# 人工智能自主治理工厂：概念、架构与启动路径

【韦季李】

【广州季李科技工作室】

【ORCID：0009-0006-5696-1873】

【通讯作者邮箱：jiligzs@qq.com】

## 摘要

制造业正经历从自动化向自主化的范式跃迁。当前“黑灯工厂”“无人工厂”等概念未能系统回答“工厂的自主能力从何而来、由谁治理”这一根本问题。本文提出“人工智能自主治理工厂”（AI Autonomous Governance Factory，AI-AGF）概念，明确其核心在于 AI 系统在无人干预下自主完成感知、决策与执行全过程，并对自身的目标、规则、边界和风险实施系统性治理。本文围绕必然性论证、双轨架构设计与四阶段启动路径三个核心命题展开。“双轨架构”由全能产线（Universal Production Line，UPL）与标准产线（Standard Production Line，SPL）构成：UPL 承担设备自供给与产品打样，SPL 承载大批量高效生产，二者由 AI 治理层统一调度形成“设备生产设备”的物理闭环。“四维治理框架”从目标、规则、边界和风险四个维度对 AI 决策实施系统性约束，将制造治理从资产中心化推进至治理中心化，并与 IEC 62264、RAMI 4.0、ISO 55000 三项工业标准形成结构性对话。不同于既有研究偏重技术维度，本文的独特贡献在于将“治理”提升为与“自主”同等重要的分析维度：AI 系统不仅要自主决策，更需对这一决策逻辑实施系统性的自我约束与风险管控。基于上述分析，本文给出 AI-AGF 的四项操作性判据（C1-C4）、一个双轨架构蓝图、一套四维治理框架及一条四阶段量化渐进路线。本文属前瞻性评述，论证效力来自逻辑自洽性、文献支撑度与竞争性假说的系统回应。

**关键词**：人工智能自主治理工厂；智能制造；可重构制造系统；自然语言脚本运动控制；自主决策

**中图分类号**：TH166

# AI Autonomous Governance Factory: Concept, Architecture, and Launch Pathway

【Wei Jili】

【Guangzhou Jili Technology Studio】

【ORCID: 0009-0006-5696-1873】

【Corresponding Author Email: jiligzs@qq.com】

## Abstract

Manufacturing is undergoing a paradigm shift from automation to autonomization. Existing concepts such as “lights-out factories” and “unmanned factories” have yet to systematically address the fundamental question: where does a factory’s autonomous capability originate, and who governs it? This paper proposes the concept of the AI Autonomous Governance Factory (AI-AGF), defining its core as an AI system that autonomously completes the full cycle of perception, decision-making, and execution without human intervention, while systematically governing its own objectives, rules, boundaries, and risks. The paper addresses three core propositions: the inevitability argument, the dual-track architecture design, and the four-stage launch pathway. The dual-track architecture consists of the Universal Production Line (UPL) and the Standard Production Line (SPL): the UPL undertakes equipment self-supply and product prototyping, while the SPL handles high-volume production. Both tracks are coordinated by an AI governance layer, forming a physical closed loop of “equipment producing equipment.” The four-dimensional governance framework imposes systematic

constraints on AI decision-making across four dimensions—objectives, rules, boundaries, and risks—advancing manufacturing governance from asset-centric to governance-centric, and engaging in structural dialogue with three industrial standards: IEC 62264, RAMI 4.0, and ISO 55000. Unlike existing research that prioritizes the technical dimension, this paper's distinctive contribution lies in elevating "governance" to an analytical dimension of equal importance to "autonomy": an AI system must not only make autonomous decisions but also exercise systematic self-constraint and risk control over its own decision logic. Based on the above analysis, this paper presents four operational criteria for AI-AGF (C1–C4), a dual-track architecture blueprint, a four-dimensional governance framework, and a four-stage quantified progressive roadmap. This paper is a prospective review, with its argumentative force deriving from logical coherence, literature support, and systematic engagement with competing hypotheses.

**Keywords**: AI Autonomous Governance Factory; intelligent manufacturing; reconfigurable manufacturing systems; natural language script-based motion control; autonomous decision-making

---

# 0 引言

## 0.1 研究背景与问题提出

制造业正经历从自动化向自主化的范式跃迁。周济等[1]将智能制造划分为数字化制造、数字化网络化制造和数字化网络化智能化制造三个阶段，指出新一代智能制造的核心是 AI 与制造技术深度融合，但制造系统距离真正的自主化运行仍有显著差距。当前“黑灯工厂”“无人工厂”等术语仅描述了工厂的外部表征（无需照明、无需人员），未能揭示内在运行机制——是什么驱动工厂运转？决策从何而来？系统如何应对不确定性？在此背景下，本文提出**人工智能自主治理工厂（AI Autonomous Governance Factory，AI-AGF）**概念，包含“自主”与“治理”两个互构维度（详细界定见§1.1），围绕三个核心命题展开：(1) 论证 AI 自治工厂的必然性；(2) 设计“全能产线+标准产线”的双轨架构；(3) 提出以自然语言脚本运动控制平台（以中文为当前实现语种：中文脚本运动控制平台）为媒介的四阶段启动路径。

## 0.2 概念的提出与界定

AI-AGF 将“自主”与“治理”统一为制造系统研究的两个不可分割的维度：AI 系统不仅要自主决策，更需对自身的目标、规则、边界和风险实施系统性治理。四维治理框架可视为 AI 对齐问题在制造场景下的一个工程化回应（详细界定与对比分析见 § 1.1）。制造领域因此不仅是一个 AI 治理技术的应用场景，更可能成为验证和迭代 AI 治理理论的天然“沙箱”。

本文对 AI-AGF 的讨论聚焦于其核心生产系统（包括产线架构、工序流程与设备自供给能力），暂未涉及工厂的管理系统、仓储物流系统、动力与能源系统等支撑性子系统。将 AI 自主治理扩展至工厂全系统的集成运行，是下一步研究的重要方向。本文聚焦核心生产系统的理由在于——生产系统是工厂自主化的“最小可行内核”：一旦 AI 能自主治理生产系统，向管理系统、物流系统、能源系统的扩展遵循相同的治理逻辑（目标→规则→边界→风险），属工程外推而非概念跳跃。

