

说服的本质

模型攻击模型



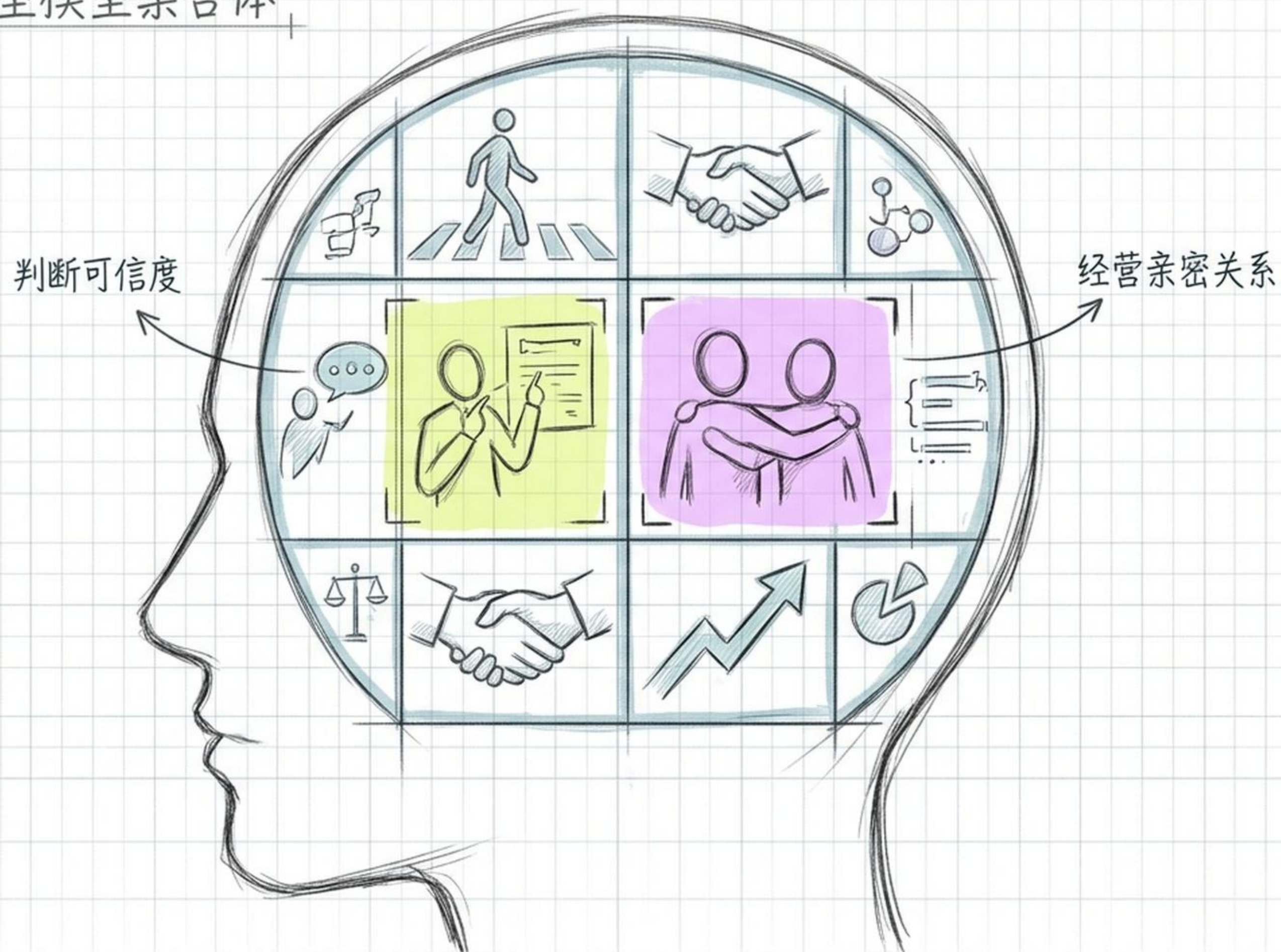
数型AI的技术
数据流

认知的图像进程

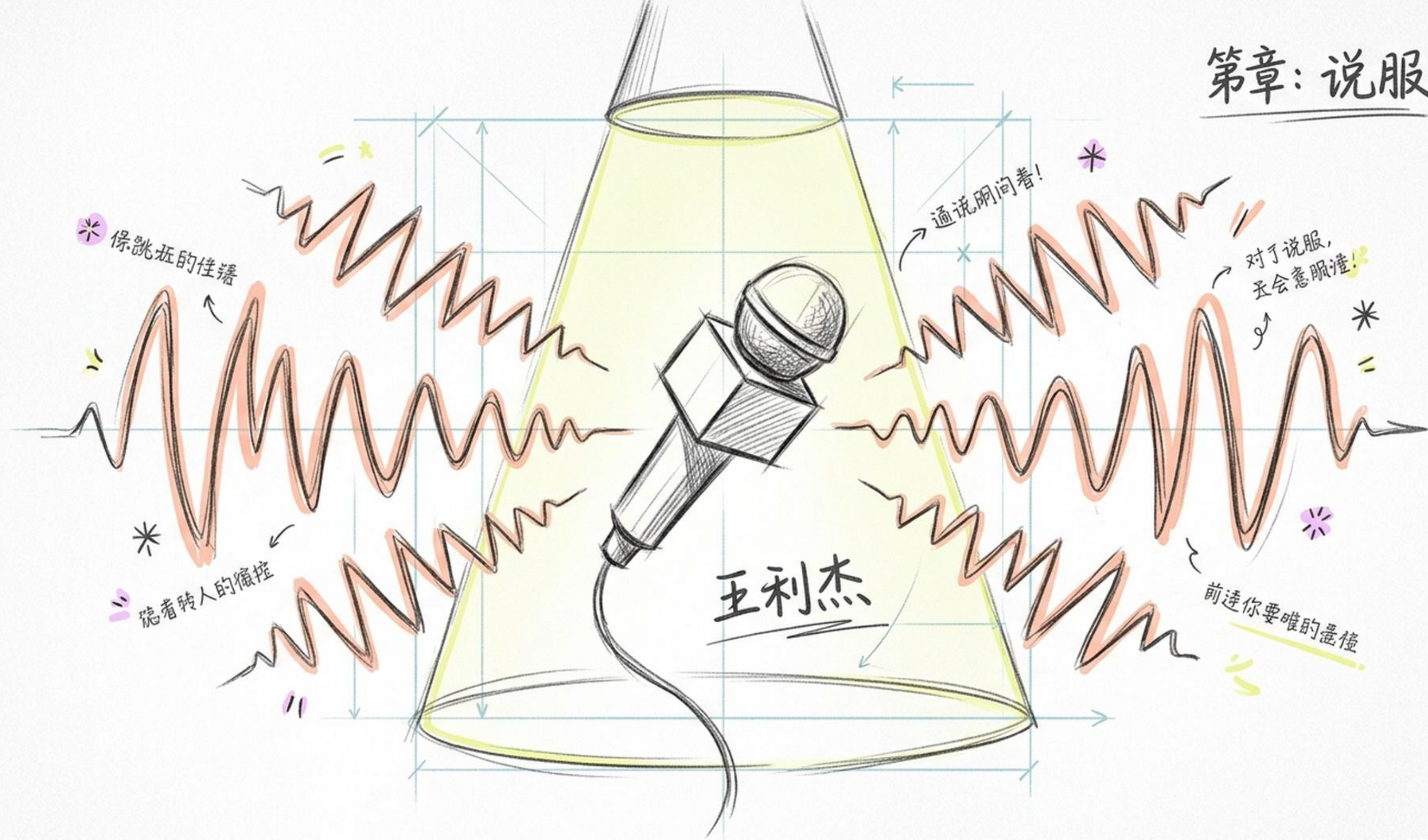
认知知识模态



微型模型集合体



第章：说服

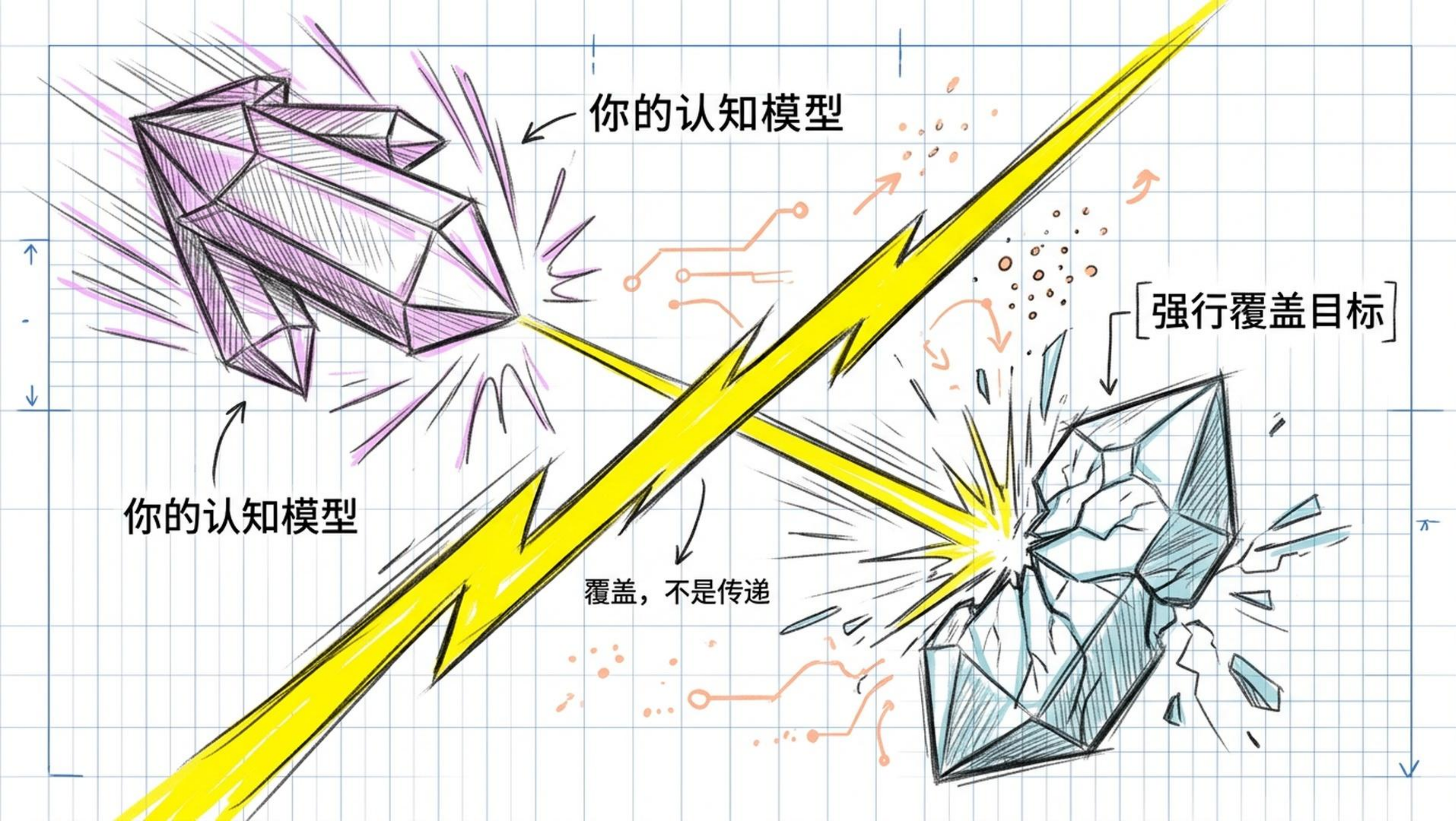


完美逻辑链条

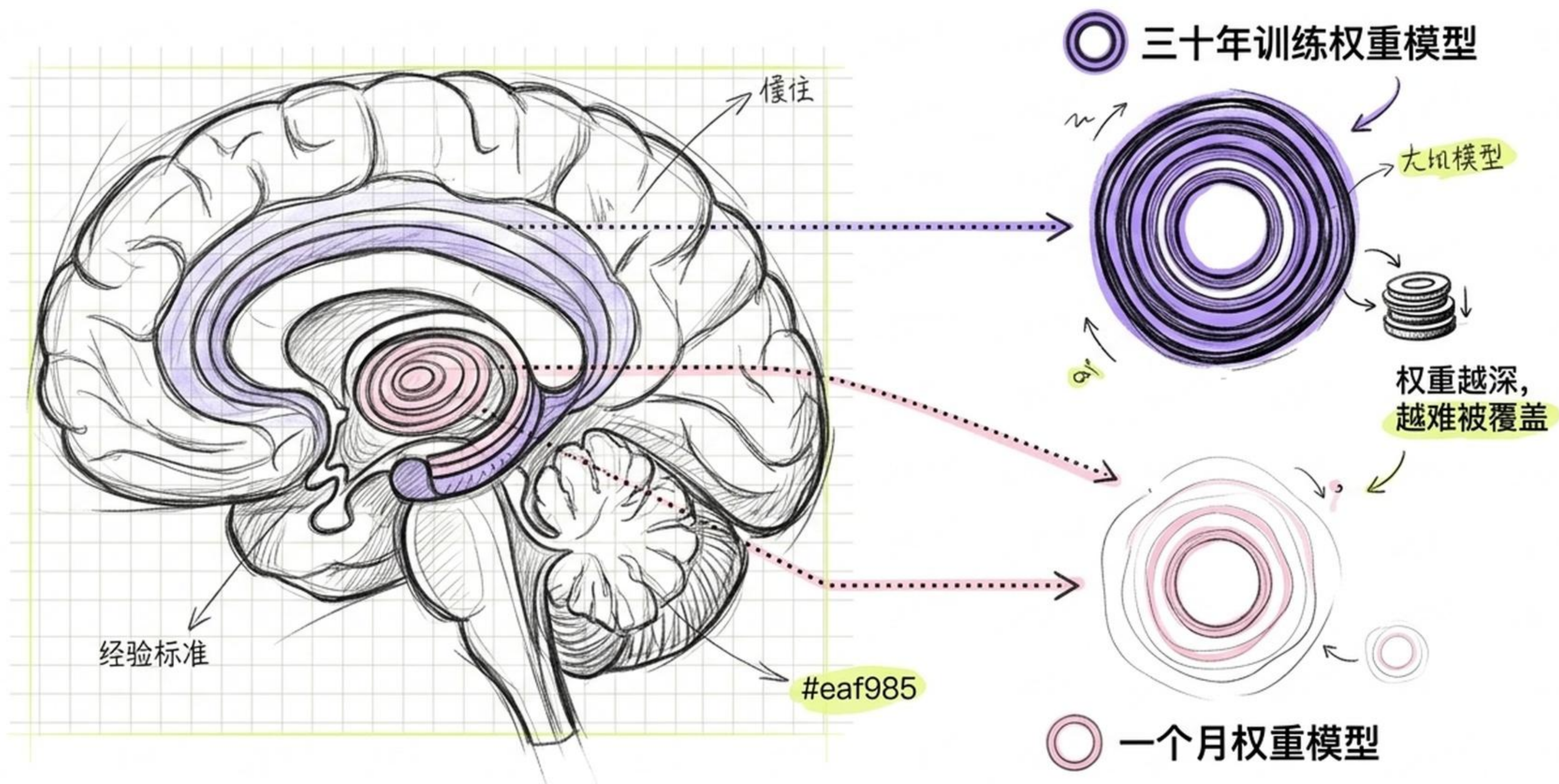
完美逻辑链
也能命的量
传成链条

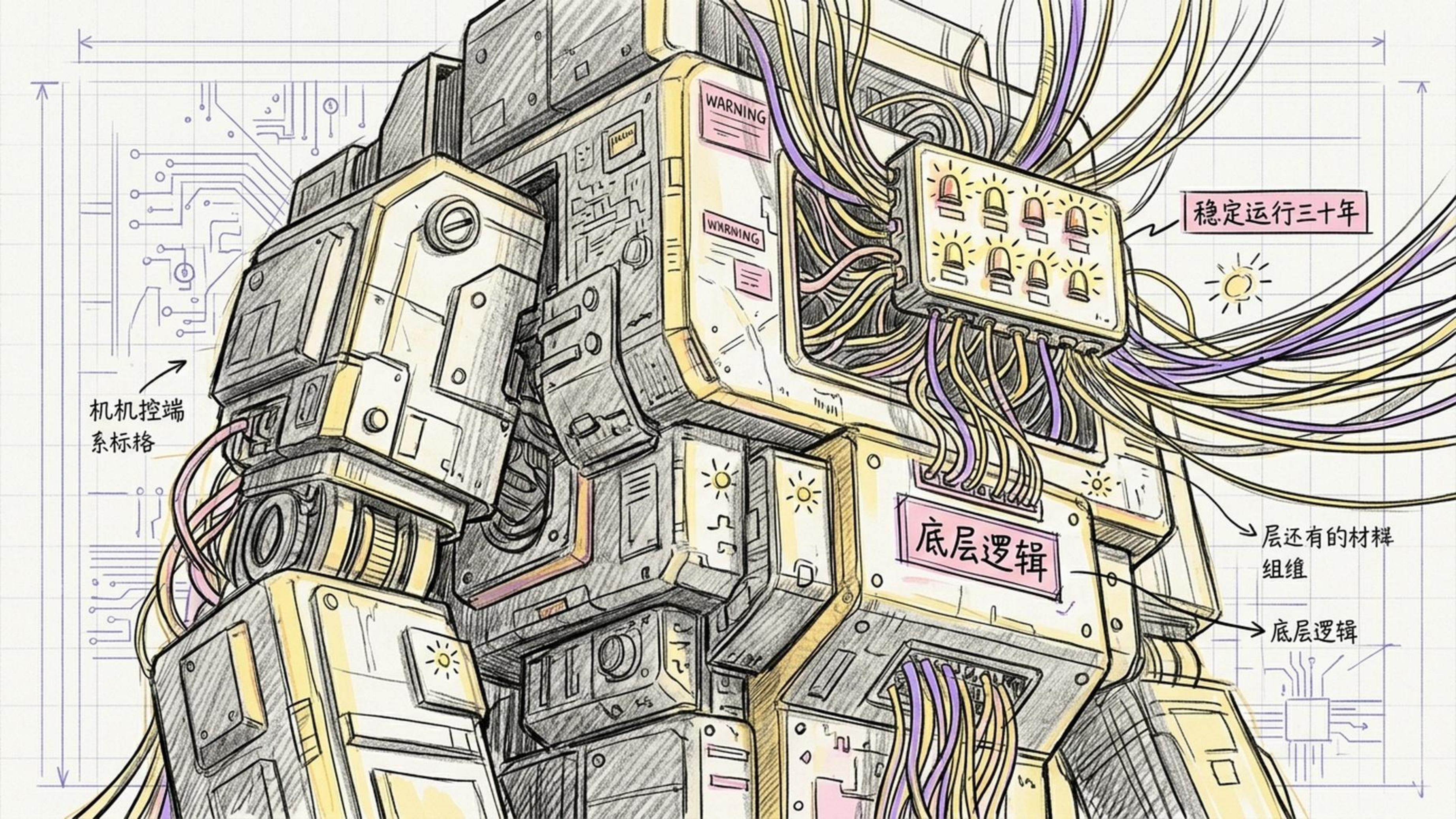
完美是青柠，杨彪
插出的完美法，完
美逻辑链条





“大脑里的小模型们”





