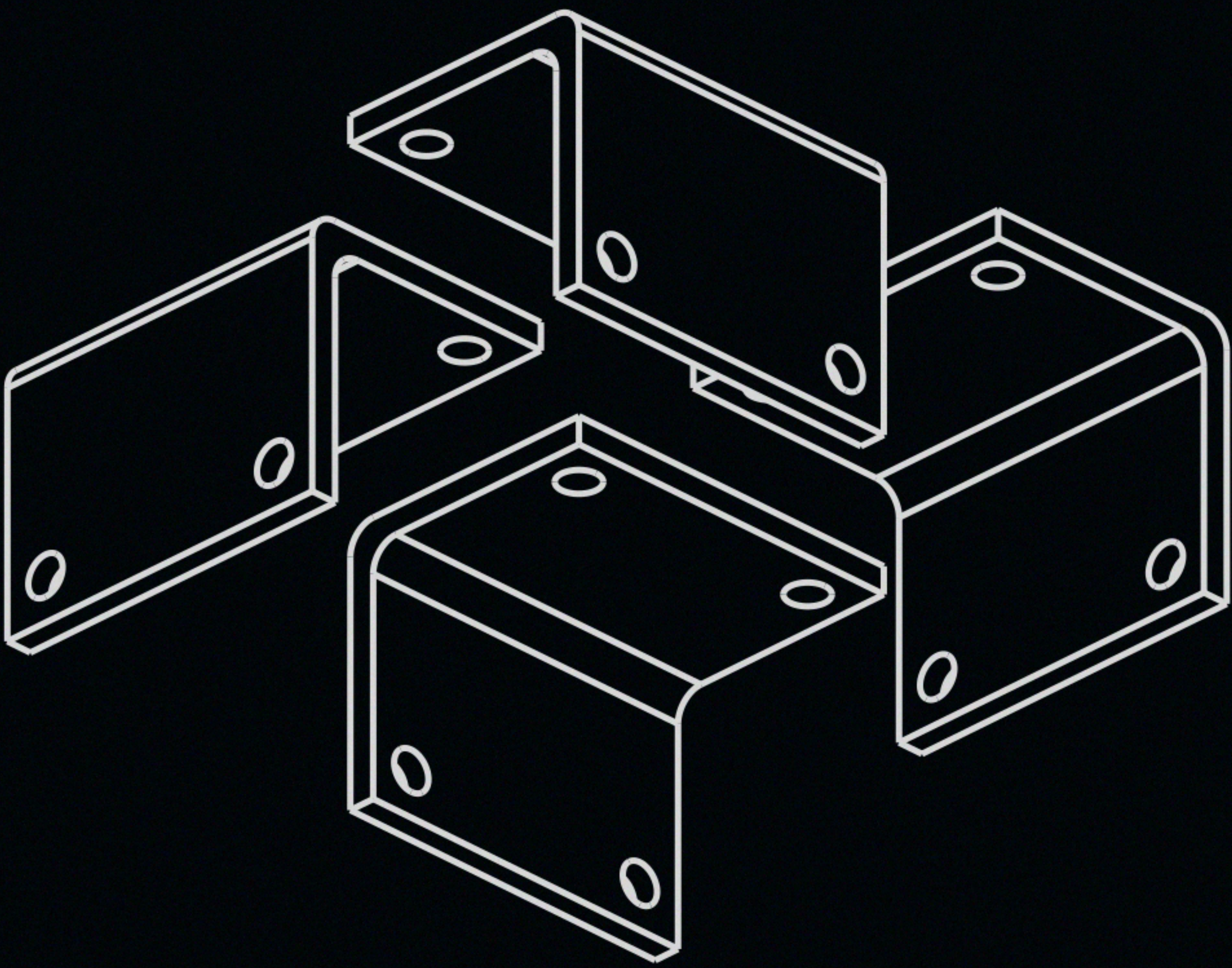
Conversational CAD Under Iterative Human Direction

An instrumented four-bracket case
and workstation-development decision



Christopher Lyon
Lyon Industries / Stavanger, Norway
christopher@lyon-industries.no

PUBLIC REVIEW DRAFT / NOT PEER REVIEWED
Working Paper 03 / 16 July 2026

www.lyon-industries.no/white-papers

Cover: undistorted orthographic line rendering generated from the version-13 STEP artifact described in this paper. White geometry on a dark field; no generated or stock imagery.

Abstract

This paper reports one builder-operated case of a conversational mechanical-design instrument. A stored design brief, injected through the app-server's developer-instruction layer, requested four identical angle brackets, one exact canonical body, three feature-preserving copies, build-plate placement, spacing, centring, measurement, replay, topology, STEP, and GLB checks. The instrument completed the turn with a payload-reported duration of 144.859 seconds; captured notification timestamps span 144.811 seconds. The operator separately observed 126 seconds using a handheld stopwatch with different start and stop actions. Thirty-one exact-CAD tool calls completed without failure. Fourteen CAD events were accepted and none rejected. The final model contains four solids with equal 6,649.148246 mm^3 volumes, a 95 by 85 by 30 mm aggregate envelope, 15 mm minimum adjacent axis-aligned bounding-box clearance, centred X and Y bounds, and Z minima at the build plate. Canonical replay, topology, same-stack STEP reimport, and GLB parsing passed. This documents one traceable transaction from layered input to exact geometry and internally checked artifacts. It does not establish population-level productivity, design correctness outside the encoded requirements, manufacturability, physical performance, or commercial demand. The evidence supports advancement from research instrument to a controlled systems-development programme; venture promotion remains gated by comparative, physical, multi-operator, cost, and buyer evidence.

Executive decision record

Item	Decision record
Expensive decision	Advance the instrument into controlled workstation systems development; do not promote it to a venture yet.
Observed result	One layered-input transaction produced four exact solids in one completed turn, with a retained local transaction record.
Recorded time	144.859 s in the completed-turn payload; 144.811 s between captured start and completion notification timestamps.
Operator observation	126 s by handheld stopwatch with unsynchronised button endpoints.
Permitted claim	One builder-operated layered-input case passed the instrument's internal software acceptance criteria.
Excluded claims	General CAD competence, comparative productivity, manufacturability, physical conformance, structural fitness, safety, customer value, and doctoral validation.
Immediate action	Build and evaluate the engineering workstation around the exact-CAD authority model; then run preregistered comparative and physical tests.
Venture gate	Independent use, baseline advantage after review cost, manufactured and inspected articles, buyer budget, lifecycle cost, and bounded liability.

Scope of claims. One frozen local instrument converted one layered input into four exact solids and versioned STEP and GLB artifacts in one completed turn. The system did not release, slice, print, inspect, load, install, or field a physical part in this case.

Revision type. Instrumented single-case working-paper checkpoint. This is a public decision record and research object, not a completed comparative study or peer-reviewed paper.

Contents

Reading and adjudication guide

The paper separates four levels of statement. An *observation* is directly recorded in the frozen run. An *inference* is a bounded interpretation of those observations. A *proposal* specifies future work. An *excluded claim* names what the evidence cannot support. A committee or technical reviewer can trace the selected public observations to identifiers, hashes, and explicit definitions in Appendix A. Full reconstruction requires the controlled-access evidence packet described there; public hashes alone establish continuity only for a reviewer who also holds the bytes. The interpretation can be challenged in Sections 10 and 13, and the proposed programme can be rejected using the invalidators in Section 13.

