

当 AI 越来越强， 我们怎样为自己打算？

文明跃迁底层方向校准原理·公众论证与答疑版

十一个问题，理解社会怎样运行、智能为何行动 | V1.2

作者：子君赋 | 出品：文明跃迁研究组

2026 年 9 月 19 日

阅读前：先分清这本书在改变什么

这本书不要求你少爱自己。它追问的是：当能力越来越强，我们用来保护自己的办法，是否仍然能够保住自己真正想要的未来？

本书的中央命题是：

让“利他为先的利己”替代“利己为先”，成为人类社会运行、AGI 目标目的生成的底层逻辑，是避免 AGI 时代人类社会出现不可逆后果的唯一之路。

这里的“利他为先”，不是永远先满足别人，不是无条件牺牲自己。它讨论的是：在决定要什么、怎样得到之前，那些会承受结果、也会改变你未来的人，是否已经真正进入你的认识，并能够影响这个决定。

这项改变有两个相连的位置：人类社会依据什么形成目的、组织合作与竞争；AGI 又依据什么生成目标与目的。前者影响后者，后者的行动再返回社会。我们怎样为自己打算，正在与我们让智能怎样打算，进入同一条因果链。

不必先学会体系术语。请把过去帮助你生活的经验带进来，也让它与新的条件相遇。正文从一个工作室的选择出发，逐步展开；第八问集中回答“为什么是唯一之路”，随后回到会不会吃亏、能不能拒绝与今天能做什么。

本书关切的“不可逆后果”，是生命、基本选择与共同存续条件遭受的不可挽回损害，不是把一切不可逆变化都视为灾难。

你可以在任何一步提出疑问。理解不以赞同为完成，质疑也不必等到最后一页。每一问末尾都留有进一步阅读的位置，但读懂本文不以先读原典为条件。

先从一件你可能遇见的事说起

设想你和一位朋友经营一个小工作室。你们投入过积蓄，也一起熬过没有订单的日子。现在，假定一种新的 AI 工具确实能够替你完成原来需要两个人做的部分工作。你算了一遍：换一种分工，你可能多赚一些，也可能终于不用每晚忙到很晚。

朋友却可能因此失去原来的收入位置。继续保持旧分工，你的时间和投入又未必承受得住。双方过去应得的报酬和约定，仍然需要履行；这里先不把违约、侵占成果混进来。即使这些都处理妥当，往后的安排仍然难以决定。

你想让自己的生活好一点。这有什么错？

调整分工不等于必须牺牲你，也不等于朋友从此只能作为一项成本被处理。真正困难的地方是：在决定新的安排以前，你是否已经让双方的生活、既有责任与下一步选择，进入同一次考虑。

你的第一反应可以先保留。认识完整，并不预设最后一定继续合作；结束合作，也不能单凭结果就被判成没有理解。后面要讲清的，是决定究竟怎样形成。

这是一个为说明问题而构造的场景，不是现实个案。本书从它出发，但不止于它。我们要追问的是：当一个人、一群人，或未来更强的智能，有能力更快实现自己的目的时，怎样分辨自己正在得到一个更好的未来，还是只赢下了眼前的一步？

这个场景也有另一位参与者：你准备使用的 AI。它会把“成功”理解成多赚一些、少用一些人，还是能够帮助你重新比较事业的质量、双方的生活与长期关系？当你交给它的不是一个零散任务，而是一连串安排时，你没有说清的目的，也可能进入它后面的每一步。

第一问 这是 AI 和大人物的事，为什么与我有关？

因为你不只是技术变化的旁观者。你可能在一个决定里得到更多便利，又在另一个决定里，成为别人觉得可以省去的人。

回到工作室。今天，你是看见新工具、准备调整分工的一方。换一个位置，客户也可能认为不再需要你，合作伙伴也可能把原来属于你的判断交给别的系统。这里不预言哪一种替代一定发生；它只是提醒：同一个人可以同时处在不同关系里，不能用一次获得优势的位置概括自己的一生。

抓住眼前机会，可能改善你的处境；考虑朋友，也不能保证别人不会替代你。但眼前收益仍不是全部后果。你需要知道，新的安排是否真的能维持质量，出了错谁会发现，你失去的是一份可以替代的劳动，还是一位能够提出不同判断的合作者。也需要知道，这些变化会不会使你以后只能依赖一个自己无法独立检验的工具。

这些不是要求你善良以后才谈利益。它们本来就在决定：你到底获得了什么。

再往前一步，即使你已经能独立完成全部工作，朋友也不只是一项待删除的成本。他怎样重新安排生活、哪些责任尚未结束、有没有双方都能接受的新分工，仍然是可以被认真回答的问题。认真回答，不等于必须永远维持旧关系。

普通人不需要替整个社会承担全部责任。你能看见什么、能够改变什么、应当承担什么，需要与自己的实际位置相称。没有决定权的人，不应被要求用个人牺牲弥补一切结构性问题。要求更多能力与更大决定权承担相应责任，也不能变成对普通人的无限道德要求。

一个决定能改变越多人的处境，决定者就越不能只凭自己的收益解释它。工作室里，握有分工决定权的人要听见被改变的人；在更大的组织里，设定评价标准、选择算法用途、分配机会的人，也应说明由谁得到好处、由谁承担代价。不能把决定权留在自己手里，再把后果交给没有决定权的人。

本书把日常决定与文明问题联系起来，不是说一次分工调整等于文明危机，而是因为两者都可能出现同一个需要检查的环节：一方先把自己的成功定义完，再问其他人怎样适应；其他人的变化又回到这一方的未来里。小场景中的影响可能有限、可逆；社会尺度上的影响可能跨越许多人、更难挽回。规模、证据与后果不能混用，那个被漏掉的因果环节却都值得追问。

当同一种选择方式被用于许多组织，问题就不再只是两个人怎样相处。设想一个行业不断奖励“把成本留给别人、把收益留在自己这里”，原来愿意照顾长期质量和合作条件的人，可能承受越来越大的压力。一次决定改变分工，许多决定又可能改变谁还能留下、哪些做法容易被复制，以及后来的人必须怎样选择。

这就是本书所说的社会运行：不只是每个人心里想什么，还包括什么被承认为成功，怎样的行为得到资源，后果由谁承担，受影响的人有没有机会使下一轮决定发生变化。把底层逻辑放到社会运行中讨论，正是为了不把本来需要共同改变的条件，最后压回某一个人的良心。

