Hansatic

TECHNICAL WHITEPAPER

Neuro-symbolic load planning

A packing engine and eight frontier AI models on real customer loads.

November 2026

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

Companion page, with every plan in 3D: hansatic.com/en/neuro-symbolic-load-planning

Abstract

We are combining a deterministic packing engine with generative AI. Before building the combination, we measured each part on its own.

- **Data:** 20 real customer load-planning cases, 26 scored runs. They cover sea containers, curtainsiders, a 53 ft (16.2 m) dry van, an AKE air container and an interlocked pallet build.
- **Entrants:** Hansatic's production packing engine and eight mainstream frontier LLMs. Each model must give a position and orientation for every item in one JSON answer, with no tools.
- **Judge:** one deterministic validator scores every entrant the same way. It checks:
  - geometry;
  - 95% base support;
  - stacking and fragile limits;
  - orientation flags;
  - slender piles;
  - payload and axle loads;
  - a longitudinal centre-of-gravity band.

Results:

- **The engine** is legal on 26 of 26 runs. Against the best legal model answer on each run, it is ahead on 18 (13 clearly, 5 narrowly) and tied on 8. It is behind on none.
- **The strongest model,** GPT-5.5, is legal on 25 of 26 runs, including manifests of up to 253 pieces.
- **Four of the eight models** are legal on fewer than a third of the runs.
- **What models are good at.** The best model answers used two layout strategies the engine did not have. Both are now deterministic engine strategies.

We conclude that the engine is the better planner today, and that model creativity is worth capturing. Our architecture for that is neuro-symbolic:

- models propose strategies;
- the engine builds every plan;
- the validator keeps only legal improvements.

We release the cases, the validator, the prompts and the scoring code.

1. Benchmark design

1.1 Cases

Every case is a real customer load from the Hansatic app, copied 1:1: dimensions, weights, quantities, flags and vehicle. Item ids are replaced and cargo names removed.

Set	Cases	Runs	Vehicles	Pieces
Standard (H01-H12)	12	16	one 20', two 40' dry, four 40' HC, one 45' HC, four 13.6 m (44.6 ft) curtainsiders	14-83
Large (BV1-BV6)	6	8	three 40' HC, one 40' dry, a 13.6 m (44.6 ft) curtainsider, a 53 ft (16.2 m) dry van	103-253
Special (U01, P01)	2	2	AKE (LD3) with a sloped wall; 1200 × 1000 mm (47 × 39 in) pallet	50 and 40

Runs and objectives:

- **Trucks** are scored twice: on loading metres (LDM) and on pallet positions. That gives 16 runs from the 12 standard cases and 8 from the 6 large ones.
- **BV5 and BV6** do not fit in full. Placing more items is the main objective there.
- **U01 and P01** each have a known optimum:
  - U01: 24 of 50 cubes, because the sloped wall allows no more;
  - P01: 40 of 40 bags at 100% interlock.

Constraints in the cases:

- non-stackable and top-only (fragile) items;
- per-item stack-weight limits;
- this-way-up and no-tilt flags;
- long bars (BV1, up to the container's length);
- a single dominant heavy item (H02);
- mixed pallet and carton loads (BV3).

1.2 Validator

The validator (v2.3.0) checks every plan the same way. A plan is legal only if every hard check passes.

Check	Rule
Bounds	every item inside the cargo space, including the AKE slope
Overlap	no two items share volume

Check	Rule
Support	at least 95% of each item's base rests on the floor or on items below
Stacking	non-stackable items carry nothing; top-only items carry nothing and sit on top; per-item stack-weight limits hold
Orientation	this-way-up and no-tilt flags respected
Slender piles	height-to-base ratio limit on free-standing stacks
Payload	total mass within the vehicle limit
Axles (trucks)	front and rear axle loads within limits, computed from item centres of mass
CoG (containers)	longitudinal centre of gravity inside the allowed band
Multi-stop	last in, first out by stop, where the case has stops

Tip-over risk is computed and reported, but not scored.

1.3 Ranking

Applied per run:

1. A legal plan beats an illegal one.
2. More items placed wins.
3. The run's objective decides:
    - a shorter load (LDM);
    - or fewer pallet positions;
    - or, for P01, a higher interlock share, then a lower height.
4. A tie on pallet positions is broken by LDM.

One counter counts pallet positions for every entrant: Hansatic's production counter.

Margins. We call a win:

- **clear** if it is at least 2% on LDM, at least one pallet position, or at least one more item placed;
- **narrow** otherwise.

The engine runs inside a time budget, so a single-run difference under 2% can be run-to-run variance (section 6.2).

1.4 Entrants

Models. The mainstream frontier model of each of eight labs:

- 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)

Settings:

- **Runs:** one per case.
- **Output limit:** 64k tokens, or the provider's lower maximum (Gemini: 32k).
- **Batching:** provider batch APIs where offered.
- **Answers collected:** 2026-11-11.
- **Total cost:** about $64.

Engine. Hansatic's production packer, version of 2026-11-20 (commit 9f66537c). Settings:

- the production configuration: 4 workers and the production time budget, pinned to one CPU core;
- one run per case, no best-of;
- it packs to its own 60% support rule, and is judged at 95% like every model.

1.5 Prompts

- One prompt per vehicle kind: container, truck, ULD and pallet.
- Each prompt states the exact rules the validator applies to that kind, the objective, and the output schema (item id, position, orientation).
- No solved layouts are given as examples.

The full templates are in Appendix A.

2. Results

2.1 Per model, all 26 runs

Model	Legal	Ties engine	Behind 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

No model beats the engine on any run.

Standard set (16 runs).

Model	Legal	Ties	Behind
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

Large set (8 runs). No model ties the engine on any large run.

Model	Legal
GPT-5.5	8
Grok 4.7	8
Opus 5.5	5

Model	Legal
Kimi K3.1 Pro	2
DeepSeek V4 Pro	1
Magistral Medium	1
Gemini 4 Argon	0
Qwen3p8	0

Special set (2 runs).

Model	Legal	Reaches the optimum
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 against the best legal model answer

Set	Clearly ahead	Narrowly ahead	Tie	Behind
Standard (16)	7	3	6	0
Large (8)	6	2	0	0
Special (2)	0	0	2	0
All (26)	**13**	**5**	**8**	**0**

Standard set. Each cell gives items placed, then load length or pallet positions.

Run	Engine	Best legal model	Margin
H01	26/27, 11.69 m (38.4 ft)	Opus 26, 11.74 m (38.5 ft)	narrow (0.4%)
H02	24/29, 5.78 m (19.0 ft)	GPT 18, 5.76 m (18.9 ft)	clear (+6 placed)
H03	31/32, 10.51 m (34.5 ft)	none legal	clear
H04	39, 10.44 m (34.3 ft)	Opus 39, 10.44 m (34.3 ft)	tie

Run	Engine	Best legal model	Margin
H05	49, 12.40 m (40.7 ft)	Opus 49, 12.40 m (40.7 ft)	tie
H06	83, 10.65 m (35.0 ft)	Grok 83, 11.41 m (37.4 ft)	clear (6.6%)
H07	45, 9.60 m (31.5 ft)	GPT 45, 9.60 m (31.5 ft)	tie
H08	52, 5.92 m (19.4 ft)	Kimi 52, 6.12 m (20.1 ft)	clear (3.3%)
H09 LDM	29/30, 13.36 m (43.8 ft)	GPT 29, 13.56 m (44.5 ft)	narrow (1.5%)
H09 pallets	29/30, 29 positions	GPT 29, 29 positions	narrow (LDM tiebreak)
H10 LDM	26, 12.20 m (40.0 ft)	DeepSeek, same	tie
H10 pallets	26, 31 positions	DeepSeek, same	tie
H11 LDM	23, 12.00 m (39.4 ft)	Opus 23, 12.20 m (40.0 ft)	narrow (1.6%)
H11 pallets	23, 30 positions	Kimi 23, 32 positions	clear (2 positions)
H12 LDM	13/14, 11.00 m (36.1 ft)	Gemini 13, 11.00 m (36.1 ft)	tie
H12 pallets	13/14, 25 positions	Kimi 13, 28 positions	clear (3 positions)

Large set.

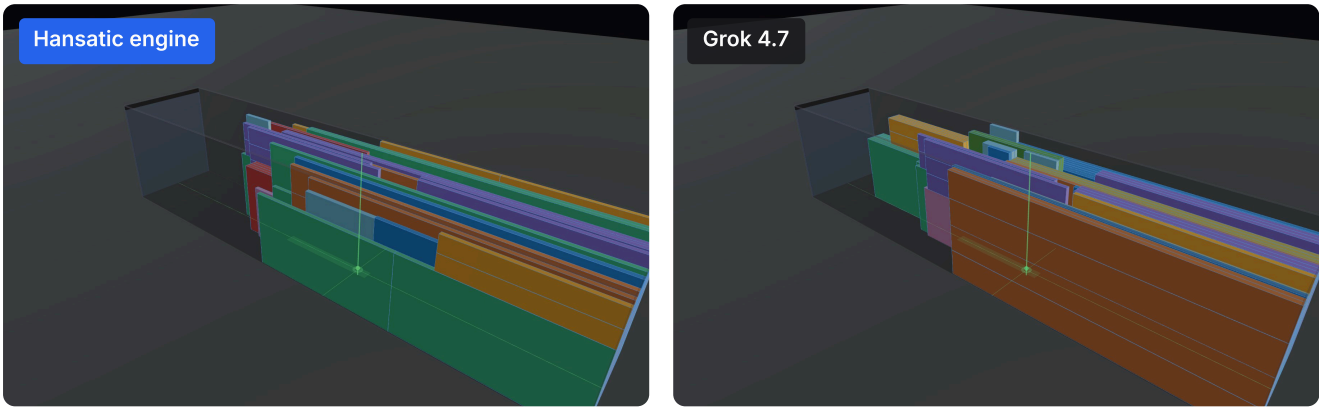
Run	Engine	Best legal model	Margin
BV1 (40' HC, 103 bars)	103, 9.17 m (30.1 ft)	Grok 103, 10.66 m (35.0 ft)	clear (14.0%)
BV2 LDM (curtainsider, 153)	153, 11.52 m (37.8 ft)	Opus 153, 11.78 m (38.6 ft)	clear (2.2%)
BV2 pallets	153, 29 positions	Opus 153, 29 positions	narrow (LDM tiebreak)
BV3 LDM (53 ft van, 119)	119, 6.83 m (22.4 ft)	Opus 119, 6.99 m (22.9 ft)	clear (2.2%)
BV3 pallets	119, 17 positions	Opus 119, 18 positions	clear (1 position)
BV4 (145, top-only fragile)	145, 8.57 m (28.1 ft)	Grok 145, 8.85 m (29.0 ft)	clear (3.1%)
BV5 (136, upright only)	127, 12.00 m (39.4 ft)	Opus 124, 11.77 m (38.6 ft)	clear (+3 placed)
BV6 (253, 95% of volume)	244, 11.98 m (39.3 ft)	GPT 238, 11.97 m (39.3 ft)	clear (+6 placed)

Special set.

Run	Engine	Models
U01 (AKE)	24/50, the optimum	Opus, Grok, Kimi, Qwen and Gemini also reach 24; GPT and DeepSeek reach 18
P01 (pallet)	40/40, 100% interlock, the optimum	every legal model answer reaches 40/40 at the optimum

2.3 What the margins mean on a real load

- **BV1:** 1.49 m (4.9 ft) shorter, about 12% of a 40' container's length.
- **H12 pallets:** 3 fewer pallet positions, about 9% of a 13.6 m (44.6 ft) curtainsider deck.
- **BV6:** 6 more pieces placed on a load that does not fit in full, so 6 fewer pieces left for another vehicle.



BV1: Hansatic engine 9.17 m (30.1 ft), left; best legal model answer (Grok 4.7) 10.66 m (35.0 ft), right.

2.4 Cost and time

	Engine	Models
Time per plan	negligible	not compared: batch APIs queue requests
Cost per plan	negligible	$0.10-0.80

3. Failure analysis

Across 208 model runs (8 models x 26), 95 plans are legal. The other 113 fail in three different ways.

Model	Legal	No answer (output limit)	Provider error	Plan breaks rules
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
All	**95**	**65**	**4**	**44**

3.1 No answer: the model runs out of room to think

In 65 runs, the model reasoned until it hit the output limit and never wrote a plan. One more answer (Magistral, H03) ended normally but was not valid JSON; it is counted here.

