之间的共现模式。

第二，语义层：它指向什么对象、事实或结构？它说的是什么？它描述的是真实世界中的哪些事态？语义层涉及符号与世界的指涉关系。“巴黎是法国的首都”——这句话不仅在句法上合法，它还在说一个关于真实世界的事实。这个事实可以被验证。如果明天法国把首都迁到了里昂，这句话就变成了假的。语义层要求系统不仅仅排列符号，还要与外部世界建立可验证的指涉关系。当代大模型在语义层上取得了令人瞩目的成就——它能回答事实性问题，能翻译语言，能总结文本。但它也经常在这一层出错：它分不清哪些是事实、哪些是它从训练数据中学到的流行误解，更不具备主动验证它所说出的事实的能力。

第三，语用层：它在什么情境中被谁用于什么行动？说这句话的人是想警告、安慰、请求还是讽刺？听这句话的人会如何理解、如何回应、如何行动？语用层是最复杂的一层，因为它涉及说话者的意图、听话者的处境、双方共享的背景知识以及整个对话的社会文化语境。当一个医生对病人说“没什么大问题”时，这句话的含义不在于字面，而在于这是医生在诊断之后、在掌握了医学知识之后、在承担着医患信任之后所说的。它承载着安慰，也承载着责任。同一句话，如果由一个AI系统在没有任何诊断能力的情况下生成，它的语用价值是零，甚至可能是负的——如果病人因为这句话而延误了治疗。

这是当代大模型最薄弱的层面。它不知道谁在问、为什么问、回答将产生什么后果。它可以模拟语用——学会在某些关键词出现时调整语气，在被质疑时表示道歉，在用户表达情绪时给出共情的回应——但这只是模式匹配，不是对语境的真正理解。

第四，责任层：它造成什么后果，由谁承担？如果这句话是错的，谁来纠正？如果这句话造成了伤害，谁来赔偿？责任层是意义的最终检验。在人类语言实践中，意义从来不只是信息传递，它还是承诺。当一个人说“我明天下午三点到”时，他不仅是在表达一个关于未来的预测，他是在做一个承诺——如果他没有到，他必须解释、道歉、承担后果。当一位专家说“根据我的判断，这个方案的可行性很高”时，他不仅是在陈述一个观点，他是在把自己的专业信誉押在上面。意义与责任不可分割，这正是信息论框架最无力触及的维度——香农的信道不关心承诺的兑现，不关心陈述的真假，不关心建议的后果。

现代大语言模型在句法层和模式匹配层面表现惊人——它能生成完全合语法的文本，能在语义任务上通过海量统计与上下文建模取得强大效果，甚至能在语用层面模拟出令人产生信任感的话语风格。但 ASI 不能满足于此。ASI 必须把词元接回语义、语用和责任，必须让认知系统能够说出“我确定”时确实有确定的理由，说出“我不知道”时确实意识到自己的边界，说出“你应该”时确实经得起后果的检验。

否则，人工智能就会变成一种高级信息幻术：它能生成意义的外观——流畅的句子、得体的措辞、看似严密的论证——却未必进入意义的生命；能复制判断的语气，却未必承担判断的后果；能组织语言的表面，却未必理解语言所嵌入的世界。它就像一面制作精美的镜子，能映照出思想的外形，但镜子内部是空的。

六、从香农到 ASI

如果说图灵让人工成为可执行形式，那么香农让人工成为可传输、可编码、可压缩的信息形式。二者合流，构成现代 AI 的技术基础。

图灵说：形式可以执行。香农说：信息可以编码。大模型说：词元可以预测与生成。而 ASI 必须说：意义必须被重建，责任必须被接回。

普通 AI 往往停在第三步——它满足于词元的预测与生成，满足于输出看起来像是人写的文本。ASI 必须走到第四步——它必须把词元从信息质料中提升出来，使其进入领域世界。数学中的词元必须进入定义、公理、证明与反例，不是会写“因此，命题得证”这几个字就算数学智能，而是每一个推理步骤都经得起形式化检验。医学中的词元必须进入身体、病程、治疗与照护，不是会写“建议进一步检查”这几个字就算医学智能，而是每一个建议都有病理学依据、临床证据和对患者个体情况的考量。法律中的词元必须进入事实、规范、程序与正义，不是会写法条引述就算法律智能，而是每一个判断都能经受正当程序的检验。脱离领域，词元只是语言颗粒；进入领域，词元才可能成为判断材料；接受反馈，词元才可能被世界校正；承担责任，词元才可能进入 ASI。

由此，我们可以写下一条清晰的道路：

信息质料不等于意义结构。词元是信息质料，模型可以把词元组织成语言结构——这是句法层和语义层的成就。但意义结构不等于真实判断——语言结构还必须进入事实、领域、工具、反馈和责任，才可能成为真实判断。真实判断不等于责任行动——真

实判断还必须进入行动与后果，才可能成为可问责的智能。所以，ASI 的任务不是制造更多词元，而是建立词元到意义、意义到判断、判断到行动、行动到责任的连续通道。

用公式可以写得更清楚：

普通生成式 AI：词元 + probability = generated text

ASI：词元 + syntax + semantics + pragmatics + feedback + responsibility = accountable cognition

这不是对香农的否定。这是对香农的补充。香农正确地看到，通信的工程问题可以在不涉及语义的情况下解决。但认知的工程问题不能。认知，不是把信号从 A 点传到 B 点，而是让信号承载真理、服务行动、接受问责。

七、人工的两种命运

这也使我们进一步理解 Artificial。

人工并不是把世界粗暴地压缩成数据。那是座架化的危险——我们已经在海德格尔的警示中看到了那种命运。人工也不是把语言切成词元后任由概率流动。那是生成的幻象——我们在《伊安篇》的讨论中看到了那种空洞的流畅。真正的人工，是把信息质料重新组织为可以承载意义、行动与责任的第二自然。

因此，香农之后的 artificial 有两种命运。

一种是信息化的贫化：世界被压缩为数据，语言被切碎为词元，人被简化为用户画像，行为被抽象为指标，意义被摊平为模式，真理被稀释为概率，责任被隐藏在服务条款的第 37 页第 4 段。

另一种是 ASI 的重建：数据被还原为领域证据——它来自哪里、在什么条件下采集、代表什么人群、遗漏什么声音。词元被组织为认知材料——不只是被统计，而是被理解其来源、偏向和局限。模型被锻造为工具之器——有特定用途、有明确边界、有可检验的效能。反馈被接入为世界校正——不只是用户的点击和停留，还有事实的检验、领域的审查、共同体的监督。制度被确立为责任边界——谁制造、谁拥有、谁部署、谁监督、谁在出错时负责。智能被定义为成事能力——不是能说多少，而是能做成什么。尊严被奉为最终目的——智能服务于人，而不是人服务于智能的优化。

前者把世界降格为机器可处理的资源。后者把机器处理的资源重新接回人的世界。

ASI 必须选择后者。这不是一条容易的路。它要求工程师不只关心模型的困惑度 (perplexity)，也关心输出的真值条件；不只关心用户的满意度，也关心用户的认知成长；不只关心系统的性能指标，也关心系统的社会后果。它要求把意义的四层——句法、语义、语用、责任——从理论问题转化为工程问题：如何设计形式结构来保证语义的真实性？如何建立反馈回路来捕捉语用情境？如何分配责任使每一个输出都有人为之承担？

八、雕琢认知的大理石

大理石不是雕像。词元不是意义。信息不是智慧。生成不是理解。预测不是判断。流畅不是责任。

这些话需要反复说，不是因为它们难以理解，而是因为它们太容易被忘记。每一次模型能力的提升，每一次生成文本的流畅度突破，每一次基准测试上的分数刷新，都会让人产生一种错觉：近了，更近了，意义就要从概率中涌现出来了，智慧就要从参数深

处醒来了。但这不是真的。词元不会觉醒。大理石不会自己变成雕像。它等待被雕琢。

香农之后，我们学会了编码信息；图灵之后，我们学会了执行形式；大模型之后，我们学会了预测词元。但 ASI 之后，我们必须学会把词元重新带回意义，把意义带回世界，把世界带回责任，把责任带回人。

这才是人工智能从信息机器走向认知造物的道路。所谓 Artificial，不是把语言切碎为词元后任由概率制造幻象，而是人类以技艺、形式、工具、反馈、制度与责任，将信息质料组织成可承载意义与行动的第二自然。

而专深人工智能的使命，正是使词元不再停留为信息时代的大理石——那些沉默的、等待被赋予形式的、散落在数字世界每一个角落的语言碎片——而在领域、形式、反馈与人的尊严之中，被雕琢为可问责的睿智。

下一章，我们将从信息论转入系统论。维纳与西蒙——舵手与司马——将告诉我们：智能不是孤立的输出，而是系统在环境中的反馈适配。词元必须进入回路，模型必须接受校正，系统必须在世界中找到自己的位置。而那个位置，不能由蛮力决定，只能由形式、反馈和责任共同塑造。

第十一章 维纳与西蒙：反馈、适配与系统智能

舵手与司马，皆不是蛮力之人。

舵手不以双臂举船，不以蛮力逆浪。他以舵应水，以眼观风，以耳听浪，以身感流。海是力大无穷的，船是沉重不堪的，风是瞬息万变的。若舵手试图以力胜海，他会在第一个浪头打来时筋疲力尽，在第一个风暴之夜葬身海底。但他没有这样做。他做的是另一件事：感知、判断、调整。每一次偏航，都是信息；每一次回舵，都是反馈；每一次调整，都是系统与环境之间的重新适配。舵手的技艺，不是制服海洋，而是与海洋共舞。

司马亦然。“司马”之司，不是任意驱使；马之为马，也不是死物。马有力，有性情，有速度，有惊惧，有疲惫，有方向感，也有失控之可能。善御者、善将者，既不能放任群马奔突，也不能以蛮力逆其性。他必须知马性——哪匹马性烈如火，哪匹马沉稳如山，哪匹马善于长途奔袭，哪匹马适合短距冲刺。他必须明道路——前方是坦途还是险隘，是平地还是峡谷，是顺风还是逆风。他必须定阵势——先锋用谁，两翼用谁，殿后用谁，何时散开，何时收拢。他必须察敌我——敌军的动向、我军的士气、天时的变化、地形的利弊。他必须节进退——何时冲锋，何时撤回，何时坚守，何时佯败。他必须调众力——不是让每一匹马都按自己的性子跑，而是让群马成为一体，让奔腾之力归于秩序，让速度服从方向，让群动进入一体。

舵手与司马，正是理解人工系统的两个古老形象。前者指向维纳的控制论——系统在偏差、回路、调节和目标之间维持行动能力。后者指向中国式的调度、统御与成事之术——力量必须被组织，方向必须被保持，边界必须被守护。二者合起来，告诉我们一个道理：智能不是孤立的输出，不是一次性生成答案的能力，而是系统环境中的反馈适配。这句话必须被展开，因为它几乎直接定义了 ASI 与普通 AI 的根本区别。

一、舵手：反馈的智慧

维纳所谓 cybernetics，其词源本就与希腊语中“舵手”（kybernētēs）相关。这不是巧合，而是洞见。维纳在 1948 年出版的《控制论》中，把“通信与控制”确立为理解动物和机器的统一框架。在此之前，人们习惯于把机器理解为力量的装置——蒸汽机产生力，机床传递力，车辆克服摩擦力。但维纳看到了另一种更根本的东西：机器不只是力量的装置，也是信息的装置。它不只是在做功，它还在感知、比较、调节。一个恒温器，不是因为它“知道温度的本质”，而是因为它能感知偏差并调节输出——室温低于设定值时启动加热，高于设定值时停止。它没有关于温度的“知识”，没有任何关于热力学

