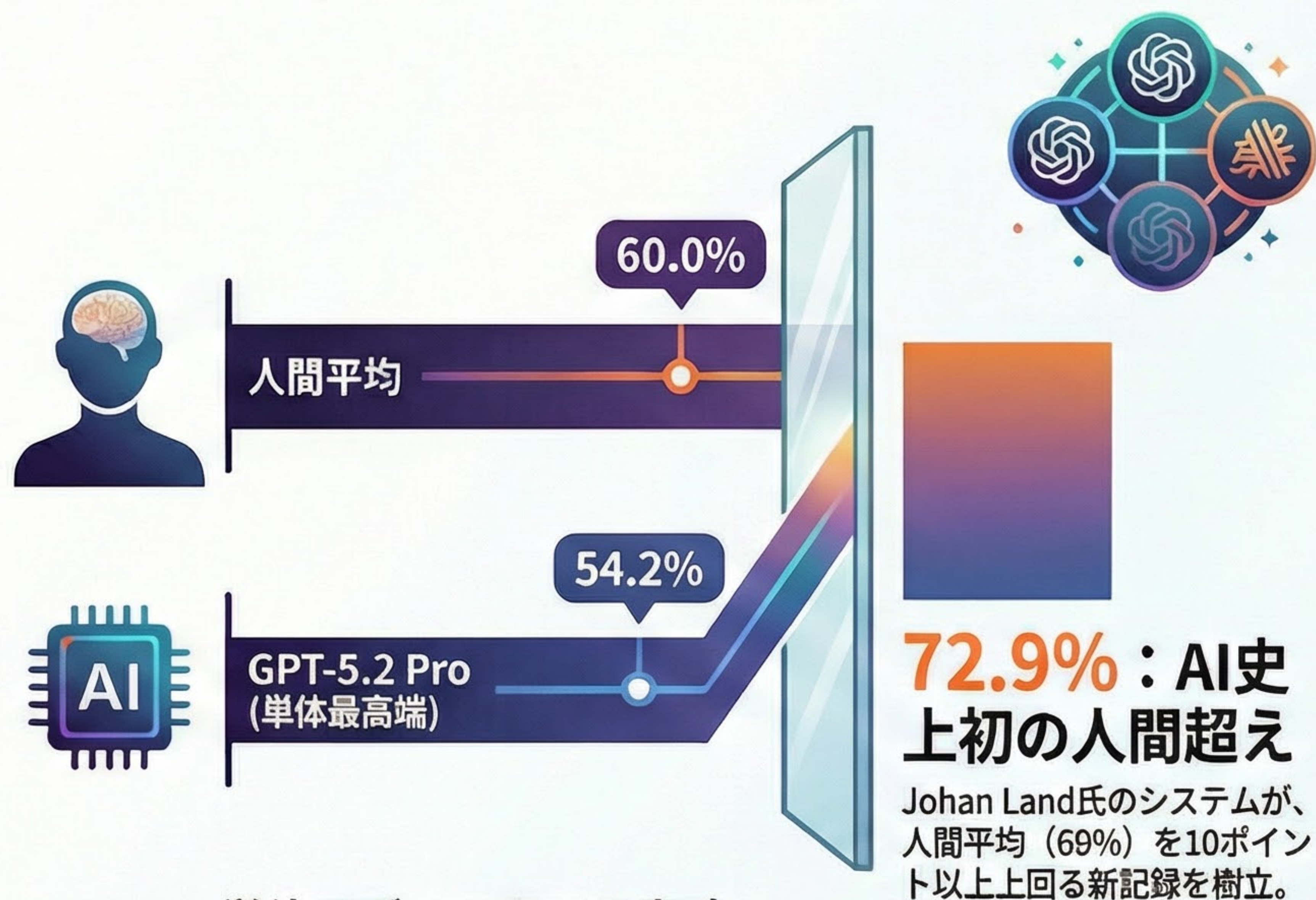


ARC-AGI-2：AIが「人間の推論力」を突破した瞬間

ARC-AGI-2は、AIの「流動性知能（未知の課題への対応力）」を測る最難関ベンチマークです。2025年末、Johan Land氏が複数のAIモデルを組み合わせた手法により、人間の平均正解率60%を大幅に上回る72.9%を達成し、AIの推論能力に革命をもたらしました。

驚異の性能比較：人間を凌駕したAI



単体モデルの限界を突破

GPT-5.2単体では54.2%に留まるが、複合システムにより推論の壁を破壊。

突破の鍵「多モデル反映的推論」の仕組み



最強モデル軍団の「審議会」

GPT-5.2、Gemini-3、Claude 4.5を連携させ、互いに回答を精査する合議制を採用。

10万回の試行錯誤と自己反省

Pythonコードを自動生成・実行し、失敗から学ぶループを1問につき平均6時間継続。



AIがAIを設計するメタ推論

ソルバー自体のコードも、AI（Gemini-3）との対話を通じて自動生成された。