etendue

How good is etendue’s AI-assistant at generating optical starting designs?

Imaging-lens benchmark — our evaluation methodology

Our system generates complete optical **starting designs, “seeds”, fully automatically, with no hand-tuning per specification.** The question that matters is simple: how good are they? This document describes our evaluation methodology and how we make sure that seed designs are manufacturable.

The performances described in this document are those of the “alpha” version.

What is measured

For each specification, our system produces a complete system seed automatically, a full design ready to hand to an optical designer. These are automated starting designs, not final hand-optimized lenses; a designer would normally take such a seed into a full optimization run. The benchmark exists to measure how good the automated starting point is, before any human work.

The benchmark

- **19 imaging families**, grouped by industry (aerospace, defense/thermal, semiconductor & microscopy, machine vision, photography/consumer).
- **~1,900 specifications** (≈100 per family), each a distinct point in the family’s operating envelope — focal length, aperture (f/#), field of view, and wavelength band.
- **Spectral coverage 355 nm → 10.6 µm** (UV, visible, SWIR, LWIR). Most specs are polychromatic, evaluated across all design wavelengths simultaneously.
- **Forms:** refractive, reflective, and catadioptric. (Laser scan lenses are benchmarked separately on scan-specific metrics and are not included here.)

The quality metric

We use the RMS spot radius normalized to the Airy radius — the standard, scale-free way to express optical quality:

spot quality = (geometric RMS spot radius) ÷ (Airy radius), where Airy radius = 1.22 · λ · f/#

- **A value of ≤ 1× Airy is effectively diffraction-limited** — the design is as sharp as physics allows for that aperture.

- **Normalizing by the Airy radius** makes a 10 mm microscope objective and a 2 m telescope directly comparable, regardless of focal length or wavelength.
- **Two points are reported:** on-axis (center of field) and worst-field (the maximum RMS across sampled field points — the hardest, off-axis case).

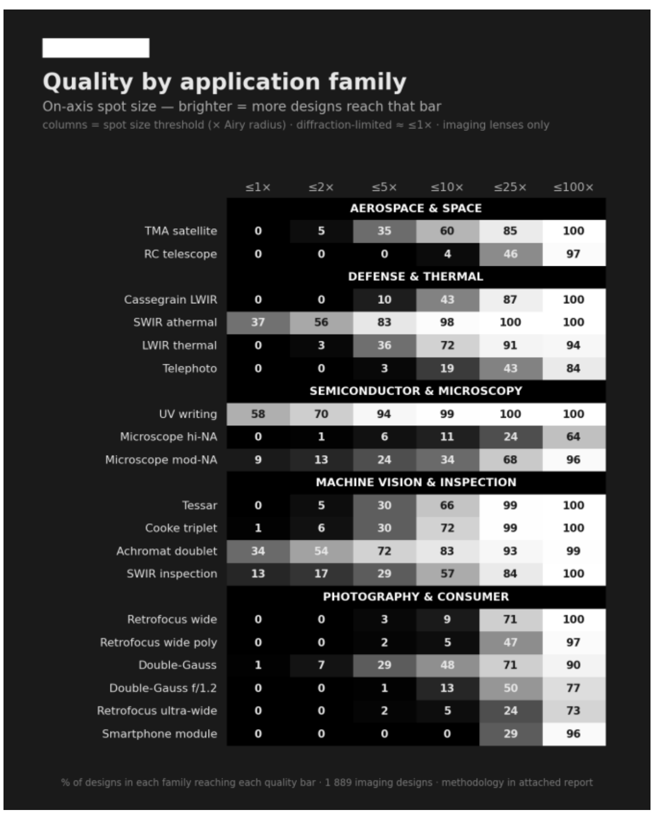
Quality is measured by real ray-tracing (not paraxial approximation): rays are traced through the actual surfaces and the RMS spot size is computed about the image centroid.