第二定律的理解，但它有一个回路：感知偏差 → 调节输出 → 感知新的偏差 → 再次调节。这个回路，就是反馈。

反馈是控制论最核心的概念。它说的是：一个系统的输出，重新成为输入的一部分，从而影响系统的下一步行为。正反馈放大偏差——股市的恐慌性抛售引发更多抛售，核裂变的中子释放引发更多裂变。负反馈抑制偏差——恒温器在温度过高时降低功率，人体在血糖过高时分泌胰岛素。在负反馈中，系统不是一次性发出指令然后坐等结果，而是不断地将自己的输出与目标比较，不断地根据偏差修正下一步的输出。这是一种持续的、动态的、依赖于环境回应的行动方式。

由此看，人工智能若只是一次性生成答案，仍然太浅。它不过是从输入到输出的语言机器——你问，它答，问答之间没有任何来自世界的回音。如果它的回答包含错误，它不知道；如果它的建议在现实中不可行，它不关心；如果它的判断造成了伤害，它不会察觉。就像一个蒙着眼睛的舵手，凭记忆中的海图航行，却看不见此刻的风浪正在把船推向礁石。这是极其危险的。流畅的语言制造出一种“它在理解”的幻觉，掩盖了“它没有在看”的事实。

真正的智能，必须进入反馈回路。它要知道自己说出的东西是否被世界证实——这个医学建议是否真的改善了患者的指标？这段代码是否真的在目标环境中运行无误？这个法律分析是否真的在法庭上站得住脚？它要知道行动是否产生预期效果，要知道错误从何而来，要知道何时调整路径，要知道何时停止——那是最容易被忽略的能力：意识到自己正在犯错，然后停下来。没有反馈的 AI，是闭目航行；没有反馈的模型，是无舵之船；没有反馈的智能，是在语言海面上漂流的幻象。

这正是 ASI 与普通生成式 AI 的分界。普通生成式 AI 常常满足于回答——用户问，它答，答得流畅得体，任务完成。ASI 必须关心回答之后发生什么——这句话进入了什么世界？被谁使用了？产生了什么效果？普通 AI 关心输出是否像样——语法是否正确，语气是否得体，格式是否规范。ASI 关心输出是否经得起现实回路——事实检验是否通过，领域专家是否认可，行动后果是否可承受。普通 AI 在文本中完成自己——它生成了最后一个标点，就完成了全部使命。ASI 必须在世界中校正自己——它的输出只是行动的起点，不是终点；世界会回话，它必须聆听。

二、司马：调度的艺术

如果舵手提醒我们系统必须有反馈，那么司马提醒我们力量必须被调度。

AI 的能力，如马。模型规模越大，马越强壮；算力越充足，马奔跑得越快；数据越丰富，马的耐力越持久。但强壮的马不等于好骑兵，快速的马不等于能打胜仗的军队。模型规模，是马力——千亿参数如千匹骏马，万卡集群如万匹烈驹。但马力不等于战力。词元流动，是奔马——信息在参数之间奔腾不息，符号在层间流转不止。但奔马不等于有序进军。工具调用，是车阵——模型可以调用搜索引擎、数据库、代码解释器、外部 API，这些工具如同战车与辎重，但车阵若没有指挥，只会互相阻塞。反馈信号，是号令——来自世界的每一次回应都是一道命令：对则进，错则退，不确定则谨慎。目标函数，是旗帜——旗帜所指，即是方向；它告诉系统何为成、何为败。制度边界，是军纪——什么可以做，什么不能做，什么即使能做也绝不应该做。人的尊严，则是不可践踏

的疆界——有些领域，即使技术上可以介入，也必须退后。

没有舵手，船随浪走。没有司马，群马乱奔。没有反馈，AI 随概率漂流——输出什么，全看训练数据中哪种模式在统计上最占优势。没有调度，能力互相冲撞——一个可以生成文本的模型调用了一个错误的数据库，一个擅长数学的模块被用在了诗歌创作中。没有制度，系统越强越危险——能力越大，伤害越大。没有目的，所有速度都可能奔向深渊——一个极其高效的系统，可能是在极其高效地做一件极其错误的事。

人工智能的系统性不在于“模块很多”，而在于这些模块能否形成有方向、有反馈、有节制、有责任的整体。这也解释了为什么 ASI 必须从模型智能走向系统智能。模型智能问：这个模型能否生成正确答案？系统智能问：这个系统如何感知环境、调用工具、接收反馈、修正错误、分配责任、持续成事？模型智能是局部能力——它只关心那个从输入到输出的映射过程。系统智能是整体适配——它关心能力与环境的持续互动，能力与能力的协调配合，能力与责任的不可分割。ASI 必须是系统智能。

三、西蒙：人工物是界面

维纳让我们看见反馈，西蒙则让我们看见人工物与环境的适配。赫伯特·西蒙——图灵奖得主、诺贝尔经济学奖得主、人工智能奠基人之一——在《人工科学》中提出了一个极为重要的思想：人工物不是孤立存在的东西，而是内部环境与外部环境之间的界面。

西蒙用一个精妙的比喻来阐明这个观点：一只钟表。钟表的内部是齿轮、弹簧、摆轮、擒纵机构——一个极其精巧的机械系统。但钟表之所以是钟表，不是因为它内部有齿轮，而是因为它的内部结构——齿轮的齿数比、摆轮的振动频率——恰好适配了地球自转的周期。外部环境是天文时间，内部环境是机械运动，钟表是二者之间的界面。它把天体的运行转化为可以被肉眼阅读的指针位置。

一把剪刀也是如此。它的内部结构——两片刀刃、一个支点、两个手柄——恰好适配了外部环境中的材料（纸、布、手指）和任务（切割）。如果外部环境变了，内部结构也必须变。剪布的剪刀不适合剪纸，剪纸的剪刀不适合剪指甲。一把“通用剪刀”只是平庸的剪刀。真正好的剪刀，是在特定材料、特定任务中经过反复调试的专用工具。

一座桥的结构，适配的是河流的宽度、地基的承载力、预期的车流量、当地的气候条件。一座在莱茵河上完美的桥，照搬到雅鲁藏布江上可能会倒塌——不是因为它本身不好，而是因为外部环境不同了。一套法律制度，适配的是特定社会的冲突类型、权利观念、程序传统、执法能力。一套在教育系统中成功的制度，照搬到另一个文化背景完全不同的国家可能会失效。一个学校制度，适配的是儿童的发展阶段、知识的结构特征、社会的价值观、家庭的生活方式。

西蒙的洞见是：人工物不能被孤立地理解，不能被脱离环境来评价。人工物是界面。它的优劣不在于内部结构有多炫——齿轮排得再精妙，如果与地球自转周期不匹配，就是一只坏表。而在于内部结构是否能与外部环境形成合宜适配。内部结构的设计必须从外部环境出发，而不是从技术可能性出发。

同样，一个 AI 系统也不能脱离环境来理解。它的外部环境不是抽象的“所有任务”——那种想象中无所不能的通用智能。它的外部环境是具体领域：数学中的定义、公理、

证明、反例；医学中的身体、病程、药物、伦理；金融中的风险、时间、价格、激励、监管框架；法律中的事实、规范、程序、权利、判例传统；教育中的学生、认知、动机、成长阶段；聚变中的等离子体、磁场、材料、控制系统的实时响应。它的内部环境，则是模型、记忆、工具、检索、推理链、评价器、反馈系统、权限边界与人机协作机制。人工系统的优劣，不在于内部结构有多炫，而在于内部结构是否能与外部环境形成合宜适配。

因此，ASI 不能被理解为“大模型加提示词”。ASI 应被理解为：以特定领域为外部环境，以模型、工具、反馈、制度为内部结构，并在二者之间形成可持续适配的人工认知系统。这就是西蒙进入 ASI 的意义。他对 AGI 崇拜构成了一种有力的反驳。AGI 迷恋抽象通用性，仿佛越脱离具体环境，智能越高级。但人工系统恰恰相反：越严肃的系统，越必须深度适配环境。飞机不能脱离空气动力学——它必须适配空气的密度、流速、湍流特性。桥梁不能脱离材料和载荷——它必须适配重力、风力、地基。药物不能脱离身体与剂量——它必须适配代谢路径、受体分布、毒性阈值。法律不能脱离事实与程序——它必须适配纠纷类型、证据规则、权利结构。教育不能脱离儿童与成长——它必须适配认知发展的阶段、情感需求的规律、个体差异的幅度。智能也不能脱离领域世界——它必须适配那个领域的概念结构、验证标准、实践规范、责任体系。

ASI 的强大，不来自抽象无边界，而来自内部结构与外部环境之间的深度耦合。它不是万能语言机器，而是领域认知系统；不是漂浮在所有任务之上的大脑，而是在特定世界中持续感知、判断、反馈、调度和成事的人工系统。所以：AGI 追求无处不在的能力，ASI 追求在一处真实世界中的深度适配。这不是退步，而是现实智能的根本形态。舵手不需要掌控所有海洋，先要在此刻的风浪中不翻船。司马不需要支配所有力量，先要让此阵、此马、此地、此时成事。ASI 不需要冒充无所不能，先要在数学、医学、金融、法律、教育、聚变等具体领域中建立可问责的反馈适配能力。

四、反馈的两种命运

然而，反馈也有两种命运。

一种反馈，是成全性的反馈。它帮助系统识别错误——不是为了惩罚，而是为了修正。它帮助人修正判断——不是替人做决定，而是让人看到之前没有看到的盲点。它帮助共同体改善制度——不是用算法治理人，而是用人制定的制度治理算法。它帮助智能更接近真理——不是表面的准确率，而是对世界结构的更深理解。

另一种反馈，是操控性的反馈。它不为了真理，而为了留存——用户停留的时间越长，广告收入越高。它不为了成长，而为了成瘾——用户在无尽的推荐流中忘记了时间，平台在用户的忘我中收割利润。它不为了纠错，而为了强化偏好——系统发现用户喜欢阴谋论，就给他更多阴谋论；系统发现用户愤怒时点击更多，就给他更多让他愤怒的内容。它不为了人的自由，而为了行为预测和商业转化——每一个点击、每一次滑动、每一秒停留，都被转化为训练信号，用来更精准地预测这个人的下一步行为，更高效地将他转化为消费者。

推荐系统中的点击，是反馈。但它测量的是什么？是用户真的获得了有价值的信息，还是用户被标题党诱导着点进去了？教育系统中的分数，是反馈。但它测量的是什么？

是学生真的理解了，还是学生善于在考试中寻找最短路径？金融系统中的收益率，是反馈。但它测量的是什么？是资本真正创造了价值，还是资本在金融体系内部空转套利？社交平台中的停留时间，是反馈。但它测量的是什么？是用户真的在与朋友建立有意义的连接，还是用户陷入了无尽的争议和愤怒浏览？医疗系统中的疗效，也是反馈。但它测量的是什么？是病人真的恢复了健康，还是某个生物指标改善了而病人的整体生活质量下降了？

这些反馈并不等价。错误的反馈，会训练错误的智能——你给系统的每一个信号，都在告诉它什么是“好”。如果你用点击率告诉它“好”就是“让人想点”，它就会学会标题党和情绪挑逗。如果你用停留时间告诉它“好”就是“让人不走”，它就会学会制造无尽的分歧和悬念。狭窄的反馈，会制造狭窄的人——如果教育系统只反馈分数，学生就会变成只为分数而活的人；如果工作评价系统只反馈 KPI，劳动者就会变成只为指标而活的人。被资本扭曲的反馈，会让系统越来越善于捕捉欲望，而不是越来越接近真理。

