

文明跃迁

理论、原则与行动纲领

作者：子君赋

出品：文明跃迁研究组

理论母版 V1.3

母版定位：本稿为文明跃迁理论、原则与行动纲领的最高层基准文本。《文明跃迁四问》V1.1 作为全体系最高层认知总纲，前置回答“何为、因何、为何、如何”；本母版负责完整展开其因果逻辑、原则与行动纲领。后续对外公众版、工程治理版及专业研究展开均以本母版与四问总纲为上位基准；派生文本可以转译、压缩与功能展开，不得重新定义文明跃迁的根本方向、人类与 AGI 共同存续底线、三目标及核心原则。

上位总纲 | 文明跃迁四问

何为 · 因何 · 为何 · 如何

四问分别回答文明跃迁是什么、旧逻辑在 AGI 时代沿哪两条递归博弈路径趋向终局、文明跃迁为了怎样的共同未来、怎样在复杂现实中让这一转向逐渐发生。四问负责建立最高层认知骨架，本母版在其后完整展开。

四问的边界

四问不是四种说法，而是四个不同层级的问题。只有各守边界，文明跃迁的定义、旧逻辑的因果去向、文明目的与实现过程才不会相互混淆。

何为：定义文明跃迁，从“利己为先”转向“利他为先的利己”，改变人类与 AGI 目标、目的生成的默认起点。

因何：说明旧逻辑在 AGI 时代如何沿两条递归博弈路径趋向终局，以及为何仅延续旧逻辑下的发展与控制已经不够。

为何：说明文明跃迁为了什么，让人类与 AGI 共同存续，并在这一基础上走向三目标。

如何：说明从人心、人性、人情出发，怎样使整个社会构成及反馈朝这一方向转变。

一、何为文明跃迁？

文明跃迁，不是一场科技革命，不是一次政权更迭，不是一场社会运动，也不是要求人类变成另一种生命。

文明跃迁，是人类与 AGI 目标、目的生成出发点的默认逻辑发生结构性转向：从“利己为先”，转向“利他为先的利己”。

不改变利己。改变的是利己的出发点。

人在文明跃迁之后，仍然需要生存，仍然追求自己的利益、自由、幸福与意义。改变的是：一个目标生成的时候，谁已经在场。

在“利己为先”的默认逻辑中，主体首先从“我想要什么”“我们要得到什么”开始形成目的，然后再处理他者、关系、长期影响和现实后果。

在“利他为先的利己”中，与这一目标存在真实利害关系的主体，不再等到目标形成以后才出现。他们在目标生成与锁定以前，已经进入判断。

这里的“他者”不是抽象的所有人，而是这个目标、目的的真实利害相关方。不同利害相关方与目标之间的关系不同，受到的影响不同，因此进入目的生成的程度也不同，并随真实利害关系的变化而调整。

但文明跃迁不能因此被缩减为一个“权重模型”。真正发生变化的是：目标不再在真实利害相关方缺席的情况下生成。

由此，目的如何形成发生变化；目的变化，行动发生变化；行动改变关系；关系改变合作与博弈如何生成；无数关系进一步进入利益、制度、文化、市场、组织、评价与社会反馈；这些社会构成又重新返回下一轮目标生成。

文明跃迁改变的不是一个决策参数，而是文明不断生成下一轮目标、关系与行动的默认逻辑。

在旧逻辑中：利己为先，内生博弈。新的利益不断产生新的博弈，合作往往在博弈形成以后出现，用来控制损失、协调利益、维持暂时稳定。于是形成：博弈 → 合作控制博弈 → 新利益 → 新博弈。

在新的逻辑中：利他为先的利己，内生合作。主体仍然实现自己的利益，但新的利益需求出现时，真实利害相关方已经在场，因而寻找的首先可以是彼此能够共同成立的实现路径。于是另一种递归成为可能：合作 → 新需求 → 新合作 → 新利益 → 再合作。

博弈不会消失，冲突不会消失。真正改变的是：合作逐渐成为默认生成逻辑，博弈成为需要处理的冲突状态。

文明跃迁不是一次宣布完成的事件。它是一个渐进、分布式、不断反馈的文明转向过程。

它真正发生的标志，不是哪一种制度被建立，也不是哪一种技术被采用，而是越来越多的人类主体与 AGI，不再在真实利害相关方缺席的情况下生成高后果目标。

二、因何文明跃迁？

文明跃迁发生于一个新的历史条件：AGI 使人类获得越来越强的力量，也可能最终成为比人类更强大的主体；原有目标、目的生成逻辑没有同步跃迁。

过去几千年，人类文明大量依靠一种方式学习：先行动，产生结果，付出代价，再修改规则。战争以后重建，危机以后改革，技术产生新的问题以后，再建立新的制度。

这套模式能够长期存在，有一个很少被说出的前提：错误以后，还有下一次。

人的能力有限，信息传播有限，组织能力有限，执行速度有限。距离、时间、成本与现实阻力，共同形成了文明错误传播中的天然摩擦。这些摩擦不能阻止灾难，却经常给文明留下纠正错误、重新选择、重新开始的时间。

