【摩杰斯AI-APS技术白皮书】拒绝“盲目自动化”：揭秘基于 Google OR-Tools 与 Qwen-4B 的真实工业智能

【摩杰斯AI-APS技术白皮书】拒绝“盲目自动化”：揭秘基于 Google OR-Tools 与 Qwen-4B 的真实工业智能

在当前的智能制造软件市场，概念炒作往往掩盖了技术的真相。许多系统标榜“AI自动排程”、“大模型决策”，实则内部逻辑混乱：要么是用简单规则伪装的“假APS”，导致工序错乱、工件报废；要么是试图用大模型直接计算排程，导致**“胡编乱造”且“响应极慢”**；更有甚者，面对产能不足的“死局”束手无策，只报错误不給方案。

对于模具机加工行业，这种“盲目自动化”是致命的。

昆山摩杰斯信息科技有限公司（MOGES） 坚持**“技术透明、架构务实”**的原则。我们的生产管理系统不玩弄概念，而是构建了一套**“数学求解 + 轻量AI + 人机协作 + 实时动态重排”**的四层防御体系，打造真正适合模具行业的**“真实智能”**。

今天，我们首次公开摩杰斯系统的核心架构：

1. **底层引擎**：基于 **Google OR-Tools** 的刚性APS排程内核。
2. **交互大脑**：基于 **Qwen-4B + RAG** 的轻量级工业记忆助手。
3. **异常处理**：独创的 **“AI辅助人工决策”** 闭环机制。
4. **动态响应**：**事件驱动的实时重排**与**可视化甘特图**。

一、技术辟谣：为什么我们坚决不用AI直接做APS?

在市场上，常有厂商宣称“用大模型一键生成排程”。在摩杰斯看来，这不仅是不专业的表现，更是对生产安全的不负责任。我们经过严格的实测与论证，发现**大语言模型（LLM）天生不适合做APS排程**，原因有二：

1. 致命缺陷：AI会“一本正经地胡说八道” (Hallucination)

- **原理限制**：大模型是基于概率预测下一个字的，它**不懂数学逻辑**，也不懂物理约束。
- **实际后果**：如果你让AI排程，它可能会为了“满足交期”，**凭空捏造**出不存在的设备空闲时间，或者**随意颠倒**工艺顺序（例如先精修后开粗）。
- **风险**：在模具行业，工序顺序错误意味着**工件直接报废**，设备冲突意味着**撞机风险**。AI的“幻觉”在工业现场是绝对不可接受的。

- **摩杰斯方案**：排程必须交给**数学求解器**。只有基于严格逻辑约束的算法（如CP-SAT），才能保证生成的每一个时间点、每一道工序都**绝对真实、绝对可行**。

2. 性能瓶颈：AI回复“超级慢”，无法应对动态变化

- **速度问题**：大模型处理复杂逻辑推理时，延迟极高。生成一个包含几千道工序的排程，大模型可能需要几分钟甚至更久，且结果往往不可用。
- **实时性要求**：模具车间瞬息万变（插单、设备故障、刀具断裂）。APS系统需要在**秒级**内重新计算出最优解。
- **摩杰斯方案**：Google OR-Tools 可以在**几秒钟内**完成数百万种可能性的遍历与求解。而AI，连“算对”都难，更别提“算快”了。

结论：**AI擅长的是“理解、交互与建议”，数学求解器擅长的才是“计算与优化”。** 摩杰斯坚持**“专业的人做专业的事”**：用OR-Tools负责排程，用AI负责辅助。

二、APS真相：刚性约束、动态优先级与资源最大化的完美平衡

模具机加工的本质是**“带强约束的多目标资源调度问题”。**摩杰斯APS引擎不仅要保证安全（不报废）、准时（不延期），更要实现**效益最大化（不闲置）**。

1. 铁律：工艺顺序不可随意变动（安全底线）

- **风险**：如果算法为了“缩短工期”而擅自将“精修”提到“热处理”之前，工件受热变形，直接**报废**。
- **摩杰斯的策略**：**刚性锁定**。
  在摩杰斯系统中，**工艺路线（Routing）是绝对硬约束**。APS引擎**不会**、也**不允许**自动调整工序的先后顺序。这是我们对客户资产安全的底线承诺。
- **例外处理**：针对极个别需要调整顺序的特殊情况，系统提供**“人工干预接口”**，只有经验丰富的工艺主管才能手动标记并指定新路径。**高风险决策留给人，高重复计算交给机器。**