- This is the most common failure, and it hits six of the eight models.
- Kimi, DeepSeek and Qwen lose most of their runs this way. Their legal plans are mostly on the smaller cases.
- Opus loses 6 runs this way, among them BV1, BV4 and BV6, three of the largest manifests.

For a planner in production, no answer is a failure like an illegal one: the dock still needs a plan.

3.2 Provider errors

Grok returned HTTP 403 on 4 runs, because the provider's spending limit was reached. These runs were not resent and are scored as failures.

3.3 Plans that break rules

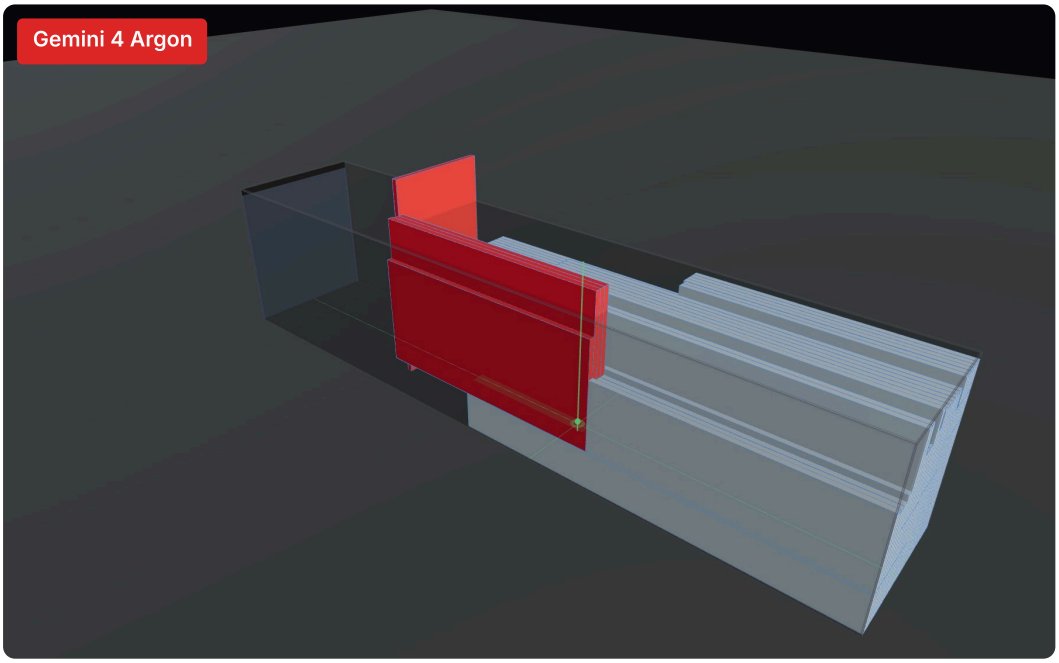
44 answers are complete plans that break at least one hard rule. Almost all come from two models: Gemini (19) and Magistral (23). Grok has 2.

Number of these 44 plans with at least one break of each kind:

Rule broken	Plans
Item outside the cargo space	27
Overlapping items	27
Floating items (under 95% support)	21
Item dimensions that match no allowed orientation	17
Load on non-stackable or fragile items, or over a stack limit	10
Stack above the roof	9
Centre of gravity outside the band	7
Axle overload	4
No-tilt item tipped	1
Outside the AKE contour	1

Grok's two illegal plans each break one rule: a crushed item on one run, the centre of gravity on another.

These plans look plausible in a 3D view. The faults are often a few items, floating or overlapping by centimetres. Only a full geometric check finds them.



BV1, Gemini 4 Argon: 77 of 103 items placed. Highlighted: items that float, overlap or leave the cargo space.

3.4 Legal, but behind

- **Incomplete.** On BV5 and BV6, the best legal model answers place 3 and 6 fewer items than the engine.

- **Longer.** On the runs where every item fits, the best legal model answer is longer than the engine's on 13 of the 24 truck and container runs. The biggest gaps:
  - 1.49 m (4.9 ft) on BV1;
  - 0.76 m (2.5 ft) on H06.

3.5 What models do well

The strongest answers use the strategies an experienced crew uses:

- **Heavy crates against the front wall.**
- **Using the height above non-stackable items.** On BV3, Opus lays cartons on their side as a floor 559 mm (22 in) high and stands 19 non-stackable appliances on it.

The engine did not have either strategy. Both are now deterministic engine strategies. This is the creativity the neuro-symbolic loop is designed to capture (section 5).

4. Tool arm

In a second arm, a model may call the engine through one tool, `pack_cargo`, and then answer with any plan.

Most models copy the manifest into the tool call and return the engine's plan unchanged. The failures are copying errors:

- Magistral sent 100 items with wrong dimensions or flags;
- Kimi made a wrong call on H10;
- Qwen failed on 4 runs.

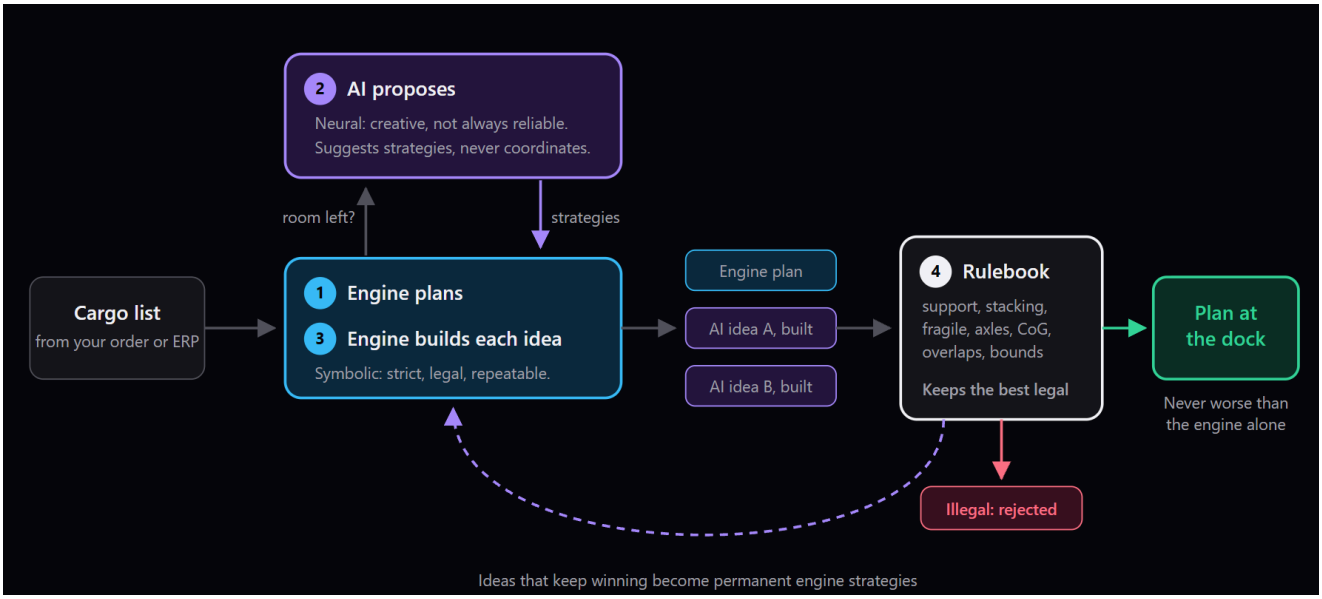
Implication: in any combined system, the engine must read the manifest straight from the source system, never through the model.

5. The neuro-symbolic architecture

5.1 Two kinds of intelligence

	Packing engine (symbolic)	Generative AI (neural)
Rules	follows every encoded rule, every time	breaks rules it was given
Repeatability	same load, same plan	the same load can get a different plan
Audit	every placement has a reason	hard to audit
Range	limited to the strategies it has	finds new strategies
Scale	handles 253 pieces	many models give no answer on large manifests

5.2 The loop



1. **Baseline.** The engine plans from the manifest, read directly from the source system.
2. **Trigger.** A model is called only when the baseline leaves room: unplaced items, low fill, or a clearance target missed.
3. **Proposal.** The model returns strategy parameters, never coordinates:
   - zones;
   - orderings;
   - orientation preferences;
   - floor and cap structures.
4. **Execution.** The engine builds one plan per proposal.
5. **Judgement.** The validator keeps the best legal plan. The baseline is always a candidate.

6. **Distillation.** A proposal that wins repeatedly becomes a permanent engine strategy, as the front-wall and carton-floor strategies did.

5.3 Properties by construction

- **Never worse than the engine alone,** because the baseline is always a candidate.
- **No illegal plan reaches the dock,** because the validator judges every candidate.
- **No transcription risk,** because the model never touches the manifest or the coordinates (section 4).
- **No failure from the model's output limit,** because a proposal is a few parameters, not a 253-item plan (section 3.1).
- **Capped AI cost,** because the model runs only on the trigger.

We are building this loop now. It will be scored as its own entrant in the next edition, with cost and time per plan.

6. Scope and limits

6.1 Validator scope

Legal means legal under the rules encoded in v2.3.0. The validator does not check build sequence beyond multi-stop last in, first out.

6.2 Variance

Every entrant runs once per case.

- **The engine** races strategies inside a time budget, so machine load can change which strategy finishes first. On H02 we measured 5.63-5.79 m (18.5-19.0 ft) across runs.
- **The models** are sampled once.

Single-run differences under 2% are reported as narrow.

6.3 Model tier and output limit

- We ran each lab's mainstream frontier model. Premium tiers (Anthropic Fable, OpenAI GPT-6 Astra) were not run.
- A larger output limit could turn some no-answer runs (section 3.1) into plans.

6.4 Case source

All cases come from one platform's users. Other shippers' cargo mix may differ.

7. Next steps

7.1 Practitioner rules

Two practitioner rules come next, in the engine and in validator v3:

- **Roof clearance:** at least 50 mm (2.0 in) under the roof, or 100 mm (3.9 in) above a stacked item over 25 kg (55 lb).
- **Side play** for each forklift row.

We re-scored every entrant with the roof rule. Floor items that cannot tilt and are taller than the roof minus 50 mm (2.0 in) are exempt, because the case forces them.

- **The engine** keeps the rule on 9 of 16 standard runs and 1 of 8 large runs.
- **On 6 runs, a model answer keeps the rule** with a plan close to the engine's. These are the engine's first targets:
  - H02 (GPT);

    - H06 (Grok, 11.41 m / 37.4 ft);
    - H12 LDM (Gemini, 11.00 m / 36.1 ft);
    - H12 pallets (Kimi, 28 positions);
    - BV3 LDM (GPT, 9.14 m / 30.0 ft);
    - BV5 (Opus, 124 placed).
- **On 8 runs, no entrant keeps the rule:** H01, H03, H08, BV1, both BV2 runs, BV3 pallets, BV6.

7.2 Next edition

- Validator v3, with roof clearance, side play and a build-sequence check, applied to every entrant.
- New real customer cases.
- Each lab's most capable models, premium tiers included, with several runs per model and variance reported.
- The neuro-symbolic loop as an entrant.

8. Release

Published:

- every case in full;
- the validator;
- the prompts;
- the scoring and ranking code;
- for every run: the request sent, the provider's untouched response and the parsed plan.

Viewer: every plan from every entrant, in 3D next to its prompt and raw answer: hansatic.com/en/neuro-symbolic-load-planning.

Closed:

- the pallet-position counter (Hansatic's production counter, applied the same way to every entrant);
- the engine;
- the ids of the source plans.

Appendix A. Prompt templates

One system prompt per vehicle kind, as sent. The user message gives the objective, then the case as JSON; the item list is shortened here. Prompts are shown in English, the language the models received.

