

Tabla de Benchmarks de IA

Líderes actualizados al 20 de febrero de 2026

🔍 Da clic en el nombre del benchmark para ir directamente a su ranking

Benchmark	Qué mide	Líder actual	Score
 Arena (LMArena)	Preferencia humana general	Claude Opus 4.6 Thinking	1506 Elo
 MMLU-Pro	Conocimiento general (14 áreas, nivel posgrado)	Gemini 3 Pro	89.80%
 GPQA Diamond	Ciencia nivel doctorado (física, química, biología)	Gemini 3.1 Pro	90.80%
 AIME 2025	Matemáticas de olimpiada	GPT-5.2	100%
 Humanity's Last Exam	Lo más difícil del conocimiento humano	Gemini 3 Pro	37.20%
 SWE-bench Verified	Ingeniería de software real	Claude Opus 4.5	80.90%
 OSWorld	Uso real de computadoras para tareas reales	Kimi K2.5	63.30%
 ARC-AGI-2	Razonamiento abstracto e inteligencia fluida	Gemini 3 DeepThink	84.60%
 Image Arena	Generación de imagen (preferencia humana)	GPT Image 1.5 (OpenAI)	1264 Elo
 Video Arena	Generación de video (preferencia humana)	Kling 3.0 Pro	1249 Elo
 AAII	Rendimiento general (índice compuesto)	Gemini 3.1 Pro	57%