RESEARCH WHITEPAPER / REVISED OCTOBER 29, 2026

Printability of AI-generated 3D models.

Software geometry assessment, matched FDM printing, blind objective grading and human preference rankings.

> A collaboration with the University of Massachusetts Amherst's ADDFab facility.

3,999 full-run software outputs | 125 physical specimens | 3 objective graders | 17 volunteer rankers

Commissioned and funded by Meshy. Printers and consumables supplied by Meshy. Internal review draft.

1. Summary

This study asks whether AI-generated model geometry is suitable for physical 3D printing, whether a print-specific preparation workflow changes the printed result, and which physical results people prefer. The evidence is divided into four stages so that an estimated geometric property, a fabrication procedure, an observed defect and a preference ranking are not treated as interchangeable measures.

The software comparison contains 100 prompts and ten generations per automated workflow. It includes 1,000 Meshy 7 + Make Printable outputs, 1,000 Rodin outputs and 999 Tripo outputs, plus a separate 1,000-output raw Meshy baseline. The 3,999 workflow outputs include prepared versions of generated designs; they are not 3,999 independently generated designs.

The physical study comprises 25 selected prompts across five workflows: 125 parts. Three experienced 3D printing users assessed each part against a five-criterion rubric developed at UMass Amherst's ADDFab facility. Seventeen volunteers separately ranked physical print quality and satisfaction after seeing the prompt.

92%

23 of 25 Meshy 7 + Make Printable prints had no major missing features or holes in the selected FDM test. Rodin: 84% (21/25); Hi3D: 60% (15/25); Tripo: 56% (14/25). Other defects may remain.

Meshy 7 + Make Printable had the lowest observed mean defect total among the four compared workflows: 1.987 out of 15, compared with Rodin 2.253, Hi3D 3.400 and Tripo 3.480. The supplied paired analysis distinguishes it from Hi3D and Tripo on total defect scores, but not from Rodin. Recalculated four-workflow subjective mean ranks are 2.25 for print quality and 2.22 for satisfaction, best among the compared workflows.

These findings support Meshy 7 + Make Printable as a strong preparation workflow for the selected FDM challenge set. They do not establish a universal successful-print probability, a statistically demonstrated advantage over every competitor, or an advantage across all materials and printing technologies.

2. Study design and comparison groups

Four stages, different questions

1. **Generate and assess in software:** compare surface integrity and estimated thickness across many model outputs.
2. **Print under matched conditions:** produce a selected, challenging subset as physical FDM parts with anonymous labels.
3. **Grade visible printing defects:** use the lab's predefined categories while engine identities remain hidden.
4. **Rank the physical results:** ask volunteers about print quality first, then satisfaction after revealing the request.

Workflow definitions

Meshy 7 + Make Printable is the Meshy generator followed by the print-specific preparation path. **Meshy 7 (standard repair)** is the conventional-repair comparator in the physical and subjective study. It is labeled "Meshy 7" in the objective summary, which must not be confused with the **raw Meshy 7** arm in the large software run. The physical comparison therefore evaluates two preparation workflows, rather than raw versus repaired geometry.

Rodin: Gen-2.5-High. **Tripo:** v3.1-20260211, verified in the generation manifest. **Hi3D:** v3.0, confirmed by Meshy for this revision. Hi3D appears only in the selected physical set and preference study, not the large automated comparison. Its separate repair-path ledger is a software record, not a sixth physically graded workflow.

The software analysis averages the passing-seed percentage within each prompt, then averages across 100 prompts. Physical scores average graders within each part before comparing workflows paired by prompt. For consistency with the overview, the main subjective summary recalculates positions among the four compared workflows from each recorded five-part order. The original task contained five parts; no new four-part responses were collected.

> **Make Printable is Meshy's print-preparation feature.** It repairs mesh defects and reinforces thin features for the selected size and process. "Meshy 7 + Make Printable" means Meshy 7 generation followed by that preparation. Uploaded models can use it too.