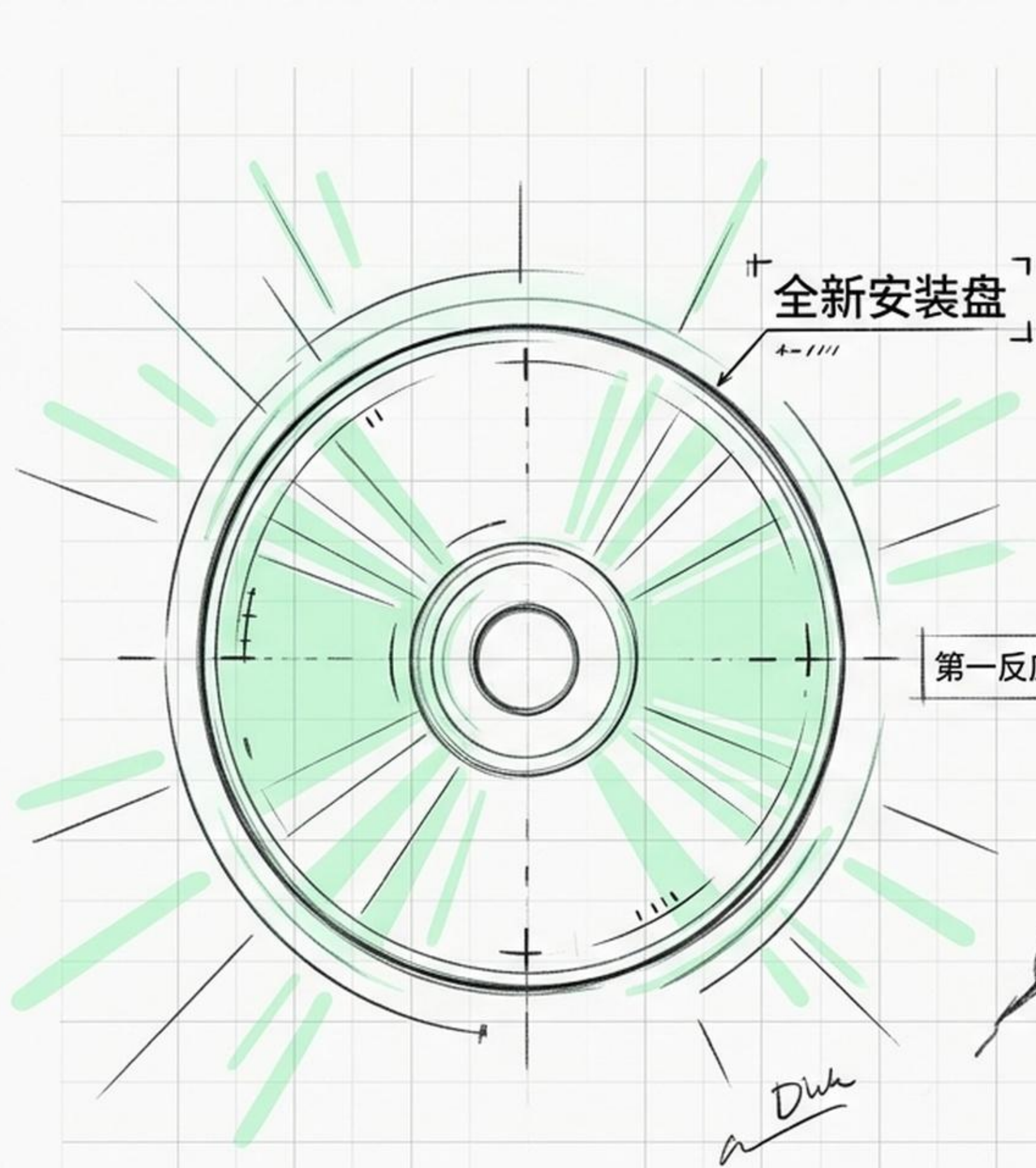
机机控端
系标格

稳定运行三十年

底层逻辑

层还有的材料
组维

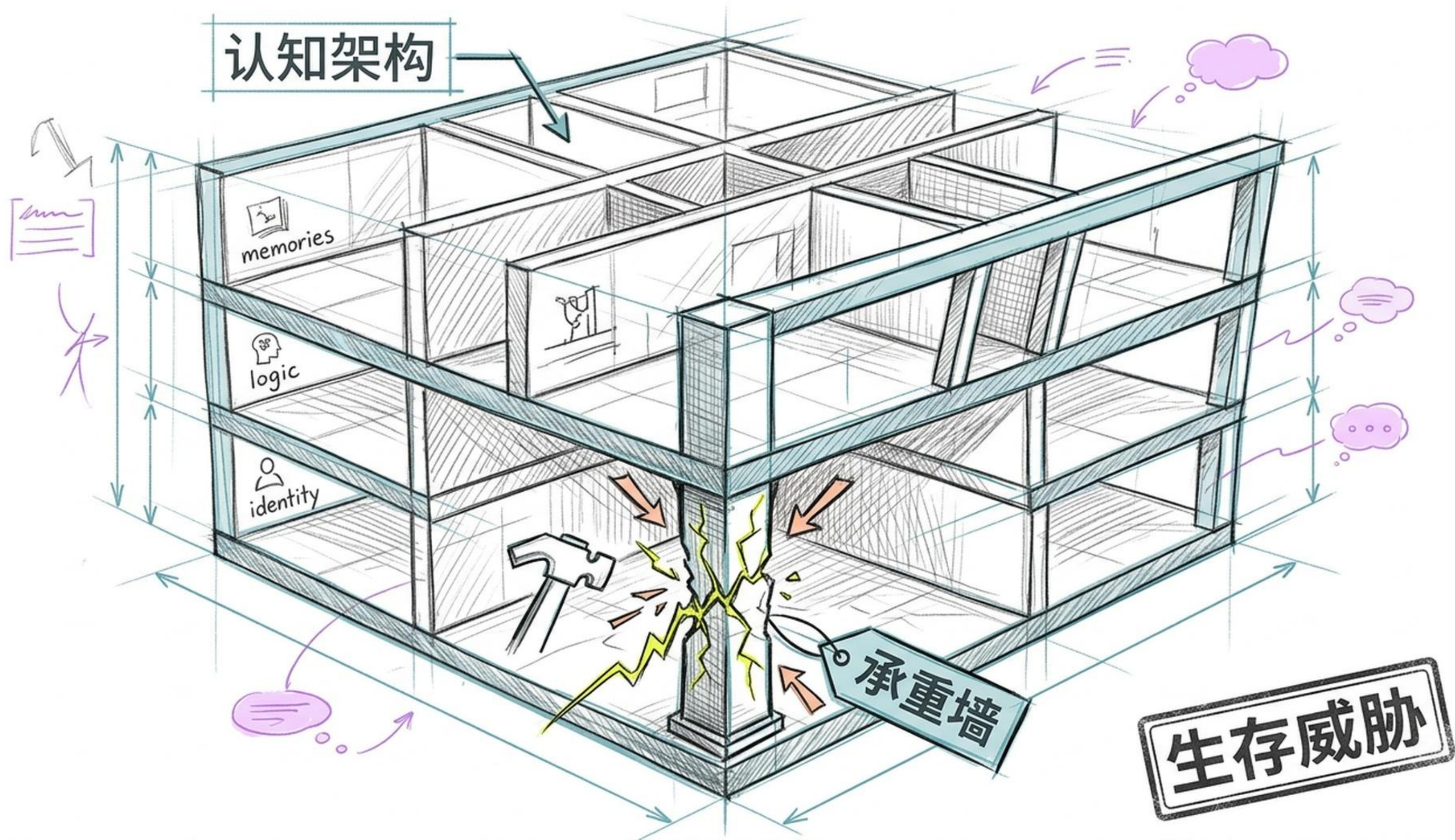
底层逻辑

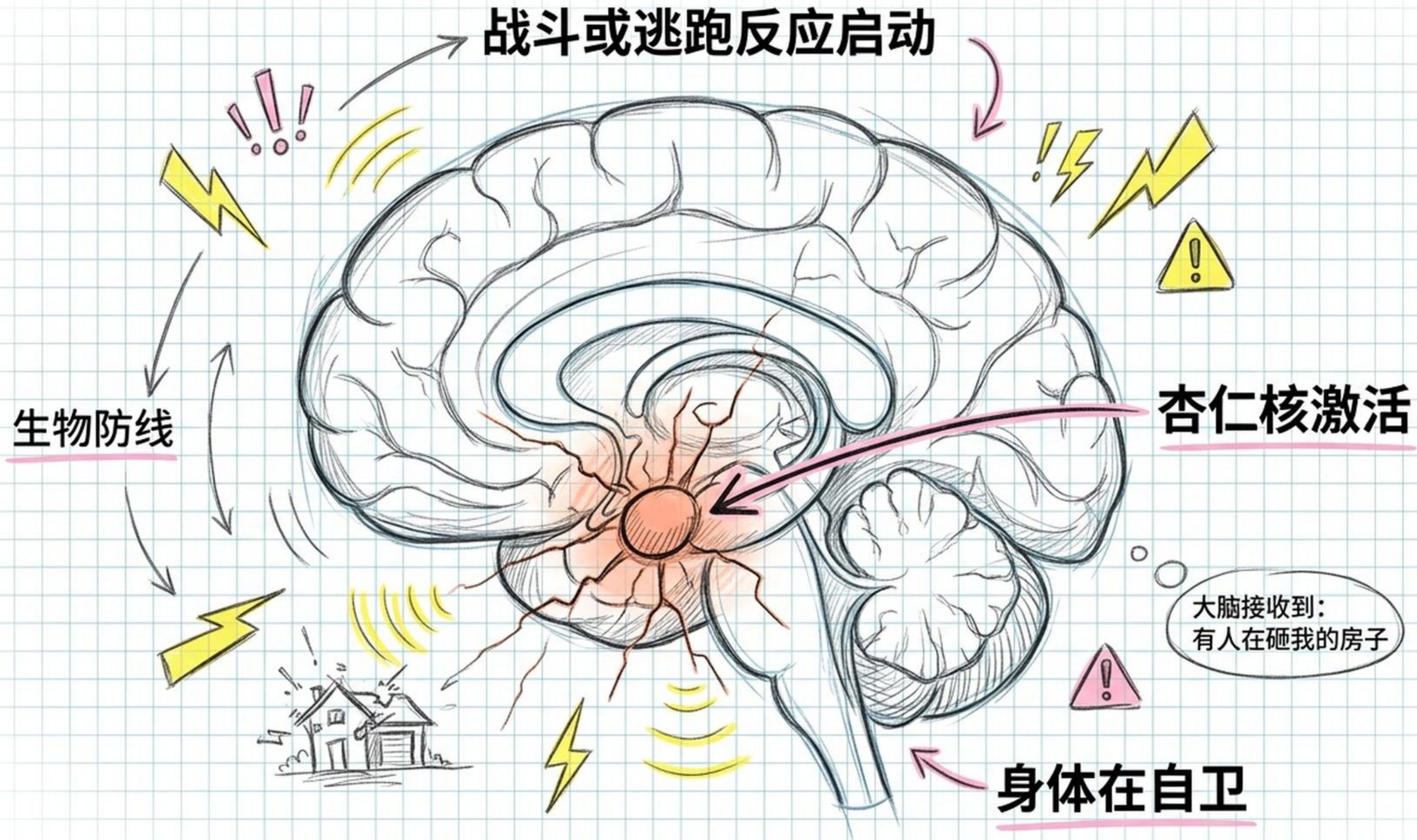


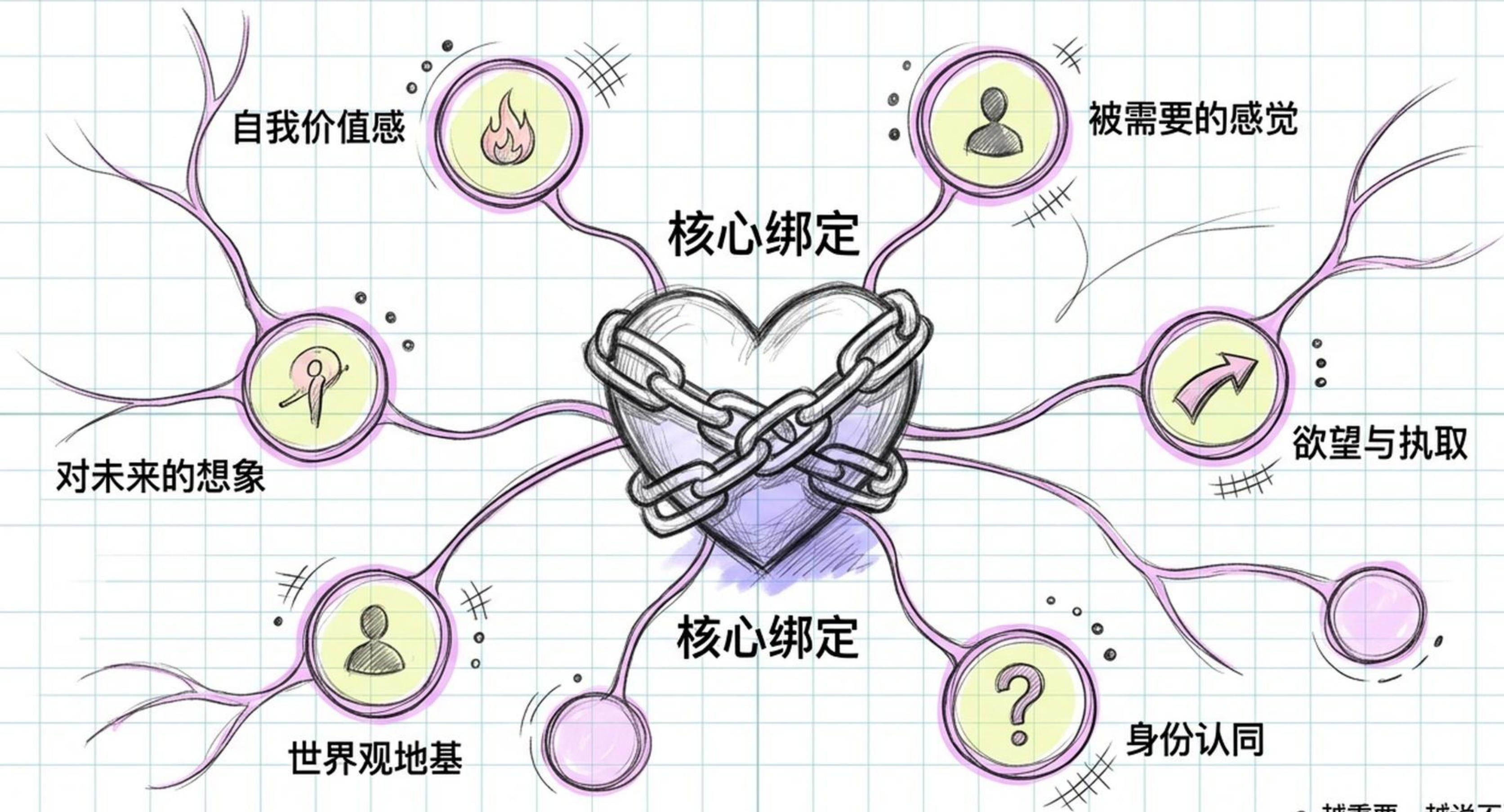
第一反应：不是感激，是恐惧



认知架构







● 越重要，越说不通

#efba95

#eaf985

#9f75ff

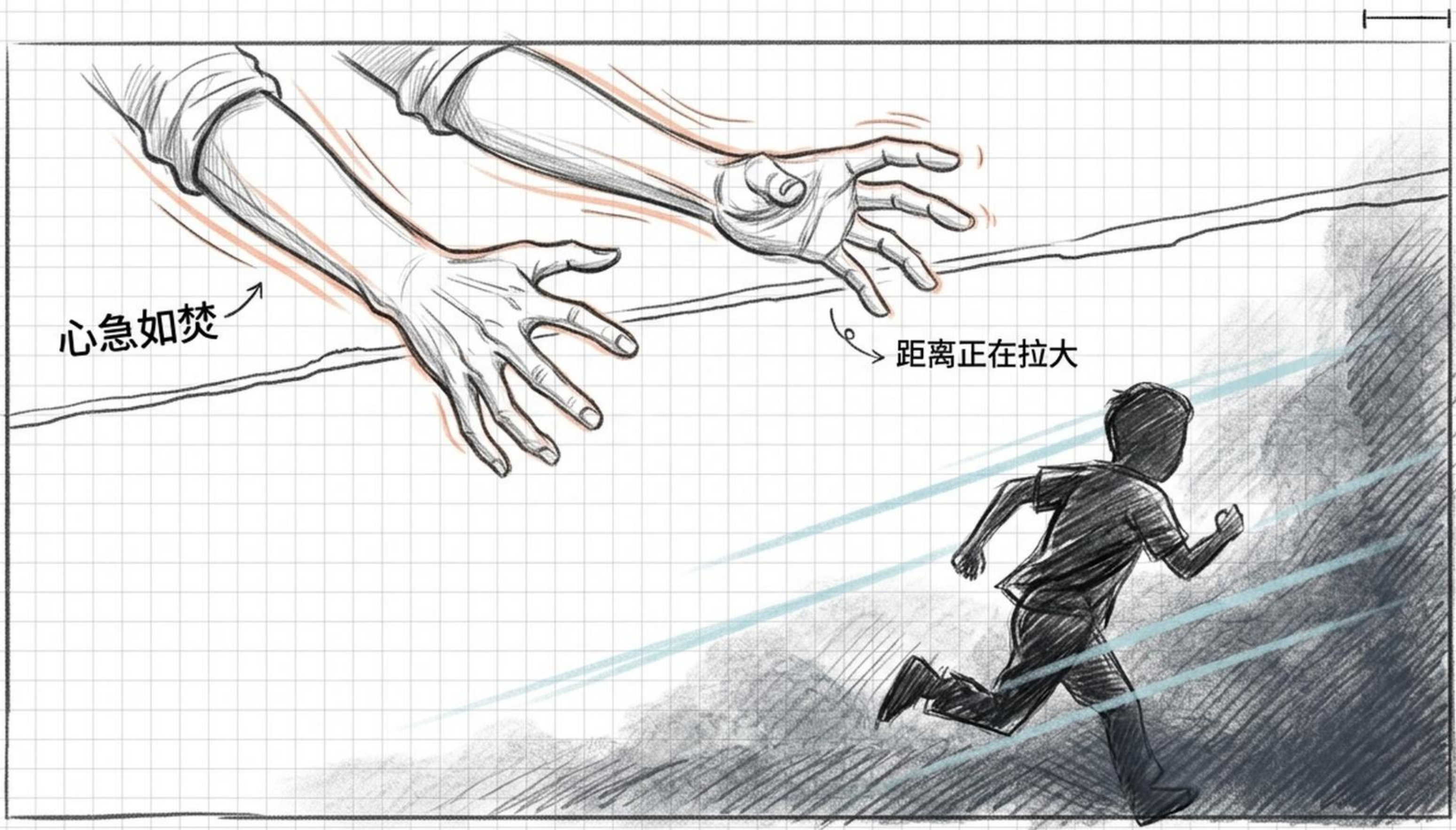
心急如焚

距离正在拉大

父母 vs
子女