2. 顶级算法背书：Google OR-Tools (CP-SAT)

既然顺序不能动，那APS算什么？**算“排队逻辑”、“资源匹配”与“效率极致”**！
当几百个订单、几千道工序争夺有限的CNC设备时，摩杰斯直接采用了 **Google OR-Tools** 中的 **CP-SAT (Constraint Programming - Satisfiability)** 求解器作为底层引擎。

- **什么是OR-Tools?** 它是Google开源的世界级运筹学工具包，其CP-SAT求解器在解决大规模整数规划和约束满足问题上，代表了全球最高水平。
- **为什么选择它?** 因为它能同时处理**硬约束**（工艺顺序、设备互斥）和**软目标**（优先级、利用率），并在几秒钟内找到**全局最优解**。

3. 核心目标一：四维动态优先级排序（保障交付）

利用OR-Tools，我们将模具生产的现实逻辑量化为**多目标优化函数**的第一层权重，确保重要订单优先交付：

1. **客户等级权重 (W_1)**：VIP客户的订单自动获得高优先级插队权。

2. **交期紧迫度 (W_2)**：基于 `剩余时间 / 剩余工时` 的动态系数，交期越近，优先级指数级上升。
3. **工件总工时 (W_3)**：短工件优先策略（SPT），快速清空小单，提升设备周转率。
4. **剩余工时负荷 (W_4)**：识别那些“已经做了很久但还剩很多”的滞留订单，防止其被无限期推迟。

4. 核心目标二：资源利用率最大化（降低成本）—— 摩杰斯的独特优势

仅仅满足交期是不够的。模具行业的CNC设备动辄数百万，高级技工薪资高昂。**设备闲置一分钟，就是真金白银的流失。**
摩杰斯APS引擎将**“最大化资源利用率”**作为核心优化目标（Objective Function）：

- **设备零闲置算法**：
  - OR-Tools会自动寻找工序间的**“微小缝隙”**。例如，在等待长工件加工完成的间隙，自动插入一个短工件的装夹或检测工序，填补时间空洞。
  - **连续加工优化**：算法倾向于将相同材质、相同刀具要求的工序连续排列，减少换刀和调机时间（Setup Time），从而变相增加有效加工时间。
- **人员技能匹配**：
  - 系统不仅排设备，还排**人**。它会识别技工的技能标签（如“擅长五轴”、“擅长精密磨削”），将复杂工件分配给最合适的人，避免“大材小用”或“小材大用”造成的效率损失。
  - **负载均衡**：防止某位加工人员忙死，而另一位闲死，确保整体人力产能最大化。
- **外协时间窗的精准嵌入**：
  - 对于热处理、电镀、发黑等外协工序，系统强制预留固定的“发出 - 返回”时间窗口，并利用这段“真空期”安排其他内部工序，确保设备和人员在等待外协回厂期间**不停工**。

算法公式示意（简化版）：

结果：摩杰斯排出的计划，不仅**交得出**，而且**做得满**。每一台设备、每一位加工人员的时间都被压榨到了极致，直接转化为工厂的**净利润**。

三、突破边界：当APS“算不出来”时，AI如何辅助人工决策?