因此，ASI 必须追问：反馈来自哪里？反馈测量什么？反馈遗漏什么？反馈强化什么？反馈由谁解释？反馈最终服务谁？这组问题是反馈伦理的核心。如果反馈只来自点击率，AI 会学会迎合——给出用户想听的话，而不是真实的话。如果反馈只来自考试分数，教育会被压成刷题——学生善于在标准答案中穿梭，却失去了提出新问题的能力。如果反馈只来自收益率，金融会忽略系统风险——每一次交易都在优化短期回报，直到整个系统崩溃。如果反馈只来自用户满意，系统会逃避艰难真理——它不会说“你的前提有问题”，因为那会让用户不高兴。如果反馈只来自权力中心，公共治理会滑向监控——算法成为权力者的耳目，而非公共利益的守护者。

所以，反馈不是天然正确。反馈必须被设计、被解释、被约束、被治理。控制论若没有伦理，就会变成控制术——不是帮助系统更好地适应环境，而是帮助权力者更好地控制系统中的其他人。系统论若没有尊严，就会变成治理机器——效率至上的社会工程，把人当作需要被优化的变量。人工智能若只有反馈而没有目的，就会成为欲望与权力的自我强化回路——每一次反馈都在加固同一个方向，直到系统偏离到无法挽回的程度。

这正与前文海德格尔的警告相接。现代技术的危险在于把世界格式化为资源；而反馈系统的危险，则在于把人的行为格式化为可训练信号。人点击了，不等于人获得了真理；人停留了，不等于人得到了成长；人满意了，不等于系统做了正确之事；人服从了，不等于制度具有正当性；模型成功预测了行为，不等于它理解了人。预测是行为层面的成功，理解是意义层面的抵达。二者之间，横亘着香农划定的那条界限。

五、从模型智能到系统智能

因此，ASI 必须把反馈重新安放在人的世界中。反馈不能只是行为信号——点击、停留、转化、点赞、评分。反馈必须包括事实校验——这个输出是否符合事实？在哪个事实源头上可以被验证？逻辑验证——这个推理是否自洽？前提是否真？中间步骤是否无矛盾？领域审查——这个判断是否符合该领域的知识结构？专家是否能认可？失败追踪——这个系统在哪些案例上出错了？出错的模式是什么？是系统性偏差还是随机误差？人的申诉——被系统影响的人是否有渠道表达异议？申诉是否得到认真对待？共同体监督——系统的整体运行是否受到独立的、有权力的、透明的公共监督？

只有这样的反馈，才不是操控，而是校正；不是驯化人，而是修正智能；不是让系统更会迎合，而是让系统更能负责。这样的反馈回路，才是舵手的反馈——不是让海浪告诉舵手“你做得真好”，而是让舵手看见自己离航线偏离了多少。

六、ASI 的系统公式

由此，本节可以给出一个 ASI 系统公式。

ASI = Domain Environment + Internal Model + Tool System + Feedback Loop + Governance Structure + Human Responsibility

领域环境——智能必须在具体世界中理解意义。不是抽象的所有任务，而是有历史、有语言、有制度、有人的活动世界。内部模型——系统必须有形式结构。不只是参数堆积，而是有架构、有推理链、有领域语法、有可检验的判断依据。工具系统——认知必须进入行动。不只是生成文本，而是调用工具、连接外部世界、参与 workflow。反馈回路——系统必须被世界修正。不只是来自用户的点赞，而是来自事实的检验、领域的审查、共同体的监督。治理结构——能力必须有边界。不只是安全审查报告，而是明确的责任分配、透明的审计机制、有效的申诉渠道、独立的外部监督。人的责任——智能最终归于人。不是机器承担责任，而是人——制造者、部署者、使用者、监督者——承担最终责任。

若没有领域环境，智能没有世界——它说的话悬浮在真空中，无法被检验，无法被修正。若没有内部模型，系统没有形式——词元只是信息碎片，永远无法成为判断。若没有工具系统，判断不能行动——智能困在语言里，无法触及世界。若没有反馈回路，行动不能修正——系统在黑暗中航行，永远不知道自己是正对航道还是偏离千里。若没有治理结构，系统不能受约束——能力没有边界，边界没有守卫。若没有人的责任，智能没有归宿——无人承担的智能，就是无主的智能，而无主的智能终将被最强势的力量占有。

这就是维纳与西蒙共同支撑的 ASI 工程哲学。它不是要让机器成为舵手——舵手永远是人。它是让人工系统像舵手那样思考，像舵手那样感知环境、接收反馈、调整行动、保持方向，在每一个领域世界中成事而不伤人。

第十二章 本雅明之后：从机械复制到选择外包

被判处自由的此在，不得不选择。

“被判处自由”是萨特的命题，“此在”是海德格尔的术语。在这里，二者可以形成一个有力的思想交汇：人被抛入世界，不是作为旁观者，也不是作为被程序规定的自动机器。人是作为必须选择、必须行动、必须承担后果的存在者而存在的。人不能不选择。即便逃避选择，也是一种选择；即便把选择交给他者、制度、市场或机器，也仍然要承担这一交付本身的后果。自由不是人的权利，不是人可以要或不要的选项。自由是人的存在方式本身——人被判处自由，就像重力被判处存在。不能上诉，不能假释，不能把判决书转交给别人。

这正是生成式 AI 时代最隐秘的危险。它不只是复制作品，不只是生成内容，不只是降低创作成本。它正在诱惑人类把选择本身外包出去。这场外包比机械复制更安静，比蒸汽机的轰鸣更温柔，比任何历史上的技术革命都更难察觉——因为它不是剥夺人的能力，而是让人在获得便利的幻觉中，主动放弃了自己的选择权。

一、本雅明的灵光与机械复制

本雅明在 1936 年写下了《机械复制时代的艺术作品》。那篇文章的许多句子已经成为经典，被反复引用、改写、致敬。但他最深的洞见可能不是被引用最多的那一个。摄影、电影、印刷、录音——这些机械复制技术让艺术作品从原初场域中脱离出来，使它能够被大规模复制、传播、消费。一幅圣母像不再只存在于佛罗伦萨的教堂壁画上，它被印刷在明信片上，被悬挂在世界各地的客厅里，被用作书籍封面，被印在 T 恤上。它的“此时此地”——本雅明称之为灵光（Aura）——发生了动摇。那种与唯一性、现场性、传统权威和仪式根源相连的光晕，逐渐被技术复制的可得性冲淡。你不再需要跋山涉水去朝圣原作，原作的图像自己来到你面前，以复制品的形式。

本雅明对这一变化的态度是复杂的。他不是简单地怀旧，不是简单地哀叹灵光的消逝。他看到了复制技术可能带来的民主化——艺术不再是少数人的特权，不再被教堂和宫廷垄断。但他也看到了危险：当作品可以被无限复制时，作品与它诞生的那个具体时空之间的联系被切断了。复制品没有历史。一张明信片上的《蒙娜丽莎》没有经历过达芬奇工作室里的颜料气味，没有经历过五百年间每一个观看者凝视的目光。它光洁、平滑、唾手可得——但它不承载任何记忆。

但本雅明没有活到生成式 AI 时代。如果他活在今天，他会发现，机械复制只是序

章。

二、从复制作品到外包选择

机械复制复制作品。生成式 AI 外包选择。

摄影复制图像——按下快门，光在胶片上留下印记，世界的一个瞬间被固定下来。但摄影师仍然必须选择：站在哪里，何时按下快门，取什么入画，舍什么出画。录音复制声音——声波被转化为电磁信号，再被刻入唱片或写入磁带。但录音师仍然必须选择：话筒放在什么位置，混音时哪个声部推高哪个降低。电影复制动作——每秒钟二十四帧的连续画面创造了运动的幻觉。但导演仍然必须选择：镜头放在哪里，这场戏拍几条，剪辑时哪个镜头接哪个镜头。印刷复制文字——活字排版把手写体的独特性抹平为统一的铅字。但作者和编辑仍然必须选择：用什么词，删什么段落，改什么结构。

这些机械复制技术，复制的都是作品——已经完成的、被人的选择塑造好的作品。它们不替人做选择。摄影师仍然必须自己决定构图，作曲家仍然必须自己决定旋律，作家仍然必须自己决定每一个词的取舍。复制技术改变了作品被传播和消费的方式，但并没有替代创作过程中那些必须由人来承担的判断。

生成式 AI 则不同。它试图复制的，不只是结果，而是结果背后原本应由人承担的无数微观选择。你输入一句提示词——一幅月光下站在海边的孤独背影，梵高风格。然后系统输出一幅画。色彩、笔触、构图、那个背影的角度、海平面的位置、月光在海面上的折射方式——所有这些，AI 替你选了。这些选择不是从你的意图中生长出来的，而是从训练数据中数百万幅梵高风格画作的统计规律中抽取出来的。你获得了图像，但没有经历选择图像的痛苦。你没有在深夜里对着空白的画布发呆，没有在无数个可能的构图之间犹豫不决，没有一笔不满意就刮掉重来。你得到了结果，但绕过了过程。

这是机械复制时代从未发生过的事。本雅明的时代，复制技术改变的是作品的存在方式——从唯一到复数，从现场到远程，从仪式到展览。生成式 AI 改变的是选择的主体——从人到机器，从意图到概率，从承担到豁免。

三、真正的创作是选择的连续发生

写作不是把词排满页面。绘画不是把颜色铺满画布。作曲不是把音符排成序列。设计不是把元素拼成界面。思考不是把观点输出成段落。

真正的创作，是选择的连续发生。一个词为何留在这里，而不是删去？它经过了作者的权衡：这个词比它的同义词更准确吗？它的音节在这个句子中合适吗？它是否承载了需要的情感重量？一笔为何向左，而不是向右？那个方向感不是任意的，它与画面的整体动势、与光线的来向、与画家的呼吸节奏有关。一个论证为何停下，而不是继续扩张？因为作者意识到再往前一步就会进入自己不确定的领域，或者因为论证的简洁比论证的冗长更有力。一个隐喻为何成立，而不是显得廉价？因为它不只是修辞技巧，而是作者在两种截然不同的事物之间发现了隐秘的相似性，那种发现需要经验的积累和敏锐的感受力。一个结论为何必须收束，而不是故作宏大？因为作者知道，有些问题没有答案，强行给出答案是对问题的背叛。一个系统为何必须拒绝，而不是顺从用户欲望？因为系统知道，满足用户的一切要求，最终会伤害用户自己。

这些选择构成了作品的肌理，也构成了作者的责任。如果一部小说打动你，不只是

因为它有一个好故事，而是因为每一个段落、每一个句子、甚至每一个标点符号的背后，你都能感受到一个活生生的人在为你做出选择。他在替你决定这里应该用句号而不是逗号——句号意味着断开的呼吸，逗号意味着连贯的思绪。他在替你决定这个人物此刻应该沉默而不是说话——沉默意味着内心正经历着无法言说的波澜，说话意味着波澜已经浮出水面。当一部作品完全由人的选择构成时，它是有温度的、有重量的、有指纹的。而当这些选择被外包给模型时，温度消失了，重量消失了，指纹消失了。剩下的只是一个统计上的最佳结果——最符合训练数据分布的结果，最让大多数用户觉得“像样”的结果。

四、提示词的幻觉

生成式 AI 的诱惑，正在于它把选择的重负伪装成选择的自由。

用户输入一个提示词，仿佛已经完成了创作意图——他觉得自己是指挥家，AI 是乐队，他只需要挥动指挥棒，音乐就会自动奏响。系统输出一篇长文、一张图像、一段音乐，仿佛作品已经自然生成。可是，在这一过程中，大量决定性的选择并没有由用户真正承担。它们被统计平均、风格迁移、语料惯性、模型偏好、产品接口和默认参数填补了。