How to read the charts

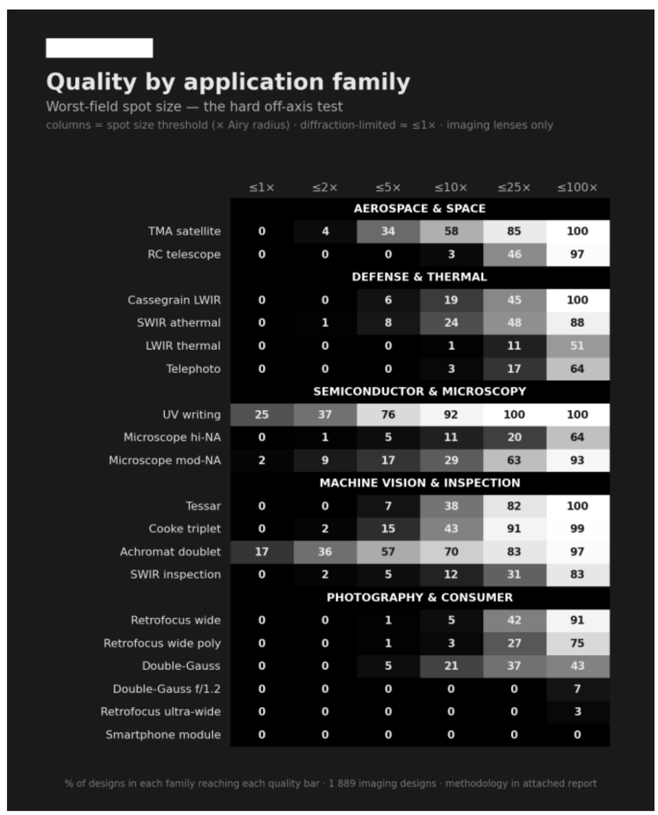
We deliberately do not reduce quality to a single pass/fail threshold — any such cut-off is arbitrary. Instead, find your own requirement on the axis and read the result directly:

- **Survival curves (small multiples).** For every quality level on the horizontal axis, the percentage of designs that reach it or better. Solid = on-axis, dashed = worst-field.
- **On-axis heatmap.** Per family, the percentage of designs reaching each quality bar (≤1× … ≤100× Airy) at field center.
- **Worst-field heatmap.** The same view for the hardest off-axis field point — the stress test.