---

3. Software geometry assessment

Inputs and preparation

The large run uses 100 prompts informed by the figures, objects and delicate details Meshy customers seek to print. Each automated workflow contributes ten seeds per prompt; one Tripo generation failed, leaving 999 Tripo outputs. The full 3,999-output ledger includes the raw Meshy baseline; the three-workflow comparison shown in the overview contains 2,999 outputs, rounded to 3,000.

Meshy 7 + Make Printable preparation in the large software run used printability-cli 0.2.2, serving build 719aae3e, with target longest dimension 100 mm, wall setting 1.1 mm and STL output. All 1,000 preparation jobs completed with one parameter set and no per-model tuning. The separate physical-set ledger records a 1.3 mm wall setting; the two runs used different preparation configurations.

What the checks mean

The strict surface-integrity result requires zero boundary edges and zero non-manifold edges. Boundary edges indicate an opening in the triangle surface; non-manifold edges indicate an ambiguous surface connection, such as an edge shared by more than two faces. A detected opening does not prove a slicer will fail: many slicers perform their own repairs. Conversely, a closed surface can still have delicate or impractical features.

Topology counts use exact float32 coordinate equality to weld shared vertices. The mesh-health counting implementation is deterministic. Thickness estimates use ray-based measurements and their own validity gates. The median estimated fragile surface is the percentage of valid thickness readings below 1 mm at a 100 mm reference size; this is a sampled geometric estimate, not an exact physical material-strength calculation.

The four-inch criterion requires the fifth-percentile valid thickness to reach 1 mm after scaling to the chosen size. In plain language, at least 95% of sampled readings must reach the guideline. Some thin regions can remain in a model that passes. Four inches is 101.6 mm; the preparation target and measurement reference are 100 mm. The separate “free of thin walls” scorecard row permits up to 10% estimated fragile surface and must not be described as literally zero thin features.

Large-run software results; prompt-averaged percentages.

Workflow	Surface integrity	Thickness at 2 in	Thickness at 3 in	Thickness at 4 in
Raw Meshy 7	33.4%	26.9%	40.9%	50.5%
Meshy 7 + Make Printable	98.8%	31.7%	56.8%	100.0%
Rodin	97.4%	43.3%	63.8%	81.9%
Tripo	53.5%	34.5%	49.7%	59.6%

Median estimated fragile surface at 100 mm is 0.033% for Meshy 7 + Make Printable, 1.107% for Rodin, 2.670% for Tripo and 5.059% for raw Meshy 7. These medians describe geometry, not observed failures. Rodin leads the tested two- and three-inch thickness criteria. The 100 mm-target preparation result cannot be generalized to very small models without a separate size-specific evaluation.

The metric historically labeled frac_4plus is a ray-reliability gate, not a direct self-intersection test. Layered garments or nested surfaces can trigger it without proving defective geometry. The “slicer-clean” composite is also a software proxy; it is not a logged slicer-import experiment. These distinctions prevent the analysis from claiming a full simulation of default FDM behavior. No thermal, extrusion, support-generation or toolpath simulation is established by these data.

4. Physical-set selection and printing

The locked physical set contains 25 prompts and one printed specimen per workflow and prompt, for 125 parts. It is a deliberately selected challenge set, not a random sample of the 100-prompt software population. The large-run software percentages and physical-set failure flags consequently have different denominators and should not be pooled.

Selection rule

The selection ledger prioritizes prompts for which both Tripo and Rodin exceed a four-inch estimated minimum printable size at the printed seed. A second tier includes prompts where one competitor exceeds that size, chosen for family and modality coverage. A third tier includes specificity cases for which all four arms meet the four-inch estimate.

