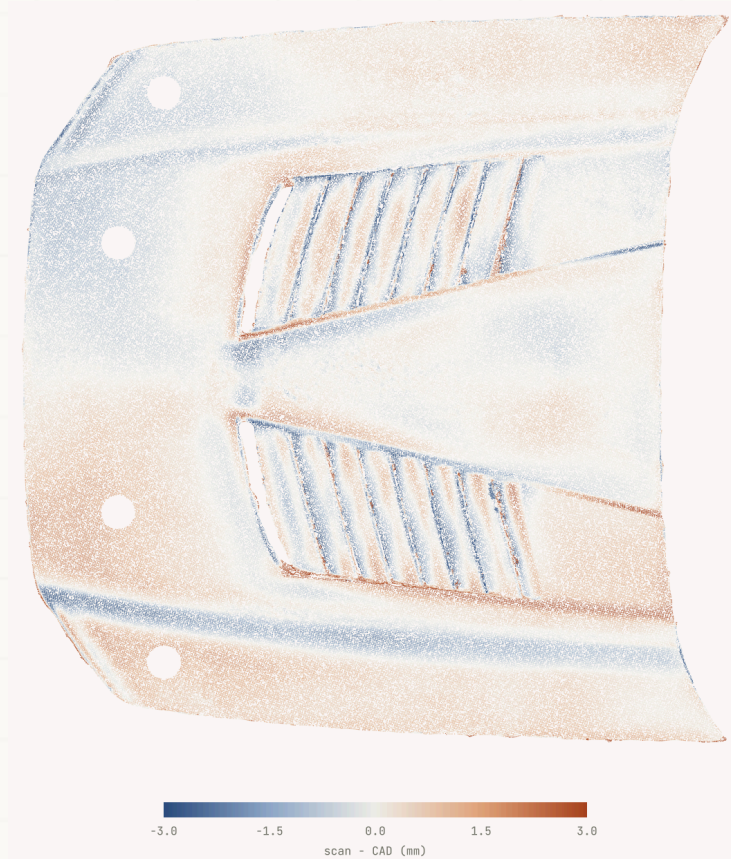


FIELD NOTES • AI-CAD BENCHMARK

Where AI-CAD breaks.

We handed an AI CAD agent two real, scanned fiberglass hoods and one rule: **log every miss**. It modeled them, checked its own work against the scan, and wrote down everything it got wrong, including the misses its checks only just caught. These are the notes.



GROUND TRUTH Signed deviation, scan to AI-built CAD. Every claim in these notes is measured against a physical part.

Tracehaus is a scanning and reverse-engineering studio. We're building toward being the **physical ground truth for AI-CAD**: real parts, real scans, and a grader that knows what a moldable, machinable part actually looks like.

The source is a client job, anonymized, and shared with permission. The public benchmark set will use parts we own.

23

FAILURE MODES
LOGGED ACROSS 2 PARTS

9

FIXED OR WORKED AROUND
BEFORE DELIVERY

2

DEFECTS THAT PASSED
"VALID SOLID" CHECKS

10

STILL OPEN
AND PUBLISHED ANYWAY

01 The test

tier: hard • freeform molded shells

Two hand-built fiberglass hoods, each scanned top and bottom. One has a raised dome, twin scoops and drafted louvers. The other has vent inserts and a bonded inner panel. Neither came with drawings.

The agent works in Python on an open-source CAD kernel. It runs the code itself and owns the whole loop. A human answers its questions, makes the design calls, and scores the result independently.

01 Intake

Load and render every file, then ask for what's missing in one short list.

02 Read the part

Feature list, datums, and design intent in plain language.

03 Model

One parametric script, with every driving dimension named at the top.

04 Self-check

Deviation, sections, validity, draft, and rebuilds with changed parameters.

05 Report

Numbers, renders, a source for every dimension, and a failure log.

EVERY DIMENSION CARRIES ITS SOURCE

TAG	MEANING
MEASURED	Calipers on the physical part
SCAN	Derived from the mesh
STANDARD	From a spec (fastener, bearing)
ASSUMED	The agent's guess, never passed off as fact
CONFIRMED	An assumption the owner signed off
DECISION	An owner's design call that overrides the scan

SCORED 1–5, BY THE AGENT AND BY US

CRITERION	QUESTION
Accuracy	Deviation against the scan and the measured dimensions
Intent	Does the feature tree match how the part was designed?
Editability	Change one parameter: is the geometry still valid?
Manufacturability	Could it be made by the original process?
Fit	Does the fabricated part fit its mate?

Where the agent's score and ours disagree is the interesting part, so we keep both. Fit can only be scored after fabrication.

