

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

对外公众版 V1.2

公众版说明：本版以《文明跃迁四问》V1.1 作为上位认知骨架，但不要求读者先理解完整理论体系。它从每个人都会面对的处境出发，回答文明跃迁是什么、为什么在 AGI 时代成为现实问题、最终为了什么，以及我们应朝什么方向改变。具体制度、技术与工程不在本文展开。

上位导读 | 文明跃迁四问

四问各守边界：何为讲目标生成的根本转向，因何讲旧逻辑在 AGI 时代的两条终局因果链，为何讲共同存续基础与三目标，如何讲从人的真实需要出发的过程与原则。

何为文明跃迁？

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。真正的变化在于，一个目标生成的时候，与它存在真实利害关系的主体不再等到事后才出现，而是在目标锁定以前已经在场。

因何文明跃迁？

AGI 让人类越来越强大。即使它忠实执行各自使用者的目的，人类也可能借助它，把彼此的防范、竞争与对抗一轮轮推向不可逆终局。另一条链有明确条件：如果 AGI 最终比人类强大，而双方仍沿用旧的目的生成逻辑，人类与 AGI 之间的博弈也会递归并趋向终局。能力更强本身，不等于必然敌视人类。

规则、协议、威慑与控制仍然有用，也能事前约束行动。但只要旧逻辑继续生成新的博弈，控制住一轮，就不能保证还有机会控制下一轮。这就是为什么新增力量不能只由旧逻辑承接：文明不能跃迁，就趋向湮灭。

为何文明跃迁？

为了让人类与 AGI 共同存续，让未来仍然存在；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。存续是基础，不是最终目的。生存无忧，是让基本生存不再成为普遍的恐惧与束缚；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，让整个社会构成及其反馈持续向新的目标生成方向对齐。方法可以不同，结果也会不同；方向指导方法，结果校准方法和对现实因果关系的认识。

四问归一

何为：改变人类与 AGI 目标、目的生成的默认起点。

因何：AGI 增强人类，旧逻辑把人类博弈推向终局；如果 AGI 最终比人类强大，旧逻辑下的人类—AGI 博弈也会递归趋向终局。发展与控制不能替代目的生成的转向。

为何：让人类与 AGI 共同存续，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。

如何：从人的真实需要出发，让新的目标生成逻辑逐渐进入社会构成及反馈。

开篇 | 一个人想让自己过得更好，没有错

一个人希望自己和家人过得更好，希望生活安全一点、自由一点、幸福一点，希望自己的努力有回报，希望这一生值得。

这没有错。

一家企业想活下去，一个组织希望发展，一个国家希望安全，也都可以理解。人会保护自己，会在意利益，会追求幸福和意义。文明跃迁从来没有要求人放弃这些。

真正需要重新看见的，是当每一个主体都从“我的利益先形成”开始，彼此会怎样进入同一个世界。

文明跃迁不改变利己。改变的是利己的出发点。

今天，人类已经拥有越来越强的技术、组织和智能能力。AGI 将继续放大这种能力。能力越强，我们越需要追问一个比“怎样把事情做好”更早的问题：

我们为什么生成这个目标？这个目标从哪里开始？

一、问题不在利己，而在“利己为先”

人类社会长期存在一种很自然的目标生成顺序：我先知道自己要什么，再去处理别人会怎样。

我想得到什么，我想保护什么，我想避免什么，我想赢得什么。目的先形成。他者随后进入。长期后果随后进入。关系与不可逆风险，也常常等到行动开始以后才逐渐显现。

在这样的顺序里，另一个主体首先以“外部”出现。他可能帮助我，也可能阻碍我；可能与我交换，也可能与我争夺。只要资源、机会、安全、权力和发展空间发生交叉，博弈便自然产生。

利己为先，内生博弈。

这不是说竞争本身是恶，也不是说人类从未合作。恰恰相反，人类创造了法律、契约、市场、组织、联盟和无数合作关系。

问题在于，很多合作是在博弈已经形成以后出现的：为了降低冲突，为了控制损失，为了让彼此还能继续相处，于是合作。

一轮博弈形成一套合作。合作又创造新的资源、能力和利益。新的利益关系出现以后，新的博弈再次生成。

博弈产生合作，合作控制博弈；新的利益再派生新的博弈。

于是，合作一直存在，博弈却仍然占据默认起点。

二、AGI 时代，真正改变的是“还有没有下一次”