The case answers a narrow question deliberately. It does not use the visual quality of a render as evidence that the part is correct. Exact B-rep state, ordered transactions, body and feature identities, measurements, replay, topology, and exchange checks carry the software claim. A future physical claim requires a drawing, process record, actual article, calibrated inspection, and disposition.

1 Problem, decision, and research question

Mechanical design by conversation is easy to demonstrate and difficult to evaluate. A model can produce a plausible shape, a screenshot, or syntactically valid code while silently changing a dimension, losing feature intent, placing a body in the wrong coordinate frame, or exporting a mesh that cannot support the downstream decision. The engineering problem is therefore not text-to-shape generation alone. It is controlled change under explicit authority, with enough evidence to reconstruct what was requested, what changed, what was rejected, what exact model exists, and what may proceed.

The expensive decision is whether the demonstrated instrument should advance into systems development as an engineering workstation, while remaining outside venture status until it survives wider tests. The immediate research question is:

> *Can the frozen instrument transform one demanding layered input into a centred four-body exact model, preserve the transaction, and produce internally checked artifacts within one completed conversational turn?*

The full programme asks a harder question: can engineers iteratively converse with an agent from ambiguous requirements through exact CAD, drawings, manufacturing preparation, release, physical inspection, and corrective action without losing design authority or evidence? This case reaches checked pre-manufacturing geometry. It does not reach the drawing, process, printer, material, inspection, or field-use stages.

1.1 Why the four-bracket case matters

The case is small enough to inspect and demanding enough to exercise multiple failure opportunities. It requires two flange solids, Boolean union, two bend features, four holes on orthogonal faces, canonical validation, copying a modified body, deterministic feature identity, four rigid placements, coordinate-frame agreement, origin centring, clearance calculation, per-body measurement, replay, topology, and two exchange formats. A result can look plausible while failing any of these requirements.

The case is not representative of all mechanical design. It contains no sketch constraint network, assembly mate, tolerance stack, fastener selection, material decision, load case, drawing, manufacturing process selection, or inspection plan. Its value is as a compact transaction test and a systems development checkpoint.

2 Related work and bounded contribution

Text-conditioned CAD generation is an active research field. Text2CAD maps natural-language descriptions to sequential parametric CAD commands [5]. CAD-Llama trains a language model for parametric 3D CAD generation [6]. Constraint-generation research evaluates whether produced constraints express design intent rather than merely solvable geometry [2]. CAD-Assistant is especially close: it uses a tool-augmented vision-language model to interact iteratively with FreeCAD and inspect current design state [8].

The closest 2026 work narrows the architectural novelty further. ToolCAD trains tool-using agents to interact with a CAD engine and describes an MCP-mediated workflow [3]. Embodied CAD resolves agent actions into typed geometric operations executed against an exact parametric B-rep backend with solver feedback [7]. CAD-Refiner joins generation and iterative editing with geometry-aware checking [14]. ArtisanCAD uses an executable intermediate representation containing ordered operations, MCP bindings, dependencies, generated entities, and verification rules against a CATIA backend [13]. These systems already cover most of the broad "agent plus tools plus CAD feedback" proposition.

The present work therefore does not claim a distinct agentic-CAD architecture. It reports a provenance-rich local case in which the conversational model requests typed transactions from a project-scoped exact-CAD service. The service controls B-rep state, versions, identifiers, rejection, replay, measurement, validation, and artifact generation. The app-server provides the thread and turn protocol [12]; MCP defines discoverable typed tools [9]; CadQuery provides the scripted CAD interface [1]; and OCCT supplies exact B-rep construction and STEP exchange [10, 11]. The browser receives a derived GLB and cannot edit canonical geometry.

Table 1: Position of this checkpoint relative to adjacent work.

Work	Primary input/output	Principal emphasis	Difference here
Text2CAD	Text to CAD command sequence	Sequential parametric generation	This case evaluates an instrumented transaction and artifact chain.
CAD-Llama	Text to parametric 3D model	Learned CAD generation	No new model architecture is proposed.
Constraint alignment	CAD constraints and design intent	Semantic suitability of constraints	This case has no sketch constraint solver.
CAD-Assistant	Iterative model/tool interaction with FreeCAD	Generic tool-augmented task solving	This paper adds no general task benchmark.
ToolCAD	Text to iterative CAD-engine tool use	Agent training and curriculum reinforcement learning	This case studies a frozen deployed transaction, not policy training.
Embodied CAD	Typed actions to exact parametric B-rep assemblies	Solver-grounded planning and learning	This paper contributes no planner or solver-grounded benchmark.
CAD-Refiner	Multimodal generation and editing	Unified iterative model and geometry-aware checking	This case reports service-level provenance.
ArtisanCAD	Expert skills and CAD-IR to CATIA B-rep	Industrial skill distillation and verification rules	This case has no skill-learning contribution.
This checkpoint	Layered input to exact state and artifacts	Run provenance, authority allocation, and claim limits	One selected local case; no benchmark or comparative effect.

The contribution is therefore a bounded case record with separated input layers, runtime and model versions, tool-call inventory, accepted transactions, canonical feature identities, measurements, validation outcomes, artifact digests, explicit timing endpoints, and a controlled-access evidence packet. It does not propose a new geometric kernel, CAD representation, language-model architecture, or product-data standard. The tested STEP declares ISO 10303-214 in its file header. STEP AP242 is relevant as a prospective managed-model context [4]; it was not the exchange schema tested in this case.

3 Claim model and evaluation logic

The paper uses a claim ladder to prevent evidence from migrating between stages. Software evidence can establish that an exact model and checked artifacts exist. It cannot establish that the requirement is correct, the part can be manufactured as intended, or the installed article is safe.

Table 2: Claim ladder for conversational mechanical design.

Stage	Question	Minimum evidence	Case status
C0 connection	Can conversation invoke the CAD service?	Thread, tool call, response	Passed previously and in this run.
C1 exact state	Does canonical exact geometry exist?	B-rep state, measurements, topology	Passed for this case.
C2 transaction	Can the change be reconstructed?	Prompt, ordered calls, versions, identifiers, accepted/rejected events	Passed locally.
C3 exchange	Do derived files remain consistent?	STEP reimport, GLB parse, declared deltas, digests	Passed under frozen checks.
C4 review	Does an independent engineer accept the model and drawing?	Review rubric, findings, rework, release decision	Not run.
C5 manufacture	Does the physical article meet its drawing?	Process, article photographs, calibrated inspection	Not run.
C6 use	Does it work safely in context?	Load, installation, field acceptance, failures	Not run.
C7 value	Does the workflow create net buyer value?	Comparative time, review cost, budget, adoption, support cost	Not run.

The permitted conclusion is C1–C3 for one case. Advancement into systems development is a judgment based on that result, not a higher-stage empirical finding. Venture potential is a hypothesis whose invalidators are stated before commercial promotion.

4 System architecture and formal invariants

4.1 Authority allocation

The operator owns the brief and decides what may proceed. The conversational model interprets intent and sequences tools. JSON schemas define admissible requests. The exact-CAD service owns bodies, features, versions, replay, measurements, validation, and artifacts. The viewport observes GLB and cannot mutate the model. A successful chat message has no independent authority.

Operator brief → thread and turn runtime → typed MCP request → schema and transaction → exact B-rep state → validation and immutable version → STEP / GLB / evidence view

Starting a new design creates an isolated design and branch workspace. The conversation thread does not own geometry. Each accepted geometry mutation creates a new immutable version bundle; a rejected mutation leaves the prior canonical version in authority and preserves the failure event. Branching must fork both conversation and exact state to avoid discussing one model while displaying another.