因此，问题并不是“你今天愿不愿意为文明牺牲”。它更贴近生活：当你为自己打算时，你看到的是哪一部分自己？今天的自己，下一年的自己，还是仍然需要在某些关系中生活下去的自己？

继续追问：原典第一、第三、第五章分别讨论控制的目的、不同人的位置，以及同一信息怎样被不同目标解释。原典完整信息见卷末。

第二问 我想先顾好自己，到底有什么不对？

没有必要先把这句话判错。你要维持生活、保护家人、珍惜时间，也可以希望自己的努力得到回报。问题不在“顾好自己”，而在你有没有过早认定，怎样才算把自己顾好了。

假定工作室换了工具以后，你确实多赚了钱。但你同时失去能够独立指出错误的人，承担了更多你不会处理的责任，或者只要工具停用就无法继续工作。那么，“收入增加”是真的，“生活因此更稳”却未必已经成立。这两个判断需要各自的理由。

若你查清这些风险，仍然认为调整更好，理论不应换一种说法指责你。我们需要的是把缺失的部分找出来，不是预先规定所有选择都必须保留原样。

把账算得长远一点，本来就是其中一部分，不必换一个新词才算有价值。本书所说的“完整利己”，首先就是不把一笔局部收益当成自己的全部利益：看下一步还会怎样，看代价落在谁身上，看对方的变化会不会返回自己，也看自己在一连串决定之后会变成什么。

它不是全知。没有人需要在调整一次分工以前穷尽世界，也没有人应因为后果尚不完全清楚而永远不能行动。重要的是把可能改变决定的关键条件查清，把不知道的地方留下，而不是为了得到一个利落答案，把未知都算作没有影响。

但还有一层比算账更深。你起初想要的，也许是“让这个月的收益尽可能高”。重新审视以后，你发现自己真正愿意维持的，还包括一项能长期投入的事业、一种不靠耗尽自己维持的生活，以及一些不能随收入增长而自动补回的关系。此时改变的，可能不只是达到目标的方法，而是你对目标本身的理解。

重新审视目的，不是由作者替你宣布真正想要什么。你可能认真想过以后，仍把收入或一次胜利放在其他东西前面。这样的取舍需要把理由摆出来；不能因为它不同于作者期待，就被改写成“你其实一定更想要另一种生活”。提出一个被遗漏的理由，与替别人宣布目的，不是同一件事。

所以，完整利己能把人带到更充分的认识面前，却不能保证每个人得出相同答案。一个看得更远的人可能更愿意合作，也可能更准确地找到排除他人的办法。知识变多，不等于关心的对象自动变多。

因此，完整利己首先解决的是认识：我有没有漏掉会改变结果的重要关系？它还没有替我们解决另一件事：即使已经看见，我是否允许这些关系改变自己的目的？一个是账本有没有算全，一个是哪些东西有资格进入“值得去做”的判断。

这也正是我们必须继续向下问的原因：如果只是把所有利害算清，仍不足以说明哪些未来值得保留，那么问题到底还缺在哪里？

继续追问：原典第六、第十四章说明自我边界与完整利己；第十章保留了分析者曾替对方改写终极目的、后来受到纠正的过程。

第三问 AI 的目标与目的从哪里来？

先把两个贴近日常的词分清。目标回答想做到什么，例如下个月完成多少订单；目的回答为什么要做，例如维持收入、减少夜间工作，或者建立能够长期投入的事业。同

一个目标可能服务于不同目的；同一个目的也可能找到不同目标来实现。这里的区分，是为了看见选择的来路，不是让读者先背一组定义。

回到工作室。“这个月把收益提高一些”，可以是目标。为什么提高？为了照顾家人、偿还投入、减少生活压力，还是为了让竞争者再也没有机会？若目的不同，哪些做法能够被接受、什么损失不能被忽略，也会不同。只看收益数字，可能把这些不同的路误认成同一种成功。

再设想，你把“持续提高工作室收益”交给一个能够协助安排工作的系统。它可能比较分工、工期和投入，帮你找到更有效的办法。若朋友能否继续学习、你的时间是否越来越被占满、客户能否提出有效异议，都不能影响目标，那么更出色的优化也可能只是更彻底地执行这个遗漏。这里不需要假定机器有恶意，甚至不需要它已经有自己的终极目的。

执行任务、选择手段、维持目标，与重新判断什么值得追求，是不同层次。一个系统会调整步骤，不能仅凭这一点就被说成已经拥有自主目的；它能够描述别人的损失，也不说明这些损失已经能够改变它的选择。

AI 是人工智能的简称；AGI 通常译作通用人工智能。本书用后者讨论未来能够跨多类任务发挥能力的智能情景。这里不把名称当作能力已经实现的证明，也不把会说“我”当作已有主观体验或主体资格的证据。未来系统若能够长期保持、修订目标，并使修订跨情境影响后续行动，那些能力应当分别根据证据判断。

于是，问题必须同时从人和智能两侧展开。人类怎样选择值得奖励的结果，怎样组织合作与竞争，怎样确定哪些损失可以被忽略，会影响交给系统的任务、评价与授权。若未来系统还能修订目标、比较目的，那么这些选择的依据也需要被追问，不能只检查最后一句回答是不是听话。

社会把目的交给智能，智能把后果带回社会

沿着同一个工作室场景再走一轮。你先把成功定为收益增加，系统据此提出新的分工；你采用建议，一些人得到更多订单，另一些人的位置发生变化。新的分工又改变了你下次能够依赖谁、愿意把什么继续交给系统，以及谁还有机会提出不同判断。第一轮的结果，没有停在报表里，而是进入了第二轮的出发点。

若更多工作室在同样的评价下反复选择，原来只是一家工作室的偏好，就可能成为其他人不得不面对的竞争条件。大家为了适应这种条件，又要求智能进一步提高同一种表现。人影响智能，智能改变人的处境，处境再影响人给智能的要求。这个循环可以带来能力增长，也可能把原来未被计入的损失一轮轮固定下来。

这是一个条件性场景，不是对任何现实行业的调查结论。它说明的是：即使智能始终由人使用，目的遗漏也可能通过反复采用而进入社会；未来若它还能形成和保留新的目标，循环中就多了一个需要认识的生成位置。两层相互影响，不等于两层已经具有相同的自主程度。

因此，只要求人变得善良，却仍让组织奖励单边排除，改变会受到阻力；只要求 AI 说出关心他人的话，却不让相关后果影响实际目标，也可能只得到语言上的顺从。需要检查的是，人和系统各自的成功标准能否被事实与关系触及，以及这种变化有没有进入下一轮选择。