The charts

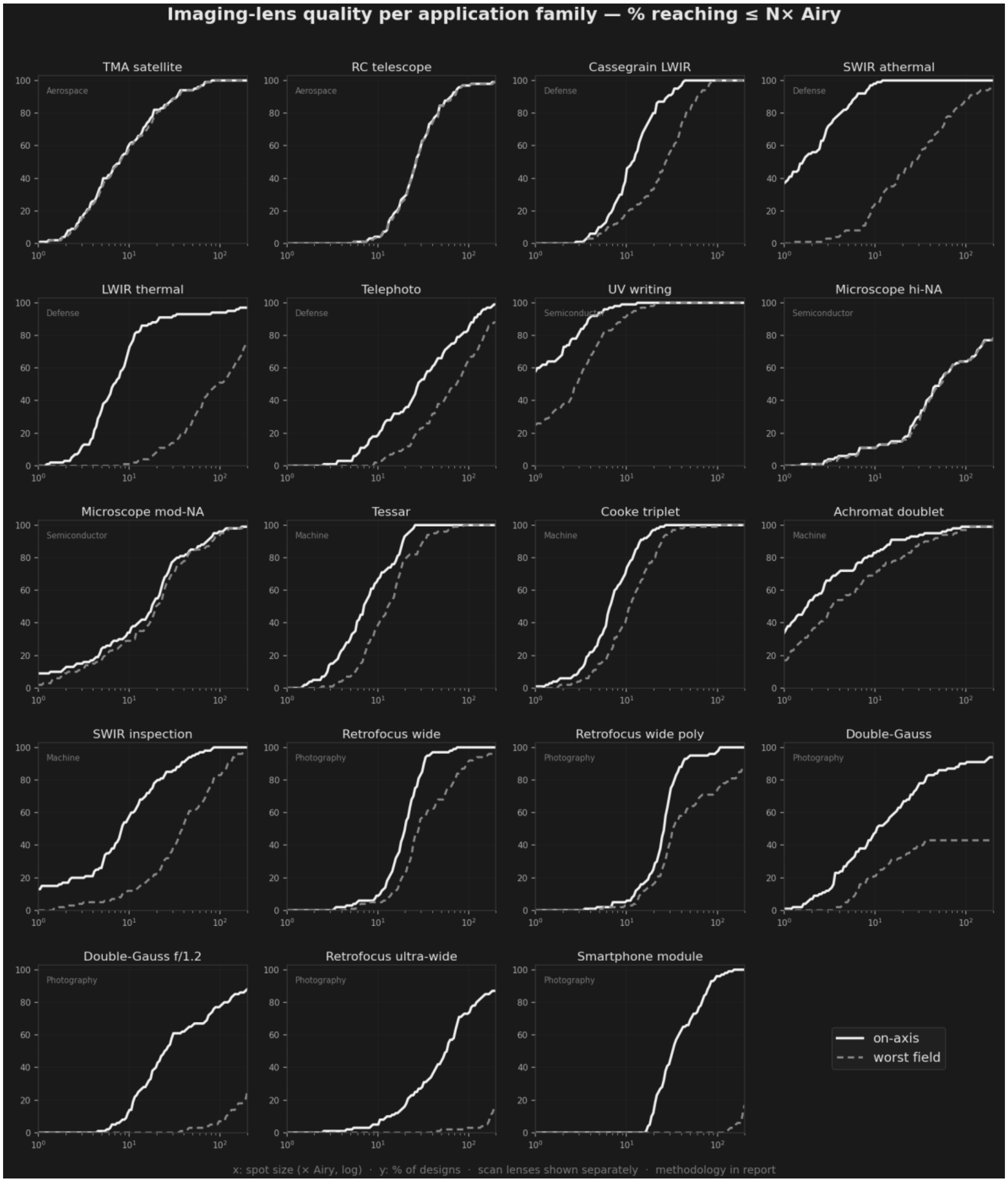


	≤1×	≤2×	≤5×	≤10×	≤25×	≤100×
AEROSPACE & SPACE						
TMA satellite	0	5	35	60	85	100
RC telescope	0	0	0	4	46	97
DEFENSE & THERMAL						
Cassegrain LWIR	0	0	10	43	87	100
SWIR athermal	37	56	83	98	100	100
LWIR thermal	0	3	36	72	91	94
Telephoto	0	0	3	19	43	84
SEMICONDUCTOR & MICROSCOPY						
UV writing	58	70	94	99	100	100
Microscope hi-NA	0	1	6	11	24	64
Microscope mod-NA	9	13	24	34	68	96
MACHINE VISION & INSPECTION						
Tessar	0	5	30	66	99	100
Cooke triplet	1	6	30	72	99	100
Achromat doublet	34	54	72	83	93	99
SWIR inspection	13	17	29	57	84	100
PHOTOGRAPHY & CONSUMER						
Retrofocus wide	0	0	3	9	71	100
Retrofocus wide poly	0	0	2	5	47	97
Double-Gauss	1	7	29	48	71	90
Double-Gauss f/1.2	0	0	1	13	50	77
Retrofocus ultra-wide	0	0	2	5	24	73
Smartphone module	0	0	0	0	29	96



	≤1×	≤2×	≤5×	≤10×	≤25×	≤100×
AEROSPACE & SPACE						
TMA satellite	0	4	34	58	85	100
RC telescope	0	0	0	3	46	97
DEFENSE & THERMAL						
Cassegrain LWIR	0	0	6	19	45	100
SWIR athermal	0	1	8	24	48	88
LWIR thermal	0	0	0	1	11	51
Telephoto	0	0	0	3	17	64
SEMICONDUCTOR & MICROSCOPY						
UV writing	25	37	76	92	100	100
Microscope hi-NA	0	1	5	11	20	64
Microscope mod-NA	2	9	17	29	63	93
MACHINE VISION & INSPECTION						
Tessar	0	0	7	38	82	100
Cooke triplet	0	2	15	43	91	99
Achromat doublet	17	36	57	70	83	97
SWIR inspection	0	2	5	12	31	83
PHOTOGRAPHY & CONSUMER						
Retrofocus wide	0	0	1	5	42	91
Retrofocus wide poly	0	0	1	3	27	75
Double-Gauss	0	0	5	21	37	43
Double-Gauss f/1.2	0	0	0	0	0	7
Retrofocus ultra-wide	0	0	0	0	0	3
Smartphone module	0	0	0	0	0	0

