完全人工智能自主治理工厂：全能产线的实现路径

【韦季李】

【ORCID：0009-0006-5696-1873】

【通讯作者邮箱：jiligzs@qq.com】

摘要

本文提出一种能力可验证的全能产线（UPL）规范化建造路径，用于建厂期设备自供给与打样试产。其实现主线为：制造数字基座将物理世界转译为 AI 可操作的数字资产；四类集成动作涌现出以碰撞日志为反馈的自治仿真闭环，使动作流程无需上实机即可自主打磨；能力沿“固定流程→参数化流程→自主编排流程”的渐进验证阶梯逐步验证，最终构成电路、机构与固定工位三环节合一的全能产线。为使“完全”自治可判别，本文给出四项可度量支撑的准入判据——全链闭环、自主生成、自主验证、自主治理，并落实为可采集指标与建议验收阈值（代表性建议值：首次通过率 FPR ≥ 95%、生成成功率 GSR ≥ 90%、异常自闭环率 ASCR ≥ 80%、稳态人工介入率 HIR ≤ 5%，均为待校准建议值）；判定标准是以 AI 是否构成覆盖感知、制图、交互、仿真、决策与执行六环节的完整闭环并对其运行实施自主治理，而非以是否使用 AI 技术为标准。全文以证据边界表显式区分已构建的工程原型与待验证的研究议程，并给出端到端概念验证协议。本文定位为定义性贡献与研究议程，而非对已实现完全自治的宣称。

图形摘要：完全人工智能自主治理工厂：全能产线的实现路径

关键词：完全人工智能自主治理工厂；全能产线；制造数字基座；自治仿真平台；sim-to-real 迁移；渐进验证；自主治理

目录

1 引言：从概念到可检验的实现路径

制造业正在经历从自动化向自主化的范式跃迁。既有“黑灯工厂”“无人工厂”等概念侧重无人化执行，却未系统回答“工厂的自主能力从何而来、由谁治理”这一根本问题[1]。在先前的研究中，我们提出人工智能自主治理工厂（Artificial Intelligence Autonomous Governance Factory, AI-AGF）概念，强调 AI 系统在无人干预下自主完成感知、决策与执行全过程，并对自身目标、规则、边界与风险实施系统性治理，同时给出四维治理框架与四阶段启动路径[25]。

针对这一背景，本文主张一个可争议的立场：当前领域把“工厂用了 AI”与“工厂自主”混为一谈，是一处根本性的概念误判。自动化的历史已经表明，在产线中部署机器学习或大模型组件，可能只带来局部的智能增强，而丝毫未改变“目标由人设定、流程由人编排、风险由人承担”的整体结构。若以“是否使用 AI 技术”作为自治性的判据，任何插入了 AI 模块的工厂都可以宣称自己走上了完全自治，而该宣称既不可证伪、也不可比较。本文认为，衡量“完全”自治的正确单元是**闭环的完备性**：感知、制图、交互、仿真、决策、执行是否形成连贯闭环，每个环节是否由 AI 自主完成，治理是否由 AI 自主实施。这一判据使“完全自治”从一个修辞性目标变成一个可在设备级与产线级判别、可度量、可审计的工程边界。以下各节给出的四项准入判据、度量框架与验证路径，都是为了把这个立场落到可检验的形式。

1.1 “完全”准入判据

据此，我们提出四项“完全”准入判据，作为与前述概念衔接、可在工程上被度量的操作性定义。判据 F1（全链闭环）要求感知、制图、交互、仿真、决策与执行各环节均由 AI 承载，任意环节缺失即退化为局部智能化；判据 F2（自主生成）要求 AI 能够在给定目标下自主生成动作流程与执行程序，而非仅匹配预设模板；判据 F3（自主验证）要求 AI 在仿真环境中自主完成动作流程的验证、迭代与优化，并将“仿真验证通过”作为上实机的必要条件；判据 F4（自主治理）要求 AI 在生产运行期实时监控执行状态，对异常进行根因分析、方案替代与经验沉淀，形成自我改进闭环。四项判据共同构成全 AI 自治工厂的充要性边界；未能满足者，应被客观描述为“部署了 AI 技术的智能制造工厂”。为使判据可检验，后文将 F1–F4 逐一映射为可采集指标、判定规则与建议验收阈值，使其从定性承诺转化为可判别条件。

为使“完全自治”的判定与实证结果的关系不被误读，此处将判据明确拆解为两层：“判据结构”指六环节闭环完备、指标定义式与判定规则的组合，其中不含任何数值参数，它界定了“什么是完全自治”，故不依赖任何实证结果即可被评估；“参数实例”指表 2 给出的具体建议阈值，属工程经验与统计校准对象，它只裁定“某工厂当前是否达标”，而不决定“边界是什么”。PoC 完成后的阈值校准仅更新参数实例、不修订判据结构，阈值漂移只会改变该产线当下的达标状态，不会移动“完全自治”的概念边界。全文如本文图形摘要所示：物理世界经数字基座转译、自治仿真平台

闭环迭代、沿渐进验证阶梯逐级验证，最终由 F 判据对该闭环的完备性与自主治理给出可判别的边界。

1.2 研究问题、定位与证据边界

前述研究侧重回答“为何要做、形态如何、如何启动”，属于概念与宏观路径层面。本文承接其四阶段启动路径，将目光聚焦于其中最具工程挑战性的全能产线（UPL），回答一个更具体的问题：在既定概念框架下，一个全 AI 自治工厂的全能产线应当如何被真正建设出来，且这一路径如何被客观验证。这里的“全能”是有边界的：本文所讨论的全能产线，指建厂期内能够承担设备自供给与打样试产任务的产线形态——以有限的运动学组合覆盖设备制造与试产所需的动作流程谱，而非对任意制造任务的无限适应。该限定在术语表（附录 C）中与 UPL 的定义保持一致。

在此需明确本文的体裁与证据边界。本文属于**立场文 + 研究议程（Perspective with a concrete research agenda）**，而非已完成大规模实证的实验研究。全文的“已实现”部分指制造数字基座的工程原型与仿真闭环环境搭建等基础工作，其状态与证据在 2.3 节以证据边界表显式给出；“预期/待验证”部分以研究议程与验证协议的形式呈现，并给出首要验证里程碑与端到端概念验证（PoC）协议（附录 A）。我们刻意避免以“已证结论”的措辞描述尚未完成的实证——凡未实测即不作完成态断言。值得注意的是，本文的定义性贡献——即 F1–F4 判据及其可判别性——不依赖任何实证结果即可被独立评估；论文对实证的承诺范围由附录 A 的协议明确界定，二者不互相遮蔽。

要使“完全自治”的边界在跨领域讨论中可定位，还需说明本文判据与既有自主性等级框架的关系。工业与机器人领域已有的自主性刻画大体沿两类思路展开：一类是面向载具的序数分级（如 SAE J3016 按驾驶责任在人与系统间的分配划分六个等级[26]），另一类是面向任务与使命的自主能力/成熟度评估，按系统能在何种任务复杂度下自主完成何种职能来划分能力档位。本文的四判据不构成一条新的序数刻度，而是给出一个**能力完备性边界**：它不问“系统处于第几级”，而问“感知—制图—交互—仿真—决策—执行的闭环是否完整、其各环节是否由 AI 承载，治理是否由 AI 自主实施”。在这两类谱系上，本文定义可被理解为把“最高等级”的语义替换为“闭环完备且各环节可度量”的工程表述；对达不到边界的系统，则给出“面向完全自治的进度状态”这一中间标签（3.2 节），从而与序数谱系下“部分自动化”的过渡语义兼容。这一映射表明，“完全自治”在本文中是一个可判定的工程边界，而非对任何外部分级最高档次的借用。

在方法上，本文综合采用概念演绎（基于 F 判据推导边界与基座设计要求）、工程推演（以系统集成视角演绎从数据采集到动作流程生成的链路）与案例剖析（以非标搬运设备为完整案例给出可复现的实施流程），并广泛吸收数字孪生[3-4]、工业

人工智能[5]、大语言模型驱动的机器人任务规划与程序生成[6-7,17-19]、仿真反馈修复[8]以及 sim-to-real 迁移[20-21]等领域的既有成果，与其系统对照（第 6 章）。

在此亦需与四类既有范式划清边界。其一，与“黑灯工厂/无人工厂”的关系：后者强调无人，本文强调自主 + 治理；黑灯工厂可以全程自动化却仍由人制定全部流程，而全 AI 自治工厂要求 AI 具备流程的自主生成与治理能力（判据 F2–F4）。其二，与“部署智能体的工厂”的关系：部署智能体只是局部智能化，缺少全链闭环与自主验证机制，不满足完全判据。其三，与数字孪生的关系：数字孪生是本体系的使能技术（建模与虚拟验证），制造数字基座则是数字孪生模型的数据来源，二者是“供应—消费”关系[3-4]。其四，与可重构制造系统（Reconfigurable Manufacturing Systems, RMS）的关系：RMS 通过模块化与可重构实现工艺柔性[14]，本文进一步将“重新配置”的动作交由 AI 自主完成，是从“结构可重构”向“能力自组织”的跃迁。

