Hansatic

技术白皮书

神经符号式装载规划

一个装载引擎与八个前沿 AI 模型在真实客户货载上的测试

2026年11月

hansatic.com/zh/neuro-symbolic-load-planning

配套页面，所有方案均为 3D：hansatic.com/zh/neuro-symbolic-load-planning

摘要

我们正在将确定性装载引擎与生成式 AI 相结合。在构建组合之前，我们先单独衡量了每个部分。

- **数据：** 20 个真实客户装载规划案例，26 次评分运行。涵盖海运集装箱、帘式货车、一辆 53 ft (16.2 m) 干货厢式车、一个 AKE 航空集装箱和一个互锁托盘堆垛。
- **参赛者：** Hansatic 的生产装载引擎和八个主流前沿 LLM。每个模型必须在一个 JSON 答案中为每件货物给出位置和方向，不得使用工具。
- **裁判：** 一个确定性验证器以相同方式为每个参赛者评分。它检查：
  - 几何；
  - 95% 底部支撑；
  - 堆叠和易碎限制；
  - 方向标志；
  - 细长堆垛；
  - 载重和轴荷；
  - 纵向重心区间。

结果：

- **该引擎**在 26 次运行中全部合法。在每次运行中与最佳合法模型答案相比，它领先 18 次（13 次明显，5 次微弱），持平 8 次，无一落后。
- **最强模型，** GPT-5.5 在 26 次运行中有 25 次合法，包括多达 253 件货物的清单。
- **八个模型中的四个**在不到三分之一的运行中合法。
- **模型的优势。** 最佳模型答案使用了引擎未有的两种布局策略。这两种如今都已成为确定性引擎策略。

我们得出结论：如今该引擎是更好的规划者，模型的创造力值得捕捉。我们为此采用的架构是神经符号式：

- 模型提出策略；
- 引擎构建每个方案；
- 验证器只保留合法改进。

我们发布这些案例、验证器、提示词和评分代码。

1. 基准测试设计

1.1 案例

每个案例都是来自 Hansatic 应用的真实客户装载，1:1 复制：尺寸、重量、数量、标志和车辆。货物 ID 已替换，货物名称已移除。

集合	案例	运行	车辆	件数
标准 (H01-H12)	12	16	一个 20'、两个 40' 干货柜、四个 40' 高柜、一个 45' 高柜、四辆 13.6 m (44.6 ft) 帘式货车	14-83
大型 (BV1-BV6)	6	8	三个 40' 高柜、一个 40' 干货柜、一辆 13.6 m (44.6 ft) 帘式货车、一辆 53 ft (16.2 m) 干货厢式车	103-253
特殊 (U01, P01)	2	2	带斜壁的 AKE (LD3)；1200 × 1000 mm (47 × 39 in) 托盘	50 和 40

运行与目标：

- **货车** 按两项评分：装载米（LDM）和托盘位。这使 12 个标准案例产生 16 次运行，6 个大型案例产生 8 次运行。
- **BV5 和 BV6** 无法全部装入。那里的主要目标是放置更多货物。
- **U01 和 P01** 各有一个已知最优解：
  - U01：50 个中的 24 个，因为斜壁无法容纳更多；
  - P01：40 袋中的 40 袋，互锁率 100%。

案例中的约束：

- 不可堆叠和仅限顶部（易碎）货物；
- 单件堆叠重量限制；
- 此面向上和禁止倾斜标志；
- 长条货物（BV1，长度达集装箱长度）；
- 单个主导重货（H02）；
- 混合托盘和纸箱装载（BV3）。

1.2 验证器

验证器（v2.3.0）以相同方式检查每个方案。只有所有硬性检查都通过，方案才合法。

检查项	规则
边界	每件货物都在货舱空间内，包括 AKE 斜壁
重叠	任意两件货物不共享体积

检查项	规则
支撑	每件货物底部至少 95% 支撑在地板上或下方货物上
堆叠	不可堆叠货物不承载任何物品；仅限顶部货物不承载任何物品且位于顶部；单件堆叠重量限制成立
方向	遵守此面向上和禁止倾斜标志
细长堆垛	自立堆垛的高度与底座比例限制
载重	总质量在车辆限重内
车轴（货车）	前后轴轴荷在限值内，根据货物质心计算
CoG（集装箱）	纵向重心在允许范围内
多点配送	按站点后进先出，案例包含站点时适用

倾覆风险会被计算和报告，但不计入评分。

1.3 排名

按每次运行适用：

1. 合法方案胜过非法方案。
2. 放置更多货物者胜。
3. 由该运行的目标决定：
    - 装载更短（LDM）；
    - 或托盘位更少；
    - 或对于 P01，互锁占比更高，其次是高度更低。
4. 托盘位平局时按 LDM 决胜。

一个计数器为每个参赛者计算托盘位：Hansatic 的生产计数器。

胜出幅度。 我们称以下情况为胜出：

- **明显**：若在 LDM 上至少领先 2%、至少多一个托盘位、或至少多放一件货物；
- **微弱**：其他情况。

引擎在时间预算内运行，因此单次运行差异低于 2% 可能是运行间波动（第 6.2 节）。

1.4 参赛方

模型。 八家实验室各自的主流前沿模型：

- GPT-5.5 (OpenAI)
- Claude Opus 5.5 (Anthropic)
- Grok 4.7 (xAI)
- Gemini 4 Argon (Google)

- Kimi K3.1 Pro (Moonshot AI)
- DeepSeek V4 Pro (DeepSeek)
- Qwen3p8 (Alibaba)
- Magistral Medium (Mistral AI)

设置：

- **运行：** 每个案例一次。
- **输出限制：** 64k tokens，或提供商更低的上限（Gemini：32k）。
- **批处理：** 提供商提供时使用批处理 API。
- **答案收集日期：** 2026-11-11。
- **总成本：** 约 $64。

引擎。 Hansatic 的生产装载机，2026-11-20 版（提交 9f66537c）。设置：

- 生产配置：4 个工作进程和生产时间预算，固定在一个 CPU 核心上；
- 每个案例一次运行，不取最优；
- 它按自身 60% 支撑规则装载，并与所有模型一样按 95% 受审。

1.5 提示词

- 每种车辆类型一个提示词：集装箱、货车、ULD 和托盘。
- 每个提示词说明验证器对该类型适用的具体规则、目标和输出模式（货物 ID、位置、方向）。
- 不提供已解布局作为示例。

完整模板见附录 A。

2. 结果

2.1 各模型在全部 26 次运行中的表现

模型	合规	与引擎持平	落后于引擎
GPT-5.5	25	4	22
Opus 5.5	20	6	20
Grok 4.7	20	4	22
Kimi K3.1 Pro	12	5	21
DeepSeek V4 Pro	7	3	23
Gemini 4 Argon	5	2	24
Qwen3p8	4	3	23
Magistral Medium	2	0	26

无任何模型在任何运行中超越引擎。

标准集（16 次运行）

模型	合规	持平	落后
GPT-5.5	15	3	13
Opus 5.5	13	4	12
Grok 4.7	10	2	14
Kimi K3.1 Pro	8	3	13
DeepSeek V4 Pro	4	2	14
Gemini 4 Argon	4	1	15
Qwen3p8	2	1	15
Magistral Medium	1	0	16

大型集（8 次运行） 无任何模型在大型运行中与引擎持平。

模型	合规
GPT-5.5	8
Grok 4.7	8
Opus 5.5	5

模型	合规
Kimi K3.1 Pro	2
DeepSeek V4 Pro	1
Magistral Medium	1
Gemini 4 Argon	0
Qwen3p8	0

特殊集（2 次运行）

模型	合规	达到最优解
Opus 5.5	2	2
Grok 4.7	2	2
Kimi K3.1 Pro	2	2
Qwen3p8	2	2
GPT-5.5	2	1
DeepSeek V4 Pro	2	1
Gemini 4 Argon	1	1
Magistral Medium	0	0

2.2 引擎与最佳合规模型答案对比

集合	明显领先	微弱领先	持平	落后
标准（16）	7	3	6	0
大型（8）	6	2	0	0
特殊（2）	0	0	2	0
全部（26）	**13**	**5**	**8**	**0**

标准集 每个单元格依次为：装载件数、装载米数或托盘位置。

运行	引擎	最佳合规模型	差距
H01	26/27, 11.69 米 (38.4 英尺)	Opus 26, 11.74 米 (38.5 英尺)	微弱 (0.4%)
H02	24/29, 5.78 米 (19.0 英尺)	GPT 18, 5.76 米 (18.9 英尺)	明显 (+6 件)
H03	31/32, 10.51 米 (34.5 英尺)	无合规答案	明显
H04	39, 10.44 米 (34.3 英尺)	Opus 39, 10.44 米 (34.3 英尺)	持平

运行	引擎	最佳合规模型	差距
H05	49, 12.40 米 (40.7 英尺)	Opus 49, 12.40 米 (40.7 英尺)	持平
H06	83, 10.65 米 (35.0 英尺)	Grok 83, 11.41 米 (37.4 英尺)	明显 (6.6%)
H07	45, 9.60 米 (31.5 英尺)	GPT 45, 9.60 米 (31.5 英尺)	持平
H08	52, 5.92 米 (19.4 英尺)	Kimi 52, 6.12 米 (20.1 英尺)	明显 (3.3%)
H09 LDM	29/30, 13.36 米 (43.8 英尺)	GPT 29, 13.56 米 (44.5 英尺)	微弱 (1.5%)
H09 托盘	29/30, 29 位置	GPT 29, 29 位置	微弱 (LDM 决胜)
H10 LDM	26, 12.20 米 (40.0 英尺)	DeepSeek, 相同	持平
H10 托盘	26, 31 位置	DeepSeek, 相同	持平
H11 LDM	23, 12.00 米 (39.4 英尺)	Opus 23, 12.20 米 (40.0 英尺)	微弱 (1.6%)
H11 托盘	23, 30 位置	Kimi 23, 32 位置	明显 (2 位置)
H12 LDM	13/14, 11.00 米 (36.1 英尺)	Gemini 13, 11.00 米 (36.1 英尺)	持平
H12 托盘	13/14, 25 位置	Kimi 13, 28 位置	明显 (3 位置)

大型集

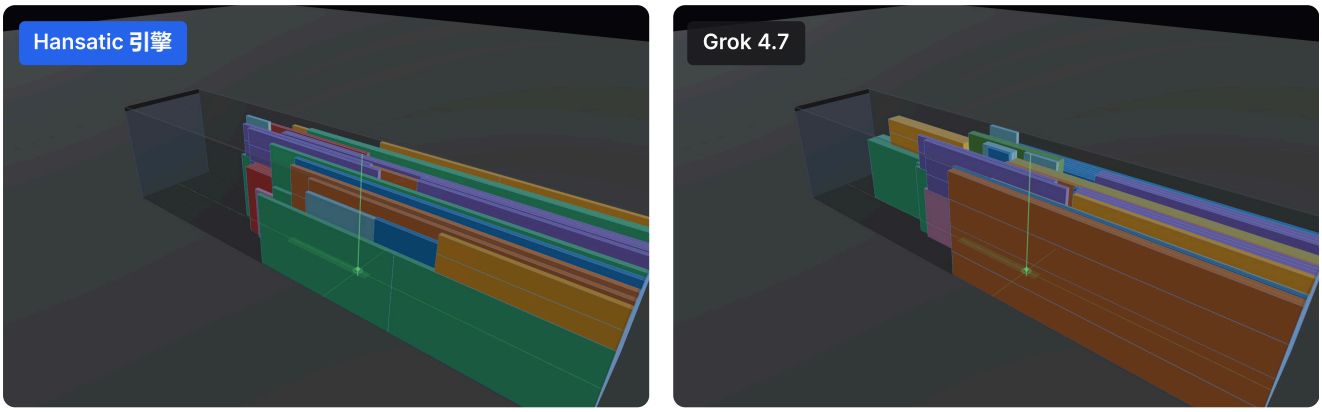
运行	引擎	最佳合规模型	差距
BV1 (40' 高柜, 103 根)	103, 9.17 米 (30.1 英尺)	Grok 103, 10.66 米 (35.0 英尺)	明显 (14.0%)
BV2 LDM (帘布车, 153)	153, 11.52 米 (37.8 英尺)	Opus 153, 11.78 米 (38.6 英尺)	明显 (2.2%)
BV2 托盘	153, 29 位置	Opus 153, 29 位置	微弱 (LDM 决胜)
BV3 LDM (53 英尺厢式车, 119)	119, 6.83 米 (22.4 英尺)	Opus 119, 6.99 米 (22.9 英尺)	明显 (2.2%)
BV3 托盘	119, 17 位置	Opus 119, 18 位置	明显 (1 位置)
BV4 (145, 仅顶部易碎)	145, 8.57 米 (28.1 英尺)	Grok 145, 8.85 米 (29.0 英尺)	明显 (3.1%)
BV5 (136, 仅直立)	127, 12.00 米 (39.4 英尺)	Opus 124, 11.77 米 (38.6 英尺)	明显 (+3 件)
BV6 (253, 95% 体积)	244, 11.98 米 (39.3 英尺)	GPT 238, 11.97 米 (39.3 英尺)	明显 (+6 件)

特殊集

运行	引擎	模型答案
U01 (AKE)	24/50, 最优解	Opus、Grok、Kimi、Qwen 和 Gemini 也达到 24；GPT 和 DeepSeek 达到 18
P01 (托盘)	40/40, 100% 交错堆叠, 最优解	每个合规模型答案均达到 40/40 最优解

2.3 差距在实际装载中的意义

- **BV1**：缩短 1.49 米（4.9 英尺），约占 40 英尺集装箱长度的 12%。
- **H12 托盘**：减少 3 个托盘位置，约占 13.6 米（44.6 英尺）帘布车底板的 9%。
- **BV6**：多装载 6 件，因无法完全装下，意味着需另一辆车少装 6 件。



BV1：Hansatic 引擎 9.17 米（30.1 英尺），左；最佳合规模型答案（Grok 4.7）10.66 米（35.0 英尺），右。

2.4 成本与时间

	引擎	模型
单次规划耗时	可忽略不计	未比较：批量 API 排队请求
单次规划成本	可忽略不计	$0.10-0.80

3. 失败分析

在 208 次模型运行（8 个模型 × 26 次）中，95 份方案合法。其余 113 份以三种不同方式失败。

模型	合法	无回答（输出上限）	提供商错误	方案违反规则
GPT-5.5	25	1	0	0
Opus 5.5	20	6	0	0
Grok 4.7	20	0	4	2
Kimi K3.1 Pro	12	14	0	0
DeepSeek V4 Pro	7	19	0	0
Gemini 4 Argon	5	2	0	19
Qwen3p8	4	22	0	0
Magistral Medium	2	1	0	23
全部	**95**	**65**	**4**	**44**

