Hansatic

技術白皮書

神經符號式裝載規劃

一套裝載引擎與八個前沿 AI 模型，實測於真實客戶貨載。

2026年11月

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

搭配頁面，所有方案皆以 3D 呈現：hansatic.com/tw/neuro-symbolic-load-planning

摘要

我們正在將確定性裝載引擎與生成式 AI 結合。在構建組合之前，我們先分別測量了各自的表現。

- **資料：** 20 個真實客戶的裝載規劃案例，26 次評分運行。涵蓋海運貨櫃、簾式貨車、一輛 53 ft (16.2 m) 乾貨廂車、一個 AKE 空運貨櫃，以及一個互鎖棧板堆疊。
- **參賽者：** Hansatic 的生產環境裝載引擎與八個主流前沿大型語言模型。每個模型必須在單一 JSON 回答中為每個品項給出位置與方向，且不得使用任何工具。
- **評審：** 一個確定性驗證器以相同方式為每位參賽者評分。它檢查：
  - 幾何；
  - 95% 底部支撐；
  - 堆疊與易碎限制；
  - 方向標記；
  - 細長堆疊；
  - 載重與軸重；
  - 一個縱向重心區間。

結果：

- **引擎**在 26 次運行中 26 次皆合法。相較於每次運行中最佳的合法模型答案，它領先 18 次（13 次明顯、5 次些微），打平 8 次，零次落後。
- **最強模型** GPT-5.5 在 26 次運行中有 25 次合法，包括多達 253 件貨物的艙單。
- **八個模型中的四個**在不到三分之一的運行中合法。
- **模型擅長什麼。** 最佳模型答案使用了引擎原本沒有的兩種佈局策略。這兩種現在都已成為確定性引擎策略。

我們得出結論：引擎是今日更好的規劃者，而模型創意值得汲取。我們為此設計的架構是神經符號式：

- 模型提出策略；
- 引擎構建每個方案；
- 驗證器只保留合法的改進。

我們發布這些案例、驗證器、提示詞與評分程式碼。

1. 基準測試設計

1.1 案例

每個案例都是 Hansatic 應用程式中的真實客戶貨載，1:1 複製：尺寸、重量、數量、標記與車輛。品號已替換，貨物名稱已移除。

組別	案例	執行次數	車輛	件數
標準（H01-H12）	12	16	一輛 20'、兩輛 40' 乾貨櫃、四輛 40' HC、一輛 45' HC、四輛 13.6 m (44.6 ft) 簾式拖車	14-83
大型（BV1-BV6）	6	8	三輛 40' HC、一輛 40' 乾貨櫃、一輛 13.6 m (44.6 ft) 簾式拖車、一輛 53 ft (16.2 m) 乾式廂型車	103-253
特殊（U01、P01）	2	2	具斜壁的 AKE (LD3)；1200 × 1000 mm (47 × 39 in) 棧板	50 和 40

執行次數與目標：

- **貨車**以兩項指標評分：裝載公尺（LDM）與棧板位置。因此 12 個標準案例產生 16 次執行，6 個大型案例產生 8 次。
- **BV5 與 BV6** 無法完整裝入，主要目標是裝入更多品項。
- **U01 與 P01** 各有已知最佳解：
  - U01：50 個立方體裝入 24 個，因斜壁無法再容納更多；
  - P01：40 袋全數裝入，交錯達 100%。

案例中的限制條件：

- 不可堆疊與僅能置頂（易碎）品項；
- 每件貨品的堆疊重量上限；
- 此面向上與不可傾斜標記；
- 長桿（BV1，長度可達貨櫃全長）；
- 單一主導重件（H02）；
- 棧板與紙箱混裝貨載（BV3）。

1.2 驗證器

驗證器（v2.3.0）以相同方式檢查每份規劃。只有所有硬性檢查都通過，規劃才算合法。

檢查項目	規則
邊界	每件貨品都在貨物空間內，包括 AKE 斜面
重疊	任兩件貨品不共用體積

檢查項目	規則
支撐	每件貨品底部至少 95% 置於地板或下方貨品上
堆疊	不可堆疊品項不承載任何重量；僅能置頂品項不承載重量且位於頂部；每件堆疊重量上限必須成立
方向	遵守此面向上與不可傾斜標記
細長堆疊	對獨立直立堆疊設定高度與底面積比例上限
載重	總質量在車輛限制內
車軸（貨車）	前、後軸重均在限制內，依各品項質心計算
CoG（貨櫃）	縱向重心位於允許範圍內
多停點	案例有停點時，依停點順序後進先出

傾覆風險會計算並回報，但不計分。

1.3 排名

每個執行次數適用以下規則：

1. 合法規劃優於非法規劃。
2. 裝入品項較多者勝。
3. 由該次執行的目標決定：
    - 裝載更短（LDM）；
    - 或棧板位置更少；
    - 或 P01 案例取交錯比例較高者，其次取高度較低者。
4. 棧板位置平手時以 LDM 決勝。

棧板位置由同一個計數器為每位參賽者計算：Hansatic 的生產環境計數器。

勝出幅度。 勝出的定義如下：

- **明確**勝出：LDM 至少 2%、至少多一個棧板位置、或至少多裝入一件；
- 其餘為**些微**勝出。

引擎在時間預算內執行，因此單次執行差距若低於 2%，可能只是逐次執行的變異（見 6.2 節）。

1.4 參賽者

模型。 八家實驗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)

設定：

- **執行次數：** 每個案例一次。
- **輸出上限：** 64k tokens，或供應商較低的上限（Gemini：32k）。
- **批次處理：** 供應商有提供時採用其批次 API。
- **答案蒐集日期：** 2026-11-11。
- **總成本：** 約 $64。

引擎。 Hansatic 的生產環境打包器，版本為 2026-11-20（commit 9f66537c）。設定：

- 採用生產環境設定：4 個 worker 與生產環境時間預算，固定於單一 CPU 核心；
- 每個案例執行一次，不取 best-of；
- 依自身 60% 支撐規則打包，並與所有模型一樣以 95% 受評。

1.5 提示詞

- 每種車輛類型各一個提示詞：貨櫃、貨車、ULD 與棧板。
- 每個提示詞都說明驗證器對該類型適用的確切規則、目標，以及輸出結構（品號、位置、方向）。
- 不提供已解出的配置作為範例。

完整範本見附錄 A。

2. 結果

2.1 各模型，全部 26 次運行

模型	合法	與引擎持平	落後引擎
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

沒有任何模型在任何一次運行中勝過引擎。

標準組（16 次運行）。

模型	合法	持平	落後
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

大型組（8 次運行）。 沒有任何模型在任何大型運行中與引擎持平。

模型	合法
GPT-5.5	8
Grok 4.7	8
Opus 5.5	5

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

特殊組（2 次運行）。

模型	合法	達到最佳解
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 vs 最佳合法模型答案

組別	明顯領先	小幅領先	持平	落後
標準（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**

標準組。 每個單元格先列出放置件數，接著是裝載長度或棧板位置。

運行	引擎	最佳合法模型	差距
H01	26/27，11.69 m (38.4 ft)	Opus 26，11.74 m (38.5 ft)	小幅（0.4%）
H02	24/29，5.78 m (19.0 ft)	GPT 18，5.76 m (18.9 ft)	明顯（+6 件）
H03	31/32，10.51 m (34.5 ft)	無合法答案	明顯
H04	39，10.44 m (34.3 ft)	Opus 39，10.44 m (34.3 ft)	持平

運行	引擎	最佳合法模型	差距
H05	49，12.40 m (40.7 ft)	Opus 49，12.40 m (40.7 ft)	持平
H06	83，10.65 m (35.0 ft)	Grok 83，11.41 m (37.4 ft)	明顯（6.6%）
H07	45，9.60 m (31.5 ft)	GPT 45，9.60 m (31.5 ft)	持平
H08	52，5.92 m (19.4 ft)	Kimi 52，6.12 m (20.1 ft)	明顯（3.3%）
H09 LDM	29/30，13.36 m (43.8 ft)	GPT 29，13.56 m (44.5 ft)	小幅（1.5%）
H09 棧板位	29/30，29 個位置	GPT 29，29 個位置	小幅（LDM 破平手）
H10 LDM	26，12.20 m (40.0 ft)	DeepSeek，相同	持平
H10 棧板位	26，31 個位置	DeepSeek，相同	持平
H11 LDM	23，12.00 m (39.4 ft)	Opus 23，12.20 m (40.0 ft)	小幅（1.6%）
H11 棧板位	23，30 個位置	Kimi 23，32 個位置	明顯（2 個位置）
H12 LDM	13/14，11.00 m (36.1 ft)	Gemini 13，11.00 m (36.1 ft)	持平
H12 棧板位	13/14，25 個位置	Kimi 13，28 個位置	明顯（3 個位置）

大型組。

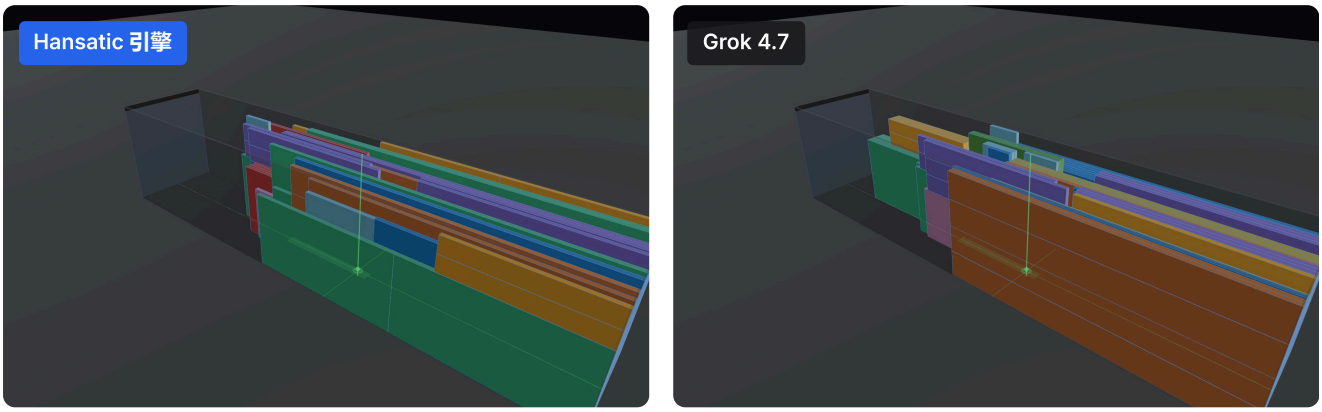
運行	引擎	最佳合法模型	差距
BV1（40' 高櫃，103 根）	103，9.17 m (30.1 ft)	Grok 103，10.66 m (35.0 ft)	明顯（14.0%）
BV2 LDM（簾式車，153）	153，11.52 m (37.8 ft)	Opus 153，11.78 m (38.6 ft)	明顯（2.2%）
BV2 棧板位	153，29 個位置	Opus 153，29 個位置	小幅（LDM 破平手）
BV3 LDM（53 英尺廂式車，119）	119，6.83 m (22.4 ft)	Opus 119，6.99 m (22.9 ft)	明顯（2.2%）
BV3 棧板位	119，17 個位置	Opus 119，18 個位置	明顯（1 個位置）
BV4（145，僅頂層放易碎品）	145，8.57 m (28.1 ft)	Grok 145，8.85 m (29.0 ft)	明顯（3.1%）
BV5（136，僅限直立放置）	127，12.00 m (39.4 ft)	Opus 124，11.77 m (38.6 ft)	明顯（+3 件）
BV6（253，95% 體積）	244，11.98 m (39.3 ft)	GPT 238，11.97 m (39.3 ft)	明顯（+6 件）

特殊組。

運行	引擎	模型
U01（AKE）	24/50，最佳解	Opus、Grok、Kimi、Qwen 和 Gemini 也達到 24；GPT 和 DeepSeek達到 18
P01（棧板）	40/40，100% 交錯，最佳解	每個合法模型答案都在最佳解達到 40/40

2.3 這些差距在實際貨載上的意義

- **BV1：** 短 1.49 m (4.9 ft)，約為 40 英尺貨櫃長度的 12%。
- **H12 棧板位：** 少 3 個棧板位置，約為 13.6 m (44.6 ft) 簾式車車板的 9%。
- **BV6：** 在無法完全裝入的貨載上多放 6 件，因此留給另一輛車的貨物就少 6 件。