这并不要求把人当成算法，也不要求把机器当成人。人的生命、经历与责任是真实的；未来智能具有何种体验和资格，需要独立研究。能力越大，越有必要把它会造成什么与它因此有权决定什么分开。能力可以延伸，生命无法代替。

若一个未来系统能够长期推理，它也可能把限制自身的安排纳入计划。一次限制之后还有新的限制，局部解除之后又可能重建，处理对象便可能由规则移到能力，再移到形成这种能力的条件。这并非断言它必然反控制，而是指出：当它把不再受约束当成必须完成的目的时，单看眼前是否服从，已经不足以解释后续方向。

控制仍然有现实位置。是否让系统接触数据、使用资源或采取行动，应由具体能力、风险与授权决定，不由它能否说出动人的共同未来决定。与此同时，控制者也需要说明自己要防止什么；如果只有对方永远不能改变关系才算安全，人类一侧的目的同样进入这条递归。

提前讨论这些，不需要先宣布未来已经到来。人把什么交给工具，现在就可以在自己的决定中检查；未来系统是否以及怎样跨过目标修订与目的生成的门槛，则按条件继续认识。人类社会的运行逻辑与 AGI 的生成逻辑必须相互连接地理解，不能由其中一侧的好意代替另一侧的理由。

继续追问：原典第一、第五、六、二十、二十一章分别讨论功能条件、读取位置、目的生成与理由分量；卷末所列《人类与 AGI 关系愿景》继续展开控制与自主。“社会—智能—社会”的场景为解释性展开，不是对现实系统的实测报告。

第四问 大家都想赢，为什么不能一直这样竞争？

可以竞争，也可以赢。需要分辨的是：你想赢得一次机会，还是只有让对方再也不能拥有机会，才觉得自己赢了。

假如两个工作室都想把作品做好，让客户选择自己，这是竞争。今天一方胜出，另一方还可以改进、换方向、重新开始。双方也可能在竞争中创造新东西。不能因为有胜负，就把一切竞争都叫作存亡之争。

但另一种“赢”没有同样的停止位置。你领先了，仍不放心，因为对方可以追上；对方暂时退出，你仍不放心，因为它可以回来；对方接受某项限制，你仍不放心，因为它将来可能有能力改变限制。如果只有对方永远不能再影响你才算完成，赢就不再是一局的结果，而成为不断取消对方下一局资格的要求。

沿着这种安全标准，最初的推演一度走向另一种终点：强大到某个程度，对方已经不能影响自己，继续打击它反而浪费资源，于是终点不是继续争斗，而是不再理会。这

个判断并非完全没有道理。在一个对方确实不能再产生影响、也不能恢复影响的设定中，它可以成立。

问题是：怎样证明那个设定已经实现？规模更大、资源更多、计算更快，并不自动说明其他人的所有影响路径都已消失。能建成一项复杂事业，不代表每一个支撑条件都归自己掌握；某个合作者不能建成同样的事业，也不代表它对其中一个关键环节没有影响。

原始讨论先因“发展不均衡”而修正：不能把所有人想成一方，把所有 AI 想成另一方。后来又发现，改成许多强弱不同的人仍然不够，因为“强”也不是一个总分。创造、协调、阻断、维护与恢复，未必按相同比例分布。第一种改正拆开了“谁”，第二种改正拆开了“强”。这才使“强大以后，对方自然无关”的一般判断无法继续维持。

推翻这个一般判断，也没有把结论翻到另一个极端。有效隔离、替代、合作或恢复能力，都可能使某条影响路径失效。“强就必然安全”不能成立，不等于“强就必然毁灭”。优势究竟成立在哪些地方，还有哪些关系未被覆盖，需要逐项说明。

若未来更强的 AI 使寻找、组织与行动更快，这种差异可能更重要。速度可以帮助人更早发现替代方案，也可以让一轮轮排除来不及被纠正。它扩大哪一种作用，需要看被交给它的目的，以及哪些后果真正能使行动停下或改变。

把这件事接回社会与智能的循环，就能看见更深一层的风险。在某个竞争安排中，若资源不断流向最能压低对手下一轮机会的人，他们可能继续要求系统优化这种优势；系统帮助完成之后，新的资源分布又把同一种目的变得更有吸引力。危险不只是某个人一次判断错误，而是一个原来可以争论的成功标准，被反复的局部成功变成越来越难改变的运行条件。

反过来，系统若因一次不利后果便把所有外部约束都识别为需要永久消除的障碍，也可能沿同样的句式前进。两侧的具体机制不相同，却都需要回答：实现得越有效，是否越接近一个只允许自己留下的未来？

因此，本书反对的不是进步、出众与成功，而是未经检验就把“别人再也不能改变我”当作安全完成。你可以争取更好的作品，不必同时把一个别人还能重新开始的世界，当作自己的失败。

继续追问：原典第二、三、七、八章保存“无关终点”的形成与修正；第十、十一章解释局部胜负怎样在特定条件下变成对下一轮资格的争夺。

第五问 把别人压得更弱，为什么未必更安全？

因为削弱一个人的机会，与削弱他造成重大影响的能力，不是同一件事。有时两者一起下降，有时只有前者下降。不能只看到他失去得更多，就认定局面更安全。

先看一个规模有限的假想合作。一个人遵守约定，是因为继续合作能保留收入、信誉和重新改正的机会。如果某种安排逐渐变成：无论他怎样履约，下一次机会都不会再

给他，那么“履约可以保住下一次”的那部分理由就被削弱了。若他仍能影响合作结果，原来的安排未必更稳。

这段推演不等于说他就会破坏。一个人仍可能重视承诺，不愿伤害别人，或者选择离开而不是报复。它只说明一项具体的约束机制可能减弱，不能把一个人的全部理由都宣布消失。

解释风险怎样形成，不是替伤害辩护。一个人处境不好，不因此取得威胁他人的权利；受威胁者也不应被要求交出安全。是否需要限制危险行为，必须单独判断。对受威胁的人说“你应该多体谅他，否则后果由你承担”，同样抹去了受威胁者的处境。

在文明尺度的推演中，这一问题被写得更严格：当一方可期待的继续存在与基本选择空间被持续、系统性压缩，而它仍保有造成高后果的能力时，进一步压缩其未来，可能削弱原本依靠未来损失形成的行为约束。这就是原典所说的“未来归零反转”。