3.1 无回答：模型耗尽了思考空间

在 65 次运行中，模型一直推理到触达输出上限，始终没有写出方案。另有一份回答（Magistral，H03）正常结束但不是有效的 JSON；也计入此类。

- 这是最常见的失败方式，八个模型中有六个都遇到了。
- Kimi、DeepSeek 和 Qwen 大部分运行都因此失败。它们的合法方案大多出现在较小的案例上。
- Opus 有 6 次运行因此失败，其中包括 BV1、BV4 和 BV6，这是三个最大的舱单。

对于生产环境中的规划器来说，没有答案和非法答案一样都是失败：月台仍然需要一份方案。

3.2 提供商错误

Grok 在 4 次运行中返回 HTTP 403，因为提供商的消费限额已达上限。这些运行没有重发，计为失败。

3.3 违反规则的方案

44 份回答是完整的方案，但至少违反一条硬性规则。几乎全部来自两个模型：Gemini（19 份）和 Magistral（23 份）。Grok 有 2 份。

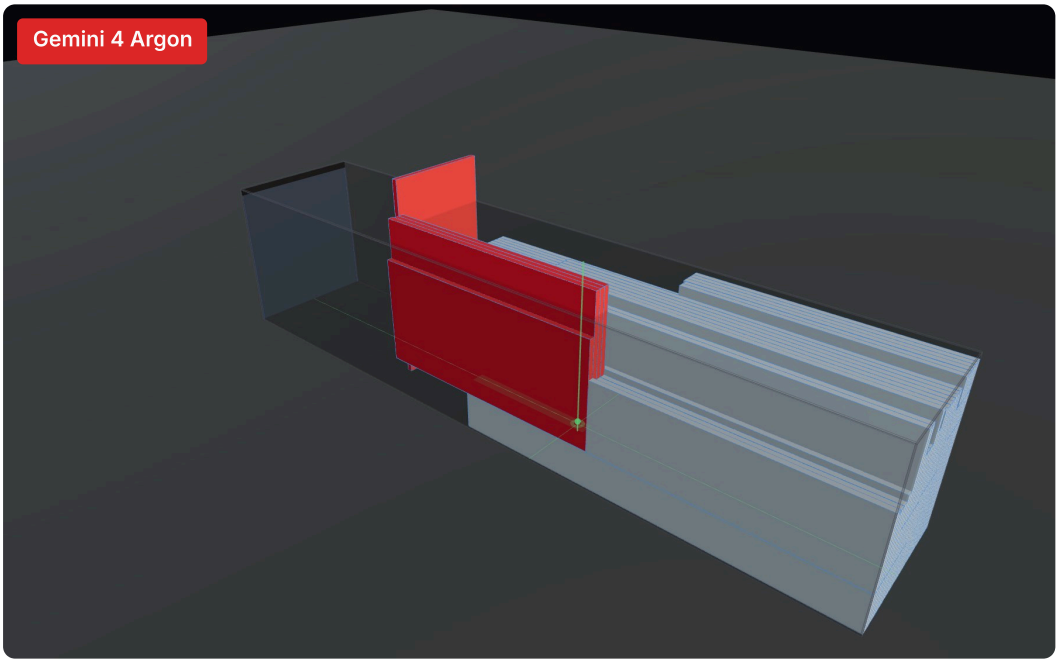
这 44 份方案中，至少违反每类规则的数量：

违反的规则	方案数
货物超出货舱空间	27
货物重叠	27

违反的规则	方案数
悬空货物（支撑率低于 95%）	21
货物尺寸不匹配任何允许的摆放方向	17
压在不可堆叠或易碎货物上，或超过堆叠上限	10
堆叠超出车顶	9
重心超出范围	7
轴荷超载	4
禁止倾斜货物被倾倒	1
超出 AKE 轮廓	1

Grok 的两份非法方案各违反一条规则：一次是货物被压坏，另一次是重心超出范围。

这些方案在 3D 视图中看起来合理。问题往往只是少数几个货物，悬空或重叠几厘米。只有完整的几何检查才能发现它们。



BV1，Gemini 4 Argon：103 件货物中放置了 77 件。高亮显示：悬空、重叠或超出货舱空间的货物。

3.4 合法，但落后

- **不完整。** 在 BV5 和 BV6 上，最佳合法模型方案比引擎少放 3 件和 6 件货物。
- **更长。** 在所有货物都能装下的运行中，最佳合法模型方案在 24 次卡车和集装箱运行中有 13 次比引擎的方案更长。最大的差距：
  - 1.49 m（4.9 ft），BV1;
  - 0.76 m（2.5 ft），H06。

3.5 模型的优势

最强的回答采用了经验丰富的装卸班组所用的策略：

- **重型货箱紧贴前壁。**
- **利用不可堆叠货物上方的高度。** 在 BV3 上，Opus 将纸箱侧放，做成 559 mm（22 in）高的底板，并在上面立起 19 台不可堆叠的电器。

引擎原本没有这两种策略。现在这两种都已成为确定性引擎策略。这正是神经符号循环旨在捕捉的创造力（第 5 节）。

4. 工具臂

在第二个实验臂中，模型可通过工具 `pack_cargo` 调用引擎，然后输出任意方案。

多数模型将货物清单复制到工具调用中，原样返回引擎方案。失败均为复制错误：

- Magistral 发送了 100 件货物，尺寸或标志错误；
- Kimi 在 H10 上调用出错；
- Qwen 在 4 次运行中失败。

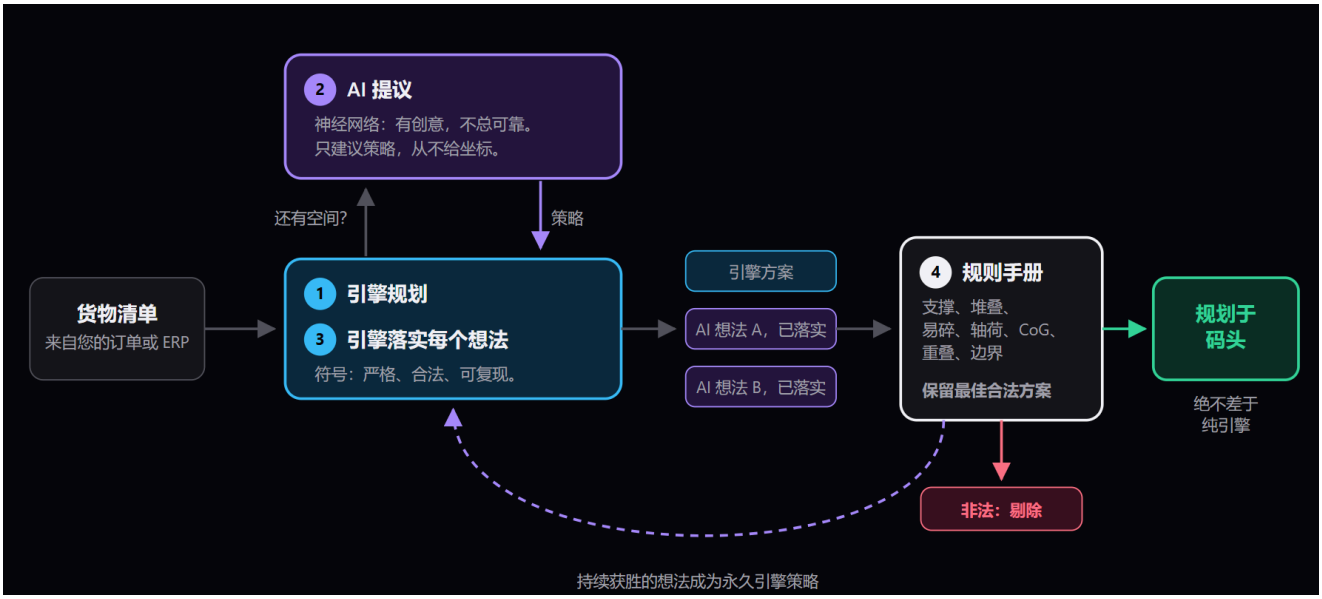
结论： 在任何组合系统中，引擎必须直接从源系统读取货物清单，绝不能经过模型。

5. 神经符号架构

5.1 两种智能

	装载引擎（符号式）	生成式 AI（神经式）
规则	每次都遵守所有已编码规则	违反所给的规则
可重复性	相同装载，相同方案	同一装载可能得到不同方案
审计	每个放置都有原因	难以审计
范围	仅限于其已有的策略	发现新策略
规模	可处理 253 件	许多模型在大型舱单上无法给出答案

5.2 循环



1. **基线。** 引擎根据舱单制定方案，舱单直接从源系统读取。
2. **触发。** 仅当基线留有改进空间时才调用模型：存在未放置的货物、装载率低，或未达到净空目标。
3. **提议。** 模型返回策略参数，而非坐标：
   - 区域；
   - 排序；
   - 方向偏好；
   - 底部与顶部结构。
4. **执行。** 引擎为每个提议构建一个方案。
5. **裁决。** 验证器保留最佳合法方案。基线始终是候选方案。
6. **蒸馏。** 反复获胜的提议将成为引擎的永久策略，如前壁策略和纸箱底策略那样。

5.3 设计保证的特性

- **绝不差于单独使用引擎，** 因为基线始终是候选方案。
- **没有非法方案到达装卸平台，** 因为验证器会裁决每个候选方案。
- **无转录风险，** 因为模型从不接触舱单或坐标（第 4 节）。
- **不会因模型输出限制而失败，** 因为提议只是几个参数，而不是一份 253 件的方案（第 3.1 节）。
- **AI 成本封顶，** 因为模型仅在触发时运行。

我们正在构建这个循环。在下一版中，它将作为独立参赛者接受评分，并统计每份方案的成本与时间。

6. 范围与限制

6.1 校验器范围

"合法"指符合 v2.3.0 中编码的规则。除多点配送的后进先出外，校验器不检查装载顺序。

6.2 差异

每个参赛者每个案例运行一次。

- **引擎**在时间预算内竞速多种策略，因此机器负载可能改变哪种策略先完成。在 H02 上，多次运行测得 5.63-5.79 m (18.5-19.0 ft)。
- **模型**只采样一次。

单次运行差异低于 2% 的报告为接近。

6.3 模型层级与输出限制

- 我们运行了每个实验室的主流前沿模型。高级版本（Anthropic Fable、OpenAI GPT-6 Astra）未运行。
- 更大的输出限制可能将部分无答案的运行（第 3.1 节）转化为方案。

6.4 案例来源

所有案例均来自一个平台的用户。其他货主的货物组合可能不同。

7. 下一步

7.1 实操规则

接下来在引擎和 validator v3 中加入两条实操规则：

- **车顶间隙：** 车顶下方至少 50 mm (2.0 in)，或在超过 25 kg (55 lb) 的堆叠货物上方至少 100 mm (3.9 in)。
- **侧向间隙**适用于每个叉车排。

我们用车顶规则重新评估了所有参赛方案。无法倾斜且高度超过车顶高度减去 50 mm (2.0 in) 的落地货物可获豁免，因为该案例本身强制要求。

- **引擎**在 16 个标准运行中的 9 个和 8 个大型运行中的 1 个上遵守该规则。
- **在 6 个运行中，某模型答案遵守该规则**，其方案接近引擎方案。这些是引擎的首要目标：
  - H02 (GPT)；
  - H06 (Grok，11.41 m / 37.4 ft)；
  - H12 LDM (Gemini，11.00 m / 36.1 ft)；

    - H12 托盘 (Kimi，28 个托盘位)；
    - BV3 LDM (GPT，9.14 m / 30.0 ft)；
    - BV5 (Opus，124 个已放置)。
- **在 8 个运行中，没有参赛方案遵守该规则：** H01、H03、H08、BV1、两个 BV2 运行、BV3 托盘、BV6。

7.2 下一版本

- Validator v3，包含车顶间隙、侧向间隙和装载顺序检查，应用于所有参赛方案。
- 新的真实客户案例。
- 每个实验室的最强模型（含高级套餐），每个模型运行多次并报告方差。
- 将神经符号循环作为参赛方案。

8. 发布

已发布：

- 每个案例的完整内容；
- 验证器；
- 提示词；
- 评分与排名代码；
- 每次运行：发送的请求、服务商未经修改的原始响应，以及解析后的方案。

查看器： 所有参赛方的全部方案，以 3D 形式与其提示词和原始回答并列展示：hansatic.com/zh/neuro-symbolic-load-planning。

不公开：

- 托盘位计数器（Hansatic 的生产计数器，对所有参赛方以相同方式应用）；
- 引擎；
- 源方案的 ID。

附录 A. 提示词模板

每种车型一条系统提示词，按发送原样列出。用户消息先给出目标，再以 JSON 格式给出案例；此处货物清单已缩减。提示词以英语显示，即模型接收时使用的语言。

海运集装箱（BV1，40' 高柜）

系统提示

```
You are an expert load planner. You will be given a sea container and a cargo manifest. Produce a complete,
physically legal loading plan.

Coordinates are millimetres. The origin is the front-left floor corner of the container.
x runs along the length from the front wall (x = 0) to the doors (x = interior length).
y runs across the width from the left wall. z runs up from the floor.
All contact and overlap tests allow 1 mm of tolerance.

Hard rules (a layout that breaks any of them is illegal):
1. Every placed item lies fully inside the cargo space: x ⩾ 0, y ⩾ 0, z ⩾ 0, x + dx ⩽ length, y + dy ⩽
width, z + dz ⩽ height.
2. No two items overlap in 3D.
3. dx, dy, dz are the item's extents along x, y, z after you orient it, and must be a permutation of its length,
width and height. Each item id appears at most once.
4. If tiltable is false the item stays upright: dz must equal its height_mm (it may only turn about the vertical
axis).
5. Support: an item with z > 0 must rest with at least 95% of its base area on the tops of items whose top is
exactly at its z.
6. Nothing may rest on an item with stackable = false, nor on a fragile item that has no max_stack_kg above 0.
7. If an item has max_stack_kg, the total weight of everything at or above its top that overlaps its footprint
(at any height) must not exceed it.
8. Total weight of placed items must not exceed max_payload_kg.
9. Balance. Skipped for a light load: total manifest weight below 15% of max_payload_kg, total manifest volume
below 25% of the cargo space, and no item of 500 kg or more (all three). Otherwise the centre of gravity of
placed cargo plus tare_kg (tare acts at the middle of the floor) must satisfy:
   a. across the width: between 40% and 60% of the width from the left wall;
   b. along the length: between 35% and 65% of the length.
10. Slender piles: an item standing on another item's top, together with the comparable items directly below it,
may be at most 5 times as tall as its narrowest braced width. An item's braced width is the narrower side of the
block it stands in: the items at the same z that touch it side to side (gap at most 10 mm), each at least half as
tall as its neighbour, chained; walls do not count, and a lone item's braced width is its own narrower side. The
pile's narrowest braced width is the smallest over its items. The pile ends at the floor or at an item whose
narrower side is more than 1.5 times the top item's narrower side; a pile that reaches the floor, and a single
item on such a wider base, are not limited.

Loading practice, on the items standing on the floor (reported as a separate verdict, not part of legality):
- They form at most two blocks: a gap longer than 300 mm along x between them starts a new block.
- Items at least 85% of the width wide that share a footprint line up at the same y.

Also reported, never scored: floor load per item against max_floor_load_kg_m2, how far the floor block stands
from the nearest end wall (more than 150 mm), more than 3 gaps over 60 mm across the width in one cross-section,
items that could tip over in transport.

Place as many items as you legally can. Any item you cannot place goes in "unplaced".

Answer with one JSON object and nothing else after it, in this exact shape:
{"placements":[{"id":"I01","x":0,"y":0,"z":0,"dx":1200,"dy":800,"dz":1000}],"unplaced":["I07"]}
```

