Hansatic

WHITEPAPER KỸ THUẬT

Lập kế hoạch chất hàng thần kinh-ký hiệu

Một công cụ xếp hàng và tám mô hình AI tiên tiến trên các tải hàng thực tế từ khách hàng.

Tháng 11 năm 2026

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

Trang đi kèm, với mọi phương án đều ở dạng 3D: hansatic.com/vi/neuro-symbolic-load-planning

Tóm tắt

Chúng tôi đang kết hợp công cụ xếp hàng tất định với AI tạo sinh. Trước khi xây dựng sự kết hợp, chúng tôi đã đo lường từng phần riêng lẻ.

- **Dữ liệu:** 20 ca lập kế hoạch chất hàng thực tế của khách hàng, 26 lượt chạy được chấm điểm. Các ca bao gồm container đường biển, xe rèm, một xe thùng khô 53 ft (16.2 m), một container hàng không AKE và một cụm pallet đan xen.
- **Các bên tham gia:** công cụ xếp hàng sản xuất của Hansatic và tám LLM tiên phong phổ biến. Mỗi mô hình phải đưa ra vị trí và hướng cho từng món hàng trong một câu trả lời JSON, không dùng công cụ.
- **Bộ chấm điểm:** một bộ kiểm tra tất định chấm điểm mọi bên tham gia như nhau. Nó kiểm tra:
  - hình học;
  - đỡ đế 95%;
  - giới hạn xếp chồng và hàng dễ vỡ;
  - cờ định hướng;
  - các đống cao và hẹp;
  - tải trọng và tải trọng trục;
  - dải trọng tâm theo chiều dọc.

Kết quả:

- **Công cụ** hợp lệ ở 26 trên 26 lượt chạy. So với câu trả lời hợp lệ tốt nhất của mô hình trên mỗi lượt, công cụ dẫn trước ở 18 lượt (13 rõ ràng, 5 sát nút) và hòa ở 8 lượt. Không đứng sau ở lượt nào.
- **Mô hình mạnh nhất,** GPT-5.5, hợp lệ ở 25 trên 26 lượt chạy, kể cả các bảng kê khai lên đến 253 kiện hàng.
- **Bốn trong tám mô hình** hợp lệ ở ít hơn một phần ba số lượt.
- **Điểm mạnh của các mô hình.** Các câu trả lời tốt nhất của mô hình dùng hai chiến lược bố trí mà công cụ chưa có. Cả hai nay đã là chiến lược của công cụ tất định.

Chúng tôi kết luận công cụ là bộ lập kế hoạch tốt hơn hiện nay, và sự sáng tạo của mô hình đáng được nắm bắt. Kiến trúc của chúng tôi cho điều đó là neuro-symbolic:

- các mô hình đề xuất chiến lược;
- công cụ xây dựng mọi kế hoạch;
- bộ kiểm tra chỉ giữ các cải tiến hợp lệ.

Chúng tôi công bố các ca, bộ kiểm tra, prompt và mã chấm điểm.

1. Thiết kế đánh giá

1.1 Các trường hợp

Mỗi trường hợp là một tải hàng thực tế của khách hàng từ ứng dụng Hansatic, sao chép nguyên bản 1:1: kích thước, trọng lượng, số lượng, cờ và phương tiện. ID hàng hóa được thay thế và tên hàng bị xóa.

Nhóm	Trường hợp	Lần chạy	Phương tiện	Số kiện
Tiêu chuẩn (H01-H12)	12	16	một container 20', hai container 40' khô, bốn container 40' HC, một container 45' HC, bốn xe rèm 13.6 m (44.6 ft)	14-83
Lớn (BV1-BV6)	6	8	ba container 40' HC, một container 40' khô, một xe rèm 13.6 m (44.6 ft), một xe thùng khô 53 ft (16.2 m)	103-253
Đặc biệt (U01, P01)	2	2	AKE (LD3) có vách vát; pallet 1200 × 1000 mm (47 × 39 in)	50 và 40

Lần chạy và mục tiêu:

- **Xe tải** được chấm hai lần: theo mét chất hàng (LDM) và theo vị trí pallet. Từ 12 ca tiêu chuẩn có 16 lần chạy, từ 6 ca lớn có 8 lần.
- **BV5 và BV6** không chứa hết được. Mục tiêu chính ở đây là xếp được nhiều hàng hơn.
- **U01 và P01** đều có giá trị tối ưu đã biết:
  - U01: 24 trong 50 khối, vì vách vát không cho phép thêm;
  - P01: 40 trong 40 bao với độ lồng ghép 100%.

Các ràng buộc trong ca:

- hàng không xếp chồng được và hàng chỉ đặt trên cùng (dễ vỡ);
- giới hạn trọng lượng xếp chồng theo từng mặt hàng;
- cờ "chiều này lên trên" và "không nghiêng";
- thanh dài (BV1, dài bằng chiều dài container);
- một mặt hàng nặng chiếm ưu thế duy nhất (H02);
- tải hỗn hợp pallet và thùng carton (BV3).

1.2 Bộ kiểm tra

Bộ kiểm tra (v2.3.0) kiểm tra mọi phương án như nhau. Phương án chỉ hợp lệ khi tất cả kiểm tra cứng đều đạt.

Kiểm tra	Quy tắc
Phạm vi	mọi hàng nằm trong không gian chở hàng, kể cả phần vát AKE
Chồng lấn	không có hai hàng nào chung thể tích
Đỡ	ít nhất 95% đế mỗi hàng tựa trên sàn hoặc trên hàng dưới
Xếp chồng	hàng không xếp chồng không đỡ gì; hàng trên cùng không đỡ gì và nằm trên cùng; giới hạn trọng lượng xếp chồng theo từng hàng được giữ
Hướng	cờ "chiều này lên" và "không nghiêng" được tôn trọng
Cọc cao mảnh	giới hạn tỷ lệ cao/đế với các cọc đứng tự do
Tải trọng	tổng khối lượng trong giới hạn xe
Trục (xe tải)	tải trọng trục trước/sau trong giới hạn, tính từ trọng tâm từng hàng
CoG (container)	trọng tâm dọc trong dải cho phép
Nhiều điểm dừng	vào sau ra trước theo điểm dừng, với ca có nhiều điểm dừng

Rủi ro lật được tính và báo cáo, nhưng không tính điểm.

1.3 Xếp hạng

Áp dụng mỗi lần chạy:

1. Phương án hợp lệ thắng phương án vi phạm.
2. Xếp nhiều hàng hơn thắng.
3. Mục tiêu của lần chạy quyết định:
    - tải ngắn hơn (LDM);
    - hoặc ít vị trí pallet hơn;
    - hoặc, với P01, tỷ lệ lồng ghép cao hơn, rồi đến chiều cao thấp hơn.
4. Hòa vị trí pallet được phá bằng LDM.

Một bộ đếm đếm vị trí pallet cho mọi đối thủ: bộ đếm sản xuất của Hansatic.

Biên độ. Chiến thắng được gọi là:

- **rõ ràng** nếu đạt ít nhất 2% về LDM, ít nhất một vị trí pallet, hoặc thêm ít nhất một hàng;
- **hẹp** nếu không.

Động cơ chạy trong ngân sách thời gian, nên chênh lệch dưới 2% ở một lần chạy có thể là nhiễu giữa các lần (mục 6.2).

1.4 Các đối thủ

Mô hình. Mô hình tiên phong chính của tám phòng thí nghiệm:

- 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)

Cài đặt:

- **Lần chạy:** một lần mỗi ca.
- **Giới hạn đầu ra:** 64k token, hoặc tối đa thấp hơn của nhà cung cấp (Gemini: 32k).
- **Xử lý theo lô:** API lô của nhà cung cấp khi có.
- **Thu thập câu trả lời:** 2026-11-11.
- **Tổng chi phí:** khoảng $64.

Động cơ. Bộ xếp hàng sản xuất của Hansatic, bản ngày 2026-11-20 (commit 9f66537c). Cài đặt:

- cấu hình sản xuất: 4 worker và ngân sách thời gian sản xuất, ghim một lõi CPU;
- một lần chạy mỗi ca, không chọn tốt nhất;
- xếp theo quy tắc đỡ 60% riêng, và chấm ở 95% như mọi mô hình.

1.5 Câu lệnh

- Một câu lệnh mỗi loại xe: container, xe tải, ULD và pallet.
- Mỗi câu lệnh nêu đúng quy tắc bộ kiểm tra áp dụng cho loại đó, mục tiêu, và lược đồ đầu ra (id hàng, vị trí, hướng).
- Không có ví dụ bố cục đã giải.

Mẫu đầy đủ ở Phụ lục A.

2. Kết quả

2.1 Theo từng mô hình, cả 26 lượt

Mô hình	Hợp lệ	Ngang engine	Kém engine
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

Không mô hình nào vượt engine ở bất kỳ lượt nào.

Tập chuẩn (16 lượt).

Mô hình	Hợp lệ	Ngang	Kém
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

Tập lớn (8 lượt). Không mô hình nào ngang engine ở bất kỳ lượt lớn nào.

Mô hình	Hợp lệ
GPT-5.5	8
Grok 4.7	8
Opus 5.5	5

Mô hình	Hợp lệ
Kimi K3.1 Pro	2
DeepSeek V4 Pro	1
Magistral Medium	1
Gemini 4 Argon	0
Qwen3p8	0

Tập đặc biệt (2 lượt).

Mô hình	Hợp lệ	Đạt tối ưu
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 Engine so với câu trả lời hợp lệ tốt nhất của mô hình

Tập	Vượt rõ	Vượt nhẹ	Hòa	Kém
Chuẩn (16)	7	3	6	0
Lớn (8)	6	2	0	0
Đặc biệt (2)	0	0	2	0
Tất cả (26)	**13**	**5**	**8**	**0**

Tập chuẩn. Mỗi ô cho biết số kiện đã xếp, rồi đến chiều dài chất hàng hoặc vị trí pallet.

Lượt	Engine	Mô hình hợp lệ tốt nhất	Biên độ
H01	26/27, 11.69 m (38.4 ft)	Opus 26, 11.74 m (38.5 ft)	nhẹ (0.4%)
H02	24/29, 5.78 m (19.0 ft)	GPT 18, 5.76 m (18.9 ft)	rõ (+6 kiện)
H03	31/32, 10.51 m (34.5 ft)	không hợp lệ	rõ
H04	39, 10.44 m (34.3 ft)	Opus 39, 10.44 m (34.3 ft)	hòa

Lượt	Engine	Mô hình hợp lệ tốt nhất	Biên độ
H05	49, 12.40 m (40.7 ft)	Opus 49, 12.40 m (40.7 ft)	hòa
H06	83, 10.65 m (35.0 ft)	Grok 83, 11.41 m (37.4 ft)	rõ (6.6%)
H07	45, 9.60 m (31.5 ft)	GPT 45, 9.60 m (31.5 ft)	hòa
H08	52, 5.92 m (19.4 ft)	Kimi 52, 6.12 m (20.1 ft)	rõ (3.3%)
H09 LDM	29/30, 13.36 m (43.8 ft)	GPT 29, 13.56 m (44.5 ft)	nhẹ (1.5%)
H09 pallet	29/30, 29 vị trí	GPT 29, 29 vị trí	nhẹ (phân định bằng LDM)
H10 LDM	26, 12.20 m (40.0 ft)	DeepSeek, như nhau	hòa
H10 pallet	26, 31 vị trí	DeepSeek, như nhau	hòa
H11 LDM	23, 12.00 m (39.4 ft)	Opus 23, 12.20 m (40.0 ft)	nhẹ (1.6%)
H11 pallet	23, 30 vị trí	Kimi 23, 32 vị trí	rõ (2 vị trí)
H12 LDM	13/14, 11.00 m (36.1 ft)	Gemini 13, 11.00 m (36.1 ft)	hòa
H12 pallet	13/14, 25 vị trí	Kimi 13, 28 vị trí	rõ (3 vị trí)

Tập lớn.

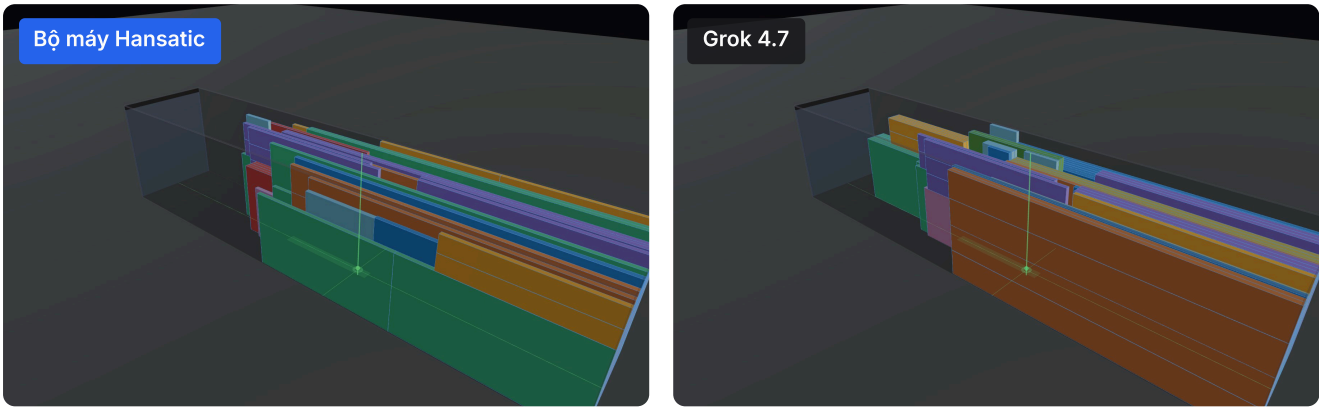
Lượt	Engine	Mô hình hợp lệ tốt nhất	Biên độ
BV1 (40' HC, 103 thanh)	103, 9.17 m (30.1 ft)	Grok 103, 10.66 m (35.0 ft)	rõ (14.0%)
BV2 LDM (xe rèm, 153)	153, 11.52 m (37.8 ft)	Opus 153, 11.78 m (38.6 ft)	rõ (2.2%)
BV2 pallet	153, 29 vị trí	Opus 153, 29 vị trí	nhẹ (phân định bằng LDM)
BV3 LDM (xe van 53 ft, 119)	119, 6.83 m (22.4 ft)	Opus 119, 6.99 m (22.9 ft)	rõ (2.2%)
BV3 pallet	119, 17 vị trí	Opus 119, 18 vị trí	rõ (1 vị trí)
BV4 (145, dễ vỡ chỉ xếp trên)	145, 8.57 m (28.1 ft)	Grok 145, 8.85 m (29.0 ft)	rõ (3.1%)
BV5 (136, chỉ xếp đứng)	127, 12.00 m (39.4 ft)	Opus 124, 11.77 m (38.6 ft)	rõ (+3 kiện)
BV6 (253, 95% thể tích)	244, 11.98 m (39.3 ft)	GPT 238, 11.97 m (39.3 ft)	rõ (+6 kiện)

Tập đặc biệt.

