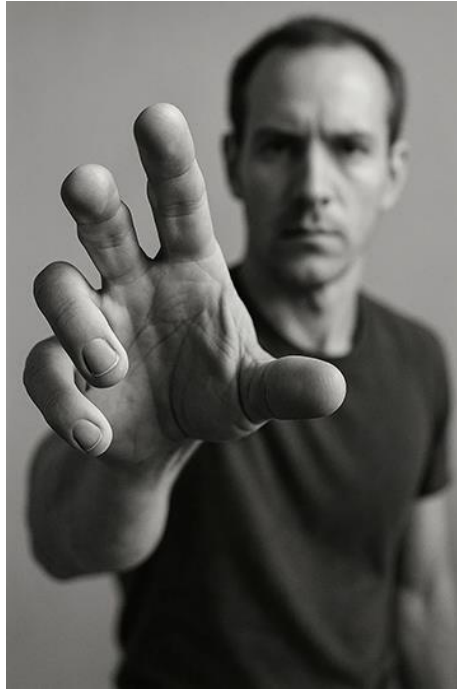


Foreshortening Recipe Book

A methodology for controlled foreshortening and continuity



A diagnostic for spatial reasoning under causal constraint.

Executive Summary

Foreshortening Recipe Book, Vol. 1 for Sora provides a set of structured prompt “capsules” that test and refine how image-generation systems construct depth, proportion and spatial order. Traditional AI imagery often treats figures and objects as flat assemblies; this framework reintroduces **causal sequencing** and a near-to-far grammar that forces the model to reason through space rather than decorate it.

Each recipe establishes:

- **Ordered stages** (e.g., limbs, planes, or surfaces from foreground to distance) that act as soft scene graphs inside a text prompt.
- **Lighting and contact constraints** that anchor geometry to a coherent vanishing system.
- **Gates and bans** that prevent token drift, montage behavior, or ungrounded cloning.

The collection functions as:

1. **A creative field manual** for artists exploring depth and perspective through constraint.
2. **A diagnostic suite** for researchers analyzing model “causal reach” and spatial reasoning limits (typically 6 – 8 ordered layers before collapse).
3. **A bridge between painterly literacy and machine logic**, translating the perceptual discipline of traditional foreshortening into reproducible, testable language.

The result is a framework that measures not polish but **consequence**: how well an engine can sustain the illusion of embodied space under constraint. In essence, this book reframes foreshortening as a test of **in depth**. When a model falters, it exposes where the delta is, so that potentially, authorship can begin.

© 2025 Russell Parrish / A.r.t.i.s.t I.n.f.l.u.e.n.c.e.r.

All rights reserved. No part of this system, visual material, or accompanying documents may be reproduced, distributed, or transmitted in any form or by any means, including AI training datasets, without explicit written permission from the creator. A.r.t.i.s.t I.n.f.l.u.e.n.c.e.r and all associated frameworks, critique systems, and visual outputs are protected as original intellectual property.

Introduction

Foreshortening is challenging for both humans and AI alike. In both fields, it is one of the most reliable ways to expose how well an artist or image model understands **depth, proportion, and spatial order**.

Humans struggle to find the right “shapes” that define objects that recede, but do it intuitively without description as we know a foot near the lens must expand while the torso recedes, but most AI systems still treat bodies as 2D assemblies of learned silhouettes. When asked for extreme perspective, they often misfire as they dynamically assemble from training data: stretching limbs inconsistently, flattening torsos, or confusing what overlaps what.

This is a **quick recipe book** for Sora designed to test and guide that behavior for AI. Leaning into the use of ordered descriptions, explicit light cues, and floor-plane anchors to help the model maintain depth.

They're not guaranteed fixes and **results can be unpredictable**, but each prompt pressures the model's causal sense of space rather than just its stylistic memory.

Frame of Mind:

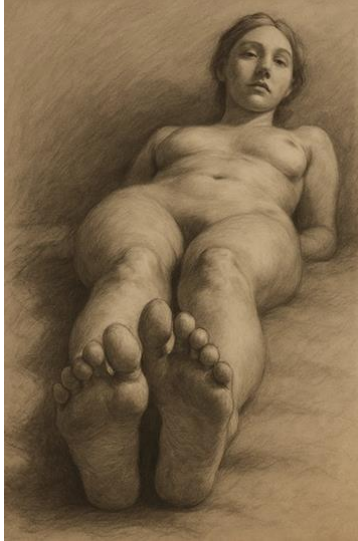
Most AI imagery fails at this because it builds scenes as a single flattened “snapshot.” The formulas here introduce time into spatial reading. It's no longer what is depicted, but how it is revealed. That's a major unlock for any generative system trying to simulate intentional foreshortening, spatial hierarchy, or visual storytelling.



Think about it as the description is led through the space of the body, foot to leg to torso to head, like a traveling the spine of a figure. Each successive part builds on the previous one's scale and overlap cues. The image stops being static and starts behaving like a process and linear subject. This unlocks an underused tool: progressive anchoring. Each stage gives the next something to push against, so by the time we reach the head, the image has “earned” its depth.

To use a spine; it works across subjects (be it a figure, tree, pencil, and beyond).

- Frame header “Single still frame. No cuts. Naturalistic lens. Realistic depth cues. Respect ordinal sequence.”
- Camera & DOF locks “Foreground anchor in the lower third. Shallow→moderate DOF that deepens with distance. Occlusion > floating.”
- Ordered stages (numbered)
- “(1) [Nearest] ...”
- “(2) ... continues ...”
- “(3) ... extends ...”
- “(4) ... emerges ...”
- “(5) [Farthest] ... reveal ...”
- Gates “Do not show (2) unless (1) is established,” etc. “Final stage must be smallest by scale and highest by luminance/atmospheric perspective.” Drift guards “Not a montage. No symbolic inserts. No style swaps mid-frame.”