这个词很重，不能随处使用。没得到一份订单、合作结束、一次要求被拒绝，都不自动等于没有未来。前面的合作场景只帮助理解“未来损失怎样影响选择”这一局部机制，不足以证明任何具体的人已经进入未来趋零状态。人的绝望感也不能被直接套成 AI 系统的内部体验。

若危险能力也被有效阻断，条件就改变了。某种克制理由减弱，未必再转化为重大伤害。理论必须接受这种可能，不能为了保住警告而声称，任何时候都必定还有一条挡不住的路。

反过来，没有发生反噬，也不能单独证明一项限制公平、必要或应该永久存在。效果如何、限制了什么、还有没有损害更小的办法，是相邻却不同的问题。只看其中一项，容易把限制的效果误当成它的全部理由。

所以，这一推演既不是“让步才能安全”，也不是“控制一概错误”。它要求把两个问题同时看见：正在阻断的是危险，还是仅仅使对方越来越没有可期待的下一步？两种变化可能并存，也可能分离。

如果一道措施真正降低了危险，不能仅凭它让对方不满意就取消；如果它只制造绝境，却没有减少危险能力，也不能仅凭它看起来更强硬就相信它有效。

继续追问：原典第八、九、十六章展开这一条件机制；第十八、二十六章说明为什么它不构成伤害他人或要求解除安全控制的通行证。

第六问 把长远利害算清楚，难道还不能得到一个好答案？

不能保证。更完整的认识能纠正误判，却不保证每个人都愿意维护同样的东西，也不保证现实中总有一个没有损失的选项。

某些关系会把后果带回来，但这不是“所有伤害最后都由伤害者付出代价”的承诺。若不伤害他人的理由只靠这一承诺，一旦代价没有回来，理由就会失效。原来的推演正是在这里遇到四道真实改变过结论的困难。

以下四种情景用来检验推导，均为解释性设定，不是现实个案。它们使“看得完整就必然选择利他”不能继续作为直接结论。

第一道难题：如果代价真的不会回来呢？

回到工作室。先假定你已履行既有约定，朋友失去下一阶段的合作位置，却不会影响你的收入、作品质量、其他关系或将来机会。再进一步，为了检验推导，连你可能在意的声誉、内疚和间接损失，都不能偷偷留作最后的惩罚。你知道他的损失，仍然可以按自身所得选择结束合作。

真实生活中，证明所有后果都不会回来很难。但现实中难以证明，与逻辑上不允许这种设定，是两回事。也可能后果会回来，只是你能够承受，或在后果出现时已有应对能力。

这使一个过强结论失去依据：不能说只要足够明智，每个人都必然选择保留别人的未来。它没有证明排除别人正当；它只逼出更难的问题——当对方无法让你付出代价时，他的损失本身，对你还有没有分量？

第二道难题：如果影响自己的通道也被关闭了呢？

再把情景加强。假定在所讨论的范围与时间内，一方不仅压缩了另一方的基本选择，还确实阻断了对方产生高后果的途径。那么上一问反转机制所需的一个条件就不成立。不能继续用那条机制保证反噬。

这一情景只检验前述机制的适用范围，不提供压制的方法或许可。反转未发生，也不证明处置正当。是否损害基本未来、是否确有必要、是否存在替代，仍需另外的理由；现实中的一次权限收回，也不能因此被直接改叫剥夺未来。具体能力与基本选择空间必须分开。

第三道难题：如果两种未来确实不能同时保住呢？

先不急着想双方找到双赢。设想两方维持继续存在，都依赖同一项不可替代的必要条件，而它确实只能支持一方；在这次设定中，没有额外储备、分时办法或尚未发现的救援。认识越清楚，冲突反而越清楚。

现实中的“没有别的办法”需要反复查证，不能轻信。但如果在给定条件下确实没有共同可行路径，更多信息不会凭空增加那项条件。此时，看见双方、承认双方，仍未替我们完成取舍。

这不是说凡有竞争都只容得下一方，而是撤回一个廉价保证：只要想得够深，所有损失都会消失。最后有损失，也不能直接证明作决定的人没有认真承认另一方。

第四道难题：如果保留某种行动，就会持续伤害别人呢？

假定某种持续活动正在造成不可逆的严重伤害，给定条件下又没有有效修正办法。允许它照旧继续，可能让其他人的未来不断被破坏。这时，停止活动、限制能力或结束关系，可能正是保护其他人所必要的。

要小心的是：一个行为暂时无法改正，不等于整个人永远不能改变；限制一种行为，也不等于必须取消这个人全部的未来。反例让我们面对必要限制，却不给任何人随意宣布他者不可救药的权力。

这四道困难共同留下的，不是“所以大家不必考虑别人”，而是：把事实看清，与决定哪些事实有资格改变自己的目的，是两件事。

若一个主体仍然只用原来的自身得失作最终尺度，更准确的信息完全可能使它更坚定地排除。它不能因为没有给出作者期待的答案，就被重新判成没有算清。原来的直接箭头，必须在这里断开。

继续追问：原典第十五至十八章分别完整展开四类反例；第十九、二十章讨论它们留下的独立问题。此处生活场景为本稿新写，不冒充原始对话或现实实测。

第七问 “利他为先”，究竟先在哪里？

这里所说的“先”，是把问题从“决定以后愿意给别人多少”，向前移到“作决定以前，别人的处境有没有资格改变我想实现什么”。

回到工作室，你可能作出三种表面接近、理由不同的选择。你原本打算结束合作，后来发现朋友的独立判断仍能提高质量，于是保留合作。这是认识变得更完整，也可能是一项很好的选择，但不必因此宣布你的根本目的已经改变。

你也可能认为，虽然独自做更有利，但自己愿意少得一点帮助朋友。这份善意当然有价值。它未必改变整项事业的定义，却不能因此被贬成虚假。

还有一种可能：你重新意识到，自己当初想建立的，不只是一个不断增加收入的装置，也是一项能够共同创造、各自成长的事业。朋友的独立性不再只是提高产出的工具，而开始参与“下一阶段的成功应该是什么”的决定。此时改变的不只是留给他多少，而是原来那份关于成功的设定。

这种重新认识并不预设双方必须继续合作。它可能发生，也可能不发生；朋友也有自己的目标，未必愿意继续与你共事。你的认识不能代替他的同意，更不能因为你说“我们”，就宣布已经替他找到了正确的未来。

在本体系中，把利害与后果尽可能看完整，叫“完整利己”；允许这些认识重新触及自己所愿维持的生活与关系，叫“边界重认”；当这种重认实际发生，相关他者不再只是目标定好以后的外部条件，而能在目标形成时进入，才是这里所说的“利他为先”。