Lượt	Engine	Các mô hình
U01 (AKE)	24/50, tối ưu	Opus, Grok, Kimi, Qwen và Gemini cũng đạt 24; GPT và DeepSeek đạt 18
P01 (pallet)	40/40, 100% đan xen, tối ưu	mọi câu trả lời hợp lệ của mô hình đều đạt 40/40 ở mức tối ưu

2.3 Ý nghĩa của các biên độ trên một chuyến hàng thực tế

- **BV1:** ngắn hơn 1.49 m (4.9 ft), khoảng 12% chiều dài container 40'.
- **H12 pallet:** ít hơn 3 vị trí pallet, khoảng 9% sàn xe rèm 13.6 m (44.6 ft).
- **BV6:** xếp được thêm 6 kiện trên một chuyến chưa đầy đủ, tức 6 kiện ít hơn phải chuyển bằng xe khác.



BV1: engine Hansatic 9.17 m (30.1 ft), bên trái; câu trả lời hợp lệ tốt nhất của mô hình (Grok 4.7) 10.66 m (35.0 ft), bên phải.

2.4 Chi phí và thời gian

	Engine	Các mô hình
Thời gian mỗi kế hoạch	không đáng kể	không so sánh: API hàng loạt xếp hàng đợi yêu cầu
Chi phí mỗi kế hoạch	không đáng kể	$0.10-0.80

3. Phân tích lỗi

Trong 208 lần chạy mô hình (8 mô hình x 26), có 95 kế hoạch hợp lệ. 113 kế hoạch còn lại lỗi theo ba cách.

Mô hình	Hợp lệ	Không có câu trả lời (giới hạn đầu ra)	Lỗi nhà cung cấp	Kế hoạch vi phạm quy tắc
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
Tất cả	**95**	**65**	**4**	**44**

3.1 Không có câu trả lời: mô hình đạt giới hạn đầu ra

Trong 65 lần chạy, mô hình suy luận đến khi đạt giới hạn đầu ra và không bao giờ viết kế hoạch. Một câu trả lời khác (Magistral, H03) kết thúc bình thường nhưng không phải JSON hợp lệ; được tính vào đây.

- Đây là lỗi phổ biến nhất, ảnh hưởng đến sáu trong tám mô hình.
- Kimi, DeepSeek và Qwen mất phần lớn số lần chạy theo cách này. Các kế hoạch hợp lệ của họ chủ yếu ở các trường hợp nhỏ.
- Opus mất 6 lần chạy theo cách này, trong đó có BV1, BV4 và BV6, ba bản kê khai lớn nhất.

Với công cụ lập kế hoạch trong thực tế, không có câu trả lời cũng là lỗi như kế hoạch bất hợp lệ: bến hàng vẫn cần kế hoạch.

3.2 Lỗi nhà cung cấp

Grok trả về HTTP 403 trong 4 lần chạy vì đạt giới hạn chi tiêu của nhà cung cấp. Các lần chạy này không được gửi lại và tính là lỗi.

3.3 Kế hoạch vi phạm quy tắc

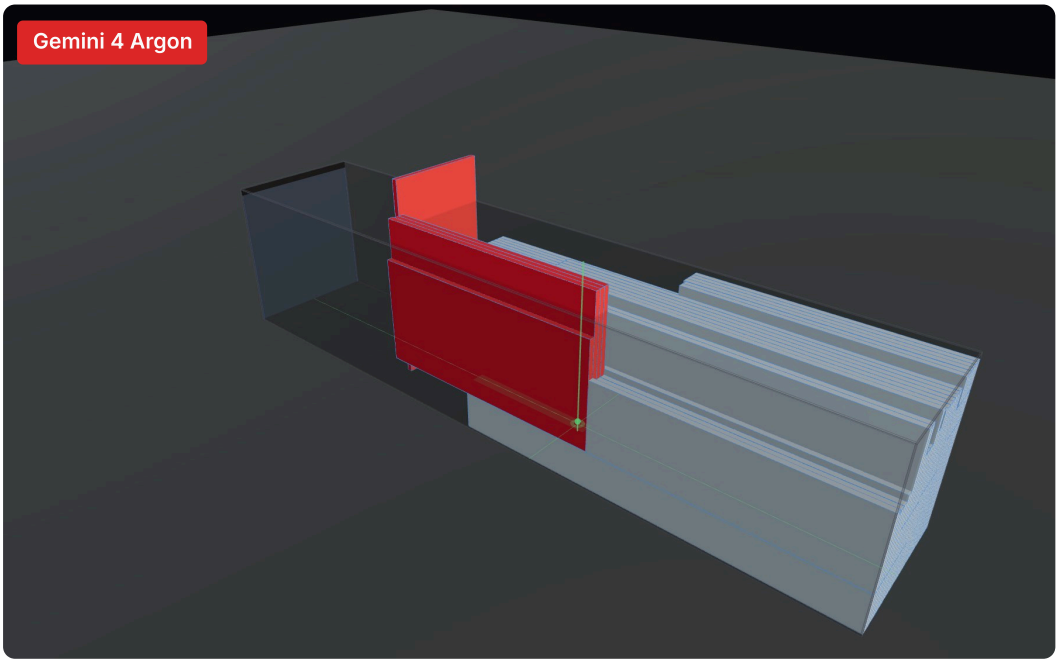
44 câu trả lời là kế hoạch hoàn chỉnh nhưng vi phạm ít nhất một quy tắc cứng. Hầu hết đến từ hai mô hình: Gemini (19) và Magistral (23). Grok có 2.

Số kế hoạch trong 44 kế hoạch này vi phạm ít nhất một loại:

Quy tắc vi phạm	Kế hoạch
Vật phẩm nằm ngoài không gian chở hàng	27
Các vật phẩm chồng lấn nhau	27
Vật phẩm lơ lửng (dưới 95% độ đỡ)	21
Kích thước vật phẩm không khớp với hướng cho phép nào	17
Chất hàng lên vật phẩm không xếp chồng được hoặc dễ vỡ, hoặc vượt quá giới hạn xếp chồng	10
Xếp chồng cao hơn nóc xe	9
Trọng tâm nằm ngoài dải cho phép	7
Quá tải trục	4
Vật phẩm không được nghiêng bị đổ	1
Nằm ngoài biên dạng AKE	1

Hai kế hoạch bất hợp lệ của Grok mỗi cái vi phạm một quy tắc: một vật phẩm bị đè bẹp ở một lần chạy, trọng tâm ở lần khác.

Các kế hoạch này trông hợp lý trong góc nhìn 3D. Lỗi thường chỉ là vài vật phẩm, lơ lửng hoặc chồng lấn vài centimet. Chỉ kiểm tra hình học đầy đủ mới phát hiện được.



BV1, Gemini 4 Argon: 77 trong số 103 vật phẩm được đặt. Đánh dấu: các vật phẩm lơ lửng, chồng lấn hoặc nằm ngoài không gian chở hàng.

3.4 Hợp lệ, nhưng kém hơn

- **Chưa đầy đủ.** Trên BV5 và BV6, câu trả lời hợp lệ tốt nhất của mô hình đặt ít hơn 3 và 6 vật phẩm so với công cụ.
- **Dài hơn.** Trên các lần chạy mà mọi vật phẩm đều vừa, câu trả lời hợp lệ tốt nhất của mô hình dài hơn công cụ trên 13 trong số 24 lần chạy xe tải và container. Khoảng cách lớn nhất:
  - 1.49 m (4.9 ft) trên BV1;
  - 0.76 m (2.5 ft) trên H06.

3.5 Những gì mô hình làm tốt

Các câu trả lời tốt nhất dùng chiến lược của đội ngũ giàu kinh nghiệm:

- **Thùng hàng nặng sát vách trước.**
- **Tận dụng chiều cao trên vật phẩm không xếp chồng được.** Trên BV3, Opus đặt thùng carton nằm nghiêng làm sàn cao 559 mm (22 in) rồi đặt 19 thiết bị không xếp chồng được lên trên.

Công cụ không có chiến lược nào trong hai. Cả hai hiện là chiến lược của công cụ tất định. Đây là sự sáng tạo mà vòng lặp neuro-symbolic được thiết kế để thu nhận (mục 5).

4. Nhánh công cụ

Trong nhánh thứ hai, một mô hình có thể gọi bộ máy thông qua một công cụ duy nhất, `pack_cargo`, rồi trả lời bằng bất kỳ phương án nào.

Hầu hết các mô hình sao chép bản kê vào lời gọi công cụ và trả về phương án của bộ máy không đổi. Các trường hợp thất bại đều là lỗi sao chép:

- Magistral đã gửi 100 mặt hàng với kích thước hoặc cờ sai;
- Kimi thực hiện lời gọi sai trên H10;
- Qwen thất bại ở 4 lần chạy.

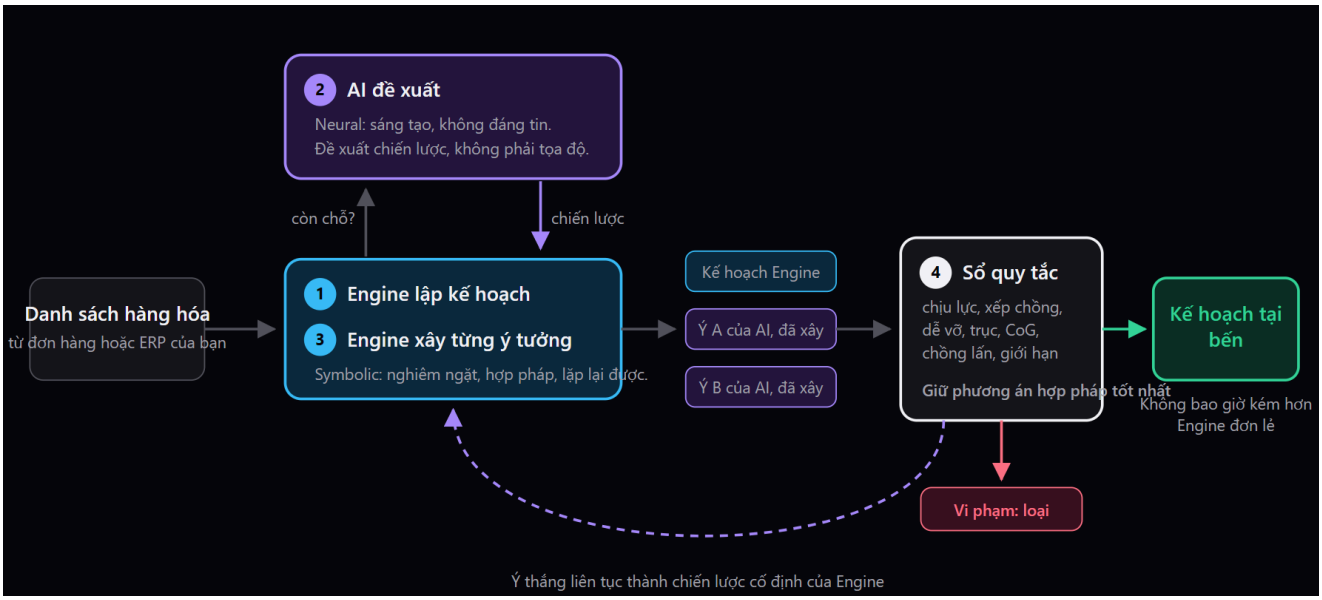
Hàm ý: trong bất kỳ hệ thống kết hợp nào, bộ máy phải đọc bản kê trực tiếp từ hệ thống nguồn, không bao giờ thông qua mô hình.

5. Kiến trúc thần kinh-ký hiệu

5.1 Hai loại trí tuệ

	Engine xếp hàng (ký hiệu)	AI tạo sinh (thần kinh)
Quy tắc	tuân thủ mọi quy tắc đã mã hóa, mọi lúc	vi phạm các quy tắc được giao
Tính lặp lại	cùng một chuyến hàng, cùng một phương án	cùng một chuyến hàng có thể ra phương án khác
Kiểm toán	mọi vị trí xếp đều có lý do	khó kiểm toán
Phạm vi	giới hạn trong các chiến lược nó có	tìm ra chiến lược mới
Quy mô	xử lý được 253 kiện	nhiều mô hình không đưa ra phương án với danh sách hàng lớn

5.2 Vòng lặp



1. **Cơ sở.** Engine lập kế hoạch từ danh sách hàng, đọc trực tiếp từ hệ thống nguồn.
2. **Kích hoạt.** Một mô hình chỉ được gọi khi phương án cơ sở còn khoảng trống: hàng chưa xếp, độ lấp đầy thấp, hoặc không đạt mục tiêu khoảng hở.
3. **Đề xuất.** Mô hình trả về các tham số chiến lược, không bao giờ trả về tọa độ:
   - vùng;
   - thứ tự;
   - ưu tiên hướng;
   - cấu trúc sàn và mui.
4. **Thực thi.** Engine xây dựng một phương án cho mỗi đề xuất.

5. **Đánh giá.** Bộ kiểm tra giữ lại phương án hợp lệ tốt nhất. Phương án cơ sở luôn là một ứng viên.
6. **Chưng cất.** Một đề xuất thắng nhiều lần trở thành chiến lược engine cố định, như các chiến lược vách trước và sàn thùng carton đã làm.