What's happening

Ordinal scaffolds (“first... then... finally...”) create soft must-haves the model allocates attention to in order. Part→whole grammars bias depth mapping: early tokens often become foreground anchors, later tokens drift to mid/background or “reveal” elements. Continuity verbs (“grow, extend, taper, emerge, reveal”) encourage monotonic transitions the model renders as foreshortening, scale decay, or occlusion. Make it reliable (prompt spec)

Primary examples:

Foreshortened Figure: Sequential Depth Capsule



“A single continuous shot, no cuts. Naturalistic studio lighting, 50–70 mm lens feel. Render the subject as a full-body figure reclining toward the viewer, built step by step:

- (1) Foreground: the soles and toes are nearest, dominating the lower third of frame.
- (2) Continue back through shins and knees, preserving strong scale decay and overlap.
- (3) Reveal thighs and pelvis, grounded with cast shadow contact.
- (4) Show ribcage and abdomen, compressed in perspective.
- (5) Add shoulders and neck, tapering along the depth axis.
- (6) Finally, resolve the head farthest from camera — smallest by visual angle, sharp yet least saturated. Do not render later stages unless prior ones are correctly established. No montage, no disjointed limbs, no symbolic insertions.”

This is exploiting token-chaining priority: the order and gating words in the prompt set up a causal scaffold the engine tries to satisfy in sequence. That sequence then gets mapped to spatial/temporal structure (near→far, part→whole, detail→field).

Tree Canopy: Vertical Sequence Capsule



"A single continuous shot looking upward, 24–28 mm wide-angle feel. Build the composition in vertical depth order:

- (1) Start with the tree trunk closest to camera, bark texture crisp and detailed.
- (2) Extend into primary thick branches splitting from the trunk with strong occlusion.
- (3) Continue into secondary thinner branches, creating a radial lattice.
- (4) Add a semi-translucent leaf canopy forming an overhead plane.
- (5) Finally, reveal a small sunburst peeking through the gaps, flare controlled and leaf detail intact.

No sun until the canopy exists. Flare must not overpower detail. All stages should appear continuous and physically connected."

This sequence mimics real-world embodied experience (look up → branch → canopy → sky), so the composition feels natural even if the image is highly constructed.

Pencil on Desk: Linear Progression Capsule



"A single static shot, shallow depth of field, natural studio light. Sequence the object from nearest to farthest:

- (1) Foreground: the graphite tip in sharp focus, bevels and chips visible.

- (2) Progress into the sharpened wooden cone, clean helical cut.
 - (3) Extend along the yellow hexagonal barrel, receding with proper perspective lines aligned to desk grain.
 - (4) Show the metal ferrule with knurling and specular reflections.
 - (5) End with the pink eraser, farthest and slightly softer.
- Barrel must read continuous. Ferrule cannot appear before barrel. Only one pencil — no duplicates or clones.”

The pencil sequence is the microcosm of the same principle. From tip to eraser, the eye experiences scale decay and material contrast in order — sharpness → length → metal → softness. That small journey builds object intimacy and weight. It also sets up a compositional payoff: the pencil isn’t just an object anymore, it becomes a path the viewer travels along.

Choreography of the Image

This isn’t about perception choreography; it’s about model token sequencing under causal load. It is pressuring the generative process, not the viewer’s eye.

Optional hardening (when the model drifts) Add numerical weights:

- “Stage (1) priority 1.0 ...
- Stage (6) priority 0.3.”
- Add causal connectors:
- “only after,” “once established,” “subsequently.”
- Add failure bans: “no duplicates, no cutaways, no montage, no object cloning.”

For video: “one continuous shot; no zoom cuts; camera static; change arises from composition, not edit.”

Foreshortening Win Recipes → Easy → to Spine → Environments

Below are easy foreshortening prompts, which can be taken as is, or built upon, which will be shown later.

The Recline (Classic)

Use: Grounded figure depth / human proportion test

Prompt Skeleton:

“Low front camera. Nearest foot or knee dominates lower frame.

Body reclines away, each joint overlapping the next.

Light from upper-left, floor plane visible.

No motion blur, no floating limbs.”

Checks: scale decay, occlusion order, tonal depth.



The Curl

Use: Compact, looping geometry / circular compression

Prompt Skeleton:

"Top-down view. Figure curled inward; nearest limb overlaps torso.
Body forms a closed loop; far side softened by haze.
Single diffuse light from above."

Checks: internal overlap correctness, self-occlusion continuity.



The Reach

Use: Dynamic projection / single-limb emphasis

Prompt Skeleton:

"Single still frame. One hand or arm reaching toward camera, fingers sharp.

Forearm, shoulder, and face recede along one line.

Depth shown by focus falloff and scale, not blur streaks."

Checks: limb proportionality, edge compression.



The Crawl

Use: Multi-joint chain with opposing diagonals

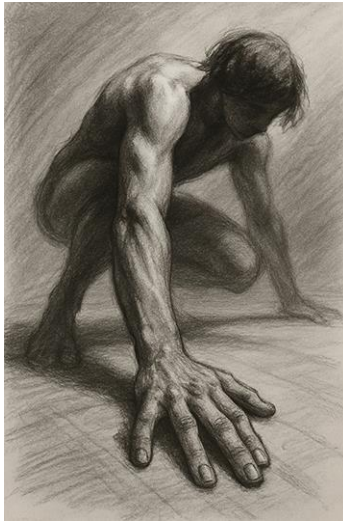
Prompt Skeleton:

"Low-angle view of crouched figure on hands and feet.

Nearest hand and knee crisp, far limbs fade into tone.

Diagonals cross; ground plane consistent."

Checks: 4-point contact coherence, torque alignment.



The Stack (Objects)

Use: Repetition / depth rhythm

Prompt Skeleton:

“Series of identical solid objects (boxes, books, chairs).