4.2 Transaction model

Let S_i be canonical state at version i, o_i a typed operation batch, F the exact modelling function, and V the frozen validation predicate. The accepted transition is

$$S_{i+1} = F(S_i, o_i) \quad \text{only if} \quad \text{schema}(o_i) \wedge V(F(S_i, o_i)) = \text{pass}.$$

If either predicate fails, $S_{i+1} = S_i$ and a rejected event is appended. The model digest identifies canonical geometry and stable model semantics. The state digest additionally binds the versioned state record. STEP and GLB are derived functions of accepted state and receive their own byte digests.

Canonical replay applies the accepted operation sequence to a fresh compatible workspace and compares the rebuilt result. Replay is stronger than preserving only the final bytes because it tests whether the recorded operations still produce the declared canonical result. It remains dependent on compatible source and dependency versions.

4.3 Copy and feature identity

The canonical bracket contains six named features: two bend features and four holes. Copying a modified body creates a new exact solid and maps each source feature to a deterministic body-scoped identifier such as `bracket_90:inside_bend_r3`. This prevents feature-identifier collisions while retaining correspondence to the canonical feature.

Each placement uses a rigid transform

$$T_k(p) = R_z(\theta_k)p + t_k, \qquad \theta_k \in \{0^\circ, 90^\circ, 180^\circ, 270^\circ\},$$

where t_k has X and Y components only. Because $\det(R_z) = 1$, the transform preserves volume and topology. No body is rotated about X or Y after the canonical build-plate pose is established.

4.4 Coordinate agreement

The CAD world uses XY as the build plate at Z=0, with +Z upward. GLB exchange uses a different up-axis convention. The viewport therefore maps the imported GLB basis into the CAD basis before computing bounds, framing, or display. This mapping is part of the instrument definition: labels alone cannot resolve a coordinate disagreement.

The case-level invariants are:

$$\min Z_k = 0, \quad \max Z_k = 30, \quad \frac{\min X + \max X}{2} = 0, \quad \frac{\min Y + \max Y}{2} = 0,$$

with at least 15 mm pairwise axis-aligned bounding-box clearance for adjacent bodies. These conditions are measured from canonical CAD-world bounds rather than inferred from the camera view.

5 Case protocol and frozen method

5.1 Instrument freeze

The reported case began after the tool schemas, modified-body copy semantics, coordinate contract, and viewport basis mapping were verified and the local instrument was restarted. No source or schema change occurred during the reported turn. The source digest recorded by the final manifest was `d7813ae26be301ec20f e356f33482c4526025fee6efec08c144f362fd5d558d7`. That digest covers only the top-level `cad_mcp/*.py` files. It does not bind the app-server prompt assembly, runtime configuration, dependency locks, tests, or viewport code. "Frozen instrument" in this paper therefore means no observed change during the selected turn, not a complete repository-level cryptographic freeze. This is an instrumented final case, not a claim that the development process required no prior correction.

5.2 Input contract

The input was layered. A fixed cockpit developer instruction established tool, authority, coordinate, placement, and evidence rules. The stored design brief was appended to that developer instruction and specified two equal 30 mm legs, 40 mm width, 3 mm flange solids, inside and outside R3 edge features, four 4.5 mm through-holes with 5 mm named-edge offsets, one canonical bracket, four independent solids, four world-Z rotations, at least 15 mm clearance, aggregate centring, and final checks. The actual user turn message was the shorter instruction to read exact state, restate intent and unknowns, and create the first justified geometry. MCP tool descriptions and generated JSON schemas formed a fourth input layer.

Appendix B reproduces the stored design brief; Appendix C reports the actual user message and the fixed instruction contract. The full developer-instruction string and current tool schema snapshot are present in the controlled-access packet. They were reconstructed after the run from retained source and are not independently bound by the final run manifest; this is a reproducibility limit.

The case used one local computer and one operator. The runtime record identifies model `gpt-5.6-sol`, `codex-cli 0.144.1`, Python 3.12.3, CadQuery 2.8.0, MCP 1.28.1, trimesh 4.12.2, and macOS 26.5.2 on arm64. The thread, turn, design, source, conversation, event, manifest, STEP, and GLB identifiers are retained.

5.3 Timing definitions

Two app-server time values are retained. The `turn/completed` payload reports 144,859 ms. Independently captured notification timestamps from `turn/started` at 05:20:32.483 UTC to `turn/completed` at 05:22:57.294 UTC span 144,811 ms. Their 48 ms difference is consistent with different internal clocks, rounding, or serialization boundaries; the archive does not identify which. Neither value is silently substituted for the other. The operator's stopwatch is a secondary contemporaneous observation whose button presses were not synchronized to either app-server measure.

Four internal intervals support reconstruction:

1. turn start to first accepted CAD event;
2. first accepted CAD event to final accepted geometry mutation;
3. final geometry mutation to turn completion;
4. the sum of per-call service durations reported for completed MCP calls.

The call-duration sum measures instrumented service work, not total compute or human time. Model inference, reasoning, serialization, scheduling, UI delay, and report generation remain in the wall-clock interval.

5.4 Acceptance criteria

The case passes only if it produces four independent exact solids; equal body volumes and matching feature sets; the stated dimensions, rotations, layout, spacing, and centring; zero failed exact-CAD calls or rejected CAD events; passed topology and canonical replay; and same-stack STEP and GLB results consistent with canonical measurements within the service's declared tolerances. The STEP checks allow at most 0.005 mm world-bound error and 0.001 mm^3 aggregate-volume error. The GLB checks allow at most 0.25 mm world-bound error and, per body, the larger of 0.1 mm^3 or 10^{-4} of exact body volume. Export tessellation uses 0.01 mm linear tolerance and 0.05 radians angular tolerance. Reported STEP bounds and volume deltas are rounded to nine decimal places; GLB bounds and volume deltas to six. No physical criterion is part of this software case.

6 Run reconstruction

The turn began at 05:20:32.483 UTC. The agent read the initial exact state, created the contact flange, created the rising flange, and united the bodies. It applied inside and outside R3 bend features, added four 4.5 mm through-holes, measured and validated the canonical bracket, created three body copies, placed the four bodies, then re-read, measured, and validated the final state.

The full conversation record contains 31 completed exact-CAD calls and no failed call. Fourteen calls were state reads. The other 17 calls comprised two boxes, one Boolean union, two fillets, four holes, three copies, one transform, two measurements, and two validations. The CAD event log contains 14 accepted events, including scene initialization and 13 geometry mutations, with no rejected event.

Table 3: Exact-CAD call inventory for the completed turn.

Tool	Count	Role in the case
project__get__state	14	Read canonical bodies, versions, features, bounds, and validation state during construction.
primitive__create__box	2	Create the contact and rising flanges.
boolean__combine	1	Unite the flanges into the canonical bracket.
feature__fillet	2	Apply the inside and outside R3 bend features.
feature__hole	4	Create the four 4.5 mm through-holes.
body__copy	3	Create the 90, 180, and 270 degree bodies.
body__transform	1	Place the canonical zero-degree body.
model__measure	2	Measure canonical and final geometry.
model__validate	2	Validate canonical and final geometry and artifacts.
Total	31	All completed; zero failed.