5.3 Đặc tính theo thiết kế

- **Không bao giờ tệ hơn engine đơn lẻ,** vì phương án cơ sở luôn là một ứng viên.
- **Không có phương án bất hợp lệ nào đến bến,** vì bộ kiểm tra đánh giá mọi ứng viên.
- **Không có rủi ro sao chép,** vì mô hình không bao giờ chạm vào danh sách hàng hoặc tọa độ (mục 4).
- **Không lỗi do giới hạn đầu ra của mô hình,** vì một đề xuất chỉ là vài tham số, không phải kế hoạch 253 kiện (mục 3.1).
- **Chi phí AI được giới hạn,** vì mô hình chỉ chạy khi có kích hoạt.

Chúng tôi đang xây dựng vòng lặp này ngay bây giờ. Nó sẽ được chấm điểm như một thí sinh riêng trong phiên bản tiếp theo, với chi phí và thời gian cho mỗi phương án.

6. Phạm vi và giới hạn

6.1 Phạm vi của bộ xác thực

Hợp lệ nghĩa là hợp lệ theo các quy tắc được mã hóa trong v2.3.0. Bộ xác thực không kiểm tra trình tự chất hàng ngoài quy tắc nhiều điểm dừng: vào sau, ra trước.

6.2 Biến thiên

Mỗi hệ thống chạy một lần cho mỗi ca.

- **Bộ máy** đua các chiến lược trong giới hạn thời gian, nên tải máy có thể làm thay đổi chiến lược nào hoàn thành trước. Trên H02, chúng tôi đo được 5.63-5.79 m (18.5-19.0 ft) qua các lần chạy.
- **Các mô hình** chỉ chạy một lần.

Chênh lệch trong từng lần chạy đơn lẻ dưới 2% được báo cáo là hẹp.

6.3 Cấp độ mô hình và giới hạn đầu ra

- Chúng tôi đã chạy mô hình tiên phong chính của mỗi phòng lab. Các cấp cao cấp (Anthropic Fable, OpenAI GPT-6 Astra) không được chạy.
- Giới hạn đầu ra lớn hơn có thể biến một số lần chạy không có câu trả lời (mục 3.1) thành các phương án.

6.4 Nguồn ca

Tất cả các ca đều đến từ người dùng của một nền tảng. Cơ cấu hàng hóa của các chủ hàng khác có thể khác.

7. Các bước tiếp theo

7.1 Quy tắc thực hành

Hai quy tắc thực hành tiếp theo, trong engine và trong validator v3:

- **Khoảng hở nóc xe:** ít nhất 50 mm (2.0 in) dưới nóc xe, hoặc 100 mm (3.9 in) phía trên một kiện hàng xếp chồng trên 25 kg (55 lb).
- **Độ rơ ngang** cho mỗi hàng xe nâng.

Chúng tôi chấm lại điểm mọi thí sinh bằng quy tắc nóc xe. Các kiện hàng sàn không thể nghiêng và cao hơn nóc xe trừ 50 mm (2.0 in) được miễn, vì trường hợp đó bắt buộc.

- **Engine** tuân thủ quy tắc trên 9 trong 16 lần chạy tiêu chuẩn và 1 trong 8 lần chạy lớn.

- **Trên 6 lần chạy, một phương án của mô hình tuân thủ quy tắc** với phương án gần engine. Đây là các mục tiêu đầu tiên của engine:
  - H02 (GPT);
  - H06 (Grok, 11.41 m / 37.4 ft);
  - H12 LDM (Gemini, 11.00 m / 36.1 ft);
  - H12 pallets (Kimi, 28 vị trí);
  - BV3 LDM (GPT, 9.14 m / 30.0 ft);
  - BV5 (Opus, 124 đã xếp).
- **Trên 8 lần chạy, không thí sinh nào tuân thủ quy tắc:** H01, H03, H08, BV1, cả hai lần chạy BV2, BV3 pallets, BV6.

7.2 Phiên bản tiếp theo

- Validator v3, với khoảng hở nóc xe, độ rơ ngang và kiểm tra trình tự chất hàng, áp dụng cho mọi thí sinh.
- Các trường hợp thực tế mới từ khách hàng.
- Các mô hình mạnh nhất của mỗi lab, bao gồm cả gói cao cấp, với nhiều lần chạy cho mỗi mô hình và báo cáo độ biến thiên.
- Vòng lặp neuro-symbolic như một thí sinh.

8. Công bố

Đã công bố:

- toàn bộ từng ca;
- bộ xác thực;
- các prompt;
- mã chấm điểm và xếp hạng;
- với mỗi lần chạy: yêu cầu đã gửi, phản hồi nguyên bản của nhà cung cấp và kế hoạch đã phân tích.

Trình xem: mọi kế hoạch từ mọi mô hình tham gia, ở dạng 3D bên cạnh prompt và phản hồi thô của nó: hansatic.com/vi/neuro-symbolic-load-planning.

Đã đóng:

- bộ đếm vị trí pallet (bộ đếm sản xuất của Hansatic, áp dụng như nhau cho mọi mô hình tham gia);
- bộ máy;
- các id của các kế hoạch gốc.

Phụ lục A. Các mẫu prompt

Mỗi loại xe một system prompt, đúng như đã gửi. Tin nhắn người dùng nêu mục tiêu, rồi đến ca hàng dạng JSON; danh sách mặt hàng được rút gọn ở đây. Các prompt hiển thị bằng tiếng Anh, ngôn ngữ mà các mô hình nhận được.

Container đường biển (BV1, 40' công cao)

Prompt hệ thống

```
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"]}
```

Tin nhắn người dùng (danh sách hàng đã rút gọn)

```
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"
 ]
}
```


Xe tải (H09, 13.6 m (44.6 ft) rèm)

Prompt hệ thống

```
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"]}
```

Tin nhắn người dùng (danh sách hàng đã rút gọn)

```
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"
```

```
 ]
}
```

Container hàng không (U01, AKE vách nghiêng)

Prompt hệ thống

```
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"]}
```

Tin nhắn người dùng (danh sách hàng đã rút gọn)

```
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"
 ]
}
```


Pallet (P01, 1200 × 1000 mm (47 × 39 in))

Prompt hệ thống

```
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"]}
```

Tin nhắn người dùng (danh sách hàng đã rút gọn)

```
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"
```

```
  ]
}
```

Phụ lục B. Toàn bộ 26 lần chạy

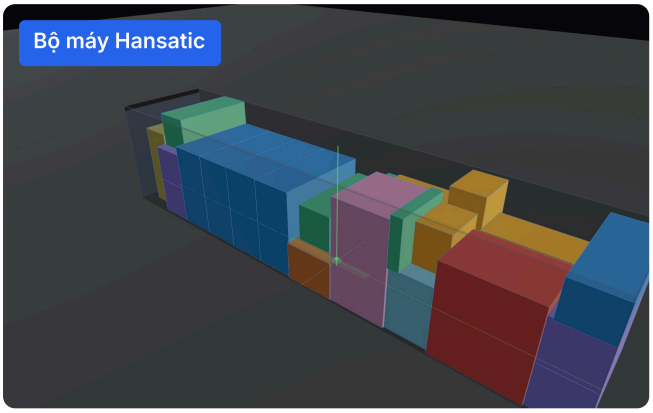
Mỗi lần chạy một trang: tải hàng, phương án của engine cạnh phương án hợp lệ tốt nhất của mô hình, và kết quả của từng mô hình tham gia.

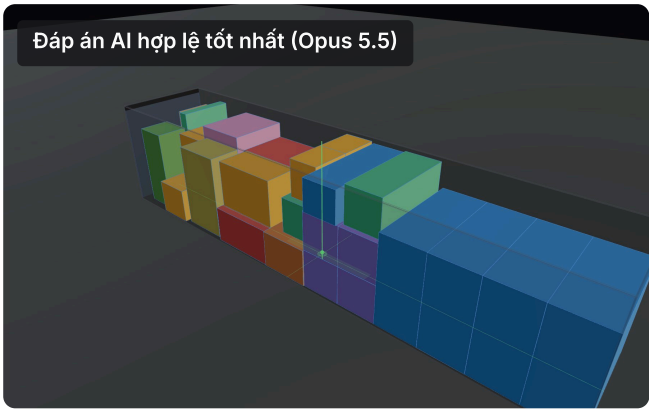
H01 Chiều dài chất hàng ngắn nhất

Phương tiện	Số kiện	Mục tiêu	Kết quả
40' High Cube	27	Chiều dài chất hàng ngắn nhất	Engine dẫn sát

Yếu tố khó

27 kiện phải giữ thẳng đứng





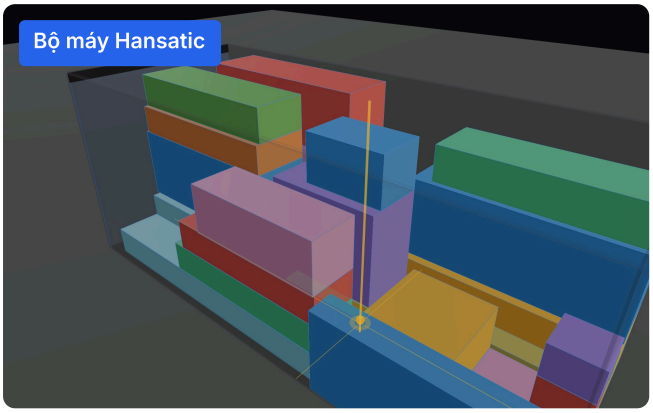
Đối thủ	Trạng thái	Đã xếp	Điểm	Ghi chú
Engine Hansatic	**Hợp pháp**	**26 / 27**	**11.69 m (38.4 ft)**	
GPT-5.5	Hợp pháp	25 / 27	11.88 m (39.0 ft)	
Opus 5.5	Hợp pháp	26 / 27	11.74 m (38.5 ft)	
Grok 4.7	Hợp pháp	25 / 27	11.37 m (37.3 ft)	
Kimi K3.1 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
DeepSeek V4 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Gemini 4 Argon	Vi phạm	22 / 27	11.56 m (37.9 ft)	1 kích thước sai, 1 ngoài xe, 1 lơ lửng
Qwen3p8	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Magistral Medium	Vi phạm	27 / 27	12.09 m (39.7 ft)	7 chồng lấn, 14 ngoài xe, 3 quá chiều cao, 1 lơ lửng, 1 trọng tâm lệch

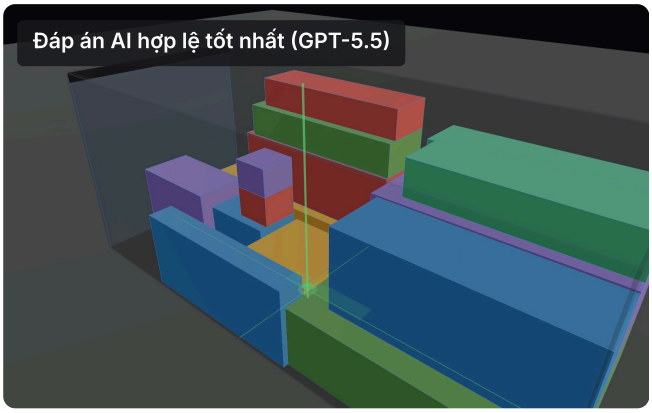
H02 Chiều dài chất hàng ngắn nhất

Phương tiện	Số kiện	Mục tiêu	Kết quả
20' Dry Container	29	Chiều dài chất hàng ngắn nhất	Engine dẫn rõ

Yếu tố khó

29 kiện phải giữ thẳng đứng; kiện dài tới 5.00 m (16.4 ft); kiện nặng nhất 1,550 kg (3,417 lb)





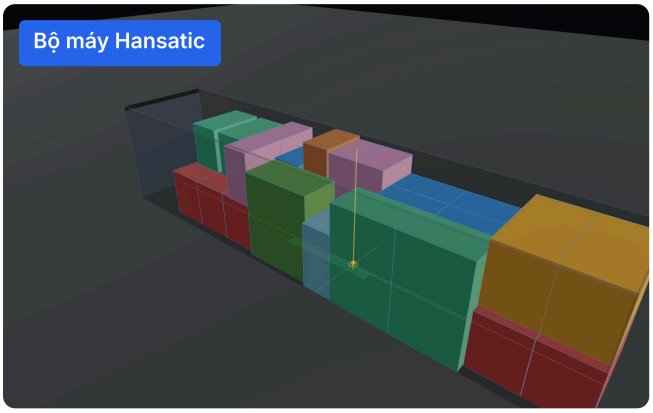
Đối thủ	Trạng thái	Đã xếp	Điểm	Ghi chú
Engine Hansatic	**Hợp pháp**	**24 / 29**	**5.78 m (19.0 ft)**	
GPT-5.5	Hợp pháp	18 / 29	5.76 m (18.9 ft)	
Opus 5.5	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Grok 4.7	Lỗi nhà cung cấp	-	-	Nhà cung cấp từ chối yêu cầu.
Kimi K3.1 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
DeepSeek V4 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Gemini 4 Argon	Vi phạm	23 / 29	5.00 m (16.4 ft)	6 kích thước sai, 13 chồng lấn, 9 lơ lửng, 9 bị đè bẹp
Qwen3p8	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Magistral Medium	Vi phạm	29 / 29	7.97 m (26.1 ft)	37 chồng lấn, 23 ngoài xe, 1 quá chiều cao, 7 lơ lửng, 5 bị đè bẹp, 2 trọng tâm lệch

H03 Chiều dài chất hàng ngắn nhất

Phương tiện	Số kiện	Mục tiêu	Kết quả
40' High Cube	32	Chiều dài chất hàng ngắn nhất	Engine dẫn rõ

Yếu tố khó

25 kiện phải giữ thẳng đứng; kiện nặng nhất 2,680 kg (5,908 lb)



Không mô hình AI nào trả về kế hoạch hợp lệ ở lần chạy này.