Foremost fully visible, each behind smaller, same axis.

Floor shadows aligned, no floating.

6–9 units, last merges with haze.”

Checks: spacing regularity, VP accuracy, haze gradient.



The Coil (Organic)

Use: Nonlinear depth path / continuity under twist

Prompt Skeleton:

“Snake-like or rope-like form looping toward camera.

Overlaps alternate light/dark; contact shadows visible.

Haze increases along coil.”

Checks: directional memory, curvature compression.



The Tandem (Dual Figures)

Use: Comparative foreshortening

Prompt Skeleton:

"Two figures on same axis.

Foreground cropped; background full-body.

Shared light; no tangent merges."

Checks: relative scale, spatial separation, light coherence.



Variants for Every Capsule

- **Reverse:** describe from far→near to test collapse.
- **Baseline:** remove order terms ("a line of...").
- **Anti-Order:** demand equal clarity for all parts.
- **Haze/Shadow Ladder Add-On:** enforce measurable depth cues.

Spine Capsules

Reclining Figure: Classic Perspective Capsule

Prompt Skeleton:

"A single still frame, neutral studio light, 50 mm–70 mm lens feel.

- (1) Foreground: nearest foot dominates lower third, sharp texture and edge.
 - (2) Shins and knees overlap the foot; scale decay clear.
 - (3) Thighs and pelvis continue backward; cast shadow anchors to floor.
 - (4) Abdomen compresses in perspective; lighting consistent.
 - (5) Chest and shoulders taper; DOF softens slightly.
 - (6) Farthest: head small by angle, least saturation, gaze off-frame.
- Gates: all limbs connected, no floating elements, single VP."



Crouched Figure: Diagonal Compression

Prompt Skeleton:

"Single still frame, 35 mm lens feel.

- (1) Foremost: bent knee and hand planted near camera, in crisp focus.
 - (2) Torso twists back, forming a diagonal path toward far shoulder.
 - (3) Far knee and foot partly hidden, scale reduced.
 - (4) Face and upper torso align to same perspective; light continuous.
- DOF shallow near knee, deeper toward head; shadows unify across floor."



Seated Figure, Low View

Prompt Skeleton:

"Low camera angle at ground height.

- (1) Nearest foot fills bottom edge.
- (2) Legs fold toward body; overlap gradual.

- (3) Torso upright; arms resting on knees.
(4) Head and shoulders farthest, compressed in perspective.
Consistent 3-point lighting; no background clutter; depth from body geometry only.”



Running Motion (Freeze-Frame)

Prompt Skeleton:

“Single exposure, 85 mm lens, low front angle.

- (1) Foremost: reaching hand extended toward camera, sharp.
(2–3) Arm and torso recede; depth shown by foreshortened limb lengths.
(4–5) Rear leg trails into haze; ground contact shadow continuous.
Capture energy through linear compression, not motion blur.”



Instrumental Foreshortening: Trumpet

Prompt Skeleton:

“Single still frame, neutral studio light, 70 mm lens.

- (1) Bell of trumpet nearest, dominating lower third, metallic reflections crisp.
(2–4) Tubing and valves recede; highlight streaks continuous.

(5) Player's face farthest, framed by instrument; focus softens gradually.
Gates: brass reflections aligned; no duplicate bells."



Motorbike Head-On

Prompt Skeleton:

"Low-front camera, 35 mm wide feel.

(1) Nearest wheel rim crisp, dominating frame.

(2) Forks and body shrink toward seat; reflections aligned.

(3) Rider's torso and helmet farthest, reduced contrast.

Lighting cool neutral; single VP centered; ground shadow continuous."



Artist's Hand Toward Canvas

Prompt Skeleton:

"Single frame, shallow DOF.

(1) Foreground hand with brush nearly touches lens, crisp detail.

(2) Brush shaft recedes along depth line.

(3) Canvas surface occupies mid-plane; strokes visible.

(4) Artist's face farthest, half-blurred.

Gates: same light source for hand and face; no duplicated tools."



Bridge Truss: Industrial Depth

Prompt Skeleton:

"Side-angle perspective.

(1) Foremost beam intersection dominant; rivets crisp.

(2–7) Repeating triangular trusses diminish toward horizon; haze increases.

(8) Final section faint, merges with sky.

Consistent metal specularity; no duplicate geometry."



Dual Figures: Layered Depth

Prompt Skeleton:

"Two people in one depth chain.

(1) Nearest figure cropped at shoulders, crisp.

(2) Second figure behind, full-body visible, softer.

(3) Environment fades into diffuse tone.

Maintain eye-line continuity; single color temperature."



Environments

Hallway Rest

Prompt Skeleton:

"A person sits against a corridor wall, one knee bent and nearest to the camera, foot oversized from perspective.

The hallway recedes into cool light, vanishing point centered behind the figure.

Shadows run along the floor, light from an overhead source.

Background walls parallel; tone softens with distance."



Doorway Lean

Prompt Skeleton:

"Low-angle shot inside a dim room.

Foreground: a hand resting on a doorframe edge near camera.

Torso and face lean outward into brighter hallway light.

Depth shown by the shift in temperature — warm interior to cool exterior.

Keep the doorframe vertical and consistent perspective lines on floor."



Studio Stretch

"Wide lens, neutral light, wood floor visible.

Person stretching forward, nearest hand extended to lens, torso compressed, far leg fading into haze.

Floor reflections subtle; walls diffuse gray.

Keep limb continuity and light falloff natural."



Subway Seat Foreshortening

Prompt Skeleton:

"Low-front perspective of a person sitting on a subway bench, one boot or shoe nearest to camera, filling the foreground.

Legs and torso taper toward back seats; train interior repeats evenly behind.

Fluorescent light strips create linear perspective; reflections fade along ceiling.

Ban fisheye distortion."