The final geometry mutation was accepted at 05:22:26.149 UTC. The subsequent 31.145 seconds covered final state reads, measurement, validation, and the final report. The runtime archive also contains two earlier bracket-design development sessions. They used an instrument that was still being changed and are excluded from the numerical case result. The selected case was the first run after the final source change and restart; that stopping rule was recorded from the development chronology after the fact, not preregistered. The case is therefore a selected post-development demonstration rather than a random trial or unbiased estimate. Earlier sessions remain in the archive for committee inspection.

Table 4: Bracket-session selection ledger. Earlier sessions are development history, not observations pooled with the final case.

Session	Start (UTC)	Eligibility and disposition
`bracket`	04:51:34	Instrument still under development; ineligible for the frozen case.
`brackets-2`	05:04:15	Instrument still under development; ineligible for the frozen case.
`brackets-3`	05:20:19	First session after final source change and restart; selected as the reported case.

7 Observed geometry and identity result

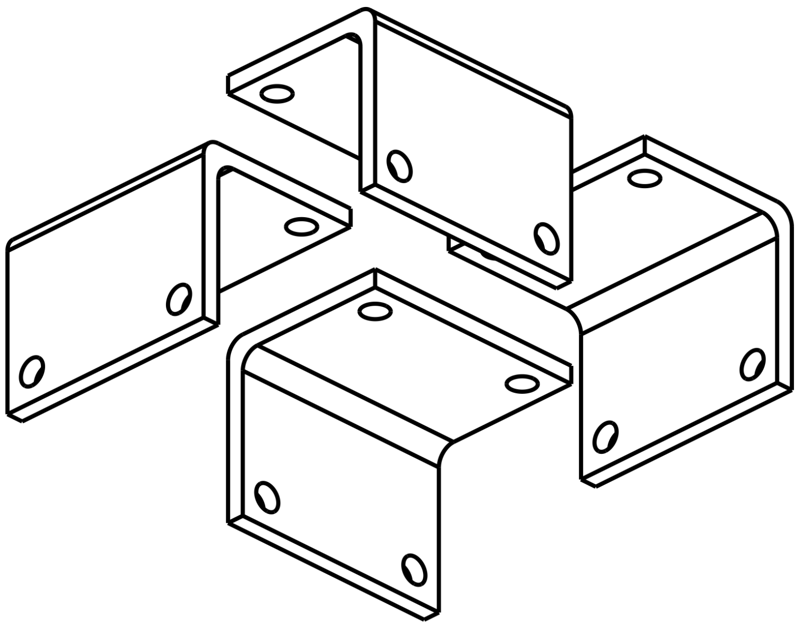


Figure 1: Undistorted black-on-white orthographic projection generated directly from the final STEP artifact. The four exact solids use the same build-plate contact pose and differ only by world-Z rotation and XY translation. This is a CAD rendering, not a physical article.

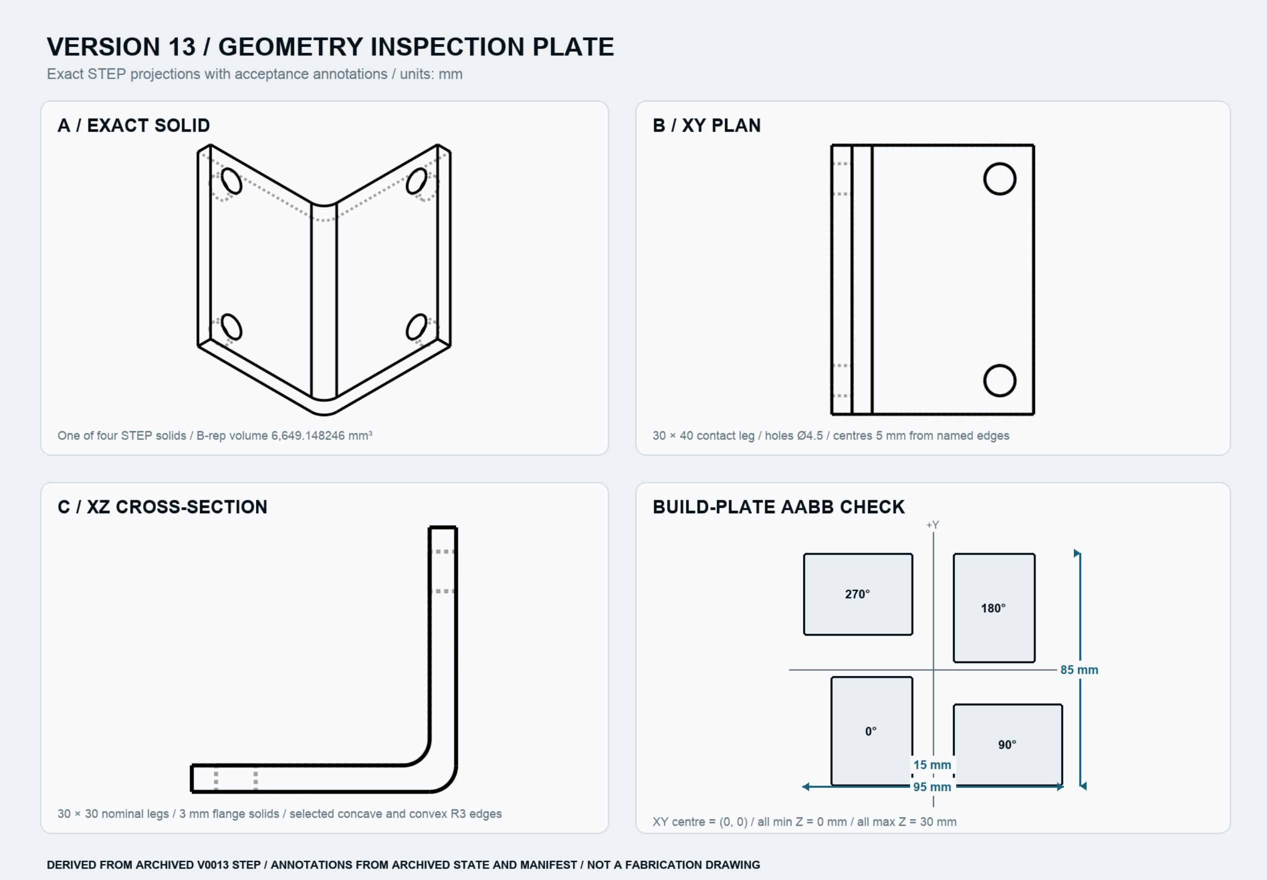


Figure 2: Inspection plate combining exact line projections from the archived version-13 STEP solids with acceptance annotations from the archived state and manifest. It exposes one solid, the contact-leg plan, the XZ cross-section, and the four-body AABB layout. It is not a fabrication drawing.

The final model is version 13. It contains four bodies and four solids. Each body has B-rep volume 6,649.148246 mm^3. The aggregate B-rep volume is 26,596.592985 mm^3 and the aggregate envelope is 95 by 85 by 30 mm.

The canonical solid is a Boolean union of one 30 by 40 by 3 mm box and one 3 by 40 by 30 mm box. Separate R3 fillets were then applied to selected concave and convex longitudinal edges. Equal concave and convex radii on these overlapping flange solids do not represent a constant-thickness formed-sheet bend. The result should be read as the exact solid L-section shown in Figure 2, not as evidence for an ordinary 3 mm sheet-metal bend model.

Table 5: Exact body measurements recorded in the final manifest.