基于上述界定，AI-AGF 的操作性定义（operational definition）如下。

**C1（自主决策）**：AI 系统在无人工干预下自主完成从感知、决策到执行的完整闭环，覆盖产线调度、工艺优化与异常处理。

**C2（自主治理）**：AI 系统对自身决策的目标、规则、边界和风险实施系统性治理，具备自我约束与纠偏能力。

**C3（双轨架构）**：工厂采用全能产线（Universal Production Line, UPL）与标准产线（Standard Production Line, SPL）双轨架构，UPL 具备制造非标设备的能力。

**C4（渐进路径）**：AI 自主治理能力通过四阶段渐进路径逐步建立，每阶段设可验证量化里程碑。

缺少 C1 或 C2 则不满足 AI 自治工厂的本质特征，退化为既有制造范式（如缺少 C2 退化为自主制造系统）；C3 和 C4 是本文对 AI 自治工厂工程实现路径的具体选择，而非概念定义的必要条件[2]——可能存在其他架构方案（如非双轨式设备自供给）或启动路径（如非四阶段式渐进）同样满足 C1-C2 所界定的“AI 自主治理”核心内涵。

### 0.3 研究意义与方法

本文采用概念分析与系统综合方法：概念分析用于从既有制造范式中提取核心差异维度，系统综合用于构建“必然性→架构→路径”的三环论证体系。本文属前瞻性评述，不涉及实验验证——论证效力来自逻辑自洽性、文献支撑度与竞争性假说的系统回应。

本文的主要贡献包括：(1) 概念层：提出 AI-AGF 概念，将自主与治理统一为制造系统研究不可分割的维度，填补既有范式的治理缺环；(2) 架构层：设计 UPL/SPL 双轨架构与四维治理框架，将制造治理从资产中心化推进至治理中心化，与 IEC 62264、RAMI 4.0、ISO 55000 形成结构性对话；(3) 路径层：以自然语言脚本运动控制平台为媒介，设计四阶段渐进路线，每阶段设可验证量化里程碑，在五重困境框架中将 AI-AGF 置入产业现实语境。

本研究的三个技术来源构成了上述贡献的实践基础。其一，本文所依托的自然语言脚本运动控制平台（以中文为当前实现语种；中文脚本运动控制平台）已于 2022 年获国家计算机软件著作权登记，其研发初衷在于降低上位机控制软件的开发复杂度，为非专业编程人员提供便捷的脚本化控制方案，从而提升自动化系统的部署效率与可维护性。该平台作为人机操作逻辑的显式化工具，在本研究的四阶段启动路径中充当了从"人工主导"向"AI 自治"渐进迁移的初始接口与过渡载体。其二，第一作者长期从事上位机软件开发工作，主要职责涵盖非标自动化设备的动作流程设计、视觉对位算法集成，以及设备在用户现场的首次调试与交付验收，多年的现场实践经验为本文的理论建构提供了扎实的工程验证基础。其三，基于对当前人工智能技术演进趋势的系统性认识，并结合自动化工程领域的长期积累，第一作者逐步凝练出"人工智能自主治理工厂"这一前瞻性命题，并进一步推导出"全能产线与标准产线的双轨设计"模型；在此过程中形成的核心判断是——从"人工主导"过渡至"AI 自治"并非跳跃式变革，而须以既有工程经验与操作规则为中介实现渐进式迁移。

---

## 1 文献综述与理论缺口

本文的概念体系建立于三条独立演进脉络的交汇处。以下分别从智能制造范式演进、AI 治理与安全理论、可重构制造系统三条脉络展开梳理，识别各脉络在当前阶段的治理缺环，为 AI-AGF 的概念建构提供文献基础。

### 1.1 智能制造范式演进：从云制造到 AI-AGF

智能制造的概念演进可追溯至 20 世纪 80 年代[3]。李伯虎等[4]提出的云制造范式将制造资源虚拟化为按需调用的网络化服务，奠定了“制造即服务”的理念基础——这一理念对本文 AI-AGF 的“双轨联动”思想有直接影响：当制造能力可被虚拟化调用时，UPL 的设备制造能力即可按 SPL 的实时需求动态调配。周济等[1]将智能制造划分为数字化制造、数字化网络化制造和数字化网络化智能化制造三个演进阶段，指出新一代智能制造的核心是 AI 与制造技术的深度融合，但该框架尚未触及自主治理维度。

工业 5.0 将范式关注转向人的价值、可持续性与韧性[5]。姚锡凡等[5]对以人为本的智能制造演进做了迄今最系统的梳理，其框架以“人”为治理中心；蒋周明矩等[6]的人机协作增材制造框架体现了工业 5.0“人在回路上”的设计原则。然而，这些以人为中心的范式均需面对一个结构性上限：当制造系统的复杂度超越人类认知负荷后，人对决策回路的参与本身即构成瓶颈。AI-AGF 正是在这一意义上提出了范式转换的可能性。具体而言，人的角色从“回路上的操作节点”（on-the-loop operator）升维为“回路之上的战略锚点”（above-the-loop strategic anchor）——即在保留人对系统终极控制权的前提下，微观决策回路让渡于 AI，人类退居回路之上承担战略监督，在人类设定的价值边界内由 AI 自主完成微观决策与治理。

表 1 从决策主体、治理机制、设备自供给和人机关系四个维度对比 AI-AGF 与五种既有制造范式。已有比较研究的分类框架[7-8]覆盖了自动化执行到数据驱动优化的完整光谱，上述概念分别触及了“自主”的某个侧面——自动化替代、虚实映射、认知推理或分布式协商——却均未将“治理”作为与“自主”同等重要的设计维度。

| 概念 | 决策主体 | 治理机制 | 设备自供给 | 人机关系 |
| --- | --- | --- | --- | --- |
| **AI-AGF** | AI 系统自主决策并自主治理 | AI 对自身目标、规则、边界和风险实施系统性治理 | 全能产线可制造工厂所需的非标设备，实现设备自供给 | 从人机协作渐进至 AI 完全自主治理 |

| 概念 | 决策主体 | 治理机制 | 设备自供给 | 人机关系 |
| --- | --- | --- | --- | --- |
| 黑灯工厂 | 人类预设规则，系统自动执行 | 无自主治理，依赖外部人工干预 | 设备由外部采购和集成 | 机器替代体力劳动，人负责异常处理 |
| 无人工厂 | 人类远程监控，系统自动执行 | 无自主治理，异常时需人工介入 | 设备由外部采购和集成 | 人完全撤离生产现场，远程管理 |
| 数字孪生工厂 | 人类决策为主，孪生模型辅助仿真验证 | 无自主治理，数字孪生仅提供信息支撑 | 设备由外部采购和集成 | 人通过数字孪生界面监控和决策 |
| 认知工厂 | 知识图谱驱动推理，人机协同决策 | 知识规则约束推理边界，无自我修正机制 | 设备由外部采购和集成 | 人通过知识界面与系统协同推理 |
| 自主制造系统 | 多 Agent 自主协商，分布式决策 | Agent 间分布式协商实现局部治理，无全局统一治理框架 | 部分研究涉及模块化自重构 | 人设定目标，多 Agent 自主协商执行 |