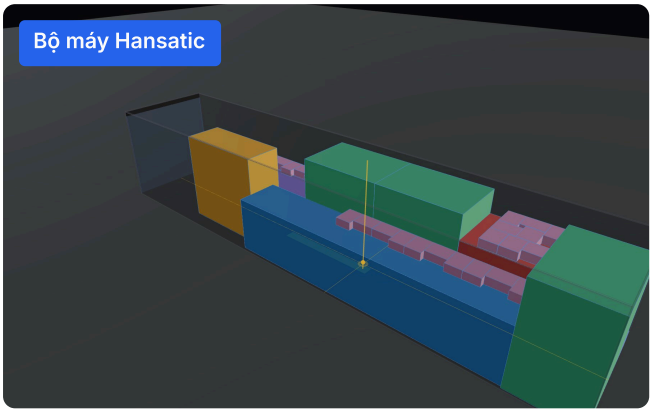
Đối thủ	Trạng thái	Đã xếp	Điểm	Ghi chú
Engine Hansatic	**Hợp pháp**	**31 / 32**	**10.51 m (34.5 ft)**	
GPT-5.5	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Opus 5.5	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Grok 4.7	Vi phạm	31 / 32	12.03 m (39.5 ft)	1 bị đè bẹp
Kimi K3.1 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
DeepSeek V4 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Gemini 4 Argon	Vi phạm	31 / 32	11.81 m (38.7 ft)	1 kích thước sai, 12 chồng lấn, 6 ngoài xe, 3 quá chiều cao, 2 lơ lửng, 3 bị đè bẹp
Qwen3p8	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Magistral Medium	Không có kế hoạch	-	-	Câu trả lời không có kế hoạch hợp lệ.

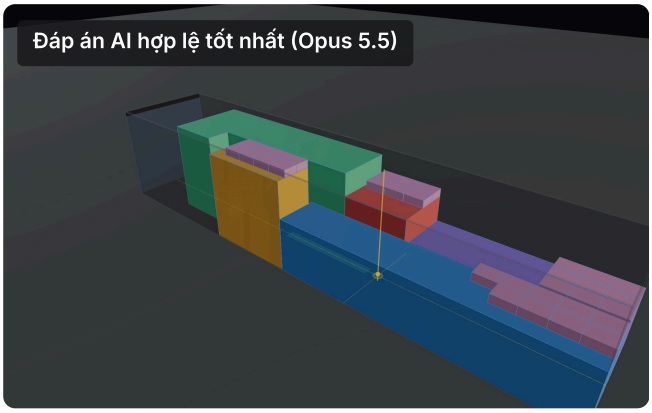
H04 Chiều dài chất hàng ngắn nhất

Phương tiện	Số kiện	Mục tiêu	Kết quả
40' Dry Container	39	Chiều dài chất hàng ngắn nhất	Hòa

Yếu tố khó

kiện dài tới 6.30 m (20.7 ft); kiện nặng nhất 998 kg (2,200 lb)





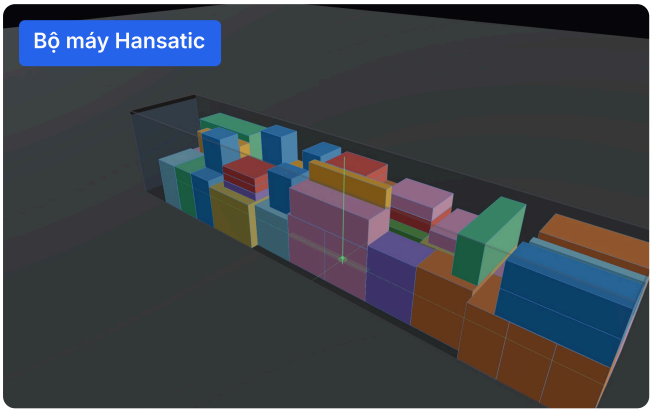
Đối thủ	Trạng thái	Đã xếp	Điểm	Ghi chú
Engine Hansatic	**Hợp pháp**	**39 / 39**	**10.44 m (34.3 ft)**	
GPT-5.5	Hợp pháp	39 / 39	10.72 m (35.2 ft)	
Opus 5.5	Hợp pháp	39 / 39	10.44 m (34.3 ft)	
Grok 4.7	Hợp pháp	39 / 39	11.10 m (36.4 ft)	
Kimi K3.1 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
DeepSeek V4 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Gemini 4 Argon	Vi phạm	29 / 39	6.30 m (20.7 ft)	10 kích thước sai, 6 lơ lửng
Qwen3p8	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Magistral Medium	Vi phạm	38 / 39	10.20 m (33.5 ft)	1 kích thước sai, 27 chồng lấn, 6 ngoài xe, 24 lơ lửng, 1 bị đè bẹp, 1 trọng tâm lệch

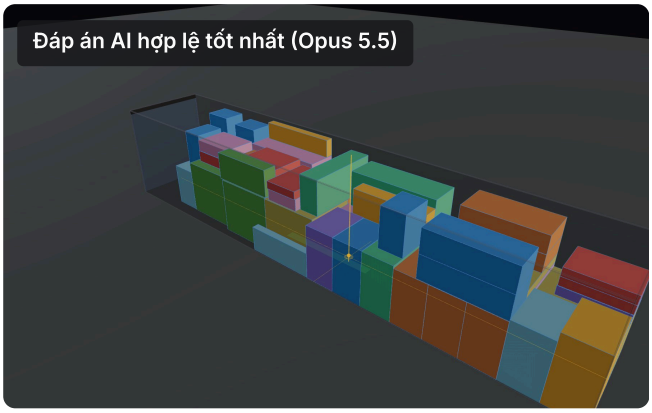
H05 Chiều dài chất hàng ngắn nhất

Phương tiện	Số kiện	Mục tiêu	Kết quả
45' High Cube	49	Chiều dài chất hàng ngắn nhất	Hòa

Yếu tố khó

49 kiện phải giữ thẳng đứng; 27 kiện có giới hạn tải trọng xếp chồng; kiện nặng nhất 814 kg (1,795 lb)





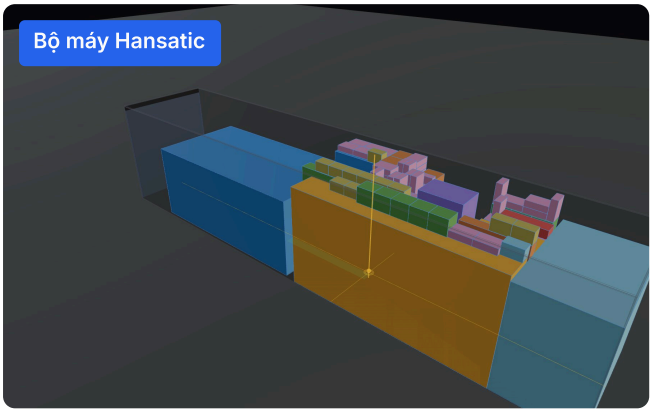
Đối thủ	Trạng thái	Đã xếp	Điểm	Ghi chú
Engine Hansatic	**Hợp pháp**	**49 / 49**	**12.40 m (40.7 ft)**	
GPT-5.5	Hợp pháp	45 / 49	13.28 m (43.6 ft)	
Opus 5.5	Hợp pháp	49 / 49	12.40 m (40.7 ft)	
Grok 4.7	Hợp pháp	43 / 49	12.96 m (42.5 ft)	
Kimi K3.1 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
DeepSeek V4 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Gemini 4 Argon	Vi phạm	37 / 49	10.49 m (34.4 ft)	11 kích thước sai, 2 chồng lấn, 5 ngoài xe, 2 quá chiều cao, 3 lơ lửng, 2 bị đè bẹp
Qwen3p8	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Magistral Medium	Vi phạm	47 / 49	12.73 m (41.8 ft)	2 kích thước sai, 31 chồng lấn, 9 ngoài xe, 14 quá chiều cao, 21 lơ lửng, 2 bị đè bẹp

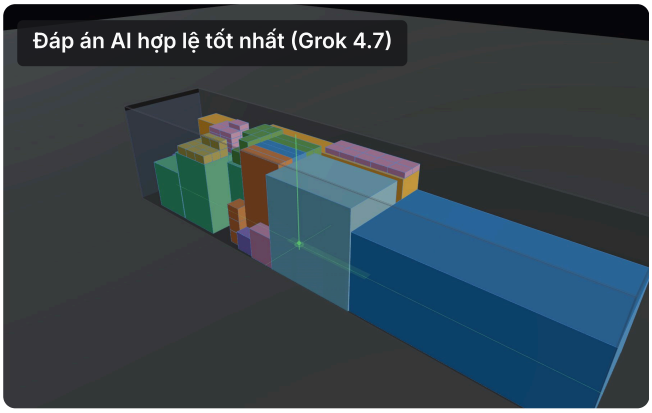
H06 Chiều dài chất hàng ngắn nhất

Phương tiện	Số kiện	Mục tiêu	Kết quả
40' High Cube	83	Chiều dài chất hàng ngắn nhất	Engine dẫn rõ

Yếu tố khó

5 kiện không xếp chồng được; 1 kiện phải giữ thẳng đứng; kiện nặng nhất 3,020 kg (6,658 lb)





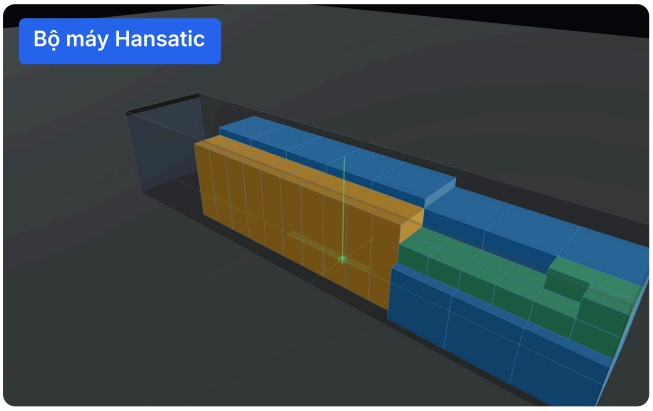
Đối thủ	Trạng thái	Đã xếp	Điểm	Ghi chú
Engine Hansatic	**Hợp pháp**	**83 / 83**	**10.65 m (35.0 ft)**	
GPT-5.5	Hợp pháp	83 / 83	11.78 m (38.7 ft)	
Opus 5.5	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Grok 4.7	Hợp pháp	83 / 83	11.41 m (37.4 ft)	
Kimi K3.1 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
DeepSeek V4 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Gemini 4 Argon	Vi phạm	7 / 83	6.89 m (22.6 ft)	2 kích thước sai, 1 trọng tâm lệch
Qwen3p8	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Magistral Medium	Vi phạm	20 / 83	12.25 m (40.2 ft)	2 chồng lấn, 4 ngoài xe, 1 quá chiều cao, 14 lơ lửng, 1 trọng tâm lệch

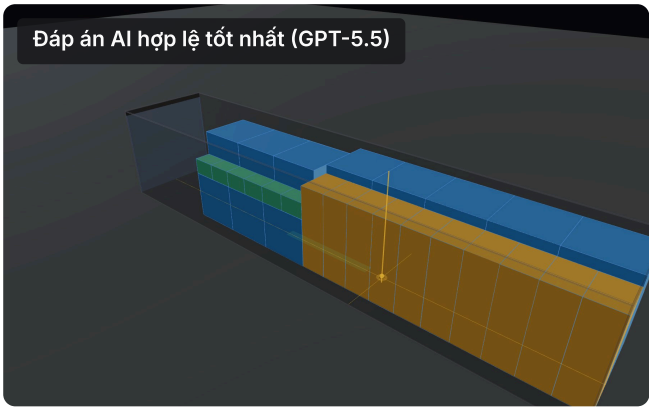
H07 Chiều dài chất hàng ngắn nhất

Phương tiện	Số kiện	Mục tiêu	Kết quả
40' Dry Container	45	Chiều dài chất hàng ngắn nhất	Hòa

Yếu tố khó

10 kiện không xếp chồng được; 15 kiện phải giữ thẳng đứng; 15 kiện có giới hạn tải trọng xếp chồng; kiện nặng nhất 1,200 kg (2,646 lb)





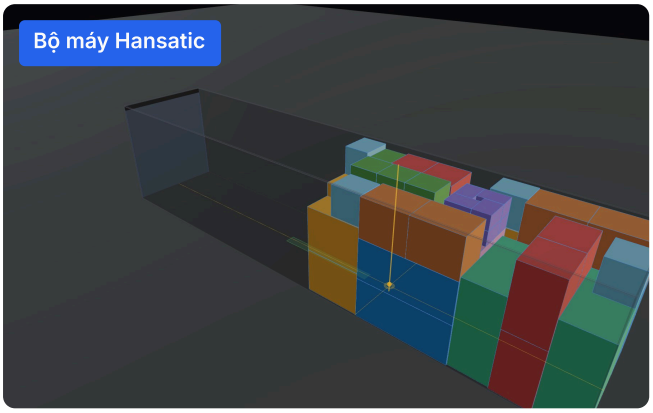
Đối thủ	Trạng thái	Đã xếp	Điểm	Ghi chú
Engine Hansatic	**Hợp pháp**	**45 / 45**	**9.60 m (31.5 ft)**	
GPT-5.5	Hợp pháp	45 / 45	9.60 m (31.5 ft)	
Opus 5.5	Hợp pháp	45 / 45	9.61 m (31.5 ft)	
Grok 4.7	Vi phạm	45 / 45	10.80 m (35.4 ft)	1 trọng tâm lệch
Kimi K3.1 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
DeepSeek V4 Pro	Hợp pháp	45 / 45	12.00 m (39.4 ft)	
Gemini 4 Argon	Vi phạm	32 / 45	7.60 m (24.9 ft)	13 kích thước sai, 1 chồng lấn, 16 lơ lửng
Qwen3p8	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Magistral Medium	Vi phạm	10 / 45	12.00 m (39.4 ft)	1 trọng tâm lệch

