

Baiduの次世代AI戦略：ERNIE 5.0とERNIE-4.5-VLの深層分析

発行日: 2025年11月16日 作成者: Manus AI

1. はじめに

2025年11月、中国のテクノロジー大手Baidu（百度）は、年次イベント「Baidu World 2025」において、次世代基盤モデル「ERNIE 5.0」を発表し、AI業界に大きな衝撃を与えました。この発表のわずか数日前には、オープンソースの高性能マルチモーダルモデル「ERNIE-4.5-VL-28B-A3B-Thinking」を公開しており、Baiduが商用とオープンソースの両輪で市場を攻略する「二刀流戦略」を本格化させたことが明らかになりました。本レポートでは、並列処理による深層調査で得られた情報に基づき、これら二つのモデルの技術的詳細、性能、そしてBaiduの包括的なAI戦略を分析し、その将来性を展望します。

2. ERNIE 5.0: 商用フラッグシップモデルの全貌

ERNIE 5.0は、Baiduが「ネイティブ・オムニモーダル」と位置づける、商用のフラッグシップ基盤モデルです。テキスト、画像、音声、動画といった多様なモダリティを、単一の統一されたアーキテクチャで処理する能力を持ち、特に複雑なエンタープライズデータの解析において競合を凌駕することを目指しています¹。

2.1. 技術アーキテクチャ

ERNIE 5.0の核となるのは、超疎混合専門家（Ultra-Sparse Mixture-of-Experts, MoE）アーキテクチャです。総パラメータ数は2.4兆を超えると報告されていますが、推論時に実際に計算に使用されるアクティブパラメータは全体の3%未満（約720億）に抑えられています²。この超疎性により、モデル全体の巨大な知識保持能力を維持しつつ、推論時の計算負荷とコストを劇的に削減することに成功しています。Baiduは、このアーキテクチャにより推論性能が2倍向上し、コストが半減したと主張しています²。

また、本モデルはBaiduが開発したオープンソースのディープラーニングフレームワークであるPaddlePaddle上で構築されており、モデル開発から実運用までの一貫したエコシステムが提供されています²。

特徴	詳細
総パラメータ数	2.4兆超
アーキテクチャ	超疎混合専門家 (Ultra-Sparse MoE)
アクティブパラメータ率	3%未満 (約720億)

モダリティ	ネイティブ・オムニモーダル(テキスト, 画像, 音声, 動画)
フレームワーク	PaddlePaddle

2.2. 性能とベンチマーク

ERNIE 5.0は、40以上の権威あるベンチマークで、OpenAIのGPT-5やGoogleのGemini 2.5 Proといった世界のトップモデルと競合する性能を達成したと主張されています³。特に、文書認識(OCR Bench)、視覚的質問応答(DocVQA)、チャート解釈(ChartQA)といったタスクで優れたスコアを記録しています⁴。また、ユーザー投票による評価サイトLM Arenaのテキスト性能ランキングでは、レビュー版である「ERNIE-5.0-Preview-1022」が世界第2位タイにランクインし、その高い言語能力を証明しました⁵。

2.3. 提供形態とビジネスモデル

ERNIE 5.0は、Baidu AI CloudのMaaS(Model as a Service) プラットフォーム「Qianfan(千帆)」を通じて、APIサービスとして提供されます⁶。価格体系は、コンテキスト長と入出力のトークン数に応じた従量課金制が採用されており、大企業から中小企業まで、幅広い顧客層が利用できる設計となっています⁷。

3. ERNIE-4.5-VL: オープンソースの戦略的価値

ERNIE 5.0の発表に先立ち公開されたERNIE-4.5-VL-28B-A3B-Thinkingは、BaiduのAI戦略におけるもう一つの重要な柱であるオープンソースモデルです。Apache 2.0ライセンスの下で公開され、商用利用も可能です⁸。

3.1. 技術的特徴と効率性

このモデルもMoEアーキテクチャを採用しており、総パラメータ数280億に対し、推論時にアクティブになるのはわずか30億です⁸。この設計により、**単一の80GB GPU**という比較的アクセスしやすいハードウェア環境での動作を可能にしています⁹。これにより、高性能なマルチモーダルAIの導入ハードルを大幅に下げ、開発者コミュニティやコスト効率を重視する企業への普及を狙っています。

特徴	詳細
総パラメータ数	280億 (28B)
アクティブパラメータ数	30億 (3B)
アーキテクチャ	混合専門家(MoE)