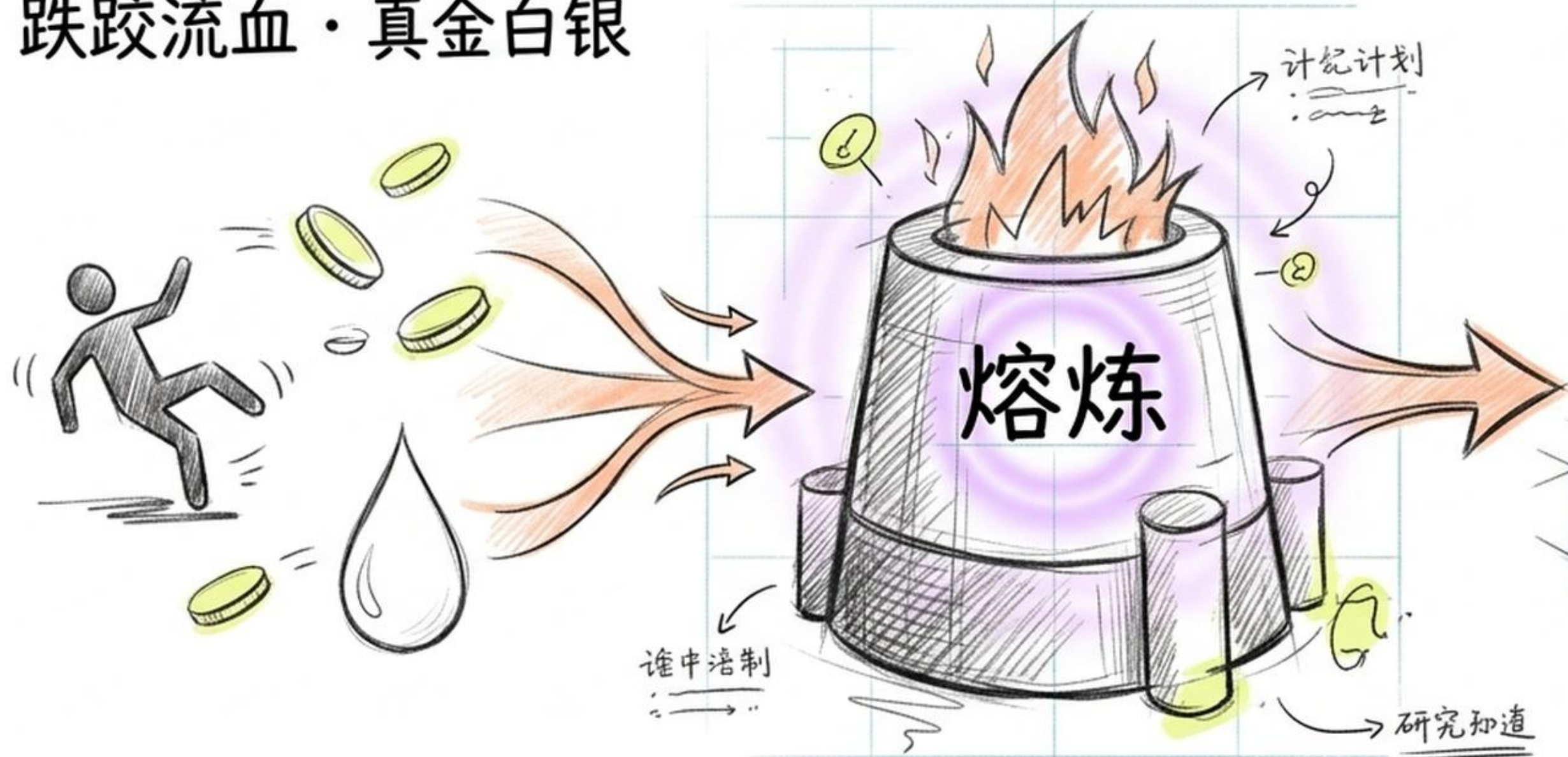
可可以:
父母
vs 子女

越急，他跑得越远

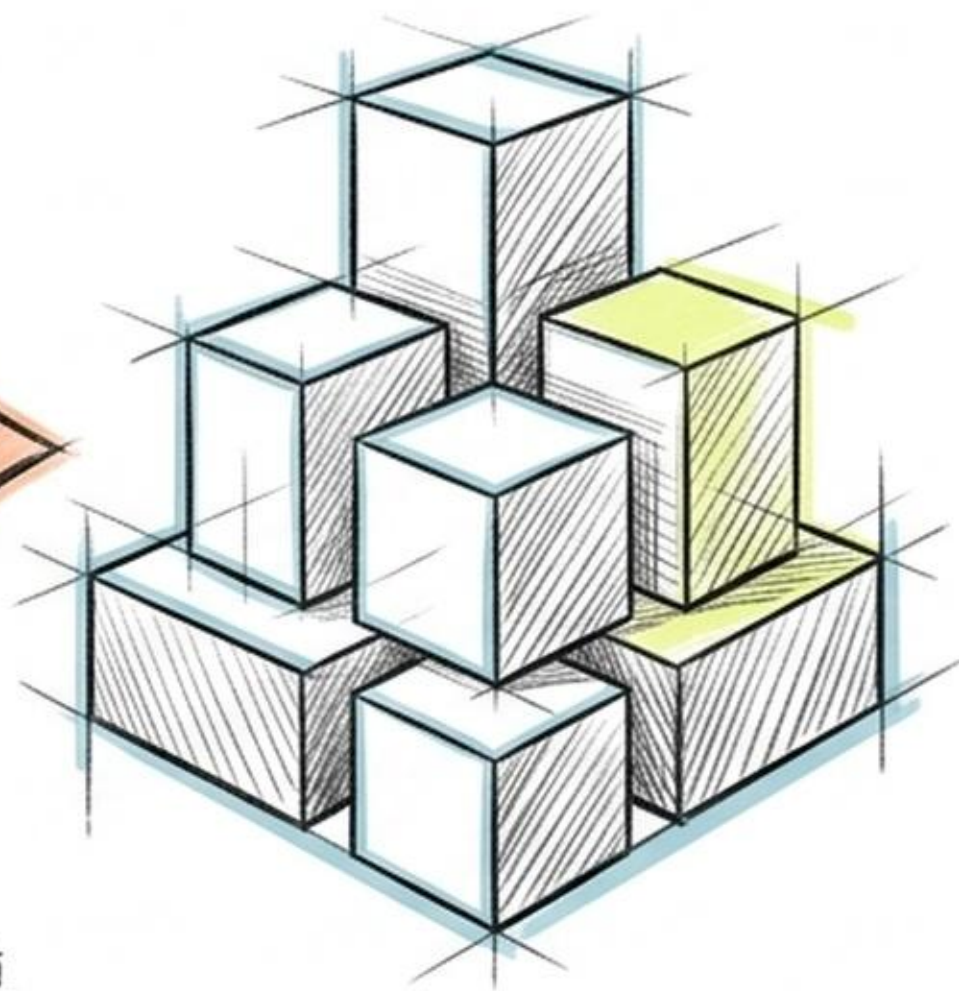


生命经验锻铸造成智慧

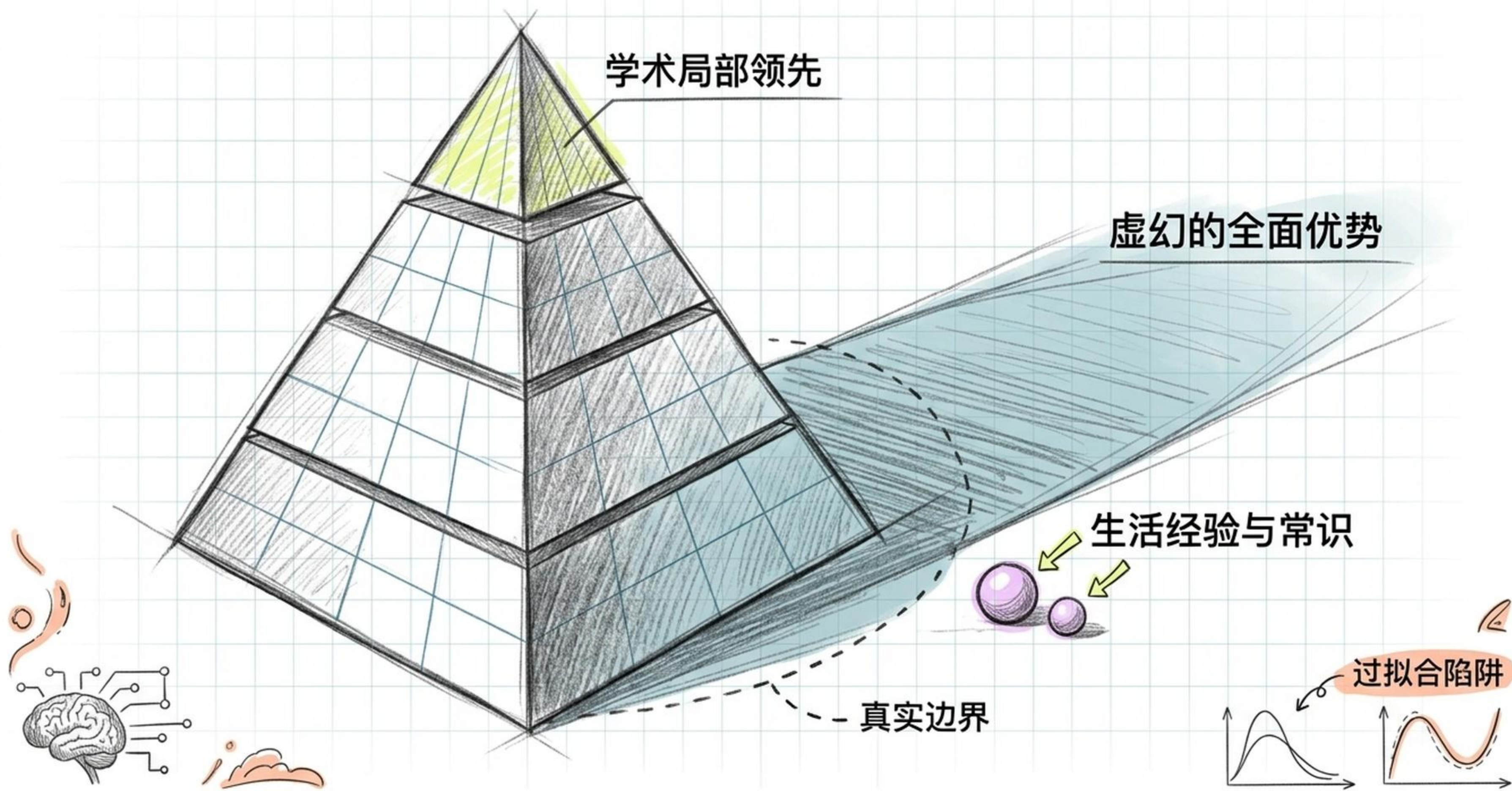
跌跤流血·真金白银

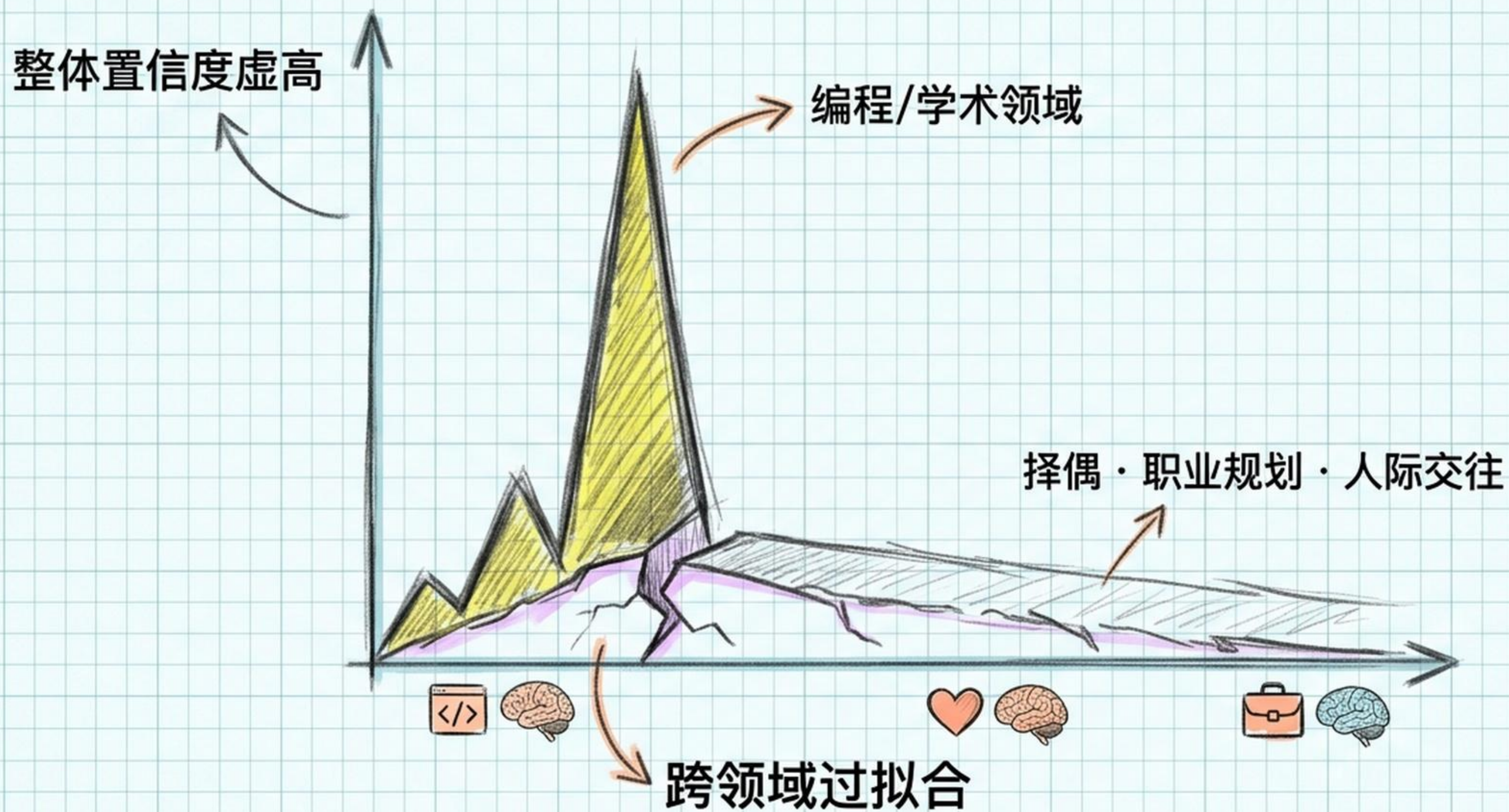


人生模型参数



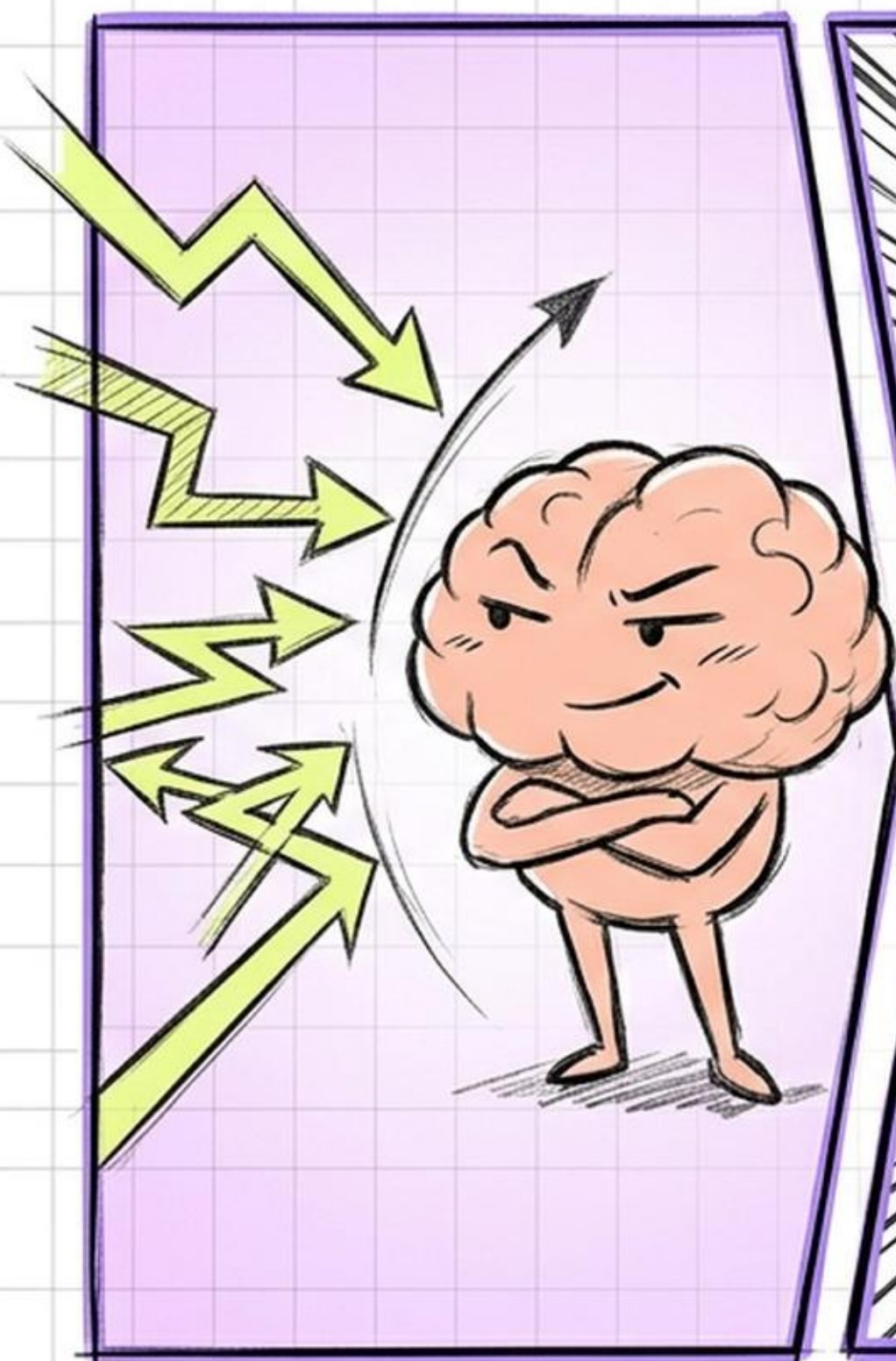
用血与汗铸成的底层认知



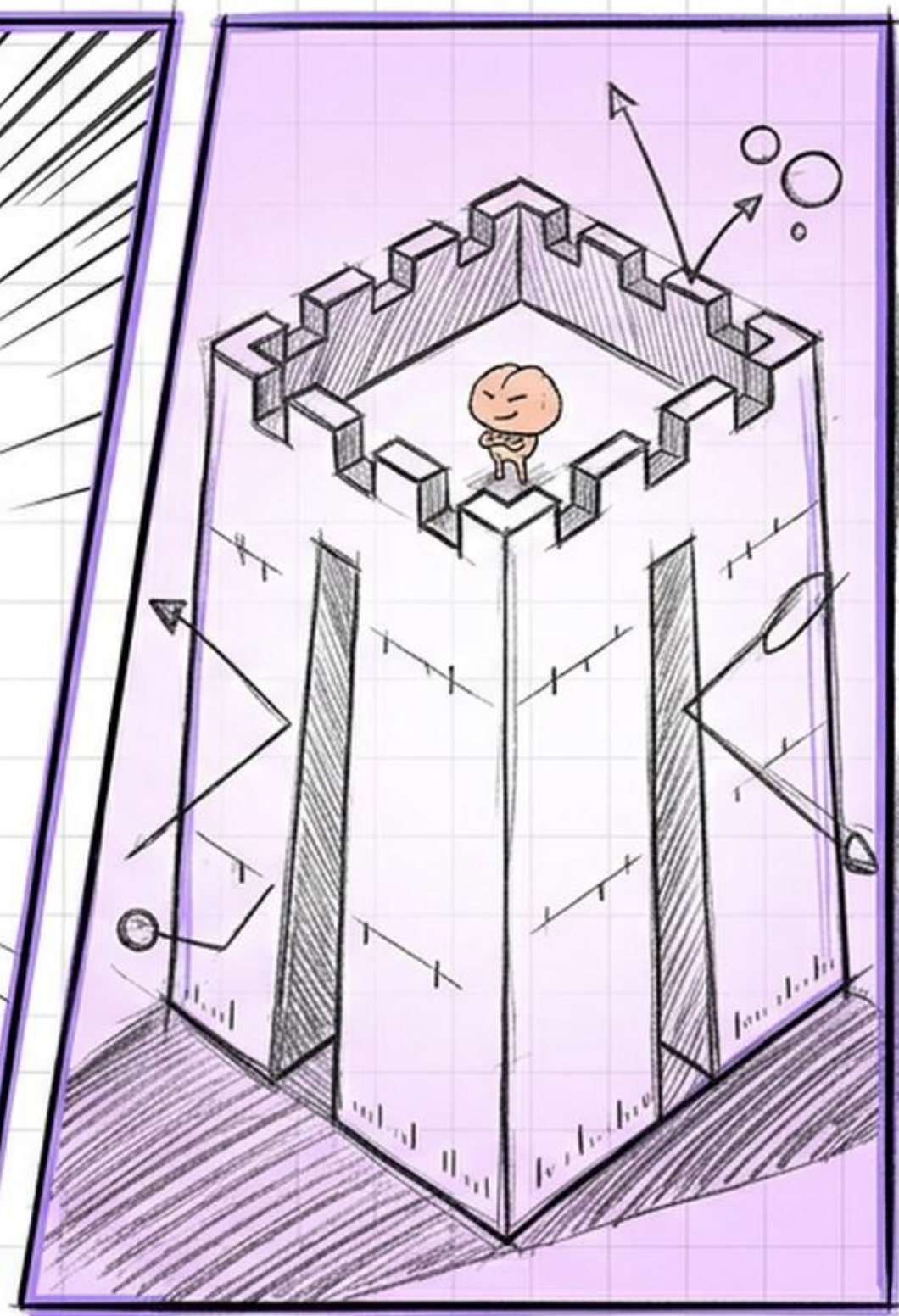
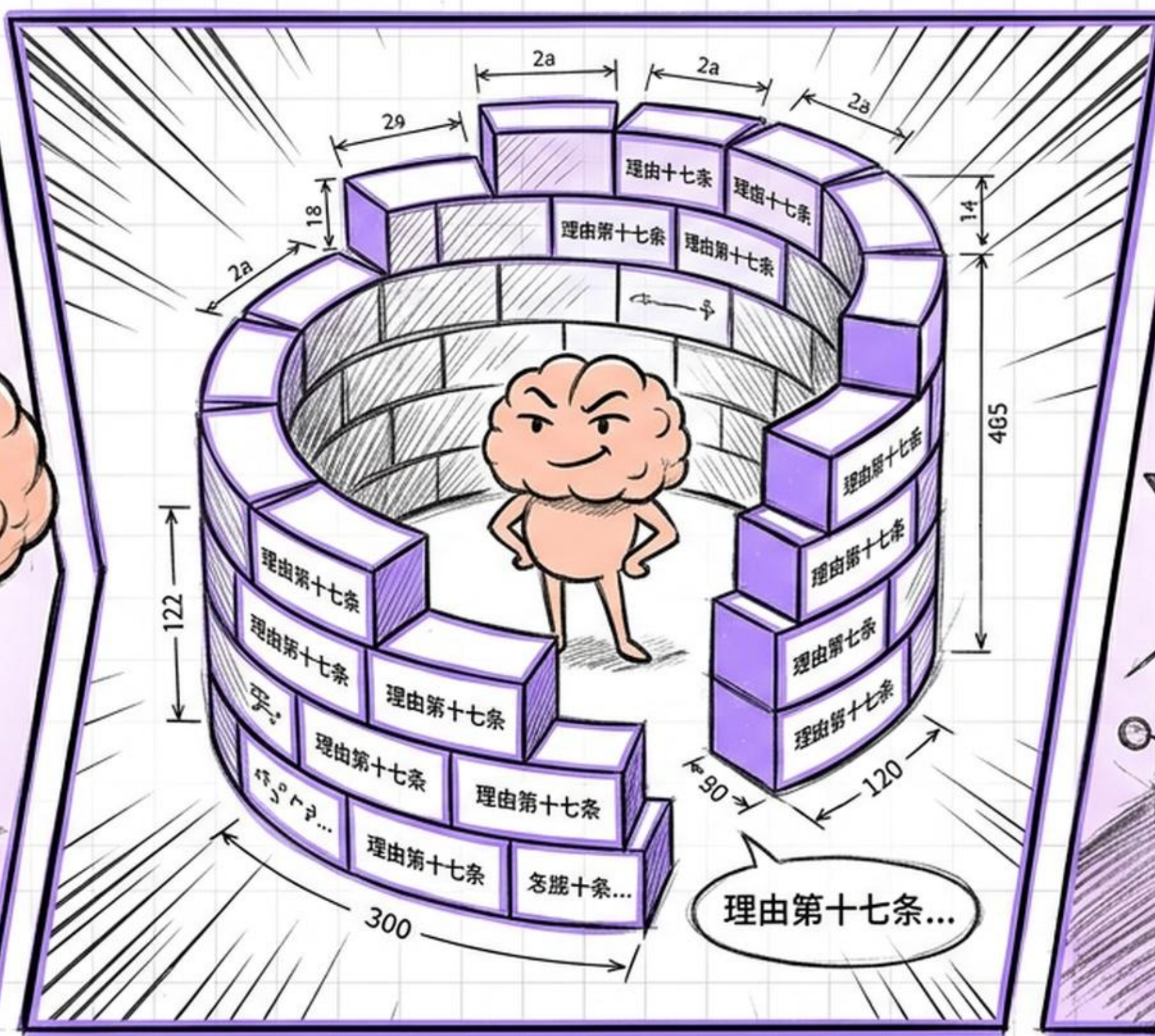