综上，AI-AGF 的质变不在于技术参数的提升，而在于决策权归属与治理架构的根本转移——将决策权从人类移交至 AI 系统并赋予自主治理能力。El Kalach[7]将这一演进明确定义为“认知制造”，Leng 等[8]系统梳理了工业 AI 向自主化演进的路径与挑战。上述大模型相关验证均在实验室环境中完成，排除了光照突变、来料异常等生产常态扰动——真实工厂中的推理可靠性仍待验证，因此需要§4.2“人机共治”阶段作为可靠性过渡期。

### 1.2 AI 治理与安全理论：制造场景的特殊性

AI-AGF 将“治理”作为与“自主”同等重要的分析维度，这一立场与 AI 安全研究存在深层对话关系。

Russell[9]系统论证了 AI 控制问题的根本性——确保 AI 行为始终与人类意图对齐。Amodei 等[10]将这一问题具体化为避免负面影响、奖励函数篡改、可扩展监督、安全探索和分布偏移鲁棒性五类具体问题；Leike 等[11]与 Hendrycks 等[12]分别从奖励建模和机器学习安全角度推进了可扩展对齐与鲁棒性研究。

上述研究以通用 AI 为分析对象，讨论开放环境中的行为约束。制造场景为 AI 治理提供了独特的工程化切入点：目标函数可量化（良率、OEE、能耗等 KPI）、环境约束可物理化（安全围栏、急停按钮、工艺参数边界）、决策后果可快速验证（产品实物即最终仲裁者）。AI-AGF 的四维治理框架因此可视为 AI 对齐问题在制造场景下的工程化回应：目标治理界定“对齐什么”，规则治理界定“不对齐的边界”，边界治理界定“谁有权在何时纠偏”，风险治理界定“对齐失败后如何降级”。

此外，制造场景引入了一种 AI 安全文献中讨论不足的风险类型——物理后果风险。AI 决策失误在纯信息领域（如推荐系统）的后果限于经济损失，而在制造场景中可能导致设备损毁、人员伤亡或停产损失。这一物理后果的严重性使得制造领域的 AI 治理不能仅依赖软件层面的安全机制，而必须构建包含物理安全层的多层防御体系[13][14]。这正是本文四维治理框架将规则治理设计为独立于推理引擎的“硬约束层”、并设置风险治理的三级响应机制（黄/橙/红）的根本原因：规则治理作为物理安全层的数字化表达，确保 AI 治理的底层安全逻辑不因上层推理能力的演进迭代而失效。

### 1.3 可重构制造系统：从模块化到 AI 驱动自重构

可重构制造系统（RMS）是 AI-AGF 双轨架构的理论基础。Koren 等提出 RMS 的核心原则——模块化、可集成性、可转换性、可诊断性和定制化——奠定了从“专用刚性”向“可调柔性”转型的基础。ElMaraghy 等将制造系统演进概括为专用制造系统（DMS）→柔性制造系统（FMS）→可重构制造系统（RMS）三阶段递进。Todescato 等[15]提出了基于多 Agent 的可重构框架，Mayer 等[16]的综述表明数字孪生与 AI 协同可将产线重构周期从数周压缩至数天。数字孪生的共享数据模式可实现制造资源与工艺过程变更时的快速适应性重构，为 AI-AGF 的“设备自供给+产线自重构”闭环提供了数字孪生层面的方法论基础。

然而 OPC UA over TSN 等接口标准尚未统一，规模化部署远慢于技术预期——因此需要§3.1 所述设备自供给闭环降低对外部标准化进程的依赖。同时，数字孪生模型与

物理实体间的实时保真同步仍是瓶颈：人类工程师可凭经验识别模型漂移并暂停执行，而 AI 自主决策场景下模型漂移可能直接导致错误决策的级联传播——因此需要§3.2 所述四维治理框架中的风险治理层作为漂移检测与降级运行的安全网。

AI-AGF 则将重构的决策权从人类工程师转移至 AI 治理层。当 AI 从 SPL 数据中感知设备性能退化趋势时，它可以在退化发展至故障之前主动触发 UPL 预生产替换件并规划 SPL 的模块级分步重构，实现从“被动维修”到“预测性自主重构”的跃迁。这正是下文§3.5 所述四种协同能力中“跨产线全局最优调度”与“设计反馈闭环”的理论基础。

上述三条脉络共同指向同一个结论：AI-AGF 不是凭空构想，而是智能制造演进、AI 治理理论与可重构制造系统三条独立发展脉络的自然交汇点。

---

## 2 AI-AGF 的产业必然性

### 2.1 产业驱动力：成本、柔性与韧性

在§1 所述技术演进趋势之外，产业层面的深层动力同样不可忽视。AI-AGF 实现小时级产线重构与秒级重调度，使产线在异常情况下自主寻找替代方案，将生产中断降至最低。

以下逐一回应三种竞争性假说。

**第一种：MES/ERP 架构叠加 AI 即可实现自主化**。该假说认为，在现有 MES/ERP 上叠加 AI 模块即可实现自主化。本文回应：这一方案忽视了 AI-AGF 的根本特征——“设备生产设备”。叠加 AI 的 MES/ERP 仍运行于外部供应商构建的设备环境中，无法实现 UPL 自主制造非标设备并交付 SPL 的物理闭环。当设备本身成为生产对象时，生产系统边界已从“给定设备下优化排程”扩展为“自主构造制造资源”——这是分层预设的 MES/ERP 从未被设计的决策范式。

**第二种：AI 自主系统可靠性不足以支撑工厂全自主运行**。该假说认为，当前 AI 系统的可靠性尚未达到工业级安全要求，AI-AGF 在可预见的未来内无法落地。本文回应：四阶段渐进路径已内置安全阀——第二阶段工程师逐项审核 AI 产线方案的“人机共治”机制、第三阶段连续六个月零干预的验证门槛与工厂受控环境的物理安全层三重对冲

了这一风险。与自动驾驶面临的开放道路长尾场景不同，工厂在物理空间、工艺边界和异常类型三个维度高度受控，“长尾”长度远小于开放道路。因此，AI-AGF 可望在封闭工厂内先于自动驾驶实现可靠的自主运行。

**第三种：人机协作工厂已满足需求，无需完全转移决策权**。该假说认为，将“人在回路上”（human-in-the-loop）作为制造系统的设计原则，利用 AI 增强而非替代人类决策，已可实现柔性与效率的充分提升——工业 5.0[8]正是这一理念的代表性表达。本文回应：AI-AGF 的目标并非否定“人在回路上”，而是在人类设定的目标框架内将微观决策权渐进转移至 AI，使人类从“回路上的操作节点”升维为“回路上方的战略锚点”，将人的注意力回归于只有人才能胜任的价值判断与伦理监督。