用户消息（货物清单已缩短）

```
Objective: Minimize loading meters: the occupied length, from the frontmost to the rearmost face of any cargo
along x. Fewer meters is better.

{
```

```
 "vehicle": {
  "code": "40HC",
  "type": "container",
  "interior_mm": {
   "length": 12032,
   "width": 2352,
   "height": 2698
  },
  "max_payload_kg": 26580,
  "tare_kg": 3900,
  "max_floor_load_kg_m2": 5400
 },
 "objective": "ldm",
 "items": [
  {
   "id": "I001",
   "length_mm": 6344,
   "width_mm": 90,
   "height_mm": 638,
   "weight_kg": 31.34,
   "stackable": true,
   "max_stack_kg": 1000,
   "fragile": false,
   "tiltable": true
  },
  {
   "id": "I002",
   "length_mm": 2706,
   "width_mm": 90,
   "height_mm": 354,
   "weight_kg": 12.72,
   "stackable": true,
   "max_stack_kg": 1000,
   "fragile": false,
   "tiltable": true
  },
  "... 101 more items"
 ]
}
```

货车（H09，13.6 m（44.6 ft）侧帘车）

系统提示

```
You are an expert load planner. You will be given a truck and a cargo manifest. Produce a complete, physically
legal loading plan.

Coordinates are millimetres. The origin is the front-left floor corner of the cargo space.
x runs along the length from the front wall (x = 0) to the doors (x = interior length).
y runs across the width from the left wall. z runs up from the floor.
All contact and overlap tests allow 1 mm of tolerance.

Hard rules (a layout that breaks any of them is illegal):
1. Every placed item lies fully inside the cargo space: x ⩾ 0, y ⩾ 0, z ⩾ 0, x + dx ⩽ length, y + dy ⩽
width, z + dz ⩽ height.
2. No two items overlap in 3D.
3. dx, dy, dz are the item's extents along x, y, z after you orient it, and must be a permutation of its length,
width and height. Each item id appears at most once.
4. If tiltable is false the item stays upright: dz must equal its height_mm (it may only turn about the vertical
axis).
5. Support: an item with z > 0 must rest with at least 95% of its base area on the tops of items whose top is
exactly at its z.
6. Nothing may rest on an item with stackable = false, nor on a fragile item that has no max_stack_kg above 0.
7. If an item has max_stack_kg, the total weight of everything at or above its top that overlaps its footprint
(at any height) must not exceed it.
8. Total weight of placed items must not exceed max_payload_kg.
9. Balance. Skipped for a light load: total manifest weight below 15% of max_payload_kg, total manifest volume
below 25% of the cargo space, and no item of 500 kg or more (all three). Otherwise the centre of gravity of
placed cargo plus tare_kg (tare acts at the middle of the floor) must satisfy:
   a. across the width: between 40% and 60% of the width from the left wall;
```

```
   b. the total weight (cargo + tare) splits between kingpin and bogie like a beam on two supports at
kingpin_from_front_mm and bogie_from_front_mm; neither may exceed its max (not checked when the total is within
both limits).
10. Slender piles: an item standing on another item's top, together with the comparable items directly below it,
may be at most 5 times as tall as its narrowest braced width. An item's braced width is the narrower side of the
block it stands in: the items at the same z that touch it side to side (gap at most 10 mm), each at least half as
tall as its neighbour, chained; walls do not count, and a lone item's braced width is its own narrower side. The
pile's narrowest braced width is the smallest over its items. The pile ends at the floor or at an item whose
narrower side is more than 1.5 times the top item's narrower side; a pile that reaches the floor, and a single
item on such a wider base, are not limited.

Loading practice, on the items standing on the floor (reported as a separate verdict, not part of legality):
- They form one block: a gap longer than 300 mm along x between them starts a new block.
- Items at least 85% of the width wide that share a footprint line up at the same y.

Also reported, never scored: floor load per item against max_floor_load_kg_m2, how far the floor block stands
from the nearest end wall (more than 150 mm), more than 3 gaps over 60 mm across the width in one cross-section,
items that could tip over in transport.

Place as many items as you legally can. Any item you cannot place goes in "unplaced".

Answer with one JSON object and nothing else after it, in this exact shape:
{"placements":[{"id":"I01","x":0,"y":0,"z":0,"dx":1200,"dy":800,"dz":1000}],"unplaced":["I07"]}
```

用户消息（货物清单已缩短）

```
Objective: Minimize loading meters: the occupied length, from the frontmost to the rearmost face of any cargo
along x. Fewer meters is better.

{
 "vehicle": {
  "code": "CURTAINSIDER_136",
  "type": "truck",
  "interior_mm": {
   "length": 13620,
   "width": 2480,
   "height": 2700
  },
  "max_payload_kg": 24000,
  "tare_kg": 7200,
  "max_floor_load_kg_m2": 5000,
  "kingpin_from_front_mm": 840,
  "max_kingpin_kg": 12000,
  "bogie_from_front_mm": 12200,
  "max_bogie_kg": 21000
 },
 "objective": "ldm",
 "items": [
  {
   "id": "I01",
   "length_mm": 1000,
   "width_mm": 600,
   "height_mm": 800,
   "weight_kg": 100,
   "stackable": true,
   "max_stack_kg": 1000,
   "fragile": true,
   "tiltable": false
  },
  {
   "id": "I02",
   "length_mm": 3800,
   "width_mm": 1300,
   "height_mm": 1600,
   "weight_kg": 1000,
   "stackable": true,
   "max_stack_kg": 1000,
   "fragile": false,
   "tiltable": false
  },
  "... 28 more items"
```

```
 ]
}
```

航空集装器（U01，带斜壁的 AKE）

系统提示

```
You are an expert load planner. You will be given an air container (ULD) and a cargo manifest. Produce a
complete, physically legal loading plan.

Coordinates are millimetres. The origin is the front-left floor corner of the container.
x runs along the length from the front wall (x = 0) to the doors (x = interior length).
y runs across the width from the left wall. z runs up from the floor.
All contact and overlap tests allow 1 mm of tolerance.

Hard rules (a layout that breaks any of them is illegal):
1. Every placed item lies fully inside the cargo space: x ⩾ 0, y ⩾ 0, z ⩾ 0, x + dx ⩽ length, y + dy ⩽
width, z + dz ⩽ height.
   Sloped sides: a sloped side {side, overhang_mm, height_mm} rises from the edge of the flat floor, overhang_mm
in from that wall, to height_mm up at the wall ("y+" is the wall at y = width, "y-" the wall at y = 0). An item
whose bottom is at z may come no closer to that wall than overhang_mm x (1 - z / height_mm); above height_mm the
full width is free.
2. No two items overlap in 3D.
3. dx, dy, dz are the item's extents along x, y, z after you orient it, and must be a permutation of its length,
width and height. Each item id appears at most once.
4. If tiltable is false the item stays upright: dz must equal its height_mm (it may only turn about the vertical
axis).
5. Support: an item with z > 0 must rest with at least 95% of its base area on the tops of items whose top is
exactly at its z.
   Exception for sloped sides: an item may instead lean on the slope with its outer bottom edge, nothing under
it, at z equal to the slope's height under that edge (height_mm x reach / overhang_mm, where reach is how far the
item extends past the edge of the flat floor; within 1.5 mm), provided that z is at most 15% of dy and at least
60% of dy, and the item's centre, lie over the flat floor.
6. Nothing may rest on an item with stackable = false, nor on a fragile item that has no max_stack_kg above 0.
7. If an item has max_stack_kg, the total weight of everything at or above its top that overlaps its footprint
(at any height) must not exceed it.
8. Total weight of placed items must not exceed max_payload_kg.
9. Balance. Skipped for a light load: total manifest weight below 15% of max_payload_kg, total manifest volume
below 25% of the cargo space, and no item of 500 kg or more (all three). Otherwise the centre of gravity of
placed cargo plus tare_kg (tare acts at the middle of the floor) must satisfy:
   a. across the width: between 40% and 60% of the width from the left wall;
   b. along the length: between 35% and 65% of the length.
10. Slender piles: an item standing on another item's top, together with the comparable items directly below it,
may be at most 5 times as tall as its narrowest braced width. An item's braced width is the narrower side of the
block it stands in: the items at the same z that touch it side to side (gap at most 10 mm), each at least half as
tall as its neighbour, chained; walls do not count, and a lone item's braced width is its own narrower side. The
pile's narrowest braced width is the smallest over its items. The pile ends at the floor or at an item whose
narrower side is more than 1.5 times the top item's narrower side; a pile that reaches the floor, and a single
item on such a wider base, are not limited.

Loading practice, on the items standing on the floor (reported as a separate verdict, not part of legality):
- They form at most two blocks: a gap longer than 300 mm along x between them starts a new block.
- Items at least 85% of the width wide that share a footprint line up at the same y.

Also reported, never scored: floor load per item against max_floor_load_kg_m2, how far the floor block stands
from the nearest end wall (more than 150 mm), more than 3 gaps over 60 mm across the width in one cross-section,
items that could tip over in transport.

Place as many items as you legally can. Any item you cannot place goes in "unplaced".

Answer with one JSON object and nothing else after it, in this exact shape:
{"placements":[{"id":"I01","x":0,"y":0,"z":0,"dx":1200,"dy":800,"dz":1000}],"unplaced":["I07"]}
```

用户消息（货物清单已缩短）

```
Objective: Place as many items as you legally can. Only the number of legally placed items is ranked.

{
```

```
 "vehicle": {
  "code": "AKE",
  "type": "uld",
  "interior_mm": {
   "length": 1500,
   "width": 1905,
   "height": 1550
  },
  "max_payload_kg": 1506,
  "tare_kg": 82,
  "max_floor_load_kg_m2": 976,
  "sloped_sides": [
   {
    "side": "y+",
    "overhang_mm": 445,
    "height_mm": 610
   }
  ]
 },
 "objective": "count",
 "items": [
  {
   "id": "I01",
   "length_mm": 500,
   "width_mm": 500,
   "height_mm": 500,
   "weight_kg": 1,
   "stackable": true,
   "max_stack_kg": 1000,
   "fragile": false,
   "tiltable": true
  },
  {
   "id": "I02",
   "length_mm": 500,
   "width_mm": 500,
   "height_mm": 500,
   "weight_kg": 1,
   "stackable": true,
   "max_stack_kg": 1000,
   "fragile": false,
   "tiltable": true
  },
  "... 48 more items"
 ]
}
```


托盘（P01，1200 × 1000 mm（47 x 39 in））

系统提示

```
You are an expert load planner. You will be given a pallet and a cargo manifest. Produce a complete, physically
legal loading plan.

Coordinates are millimetres. The origin is one corner of the pallet deck, on its top surface.
x runs along the deck length, y across the deck width, z up from the top of the deck.
All contact and overlap tests allow 1 mm of tolerance.

Hard rules (a layout that breaks any of them is illegal):
1. Every placed item stays on the deck: cargo may hang past each deck edge by at most 25 mm, so -25 ⩽ x, x + dx
⩽ length + 25, -25 ⩽ y, y + dy ⩽ width + 25. And z ⩾ 0, z + dz ⩽ max_cargo_height_mm.
2. No two items overlap in 3D.
3. dx, dy, dz are the item's extents along x, y, z after you orient it, and must be a permutation of its length,
width and height. Each item id appears at most once.
4. If tiltable is false the item stays upright: dz must equal its height_mm (it may only turn about the vertical
axis).
5. Support: an item with z > 0 must rest with at least 95% of its base area on the tops of items whose top is
exactly at its z.
6. Nothing may rest on an item with stackable = false, nor on a fragile item that has no max_stack_kg above 0.
7. If an item has max_stack_kg, the total weight of everything at or above its top that overlaps its footprint
```

```
(at any height) must not exceed it.
8. Total weight of placed items must not exceed max_payload_kg.
9. Balance. Skipped for a light load: total manifest weight below 15% of max_payload_kg, total manifest volume
below 25% of deck length x deck width x max_cargo_height_mm, and no item of 500 kg or more (all three). Otherwise
the centre of gravity of placed cargo plus tare_kg (tare acts at the middle of the deck) must satisfy:
   a. across the width: between 40% and 60% of the deck width;
   b. along the length: between 35% and 65% of the deck length.
10. Slender piles: an item standing on another item's top, together with the comparable items directly below it,
may be at most 5 times as tall as its narrowest braced width. An item's braced width is the narrower side of the
block it stands in: the items at the same z that touch it side to side (gap at most 10 mm), each at least half as
tall as its neighbour, chained; nothing outside the load counts, and a lone item's braced width is its own
narrower side. The pile's narrowest braced width is the smallest over its items. The pile ends at the deck or at
an item whose narrower side is more than 1.5 times the top item's narrower side; a pile that reaches the deck,
and a single item on such a wider base, are not limited.

Loading practice, on the bottom layer (items at z = 0) (reported as a separate verdict, not part of legality):
- They form one block: a gap longer than 300 mm along x between them starts a new block.
- Items at least 85% of the deck width wide that share a footprint line up at the same y.

Also reported, never scored: how far the bottom layer stands from the nearer deck end (more than 150 mm), more
than 3 gaps over 60 mm across the width in one cross-section, items that could tip over in transport.

Place as many items as you legally can. Any item you cannot place goes in "unplaced".

Answer with one JSON object and nothing else after it, in this exact shape:
{"placements":[{"id":"I01","x":0,"y":0,"z":0,"dx":1200,"dy":800,"dz":1000}],"unplaced":["I07"]}
```