ハードウェア要件	単一の80GB GPUで動作可能
ライセンス	Apache 2.0 (商用利用可)

3.2. 性能とベンチマーク

ERNIE-4.5-VLは、その軽量さにもかかわらず、多くのマルチモーダルベンチマークで競合のフラッグシップモデルを上回る性能を示しています。特に、チャート解釈（ChartQA）や視覚的数学推論（MathVista）といったタスクで、Gemini 2.5 ProやGPT-5-Highを凌駕するスコアを記録しており、その高い推論能力が際立っています¹⁰。また、画像の詳細を検査する「Thinking with Images」機能や、高度な強化学習戦略（GSPO + IcePop）の採用も、その性能を支える重要な技術的革新です¹¹。

4. Baiduの二刀流戦略と市場への影響

Baiduが展開する商用フラッグシップと高性能オープンソースモデルの「二刀流戦略」は、AI市場における包括的なリーダーシップを確立するための巧みなアプローチです。

- **高収益市場の確保:** ERNIE 5.0は、その圧倒的な性能とエンタープライズ対応機能により、金融、製造、医療といった高付加価値市場のニーズに応え、Baidu AI Cloudの収益基盤を強化します¹。
- **エコシステムの拡大:** ERNIE-4.5-VLは、そのアクセシビリティと高いコストパフォーマンスにより、広範な開発者や中小企業を取り込みます。これにより、ERNIE技術のデファクトスタンダード化を促進し、将来的にQianfanプラットフォームへの誘導路となることが期待されます¹²。

この戦略により、Baiduは技術革新の最前線を走り続けると同時に、AI技術の民主化を推進することで、市場全体における影響力を最大化しようとしています。中国国内では、ByteDanceの「Douyin」と「Douyin」などの激しい競争に直面していますが、技術的優位性とコスト競争力を武器に、グローバル市場での存在感を高めていくことが予測されます¹³。

5. 結論

BaiduによるERNIE 5.0とERNIE-4.5-VLの発表は、同社が単なる検索エンジン企業から、チップ、モデル、アプリケーションを垂直統合するフルスタックAI企業へと完全に移行したことを示すものです。商用モデルで市場の頂点を狙い、オープンソースモデルで市場の裾野を広げる二刀流戦略は、今後のAI業界における競争のあり方に大きな影響を与えるでしょう。特に、超大規模モデルの性能と、MoEアーキテクチャによる運用効率を両立させた点は、AI技術の実用化を加速させる重要なマイルストーンと言えます。Baiduの今後の動向は、グローバルなAIエコシステムの発展を占う上で、引き続き注視していく必要があります。

参考文献

- [1] PR Newswire. (2025, November 13). Baidu Unveils ERNIE 5.0 and a Series of AI Applications at Baidu World 2025, Ramps Up Global Push.
- [2] weel. (2025, November 14). Baiduが次世代基盤モデル「ERNIE 5.0」を発表！ .
- [3] VentureBeat. (2025, November 13). Baidu unveils proprietary ERNIE 5 beating GPT-5 performance on charts.
- [4] i10x.ai. (2025, November 14). Baidu ERNIE Benchmarks: Challenging GPT-4o and Gemini.
- [5] ERNIE Blog. (2025, November 7). ERNIE-5.0-Preview-1022, ready for testing in LM Arena!.
- [6] Baidu Cloud. 模型列表 - 百度千帆 · 大模型服务及Agent开发平台.
- [7] Baidu Cloud. 价格 - 千帆AI应用开发者中心.
- [8] Hugging Face. baidu/ERNIE-4.5-VL-28B-A3B-Thinking.
- [9] Artificial Intelligence News. (2025, November 12). Baidu ERNIE multimodal AI beats GPT and Gemini in benchmarks.
- [10] dev.to. (2025, November 12). In-Depth Analysis of ERNIE-4.5-VL-28B-A3B-Thinking Multimodal AI Model.
- [11] ERNIE Blog. (2025, November 11). ERNIE-4.5-VL-28B-A3B-Thinking: A Breakthrough in Multimodal AI.
- [12] Dera.ai. (2025, November 14). Baidu's Two-Pronged AI Strategy: ERNIE 5.0 and Open Source.
- [13] DIGITIMES. (2025, January 15). ByteDance's Doubao surpasses Baidu's Ernie Bot in China's chatbot market.