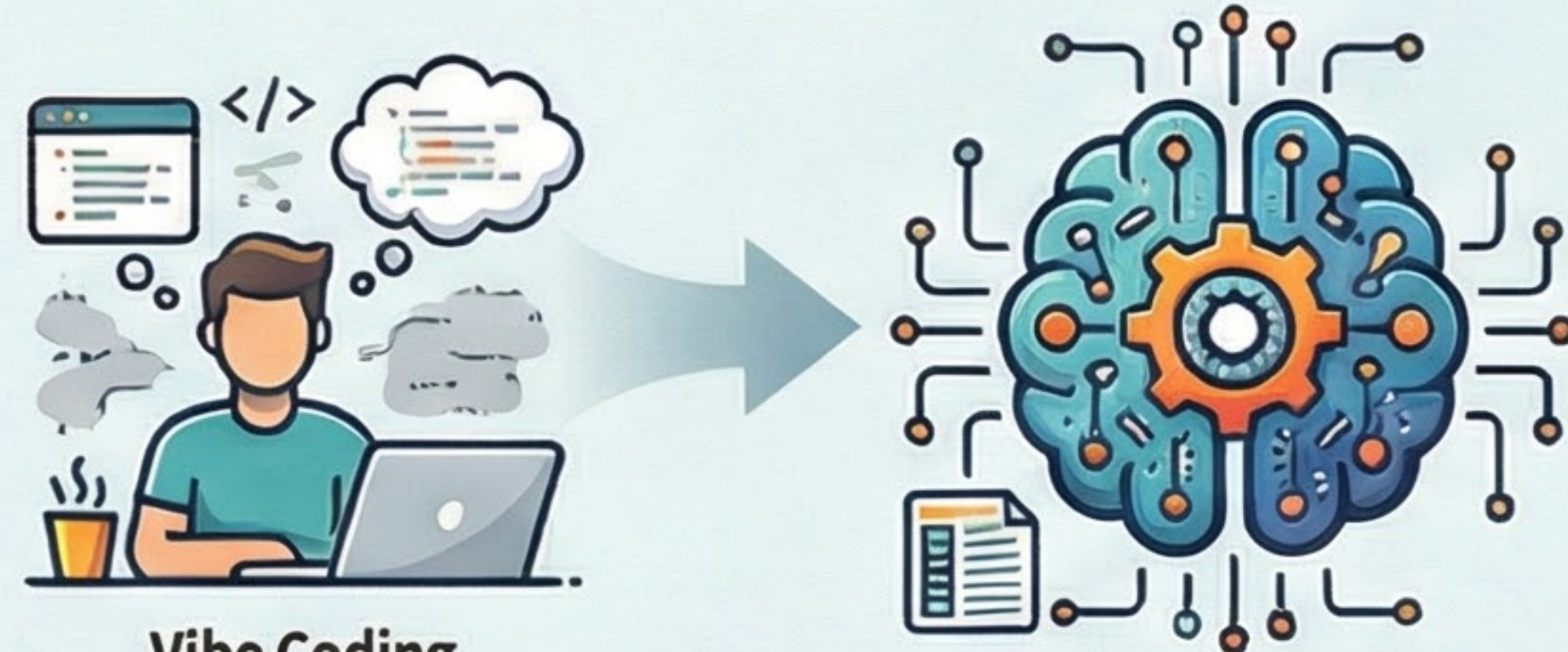


GLM-5：チャットから「自律的エンジニアリング」へのパラダイムシフト

Agentic Engineering（自律的エンジニアリング）



Vibe Coding
(雰囲気コード)

Agentic Engineering

「Vibe Coding」から「Agentic Engineering」へ
雰囲気コードを書く経験を卒業し、AIが過剰的に計画を立て、
ツールを駆使して活動するアプリやドキュメントを基礎生成。