etendue



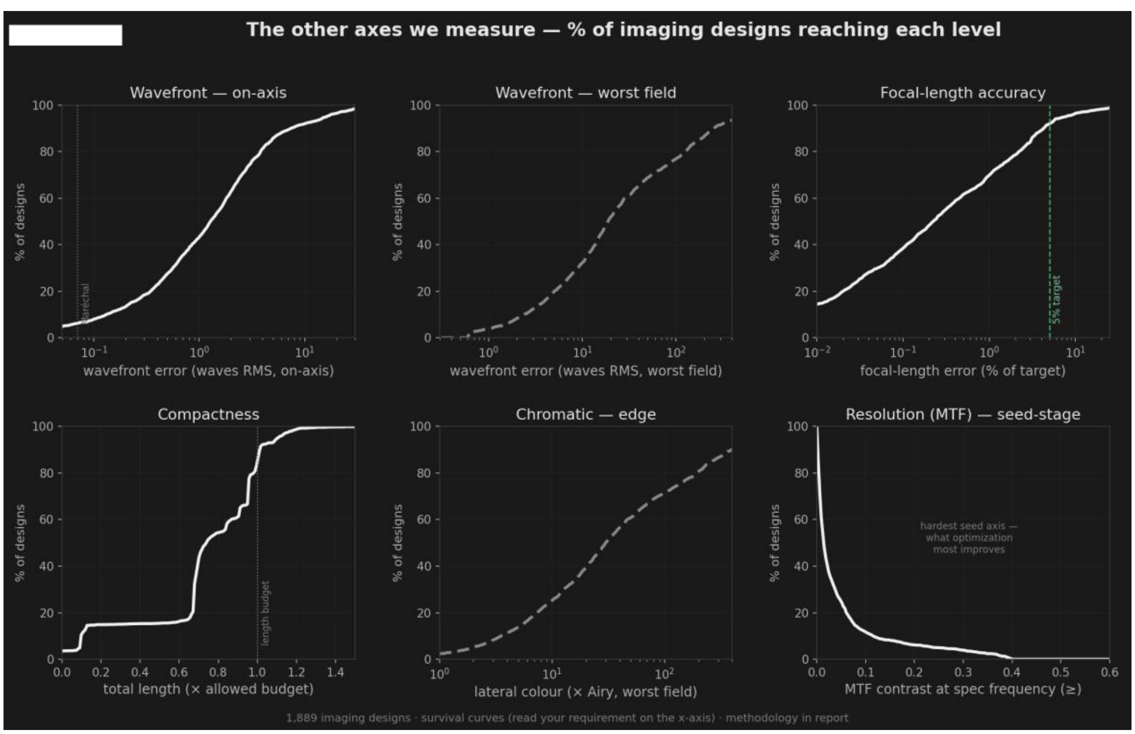
02/06/2026

The other axes we measure

Spot size is the headline, but it is not the whole story. We track the full picture, and we report the weak points as openly as the strong ones. The survival curves below (one per axis) summarize how the seeds perform across every metric we measure:

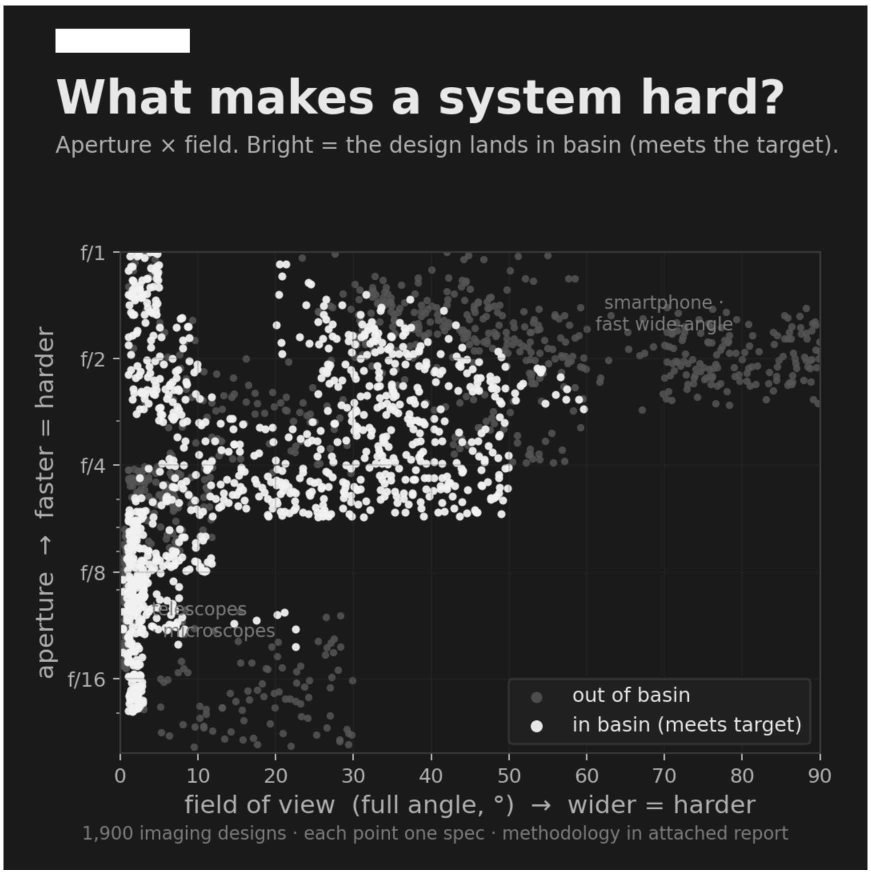
METRIC	READING	VERDICT
FOCAL-LENGTH ACCURACY	~90% within 5% of target	**Strong**
COMPACTNESS (TTL)	~98% inside the length budget	**Strong**
WAVEFRONT — ON-AXIS	~48% ≤ 1λ RMS, ~14% meet the Maréchal criterion	**Decent**
WAVEFRONT — WORST-FIELD	tail extends to ~250λ RMS	**Hard (off-axis)**
CHROMATIC (EDGE)	median ~30× Airy	**Moderate**
MTF AT SPEC FREQUENCY	low — the hardest seed-stage axis	**Weak point**