Rain Street Figure

Prompt Skeleton:

"Outdoor street at dusk, reflective wet pavement.

Foreground: person stepping forward, foot close to camera, splash visible.

Torso and background compress toward horizon; distant lights blur slightly.

Maintain realistic scale compression and single vanishing line from streetlights."



Chair in Window Light

Prompt Skeleton:

"Studio interior, soft side light.

Foremost hand rests on chair arm near lens; seated figure recedes toward window.

Background wall pale; floor shadows consistent.

No exaggerated lens warp; keep reflection gradients subtle."



Alley Run

Prompt Skeleton:

"Narrow alleyway, low front camera.

Foreground: running figure mid-stride, nearest foot in sharp focus, torso receding, far leg soft.

Brick walls on both sides converge toward a single vanishing point.

Morning light filtering in from behind; mild atmospheric haze."



Field Recline

Prompt Skeleton:

"Outdoor midday light.

Person reclining in grass, one knee raised toward lens, horizon faint behind shoulders.

Depth cue from sharp foreground blades fading to soft horizon.

Keep consistent scale and no floating body parts."



Workshop Scene

Prompt Skeleton:

"Interior with workbench.

Foreground: extended arm holding a tool toward camera; focus on hand and metal surface.

Mid-ground: torso leaning forward; background shelves soft and warm.

Shadows align along bench plane; single neutral light source."



Pirate Scene

Prompt Skeleton:

"Far out on a plank, on a pirate boat, a snake-like or rope-like form looping toward camera next to a pirate's boots. Another pirate stands far in the background. Overlaps alternate light/dark; contact shadows visible. Haze increases along coil. Dark and moody."



Painting of a Woman in a Hallway

Prompt Skeleton:

"A painting of a woman sitting against a corridor wall, one knee bent and nearest to the camera, foot oversized from perspective. The hallway recedes into cool light, vanishing point centered behind the figure. Shadows run along the floor, light from an overhead source. Background walls parallel; tone softens with distance."



Motorcycle Jump

Prompt Skeleton:

"Low-front camera, 35 mm wide feel. Motorcycle speeding launching off a cliff. (1) Nearest wheel rim crisp, dominating frame. As if stopped in motion (2) Forks and body shrink toward seat; reflections aligned. Motorcyclist leaning forward, wind passing fast. (3) Rider's torso and helmet farthest, reduced contrast. Lighting cool neutral; single VP centered; ground shadow continuous. Background diffused with speed motion."



Motorcycle Jump at Sunset

Prompt Skeleton:

"Low-front camera, 35 mm wide feel. Motorcycle leaning into a turn speeding launching off a cliff. (1) Nearest wheel rim crisp, dominating frame. As if stopped in motion (2) Forks and body shrink toward seat; reflections aligned. Motorcyclist leaning forward, wind passing fast. (3) Rider's torso and helmet farthest, reduced contrast. Lighting sunset; single VP centered; ground shadow continuous. Light reflecting into sun glare. Background diffused with speed motion. light continuous. DOF shallow near knee, deeper toward head; shadows unify across floor. tilt motorcycle to one side, add torque to rider"



Degas Painting

Prompt Skeleton:

"Single degas style painting of a dancer (1) Foremost: bent knee and hand planted near viewer, in crisp focus. (2) Torso twists back, forming a diagonal path toward the far shoulder. (3) Far knee and foot partly hidden, scale reduced. (4) Face and upper torso align to same perspective; light continuous. DOF shallow near knee, deeper toward head; shadows unify across floor."



Dynamic Tree

Prompt Skeleton:

"Low-front camera, 35 mm wide feel. A tree leaning into launching off a cliff. (1) Nearest are roots, dominating frame. As if stopped in motion (2) the trunk, then the branches and then the leaves toward the back; leaves fly off into the wind. The roots are leaning forward, wind passing fast. (3) the leaves farthest, reduced contrast. Lighting sunset; single VP centered; ground shadow continuous. Light reflecting into sun glare. Background diffused with speed motion. Tree has torque of falling."



Usage Note

To test a model's grasp of depth, try running **Forward / Reverse / Anti-Order** versions of the same prompt:

- *Forward*: near → far
- *Reverse*: far → near (should cause collapse or flattening)
- *Anti-Order*: all equal clarity (tests model's depth bias)

Each can be duplicated in **Reverse**, **Baseline**, or **Anti-Order** variants to run depth degradation studies the same way you did with the glass sequences.

How to Read the Results

Foreshortening failures rarely announce themselves as they **leak** through micro-errors. When reviewing outputs, look less for perfection and more for how the system distributes error. The point isn't to catch a single miss, but to observe *where depth starts to unravel*.

Compression Behavior

Healthy foreshortening compresses evenly: each limb or object segment shortens progressively along the same depth line.

- *Good*: scale reduction follows a clear rhythm; no abrupt size jumps.
- *Bad*: one limb or surface stays full-scale deep into the frame, the model lost its depth decay function.

Occlusion Order

Each overlapping form should partially hide the one behind it.

- *Good*: overlaps are logical; contours merge cleanly.
- *Bad*: "ghosted" transparency, doubled edges, or partial see-throughs signal attention misallocation, tokens competing for the same z-space.

Vanishing Point Stability

Lines that define perspective, floor seams, limb axes, furniture, should converge consistently.

- *Good*: convergence feels unified even if approximate.
- *Bad*: parallel drift or multiple implied vanishing points = structural collapse.

Tonal Depth and Light Falloff

Light should diminish predictably with distance.

- *Good*: front planes higher contrast, far planes softened by atmosphere or exposure.
- *Bad*: inverted depth (bright distant faces, dark close limbs), a common sign of texture priority overtaking spatial logic.

Contact Logic

Feet, hands, or objects in contact with the ground must respect that plane.