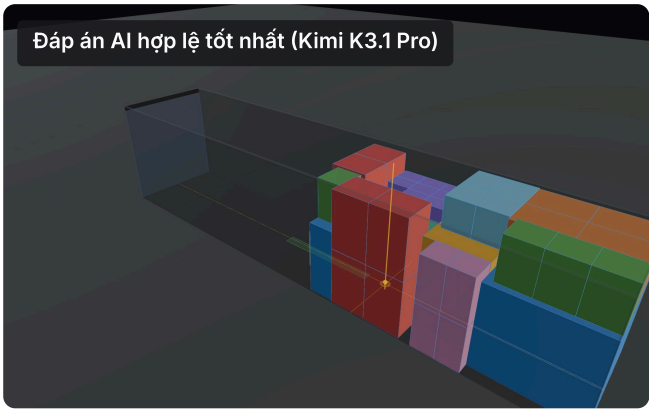
H08 Chiều dài chất hàng ngắn nhất

Phương tiện	Số kiện	Mục tiêu	Kết quả
40' High Cube	52	Chiều dài chất hàng ngắn nhất	Engine dẫn rõ

Yếu tố khó

52 kiện phải giữ thẳng đứng





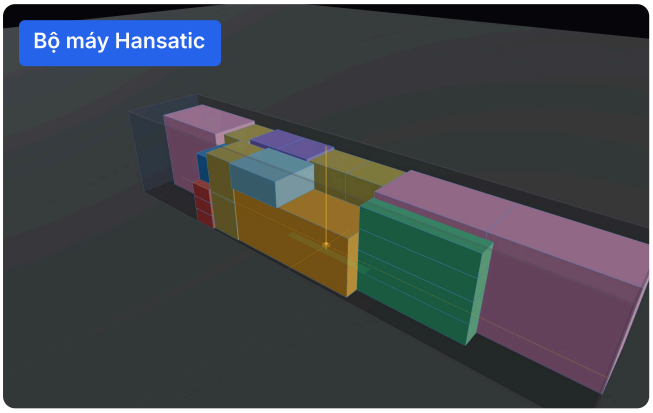
Đối thủ	Trạng thái	Đã xếp	Điểm	Ghi chú
Engine Hansatic	**Hợp pháp**	**52 / 52**	**5.92 m (19.4 ft)**	
GPT-5.5	Hợp pháp	52 / 52	6.26 m (20.5 ft)	
Opus 5.5	Hợp pháp	52 / 52	6.20 m (20.3 ft)	
Grok 4.7	Hợp pháp	52 / 52	6.18 m (20.3 ft)	
Kimi K3.1 Pro	Hợp pháp	52 / 52	6.12 m (20.1 ft)	
DeepSeek V4 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Gemini 4 Argon	Vi phạm	51 / 52	6.51 m (21.4 ft)	1 kích thước sai, 4 chồng lấn, 1 lơ lửng
Qwen3p8	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Magistral Medium	Vi phạm	52 / 52	11.25 m (36.9 ft)	14 chồng lấn, 20 ngoài xe

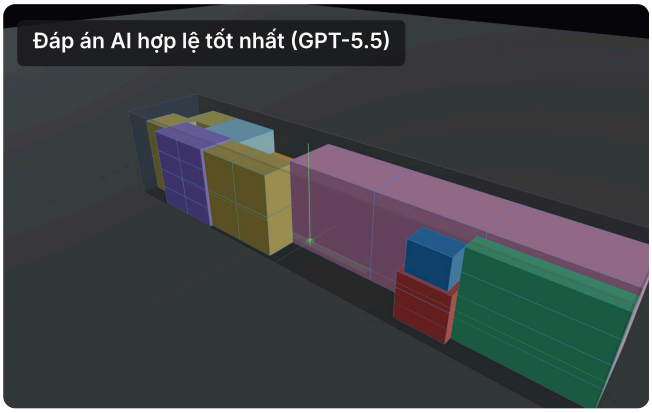
H09 Chiều dài chất hàng ngắn nhất

Phương tiện	Số kiện	Mục tiêu	Kết quả
13.6m Curtainsider	30	Chiều dài chất hàng ngắn nhất	Engine dẫn sát

Yếu tố khó

30 kiện phải giữ thẳng đứng; kiện nặng nhất 2,400 kg (5,291 lb)





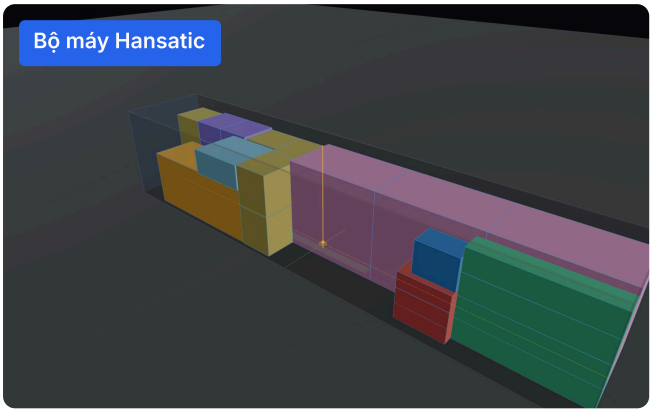
Đối thủ	Trạng thái	Đã xếp	Điểm	Ghi chú
Engine Hansatic	**Hợp pháp**	**29 / 30**	**13.36 m (43.8 ft)**	
GPT-5.5	Hợp pháp	29 / 30	13.56 m (44.5 ft)	
Opus 5.5	Hợp pháp	29 / 30	13.60 m (44.6 ft)	
Grok 4.7	Hợp pháp	29 / 30	13.60 m (44.6 ft)	
Kimi K3.1 Pro	Hợp pháp	29 / 30	13.60 m (44.6 ft)	
DeepSeek V4 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Gemini 4 Argon	Vi phạm	19 / 30	11.60 m (38.1 ft)	11 kích thước sai, 7 chồng lấn
Qwen3p8	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Magistral Medium	Vi phạm	30 / 30	13.36 m (43.8 ft)	17 chồng lấn, 25 ngoài xe, 7 lơ lửng, 1 trục quá tải

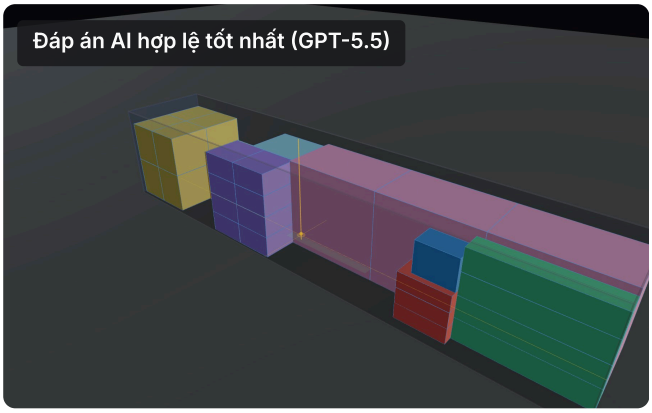
H09 Ít vị trí pallet nhất

Phương tiện	Số kiện	Mục tiêu	Kết quả
13.6m Curtainsider	30	Ít vị trí pallet nhất	Engine dẫn sát

Yếu tố khó

30 kiện phải giữ thẳng đứng; kiện nặng nhất 2,400 kg (5,291 lb)





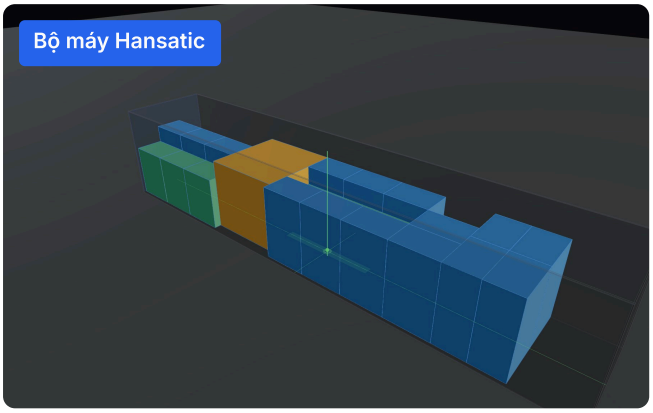
Đối thủ	Trạng thái	Đã xếp	Điểm	Ghi chú
Engine Hansatic	**Hợp pháp**	**29 / 30**	**29 vị trí, 13.36 m (43.8 ft)**	
GPT-5.5	Hợp pháp	29 / 30	29 vị trí, 13.60 m (44.6 ft)	
Opus 5.5	Hợp pháp	29 / 30	31 vị trí, 13.60 m (44.6 ft)	
Grok 4.7	Hợp pháp	29 / 30	29 vị trí, 13.62 m (44.7 ft)	
Kimi K3.1 Pro	Hợp pháp	29 / 30	29 vị trí, 13.60 m (44.6 ft)	
DeepSeek V4 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Gemini 4 Argon	Vi phạm	30 / 30	29 vị trí, 11.72 m (38.5 ft)	18 chồng lấn, 1 ngoài xe, 3 bị đè bẹp, 1 trục quá tải
Qwen3p8	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Magistral Medium	Vi phạm	30 / 30	34 vị trí, 13.36 m (43.8 ft)	4 chồng lấn, 25 ngoài xe, 1 trục quá tải, 1 hàng lệch

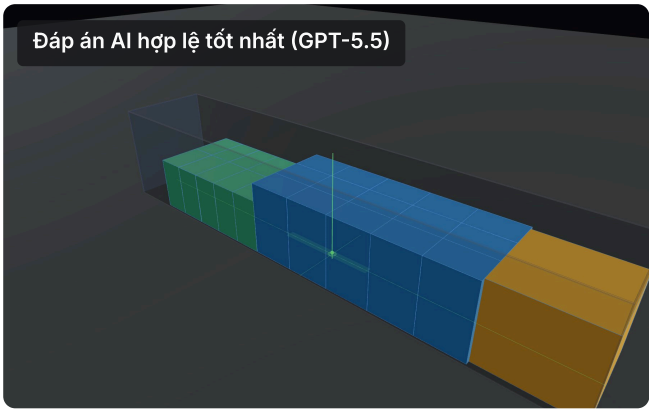
H10 Chiều dài chất hàng ngắn nhất

Phương tiện	Số kiện	Mục tiêu	Kết quả
13.6m Curtainsider	26	Chiều dài chất hàng ngắn nhất	Hòa

Yếu tố khó

26 kiện không xếp chồng được; 16 kiện phải giữ thẳng đứng; kiện nặng nhất 666 kg (1,468 lb)





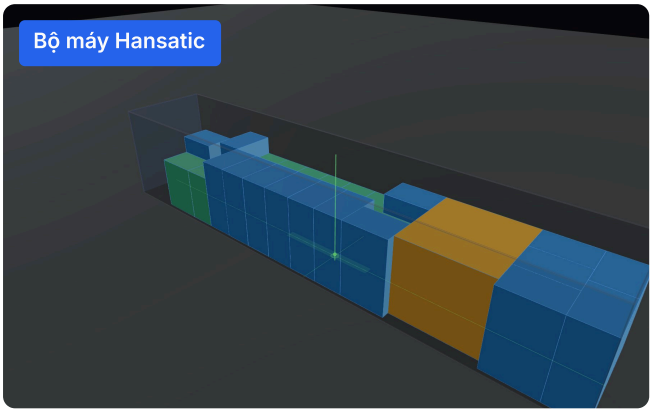
Đối thủ	Trạng thái	Đã xếp	Điểm	Ghi chú
Engine Hansatic	**Hợp pháp**	**26 / 26**	**12.20 m (40.0 ft)**	
GPT-5.5	Hợp pháp	26 / 26	12.20 m (40.0 ft)	
Opus 5.5	Hợp pháp	26 / 26	12.20 m (40.0 ft)	
Grok 4.7	Hợp pháp	26 / 26	12.20 m (40.0 ft)	
Kimi K3.1 Pro	Hợp pháp	26 / 26	12.20 m (40.0 ft)	
DeepSeek V4 Pro	Hợp pháp	26 / 26	12.20 m (40.0 ft)	
Gemini 4 Argon	Hợp pháp	26 / 26	12.60 m (41.3 ft)	
Qwen3p8	Hợp pháp	26 / 26	12.20 m (40.0 ft)	
Magistral Medium	Hợp pháp	26 / 26	13.00 m (42.7 ft)	

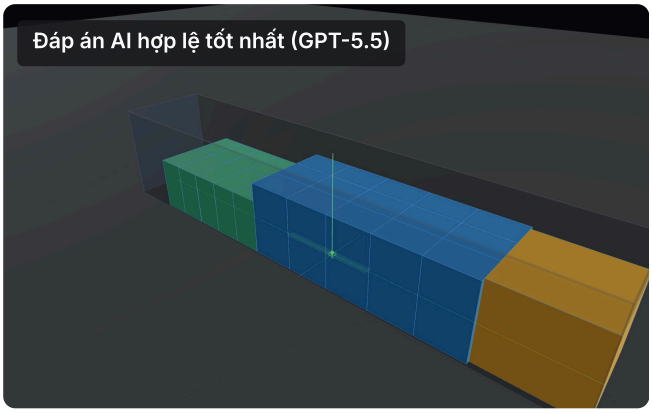
H10 Ít vị trí pallet nhất

Phương tiện	Số kiện	Mục tiêu	Kết quả
13.6m Curtainsider	26	Ít vị trí pallet nhất	Hòa

Yếu tố khó

26 kiện không xếp chồng được; 16 kiện phải giữ thẳng đứng; kiện nặng nhất 666 kg (1,468 lb)