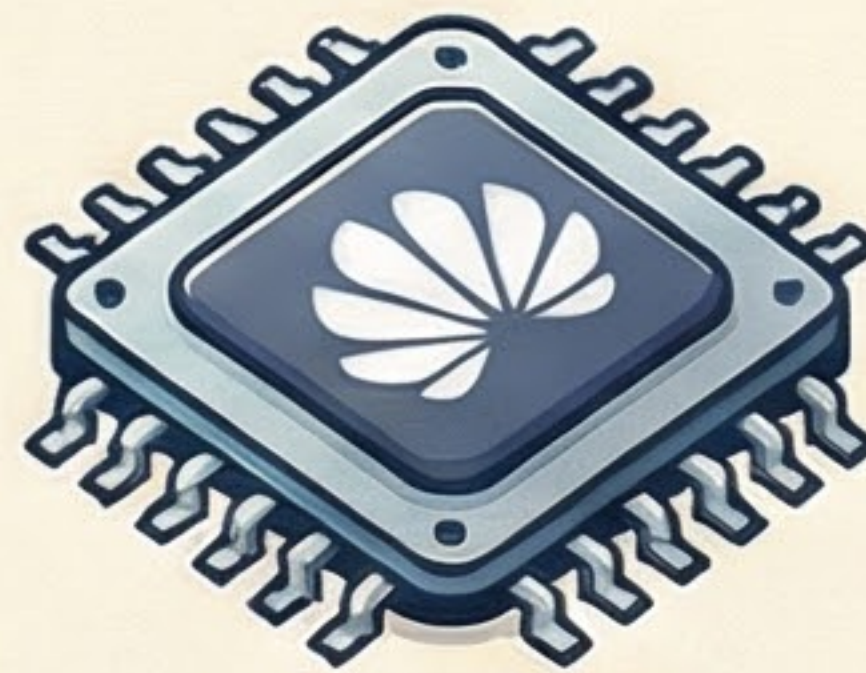


自律的なデバッグ・ループの実現
エラーが発生しても、AIが自らスタック
トレースを読み、修正と再実行を繰り
返してタスクを遂行する「粘り強さ」。



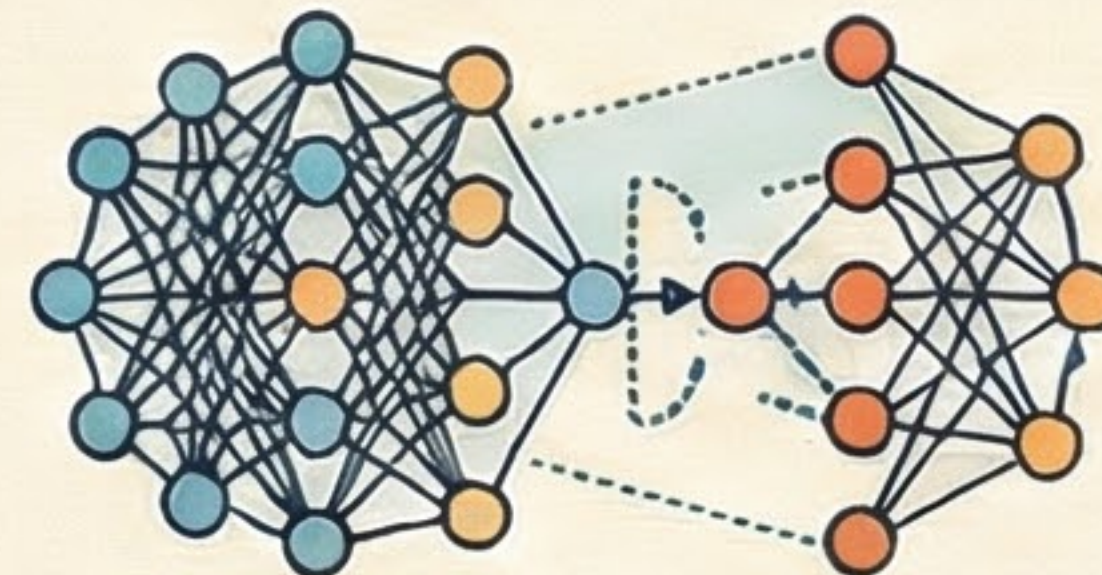
実務に即した最終成果物の直接出力
テキスト回答だけでなく、レイアウトや配色
が整ったWord、Excel、PDF、さらには完全
なソースコードリポジトリを直接生成可能。