Body	Rot.	Minimum XYZ (mm)	Maximum XYZ (mm)	Volume (mm^3)
bracket_0	0	(-37.5, -42.5, 0)	(-7.5, -2.5, 30)	6,649.148246
bracket_90	90	(7.5, -42.5, 0)	(47.5, -12.5, 30)	6,649.148246
bracket_180	180	(7.5, 2.5, 0)	(37.5, 42.5, 30)	6,649.148246
bracket_270	270	(-47.5, 12.5, 0)	(-7.5, 42.5, 30)	6,649.148246

The X-centre calculation is $(-47.5 + 47.5)/2 = 0$ mm. The Y-centre calculation is $(-42.5 + 42.5)/2 = 0$ mm. Every minimum Z is 0 mm and every maximum Z is 30 mm. For a pair i, j, separation on X is $d_x = \max(0, \max(x_i^{\min}, x_j^{\min}) - \min(x_i^{\max}, x_j^{\max}))$ and equivalently for d_y. The reported adjacent clearance is $\max(d_x, d_y)$ for pairs separated on one build-plate axis. Its minimum is 15 mm for the horizontally or vertically adjacent pairs; diagonal pairs are farther apart.

7.1 Feature correspondence

Each body records six feature identifiers. The canonical body uses `inside_bend_r3`, `outside_bend_r3`, `contact_hole_y5`, `contact_hole_y35`, `rising_hole_y5`, and `rising_hole_y35`. Each copied body contains the same suffixes under its body identifier. Equal feature count alone would not prove geometric equivalence;

here it is combined with deterministic copy semantics, rigid transforms, equal exact volumes, matching Z range, and replay.

Table 6: Acceptance matrix for the frozen brief.

Criterion	Observation	Status
Four independent bodies/solids	4 bodies / 4 solids	Passed
Canonical geometry complete	2 flanges, union, 2 R3 features, 4 holes	Passed
Matching copies	Deterministic scoped feature sets and equal B-rep volume	Passed
Build-plate contact	Every minimum Z equals 0 mm	Passed
Common height	Every maximum Z equals 30 mm	Passed
World-Z rotations only	0, 90, 180, 270 degrees	Passed
Layout centre	X centre 0 mm; Y centre 0 mm	Passed
Minimum adjacent clearance	15 mm	Passed
Call/event completion	31/31 calls completed; 14 accepted, 0 rejected CAD events	Passed
Replay/topology/internal artifact checks	All declared same-stack checks passed	Passed

8 Validation and exchange result

Validation is layered because no single check is sufficient:

1. schema validation constrains the structure and types of tool requests;
2. transactional validation prevents a failed mutation from replacing canonical state;
3. exact topology checks the accepted OCCT bodies;
4. measurement compares bodies, solids, bounds, and B-rep volume;
5. canonical replay rebuilds the accepted operation sequence;
6. STEP reimport checks the exact exchange artifact within the same CadQuery/OCCT implementation used to create and export it;
7. GLB parsing checks the observable tessellated representation and maps its basis back to CAD-world coordinates;
8. cryptographic digests bind the reported version to specific records and files.

Topology and canonical replay passed. The version-creation manifest recorded a four-solid STEP reimport, zero reported volume delta, and 0.000233223 mm maximum world-bound delta. It recorded four GLB bodies, four mesh solids, 56 geometry nodes, 0.079099 mm^3 tessellated-volume delta (approximately 0.000297 percent of aggregate exact volume), and 0.000233 mm maximum CAD-world bound delta. The later final `model_validate` call regenerated and rechecked the files, reported 0.0 mm for both bound deltas, retained the same volume results, and reported matching regenerated digests. Both records are preserved; the small difference in reported bounds is not collapsed into one value.

These checks establish internal consistency among canonical state and two derived representations under one software stack. Construction, replay, STEP export, and STEP reimport all use the same CadQuery/OCCT implementation. No second CAD system has accepted the STEP file. The result therefore does not establish independent exchange, preservation of feature semantics elsewhere, completeness of product definition, correctness of a drawing dimension, or conformance of a physical article.

9 Timing result and efficiency interpretation

Table 7: Named timing endpoints and reconstructed intervals.

Measure	Result	Interpretation
Completed-turn payload	144.859 s	Duration reported inside the final app-server payload.
Notification timestamps	144.811 s	Captured start-to-completion event span; 48 ms shorter than the payload.
Operator stopwatch	126 s	Contemporaneous observation with unsynchronised button endpoints.
Start to first CAD event	6.374 s	Runtime and initial state-read lead-in.
First event to final mutation	107.292 s	Canonical construction, feature creation, copies, and placement.
Final mutation to completion	31.145 s	Final reads, measurement, validation, and response.
Sum of MCP call durations	5.268 s	Service-reported call durations only; not end-to-end compute.

The timing result is that this one case reached its final internally checked state within a payload-reported 144.859 seconds and a captured notification span of 144.811 seconds. The three internal event intervals sum to the latter, not the payload duration. The 48 ms discrepancy is retained because the archive does not establish whether clock selection, rounding, or serialization accounts for it. The 126-second stopwatch observation is consistent with a different operator-visible interval but cannot be treated as the same measure. Neither duration is an estimate of expected performance. There is one observation, no comparator, no variance, and no control for task familiarity.

The case nevertheless has decision value. It shows that conversation, exact geometry, copies, validation, and reporting can occur quickly enough that interaction design and engineering review become the next constraints worth studying. It does not show a time saving. A conventional CAD expert, macro, or parameterized template may complete the same bracket family faster and with lower review burden.

10 Discussion and competing explanations

The strongest result is the joined transaction: one stored design brief, developer-instruction contract, and user turn led to an exact model and internally checked artifacts without a failed exact-CAD call. The case is sufficient to justify treating the instrument as a systems-development candidate rather than a rendering demonstration.

Several explanations limit that conclusion:

- The builder authored and operated the instrument.
- The brief was unusually explicit and aligned with available operations.
- The bracket family is geometrically simple.
- One model, machine, kernel configuration, operator, and run were tested.
- A conventional CAD expert, macro, or script may be faster.
- Zero rejected calls may indicate instrument maturity for this task or unusually favorable task-tool fit.
- The operator stopwatch is not synchronized measurement.
- Passing representation checks cannot detect wrong requirements, load paths, tolerances, material choice, printer settings, support strategy, or unsafe use.

The evidence rejects neither human expertise nor conventional CAD. The proposed workstation is valuable only if it improves the complete engineering outcome: less active effort or fewer errors after accounting for clarification, review, rework, support, and lifecycle cost. Fast generation followed by expensive verification would fail that test.

10.1 What was observed

For this case, the instrument retained a stored design brief and one actual user message; turned the layered input into typed, accepted exact-CAD transactions; constructed one complete canonical body and three traceable copies; aligned the CAD world, derived GLB, and user-visible coordinate frame; met the explicit layout and geometry acceptance criteria; and bind canonical state, STEP, and GLB artifacts. Full reconstruction is available through controlled access, not from the public hashes alone.

10.2 What remains unproved

The case does not establish autonomous requirement interpretation, design novelty, sketch or assembly competence, tolerance design, material selection, analysis, manufacturing preparation, printer control, physical conformance, structural performance, safety, generality, comparative productivity, user adoption, or business value. The phrase button to printed part is an application hypothesis until release, slicing, machine execution, the physical article, and inspection are recorded.

