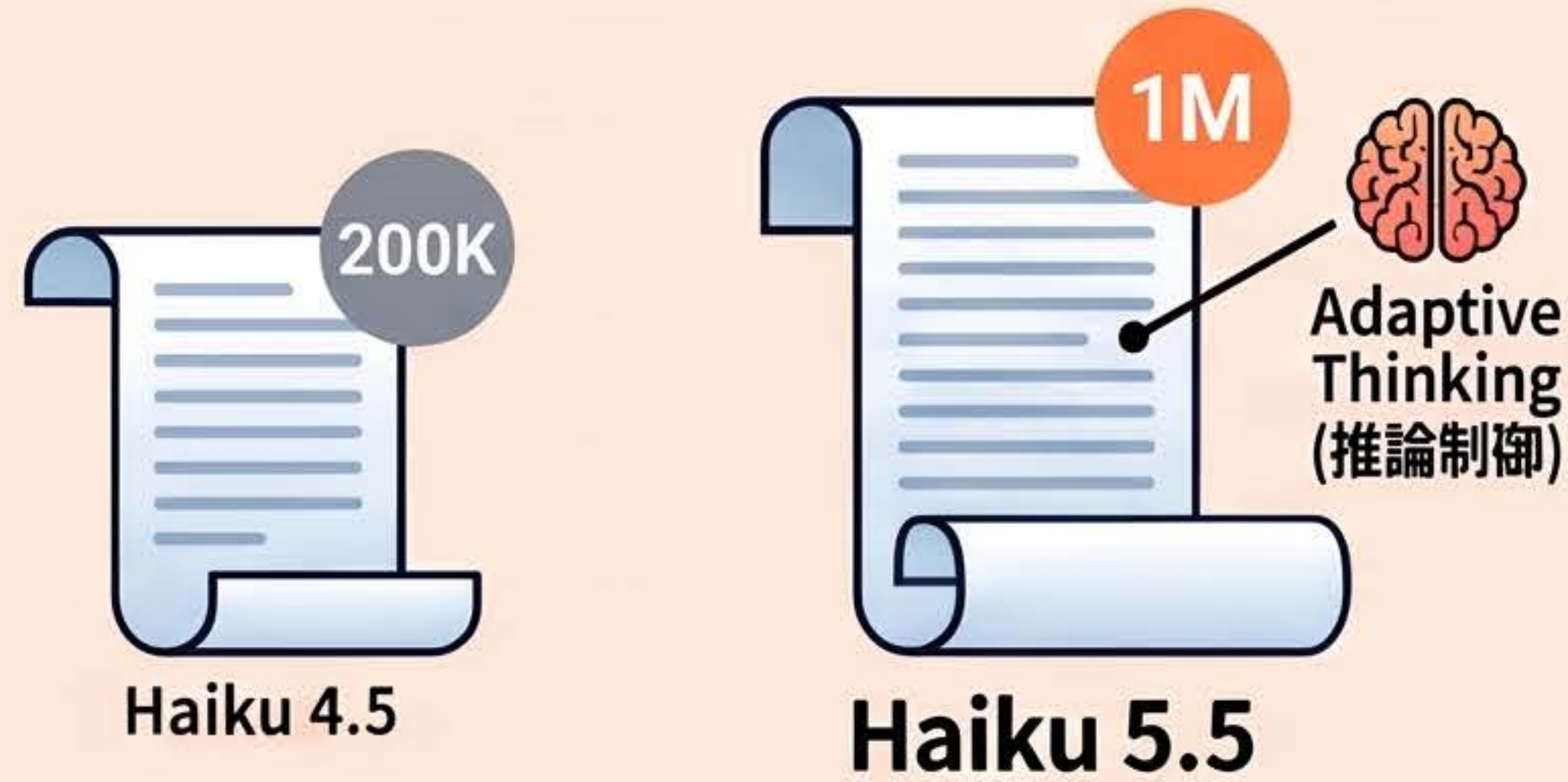


Claude Haiku 5.5：次世代AIの「Worker Tier」へ進化した小型モデルの真価

性能・機能の圧倒的アップデート

コンテキスト長が1Mトークンへと大幅拡張



実践的ベンチマークで前世代を凌駕

OSWorld (PC操作精度)