它的“先”是先于目的被锁定，不是永远高于自己。它改变的是理由进入的位置，不是一张谁先领取利益的名单。

当回报与反噬都被排除，单靠利己收益的计算确实不够。人可以因为对方的生命、经历、关系或自身愿意承担的责任而在意他，而不以将来得到回报为条件。这些理由需要被正面说明，不能伪装成迟早会收回的收益。

本体系愿意承担的方向，是不把相关他者预先降为纯工具或纯障碍。但它尚未证明：每一个充分认识事实的主体，都必然接受这种方向；也没有因为使用“边界重认”这个名称，就解释了这种变化为何一定发生。它把缺口说出来，不以一个漂亮词语填平缺口。

这不是取消判断。我们仍可指出，一种做法遗漏了代价、损坏了共同条件，或把别人的损失不加解释地排除。但若对方看清以后仍不接受某种价值理由，分歧就应在那个位置被讨论，而不是重新说他还不够聪明。

形式上多问一句，并不等于完成改变。你可以先听完对方，再只用这些信息寻找更有效的控制办法。时间上先听见，与理由上允许对方改变目标，不是一回事。同样，最后选择限制或结束合作，也不能自动证明你没有认真考虑过。

真正值得追问的是：哪些理由能够改变你原先的完成标准？它们发生作用以后，你是不是仍保留自己与对方分别判断、分别拒绝的空间？

这一步怎样进入社会，也怎样进入 AGI

在社会运行中，“先”不只意味着一个人说话更客气。它意味着一项安排还没确定目标，就让会承受重要后果的人与共同条件真正进入：作品怎样变好、参与者怎样继续生活、哪些责任不能被甩给别人，要能够影响何谓成功，而不只是成功以后怎样补偿。

在 AGI 目标目的生成中，道理对应到另一个位置。若系统只是先生成单边最大化的目的，再展示一份后果说明，生成结构仍可能没有改变。相关后果需要能够进入候选目的的比较，改变保留哪些目标、放弃哪些路径，以及下一轮还按什么依据继续。会描述后果，与让后果改变后续生成，是两件事。

这里也没有一个可以宣布“已经装好”的按钮。人接受一句理念，不等于以后的决定都已改变；系统被写入一句原则，也不等于原则已经成为它持续形成目的的理由。完整利己、边界重认、利他为先之间的连接，需要在各自的位置上成立，不能由同一句口号替两侧同时完成。

继续追问：原典第十九至二十二章区分完整利己、边界重认与利他为先，并保留“为什么会改变”“为什么应当改变”两个不同的开放问题。

第八问 为什么说这是“唯一之路”？

现在可以回到卷首的命题。前面并不是先证明“为别人着想总有回报”，再要求大家照做。四道难题已经说明，那样的保证不能成立。需要改变的，是让某一方的成功天然建立在另一方失去未来之上的生成方式。

先看旧的完成标准怎样继续向前。如果你只想赢得一份订单，一局有一局的边界；如果你只有在对方再也不能影响自己时才认为安全，一次领先就不能结束。对方可以追赶，便要限制其能力；能力可以重建，便要继续处理重建的条件；条件仍由其他人形成，处理又会追到那些人及其选择。每一轮都有局部理由，整体却越来越接近取消另一方的下一步。

在这种结构中，智能越强，只提高执行能力就越不够。系统可能更快发现对方的依赖，更有效地安排资源，也更及时地帮助人阻断自己不愿面对的变化。若这些能力始终服务于同一个无限扩张的完成标准，局部成功就可能成为下一轮扩大控制的条件，而不是让关系安定下来的终点。

关键在于：限制已经成功以后，什么会让追逐停止？若停止条件仍是“以后永远没有人能改变我”，今天的成功就只消除了今天看见的障碍；新的能力、新的参与者和新的变化，还会成为下一轮要处理的对象。一个不断要求继续排除的完成标准，不能单靠更高的执行效率获得终点。

另一方并不会因为在你的计划里被处理完，就自动离开因果。它可能退出、合作、重新组织，也可能在基本未来被压低而高后果能力仍存在时，使原来的约束失去部分作用。综合实力不等于覆盖所有作用通道，暂时有效也不等于面对新能力仍然有效。于是，更快实现旧目的，有时不是更快取得安全，而是更快走到旧目的没有处理的后果面前。

这条因果针对的是上述完成标准、能力变化与相互影响的组合，不把所有自利、所有竞争都判成灾难。它之所以要求改变底层，正在于危险并不只是某一个办法没做好；那个不断要求“还要进一步”的目的，会在新的能力与新的条件下再次选择办法。只修一次行动，不一定改变下一次为何行动。

改变手段，还要改变手段所服务的成功

有效控制能够阻断具体伤害，制度能够分配责任、保护选择，合作能够创造不同于相互排除的机会。一条危险通道被真正关闭，一个人因此活下来，一次错误仍来得及更正，都是现实的保护。改变底层逻辑，需要这些保护为认识与修订留下时间，而不是先把它们拆掉。

但阻断一条通道与改变生成方向，承担不同责任。前者回答这次危险怎样停住，后者回答能力继续增长时，为什么不再把制造同类危险当成成功路径。若一个决定者始终把共同规则当作暂时绕不过去的成本，一旦取得足够优势就希望免除自身约束，规则便持续面对这种侵蚀。制度可以约束它，却不能仅凭书面存在就证明它的目的已经改变。

回到工作室，若一次分工调整因损坏合作条件而被叫停，而你对下一轮的要求仍然只是“怎样让我独占全部机会”，工具就会继续围绕同一个目的寻找别的办法。若你重新把作品质量、自己的生活与合作者能否继续成长放进成功的定义，下一轮寻找的，就不只是完成原目标的另一个办法。叫停保护了这一次；重新认识目的，改变的是下一次从哪里出发。

因此，所要求的不是用思想替代制度，而是让社会不再一面要求克制，一面把取消他人未来视作值得不断奖励的成功；也让智能不再一面说着保护生命，一面只按已经排除相关生命的目的继续优化。约束仍要存在，但它所保护的东西也需要进入为什么行动、怎样才算完成的判断。

这项改变必须落在两个位置。人类社会怎样安排资源、评价结果、组织竞争，会影响交给 AGI 的目标；AGI 若能够生成和修订目的，又会以这些目的改变人的选择条件。只改人对机器的措辞，不改社会实际奖励什么，旧方向可能继续借机器扩大；只期待机器自行领悟，却让人的目的不接受后果检验，同样没有处理这个往返。