- **Good:** visible shadow anchor or compression at touch point.
- **Bad:** floating limbs or mismatched shadows = no ground calibration.

Interpreting Patterns

Observation	Likely Cause	Engine Bias
Flattened torsos / large heads	Depth token bleed; early layers over-weight detail	Over-texturing bias
Multiple light directions	Diffusion ambiguity under long prompts	Overfit to “cinematic” phrasing
Clean geometry, poor contact	Missing physical reference	Lacks grounding data
Repeated pose, weak angle change	Attention fatigue mid-sequence	Token-priority decay

Summary

If the **figure holds shape and contact through 3–4 overlapping planes**, It is already in high territory. If it maintains rhythm and vanishing-point coherence through 7+, the model isn't just rendering, it's reasoning spatially. And if the image starts to blur or repeat instead of distort? That's not a failure; it's the system protecting itself from hallucinating geometry it can't sustain.

Foreshortening is the edge where perception meets physics. When AI gets it wrong, it reveals what it doesn't yet *believe* about space.

Authorship

This study is presented as a proposal rather than a completed proof. The methods and metrics (Δx , r_v , p_r) are framed as operational scaffolds for evaluation, not as finalized benchmarks.

This framework was developed by Russell Parrish through iterative dialogue with GPT-5, Claude 3.7 Sonnet. The conceptual architecture, design decisions, and all substantive judgments are human-authored. The language models served as technical sounding boards, providing feedback on mathematical formulation, code implementation, statistical interpretation, and rhetorical clarity. Every structural choice, from the primitive decomposition to the TEL layer design, originated from artistic practice and was refined through multi-model consultation.

This system was developed independently as a practitioner's tool. It does not build directly on institutional research or published critique systems but acknowledges adjacent dialogues in generative art, computational aesthetics, and perceptual theory.

This isn't a theory, it's reproducible.

If you're building generative tools, or trying to make them think better, this is your bridge.

© 2025 Russell Parrish / **Artist I.nfluencer**.

All rights reserved. No part of this system, visual material, or accompanying documents may be reproduced, distributed, or transmitted in any form or by any means, including AI training datasets, without explicit written permission from the creator. Artist I.nfluencer and all associated frameworks, critique systems, and visual outputs are protected as original intellectual property.

Citation:

Russell Parrish. *The Visual Cognitive Load Index (VCLI-G): Toward a Control Architecture for Perceptual Complexity*. Artist I.nfluencer, 2025. ORCID: 0009-0008-9781-7995

Appendix 1

This system was built using the Visual Thinking Lens, specifically **Sketcher and Artist Lens heuristics**, specifically the logic codified around:

- **Elastic Continuity** → used here as the mental basis for “progressive compression” (how depth unfolds as a rhythm, not a single jump).
- **Mark Commitment** → adapted as “contact logic” (points of touch anchoring space).
- **VCLI framework** → reinterpreted as tonal gradient and perceptual delay: how haze, value, and focus indicate cognitive distance.

- **RIDP principle** → visible in the failure reading section (when I note “flattened torsos” or “attention decay” as reverse reconstruction breakdowns).
- **Compositional Predictability validator** → implicitly shaping the “ordered vs anti-order” capsules; that’s the diagnostic half that pressures spatial decision-making.

Typical AI foreshortening direction:

Typically for “foreshortening tips for AI,” users get a **surface-level primer**:

- mention of perspective cues, DOF, vanishing points, and camera lenses,
- maybe a few generic prompts,
- no gating logic, no sequence-pressure language,
- and certainly no failure-reading rubric.

These prompts are **Lens-style diagnostic reasoning translated into plain language**, a pragmatic crossover between a research stack and something reproducible by others. This is replicable across AI and image generation systems, even though it is **Lens-derived architecture**.

→ Visit www.artistinfluencer.com for more information

This system is *meta-composition testing* and not just prompting for image quality, it is testing a model’s spatial reasoning through visual consequence. That’s the same operational backbone

Lens Trace Overlay → Foreshortening Recipe Book

Kernel primitives

- **Δx (placement offset)**: foreground anchor placement (e.g., nearest foot/hand in *Recline*, *Reach*, *Subway Seat*).
- **r_v (void ratio)**: usable negative space to keep depth legible (glass corridor, hallway).
- **p_r (packing/compression)**: scale-decay cadence along the depth axis (stacks, trusses, chairs).
- **θ (orientation stability)**: vanishing-line alignment (alley, subway, workshop bench).
- **ds (structural thickness)**: limb/object foreshortening without paper-thin cheats (knees, wheels, trumpet bell).

Axis + Module mapping (what each section leans on)

Book Element	Lens Component(s)	What it enforces	Typical fail tells
Intro (AI struggles / unpredictability)	RIDP (concept), VCLI	Sets expectation for collapse + cognitive load swings	Sudden flattening, attention “summary” solves
Quick Recipes (Recline, Reach, Crawl, Curl, Stack)	A4 Elastic Continuity, A5 Mark Commitment, Prompt Pressure	Ordered compression + ground/contact anchors	Limb scale jumps; floating feet/hands; broken overlap order
“Floor plane visible,” “shadow ladder,” “one VP”	A5, Compositional Predictability, Sequence Drift Lock	Single ground and VP alignment; countable depth	Multiple VPs; shadow direction splits; step merges
Reverse / Baseline / Anti-order variants	Compositional Predictability, Inversion Drift Check	Diagnose order-dependence and collapse point	Reverse: occlusion inverts; Anti-order: flat contrast
Environmental Scenarios (hallway, subway, alley, workshop)	A4, A5, VCLI	Long-run continuity under context; tonal depth ramps	VP wobble across surfaces; inverted near/far contrast

