

# GPT-4

- 2023年3月にリリースされた入力をマルチモーダルに拡張したモデル
- GPT-3.5よりも複雑なタスクを解く能力が高く、多くの評価タスクで大幅な性能向上を示している
- 6ヶ月間の反復調整により、悪意や挑発的な意図のある質問により安全に応答できるようになった
- Red Teamを組んでバイアスや憎悪に満ちた出力を生成させるなどの攻撃をさせ、有害なコンテンツ生成を減らした

# 3. LLMのリソース