即使是最强大的数学求解器，也会遇到**“无解”（Infeasible）的情况。例如：订单总量超过了工厂的最大产能（任务溢出），或者关键设备突发故障导致工期必然延误。**
此时，APS会返回“Infeasible（不可行）”**状态。很多系统到此就卡住了，只报错不给方案。

摩杰斯的创新在于：引入 AI 作为“谈判专家”，构建“人机协作”闭环。

1. 场景：任务溢出（产能不足）

当OR-Tools检测到无论如何排程都无法满足所有交期时：

- **传统系统**：弹出红色警告“排程失败”，然后束手无策，留给计划员一堆Excel表格去手动调整。
- **摩杰斯方案**：**Qwen-4B AI 主动介入对话。**

- **AI 报告**：“警告：当前订单总量超出产能15%，预计有3个订单（#A01, #B05, #C12）无法按期交付。”
- **AI 分析**：“经检索历史数据，订单#B05的客户对交期容忍度较高；订单#C12的部分工序可外协。”
- **AI 建议**：“建议方案：1. 与客户沟通将#B05延期2天；2. 将#C12的粗加工工序外协。采纳任一方案即可解除溢出。”

2. 交互：自然语言驱动决策

计划员无需在复杂的表格中拖拽，只需通过**对话框**与AI互动：

- **计划员**：“把#C12转为外协，重新排程。”
- **AI 执行**：
  1. 理解指令，调用后台接口修改#C12的工艺路线（标记为外协）。
  2. 触发 **Google OR-Tools** 重新运行排程算法。
  3. **秒级反馈**：“已重新排程。所有订单现在均可按期交付，设备利用率保持在92%。”

3. 价值：让机器做计算，让人做决策

- **APS**：负责在**可行域**内找最优解。
- **AI**：负责在**不可行域**内找**解决方案**，并充当人与系统的翻译官。
- **人**：负责做最终的**商业决策**（是否延期、是否外协、是否加班）。

这种架构既保留了数学算法的**严谨性**，又赋予了系统处理**极端异常**的**灵活性**。

四、动态响应：事件驱动的实时重排与可视化信任

模具车间是动态的，计划永远赶不上变化。摩杰斯系统通过**事件驱动机制**和**可视化解释**，确保计划始终“鲜活”且“可信”。

1. 事件驱动的“实时动态重排” (Event-Driven Rescheduling)

排程不是一次性的静态任务，而是**随事件触发的动态过程**。

- **触发机制**：系统实时监控并监听多种**“扰动事件”**：
  - **设备层**：IoT信号检测到主轴报警、停机或维护完成。
  - **执行层**：工人在终端扫码报工“提前完成”、“发生报废”或“请假”。
  - **管理层**：销售插入“急单”或客户变更需求。
- **自动响应**：
  - 一旦事件发生，系统**自动触发**OR-Tools进行**增量重排（Incremental Rescheduling）**。
  - **局部优化**：算法仅调整受影响的工序链，保持其他未受影响部分的稳定性，避免“牵一发而动全身”的大规模震荡。
- **效果**：从“事后补救”变为**“即时自愈”**，确保计划表永远反映车间的最新真实状态。

2. “白盒化”的可视化甘特图

计划员不敢用“黑盒”算法。摩杰斯让每一步排程都**有据可依**，直观呈现。

- **交互式甘特图**：清晰展示设备占用、工序流转、外协时间窗及人员分配。
- **AI 智能批注**：鼠标悬停在任意工序块上，Qwen-4B 即时生成解释，消除疑虑：
  - *“此工序被安排在周二上午，是因为：1. 周一刀具库满；2. 需预留周三的热处理时间窗；3. 该设备当前负载最低，且王师傅（擅长此类工件）在岗。”*
- **价值**：将复杂的数学逻辑转化为**人类可读的业务语言**，建立计划员对系统的**深度信任**。

五、AI底座：Qwen-4B + RAG，精准记忆的超级助理

为了支撑上述复杂的交互，我们摒弃了盲目追求大参数，选择了最务实的架构：**Qwen-4B + RAG**。

1. 务实之选：Qwen-4B 超小模型

- **极速响应**：毫秒级延迟，满足车间实时对话。
- **低成本私有化**：单卡/CPU即可部署，数据不出内网。
- **可控性强**：专注于工业术语，减少幻觉。