技術的イノベーションと独立性



国産チップ（Huawei Ascend）による
完全自立

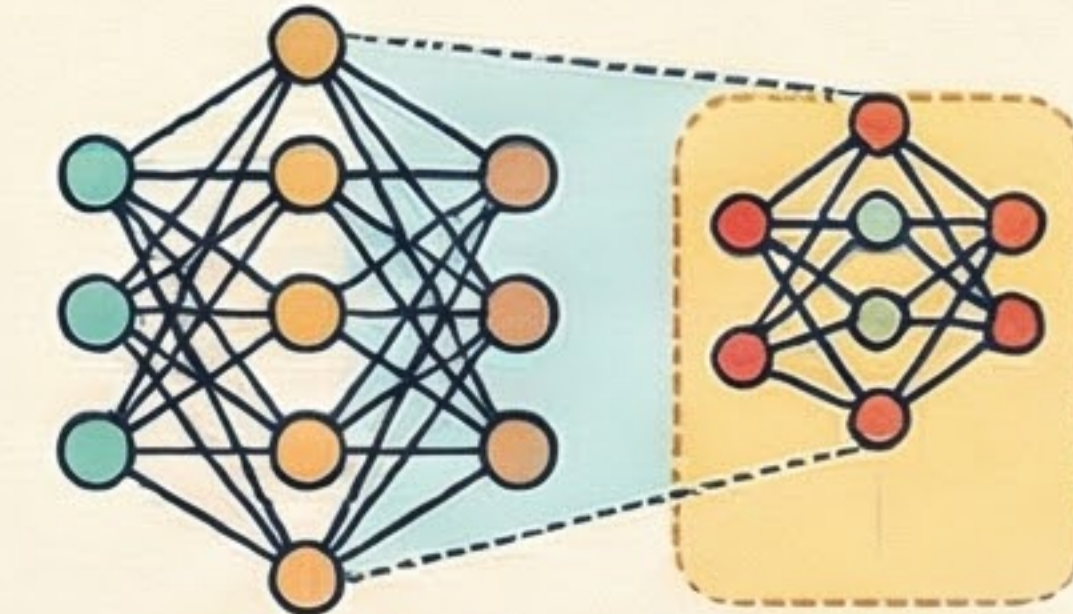
NVIDIA製GPUに依存せず、HuaweiのAscend
910とMindSporeフレームワークのみでトレー
ニングを完了し、技術的デカップリングを証明。



745B巨大モデルと効率性の両立
7450億パラメータを持ちながら、推論時
に動くのは13440億（MoE採用）のみ。
これにより高知能と低コストを同時に実現。



MITライセンスによるオープンウェイト公開
モデルの基盤が公開されており、全額がプライ
ベートクラウドやオンプレミス環境で機密デ
ータを保護しながら運用可能。



745B巨大モデル 440B MoE（推論時）

745B巨大モデルと効率性の両立
7450億パラメータを持ちながら、推論時に
動くのは440種（MoE採用）のみ。これにより
高知能と低コストを同時に実現。

主要モデル比較

モデル名	入力コル (G/M)	SWE- bench	アラウジ能力
Actor (指論)	\$0.11 (推定)	77.8%	Learner (学習)
Anthropic Claude Opus 4.5	\$5.00	80.9%	-

非同期強化学習フレームワーク「slime」
建築を行う「Actor」と学習を行う「Learner」
を分離・非同期化することで、膨大な実行履歴
データを高速に処理し、エージェント能力を両立。