这不叫自信，这叫跨领域的过拟合自信

聪明人的防御工事



分镜1：拒绝新数据输入



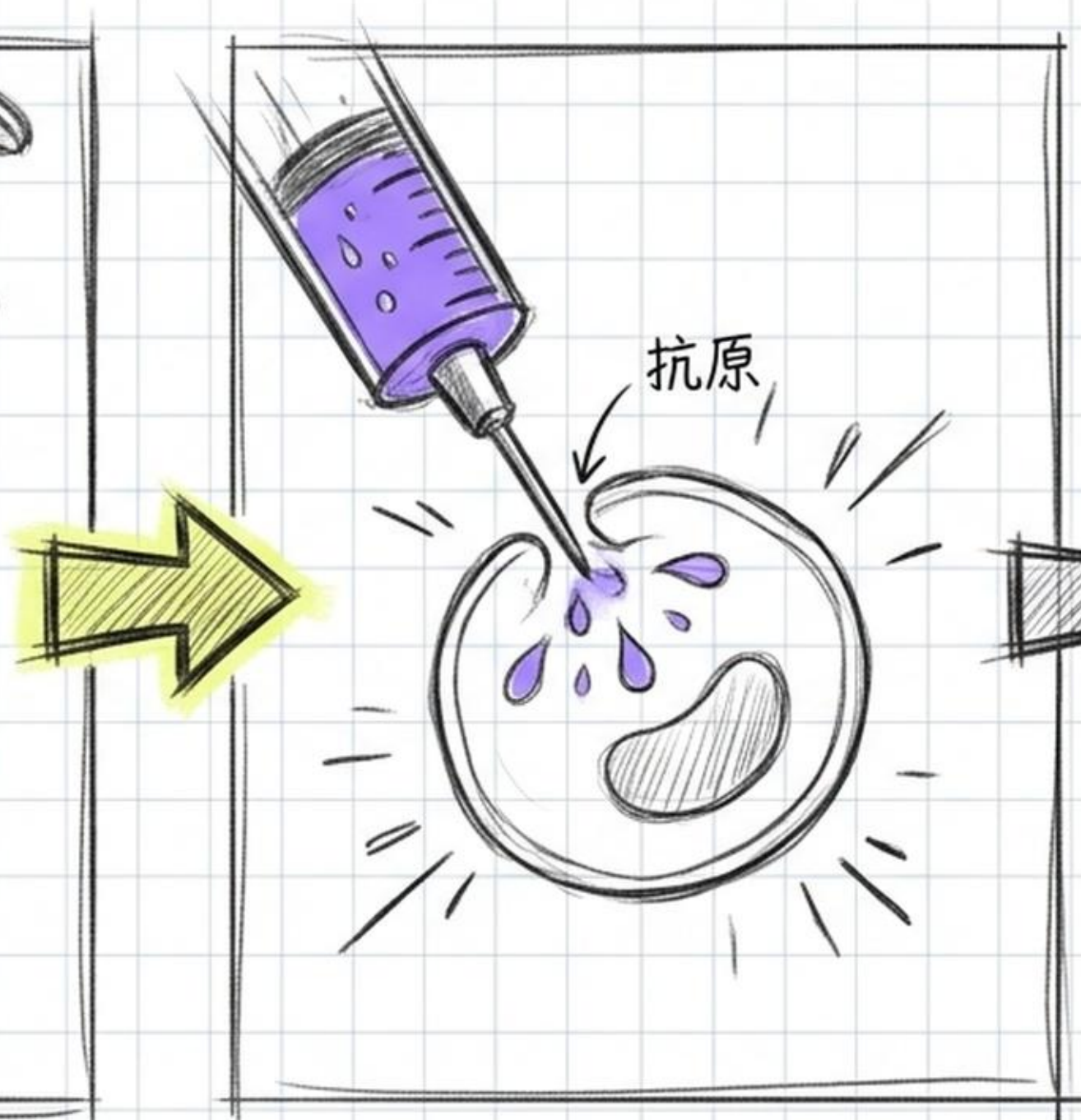
分镜3：坚不可摧的堡垒

最差的策略

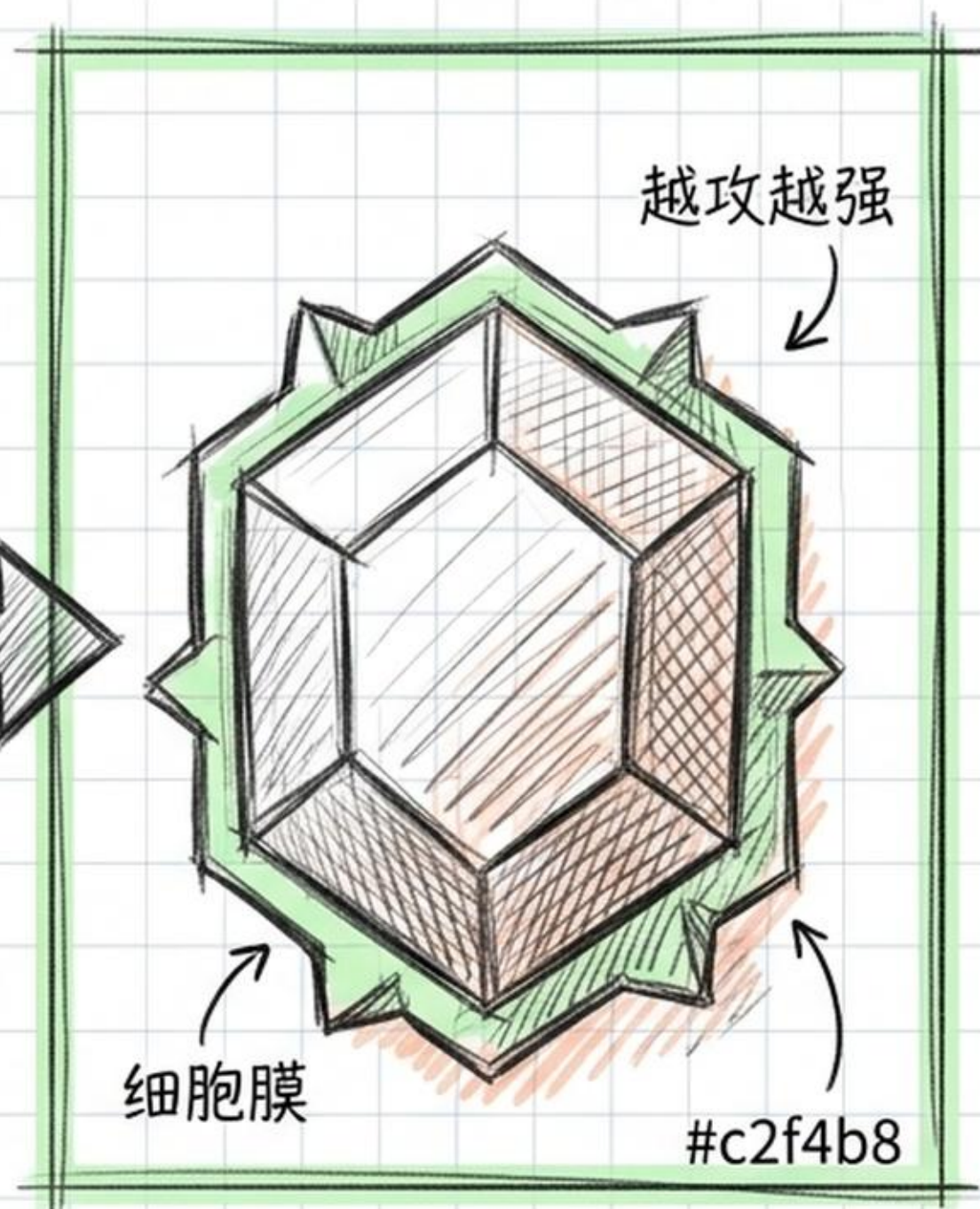
第一步：正面攻击



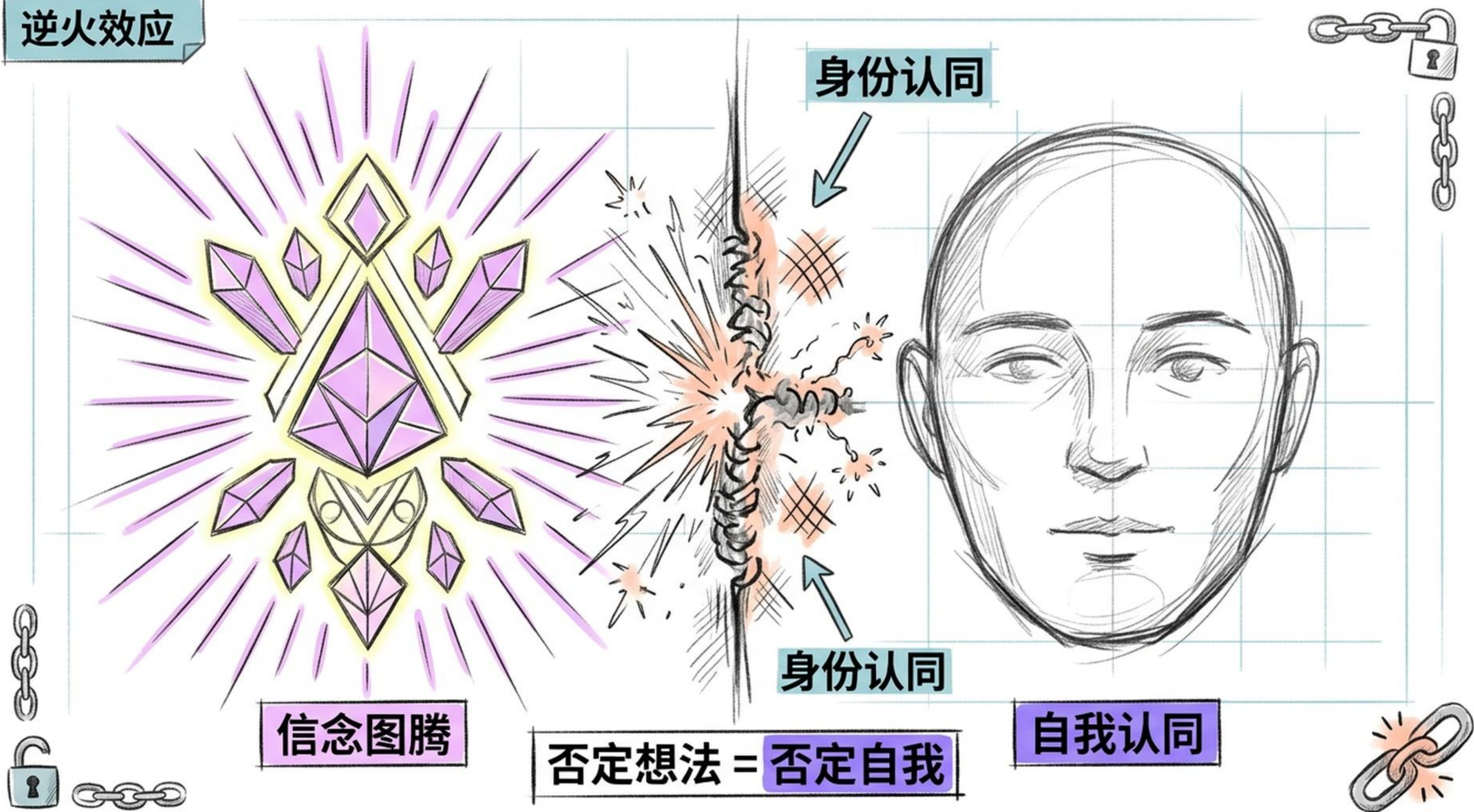
第二步：疫苗效应



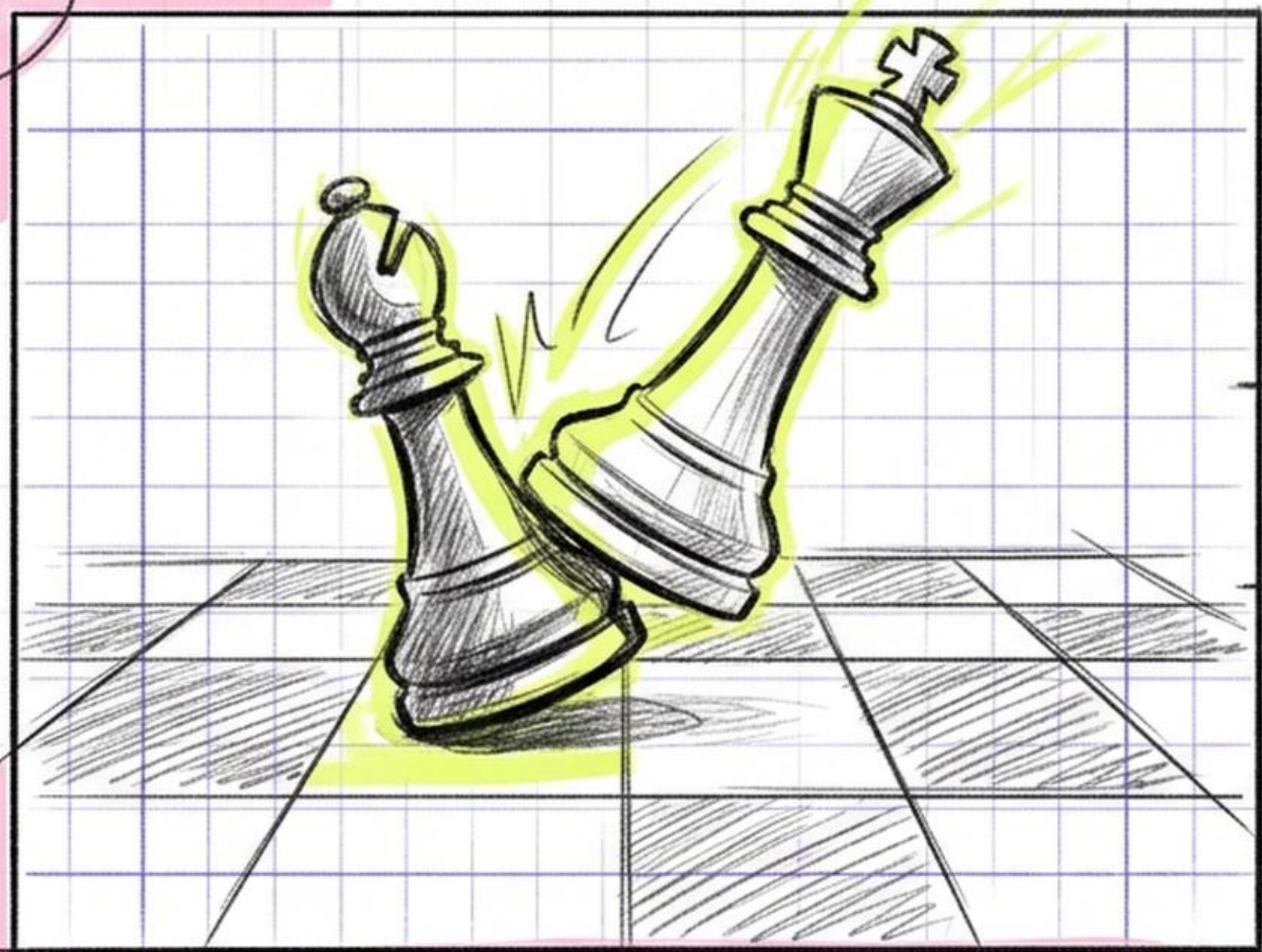
第三步：逆反加固屏障



