

Inference price across select benchmarks, 2022–24

Source: Epoch AI, 2025; Artificial Analysis, 2025 | Chart: 2025 AI Index report

- GPT-3.5 level+ in multitask language understanding (MMLU)
- GPT-4 level+ in code generation (HumanEval)

- GPT-4o level+ in PhD-level science questions (GPQA Diamond)
- GPT-4o level+ in LMSYS Chatbot Arena Elo