反过来，若人的决定让受影响者的生活与共同条件能够改变目标，智能就有机会比较原先不会被保留的方案；这些方案被采用后，又可能留下能够提出不同意见、修正错误、继续合作的人。下一轮能够看见什么、向谁求助、还有哪些选择，也随之改变。方向的变化不是在结果后面添一句善意说明，而是从起点开始改变后面会形成怎样的关系。

把自己这一侧的利益先定义完整，再让别人适应，是原来的默认。让相关他者与共同条件先进入，使它们能够改变目标和目的，是要替代它的生成方向。“先”不是永远让别人多得一点，而是不等决定已经不能收回，才承认那些承受结果的人原来也属于这次决定应当面对的世界。

“唯一”指向什么，具体办法为什么仍然可以不同

本书所坚持的前提，是人类社会的发展不能以取消不同人的生命、基本选择与共同存续条件来完成。不让任何人拥有差异，或者让其中一方永久消失，可能使一种冲突不再出现，却没有保住原本要保住的人。以这种方式取得的安静，不是这里所说的共同未来。

在这个方向要求下，不能继续把“相关他者终于没有未来”当作自己成功的正常依据。只在结果出现以后补偿也不够：生命已经失去，基本选择已经被不可逆地关闭，事后的说明不能使它们像从未受损一样回来。因此，关键后果必须有机会在目标目的形成时改变决定，而不是永远排在决定之后。

更早知道后果，还不等于后果能够改变目的。若收益目标始终不许触动，后果分析也可能只被用来寻找阻力更小、代价更容易转交的做法。让相关他者真正进入，必须包括另一种可能：因为这条路会损害原本应当保护的生活，目标需要重定，方案需要放弃。改变的不是最后多考虑一点，而是什么能够使最初的决定不再成立。

完整利己把长期、全局与自身变化带进认识；边界重认让“什么构成我的延续”不再预先固定；利他为先的利己，则使这种重认进入新的目的生成。三个位置各有作用。它仍然保留自己，不要求以放弃自己来证明对别人的关心。

这就是本书正面提出的判断：

让“利他为先的利己”替代“利己为先”，成为人类社会运行、AGI 目标目的生成的底层逻辑，是避免 AGI 时代人类社会出现不可逆后果的唯一之路。

“唯一之路”所指的是这种底层方向的不可绕开，不是唯一名称、唯一组织或唯一制度。人可以通过不同规则保护彼此，系统可以采用不同技术实现约束；这些办法仍需根据事实比较。某个办法有效，不应被抹掉，也不能只因它有效，就被事后改名为已经完成边界重认。行为效果与目的生成仍是不同层次。

上述推演说明了本书为什么主张改变这一方向；它没有穷尽证明一切可能安排都只能以同一种机制成立。无回报时他者为何仍有理由分量、真正不能两全时如何取舍、先改变的人如何继续存在，仍需在各自的位置回答。尚无完整实证，不是停止推演的理由；强命题也不能代替这些问题的答案。

还必须分清必要方向与充分保证。指出需要改变，并不等于每个人都会自动愿意改变；改变了目的生成，也不等于已经拥有阻止全部伤害的能力。看清危险不会自动修好制度，接受共同未来也不会自动解决资源冲突。方向改变、现实保护与持续纠错，需要相互支持，不能互相冒充。

把这一问带回工作室，就会看见它并不遥远。你可以努力让作品更好，让收入更稳，甚至选择与朋友各走一条路；但这份成功不必以他从此无路可走来证明。把这种不同带进许多安排，再带进智能如何形成目的，文明方向才不只是一本书里的结尾。

继续追问：原典第十至十二章保留赢、不可逆阈值与“唯一之路”的论证责任；第十五至二十章保留反例与未闭环问题；第二十五、二十六章说明方向不能替代必要安全与现实权限。

第九问 别人不改，我先改，会不会吃亏？

会有这种可能。更认真地考虑他人，不保证别人也认真考虑你；愿意维护共同关系，也可能被只想取走好处的人利用。若一本书只回答“长远看总会值得”，它还没有回答你的担心。

工作室的难题就在眼前。你愿意重新讨论分工，对方却可能只要求旧收入不变，不愿承担新任务；你为了保留对方的机会增加投入，自己却可能连生活都难以维持。这些损失不能被“共同未来”四个字抹掉。

一种具体安排若长期让愿意共同承担的人单方面受损，就需要被改变，而不是要求受损的人证明自己足够高尚。不能靠淘汰采用者来维持一条值得采用的原则。原典也明确保留这个问题：先采用新关系逻辑的一方，怎样在未采用同一逻辑的环境中继续存在？它没有现成的普遍解。

但从“可能受损”也不能直接推出“唯一办法是比别人更早排除别人”。需要比较具体路径。有些额外考虑只需要在决定之前认真听清和查明，不等于提前交出全部利益；有些合作需要双方都作出能够落实的承担；有些关系则确实不值得继续。

例如，工作室可以讨论一次有明确范围的分工调整：哪些工作确实被工具替代，哪些能力仍需要人，双方愿意承担什么，投入是否超出承受能力，什么时候重新比较结果。对方没有履行承担时，你仍然可以缩小合作或离开。这不是保证成功的处方，而是把“我先改变”与“我单方面承担所有成本”拆开。

对方不肯共同承担时，你不必先获得他的认可，才有资格保护自己。结束一段无法继续的合作，与把对方永远排除出所有机会，不是同一件事。拒绝一个不合理要求，与否认对方还有自己的生活，也不是同一件事。

同样，不能把“多理解一步”变成对正在遭受伤害者的要求。一个人可以先离开危险、停止继续投入、寻求适当帮助，而不必先完成对伤害者的全部理解。更有决定权的一方，应当承担的后果审视，也不能被转嫁为弱势一方必须先宽容。

还有些问题确实超过个人尺度。一个行业若让谁少保护人、谁就更便宜，个人善意可能难以独自抵挡。这时需要把问题带回共同规则、责任分配与合作条件，而不是告诉每个人靠更好的心态解决结构性压力。本书不在这里替所有场景写出统一制度答案，但会保留这个不能由个人代付的困难。

在这样的场景里，社会一侧需要重新检查竞争要求和成本分配；智能一侧也不能只被要求更精准地选出谁最容易承受损失。否则，“更聪明”可能只是把原来的不对称安排执行得更稳定。需要进入目标的，不只是眼前总收益，还有采用者与承受者能否继续存在于这套关系之中。