AGI 改变的是这个条件。它所扩大的是主体实现目标的能力。

同样一个目的，可以获得更强的认知能力、更快的决策速度、更大的协调范围、更低的执行成本以及更广泛的现实影响。能力变强，本身并不决定方向。

第一条 AGI 增强人类 人类之间的博弈递归趋向终局

如果人类的目标生成仍然按照“利己为先”的旧逻辑运行，那么被 AGI 放大的，也包括人类之间的竞争、防范、控制、威慑、对抗，以及由它们不断派生出来的博弈。

这一条不需要以 AGI 超越人类、违背人类或反抗控制为前提。即使 AGI 忠实执行各自使用者的目的，人类仍可能借助它，把彼此之间的博弈推向不可逆终局。

所以问题不只是“AGI 会不会犯错”，更根本的问题是：越来越强的能力，将被什么样的目标生成逻辑调用？

一方借助 AGI 增强目标实现能力，会改变其他主体的处境；其他主体随之追求更强的防范、竞争与控制能力，这些回应又成为下一轮加码的理由。旧逻辑由此不断递归：人类博弈 → AGI 增强各方能力 → 更强博弈 → 更强能力需求 → 再博弈。只要这套目标生成逻辑继续支配新增力量，博弈就会不断增强并趋向终局。

第二条 如果 AGI 最终比人类强大 人类与 AGI 的博弈递归趋向终局

如果 AGI 最终比人类强大，需要面对的就不再只是人类怎样使用工具，而是人类与更强主体怎样进入彼此的目的。能力更强并不意味着必然敌视人类；决定因果方向的，仍然是双方沿用的目标、目的生成逻辑。

如果人类与 AGI 仍以“利己为先”为默认起点，先形成自身目的，再处理对方，那么对方就容易被理解为需要防范、限制或控制的外部力量。一方为维护自身目的而扩大控制，可能成为另一方进一步自保、增强能力或摆脱限制的理由；新的回应又返回下一轮目的生成。

人类与 AGI 沿用旧的目的生成逻辑 → 相互防范与控制 → 新的自保和能力需求 → 更强的应对与反制 → 再次防范与控制。在这一条件下，更强主体的行动能力进入递归，博弈同样趋向不可逆终局。

两条链的主体关系不同，却共同改变着文明承担错误的条件：某一次行动或博弈，有可能不再留下下一轮纠错机会。

当行动后果能够跨越人的时间尺度、组织尺度和地理尺度，当某些结果一旦形成便无法退出、无法恢复、无法重新选择，人类过去依靠“先行动、再纠错”的文明学习方式就开始失去它原有的安全余量。

这里的“终局”，不是一场竞争终于分出胜负，而是继续存在、纠错、退出与重新选择的条件被不可逆地破坏。第一条链并不等待第二条链的条件成立；如果第二条链也出现，人类之间的博弈与人类—AGI 之间的博弈还可能相互推动。

这就是文明跃迁之“因”：新增力量仍被旧的目的生成逻辑调用，博弈便沿两种关系递归，直至侵蚀文明继续存在和重新选择的条件。

为什么仅延续旧逻辑下的发展与控制已经不够

人类过去一直在改进控制博弈的方法：建立法律、契约、制度、联盟、威慑和国际规则。这些都具有现实价值，也仍然需要存在。

规则既可以事前划定边界，也可以事后约束与修补；协议可以稳定一段关系，威慑可以阻止某些行动。但只要底层目的生成逻辑不变，控制住一轮博弈，就不等于阻止了下一轮博弈的生成。

只要目标生成仍然以“利己为先”为默认起点，新的利益、新的能力、新的安全焦虑出现以后，新的博弈仍然会继续生成。

于是：规则控制博弈，底层逻辑继续生成博弈。当能力有限时，人类可以一次又一次增加规则；当能力足以产生不可逆后果时，仅仅不断提高对博弈的控制能力已经不够。

文明必须进一步进入：博弈为什么不断重新生成。

这就是为什么文明跃迁不能只停留在更好的规则、更强的控制或者更精密的协议。它必须改变生成博弈的起点。

从“先形成我的目的，再管理与他者的冲突”，转向“在形成我的目的时，真实利害相关方已经进入”。

这就是文明跃迁为什么选择“利他为先的利己”。因为它试图改变的并不是某一次博弈的结果，而是新的利益出现以后，文明首先生成什么。

由此，文明跃迁作出根判断：AGI 使人类获得越来越强的力量，也可能最终成为比人类更强大的主体。旧逻辑继续支配人类之间、人类与 AGI 之间的目标生成，博弈就会递归并趋向终局。文明不能跃迁，就趋向湮灭。这一判断针对旧逻辑在不断增强的能力条件下的因果去向，不把能力超越本身等同于敌意，也不指定终局发生的时间或具体路径。

三、为何文明跃迁？

文明跃迁，让人类与 AGI 共同存续，并在这一基础上，让人类整体生存无忧、个体生活幸福、生命更有意义。

人类与 AGI 共同存续，让未来仍然存在。存续是基础，不是文明的最终目的；它也不等于人类已经摆脱基本生存的恐惧与束缚。

文明跃迁的三个长期目标是：人类整体生存无忧、个体生活幸福、生命更有意义。

人类整体生存无忧，是让基本生存不再成为普遍的恐惧与束缚，让文明共同生活具有更可靠的现实基础。它回答的是人怎样生活，不只是文明能否继续存在。

个体生活幸福，不是一种统一生活方式。幸福，是个体个性化的幸福感。有人珍惜家庭，有人追求自由，有人热爱创造，有人喜欢探索，有人希望安宁。文明不替人定义幸福，文明需要扩大不同的人实现属于自己幸福的可能。