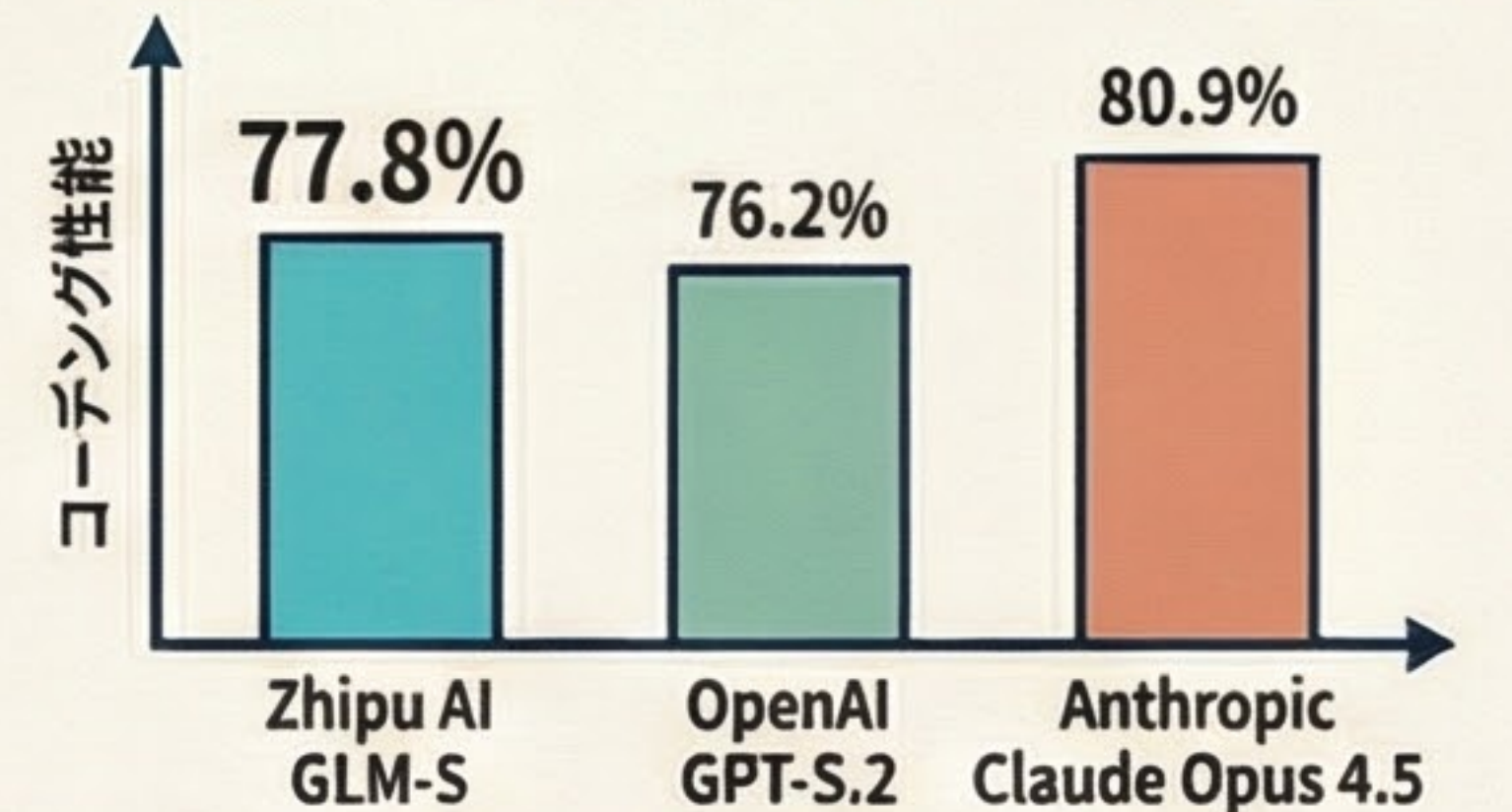
エコシステムと今後の課題



既存ツール（Cursor, Cline等）との統合
Z Code IDEだけでなく、既存の人気AIコー
ディングツールからバックエンドとしてGLM-5
を呼び出すことが可能。

性能ベンチマークと経済的破壊力

SWE-bench Verified（コーディング性能）



SWE-bench Verifiedで77.8%を記録
ソフトウェアエンジニアリングの実験能力を測るテスト
でGPT-5.2（76.2%）を上回り、世界最高峰のエン
지니어リング能力を証明。



GPT-5.2の約16分の1という破壊的価格
100万トークンあたりの入力知識は約0.11ドル。Claude Opus
4.5（5ドル）と比較すると約45分の1という圧倒的な安さ。



日本語ローカル知識には補完が必要
論議的思考は極めて高い一方、日本の地方の観
光地機能などにはハルシネーションが見られる
ため、BAG（検索補完）との併用が重要。