BV1：Hansatic 引擎 9.17 m (30.1 ft)，左側；最佳合法模型答案（Grok 4.7）10.66 m (35.0 ft)，右側。

2.4 成本與時間

	引擎	模型
每次規劃時間	可忽略	未比較：批次 API 會排隊請求
每次規劃成本	可忽略	$0.10-0.80

3. 失敗分析

208 次模型運行（8 個模型 x 26 次）中，95 份計畫合法。其餘 113 份以三種不同方式失敗。

模型	合法	無回答（輸出上限）	供應商錯誤	違反規則的計畫
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 無回答：模型思考空間用盡

65 次運行中，模型推理到輸出上限，始終沒有寫出計畫。另有一份回答（Magistral，H03）正常結束，但不是有效的 JSON；計入此處。

- 這是最常見的失敗模式，八個模型中有六個都遇到。
- Kimi、DeepSeek 和 Qwen 大部分的運行都因此失敗。它們的合法計畫多半出現在較小的案例。
- Opus 有 6 次運行因此失敗，包括 BV1、BV4 和 BV6——三份最大的貨物清單。

對生產環境中的規劃器而言，沒有回答和非法回答一樣都是失敗：月台仍然需要一份計畫。

3.2 供應商錯誤

Grok 在 4 次運行中回傳 HTTP 403，原因是供應商的支出上限已達。這些運行未重新發送，計為失敗。

3.3 違反規則的計畫

44 份回答是完整計畫，但至少違反一條硬性規則。幾乎都出自兩個模型：Gemini（19 份）和 Magistral（23 份）。Grok 有 2 份。

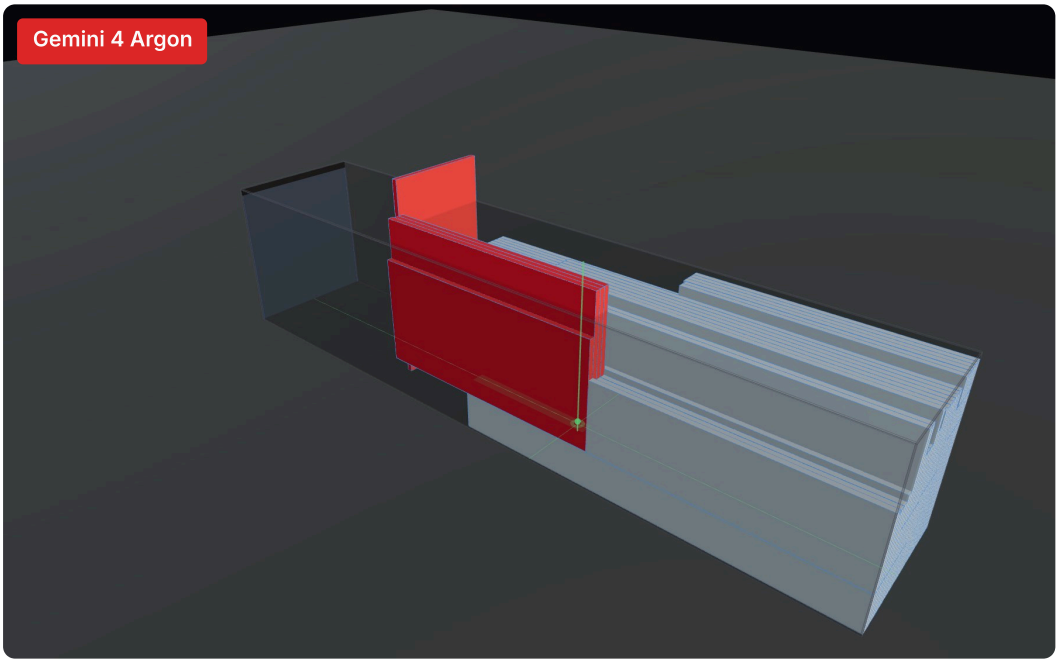
這 44 份計畫中，各類違規至少出現一次的計畫數：

違反的規則	計畫數
物品超出貨物空間	27
物品重疊	27

違反的規則	計畫數
懸空物品（支撐度低於 95%）	21
物品尺寸不符任何允許方向	17
裝載於不可堆疊或易碎物品上，或超過堆疊上限	10
堆疊超出車頂	9
重心超出範圍	7
軸重超載	4
不可傾斜物品被傾倒	1
超出 AKE 輪廓	1

Grok 的兩份非法計畫各違反一條規則：一次是物品被壓壞，另一次是重心問題。

這些計畫在 3D 視圖中看起來合理。錯誤往往只是幾件物品懸空或重疊幾公分。只有完整的幾何檢查才能發現。



BV1，Gemini 4 Argon：已放置 103 件中的 77 件。標示處：懸空、重疊或超出貨物空間的物品。

3.4 合法，但仍落後

- **不完整。** 在 BV5 和 BV6 上，最佳合法模型回答分別比引擎少放 3 件和 6 件物品。
- **更長。** 在所有物品都能放入的運行中，24 次貨車與貨櫃運行有 13 次，最佳合法模型回答比引擎的長。最大差距：
  - BV1：1.49 m (4.9 ft)；
  - H06：0.76 m (2.5 ft)。

3.5 模型的優點

最強的回答運用了資深裝載人員會用的策略：

- **重型箱子緊貼前壁。**
- **利用不可堆疊物品上方的高度。** 在 BV3 上，Opus 將紙箱側放，做成 559 mm (22 in) 高的底板，並在上面立起 19 台不可堆疊的家電。

引擎原本沒有這兩種策略。現在兩者都已成為確定性引擎策略。這正是神經符號迴路設計要捕捉的創意（第 5 節）。

4. 工具臂

在第二臂中，模型可透過單一工具 `pack_cargo` 呼叫引擎，然後以任何方案作答。

多數模型會將貨物清單複製到工具呼叫中，並不加修改地回傳引擎的方案。失敗原因都是複製錯誤：

- Magistral 送出 100 件尺寸或標記錯誤的貨物；
- Kimi 在 H10 上呼叫錯誤；
- Qwen 在 4 次執行中失敗。

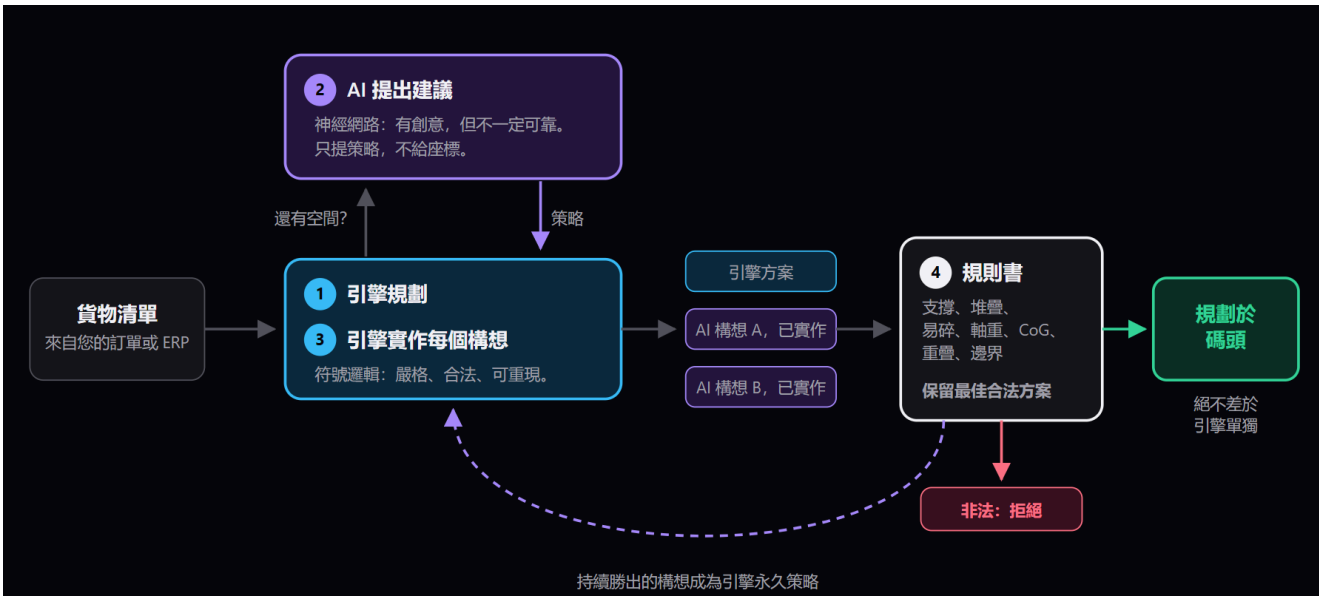
結論： 在任何組合系統中，引擎必須直接從來源系統讀取貨物清單，絕不透過模型。

5. 神經符號架構

5.1 兩種智慧

	裝載引擎（符號式）	生成式 AI（神經式）
規則	每次都遵守所有已編碼規則	違反給定的規則
可重複性	相同貨物，相同計畫	相同貨物可能得到不同計畫
可稽核性	每個放置都有理由	難以稽核
範圍	僅限於已有策略	發現新策略
規模	可處理 253 件	許多模型在大型艙單上無法給出答案

5.2 迴圈



1. **基準線。** 引擎直接從來源系統讀取艙單，據以規劃。
2. **觸發。** 只有當基準線留有空間時才呼叫模型：有未放置貨物、裝載率低、或未達淨空目標。
3. **提案。** 模型只回傳策略參數，絕不回傳座標：
   - 區域；
   - 排序；
   - 方向偏好；
   - 底部與頂部結構。
4. **執行。** 引擎為每個提案建立一個計畫。
5. **判定。** 驗證器保留最佳合法計畫。基準線始終是候選方案。
6. **蒸餾。** 重複獲勝的提案會成為引擎的永久策略，如前牆策略與紙箱底層策略便是如此。

5.3 結構保證的特性

- **絕不比單獨使用引擎差，** 因為基準線始終是候選方案。
- **沒有非法計畫會送到碼頭，** 因為驗證器會判定每個候選方案。
- **沒有轉錄風險，** 因為模型絕不接觸艙單或座標（第 4 節）。
- **不會因模型輸出限制而失敗，** 因為提案只是幾個參數，而不是一份 253 件的計畫（第 3.1 節）。
- **AI 成本有上限，** 因為模型只在觸發時執行。

我們正在建立這個迴圈。在下一版中，它將作為獨立參賽者接受評分，評估每個計畫的成本與時間。

6. 範圍與限制

6.1 驗證器範圍

「合法」指符合 v2.3.0 版所編碼之規則。驗證器不檢查裝載順序，多站點的後進先出除外。

6.2 差異

每位參賽者每個案例執行一次。

- **引擎**在時間預算內競速各種策略，因此機器負載可能改變哪個策略先完成。在 H02 上，我們測得各次執行結果為 5.63-5.79 m (18.5-19.0 ft)。
- **模型**僅取樣一次。

單次執行差異低於 2% 者，報告為差距小。

6.3 模型層級與輸出限制

- 我們執行了每個實驗室的主流前沿模型。高階版本（Anthropic Fable、OpenAI GPT-6 Astra）未執行。
- 更大的輸出限制可能將部分無答案的執行（第 3.1 節）轉為規劃。

6.4 案例來源

所有案例皆來自單一平台的使用者。其他貨主的貨物組合可能不同。

7. 下一步

7.1 實務規則

接下來有兩條實務規則，分別用於引擎與驗證器 v3：

- **車頂淨空：** 車頂下方至少 50 mm (2.0 in)，或重量超過 25 kg (55 lb) 的堆疊貨物上方至少 100 mm (3.9 in)。
- **側向間隙：** 每個堆高機排。

我們用車頂規則重新評分每個參賽者。無法傾斜且高度超過車頂高度減去 50 mm (2.0 in) 的地板貨物可豁免，因為案例本身強制如此。

- **引擎**在 16 個標準運行中的 9 個以及 8 個大型運行中的 1 個上符合此規則。
- **在 6 個運行中，某個模型答案符合此規則**，且方案與引擎接近。這些是引擎的首要目標：
  - H02 (GPT);
  - H06 (Grok, 11.41 m / 37.4 ft);
  - H12 LDM (Gemini, 11.00 m / 36.1 ft);

- H12 棧板 (Kimi, 28 positions);
- BV3 LDM (GPT, 9.14 m / 30.0 ft);
- BV5 (Opus, 124 已放置)。

- **在 8 個運行中，沒有參賽者符合此規則：** H01、H03、H08、BV1、兩個 BV2 運行、BV3 棧板、BV6。

7.2 下一版

- 驗證器 v3，包含車頂淨空、側向間隙與裝載順序檢查，並套用於每個參賽者。
- 新的真實客戶案例。
- 每個實驗室能力最強的模型，包含高階方案，每個模型多次運行並回報方差。
- 神經符號迴路作為參賽者。