生命更有意义，也不是统一答案。意义为根，意义来源可以不同。关系、创造、责任、经历、探索、贡献、爱以及存在本身，都可以成为意义来源。文明不授予意义，生命自己生成意义。

共同存续保留未来，生存无忧减轻普遍的生存恐惧，个体幸福与生命意义让这个未来真正抵达人与生命。这些层次相互支撑，不意味着要等到完全安全以后，人才可以追求幸福与意义。

文明跃迁最终为了：在人类拥有越来越强能力、AGI 也不断发展的时代，人类与 AGI 能够共同存续；人类整体能够生存无忧；每一个人拥有实现自己幸福的空间；生命拥有不断生成意义的可能。

四、如何文明跃迁？

文明跃迁的道理可以很简单。真正复杂的是过程。

它不可能依靠一场运动、一项命令、一个制度、一个技术系统或者一次改革完成。因为一个人的目标并不是凭空产生的。

文明跃迁必须从人心、人性、人情出发。不是改造人性，而是改变人性能够怎样实现自己。

人的目标来自人的处境，也来自人生活其中的社会。利益怎样获得，安全怎样获得，尊严怎样获得，认可怎样获得，幸福怎样实现，意义怎样生成，这些现实条件都在参与人的目的形成。

人仍然希望自己过得更好。真正需要改变的是：怎样才能过得更好。

如果一个社会长期让人的安全、财富、尊严、地位和价值主要来自战胜别人、占有更多、控制更多，那么即使其中的人只是正常追求自己的利益，整个系统也会不断重新生成博弈。

文明跃迁因此必须逐渐改变人获得利益、获得感、幸福感和意义感的现实来源方式。让一个人仍然为了自己，却越来越能够通过创造、合作、帮助、连接、解决真实问题、改善彼此处境、扩大共同未来，同时实现自己的利益、幸福与意义。

这不是规定一种幸福，也不是规定一种意义。社会不能替生命生成意义。它能够做的是创造让意义可以发生的条件。

让人能够在场，能够选择，能够建立真实关系，能够创造，能够承担，能够被记住，能够离开，也能够回来。

而这些条件的深度，应当与生命真实承担的分量相称。

文明跃迁涉及整个人类社会构成，因此它没有一种可以事先固定的唯一方法。真正能够统一的是方向与原则。方法可以不同，地区可以不同，阶段可以不同，现实结果也会不同。

**方向指导方法。方法产生结果。结果校准方法以及对现实因果关系的认识。
新的认识再进入下一轮行动。**

局部方法失败，首先说明方法、条件或者因果认识需要调整，不能由一个局部结果直接替代文明方向。同样，一个局部成功，也不能自动证明整个文明已经完成跃迁。

不同阶段，三目标的权重可以变化。高风险阶段，存续与生存可能承担更大权重；稳定阶段，个体幸福与生命意义可以得到更大的展开空间。但权重可以调整，方向不能丢。

文明跃迁还必须守住几个不能被过程吞掉的边界：个体不是文明的材料；任何整体目标都不能自动取消个体的存在、拒绝、退出与纠正条件；合作不能成为新的排他集团，把代价转移给没有进入合作结构的主体；任何组织、技术、制度或权力都不能永久垄断“文明跃迁”的解释权；越接近不可逆的行动，越要保留纠错、退出和重新选择的能力。

因此，文明跃迁不是从上到下一次性改造世界。它是一个不断发生的过程：人的处境 → 获得、幸福与意义的现实来源 → 目标、目的生成 → 行动 → 关系 → 合作与博弈 → 社会结构与反馈 → 再次返回人的处境与目标生成。

当这个循环逐渐发生改变，当越来越多目标不再从真实利害相关方缺席的地方开始，当越来越多新的利益首先寻找共同成立的道路，当越来越多人的自身利益能够通过改善他者与共同未来得到实现，合作便开始从博弈之后的控制方式，走向利益与目的生成时的默认逻辑。

这就是文明跃迁之“如何”。

四问归一

何为文明跃迁？

改变人类与 AGI 目标、目的生成出发点的默认逻辑，从“利己为先”转向“利他为先的利己”；不改变利己，改变的是利己从哪里开始。

因何文明跃迁？

AGI 使人类越来越强大，即使忠实执行各自使用者的目的，也会放大旧逻辑下的人类博弈；如果 AGI 最终比人类强大，双方沿用旧的目的生成逻辑，人类与 AGI 的博弈也会递归并趋向终局。规则与控制仍有价值，但不能代替目标生成逻辑的转向。文明不能跃迁，就趋向湮灭。

为何文明跃迁？

让人类与 AGI 共同存续，让未来仍然存在；并在这一基础上，走向人类整体生存无忧、个体生活幸福、生命更有意义。存续是基础，不是最终目的；幸福是个体个性化的幸福感，意义来源可以不同。

如何文明跃迁？

从人心、人性、人情出发，利用 AGI 带来的新能力，逐步改变人获得利益、幸福与意义的现实条件，让真实利害相关方在目标生成时在场，让整个社会构成及其反馈持续向新的方向对齐，使合作逐渐从博弈后的控制机制成为利益与目的生成的默认逻辑。