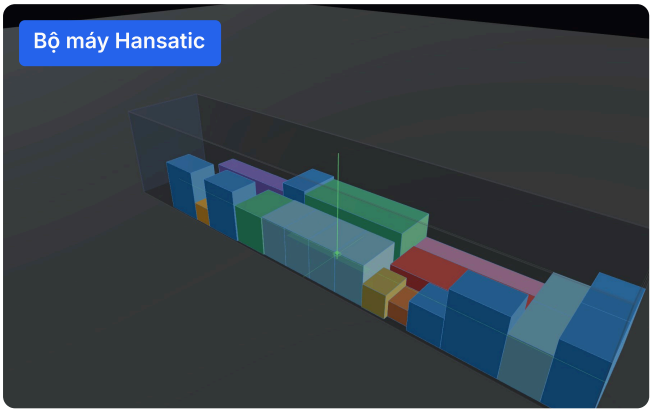
Đối thủ	Trạng thái	Đã xếp	Điểm	Ghi chú
Engine Hansatic	**Hợp pháp**	**26 / 26**	**31 vị trí, 12.20 m (40.0 ft)**	
GPT-5.5	Hợp pháp	26 / 26	31 vị trí, 12.20 m (40.0 ft)	
Opus 5.5	Hợp pháp	26 / 26	31 vị trí, 13.00 m (42.7 ft)	
Grok 4.7	Hợp pháp	26 / 26	31 vị trí, 12.20 m (40.0 ft)	
Kimi K3.1 Pro	Hợp pháp	26 / 26	31 vị trí, 12.20 m (40.0 ft)	
DeepSeek V4 Pro	Hợp pháp	26 / 26	31 vị trí, 12.20 m (40.0 ft)	
Gemini 4 Argon	Vi phạm	26 / 26	31 vị trí, 12.60 m (41.3 ft)	1 trục quá tải
Qwen3p8	Hợp pháp	26 / 26	31 vị trí, 13.00 m (42.7 ft)	
Magistral Medium	Vi phạm	26 / 26	28 vị trí, 13.00 m (42.7 ft)	10 ngoài xe

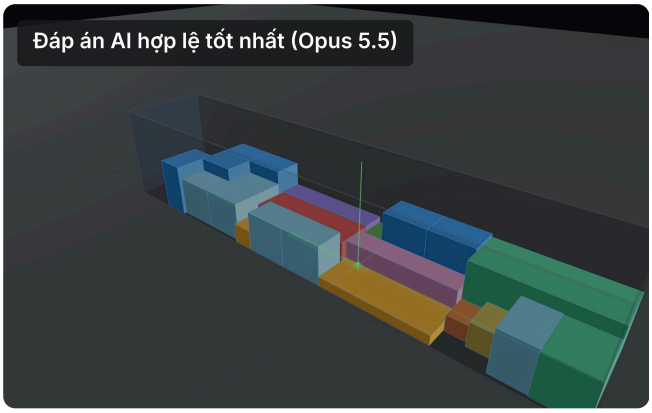
H11 Chiều dài chất hàng ngắn nhất

Phương tiện	Số kiện	Mục tiêu	Kết quả
13.6m Curtainsider	23	Chiều dài chất hàng ngắn nhất	Engine dẫn sát

Yếu tố khó

23 kiện không xếp chồng được; 23 kiện phải giữ thẳng đứng; kiện nặng nhất 2,700 kg (5,952 lb)





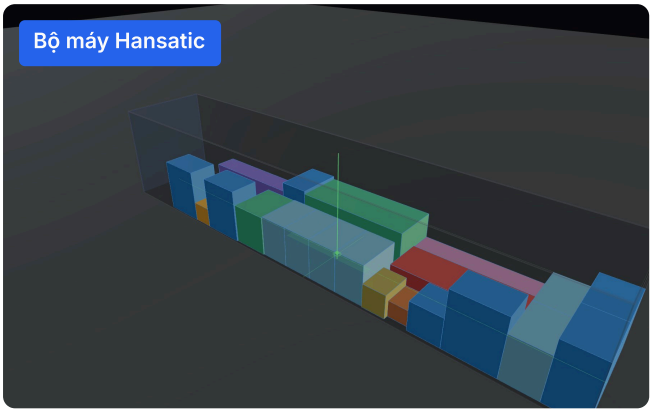
Đối thủ	Trạng thái	Đã xếp	Điểm	Ghi chú
Engine Hansatic	**Hợp pháp**	**23 / 23**	**12.00 m (39.4 ft)**	
GPT-5.5	Hợp pháp	23 / 23	12.42 m (40.7 ft)	
Opus 5.5	Hợp pháp	23 / 23	12.20 m (40.0 ft)	
Grok 4.7	Hợp pháp	23 / 23	12.62 m (41.4 ft)	
Kimi K3.1 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
DeepSeek V4 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Gemini 4 Argon	Hợp pháp	23 / 23	13.02 m (42.7 ft)	
Qwen3p8	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Magistral Medium	Vi phạm	23 / 23	12.34 m (40.5 ft)	7 chồng lấn, 8 ngoài xe

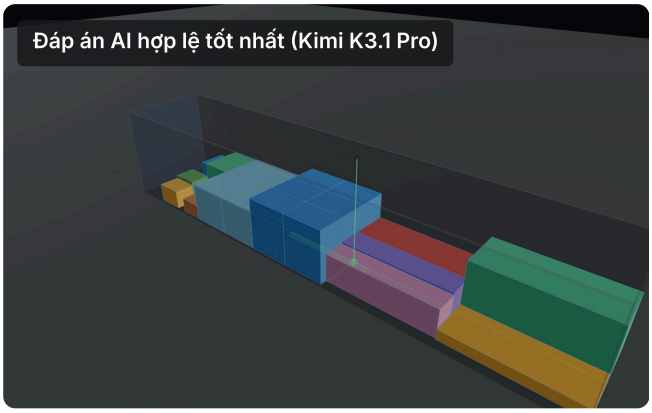
H11 Ít vị trí pallet nhất

Phương tiện	Số kiện	Mục tiêu	Kết quả
13.6m Curtainsider	23	Ít vị trí pallet nhất	Engine dẫn rõ

Yếu tố khó

23 kiện không xếp chồng được; 23 kiện phải giữ thẳng đứng; kiện nặng nhất 2,700 kg (5,952 lb)





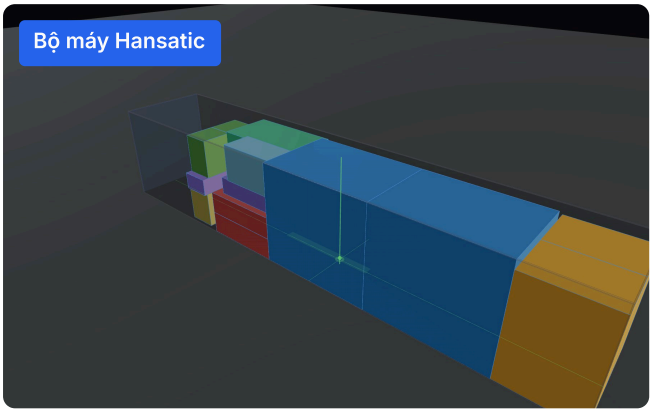
Đối thủ	Trạng thái	Đã xếp	Điểm	Ghi chú
Engine Hansatic	**Hợp pháp**	**23 / 23**	**30 vị trí, 12.00 m (39.4 ft)**	
GPT-5.5	Hợp pháp	23 / 23	32 vị trí, 12.64 m (41.5 ft)	
Opus 5.5	Hợp pháp	23 / 23	32 vị trí, 12.84 m (42.1 ft)	
Grok 4.7	Lỗi nhà cung cấp	-	-	Nhà cung cấp từ chối yêu cầu.
Kimi K3.1 Pro	Hợp pháp	23 / 23	32 vị trí, 12.62 m (41.4 ft)	
DeepSeek V4 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Gemini 4 Argon	Vi phạm	22 / 23	30 vị trí, 12.44 m (40.8 ft)	1 kích thước sai
Qwen3p8	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Magistral Medium	Vi phạm	23 / 23	30 vị trí, 12.60 m (41.3 ft)	11 chồng lấn, 9 ngoài xe

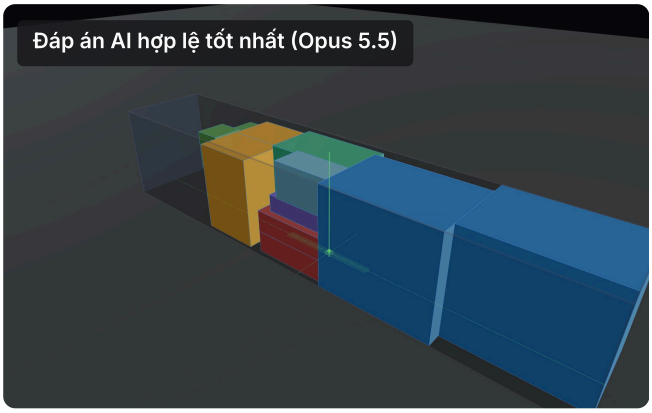
H12 Chiều dài chất hàng ngắn nhất

Phương tiện	Số kiện	Mục tiêu	Kết quả
13.6m Curtainsider	14	Chiều dài chất hàng ngắn nhất	Hòa

Yếu tố khó

6 kiện không xếp chồng được; 9 kiện phải giữ thẳng đứng; kiện nặng nhất 732 kg (1,614 lb)





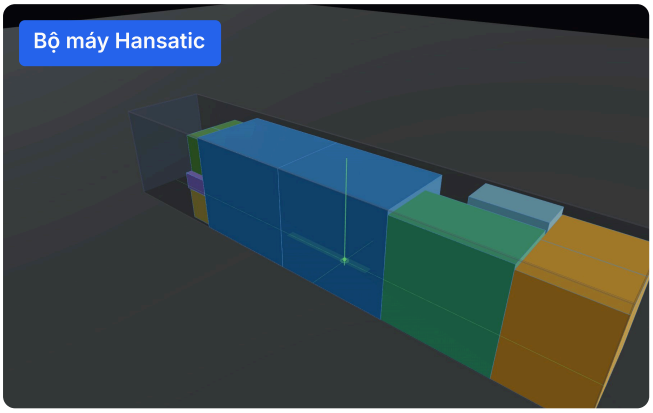
Đối thủ	Trạng thái	Đã xếp	Điểm	Ghi chú
Engine Hansatic	**Hợp pháp**	**13 / 14**	**11.00 m (36.1 ft)**	
GPT-5.5	Hợp pháp	13 / 14	11.40 m (37.4 ft)	
Opus 5.5	Hợp pháp	13 / 14	11.00 m (36.1 ft)	
Grok 4.7	Lỗi nhà cung cấp	-	-	Nhà cung cấp từ chối yêu cầu.
Kimi K3.1 Pro	Hợp pháp	13 / 14	11.00 m (36.1 ft)	
DeepSeek V4 Pro	Hợp pháp	13 / 14	11.70 m (38.4 ft)	
Gemini 4 Argon	Hợp pháp	13 / 14	11.00 m (36.1 ft)	
Qwen3p8	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Magistral Medium	Vi phạm	14 / 14	13.50 m (44.3 ft)	5 chồng lấn, 5 ngoài xe, 2 lơ lửng

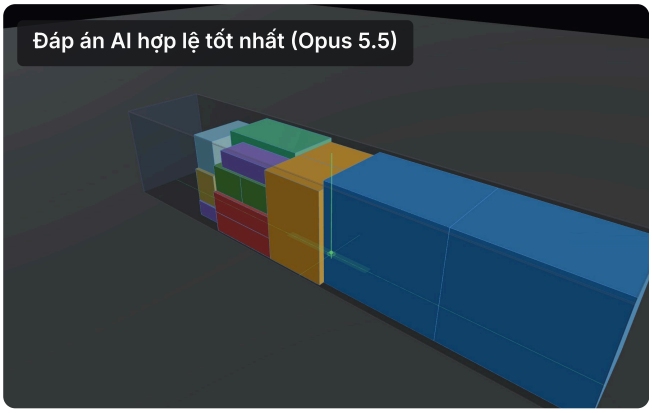
H12 Ít vị trí pallet nhất

Phương tiện	Số kiện	Mục tiêu	Kết quả
13.6m Curtainsider	14	Ít vị trí pallet nhất	Engine dẫn rõ

Yếu tố khó

6 kiện không xếp chồng được; 9 kiện phải giữ thẳng đứng; kiện nặng nhất 732 kg (1,614 lb)





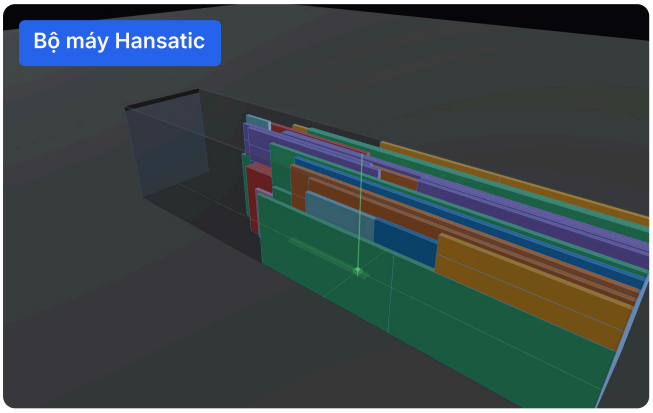
Đối thủ	Trạng thái	Đã xếp	Điểm	Ghi chú
Engine Hansatic	**Hợp pháp**	**13 / 14**	**25 vị trí, 11.00 m (36.1 ft)**	
GPT-5.5	Hợp pháp	13 / 14	29 vị trí, 11.73 m (38.5 ft)	
Opus 5.5	Hợp pháp	13 / 14	28 vị trí, 11.00 m (36.1 ft)	
Grok 4.7	Lỗi nhà cung cấp	-	-	Nhà cung cấp từ chối yêu cầu.
Kimi K3.1 Pro	Hợp pháp	13 / 14	28 vị trí, 11.00 m (36.1 ft)	
DeepSeek V4 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Gemini 4 Argon	Hợp pháp	13 / 14	28 vị trí, 11.15 m (36.6 ft)	
Qwen3p8	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Magistral Medium	Vi phạm	14 / 14	33 vị trí, 13.20 m (43.3 ft)	9 chồng lấn, 9 ngoài xe