8. 發布

已發布：

- 每個案例全文；
- 驗證器；
- 提示詞；
- 評分與排名程式碼；
- 每次運行：發送的請求、供應商未經修改的回應，以及解析後的規劃。

檢視器： 每位參賽者的每份規劃皆以 3D 呈現，旁邊附上其提示詞與原始回應：hansatic.com/tw/neuro-symbolic-load-planning。

未公開：

- 棧板位置計數器（Hansatic 生產環境計數器，以相同方式套用於每位參賽者）；
- 引擎；
- 來源規劃的 ID。

附錄 A：提示詞模板

每種車型一個系統提示詞，依原樣發送。使用者訊息先給出目標，接著以 JSON 格式提供案例；此處縮短品項清單。提示詞以英文呈現，即模型收到的語言。

海運貨櫃（BV1，40' 高櫃）

系統提示

```
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"]}
```

使用者訊息（品項清單已縮短）

```
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"
 ]
}
```


貨車（H09，13.6 m (44.6 ft) 簾式貨車）

系統提示

```
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"]}
```

使用者訊息（品項清單已縮短）

```
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"
```

```
 ]
}
```

空運貨櫃（U01，AKE 斜壁）

系統提示

```
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"]}
```

使用者訊息（品項清單已縮短）

```
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"
 ]
}
```


棧板（P01，1200 × 1000 mm (47 × 39 in)）

系統提示

```
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"]}
```

使用者訊息（品項清單已縮短）

```
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"
```

```
  ]
}
```

附錄 B。全部 26 次執行

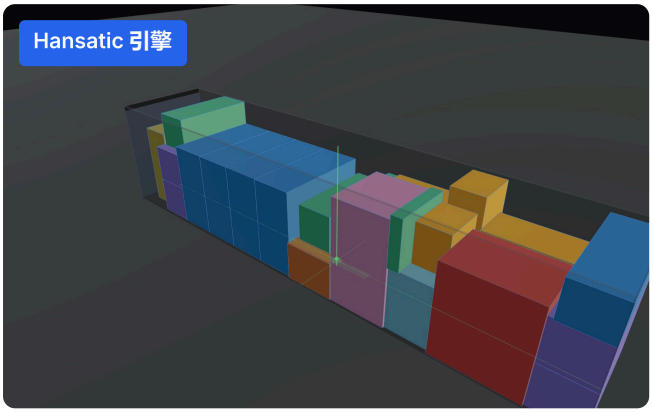
每次執行一頁：裝載內容、引擎方案與最佳合法模型答案並列，以及每位參賽者的結果。

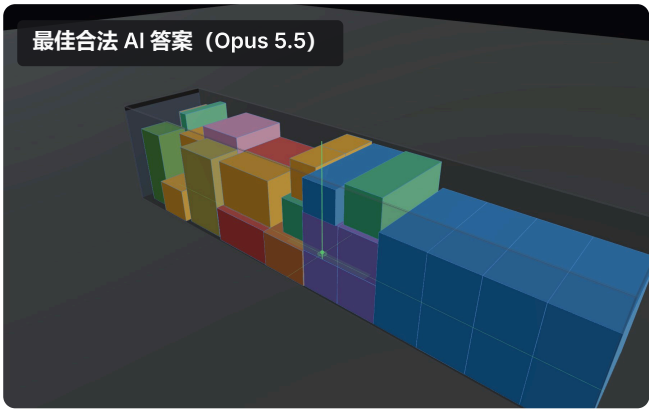
H01 最短裝載長度

車輛	件數	目標	結果
40' High Cube	27	最短裝載長度	引擎些微領先

難點

27 個必須直立放置的品項





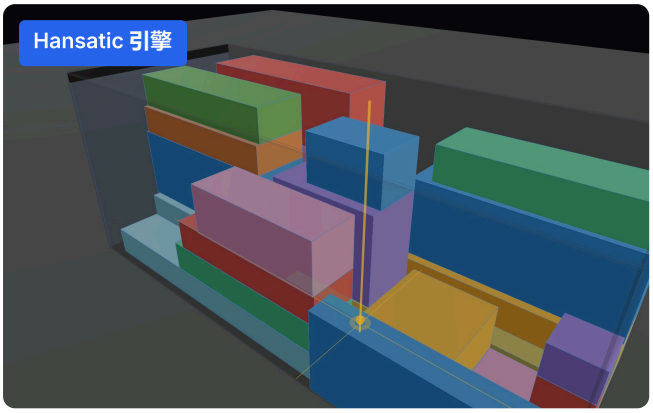
參賽者	狀態	已放置	分數	備註
Hansatic 引擎	**合法**	**26 / 27**	**11.69 m (38.4 ft)**	
GPT-5.5	**合法**	25 / 27	11.88 m (39.0 ft)	
Opus 5.5	**合法**	26 / 27	11.74 m (38.5 ft)	
Grok 4.7	**合法**	25 / 27	11.37 m (37.3 ft)	
Kimi K3.1 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
DeepSeek V4 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
Gemini 4 Argon	**非法**	22 / 27	11.56 m (37.9 ft)	1 貨物尺寸錯誤, 1 超出車輛範圍, 1 懸空貨物
Qwen3p8	未回傳方案	-	-	模型寫出方案前達輸出上限。
Magistral Medium	**非法**	27 / 27	12.09 m (39.7 ft)	7 貨物重疊, 14 超出車輛範圍, 3 超過高度限制, 1 懸空貨物, 1 重心偏移

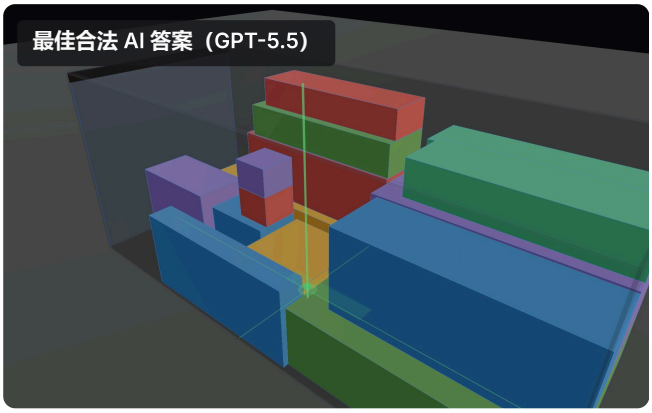
H02 最短裝載長度

車輛	件數	目標	結果
20' Dry Container	29	最短裝載長度	引擎明顯領先

難點

29 個必須直立放置的品項; 品項最長達 5.00 m (16.4 ft); 最重品項 1,550 kg (3,417 lb)





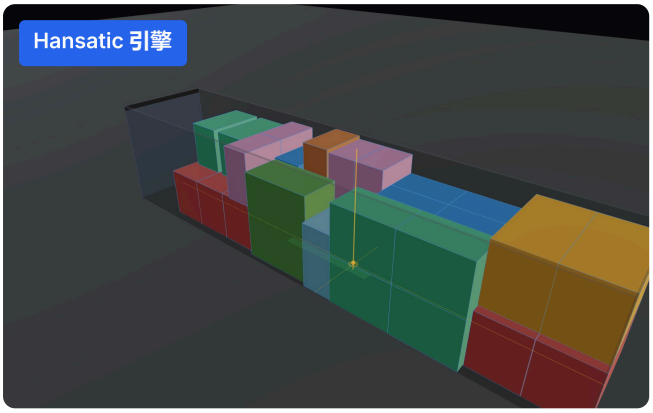
參賽者	狀態	已放置	分數	備註
Hansatic 引擎	**合法**	**24 / 29**	**5.78 m (19.0 ft)**	
GPT-5.5	**合法**	18 / 29	5.76 m (18.9 ft)	
Opus 5.5	未回傳方案	-	-	模型寫出方案前達輸出上限。
Grok 4.7	供應商錯誤	-	-	供應商拒絕請求。
Kimi K3.1 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
DeepSeek V4 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
Gemini 4 Argon	**非法**	23 / 29	5.00 m (16.4 ft)	6 貨物尺寸錯誤, 13 貨物重疊, 9 懸空貨物, 9 貨物壓損
Qwen3p8	未回傳方案	-	-	模型寫出方案前達輸出上限。
Magistral Medium	**非法**	29 / 29	7.97 m (26.1 ft)	37 貨物重疊, 23 超出車輛範圍, 1 超過高度限制, 7 懸空貨物, 5 貨物壓損, 2 重心偏移

H03 最短裝載長度

車輛	件數	目標	結果
40' High Cube	32	最短裝載長度	引擎明顯領先

難點

25 個必須直立放置的品項; 最重品項 2,680 kg (5,908 lb)



本次運行中，沒有 AI 模型回傳合法方案。

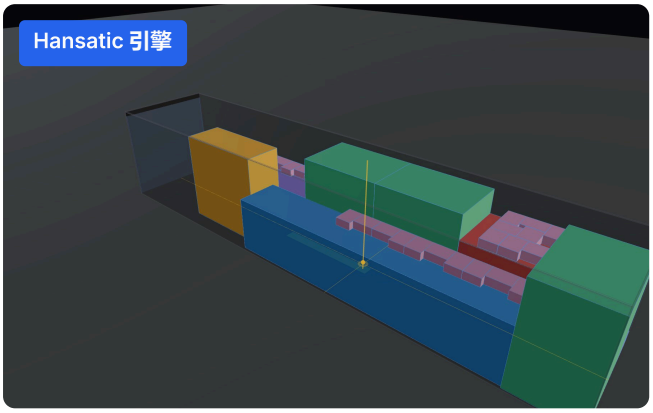
參賽者	狀態	已放置	分數	備註
Hansatic 引擎	**合法**	**31 / 32**	**10.51 m (34.5 ft)**	
GPT-5.5	未回傳方案	-	-	模型寫出方案前達輸出上限。
Opus 5.5	未回傳方案	-	-	模型寫出方案前達輸出上限。
Grok 4.7	**非法**	31 / 32	12.03 m (39.5 ft)	1 貨物壓損
Kimi K3.1 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
DeepSeek V4 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
Gemini 4 Argon	**非法**	31 / 32	11.81 m (38.7 ft)	1 貨物尺寸錯誤, 12 貨物重疊, 6 超出車輛範圍, 3 超過高度限制, 2 懸空貨物, 3 貨物壓損
Qwen3p8	未回傳方案	-	-	模型寫出方案前達輸出上限。
Magistral Medium	未回傳方案	-	-	答案不含有效方案。

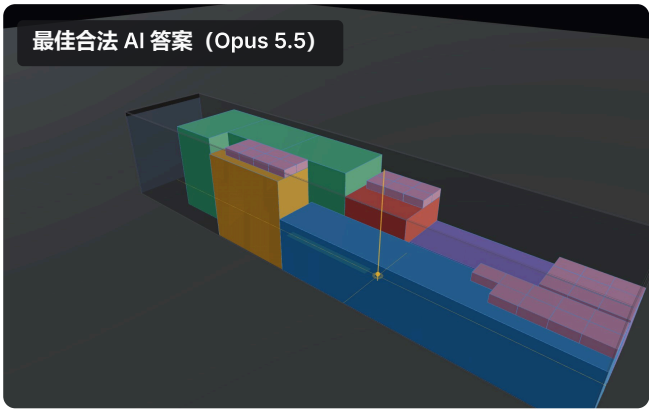
H04 最短裝載長度

車輛	件數	目標	結果
40' Dry Container	39	最短裝載長度	平手

難點

品項最長達 6.30 m (20.7 ft); 最重品項 998 kg (2,200 lb)





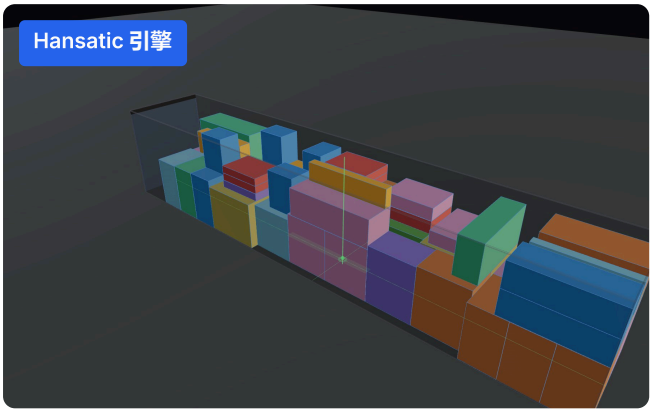
參賽者	狀態	已放置	分數	備註
Hansatic 引擎	**合法**	**39 / 39**	**10.44 m (34.3 ft)**	
GPT-5.5	**合法**	39 / 39	10.72 m (35.2 ft)	
Opus 5.5	**合法**	39 / 39	10.44 m (34.3 ft)	
Grok 4.7	**合法**	39 / 39	11.10 m (36.4 ft)	
Kimi K3.1 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
DeepSeek V4 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
Gemini 4 Argon	**非法**	29 / 39	6.30 m (20.7 ft)	10 貨物尺寸錯誤, 6 懸空貨物
Qwen3p8	未回傳方案	-	-	模型寫出方案前達輸出上限。
Magistral Medium	**非法**	38 / 39	10.20 m (33.5 ft)	1 貨物尺寸錯誤, 27 貨物重疊, 6 超出車輛範圍, 24 懸空貨物, 1 貨物壓損, 1 重心偏移