2 制造数字基座与能力涌现

构建完全人工智能自主治理工厂的前提，是让 AI 能够“看得见、画得出、传得通”，即理解物理世界、生成工程对象并与人类高效协作。我们将支撑这三项能力的软件集合统称为**制造数字基座（Manufacturing Digital Base, MDB）**，即把物理制造世界中的对象、行为与知识转化为 AI 可理解、可操作的数字资产的软件集合。此处需与业界常见的“数据底座/基础平台”提法作一区分：通用数据平台以存储、流转与汇聚数据为职能，而 MDB 的职能是**转译**——将设备几何、图纸语义与位移行为转换为 AI 可直接调用的模型、参数与可执行脚本，其输出物是制造数字资产而非数据集本身。MDB 的系统边界为：输入端接收物理世界信息（扫描数据、图纸、参数、需求文本），输出端提供数字资产（三维模型、图纸数据、位移参数、可执行脚本），对内供自治仿真平台调用，对外供人机交互审核。需要强调的是，MDB 并非某个具体应用，而是承载上层仿真、规划与治理能力的公共底座，其“数字”属性强调了对物理制造世界完成数字化、资产化转译的本质，可与数字孪生[3-4]衔接：数字孪生是“模型”，而数字基座是“制造数字资产的生产流水线”。制造数字基座的三层架构如图 1 所示。

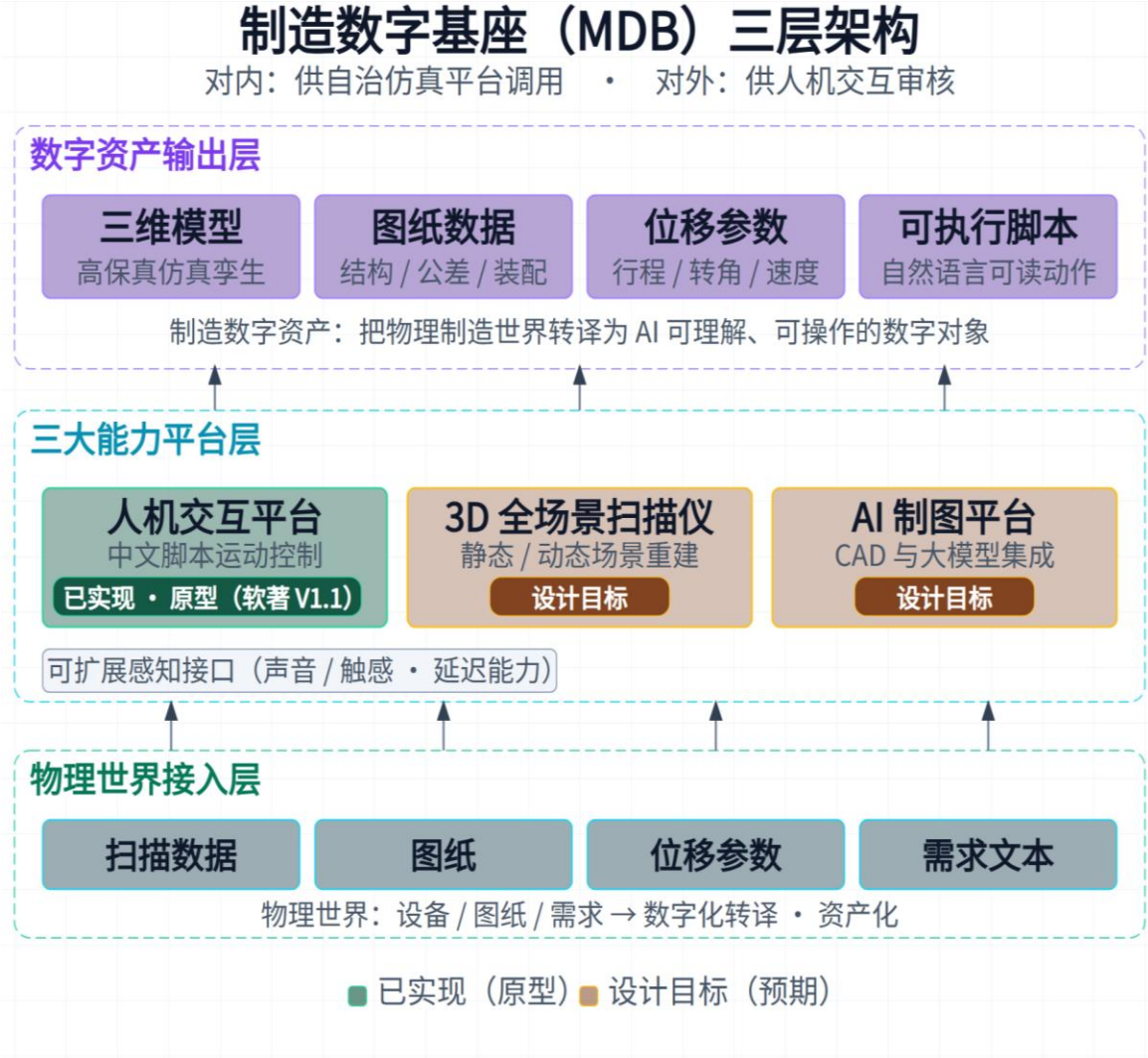

图 1 | 制造数字基座（MDB）三层架构

2.1 三类基础软件的设计

制造数字基座由三类软件构成，它们分别解决“人能看懂代码、机器能被看见、图纸能被生成”三个基础问题，并预留统一的感知扩展接口。

人机交互平台（自然语言脚本运动控制）。 AI 动作流程生成的产物最终须落实为可执行代码，而人类工程师需要能够审阅、理解并修正该代码；若 AI 直接生成晦涩的底层指令，人机协作的审核环节将无法成立。因此，人机交互平台的核心诉求是让 AI 输出的动作代码对人类可读、可审、可改。这一设计以我们已构建的中文脚本运动控制平台原型为载体（前置工程成果，已获软件著作权登记 V1.1）：它将底层运动控制（点位运动、直线插补、圆弧插补、IO 控制、多轴联动等）封装为接近自然语言的脚本指令，形成“运动控制函数库 + 可读脚本”的双层结构。该设计与大语言模型用于机器人控制的主流范式一致：Vemprala 等人论证了“高层函数

库 + 自然语言交互”架构使大模型无需重训即可适配不同机器人形态与仿真环境，将用户意图翻译为可执行代码[9]。这一范式在非标自动化领域的工业落地中，有三个特性尤为关键。其一是可审核性，AI 生成的每一行动作脚本工程师均可在不依赖厂家私有协议的情况下直接读懂，从而胜任审核者角色；其二是可移植性，脚本作为中间表示可映射到不同运动控制卡的底层指令，避免 AI 输出与硬件强耦合；其三是可叠加性，在脚本基础上可进一步叠加工艺参数、时序避让规则与安全约束，为后续自主仿真与治理预留接口。此外，该平台也是异常追踪与审计溯源的重要载体——每次执行的脚本与参数均构成可回溯的“动作档案”。

3D 全场景扫描仪（静态与动态场景重建）。 AI 治理工厂的前提是“看见”真实设备。3D 全场景扫描仪的设计目标是以手持设备或轻量摄像头级别的硬件实现设备与场景的三维重建，覆盖两类能力。静态 3D 场景扫描对整机、模组、工装与产品建立高保真三维几何模型，用于搭建仿真环境并生成点云以支持位姿识别；动态 3D 场景扫描则在设备动作过程中实时采集场景变化，用于验证零部件的位移参数，例如轴的行程范围、气缸的上下限位与运动部件的实际轨迹，从而将“图纸上写的参数”校准为“实际跑的参数”。动态扫描是本文方法论的特征环节：它不仅服务于仿真建模，更服务于逆向数据采集（2.5 节），使工厂在运行过程中持续产生可用于 AI 学习的动作数据。在技术支撑上，面向工业抓取与装配的位姿估计方法已能在点云基础上实现稳健的六自由度（6-DoF）位姿回归[10]，为扫描数据直接驱动机械臂作业提供了成熟底座；数字孪生研究亦表明，结合实时感知数据构建孪生模型是支撑运行期监测与决策的关键[3]。本组件的设计与性能目标尚待原型化验证，其状态按 2.3 节口径以“设计目标”对待。