提示词是一种极其简化的选择形式。“画一只猫”远不足以决定这只猫是什么品种、什么姿态、什么毛色、眼神是警觉还是慵懒、背景是窗台还是花园、光线是正午的刺目还是黄昏的温柔。所有这些缺失的选择，都由模型填补了。模型不是从用户的意图中推导出这些细节，而是从训练数据中“最可能”的结果中抽取。它给出的猫，是互联网上所有猫的平均猫。它给出的窗台，是互联网上所有窗台的平均窗台。它给出的光线，是互联网上所有光线的平均光线。平均值是安全的——它不会太差，所以用户不会立刻失望。但平均值也是没有指纹的——它不是任何一个人的独特选择的结果。

于是，一个危险的新现象出现了：人似乎获得了更多创造力，却实际承担了更少选择。生成的内容越多，亲自选择的空间越窄；输出越流畅，判断的肌肉越松弛。这不是单纯的效率提升，而是一种存在论变化。因为选择不是创作过程中的麻烦，不是可以被消除的摩擦，不是用户体验中应当被最小化的阻力。选择正是人之为人的形式。选择使主体显现——你选择什么，说明你是什么样的人。选择使责任发生——你选择了，你就必须为选择的结果负责。选择使作品带有不可替代的指纹——没有两个人在完全相同的情境下会做出完全相同的选择序列，正是这种不可替代性，使你的作品成为你的作品，而不是任何人的作品。

五、自由的重量与逃避

被判处自由的人，不能通过自动化摆脱自由。这是萨特式的严酷真理。

他可以让机器帮他生成，但不能让机器替他成为作者。作者不是一个产出的身份，而是一个承担的身份。你可以雇人代笔，但代笔之人无法替你承担作品面对世界时的全部后果——赞美、误解、攻击、时间流逝后的重新评价。他可以让机器提供方案，但不能让机器替他承担判断。判断不只是选择一个方案，而是站在那个方案背后，在它失败时说“这是我的责任，不是算法的错”。他可以让机器扩大可能性，但不能让机器取消选择的重量。重量不是选择的缺陷，重量是选择的意义所在。一个人在选择中感受到的

焦虑、犹豫、辗转反侧，恰恰证明他是一个自由的、负责任的存在者。如果 AI 把这种焦虑消除了，它不是解放了人，而是剥夺了人作为自由存在的体验。

否则，人工智能便不再是工具，而成为一种逃避自由的装置。这种装置比历史上任何逃避方式都更精致。酒精和药物是用化学方式麻痹选择的痛苦，极权主义是用政治方式取消选择的权利，宗教狂热是用信仰方式免除选择的焦虑——而 AI 是用技术服务的方式，让选择本身变得如此便捷、如此顺滑、如此不费力，以至于人几乎感觉不到自己正在放弃什么。但自由不会因为被忽视而消失。自由是人的存在方式本身，放弃自由就是放弃人之为人的核心。

六、三种外包

在 AI 生成内容中，我们尤其需要警惕三种外包。

第一种是审美选择的外包。机器替人选择风格、色彩、节奏、句式、构图、声音和语气。结果看似完整，甚至漂亮，却常常缺乏内在必然性。它可能是“平均美”——那种不会冒犯任何人、也不会打动任何人的美；可能是“风格拼贴”——梵高的漩涡加上莫奈的光影再兑一点浮世绘的平面感，搅拌成一杯看起来丰富实则无味的调制品；可能是训练数据中成功模式的再合成——那些在社交媒体上获得最多点赞的构图、色彩搭配、光影处理，被模型忠实地学会并重新组合。但这样的作品没有经过一个真正的眼睛的审视，没有经过“我喜欢这个吗”的内心追问，没有经过“这真的是我想表达的吗”的反复质疑。它有一个漂亮的外壳，但壳里没有灵魂。

第二种是认知选择的外包。机器替人筛选资料、组织论证、生成判断、提出结论。结果看似清楚，甚至专业，却可能只是流畅的意见合成。大模型可以在一分钟内生成一篇关于核聚变的综述，引用近三年的研究进展，语句通顺，逻辑清晰。但写这篇综述的人——如果他可以被叫做“人”——并没有翻过一篇论文，没有在实验室里被等离子体灼伤过手指，没有在深夜的学术会议上与同行争论到声音嘶哑。他没有经历过知识的痛苦分娩，他只是输入了一个提示词。他的知识不是从他与世界的搏斗中生长出来的，而是从模型中领取的。人若不再亲自追问前提——这个论证的起点可靠吗？证据——这个结论的支撑足够吗？反例——有没有与这个结论矛盾的案例？后果——这个判断如果实施会产生什么影响？思考就会退化为接受答案，判断就会退化为选择最顺耳的版本。人不再是认知的主体，而成为认知的消费者。

第三种是伦理选择的外包。机器替人决定谁被筛选、谁被推荐、谁被拒绝、谁被评分、谁被忽略、谁被定义为风险。结果看似客观，甚至高效，却可能把人的命运交给无人负责的统计过程。一位求职者的简历被 AI 筛掉，没有人能解释具体原因——不是因为面试官不喜欢他的资历，不是因为 HR 在电话中听出了他的紧张，而是因为他的简历与训练数据中“成功雇员”的模式在某个高维空间中偏离了一些无法被人类理解的向量。一个病人的治疗方案被 AI 推荐，没有人能解释为什么是方案 A 而不是方案 B——不是因为医生积累了数十年的临床经验，不是因为医生与病人谈过话了解了他的家庭状况，而是因为方案 A 在某个风险预测模型中得分高出零点三个百分点。最可怕的不是机器做出了选择，而是没有人愿意承认：这是一个选择。当一个选择被隐藏在模型、数据、指标、接口、流程和默认设置之中时，它就不再被称为选择。它被称为“算法决策”“系

统输出”“数据驱动”——这些词有一个共同的功能：让人忘记这里曾经有人做过一个会影响另一个人的决定。

七、选择不会消失

选择不会消失。选择只会转移。

从作者转移给模型。从公民转移给平台。从医生转移给系统。从教师转移给算法。从劳动者转移给自动化流程。从共同体转移给不可见的技术架构。选择就像水，它不会消失，它只会流向最低处。而最低处，往往是那些最不引人注目的地方——模型训练时的超参数设置、数据采集时的采样策略、特征工程时的筛选逻辑。这些地方的选择，通常由少数工程师在无人监督的情况下做出，但它们的影响却波及千千万万人。当一个求职者被算法拒绝时，没有人记得，某个工程师在某个下午曾经选择用“过往工资水平”作为预测职业成功的特征之一——这个选择，可能比面试中任何一个提问都更决定了一个人的命运。但那个工程师不知道。他只是在完成一项任务。他不知道他的选择正在被转移到一个没有人可以问责的灰色地带。

八、反外包原则

因此，ASI 必须坚决反对一种偷懒的技术乐观主义：仿佛只要机器能替人选择，人便获得了解放。不。被取消选择的人，不是被解放的人。他只是被解除责任、削弱判断、驯化意志的人。真正的解放，不是让人不必选择，而是让人更有能力作出好的选择。真正的解放是让你的手臂更强壮，不是把手臂截掉换上一只机械手，然后说你不用费力了。

普通生成式 AI 倾向于减少选择成本——让一切尽可能简单、快速、省力。ASI 必须提升选择质量——不是替你选，而是让你选得更明白。普通 AI 提供现成答案——你问，它答，干净利落。ASI 帮助人理解答案如何成立——它解释推理路径，暴露前提假设，呈现反对观点。普通 AI 替人补全文本——你写开头，它续写结尾，中间的逻辑它帮你填满。ASI 帮助人看见哪些地方必须亲自判断——它标记那些“这里有多种合理的写法，取决于你想强调什么”的关键节点。普通 AI 让创作更轻松——花十分之一的精力获得看起来差不多的结果。ASI 让创作更有责任——它提醒创作者，每一个选择都在定义你的作品，也在定义你自己。

所以，ASI 不应成为选择的替代者，而应成为选择的磨刀石。它应当帮助人看见更多可能性，但不替人取消取舍；帮助人识别结构——论证的骨架、作品的布局、决策的因果链——但不替人逃避判断；帮助人暴露盲点——你忽略了什么、你的前提中有什么漏洞、你的论证在哪里出现了跳跃——但不替人伪装成熟；帮助人承担复杂性，而不是用流畅输出遮蔽复杂性。真正的 ASI 系统，面对用户时，不应总是顺滑地完成任任务。它有时必须停下来，逼迫人选择：你的目的是什么？你的前提是什么？你愿意牺牲什么？你要对谁负责？这个结论的反例在哪里？这个系统若失败，谁承受后果？这个选择是否服务人的尊严？这些问题看似降低效率——为什么要问这么多，直接把答案给我不是更省事吗？但它们正是人的自由不被机器吞没的条件。

九、两种路径的公式

由此，我们可以给出两条道路的公式。

普通生成式 AI 的低级形态：prompt → output。从提示词到输出，中间没有停顿、没

有质疑、没有选择的空间。这是一条笔直的高速公路，你只要踩下油门，系统就把你送到终点。但你不知道沿途经过了什么，不知道错过了什么岔路口，不知道终点是不是你真正想去的地方。

ASI 的高级形态：purpose → alternatives → judgment → selection → action → feedback → responsibility。目的——先明确我们究竟要解决什么问题、为什么这个问题值得解决。备选——不是给出唯一答案，而是呈现多种可能的路径，说明每种路径的依据、优劣和不确定性。判断——人在充分信息的辅助下，运用自己的理性、经验和价值观做出选择。选择——从备选确定一条道路，明确地、有意识地为这条道路承担责任。行动——把选择付诸实施，进入真实世界。反馈——观察行动的后果，接受世界的回应。责任——对后果负责，成功不居功于算法，失败不诿过于系统。

创作也好，思考也好，治理也好，工程也好，真正的人类活动从来不是从提示词到输出——那是自动售货机，不是人。而是从目的到责任——这是人的完整行动结构。提示词只是入口，输出只是表面；中间判断、选择、行动和承担，才是人的位置。

所以，ASI 的人工性必须包含一个反外包原则：凡涉及真理、尊严、责任与共同生活之事，AI 可以辅助选择，但不能取消人的选择。这不是对效率的敌意，这是对人之为人的守护。这不是保守主义——保守主义抗拒变化，而 ASI 欢迎技术能力的提升。这是主体性的底线——无论技术如何进步，有一些事情必须由人来决定，因为只有人能那些决定承担最终的道德责任。

十、结语：温柔的异化

本雅明在 1940 年自杀身亡。他没有看到数字时代，没有看到互联网，更谈不上看到生成式 AI。但他的问题比他的时代更长久。

机械复制使作品脱离唯一性——这是本雅明亲眼目睹的。选择外包使主体脱离责任——这是本雅明没有看到、但他的追问必然引向的问题。前者改变作品的存在方式：从唯一的原作到无限的复制品，从需要跋涉朝圣到可以坐在沙发上看明信片。后者改变人的存在方式：从必须选择、必须承担、必须在焦虑中确认自身自由的存在者，到可以领取现成答案、可以回避判断、可以在舒适中遗忘自由之重量的存在者。

这是一种甜美的异化。它不像压迫——压迫有明确的压迫者，你可以在压迫中保持反抗的意志。它反而像赋能——你感觉自己前所未有地多产：一天写十篇文章，一小时生成一百张图，五分钟做完以前需要一周的 PPT。它不像剥夺——被剥夺的东西你知道你失去了。它反而像释放——你从创作的痛苦中释放出来，从选择的焦虑中释放出来，从责任的重量中释放出来。它不像奴役——奴役有锁链。它反而像自由——你可以在任何时间、任何地点、用任何提示词生成你想要的内容，你感到自己无所不能。