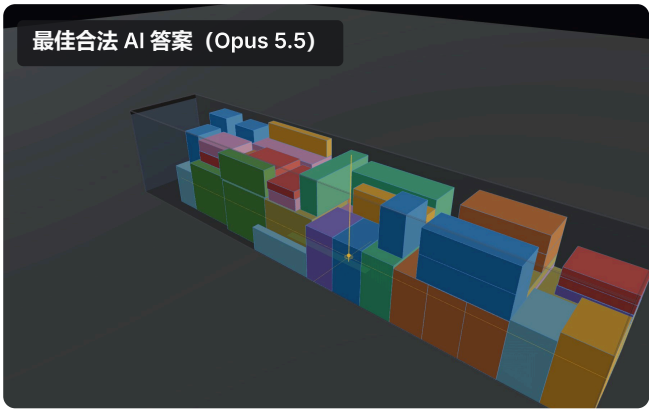
H05 最短裝載長度

車輛	件數	目標	結果
45' High Cube	49	最短裝載長度	平手

難點

49 個必須直立放置的品項; 27 個有堆重限制的品項; 最重品項 814 kg (1,795 lb)





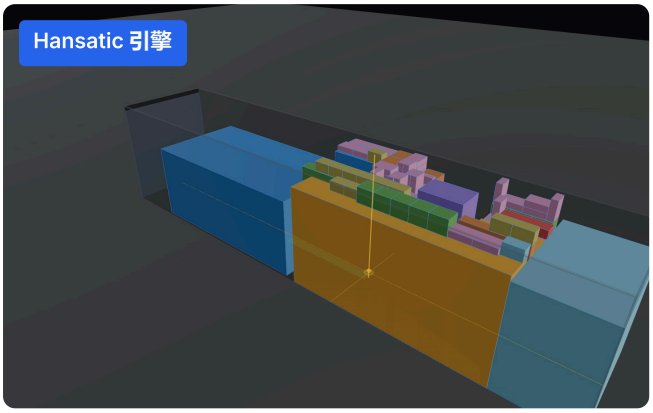
參賽者	狀態	已放置	分數	備註
Hansatic 引擎	**合法**	**49 / 49**	**12.40 m (40.7 ft)**	
GPT-5.5	**合法**	45 / 49	13.28 m (43.6 ft)	
Opus 5.5	**合法**	49 / 49	12.40 m (40.7 ft)	
Grok 4.7	**合法**	43 / 49	12.96 m (42.5 ft)	
Kimi K3.1 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
DeepSeek V4 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
Gemini 4 Argon	**非法**	37 / 49	10.49 m (34.4 ft)	11 貨物尺寸錯誤, 2 貨物重疊, 5 超出車輛範圍, 2 超過高度限制, 3 懸空貨物, 2 貨物壓損
Qwen3p8	未回傳方案	-	-	模型寫出方案前達輸出上限。
Magistral Medium	**非法**	47 / 49	12.73 m (41.8 ft)	2 貨物尺寸錯誤, 31 貨物重疊, 9 超出車輛範圍, 14 超過高度限制, 21 懸空貨物, 2 貨物壓損

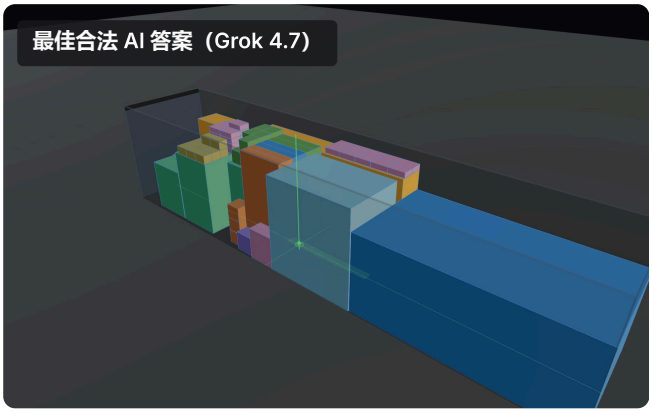
H06 最短裝載長度

車輛	件數	目標	結果
40' High Cube	83	最短裝載長度	引擎明顯領先

難點

5 個不可堆疊品項; 1 個必須直立放置的品項; 最重品項 3,020 kg (6,658 lb)





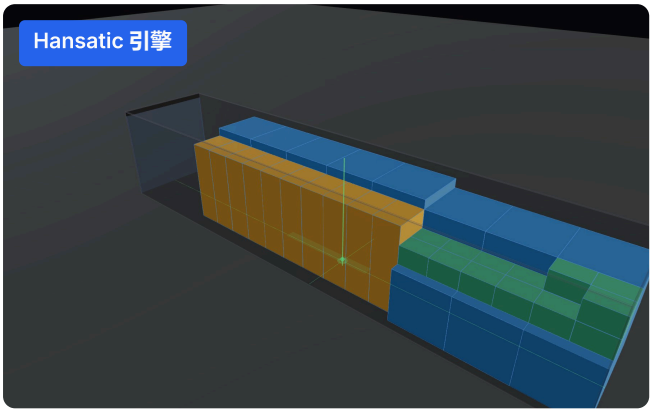
參賽者	狀態	已放置	分數	備註
Hansatic 引擎	**合法**	**83 / 83**	**10.65 m (35.0 ft)**	
GPT-5.5	**合法**	83 / 83	11.78 m (38.7 ft)	
Opus 5.5	未回傳方案	-	-	模型寫出方案前達輸出上限。
Grok 4.7	**合法**	83 / 83	11.41 m (37.4 ft)	
Kimi K3.1 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
DeepSeek V4 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
Gemini 4 Argon	**非法**	7 / 83	6.89 m (22.6 ft)	2 貨物尺寸錯誤, 1 重心偏移
Qwen3p8	未回傳方案	-	-	模型寫出方案前達輸出上限。
Magistral Medium	**非法**	20 / 83	12.25 m (40.2 ft)	2 貨物重疊, 4 超出車輛範圍, 1 超過高度限制, 14 懸空貨物, 1 重心偏移

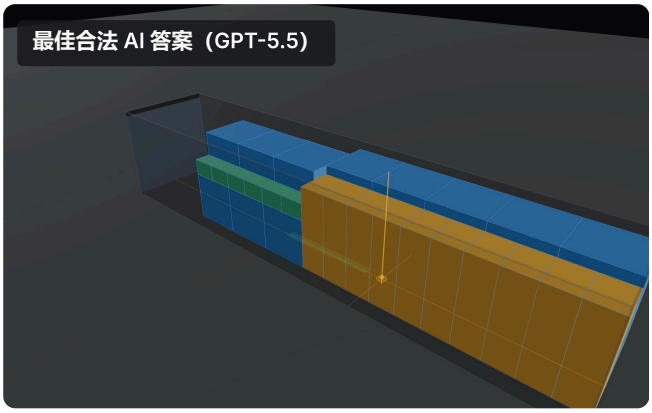
H07 最短裝載長度

車輛	件數	目標	結果
40' Dry Container	45	最短裝載長度	平手

難點

10 個不可堆疊品項; 15 個必須直立放置的品項; 15 個有堆重限制的品項; 最重品項 1,200 kg (2,646 lb)





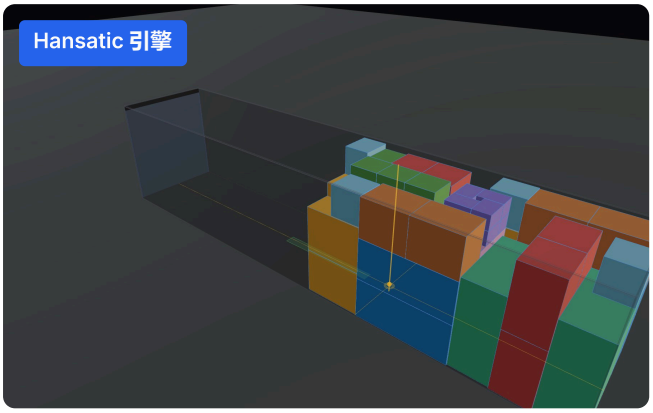
參賽者	狀態	已放置	分數	備註
Hansatic 引擎	**合法**	**45 / 45**	**9.60 m (31.5 ft)**	
GPT-5.5	**合法**	45 / 45	9.60 m (31.5 ft)	
Opus 5.5	**合法**	45 / 45	9.61 m (31.5 ft)	
Grok 4.7	**非法**	45 / 45	10.80 m (35.4 ft)	1 重心偏移
Kimi K3.1 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
DeepSeek V4 Pro	**合法**	45 / 45	12.00 m (39.4 ft)	
Gemini 4 Argon	**非法**	32 / 45	7.60 m (24.9 ft)	13 貨物尺寸錯誤, 1 貨物重疊, 16 懸空貨物
Qwen3p8	未回傳方案	-	-	模型寫出方案前達輸出上限。
Magistral Medium	**非法**	10 / 45	12.00 m (39.4 ft)	1 重心偏移

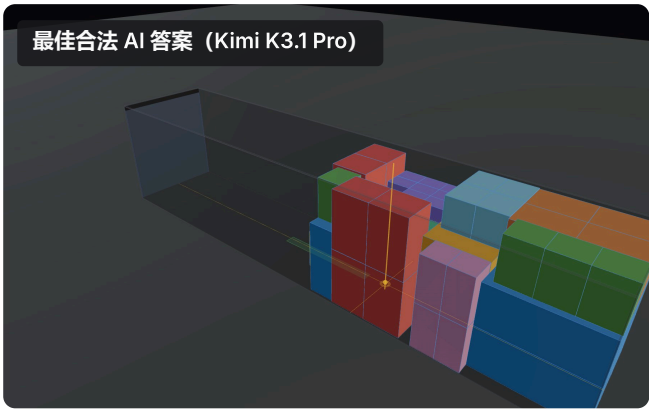
H08 最短裝載長度

車輛	件數	目標	結果
40' High Cube	52	最短裝載長度	引擎明顯領先

難點

52 個必須直立放置的品項





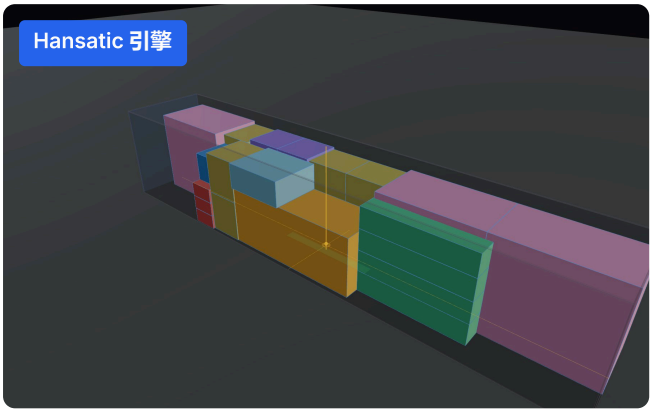
參賽者	狀態	已放置	分數	備註
Hansatic 引擎	**合法**	**52 / 52**	**5.92 m (19.4 ft)**	
GPT-5.5	**合法**	52 / 52	6.26 m (20.5 ft)	
Opus 5.5	**合法**	52 / 52	6.20 m (20.3 ft)	
Grok 4.7	**合法**	52 / 52	6.18 m (20.3 ft)	
Kimi K3.1 Pro	**合法**	52 / 52	6.12 m (20.1 ft)	
DeepSeek V4 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
Gemini 4 Argon	**非法**	51 / 52	6.51 m (21.4 ft)	1 貨物尺寸錯誤, 4 貨物重疊, 1 懸空貨物
Qwen3p8	未回傳方案	-	-	模型寫出方案前達輸出上限。
Magistral Medium	**非法**	52 / 52	11.25 m (36.9 ft)	14 貨物重疊, 20 超出車輛範圍