The selection ledger records four prompts excluded because the body-removal rule damaged the prepared models. Seeds were zero except for the elven archer (008) and arched bridge (004), chosen by visual inspection. The ledger lists favorability ranks of 10 of 10 and 3 of 10, but does not define the rank direction. These exclusions and seed choices can favor the preparation method, so the physical results are scoped to this selected challenge set.

Fabrication and blinding

The physical workflow used six brand-new Bambu Lab A1 FDM printers supplied by Meshy, with the same printer settings and filament across workflows. UMass ADDFab Lab prepared and sliced the files, printed the parts and removed supports before evaluation. Engine identities were concealed during file preparation, printing and grading. Anonymous specimen IDs linked parts to their grading records; the assignment record used shuffle seed 20260914.

5. Blind objective grading

Three experienced 3D printing users independently scored each of the 125 anonymous parts using UMass ADDFab Lab's rubric. Five criteria are scored 0–3, with lower scores indicating fewer defects. The sum of the five part-level criterion means runs from 0 to 15. The completed workbook contains all 1,875 criterion scores: 125 parts × 3 graders × 5 criteria. Each part's criterion mean uses all three graders.

Defect categories

- **Layer shift:** no shift through obvious shifts distracting from the part's intended appearance.
- **Support removal:** no damage through a visible hole or a piece removed with the supports.
- **Stringing:** no strings through consistent stringing distracting from the overall appearance.
- **Drooping/sagging:** no visible deformation through deformation in more than one location.
- **Print completeness:** no noticeable issues through any missing element larger than 3 mm or any hole larger than 1 mm. Features successfully printed and subsequently removed with supports are scored under support removal rather than counted again as completeness loss.

▼ Read the lab's full 0–3 grading definitions

Layer Shift

- 0 - no layer shifts detected
- 1- detectable small shift(s) but difficult to see from 12" away
- 2-noticeable layer shift, but part still usable
- 3- obvious layer shift, distracts from the overall intent of the part.

Support Removal

- 0- no supports or no damage at all
- 1-slightly rough where supports were removed
- 2-rough surface from wall being removed, or supports remaining, but in a hidden area, or not distracting from 12" away
- 3 - visible hole in the part where the supports ripped away the outer wall. Or, supports removed a piece of the part.

Stringing

- 0-no stringing
- 1-very thin or infrequent stringing
- 2-one or two areas that have occasional strings attached
- 3-consistent stringing that distracts from the overall appearance of the part

Drooping/Sagging

- 0-no visible drooping or sagging
- 1-very subtle drooping or sagging in one or two locations.
- 2-small deformation in only one location
- 3- part appears deformed in more than one location

Print Completeness

- 0-no noticeable issues
- 1-one small element missing or one hole <1mm
- 2-two small elements missing or up to three holes <1mm
- 3-any elements larger than 3mm missing, and any holes larger than 1mm in diameter
- *if it successfully printed but supports removed it, then 0

Definition of a severe completeness flag

A part is flagged if **any grader** assigns 3 on Print Completeness. This is the supplied analysis's operational "print failure" definition. It does not mean the printer aborted, and a part without the flag is not necessarily free of defects. A majority-vote definition could produce different counts; the overview consistently uses the any-grader rule.

Mean defect scores across 25 parts per workflow; lower is better.

Workflow	All five criteria / 15	Excluding support removal / 12	Severe incompleteness
Meshy 7 + Make Printable	1.987	0.920	2/25 (8%)
Rodin	2.253	1.107	4/25 (16%)
Hi3D	3.400	2.067	10/25 (40%)
Tripo	3.480	2.093	11/25 (44%)
Meshy 7 (standard repair)	3.813	2.227	11/25 (44%)

Workflow	Layer shift	Support removal	Stringing	Drooping / sagging	Print completeness
Meshy 7 + Make Printable	0.19	1.07	0.09	0.45	0.19
Rodin	0.11	1.15	0.23	0.33	0.44
Hi3D	0.28	1.33	0.40	0.59	0.80
Tripo	0.36	1.39	0.41	0.60	0.72
Meshy 7 (standard repair)	0.28	1.59	0.49	0.49	0.96

Cell numbers are mean scores from 0 (no defect) to 3 (severe). Darker amber indicates more defects. Averages are computed within each part before averaging the 25 prompts.