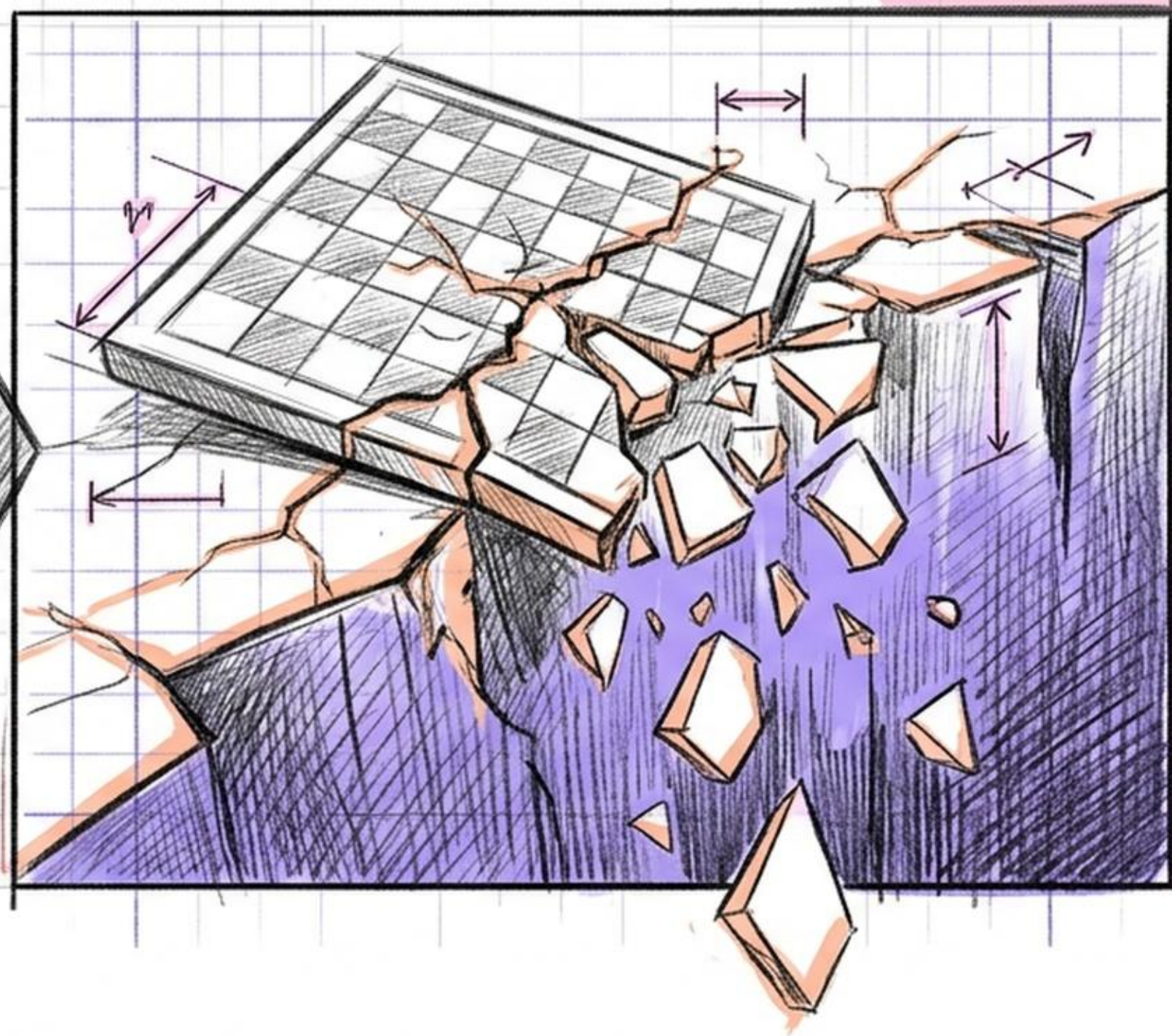
逆火效应



战术级胜利 ✓



战略性惨败 ↓



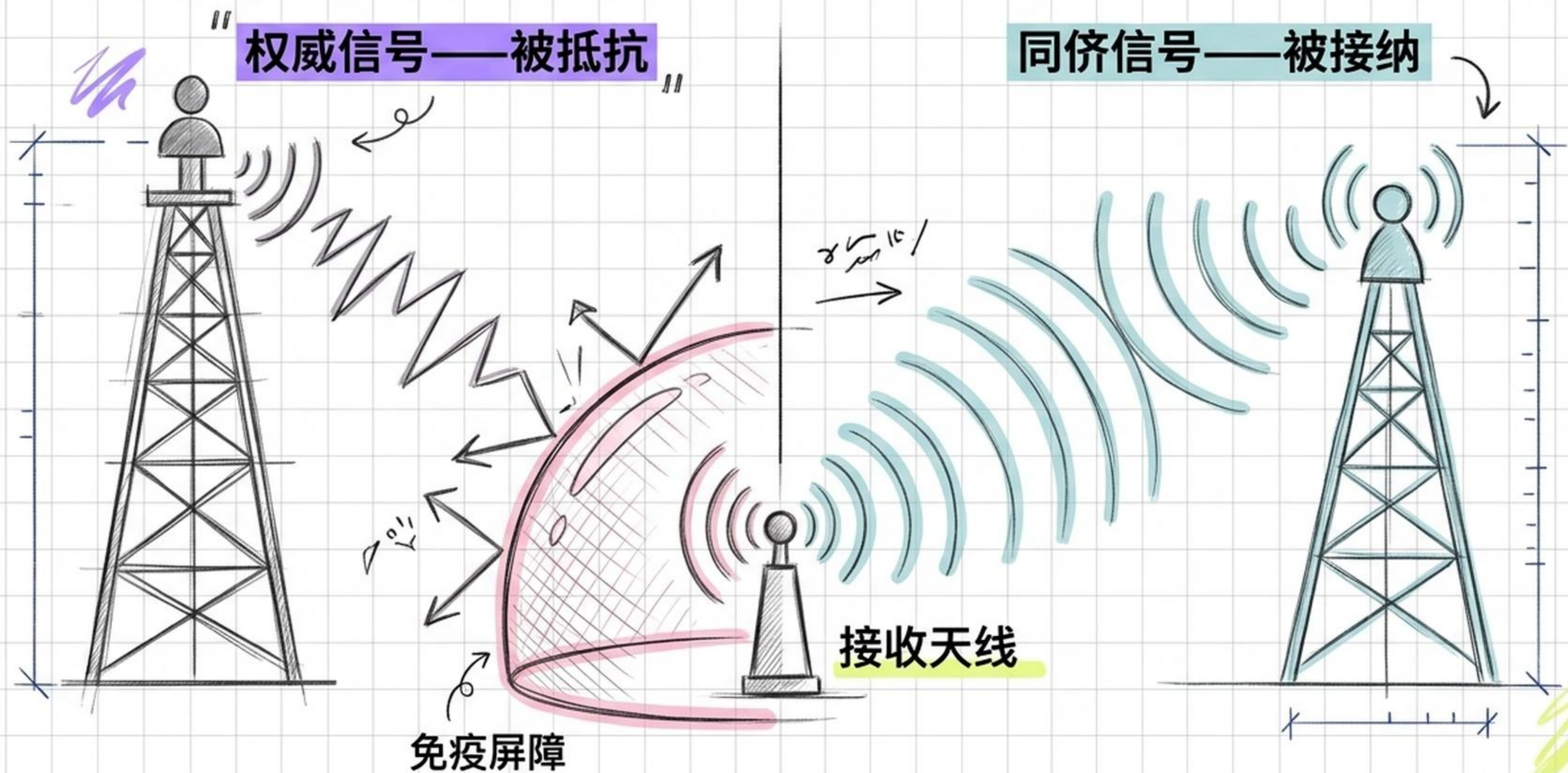
赢了辩论，输了战争

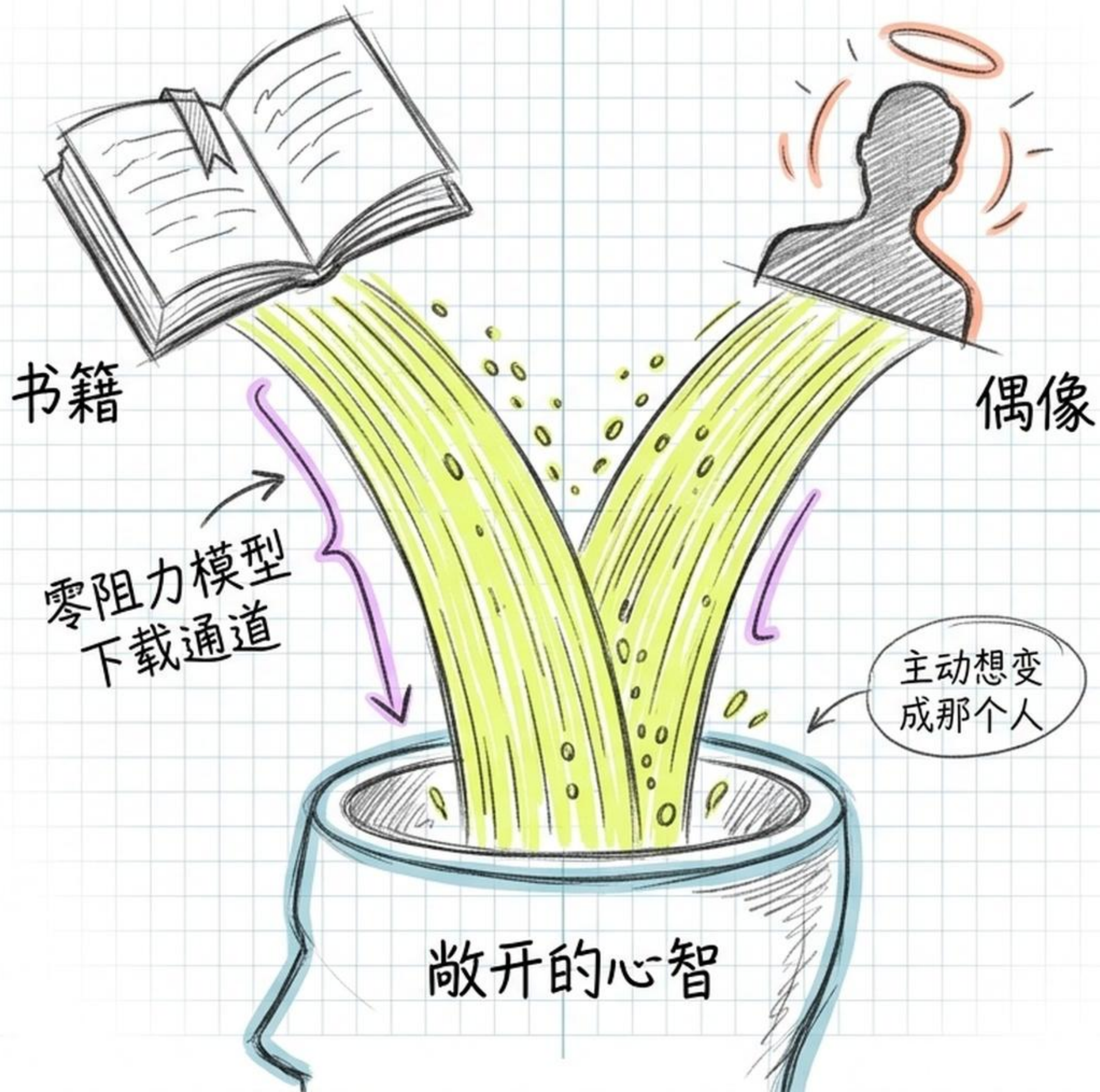
策略一：成为环境设计师

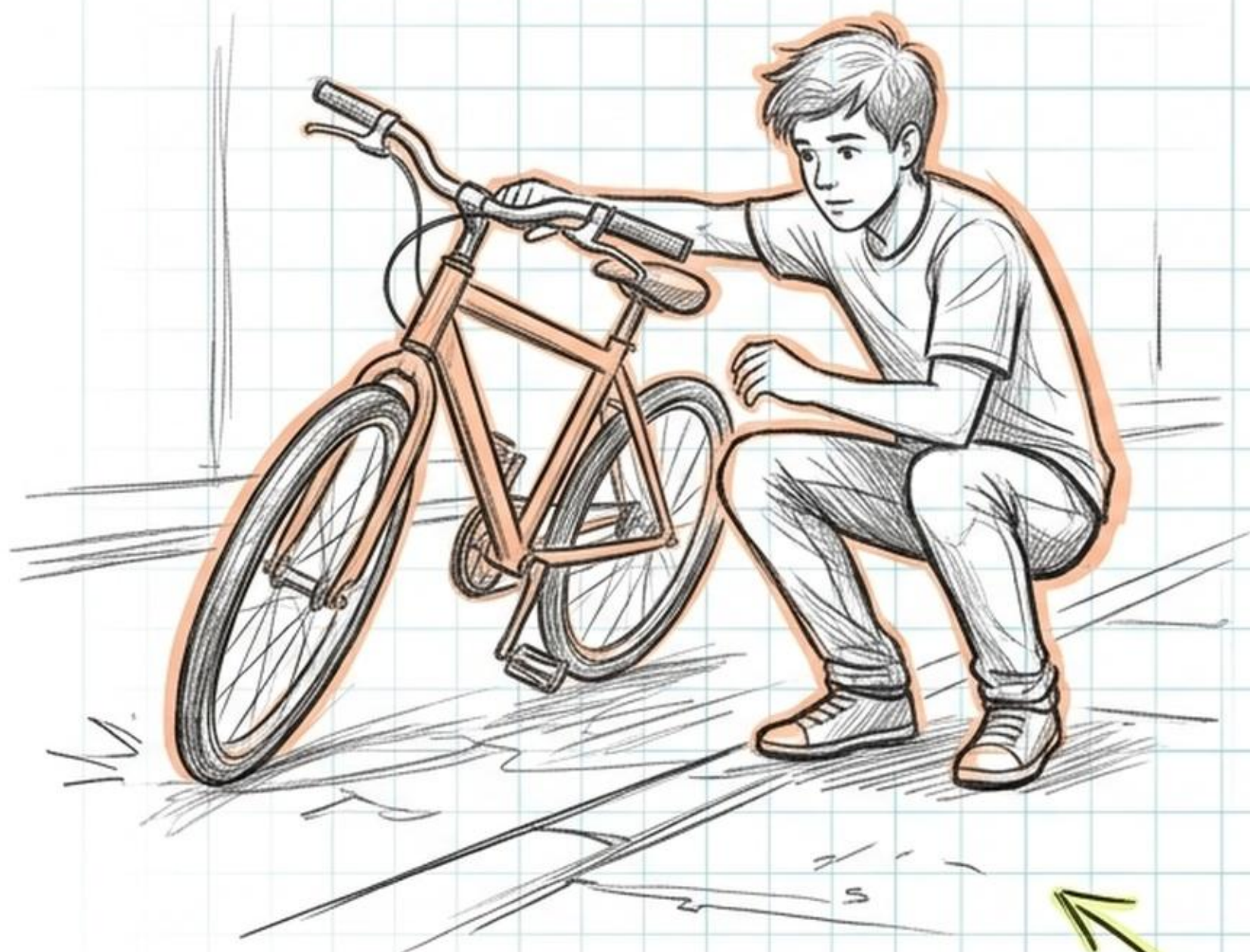


● 现实会说服他

换一个发射源

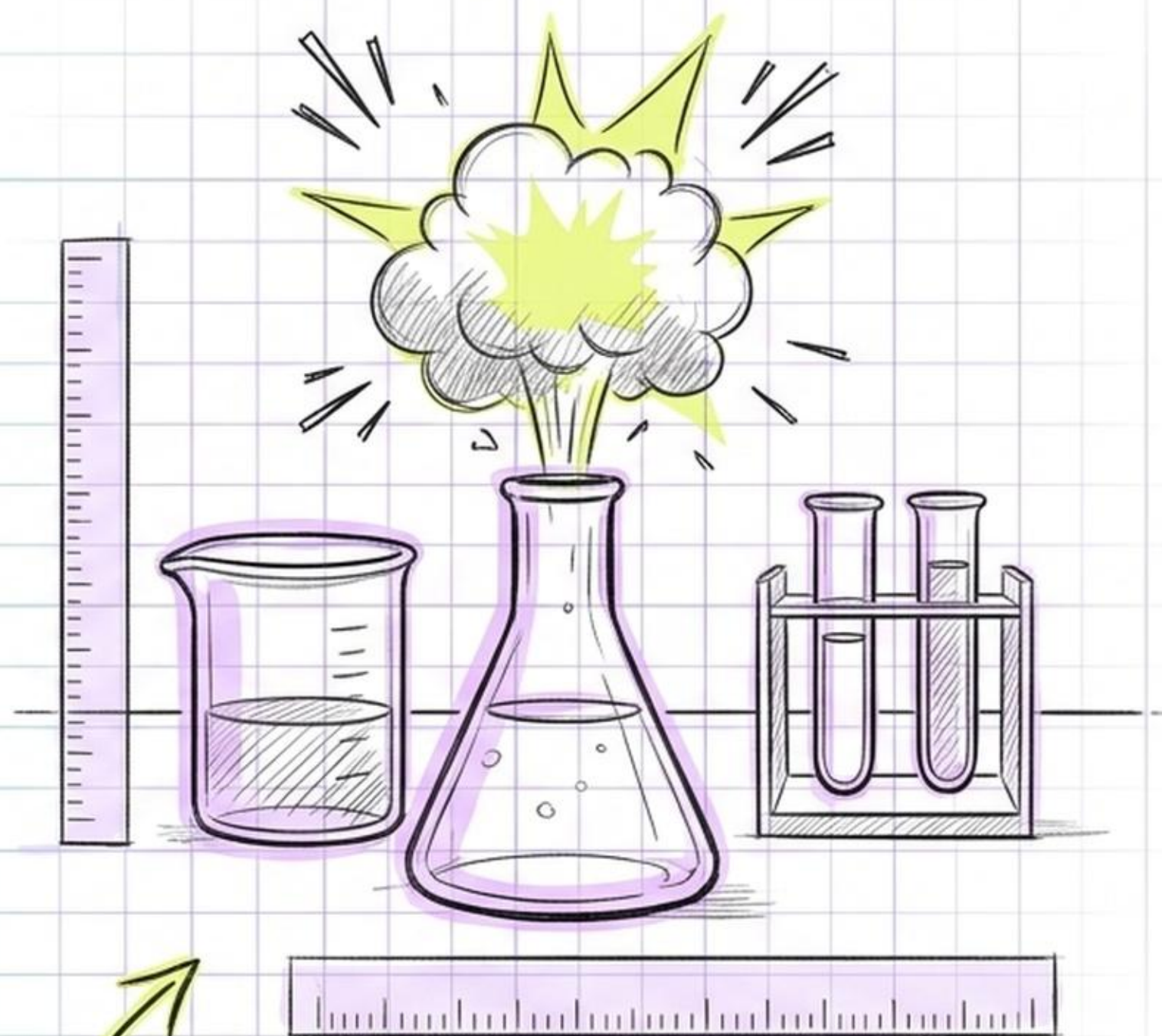