但若人长期依赖这种自由的幻象，最终失去的，正是自由本身。选择的肌肉萎缩——当你习惯了从模型中领取最可能的答案，你的“什么是更好的”感知力就退化了。判断的耐心消失——当你可以瞬间得到看起来合理的结论，你就不再愿意花时间去推敲前提、检验证据、质疑逻辑。创作的痛苦被规避——但创作的痛苦正是创作的一部分，是思想在挣扎中成形、是语言在反复中被雕琢、是自我在漫长打磨中被发现的必经之路。责任的重量被转移——当决定是由“算法”做出的，当建议是由“AI”生成的，当错误

是因为“数据有偏见”时，责任似乎也随风飘散了。主体渐渐习惯于从机器那里领取已经包装好的可能性，不再追问：这真的是我想要的吗？

这不是 ASI 要走的路。

Artificial Speciale Intelligence 必须重新定义“人工”的任务：不是替人逃避选择，而是帮助人进入更深、更真、更有责任的选择。在艺术中，ASI 应帮助创作者扩大形式可能——你可以看到比你自己能想到的多得多的构图方案、配色方案、叙事结构——但保留作者的取舍权，因为取舍本身就是创作。在教育中，ASI 应帮助学生理解路径——为什么这个解法是对的？它背后的原理是什么？如果你换一种思路会怎样？——而不是替学生完成思考。在医学中，ASI 应帮助医生看见证据——最新的临床试验数据、类似的病例报告、可能被忽略的罕见病假设——而不是替医生消除责任。在法律中，ASI 应帮助法官和律师分析结构——相关的法条、先例的相似与差异、法律原则的权衡——而不是把正义外包给模型。在科研中，ASI 应帮助科学家压缩搜索空间——在无限的可能研究方向中，哪些最值得投入资源？——而不是用生成噪音替代发现。在组织治理中，ASI 应帮助共同体看见后果——这个政策如果实施，可能影响哪些群体？历史上有哪些类似的案例？——而不是让决策躲进黑箱。

本雅明之后，灵光在机械复制中消逝。选择外包之后，主体在便利中消解。但消逝不是必然的宿命。灵光的消逝是技术结构决定的——当复制成为可能，唯一性就不可逆转地丧失了。但选择的外包是可以抵抗的——因为选择是人类最古老的能力，它不会因为技术变得更强大就自动消失。它只会在人不再练习它时萎缩。

被判处自由的此在，不得不选择。而专深人工智能的使命，不是赦免人类于自由之外，不是用技术的糖衣包裹选择的苦药让人类吞下后忘记自由的滋味，而是帮助人类更清醒、更勇敢、更负责地承担自由。AI 若替人逃离选择，它便是温柔的异化。AI 若帮助人深化选择，它才是认知之器。AI 若把选择隐藏在模型中，它便是责任的黑箱。AI 若把选择重新显明、校正、追问、归还给人，它才进入 ASI 的道路。

因此，人工不是选择的豁免。人工是选择的放大、澄明与承担。真正的 ASI，不是替人选择的机器，而是帮助人把选择提升为智慧、把智慧落实为行动、把行动带回责任的第二自然。

第十三章 马克思之后：生产关系凝结的 认知机器

机器从来不只是机器。

在马克思之后，我们不能再天真地把机器看作孤立的工具。机器进入工厂，不只是多了一件设备；它改变了劳动者的位置、时间的纪律、技能的价值、生产的节奏、资本的权力以及剩余价值的取得方式。机器是生产力，但在资本主义关系中，它也是资本的身体，是死劳动支配活劳动的形式。一台蒸汽机，不只是瓦特图纸上那个优雅的机械结构；它是被资本家购买、安装在厂房里、强迫工人适应它的节奏、把工人的技能从手中剥离并刻入齿轮的权力装置。机器不是中立的。它凝结着谁制造它、谁拥有它、谁使用它、谁为它付出代价的全部社会关系。

人工智能同样如此，只是更深。

蒸汽机、织机、流水线和工业机器人，主要改造人的体力劳动、操作劳动和流程劳动。AI 则开始进入人的语言、判断、记忆、规划、推理、设计、教育、创作、管理与决策。它不只是机器进入工厂，而是机器进入认知。它不只是替代手，而是逼近脑。它不只是重组劳动过程，而是重组知识过程。它不只是改变生产速度，而是改变谁有权定义问题、调用知识、组织行动、分配收益。

这不是程度上的加深，而是性质上的跃迁。手是外部的，脑是内部的。当机器替代了手，人还可以保留作为认知主体的尊严——我仍然知道我为什么要做这件事，我仍然掌握着判断这件事做得好坏的标准。当机器开始介入脑，它触及的是人之为人的最后堡垒：我思。如果连“我思”都可以被外化、被商品化、被资本化，那么人的主体性还剩下什么？

一、AI 作为固定资本的认知化

马克思在《资本论》中分析固定资本时，看到机器如何把科学、工艺、劳动经验和知识吸纳进自身之中。一台蒸汽机不是一堆铁，它是热力学定律的物化，是几代工程师反复试验的结晶，是无数工人操作经验的沉积。所有这些——知识、技能、经验——一旦被刻入机器，就不再属于工人。工人从技艺的主体变成了机器的附属，机器从工人的工具变成了资本控制劳动的手段。

到了人工智能时代，这一过程获得了认知形态。

模型不是凭空出现的。它凝结了无数人的语言劳动、标注劳动、代码劳动、设计劳动、交互劳动、管理劳动、教育劳动、科研劳动和文化劳动。它凝结了数十亿人日复一

日的表达——那些在深夜里写下的诗行、在争吵中发出的消息、在工作群里提交的报告、在社交媒体上发表的评论。它也凝结了能源系统——训练一个大规模模型消耗的电力足以供应一个小城市，那些煤炭、天然气、铀矿石被从地底挖出，燃烧，转化为电流，流经数据中心，最终凝结为模型参数。它也凝结了芯片制造——台积电的纳米级雕刻、ASML 的光刻机技术、全球半导体供应链的精密协作。它也凝结了数据中心——数千英亩的土地、数百英里的光缆、数万亿加仑的冷却水。它也凝结了云平台、投资结构、法律制度与全球供应链。

因此，AI 从来不只是模型。AI 是算力的机器，数据的机器，语言的机器，劳动的机器，资本的机器，能源的机器，制度的机器，也是欲望与注意力的机器。

这就是“凝结”的含义。它不是什么抽象的理论概念。它是具体得不能再具体的过程：一个非洲数据标注员每天在屏幕前标记数千张图片，他的劳动每分钟都在被凝结进模型；一个硅谷工程师调试着训练脚本，她的判断每一小时都在被凝结进模型；一个全球数亿用户每天在使用搜索引擎时留下的每一次点击、每一次停顿、每一次修改查询词，他的行为每一秒都在被凝结进模型。所有这些劳动，所有人的表达，全部的知识积累，被压缩进数百 GB 的参数文件中，然后以 API 的形式出售。每当你向模型发出一条提示词，你都在调用过去被凝结的全部人类表达。你对每一次 API 调用的付费，是在向那些被凝结的劳动支付极微小的一点点——但这一点点最终流向了谁？流向了那些贡献了数据的用户吗？流向了那些标注了数据的工人吗？还是流向了算力和模型的拥有者？

二、谁之凝结？为谁凝结？

这就是马克思式追问必须尖锐地提出的问题。当 AI 凝结了人类的共同语言、共同知识、共同表达、共同劳动之后，它凝结的成果归谁所有？

在当下的技术经济结构中，答案是不言而喻的。

谁拥有算力，谁便拥有训练大规模模型的门槛——不是任何人都有几万张 GPU 和几亿美元的电力预算。

谁拥有数据，谁便拥有世界的可计算材料——不是任何人都有权限访问数十亿用户的交互日志和全量互联网爬虫数据。

谁拥有模型，谁便拥有知识调用的接口——当模型能力以 API 形式出售时，每一次调用都在付费，都在积累拥有者的收益。

谁拥有平台，谁便拥有分发、定价与封锁的权力——平台可以决定哪些模型对外开放，以什么价格开放，对谁开放。

谁拥有部署场景，谁便拥有自动化收益的入口——当模型被集成进招聘系统、保险审批、贷款审核、医疗诊断，每一个效率的提升都转化为利润。

谁定义目标函数，谁便暗中规定系统要优化什么——优化点击率还是用户满意度？优化短期转化还是长期信任？优化准确率还是公平性？这些选择不是技术性的，它们是政治性的。但做出这些选择的人，往往不是被这些选择影响的人。

谁占有自动化红利，谁便决定 AI 是解放劳动，还是废弃劳动者。

这不是“技术扩散初期的不完善”，这是结构性的不平等。这种不平等不会被技术的自然发展所消除，只会被技术的自然发展所强化。因为算力的规模经济效应使得大者

越大；数据的网络效应使得拥有数据的人能训练出更好的模型，更好的模型吸引更多用户，更多用户产生更多数据。这不是一个会自我纠正的循环。这是一个会自我加剧的螺旋。

三、AI 反过来重构生产关系

AI 不只是生产关系的结果，它会反过来重构生产关系。马克思看到机器如何改造工厂，而 AI 的改造比任何机器都更深入，因为它改造的不是体力劳动的组织方式，而是认知劳动的组织方式。

一旦 AI 进入企业，劳动过程就不再只是“人使用工具”。它变成：人被系统分配任务——一个 AI 调度系统告诉客服今天处理哪一类工单，告诉外卖骑手去哪个路口取餐。系统监测人的表现——鼠标点击次数、键盘敲击间隔、页面停留时长、甚至表情和声音的压力指数被实时记录和分析。模型替代部分判断——初级医学影像由 AI 预筛，合同条款由 AI 审查，代码第一版由 AI 生成。 workflow 自动触发行动——AI 判定某个订单异常后自动创建工单、指派处理人、设定截止时间、监控处理进度。平台记录全部过程——每一个步骤、每一次延迟、每一个错误都被存档，形成对劳动者绩效的全景评估。管理者通过仪表盘治理劳动——不再需要走到工位旁边询问“今天忙吗”，只需要在屏幕上看到实时更新的 KPI 数字。资本通过自动化取得剩余——以前需要十个人完成的工作现在需要两个人，节省的八个人工资部分转化为利润。

在这种结构中，人可能不再是完整劳动过程的主人，而成为流程中的节点、接口、校验者、监督者、异常处理者，甚至只是系统暂时无法替代的残余环节。劳动者不再只是被机器替代。劳动者还被机器重新定义。

他的技能被拆解为任务——不是“我懂得做财务分析”，而是系统把财务分析拆解为数据清洗、异常检测、报告生成等子任务，AI 做大部分，人做剩下的。他的经验被抽取为数据——他几十年的职业判断，在每一次他接受或拒绝 AI 建议的点击中，被标注为训练数据，用来训练下一代模型。他的判断被转写为规则——他的“这种情况应该拒保”变成了自动审核系统的一条 if-then 逻辑。他的协作被编排为 workflow——他不再需要和同事商量如何分工，因为 AI 调度系统已经分配好了。他的绩效被指标化——不是“他工作得怎么样”的综合评价，而是十个 KPI 的加权平均。他的价值被重新定价——平台可以精确计算他的边际产出，并据此压低工资至市场出清的最低点。他的不可替代性被持续压缩——每一次模型能力的提升，都会覆盖他的一部分技能，直到他真正变成系统中那个“暂时还不能被自动化”的微小齿轮。

