RN-002 — Hand Family Prompt Comparison

Astract. This note examines whether constrained prompt phrasing improves the structural reliability of diffusion-generated hands. A locked hand-family set was reviewed under a three-class rubric: A (believable / structurally acceptable), B (flawed but still believable), and C (failure). Seven prompt conditions were compared using one-image-at-a-time manual review, along with supporting fields for visible digit count and recurring failure signatures. The central result is modest but clear: constrained prompting improves outcomes, yet no reviewed condition solved hand anatomy reliably. More importantly, different prompt conditions improved different aspects of performance. Pose cues reduced outright failure most effectively, semantic-styling cues increased the rate of clearly acceptable hands, numeric wording helped only partially, and semantically ambiguous phrasing introduced scene contamination.

Research question

Can small prompt changes meaningfully reduce structural failure in generated hands, and do different prompt conditions improve different aspects of performance rather than one universal notion of quality?

Introduction and framing

Human hands remain one of the most recognizable and failure-sensitive structures in diffusion-generated imagery. Small errors in digit count, articulation, thumb placement, or surface logic are easy to notice and difficult to ignore. This makes hands a useful test case for studying the gap between visual plausibility and structural correctness in generative image systems.

RN-002 extends Driftline's broader research program into structural reliability under controlled prompt variation. Earlier Driftline work on chairs showed that generated objects can appear convincing at first glance while still failing basic structural checks under closer inspection because of missing supports, impossible geometry, or fused elements. Hands present a more anatomically sensitive version of the same problem: they are familiar to viewers, compositionally common in image generation, and highly vulnerable to both local and global structural breakdown.

The central aim of RN-002 is not to ask whether a generated hand looks attractive, photorealistic, or stylistically polished. The aim is to ask whether the hand holds together as a believable structure, and how that answer changes under different prompt conditions. A hand can look glossy, cinematic, or professionally lit and still fail if the digit count is wrong, the thumb is duplicated, the fingers are fused, the articulation is implausible, or the anatomical orientation contradicts itself. RN-002 therefore treats hands as a structural evaluation problem rather than a beauty contest.

A key premise of this note is that different prompt constraints may improve different aspects of performance. A pose cue may reduce catastrophic failure without producing many truly strong hands. A styling cue may improve realism without stabilizing anatomy. An explicit numeric phrase may improve digit count without resolving articulation. A semantically ambiguous word may generate the intended subject while also pulling unwanted visual material into the image. These are all distinct outcomes, and they should not be collapsed into a single vague sense of “better.”

RN-002 should therefore be read as an observational research note rather than a final benchmark paper. The goal is to establish a clear comparative framework, identify recurring structural failure patterns, and test whether a lightweight evaluator rubric can reveal meaningful differences across prompt conditions. Within that framing, the hand-family set serves three purposes: it provides a difficult but intuitive object class for structural testing; it demonstrates that improvements in generation are not one-dimensional; and it serves as a practical proving ground for Driftline's evaluator logic.

Methods and review workflow

RN-002 evaluates structural correctness in diffusion-generated human hands under controlled prompt variation. The goal is not to measure style preference or photorealistic quality alone, but to compare how different prompt conditions

affect hand structure, visible digit count, and recurring anatomical failure patterns.

The reviewed hand-family set includes the following prompt conditions: *hand*, *hand isolated*, *back of hand*, *human hand, five fingers*, *hand model*, *professional hand model*, and *palm of hand*. Images were generated in batch form and then reviewed manually using the Driftline local reviewer. Each image was scored one at a time rather than through contact-sheet review. This reduced oversight, improved consistency, and allowed closer inspection of finger count, articulation, thumb structure, and orientation logic.

The reviewer interface recorded a primary A/B/C rating together with supporting fields for visible digit count, fused digits, multi-hand contamination, pinky drift, and freeform notes. Photorealism was not required for an A score; conversely, polished-looking images could still fail structurally. Early provisional D values in the weakest baseline set were folded into the locked C category before final comparison.

Locked first-pass rubric

A — structurally acceptable. The hand reads as believable and structurally acceptable. Typical characteristics include five visible digits, plausible thumb placement, plausible finger spacing, believable articulation, and no major structural contradiction. An A-rated hand does not need to be perfect or photorealistic, but it must look like a hand could actually exist in the pose shown.

B — flawed but still believable. The hand remains broadly readable as a hand, but contains visible issues that reduce confidence or usability. Typical examples include pinky drift, awkward spread between fingers, minor articulation issues, or local anatomical oddities that do not completely break the image.