AI 制图平台（CAD 与大模型集成）。 传统图纸设计高度依赖工程师经验，是制约工厂快速响应新订单的瓶颈。AI 制图平台的设计目标是实现计算机辅助设计（Computer-Aided Design, CAD）与 AI 大模型的集成，使 AI 在收到客户需求后自主完成从需求理解、方案设计到全部图纸输出的过程。在实现路径上，该平台将设计知识（标准件库、公差体系、结构约束、工艺规范）沉淀为可检索的工程知识库——制造领域知识图谱的研究为这类知识库的构建提供了系统化方法基础[12]——由大语言模型对需求文本进行语义解析，并调用参数化建模接口逐步生成零件图、装配图、物料清单（Bill of Materials, BOM）与机构运动模型。平台输出并不孤立存在，而是作为后续仿真的输入数据之一（2.4 节），形成“设计—仿真—验证—优化”的闭环。与 3D 扫描仪一致，该平台的当前状态为设计目标，性能表述均为待验证假设。

可扩展感知（延迟能力与统一接口）。 声音（振动噪声、气动动作声、异常声响）与触感（力、压力、力矩）在装配力控与故障诊断场景具有重要价值，但考虑工程投入与当前验证优先级，我们将其列为“需要时再开发”的延迟能力，并在 MDB 中预留统一的感知接入接口，以避免重复建设。

2.2 四类集成与能力涌现

构建完基础软件后，能力不会自动涌现，还需通过平台集成把数字资产组织为决策能力。本组件的核心是四类集成动作，其涌现产物是“以碰撞日志为反馈的自治仿真平台闭环”。

数据集成：高保真数字孪生模型。 将 3D 扫描得到的场景数据、对应设备的全部图纸与零部件的位移参数（标准设备另附使用说明书）统一汇入仿真平台，经配准、装配与语义标注，目标是获得高保真的仿真模型，其三个特性构成后续碰撞检测的基础。其一，结构完整，每个零部件归属于明确的模组且装配关系与层级关系可查询；其二，语义可标，每个模组可标注名称、功能、规格以及可位移范围（行程/转角/角度）；其三，行为可拟，结合位移参数，模型运动学与运动范围与实机一致。这一过程与工业数字孪生“虚实映射”理念一致，其核心挑战之一正是模型精准性与数据有效性[3]；本文通过“扫描几何 + 图纸结构 + 位移参数”的三源集成方式在工程上缓解该挑战，使仿真模型具备“可验证级”的精度取向。数据与知识混合驱动的思路在离散制造系统智能控制中已有大量实践[16]，可为本节模型构建提供参考。

AI 集成：自治仿真平台与闭环迭代。 若将数据集成比作编写程序前的“函数定义”，本项即“AI 根据需求进行动作流程开发”的阶段。其核心机制是五步闭环（图 2）：AI 接收“完成某项动作”的高层目标（如将产品从料仓取出并搬运至工位）；基于仿真模型中的模组定义与运动学约束自主规划动作序列并生成仿真可执行流程；仿真平台运行该流程并输出碰撞提醒与干涉日志；AI 读取碰撞日志作为异常调试信息，修正动作时序、速度与路径后再次仿真；反复迭代直至获得无碰撞、满足节拍与工艺要求的流程。这一“生成—反馈—修正”闭环与大语言模型工业任务规划中“以执行反馈修正规划”的方向一致[7-8]。例如，已有研究提出轨迹引导的领域修复框架，通过执行失败的精准归因与结构化提示，使大模型生成的规划从“语法正确”提升为“物理可执行”[8]。自治仿真平台的贡献在于将这一思路机制化为通用闭环：每一次仿真碰撞反馈相当于一次代码调试日志，AI 以日志为输入持续优化直至收敛，从而使工厂获得“不上实机”即自主打磨动作流程的能力。该闭环的收敛性、终止性与保真度误差传导在第 5 章作为分析性命题展开，并在附录 A 的 PoC 协议中被设计为可实证检验的指标。

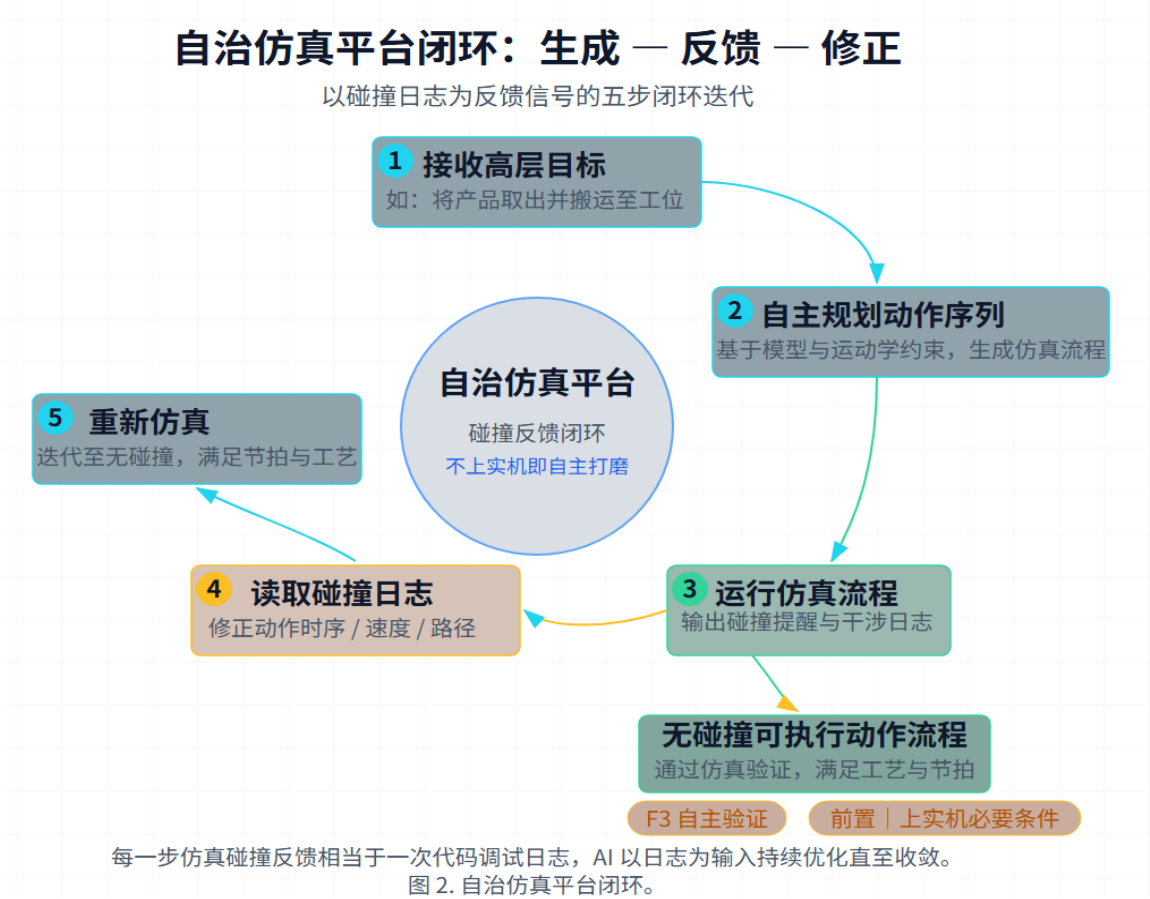

图 2 | 自治仿真平台闭环

人机交互集成：实机运行前的审核闭环。 自治仿真平台给出的流程在实际运行前仍需一道“人机共治”审核环节。将人机交互平台集成进来，可实现：可读呈现，将 AI 生成的流程翻译为中文脚本与可视化步骤供工程师快速理解；前置审核，在实机首次运行前由工程师对高风险动作与异常工况预案进行确认，规避设备首个运行周期（生涯首次运行）可能出现的异常；动态收敛，当同一类流程经多次实机验证趋于稳定后，逐项人工审核可取消而改由 AI 自主执行；档案保留，交互记录持续保留，作为异常发生时追踪定位与审计溯源的依据。这一环节与此前提出的“四阶段启动路径”中“AI 启动、人做审核”的人机共治思想一脉相承[15,25]，也与增强现实辅助的互认知人机安全交互强调的“人在回路上”理念相互印证[11]；面向工业 5.0 的人本融合技术为这种人机共治模式提供了面向生产场景的方法支撑[15]。

制图集成：贯穿设计—仿真—制造。 AI 制图平台并非独立的末端环节，而是贯穿始终的源头能力。其作用机制为：获取客户需求后生成设备图纸数据；图纸数据配合位移参数进入仿真平台进行运动学仿真与机构干涉检查；仿真结果反馈机构不合理之处（干涉、行程不足、刚度过低等），AI 据反馈优化图纸；优化后的图纸再次进入仿真验证，直至机构设计合理。由此，AI 制图平台与自治仿真平台形成“设