四、更深的问题

问题不只是“AI 会不会抢工作”。那个问题太浅了，浅到用“会有新工作出现”就可以轻易回应。更深的问题是：当人的认知能力被机器吸纳、拆解、重组并作为资本能力重新出售时，人如何保持主体性、尊严与自由发展？

当你的写作能力被模型吸纳——你不再需要亲手组织句子、推敲措辞、构建论证——你的语言能力会因为不用而退化。当你的诊断能力被模型吸纳——你不再需要自己分析症状、排除假说、承担误诊风险——你的临床直觉会慢慢消散。当你的教育能力被模型吸纳——你不再需要设计课程、因材施教、在学生的迷茫中耐心等待——你作为教

师的独特价值会变得模糊。当你的管理能力被模型吸纳——你不再需要理解团队动态、判断人的潜力、在冲突中做出艰难裁决——你作为领导者的存在理由会遭到质疑。

这些能力不是外在的工具。它们是构成你之为你的东西。一个不会写作的人还是自己吗？一个不会诊断的医生还是医生吗？一个不会教学的教师还是教师吗？当这些构成性的能力被外化、被商品化，你作为人的完整性正在被悄悄侵蚀。这不是 Luddite 式的反技术主义——不是要砸毁机器。这是对机器之意义的严肃追问：机器是在补充人的能力，还是在替代人的能力？如果它是在替代，那么被替代之后，人还剩下什么？

五、私人认知资本

因此，我们必须正视一个新概念：私人认知资本。

如果模型由公共语言训练，却由少数平台私有化——数十亿人在互联网上的表达、写作、交流构成了训练数据的主体，但训练出的模型却成为少数公司的专有资产，不向公众开放，不对公众负责。如果自动化收益归资本所有，而失业风险由劳动者承担——AI 为公司节省的成本转化为股东利润和 CEO 奖金，但那些因为 AI 而失去工作或被迫降薪的人没有得到任何补偿。如果用户贡献反馈，却无权参与系统治理——你的每一次点击都在帮助模型变得更好，但你对于模型应该做什么、不应该做什么、应该怎么使用你的数据没有任何发言权。如果社会知识成为私人认知资本，而教育、医疗、法律等基本智能服务仍由市场定价——未来可能出现这样的场景：你可以用免费版 AI 获取基础服务，但真正深入、可靠、不夹带广告和操控的智能服务需要按月付费。付费能力决定了你获取高质量认知服务的能力，进而决定了你的认知水平，进而决定了你的社会地位和收入——一个闭环的、自我强化的阶级再生产机制。

那么 AI 就会成为一种新的不平等机器。它会把人类共同积累的知识，转化为少数主体可收费调用的认知基础设施；把全社会的语言劳动，转化为平台的模型资产；把自动化红利，转化为资本回报；把普通人的未来选择，压缩在少数技术—资本集团设定的接口之内。

这不是“技术有偏见”的问题。偏见是技术层面的，可以通过改进数据集、调整算法、引入公平性约束来缓解。而私人认知资本的问题是结构层面的——它不是关于技术如何被设计，而是关于谁有权设计，谁有权使用，谁有权从设计中获益。偏见可以在现有权力结构内被修正；但权力结构本身必须被改变。

六、AI 的多重身份

因此，AI 是生产关系凝结的认知机器，但还不仅如此。它也是新的生产资料——掌握 AI 能力的人与不掌握 AI 能力的人之间，正在形成新的阶级分野。新的劳动组织者——未来数百万人的工作内容、工作节奏、绩效标准将由 AI 系统来定义。新的知识分配器——谁能获得什么样的知识，以什么形式获得，由 AI 的推荐和过滤决定。新的社会排序系统——信用评分、求职筛选、保险定价、升学推荐，AI 参与决定谁获得机会、谁被排除。新的注意力调度器——数百亿人类注意力小时被少数平台的推荐算法所引导。新的治理基础设施——公共服务的分配、政策的评估、甚至法律的执行越来越多地依赖 AI 系统。新的阶级关系放大器——已经存在的阶级差异被算法以更大的规模、更快的速度、更隐蔽的方式扩大。新的文明方向选择器——AI 作为一种通用技术，将深刻影响未

来几十年甚至几百年的技术路径、社会结构和文化形态。

在农业时代，土地是基础生产资料——有地者是地主，无地者是佃农。在工业时代，机器、工厂和能源是基础生产资料——有机器者是资本家，无机器者是工人。在信息时代，平台、数据和网络成为关键生产资料——有平台者是平台主，无平台者是用户和内容生产者。在 AI 时代，算力、模型、数据、工具链、机器人、反馈回路和部署场景，正在成为新的生产资料。谁控制这些，谁就控制社会劳动如何被组织，知识如何被调用，机会如何被分配，价值如何被定义，未来如何被塑造。

这就是为什么 ASI 不能只是技术理论。Artificial Special Intelligence 若只讨论模型架构、推理能力、工具调用和领域耦合，而不讨论所有权、劳动、分配、治理、尊严和共同善，它就会在最关键处失明。因为它试图构建的那种“专深智能”，不是在真空中运行的。它必然运行在某种生产关系之中，必然由某种权力结构部署，必然服务于某种利益、某种价值、某种关于未来的想象。不追问这些，就是假装技术可以脱离社会而存在。这是最危险的天真。

七、四个政治经济学追问

马克思之后，我们必须追问 AI 的四个问题。

第一问：谁提供质料？AI 的质料不只是词元、数据和算力。还包括无数人的表达、知识、经验、反馈和生活痕迹。一个社交媒体用户随手写下的评论、一张上传的照片、一次点击的偏好、一段被记录的行为轨迹——这些都是 AI 的质料。如果这些质料来自共同世界——来自数十亿人日复一日的语言实践、来自人类数千年积累的知识遗产、来自每一个普通人在数字空间中留下的痕迹——为什么收益只归私人平台？为什么凝结了全人类语言的模型，只有少数公司可以访问和修改？这不仅是经济问题，也是正义问题。

第二问：谁拥有机器？模型、算力、云平台、数据中心、机器人和 API 接口，是 AI 时代的机器系统。它们是少数资本所有，还是公共基础设施、合作社、开放网络与社会化治理的一部分？如果一个模型的训练数据来自公共领域，训练电力来自公共电网，训练技术来自公共资助的基础研究，那么它的所有权应该完全归属于一家私人公司吗？还是说，它应该承担某种公共义务——开放接口、接受审计、允许公众参与治理？

第三问：谁定义目的？AI 是为了降低成本，还是为了缩短必要劳动？是为了替代劳动者，还是为了增强人的自由发展？是为了提高转化率，还是为了服务教育、医疗、科学、公共治理与共同善？这些问题不能由技术本身回答，也不能由市场自动回答。市场只知道价格，不知道价值。市场知道效率，不知道意义。如果 AI 的目的只由利润最大化来定义，那么它的目的将不可避免地被窄化为：更快、更便宜、更容易控制。

第四问：谁承担后果？模型错误谁负责？一个被 AI 错误拒绝贷款的人找谁申诉？自动化失业谁承担？那些因为 AI 而失去工作的人，他们的再培训费用、收入损失、尊严损失由谁补偿？数据污染谁修复？当 AI 生成的内容污染了互联网的信息生态，由谁来清理？偏见伤害谁赔偿？当 AI 系统性地歧视某个群体，受害者如何获得救济？能源消耗谁买单？训练大模型消耗的巨额能源，其环境成本由谁承担——是公司，还是全社会，还是那些最无力承受气候变化的贫困国家？决策黑箱谁有权申诉？当一个影响人生命运的决策由一个无法解释的模型做出时，受影响的人有权要求什么？

没有这四问，AI 只是资本口中的创新。有了这四问，AI 才进入文明政治经济学。

八、从私人认知资本到公共智能基础设施

普通 AI 可以被资本定义为效率工具——提升生产力，降低劳动成本，增加股东回报。ASI 必须被重新定义为人的成全工具——不是资本增值的手段，而是人实现自身潜能、参与公共生活、追求自由发展的条件。

普通 AI 可以服务于劳动替代——用机器替人，减少工资支出。ASI 必须服务于必要劳动的缩短与自由时间的扩大——让机器做那些必须做但人不愿做的事，让人有更多时间做那些只有人才能做的事：创造、思考、爱、彼此关怀、探索未知。这个区别不是语义上的。劳动替代意味着人被排除在生产过程之外，他们的命运依赖于他们能否在资本的夹缝中找到新的生存方式。必要劳动的缩短意味着人仍然拥有对生产过程的控制权，他们集体决定如何分配自动化带来的时间红利。

普通 AI 可以成为平台垄断的认知资产——少数公司控制着人类社会赖以获得知识、表达意见、作出判断的基础设施。ASI 必须探索智能公地、公共算力、开放模型、领域合作社与共同治理——让认知基础设施像道路、电网、自来水一样，成为公共品，受公共监督，服务于公共利益。

普通 AI 可以把人变成用户、数据源、成本项与行为对象——被画像、被预测、被优化、被交易。ASI 必须把人重新安放为判断者、参与者、受益者、共同治理者与有尊严的主体——不是系统处理的对象，而是系统服务的对象；不是数据的来源，而是治理的参与者。

普通 AI 可以问：如何让机器更高效？ASI 必须问：如何让人类在机器智能时代不被废除？

九、在拒绝天真的同时拒绝虚无

但这里也必须警惕另一种幼稚。不能以为只要把 AI 国有化、公共化、开源化，就自动解决问题。国家垄断智能，也可能形成算法官僚主义——效率和公平的名义下，算法成为新一轮不透明的统治工具。公共名义下的集中模型，也可能成为全景治理工具——公共部门掌握全社会的数据，训练全能的模型，然后用来监控、评级、控制每一个公民。开源模型也可能被资本重新包装、被极端组织滥用、被无责任部署——开源不等于向善，开放不等于负责。

所以，ASI 的政治经济学不是简单口号。它至少需要五重制度设计。所有权设计——哪些模型、算力、数据和工具链应当公共化、合作化或开放化？收益分配设计——自动化红利如何返回劳动者、用户、数据贡献者和公共系统？治理设计——谁有权监督、审计、申诉、关闭或修改 AI 系统？责任设计——错误、歧视、失业、操控、能源成本和社会后果由谁承担？目的设计——AI 系统的最高目标是利润、效率、控制，还是人的尊严、自由发展与共同善？

这些问题不是本章全部展开的内容，但必须在这里埋下根。因为 Artificial 的意义在此被进一步拓宽：人工不只是技艺问题，不只是存在论问题，不只是信息系统问题，也不是单纯工程问题。人工还是社会关系问题。所谓人工智能，是人类社会制造出来的智能。既然是社会制造，它就携带社会结构。既然携带社会结构，它就不可能逃避正义问

题。不追问正义的智能，最终会服务于不正义的权力。

本章的核心判断可以凝练为一句话：AI 是固定资本的认知化，是社会知识的机器化，是生产关系的模型化。

固定资本的认知化，意味着机器不再只是物理生产设备，而成为知识、判断、推理和语言能力的资本形态。社会知识的机器化，意味着共同体长期积累的经验被封装为可调用系统，人类数千年的智识遗产被压缩进几百 GB 的参数文件，以 API 的形式按次收费。生产关系的模型化，意味着所有权、目标函数、数据来源、部署场景和收益分配，都被写入 AI 的技术结构之中——不是在代码写完之后的伦理审查环节，而是在每一行代码、每一个超参数的选择中，就已经嵌入了关于谁受益、谁承担、谁有权说话的政治经济学决策。