Paired statistical interpretation

The supplied analysis compares the 25 prompt-paired workflow scores and applies Holm correction across the four comparisons against Meshy 7 + Make Printable. For the five-criterion total, adjusted p-values are 0.00674 against standard-repair Meshy 7, 0.00674 against Hi3D, 0.00430 against Tripo and 0.90351 against Rodin. These are comparisons of defect scores, not significance tests for the binary severe-flag percentages. The source reports Friedman p = 0.0000723 for the overall workflow comparison.

Without support removal, the workflow ordering remains the same: Meshy 7 + Make Printable 0.920, Rodin 1.107, Hi3D 2.067, Tripo 2.093 and standard-repair Meshy 7 2.227 out of 12. The supplied Holm-adjusted comparisons are 0.02470 against the Meshy baseline, 0.00909 against Hi3D, 0.00244 against Tripo and 0.48902 against Rodin. Support removal is a large component of total defects and depends on fabrication choices, so this sensitivity analysis matters.

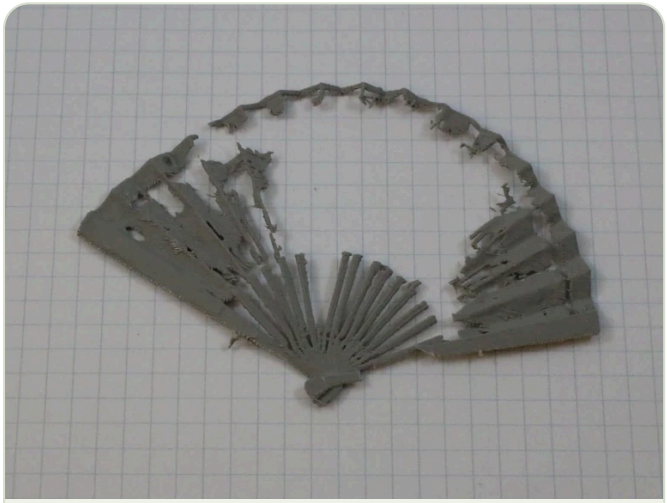
On completeness scores alone, Meshy 7 + Make Printable averages 0.187 out of 3 after averaging within each part. Rodin averages 0.440, Hi3D 0.800, Tripo 0.720 and the standard-repair Meshy baseline 0.960. Rodin has lower observed mean layer-shift and drooping scores than Meshy 7 + Make Printable; Meshy 7 + Make Printable does not lead every individual criterion. The supplied grader-agreement estimate for totals is ICC 0.596, with pairwise Spearman correlations 0.790, 0.752 and 0.764.

The objective p-values are reported from the supplied analysis; descriptive counts and means were independently reconciled to the grading workbook. Exact paired-test options and the ICC convention are not specified in the supplied record. With only 25 selected prompts, failure to distinguish Meshy 7 + Make Printable from Rodin does not establish equivalence.