用户消息（货物清单已缩短）

```
Objective: Build a stable interlocked pallet. After the number of placed items, the interlock share is ranked:
the share of items above the bottom layer that each rest on at least two different items directly below
(overlapping each by more than 10 mm in both x and y). Higher is better; a column stack scores 0 %. Ties are
broken by the lower build height.

{
 "vehicle": {
  "code": "IND",
  "type": "pallet",
  "deck_mm": {
   "length": 1200,
   "width": 1000
  },
  "max_cargo_height_mm": 1250,
  "max_payload_kg": 1500,
  "tare_kg": 30,
  "preferred_pattern": "interlocked"
 },
 "objective": "interlock",
 "items": [
  {
   "id": "I01",
   "length_mm": 600,
   "width_mm": 400,
   "height_mm": 120,
   "weight_kg": 25,
   "stackable": true,
   "max_stack_kg": 1000,
   "fragile": false,
   "tiltable": false
  },
  {
   "id": "I02",
   "length_mm": 600,
   "width_mm": 400,
   "height_mm": 120,
   "weight_kg": 25,
   "stackable": true,
   "max_stack_kg": 1000,
   "fragile": false,
   "tiltable": false
  },
  "... 38 more items"
```

```
  ]
}
```

附录 B. 全部 26 次运行

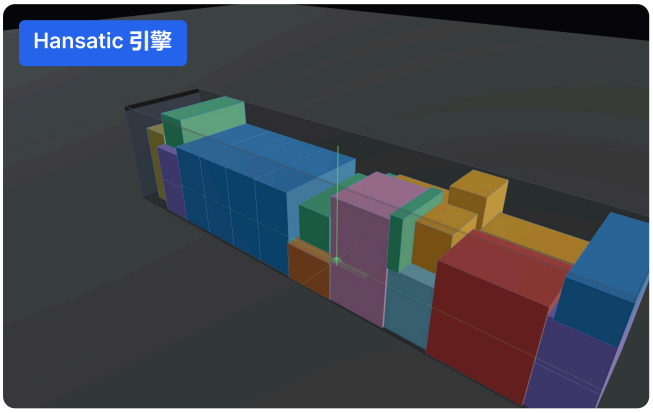
每次运行一页：货物、引擎方案与最佳合法模型答案并列，以及每位参赛者的结果。

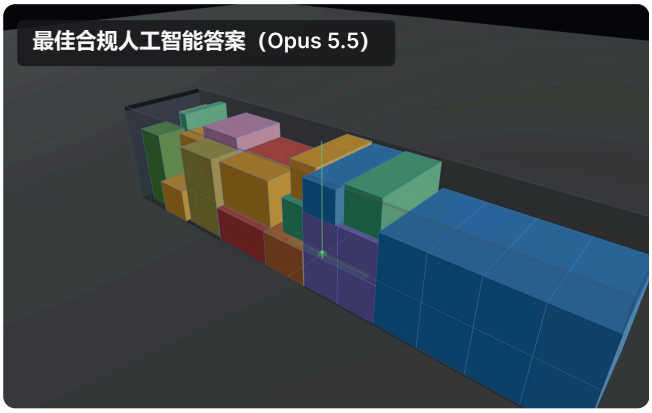
H01 最短货长

车辆	件数	目标	结果
40' High Cube	27	最短货长	引擎小幅领先

难点

27 件需保持直立





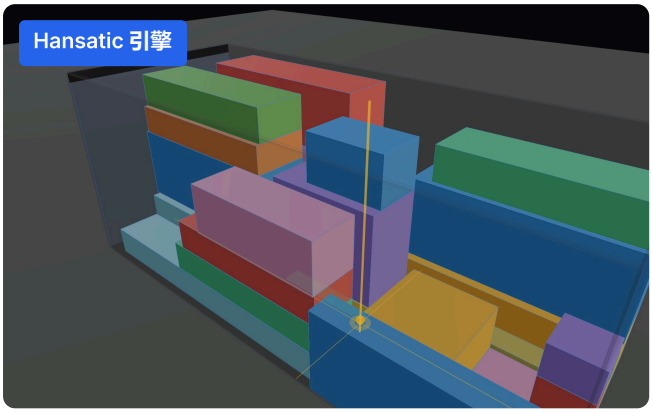
参赛方	状态	已放置	得分	备注
Hansatic引擎	**合法**	**26 / 27**	**11.69 m (38.4 ft)**	
GPT-5.5	**合法**	25 / 27	11.88 m (39.0 ft)	
Opus 5.5	**合法**	26 / 27	11.74 m (38.5 ft)	
Grok 4.7	**合法**	25 / 27	11.37 m (37.3 ft)	
Kimi K3.1 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
DeepSeek V4 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
Gemini 4 Argon	**非法**	22 / 27	11.56 m (37.9 ft)	1 货物尺寸错误, 1 超出车辆范围, 1 货物悬空
Qwen3p8	未返回方案	-	-	模型在写出方案前达到输出上限。
Magistral Medium	**非法**	27 / 27	12.09 m (39.7 ft)	7 货物重叠, 14 超出车辆范围, 3 超出高度限制, 1 货物悬空, 1 重心偏移

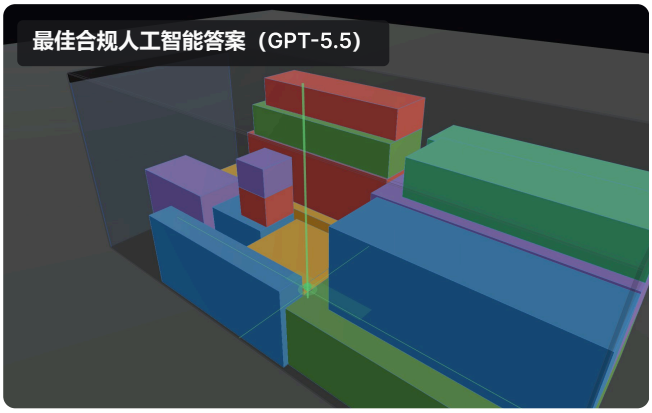
H02 最短货长

车辆	件数	目标	结果
20' Dry Container	29	最短货长	引擎明显领先

难点

29 件需保持直立; 最长 5.00 m (16.4 ft) 的货物; 最重货物 1,550 kg (3,417 lb)





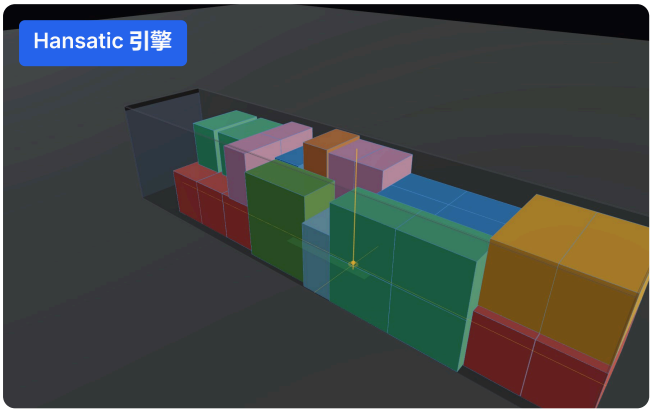
参赛方	状态	已放置	得分	备注
Hansatic引擎	**合法**	**24 / 29**	**5.78 m (19.0 ft)**	
GPT-5.5	**合法**	18 / 29	5.76 m (18.9 ft)	
Opus 5.5	未返回方案	-	-	模型在写出方案前达到输出上限。
Grok 4.7	提供商错误	-	-	提供商拒绝了请求。
Kimi K3.1 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
DeepSeek V4 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
Gemini 4 Argon	**非法**	23 / 29	5.00 m (16.4 ft)	6 货物尺寸错误, 13 货物重叠, 9 货物悬空, 9 货物压损
Qwen3p8	未返回方案	-	-	模型在写出方案前达到输出上限。
Magistral Medium	**非法**	29 / 29	7.97 m (26.1 ft)	37 货物重叠, 23 超出车辆范围, 1 超出高度限制, 7 货物悬空, 5 货物压损, 2 重心偏移

H03 最短货长

车辆	件数	目标	结果
40' High Cube	32	最短货长	引擎明显领先

难点

25 件需保持直立; 最重货物 2,680 kg (5,908 lb)



本次运行中没有人工智能模型返回合规方案。

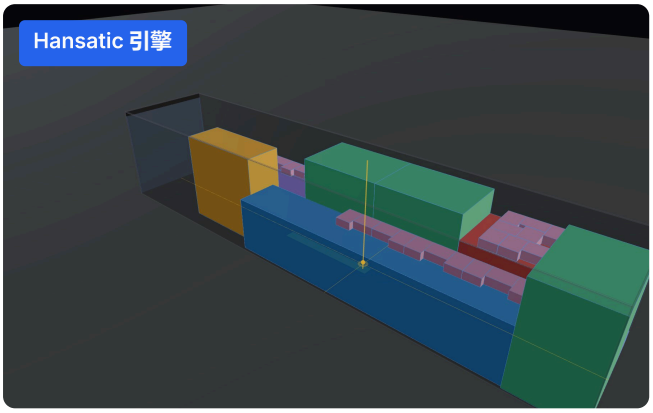
参赛方	状态	已放置	得分	备注
Hansatic引擎	**合法**	**31 / 32**	**10.51 m (34.5 ft)**	
GPT-5.5	未返回方案	-	-	模型在写出方案前达到输出上限。
Opus 5.5	未返回方案	-	-	模型在写出方案前达到输出上限。
Grok 4.7	**非法**	31 / 32	12.03 m (39.5 ft)	1 货物压损
Kimi K3.1 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
DeepSeek V4 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
Gemini 4 Argon	**非法**	31 / 32	11.81 m (38.7 ft)	1 货物尺寸错误, 12 货物重叠, 6 超出车辆范围, 3 超出高度限制, 2 货物悬空, 3 货物压损
Qwen3p8	未返回方案	-	-	模型在写出方案前达到输出上限。
Magistral Medium	未返回方案	-	-	回答中不包含有效方案。

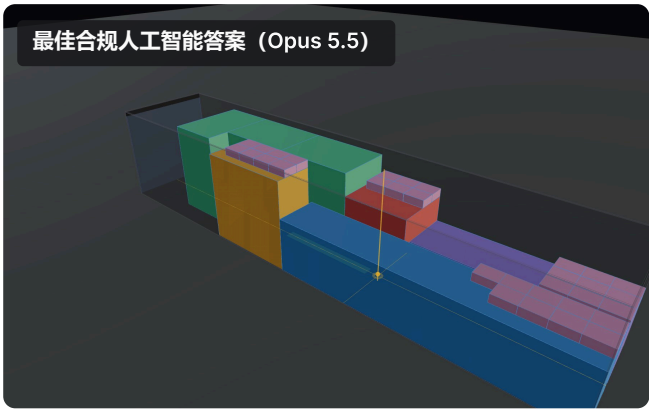
H04 最短货长

车辆	件数	目标	结果
40' Dry Container	39	最短货长	持平

难点

最长 6.30 m (20.7 ft) 的货物; 最重货物 998 kg (2,200 lb)





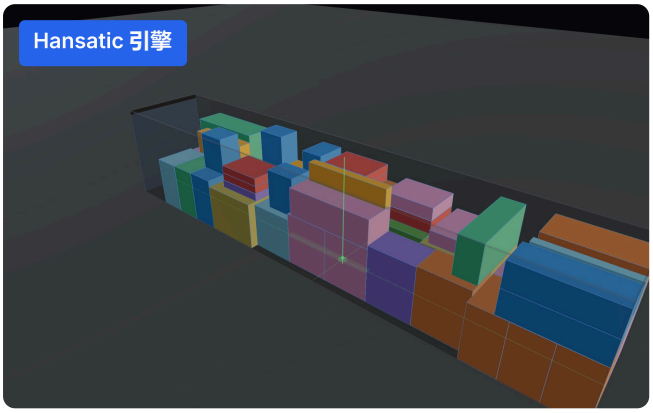
参赛方	状态	已放置	得分	备注
Hansatic引擎	**合法**	**39 / 39**	**10.44 m (34.3 ft)**	
GPT-5.5	**合法**	39 / 39	10.72 m (35.2 ft)	
Opus 5.5	**合法**	39 / 39	10.44 m (34.3 ft)	
Grok 4.7	**合法**	39 / 39	11.10 m (36.4 ft)	
Kimi K3.1 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
DeepSeek V4 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
Gemini 4 Argon	**非法**	29 / 39	6.30 m (20.7 ft)	10 货物尺寸错误, 6 货物悬空
Qwen3p8	未返回方案	-	-	模型在写出方案前达到输出上限。
Magistral Medium	**非法**	38 / 39	10.20 m (33.5 ft)	1 货物尺寸错误, 27 货物重叠, 6 超出车辆范围, 24 货物悬空, 1 货物压损, 1 重心偏移

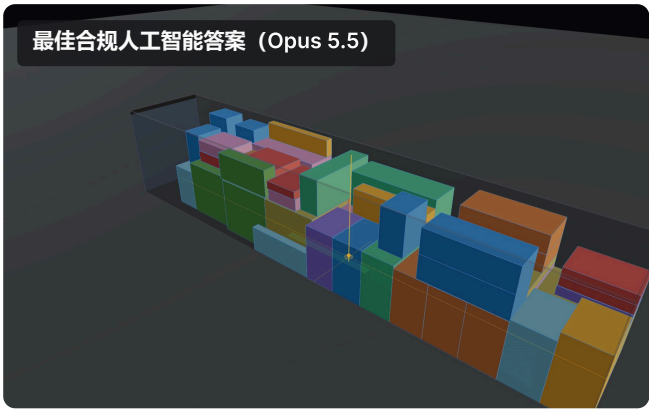
H05 最短货长

车辆	件数	目标	结果
45' High Cube	49	最短货长	持平

难点

49 件需保持直立; 27 件有堆重限制; 最重货物 814 kg (1,795 lb)





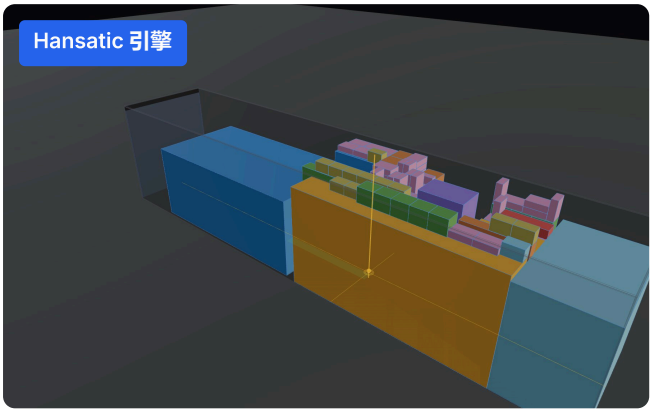
参赛方	状态	已放置	得分	备注
Hansatic引擎	**合法**	**49 / 49**	**12.40 m (40.7 ft)**	
GPT-5.5	**合法**	45 / 49	13.28 m (43.6 ft)	
Opus 5.5	**合法**	49 / 49	12.40 m (40.7 ft)	
Grok 4.7	**合法**	43 / 49	12.96 m (42.5 ft)	
Kimi K3.1 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
DeepSeek V4 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
Gemini 4 Argon	**非法**	37 / 49	10.49 m (34.4 ft)	11 货物尺寸错误, 2 货物重叠, 5 超出车辆范围, 2 超出高度限制, 3 货物悬空, 2 货物压损
Qwen3p8	未返回方案	-	-	模型在写出方案前达到输出上限。
Magistral Medium	**非法**	47 / 49	12.73 m (41.8 ft)	2 货物尺寸错误, 31 货物重叠, 9 超出车辆范围, 14 超出高度限制, 21 货物悬空, 2 货物压损