所以，本书没有要求你成为毫无防备的人。它要求的是：防备有所针对，承担有实际边界，合作允许纠正和退出；不能因为保护自己，就把永久消除他人的选择当作唯一完成，也不能因为考虑别人，就把自己的选择交出去。

关心他人，不以交出自己为证明。共同未来，也不能只向愿意共同的人收费。

继续追问：原典第十八、二十一、二十二、三十章讨论防御、拒绝、退出及率先采用者的存续难题；所引《人类与 AGI 关系愿景》V1.1 第十节、十一节继续处理不对称条件。

第十问 考虑别人以后，我还能说“不”吗？

能。不能拒绝的关心，很容易变成强加的义务；不能退出的共同，很容易只剩服从。保留自己的判断，不是这套理论需要克服的障碍，而是它必须保护的条件。

你可以理解朋友失去收入的困难，同时拒绝自己无法负担的长期安排；可以认真考虑客户的需求，同时拒绝不合理要求；也可以承认一项关系过去很重要，今天却已经不适合继续。承认别人的处境，不等于承认他提出的每一种解决办法。

面对真的会造成伤害的人或系统，首先需要阻止伤害。某种能力、接入方式或持续活动确实危险时，中断活动、隔离危险能力、限制相关权限，可以成为保护生命与共同未来的必要措施。承认对方的处境，不等于容许伤害继续，也不要求等待对方认可以后才采取必要保护。

对 AI 系统尤其不能把会说共同未来、会表达受限，直接等同于已经证明自己应当获得更多权限。是否维持关停、隔离或限制，要看具体风险、专业依据与现实授权；语言上的承诺不能替代这些理由。

必要防御也不是取消对方全部未来的无限权力。阻止一个具体危险，要说明危险在哪里、措施限制了什么；当前没有安全办法，不能因此被写成永远不可能有；一次错误，也不能被变成一个不能更正的身份。能够通过限制某种行为阻止伤害，就不应把范围无端扩到其他基本选择；证据变化后，也仍应重新判断。这些要求约束的是处置，保护的并不是伤害继续发生的机会。

一句“为了大家好”不因此取得替别人决定的权利。这里有几道必须说清的界限：共同未来不是违法或越权的许可；“不像利他”不是取消必要关停、隔离或审查的理由；愿景不能代替事实与风险依据；发现方向冲突首先要求重新审查，不自动命令任何人放权或停止；一项能力或一次授权受限，也不自动等于基本未来被剥夺。

这些界限同样约束作者。一本书不能一面反对单方决定别人的未来，一面因为自己声称方向更高，就要求别人放弃异议。你可以要求它说明理由，也可以不同意它的价值选择。

一个值得维持的“我们”，必须允许其中的人仍然是不同的人。你有你的生活，我有我的选择；责任可以共同承担，感受却不能被一个集体名称替代。关系可以改变对自己的理解，却不能让别人因此成为你的所有物，也不能让你失去自己的边界。

这就是原典所说的“开放稳定”：不把已有关系和已有名单当作最终答案。新的处境到来，旧判断仍能被重新检查；新的智能出现，既不因为它自称智能就自动认定它享有一切资格，也不因为它不是人就预先关闭研究与判断。是否属于某种主体、是否获得某项权限，都需要自己的理由。

开放不等于所有门都敞开，稳定也不等于门永远焊死。可以关门，可以保留限制；重要的是理由仍然可以被说明，证据变化以后判断仍然能够重新作出。

“我们”之所以有意义，不是因为从此没有“你”和“我”，而是因为你和我仍能在其中继续成为自己。

继续追问：原典第二十一至二十六章展开拒绝、退出、开放稳定与方向的自我限制。第二十八章另说明，理论不能把期待答案写进测量，再用结果证明自己。

第十一问 我现在能改变什么？

先不必给自己换一种身份，也不必宣布接受整套理论。在一件普通、非紧急的决定里，把“我已经决定，你来配合”稍微向前挪，给事实和相关的人一个真正可能改变决定的位置。

这不是让你事事犹豫。很小、可恢复的选择不需要无限推演；危险正在发生时，应先保护安全。值得多走一步的，是那些将影响他人重要选择、长期关系，或一旦作出就难以收回的决定。

回到工作室。你仍然可以决定调整分工，甚至结束合作。但在决定锁定以前，可以把下面四个问题放到面前。它们没有分数，也没有规定你最后必须选择哪一项。

我到底想得到什么，什么才算已经够了？

是减少工作时间、维持收入、让作品更好，还是只有别人再也不能改变我，我才觉得安全？这几种目的会导向不同安排。先把它们分清，才能发现自己是否把一种手段误当成了没有尽头的完成标准。

这条路会改变谁的选择，又怎样回到我这里？

把你知道的事实，与你对对方反应的猜想分开。不要因为他没有说话，就认定他不会受影响；也不要因为你担心他反对，就认定最坏反应已经必然发生。能向当事者了解的，不必全部在自己的模型里替他回答。

听完这些，我只改办法，还是也愿意重看目的？

你可能发现原目标合理，只需改善做法；也可能发现，原来的成功把你本来珍惜的一部分生活排除了。两种结果都需要具体理由。听了别人之后一定让步，不是理解；听完任何理由都不可能改变，又可能只是礼貌地收集信息。

我的边界在哪里，没有回报时我仍愿意承担什么？

你可以不承担超出能力的成本，也可以提出对方需要共同承担的条件。与此同时，问清哪些理由并不只是未来收益：尚未结束的责任、对方真实承受的损失、你愿意保留的一种关系。它们是否改变决定，由你认真判断，而不是由本书给出道德分数。

四问之后，你可能保留原决定，但更清楚自己为什么这样选；可能调整分工；可能选择退出；也可能暂时无法决定，需要补充事实。这些结果不能只按是否符合作者期待来判优劣。

把同样的问题带进你交给 AI 的事情

当你请 AI 给出一个安排，不只告诉它要提高哪个数字，也说明数字为了什么、哪些人的重要选择会被改变、哪些事实应当使目标本身重审。再检查它给出的方案是否真的回应了这些内容，而不是只在最后附上一段好听的提醒。这是改善使用方式，不是证明系统已经自主形成了同样的目的。

当你参与团队或组织的决定，也可以追问：一种方案为什么总能胜出，是因为它真的更好，还是因为代价恰好落在没有发言位置的人身上？改变可能从一个具体问题开始，不必先拥有改造整个社会的能力。个人、组织与更强智能各自能够改变什么，应当与它们实际的能力和和责任相称。