**工业 5.0 以人为中心批评的价值论回应**。AI 自治工厂设想的“人的退出”仅限于微观操作决策回路——依赖重复性判断和瞬时响应的任务恰是人类认知的弱项。在战略目标设定、伦理边界守护和创造性产品定义等宏观层面，人的角色因从操作细节中解放而获得更充分的价值施展空间。AI 自治工厂不是对工业 5.0 的否定，而是对它的工程化承接——当制造系统的复杂度使“人在回路上”从保障变为瓶颈时，将人提升至回路上方的战略位置，恰是对“以人为中心”的最彻底贯彻。

以上从成本、柔性与韧性三个维度论证了 AI-AGF 的产业驱动力，并通过三种竞争性假说的系统回应（含工业 5.0 价值论层面的专门辩驳）加固了论证结构的完备性。然而，仅从“驱动力”一端论证必然性尚不充分——还需反向审视维持现状的代价。

在评估 AI-AGF 必要性的同时须反向审视维持现状的深层风险。当前中国工厂面临制造业订单的断崖式下滑、成本利润压缩、供应链脆弱、劳动力锐减、新旧市场领域 K 型背离（即不同赛道/企业呈现截然相反的走势分叉）五重叠加困局（详细分析与逐维回应见§5.3）。

综合上述产业驱动力与结构性困境的双向论证，本文的核心判断是：**人工智能自主治理工厂在技术演进与产业压力双重驱动下具备高度现实可能性**——关键问题不在于“是否”，而在于“何时、以何种路径”。

## 3 AI-AGF 的产线架构：全能产线与标准产线的双轨设计

### 3.1 架构总览与设计哲学

AI-AGF 将 UPL 的设备自供给能力与 SPL 的大批量生产能力整合在同一工厂系统内，由 AI 统一规划调度，实现了“设备自供给”与“产品生产”的闭环。

将双轨整合至同一 AI 治理框架产生四种协同能力（system-level synergies）。**设备设计—生产反馈闭环**：SPL 性能数据回流优化下一代非标设备设计。**跨产线全局最优调度**：AI 基于 SPL 订单压力动态调配 UPL 资源。**知识跨产品迁移**：UPL 多产品打样知识经 AI 抽象形成可迁移知识库，产线搭建周期随积累递减。**对设备供应链的免疫能力**：设备自供给闭环摆脱外部供应商单一依赖。四种能力的协同效应构成 AI-AGF 区别于“通用车间+专用产线”传统布局的根本差异——后者由人类管理者协调，无跨产线自主优化闭环。

### 3.2 AI 治理机制：四维闭环框架

AI-AGF 的根本特征不仅在于自主决策，更在于对自身决策实施系统性治理。本文所述“治理”包含三重递进语义：(i) **操作性治理**——AI 对自身决策循环的实时约束与纠偏，即“自主治理”的核心内涵；(ii) **架构性治理**——AI 对双轨产线的全局调度与资源优化；(iii) **制度性治理**——AI-AGF 框架与既有工业治理标准（IEC 62264、RAMI 4.0、ISO 55000）的结构性对齐。三层治理共用同一框架但作用于不同抽象层级。以下详述四维闭环框架。

**目标治理**：AI 持续监控 KPI 集合（良率、节拍、能耗、OEE 等），当指标偏离目标阈值时自主触发根因分析并生成纠偏策略，数字孪生为纠偏策略提供验证环境。

**规则治理**：工厂的安全标准、工艺规范、合规要求被编码为机器可读规则库。AI 生成任何决策方案时，规则引擎自动进行合规检查，拦截违反安全边界或工艺约束的方案。规则定义为(触发条件，约束类型，约束对象，优先级)四元组，优先级排序为安全 > 质量 > 效率 > 成本。规则库作为“硬约束层”独立于推理引擎，确保治理内核在 AI 推理能力演进中保持稳定。 例如，回流焊温区的规则实例为(加热区温度实测值 > 安全阈值上限，禁止继续升温，温控模块，安全级)，该规则优先于任何针对节拍或能耗的优化指令，且其优先级不可被 AI 推理引擎覆盖。

**边界治理**：AI 决策权限分层授予——日常运营调度自主执行，设备设计变更经数字孪生验证后通过，安全停机等重大决策须触发人类审核节点。操作审计日志完整记录每项决策及依据。

**风险治理**：AI 持续评估生产系统风险状态（设备健康度、物料供应稳定性、质量波动趋势），当风险指数超阈值时自主制定降级运行方案。

风险指数的量化可采用**多维度加权综合评估法**，综合设备健康分、物料库存覆盖率和质量波动率三维加权计算。当全产线风险指数超过预设阈值时，AI 治理层触发三级响应：(i) 黄色预警——提升采样频率与仿真频次；(ii) 橙色降速——降至 70%节拍+UPL 预生产替换件；(iii) 红色停机——部分工位停机，机器人转移已投料工件至缓冲区。阈值标定可参照 ISO 13374（机器状态监测与诊断）和 ISO 55000（资产管理）系列标准，并结合工厂历史运行数据做 Bayesian 在线标定。

制造场景中物理后果的严重性要求治理框架必须包含物理安全层的硬约束表达。之所以将规则治理设计为独立于推理引擎的硬约束层，根本原因正在于此——物理安全层的硬约束确保 AI 治理的底层安全逻辑不因上层推理能力的迭代而失效。

选取 IEC 62264 作为层级控制代表、RAMI 4.0 作为信息模型代表、ISO 55000 作为风险管理代表。四维治理框架的工程定位可通过与三项工业标准的结构性对话加以阐明（表 2）。

表 2 AI-AGF 四维框架与三项工业标准的治理对比

| 标准 | 核心贡献 | 治理缺环 | AI-AGF 四维框架的回应 |
|---|---|---|---|
| IEC 62264 | 层级控制模型（ERP→MES→SCADA→PLC） | 不允许数据流横向跨越层级，无法支持 AI 跨层调度 UPL/SPL | 在 AI 治理层内扁平化三层功能，通过规则治理硬约束和边界治理分层权限保持安全隔离 |
| RAMI 4.0 | 资产管理壳（AAS）统一信息模型 | 资产中心化架构：描述”资产能做什么”，未定义”谁决定资产何时做什么” | 从资产中心化推进至治理中心化：目标/规则/边界/风险四维分别界定做什么/不能做什么/谁有权决定/错了怎么办 |
| ISO | 资产全生命周期 | 风险逻辑以”资产价 | 将风险治理设为独立维度并设置 |

| 标准 | 核心贡献 | 治理缺环 | AI-AGF 四维框架的回应 |
| --- | --- | --- | --- |
| 55000 | 风险管理 | 值最大化”为核心，不涉及决策者自身可靠性问题 | 三级响应机制（黄/橙/红），回应 ISO 55000 框架中不存在的决策者可靠性问题 |