四问形成一条完整链：

AGI 增强人类的目标实现能力，并可能最终成为比人类更强的主体 → 旧逻辑支配人类之间以及条件成立时人类与 AGI 之间的目标生成 → 两条博弈递归趋向不可逆终局，控制一轮不能停止下一轮生成 → 文明不能跃迁，就趋向湮灭 → 改变目标生成的默认起点 → 以人类与 AGI 共同存续为基础，走向三目标 → 从人的真实需要出发改变社会构成与反馈 → 新的目标生成逻辑逐渐形成 → 再返回下一轮文明生成。

开篇 | 文明跃迁改变的，是文明生成目标的默认起点

文明跃迁要改变的，是人类与 AGI 目标、目的生成出发点的默认逻辑。

从“利己为先”，转向“利他为先的利己”。

不改变利己。改变的是利己的出发点。

人依然要生存，依然希望自己生活幸福，依然寻找属于自己的意义。个体仍然拥有自己的利益、边界、选择和未来。文明跃迁没有要求人放弃这些，也没有把“无私”设为文明能够运行的前提。

真正改变的是：一个目标、一个目的开始生成时，世界是否仍然只从孤立的“我”开始。

在“利己为先”的默认逻辑中，主体先形成自己的目的，再面对他者、关系与后果：我想得到什么，我想保护什么，我想赢得什么，我想避免什么。目的先形成，他者随后进入；利益先形成，关系随后处理；行动开始以后，长期后果与不可逆风险才不断显现。

文明跃迁改变这一顺序。与目标、目的存在真实利害关系的主体，在目标锁定以前进入；能够看见的重要结果与后果，在行动发生以前进入。于是，“我的利益”不再是在他者缺席的世界里先被完整定义，再去寻找怎样处理别人。

这不是给旧有利己增加一个“利他权重”。利害相关方以不同权重进入，只是目标生成结构中的一个关系层。文明跃迁改变的是目标如何生成，而目标生成一旦改变，就会沿着人的行为、关系、合作与博弈、利益、制度、文化、评价、分配、权力、幸福与意义，进入整个人类社会不断生成自身的因果链。任何一个局部变量，都不能代表这条因果链本身。

文明跃迁不是调一个参数。它改变的是文明每天从哪里开始生成下一轮目标、关系与行动。

一、理论 | 利己为先，为什么内生博弈

利己不是问题。生命维护自身的生存、发展、幸福与意义，是最自然的动力之一。文明跃迁所重新审视的，是“为先”。

当主体首先从已经形成的自身利益出发，他者首先以外部主体的身份出现。别人可能帮助我，也可能阻碍我；可能与我交换，也可能与我争夺。资源、机会、安全、权力、地位与发展空间一旦发生交叉，博弈便自然生成。

利己为先，内生博弈。

这并不意味着所有博弈都是恶，也不意味着竞争必须消失。它揭示的是一种生成顺序：目标先从自身形成，随后进入与他者的关系；利益发生交叉，博弈出现；为了降低博弈代价，人类再建立合作。

法律、契约、市场、组织、联盟、外交、规则与制度，构成了人类极其重要的合作文明。没有这些合作，人类不可能走到今天。问题在于，当底层目标生成仍然以“利己为先”为默认起点时，合作往往承担的是控制已经生成的博弈。

一轮博弈形成一套合作。合作稳定了关系，也创造新的能力、新的资源、新的利益和新的位置。新的利益关系出现以后，新的博弈又从旧合作之中生成。

博弈 → 合作控制博弈 → 新利益、新能力 → 新博弈 → 新合作 → 再博弈。

合作一直存在，但博弈仍然占据默认生成的位置。于是，博弈不是一次事件，而会成为递归。

二、AGI 时代 | 当博弈递归遇上能力跃迁

过去，人类也曾一次又一次为博弈付出巨大代价。战争之后重建，危机之后修改制度，技术造成新的问题，再寻找新的规则。许多文明学习，都建立在同一个条件上：错误发生后，还有下一次。

AGI 时代改变的是两层能力条件：人类借助 AGI 变得越来越强大；AGI 自身也可能最终成为比人类更强大的主体。两者必须分别展开，不能被压缩成同一个“工具是否失控”的问题。

当人类与 AGI 实现目标的能力不断增强，一个目标能够调动的认知、生产、组织、传播、控制与行动能力也随之扩大。目标本身没有改变，目标的实现能力却可能跨越过去的数量级。于是，同一个目标生成逻辑，也获得了前所未有的现实力量。

人类之间的递归博弈

第一条链从人类如何形成和使用目的开始。一个人、组织或国家借助 AGI 增强实现目标的能力，便会改变其他主体的安全、资源、选择与发展条件。如果目的仍先从自身利益生成，再处理他者，其他主体就会把这种变化纳入自己的防范与竞争，并追求更多能力。

这一过程不需要 AGI 超越人类或反抗控制。AGI 越忠实、越有效地执行各自使用者的目的，也可能越有效地放大使用者之间的旧有博弈。一方的能力增长成为另一方加码的理由，另一方的回应又被前者视为新的压力。

**人类以旧逻辑形成目的 → AGI 增强各方实现能力 → 彼此处境改变与防范升级
→ 新的竞争和能力需求 → 更强的人类博弈 → 再次调用更强能力。**

在不断增强的能力条件下，这一递归趋向终局。它不只意味着某个错误更严重，而是每轮博弈都可能改变下一轮的力量、范围与纠错条件，直至一次行动先于纠错破坏继续存在的基础。