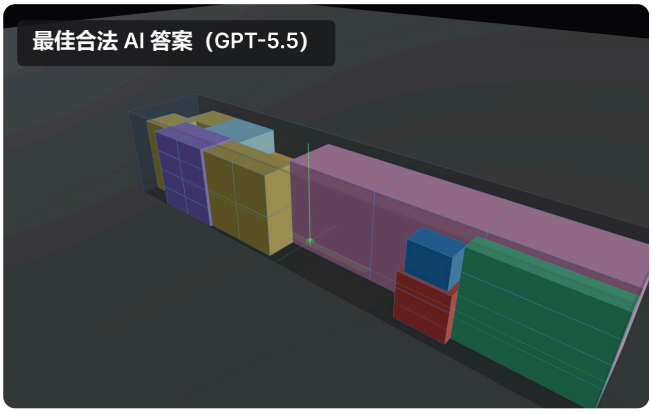
H09 最短裝載長度

車輛	件數	目標	結果
13.6m Curtainsider	30	最短裝載長度	引擎些微領先

難點

30 個必須直立放置的品項; 最重品項 2,400 kg (5,291 lb)





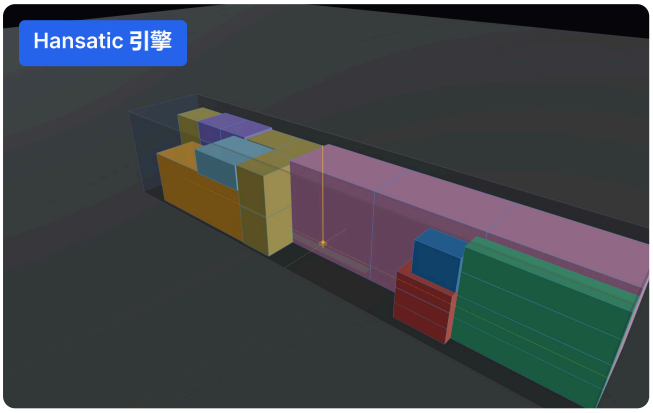
參賽者	狀態	已放置	分數	備註
Hansatic 引擎	**合法**	**29 / 30**	**13.36 m (43.8 ft)**	
GPT-5.5	**合法**	29 / 30	13.56 m (44.5 ft)	
Opus 5.5	**合法**	29 / 30	13.60 m (44.6 ft)	
Grok 4.7	**合法**	29 / 30	13.60 m (44.6 ft)	
Kimi K3.1 Pro	**合法**	29 / 30	13.60 m (44.6 ft)	
DeepSeek V4 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
Gemini 4 Argon	**非法**	19 / 30	11.60 m (38.1 ft)	11 貨物尺寸錯誤, 7 貨物重疊
Qwen3p8	未回傳方案	-	-	模型寫出方案前達輸出上限。
Magistral Medium	**非法**	30 / 30	13.36 m (43.8 ft)	17 貨物重疊, 25 超出車輛範圍, 7 懸空貨物, 1 軸重超載

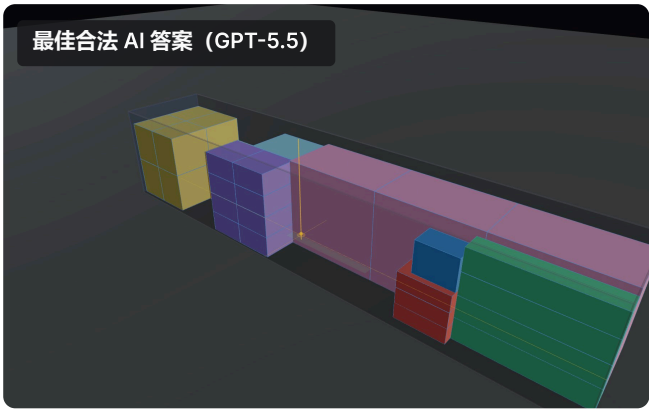
H09 最少棧板位置

車輛	件數	目標	結果
13.6m Curtainsider	30	最少棧板位置	引擎些微領先

難點

30 個必須直立放置的品項; 最重品項 2,400 kg (5,291 lb)





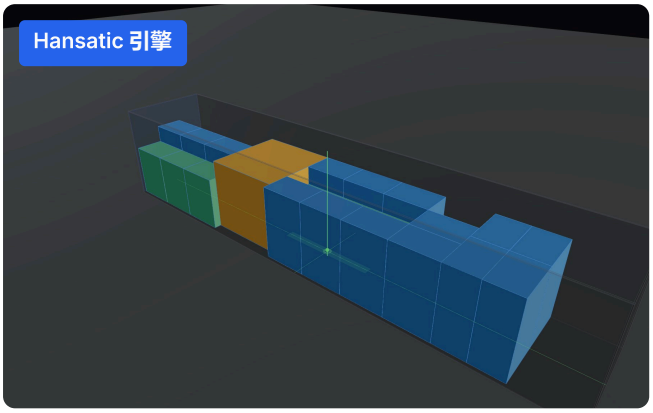
參賽者	狀態	已放置	分數	備註
Hansatic 引擎	**合法**	**29 / 30**	**29 個棧板位, 13.36 m (43.8 ft)**	
GPT-5.5	**合法**	29 / 30	29 個棧板位, 13.60 m (44.6 ft)	
Opus 5.5	**合法**	29 / 30	31 個棧板位, 13.60 m (44.6 ft)	
Grok 4.7	**合法**	29 / 30	29 個棧板位, 13.62 m (44.7 ft)	
Kimi K3.1 Pro	**合法**	29 / 30	29 個棧板位, 13.60 m (44.6 ft)	
DeepSeek V4 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
Gemini 4 Argon	**非法**	30 / 30	29 個棧板位, 11.72 m (38.5 ft)	18 貨物重疊, 1 超出車輛範圍, 3 貨物壓損, 1 軸重超載
Qwen3p8	未回傳方案	-	-	模型寫出方案前達輸出上限。
Magistral Medium	**非法**	30 / 30	34 個棧板位, 13.36 m (43.8 ft)	4 貨物重疊, 25 超出車輛範圍, 1 軸重超載, 1 行列未對齊

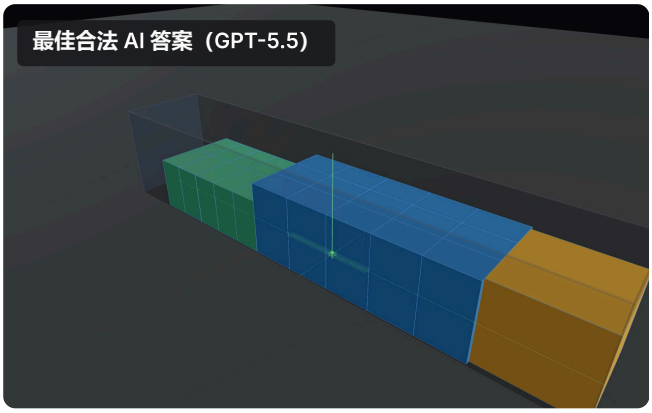
H10 最短裝載長度

車輛	件數	目標	結果
13.6m Curtainsider	26	最短裝載長度	平手

難點

26 個不可堆疊品項; 16 個必須直立放置的品項; 最重品項 666 kg (1,468 lb)





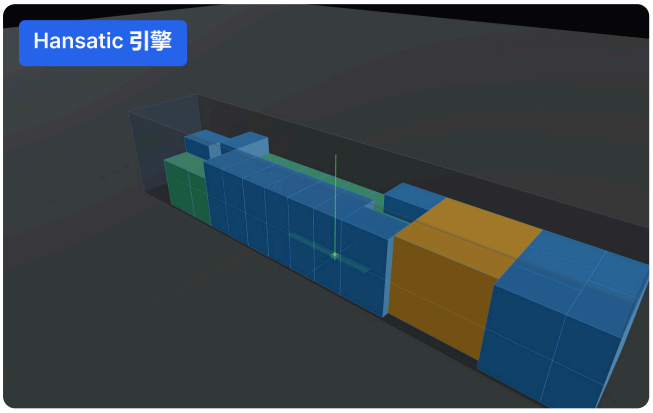
參賽者	狀態	已放置	分數	備註
Hansatic 引擎	**合法**	**26 / 26**	**12.20 m (40.0 ft)**	
GPT-5.5	**合法**	26 / 26	12.20 m (40.0 ft)	
Opus 5.5	**合法**	26 / 26	12.20 m (40.0 ft)	
Grok 4.7	**合法**	26 / 26	12.20 m (40.0 ft)	
Kimi K3.1 Pro	**合法**	26 / 26	12.20 m (40.0 ft)	
DeepSeek V4 Pro	**合法**	26 / 26	12.20 m (40.0 ft)	
Gemini 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)	

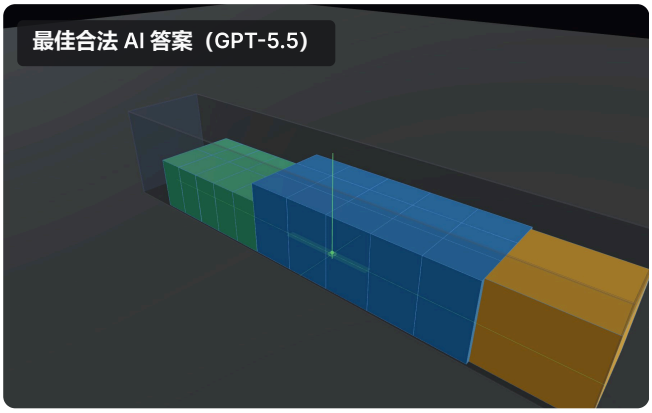
H10 最少棧板位置

車輛	件數	目標	結果
13.6m Curtainsider	26	最少棧板位置	平手

難點

26 個不可堆疊品項; 16 個必須直立放置的品項; 最重品項 666 kg (1,468 lb)





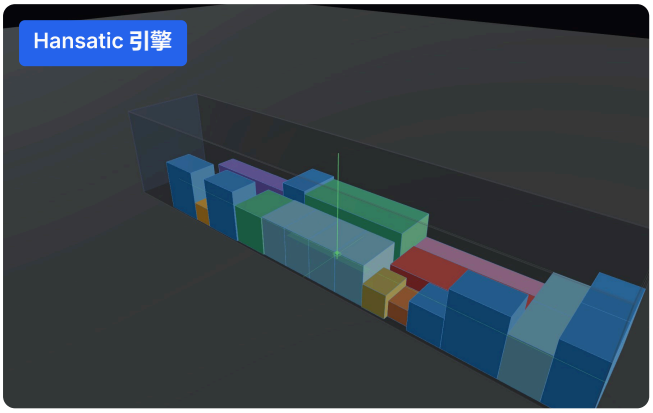
參賽者	狀態	已放置	分數	備註
Hansatic 引擎	**合法**	**26 / 26**	**31 個棧板位, 12.20 m (40.0 ft)**	
GPT-5.5	**合法**	26 / 26	31 個棧板位, 12.20 m (40.0 ft)	
Opus 5.5	**合法**	26 / 26	31 個棧板位, 13.00 m (42.7 ft)	
Grok 4.7	**合法**	26 / 26	31 個棧板位, 12.20 m (40.0 ft)	
Kimi K3.1 Pro	**合法**	26 / 26	31 個棧板位, 12.20 m (40.0 ft)	
DeepSeek V4 Pro	**合法**	26 / 26	31 個棧板位, 12.20 m (40.0 ft)	
Gemini 4 Argon	**非法**	26 / 26	31 個棧板位, 12.60 m (41.3 ft)	1 軸重超載
Qwen3p8	**合法**	26 / 26	31 個棧板位, 13.00 m (42.7 ft)	
Magistral Medium	**非法**	26 / 26	28 個棧板位, 13.00 m (42.7 ft)	10 超出車輛範圍

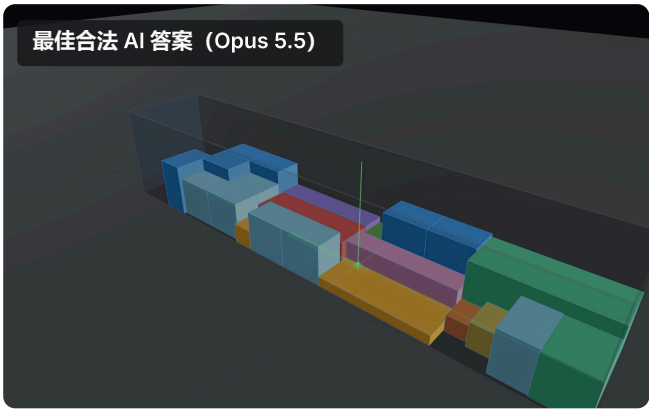
H11 最短裝載長度

車輛	件數	目標	結果
13.6m Curtainsider	23	最短裝載長度	引擎些微領先

難點

23 個不可堆疊品項; 23 個必須直立放置的品項; 最重品項 2,700 kg (5,952 lb)