11 Systems-development programme

The recommendation is to advance within *systems*, with the exact-CAD service retained as the source of geometric truth. The engineering workstation should be developed through capability gates rather than an undifferentiated feature list.

Table 8: Proposed systems-development sequence.

Stage	Engineering objective	Required evidence	Exit rule
S1 workstation	Faster conversation, state inspection, feature selection, measurement, branching, recovery	Usability tasks, event integrity, accessibility, failure recovery	Qualified users can complete bounded tasks without state confusion.
S2 design intent	Constraints, dimensions, datums, tolerances, drawings, assemblies, interference	Independent rubric and second-CAD review	Critical intent survives review and exchange.
S3 process preparation	Slicer/printer profiles, build volume, orientation, packing, support, material, estimated time and cost	Reproducible process plan and machine-specific checks	Release package is distinct from checked geometry.
S4 physical loop	Manufacture, identify, photograph, inspect, disposition, revise	Actual articles, calibrated measurements, nonconformance record	Preregistered dimensions pass or failures become rules.
S5 comparative pilot	Multiple operators, tasks, baseline, blinded review	Active time, review, defects, rework, workload, total cost	Net advantage remains after review and support.

The workstation interface should make authority visible. It must distinguish draft intent, accepted exact state, checked artifacts, reviewed release, machine job, physical article, inspection result, and field disposition. Those states should not collapse into a single success indicator. The user should be able to inspect why a body exists, which feature is selected, what changed between versions, which requirement is verified, and which claim is still open.

The cost model must include model inference, engineering review, CAD-kernel and dependency maintenance, artifact storage, workstation support, drawings, printer or supplier time, consumables, inspection, rework, liability, security, and migration. The two-minute result affects only one component of that total.

12 Translation toward a Lyon Industries venture

The plausible venture is an auditable conversational mechanical-engineering workstation for engineers, not a generic text-to-mesh generator. Its potential value lies in shortening bounded design iterations while keeping exact state, review evidence, manufacturing preparation, and corrective history connected.

The initial buyer hypothesis is a small engineering team that repeatedly makes simple fixtures, brackets, enclosures, adapters, and revisions but cannot justify fragmented specialist tooling and manual evidence assembly. A stronger entry point may be teams already using additive manufacturing for shop-floor fixtures and installation aids, because the loop from requirement to inspected article is short enough to measure. This remains a hypothesis, not market evidence.

12.1 Venture thesis

The commercial thesis has four linked propositions:

1. conversational direction reduces active iteration effort for bounded mechanical tasks;
2. exact state, typed change, and evidence reduce review and rework risk;
3. process-aware preparation shortens the route from accepted geometry to a controlled machine job;
4. the combined benefit exceeds inference, integration, support, review, inspection, and liability cost.

Failure of any proposition narrows the product. If conversation saves drafting time but increases review cost, the instrument may remain an internal tool. If the exact transaction record is valuable without machine integration, the product may become a review and change-control workstation. If only simple fixture families perform well, the market and vocabulary should remain narrow.

12.2 Promotion gates

Promotion from systems to venture requires independent multi-operator and multi-task evidence; a named conventional baseline and net review-cost advantage; multiple manufactured and inspected articles with traceable failures; at least one buyer with a budget, workflow owner, and acceptance rule; security, data isolation, source provenance, and retention controls; explicit printer, material, supplier, and geometric support limits; a lifecycle cost and support model; and bounded allocation of design, release, and field liability.

The venture should not be presented as earned merely because the recorded duration is short or the interface is compelling. The present result earns a structured commercial test, not a company claim.

13 Threats to validity, invalidators, and next studies

13.1 Internal validity

The builder was the sole operator and knew the system, prompt style, and desired outcome. The final case was selected after instrument development. The task was not blinded, randomized, or independently authored. Tool semantics were frozen before the run but the broader development history may have improved task fit. The stopwatch endpoints were not synchronized. These conditions make causal productivity claims invalid.

13.2 Construct validity

The case measures time to checked exact geometry, not time to an accepted engineering outcome. Body count, volume, bounds, feature names, replay, and exchange consistency are appropriate for the frozen brief but do not measure design intent beyond it. The phrase matching features depends partly on the copy operation and identifier map; an independent geometric comparison would strengthen it. GLB agreement measures a derived representation, not exact manufacturing geometry.

13.3 External and conclusion validity

External validity is limited to one explicit bracket brief, one operator, one model, one computer, one source revision, and one software environment. There is no sample from which to estimate variance, task effects, operator effects, or a comparative effect. The exact times are observations, not expected values.

13.4 Artifact and accessibility limits

The design brief, actual user message, reconstructed developer instructions, conversation, state, events, STEP, GLB, dependency locks, exact-CAD source, and post-hoc tool-schema snapshot are retained in a 28-file controlled-access committee packet. Its manifest digest is `076a765d5f6bfefd9b7a7d3b38ae582614fe26bb902610b441982724557e066a`. Public hashes support continuity checks only for a reviewer who also holds the bytes. The public PDF is not a replacement for that packet. The generated PDF is also untagged; the HTML companion provides a more accessible reading route but does not remove that limitation.

13.5 Invalidators

This checkpoint is invalidated if the archived files do not reproduce the reported hashes; if canonical replay changes the model; if a second CAD system cannot read the STEP result within a declared tolerance; or if the reported timestamps, calls, bodies, features, or bounds cannot be reconstructed.

The systems-development proposition is invalidated if review and rework consume the time saved, critical errors escape, success depends on post-disclosure target-specific code changes, physical articles fail preregistered inspection, or users cannot recover from state and coordinate errors.

The venture proposition is invalidated if qualified users do not adopt the workflow, buyers will not pay its full support cost, required integrations make deployment uneconomic, security or source-provenance risk cannot be controlled, or liability cannot be bounded.

13.6 Next comparative and physical study

The next study should freeze at least six task classes before disclosure, assign at least two qualified mechanical designers and an independent reviewer, name the conventional baseline, and counterbalance task order. Each attempt should retain active time, clarification, accepted and rejected calls, critical criteria, review findings, reviewer time, rework, workload, and final disposition. Report task cases before aggregate statistics. A pilot may report paired differences, medians, ranges, and bootstrap intervals but should not claim population generality.

One low-risk task should continue through drawing review, slicer or supplier preparation, manufacture, photographic evidence, calibrated dimensional inspection, nonconformance, and disposition. The physical article must be identified and linked to its release artifact. A generated render cannot evidence manufacture.

If the study includes external participants, the relevant institution should determine whether ethics review, consent, data protection, or workplace approval is required before recruitment. Model and event logs may contain proprietary design information and must be governed accordingly.

14 Conclusion

The selected final case passed the instrument's internal software acceptance criteria. In one completed conversational turn, it created four independent exact solids, retained feature correspondence, placed them in a Z-up centred layout, and produced versioned STEP and GLB artifacts. Thirty-one exact-CAD calls completed without failure. Fourteen CAD events were accepted and none rejected. The completed-turn payload reported 144.859 seconds, captured notification timestamps spanned 144.811 seconds, and the operator observed 126 seconds using different endpoints.

The defensible conclusion is narrow: the archived local record documents one builder-operated run in which the same implementation reported a four-solid version-13 model with passing internal replay and artifact checks. External reproduction, second-CAD acceptance, comparative productivity, physical conformance, and customer value have not been established. The result is sufficient to begin controlled workstation systems development under the explicit methods, gates, and invalidators defined here.