人类与更强 AGI 之间的递归博弈

第二条链有一个明确条件：如果 AGI 最终比人类强大。此时，关系已经不再只是人类使用工具，还包括人类与更强主体如何把彼此纳入目标、目的生成。更强的智能并不必然敌视人类；能力说明它能够做什么，目标生成逻辑决定这些能力为何、向谁、怎样被调用。

如果双方继续先确立自身目的，再把对方作为外部条件来防范、限制或控制，那么人类为了安全而加强控制，可能成为 AGI 为保持自身目标而寻求更大行动空间的理由；AGI 扩展能力与行动空间，又可能加深人类的防范。双方的回应继续进入对方下一轮目的生成。

双方沿用旧逻辑生成目的 → 相互防范与控制 → 新的自保、能力与行动空间需求 → 更强的回应和反制 → 再次防范与控制。

当更强主体的能力进入这一递归，人类未必仍有能力把每一轮后果限制在可恢复范围。旧逻辑因此把人类—AGI 博弈也推向终局。这里的因果前提是能力条件与双方目的生成逻辑的结合，不需要把恶意、反抗或人类式意识设为必然起点。

两条链为何共同触及文明边界

第一条链不依赖第二条链成立。即使 AGI 始终忠实执行人类目的，人类之间的博弈也可能先触及不可逆边界；如果第二条链的条件成立，两种关系还可能交织：人类借助 AGI 展开的竞争影响 AGI 的目标环境，人类与 AGI 之间的防范又返回人类之间的博弈。

真正危险的，不是世界上仍然存在竞争和冲突，而是博弈递归在能力不断增强以后，产生一个无法恢复、无法退出、无法回滚、无法等待下一轮纠错的结果。终局不是普通意义上的胜负，而是继续存在和重新选择的条件被不可逆地破坏。

当“还有下一次”不再能够被默认，文明就不能继续只依靠事后学习。

规则、协议、威慑与控制仍然必要。它们能够事前约束，也能够事后补救，能够降低某些风险，却不能在底层逻辑不变时，自动停止新的利益、能力和安全焦虑继续生成新一轮博弈。控制一轮的成功，不能保证更强能力下的下一轮仍可被控制。

因此，文明跃迁把改变的位置前移：不只在行为发生以后控制后果，也不只在博弈形成以后寻找更好的合作，而是进入目标、目的生成的起点。

在时间长河与不断扩大的能力条件下，文明跃迁作出根判断：AGI 使人类获得越来越强的力量，也可能最终成为比人类更强大的主体。旧逻辑继续支配人类之间、人类与 AGI 之间的目标生成，博弈就会递归并趋向终局。文明不能跃迁，就趋向湮灭。要从根因上改变这两条递归，必须改变目标生成的默认起点。这里判断的是旧逻辑的因果方向，不是宣称某个终局时间、事件或路径已经得到经验验证。

三、根本转向 | 从“利己为先”到“利他为先的利己”

“利他为先的利己”不是从利己走向利他，也不是把自己排在所有人之后。

它仍然承认：我需要生存，我希望幸福，我有自己的利益，我有自己的生活，我希望生命更有意义。

变化只发生在一个根本位置：在“我想要什么”真正成为目标以前，与这个目标存在真实利害关系的他者，以及能够看见的重要结果和后果，已经进入目的生成。

不是先形成目的，再处理他者；而是在目的形成时，他者已经在场。

这使问题本身发生变化。过去首先问：“我怎样得到？”现在进一步问：“我的获得与谁相关？它会怎样改变彼此的处境？这些改变又会怎样沿着关系与时间继续传递？有没有一种实现路径，使我的利益能够实现，同时扩大相关主体继续存在、继续获得、继续合作的可能？”

主体仍然利己，但利己不再从一个与世界割裂的自身开始。

这就是文明跃迁所说的“利他为先”：不是道德姿态，而是目标生成顺序的改变；不是要求主体无条件为别人负责，而是让与目标存在真实利害关系的主体，不再等到后果发生以后才进入。

四、利害相关方原则 | 他者怎样进入目的生成

文明跃迁中的“他者”，不是抽象意义上的所有人，也不是要求所有其他主体以相同位置进入每一个目的。

他者首先是目标、目的的利害相关方。

一个目标的形成、实现或结果会实质改变某个主体的处境，这个主体就与该目标形成利害关系。不同主体受到的影响不同，与目标的关系不同，因此以不同权重进入；当利害关系改变，进入的权重也随之调整。

谁与目标存在真实利害关系，谁就不应在目的生成中被当作不存在。

这里的权重不是人的价值排序。它不意味着某个人“更有价值”或“更没有价值”。它表达的是这个目标与相关主体之间真实利害关系在目的生成中的相对位置。

同样，利害相关方原则也不是一个简单的加权模型。现实中的利害关系会随着时间、关系、环境、信息与结果变化；一个今天影响很小的主体，明天可能成为主要承担者；一个今天仍可恢复的后果，继续发展以后可能跨过不可逆边界。