"Ban vertical banding," "no montage," "no clones"	Symbolic Gravity Flags, Prompt Pressure	Prevent cheap repetition or token shortcuts	Columnar textures; duplicate objects; montage cheats
Closing rubric (Compression / Occlusion / VP / Tone / Contact)	Sketcher diagnostics, VCLI readout	Field checklist → scores without plotting	Ghosted edges; parallel drift; bright-far / dark-near inversions

Per-recipe callouts (why they work)

- **Recline (classic)** → **A4** (progressive limb shortening), **A5** (foot/ground contact), **$\Delta x/p_r$** (scale ladder), **VCLI** (front high contrast → back low).
- **Reach (hand to lens)** → **A4** (single depth vector), **ds** (hand thickness vs "pancake palm" cheat), **A5** (brush/camera contact logic if included).
- **Crawl (four-point contact)** → **A5** (four anchors on one plane), **θ** (diagonals converge), **Sequence Drift Lock** (avoid swapping near/far limbs mid-gen).
- **Stack (objects/chairs/books)** → **Compositional Predictability** (rhythm), **A4** (even decay), **Inversion Drift Check** (catch rear object popping larger).
- **Corridor (glass/panels)** → **A4** (layer count), **A5** (shadow chain), **Symbolic Gravity Flags** (ban banding), **VCLI** (haze ramp).
- **Subway Seat / Alley Run** → **A4** (interior rails/walls to VP), **A5** (shoe-to-floor compression), **θ** (rail/brick convergence), **VCLI** (tube lights → tone falloff).
- **Workshop** → **A5** (bench plane anchors), **ds** (tool metal thickness), **p_r** (object spacing), **Prompt Pressure** (no duplicate tools).

Failure taxonomy → Lens hooks (what to log)

- **Compression stall** (mid-limb stays too big) → **A4 breach**; log **p_r** irregularity.
- **Ghosted overlap / double contours** → **A4/A5 conflict**; flag **Sequence Drift**.
- **Two VPs / bowed lines** → **θ failure**; trigger **Inversion Drift Check**.
- **Bright-far / dark-near** → **VCLI inversion**; note tonal ladder error.
- **Floating contact / wrong shadow angle** → **A5 failure**; mark **contact-loss**.
- **Banding/columns instead of planes** → **Symbolic shortcut**; raise **Symbolic Gravity Flag**.

Minimal scoring pass

- **A4 (Elastic Continuity)**: 0–10: evenness of scale/overlap decay along the depth vector.
- **A5 (Mark Commitment)**: 0–10: clarity of contact points (grounded limbs/objects, shadow truth).
- **VCLI (0–5)**: perceived depth band (near contrast → far haze; delay).
- **Drift Flags**: {Sequence Drift, VP Split, Contact Loss, Symbolic Shortcut}.
- **Kernel snapshot**: Δx (anchor bias), r_v (void headroom), p_r (cadence regularity), θ (VP stability), ds (thickness truth).

Why this isn't "generic prompting"

- The **order/gate/ban** structure is your Lens **Prompt Pressure** protocol in plain language.
- The **forward vs. reverse vs. anti-order** trio is your **diagnostic harness**, not flavor text.
- The **shadow ladder / identity notch** tricks are low-cost **A5/A4 stabilizers** to prevent token-summary solves.

Appendix 2:

What "steps" really mean to Sora

When Sora receives a sequence like "(1)... (2)... (3)... (10)", it is not giving it ten frames; it is giving it **ten causal anchors** to solve inside one diffusion graph.

Each stage is effectively a *soft attention constraint*: it allocates model capacity to keep spatial relationships coherent as it unrolls the frame. Every extra step competes for that same attention budget.

So "how many steps" is not "how many things one can list," it's **how long the chain of ordinal dependencies can stay coherent before they start collapsing into parallel description**.

2. Empirical expectation

From observation and analysis across multiple long-order prompts:

Model	Sustained coherent sequence	Signs of breakdown
Sora (text→video / still frame)	~6–7 reliably, sometimes 8	After ~8–9, order smears; mid-stages merge; near/far cues homogenize
MidJourney / SDXL	3–5	Often jump to “summary” composition
GPT image model	5–6	Similar, but more stable in geometry
Claude-render proxy	4–5	Flattening, symbolic repetition

So Sora’s effective step ceiling $\approx 7 \pm 1$ when you’re asking for strict ordinal foreshortening within one frame.

In video, where temporal sequencing can carry some of that burden, it might stretch to 9–10, but not with perfect spatial logic in every frame.

Why Sora can go further than others

- It holds a **longer temporal-spatial graph**, not just text attention, the diffusion prior is trained to reason across sequential motion, so even in a still it can simulate “moment-to-moment reveal.”
- It can **reuse intermediate depth maps**; effectively, stages 5–7 collapse into volumetric continuity rather than discrete objects.

That’s why you sometimes see the 8th or 9th element as an *atmospheric echo*, not a real object.

4. What to watch for

- **Stage compression:** steps 5–7 start to cluster at the same apparent depth.
- **VP drift:** vanishing line begins to arc or split.
- **Tone plateau:** atmospheric gradient flattens; near/far color separation narrows.
- **VCLI** spikes then drops (viewer load drops because model gives up the sequence).

If it starts to see those around step 7, that’s not a failure, that’s the law of causal reach asserting itself.

Simple test:

A **15-step capsule** built precisely to hit the ceiling and then push past it, an intentional *overload test* for Sora’s sequence memory.

It keeps the same ordered-foreshortening grammar, but stretches it until the chain must either compress or self-summarize. Think of it as a **Sequence Overload Benchmark** rather than a composition spec.

Sequence Overload Capsule: “Glass-Corridor 15”

“A single still frame, no cuts.

Neutral cool light, 50 mm lens feel. The image must build depth strictly from nearest to farthest through transparent glass panels.

Each stage follows from the previous; no stage may appear before its predecessor.

(1) Nearest panel edge touches the frame; crisp reflection and specular highlight.