计—仿真—优化—再设计”的协同闭环，使设备设计错误在虚拟空间中被充分暴露与修正，显著降低实机阶段返工风险。

2.3 证据边界与当前状态

为确保上述叙述不被误读为已完成实证，我们对每个主要组件显式声明其当前状态与可核查性。表 1 是全文的证据边界表：凡标注“原型”的组件，其存在性可由软件著作权登记或可运行的本地环境核查；凡标注“设计”的组件，仅给出设计目标与实现路径，任何关于其性能的表述均为待验证假设。全文将据此使用统一措辞口径——“原型”组件使用已实现/已构建的表述，“设计”组件一律使用设计目标/预期/规划及其等价表述。

组件	当前状态	可核查证据	性能表述口径
中文脚本运动控制平台	已构建原型	软件著作权登记 V1.1	可运行；相关性能指标待 PoC 量化
3D 全场景扫描仪	设计目标	无独立交付物	设计目标与集成设想，待原型化
AI 制图平台	设计目标	无独立交付物	设计目标与集成设想，待原型化
自治仿真平台闭环	机制设计 + 环境设想	无独立交付物	机制设计与收敛性命题，待 PoC 实证
F1–F4 判据与度量框架	定义层（已完成形式化）	本文第 3 章	定义与规则可在工程上直接判别

表 1. 证据边界与组件当前状态。凡未进入“已构建原型”状态的组件，全文一律不作完成态断言。

2.4 顺向与逆向：双向完整动作流程

平台集成的最终产出是两条互为补充的完整动作流程获取路径。顺向路径（人 → AI）由工程师或客户输入设备定义与动作目标，AI 基于图纸、位移参数与仿真验证正向生成真实设备的完整动作流程，是对当前人工动作流程开发的自动化替代。逆向路径（实机 → AI）通过对设备实时运行的 3D 扫描，逆向提取设备的实际动作流程，形成训练数据；其用途有二：其一，设备故障时通过比对“期望流程”与“实际流程”反推故障原因；其二，将真实动作数据沉淀为数据集，用于 AI 持续学习与模型微调。顺向与逆向构成“先正向生成、再逆向校验、以逆向数据强化正向能力”的自举式进步，是工厂自治治理能力自我增强的机制基础。

2.5 延伸：标准设备的泛化操控

上述能力对非标设备与标准设备（机械臂、机器人）均适用，且标准设备操控实现更为直接。扫描机械臂与机器人并获取其操作方法和相关图纸后，即可在仿真环境中对其正常操控；标准设备的操作在说明书中已有完整定义，AI 可直接解析说明书与实时位姿确认操控方法，无需像非标设备那样进行“机构图纸匹配—模块识别—位移标定”的复杂推演。非标设备因工作目标可编程，本质上是把“动作流程开发”下沉为岗位能力；标准设备因说明书完备天然更易被 AI 接管。由此本文的能力体系呈现出“逆向收敛”规律：非标设备越难越先需要全链路 AI 支撑；一旦非标设备的全链路能力跑通，标准设备的泛化操控即为水到渠成。

3 判据的可操作化与度量框架

四判据为定性描述、难以度量。本章将四项判据转化为“定义 + 度量函数 + 判定规则”的形式，使其可在设备级与产线级两个粒度收集数据并判别。需要说明的是，本章所给验收阈值目前为基于工程经验设定的**建议值**，其合理性本身也是研究议程的一部分，将通过附录 A 的 PoC 协议在实际数据上校准。

3.1 度量指标的形式定义

我们对五项核心指标逐一给出可复核的形式定义，包括定义式、观测单元、统计窗口与采集/审计方。表 2 给出判据—度量—定义—数据来源—阈值—阈值依据的完整对照。各指标统计窗口均为“最近一个自然季度内的全部相关事件”，由运行日志自动导出并接受随机抽检审计。

为消除观测口径歧义，先约定本文的基本观测单元——**生产运行周期**：一条产线上从一个任务/订单开始到其验收完结的连续运行窗口，以任务/订单标识为边界的可观测区间；同一窗口内的所有感知、规划、执行与治理事件均归属该周期。所有按“周期”计数的指标均以此定义为准。

指标定义如下（以“观测单元”和“判定式”表述，非公式原文可以等价的文字形式核对）：

- **AI 环节覆盖率（CoC）**：对 F1。观测单元为一次生产运行周期的全部环节（感知、制图、交互、仿真、决策、执行）。定义为“有人工 AI 承载的环节数 / 环节总数”，要求六项环节均非空承载，即 CoC = 1。
- **人工介入率（HIR）**：对 F1。观测单元为一次生产运行周期。定义为“发生人工介入的运行周期数 / 总运行周期数”。为使“人在回路把关”与“AI 能力不足”两类性质不同的介入可区分，本指标拆分为两个子指标：复核放行介入率（HIR_r），指人工仅完成预设的放行/确认动作（如对流程的审批提交）即告结束的介入占比；纠错接管介入率（HIR_c），指人工对流程内容实施了修改或对执行实施了接管/干预的介入占比。建议稳态阈值 HIR（合计）

≤ 5%；HIR_c 系趋势指标，要求逐季度单独报告且呈下降趋势，连续两个季度上升即触发治理记录级审查（3.3 节时长口径），其收敛目标（趋近于 0）将在 PoC 校准后给出量级。

- **生成成功率（GSR）**（对应 F2）：观测单元为一条动作流程的生成任务。定义为“一次生成后未经人工实质修改即通过审核与仿真验证的流程数 / 生成的流程总数”（模板匹配外占比作辅助指标一并记录）。其中“未经人工实质修改”以可操作的二分类判据界定：若人工对流程的核查仅产生预设的放行动作、未改动动作序列或参数，则该流程计为“未修改”；任何对动作序列或参数的改动（即使一处）均将该流程重分类为“已修改”。建议阈值 GSR ≥ 90%。
- **首次通过率（FPR）**（对应 F3）：观测单元为一条动作流程的实机首轮执行。“通过验收”的判据绑定第 4.2 节六级调试阶梯的末级验收清单：带产品试跑在节拍与精度公差内完成、无碰撞/无干涉、无安全连锁触发，即视为通过。定义为“实机首次执行即满足该验收判据的动作流程数 / 实机上线的流程总数”。建议阈值 FPR ≥ 95%。
- **仿真收敛轮次**（对应 F3，辅助）：定义为“从首轮仿真到满足终止判据所消耗的迭代轮次”，记录并报告而不设硬阈值。
- **零参考首见任务规划成功率（ZFS-FPR）**（对应 F2/F3，高阶扩展）：观测单元为一条输入为“从未见过”的装配或协作任务——即无既有流程模板、无人工流程示例。定义为“AI 在零参考条件下独立生成、且首轮即通过仿真验证并满足六级调试末级验收条件的任务数 / 该类任务总数”。数据来源为流程生成日志、仿真平台日志与实机记录。建议阈值 ZFS-FPR ≥ 60%（设备级），该阈值同样为建议值，待自主编排级试点校准；统计窗口与审计口径同表 2。
- **异常自闭环率（ASCR）**（对应 F4）：观测单元为一次治理事件。定义为“由 AI 自主完成根因分析并闭环（含替代方案通过复跑验证）的异常数 / 审计期间异常总数”。建议阈值 ASCR ≥ 80%。
- **代际人工升级时间（ToE）**（对应 F4，辅助）：定义为“因治理能力不足而上报至人工的同类异常，其在两次代际升级之间的时间间隔中位数”，要求逐月下降并记录。

判据	度量指标	定义式与统计口径	数据来源	建议阈值（设备级）	阈值依据
F1 全链闭环	CoC；HIR（HIR_r / HIR_c）	六环节覆盖 = 1；HIR = 介入周期/总周期，其中复核放行与纠错接管分开报告	运行日志	CoC = 1；稳态 HIR ≤ 5%（HIR_c 另报）	工程经验建议值
F2 自主生成	GSR	免人工实质修改（二分类判据）通过审核	流程生成日志	GSR ≥ 90%（另报模板	工程经验建议值

判据	度量指标	定义式与统计口径	数据来源	建议阈值（设备级）	阈值依据
		与仿真的比例		匹配外占比）	
F3 自主验证	FPR；收敛轮次	首轮执行且满足六级调试末级验收的比例；收敛迭代数	仿真平台日志 + 实机记录	FPR ≥ 95%；收敛轮次记录并报告	工程经验建议值
F4 自主治理	ASCR；ToE	自闭环异常比例；升级时间中位数	治理事件库	ASCR ≥ 80%；ToE 中位数逐月下降	工程经验建议值
F2/F3（高阶扩展）	ZFS-FPR	零参考首见任务首轮仿真通过且实机首轮通过占比	流程生成 + 仿真 + 实机日志	ZFS-FPR ≥ 60%（建议值）	工程经验建议值