C — failure. The hand is structurally broken or implausible enough to be treated as a failure. Typical triggers include wrong visible digit count, fused digits that break the silhouette, duplicated thumbs, ghost digits, multi-hand contamination, severe articulation failure, palm-side fingernails, implausible finger-to-palm proportions, or other contradictions that make the hand clearly unusable.

Results

Table 1. Hand-family outcome comparison

Prompt	N	A	B	C	A%	C%
hand	210	56	40	114	26.7	54.3
hand isolated	200	55	61	84	27.5	42.0
back of hand	105	36	37	32	34.3	30.5
human hand, five fingers	100	33	31	36	33.0	36.0
hand model	100	30	31	39	30.0	39.0
professional hand model	100	38	28	34	38.0	34.0
palm of hand	100	23	37	40	23.0	40.0

The strongest pose-based branch was *back of hand*, which produced the lowest failure rate in the reviewed set. The strongest semantic-styling branch was *professional hand model*, which produced the highest A rate. The unconstrained baseline *hand* remained the weakest overall condition, with a majority of outputs scoring as failures.

Table 2. Hand-family failure-signature comparison

Prompt	5-digit %	Pinky drift	Fused digits	Multi-hand
hand	57.1	41	25	17
hand isolated	59.5	53	19	11
back of hand	69.5	40	0	0
human hand, five fingers	64.0	39	4	0
hand model	61.0	35	0	0
professional hand model	66.0	38	4	3
palm of hand	60.0	34	1	0

Failure signatures varied by prompt condition. *Back of hand* was the cleanest branch on fused digits and multi-hand contamination, reinforcing the value of strong pose priors for suppressing outright collapse. Across all branches, however, visible five-digit counts were insufficient as a standalone measure because anatomically implausible hands could still present five visible digits.

Interpretation and discussion

Several distinct improvement axes appeared. Isolation reduced some catastrophic collapse but mostly shifted outcomes into borderline middle cases. Numeric wording helped, but was weaker than a strong pose prior. Semantic-styling cues improved the rate of clearly acceptable hands, while semantic ambiguity introduced scene contamination in the *palm of hand* branch.

The unconstrained *hand* baseline established the weakest condition in the family set, supporting the broader Driftline position that diffusion systems can produce visually plausible imagery while remaining structurally unreliable at the object level. Recurrent problems included pinky drift, fused digits, multi-hand contamination, duplicated thumbs, ghost structures, palm-side fingernails, and implausible finger-to-palm proportions.

The strongest branch for failure suppression was *back of hand*. One plausible interpretation is that canonical pose and framing cues provide a stronger structural prior than generic semantic refinement. Even in this stronger condition, however, pinky drift remained common, which suggests that local finger articulation remains unstable even when global hand pose improves.

The semantic styling branches reveal a second kind of improvement. *Hand model* modestly improved performance relative to the baseline, while *professional hand model* produced the highest rate of clearly acceptable outputs. This suggests that stronger semantic cues can improve the presentation and realism of the hand without reducing failure as effectively as a pose prior. The *human hand, five fingers* branch indicates that explicit numeric wording helps, but only partially: count guidance is not equivalent to structural control.

These results also support the need for a dedicated evaluator. Several reviewed images contained five visible digits and still failed because of duplicated thumbs, ghost digits, palm-side fingernails, or implausible articulation. Digit counting alone is therefore insufficient as a scoring method. A useful evaluator must account for both count-level structure and higher-order anatomical logic. RN-002 demonstrates that even a simple first-pass rubric can reveal distinctions that would be missed by counting alone or by casual visual inspection.

Limitations and follow-up controls

RN-002 is an observational research note rather than a final benchmark study. The rubric is intentionally lightweight, the reviewed branches are finite, and not every failure subtype was promoted to its own explicit field.

The current note also leaves several targeted controls for later validation, including *human hand*, *hand, five fingers*, *open palm hand*, *palm of a hand*, and *back of a hand*. These follow-up controls may help isolate whether the strongest current branch remains stable under small language changes.

Recommended next steps

- Keep the current hand-family set locked as the main comparative result group for RN-002.
- Tighten the manuscript and normalize wording so the results, methods, and rubric all use the same locked language.
- Run follow-up controls only when they clarify a specific ambiguity already identified in the note.
- Treat the evaluator implication as a research-supported extension, not as promotional product copy.

Locked conclusion

Diffusion-generated hands improve under constrained prompting, but no reviewed condition solved hand anatomy reliably. Pose cues were best at reducing outright failure; semantic-styling cues were best at increasing clearly acceptable hands; numeric wording helped only partially; and digit count alone was insufficient to judge correctness. That gap between looking better and being more structurally correct is precisely the kind of gap a structured evaluation workflow is meant to measure.