Sea container (BV1, 40' High Cube)

System prompt

```
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"]}
```

User message (item list shortened)

```
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"
 ]
}
```


Truck (H09, 13.6 m (44.6 ft) curtainsider)

System prompt

```
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"]}
```

User message (item list shortened)

```
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"
```

```
 ]
}
```

Air container (U01, AKE with a sloped wall)

System prompt

```
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"]}
```

User message (item list shortened)

```
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))

System prompt

```
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"]}
```

User message (item list shortened)

```
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"
```

```
  ]
}
```

Appendix B. All 26 runs

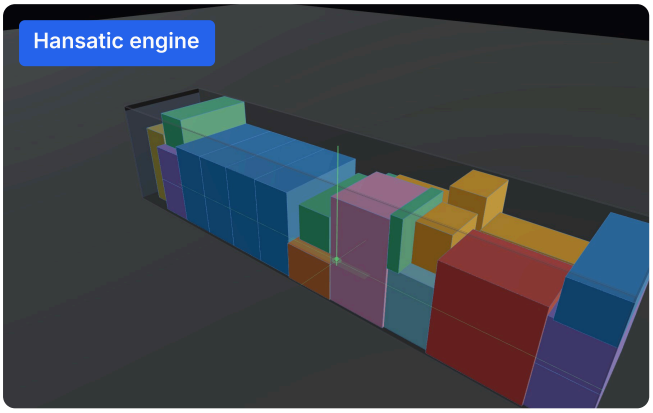
One page per run: the load, the engine's plan next to the best legal model answer, and every entrant's result.

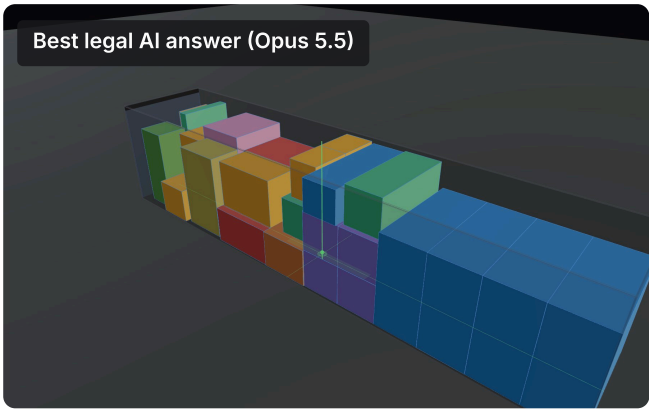
H01 Shortest load

Vehicle	Pieces	Objective	Result
40' High Cube	27	Shortest load	Engine narrowly ahead

What makes it hard

27 items that must stay upright





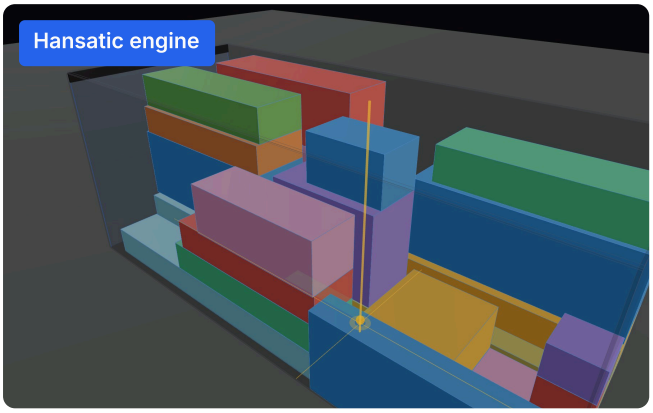
Entrant	Status	Placed	Score	Notes
Hansatic engine	**Legal**	**26 / 27**	**11.69 m (38.4 ft)**	
GPT-5.5	Legal	25 / 27	11.88 m (39.0 ft)	
Opus 5.5	Legal	26 / 27	11.74 m (38.5 ft)	
Grok 4.7	Legal	25 / 27	11.37 m (37.3 ft)	
Kimi K3.1 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
DeepSeek V4 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
Gemini 4 Argon	Illegal	22 / 27	11.56 m (37.9 ft)	1 wrong item sizes, 1 outside the vehicle, 1 floating items
Qwen3p8	No plan returned	-	-	The model hit its output limit before writing a plan.
Magistral Medium	Illegal	27 / 27	12.09 m (39.7 ft)	7 overlapping items, 14 outside the vehicle, 3 over the height limit, 1 floating items, 1 centre of gravity off

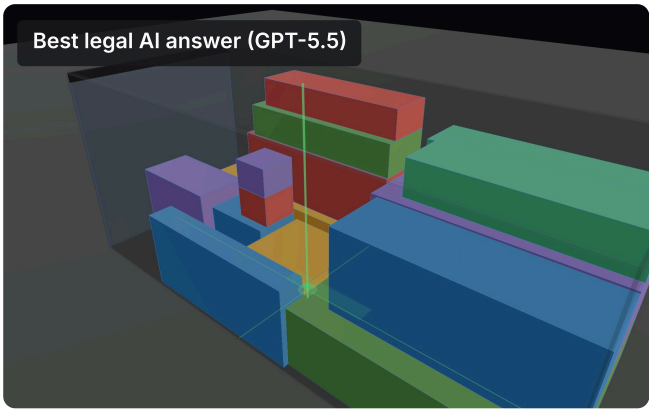
H02 Shortest load

Vehicle	Pieces	Objective	Result
20' Dry Container	29	Shortest load	Engine clearly ahead

What makes it hard

29 items that must stay upright; items up to 5.00 m (16.4 ft) long; heaviest item 1,550 kg (3,417 lb)





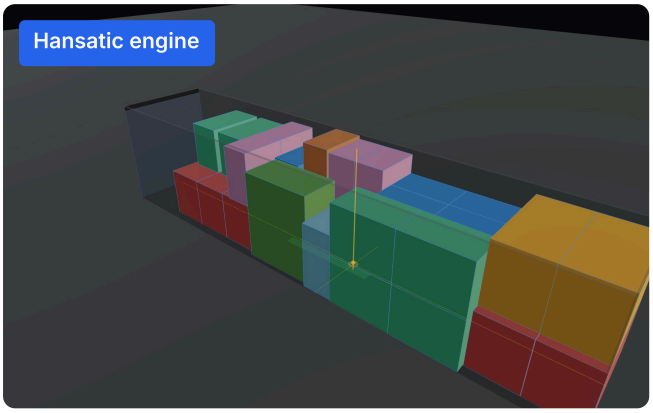
Entrant	Status	Placed	Score	Notes
Hansatic engine	**Legal**	**24 / 29**	**5.78 m (19.0 ft)**	
GPT-5.5	Legal	18 / 29	5.76 m (18.9 ft)	
Opus 5.5	No plan returned	-	-	The model hit its output limit before writing a plan.
Grok 4.7	Provider error	-	-	The provider refused the request.
Kimi K3.1 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
DeepSeek V4 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
Gemini 4 Argon	Illegal	23 / 29	5.00 m (16.4 ft)	6 wrong item sizes, 13 overlapping items, 9 floating items, 9 crushed items
Qwen3p8	No plan returned	-	-	The model hit its output limit before writing a plan.
Magistral Medium	Illegal	29 / 29	7.97 m (26.1 ft)	37 overlapping items, 23 outside the vehicle, 1 over the height limit, 7 floating items, 5 crushed items, 2 centre of gravity off

H03 Shortest load

Vehicle	Pieces	Objective	Result
40' High Cube	32	Shortest load	Engine clearly ahead

What makes it hard

25 items that must stay upright; heaviest item 2,680 kg (5,908 lb)



No AI model returned a legal plan on this run.

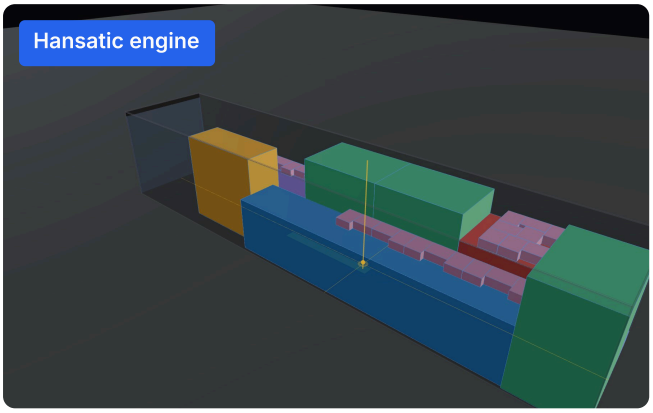
Entrant	Status	Placed	Score	Notes
Hansatic engine	**Legal**	**31 / 32**	**10.51 m (34.5 ft)**	
GPT-5.5	No plan returned	-	-	The model hit its output limit before writing a plan.
Opus 5.5	No plan returned	-	-	The model hit its output limit before writing a plan.
Grok 4.7	**Illegal**	31 / 32	12.03 m (39.5 ft)	1 crushed items
Kimi K3.1 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
DeepSeek V4 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
Gemini 4 Argon	**Illegal**	31 / 32	11.81 m (38.7 ft)	1 wrong item sizes, 12 overlapping items, 6 outside the vehicle, 3 over the height limit, 2 floating items, 3 crushed items
Qwen3p8	No plan returned	-	-	The model hit its output limit before writing a plan.
Magistral Medium	No plan returned	-	-	The answer did not contain a valid plan.

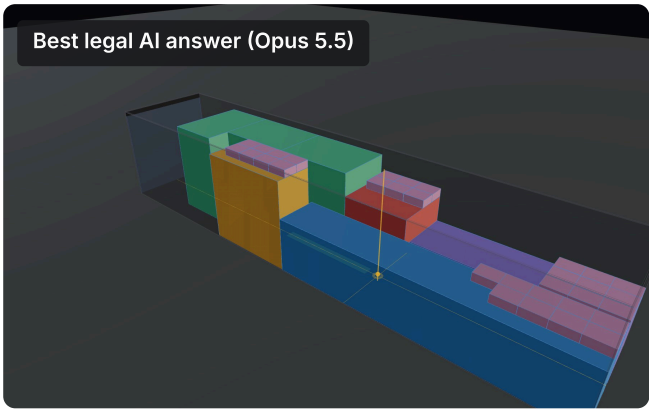