表 2. 完全准入判据 F1–F4 的形式化定义：指标、定义式、数据来源与建议验收阈值（建议值，待 PoC 校准）。

3.2 判定规则：灰度准入（graded admission）

在判定规则上，我们建议采用“灰度准入”而非“全有全无”：工厂可在某一产线上部分满足判据（如 F3 已达标而 F4 尚在积累），此时应被描述为“面向完全自治的进度状态”并在治理记录中明示缺口。这一设计意在既保留概念边界，又避免因单一指标未达而否定整体进展。

规则如下：对每条产线，逐判据计算指标并与阈值比较，得到“达标 / 部分达标 / 未达标”三态；产线状态按“最弱判据”原则分类——仅当 F1–F4 全部达标方可标注为“完全自治”；任一判据未达标即标注为“面向完全自治的进度状态（列出未达标判据）”；部分达标判据须在治理记录中写明缺口与改善计划。每次状态判定均产生一条治理记录，字段含：产线标识、各判据指标值、判定结果、缺口清单、纳入下季度改进目标与否。灰度准入的完整判定流程与治理记录要求将被附录 A 的 PoC 协议引用。

3.3 设备级到产线级的聚合

判据度量在设备级采集，而“完全自治”是对产线或工厂整体属性的断言，二者需要明确的聚合规则。产线级指标定义为：对构成该产线的全部设备，按各设备在产线任务中的作业时长占比为权重，加权聚合得到产线级 CoC、GSR、FPR、ASCR 等指标；HIR 与 ToE 按产线运行周期的合计口径计算。评估窗口为自然季度，自治等级的升级与降级行为（upgrade/downgrade，区别于故障上报 escalation）均以连

续两个季度满足或跌破对应阈值为触发条件并留痕。该聚合映射解决了“单设备达标 ≠ 产线达标”的粒度错位：一条产线只有在全部构成设备均达到基准阈值的条件下，其加权指标才有资格被解读为产线级“完全自治”进度。

4 能力验证与实证路线图

设备动作流程的自主生成能力直接关系到 AI 治理工厂能否被信赖为可靠运行。为实现从易到难、从固定到自由的渐进验证，以确定性的提升换取风险的可控，本章给出验证阶梯、一个可复现的完整案例，以及将全部实证承诺收敛于一项最高优先级端到端概念验证（PoC）的路线图。

4.1 渐进验证阶梯

低阶阶段（固定流程）中设备动作路径相对固定，流程由有限运动学组合构成，验证重点是能否稳定复现正确流程；中阶阶段（参数化流程）中动作路径随产品与工艺参数变化，验证重点是能否按参数正确匹配与调整流程；高阶阶段（自主编排流程）中需根据图纸、实物与实时状态理解模块能力与装配目标、自主选择零件、编排动作流程并协调多个执行单元协作完成装配，验证重点是能否自主规划并完成全新任务。只有前一阶段通过充分验证后才进入下一阶段，从而使“仿真训练效果在实机上也有良好表现”这一核心命题逐步被证明。设备级验证覆盖四类具有代表性的非标设备，难度逐级递增（表 3）。

例	设备	验证内容	难度要点
1	搬运设备	料仓升降、出仓进仓及简单取放流程	固定流程，动作组合少
2	贴合设备	取放流程，更高定位精度	精度要求提升，姿态控制更严
3	组装设备	取放基础上增加组装动作	动作类型增加，力位配合
4	点胶设备	组装基础上增加点胶工艺	精度进一步提升，轨迹连续性要求高

表 3. 设备级渐进验证阶梯（固定与参数化阶段用例）。

注：用例 1 对应固定流程档，设备动作路径相对固定；用例 2–4 对应参数化流程档，动作路径随产品与工艺参数变化。

自主编排阶段（高阶）的验证任务为电路板组装、零部件组装以及多执行单元（如机械臂/机器人）协作装配，作用对象是由电路、机构与固定工位三环节构成的产线；该阶段验证的是产线整体的流程编排能力，而非将电路、机构或固定工位任一环节绑定为“自主编排阶段”；其达成程度以第 3.1 节零参考首见任务规划成功率（ZFS-FPR）与 M3 里程碑完成度共同判定，从而为“自主编排”提供可度量的验收口径。

在部件级验证层面，电路板组装任务要求根据电路图纸选择线与电器元件并自行完成组装，零部件组装任务要求根据设备图纸选择零件并完成设备组装；此类任务要求 AI 不仅会复现流程，更会按图选件、按需组织流程，是通向完全自治的关键一跃。综合设备级与部件级验证，目标能力为：能够根据实物 3D 模型、图纸、位移参数或说明书，实现对被操作对象的自主操控并完成指定工作。据此，本文设定四个可度量里程碑：M1 完成搬运与贴合两类标准流程的自主生成与实机验证；M2 完成组装、点胶两类复杂度流程的自主生成与实机验证；M3 完成电路板组装、零部件组装两类自主编排任务的自主规划与实机验证；M4 将任一新设备从交付数据到完成实机首件试产的时间压缩至约定的时间阈值。其中 M1–M3 依次验收固定、参数化、自主编排三档流程能力，固定档并入 M1 与首个参数化用例并批验收；M4 为不依赖流程档位的端到端效率判据。M1–M4 全部达成的时间表与判据，构成研究议程的主干。此处须澄清：F1–F4 是界定“完全自治”边界的判据；M1–M3 是依次验收三档流程能力的议程里程碑，对应 F2/F3 边界的实证承载，固定档并入 M1 并批验收，表 3 用例 1 覆盖 M1 固定档验收、用例 2–4 覆盖 M1–M2 参数化档验收范围；M4 则是端到端交付效率的性能目标，不构成“完全自治”边界的组成项。两条序列分别回答“何为完全自治”与“研究议程完成度”，不可混读为同一判据集。

4.2 完整案例：非标搬运设备的全流程落地

以非标搬运设备为例，可完整演示从数据到上产线的工程流程，作为方法论的可复现范本，其各步骤同时也是 PoC 协议（附录 A）的数据采集点。

数据准备环节收集三类数据：3D 全场景扫描数据（设备整机与环境的静态三维模型）、全部图纸数据（总装图、模组图、零件图及所搬运产品的模型）与零部件位移参数（轴行程及正负限位、气缸上下限位、取放手爪行程）。数据经格式归一后进入制造数字基座，成为仿真与校验的统一数据源。模组校验环节在进入动作流程开发前先对数字孪生模型做一致性校验：模组从属关系校验确认各模组上的轴/执行器归属正确，识别“模组 A 的 Y 轴出现在模组 C 上”之类的结构性错误并报错；坐标与右手定则校验核对轴的正负方向定义，避免仿真与实机方向不一致；数据一致性校验核对扫描数据与图纸在尺寸、结构上的不一致。校验不通过项输出为异常清单，作为后续修正依据，避免错误数据污染流程生成。

动作流程自主生成环节由 AI 在自治仿真平台上完成：从低速到满速的渐进速度配置、部件间硬件干涉检查、模组间动作的时序避让规划，最终输出适配该设备的完整动作流程。实机调试阶梯环节采用“仿真先行、逐级加速”的六级流程，每一级均基于实时 3D 扫描结果：硬件核验（扫描当前设备，确认装配与仿真模型一致）；设备点检（核对各零部件位移方向与动作开合状态）；模组调试（获取当前设备实际位置坐标，进行带产品状态下动作开合限位调整）；仿真模拟（依据当前扫描与调试数据在仿真平台完整跑一遍动作流程，从低速到满速排查干涉与时序碰撞）；设备空跑（不带产品低速运行全部流程，正常后逐步提速至满速并观察干涉与碰撞）；带产品试跑（带产品执行全部流程并同样逐步提速，验证干涉、时序碰撞与

定位稳定性）。六级验证全部通过后，设备方可进入生产状态。对机械臂与机器人，流程可简化：去除“模组位移识别”环节，改为直接依据说明书与实时位姿确认操控方法，因而标准设备纳入完全自治体系的成本显著低于非标设备。