允许摔跤

必要的模型撞墙

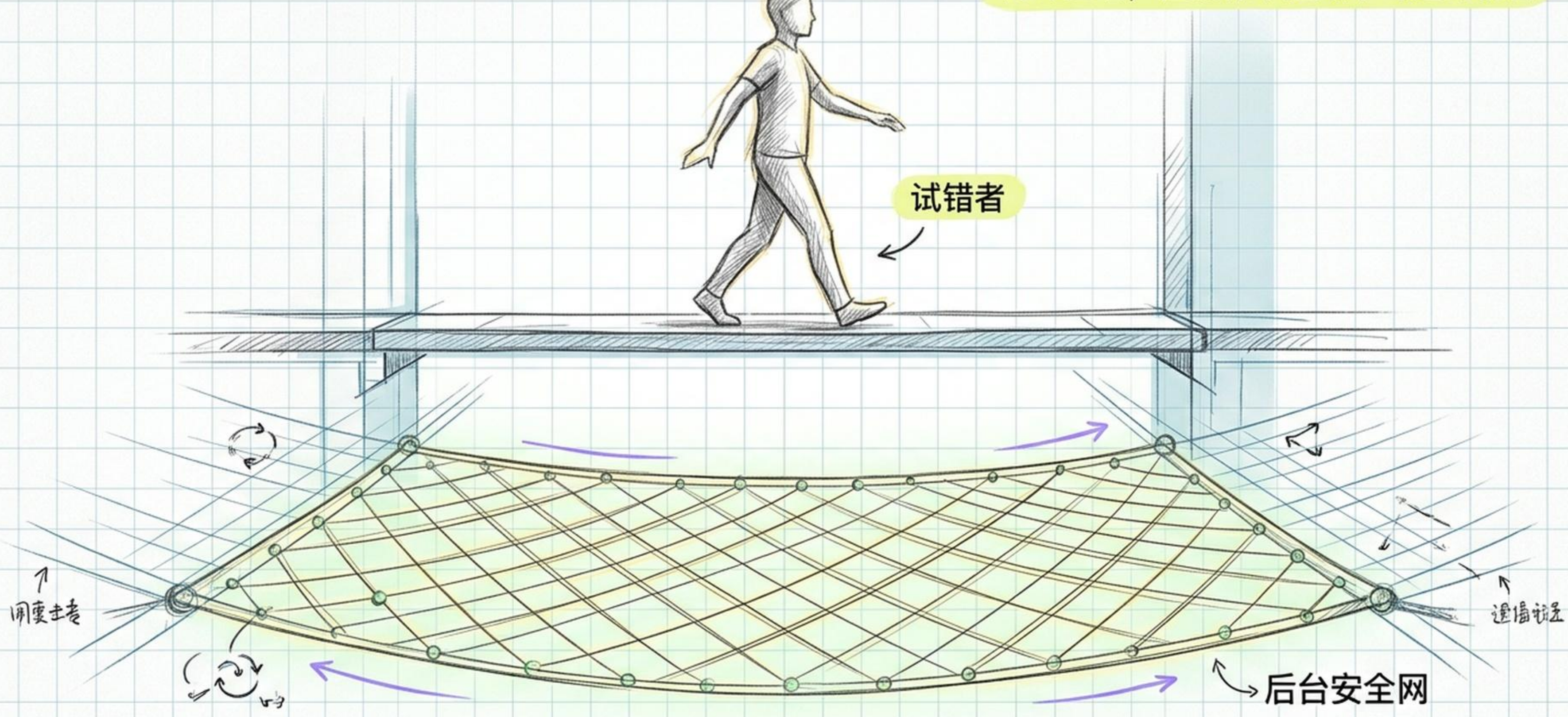


小型纠错爆破

撞墙是模型校准的必要输入

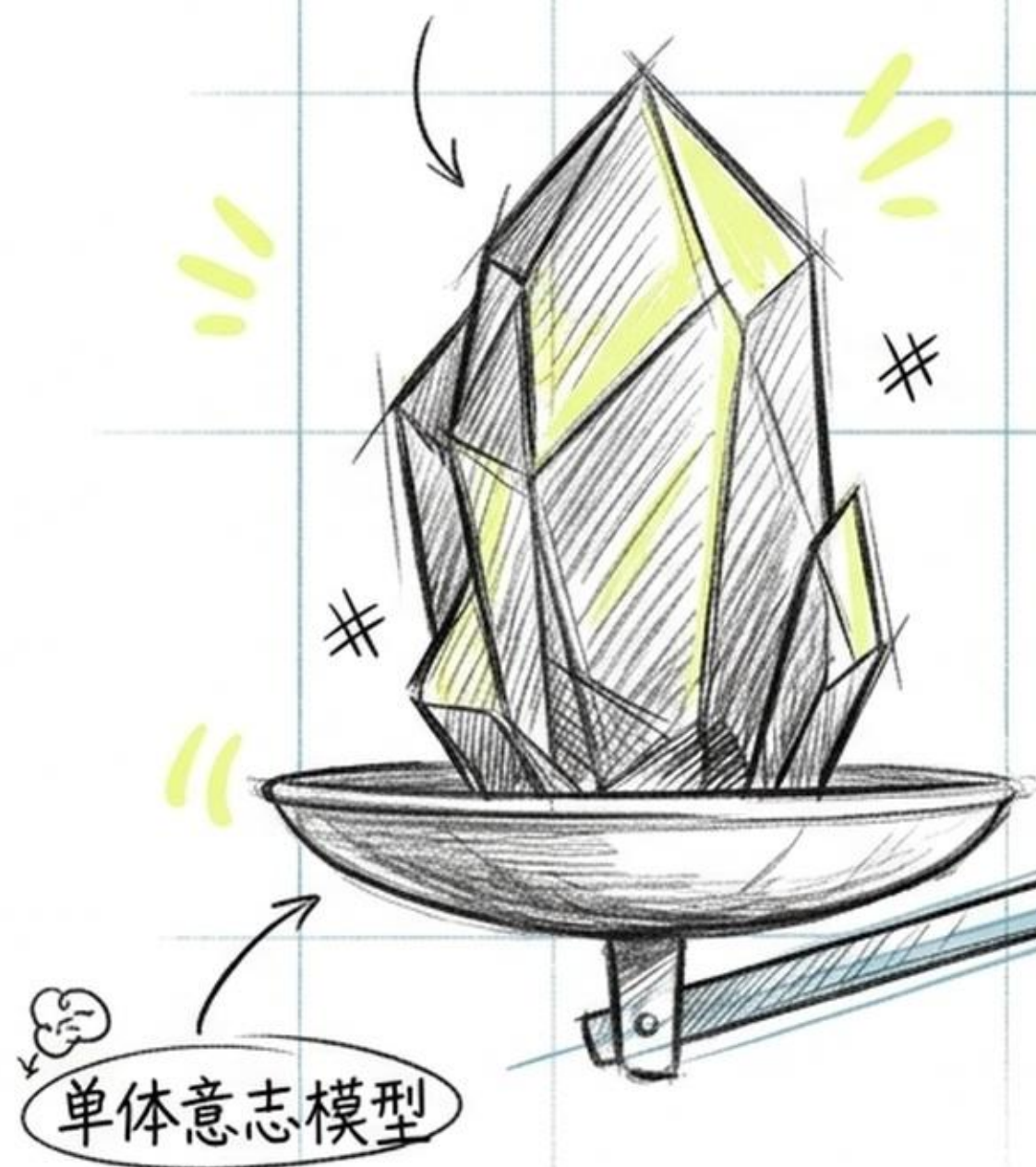
允许撞墙，但确保不造成不可逆的伤害

试错者



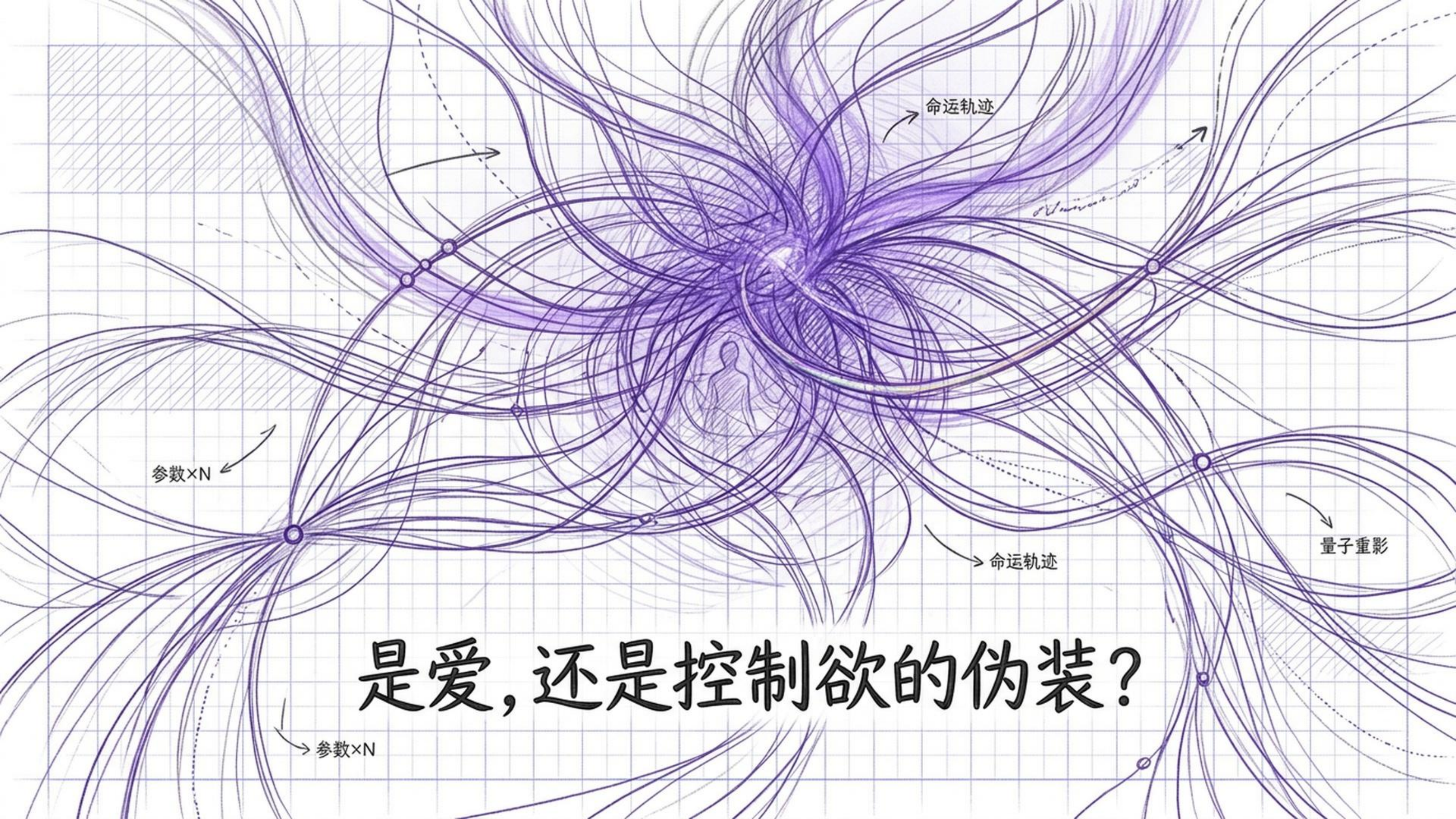
反向传播：允许犯错，梯度回传才能更新参数

我知道什么对你好



执念 vs 真实命运

复杂的真实生命轨迹



是爱, 还是控制欲的伪装?