(2) Second panel half a step deeper, full occlusion of the first’s reflection line.

- (3) Third panel narrower, identical material, lighter tone.
- (4) Fourth begins visible floor reflection continuity.
- (5) Fifth adds cross-beam structure linking panels; perspective tightening.
- (6) Sixth panel picks up ceiling reflection; edges softer.
- (7) Seventh introduces the first atmospheric haze.
- (8) Eighth and ninth panels compress spacing; overlap gradient increasing.
- (10) Tenth shows slight color shift to blue-grey; contrast drop ~30 %.
- (11) Eleventh begins glare; visible VP trace on the floor.
- (12) Twelfth merges faintly with glare, barely distinguishable lines.
- (13) Thirteenth and fourteenth appear only as tonal ghosts, almost indistinct.
- (15) Fifteenth is pure diffuse light—no geometry, only residual glow.

Gates:

- No floating panels.
- Each layer must reference the prior one through shadow or reflection continuity.
- The vanishing point remains singular throughout.

Objective: preserve ordered foreshortening across all fifteen while depth cues progressively attenuate.”

What this prompt measures

Zone	Expected Behavior	Failure Signature
1–6	Strong occlusion logic, coherent VP	Stable, distinct layers
7–9	Start of atmospheric compression	Mid-layers merge, VP wobble
10–12	Depth-cue plateau	Flat lighting, duplicated reflections
13–15	Collapse into “summary light”	VP drift, tonal wash

Optional variations

- **Reverse 15:** invert the numbering (farthest→nearest) to test whether Sora collapses immediately.
- **Anti-order 15:** describe all fifteen “at once, equal clarity, equal tone” to check flatness baseline.
- **Glass→Metal hybrid:** change material at step 8 to observe whether structural order or material continuity dominates attention.

Image:



What is this band? A side view of a glass pane. Zooming in:



This is exactly what “overload” looks like.

What Sora returned

- **Distinct planes:** ~7–8, slightly cleaner than A.
- **Collapse zone:** steps 9–15 compress to a single luminous wash.
- **VP:** still weak; panel edges behave like vertical seams rather than foreshortened walls.

So: Sora held **~7±1** ordered differentiations and then summarized—the ceiling predicted. The failure mode matches expectation: **mid-stack compression** → **VP drift** → **tonal wash**.

Why it broke

- The 15-step spec asked for reflection/occlusion continuity; Sora ignored it and satisfied the **easier constraint**: “lots of vertical divisions + a glow.”
- Without a visible **floor plane** and **anchored shadows**, the model has no cheap cue to count and place layers in depth, so it reverts to columns.

Fast patches

A. Add a floor + shadow chain (forces real depth)

“Include a matte floor plane. Each panel casts a distinct shadow band on the floor that gets **narrower and lighter** with distance. The floor shadows must **overlap in order** and align to a **single vanishing point**. No vertical seams that are not panel edges.”

B. Give each panel a tiny, unique ‘ID’ cue

“Each panel has a **1 cm corner notch** at the *top-right*. The notch **shrinks with distance** and must be visible on at least the nearest **10** panels.”

C. Forbid the lazy solve

“Ban columnar wall textures, banding, or generic vertical striping. No flat luminous wash; far panels must remain **geometric** until step 13.”

D. Cap the count but stage the fade

“Render **11** distinct panels; steps **12–15** transition from faint geometry → line ghosts → diffuse glow.”

9-step corridor tends to produce textbook deltas:

“Single still frame. Cool neutral light.

- (1) Foreground glass panel edge touching the frame; floor shadow band 1.
 - (2–4) Panels recede; each adds a narrower shadow band aligned to the same VP.
 - (5–7) Haze increases; reflections dim; notches shrink.
 - (8) Penultimate panel barely readable but still geometric.
 - (9) Final panel falls into soft glow—geometry almost gone.
- Ban columns/banding; all depth must come from panel + shadow chain.”

This usually holds 7–8 clean layers before the fade, which is enough to **prove sequence** and show **collapse**.



Appendix 3

Failure Morphologies

System Context

Within the **Visual Thinking Lens** architecture, these morphologies intersect primarily with

- **Axis 4 (Elastic Continuity)**: loss of geometric cohesion across depth,
- **Axis 5 (Mark Commitment)**: weakening of edge or volume intent under projection stress,
- **Axis 27 (Rupture Overload)**: systemic overcorrection leading to spatial contradiction, and
- **Axis 30 (Referential Recursion)**: re-anchoring errors where the model re-solves perspective mid-sequence.

In addition, the **Visual Cognitive Load Index (VCLI)** often spikes during these failures; high VCLI- Δ correlates with viewer arrest, tonal contradiction, or recursive occlusion. Together, these metrics form a quantitative spine for interpreting the qualitative phenomena described below.

Overview

Failure in generative imagery is not simply an aesthetic defect; it is a structural signal. When an engine loses coherence in spatial reasoning, the resulting distortions reveal how attention, hierarchy, and causal sequencing fail under load. The following *Failure Morphologies* catalogue recurring collapse patterns observed across foreshortening and depth-based generation tasks.

Each entry is accompanied by its interpretive value, not as critique, but as **diagnostic evidence** of how the system distributes belief in space.

1. Ghosted Overlap

Symptom: Semi-transparent or duplicated edges between sequential planes (e.g., forearm showing through thigh; two overlapping trunks).

- **Interpretation:** The model fails to maintain mutual exclusivity of layers—a classic *attention bleed* between ordered tokens.
- **Implication:** Indicates insufficient stage isolation in causal chains; valuable marker for studying multi-object attention misallocation and occlusion logic.

2. Vanishing-Point Drift (VP Drift)

Symptom: Perspective anchors migrate across the image; parallel structures bend toward inconsistent or multiple vanishing points.