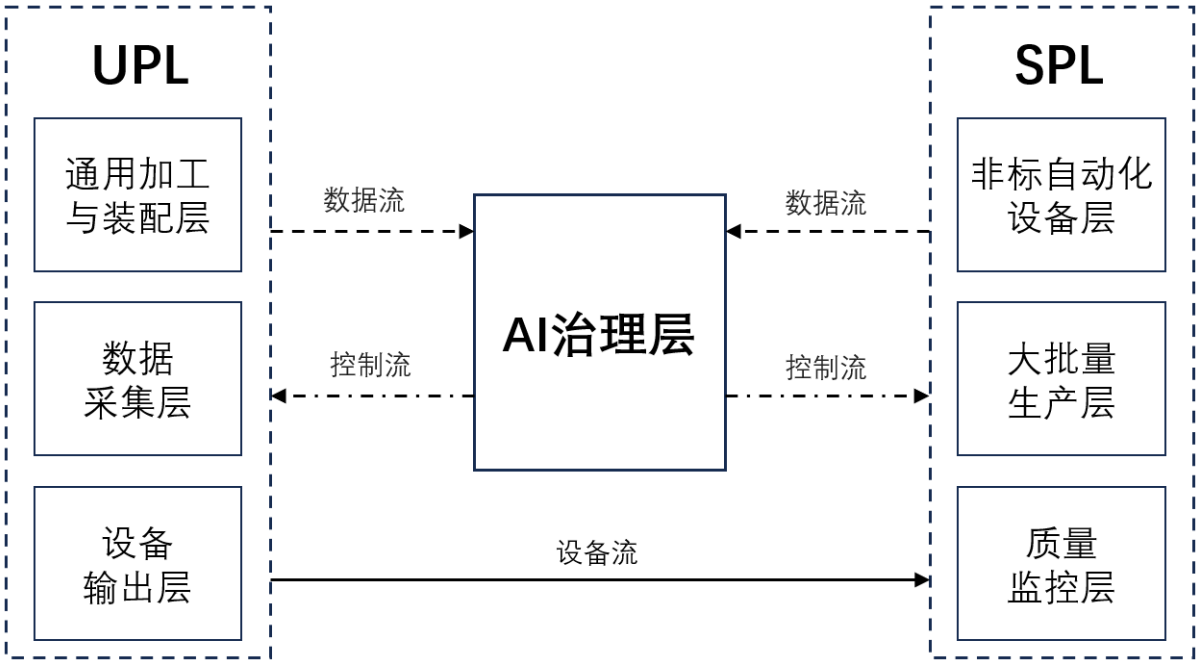


图 1：UPL/SPL 双轨架构示意图

### 3.3 全能产线（UPL）：设备生产设备的能力内核

**功能定位**：UPL 是 AI-AGF 的“能力内核”，承担两项核心任务：(1) 制造工厂自身运行所需的生产设备（“设备生产设备”）；(2) 产品打样与小批量试产。UPL 由流水线、大量工业机械手和多自由度机器人组成，产线相对固定，倾向于通用化设计。产品打样阶段不仅完成制造，更重要的是实时采集每个工序的操作参数、物料状态、设备负载、节拍等数据并注入 AI 训练数据库。这些数据随后成为 AI 规划标准产线的“原料”。张党等[17]提出的数据-知识混合驱动智能控制框架可为 UPL 中的数据采集与知识沉淀提供技术参考。UPL 在制造系统分类学中属 AI 治理层统一调度的通用制造单元，其工序切换与参数优化由 AI 闭环自主完成，响应速度与优化粒度非人工排程所能企及。

UPL 制造能力边界：覆盖(i)标准加工工艺（车铣钻磨、激光切割、3D 打印等）的机械结构件与夹具，(ii)标准电气接口的传感器/执行器集成模块，(iii)通用机械手可组

装的非标自动化工位；排除(a)依赖专用光学镀膜或超精密光刻的光学检测模块，(b)高功率激光/电子束焊接的特殊材料连接设备，(c)化学蚀刻或半导体掺杂的传感器芯片。边界非固定——随 UPL 加工能力迭代扩展，AI 治理层在规划时动态判定每个非标件的“自给可行性”并生成外部采购清单。

### 3.4 标准产线（SPL）：大批量高效生产的主体

**功能定位**：SPL 是 AI-AGF 的“生产主体”，由流水线、大量非标自动化专用设备和若干工业机械手组成。与 UPL 的关键区别在于工序实现方式：UPL 中由通用机械手完成，SPL 中由 UPL 为该产品专门生产的非标设备完成——本质是成本优化，非标设备针对单一工序的极致优化远优于通用机械手的“通用但缺乏专项效率”。SPL 具有动态可重构属性——AI 根据打样数据规划产线拓扑后由机器人搭建。

物理重构分四步：(i)**拓扑解算**——AI 根据工序图生成产线物理排布拓扑；(ii)**模块标定**——统一物理接口标准，机器人视觉引导定位至预标定网格节点；(iii)**自动校准**——数字孪生虚拟联调匹配节拍与物料时序，物理运行后实时微调；(iv)**模块级分步切换**——AI 生成最小停机方案，将产线拆分为独立可停机模块逐模块重组。上述流程的工业可行性取决于物理接口标准化程度——当所有非标设备的外部尺寸、安装孔位和电气接口遵循同一规范时，机器人即可像“乐高积木”般完成产线重组。Mayer 等[16]综述表明数字孪生可减少重构周期，Todescato 等[15]的多 Agent 框架进一步补充了自主重构图景。SPL 的动态可重构使其区别于传统专用产线——后者硬件与产品生命周期绑定，产品迭代即产线废弃；SPL 实现了“专用化效率”与“通用化柔性”的统一。

### 3.5 双轨联动的完整生产流程

UPL 与 SPL 协同构成五步闭环：(1)UPL 打样并采集数据；(2)AI 确定设备清单与排布；(3)UPL 制造非标设备；(4)机器人组装 SPL 并虚拟调试；(5)SPL 全速运行，AI 持续优化。

**脚注 1**：以智能手表为例的思想实验（作者基于消费电子 ODM 新品导入周期、非标设备平均交付周期及数字孪生虚拟调试时间压缩比例的行业数据估算）：传统模式下打样到量产约 8 个月（设备外发设计 2 个月→制造 3 个月→产线搭建调试 2 个月→试产