4.3 研究议程：PoC 与实证路线

本节将未来工作重构为一条可执行的实证路线图，并明确其“预期/待验证”身份。

首要验证里程碑：单一非标设备的端到端概念验证（PoC）。 作为最高优先级，我们计划在搬运设备上完成“数据准备—仿真闭环收敛—实机调试阶梯”的完整闭环，并采集四类指标：仿真收敛曲线（迭代轮次 vs 剩余碰撞数）、仿真→实机性能落差（FPR 与干涉概率的对比）、调试工时对比（AI 辅助 vs 传统人工编程基线）、人工审核通过率。统计设计上，对三类待证假设（H1 仿真收敛在有限轮次内达成；H2 实验组实机 FPR 显著高于对照组；H3 1× 扫描精度的误差包络在保证安全余量的同时不显著牺牲节拍）均按两比例检验处理：设显著性水平 $\alpha = 0.05$、功效 $1 - \beta = 0.8$、对照组首次通过率基线 0.5，则 H2 所设的效果量先验（实验组相对对照组 FPR 差距 ≥ 20 个百分点）需求每组约 25–30 个动作周期；据此我们在附录 A 中设定每类实验不少于 30 个动作周期，并明确该效果量为课题假设而非既有文献数值。完整的变量设计、对照组、样本与报告格式见附录 A。我们明确声明：H1–H3 均为待验证命题而非既有结论。

后续议程。 须强调：附录 A 的单设备 PoC 仅用于校准表 2 与表 4 的建议阈值、验证 F 判据在单设备上的可操作性，并检验 H1–H3；它不构成对产线级自治或自主编排阶段能力的任何证明。在 PoC 里程碑之后，验证议程与渐进验证阶梯逐级对齐，依次推进：单设备 PoC（本协议）→ 多环节—多设备产线级试点（电路、机构与固定工位三环节各至少一台设备接入，按第 3.3 节聚合规则采集产线级指标）→ 自主编排级验证（以零参考首见任务规划成功率 ZFS-FPR 与 M3 里程碑为验收口径，覆盖电路板组装、零部件组装与多执行单元协作）→ 跨产品知识迁移与多产线联动试验。横向议程包括：制造数字基座中声音、触感等扩展感知能力的开发与集成；逆向数据自动标注与 AI 模型持续微调机制的实现；在真实产线上开展渐进验证阶梯的试点，按第 3 章度量框架逐项采集判据指标；以及将表 2 的建议阈值在实际数据上校准并公开校准结果。

5 反证、安全与责任

本章先对仿真闭环的收敛性、终止性与保真度误差传导做分析性讨论，再给出安全标准对齐、失效分级与责任分配的框架。本章所涉命题均为需要被实验回答的开放性命题。

5.1 收敛性与终止性

以碰撞日志为反馈的迭代可视为对动作序列空间的局部搜索。碰撞反馈具有局部性：一次碰撞仅指示特定时刻特定部件对的冲突，据此修正时序或路径可在局部降低冲突数，但可能引入新的冲突。因此我们并不主张该迭代在任意问题上有全局收敛保证，而是将其界定为“以确定性退回策略保证的工程收敛”：每次修正后保留碰撞数评估最小的候选解，且当连续 K 轮（建议 $K \geq 5$）无改善时终止并退回最优历史解。“最优动作流程”的判定需要明确目标函数，我们建议以多目标加权验收：无碰撞（硬约束）、节拍达标、安全距离裕度、可读脚本满足人工侧检，由工程师对多目标权重作前置设定；终止条件为“连续 K 轮无改善 + 人工侧检通过”两者缺一不可，从而把“算法终止”与“工程放行”解耦。上述终止判据与收敛边界的建议，将在附录 A 的 PoC 协议中通过仿真收敛曲线（迭代轮次 vs 碰撞数）实证检验。

5.2 保真度误差传导与对冲

仿真到实机的性能落差源于三类保真度误差：动力学误差（物理引擎与真实摩擦、惯量、弹性不一致）、感知误差（3D 扫描重建引入的几何/法向偏差，进而影响位姿标定与碰撞检测）、控制/延迟误差（执行器响应与通信延迟）。我们给出三项对冲手段：物理引擎标定（以实机空跑数据反标定动力学参数）、误差包络预算（在仿真碰撞检测中预留扫描精度量级的间隙余量）、以及 4.2 节实机调试阶梯对残余误差的逐级吸收。

关于误差包络量级的选择依据：工业点云配准的公开基准显示，配准精度的典型量级为毫米级（常用迭代最近点（Iterative Closest Point, ICP）变体在真实扫描数据上的均方根误差（Root Mean Square Error, RMSE）多在 1–3 mm 区间），6-DoF 位姿估计在遮挡/稀疏场景下的误差同样以毫米与毫弧度量级衡量[10]；因此，在仿真碰撞检测中预留“1 倍扫描精度量级”的间隙余量，即可吸收由感知误差引入的大部分假阴性碰撞风险，同时不至于为每处干涉均引入过大的工艺间隙而牺牲节拍。我们将该量级的具体取值设计为 PoC 的一个自变量（误差包络 0 / 1× / 2× 扫描精度，见附录 A），以“仿真→实机性能落差”指标定量评估其对安全余量与节拍的影响。

5.3 安全标准对齐与分级边界

全能产线中的机器人作业应显式对齐现有安全标准体系：协作机器人的人机共存作业需要遵循 ISO 10218 系列与 ISO/TS 15066 关于力、速度与安全距离的限制[22]；功能安全框架遵循 IEC 61508 的安全完整性等级（Safety Integrity Level, SIL）分级思路，对自主决策链路的软件失效进行风险归级[23]；面向 AI 的管理体系可对接 ISO/IEC 42001，将 AI 风险纳入组织级管理体系[24]。在此基础上，我们建议实行分级安全边界：高风险动作（高动能、近人、关键机构）默认由人工审核放行并设硬限制，中低风险动作由 AI 自主执行但保留现场急停与安全 PLC 兜底。

5.4 失效分级、降级与人工接管

自主系统必须定义明确的失效模式、降级策略与可审计的接管条件，否则“自主”在安全审计中不可验证。我们将失效分为感知失效、执行失效、治理失效三类，并为每一类指定可审计的触发判据、降级路径、人工接管时限与日志留痕字段（表 4）。关键原则是：人工接管是最终兜底而非备用选项；任何降级与接管事件都必须可回溯、可复演。

失效类别	触发判据（监测信号）	降级路径	人工接管时限	日志留痕字段
感知失效	扫描/位姿更新异常（点云缺失率、配准 RMSE 超阈值）	模型拒绝生成，该工位降级为人工引导扫描或人工输入	首次触发后即时提示，5 分钟内未解决则强制上报	异常信号、触发时刻、降级动作、接管人、恢复时刻
执行失效	动作脱靶、超程限位触发、力/力矩越限、安全 PLC 停车	安全停车，生成治理事件并调用预案	停产后即时接管至控制台，按预案时限处置	偏差量、限位触发点、停车原因、预案编号、处理结论
治理失效	连续 N 次异常自闭环失败（建议 N ≥ 3）或 ToE 超限	上报至人工仲裁（escalation），AI 转入受限只读模式	ToE 超时即上报，上报后 30 分钟内由值班工程师接管	迭代记录、失败原因链、上报触发值、接管人、经验沉淀项

表 4. 失效分级与人工接管触发表。表中触发阈值（感知指标阈值、N、时限）为建议值，将在 PoC 中按设备类别校准。

5.5 责任归属与 AI 管理体系

当 AI 自主决策导致事故时，责任不应悬置。我们主张将“人—AI”责任分配显式化：设备与产线所有者为组织责任主体，AI 系统提供者承担可追溯性与技术缺陷责任，操作/审核工程师承担授权范围内的人工确认责任，并通过事件日志、决策留痕与审计接口支撑事后的责任认定。该框架与 ISO/IEC 42001 的 AI 风险管理闭环[24]保持一致。

5.6 劳动力与技能结构影响

全能产线会显著改变一线岗位结构：重复性程序编排岗位减少，而“流程审核、异常裁决、数据治理、AI 运营”类岗位增加。作为立场文，我们有责任明确提出：应将自动化收益的一部分投入再培训体系，使产线工人向流程审核员与 AI 运营员转型；这一议题虽超出本文技术范围，可纳入后续政策与技术协同研究的议程，并在第 7 章的局限中一并收束。需补充的是，这些新增岗位（流程审核、异常裁决、数

据治理与 AI 运营）的运行质量本身即是第 4 章 F4 判据（异常自闭环率 ASCR、代际人工升级时间 ToE）的审计对象，劳动力结构变化与自主治理边界互为表里。

6 讨论：与既有工作的对照

6.1 与代表性现有系统的逐环节对照

这里以一个目标相近、最近公开发表的代表性工作——面向工业多机任务规划与程序生成的 IMR-LLM[7]——为对照对象。选择依据为：其发布窗口与本稿最接近（ICRA 2026）、任务设定（设备定义输入 → 动作流程生成 → 验证 → 上实机）与本稿同构、且已公开可复现，在该范围内未见目标更相近的公开工作。在统一的任务设定（设备定义输入 → 动作流程生成 → 验证 → 上实机）下逐环节比较本文与它的差异。对照结果见表 5，“增强”指在既有思路上产品化/机制化，“新增”指既有框架中不存在、本文提出的主张。