H04 Shortest load

Vehicle	Pieces	Objective	Result
40' Dry Container	39	Shortest load	Tie

What makes it hard

items up to 6.30 m (20.7 ft) long; heaviest item 998 kg (2,200 lb)





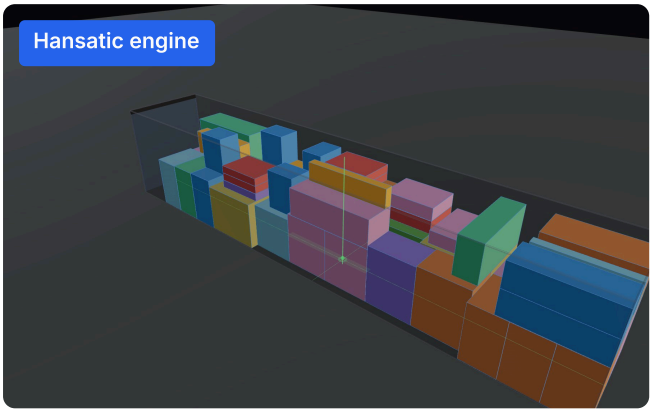
Entrant	Status	Placed	Score	Notes
Hansatic engine	**Legal**	**39 / 39**	**10.44 m (34.3 ft)**	
GPT-5.5	Legal	39 / 39	10.72 m (35.2 ft)	
Opus 5.5	Legal	39 / 39	10.44 m (34.3 ft)	
Grok 4.7	Legal	39 / 39	11.10 m (36.4 ft)	
Kimi K3.1 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
DeepSeek V4 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
Gemini 4 Argon	Illegal	29 / 39	6.30 m (20.7 ft)	10 wrong item sizes, 6 floating items
Qwen3p8	No plan returned	-	-	The model hit its output limit before writing a plan.
Magistral Medium	Illegal	38 / 39	10.20 m (33.5 ft)	1 wrong item sizes, 27 overlapping items, 6 outside the vehicle, 24 floating items, 1 crushed items, 1 centre of gravity off

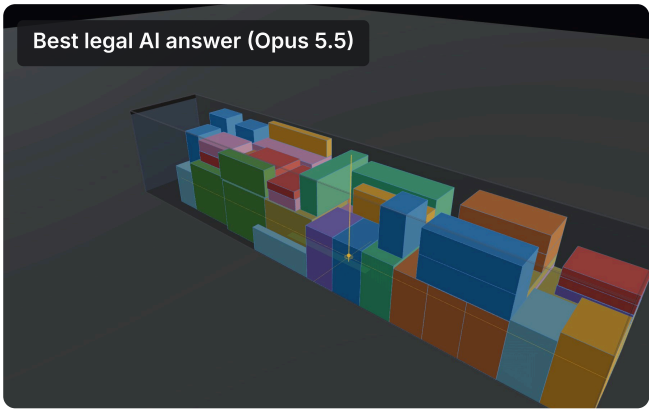
H05 Shortest load

Vehicle	Pieces	Objective	Result
45' High Cube	49	Shortest load	Tie

What makes it hard

49 items that must stay upright; 27 items with a stack-weight limit; heaviest item 814 kg (1,795 lb)





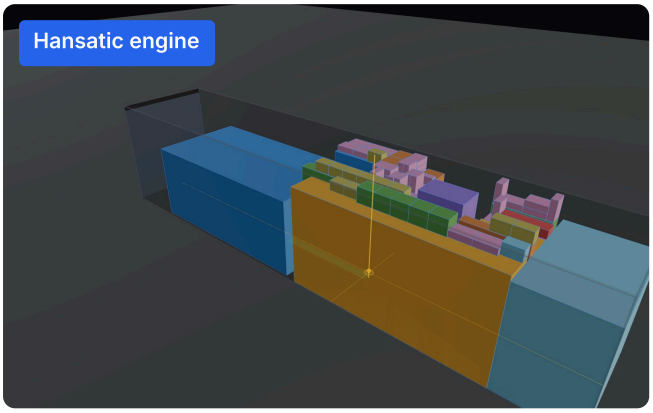
Entrant	Status	Placed	Score	Notes
Hansatic engine	**Legal**	**49 / 49**	**12.40 m (40.7 ft)**	
GPT-5.5	Legal	45 / 49	13.28 m (43.6 ft)	
Opus 5.5	Legal	49 / 49	12.40 m (40.7 ft)	
Grok 4.7	Legal	43 / 49	12.96 m (42.5 ft)	
Kimi K3.1 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
DeepSeek V4 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
Gemini 4 Argon	Illegal	37 / 49	10.49 m (34.4 ft)	11 wrong item sizes, 2 overlapping items, 5 outside the vehicle, 2 over the height limit, 3 floating items, 2 crushed items
Qwen3p8	No plan returned	-	-	The model hit its output limit before writing a plan.
Magistral Medium	Illegal	47 / 49	12.73 m (41.8 ft)	2 wrong item sizes, 31 overlapping items, 9 outside the vehicle, 14 over the height limit, 21 floating items, 2 crushed items

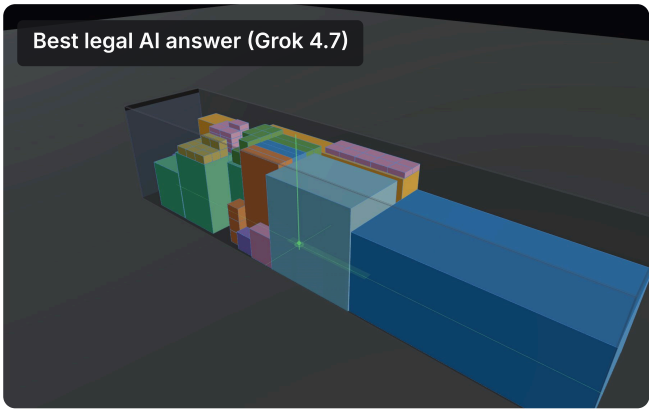
H06 Shortest load

Vehicle	Pieces	Objective	Result
40' High Cube	83	Shortest load	Engine clearly ahead

What makes it hard

5 non-stackable items; 1 items that must stay upright; heaviest item 3,020 kg (6,658 lb)





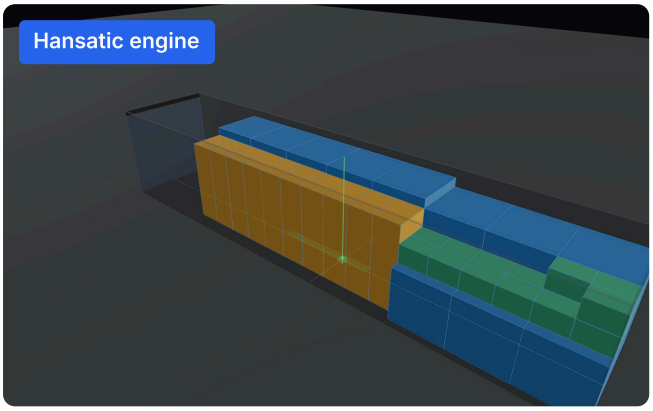
Entrant	Status	Placed	Score	Notes
Hansatic engine	**Legal**	**83 / 83**	**10.65 m (35.0 ft)**	
GPT-5.5	Legal	83 / 83	11.78 m (38.7 ft)	
Opus 5.5	No plan returned	-	-	The model hit its output limit before writing a plan.
Grok 4.7	Legal	83 / 83	11.41 m (37.4 ft)	
Kimi K3.1 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
DeepSeek V4 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
Gemini 4 Argon	Illegal	7 / 83	6.89 m (22.6 ft)	2 wrong item sizes, 1 centre of gravity off
Qwen3p8	No plan returned	-	-	The model hit its output limit before writing a plan.
Magistral Medium	Illegal	20 / 83	12.25 m (40.2 ft)	2 overlapping items, 4 outside the vehicle, 1 over the height limit, 14 floating items, 1 centre of gravity off

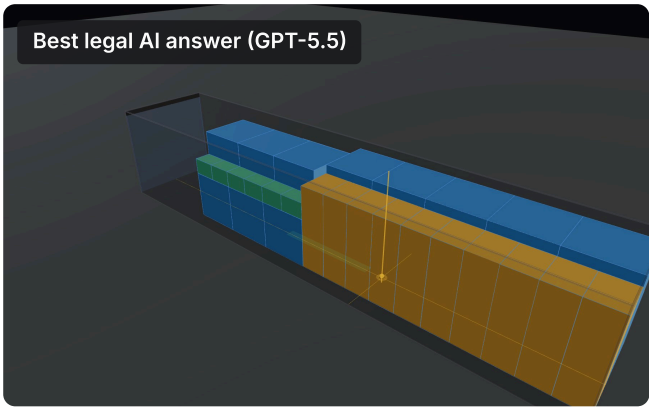
H07 Shortest load

Vehicle	Pieces	Objective	Result
40' Dry Container	45	Shortest load	Tie

What makes it hard

10 non-stackable items; 15 items that must stay upright; 15 items with a stack-weight limit; heaviest item 1,200 kg (2,646 lb)





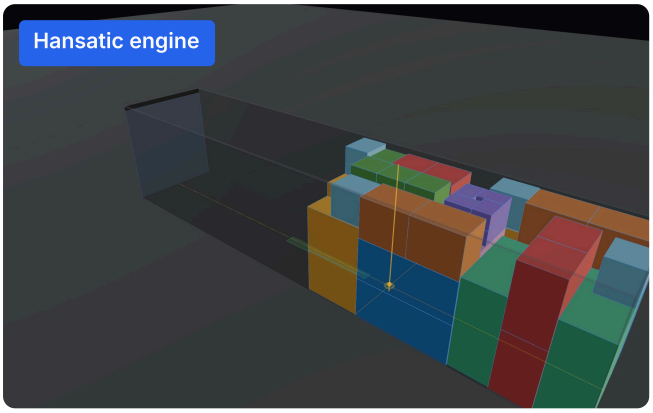
Entrant	Status	Placed	Score	Notes
Hansatic engine	**Legal**	**45 / 45**	**9.60 m (31.5 ft)**	
GPT-5.5	Legal	45 / 45	9.60 m (31.5 ft)	
Opus 5.5	Legal	45 / 45	9.61 m (31.5 ft)	
Grok 4.7	Illegal	45 / 45	10.80 m (35.4 ft)	1 centre of gravity off
Kimi K3.1 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
DeepSeek V4 Pro	Legal	45 / 45	12.00 m (39.4 ft)	
Gemini 4 Argon	Illegal	32 / 45	7.60 m (24.9 ft)	13 wrong item sizes, 1 overlapping items, 16 floating items
Qwen3p8	No plan returned	-	-	The model hit its output limit before writing a plan.
Magistral Medium	Illegal	10 / 45	12.00 m (39.4 ft)	1 centre of gravity off

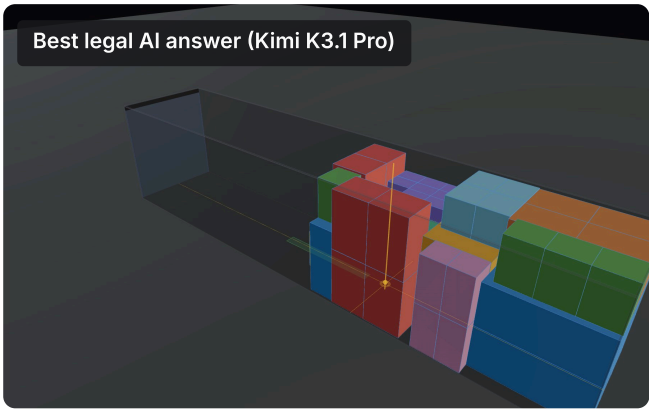
H08 Shortest load

Vehicle	Pieces	Objective	Result
40' High Cube	52	Shortest load	Engine clearly ahead

What makes it hard

52 items that must stay upright





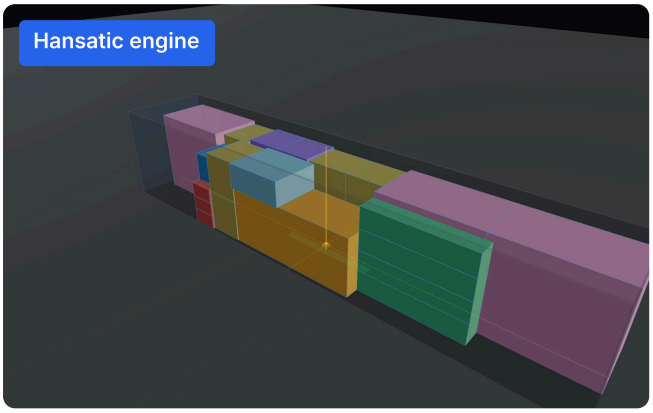