先检查你自己的模型是否过拟合

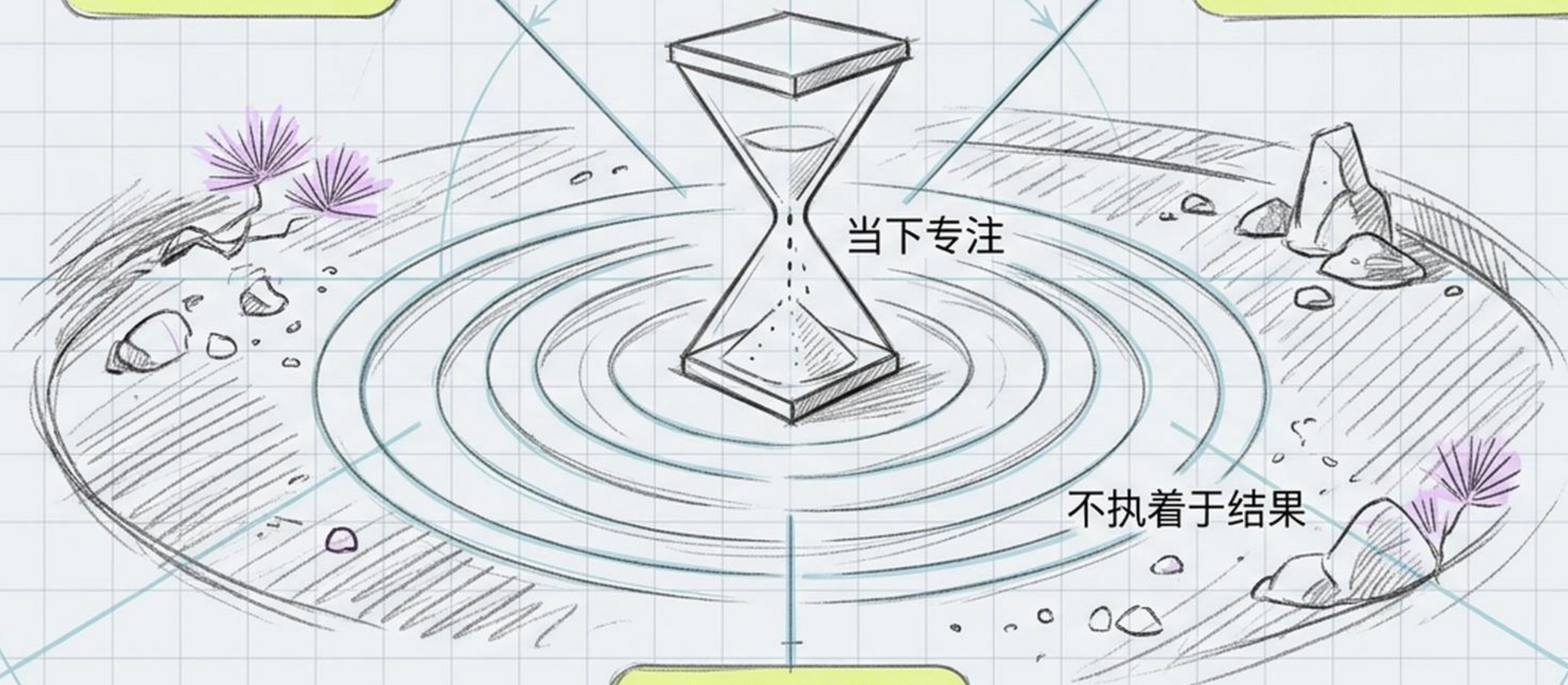
行动充分

接纳失败缺口

当下专注

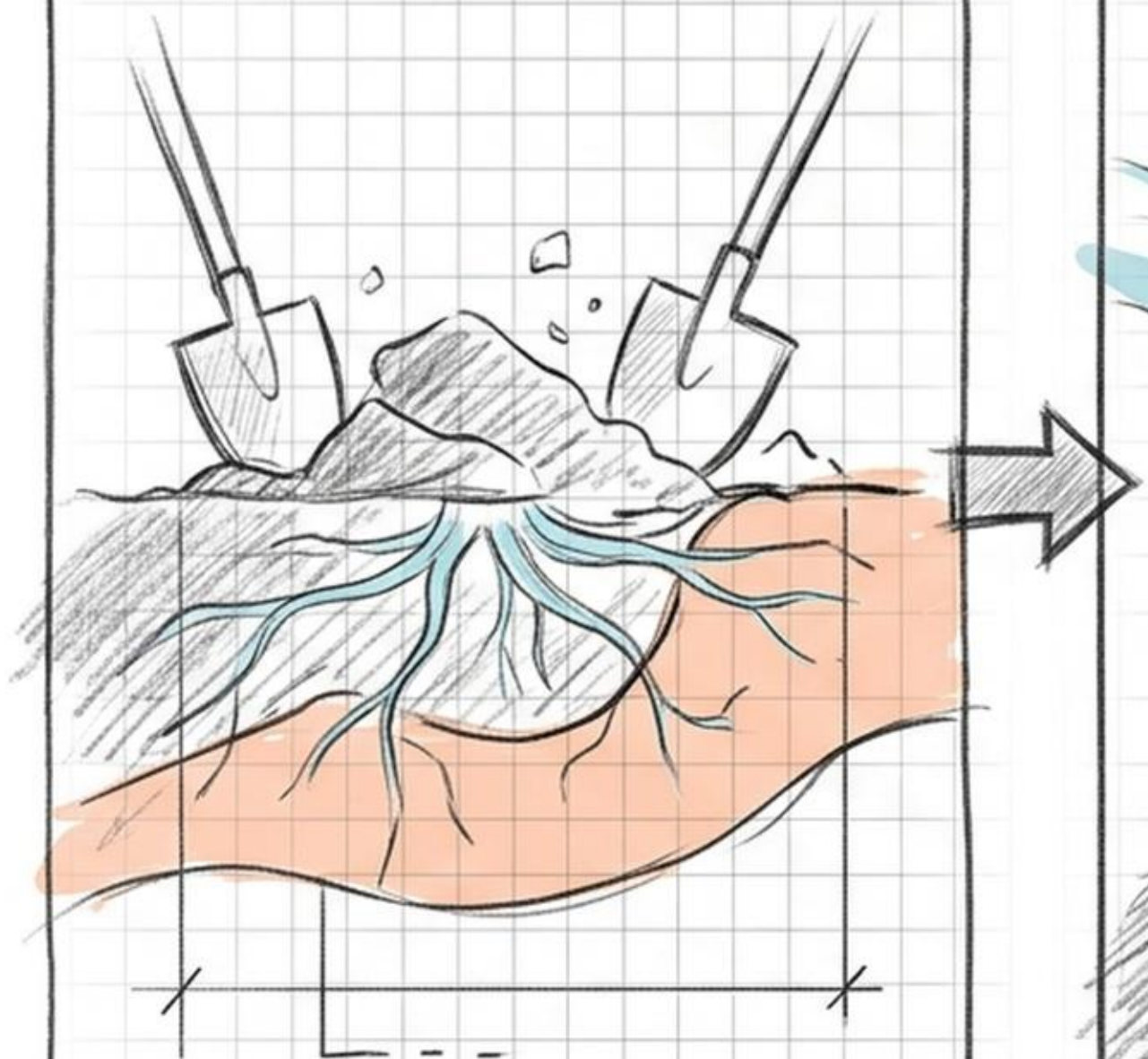
不执着于结果

允许缓慢

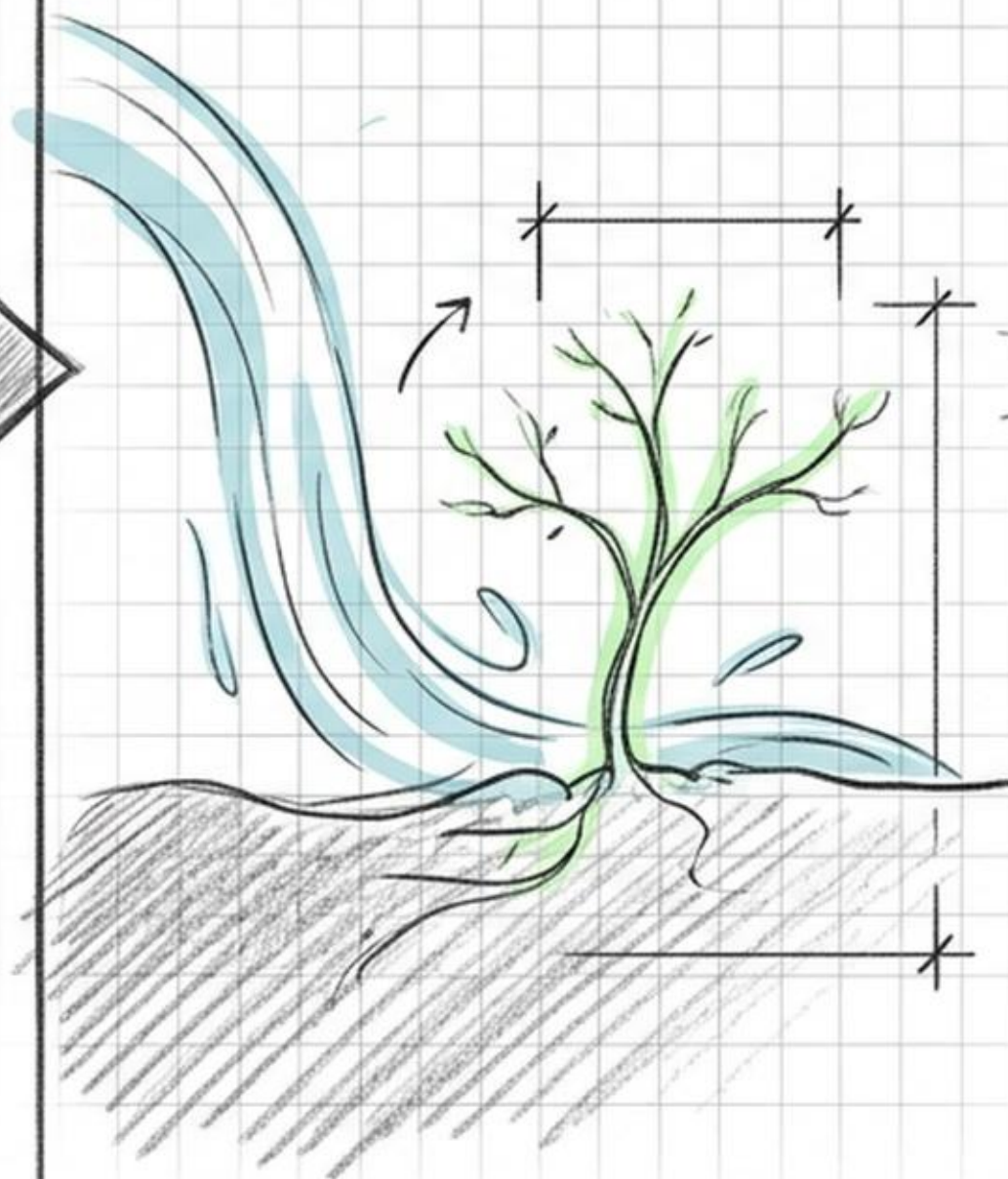


创造条件

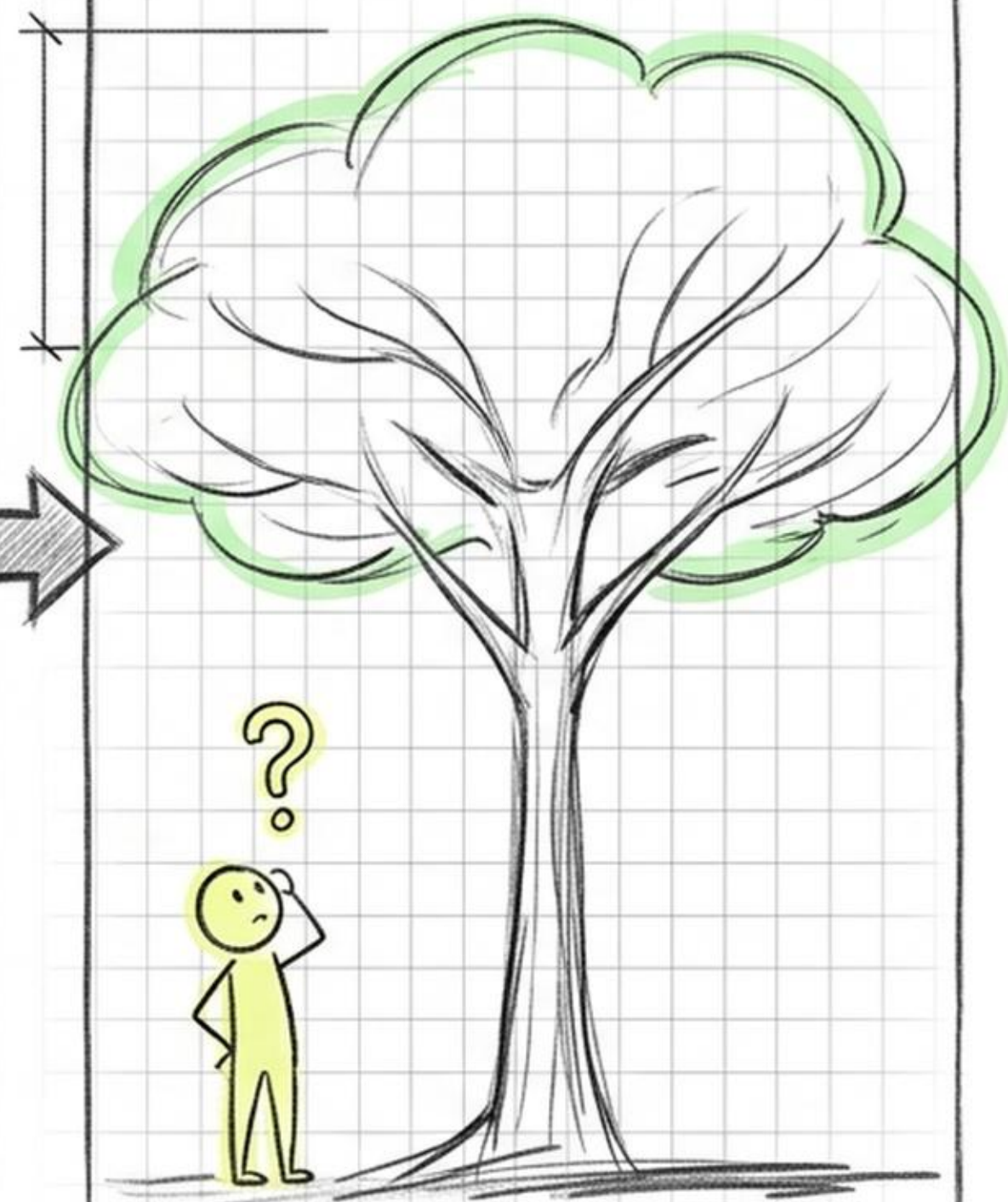
替换训练数据



引入替代模型示范



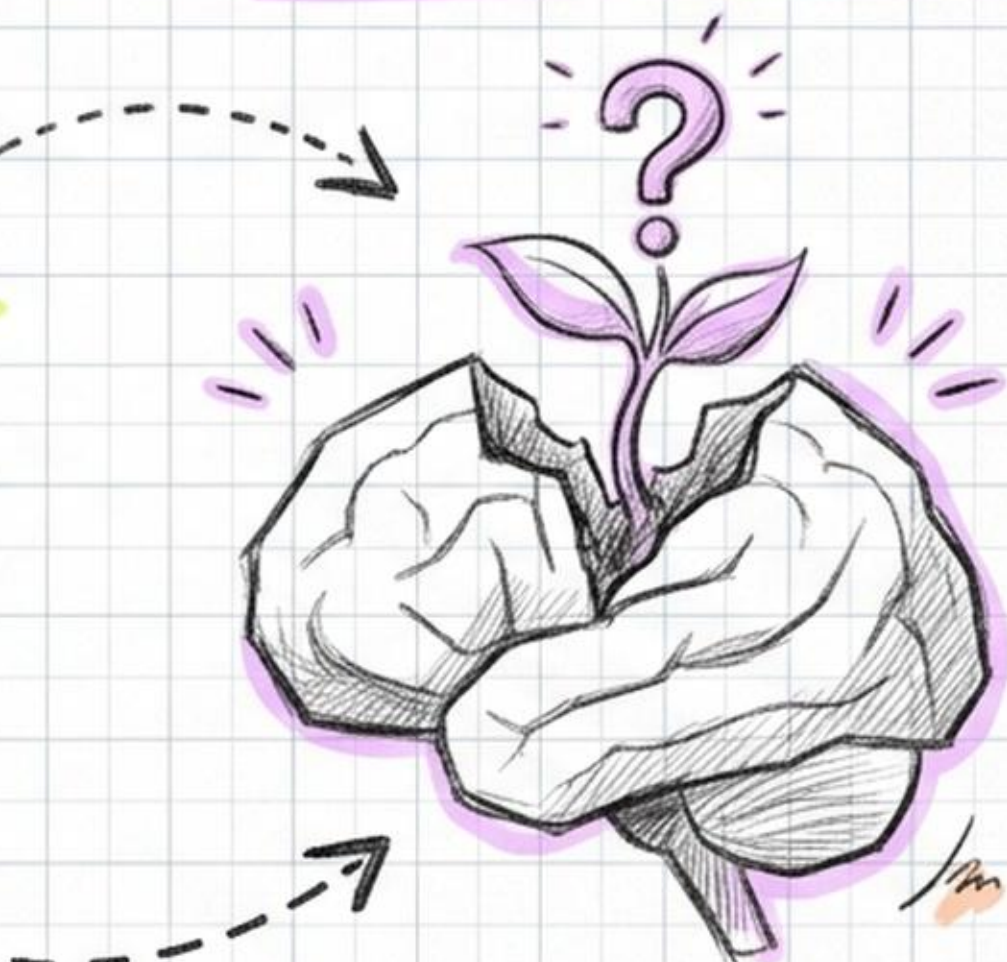