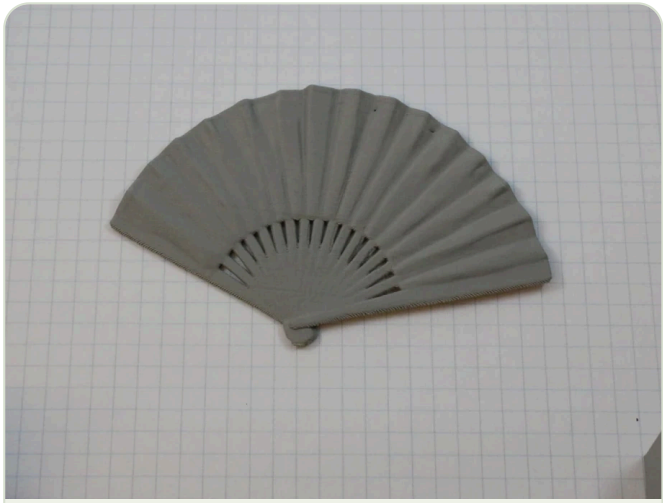
---

6. Matched physical examples

The numbered photographs link each specimen to its objective grading record. The open-hand-fan example compares conventional repair of Meshy 7 output with Make Printable preparation for the same prompt. The conventional-repair specimen had a maximum completeness score of 3 and mean total defect score of 8.00; the Make Printable specimen had a maximum completeness score of 1 and mean total of 1.00.



Meshy 7 (standard repair): severe completeness flag.

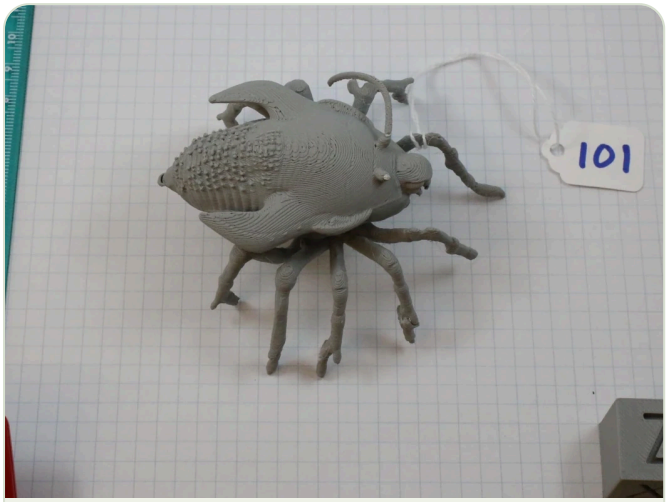


Meshy 7 + Make Printable: no severe completeness flag.

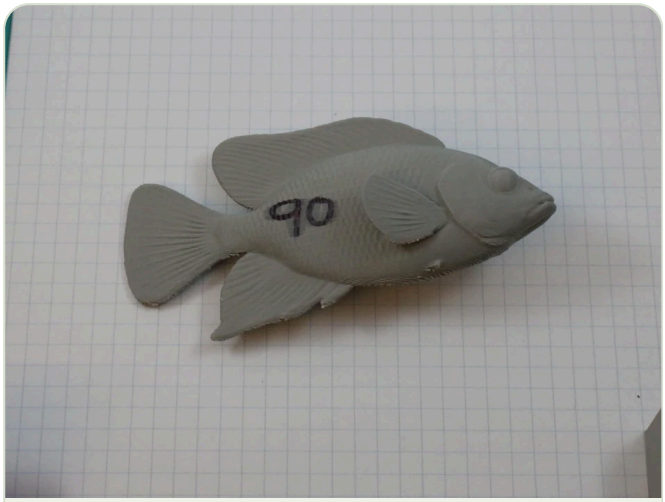
This is an illustrative selected pair, not an additional statistical sample. The photograph alone is not the scoring rule; its grade comes from the three blinded records. A maximum score of 1 also makes clear that “no severe flag” does not mean perfect.

Two Meshy 7 + Make Printable misses

The six-legged insect and long-finned tropical fish still received severe completeness flags, despite software fragile-surface estimates of 0.092% and 0.010%. Their mean total defect scores were 5.00 and 3.00. These cases demonstrate why a favorable geometry estimate cannot guarantee a successful physical result.



Six-legged insect: severe completeness flag.



Long-finned tropical fish: severe completeness flag.

7. Relationship between software and physical outcome

The supplied cross-method ledger links estimated fragile surface, estimated minimum printable size and physical scores for all 125 parts. It reports positive Spearman associations between fragile surface and mean completeness defect ($\rho = 0.532$), and between fragile surface and total defects ($\rho = 0.564$). Estimated minimum printable size is associated with completeness defects ($\rho = 0.573$) and total defects ($\rho = 0.604$).