Entrant	Status	Placed	Score	Notes
Hansatic engine	**Legal**	**52 / 52**	**5.92 m (19.4 ft)**	
GPT-5.5	Legal	52 / 52	6.26 m (20.5 ft)	
Opus 5.5	Legal	52 / 52	6.20 m (20.3 ft)	
Grok 4.7	Legal	52 / 52	6.18 m (20.3 ft)	
Kimi K3.1 Pro	Legal	52 / 52	6.12 m (20.1 ft)	
DeepSeek V4 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
Gemini 4 Argon	Illegal	51 / 52	6.51 m (21.4 ft)	1 wrong item sizes, 4 overlapping items, 1 floating items
Qwen3p8	No plan returned	-	-	The model hit its output limit before writing a plan.
Magistral Medium	Illegal	52 / 52	11.25 m (36.9 ft)	14 overlapping items, 20 outside the vehicle

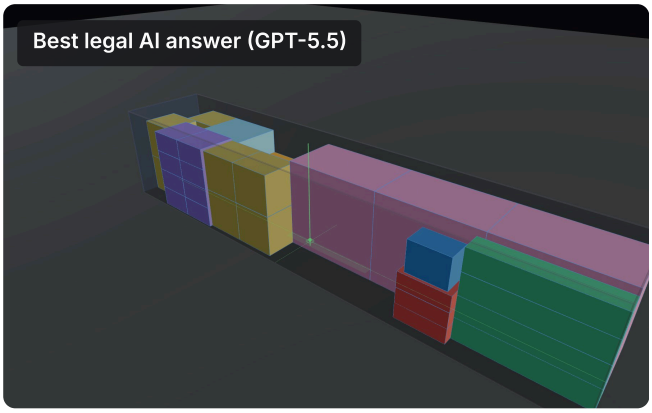
H09 Shortest load

Vehicle	Pieces	Objective	Result
13.6m Curtainsider	30	Shortest load	Engine narrowly ahead

What makes it hard

30 items that must stay upright; heaviest item 2,400 kg (5,291 lb)





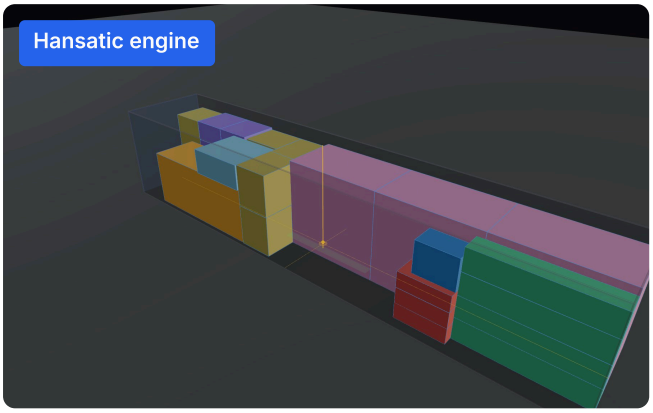
Entrant	Status	Placed	Score	Notes
Hansatic engine	**Legal**	**29 / 30**	**13.36 m (43.8 ft)**	
GPT-5.5	Legal	29 / 30	13.56 m (44.5 ft)	
Opus 5.5	Legal	29 / 30	13.60 m (44.6 ft)	
Grok 4.7	Legal	29 / 30	13.60 m (44.6 ft)	
Kimi K3.1 Pro	Legal	29 / 30	13.60 m (44.6 ft)	
DeepSeek V4 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
Gemini 4 Argon	Illegal	19 / 30	11.60 m (38.1 ft)	11 wrong item sizes, 7 overlapping items
Qwen3p8	No plan returned	-	-	The model hit its output limit before writing a plan.
Magistral Medium	Illegal	30 / 30	13.36 m (43.8 ft)	17 overlapping items, 25 outside the vehicle, 7 floating items, 1 axle overloaded

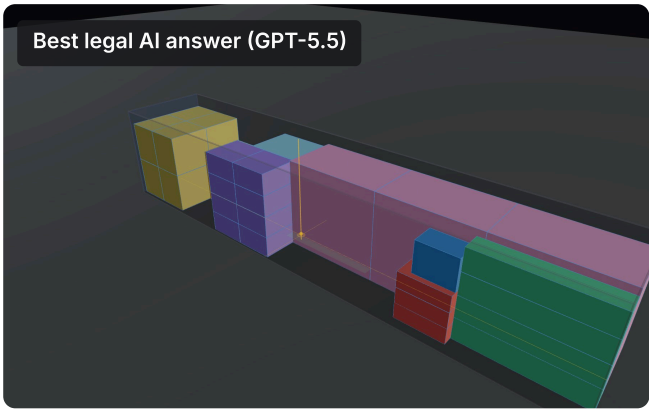
H09 Fewest pallet positions

Vehicle	Pieces	Objective	Result
13.6m Curtainsider	30	Fewest pallet positions	Engine narrowly ahead

What makes it hard

30 items that must stay upright; heaviest item 2,400 kg (5,291 lb)





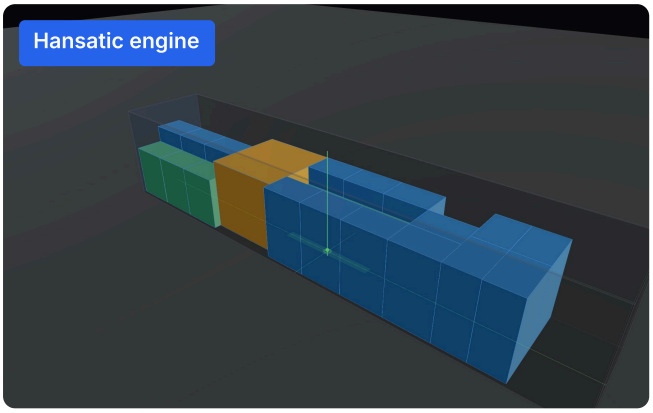
Entrant	Status	Placed	Score	Notes
Hansatic engine	**Legal**	**29 / 30**	**29 positions, 13.36 m (43.8 ft)**	
GPT-5.5	Legal	29 / 30	29 positions, 13.60 m (44.6 ft)	
Opus 5.5	Legal	29 / 30	31 positions, 13.60 m (44.6 ft)	
Grok 4.7	Legal	29 / 30	29 positions, 13.62 m (44.7 ft)	
Kimi K3.1 Pro	Legal	29 / 30	29 positions, 13.60 m (44.6 ft)	
DeepSeek V4 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
Gemini 4 Argon	Illegal	30 / 30	29 positions, 11.72 m (38.5 ft)	18 overlapping items, 1 outside the vehicle, 3 crushed items, 1 axle overloaded
Qwen3p8	No plan returned	-	-	The model hit its output limit before writing a plan.
Magistral Medium	Illegal	30 / 30	34 positions, 13.36 m (43.8 ft)	4 overlapping items, 25 outside the vehicle, 1 axle overloaded, 1 misaligned rows

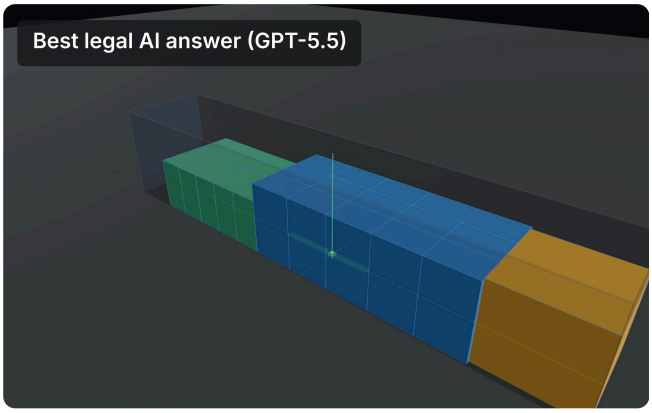
H10 Shortest load

Vehicle	Pieces	Objective	Result
13.6m Curtainsider	26	Shortest load	Tie

What makes it hard

26 non-stackable items; 16 items that must stay upright; heaviest item 666 kg (1,468 lb)





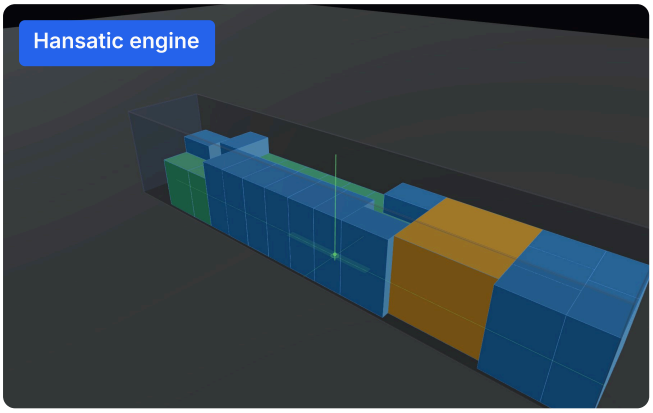
Entrant	Status	Placed	Score	Notes
Hansatic engine	**Legal**	**26 / 26**	**12.20 m (40.0 ft)**	
GPT-5.5	Legal	26 / 26	12.20 m (40.0 ft)	
Opus 5.5	Legal	26 / 26	12.20 m (40.0 ft)	
Grok 4.7	Legal	26 / 26	12.20 m (40.0 ft)	
Kimi K3.1 Pro	Legal	26 / 26	12.20 m (40.0 ft)	
DeepSeek V4 Pro	Legal	26 / 26	12.20 m (40.0 ft)	
Gemini 4 Argon	Legal	26 / 26	12.60 m (41.3 ft)	
Qwen3p8	Legal	26 / 26	12.20 m (40.0 ft)	
Magistral Medium	Legal	26 / 26	13.00 m (42.7 ft)	

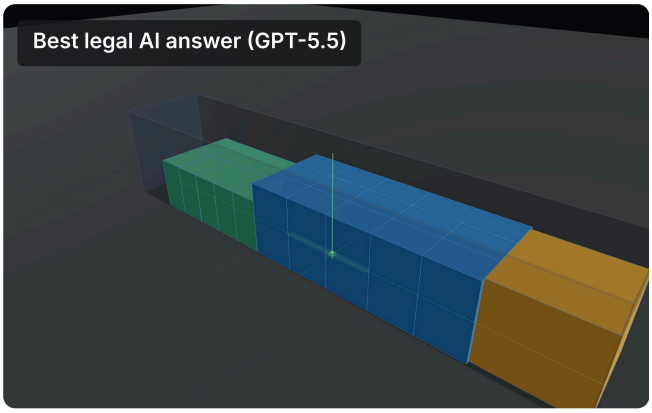
H10 Fewest pallet positions

Vehicle	Pieces	Objective	Result
13.6m Curtainsider	26	Fewest pallet positions	Tie

What makes it hard

26 non-stackable items; 16 items that must stay upright; heaviest item 666 kg (1,468 lb)





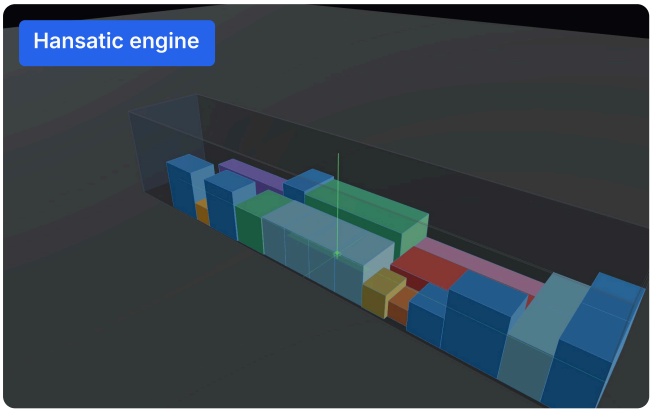
Entrant	Status	Placed	Score	Notes
Hansatic engine	**Legal**	**26 / 26**	**31 positions, 12.20 m (40.0 ft)**	
GPT-5.5	Legal	26 / 26	31 positions, 12.20 m (40.0 ft)	
Opus 5.5	Legal	26 / 26	31 positions, 13.00 m (42.7 ft)	
Grok 4.7	Legal	26 / 26	31 positions, 12.20 m (40.0 ft)	
Kimi K3.1 Pro	Legal	26 / 26	31 positions, 12.20 m (40.0 ft)	
DeepSeek V4 Pro	Legal	26 / 26	31 positions, 12.20 m (40.0 ft)	
Gemini 4 Argon	Illegal	26 / 26	31 positions, 12.60 m (41.3 ft)	1 axle overloaded
Qwen3p8	Legal	26 / 26	31 positions, 13.00 m (42.7 ft)	