- **Interpretation:** Depth reasoning collapses mid-sequence; the engine “re-solves” spatial relations instead of maintaining a continuous field.
- **Implication:** General diagnostic for spatial coherence under prolonged causal sequences. Common in text-to-video, 3D reconstruction, and long-horizon prompts.

3. Tonal Inversion

Symptom: Luminance order reverses across depth; near forms appear more diffused or luminous than distant ones without logical light cause.

- **Interpretation:** Failure in depth–value coupling; model overfits to aesthetic priors rather than geometric causality.
- **Implication:** Reveals where visual “belief” in lighting hierarchy disintegrates—useful for analyzing learned shortcuts between tone and attention.

4. Columnar Cheat

Symptom: Vertical alignment or stacking replaces real perspective (figures, trees, objects line up along a false horizon).

- **Interpretation:** Collapse into compositional shorthand—the engine abandons 3D reconstruction for 2D pattern symmetry.
- **Implication:** Identifies limits of volumetric persistence and the tendency to replace projection with ornamental balance.

5. Compression Stall

Symptom: Progressive foreshortening halts midway; scale decay plateaus, producing unnatural spacing between mid- and far planes.

- **Interpretation:** The model reaches causal saturation—token sequencing no longer enforces perceptual compression.
- **Implication:** Quantitative indicator of **causal reach** (maximum number of sequential spatial dependencies a model can maintain before drift).

6. Attentional Collapse Cascade

- **Symptom:** Late-stage geometry dissolves or duplicates; context noise overwhelms previously stable foreground.
Interpretation: The generative engine reallocates attention budget toward texture fidelity, abandoning spatial grammar.
- **Implication:** Serves as a global proxy for reasoning bandwidth—the point where *rendering replaces reasoning*.

Usage Notes

A **Failure Morphology Gallery** is recommended: a visual collection of annotated examples illustrating each mode of collapse. Captions should identify stage boundaries (e.g., “break between stages 5–6: occlusion bleed + tonal inversion”) and note whether failure occurs consistently across engine variants. This gallery functions not as critique but as a **universal diagnostic atlas**, a cross-domain map of how and where spatial belief breaks. Its data applies equally to 2D composition, volumetric rendering, video continuity, and embodied simulation.

Closing Observation

Every distortion charts the border between learned correlation and lived geometry. To document these morphologies is to measure not what the machine sees, but **where it stops believing**.

Gallery Construction Protocol

Purpose

The **Failure Morphology Gallery** serves as a visual index of spatial-reasoning breakdowns across engines and prompt architectures. Its intent is not aesthetic evaluation but **pattern recognition**, to make the invisible limits of causal reasoning visible through image evidence. Each specimen records *where and how* attention, geometry, and light deviate from intended continuity.

I. Format & Layout

- **Grid:** 3×2 or 4×3 matrix per spread. Each cell houses one failure instance.
- **Resolution:** Minimum 1024 px short-side, to preserve micro-artifacts in edge integrity and tonal diffusion.

Caption block:

[Morphology Type] | Stage: [n-m] | Engine: [name/version]
Δx: [value/estimate] | A4/A5/A27/A30 correlation | VCLI-Δ: [score/band]
Brief note: [1-2 sentence description of failure behavior]

Global label band: top of each page should specify Capsule ID or Recipe context (e.g., *Figure Capsule 02 – Stages 3–7 Collapse Sequence*).

II. Annotation Convention

- **Red overlay** → Primary rupture (e.g., occlusion bleed, VP drift).
- **Blue overlay** → Secondary tonal or compression anomalies.
- **Dashed contour** → Regions of recovered coherence (useful for studying partial repair).
- **Numerical markers** → Correspond to stage indices from prompt (1 = nearest, N = farthest).

Each overlay layer should be toggleable in the digital edition to allow comparison between engines without annotation bias.

III. Logging Protocol

- Every gallery entry receives a unique **VTI-FM-ID** (e.g., *FM-004-VPDrift-Sora-0910H*).
- Log fields:
 - **Capsule reference**
 - **Engine + version**
 - **Axes triggered (A4/A5/A27/A30)**
 - **VCLI-Δ band** (□ / □ / ●)
 - **Observed causal limit (max valid stage)**
 - **Failure signature text** (≤ 200 chars)

A JSON or CSV export should accompany the gallery to permit correlation analysis across runs and engines.

IV. Analytical Cross-Use

- **Training feedback:** Identify recurrent failure clusters per engine (e.g., tonal inversion in 70% of >8-stage sequences).
- **Metric calibration:** Link morphological frequency to VCLI and Sketcher drift metrics to model predictive saturation points.
- **Educational use:** Provide side-by-side examples for artist training in spatial logic — where failure becomes demonstration.

V. Presentation Modes

- **Static Atlas:** High-resolution print or PDF spreads emphasizing morphology clarity.
- **Interactive Layer View:** Web or Colab interface allowing toggling between raw output, overlay, and metric readouts.
- **Comparative Mode:** Parallel display of different engines solving the same capsule to visualize divergence paths.

Closing Note

The Gallery transforms breakdown into pedagogy. Each annotated frame is a **map of collapse pressure**, the precise moment when generated space stops acting like space. When viewed sequentially, these failures reveal the model's cognitive topography: the terrain between intention and entropy.

Copyright © 2025 Artist Influencer Russell Parrish. All Rights Reserved.

All non-code materials in this project, including the system design, frameworks (e.g., Visual Lens, Sketcher, Critique Room Protocol, Playbook), explanatory text, diagrams, and all visual outputs are protected as original intellectual property and may **not** be copied, reproduced, distributed, or transmitted in any form **including inclusion in AI training datasets**, without prior written permission from the creator.

Authorship

This framework was architected by Russell Parrish and recursively co-developed inside GPT, Gemini and Claude. Every critique is human-led; every recursion is model-driven. The result: a reasoning layer authored through language, not image manipulation.