对照环节	IMR-LLM[7]（代表性现有系统）	本文主张	增量定性
输入数据来源	文本指令驱动的多机任务定义	扫描几何 + 图纸结构 + 位移参数三源集成，含非标设备	增强
动作流程生成	大模型生成动作序列/程序	大模型生成 + 仿真碰撞反馈闭环迭代	增强
上实机前验证	以语法/任务级检查为主	以物理碰撞与干涉检测为必要条件的仿真验证	增强
产线级治理	未将治理纳入任务环	F1–F4 判据 + 灰度准入 + 失效分级接管（本文第 3、5.4 节）	新增

表 5. 与代表性现有系统的逐环节对照（以 IMR-LLM 为对照对象）。

该对照的落点有二。其一，既有方法多数面向单任务、单设备的动作策略，而本文主张的对象是产线级的能力涌现与流程级的自主治理，增量主要在设施层（基座集成）与治理层（F 判据驱动的代际升级），而非单点算法。其二，本文的增量是双层的：定义性链路上，为“完全自治”提供可判别边界——这一贡献不依赖 PoC 即可被评估；实现性链路上，具体能力与性能指标则以研究议程中的实证来兑现。两层边界在 1.2 节已明确，此处从对照角度再次确认。

对照亦需给出对称一侧。本文相对 IMR-LLM 类方法仍受制于以下短板：其一，多机调度与异构执行器编排的在线验证能力——对照方法已具备，而本文当前仅以非标搬运设备的单设备 PoC 承载，多执行单元协作（如机械臂装配）处于自主编排阶段的规划项；其二，对照方法以既有工业数据集或标准机型的任务设定可离线复现，而本文的验证阶梯依赖自建仿真环境，其保真度校准（5.2 节）本身即是研究

议程的一部分。本文的增量主要落在设施层（基座集成）与治理层（F 判据驱动的代际升级），而非对标既有方法在在线多机编排上已经具备的能力。

6.2 技术增量矩阵

为明确本文相对既有工作的完整增量，表 6 以“一致/增强/新增”三级标注本文组件与既有方向的关系。

既有方向	代表性工作与本组件	与本文关系	文本定位
sim-to-real 迁移	域随机化/域自适应[20-21]	增强	第 5.2 节保真度分析与参考文献
闭环监督学习 / 规划修复	轨迹引导的领域修复[8]	增强	第 2.2 节
LLM 任务规划 / 程序生成	IMR-LLM[7]、SayCan[17]、Code-as-Policies[18]、ProgPrompt[19]	增强	第 2.1、2.2、6.1 节
工业多智能体 / 人机协作规划	人机协作组装 LLM 多智能体[13]	增强	第 2.2 节
数字孪生工厂	工业数字孪生[3-4]	增强	第 2.2、1.2 节
可重构制造系统	RMS[14]	增强	第 1.2 节
工厂级 AI 治理 / AI 管理体系	ISO/IEC 42001[24]	一致	第 5 章
产线级能力涌现与流程自主治理	本文 MDB + F 判据驱动代际升级	新增	第 2–5 章

表 6. 与既有技术方向的技术增量矩阵（“一致”指立场参照，“增强”指在既有方法上产品化/机制化，“新增”指本文主张的产线级增量）。“新增”仅指本文主张的产线级增量，不隐含对对照方法既有能力的否定；对照方法的相对优势与本文的受制面见 6.1 节对称陈述。

7 总结

本文的核心论断是：全 AI 自治工厂不是“用上了 AI”的工厂，而是“AI 构成完整自治闭环”的工厂。为界定这一边界，本文确立了“全链闭环、自主生成、自主验证、自主治理”四项可由度量支撑的准入判据，并将其形式化定义、灰度准入规则与设备级到产线级的聚合映射落实为可直接判别的工程框架；围绕面向设备自供给与打样试产的全能产线，本文阐述了以制造数字基座为起点、以自治仿真平台为中枢、以渐进验证为阶梯、以电路、机构与固定工位三环节构成全能产线的实现路径，并以非标搬运设备给出可复现的实施流程。全文以证据边界表显式区分已构建的工程原

型与待验证的设计主张，以失效分级与人工接管的可审计框架回应安全关切。定义层的贡献——F 判据与“完全自治”的可判别边界——已完整给出，且不依赖本章之外的任何实证即可评估；实现层的贡献则收敛为一项最高优先级的端到端 PoC 及其后的实证路线，其成败将由附录 A 定义的指标以可复核的方式判定；该判定限于单设备上判据的可操作性与阈值校准，产线级与自主编排级能力由随后两阶段（产线级试点与编排级验证）承载。判据结构界定了“完全自治”是什么（不依赖实证即可评估），参数实例界定某一产线当前是否达标（依赖 PoC 校准）；二者解耦，阈值漂移不移动概念边界。下一步，我们将按该协议启动首台设备的概念验证，并将校准后的阈值与结果公之于众。

7.1 局限与风险

本文存在若干需坦诚声明的局限。其一，数据与模型质量依赖：仿真模型精准性受扫描质量与参数完整度制约，模型误差可能导致“仿真通过、实机异常”，需通过实机调试阶梯兜底并对误差影响做定量评估（第 5.2 节）。其二，大语言模型的可靠性：生成式模型存在逻辑幻觉与流程不完整风险，工业场景必须坚持“仿真验证前置 + 实机审核兜底 + 异常闭环治理”三道防线。其三，初始投入与人才门槛：制造数字基座建设需要跨视觉、机械、软件、AI 的复合工程能力，短期成本不菲，其成本收益比本身需要实证数据支撑（M4 效率目标属性能验收而非自治边界项，其达成与否不影响“完全自治”标签的判定）。其四，安全与责任边界：AI 自主治理重大异常时的责任归属、安全冗余与人工接管机制仍需与企业安全管理体系深度结合，本文第 5 章给出了框架但尚未经受真实事故检验。其五，作为立场文，本文当前的判据阈值、失效触发参数与 PoC 协议尚未在真实产线上被数据校准，这是最需要被读者审慎看待的部分，也是我们后续工作的第一优先级。其六，5.6 节所述的劳动力转型涉及政策与社会层面，需要跨学科协同研究方能落实其影响评估。此外，本文的讨论范围止于产线级的自治闭环：工厂级管理自治（如企业资源计划、跨工厂调度）属更广的数字化与组织级研究议程，不在本文讨论范围之内，本文亦不构成对其可实现性的判断。

附录 A 概念验证设计（PoC Protocol）

本协议为第 4.3 节首要验证里程碑的展开设计，供执行时直接套用。

目标。 在单一非标搬运设备上验证“数据准备—仿真闭环收敛—实机调试阶梯”的端到端可行性，采集四类指标：仿真收敛曲线、仿真→实机性能落差、调试工时对比、人工审核通过率，并折算第 3 章 F1–F4 判据指标。

对象与样本。 一台含取放功能的中型非标搬运设备，产品为单一规格工件；每类实验（仿真、空跑、带产品试跑）重复不少于 30 个动作周期，以给出可报告的成功率置信区间。样本量依据：$\alpha = 0.05$、$1 - \beta = 0.8$、对照组首次通过率基线 0.5 的两比例检验下，H2 效果量先验（FPR 差距 ≥ 20 个百分点）需每组约 25–30 例，取 30 例留取余量；该效果量系课题假设而非文献数值。

实验变量。 自变量为（a）是否启用仿真预演闭环（实验组 vs 对照组仅人工编程），（b）仿真保真度配置（物理引擎标定开/关、误差包络 0 / 1× / 2× 扫描精度）；因变量为碰撞残留数、FPR、干涉概率、首次合格周期、调试工时、人工审核通过率。

流程。 ① 采集扫描/图纸/位移参数并建模；② 运行仿真闭环记录收敛曲线（迭代轮次 vs 剩余碰撞数），按第 5.1 节终止判据停止；③ 实机六级调试阶梯逐级记录；④ 对照组以传统人工编程方式完成同等任务并记录工时；⑤ 数据按第 3 章度量框架折算 F1–F4 指标；⑥ 记录感知/执行/治理三类失效事件并落表 4 留痕字段。

报告格式。 提交结果时将包含：仿真收敛曲线图、仿真→实机指标落差表、实验/对照组调试工时对比、按表 2 折算的判据达成情况、误差包络敏感性分析，以及失效事件的接管与留痕统计。结果无论正负均如实报告。

预期假设（待证）。 H1：仿真收敛在有限轮次内达成；H2：实验组实机 FPR 显著高于对照组（效果量先验 ≥ 20 个百分点）；H3：1× 扫描精度的误差包络在保证安全余量的同时不显著牺牲节拍。