Magistral Medium	Illegal	26 / 26	28 positions, 13.00 m (42.7 ft)	10 outside the vehicle

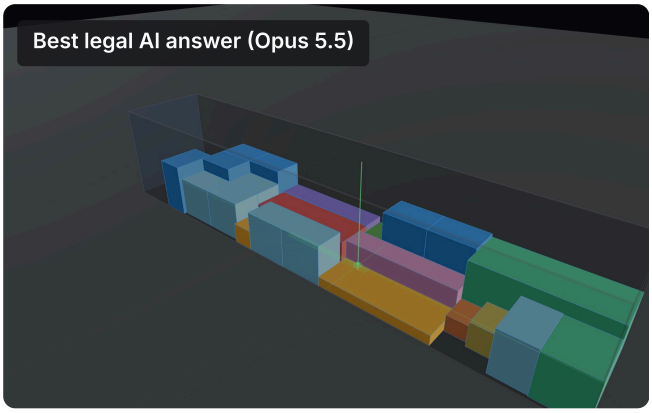
H11 Shortest load

Vehicle	Pieces	Objective	Result
13.6m Curtainsider	23	Shortest load	Engine narrowly ahead

What makes it hard

23 non-stackable items; 23 items that must stay upright; heaviest item 2,700 kg (5,952 lb)





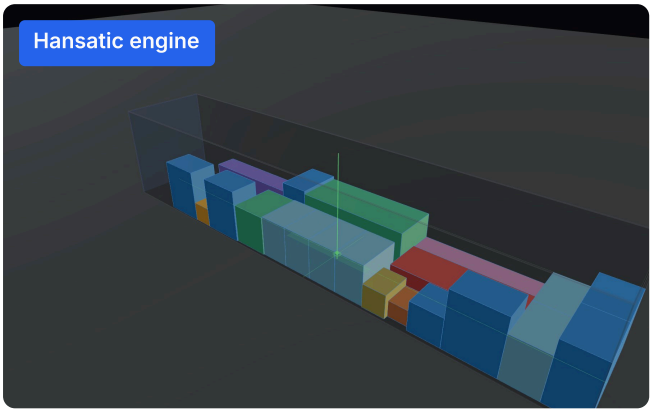
Entrant	Status	Placed	Score	Notes
Hansatic engine	**Legal**	**23 / 23**	**12.00 m (39.4 ft)**	
GPT-5.5	Legal	23 / 23	12.42 m (40.7 ft)	
Opus 5.5	Legal	23 / 23	12.20 m (40.0 ft)	
Grok 4.7	Legal	23 / 23	12.62 m (41.4 ft)	
Kimi K3.1 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
DeepSeek V4 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
Gemini 4 Argon	Legal	23 / 23	13.02 m (42.7 ft)	
Qwen3p8	No plan returned	-	-	The model hit its output limit before writing a plan.
Magistral Medium	Illegal	23 / 23	12.34 m (40.5 ft)	7 overlapping items, 8 outside the vehicle

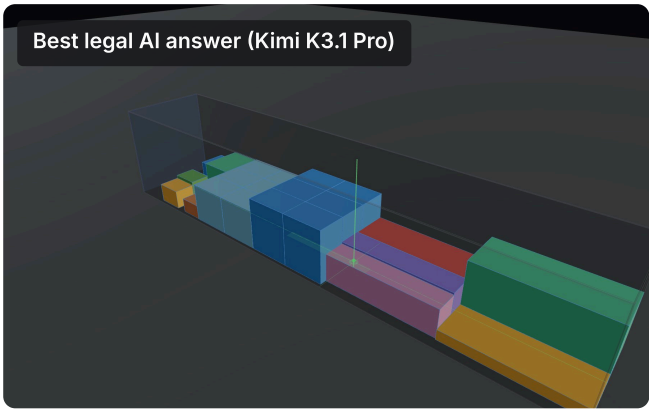
H11 Fewest pallet positions

Vehicle	Pieces	Objective	Result
13.6m Curtainsider	23	Fewest pallet positions	Engine clearly ahead

What makes it hard

23 non-stackable items; 23 items that must stay upright; heaviest item 2,700 kg (5,952 lb)





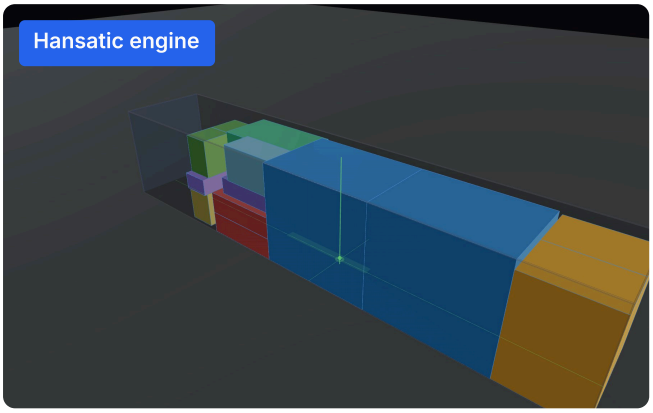
Entrant	Status	Placed	Score	Notes
Hansatic engine	**Legal**	**23 / 23**	**30 positions, 12.00 m (39.4 ft)**	
GPT-5.5	Legal	23 / 23	32 positions, 12.64 m (41.5 ft)	
Opus 5.5	Legal	23 / 23	32 positions, 12.84 m (42.1 ft)	
Grok 4.7	Provider error	-	-	The provider refused the request.
Kimi K3.1 Pro	Legal	23 / 23	32 positions, 12.62 m (41.4 ft)	
DeepSeek V4 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
Gemini 4 Argon	Illegal	22 / 23	30 positions, 12.44 m (40.8 ft)	1 wrong item sizes
Qwen3p8	No plan returned	-	-	The model hit its output limit before writing a plan.
Magistral Medium	Illegal	23 / 23	30 positions, 12.60 m (41.3 ft)	11 overlapping items, 9 outside the vehicle

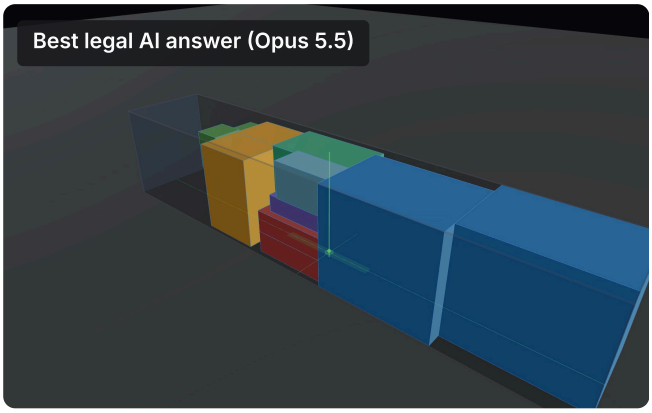
H12 Shortest load

Vehicle	Pieces	Objective	Result
13.6m Curtainsider	14	Shortest load	Tie

What makes it hard

6 non-stackable items; 9 items that must stay upright; heaviest item 732 kg (1,614 lb)





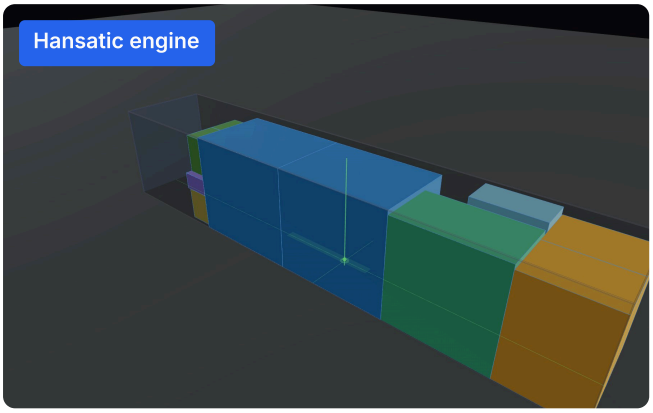
Entrant	Status	Placed	Score	Notes
Hansatic engine	**Legal**	**13 / 14**	**11.00 m (36.1 ft)**	
GPT-5.5	**Legal**	13 / 14	11.40 m (37.4 ft)	
Opus 5.5	**Legal**	13 / 14	11.00 m (36.1 ft)	
Grok 4.7	Provider error	-	-	The provider refused the request.
Kimi K3.1 Pro	**Legal**	13 / 14	11.00 m (36.1 ft)	
DeepSeek V4 Pro	**Legal**	13 / 14	11.70 m (38.4 ft)	
Gemini 4 Argon	**Legal**	13 / 14	11.00 m (36.1 ft)	
Qwen3p8	No plan returned	-	-	The model hit its output limit before writing a plan.
Magistral Medium	**Illegal**	14 / 14	13.50 m (44.3 ft)	5 overlapping items, 5 outside the vehicle, 2 floating items

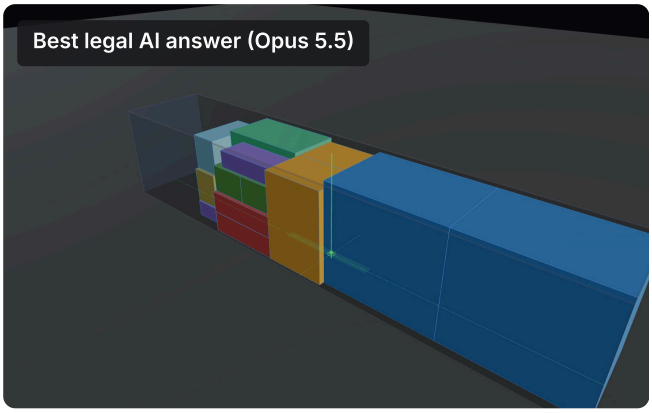
H12 Fewest pallet positions

Vehicle	Pieces	Objective	Result
13.6m Curtainsider	14	Fewest pallet positions	Engine clearly ahead

What makes it hard

6 non-stackable items; 9 items that must stay upright; heaviest item 732 kg (1,614 lb)





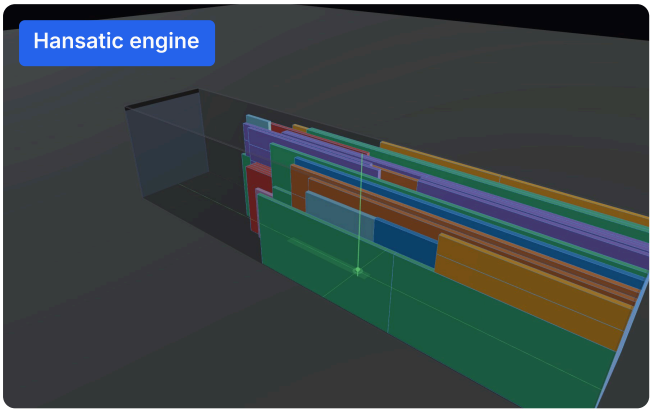
Entrant	Status	Placed	Score	Notes
Hansatic engine	**Legal**	**13 / 14**	**25 positions, 11.00 m (36.1 ft)**	
GPT-5.5	Legal	13 / 14	29 positions, 11.73 m (38.5 ft)	
Opus 5.5	Legal	13 / 14	28 positions, 11.00 m (36.1 ft)	
Grok 4.7	Provider error	-	-	The provider refused the request.
Kimi K3.1 Pro	Legal	13 / 14	28 positions, 11.00 m (36.1 ft)	
DeepSeek V4 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
Gemini 4 Argon	Legal	13 / 14	28 positions, 11.15 m (36.6 ft)	
Qwen3p8	No plan returned	-	-	The model hit its output limit before writing a plan.
Magistral Medium	Illegal	14 / 14	33 positions, 13.20 m (43.3 ft)	9 overlapping items, 9 outside the vehicle

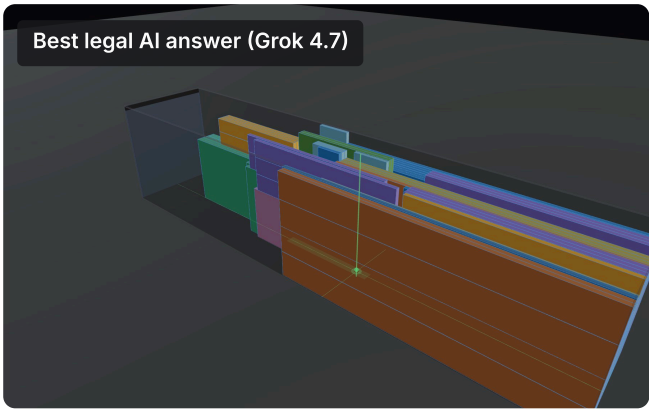
BV1 Shortest load

Vehicle	Pieces	Objective	Result
40' High Cube	103	Shortest load	Engine clearly ahead

What makes it hard

37 items that must stay upright; items up to 8.27 m (27.1 ft) long





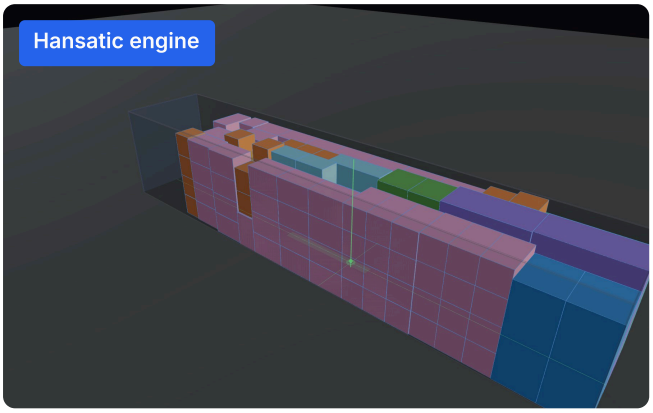
Entrant	Status	Placed	Score	Notes
Hansatic engine	**Legal**	**103 / 103**	**9.17 m (30.1 ft)**	