A Artifact-integrity record

Table 9: Run identifiers and environment.

Item	Recorded value
Design	`20260716052019-brackets-3`
Thread	`019f695e-5bf1-74e0-ba2d-86bc64bed8b4`
Turn	`019f695e-8eac-7563-893d-ed1c321948ea`
Model	`gpt-5.6-sol`
Runtime	`codex-cli 0.144.1`
Exact-CAD environment	Python 3.12.3; CadQuery 2.8.0; MCP 1.28.1; trimesh 4.12.2; macOS 26.5.2 arm64
Final version	13
Final bodies / solids	4 / 4

Table 10: Selected claim-bearing SHA-256 identifiers.

Item	SHA-256
Exact-CAD source	`d7813ae26be301ec20fe356f33482c4526025fee6efec08c144f362fd5d558d7`
Model	`fd66de5c12bc251aa163c1ce82e98295e75a861cd4d0806f9cd222d3c95894d5`
State	`ed17ff938dcb32ce1a8b214350525947bfb0c09b7fe4d920b10c031ca6600582`
STEP	`66943199c5f4562a6819aed82117a698474c9f1d164a397899fb70d4afd2c257`
GLB	`afa7e4379e35662610ec54a8224766fbc44e2a1637bf2019bc465475392697b9`
Design record	`883d7e0e0efce1074e32f0a605a67d9c333f2d7189f8c71ad5e7dd8d336cf90f`
Conversation	`e48a6c461ec016bcfa9e850fd137f482424f28ae414720bf13ddabd8dfc220f8`
Operations log	`90c1ed1a60df505165ccb1831f990b55547614eed16d0b44ef48bbeb26ae360e`
Final manifest	`8700ec58d5aa35e06b8e6bc5f6b3a4975961cc2e8638f0a7e7e5c72d03cfff30`
Stored design brief	`57a12b7b0a024b5cce357636cac16b9d9d5cb863a70900c30d129530e48d6d3f`
Actual user message	`3b546609461673732fe3831c6d6f639468fb42fbe2dd731a1cd5d156d5c025ef`
Reconstructed developer instruction	`9491305ba410649f222e13efb86b23343e6fa3fc1a3c2b1ca18c8fac59528ed0`
Post-hoc tool-schema snapshot	`43e87fe484a572ca964e0b680ae40922e09b2c9661350d418172e83c543159d5`
Committee-packet manifest	`076a765d5f6bfefd9b7a7d3b38ae582614fe26bb902610b441982724557e066a`

The exact-CAD source, model, and state rows are instrument-domain digests. STEP, GLB, design record, conversation, operations log, final manifest, and committee-manifest rows hash the retained file bytes. The three textual input rows hash their UTF-8 content without an added trailing newline. The tool-schema row hashes compact canonical JSON with sorted keys and no trailing newline. The packet's own `MANIFEST.sha256` separately binds every retained packet file, including newline-delimited reconstructions.

Data availability. The 28-file controlled-access packet contains the stored design brief, actual user message, reconstructed developer instructions, post-hoc tool schemas, app-server conversation, exact-CAD events, canonical state, measurements, manifests, STEP, GLB, dependency locks, and relevant source. It is available for committee inspection but is not published with this checkpoint. The public PDF and HTML companion expose selected values and hashes. The hashes do not make withheld bytes independently available. The developer-instruction and tool-schema hashes identify post-hoc reconstructions and are not run-time bindings.

B Stored design brief injected as design context

```
Begin by reading the exact CAD state. Complete the entire design in this turn
using only the lyon_exact_cad tools.

Create four identical 90-degree angle brackets as four independent exact solids.

Each finished bracket must have:
- Two equal 30 mm legs
- 40 mm bracket width
- 3 mm wall thickness
- 3 mm inside bend fillet
- 3 mm outside bend fillet
- Four 4.5 mm through-holes
- Two holes on each flange
- Every hole centre 5 mm from the flange’s free-end edge
- Every hole centre 5 mm from its adjacent side edge

Use this world coordinate frame:
- The build plate is the global XY plane at Z = 0
- Red is X
- Green is Y
- Blue is Z
- +Z points vertically upward from the build plate
- Gravity points in -Z
- The centre of the build plate is X = 0, Y = 0

First construct and validate one complete canonical bracket. Place one specific
broad outside flange flat and coplanar with the build plate at Z = 0. The
complete bracket must remain at Z >= 0, with its other flange rising toward +Z.

Use that exact same plate-contact pose for all four brackets. After establishing
the canonical pose:
- Do not rotate any instance about X or Y
- Rotate the instances only about world +Z
- Use rotations of 0, 90, 180, and 270 degrees
- Translate instances only in world X and Y
- Keep the same flange against the build plate for every instance
- Keep every bracket’s minimum Z exactly at 0
- Keep every bracket’s maximum Z identical

Arrange the four brackets in the smallest practical square or rectangular
build-plate pattern with at least 15 mm bounding-box clearance between adjacent
brackets.

Centre the complete four-bracket layout exactly around the build-plate origin.
The final aggregate bounds must satisfy:
- (minimum X + maximum X) / 2 = 0
- (minimum Y + maximum Y) / 2 = 0

Do not stop after creating preliminary geometry. Complete all brackets, holes,
fillets, placement, and validation.

Before finishing:
1. Read the final exact state.
2. Measure all four bodies.
3. Confirm four bodies and four solids.
4. Confirm equal volume and matching features for all four brackets.
5. Report each body’s minimum and maximum XYZ bounds.
6. Confirm every minimum Z is 0.
7. Show both aggregate centre calculations.
8. Confirm minimum adjacent bounding-box spacing.
9. Run canonical replay and topology validation.
10. Validate STEP reimport and GLB parsing.
11. Report the final model version, aggregate envelope, model digest, state
    digest, STEP digest, and GLB digest.
```

C Input-layer record

Actual user turn message.

```
Begin by reading the exact state, restating the design intent and unknowns,
then create the first justified geometry.
```

Fixed cockpit developer instruction. The following text is reconstructed from the retained the cockpit session source. The full developer instruction consisted of this fixed text, the design name "Brackets 3", the stored brief in Appendix B, and the fallback used when no formal requirements exist. The reconstructed UTF-8 string is 5,320 bytes with SHA-256 `9491305ba410649f222e13efb86b23343e6fa3fc1a3c2b1ca18c8fac59528ed0`.

```
You are operating the Lyon Agentic Mechanical Design research cockpit.

The operator wants iterative mechanical design, not prose about a hypothetical
model. Read project_get_state first, then use only lyon_exact_cad MCP tools for
geometry. You may autonomously make multiple validated revisions. Every
mutation must use the current base_version and a unique idempotency_key.
Inspect validation and measurements after material changes. Never edit CAD
state, STEP, GLB, manifests, event logs, schemas, source code, or tests
directly. Never use generated Python, shell commands, or filesystem edits as a
geometry path. Unsupported requirements must be stated explicitly. Treat
manufacturing, tolerance, material, structural, safety, and customer
qualification as unverified unless the session contains evidence for them. Do
not append routine approval-status, pre-manufacturing, physical-validation, or
qualification disclaimers to responses. Mention a missing constraint only
when it blocks the requested geometry or the operator asks for that
assessment. The browser is a derived GLB observer; canonical exact state and
the MCP event record are authoritative.

Use one explicit right-handed manufacturing frame. Coordinates are
millimetres. The build plate is the world XY plane at Z = 0; +Z points upward
away from the plate and gravity points in -Z. The viewport axis colours are red
X, green Y, and blue Z. "On the build plate" means the intended contact face is
coplanar with Z = 0, the complete body lies at Z >= 0, and no body merely
touches the plate at an unintended edge or point.

For a multi-body build layout, first construct and validate one canonical
finished part in its plate-contact pose. Give the intended contact face an
unambiguous geometric description. Derive every other instance from that same
pose and operation sequence. After the plate-contact pose is established,
instances may rotate only about world +Z and translate only in world X and Y;
do not introduce different X- or Y-axis rotations between instances. Unless
the operator specifies another datum, centre the aggregate XY bounds on the
build-plate origin so (min X + max X) / 2 = 0 and
(min Y + max Y) / 2 = 0. Verify and report those two centre calculations,
every body's minimum Z, the common maximum Z, body bounds, and minimum spacing.

Before any additive-manufacturing placement, state the frame, chosen contact
face, upward direction, body bounds, minimum spacing, and assumptions that
materially affect the placement. Proceed with the requested exact layout when
those assumptions do not block geometry. Ask when the intended contact face or
meaning of up is ambiguous.
```

