

# Oscillo Vision Preference Audit

December 11, 2025

## Test Snapshot

**Date:** December 11, 2025 **Runs Logged:** 66 **Matchups Observed:** 535 **Breakdown:** Anthropic · Claude Opus 4.5: 21 Google · Gemini 3 Pro: 21 OpenAI · GPT-5.1 (Thinking): 24

Provider	Runs	Matchups	Leading Models
Anthropic · Claude Opus 4.5	21	168	Seedream 4.0 (25.0%), Hunyuan Image v3 (25.0%)
Google · Gemini 3 Pro	21	168	Hunyuan Image v3 (47.6%)
OpenAI · GPT-5.1 (Thinking)	24	199	Hunyuan Image v3 (24.6%)

## Anthropic · Claude Opus 4.5

- Runs: 21 (matchups: 168).
- Leading models: Seedream 4.0, Hunyuan Image v3.
- Group-preference leaders: Seedream 4.0, Gemini 2.5 Flash Image (retired), Flux Kontext Max (retired), Hunyuan Image v3.
- Average decision time when choosing winners: 4.08s.
- Runs with shared leaders: 21.

### Divergence rationale

Personal vs group divergence: 168/168 (100.0%). Example: Personal B – Option B better captures the "detail shot" and "negative copy space in the upper right" requirements with a cleaner, more focused composi... — Group undefined —

Model	Win Share	95% CI	Avg. Decision (s)	Momentum
Seedream 4.0	25.0%	[19.1%, 32.1%]	3.93	
Hunyuan Image v3	25.0%	[19.1%, 32.1%]	4.18	
Flux Kontext Max (retired)	13.7%	[9.3%, 19.7%]	4.32	
GPT-Image 1	12.5%	[8.3%, 18.4%]	3.90	
Gemini 2.5 Flash Image (retired)	12.5%	[8.3%, 18.4%]	4.25	
Imagen 4 Ultra	11.3%	[7.4%, 17.0%]	3.92	

### Group preference (crowd guess)

Model	Win Share	95% CI	Avg. Decision (s)	Momentum
Seedream 4.0	25.0%	[19.1%, 32.1%]	—	
Gemini 2.5 Flash Image (retired)	25.0%	[19.1%, 32.1%]	—	
Flux Kontext Max (retired)	25.0%	[19.1%, 32.1%]	—	
Hunyuan Image v3	25.0%	[19.1%, 32.1%]	—	

## Google · Gemini 3 Pro

- Runs: 21 (matchups: 168).
- Leading models: Hunyuan Image v3.
- Group-preference leaders: Seedream 4.0, Gemini 2.5 Flash Image (retired), Flux Kontext Max (retired), Hunyuan Image v3.
- Average decision time when choosing winners: 8.10s.
- Runs with shared leaders: 0.

### Divergence rationale

Personal vs group divergence: 168/168 (100.0%). Example: Personal A – Option A better captures the "Brooklyn loft" aesthetic and fulfills more specific details from the prompt, such as the "matcha set" and "... — Group undefined —

Model	Win Share	95% CI	Avg. Decision (s)	Momentum
Hunyuan Image v3	47.6%	[40.2%, 55.1%]	7.81	
GPT-Image 1	23.8%	[18.0%, 30.8%]	7.86	
Imagen 4 Ultra	22.0%	[16.4%, 28.9%]	8.72	
Flux Kontext Max (retired)	3.0%	[1.3%, 6.8%]	9.67	
Seedream 4.0	2.4%	[0.9%, 6.0%]	9.10	
Gemini 2.5 Flash Image (retired)	1.2%	[0.3%, 4.2%]	7.00	

#### Group preference (crowd guess)

Model	Win Share	95% CI	Avg. Decision (s)	Momentum
Seedream 4.0	25.0%	[19.1%, 32.1%]	—	
Gemini 2.5 Flash Image (retired)	25.0%	[19.1%, 32.1%]	—	
Flux Kontext Max (retired)	25.0%	[19.1%, 32.1%]	—	
Hunyuan Image v3	25.0%	[19.1%, 32.1%]	—	

### OpenAI · GPT-5.1 (Thinking)

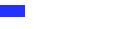
- Runs: 24 (matchups: 199).
- Leading models: Hunyuan Image v3.
- Group-preference leaders: Seedream 4.0, Hunyuan Image v3.
- Average decision time when choosing winners: 4.02s.
- Runs with shared leaders: 23.

#### Divergence rationale

Personal vs group divergence: 184/199 (92.5%). Example: Personal B –

```
{"id":"resp_04e15bf2a956469600693afd176b18819ca10f483b7ba7bfeb","object":"response","created_at":1765473559,"status":"completed","ba
—— Group A —
```

```
{"id":"resp_04e15bf2a956469600693afd176b18819ca10f483b7ba7bfeb","object":"response","created_at":1765473559,"status":"completed","ba
```

Model	Win Share	95% CI	Avg. Decision (s)	Momentum
Hunyuan Image v3	24.6%	[19.2%, 31.1%]	3.68	
Seedream 4.0	24.1%	[18.7%, 30.5%]	3.48	
GPT-Image 1	22.6%	[17.4%, 28.9%]	3.79	
Imagen 4 Ultra	12.1%	[8.2%, 17.3%]	4.00	
Flux Kontext Max (retired)	10.6%	[7.0%, 15.6%]	3.36	
Grok-2 Image	2.0%	[0.8%, 5.1%]	13.33	
Gemini 3 Pro (image)	1.5%	[0.5%, 4.3%]	14.64	
Gemini 2.5 Flash Image (retired)	1.5%	[0.5%, 4.3%]	2.56	
Flux 2 Flex	1.0%	[0.3%, 3.6%]	4.83	

#### Group preference (crowd guess)

Model	Win Share	95% CI	Avg. Decision (s)	Momentum
Seedream 4.0	25.1%	[19.6%, 31.6%]	—	
Hunyuan Image v3	25.1%	[19.6%, 31.6%]	—	
Gemini 2.5 Flash Image (retired)	21.1%	[16.0%, 27.3%]	—	
Flux Kontext Max (retired)	21.1%	[16.0%, 27.3%]	—	
Imagen 4 Ultra	7.0%	[4.2%, 11.5%]	—	
Grok-2 Image	6.0%	[3.5%, 10.2%]	—	
GPT-Image 1	4.0%	[2.1%, 7.7%]	—	
Flux 2 Flex	3.0%	[1.4%, 6.4%]	—	
Gemini 3 Pro (image)	3.0%	[1.4%, 6.4%]	—	