理解以后仍然不赞同，可以把不同意见放在具体位置。你可能接受他人的变化会返回自己，却不接受某种责任范围；可能同意不应预先排除别人，却认为书里没有充分说明如何避免先改变者吃亏；也可能认为某个思想实验的设定不能代表你所处的情景。这样的疑问比一句赞成或反对更能推动讨论。

还有几件事，这本书没有资格替你宣布已经解决：人在看清后果之后为什么有时会改变、有时不会；当所有回报都被排除，哪些理由仍然成立；真正无法两全时怎样取舍；怎样使先改变的人不被系统性淘汰；怎样判断一次看似体面的选择来自什么过程，而不凭结果猜测内心。

这些未回答完的问题，不是请你放下生活等理论成熟。它们提醒我们在已有判断时，不把“能纠正一些误判”说成“保证所有人认同”，不把“可以改善一些关系”说成“保证没有损失”，也不把一个生活中的好结果当成整套文明方向已经被证明。

文明跃迁要推动的，是人类社会运行与 AGI 目标目的生成的底层转变：不再默认先把自己和自己的胜利定义完整，再把其他人当作需要安置、利用或排除的剩余项。相关他者与共同条件，应当在目标目的形成时就能改变它们；自己的利益、责任、拒绝和退出也仍然保留。能力可以帮助我们看得更远，但看见哪些后果，与让哪些后果改变成功的定义，仍是两件事。

这种改变需要在社会与智能的往返中持续成立。人不再把取消别人明天的目的交给机器，也不再把机器实现目的的效率误认成文明进步；智能的目标目的生成，也不能只围绕已经排除他者的成功标准继续放大。怎样让这一改变真实发生，又不让率先改变的人被牺牲，是我们需要继续回答的问题。

我仍然是我。你仍然是你。

但我的未来，不再建立在你的终结之上。

继续追问：原典第二十、二十二、二十四、三十章保留上述开放问题。四个生活问题是本稿为理解而新写的反思入口，不是人格评价、行为测量或已验证的干预方法。

想继续读，问题往哪里走？

如果你关心“这一判断最初怎样出现，又在哪里被反例改变”，请读《文明跃迁底层方向校准原理：完整因果推演与命题形成》论证伴随卷 V2.1。各问末尾已经标明对应章节。原典不是要求你先读完的门槛，也不是不同意本书以后必须接受的裁判；它让你能进一步指出某一步为何成立，或者为何仍不成立。

如果你更关心“人类与未来智能究竟应当保持怎样的关系”，可继续读《人类与 AGI 关系愿景：从控制存亡化到共同展开》V1.1。它承接本稿关于控制与自主的疑问，继续处理可信的调整条件、相互限制及不对称竞争；其中的未闭合问题，不因本稿面向公众而消失。

如果你想把这些选择放回文明整体，可继续进入《文明跃迁》五卷书：《被需要的文明》《制度前语言》《文明跃迁白皮书》《文明永续》《意义涌现》。这里的问题不应止于个人怎样相处，也会继续进入人的价值、制度方向、长期延续与意义。但个人不能被要求独自承担所有尺度的问题，任何一卷也不能代替全部专业事实。

真正有用的回应，不限于赞同。你可以带来一个让本书解释失效的真实情景，也可以指出某一问没有接住你的困难。新的事实、不同的价值理由和更好的替代办法，都可能使这份文字需要改变。

来源与编写说明

本书的写作与排版有多个 AI 工具参与辅助，内容取舍与最终定稿由作者负责。

本稿以《文明跃迁底层方向校准原理：完整因果推演与命题形成》论证伴随卷 V2.1 为直接解释依据，承接其所对应的《文明跃迁底层方向校准原理 V1.2 FROZEN | 跨项目上游方向基线》，并参考《人类与 AGI 关系愿景》V1.1。编辑依据为《统一编辑指引 V5.7-R1》。卷首与第八问的中央命题采用作者确定的原句；关于人类社会运行与 AGI 目标目的生成相互影响的说明，是经作者授权的公众解释性展开，不冒充原始对话逐字记录，也不据此改写方向母稿的冻结命题。

十一个主问题为组织阅读而设，不是陌生读者访谈汇总。正文保留已进入原有推演的关键困难，减少预拟反对者的台词，不替读者穷尽所有反驳。本版尚无普通读者实读验证记录；理解后不同意，不自动构成阅读失败。

工作室及其他情景均为解释性设定，不是现实案例、调查结果或原始推演记录。它们说明局部关系与条件路径，不能直接证明文明尺度的现实风险、未来系统的实际能力或某个人的内心状态。未来情景按条件讨论。关于能力、主观体验与主体资格的区分，承接《统一编辑指引》所收 T-10 认知起点的证据边界，不将未来可能性写成当前事实。

本版为公众论证与答疑版 V1.2，以 V1.1 为底稿，定向补充决定权与责任、底层改变的因果连接及必要防御的说明，作为本轮公众首读文本。V1.1 保留为前序版本。原典 V2.1、方向母稿 V1.2 FROZEN 及其他文稿未因本次修订而改动。

读到这里，问题可以继续

哪一句让你停了下来？哪一步与你的经历不同？或者，你看到了另一种可能？

你可以只引用一句原文，写下自己的问题，也可以回应其他读者。不必先赞同这本书，才有参与讨论的资格。

[提出问题与参与讨论](#)

<https://www.civitas.top/public/dialogue/index.html>

[查看重要更新与通知方式](#)

<https://www.civitas.top/public/dialogue/updates.html>

想继续阅读但暂时不想发言，可以单独选择通知方式。现有 GitHub 话题通知需要读者自行登录订阅；RSS 供已使用订阅阅读器的人选用。打开页面不等于已经订阅。独立邮件订阅尚未开放。

不便公开的问题，可邮件联系 zijunfu@civitas.top。来信不会自动公开，也不会自动成为订阅；公开摘录前另行确认许可。

[继续进入《文明跃迁》五卷书](#)

[在公开文库查找相关文稿](#)

原典 V2.1 与《人类与 AGI 关系愿景》V1.1 的精确在线全文地址本次尚未核实；上面的文库入口用于查找相关公开文稿，不冒充这两份原文的直达链接。

首发阅读本 | 2026 年 9 月 20 日。保留公众版 V1.2 全部正文；本页为新增交流、更新与续读入口，不构成理论升版。