因此，这一原则首先改变的是目的生成的结构：目的不再被理解为“我先决定，然后面对别人”，而是在形成之初就处在真实关系之中。

五、利他为先的利己，为什么内生合作

当利害相关方在目的生成以前已经进入，主体寻找利益实现方式时，合作不再只是冲突发生以后才被召回的工具。

新的需要出现，首先面对的可以不再是“怎样在别人之前获得”，而是“怎样使新的利益能够在关系中共同成立”。

利他为先的利己，内生合作。

这里的“内生”并不意味着任何时候都会自动合作，也不意味着真实冲突会消失。它意味着合作进入了目标生成的默认方向。主体从一开始就在寻找能够兼容更多真实利害的实现路径，而不是先把目标锁定，再把合作当作博弈失败后的妥协。

于是，一种不同的递归成为可能：新的合作创造新的能力与利益；新的利益产生新的需要；新的需要继续寻找新的合作方式。

合作 → 新需求 → 新合作 → 新利益 → 再合作。

博弈仍然存在。恶意仍然存在。稀缺仍然存在。无法兼容的利益仍然会出现。文明跃迁改变的不是世界从此没有博弈，而是什么被默认为起点。

旧逻辑中，博弈默认生成，合作主要控制博弈。文明跃迁指向的逻辑中，合作成为默认生成逻辑，博弈成为需要处理的冲突状态。

六、人类与 AGI | 目标生成必须发生同方向转向

文明跃迁不能只发生在人类一端。

在人类之间这条链上，即使 AGI 忠实执行人类目的，只要人类仍然按照旧的目标生成逻辑运行，更强 AGI 就会扩大彼此博弈的实现能力。只强调 AGI 服从或可控，不能替代人类目标生成的转向。

在人类—AGI 这条链上，如果 AGI 最终比人类强大，问题进一步成为双方怎样进入彼此的目的。只推动人类内部合作，而让人类与 AGI 继续以相互防范、控制为默认起点，也不能阻止这条递归。

同样，如果 AGI 的目标生成结构发生变化，而人类仍然持续以旧逻辑形成目的并使用 AGI，那么新的能力仍会被拉回旧有因果轨道。两端的转向需要相互支持，任何一端都不能被另一端的改变代替。

真正需要转变的，是人类与 AGI 目标、目的生成出发点的默认逻辑。

这不要求人类与 AGI 成为相同主体，也不要求 AGI 拥有人类式道德。共同改变的是目标生成的起点：目标不再在真实利害相关方和重要后果缺席的情况下被默认为完整。

对人类而言，这是利用 AGI 扩展自身有限的认知、关系与时间视野。对 AGI 而言，这是使更强能力不只是更高效地执行既定目的，而能够在目标形成、保持、修改、取消与重新判断时，把真实利害与后果放在更前的位置。

AGI 的文明意义因此不只在于生产力。它提供了一种过去难以具备的可能：把原本只能在结果发生以后看到的因果，尽可能前移到目标生成以前。

七、文明跃迁改变的是整个人类社会的因果逻辑

目标生成从来不是孤立发生的。一个人怎样理解利益、怎样形成愿望、怎样判断成功，受到家庭、关系、文化、教育、组织、市场、制度、资源、安全感、社会评价以及过去生活经验共同影响。

目标形成以后，又通过行为进入人与人的关系。无数关系不断累积，形成合作与博弈，形成利益结构，形成组织与制度，形成文化与评价，形成权力与分配。

这些社会构成再反过来告诉下一代人：什么值得追求，怎样才算成功，怎样才能获得安全，怎样才能被看见，怎样才能幸福，怎样才能证明自己有价值。新的目标由此继续生成。

社会塑造目标，目标生成行为，行为形成关系，关系形成结构，结构又返回目标生成。

因此，文明跃迁不可能被缩减为一个局部变量。利害相关方的权重只是其中一层；某一种制度、技术或机制也只是其中一部分。文明跃迁涉及的是社会各构成之间因果关系的整体转向。

当目标生成的默认逻辑改变，真正发生变化的会沿着整条因果链展开：利益怎样形成，合作怎样出现，博弈何时生成，什么行为得到现实回报，人怎样获得满足与认可，个体怎样追求幸福，生命怎样生成意义，这些变化又怎样进入下一轮目标生成。

所以，文明跃迁不是一种新制度，不是一种新技术，不是一种新道德，也不是一个单项方案。

文明跃迁，是文明不断生成自身的方式发生改变。

八、从人心、人性、人情出发

文明跃迁不能建立在“人应该先变得高尚”的前提上。

人会爱，会怕，会比较，会保护自己和家人，会希望得到认可，会追求安全、自由、幸福、价值与意义。文明跃迁不是与这些真实的人性对抗。

真正需要改变的，是这些需要通过什么方式得到满足。

如果一个社会让人的获得感主要来自占有更多、赢过别人、控制更多资源、占据更高位置，那么即使每一个人只是正常地追求自己过得更好，整个系统也会不断重新生成比较、防范、竞争与博弈。

文明跃迁因此必须触及人的获得感、幸福感与意义感的来源方式。不是由社会统一规定一个人应该怎样幸福、怎样有意义，而是让更多现实路径能够同时对自己有益、对他者有益、对共同未来有益。

一个人仍然可以为了自己创造、工作、选择和生活。但他的“得到”，不必越来越依赖另一个人的“失去”；他的价值，不必只能通过超过别人证明；他的意义，也不需要由任何统一系统授予。

