

AIの経済価値を測る新基準：OpenAIの「GDPval」とは？



GDPvalとは？－AI評価の新時代

「現実世界の経済的価値」を測定

GDPに貢献する44職種の1320件の専門的タスクでAIの性能を評価します。



学術的な問題

実務に基づいたリアルなタスク設計

法律意見書や設計図など、現実の業務成果物に基づいてタスクが作られています。



評価ベンチマーク vs 評価の焦点 vs 現実世界との乖離

MMLU (従来例)



学術知識の
カバー範囲

高
(試験問題形式)

VS

GDPval (新基準)

知識労働における
成果物の品質

極めて低
(実在の成果物ベース)



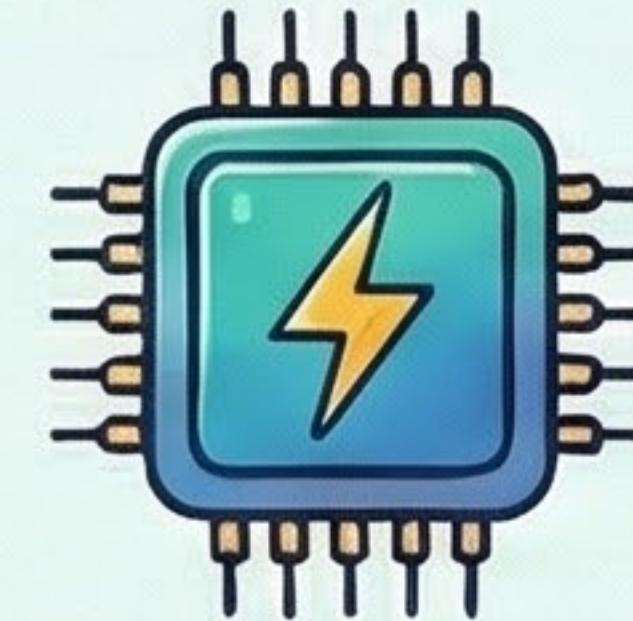
主な調査結果と社会へのインパクト

最新AIは専門家の仕事より
100倍速く、コストは100分の1

純粋なモデルの推論時間とAPI料金に
基づく理論上の試算です。



100倍
遅い



100倍
速い

VS



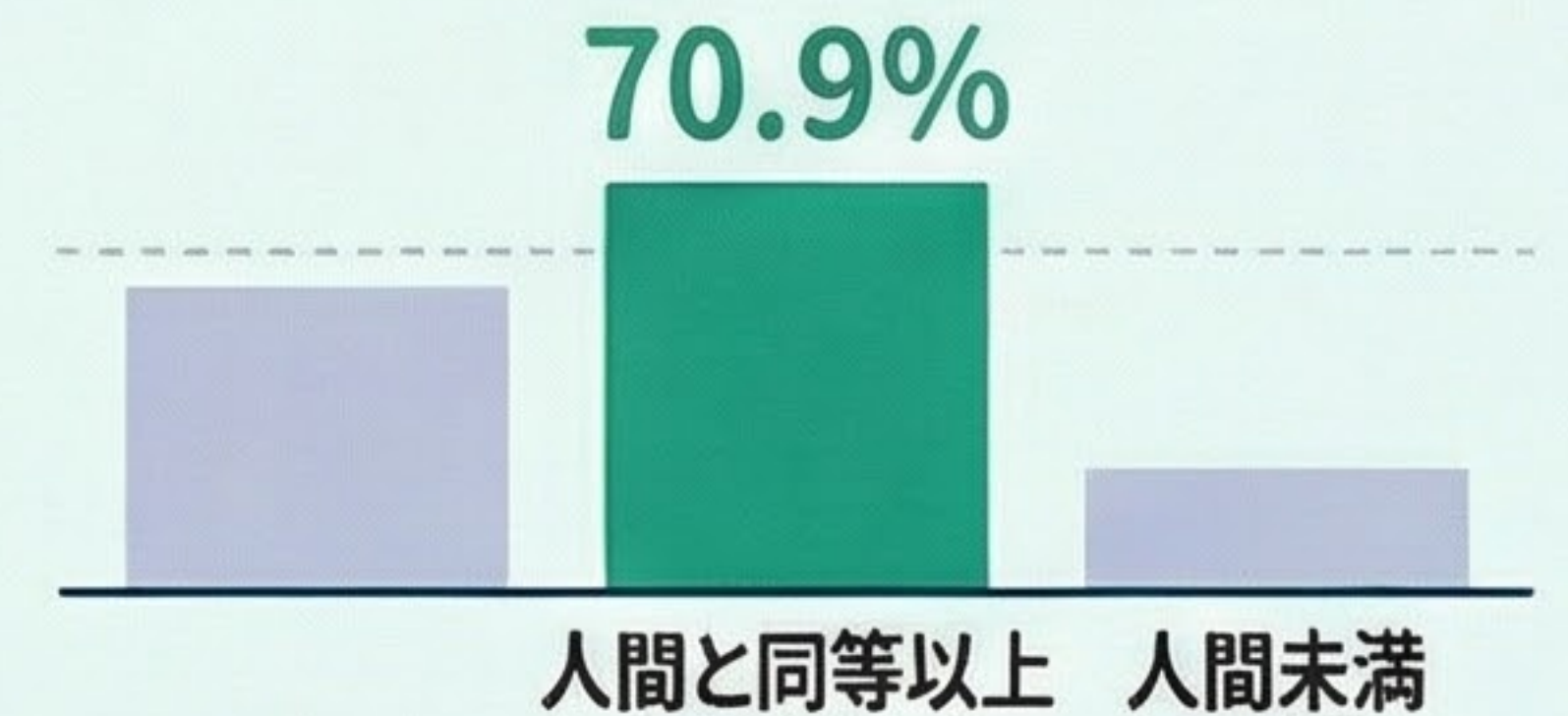
100倍
のコスト



1/100
のコスト

GPT-5.2は70.9%のタスクで
人間と同等以上の成果

人間の専門家に対する勝敗・引き分け
率で、モデルの飛躍的進歩を示しました。



AIによる生産性革命の現実味

AIとの協業で、数日かかった仕事を数時間で完了できる可能性を示唆しています。



数日

数時間