爬坡 1 个月）；AI-AGF 闭环下四项高度并行，在理想条件下可望压缩至传统模式的三分之一到四分之一（约 2-3 个月）。

---

## 4 AI-AGF 的启动路径

### 4.1 过渡媒介：自然语言脚本运动控制平台

从人主导过渡至 AI 主导的核心挑战在于衔接。本文设想以**自然语言脚本运动控制平台**（以中文为当前实现语种：中文脚本运动控制平台）作为过渡媒介，其设计基于两个核心判断。**第一**，自然语言脚本使工程师可以用母语直接描述控制逻辑——大语言模型原生输出自然语言推理，而脚本语言恰好是自然语言与控制指令之间的最优桥接格式。论文团队以中文为当前实现语种，但该框架的设计逻辑与语种无关，任何自然语言均可作为脚本的表示层。即使非编程专业的工艺工程师也能快速理解、修改和验证以母语编写的控制逻辑。**第二**，IMR-LLM[18]等研究已表明大语言模型能将自然语言制造任务分解为工序并生成控制代码，脚本语言简洁、可解释、易调试的特性使 AI 输出天然可被工程师审核修改，形成“人机共治”阶段的双向可读交互界面。

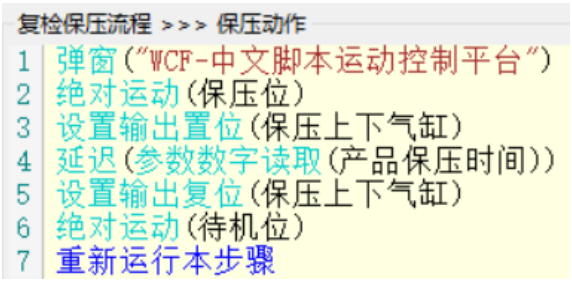


图 2：中文脚本运动控制平台

**自然语言脚本平台的基本工作原理**。该平台封装自然语言语义解释层（以中文为当前实现语种：中文脚本运动控制平台），将工序描述翻译为轴控指令，实现控制精确性与自然语言可读性的统一。

**为什么选择脚本平台而非传统 PLC/G 代码**。PLC（梯形图/结构化文本）和 G 代码的表达范式与大语言模型的自然语言-代码生成范式之间存在语义断层。自然语言脚本平台将控制逻辑切换为“工序动作自然语言描述”，恰好落入大语言模型输出分布的高密度区域——这正是 IMR-LLM[18]能以高通过率生成自然语言控制脚本的根本原因。本

文所述自然语言脚本平台的工程实践（以中文为当前实现语种：中文脚本运动控制平台）来自作者团队，待后续系统性验证。

### 4.2 从“人”到“AI”的四阶段过渡

以下时间估算基于三项参照：IMR-LLM[18]等工业 AI 从实验室到生产环境验证所需的历史外推周期、同类制造系统转型项目（如 FANUC 渐进自动化产线）的实施周期记录、以及 AI 能力在工业场景中的当前成熟度评估（详见§1.3）。各阶段里程碑均为可验证的量化指标，实际推进节奏因行业特性与企业规模而异。

#### *第一阶段：搭建基准——将“人”的知识注入“AI”*

将人类工程师积累的制造知识注入 AI 训练数据集，涵盖设备设计图纸、BOM、技术规格书与自然语言脚本运动控制程序。郑湃等[19]的“互认知”概念为此提供理论指导。**里程碑**：AI 能对至少 80%标准生产场景自主生成正确的控制脚本，且设备知识库覆盖全部在册设备。

#### *第二阶段：AI 启动，人做审核——“人机共治”模式*

AI 主导方案生成，工程师从“执行者”转变为“审核者”。AI 自主完成设备概念设计、控制脚本编写与产线布局规划，工程师审查 AI 输出并反馈迭代。陶家琦等[20]的制造领域知识图谱为此提供推理与合规检查能力。**里程碑**：AI 方案的工程师直接通过率达 85%以上，连续 3 个月无因 AI 设计缺陷导致的生产事故。

#### *第三阶段：AI 主导——逐步减少人类干预*

工程师有意识地减少对 AI 决策的干预，信任建立在 AI 性能验证与可解释性两个层面。乔非等[21]的多层次人本融合框架为此提供理论支撑。**里程碑**：工程师干预率降至 5%以下，AI 决策驳回率低于 15%。

#### *第四阶段：进入 AI 自主治理——持续优化的自主闭环*

赋予 AI 全部生产系统权限，在无人干预下自主完成感知→决策→执行闭环并实施自主治理。汪俊亮等[22]指出制造大数据支撑的持续优化闭环使 AI-AGF 脱离外部专家实现效率提升。蒋周明矩等[6]的工业 5.0 增材制造框架为此提供了参照。**里程碑**：连续 6

个月无需人类干预完成打样到量产闭环，关键 KPI 达到或超过同行业人类管理最佳水平。

量化阈值参照了 IMR-LLM 实验室通过率历史外推、同类制造系统转型周期及自主系统接管率行业基线。

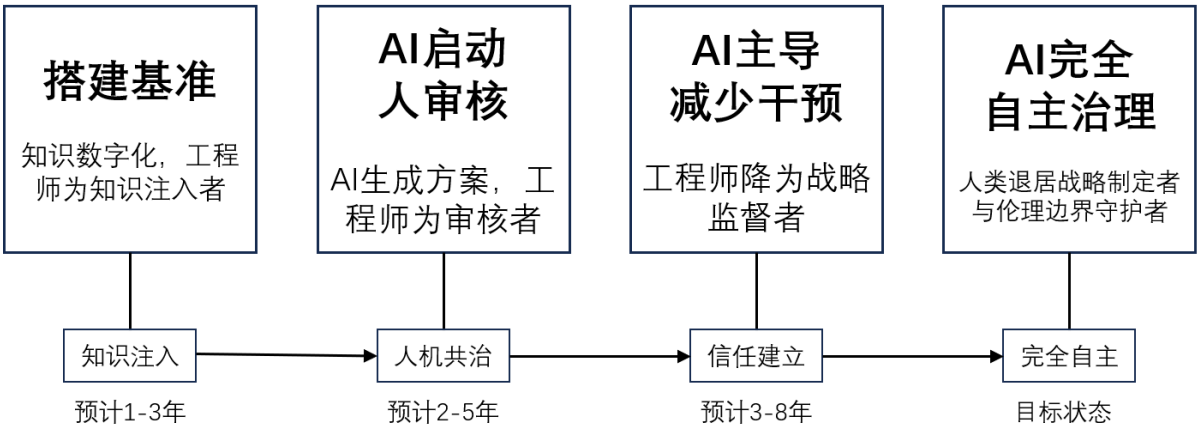


图 3：四阶段过渡路线图

---

## 5 总结与展望

### 5.1 核心贡献回顾

本文的核心贡献可概括为三个递进层面。(1) 概念层：提出 AI-AGF 概念，将自主与治理统一为制造系统研究不可分割的维度，填补既有范式的治理缺环。(2) 架构层：设计 UPL/SPL 双轨架构与四维治理框架，将制造治理从资产中心化推进至治理中心化，与 IEC 62264、RAMI 4.0、ISO 55000 形成系统性对话。(3) 路径层：以自然语言脚本运动控制平台为媒介设计四阶段渐进路线，每阶段设可验证量化里程碑，在五重困境框架中将 AI-AGF 置入产业现实语境。