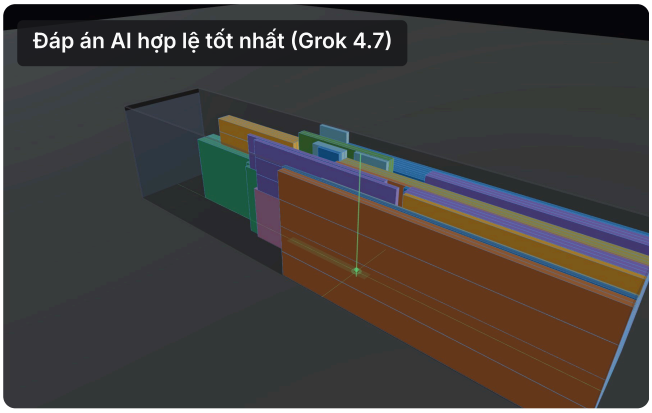
BV1 Chiều dài chất hàng ngắn nhất

Phương tiện	Số kiện	Mục tiêu	Kết quả
40' High Cube	103	Chiều dài chất hàng ngắn nhất	Engine dẫn rõ

Yếu tố khó

37 kiện phải giữ thẳng đứng; kiện dài tới 8.27 m (27.1 ft)





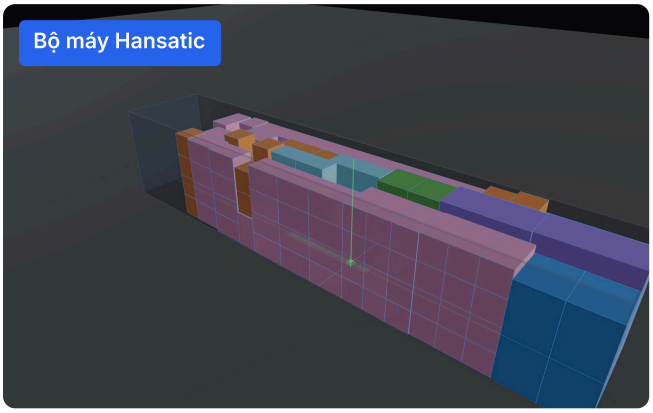
Đối thủ	Trạng thái	Đã xếp	Điểm	Ghi chú
Engine Hansatic	**Hợp pháp**	**103 / 103**	**9.17 m (30.1 ft)**	
GPT-5.5	Hợp pháp	92 / 103	8.27 m (27.1 ft)	
Opus 5.5	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Grok 4.7	Hợp pháp	103 / 103	10.66 m (35.0 ft)	
Kimi K3.1 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
DeepSeek V4 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Gemini 4 Argon	Vi phạm	77 / 103	8.36 m (27.4 ft)	1 kích thước sai, 4 quá chiều cao, 4 lơ lửng
Qwen3p8	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Magistral Medium	Vi phạm	22 / 103	16.54 m (54.3 ft)	9 ngoài xe

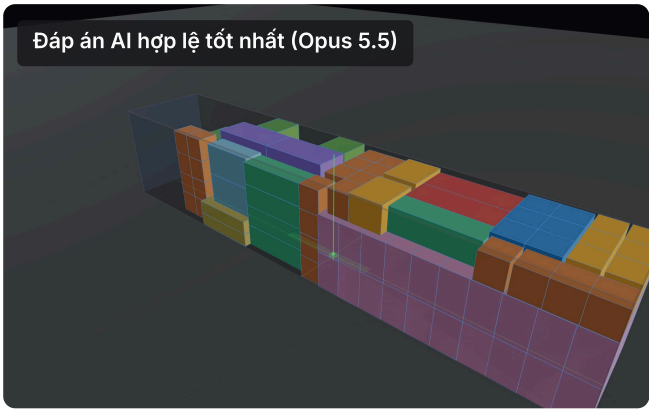
BV2 Chiều dài chất hàng ngắn nhất

Phương tiện	Số kiện	Mục tiêu	Kết quả
13.6m Curtainsider	153	Chiều dài chất hàng ngắn nhất	Engine dẫn rõ

Yếu tố khó

2 kiện không xếp chồng được; kiện nặng nhất 578 kg (1,274 lb)





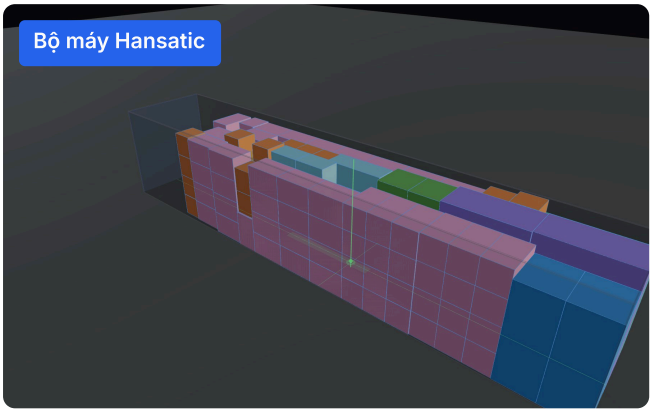
Đối thủ	Trạng thái	Đã xếp	Điểm	Ghi chú
Engine Hansatic	**Hợp pháp**	**153 / 153**	**11.52 m (37.8 ft)**	
GPT-5.5	Hợp pháp	153 / 153	13.14 m (43.1 ft)	
Opus 5.5	Hợp pháp	153 / 153	11.78 m (38.6 ft)	
Grok 4.7	Hợp pháp	153 / 153	12.30 m (40.4 ft)	
Kimi K3.1 Pro	Hợp pháp	153 / 153	11.99 m (39.3 ft)	
DeepSeek V4 Pro	Hợp pháp	153 / 153	12.79 m (42.0 ft)	
Gemini 4 Argon	Vi phạm	147 / 153	12.77 m (41.9 ft)	6 kích thước sai
Qwen3p8	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Magistral Medium	Vi phạm	56 / 153	14.02 m (46.0 ft)	39 chồng lấn, 27 ngoài xe

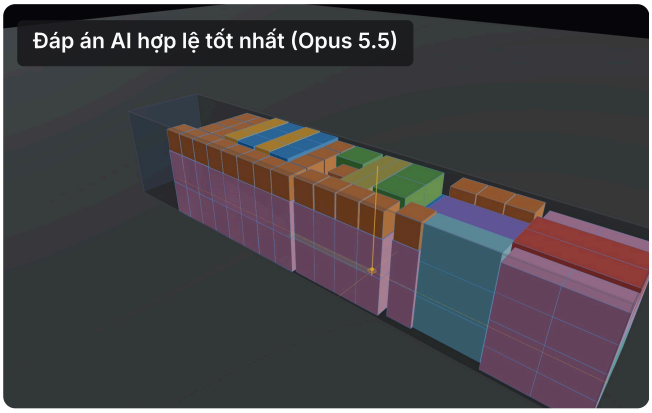
BV2 Ít vị trí pallet nhất

Phương tiện	Số kiện	Mục tiêu	Kết quả
13.6m Curtainsider	153	Ít vị trí pallet nhất	Engine dẫn sát

Yếu tố khó

2 kiện không xếp chồng được; kiện nặng nhất 578 kg (1,274 lb)





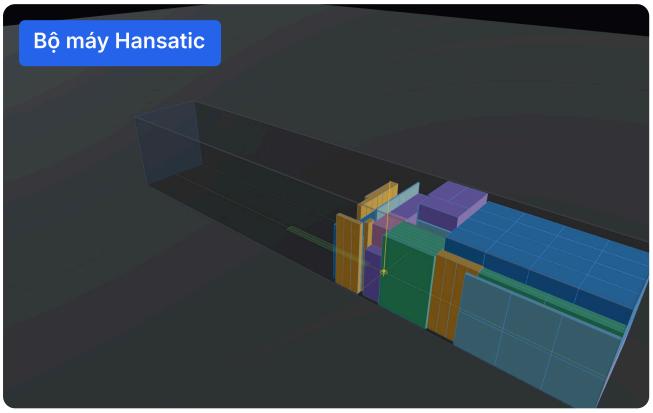
Đối thủ	Trạng thái	Đã xếp	Điểm	Ghi chú
Engine Hansatic	**Hợp pháp**	**153 / 153**	**29 vị trí, 11.52 m (37.8 ft)**	
GPT-5.5	Hợp pháp	153 / 153	32 vị trí, 12.60 m (41.3 ft)	
Opus 5.5	Hợp pháp	153 / 153	29 vị trí, 11.67 m (38.3 ft)	
Grok 4.7	Hợp pháp	153 / 153	32 vị trí, 12.78 m (41.9 ft)	
Kimi K3.1 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
DeepSeek V4 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Gemini 4 Argon	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Qwen3p8	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Magistral Medium	Vi phạm	153 / 153	34 vị trí, 14.82 m (48.6 ft)	7 chồng lấn, 31 ngoài xe, 86 quá chiều cao, 96 lơ lửng

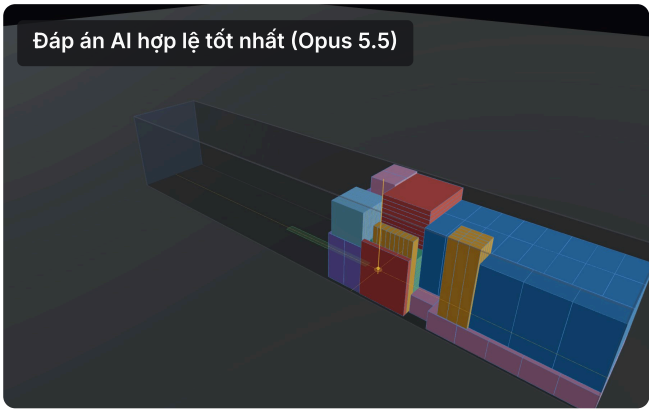
BV3 Chiều dài chất hàng ngắn nhất

Phương tiện	Số kiện	Mục tiêu	Kết quả
53ft Dry Van	119	Chiều dài chất hàng ngắn nhất	Engine dẫn rõ

Yếu tố khó

19 kiện không xếp chồng được; 35 kiện phải giữ thẳng đứng





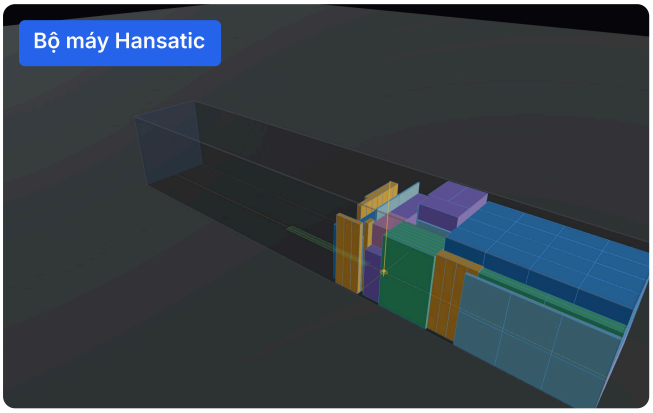
Đối thủ	Trạng thái	Đã xếp	Điểm	Ghi chú
Engine Hansatic	**Hợp pháp**	**119 / 119**	**6.83 m (22.4 ft)**	
GPT-5.5	Hợp pháp	119 / 119	9.14 m (30.0 ft)	
Opus 5.5	Hợp pháp	119 / 119	6.99 m (22.9 ft)	
Grok 4.7	Hợp pháp	119 / 119	8.10 m (26.6 ft)	
Kimi K3.1 Pro	Không có kế hoạch	-	-	Câu trả lời không có kế hoạch hợp lệ.
DeepSeek V4 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Gemini 4 Argon	Vi phạm	115 / 119	8.36 m (27.4 ft)	4 kích thước sai, 5 chồng lấn, 3 bị đè bẹp
Qwen3p8	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Magistral Medium	Vi phạm	22 / 119	16.76 m (55.0 ft)	11 ngoài xe

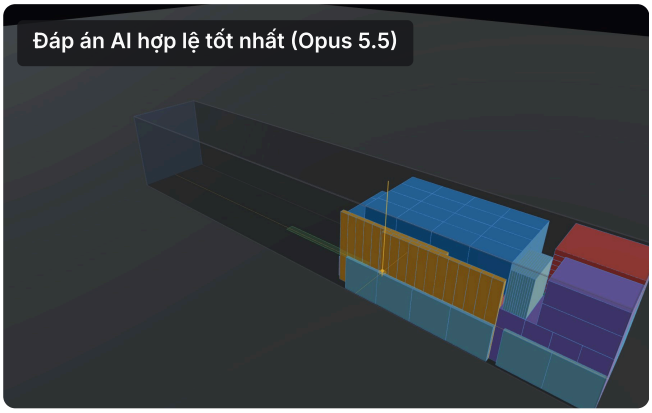
BV3 Ít vị trí pallet nhất

Phương tiện	Số kiện	Mục tiêu	Kết quả
53ft Dry Van	119	Ít vị trí pallet nhất	Engine dẫn rõ

Yếu tố khó

19 kiện không xếp chồng được; 35 kiện phải giữ thẳng đứng





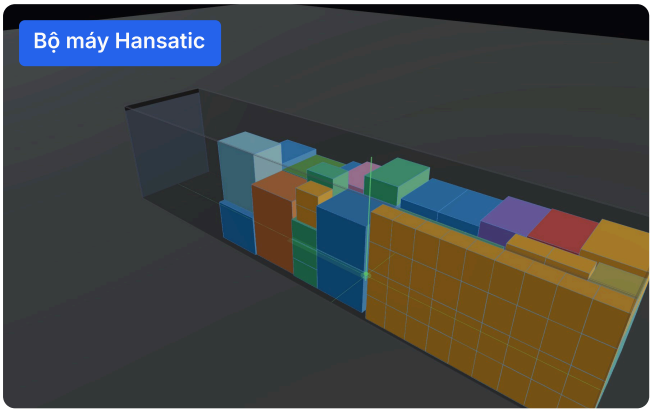
Đối thủ	Trạng thái	Đã xếp	Điểm	Ghi chú
Engine Hansatic	**Hợp pháp**	**119 / 119**	**17 vị trí, 6.83 m (22.4 ft)**	
GPT-5.5	Hợp pháp	119 / 119	23 vị trí, 9.07 m (29.8 ft)	
Opus 5.5	Hợp pháp	119 / 119	18 vị trí, 6.99 m (22.9 ft)	
Grok 4.7	Hợp pháp	119 / 119	22 vị trí, 8.64 m (28.3 ft)	
Kimi K3.1 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
DeepSeek V4 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Gemini 4 Argon	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Qwen3p8	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Magistral Medium	Vi phạm	107 / 119	21 vị trí, 16.18 m (53.1 ft)	12 kích thước sai, 66 chồng lấn, 35 ngoài xe, 60 lơ lửng, 1 bị đè bẹp, 1 nhóm rời không được chèn