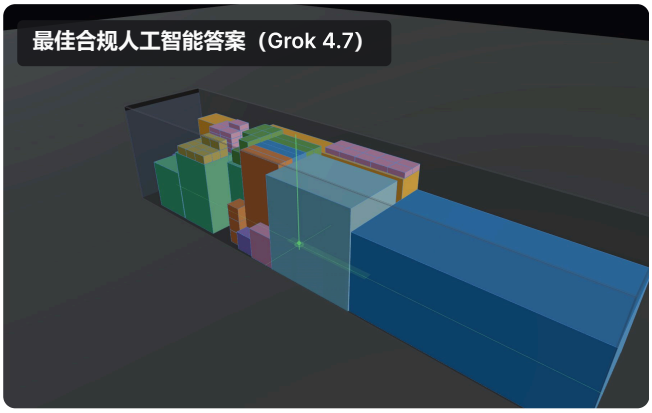
H06 最短货长

车辆	件数	目标	结果
40' High Cube	83	最短货长	引擎明显领先

难点

5 件不可堆叠货物; 1 件需保持直立; 最重货物 3,020 kg (6,658 lb)





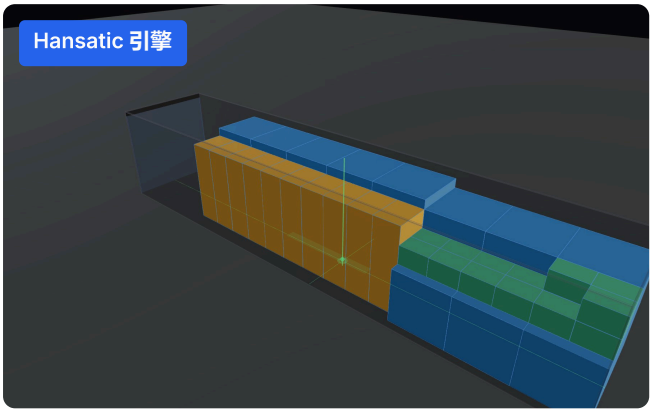
参赛方	状态	已放置	得分	备注
Hansatic引擎	**合法**	**83 / 83**	**10.65 m (35.0 ft)**	
GPT-5.5	**合法**	83 / 83	11.78 m (38.7 ft)	
Opus 5.5	未返回方案	-	-	模型在写出方案前达到输出上限。
Grok 4.7	**合法**	83 / 83	11.41 m (37.4 ft)	
Kimi K3.1 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
DeepSeek V4 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
Gemini 4 Argon	**非法**	7 / 83	6.89 m (22.6 ft)	2 货物尺寸错误, 1 重心偏移
Qwen3p8	未返回方案	-	-	模型在写出方案前达到输出上限。
Magistral Medium	**非法**	20 / 83	12.25 m (40.2 ft)	2 货物重叠, 4 超出车辆范围, 1 超出高度限制, 14 货物悬空, 1 重心偏移

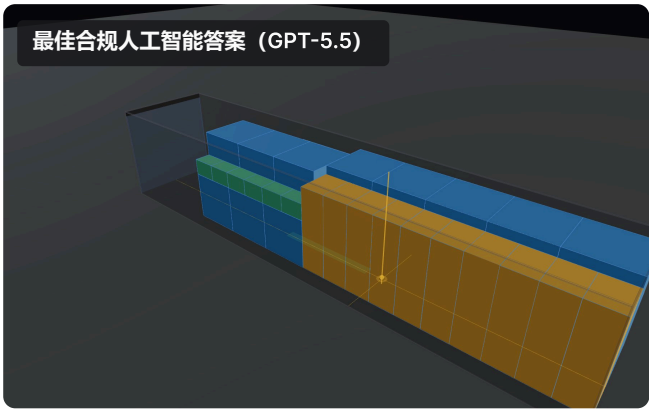
H07 最短货长

车辆	件数	目标	结果
40' Dry Container	45	最短货长	持平

难点

10 件不可堆叠货物; 15 件需保持直立; 15 件有堆重限制; 最重货物 1,200 kg (2,646 lb)





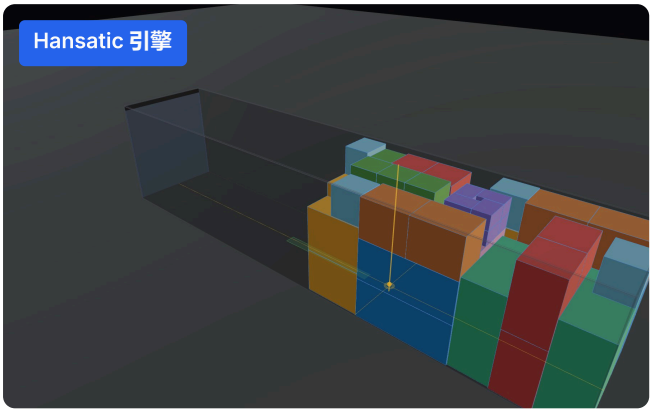
参赛方	状态	已放置	得分	备注
Hansatic引擎	**合法**	**45 / 45**	**9.60 m (31.5 ft)**	
GPT-5.5	**合法**	45 / 45	9.60 m (31.5 ft)	
Opus 5.5	**合法**	45 / 45	9.61 m (31.5 ft)	
Grok 4.7	**非法**	45 / 45	10.80 m (35.4 ft)	1 重心偏移
Kimi K3.1 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
DeepSeek V4 Pro	**合法**	45 / 45	12.00 m (39.4 ft)	
Gemini 4 Argon	**非法**	32 / 45	7.60 m (24.9 ft)	13 货物尺寸错误, 1 货物重叠, 16 货物悬空
Qwen3p8	未返回方案	-	-	模型在写出方案前达到输出上限。
Magistral Medium	**非法**	10 / 45	12.00 m (39.4 ft)	1 重心偏移

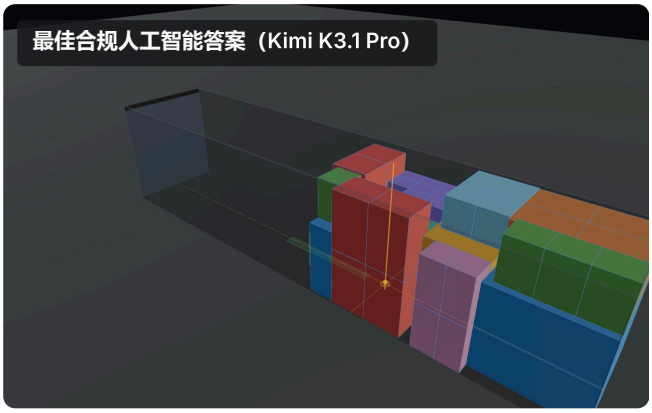
H08 最短货长

车辆	件数	目标	结果
40' High Cube	52	最短货长	引擎明显领先

难点

52 件需保持直立





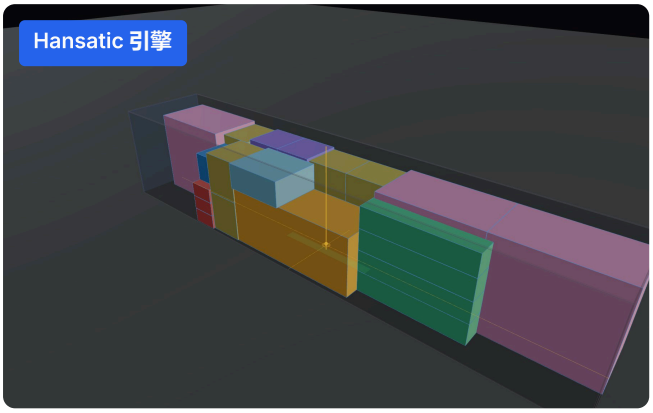
参赛方	状态	已放置	得分	备注
Hansatic引擎	**合法**	**52 / 52**	**5.92 m (19.4 ft)**	
GPT-5.5	**合法**	52 / 52	6.26 m (20.5 ft)	
Opus 5.5	**合法**	52 / 52	6.20 m (20.3 ft)	
Grok 4.7	**合法**	52 / 52	6.18 m (20.3 ft)	
Kimi K3.1 Pro	**合法**	52 / 52	6.12 m (20.1 ft)	
DeepSeek V4 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
Gemini 4 Argon	**非法**	51 / 52	6.51 m (21.4 ft)	1 货物尺寸错误, 4 货物重叠, 1 货物悬空
Qwen3p8	未返回方案	-	-	模型在写出方案前达到输出上限。
Magistral Medium	**非法**	52 / 52	11.25 m (36.9 ft)	14 货物重叠, 20 超出车辆范围

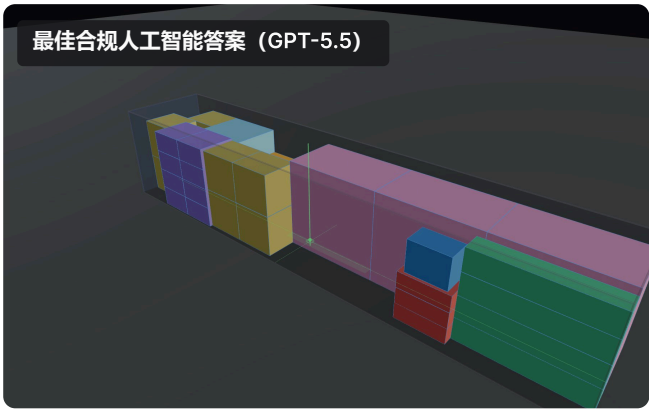
H09 最短货长

车辆	件数	目标	结果
13.6m Curtainsider	30	最短货长	引擎小幅领先

难点

30 件需保持直立; 最重货物 2,400 kg (5,291 lb)





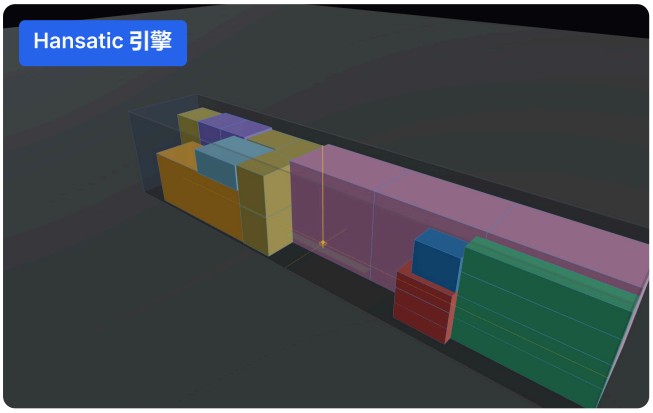
参赛方	状态	已放置	得分	备注
Hansatic引擎	**合法**	**29 / 30**	**13.36 m (43.8 ft)**	
GPT-5.5	**合法**	29 / 30	13.56 m (44.5 ft)	
Opus 5.5	**合法**	29 / 30	13.60 m (44.6 ft)	
Grok 4.7	**合法**	29 / 30	13.60 m (44.6 ft)	
Kimi K3.1 Pro	**合法**	29 / 30	13.60 m (44.6 ft)	
DeepSeek V4 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
Gemini 4 Argon	**非法**	19 / 30	11.60 m (38.1 ft)	11 货物尺寸错误, 7 货物重叠
Qwen3p8	未返回方案	-	-	模型在写出方案前达到输出上限。
Magistral Medium	**非法**	30 / 30	13.36 m (43.8 ft)	17 货物重叠, 25 超出车辆范围, 7 货物悬空, 1 轴荷超载

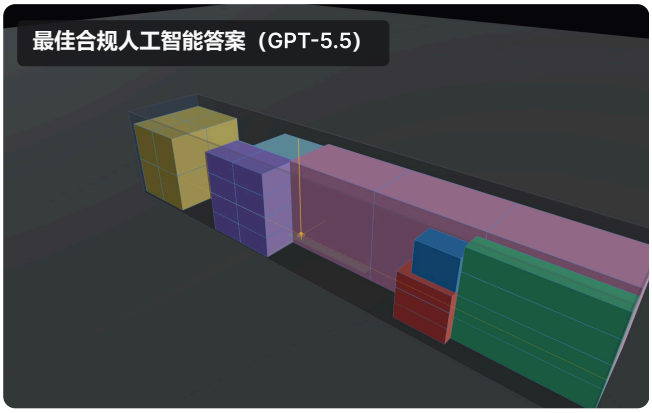
H09 最少托盘位

车辆	件数	目标	结果
13.6m Curtainsider	30	最少托盘位	引擎小幅领先

难点

30 件需保持直立; 最重货物 2,400 kg (5,291 lb)





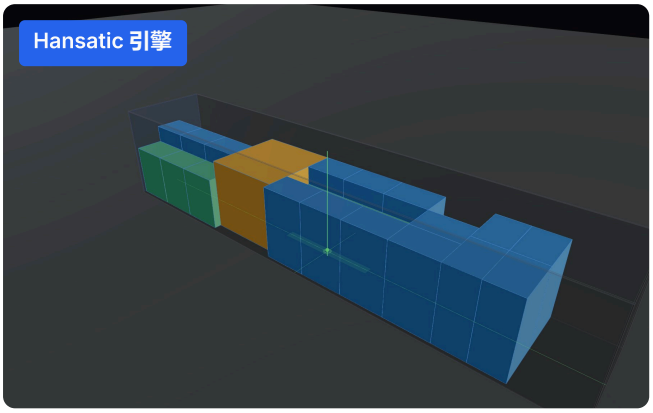
参赛方	状态	已放置	得分	备注
Hansatic引擎	**合法**	**29 / 30**	**29 个托盘位, 13.36 m (43.8 ft)**	
GPT-5.5	**合法**	29 / 30	29 个托盘位, 13.60 m (44.6 ft)	
Opus 5.5	**合法**	29 / 30	31 个托盘位, 13.60 m (44.6 ft)	
Grok 4.7	**合法**	29 / 30	29 个托盘位, 13.62 m (44.7 ft)	
Kimi K3.1 Pro	**合法**	29 / 30	29 个托盘位, 13.60 m (44.6 ft)	
DeepSeek V4 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
Gemini 4 Argon	**非法**	30 / 30	29 个托盘位, 11.72 m (38.5 ft)	18 货物重叠, 1 超出车辆范围, 3 货物压损, 1 轴荷超载
Qwen3p8	未返回方案	-	-	模型在写出方案前达到输出上限。
Magistral Medium	**非法**	30 / 30	34 个托盘位, 13.36 m (43.8 ft)	4 货物重叠, 25 超出车辆范围, 1 轴荷超载, 1 排列错位