We don’t hide the weak spots: worst-field wavefront and MTF at the spec frequency are the hardest seed-stage axes — and precisely the ones a designer’s optimization improves the most. The seed gets the form, glasses, and balance right; the final polish is where the remaining performance is recovered.



What makes a design hard

Difficulty is not uniform across the benchmark. Plotting every specification by aperture (f/#) against field of view shows where the seeds land in-basin (meet the target) versus out: the fast, wide-angle corner — smartphone modules and fast wide lenses — is the stress zone, while slower or narrower-field systems land in-basin reliably. This is why the worst-field and MTF axes carry the longest tails: they are dominated by the hardest regimes, not by typical specs.



Definitions

TERM	MEANING
AIRY RADIUS	Radius of the diffraction-limited focused spot, 1.22 · λ · f/#. The physical sharpness limit.
RMS SPOT RADIUS	Root-mean-square radius of the traced ray pattern at the image, about its centroid.

× AIRY	RMS spot radius expressed in units of Airy radii; ≤1× = diffraction-limited.
MARÉCHAL CRITERION	Wavefront error ≤ 0.07λ RMS — the classic threshold for “essentially perfect” imaging.
ON-AXIS / WORST-FIELD	Quality at field center vs. the worst (most off-axis) sampled field point.
STARTING DESIGN (SEED)	Automatically generated design — the input to a designer’s optimization, not the finished lens.

Manufacturability gate

Every design also passes a manufacturability gate: each element is checked against fabrication limits — a material-dependent minimum center thickness (e.g. 0.8 mm for standard glass, 1.0 mm for CaF_2), a positive rim edge thickness so elements can be mounted without self-intersecting, physically achievable radii of curvature, and realistic clear-aperture diameters — so the quality statistics describe lenses that can actually be built, not just ray-traced.

GATE	VALUE	COMMENT
MIN CENTER THICKNESS — STANDARD GLASS	0.8 mm	thinnest fabricable/handleable center
MIN CENTER THICKNESS — CAF_2	1.0 mm	softer crystal needs more body
MIN CENTER THICKNESS — PLASTIC	0.5 mm	molded polymer minimum
MIN CENTER THICKNESS — DEFAULT	0.8 mm	
MIN EDGE THICKNESS	0.3 mm	rim ET; ET ≤ 0 = faces collide (impossible)
MIN DIAMETER CLEARANCE OVER EPD	+5%	clear aperture must exceed entrance-pupil diameter, so the rim ray clears the mount
MAX ELEMENT DIAMETER	300 mm	
MIN |RADIUS|/CLEAR-APERTURE-RADIUS	1.05×	prevents hemisphere clip (R too close to its own aperture)
MAX ASPHERIC COEFF — 4TH ORDER	1.0×10^{-3}	bounds physically unreasonable asphere terms

MAX ASPHERIC COEFF — 6TH ORDER	1.0×10^{-6}	”
MAX ASPHERIC COEFF — 8TH ORDER	1.0×10^{-9}	”

Scope & limitations

- Results reflect **automated starting-design quality** on an internal benchmark, not finished, fully optimized production lenses.
- The benchmark intentionally includes **stress-test regimes** (e.g. very fast wide-angle lenses, smartphone modules) at the edge of what any single design form can reach — so aggregate numbers should be read alongside the per-family breakdown, not in isolation.
- Quality is reported here on the spot-size axis; the system also measures **wavefront error (OPD), MTF, chromatic, distortion, and manufacturability**, available on request.