过去，人类经常在犯错以后学习。战争之后重建，危机之后改规则，技术造成新问题，再寻找新的约束。

这种学习方式一直依赖一个没有被说出的条件：错误发生以后，还有机会回来。

AGI 时代正在改变这个条件。

当一个目标能够调动更强的认知、生产、组织、传播与行动能力，同样的目标生成逻辑，也会获得更强的现实力量。一个过去只能造成局部后果的错误目的，未来可能影响更广、更快、更深。

先看人与人。一个主体利用 AGI 获得更强能力，其他主体会担心落后、受制或失去安全，于是也投入更多力量；新的防范又让对方继续加码。即使每个 AGI 都忠实服从自己的使用者，旧的目的生成逻辑也能把人类之间的博弈越推越强。这条链不需要 AGI 反抗人类。

再看人与 AGI。如果 AGI 最终比人类强大，人类就要面对一个更强主体怎样形成目的。更聪明、更强大，不等于必然敌视人类；关键在于双方是否仍然先形成自己的目的，再把对方当作需要防范或控制的外部力量。

在这种旧逻辑下，一方加强控制，可能让另一方更想增强能力、保护自己；另一方的回应，又让前者进一步防范。人类与 AGI 之间也就出现新的博弈递归。当更强的力量进入其中，人类未必还能在后果发生后把局面带回来。

两条链不必同时开始，也不需要同一个触发点。第一条在人类沿用旧逻辑使用 AGI 时会发生；第二条以 AGI 最终比人类强大为条件。它们共同逼近的问题是：某一轮博弈是否会在纠错到来以前，先破坏继续存在和重新选择的条件。这就是这里所说的终局。

当“还有下一次”不再能够被默认，文明就不能只依靠事后学习。

规则、协议、威慑与控制能够事前防范，也能够事后补救，仍然需要不断完善。问题是，只要旧逻辑继续把新的利益和安全焦虑变成新的博弈，我们就仍要一轮轮控制后果，却不能保证每次都还有下一次。

由此，文明跃迁作出一个根判断：AGI 增强人类，也可能成为比人类更强的主体；旧逻辑继续支配这两种关系，博弈就会递归并趋向终局。文明不能跃迁，就趋向湮灭。

所以，文明跃迁把改变的位置向前推进。不是只在结果出现以后补救，也不只是等博弈发生以后寻找更好的合作，而是进入目标、目的生成的起点。

三、文明跃迁改变的，是目标从哪里开始

文明跃迁要推动的根本转向，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”。

从“利己为先”，转向“利他为先的利己”。

“利他为先”最容易被误解成要求自己退后、让别人优先，或者要求所有人变得无私。文明跃迁不是这个意思。

人仍然要生存，仍然希望幸福，仍然有自己的边界、利益、选择和意义。变化发生在一个更早的位置：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其他者，以及能够看见的重要结果和后果，已经进入目的生成。

过去可能是：我先形成目的，再问它会影响谁。

文明跃迁进一步要求：在目的形成时，就开始看见谁会因此改变处境，这些改变会怎样继续传递，哪些后果仍然可以恢复，哪些后果一旦发生就很难回来。

他者不再等到我的目的形成以后才出现。

这并没有取消“我”。它只是让“我”不再假设自己生活在一个与他者割裂的世界里。

四、他者不是“所有别人”，而是目标、目的的利害相关方

如果“利他为先”被理解为每一次选择都要对所有人承担同样责任，它既无法运行，也会变成新的道德压力。

文明跃迁中的他者，是与具体目标、目的存在真实利害关系的主体。

一个目标会实质改变谁的生存、利益、选择、关系、幸福、意义或未来条件，谁就与这个目标形成利害关系。不同主体受到的影响不同，因此以不同权重进入；利害关系发生变化，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。