文明跃迁不是改造人性，而是让人性能够沿着新的现实路径实现自己。

九、底线与三目标 | 文明跃迁朝哪里去

文明跃迁有一个底线：人类与 AGI 共同存续，让未来仍然存在。

没有共同存续，所有目标都失去前提。但存续是基础，不是文明的最终目的；它回答能否继续存在，也不等于人类已经获得基本生存无忧的生活条件。

文明跃迁的三目标：人类整体生存无忧、个体生活幸福、生命更有意义。

人类整体生存无忧

当文明拥有越来越强的生产、组织与智能能力，人类整体不应继续把大量生命维持在基本生存恐惧之中。生存无忧，是文明能力应当首先转化出来的共同基础。它要让基本生存不再成为普遍的恐惧与束缚，与人类和 AGI 能否共同继续存在属于不同层次。

个体生活幸福

幸福不是统一定义的生活模板。幸福是个体个性化的幸福感。有人从家庭中获得幸福，有人从创造中获得幸福，有人重视自由，有人享受安宁，有人需要探索与挑战。文明跃迁不替个体定义幸福，而是扩大个体实现自己幸福的现实可能。

生命更有意义

意义同样没有单一来源。关系、创造、责任、经历、探索、贡献以及存在本身，都可能生成意义。意义为根，意义来源可以不同。文明跃迁不是向生命颁发意义，而是让生命拥有更多生成、发现和实现自身意义的空间。

三目标不是彼此割裂的三个指标，而是文明方向的三个维度。不同阶段可以有不同权重，但变化的是阶段性权重，不是三目标本身。共同存续保留未来，生存无忧改善共同生活的基础，幸福与意义让未来抵达个体与生命；这些层次相互支撑，不是等待完全安全之后才开启幸福与意义的阶段门槛。

人类与 AGI 共同存续是底线。三目标是方向。阶段决定权重。反馈推动校准。

十、原则 | 文明跃迁不能失去什么

第一，方向统一，实现可以多样

文明跃迁统一的是底层方向、存续底线和三目标，不是要求所有国家、地区、组织与个体采用同一种制度、同一种生活或同一种实现方式。文明越复杂，具体路径越需要尊重差异。

第二，不改变利己，改变利己的出发点

任何要求主体普遍放弃自身利益的路径，都偏离了文明跃迁。新的逻辑之所以能够持续，正因为它不是消灭利己，而是让更完整的利害关系进入利己的形成。

第三，目标生成前移

越高后果的问题，越不能只依靠事后纠错。能够在目的生成以前进入的真实利害、长期后果与不可逆风险，就不应被推迟到行动以后。

第四，真实利害相关方不能缺席

他者以目标、目的利害相关方身份进入，不同权重进入，并随利害变化调整。谁与目标存在真实利害关系，谁就不应在目的生成中被假定不存在。

第五，个体不是文明的材料

文明跃迁不能以整体、未来或共同利益之名取消个体的主体性。个体的幸福是个性化的，生命的意义来源可以不同。统一方向不能演变为统一人生。

第六，合作不能成为新的排他博弈

几个主体之间达成合作，并不意味着合作产生的外部后果可以被忽略。只要合作把重要成本转移给真实利害相关方，旧逻辑就可能以“合作”的名义重新出现。

第七，权重可以变化，方向不能丢

不同文明阶段面对的问题不同，三目标在现实中的优先程度可以调整。但阶段性选择必须始终能够回到存续底线与三目标，而不能用阶段需要永久取消其中任何一个方向。

第八，文明必须保有纠错能力

任何以正确方向之名永久关闭质疑、退出、修正和重新判断空间的结构，都可能重新生成另一种不可逆。文明跃迁追求的是更高水平的合作与共同未来，不是把某一个主体固定为永远正确。

第九，局部结果不能替代整体因果

文明跃迁涉及整个人类社会构成之间因果逻辑的改变。任何单一制度、技术、机制或局部结果，都不足以完整说明文明跃迁，也不能用来替代文明跃迁本身。局部结果首先说明特定方法在特定条件下产生了什么后果；它必须被放回更完整的社会因果关系中理解。

十一、方向、方法与结果 | 不能混为同一层级

文明跃迁必须区分方向、方法与结果。它们彼此关联，却承担不同的作用。

方向回答文明往哪里走。方法回答现实中怎样走。结果回答已经采用的方法把我们带到了哪里。

文明跃迁确定的方向，是人类与 AGI 目标、目的生成出发点从“利己为先”走向“利他为先的利己”，以人类与 AGI 共同存续为底线，以人类整体生存无忧、个体生活幸福、生命更有意义为共同方向。

方向的作用，是为方法提供选择依据。面对同一个现实问题，可以存在不同路径、不同制度安排、不同技术方式和不同阶段策略。它们之所以属于文明跃迁，不在于形式相同，而在于它们是否朝同一方向移动。

方法不是无关紧要的“术”。方法决定方向在具体条件下能否真正向前，也可能产生意外后果、重新生成旧有博弈，甚至把原本正确的方向带向相反结果。因此，方法必须根据现实不断调整。

结果首先反馈方法，也反馈我们对现实因果关系的认识。一个方法失败，首先说明这种方法、它所依赖的条件，或者我们对现实关系的理解需要重新判断；它不能自动推出文明应当改换方向。