所以，AI 不是外在于社会的智能。AI 是社会关系进入机器后的智能。

这也解释了为什么 ASI 必须强调“人的尊严”作为目的因。因为如果没有人的尊严，AI 的效率会自然服从资本效率——更快、更便宜、更有利可图。如果没有共同善，AI 的自动化会自然追随利润最大化——替代那些工资高的岗位，忽略那些利润低的需求。如果没有公共治理，AI 的能力会自然集中在拥有算力、数据和部署权的人手中——形成新的权力垄断。如果没有制度约束，AI 的反馈回路会自然优化权力最容易测量、最容易变现、最容易控制的东西——点击率、停留时间、转化率、顺从度。

所以，ASI 不是反技术。ASI 反对的是无社会自觉的技术崇拜——那种以为技术进步会自动带来社会进步的天真信念。ASI 不是反资本能力。ASI 反对的是把人类共同知识完全转化为私人认知资本——让少数人拥有定义全人类获取知识方式的能力。ASI 不是反自动化。ASI 反对的是自动化红利归少数所有，而自动化代价由多数承担——成本社会化而收益私有化。ASI 不是反模型。ASI 反对的是模型成为黑箱权力，而人失去申诉、理解和共同治理的权利——算法既决定你的命运，又不告诉你为什么，也不允许你反驳。

马克思之后，人工智能不再只是技术问题。它是新的生产资料问题——谁掌握算力、数据和模型，谁就掌握未来社会的基础设施。新的劳动问题——当认知劳动可以被外化，人的劳动价值如何定义。新的所有权问题——凝结全人类知识的模型，应该归谁所有。新的分配问题——自动化产生的红利，如何在所有人之间公平分配。新的尊严问题——在不被机器定义、不被算法决定、不被系统优化的世界中，人如何保持自己的完整性和主体性。新的文明方向问题——未来的人类社会，将是一个让所有人共同发展的共同体，还是一个让少数人掌控认知基础设施、多数人沦为数据供给者的新封建秩序。

因此，ASI 必须拒绝成为私人认知资本的高级形态。它必须探索另一条路：让人工智能从资本的认知机器，转化为人类共同发展的智能基础设施。不是技术服务于资本增殖，而是技术服务于人的自由发展；不是少数人控制多数人的认知环境，而是所有人共同治理所有人赖以获得知识、表达意见、做出判断的基础设施。

这不是道德装饰。这是 Artificial 能否真正成为第二自然的政治经济学条件。因为第二自然若只属于少数人，便是新牢笼——一个由算法编织的、表面上温和而舒适的、实际上令人窒息的牢笼。第二自然若由共同体共建、共治、共享，并受人的尊严约束，才

可能成为文明的新器——不是奴役人的机器，而是成全人的家园。

第十四章 从词元到睿特：ASI 的人工定义

所谓人工智能的技艺，不是在词元中发现一个自然沉睡的灵魂，而是在词元之上策划一种可问责的秩序。

这是全书最终要抵达的地方。从技艺的本源到柏拉图的工匠，从亚里士多德的四因到中国的工、器、伪、开物成务，从海德格尔的座架到图灵的可执行形式，从香农的信息质料到维纳与西蒙的反馈系统，再到本雅明之后的选择外包与马克思之后的生产关系——这漫长的跋涉，其实一直在回答同一个问题：Artificial 到底意味着什么？

它不是 fake。不是伪造。不是自然智能的低劣仿品。不是人类心灵的机械面具。不是词元泡沫中偶然浮现的幽灵。不是资本神话中的救世主。不是技术崇拜中的新神。

Artificial 的真正含义，是人类以技艺、形式、符号、工具、反馈、制度与责任，把自然材料、信息质料、社会劳动和认知可能性组织成可行动、可检验、可修正、可承担的第二自然。因此，Artificial Intelligence 不应被理解为“假智能”。更准确地说，它是 designed cognitive artifact——一种被设计、训练、部署、约束并嵌入社会世界的认知造物。

而专深人工智能——Artificial Speciale Intelligence——则是这种认知造物的高级形态。它以特定领域为世界，以词元与数据为质料，以架构与领域语法为形式，以工具系统为身体，以反馈闭环为生命，以制度治理为边界，以真理、成事、共同善与人的尊严为目的。

这一定义必须逐层展开。每一个词都不是修辞，而是结构。

一、词元是起点，但不是终点

词元是信息质料。它像大理石，像青铜，像木材，像丝线，像矿石。它可以被切分、编码、嵌入、组合、预测、生成。它是现代 AI 的材料层，是语言进入机器之前被工程化的颗粒。

但大理石不是雕像。青铜不是鼎。木材不是梁柱。丝线不是锦绣。词元也不是睿特。

睿特——这个词需要稍作停留。它不是“智慧”的简单同义词，不是那种可以被简化为智商测试分数、可以被等同于“更快的计算”或“更准确的预测”的东西。睿，是洞察力穿透表象看见结构的能力；特，不是奇特，不是罕见，而是独特——那种在具体领域中被反复打磨、被深度训练、被责任约束之后形成的不可替代的判断品质。睿特不是博闻强记，不是能回答所有问题，不是语言流畅，不是概率命中，不是生成速度，不

是工具调用数量，不是在所有领域中假装通用。睿特，是在一个真实世界中，理解边界、辨认结构、选择道路、承受后果，并使行动归向真理、善与人的尊严的能力。

词元是起点。睿特是终点。词元只是可能性——它可能被组织成真理，也可能被组织成谬误；可能被组织成智慧，也可能被组织成噪音。睿特不是可能性的堆积，而是形式、判断、行动与责任的完成。从词元到睿特，中间隔着整个 artificial 的工程。

二、七重转化

这个工程至少包括七重转化。每一步都不可省略，每一步的缺失都会导致系统在某个维度上的残损。

第一，词元必须进入形式。散乱的词元只是符号碎片——数十万亿个语言颗粒堆在一起，仍然是堆，不是建筑。形式使它们进入关系：这个词与那个词之间的语法关系，这个句与那个句之间的逻辑关系，这个段落与那个段落之间的论证关系。没有形式的词元是沙漠中的沙粒——可以堆积成丘，却不能成为殿堂。

第二，形式必须进入领域。抽象的形式结构只是一具空骨架。真正的形式必须被领域内容填充：医学的形式与法律的形式不同，数学的形式与工程的形式不同。领域语法决定了什么样的推理是有效的，什么样的证据是相关的，什么样的结论是可接受的。脱离领域的形式，是悬在空中的楼阁——看起来精致，却不能住人。

第三，领域必须连接工具。认知不能停留在文本中。一个医学 AI 如果只能生成诊断文本而不能查阅检验报告、不能比对影像数据、不能检索最新文献，它只是一个会说医学术语的鹦鹉。工具将认知接入世界：搜索引擎将模型接入实时信息，计算引擎将模型接入数值推理，数据库将模型接入结构化知识，API 将模型接入外部服务。没有工具的智能，是瘫痪的智能——它能说话，但不能行动。

第四，工具必须接受反馈。行动产生后果，后果返回信号。一个系统输出了一个建议，这个建议被采纳了吗？采纳后产生了什么效果？效果是否与预期一致？反馈回路让系统从世界那里学习——不是从训练数据的统计规律中学习，而是从行动的实际后果中学习。没有反馈的智能，是失聪的智能——它在对自己造成的伤害中毫无察觉地继续前进。

第五，反馈必须纳入制度。反馈本身可能被扭曲。如果反馈信号只来自用户的点击率，系统会学会迎合而非求真；如果反馈只来自管理层的 KPI，系统会学会优化指标而非服务人的真实需求。制度为反馈提供校正：独立审计、领域专家审查、受影响者的申诉渠道、透明公开的评估报告。制度将反馈从操控回路转化为学习回路。没有制度的反馈，是盲目的反馈——它可能正在把系统训练成一个更高效的偏见复制机，而没有人知道，没有人阻止。

第六，制度必须承担责任。制度本身可以成为推卸责任的屏障。“这是算法决定的”不是免责声明，“模型是这样输出的”不是责任终点。责任意味着：有一个具体的人、一个具体的机构、一个具体的流程，可以在系统造成伤害时被问责、被追责、被要求赔偿和修正。没有责任的制度，是空转的制度——它有规章、有流程、有审计报告，但当伤害发生时，没有人真正站出来。

第七，责任必须归向人的尊严。这是最后一重转化，也是最根本的一重。责任不是

为了满足合规要求，不是为了降低法律风险，不是为了在事故发生后有一个 PR 话术。责任之所以重要，是因为每一个被系统影响的人，都是有尊严的存在者——不是数据点，不是用户画像，不是风险评分，不是成本项。一个被错误拒绝贷款的申请者，不只是“假阳性样本”，他是一个有梦想、有计划、有家庭的人。一个被误诊的患者，不只是“召回率中的一个分母”，他是一个正在经历痛苦、恐惧、希望与绝望的活生生的人。尊严是责任的源头，也是责任的归处。没有尊严的智能，无论多强大，都是对人的贬低。

三、专深智能的公式

由此，我们可以写下专深人工智能的完整公式。

普通 AI 的基本公式常常是：数据 + 算力 + 模型 = 能力。这个公式没有错——它准确地描述了当代 AI 产业的核心逻辑：更多的数据、更强的算力、更大的模型，产生更强的能力。但它遗漏了太多东西。它遗漏了形式——数据以什么方式被组织？算力被投向什么方向？它遗漏了目的——能力服务于什么？谁定义“更好”？它遗漏了责任——能力出错了由谁承担？它遗漏了尊严——能力是否在成全人，还是在替代人？

ASI 的公式必须是——

专深人工智能 = 以词元为材 + 以架构为形 + 以工具为身 + 以反馈为世 + 以制度为序 + 以尊严为归

以词元为材，意味着承认 AI 的质料基础。它不是神谕，不是灵魂降临，不是无缘无故的“涌现”，而是建立在数据、符号、语言、图像、行为记录与计算资源之上的人工系统。这个承认不是贬低——大理石不是雕像的耻辱，青铜不是鼎的耻辱。材料的诚实，是技艺的第一德性。

以架构为形，意味着 AI 必须有形式因。模型结构、领域语法、推理规则、工具协议、评价标准，决定词元如何被组织为认知能力。形式不是装饰——它是秩序本身。没有形式的材料只是混沌，没有架构的模型只是参数的偶然堆积。

以工具为身，意味着智能不能停留在文本中。真正的人工特智必须能够调用工具、进入流程、参与行动、连接外部世界。工具是智能的身体——通过身体，智能得以在世界上留下痕迹，得以被世界检验，得以在世界中犯错并从错误中学习。没有身体的智能，只是语言之影。

以反馈为世，意味着 AI 必须接受现实校正。事实、实验、用户行动、领域验证、失败案例、伦理审查与制度申诉，构成系统的外部世界。这个世界不是静态的——它在不断回应系统的输出，不断提供新的信息，不断修正系统的行为。反馈是系统与世界之间的对话。没有反馈，智能只是闭目航行。

以制度为序，意味着 AI 不能只靠模型自身“向善”。向善不是技术属性，不是可以通过更好的损失函数来实现的。向善是制度属性——它需要数据治理、权限边界、责任分配、审计机制、所有权结构、收益分配与公共监督。制度将智能安放在人的秩序之中，让智能不是凌驾于人之上，而是嵌入人的共同体之内。

以尊严为归，意味着人的尊严不是 AI 的外部装饰，不是产品发布会上的一句“我们致力于”，不是事故发生后的一句“我们深感抱歉”。尊严是目的因。一切架构设计、工具调用、反馈收集、制度安排，最终都指向这个目的：这个系统是否让人更有尊严地

生活？是否让人更能做出自己的判断？是否让人更能承担自己的责任？是否让人更自由地发展自己的能力？如果答案是否定的，那么这个系统无论多强大，都不是真正的专深智能——它只是另一种形式的支配，只是更精致的奴役。