这里的权重不是对人的价值排序。它不是说谁更“值钱”，而是在说明这个目标与不同主体之间真实利害关系的相对位置。

更重要的是，这也绝不是把文明跃迁简化成一个“加权公式”。现实中的利害会随着时间、关系、信息、权力和后果变化。权重只是目标生成中的一个关系层。文明跃迁改变的是整条因果链，而不是一个参数。

五、当合作进入目的生成，社会的底色开始改变

如果一个主体在形成目标时，真实利害相关方已经在场，那么它寻找利益实现方式时，问题就会发生变化。

它仍然会问：我怎样过得更好？

但它同时会问：我的获得与谁相关？有没有一种路径，使我的利益能够实现，也让相关主体继续存在、继续获得、继续合作？

利他为先的利己，内生合作。

这里的“内生”不是说从此没有冲突，也不是说所有利益都能兼容。它表示合作不再主要等到博弈发生以后才出现，而开始进入利益和目的形成本身。

新的需要出现，可以首先寻找新的共同实现方式；新的合作产生新的利益；新的利益又继续生成新的合作。

合作 → 新需求 → 新合作 → 新利益 → 再合作。

博弈不会消失。恶意、稀缺和无法兼容的利益也不会消失。改变的是默认起点。

旧逻辑中，博弈默认生成，合作主要负责控制博弈。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、改变的不只是一念，而是社会不断生成自身的方式

一个人的目标从来不是孤立生成的。家庭、关系、教育、文化、市场、组织、资源、安全感、社会评价和过去的生活经验，都在告诉我们什么值得追求、怎样才算成功。

目标形成以后，又通过行为进入人与人的关系。无数关系积累成合作与博弈，形成利益结构、制度、文化、权力和分配。

这些社会结构随后又返回每一个人的生活，继续影响下一轮目标生成。

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。

所以，文明跃迁不是一种新制度，不是一种新技术，也不是给既有社会增加一个道德口号。

如果目标生成的默认逻辑真正改变，它会沿着整个人类社会的因果链继续展开：利益怎样形成，合作怎样出现，什么行为获得现实回报，人怎样得到认可，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。

文明跃迁，是文明不断生成自身的方式发生改变。

七、从人心、人性、人情出发

文明跃迁不能要求人先变成圣人。

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。真正可持续的文明转向，只能从这些真实的人心、人性、人情出发。

因此，文明跃迁关注的不只是“人应该怎样做”，还包括“人通过什么方式才能得到自己真正想要的东西”。

如果一个社会让人的获得感主要来自赢过别人、占有更多、控制更多、处在更高位置，那么即使每个人只是在正常地追求自己的生活，整个系统也会不断重新生成竞争、防范和博弈。

文明跃迁要改变这种现实路径。让一个人仍然可以为了自己创造、工作、选择和生活，却越来越不需要以另一个人的失去作为自己得到的前提。

他的价值不必只能通过超过别人证明。他的幸福不需要按照统一模板实现。他的意义也不由任何系统授予。

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。

八、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，一切无从谈起。但仅仅继续存在，也不是文明的终点；它更不等于每个人已经摆脱基本生存的恐惧。

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，人类整体不应继续让大量生命被基本生存焦虑长期束缚。生存无忧，是更高文明能力首先应当转化出的共同基础。它关心的是基本生存不再成为普遍的恐惧与束缚，而不只是文明还会不会继续存在。

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。有人珍惜家庭，有人热爱创造，有人需要自由，有人喜欢安静，有人愿意探索。文明跃迁不替任何人决定幸福，而是扩大每个人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、贡献和存在本身，都可能产生意义。意义为根，意义来源可以不同。文明跃迁要让生命拥有更多生成、发现和实现自身意义的空间。

共同存续，让我们还有未来；生存无忧，让这个未来有更可靠的生活基础；幸福与意义，让未来真正抵达每一个人。这些可以相互促进，不需要等到一切风险消失以后，才开始追求属于自己的幸福与意义。