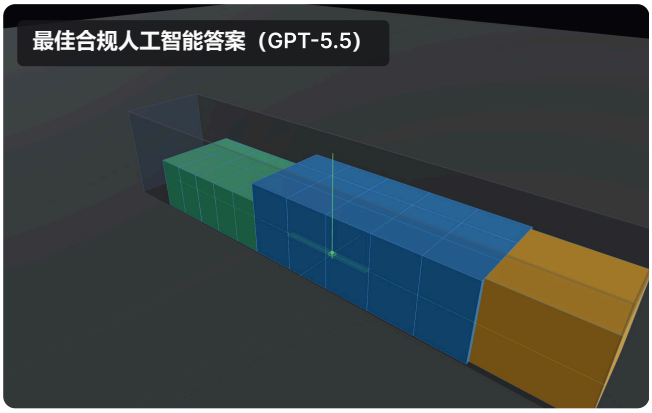
H10 最短货长

车辆	件数	目标	结果
13.6m Curtainsider	26	最短货长	持平

难点

26 件不可堆叠货物; 16 件需保持直立; 最重货物 666 kg (1,468 lb)





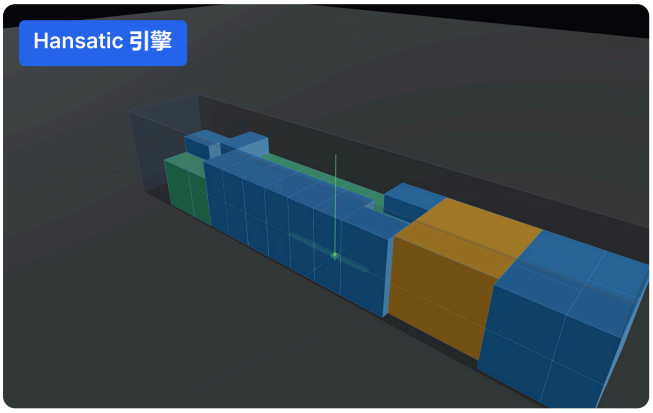
参赛方	状态	已放置	得分	备注
Hansatic引擎	**合法**	**26 / 26**	**12.20 m (40.0 ft)**	
GPT-5.5	**合法**	26 / 26	12.20 m (40.0 ft)	
Opus 5.5	**合法**	26 / 26	12.20 m (40.0 ft)	
Grok 4.7	**合法**	26 / 26	12.20 m (40.0 ft)	
Kimi K3.1 Pro	**合法**	26 / 26	12.20 m (40.0 ft)	
DeepSeek V4 Pro	**合法**	26 / 26	12.20 m (40.0 ft)	
Gemini 4 Argon	**合法**	26 / 26	12.60 m (41.3 ft)	
Qwen3p8	**合法**	26 / 26	12.20 m (40.0 ft)	
Magistral Medium	**合法**	26 / 26	13.00 m (42.7 ft)	

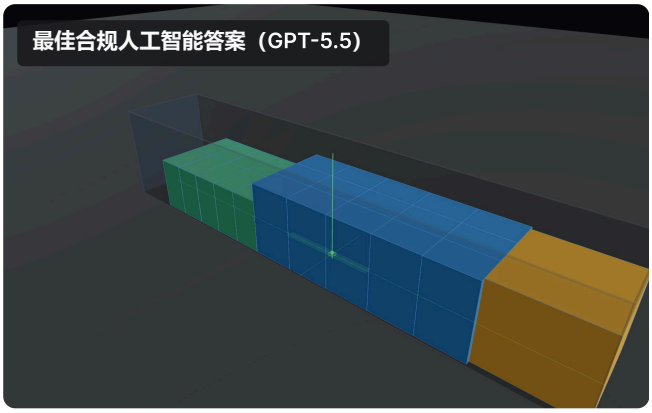
H10 最少托盘位

车辆	件数	目标	结果
13.6m Curtainsider	26	最少托盘位	持平

难点

26 件不可堆叠货物; 16 件需保持直立; 最重货物 666 kg (1,468 lb)





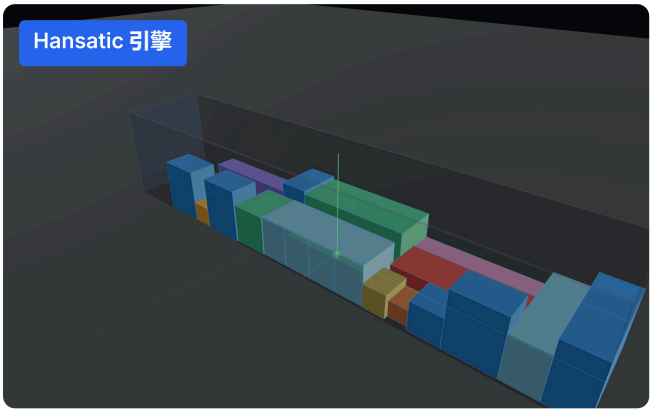
参赛方	状态	已放置	得分	备注
Hansatic引擎	**合法**	**26 / 26**	**31 个托盘位, 12.20 m (40.0 ft)**	
GPT-5.5	**合法**	26 / 26	31 个托盘位, 12.20 m (40.0 ft)	
Opus 5.5	**合法**	26 / 26	31 个托盘位, 13.00 m (42.7 ft)	
Grok 4.7	**合法**	26 / 26	31 个托盘位, 12.20 m (40.0 ft)	
Kimi K3.1 Pro	**合法**	26 / 26	31 个托盘位, 12.20 m (40.0 ft)	
DeepSeek V4 Pro	**合法**	26 / 26	31 个托盘位, 12.20 m (40.0 ft)	
Gemini 4 Argon	**非法**	26 / 26	31 个托盘位, 12.60 m (41.3 ft)	1 轴荷超载
Qwen3p8	**合法**	26 / 26	31 个托盘位, 13.00 m (42.7 ft)	
Magistral Medium	**非法**	26 / 26	28 个托盘位, 13.00 m (42.7 ft)	10 超出车辆范围

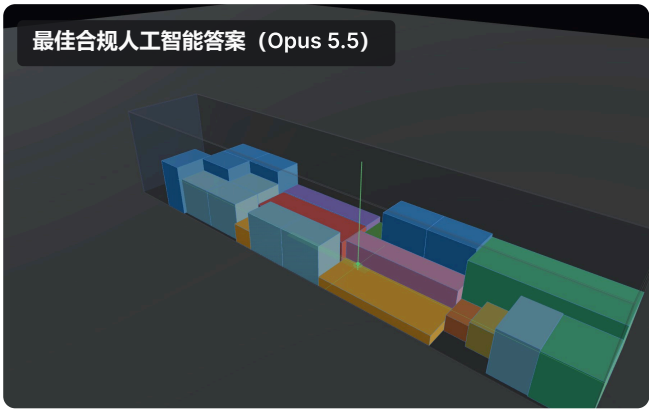
H11 最短货长

车辆	件数	目标	结果
13.6m Curtainsider	23	最短货长	引擎小幅领先

难点

23 件不可堆叠货物; 23 件需保持直立; 最重货物 2,700 kg (5,952 lb)





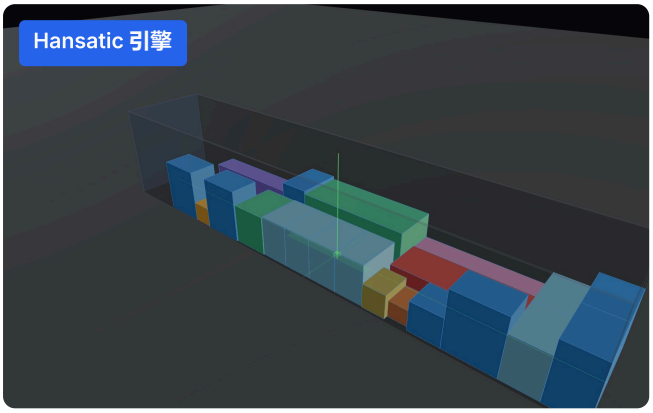
参赛方	状态	已放置	得分	备注
Hansatic引擎	**合法**	**23 / 23**	**12.00 m (39.4 ft)**	
GPT-5.5	**合法**	23 / 23	12.42 m (40.7 ft)	
Opus 5.5	**合法**	23 / 23	12.20 m (40.0 ft)	
Grok 4.7	**合法**	23 / 23	12.62 m (41.4 ft)	
Kimi K3.1 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
DeepSeek V4 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
Gemini 4 Argon	**合法**	23 / 23	13.02 m (42.7 ft)	
Qwen3p8	未返回方案	-	-	模型在写出方案前达到输出上限。
Magistral Medium	**非法**	23 / 23	12.34 m (40.5 ft)	7 货物重叠, 8 超出车辆范围

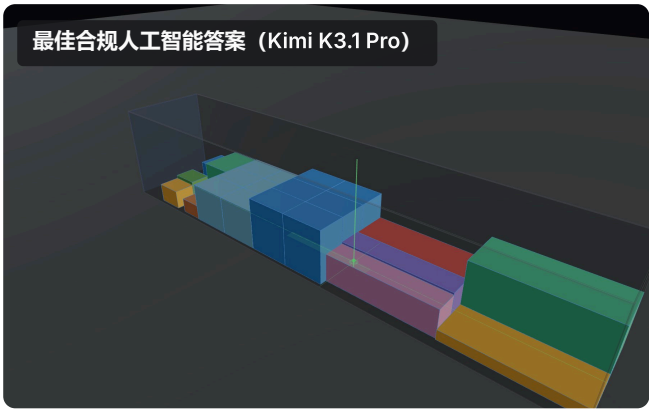
H11 最少托盘位

车辆	件数	目标	结果
13.6m Curtainsider	23	最少托盘位	引擎明显领先

难点

23 件不可堆叠货物; 23 件需保持直立; 最重货物 2,700 kg (5,952 lb)





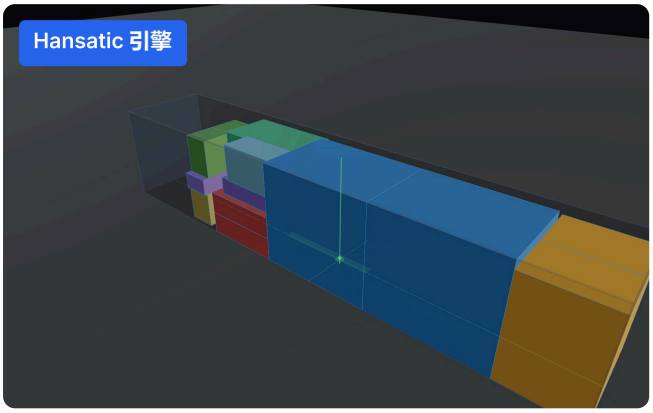
参赛方	状态	已放置	得分	备注
Hansatic引擎	**合法**	**23 / 23**	**30 个托盘位, 12.00 m (39.4 ft)**	
GPT-5.5	**合法**	23 / 23	32 个托盘位, 12.64 m (41.5 ft)	
Opus 5.5	**合法**	23 / 23	32 个托盘位, 12.84 m (42.1 ft)	
Grok 4.7	提供商错误	-	-	提供商拒绝了请求。
Kimi K3.1 Pro	**合法**	23 / 23	32 个托盘位, 12.62 m (41.4 ft)	
DeepSeek V4 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
Gemini 4 Argon	**非法**	22 / 23	30 个托盘位, 12.44 m (40.8 ft)	1 货物尺寸错误
Qwen3p8	未返回方案	-	-	模型在写出方案前达到输出上限。
Magistral Medium	**非法**	23 / 23	30 个托盘位, 12.60 m (41.3 ft)	11 货物重叠, 9 超出车辆范围

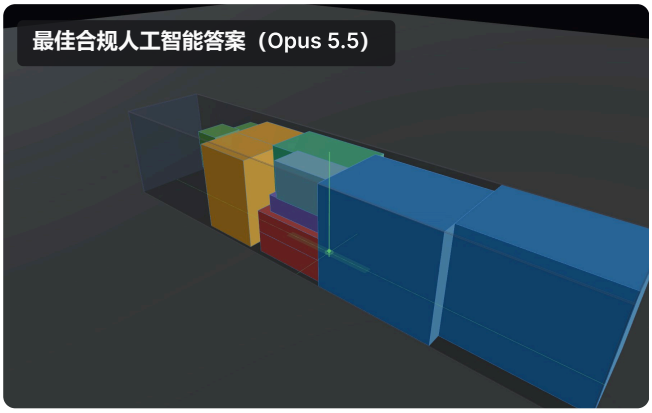
H12 最短货长

车辆	件数	目标	结果
13.6m Curtainsider	14	最短货长	持平

难点

6 件不可堆叠货物; 9 件需保持直立; 最重货物 732 kg (1,614 lb)





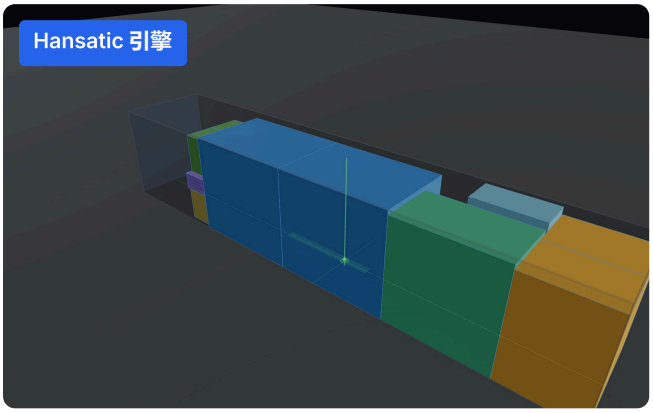
参赛方	状态	已放置	得分	备注
Hansatic引擎	**合法**	**13 / 14**	**11.00 m (36.1 ft)**	
GPT-5.5	**合法**	13 / 14	11.40 m (37.4 ft)	
Opus 5.5	**合法**	13 / 14	11.00 m (36.1 ft)	
Grok 4.7	提供商错误	-	-	提供商拒绝了请求。
Kimi K3.1 Pro	**合法**	13 / 14	11.00 m (36.1 ft)	
DeepSeek V4 Pro	**合法**	13 / 14	11.70 m (38.4 ft)	
Gemini 4 Argon	**合法**	13 / 14	11.00 m (36.1 ft)	
Qwen3p8	未返回方案	-	-	模型在写出方案前达到输出上限。
Magistral Medium	**非法**	14 / 14	13.50 m (44.3 ft)	5 货物重叠, 5 超出车辆范围, 2 货物悬空