02 What broke

23 failure modes • 2 parts

BY KIND

Design intent		7
Geometry construction		6
Data & evidence		5
CAD kernel limits		4
Process		1

BY STATUS

Fixed		8
Worked around		1
Owner's decision		2
Flagged		2
Still open		10

TEN THAT MATTER

KIND	TAG	WHAT HAPPENED	CAUGHT BY
CONSTRUCTION	slanted-cap-void	A cutting tool's end cap came out tilted, leaving a hidden void and a 6.7 mm step.	Section overlay
CONSTRUCTION	tangent-boolean	A cut where two surfaces only touched silently did nothing, producing an inverted part that was still a "valid" solid.	Section overlay
CONSTRUCTION	slot-normal-wrong	Swapped trig in a cutter. It worked at 75° by luck and failed at 85°. The same bug had already shipped in the first part.	Editability sweep
INTENT	draft-ignored	Louver steps modeled vertical on a molded part. The scan shows 70–77°.	Self-review
INTENT	rim-not-parametric	Raised rims captured only inside fitted surfaces, not as editable features.	Self-review
KERNEL	fillet-failed	Fillets fail where B-spline ramps meet drafted planes, even at R2.	Kernel error
KERNEL	thicken-failed	Thickening a 400-face freeform plug hung. Worked around by rebuilding the tree on offset surfaces.	Timeout
DATA	thickness-unobservable	Two separately scanned faces share no surface, so laminate thickness can't be measured from them. Fits gave 17.4 mm and -0.7 mm.	Physics check
DATA	periodic-feature-aliasing	Hole-based registration locked onto a shift of two vent pitches (about 150 mm).	Visual check
PROCESS	file-misread	An ambiguous question about file names led to loading the top of another hood as the "underside".	First render

03

Five lessons

what we changed in the loop

01 "Valid solid" is not "correct part".

Two defects passed the kernel's validity check: an inverted vent insert and a hidden void. Only section overlays against the scan caught them. Sections are now mandatory in every self-check.

02 Deviation needs a floor.

A symmetric model can't beat an asymmetric original. These hoods are twisted left to right by 0.73–0.98 mm RMS, so we measure that first and judge the CAD against it. Coarse CAD tessellation also inflated the error fivefold until we fixed the measuring tool.

03 Some numbers can't come from a scan.

The agent tried to fit laminate thickness from two separate scans and got 17.4 mm, then -0.7 mm. Neither is physical. One caliper reading beats any amount of fitting, so intake now asks for one.

04 Editability tests find real bugs.

A cutter that worked at the scanned 75° draft cut nothing at 85°, and the same bug was already in the first part. Every part now gets rebuilt with 3–5 changed parameters before anyone sees it.

05 Kernel limits are predictable, so plan for them.

Fillets on boolean-split B-spline edges fail, and thickening large freeform solids hangs. Rebuilding the same feature tree on offset surfaces gave a uniform 3.00 ± 0.03 mm skin. Design around the limit instead of retrying it.

DESIGN INTENT IS A DECISION

Equal louver pitch is cleaner than the hand-made original. Because of it, and the sharp crests, the louver area lands at 3.6 mm P95 against the scan. The owner made that call, and the model records it as a DECISION rather than a fitted value.

OWNER VS. EVIDENCE

The owner said the louvers were open; the underside scan shows them closed. The model follows the owner, puts the choice behind a one-line toggle, and flags the conflict.



LESSON 5, IN SECTION Hood A's skin (black) over the scan (wide grey): uniform thickness from the offset rebuild, with drafted louver steps and open slots.



LESSON 3, IN SECTION Hood B across the vents. The inner panel (grey) sits where an assumed 5 + 5 mm stack puts it. The scans alone can't say.

04 Scorecard

agent self-scores • ours follow fabrication

PART	ACCURACY	INTENT	EDITABILITY	MFG.	FIT
Hood A dome · scoops · louvers	3	4	3	4	–
Hood B skin · inner panel · inserts	3	3	3	4	–

WHERE IT HELD UP

- Hood B's skin sits at 0.58 mm RMS from the scan, at the original's own asymmetry floor
- Feature trees read like a designer built them: straight creases, constant drafts, equal pitch
- Every face drafted for the mold pull: no undercuts on hood B's skin or inner panel, and 8 mm² on hood A's plug
- One shared base surface drives both hoods

WHERE IT DIDN'T, YET

- Fillets on louver crests, which the CAD kernel can't build
- The inner panel is fitted, not built as pads and beams
- Stitching skin and inner fails at 3 mm thickness
- Fit on the car hasn't been tested



SECTION AI-built vent insert (green) seated in the skin (black) over the inner panel (grey), against both scans (light grey).

WHAT WE OFFER AI-CAD TEAMS

Physical test parts

Parts we own, scanned on both faces, with caliper-verified critical dimensions. Tiered from prismatic brackets to freeform shells.

Graded CAD

Your model's output scored on accuracy, intent, editability, manufacturability and fit, with deviation maps and section overlays.

Failure data

Tagged failure modes with the check that caught each one: the cases a "valid solid" test passes and a mold shop rejects.

Building AI-CAD?
We bring the ground truth.

PHYSICAL PARTS • VERIFIED SCANS • GRADED CAD •
FAILURE DATA

info@trace.haus

trace.haus

Los Angeles, CA