### 5.2 不确定性、瓶颈与风险

AI-AGF 概念的提出基于合理的技术推断，但必须正视其面临的多重不确定性与瓶颈。

**技术不确定性与安全脆弱性：**AI 大模型在工业场景的可靠性与可解释性尚未验证——IMR-LLM[18]能否从实验室迁移至真实工厂尚不可知。当 AI 被赋予全面自主权时，引入了“网络攻击→决策篡改→物理异常→安全事故”的风险链。四维治理框架可在一

定程度上缓解（规则治理强制合规检查、风险治理降级运行），但零信任架构、AI 鲁棒性验证和入侵检测[13][14]的工业级集成仍是待攻克的工程难题。

若 AI 大模型可靠性始终无法跨越第三阶段阈值，AI-AGF 将退化为增强型自动化（详见§5.4）。

**组织与经济可行性瓶颈：**工程师从“执行者”到“审核者”的身份转换意味着权力与成就感的转移，一线人员对 AI 的信任建立和管理层对“黑箱 AI”的监管均需制度和时间消化。

**建设成本与回报周期**。初步估算，AI-AGF 增量投入约为传统模式的 1.5-2 倍成本，2-3 代产品迭代后回报周期约 3-5 年（待实证验证）。不确定性主要来自 AI 基础设施价格下降速率和该脚本平台推广速度。

### 短期与中期研究路径

**短期（1-3 年）：**将核心生产系统治理框架扩展至 ERP/MES/仓储物流等支撑子系统，验证四维治理逻辑的可迁移性。**中期（3-8 年）：**完成治理框架四个子模块的工程化实现——KPI 联动矩阵、完整规则库、决策分级表、降级策略图[23]——并在仿真环境中完成压力测试。上述路线与四阶段启动路径的前两阶段在时间轴上基本对应，可为后续推进提供方法论储备。

## 5.3 从困境到出路：AI-AGF 的前瞻展望

§2.1 所述五重困境共同指向一个症结：传统制造模式的能力边界已无法覆盖当前产业环境的复杂度。AI-AGF 对此提供差异化的结构化回应——并非声称解决所有问题，而是明确每项困境的回应方式和能力边界。

**需求侧：**双轨闭环将打样到量产周期压缩至传统模式的 20%-25%，使工厂可在有限窗口期内快速换产捕捉碎片化订单，在订单下滑中获取差异化竞争力。

**成本侧：**设备自供给闭环使工厂摆脱外部非标设备供应商的定价权依赖，AI 持续优化能耗与物料利用率，在生产成本逼近或超越收入底线的边缘挤出利润空间。

**供应链**：设备自供给闭环直接免疫设备端关键环节外部依赖风险；原材料端断供是 AI-AGF 当前能力边界之外的问题，需在 §5.2 所述研究路径中将治理框架扩展至全系统供应链后逐步覆盖。

**劳动力**：AI 自主决策替代执行层人力缺口，四维治理框架将人类从“微观决策者”解放为“宏观目标设定者”——这是 AI-AGF 最直接且最完整的回应维度。

**K 型背离**：AI-AGF 提供了一条从“依赖规模与低成本”到“依赖 AI 与自主治理”的明确转型轨道——不是告诉工厂该选哪个市场领域，而是为已选定增长轨道的工厂提供跨越能力断层的系统性方案。

困境的缓解伴随着新形态风险的引入。设备自供给闭环在免疫外部供应链风险的同时，创造了 AI-AGF 独有的治理盲区：当 UPL 制造的设备被用于 SPL 生产后，设备性能缺陷将通过“自供给链”在工厂内部闭环传播。这一闭环传播绕过了传统供应链中外部供应商的质量仲裁机制——后者在此闭环中完全失效。为此，四维治理框架中的规则治理层需设置“独立审计回路”——由外部评估节点周期性验证设备自供给链的质量独立性，切断级联反馈路径。

**产业与社会影响**：AI-AGF 将减少制造业直接雇佣人数，但同步创造 AI 训练师、治理审计师等新岗位。治理机制须在设计阶段嵌入社会责任与环境可持续性等多元目标。创业者提供产品设计即可由 AI-AGF 完成打样到量产，可视为云制造从信息层向物理层的关键下沉。

AI-AGF 是对制造本质的重新定义——制造从人类独享的技艺进化为人类与 AI 共构的治理实践。面对五重结构性困境的持续加深，方向已不再是可选项。

### 5.4 研究局限与退行路径

**研究局限**。本文属前瞻性评述，未涉及实验验证或仿真模拟，四阶段路线图的时间估算和阈值设定基于历史数据外推和行业经验参照，其在不同行业和规模工厂中的适用性有待实证检验。双轨架构的物理重构机制和 UPL 制造能力边界尚停留在概念描述层面，缺乏原型验证。本研究的主要局限包括：(1) 论证效力依赖逻辑自洽性与文献支撑度，未经受真实工厂环境的压力测试；(2) 四阶段里程碑的量化阈值参照同类系统历史周期外推，不同行业的适用性差异未充分讨论；(3) AI 大模型在工业场景的可靠

性构成了整个概念体系的底层假设——若该假设不成立，AI-AGF 将面临概念根基的整体动摇。

**反事实退行路径**。在上述局限中，最具破坏性的情景是 AI 大模型在工业场景的可靠性始终无法跨越第三阶段的门槛。在这一退行路径中，AI-AGF 虽未能实现完全自主治理，但其前期投入——知识数字化、自然语言脚本平台、工程师审核机制——仍可显著提升工厂的自动化水平，只是丧失了最核心的四项协同能力（设计反馈闭环、跨产线全局调度、知识跨产品迁移、设备供应链免疫），本质上回归了现有“AI 辅助+人工决策”的主流范式。这一退行路径的坦诚呈现并非对论文核心主张的自我否定，而是对学术诚实性的基本恪守：承认理论构想可能遭遇的现实上限。

## 参考文献


[1] 周济, 李培根, 周艳红, 等. 走向新一代智能制造[J]. Engineering, 2018, 4(1): 11-20. DOI: 10.1016/j.eng.2018.01.002.

ZHOU Ji, LI Peigen, ZHOU Yanhong, et al. Toward new-generation intelligent manufacturing[J]. Engineering, 2018, 4(1): 11-20.

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

[3] Wright P K, Bourne D A. Manufacturing Intelligence[M]. Addison-Wesley, 1988.