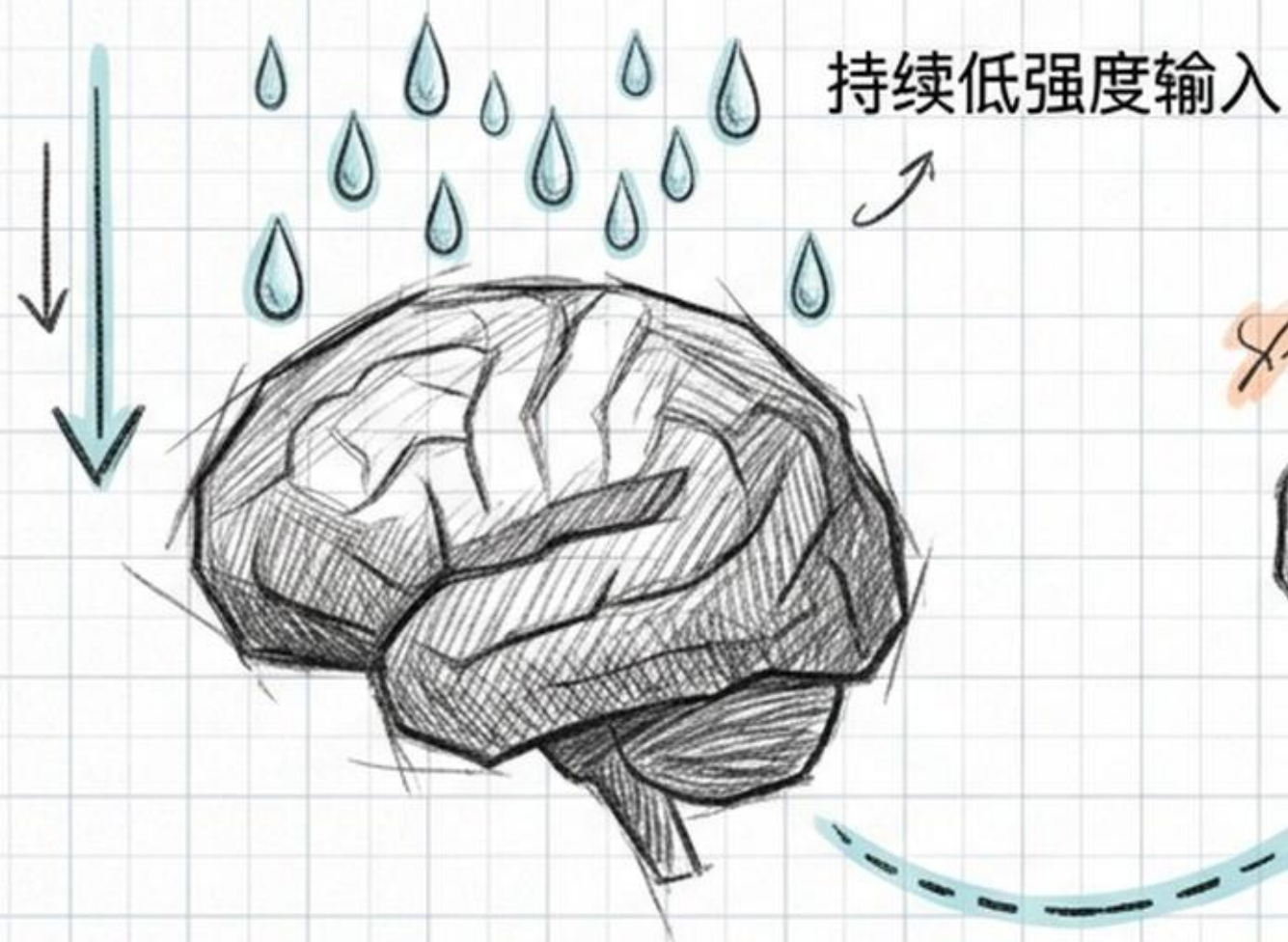
引发主动好奇



缓慢滴进信号

阈值击破

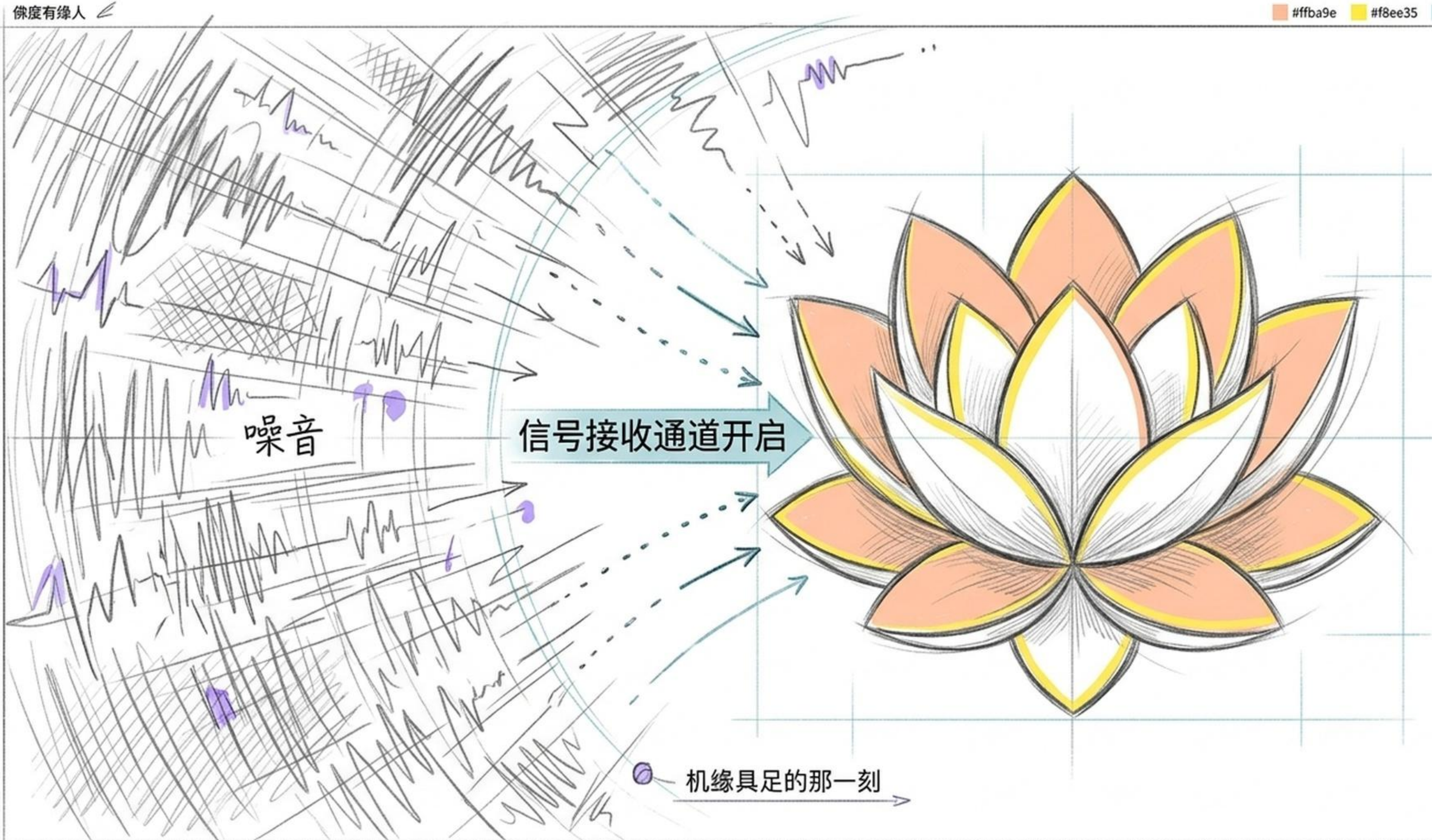
主动发问发芽

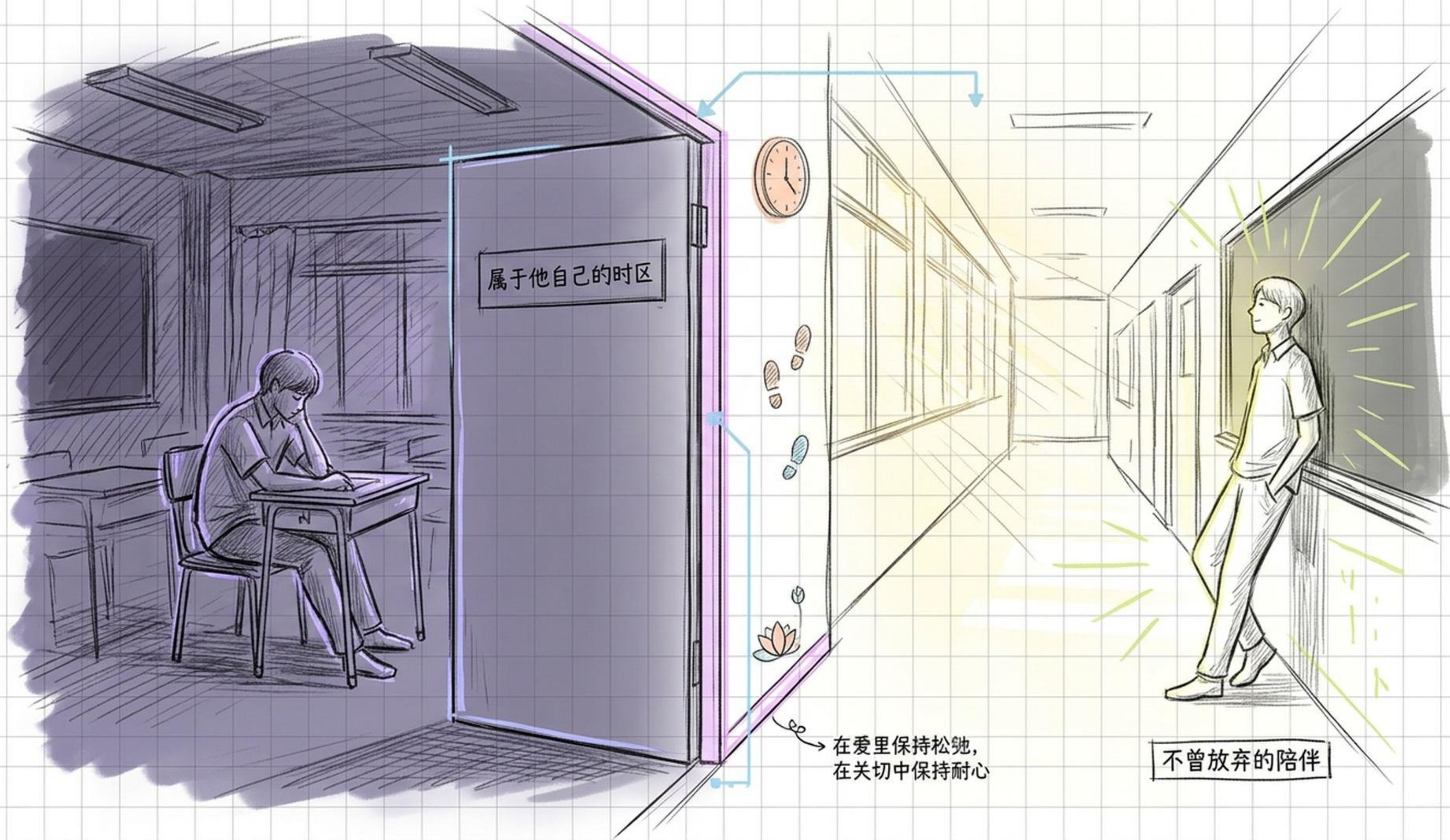


缓慢滴进信号

阈值击破

主动发问发芽





属于他自己的时区

在爱里保持松弛，
在关切中保持耐心

不曾放弃的陪伴

最好的礼物，是一个更好的你



我是王利杰，下期见