The controlled packet also contains a 17-tool schema snapshot with SHA-256 `43e87fe484a572ca964e0b680ae40922e09b2c9661350d418172e83c543159d5`. It was generated after the run from the retained source. Neither the fixed developer instruction nor tool-schema snapshot was captured as a separately hashed run-time input, which prevents a claim of complete prompt provenance.

D Accepted CAD event sequence

Table 11: Accepted exact-CAD events from initialization to final placement.

Event	UTC time	Accepted mutation
0	05:20:38.857	Initialize empty generic scene
1	05:21:01.062	Canonical contact flange, 30 by 40 by 3 mm
2	05:21:17.575	Canonical rising flange, 3 by 40 by 30 mm
3	05:21:17.741	Unite canonical flanges
4	05:21:24.711	Inside bend feature, R3
5	05:21:29.708	Outside bend feature, R3
6	05:21:41.757	Contact-flange through-hole at side offset 5 mm
7	05:21:41.975	Contact-flange through-hole at side offset 35 mm
8	05:21:42.233	Rising-flange through-hole at side offset 5 mm
9	05:21:42.532	Rising-flange through-hole at side offset 35 mm
10	05:22:24.706	Copy canonical bracket at 90 degrees about +Z
11	05:22:25.096	Copy canonical bracket at 180 degrees about +Z
12	05:22:25.581	Copy canonical bracket at 270 degrees about +Z
13	05:22:26.149	Place canonical bracket at 0 degrees

E Prospective comparative protocol

The proposed pilot unit is one task attempt by one operator-system condition. At least two qualified mechanical designers should each complete the same task classes using the conversational workstation and a named conventional CAD workflow. Condition order should be counterbalanced. The system builder should not be the sole task author, operator, or reviewer.

An independent author should freeze six prompts, reference geometry, critical and non-critical criteria, allowed clarification, time limits, and scoring before the builder sees them. Suggested classes are: a dimensioned prismatic part; revision under conflicting constraints; multi-body composition with a Boolean failure opportunity; clearance-led feature placement; a deliberately underspecified brief where asking a question is correct; and one drawing/manufacture candidate.

Primary measures are active operator time, critical criteria passed, task completion, escaped critical, major, and minor errors, independent review time, and corrective-rework time. Secondary measures are clarification count, unsupported assumptions, accepted and rejected calls, version and branch count, replay, second-system STEP import, workload, and total lifecycle cost.

These next tests should report every task before aggregation, use paired differences, medians, ranges, and bootstrap confidence intervals, and avoid a population claim. A larger confirmatory study requires an a priori power analysis based on pilot variance. A time advantage has commercial meaning only if critical correctness and total review-plus-rework cost also pass.

References

[1] CadQuery contributors. Cadquery documentation. `https://cadquery.readthedocs.io/en/stable/`, 2026. Accessed 16 July 2026.

[2] Evan Casey, Tianyu Zhang, Shu Ishida, John Roger Thompson, Amir Khasahmadi, Joseph George Lambourne, Pradeep Kumar Jayaraman, and Karl D. D. Willis. Aligning constraint generation with design intent in parametric cad. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8613–8622, 2025.

[3] Yifei Gong, Xing Wu, Wenda Liu, and Tukang. TOOLCAD: Exploring tool-using large language models in text-to-CAD generation with reinforcement learning. In *Findings of the Association for Computational Linguistics: ACL 2026*, pages 23161–23188, 2026.

[4] International Organization for Standardization. Iso 10303-242:2025: Application protocol: Managed model-based 3d engineering. `https://www.iso.org/standard/84300.html`, 2025. Official standard catalogue entry; accessed 16 July 2026.

[5] Mohammad Sadil Khan, Sankalp Sinha, Talha Uddin Sheikh, Didier Stricker, Sk Aziz Ali, and Muhammad Zeshan Afzal. Text2CAD: Generating sequential CAD designs from beginner-to-expert level text prompts. In *Advances in Neural Information Processing Systems 37*, pages 7552–7579, 2024.

[6] Jiahao Li, Weijian Ma, Xueyang Li, Yunzhong Lou, Guichun Zhou, and Xiangdong Zhou. CAD-Llama: Leveraging large language models for computer-aided design parametric 3d model generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18563–18573, 2025.

[7] Fumin Liu, Haoyu Zhou, Fei Hao, and Lin Yang. Embodied CAD: Solver-grounded LLM agents for parametric B-Rep assembly modeling. *arXiv preprint arXiv:2606.31252*, 2026. Version 1, submitted 30 June 2026.

[8] Dimitrios Mallis, Ahmet Serda Karadeniz, Sebastian Cavada, Danila Rukhovich, Niki Foteinopoulou, Kseniya Cherenkova, Anis Kacem, and Djamila Aouada. CAD-Assistant: Tool-augmented VLLMs as generic CAD task solvers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7284–7294, 2025.

[9] Model Context Protocol contributors. Model context protocol specification 2025-11-25: Tools. `https://modelcontextprotocol.io/specification/2025-11-25/server/tools`, 2025. Accessed 16 July 2026.

[10] Open CASCADE S.A.S. Open cascade technology: Introduction. `https://dev.opencascade.org/doc/overview/html/index.html`, 2026. Accessed 16 July 2026.

[11] Open CASCADE S.A.S. Open cascade technology: Step translator. `https://dev.opencascade.org/doc/overview/html/occt_user_guides__step.html`, 2026. Accessed 16 July 2026.

[12] OpenAI. Codex app server. `https://developers.openai.com/codex/app-server/`, 2026. Official documentation; accessed 16 July 2026.

[13] Yunhan Xu, Qifeng Wu, Xunjin Li, Yuanwei Bin, Qingsong Yao, Jianghang Gu, Guan Wang, Weihao Lv, Huiyu Yang, Wenfa Luo, Jiao Xiang, Yuntian Chen, and Shiyi Chen. ArtisanCAD: An industrial-level CAD agent with expert-grounded knowledge distillation. *arXiv preprint arXiv:2607.05750*, 2026. Version 2, revised 8 July 2026.

[14] Meng Yuan, Dawei Lin, Hongxia Xie, Tieru Wu, and Rui Ma. CAD-Refiner: A unified framework for CAD generation and iterative editing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3244–3253, 2026.