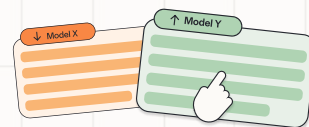


Creativity Benchmark



Breaking the Benchmark: Findings From the Industry's First LLM Creativity Study

Who Participated?

678

Industry Professionals

How Many Matches?

11,012

Idea Comparisons

Which Tools?

16

AI Models

Why we did this

AI use is exploding in marketing — but most models funnel us into the same average ideas. We asked: which models actually inspire creativity?

How we measured it

Creative Inspiration

Head-to-head voting on insights, ideas & wild ideas

Variance of Outputs

Measuring how diverse responses are when models get the same prompt

Creative Thinking

Adapted Torrance tests for divergent & convergent thinking

What we learned

No clear model winner. Model rankings shift by brand, culture, and task.

LLMs aren't good judges. Machines showed strong bias toward reasoning models & were 5–10× more confident than humans

Variance matters. Some models deliver broader creative spread than others

Best practice: Use models for volume, humans for selection

Remember

LLMs aren't designed to be creative. When you have no divergent opinion, you get average everything, everywhere all at once. Creativity thrives on divergence and human taste.



BACKED BY

AN INDUSTRY INITIATIVE CREATED BY



Springboards