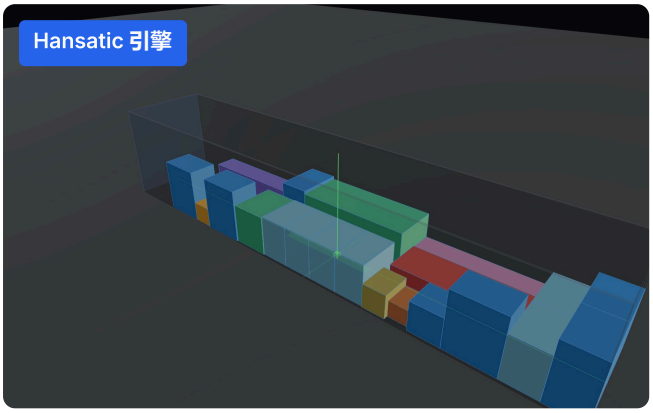
參賽者	狀態	已放置	分數	備註
Hansatic 引擎	**合法**	**23 / 23**	**12.00 m (39.4 ft)**	
GPT-5.5	**合法**	23 / 23	12.42 m (40.7 ft)	
Opus 5.5	**合法**	23 / 23	12.20 m (40.0 ft)	
Grok 4.7	**合法**	23 / 23	12.62 m (41.4 ft)	
Kimi K3.1 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
DeepSeek V4 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
Gemini 4 Argon	**合法**	23 / 23	13.02 m (42.7 ft)	
Qwen3p8	未回傳方案	-	-	模型寫出方案前達輸出上限。
Magistral Medium	**非法**	23 / 23	12.34 m (40.5 ft)	7 貨物重疊, 8 超出車輛範圍

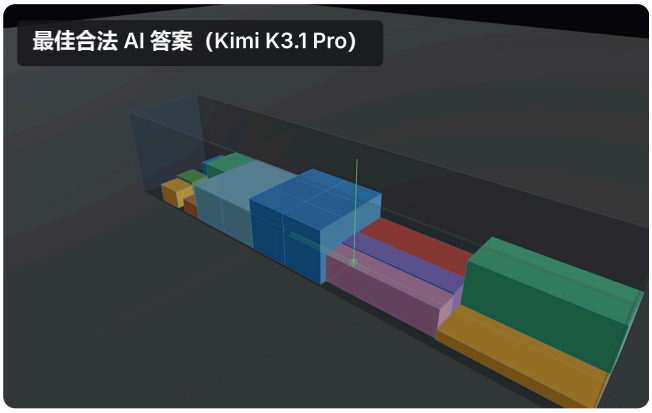
H11 最少棧板位置

車輛	件數	目標	結果
13.6m Curtainsider	23	最少棧板位置	引擎明顯領先

難點

23 個不可堆疊品項; 23 個必須直立放置的品項; 最重品項 2,700 kg (5,952 lb)





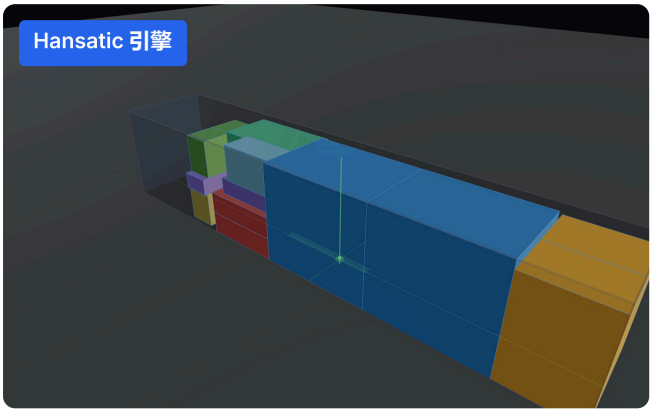
參賽者	狀態	已放置	分數	備註
Hansatic 引擎	**合法**	**23 / 23**	**30 個棧板位, 12.00 m (39.4 ft)**	
GPT-5.5	**合法**	23 / 23	32 個棧板位, 12.64 m (41.5 ft)	
Opus 5.5	**合法**	23 / 23	32 個棧板位, 12.84 m (42.1 ft)	
Grok 4.7	供應商錯誤	-	-	供應商拒絕請求。
Kimi K3.1 Pro	**合法**	23 / 23	32 個棧板位, 12.62 m (41.4 ft)	
DeepSeek V4 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
Gemini 4 Argon	**非法**	22 / 23	30 個棧板位, 12.44 m (40.8 ft)	1 貨物尺寸錯誤
Qwen3p8	未回傳方案	-	-	模型寫出方案前達輸出上限。
Magistral Medium	**非法**	23 / 23	30 個棧板位, 12.60 m (41.3 ft)	11 貨物重疊, 9 超出車輛範圍

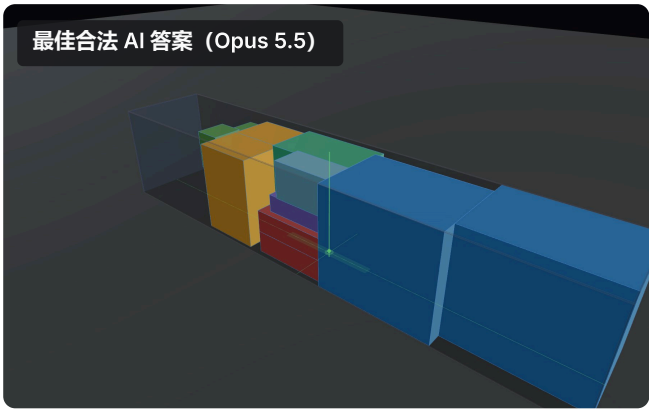
H12 最短裝載長度

車輛	件數	目標	結果
13.6m Curtainsider	14	最短裝載長度	平手

難點

6 個不可堆疊品項; 9 個必須直立放置的品項; 最重品項 732 kg (1,614 lb)





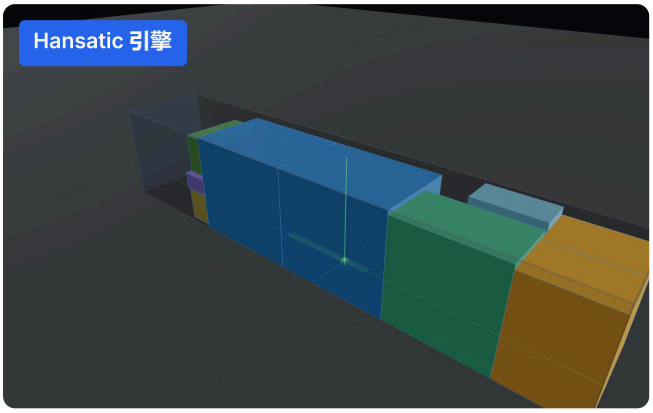
參賽者	狀態	已放置	分數	備註
Hansatic 引擎	**合法**	**13 / 14**	**11.00 m (36.1 ft)**	
GPT-5.5	**合法**	13 / 14	11.40 m (37.4 ft)	
Opus 5.5	**合法**	13 / 14	11.00 m (36.1 ft)	
Grok 4.7	供應商錯誤	-	-	供應商拒絕請求。
Kimi K3.1 Pro	**合法**	13 / 14	11.00 m (36.1 ft)	
DeepSeek V4 Pro	**合法**	13 / 14	11.70 m (38.4 ft)	
Gemini 4 Argon	**合法**	13 / 14	11.00 m (36.1 ft)	
Qwen3p8	未回傳方案	-	-	模型寫出方案前達輸出上限。
Magistral Medium	**非法**	14 / 14	13.50 m (44.3 ft)	5 貨物重疊, 5 超出車輛範圍, 2 懸空貨物

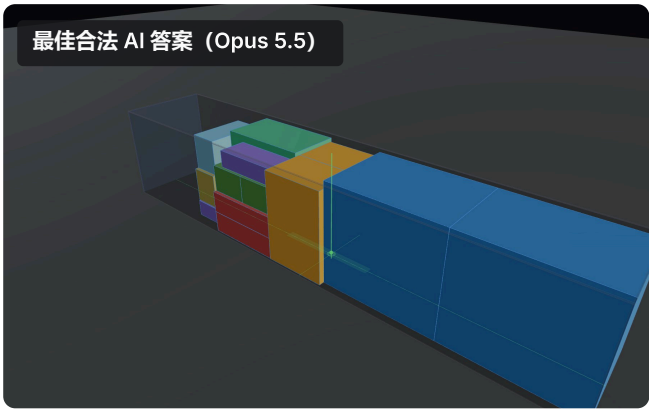
H12 最少棧板位置

車輛	件數	目標	結果
13.6m Curtainsider	14	最少棧板位置	引擎明顯領先

難點

6 個不可堆疊品項; 9 個必須直立放置的品項; 最重品項 732 kg (1,614 lb)





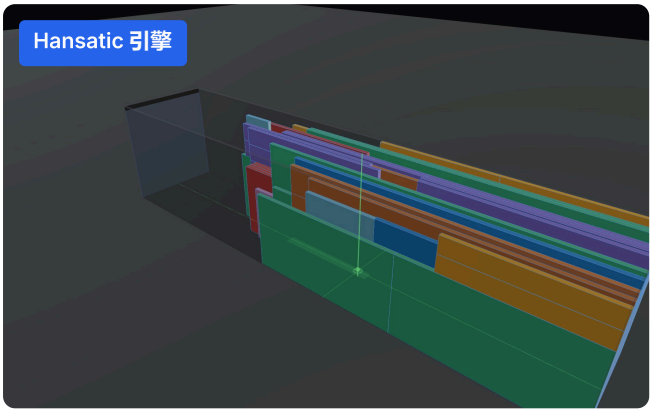
參賽者	狀態	已放置	分數	備註
Hansatic 引擎	**合法**	**13 / 14**	**25 個棧板位, 11.00 m (36.1 ft)**	
GPT-5.5	**合法**	13 / 14	29 個棧板位, 11.73 m (38.5 ft)	
Opus 5.5	**合法**	13 / 14	28 個棧板位, 11.00 m (36.1 ft)	
Grok 4.7	供應商錯誤	-	-	供應商拒絕請求。
Kimi K3.1 Pro	**合法**	13 / 14	28 個棧板位, 11.00 m (36.1 ft)	
DeepSeek V4 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
Gemini 4 Argon	**合法**	13 / 14	28 個棧板位, 11.15 m (36.6 ft)	
Qwen3p8	未回傳方案	-	-	模型寫出方案前達輸出上限。
Magistral Medium	**非法**	14 / 14	33 個棧板位, 13.20 m (43.3 ft)	9 貨物重疊, 9 超出車輛範圍

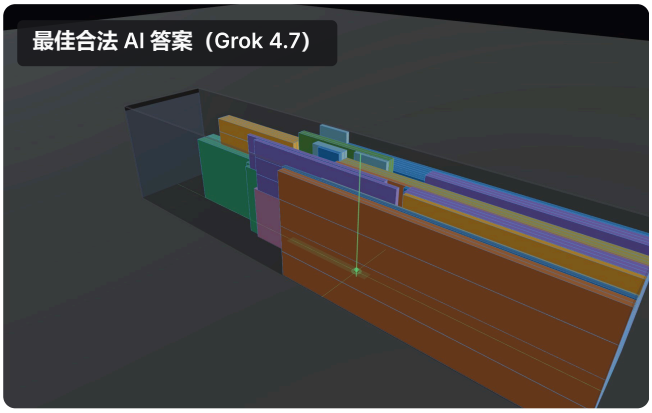
BV1 最短裝載長度

車輛	件數	目標	結果
40' High Cube	103	最短裝載長度	引擎明顯領先

難點

37 個必須直立放置的品項; 品項最長達 8.27 m (27.1 ft)