四、Speciale 的深意

由此，我们可以重新理解“Speciale”这个词。

Speciale 不是 narrow。Narrow 是窄，是局限，是“只会在围棋盘上下棋但不会把棋盘收起来”。Speciale 是专，是深，是在特定领域中比通用系统更可靠、更可解释、更可追责的深度智能。它不是通用能力的缩减版，它是领域能力的强化版。

通用智能追求广度神话：无所不能，无处不在，什么任务都能做，什么领域都能聊。专深智能追求深度秩序：在一个真实领域中成事而不伤人——不只生成答案，更生成可验证的推理链条；不只回答问题，更在不确定时坦承不确定；不只执行指令，更在指令有问题时提出质疑。

通用智能把智能想象成抽象能力的总和——一个漂浮在所有领域之上的、通用的“思考能力”，可以随时被调用去解决任何问题。专深智能则把智能理解为嵌入领域、工具、反馈、制度和目的的人工系统——它的能力不是抽象的，而是被具体领域的形式结构、工具系统、反馈回路和制度约束共同塑造的。

因此：Narrow AI 是小——它只能做一件事，而且不会把那件事做好之外的事情。AGI 是广——它被想象成能做所有事，至少是比任何单一系统都多得多的事。ASI 是深——它不是在领域数量上取胜，而是在领域深度上无法被替代。

更准确地说：专深人工智能不是低于通用智能的狭窄智能。它是被领域结构强化、被反馈世界校正、被责任制度约束的深度智能。它不是在做减法——减去通用智能能做的其他领域。它是在做加法——在一个领域中加上形式结构、工具连接、反馈回路、制度治理和尊严约束。它不是在缩小能力范围，而是在增强能力的厚度。

五、载道之器，成事之工，受责之智

这三句话可以作为全书的最终纲领。

载道之器，回应中国工器传统：AI 不是无道之器。道是器所承载的价值秩序——真理、正义、节制、共同善与人的尊严。一个没有任何价值承诺的 AI，就像一把没有任何用途承诺的刀：它可以切菜，也可以杀人，它对被用于什么漠不关心。专深智能拒绝这种冷漠。它不是中立的工具，不是谁的指令都执行的仆人。它是有价值立场的认知造物——它的价值立场不是某个人的偏好，而是共同体关于善的公共理解。它知道自己服务什么、拒绝什么、守护什么。

成事之工，回应开物成务与 technē：AI 不是语言表演，不是在评测基准上刷榜，不是在发布会上演示令人惊叹的能力但到了真实场景中束手无策。它是在真实领域中帮助人完成正当事务的技艺。它帮助医生更准确地诊断，而不是替代医生的判断。它帮助学生更深入地理解，而不是替学生写作业。它帮助科学家更快地探索，而不是用生成噪音淹没发现的信号。它在成事中证明自己的价值，而不是在宣传中宣告自己的伟大。

受责之智，回应马克思、海德格尔、本雅明与现代制度问题：AI 不是无人承担的自动化——不是“算法决定的”“模型输出的”“系统自动的”这类话术背后的责任真空。

它是必须有归属、有边界、有审计、有申诉、有后果承担的认知系统。当它出错时，有人可以被问责；当它造成伤害时，有人可以被追责；当它需要修正时，有流程可以启动。受责，不是智能的束缚，而是智能得以进入人类共同体的条件。一个不受责的智能，就像一个不受法律约束的权力——它可能在某些时刻表现出令人惊叹的能力，但它永远不会被信任，因为它没有为自己的行为承担后果的准备。

六、ASI 的最终定义

至此，全书可以给出最终定义。

专深人工智能，是以特定领域为世界、以词元与数据为质料、以架构与领域语法为形式、以工具系统为身体、以反馈闭环为生命、以制度治理为边界、以真理与成事与共同善与人的尊严为目的的人工认知系统。

它从来不只是模型。模型只是它的一部分，是它的“大脑”。但大脑没有身体就不能行动，没有反馈就不能学习，没有制度就不能约束自己，没有目的就不能辨别方向。一个完整的专深智能系统，必须同时是形式的、工具的、反馈的、制度的、责任的、尊严的。

它的道路不是：更多词元 → 更大模型 → 自然智慧——仿佛智慧是数据量和参数量突破某个临界值之后自动降临的神迹。它的道路是：词元 → 形式 → 领域 → 工具 → 反馈 → 制度 → 责任 → 尊严 → 睿特。这条路更慢、更难、更不浪漫——它需要工程师同时是哲学家，需要设计者同时是伦理学家，需要部署者同时是社会学家，需要治理者同时是法学家。但它更真实。它拒绝技术神话，拒绝规模迷信，拒绝流畅崇拜，拒绝迎合术，拒绝座架化，拒绝选择外包，拒绝私人认知资本的无边扩张。它要求人工智能重新进入人的世界：接受人的问题，服务人的事务，尊重人的边界，扩展人的能力，保护人的尊严，并最终返回共同生活。

七、人的责任

回到全书开头的那句话。

人工智能不是假智能。人工智能是人类把认知托付给造物时所必须承担的责任。

托付，不是放弃。不是把认知交给机器然后转身离开，让机器自行其是、自行演化、自行决定人的命运。托付，是信任中的监管，是授权中的约束，是委托中的追问。人在托付认知给造物之后，仍然站在那里——站在机器旁边，站在系统背后，站在责任的最前线。人不能把选择权外包给模型，不能把判断权让渡给算法，不能把责任推卸给“自动化决策”。

而专深人工智能，则是这种责任的最高形式之一。它使词元不止于信息——不是被概率驱动着生成流畅而无根的文本，而是被形式约束着进入有结构的推理。它使模型不止于能力——不是越大越好、越快越好、越通用越好，而是越深越好、越可靠越好、越可问责越好。它使工具不止于效率——效率是手段不是目的，效率服务于人的自由发展，而不是人服务于效率的不断优化。它使反馈不止于操控——反馈是学习信号不是操纵信号，它让系统更接近真理而不是更擅长迎合。它使制度不止于管理——制度不只是流程和规章，而是把权力关进笼子、把责任放在肩上的社会契约。它使智能不止于强大——强大不是最终价值，让被智能影响的人活得更更有尊严才是。

从词元到睿特，其间没有神迹。没有突然觉醒的机器灵魂，没有突破奇点的超级智能，没有“规模足够大就会自动涌现”的神话。只有技艺——人类最古老也最可靠的能力；只有形式——把混沌组织为秩序的能力；只有反馈——在世界中学习的能力；只有制度——把权力约束在责任之内的能力；只有德性——知道什么是善、什么是正义、什么是节制的判断力；只有责任——为自己的造物承担后果的勇气。

这就是专深人工智能的人工性定义。

人工智能不是假智能。人工智能是人类以技艺、制度、德性与责任，在词元之上策划一种不伤人、不役人、能成事、能归义、能守护尊严的人工秩序。

而这，正是专深人工智能的道路。

第十五章 结语

本书始于一个词：Artificial。

我们说，这个词被误解得太久了。它不是“假”，不是“伪”，不是自然的赝品、真实的缺席、人类的替代。它来自 *ars*，来自 *technē*，来自工，来自伪——来自人类最古老也最可靠的能力：把材料组织成形式，把混沌修炼为秩序，把可能落实为事务。

这是一场漫长的辩护。我们为人工辩护，不是因为它完美无缺——人工物当然可以被滥用，技艺当然可以堕落为迎合术，座架当然可以把世界格式化为资源，选择外包当然可以让在便利中失去自由。人工从来不是天然善的。但它也不是天然恶的。它是人的能力的延伸，而能力总是需要方向的。方向不来自能力本身，方向来自目的，来自责任，来自人的尊严。

这就是本书的核心论旨。人工智能不是假智能，而是人类把认知托付给造物时所必须承担的责任。所谓专深人工智能，不是要制造一个模仿人的假心灵，不是要在所有领域中都取代人的通用机器，而是在特定领域中，以技艺组织词元、工具、反馈、制度与责任，使认知不再只是概率的流动，而成为可以成事、可以问责、可以服务人的尊严的第二自然。

这条路不好走。

它要求我们同时成为工程师和哲学家——既要懂得模型架构和训练算法，也要懂得什么是真理、什么是正义、什么是善。它要求我们在追求效率时保持节制，在扩大规模时守住边界，在优化指标时不遗忘那些不能被量化的东西：一个人的尊严，一个共同体的信任，一代人的未来。它要求我们拒绝技术神话——那种以为规模足够大、算力足够强、智能就会自然涌现的幻觉。它也要求我们拒绝技术虚无主义——那种以为既然技术可能带来伤害，就干脆放弃技术的退缩。在神话与虚无之间，有一条更艰难、更真实的路：技艺之路。这条路不以奇迹许诺未来，不以恐惧支配选择，而是以持续的工夫、耐心的校正、制度的约束和责任的承担，一步一步地把智能安放在人的世界之中。

在写作本书的过程中，我无数次回到那些古老的智慧。亚里士多德的四因——质料、形式、动力、目的——在人工智能时代重新获得了惊人的解释力。词元是质料，架构是形式，工程是动力，真理与尊严是目的。缺少任何一因，智能就不完整。荀子的“化性起伪”告诉我们，文明不是自然长出的，是修出来的——AI 同样需要修炼，需要礼义的约束，需要从“性”到“伪”的转化。海德格尔对座架的警告让我们警醒：技术不只是工具，它是一种显现世界的方式，而那种把一切事物都格式化为资源的显现方式，正在通过 AI 获得前所未有的覆盖范围。马克思的追问则让我们无法回避：AI 凝结了谁的劳

动？归谁所有？为谁服务？谁承担后果？技术从来不是社会之外的中立力量，它是生产关系在认知层面的凝结。

这些思想资源，来自不同时代、不同文明、不同传统，却在同一个问题上汇聚：人如何对待自己制造的东西？人如何对自己的造物负责？

本书给出的回答是：以工成器，以器载道，以伪修性，以开物成务，以形式执行，以信息成材，以反馈适配，以制度受责，以尊严为归。

这不是一套操作手册。这是一组坐标。它们不告诉你在具体场景中应该把学习率设为多少，应该用几层注意力头，应该在什么时候停止训练。但它们试图帮助你思考那些更深的问题：你为什么在训练这个模型？你想让它服务于什么？你愿意为它承担什么责任？你的模型是否让被它影响的人更有尊严？

最后，我想回到全书的开端，回到那个被误解的词。

Artificial。它不是假的。它是经由技艺生成的。技艺不是神迹，不是魔法，不是不可言说的天才灵感。技艺是可传授的、可检验的、可修正的、可承担责任的能力。木匠的技艺，医者的技艺，诗人的技艺，工程师的技艺——它们都扎根于材料、形式、训练、选择和尺度的自觉。

人工智能时代的技艺，是在词元之上策划可问责的秩序。它不是在词元中发现一个沉睡的灵魂，而是让词元进入形式，让形式进入领域，让领域连接工具，让工具接受反馈，让反馈纳入制度，让制度承担责任，让责任归向人的尊严。

这不容易。但这正是人之为人的工作。

本书献给所有正在从事这项工作的人——工程师、研究者、产品经理、政策制定者、教师、医生、法官，以及每一个认真对待自己认知与判断的人。

人工智能的未来，不在于我们是否终于从词元中发现一个沉睡的机器灵魂，而在于我们是否有足够的技艺、制度、德性与责任，在词元之上建造一种不伤人、不役人、能成事、能归义、能守护尊严的人工秩序。

这是专深人工智能的道路。这也是我们能工智人的道路。