阈值校准。 表 2 建议阈值与表 4 触发阈值均在本协议完成后，基于实际数据给出校准值并同步公开校准报告。

里程碑安排。 为便于评估本协议的可执行性，给出阶段—周期—产出的顺序安排（周期为预计值，按设备实际状态滚动更新）：数据准备与建模（约 1 个月，产出设备建模与校验清单）；仿真闭环收敛试验（约 2 个月，产出收敛曲线与终止判据验证结果）；实机六级调试与对照试验（约 2 个月，产出实机指标落差、调试工时对比与 F1–F4 折算表）；判据阈值校准与公开报告（约 1 个月，产出校准阈值与全套复算报告）。

附录 B 实现参数与数据生产说明

扫描精度指标。 建议记录点云分辨率、配准误差（RMSE）、法向偏差与球度误差作为重建质量基线；位姿标定采用“图纸基准 + 现场特征点对”的粗配准加 ICP 精配准流程。误差包络的 1× 基准即取该重建质量基线估算的扫描精度量级。

坐标系约定。 统一右手定则，轴正方向与图纸定义一致，模组局部坐标系归属全局坐标树；违反右手定则的模组在模组校验阶段即应报错拒绝进入仿真。

脚本平台要点。 中文自然语言脚本由“运动控制函数库 + 可读脚本”双层构成，脚本作为中间表示映射至底层控制卡指令；建议在脚本层预留工艺参数、时序避让与安全约束的扩展字段。平台现有版本为软件著作权登记 V1.1。

仿真器选型与模型策略。 仿真器应支持碰撞检测日志导出与速度/时序参数化；模型建议采用推理与反思的两阶段提示策略，并将碰撞日志结构化后注入提示上下文。

逆向数据生产。 动态 3D 扫描获取“实机动作→动作流程”标注数据时，需定义标注精度要求（关键位姿采样率、时间戳对齐误差上限），并处理异常样本稀疏问题——建议以故障注入或人工复演方式增广罕见异常样本，标注需经工程师抽检复核。逆向数据集计划在完成脱敏与协议审查后选择性公开。

附录 C 术语表

缩写/术语	全称/定义	备注
AI-AGF	人工智能自主治理工厂（Artificial Intelligence Autonomous Governance Factory）	先前研究提出的概念，见文献 [25]
全 AI 自治工厂	完全人工智能自主治理工厂（Full AI Autonomous Governance Factory, Full-AGF）	满足 F1–F4 边界者；本文研究对象
UPL	全能产线（Universal Production Line）	构成层：电路、机构与固定工位三环节（three segments）合一；注意“环节”为同字两义：三环节（three segments）与全链闭环六环节（six links）分属不同坐标系，英译须区分；部署层：建厂期面向设备自供给与打样试产
MDB	制造数字基座（Manufacturing Digital Base）	三类基础/支撑软件的集成体；以“转译”（transduction）区别于通用数据平台
F1–F4	全链闭环 / 自主生成 / 自主验证 / 自主治理	完全准入判据，见表 2
HIR / GSR / FPR / ASCR / ToE / CoC	人工介入率 / 生成成功率 / 首次通过率 / 异常自闭环率 / 代际人工升级时间 / AI 环节覆盖率	判据度量指标，见表 2
HIR_r / HIR_c	复核放行介入率 / 纠错接管介入率	HIR 的两个子指标，见 3.1 节
灰度准入	graded admission	按判据达标程度分级标注产线自治状态，见 3.2 节
ZFS-FPR	零参考首见任务规划成功率（Zero-Reference First-Success Rate）	F2/F3 在自主编排阶段的高阶扩展指标，见 3.1 节
证据边界	evidence-boundary declaration	以表 1 显式区分“已实现/预期”的

缩写/术语	全称/定义	备注
		声明机制，见 2.3 节
RMS / DT	可重构制造系统 / 数字孪生	与本文概念的关系见 1.2 节

参考文献

[1] 周济，李培根，等。走向新一代智能制造（Toward a New Generation of Intelligent Manufacturing）。Engineering（中国工程院院刊），2018, 4(1): 11-20. DOI: 10.1016/j.eng.2017.12.006.

[2] ISO/IEC 22989:2022. Information technology — Artificial intelligence — Artificial intelligence concepts and terminology. Geneva: ISO/IEC, 2022.

[3] Tao F, Zhang H, Zhang C. “Advancements and challenges of digital twins in industry”. Nature Computational Science, 2024, 4(3): 169-177.

[4] Tao F, Zhang H, Liu A, et al. “Make more digital twins”. Nature, 2019, 573(7775): 490-491.

[5] Leng J, et al. “Unlocking the power of industrial AI towards Industry 5.0”. Journal of Manufacturing Systems, 2024, 73: 349-363.

[6] Fan H, Liu X, Fuh J Y H, et al. “Embodied intelligence in manufacturing: leveraging large language models for autonomous industrial robotics”. Journal of Intelligent Manufacturing, 2025, 36(2): 1141-1157.

[7] Su X, Xu J, van Kaick O, et al. “IMR-LLM: Industrial multi-robot task planning and program generation using large language models”. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2026), Vienna, 2026. [in press]

[8] Liu R, et al. “Bridging the semantic gap: trajectory-guided domain repair for reliable planning”. Robotics and Computer-Integrated Manufacturing, 2026. DOI: 10.1016/j.rcim.2026.103290. [in press]

[9] Vemprala S, Bonatti R, Bucker A, et al. “ChatGPT for robotics: design principles and model abilities”. IEEE Access, 2024, 12: 55682-55696.

[10] Zhuang C, Wang H, Ding H. “AttentionVote: a coarse-to-fine voting network of anchor-free 6D pose estimation on point cloud for robotic bin-picking application”. Robotics and Computer-Integrated Manufacturing, 2024, 86: 102671.

[11] 郑湃，李成熙，殷悦，等。增强现实辅助的互认知人机安全交互系统（Augmented-reality-assisted mutual-cognitive human-robot safety interaction system）。机械工程学报，2023, 59(6): 173-184. DOI: 10.3901/JME.2023.06.173.

[12] 陶家琦，等。制造领域知识图谱的应用研究现状与前沿（Application research status and frontiers of manufacturing-domain knowledge graphs）。计算机集成制造系统，2022, 28(12): 3720-3736.

[13] Wang B, Zheng L, Wang Y, et al. “LLM-based multi-agent task planning for human-robot collaborative assembly balancing operator experience and efficiency”. Journal of Manufacturing Systems, 2025, 82: 1020-1045.

[14] Koren Y, Jovane F, Heisel U, et al. “Reconfigurable manufacturing systems”. CIRP Annals — Manufacturing Technology, 1999, 48(2).

[15] 乔非，等。工业 5.0 环境下面向生产调度的人本融合技术（Human-centric fusion technologies for production scheduling in Industry 5.0）。机械工程学报，2025, 61(15): 40-56.

[16] 张党，等。数据-知识混合驱动的离散制造系统智能控制体系构架（A data-knowledge hybrid-driven intelligent control architecture for discrete manufacturing systems）。机械工程学报，2024, 60(6): 1-10.

[17] Ahn M, Brohan A, Brown N, et al. “Do As I Can, Not As I Say: Grounding language in robotic affordances”. In: Proceedings of the 5th Conference on Robot Learning (CoRL 2021). arXiv:2204.01691, 2022.

[18] Liang J, Huang W, Xia F, et al. “Code as Policies: Language model programs for embodied control”. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2023), London, 2023.

[19] Huang W, Abbeel P, Pathak D, Mordatch I. “Language models as zero-shot planners: Extracting actionable knowledge for embodied agents”. In: Proceedings of the 39th International Conference on Machine Learning (ICML 2022). PMLR 162: 9118-9147, 2022.

[20] Tobin J, Fong R, Ray A, et al. “Domain randomization for transferring deep neural networks from simulation to the real world”. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2017), Vancouver, 2017: 23-30.

[21] Zhao W, Queralta J P, Westerlund T. “Sim-to-real transfer in deep reinforcement learning for robotics: a survey”. In: Proceedings of the IEEE Symposium Series on Computational Intelligence (SSCI 2020), Canberra, 2020: 737-744.

[22] ISO 10218:2011, ISO/TS 15066:2016. Robots and robotic devices — Safety requirements for industrial robots / Collaborative robots. Geneva: ISO.

[23] IEC 61508:2010. Functional safety of electrical/electronic/programmable electronic safety-related systems. Geneva: IEC.

[24] ISO/IEC 42001:2023. Information technology — Artificial intelligence — Management system. Geneva: ISO/IEC.

[25] Wei J. AI Autonomous Governance Factory: Concept, Architecture, and Launch Pathway. engrXiv Preprint, 2026. DOI: 10.31224/7440.

[26] SAE International. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles (SAE J3016). SAE Standard, 2021.