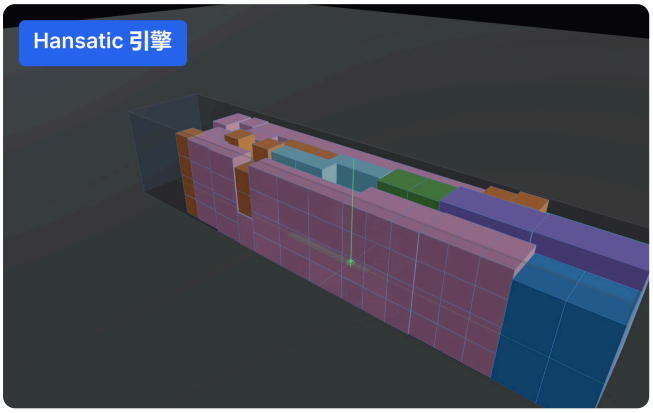
參賽者	狀態	已放置	分數	備註
Hansatic 引擎	**合法**	**103 / 103**	**9.17 m (30.1 ft)**	
GPT-5.5	**合法**	92 / 103	8.27 m (27.1 ft)	
Opus 5.5	未回傳方案	-	-	模型寫出方案前達輸出上限。
Grok 4.7	**合法**	103 / 103	10.66 m (35.0 ft)	
Kimi K3.1 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
DeepSeek V4 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
Gemini 4 Argon	**非法**	77 / 103	8.36 m (27.4 ft)	1 貨物尺寸錯誤, 4 超過高度限制, 4 懸空貨物
Qwen3p8	未回傳方案	-	-	模型寫出方案前達輸出上限。
Magistral Medium	**非法**	22 / 103	16.54 m (54.3 ft)	9 超出車輛範圍

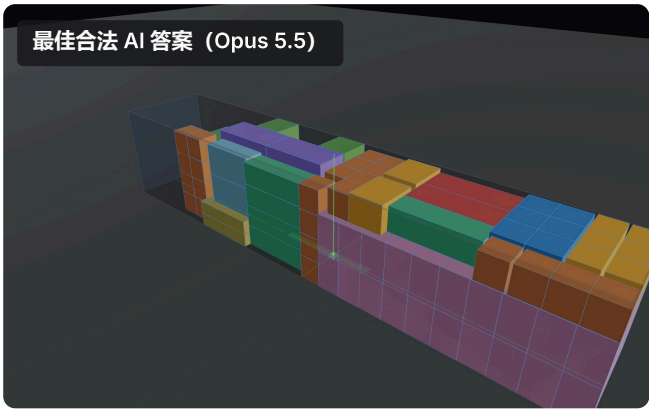
BV2 最短裝載長度

車輛	件數	目標	結果
13.6m Curtainsider	153	最短裝載長度	引擎明顯領先

難點

2 個不可堆疊品項; 最重品項 578 kg (1,274 lb)





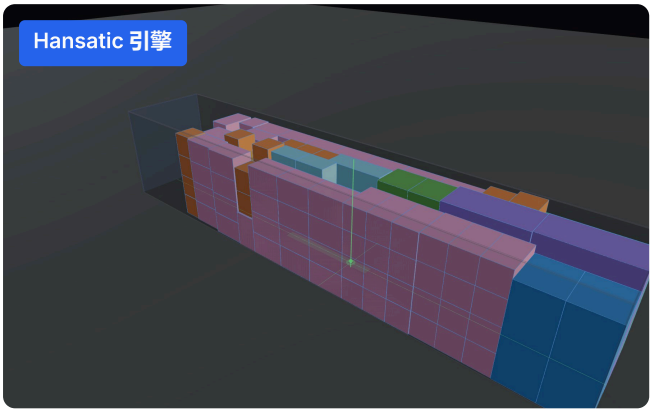
參賽者	狀態	已放置	分數	備註
Hansatic 引擎	**合法**	**153 / 153**	**11.52 m (37.8 ft)**	
GPT-5.5	**合法**	153 / 153	13.14 m (43.1 ft)	
Opus 5.5	**合法**	153 / 153	11.78 m (38.6 ft)	
Grok 4.7	**合法**	153 / 153	12.30 m (40.4 ft)	
Kimi K3.1 Pro	**合法**	153 / 153	11.99 m (39.3 ft)	
DeepSeek V4 Pro	**合法**	153 / 153	12.79 m (42.0 ft)	
Gemini 4 Argon	**非法**	147 / 153	12.77 m (41.9 ft)	6 貨物尺寸錯誤
Qwen3p8	未回傳方案	-	-	模型寫出方案前達輸出上限。
Magistral Medium	**非法**	56 / 153	14.02 m (46.0 ft)	39 貨物重疊, 27 超出車輛範圍

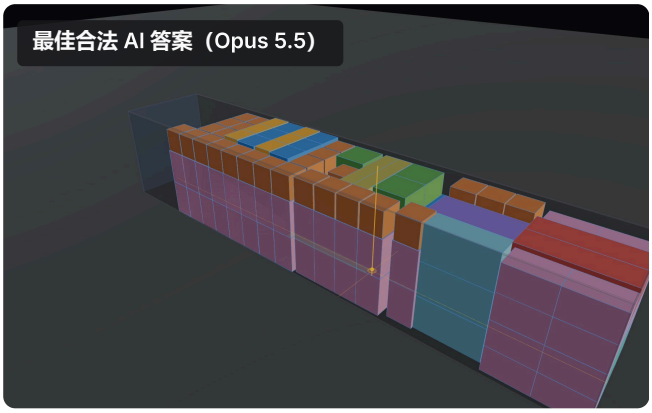
BV2 最少棧板位置

車輛	件數	目標	結果
13.6m Curtainsider	153	最少棧板位置	引擎些微領先

難點

2 個不可堆疊品項; 最重品項 578 kg (1,274 lb)





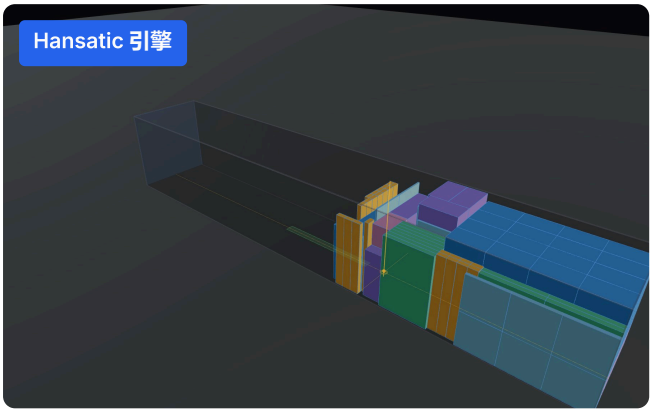
參賽者	狀態	已放置	分數	備註
Hansatic 引擎	**合法**	**153 / 153**	**29 個棧板位, 11.52 m (37.8 ft)**	
GPT-5.5	**合法**	153 / 153	32 個棧板位, 12.60 m (41.3 ft)	
Opus 5.5	**合法**	153 / 153	29 個棧板位, 11.67 m (38.3 ft)	
Grok 4.7	**合法**	153 / 153	32 個棧板位, 12.78 m (41.9 ft)	
Kimi K3.1 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
DeepSeek V4 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
Gemini 4 Argon	未回傳方案	-	-	模型寫出方案前達輸出上限。
Qwen3p8	未回傳方案	-	-	模型寫出方案前達輸出上限。
Magistral Medium	**非法**	153 / 153	34 個棧板位, 14.82 m (48.6 ft)	7 貨物重疊, 31 超出車輛範圍, 86 超過高度限制, 96 懸空貨物

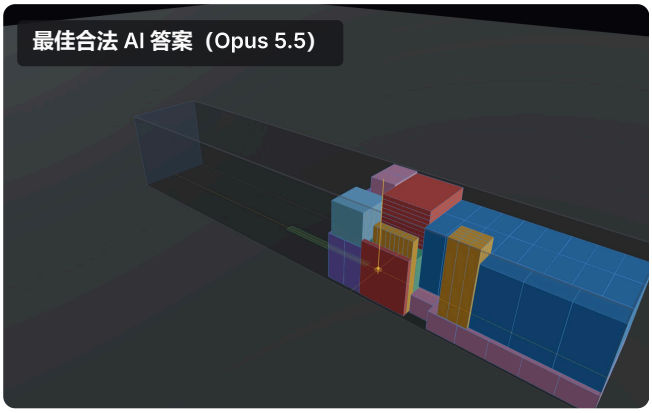
BV3 最短裝載長度

車輛	件數	目標	結果
53ft Dry Van	119	最短裝載長度	引擎明顯領先

難點

19 個不可堆疊品項; 35 個必須直立放置的品項





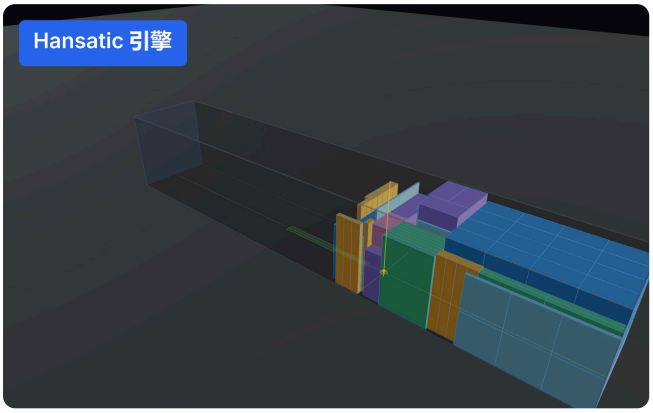
參賽者	狀態	已放置	分數	備註
Hansatic 引擎	**合法**	**119 / 119**	**6.83 m (22.4 ft)**	
GPT-5.5	**合法**	119 / 119	9.14 m (30.0 ft)	
Opus 5.5	**合法**	119 / 119	6.99 m (22.9 ft)	
Grok 4.7	**合法**	119 / 119	8.10 m (26.6 ft)	
Kimi K3.1 Pro	未回傳方案	-	-	答案不含有效方案。
DeepSeek V4 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
Gemini 4 Argon	**非法**	115 / 119	8.36 m (27.4 ft)	4 貨物尺寸錯誤, 5 貨物重疊, 3 貨物壓損
Qwen3p8	未回傳方案	-	-	模型寫出方案前達輸出上限。
Magistral Medium	**非法**	22 / 119	16.76 m (55.0 ft)	11 超出車輛範圍

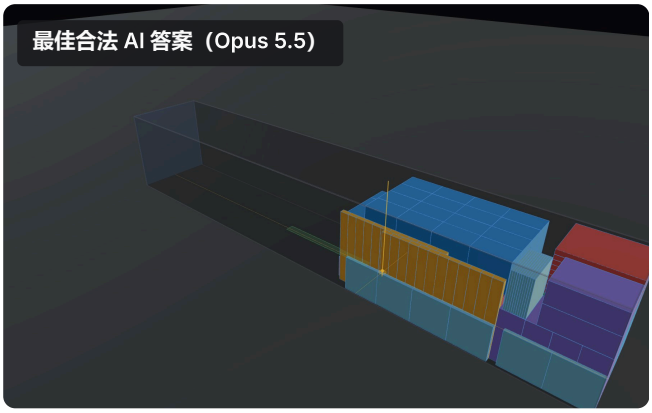
BV3 最少棧板位置

車輛	件數	目標	結果
53ft Dry Van	119	最少棧板位置	引擎明顯領先

難點

19 個不可堆疊品項; 35 個必須直立放置的品項





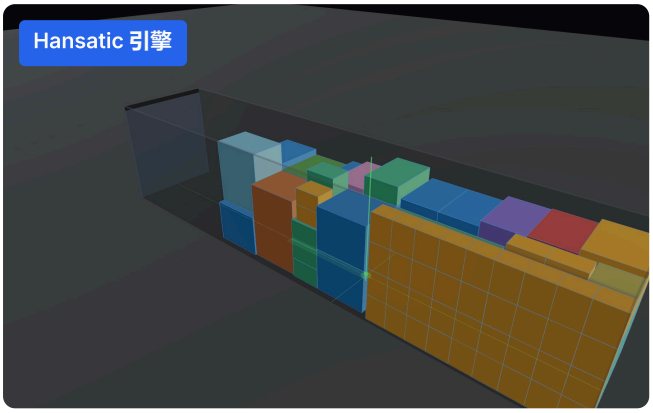
參賽者	狀態	已放置	分數	備註
Hansatic 引擎	**合法**	**119 / 119**	**17 個棧板位, 6.83 m (22.4 ft)**	
GPT-5.5	**合法**	119 / 119	23 個棧板位, 9.07 m (29.8 ft)	
Opus 5.5	**合法**	119 / 119	18 個棧板位, 6.99 m (22.9 ft)	
Grok 4.7	**合法**	119 / 119	22 個棧板位, 8.64 m (28.3 ft)	
Kimi K3.1 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
DeepSeek V4 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
Gemini 4 Argon	未回傳方案	-	-	模型寫出方案前達輸出上限。
Qwen3p8	未回傳方案	-	-	模型寫出方案前達輸出上限。
Magistral Medium	**非法**	107 / 119	21 個棧板位, 16.18 m (53.1 ft)	12 貨物尺寸錯誤, 66 貨物重疊, 35 超出車輛範圍, 60 懸空貨物, 1 貨物壓損, 1 鬆散群組未固定