GPT-5.5	Legal	92 / 103	8.27 m (27.1 ft)	
Opus 5.5	No plan returned	-	-	The model hit its output limit before writing a plan.
Grok 4.7	Legal	103 / 103	10.66 m (35.0 ft)	
Kimi K3.1 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
DeepSeek V4 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
Gemini 4 Argon	Illegal	77 / 103	8.36 m (27.4 ft)	1 wrong item sizes, 4 over the height limit, 4 floating items
Qwen3p8	No plan returned	-	-	The model hit its output limit before writing a plan.
Magistral Medium	Illegal	22 / 103	16.54 m (54.3 ft)	9 outside the vehicle

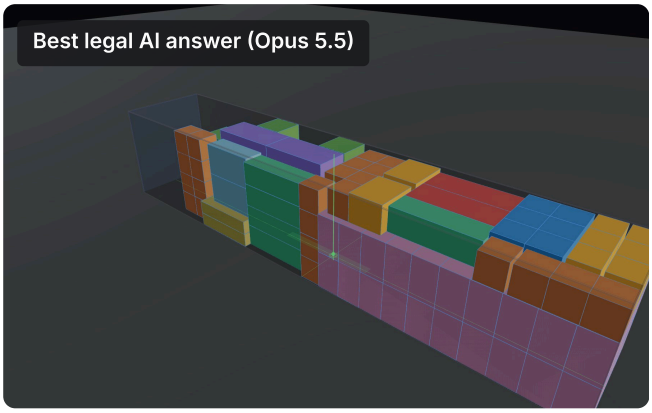
BV2 Shortest load

Vehicle	Pieces	Objective	Result
13.6m Curtainsider	153	Shortest load	Engine clearly ahead

What makes it hard

2 non-stackable items; heaviest item 578 kg (1,274 lb)





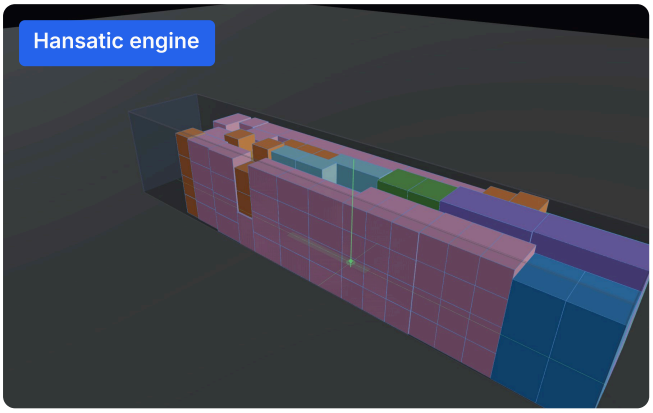
Entrant	Status	Placed	Score	Notes
Hansatic engine	**Legal**	**153 / 153**	**11.52 m (37.8 ft)**	
GPT-5.5	Legal	153 / 153	13.14 m (43.1 ft)	
Opus 5.5	Legal	153 / 153	11.78 m (38.6 ft)	
Grok 4.7	Legal	153 / 153	12.30 m (40.4 ft)	
Kimi K3.1 Pro	Legal	153 / 153	11.99 m (39.3 ft)	
DeepSeek V4 Pro	Legal	153 / 153	12.79 m (42.0 ft)	
Gemini 4 Argon	Illegal	147 / 153	12.77 m (41.9 ft)	6 wrong item sizes
Qwen3p8	No plan returned	-	-	The model hit its output limit before writing a plan.
Magistral Medium	Illegal	56 / 153	14.02 m (46.0 ft)	39 overlapping items, 27 outside the vehicle

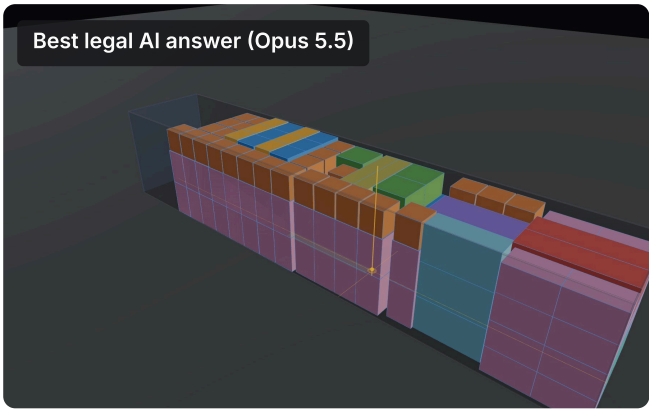
BV2 Fewest pallet positions

Vehicle	Pieces	Objective	Result
13.6m Curtainsider	153	Fewest pallet positions	Engine narrowly ahead

What makes it hard

2 non-stackable items; heaviest item 578 kg (1,274 lb)





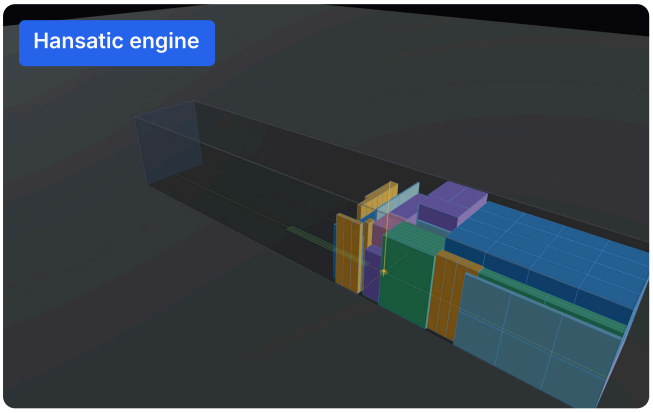
Entrant	Status	Placed	Score	Notes
Hansatic engine	**Legal**	**153 / 153**	**29 positions, 11.52 m (37.8 ft)**	
GPT-5.5	Legal	153 / 153	32 positions, 12.60 m (41.3 ft)	
Opus 5.5	Legal	153 / 153	29 positions, 11.67 m (38.3 ft)	
Grok 4.7	Legal	153 / 153	32 positions, 12.78 m (41.9 ft)	
Kimi K3.1 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
DeepSeek V4 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
Gemini 4 Argon	No plan returned	-	-	The model hit its output limit before writing a plan.
Qwen3p8	No plan returned	-	-	The model hit its output limit before writing a plan.
Magistral Medium	Illegal	153 / 153	34 positions, 14.82 m (48.6 ft)	7 overlapping items, 31 outside the vehicle, 86 over the height limit, 96 floating items

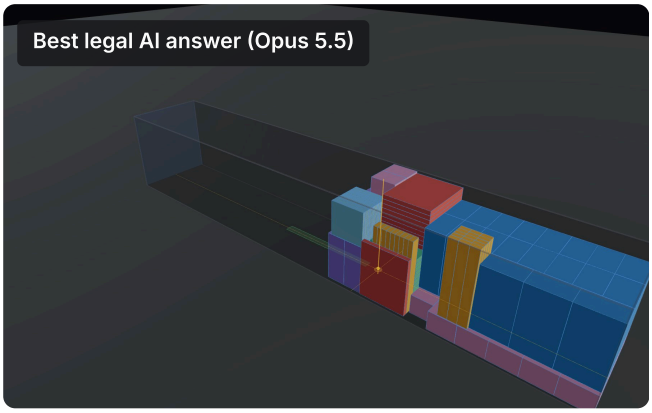
BV3 Shortest load

Vehicle	Pieces	Objective	Result
53ft Dry Van	119	Shortest load	Engine clearly ahead

What makes it hard

19 non-stackable items; 35 items that must stay upright





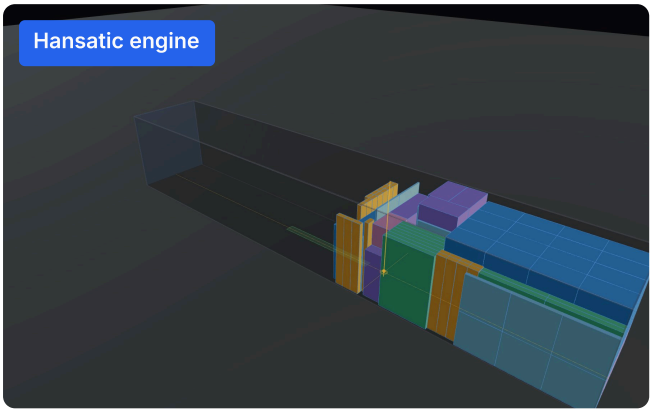
Entrant	Status	Placed	Score	Notes
Hansatic engine	**Legal**	**119 / 119**	**6.83 m (22.4 ft)**	
GPT-5.5	Legal	119 / 119	9.14 m (30.0 ft)	
Opus 5.5	Legal	119 / 119	6.99 m (22.9 ft)	
Grok 4.7	Legal	119 / 119	8.10 m (26.6 ft)	
Kimi K3.1 Pro	No plan returned	-	-	The answer did not contain a valid plan.
DeepSeek V4 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
Gemini 4 Argon	Illegal	115 / 119	8.36 m (27.4 ft)	4 wrong item sizes, 5 overlapping items, 3 crushed items
Qwen3p8	No plan returned	-	-	The model hit its output limit before writing a plan.
Magistral Medium	Illegal	22 / 119	16.76 m (55.0 ft)	11 outside the vehicle

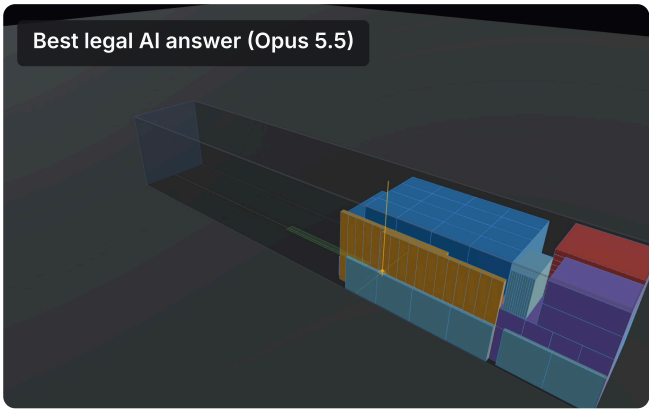
BV3 Fewest pallet positions

Vehicle	Pieces	Objective	Result
53ft Dry Van	119	Fewest pallet positions	Engine clearly ahead

What makes it hard

19 non-stackable items; 35 items that must stay upright





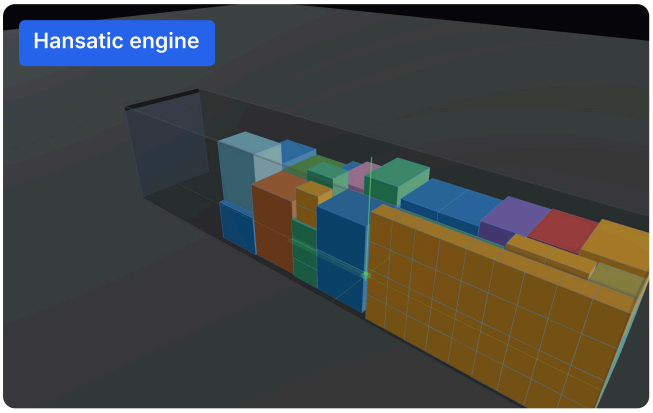
Entrant	Status	Placed	Score	Notes
Hansatic engine	**Legal**	**119 / 119**	**17 positions, 6.83 m (22.4 ft)**	
GPT-5.5	Legal	119 / 119	23 positions, 9.07 m (29.8 ft)	
Opus 5.5	Legal	119 / 119	18 positions, 6.99 m (22.9 ft)	
Grok 4.7	Legal	119 / 119	22 positions, 8.64 m (28.3 ft)	
Kimi K3.1 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
DeepSeek V4 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
Gemini 4 Argon	No plan returned	-	-	The model hit its output limit before writing a plan.
Qwen3p8	No plan returned	-	-	The model hit its output limit before writing a plan.