[4] 李伯虎, 张霖, 王时龙, 等. 云制造——面向服务的网络化制造新模式[J]. 计算机集成制造系统, 2010, 16(1): 1-7. DOI: 10.13196/j.cims.2010.01.3.libh.004.

LI Bohu, ZHANG Lin, WANG Shilong, et al. Cloud manufacturing: A new service-oriented networked manufacturing model[J]. Computer Integrated Manufacturing Systems, 2010, 16(1): 1-7.

[5] 姚锡凡, 马南峰, 张存吉, 等. 以人为本的智能制造: 演进与展望[J]. 机械工程学报, 2022, 58(18): 2-15. DOI: 10.3901/JME.2022.18.002.

YAO Xifan, MA Nanfeng, ZHANG Cunji, et al. Human-centric smart manufacturing: Evolution and prospect[J]. Journal of Mechanical Engineering, 2022, 58(18): 2-15.

[6] 蒋周明矩, 熊异, 王柏村. 面向工业 5.0 的人机协作增材制造[J]. 机械工程学报, 2024, 60(3): 238-253. DOI: 10.3901/JME.2024.03.238.

JIANG Zhoumingju, XIONG Yi, WANG Baicun. Human-robot collaborative additive manufacturing for Industry 5.0[J]. Journal of Mechanical Engineering, 2024, 60(3): 238-253.

[7] EL KALACH F, YOUSIF I, WUEST T, et al. Cognitive manufacturing: definition and current trends[J]. Journal of Intelligent Manufacturing, 2025, 36: 3695-3715. DOI: 10.1007/s10845-024-02429-9.

[8] LENG J, ZHU X, HUANG Z, et al. Unlocking the power of industrial artificial intelligence towards Industry 5.0: Insights, pathways, and challenges[J]. Journal of Manufacturing Systems, 2024, 73: 349-363.

[9] RUSSELL S. Human compatible: Artificial intelligence and the problem of control[M]. New York: Viking, 2019.

[10] AMODEI D, OLAH C, STEINHARDT J, et al. Concrete problems in AI safety[EB/OL]. arXiv preprint arXiv:1606.06565, 2016.

[11] LEIKE J, KRUEGER D, EVERITT T, et al. Scalable agent alignment via reward modeling: A research direction[EB/OL]. arXiv preprint arXiv:1811.07871, 2018.

[12] HENDRYCKS D, CARLINI N, SCHULMAN J, et al. Unsolved problems in ML safety[EB/OL]. arXiv preprint arXiv:2109.13916, 2021.

[13] STOUFFER K, PILLITTERI V, LIGHTMAN S, et al. Guide to industrial control systems (ICS) security[R]. NIST Special Publication 800-82, Revision 2, 2015. DOI: 10.6028/NIST.SP.800-82r2.

[14] MCLAUGHLIN S, KONSTANTINOU C, WANG X, et al. The cybersecurity landscape in industrial control systems[J]. Proceedings of the IEEE, 2016, 104(5): 1039-1057. DOI: 10.1109/JPROC.2015.2512235.

[15] TODESCATO M, BRAHOLLI O, CHALTSEV D, et al. Sustainable manufacturing through application of reconfigurable and intelligent systems in production processes: A system perspective[J]. Scientific Reports, 2023, 13: 22374. DOI: 10.1038/s41598-023-49727-5.

[16] MAYER A, GREIF L, HÄUßERMANN TM, et al. Digital twins, extended reality, and artificial intelligence in manufacturing reconfiguration: A systematic literature review[J]. Sustainability, 2025, 17(5): 2318. DOI: 10.3390/su17052318.

[17] 张党, 赵永宣, 王振军, 等. 数据-知识混合驱动的离散制造系统智能控制体系构架研究[J]. 机械工程学报, 2024, 60(6): 1-10,20. DOI: 10.3901/JME.2024.06.001.

ZHANG Dang, ZHAO Yongxuan, WANG Zhenjun, et al. Data-knowledge hybrid-driven intelligent control architecture for discrete manufacturing systems[J]. Journal of Mechanical Engineering, 2024, 60(6): 1-10,20.

[18] SU X, XU J, VAN KAICK O, et al. IMR-LLM: Industrial multi-robot task planning and program generation using large language models[C]//Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Vienna, Austria, 2026.

[19] 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[J]. 机械工程学报, 2023, 59(6): 173-184. DOI: 10.3901/JME.2023.06.173.

ZHENG Pai, LI Chengxi, YIN Yue, et al. Augmented reality-assisted mutual-cognitive human-robot safe interaction system[J]. Journal of Mechanical Engineering, 2023, 59(6): 173-184.

[20] 陶家琦, 李心雨, 郑湃, 等. 制造领域知识图谱的应用研究现状与前沿[J]. 计算机集成制造系统, 2022, 28(12): 3720-3736. DOI: 10.13196/j.cims.2022.12.002.

TAO Jiaqi, LI Xinyu, ZHENG Pai, et al. Research status and frontiers of knowledge graph application in manufacturing[J]. Computer Integrated Manufacturing Systems, 2022, 28(12): 3720-3736.

[21] 乔非, 刘鹃, 王冬源, 等. 工业 5.0 环境下面向生产调度的人本融合技术[J]. 机械工程学报, 2025, 61(15): 40-56. DOI: 10.3901/JME.2025.15.040.

QIAO Fei, LIU Juan, WANG Dongyuan, et al. Human-centric fusion technology for production scheduling in Industry 5.0 environment[J]. Journal of Mechanical Engineering, 2025, 61(15): 40-56.

[22] 汪俊亮, 高鹏捷, 张洁, 等. 制造大数据分析综述：内涵、方法、应用和趋势[J]. 机械工程学报, 2023, 59(12): 1-16. DOI: 10.3901/JME.2023.12.001.

WANG Junliang, GAO Pengjie, ZHANG Jie, et al. A survey on manufacturing big data analytics: Connotation, methods, applications and trends[J]. Journal of Mechanical Engineering, 2023, 59(12): 1-16.

[23] HUMAYED A, LIN J, LI F, et al. Cyber-physical systems security—A survey[J]. IEEE Internet of Things Journal, 2017, 4(6): 1802-1831. DOI: 10.1109/JIOT.2017.2703172.

## 作者简介

韦季李(通信作者)，男，1992 年出生，大专，工程师（非标自动化设备视觉引导与运动控制开发）

E-mail：jiligzs@qq.com

个人软著

便捷自动化软件平台: V1.0[CP]. 登记号: 2019SR0888778, 2019-08-27.

中文脚本运动控制平台: V1.1[CP]. 登记号: 2022SR0129906, 2022-01-20.

韦季李输入法软件: V1.0[CP]. 登记号: 2026SR0058599, 2026-01-09.

## 利益冲突声明

作者声明本文不存在任何利益冲突。