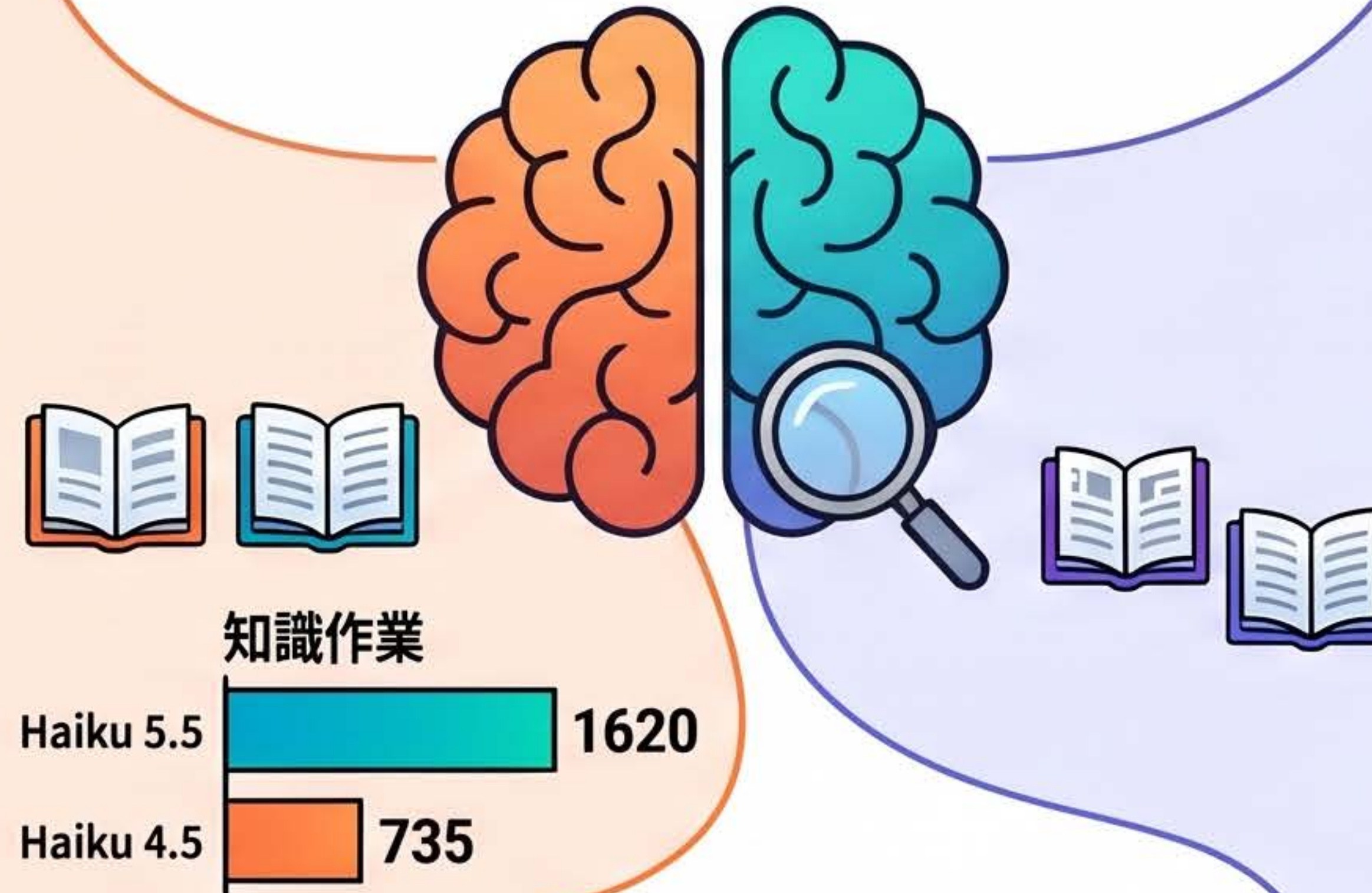


PC操作 (OSWorld) で72.4% (旧15.7%)、知識作業で1620 (旧735) を記録。

Browser & Computer Useに新規対応



Anthropicが発表したClaude Haiku 5.5は、前世代 (Haiku 4.5) から処理能力とコンテキスト長 (1M) を飛躍的に向上させつつ、標準利用で90%のコスト削減を実現。複雑な処理を行うフラグシップモデルの配下で、大量の並列タスクを実行する「Worker Tier」として設計されています。



	Haiku 4.5	Haiku 5.5	GPT-6 Luna
Context / Max Output	200K 64K	1M 128K	1.05M 128K
OSWorld (PC操作精度)	15.7%	72.4%	48.9%
標準入力/出力価格 (1M tokens)	\$1.00 \$5.00	\$0.10 \$0.50	\$0.10 \$0.50

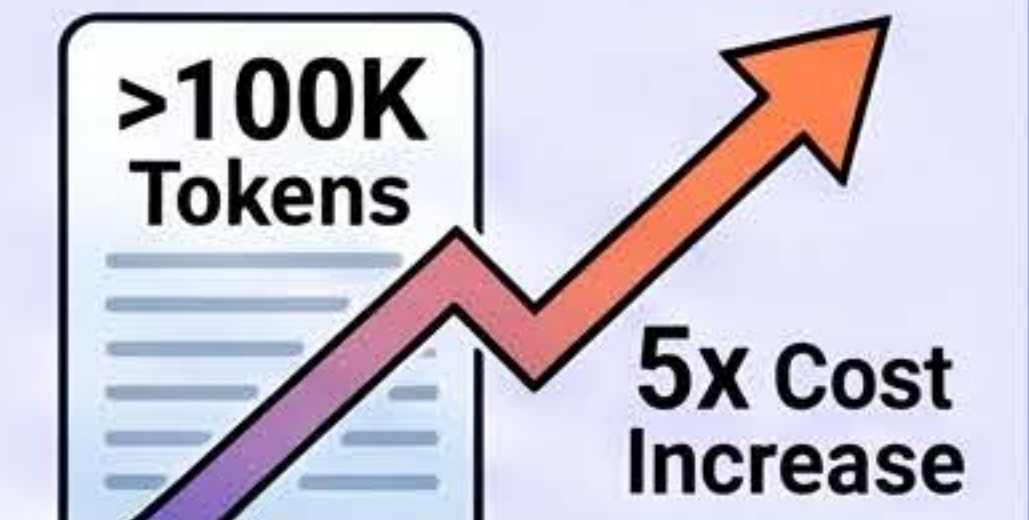
経済性と「Worker Tier」の役割分担

100K以下のプロンプト費用を90%削減



100Kトークン超えて単価が5倍に上昇

長文利用 (RAG等) ではGPT-6 Luna とのTCO (総保有コスト) 比較が必要。



司令塔 (Opus/Sonnet) と実動隊 (Haiku) の階層化

高度な計画・レビューと大量のサブエージェント処理を分相させるアーキテクチャが推奨。



司令塔 (Opus / Sonnet) 高度な計画・レビュー