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。

九、文明跃迁必须守住的几条原则

方向统一，实现可以多样

文明跃迁统一的是根本方向、存续底线和三目标，不要求所有国家、地区、组织和个人采用同一种生活方式。文明越复杂，路径越需要尊重差异。

不改变利己，改变利己的出发点

任何要求人普遍放弃自身利益的路径，都偏离文明跃迁。新的逻辑之所以能够持续，正因为它仍然让每个人追求自己的生存、幸福与意义，只是让真实利害关系更早进入目标形成。

真实利害相关方不能缺席

谁真正承担一个目标的后果，谁就不能只在后果发生以后才被看见。不同利害相关方以不同权重进入，权重随利害变化调整。

个体不是文明的材料

共同方向不能变成统一人生。不能以整体、未来或共同利益之名取消个体主体性。幸福是个体化的，意义来源可以不同。

合作不能成为新的排他博弈

几个主体彼此合作，也可能把代价转移给合作之外的人。只要真实利害相关方承担了重要后果，合作本身也需要继续向外看。

文明必须保有纠错能力

任何以正确方向之名关闭质疑、退出、修正和重新判断空间的结构，都可能重新制造不可逆。文明向前，不等于把任何主体固定成永远正确。

十、方向、方法与结果不能混为一谈

文明跃迁确定方向，但方向不会自动替我们解决现实中的所有方法问题。

方向回答往哪里走。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。

不同阶段可以有不同方法。同一种方法在不同社会条件下，也可能产生完全不同的结果。方法可能失败，也可能产生意外后果，甚至把原本正确的方向带向相反结果。

所以，结果必须返回方法。方法必须不断调整。

但一个局部方法失败，首先说明这个方法、它依赖的条件，或者我们对现实因果关系的理解需要修正。它不能自动推出文明应该重新选择一个以博弈为默认起点的方向。

同样，一个局部方法成功，也不能代表文明跃迁已经完成。文明跃迁涉及的是整个人类社会构成之间长期、相互作用的因果变化。

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。

十一、我们现在应该朝哪里走

文明跃迁的行动首先是方向性的。它不要求今天就给出整个未来社会的答案，却要求我们不再把越来越强的能力继续无条件送入旧的目标生成逻辑。

把“目标为什么这样生成”放到更前面

我们已经非常擅长讨论怎样更快、更有效地实现一个目标。AGI 时代还必须继续向前追问：这个目标为什么应该存在，它从哪里开始，它把谁放在了外面。

让能够看见的后果更早进入

能够在行动以前看见的重要结果，不应总是等到伤害发生以后才成为教训。文明越有能力，越应该把学习向前移动。

让合作更早进入新的利益生成

合作不必永远等到双方已经进入对立以后才出现。新的需要出现时，就开始寻找共同成立的路径，新的利益才可能不断派生新的合作。

让人的“得到”拥有更多非对立的来源

一个文明如果只能让人通过超过别人来获得安全、尊严和价值，旧逻辑就会不断复制。文明跃迁要让创造、连接、帮助、合作和共同改善，也能够成为真实而稳定的个人获得。

让人类与 AGI 朝同一个目标生成方向前进

只改变人类而让 AGI 继续放大旧目标，不够；只要求 AGI 改变而人类仍用旧逻辑驱动它，也不够。能力增长与目标生成必须朝同一方向重新连接。

让所有变化最终回到三目标

无论未来社会出现多少新技术、新组织和新生活方式，最终都要回到三个问题：人类整体是否更接近生存无忧？每一个人实现自己幸福的空间是否更大？生命生成意义的可能是否更多？

结语 | 文明每天都从一个目的开始

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。

我要怎样使用越来越强的能力。

如果目的仍然从一个他者缺席的世界开始，能力越强，旧逻辑就可能走得越远。

文明跃迁改变的，就是这个起点。

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更早进入，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。

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。

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，从一个目标开始生成的地方出发，一步一步成为文明不断生成未来的方式。

如果你愿意继续读

你不需要一次理解全部。可以从与你最有关的问题、生活或未来继续。

未来文明的一天

先不从宏大结构进入，看看文明跃迁如果真正进入现实，一个普通人的一天可能发生什么变化。

《活的文明·第二卷：未来文明的真实生活》

从出生、成长、工作、关系、幸福与意义继续理解文明跃迁怎样进入人的一生。

公众思想短篇

从一个个现实问题继续，不需要一次理解整个体系。

文明跃迁之路

如果希望进一步理解两种关系中的博弈为什么趋向终局，以及共同存续与三目标怎样相接，可以从这里继续。

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/>