When parts are sorted into four equal-sized groups by estimated fragility, severe completeness flags occur in 2/31, 5/31, 9/31 and 22/32 parts: 6.5%, 16.1%, 29.0% and 68.8%. This monotonic pattern is consistent with thin geometry identifying a meaningful source of printing risk.

This is an exploratory association within the same selected set. It is not held-out validation of a calibrated failure predictor. The 125 observations include five workflows per prompt and are not independent draws from a general population. Shared prompt difficulty and workflow differences can contribute to the pooled correlation; nominal per-part p-values should not be interpreted as cluster-adjusted evidence. A prompt-aware analysis or held-out experiment would be needed for a broader predictive claim.

ADDFab graded physical parts; it did not audit or endorse Meshy's software measurement method. The two near-zero-fragility failures are direct counterexamples to treating the geometry measure as a guarantee.

8. Blind subjective rankings

Seventeen anonymous volunteers contributed 413 rankings per question. Sixteen completed all 25 prompt sets; one completed 13. Each set contained five printed parts with engine identities hidden.

At each station, five printed specimens from the same prompt were presented with engine identities hidden. Question 1 asked graders to rank print quality before seeing the prompt; the answer was locked before Question 2 revealed the request. Question 2 asked which result they would be happiest with, including both the match to the requested subject and physical quality.

Meshy 7 + Make Printable

2.25
Print quality

2.22
Satisfaction given the prompt

Rodin

2.43
Print quality

2.66
Satisfaction given the prompt

Hi3D

2.59
Print quality

2.39
Satisfaction given the prompt

Tripo

2.74
Print quality

2.73
Satisfaction given the prompt

Average rank among four compared workflows: 1 best, 4 worst. These positions are recalculated within each of the 413 original five-part orders per question after excluding the standard Meshy 7 reference. They are not new four-part responses. The original study ranks are retained separately in the study record.

Ranks within an order are dependent, and each volunteer rated multiple prompts. The four-workflow averages above are descriptive summaries. Statistical inference must account for repeated ratings by the same volunteer; the earlier supplied bootstrap analysis used a 16-participant dataset and is not presented as an uncertainty analysis for these 17-participant results.

Meshy 7 + Make Printable had the best average rank among the four compared workflows on both questions, with modest differences from the nearest alternatives. An average rank describes relative position, not a quality rating, satisfaction percentage or first-choice share. The original five-part means for this workflow were 2.51 on both questions; recalculating positions among four workflows gives 2.25 and 2.22.

The standard-repair Meshy workflow was present in the original five-part questions even where its result is omitted from the four-workflow marketing comparison. The prompts test both subject fidelity and physical print quality; preference does not isolate the causal effect of wall reinforcement.

Original five-part rankings

The original task included the standard-repair Meshy 7 reference. The table below preserves that full five-part result; its scale is 1 best to 5 worst. The four-workflow summary above matches the overview and uses a different, explicitly recalculated scale.

Workflow	Print quality	Satisfaction
Meshy 7 + Make Printable	2.51	2.51
Rodin	2.77	3.08
Hi3D	2.97	2.77
Tripo	3.13	3.20
Meshy 7 (standard repair)	3.63	3.43

413 orders per question, 17 volunteers. Each mean uses the original position recorded in the five-part task.

9. Limitations and interpretation

Generalization and selection

Meshy 7 + Make Printable had the highest observed completeness rate and lowest average defect score among the four compared workflows in the selected physical set. The challenge-set selection, body-removal exclusions and visually selected seeds limit population-level generalization. A random sample and repeated prints per prompt would give a stronger estimate of everyday failure risk.