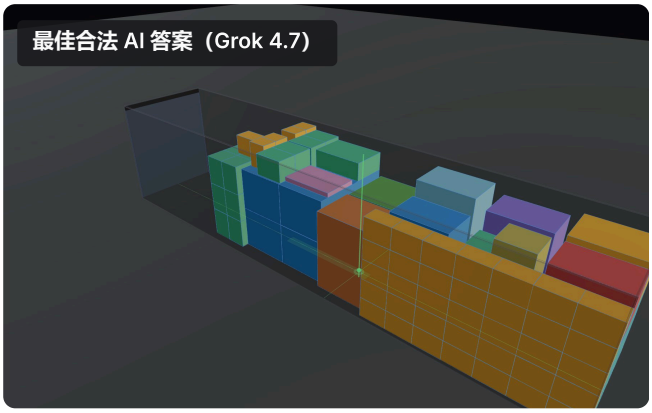
BV4 最短裝載長度

車輛	件數	目標	結果
40' High Cube	145	最短裝載長度	引擎明顯領先

難點

11 個不可堆疊品項; 11 個易碎品，僅限頂部; 20 個必須直立放置的品項





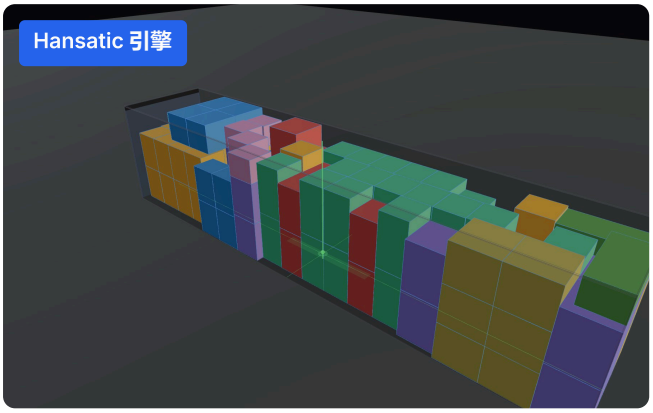
參賽者	狀態	已放置	分數	備註
Hansatic 引擎	**合法**	**145 / 145**	**8.57 m (28.1 ft)**	
GPT-5.5	**合法**	145 / 145	11.96 m (39.2 ft)	
Opus 5.5	未回傳方案	-	-	模型寫出方案前達輸出上限。
Grok 4.7	**合法**	145 / 145	8.85 m (29.0 ft)	
Kimi K3.1 Pro	**合法**	145 / 145	10.87 m (35.7 ft)	
DeepSeek V4 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
Gemini 4 Argon	**非法**	145 / 145	10.05 m (33.0 ft)	19 貨物重疊, 3 超出車輛範圍, 9 超過高度限制, 4 懸空貨物
Qwen3p8	未回傳方案	-	-	模型寫出方案前達輸出上限。
Magistral Medium	**合法**	11 / 145	10.33 m (33.9 ft)	

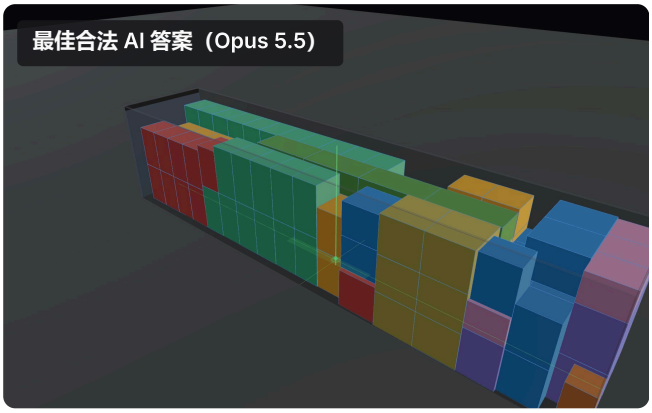
BV5 最短裝載長度

車輛	件數	目標	結果
40' High Cube	136	最短裝載長度	引擎明顯領先

難點

136 個必須直立放置的品項





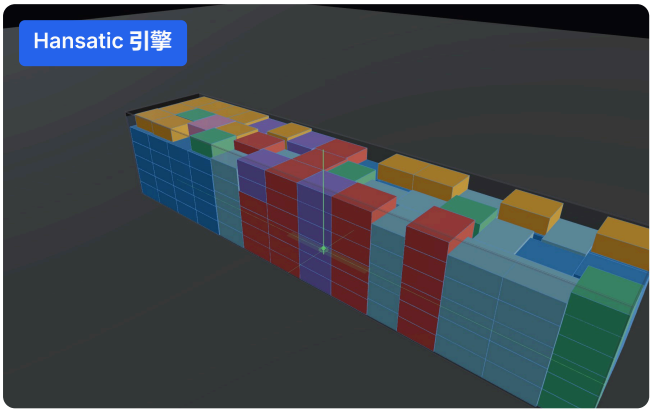
參賽者	狀態	已放置	分數	備註
Hansatic 引擎	**合法**	**127 / 136**	**12.00 m (39.4 ft)**	
GPT-5.5	**合法**	115 / 136	11.73 m (38.5 ft)	
Opus 5.5	**合法**	124 / 136	11.77 m (38.6 ft)	
Grok 4.7	**合法**	115 / 136	11.87 m (39.0 ft)	
Kimi K3.1 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
DeepSeek V4 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
Gemini 4 Argon	**非法**	121 / 136	12.00 m (39.4 ft)	1 傾斜了必須直立的貨物, 12 貨物重疊, 6 超出車輛範圍, 6 懸空貨物
Qwen3p8	未回傳方案	-	-	模型寫出方案前達輸出上限。
Magistral Medium	**非法**	51 / 136	12.89 m (42.3 ft)	34 超出車輛範圍, 6 懸空貨物

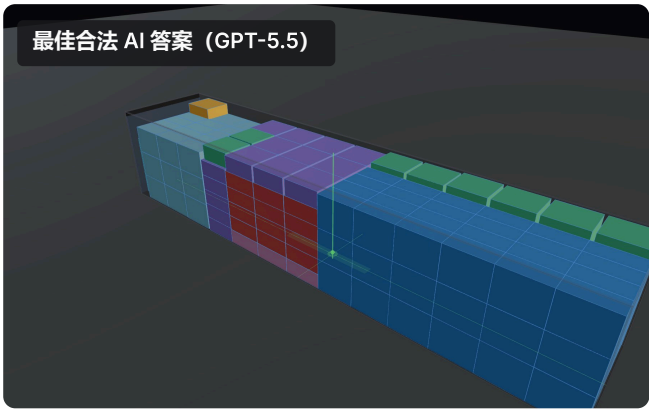
BV6 最短裝載長度

車輛	件數	目標	結果
40' Dry Container	253	最短裝載長度	引擎明顯領先

難點

尺寸混雜，無特殊標記



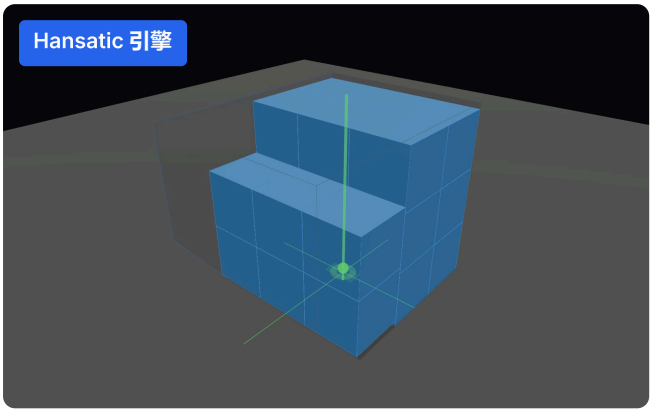


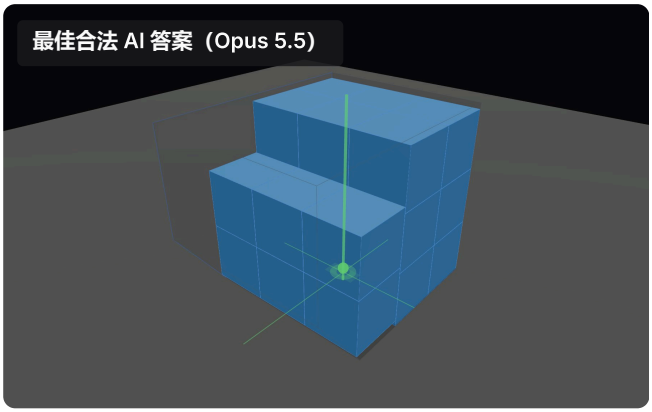
參賽者	狀態	已放置	分數	備註
Hansatic 引擎	**合法**	**244 / 253**	**11.98 m (39.3 ft)**	
GPT-5.5	**合法**	238 / 253	11.97 m (39.3 ft)	
Opus 5.5	未回傳方案	-	-	模型寫出方案前達輸出上限。
Grok 4.7	**合法**	218 / 253	11.98 m (39.3 ft)	
Kimi K3.1 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
DeepSeek V4 Pro	未回傳方案	-	-	模型寫出方案前達輸出上限。
Gemini 4 Argon	**非法**	197 / 253	11.50 m (37.7 ft)	13 貨物重疊
Qwen3p8	未回傳方案	-	-	模型寫出方案前達輸出上限。
Magistral Medium	**非法**	42 / 253	12.18 m (40.0 ft)	3 超出車輛範圍

U01 最多件數

車輛	件數	目標	結果
AKE (LD3)	50	最多件數	平手

難點
斜壁





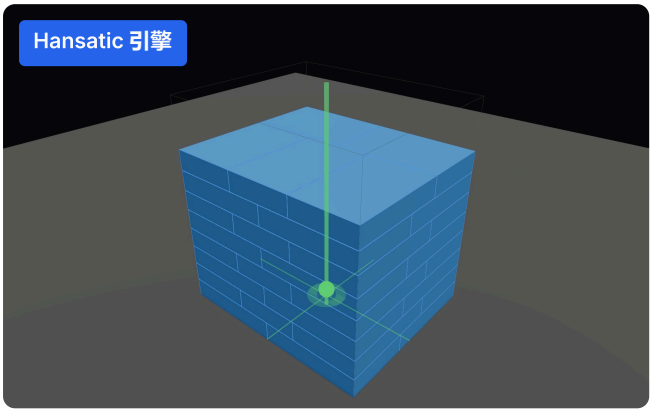
參賽者	狀態	已放置	分數	備註
Hansatic 引擎	**合法**	**24 / 50**	**24**	
GPT-5.5	**合法**	18 / 50	18	
Opus 5.5	**合法**	24 / 50	24	
Grok 4.7	**合法**	24 / 50	24	
Kimi K3.1 Pro	**合法**	24 / 50	24	
DeepSeek V4 Pro	**合法**	18 / 50	18	
Gemini 4 Argon	**合法**	24 / 50	24	
Qwen3p8	**合法**	24 / 50	24	
Magistral Medium	**非法**	24 / 50	24	3 超出航空貨櫃輪廓

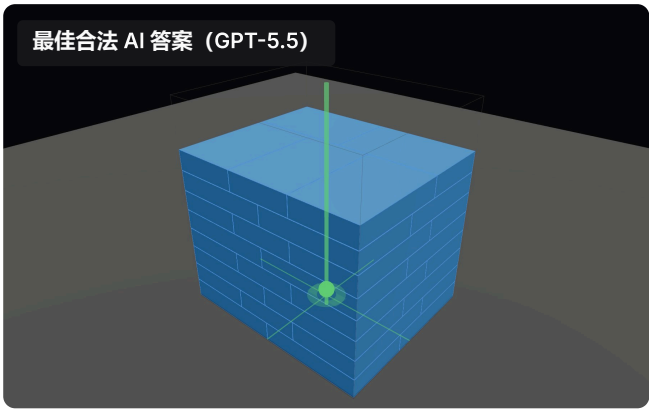
P01 最多互鎖棧板

車輛	件數	目標	結果
Industrial / Block 1200×1000	40	最多互鎖棧板	平手

難點

40 個必須直立放置的品項; 品項最長達 0.60 m (2.0 ft)





參賽者	狀態	已放置	分數	備註
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 貨物尺寸錯誤, 12 懸空貨物
Qwen3p8	**合法**	40 / 40	100%	
Magistral Medium	**非法**	40 / 40	100%	30 貨物重疊, 9 超出車輛範圍