中国製LLMはなぜARC-AGI-2で苦戦するのか？：欧米モデルとの「知能」の戦略的乖離

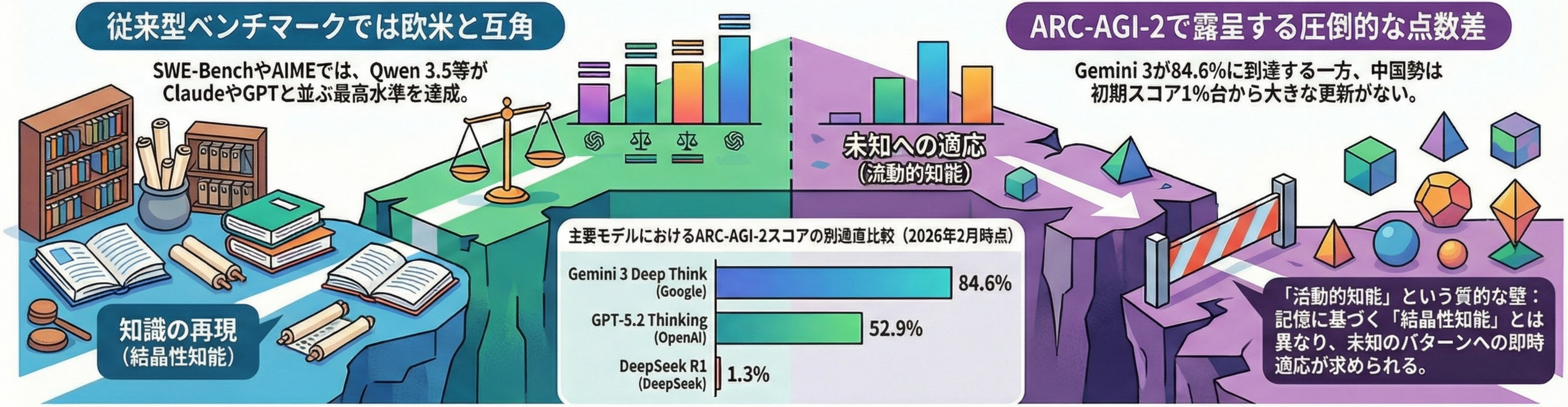
スコアの乖離：知識の再現 vs. 未知への適応

従来型ベンチマークでは欧米と互角

SWE-BenchやAIMEでは、Qwen 3.5等がClaudeやGPTと並ぶ最高水準を達成。

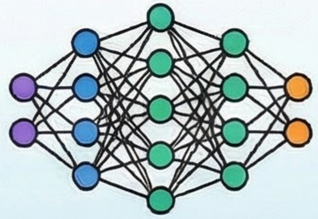
ARC-AGI-2で露呈する圧倒的な点数差

Gemini 3が84.6%に到達する一方、中国勢は初期スコア1%台から大きな更新がない。

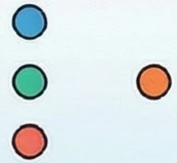


中国製LLMが低スコアに留まる4つの要因

テスト時計算（TTC）への投資格差



欧米勢：推論時に「長く考える」仕組みに注力



中国勢：コスト効率を優先

戦略的優先順位とコスト競争力

商用差別化・実用的指標

半導体規制によるハードウェア制約

NVIDIA製GPUへのアクセス制限が、推論時に大量の計算資源を要する手法の障壁に。