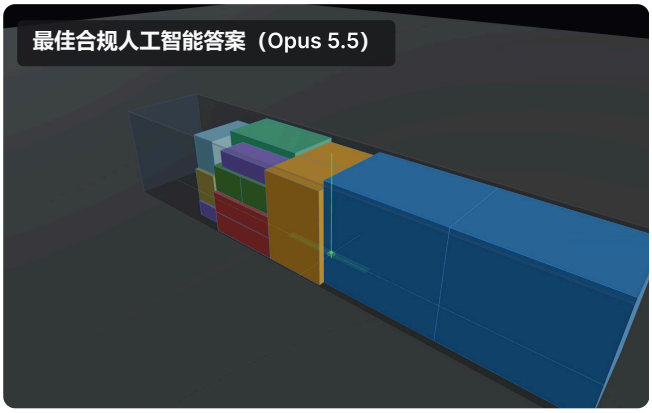
H12 最少托盘位

车辆	件数	目标	结果
13.6m Curtainsider	14	最少托盘位	引擎明显领先

难点

6 件不可堆叠货物; 9 件需保持直立; 最重货物 732 kg (1,614 lb)





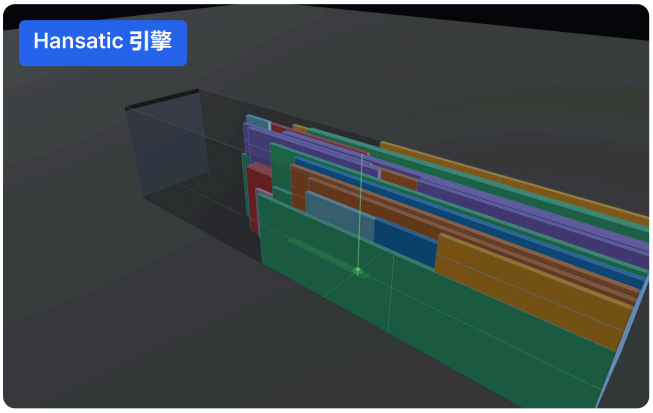
参赛方	状态	已放置	得分	备注
Hansatic引擎	**合法**	**13 / 14**	**25 个托盘位, 11.00 m (36.1 ft)**	
GPT-5.5	**合法**	13 / 14	29 个托盘位, 11.73 m (38.5 ft)	
Opus 5.5	**合法**	13 / 14	28 个托盘位, 11.00 m (36.1 ft)	
Grok 4.7	提供商错误	-	-	提供商拒绝了请求。
Kimi K3.1 Pro	**合法**	13 / 14	28 个托盘位, 11.00 m (36.1 ft)	
DeepSeek V4 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
Gemini 4 Argon	**合法**	13 / 14	28 个托盘位, 11.15 m (36.6 ft)	
Qwen3p8	未返回方案	-	-	模型在写出方案前达到输出上限。
Magistral Medium	**非法**	14 / 14	33 个托盘位, 13.20 m (43.3 ft)	9 货物重叠, 9 超出车辆范围

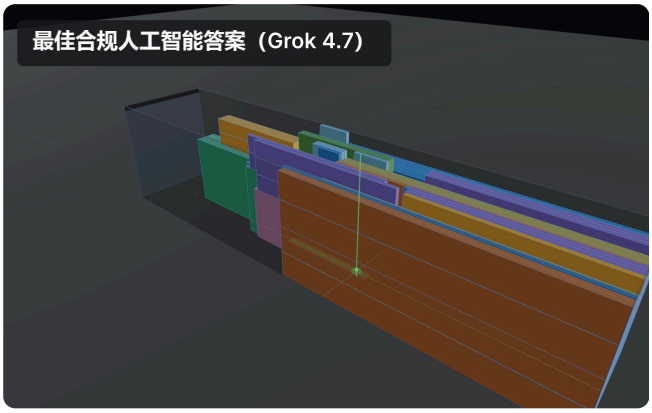
BV1 最短货长

车辆	件数	目标	结果
40' High Cube	103	最短货长	引擎明显领先

难点

37 件需保持直立; 最长 8.27 m (27.1 ft) 的货物





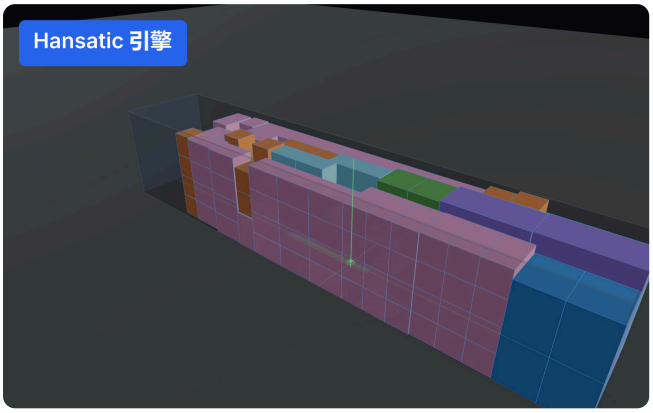
参赛方	状态	已放置	得分	备注
Hansatic引擎	**合法**	**103 / 103**	**9.17 m (30.1 ft)**	
GPT-5.5	**合法**	92 / 103	8.27 m (27.1 ft)	
Opus 5.5	未返回方案	-	-	模型在写出方案前达到输出上限。
Grok 4.7	**合法**	103 / 103	10.66 m (35.0 ft)	
Kimi K3.1 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
DeepSeek V4 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
Gemini 4 Argon	**非法**	77 / 103	8.36 m (27.4 ft)	1 货物尺寸错误, 4 超出高度限制, 4 货物悬空
Qwen3p8	未返回方案	-	-	模型在写出方案前达到输出上限。
Magistral Medium	**非法**	22 / 103	16.54 m (54.3 ft)	9 超出车辆范围

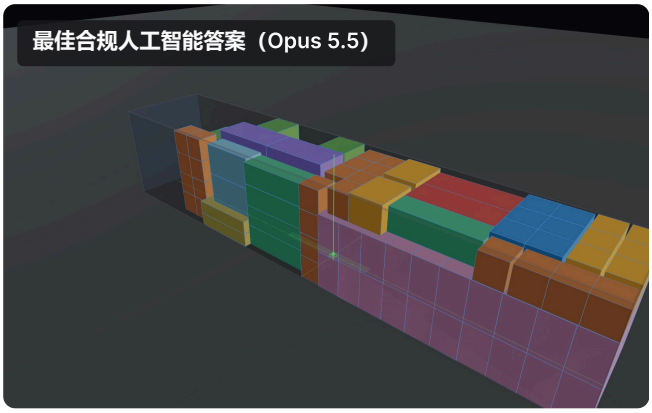
BV2 最短货长

车辆	件数	目标	结果
13.6m Curtainsider	153	最短货长	引擎明显领先

难点

2 件不可堆叠货物; 最重货物 578 kg (1,274 lb)





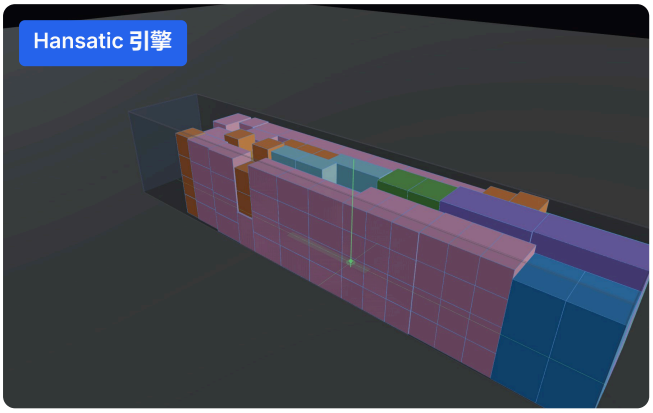
参赛方	状态	已放置	得分	备注
Hansatic引擎	**合法**	**153 / 153**	**11.52 m (37.8 ft)**	
GPT-5.5	**合法**	153 / 153	13.14 m (43.1 ft)	
Opus 5.5	**合法**	153 / 153	11.78 m (38.6 ft)	
Grok 4.7	**合法**	153 / 153	12.30 m (40.4 ft)	
Kimi K3.1 Pro	**合法**	153 / 153	11.99 m (39.3 ft)	
DeepSeek V4 Pro	**合法**	153 / 153	12.79 m (42.0 ft)	
Gemini 4 Argon	**非法**	147 / 153	12.77 m (41.9 ft)	6 货物尺寸错误
Qwen3p8	未返回方案	-	-	模型在写出方案前达到输出上限。
Magistral Medium	**非法**	56 / 153	14.02 m (46.0 ft)	39 货物重叠, 27 超出车辆范围

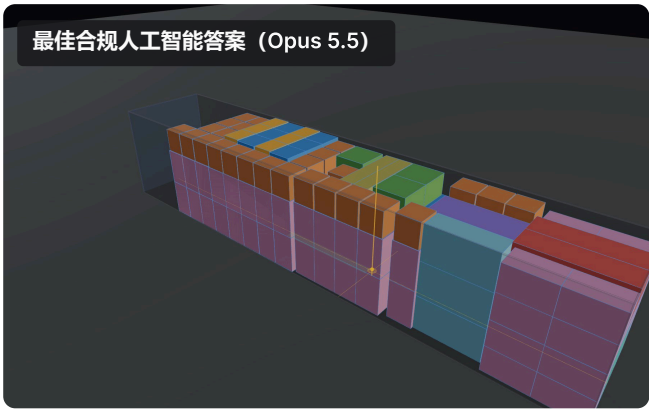
BV2 最少托盘位

车辆	件数	目标	结果
13.6m Curtainsider	153	最少托盘位	引擎小幅领先

难点

2 件不可堆叠货物; 最重货物 578 kg (1,274 lb)





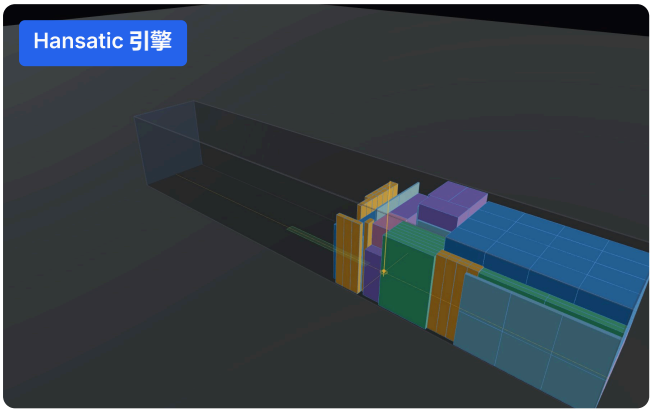
参赛方	状态	已放置	得分	备注
Hansatic引擎	**合法**	**153 / 153**	**29 个托盘位, 11.52 m (37.8 ft)**	
GPT-5.5	**合法**	153 / 153	32 个托盘位, 12.60 m (41.3 ft)	
Opus 5.5	**合法**	153 / 153	29 个托盘位, 11.67 m (38.3 ft)	
Grok 4.7	**合法**	153 / 153	32 个托盘位, 12.78 m (41.9 ft)	
Kimi K3.1 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
DeepSeek V4 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
Gemini 4 Argon	未返回方案	-	-	模型在写出方案前达到输出上限。
Qwen3p8	未返回方案	-	-	模型在写出方案前达到输出上限。
Magistral Medium	**非法**	153 / 153	34 个托盘位, 14.82 m (48.6 ft)	7 货物重叠, 31 超出车辆范围, 86 超出高度限制, 96 货物悬空

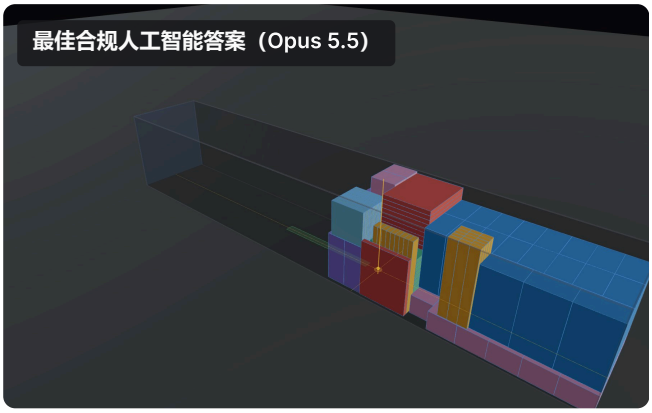
BV3 最短货长

车辆	件数	目标	结果
53ft Dry Van	119	最短货长	引擎明显领先

难点

19 件不可堆叠货物; 35 件需保持直立





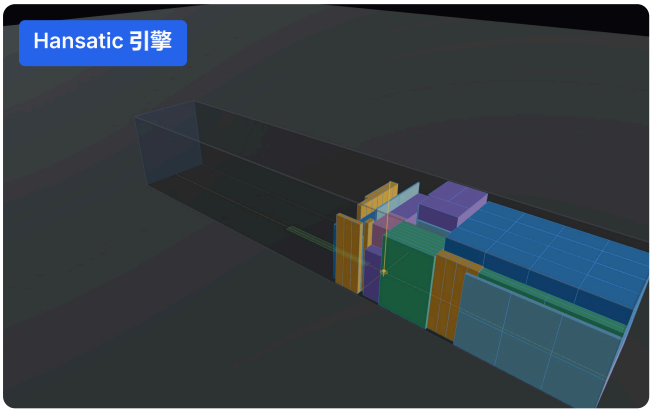
参赛方	状态	已放置	得分	备注
Hansatic引擎	**合法**	**119 / 119**	**6.83 m (22.4 ft)**	
GPT-5.5	**合法**	119 / 119	9.14 m (30.0 ft)	
Opus 5.5	**合法**	119 / 119	6.99 m (22.9 ft)	
Grok 4.7	**合法**	119 / 119	8.10 m (26.6 ft)	
Kimi K3.1 Pro	未返回方案	-	-	回答中不包含有效方案。
DeepSeek V4 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
Gemini 4 Argon	**非法**	115 / 119	8.36 m (27.4 ft)	4 货物尺寸错误, 5 货物重叠, 3 货物压损
Qwen3p8	未返回方案	-	-	模型在写出方案前达到输出上限。
Magistral Medium	**非法**	22 / 119	16.76 m (55.0 ft)	11 超出车辆范围