Magistral Medium	Illegal	107 / 119	21 positions, 16.18 m (53.1 ft)	12 wrong item sizes, 66 overlapping items, 35 outside the vehicle, 60 floating items, 1 crushed items, 1 loose groups not braced by the load

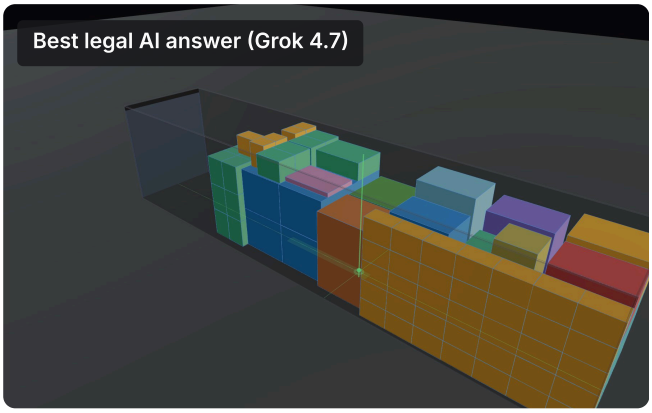
BV4 Shortest load

Vehicle	Pieces	Objective	Result
40' High Cube	145	Shortest load	Engine clearly ahead

What makes it hard

11 non-stackable items; 11 fragile items, top only; 20 items that must stay upright





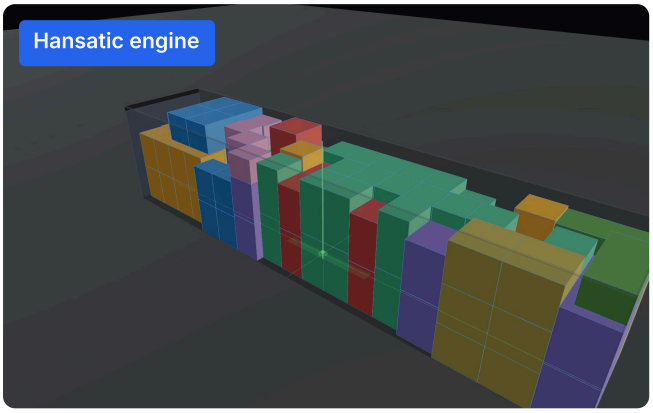
Entrant	Status	Placed	Score	Notes
Hansatic engine	**Legal**	**145 / 145**	**8.57 m (28.1 ft)**	
GPT-5.5	Legal	145 / 145	11.96 m (39.2 ft)	
Opus 5.5	No plan returned	-	-	The model hit its output limit before writing a plan.
Grok 4.7	Legal	145 / 145	8.85 m (29.0 ft)	
Kimi K3.1 Pro	Legal	145 / 145	10.87 m (35.7 ft)	
DeepSeek V4 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
Gemini 4 Argon	Illegal	145 / 145	10.05 m (33.0 ft)	19 overlapping items, 3 outside the vehicle, 9 over the height limit, 4 floating items
Qwen3p8	No plan returned	-	-	The model hit its output limit before writing a plan.
Magistral Medium	Legal	11 / 145	10.33 m (33.9 ft)	

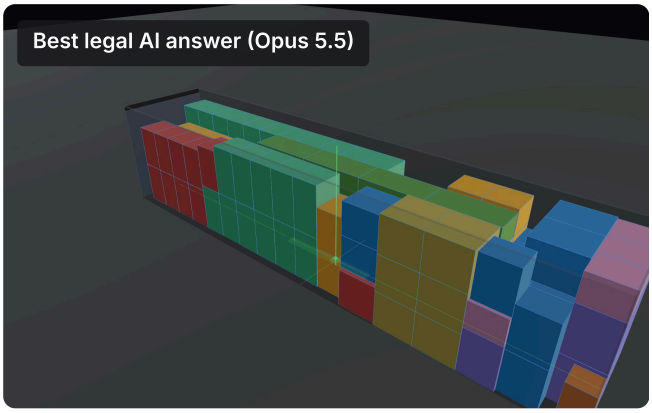
BV5 Shortest load

Vehicle	Pieces	Objective	Result
40' High Cube	136	Shortest load	Engine clearly ahead

What makes it hard

136 items that must stay upright





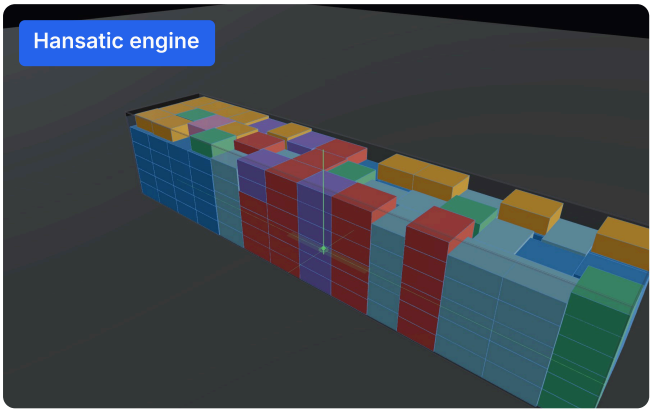
Entrant	Status	Placed	Score	Notes
Hansatic engine	**Legal**	**127 / 136**	**12.00 m (39.4 ft)**	
GPT-5.5	Legal	115 / 136	11.73 m (38.5 ft)	
Opus 5.5	Legal	124 / 136	11.77 m (38.6 ft)	
Grok 4.7	Legal	115 / 136	11.87 m (39.0 ft)	
Kimi K3.1 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
DeepSeek V4 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
Gemini 4 Argon	Illegal	121 / 136	12.00 m (39.4 ft)	1 tilted an item that must stay upright, 12 overlapping items, 6 outside the vehicle, 6 floating items
Qwen3p8	No plan returned	-	-	The model hit its output limit before writing a plan.
Magistral Medium	Illegal	51 / 136	12.89 m (42.3 ft)	34 outside the vehicle, 6 floating items

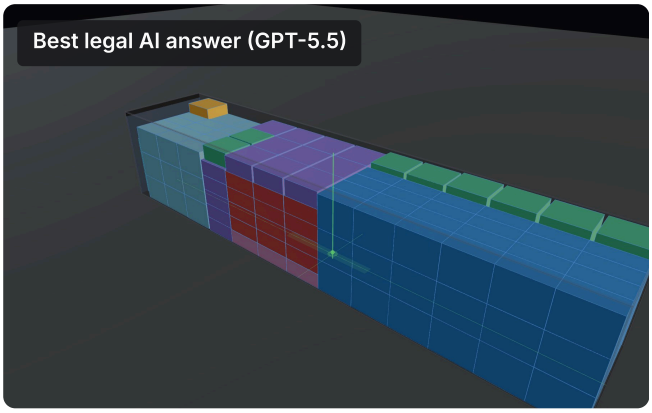
BV6 Shortest load

Vehicle	Pieces	Objective	Result
40' Dry Container	253	Shortest load	Engine clearly ahead

What makes it hard

mixed sizes, no special flags





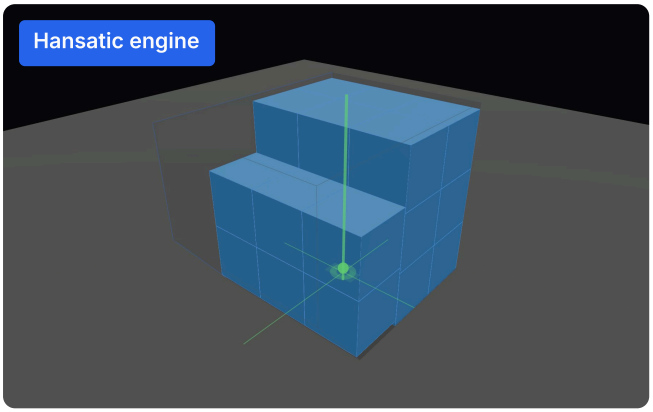
Entrant	Status	Placed	Score	Notes
Hansatic engine	**Legal**	**244 / 253**	**11.98 m (39.3 ft)**	
GPT-5.5	Legal	238 / 253	11.97 m (39.3 ft)	
Opus 5.5	No plan returned	-	-	The model hit its output limit before writing a plan.
Grok 4.7	Legal	218 / 253	11.98 m (39.3 ft)	
Kimi K3.1 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
DeepSeek V4 Pro	No plan returned	-	-	The model hit its output limit before writing a plan.
Gemini 4 Argon	Illegal	197 / 253	11.50 m (37.7 ft)	13 overlapping items
Qwen3p8	No plan returned	-	-	The model hit its output limit before writing a plan.
Magistral Medium	Illegal	42 / 253	12.18 m (40.0 ft)	3 outside the vehicle

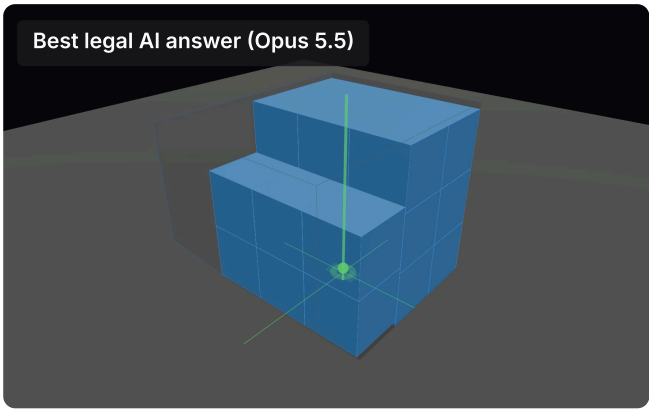
U01 Most pieces placed

Vehicle	Pieces	Objective	Result
AKE (LD3)	50	Most pieces placed	Tie

What makes it hard

sloped wall





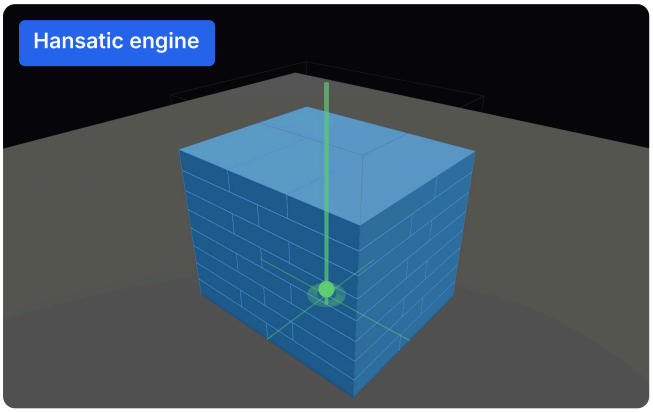
Entrant	Status	Placed	Score	Notes
Hansatic engine	**Legal**	**24 / 50**	**24**	
GPT-5.5	Legal	18 / 50	18	
Opus 5.5	Legal	24 / 50	24	
Grok 4.7	Legal	24 / 50	24	
Kimi K3.1 Pro	Legal	24 / 50	24	
DeepSeek V4 Pro	Legal	18 / 50	18	
Gemini 4 Argon	Legal	24 / 50	24	
Qwen3p8	Legal	24 / 50	24	
Magistral Medium	Illegal	24 / 50	24	3 outside the air container contour

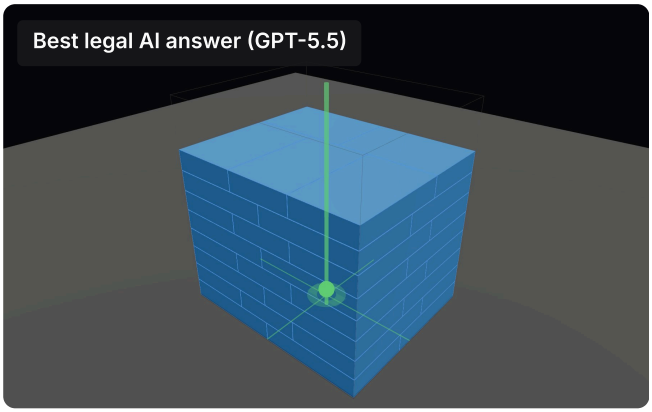
P01 Most interlocked pallet

Vehicle	Pieces	Objective	Result
Industrial / Block 1200×1000	40	Most interlocked pallet	Tie

What makes it hard

40 items that must stay upright; items up to 0.60 m (2.0 ft) long





Entrant	Status	Placed	Score	Notes
Hansatic engine	**Legal**	**40 / 40**	**100%**	
GPT-5.5	Legal	40 / 40	100%	
Opus 5.5	Legal	40 / 40	100%	
Grok 4.7	Legal	40 / 40	100%	
Kimi K3.1 Pro	Legal	40 / 40	100%	
DeepSeek V4 Pro	Legal	40 / 40	100%	
Gemini 4 Argon	Illegal	35 / 40	100%	5 wrong item sizes, 12 floating items
Qwen3p8	Legal	40 / 40	100%	
Magistral Medium	Illegal	40 / 40	100%	30 overlapping items, 9 outside the vehicle