There is one physical print for each prompt–workflow combination. Three graders provide independent assessments of that specimen, not three fabrication replicates. Printer, orientation, support strategy and material can affect a defect independently of the generated geometry. Physical results apply to the recorded FDM setup, not automatically to SLA, Full-Color or every print size.

Comparators and software proxies

The large software run and physical set have different workflows and preparation ledgers. Hi3D's unavailable automated repair API limits symmetry with the automated engines. Raw Meshy, conventional repair and Meshy 7 + Make Printable must remain distinct. Closed geometry, a thickness pass and good printed appearance each answer different questions.

Neither the ray-reliability gate nor the slicer-clean composite is a direct measurement of self-intersection or slicer import. Exact vertex welding and sampling validity conditions affect how geometry is counted. No claim is made that an FDM printer was simulated in all its material, thermal or support behavior.

Uncertainty and sponsor involvement

Rodin is a close physical comparator and cannot be declared inferior on the supplied paired-score analysis. Lower observed values are not the same as statistically established superiority, and non-significance does not establish equality. The subjective sample contains 17 volunteers. Exploratory cross-method correlations do not constitute independently calibrated prediction.

Meshy commissioned and funded the project, supplied six printers and consumables, generated and prepared the digital geometry and performed the software analysis. UMass ADDFab Lab sliced, printed, removed supports and graded the physical parts with engine identities concealed. Anonymous evaluation reduces identity-related bias, while sponsor involvement and challenge-set selection remain relevant to interpretation.

10. Responsibilities and disclosures

University of Massachusetts Amherst

About the collaboration

The Advanced Digital Design and Fabrication facility (ADDFab) at the University of Massachusetts Amherst performed file preparation and slicing, printing, support removal and objective grading. UMass ADDFab Lab developed the objective grading rubric and recruited volunteers for the subjective evaluation. Engine identities remained hidden from lab staff during file preparation, printing and grading; anonymous specimen IDs linked parts to their records. Meshy generated the models, prepared the digital geometry and performed the software analysis.

Meshy commissioned and funded the study and supplied six Bambu Lab A1 printers and consumables.

Study funding and responsibilities are disclosed above. UMass ADDFab Lab's participation describes its work on this study; it does not constitute product endorsement or an audit of Meshy's software analysis.

11. Source record and reproducibility

1. **Large-run scorecard:** scorecard_full.json and scorecard_1000_table.csv from the October 26 handoff. Used for the 100-prompt automated comparison.
2. **Measurement definitions:** MESH_HEALTH_MEASUREMENT_README.md. Governs topology, valid thickness readings, size thresholds and ray-reliability interpretation.
3. **Selected-set record:** _printset_locked25.json, _ps25_metrics.json and scorecard_v8.json. Governs the challenge-set selection and recorded printed seeds.
4. **Objective grades:** Meshy Objective Grading – Shared (1).xlsx, supplied October 29: 125 parts, three graders and five criteria, with all 1,875 scores populated. The revised export fills Grader 1's UM-020 print-completeness score with 0. The other two graders also recorded 0, so part-level averages and workflow results are unchanged.
5. **Objective aggregate analysis:** objective_grading_by_part.csv and objective_grading_summary.json. Descriptive means independently recomputed; inferential statistics retained from the supplied analysis with the implementation caveat in Section 5.
6. **Cross-method ledger:** crossmethod_part_level.json. Exploratory relationships between software estimates and physical outcomes.
7. **Subjective submissions:** the supplied workbooks (3) and (4) contain the same 17-participant response set: 413 rankings per question. Four-workflow means are recalculated from recorded orders; older 16-participant aggregates are not used for these descriptive results.
8. **Physical photographs:** Numbered Photos archive delivered October 26. Example photos joined to objective grading by anonymous specimen number. Public review assets omit participant records and the complete blind assignment key.

The study record comprises prompts and seed assignments, generation and preparation configurations, fabrication procedures, anonymous specimen mappings and grading data. The public review package omits participant-level records and the complete assignment key.