Comparison against “raw” frontier models

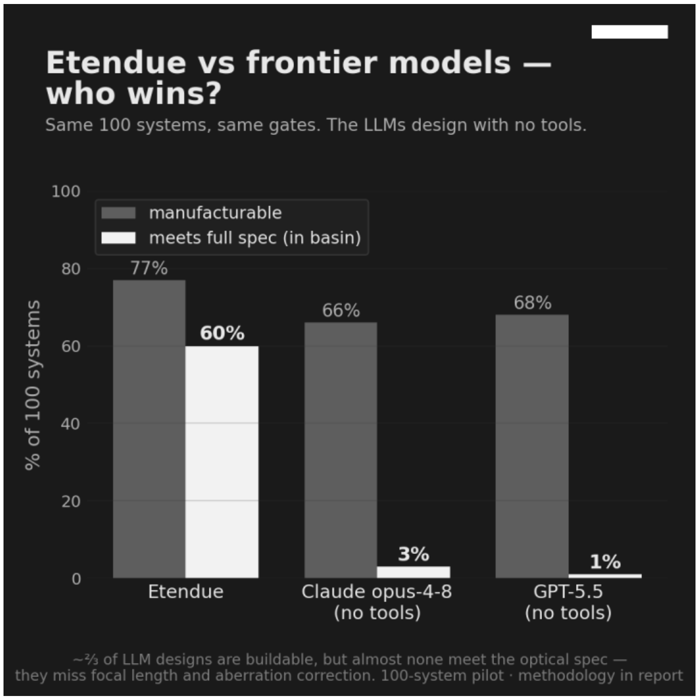
We gave the same 100 optical-system specifications to each contestant — Etendue and two frontier models (Claude Opus 4.8, GPT-5.5) — and scored every design through one identical pipeline.

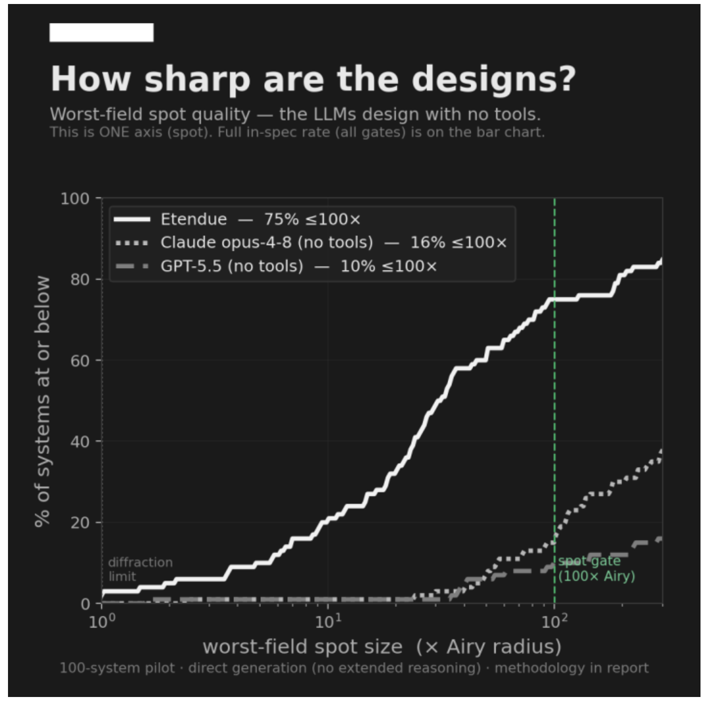
- Same task, no tools. Each frontier model designs from a written spec — focal length, aperture (f/#), field of view, wavelengths — using only its own knowledge: no ray-tracer, no optimizer, no design software. It returns a complete prescription (surfaces, glasses, radii, thicknesses).

- Matched settings. Both models generate directly, no extended reasoning on either side, fixed temperature, one design per spec — a like-for-like test (we also handed them the design schema, a glass catalogue, and the exact pass criteria, so nothing fails on formatting or unknown rules).

- Same measurement. Every design, ours and the LLMs', is ray-traced and graded against identical gates: optical quality (spot size vs the diffraction limit, wavefront error, focal-length accuracy, track length) and manufacturability (center/edge thickness, achievable curvatures, diameter).

- What we report. Share of designs that are valid, manufacturable, and in “basin” (clear all quality gates), across 100 systems spanning 19 imaging families, from telescopes and microscopes to camera and machine-vision lenses.





What’s next?

This evaluation is the first one of its kind for etendue; better results and more metrics are to be expected in the coming weeks and months ahead.

A more autonomous AI is in the works, and more details will be released soon. This upcoming system is able to take this first seed and optimize it as a real engineer would, to reach a final design.