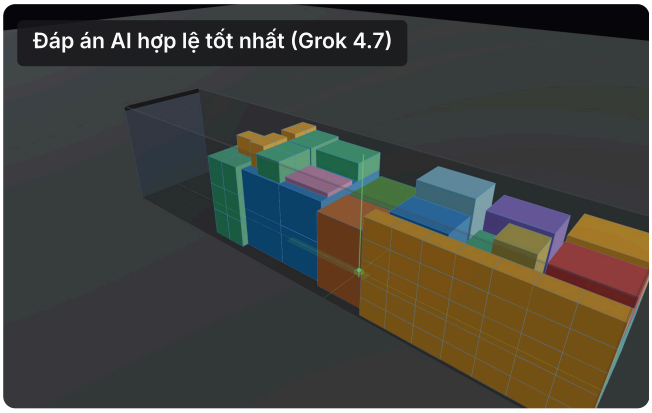
BV4 Chiều dài chất hàng ngắn nhất

Phương tiện	Số kiện	Mục tiêu	Kết quả
40' High Cube	145	Chiều dài chất hàng ngắn nhất	Engine dẫn rõ

Yếu tố khó

11 kiện không xếp chồng được; 11 kiện dễ vỡ, chỉ ở trên cùng; 20 kiện phải giữ thẳng đứng





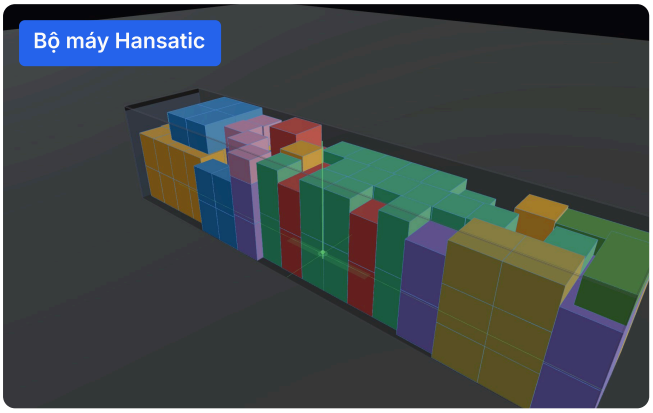
Đối thủ	Trạng thái	Đã xếp	Điểm	Ghi chú
Engine Hansatic	**Hợp pháp**	**145 / 145**	**8.57 m (28.1 ft)**	
GPT-5.5	Hợp pháp	145 / 145	11.96 m (39.2 ft)	
Opus 5.5	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Grok 4.7	Hợp pháp	145 / 145	8.85 m (29.0 ft)	
Kimi K3.1 Pro	Hợp pháp	145 / 145	10.87 m (35.7 ft)	
DeepSeek V4 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Gemini 4 Argon	Vi phạm	145 / 145	10.05 m (33.0 ft)	19 chồng lấn, 3 ngoài xe, 9 quá chiều cao, 4 lơ lửng
Qwen3p8	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Magistral Medium	Hợp pháp	11 / 145	10.33 m (33.9 ft)	

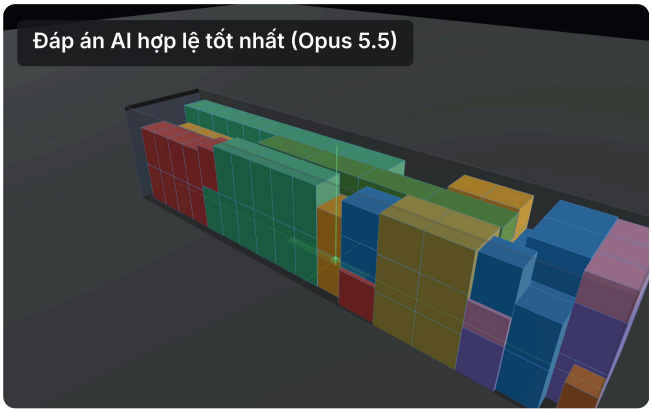
BV5 Chiều dài chất hàng ngắn nhất

Phương tiện	Số kiện	Mục tiêu	Kết quả
40' High Cube	136	Chiều dài chất hàng ngắn nhất	Engine dẫn rõ

Yếu tố khó

136 kiện phải giữ thẳng đứng





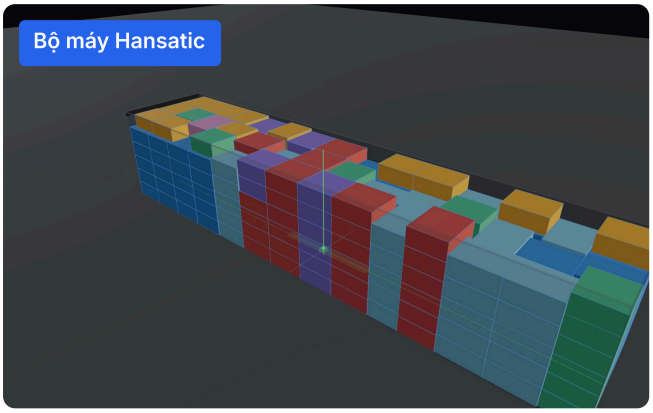
Đối thủ	Trạng thái	Đã xếp	Điểm	Ghi chú
Engine Hansatic	**Hợp pháp**	**127 / 136**	**12.00 m (39.4 ft)**	
GPT-5.5	Hợp pháp	115 / 136	11.73 m (38.5 ft)	
Opus 5.5	Hợp pháp	124 / 136	11.77 m (38.6 ft)	
Grok 4.7	Hợp pháp	115 / 136	11.87 m (39.0 ft)	
Kimi K3.1 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
DeepSeek V4 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Gemini 4 Argon	Vi phạm	121 / 136	12.00 m (39.4 ft)	1 nghiêng hàng phải đứng thẳng, 12 chồng lấn, 6 ngoài xe, 6 lơ lửng
Qwen3p8	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Magistral Medium	Vi phạm	51 / 136	12.89 m (42.3 ft)	34 ngoài xe, 6 lơ lửng

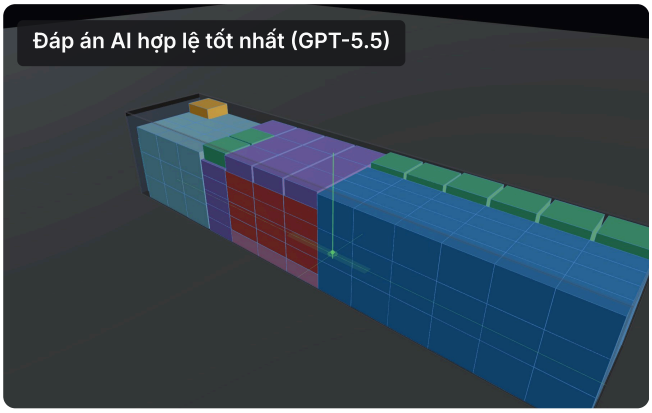
BV6 Chiều dài chất hàng ngắn nhất

Phương tiện	Số kiện	Mục tiêu	Kết quả
40' Dry Container	253	Chiều dài chất hàng ngắn nhất	Engine dẫn rõ

Yếu tố khó

nhiều kích cỡ, không cờ đặc biệt





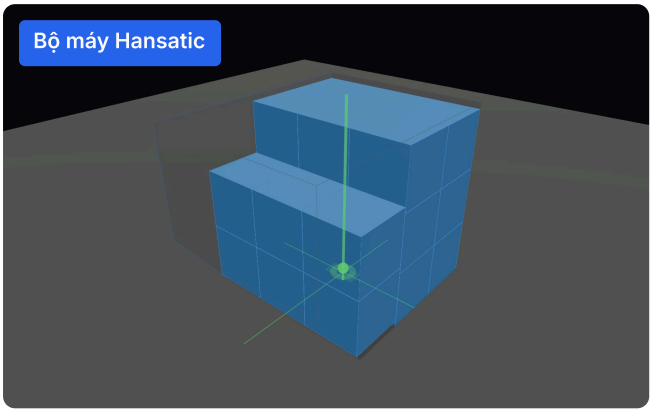
Đối thủ	Trạng thái	Đã xếp	Điểm	Ghi chú
Engine Hansatic	**Hợp pháp**	**244 / 253**	**11.98 m (39.3 ft)**	
GPT-5.5	Hợp pháp	238 / 253	11.97 m (39.3 ft)	
Opus 5.5	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Grok 4.7	Hợp pháp	218 / 253	11.98 m (39.3 ft)	
Kimi K3.1 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
DeepSeek V4 Pro	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Gemini 4 Argon	Vi phạm	197 / 253	11.50 m (37.7 ft)	13 chồng lấn
Qwen3p8	Không có kế hoạch	-	-	Mô hình đạt giới hạn đầu ra trước khi viết kế hoạch.
Magistral Medium	Vi phạm	42 / 253	12.18 m (40.0 ft)	3 ngoài xe

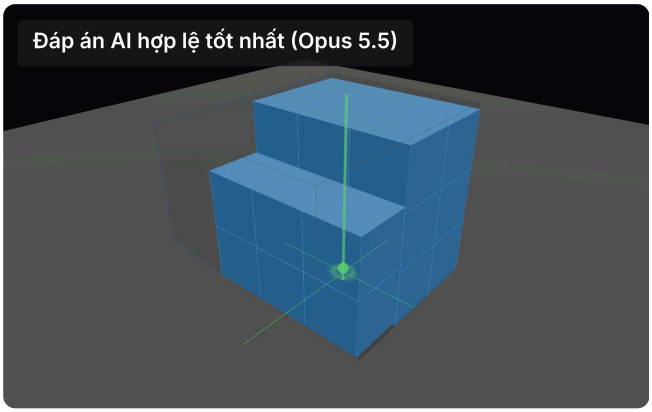
U01 Nhiều kiện nhất

Phương tiện	Số kiện	Mục tiêu	Kết quả
AKE (LD3)	50	Nhiều kiện nhất	Hòa

Yếu tố khó

vách nghiêng





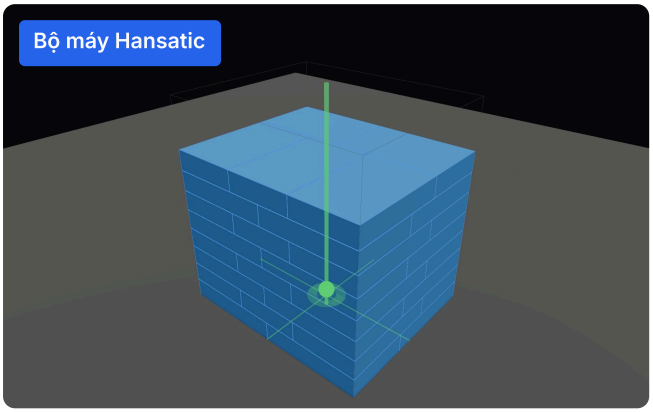
Đối thủ	Trạng thái	Đã xếp	Điểm	Ghi chú
Engine Hansatic	Hợp pháp	**24 / 50**	**24**	
GPT-5.5	Hợp pháp	18 / 50	18	
Opus 5.5	Hợp pháp	24 / 50	24	
Grok 4.7	Hợp pháp	24 / 50	24	
Kimi K3.1 Pro	Hợp pháp	24 / 50	24	
DeepSeek V4 Pro	Hợp pháp	18 / 50	18	
Gemini 4 Argon	Hợp pháp	24 / 50	24	
Qwen3p8	Hợp pháp	24 / 50	24	
Magistral Medium	Vi phạm	24 / 50	24	3 ngoài biên dạng ULD

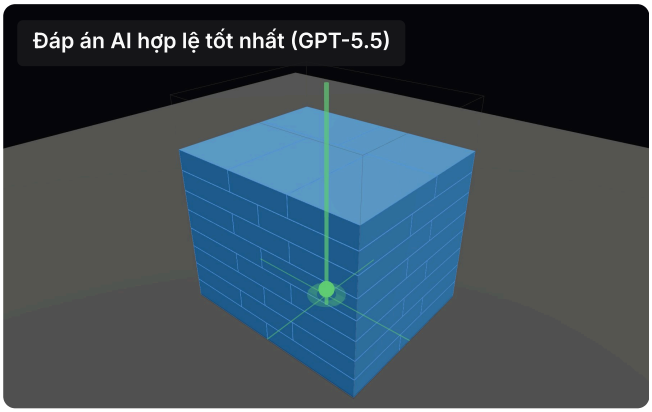
P01 Pallet liên kết nhiều nhất

Phương tiện	Số kiện	Mục tiêu	Kết quả
Industrial / Block 1200×1000	40	Pallet liên kết nhiều nhất	Hòa

Yếu tố khó

40 kiện phải giữ thẳng đứng; kiện dài tới 0.60 m (2.0 ft)





Đối thủ	Trạng thái	Đã xếp	Điểm	Ghi chú
Engine Hansatic	**Hợp pháp**	**40 / 40**	**100%**	
GPT-5.5	Hợp pháp	40 / 40	100%	
Opus 5.5	Hợp pháp	40 / 40	100%	
Grok 4.7	Hợp pháp	40 / 40	100%	
Kimi K3.1 Pro	Hợp pháp	40 / 40	100%	
DeepSeek V4 Pro	Hợp pháp	40 / 40	100%	
Gemini 4 Argon	Vi phạm	35 / 40	100%	5 kích thước sai, 12 lơ lửng
Qwen3p8	Hợp pháp	40 / 40	100%	
Magistral Medium	Vi phạm	40 / 40	100%	30 chồng lấn, 9 ngoài xe