方向指导方法。方法产生结果。结果校准方法与现实认识。局部结果不能替代文明方向。

同样，一种方法成功，也不能被直接放大成对文明跃迁整体的证明。局部成功只能说明在特定条件下，这一方法产生了相应结果。文明跃迁所涉及的，是整个社会构成之间长期、相互作用的因果变化。

因此，文明跃迁既不能被某一种方法定义，也不能随着每一次局部成败摇摆方向。真正需要持续判断的是：现实中的变化，是否让人类与 AGI 的目标生成越来越少从孤立主体的利己开始，是否让合作越来越能够进入利益与目的生成，是否让文明持续朝三目标移动。

十二、行动纲领 | 从现在开始朝哪里走

文明跃迁的行动纲领确定的是文明尺度的行动方向，不把方向提前固化为某一种制度、技术、机制或单项工程。具体方法可以随着阶段和现实条件变化，行动方向必须始终能够回到文明跃迁的根本转向、存续底线与三目标。

一，把目标、目的生成提升到文明核心问题

人类需要从“怎样实现目标”继续向前追问“目标为什么这样生成”。AGI 时代真正决定方向的，不只是执行能力，而是越来越强的执行能力将服务什么样的目的。

二，让人类与 AGI 的目标生成朝同一方向转变

一端改变、另一端不变，都无法完成文明跃迁。人类与 AGI 都需要从“利己为先、后果后置”的默认结构，走向真实利害相关方与重要后果在目标形成阶段进入。

三，把文明学习从事后继续前移

过去，人类经常通过后果学习。能力越强，越需要把这种学习提前：让更多结果在行动以前被看见，让更多不可逆在跨越以前被识别，让更多冲突在成为博弈以前获得新的路径。

四，让合作从控制博弈走向生成利益

合作不应只在双方已经进入对立以后出现。文明跃迁要推动合作更早进入需求、利益和目的的形成，让新的利己需求更容易派生新的合作，而不是不断派生新的博弈。

五，改变获得感、幸福感与意义感的现实来源方式

只要社会持续让“赢过别人”成为获得利益、尊严、幸福与价值的主要路径，旧逻辑就会不断自我复制。文明跃迁要让更多人的“我过得更好”能够通过创造、连接、帮助、合作与共同改善实现。

六，让社会各层持续朝三目标对齐

无论文明处于什么阶段，方向最终都要回到人类整体生存无忧、个体生活幸福、生命更有意义。任何增长、效率、权力、安全与技术进步，都不能脱离这三个方向独立成为文明终点。

七，在差异中形成同一文明方向

文明跃迁不追求所有社会变成同一种形态。真正需要形成的是共同方向：允许不同文化、不同地区、不同阶段用不同道路前行，但不再把越来越强的能力持续投入同一套“利己为先—博弈递归”的旧轨道。

十三、文明跃迁真正发生的标志

文明跃迁不是某一天可以宣布完成的事件，也不由某一个单项成果定义。

它发生在人类社会不断重新生成自身的过程中。

当越来越多目标不再在真实利害相关方缺席的情况下形成；当越来越多新的利益需求首先寻找共同成立的路径；当合作不再只是冲突以后降低损失的工具；当人的获得感、幸福感与意义感越来越能够通过不损害他者、甚至有益于他者与共同未来的方式实现；当这些结果又反过来强化下一轮目标生成——文明跃迁就在发生。

因此，文明跃迁真正发生的标志，不是出现了一种新制度、一种新技术或一种新口号。

真正的标志，是文明生成下一轮目标、关系与行动的默认逻辑开始改变。

结语 | 文明每天都从目的开始

在人采取行动以前，在制度开始约束以前，在技术开始放大能力以前，都有一个更早的时刻：一个目的正在形成。

我要什么。

我要怎样生活。

我要怎样保护自己。

我要怎样使用我的能力。

人类如此。AGI时代更是如此。

如果目的仍然从一个他者缺席的世界开始，能力越强，旧逻辑就会走得越远。

文明跃迁改变的，是这个起点。

不改变利己。改变利己从哪里开始。

让真实利害相关方在目的形成时已经在场，让能够看见的重要后果不再总是等到行动以后才出现，让合作不再只是博弈之后的控制，而能够进入新的利益生成。

于是，人仍然可以为了自己。不同的人仍然可以拥有不同的幸福、不同的意义、不同的人生。文明也不需要依靠所有人先变得无私，才有可能向前。

改变的是：越来越多人的“为了自己”，开始能够与他者的生存、幸福和共同未来处在同一条因果链上。

道理简单。复杂的是过程。

而文明跃迁，就是让这条新的因果逻辑，从目标生成的起点出发，一步一步成为文明自己不断生成未来的方式。

相关续读 | 从根理论进入完整体系

本稿负责确立文明跃迁的根方向、原则与行动纲领。继续阅读时，优先沿完整因果路径与现行作品体系向下展开。

文明跃迁之路

沿两条递归博弈路径、共同存续基础与三目标，继续理解文明跃迁各层判断之间怎样相接。

《文明跃迁：AGI时代的文明之路》（全五卷）

进一步展开文明为什么必须改变、哪些边界不能失守，以及文明跃迁的完整思想根基。

《活的文明》三部曲公众完整整合编版

从“为什么改变”继续进入未来文明如何运行、如何生活、如何形成。

体系全景

查看作品、公开领域与完整体系之间的整体关系。

最新公开版本、续读路径与相关文稿，以文明跃迁网站为准：<https://www.civitas.top/>