2. 灵魂所在：RAG（检索增强生成）

- **工作原理**：用户提问 -> 向量检索数据库（工单、质检、工艺） -> 注入真实数据 -> 模型生成精准回答。
- **应用场景**：
  - **记住工艺偏好**：“王师傅加工薄壁件习惯降速10%。”
  - **记住质量教训**：“供应商A的钢材去年硬度不稳定。”
  - **记住客户习惯**：“客户B验收前需额外抛光。”

结论：**Qwen-4B负责理解语言，RAG负责提供事实，OR-Tools负责执行计算。** 三者各司其职，构成摩杰斯的铁三角。

六、自主可控：从算法到代码的完全掌握

技术的先进性必须建立在**自主可控**的基础上。摩杰斯生产管理系统的所有核心逻辑，均由**昆山摩杰斯信息科技有限公司**自主研发，并已获得国家版权局认证。

- **软件名称**：摩杰斯生产管理系统
- **登记号**：**2020SR0315660**
- **著作权人**：**昆山摩杰斯信息科技有限公司**
- **开发完成日期**：2018年07月01日

这张证书代表我们的承诺：

1. **非套壳**：深入到底层算法优化（OR-Tools建模）和模型微调（Qwen-4B RAG）的技术团队。
2. **可定制**：灵活调整约束条件或训练专用语料，不受制于商业黑盒。
3. **可持续**：拥有源码级掌控力，确保系统长期演进。

七、结语：不是取代，而是最强助力

这套方案，是摩杰斯团队在模具行业深耕多年，经过无数次现场摸索、实践试错与数据验证后，提炼出的“最佳实践”。

它的核心特点在于**“扬长避短，各司其职”**：

- 它利用 **Google OR-Tools** 的算力，解决了人工无法处理的**海量组合优化**问题，让排程**更全面**（考虑所有约束与资源）；
- 它利用 **OR-Tools** 的速度与 **AI** 的实时响应，解决了人工无法做到的**秒级动态调整**，让排程**更快速**；
- 它利用 **Qwen-4B + RAG** 的记忆力，解决了人工容易遗忘的**历史经验与细节**，让决策**更精准**。

但是，我们必须郑重说明一点：

摩杰斯系统的目标，绝对不是要取代经验丰富的计划员。

模具生产充满了非标、突发与人情世故，这些是冷冰冰的算法无法完全理解的。**真正的智慧，依然在人。**

- 系统负责处理**90%的重复性计算、资源匹配与异常预警**，将计划员从繁琐的Excel表格和无尽的电话协调中解放出来。
- 计划员则专注于**10%的核心价值**：处理复杂的客户关系、做出关键的商业权衡（如是否外协、是否加班）、以及利用系统提供的“白盒”建议进行最终决策。

摩杰斯，愿做您最得力的“数字副驾驶”。
它不会抢走您的方向盘，但它能帮您看清盲区、规划最优路线、并在路况突变时瞬间给出最佳避险方案。

让我们用真实的算力和自主的技术，为您的模具工厂构建一条真正的数字化护城河，让人类智慧与机器智能完美共舞。

🏭 预约“真实场景”演示

我们不提供通用的Demo，只提供**基于您实际工艺逻辑**的沙箱测试，让您亲身体验“人机协作”的力量：

- **验证点1**：**极限施压**。输入超过产能的订单，看AI如何给出建议，您如何一键决策。
- **验证点2**：**突发响应**。模拟设备故障，观察系统如何秒级重排，为您节省多少调整时间。
- **验证点3**：**白盒信任**。随机点击工序，体验AI的透明解释，确认它是否真的“懂”您的工厂。
- **验证点4**：**效率对比**。直观感受系统排程与人工排程在**资源利用率**上的巨大差异。

- **验证点5**：**记忆测试**。询问AI关于历史订单的细节，验证其“过目不忘”的能力。

昆山摩杰斯信息科技有限公司 (MOGES)

- **官方网站**：moges.cn
- **技术咨询热线**：18625220688
- **核心理念**：让算法回归数学，让AI拥有记忆，让资源发挥极致，**让人机协作无间**，让计划实时灵动，让生产回归常识。