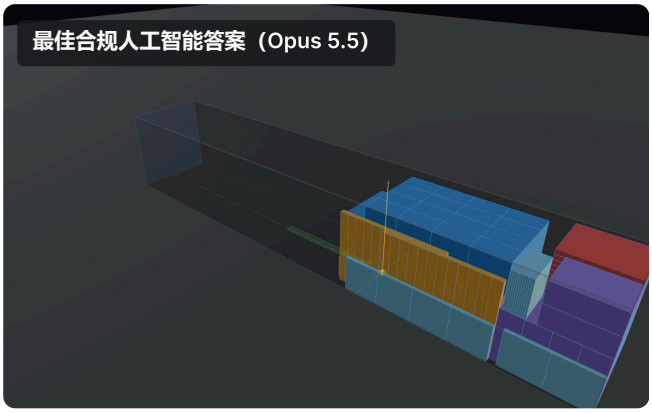
BV3 最少托盘位

车辆	件数	目标	结果
53ft Dry Van	119	最少托盘位	引擎明显领先

难点

19 件不可堆叠货物; 35 件需保持直立





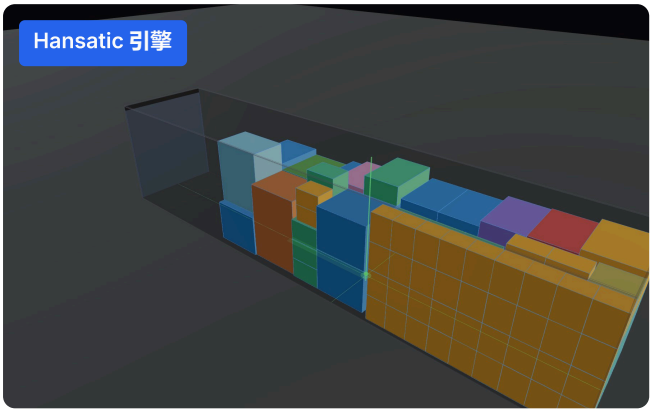
参赛方	状态	已放置	得分	备注
Hansatic引擎	**合法**	**119 / 119**	**17 个托盘位, 6.83 m (22.4 ft)**	
GPT-5.5	**合法**	119 / 119	23 个托盘位, 9.07 m (29.8 ft)	
Opus 5.5	**合法**	119 / 119	18 个托盘位, 6.99 m (22.9 ft)	
Grok 4.7	**合法**	119 / 119	22 个托盘位, 8.64 m (28.3 ft)	
Kimi K3.1 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
DeepSeek V4 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
Gemini 4 Argon	未返回方案	-	-	模型在写出方案前达到输出上限。
Qwen3p8	未返回方案	-	-	模型在写出方案前达到输出上限。
Magistral Medium	**非法**	107 / 119	21 个托盘位, 16.18 m (53.1 ft)	12 货物尺寸错误, 66 货物重叠, 35 超出车辆范围, 60 货物悬空, 1 货物压损, 1 松散货物组未固定

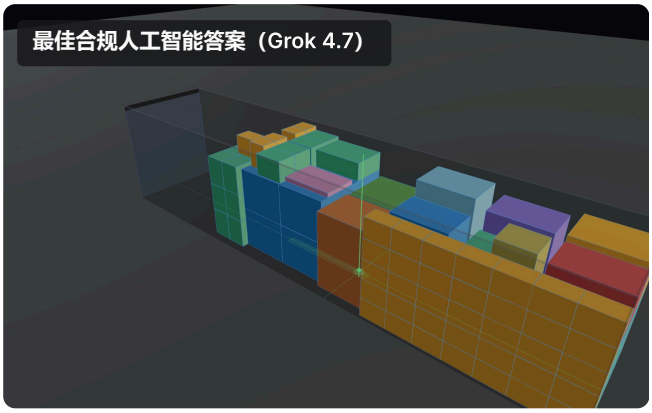
BV4 最短货长

车辆	件数	目标	结果
40' High Cube	145	最短货长	引擎明显领先

难点

11 件不可堆叠货物; 11 件易碎货物，仅限顶部; 20 件需保持直立





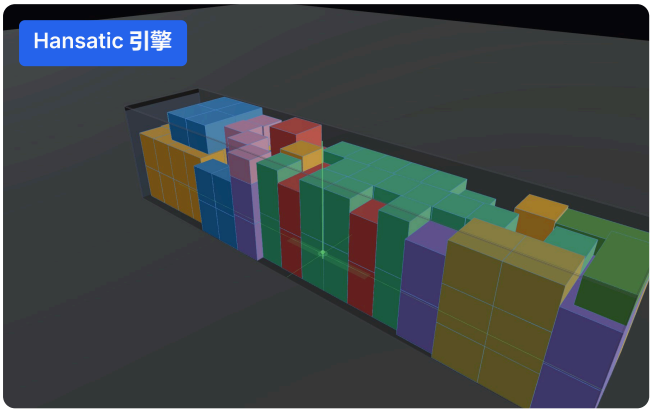
参赛方	状态	已放置	得分	备注
Hansatic引擎	**合法**	**145 / 145**	**8.57 m (28.1 ft)**	
GPT-5.5	**合法**	145 / 145	11.96 m (39.2 ft)	
Opus 5.5	未返回方案	-	-	模型在写出方案前达到输出上限。
Grok 4.7	**合法**	145 / 145	8.85 m (29.0 ft)	
Kimi K3.1 Pro	**合法**	145 / 145	10.87 m (35.7 ft)	
DeepSeek V4 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
Gemini 4 Argon	**非法**	145 / 145	10.05 m (33.0 ft)	19 货物重叠, 3 超出车辆范围, 9 超出高度限制, 4 货物悬空
Qwen3p8	未返回方案	-	-	模型在写出方案前达到输出上限。
Magistral Medium	**合法**	11 / 145	10.33 m (33.9 ft)	

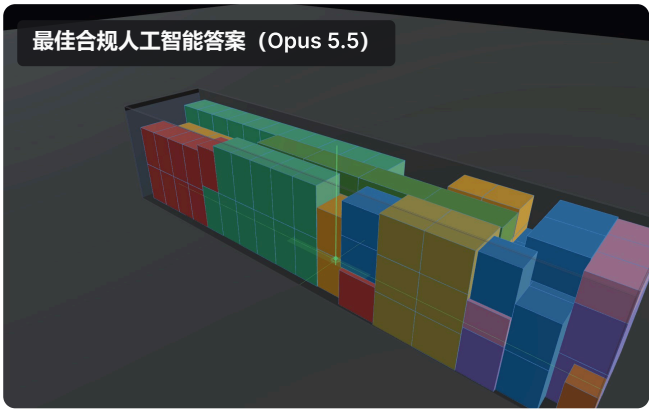
BV5 最短货长

车辆	件数	目标	结果
40' High Cube	136	最短货长	引擎明显领先

难点

136 件需保持直立





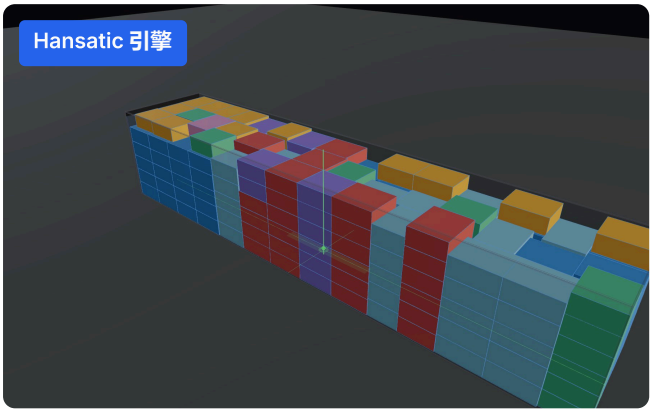
参赛方	状态	已放置	得分	备注
Hansatic引擎	**合法**	**127 / 136**	**12.00 m (39.4 ft)**	
GPT-5.5	**合法**	115 / 136	11.73 m (38.5 ft)	
Opus 5.5	**合法**	124 / 136	11.77 m (38.6 ft)	
Grok 4.7	**合法**	115 / 136	11.87 m (39.0 ft)	
Kimi K3.1 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
DeepSeek V4 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
Gemini 4 Argon	**非法**	121 / 136	12.00 m (39.4 ft)	1 倾斜了必须直立的货物, 12 货物重叠, 6 超出车辆范围, 6 货物悬空
Qwen3p8	未返回方案	-	-	模型在写出方案前达到输出上限。
Magistral Medium	**非法**	51 / 136	12.89 m (42.3 ft)	34 超出车辆范围, 6 货物悬空

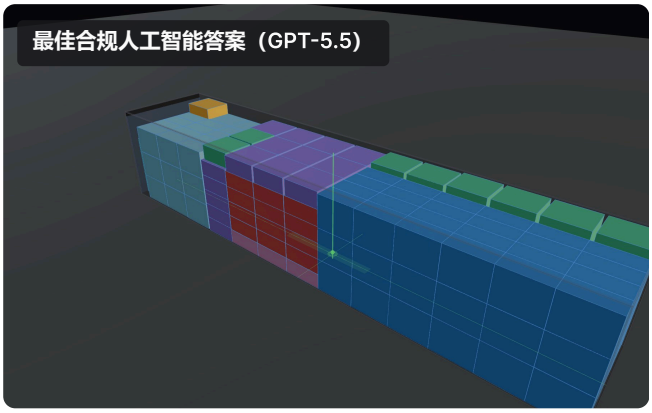
BV6 最短货长

车辆	件数	目标	结果
40' Dry Container	253	最短货长	引擎明显领先

难点

尺寸混杂，无特殊标记



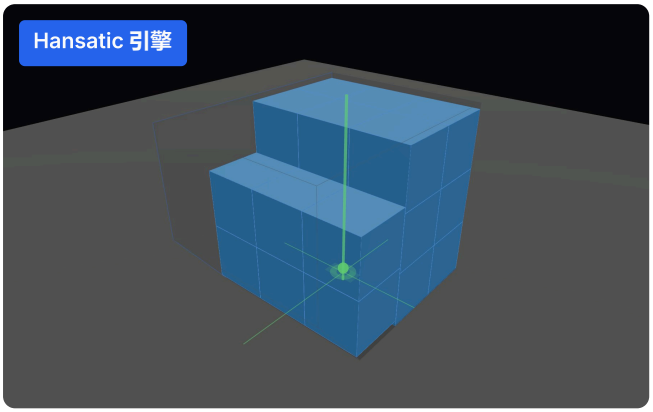


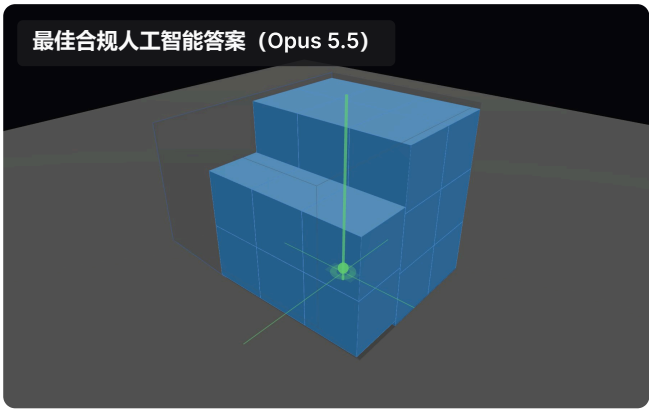
参赛方	状态	已放置	得分	备注
Hansatic引擎	**合法**	**244 / 253**	**11.98 m (39.3 ft)**	
GPT-5.5	**合法**	238 / 253	11.97 m (39.3 ft)	
Opus 5.5	未返回方案	-	-	模型在写出方案前达到输出上限。
Grok 4.7	**合法**	218 / 253	11.98 m (39.3 ft)	
Kimi K3.1 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
DeepSeek V4 Pro	未返回方案	-	-	模型在写出方案前达到输出上限。
Gemini 4 Argon	**非法**	197 / 253	11.50 m (37.7 ft)	13 货物重叠
Qwen3p8	未返回方案	-	-	模型在写出方案前达到输出上限。
Magistral Medium	**非法**	42 / 253	12.18 m (40.0 ft)	3 超出车辆范围

U01 放置最多货件

车辆	件数	目标	结果
AKE (LD3)	50	放置最多货件	持平

难点
斜壁





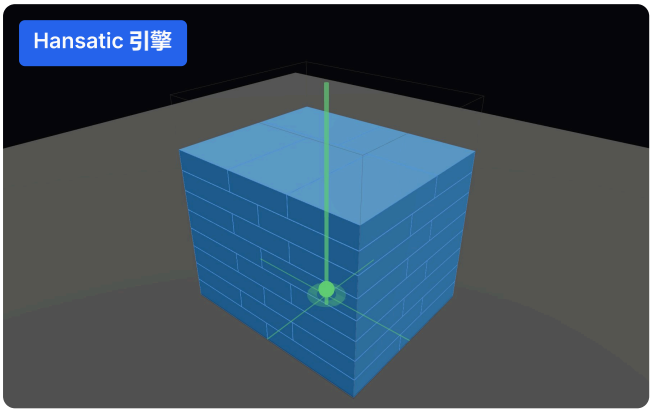
参赛方	状态	已放置	得分	备注
Hansatic引擎	**合法**	**24 / 50**	**24**	
GPT-5.5	**合法**	18 / 50	18	
Opus 5.5	**合法**	24 / 50	24	
Grok 4.7	**合法**	24 / 50	24	
Kimi K3.1 Pro	**合法**	24 / 50	24	
DeepSeek V4 Pro	**合法**	18 / 50	18	
Gemini 4 Argon	**合法**	24 / 50	24	
Qwen3p8	**合法**	24 / 50	24	
Magistral Medium	**非法**	24 / 50	24	3 超出航空集装器轮廓

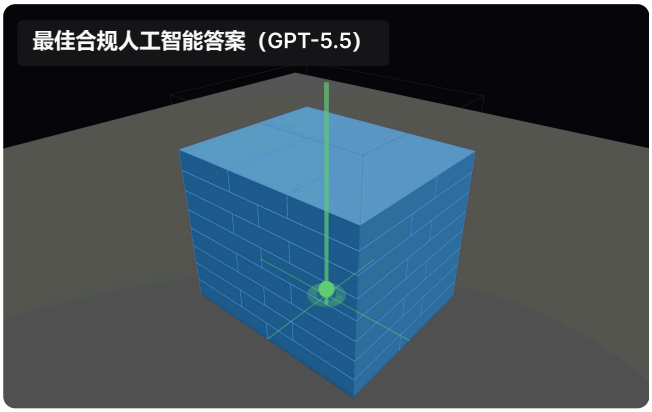
P01 互锁托盘最多

车辆	件数	目标	结果
Industrial / Block 1200×1000	40	互锁托盘最多	持平

难点

40 件需保持直立; 最长 0.60 m (2.0 ft) 的货物





参赛方	状态	已放置	得分	备注
Hansatic引擎	**合法**	**40 / 40**	**100%**	
GPT-5.5	**合法**	40 / 40	100%	
Opus 5.5	**合法**	40 / 40	100%	
Grok 4.7	**合法**	40 / 40	100%	
Kimi K3.1 Pro	**合法**	40 / 40	100%	
DeepSeek V4 Pro	**合法**	40 / 40	100%	
Gemini 4 Argon	**非法**	35 / 40	100%	5 货物尺寸错误, 12 货物悬空
Qwen3p8	**合法**	40 / 40	100%	
Magistral Medium	**非法**	40 / 40	100%	30 货物重叠, 9 超出车辆范围