

BaiduのERNIE 5.0と ERNIE-4.5-VL-28B-A3B-Thinkingの詳細分析

背景

Baiduは2025年11月13日の年次イベント **Baidu World 2025** で、新世代の基盤モデル **ERNIE 5.0** を発表しました。これは同社が掲げる“**ネイティブ・オムニモーダル**”AI構想の中核であり、テキスト、画像、音声、動画を初期段階から統合的に扱うアーキテクチャとされます ¹。一方で、発表の数日前（11月11日）には Apache 2.0ライセンスで利用できる **ERNIE-4.5-VL-28B-A3B-Thinking** をオープンソースで公開しました ²。

このレポートでは、両モデルの設計思想、技術的特徴、利用方法、ベンチマーク結果、採用面の利点と課題を整理し、Baiduの二刀流戦略の意義を検討します。

ERNIE 5.0 – ネイティブ・オムニモーダル基盤モデル

発表の概要

- **発表日:** 2025年11月13日（Baidu World 2025）。
- **位置付け:** ERNIEシリーズの最新基盤モデルであり、“**natively omni-modal**（ネイティブ・オムニモーダル）”を謳う。従来のマルチモーダルモデルが後から別々のエンコーダーを統合する“late fusion”方式であるのに対し、ERNIE 5.0はテキスト・画像・音声・動画を最初から共同でモデリングする設計を採用 ¹。
- **規模:** 複数の報道によれば総パラメータ数は約**2.4兆**に達し、命令追従・創造的文章生成・事実に基づく推論・計画立案・ツール利用などの能力が大きく向上したとされる ³。

主な特徴と能力

特徴	解説	出典
ネイティブ・オムニモーダルアーキテクチャ	テキスト、画像、音声、動画を基礎レベルで共同モデリングし、全モダリティに渡る理解と生成を実現する ⁴ 。従来の後期融合方式よりも深い文脈理解が可能とされる ¹ 。	¹ ⁴
高いベンチマーク性能	OCRBench、DocVQA、ChartQAなどの文書・チャート理解ベンチマークで GPT-5 HighやGemini 2.5 Proに匹敵あるいは上回るスコアを記録したと発表 ⁵ ⁶ 。ただし多くはBaidu自身の測定であり独立検証が求められている ⁷ 。	⁵ ⁷
マルチモーダル生成	画像生成ではGoogle Veo3と同等以上の品質を示し ⁸ 、音声認識や音声からのQAでも競争力を持つ ⁹ 。	⁹
応用指向の強化	ビジネス用途を意識したOCR・文書理解・チャート解析での性能が強調され、企業向け自動化に直結する ¹⁰ 。	¹⁰
エージェント機能	計画策定やツール呼び出しなど、AIエージェントとしての実行能力を備える。Baiduはこれを「知的主体」への進化と位置付ける ¹¹ 。	¹¹

特徴	解説	出典
API提供と価格	商用モデルとしてQianfan（千帆）クラウド経由で提供される ¹² 。入力1万トークン当たり約0.85ドル、出力約3.40ドルと競合モデルよりやや安価（VentureBeatの比較表による） ¹³ ¹⁴ 。	¹³ ¹⁴

イベントで披露された関連製品と戦略

Baidu World 2025では、ERNIE 5.0を核とした多彩なAIエコシステムも発表されました。

- **逆ピラミッド構造:** CEOのロビン・リー氏はAIの価値構造が“逆ピラミッド”に変わり、アプリケーションが基盤モデルの100倍の価値を生むと述べました¹⁵。AIを内面化し業務に統合することでインテリジェンスをコストではなく生産性の源泉に変えるという哲学です¹⁶。
- **GenFlow 3.0:** 汎用AIエージェントGenFlowはバージョン3.0にアップグレードされ、ユーザー数は2,000万人を超えました¹⁷。マルチモーダル能力と記憶機能が強化され、複雑なタスクの自動化に対応します。
- **Famou:** 世界初の商用自己進化型エージェントとして発表され、複雑な問題を抽象化・反復し最適解を自律的に探索できると説明されました¹⁸。
- **MeDo/Oreate:** ノーコード開発プラットフォームMeDoとAIワークスペースOreateが世界展開され、開発者が専門知識なしでアプリケーションを構築できるようになっています¹⁹。
- **Kunlun M100/M300チップ:** AI推論に特化したKunlun M100（2026年発売予定）と超大型モデル向けKunlun M300（2027年予定）が発表されました。これにより自給自足のAIインフラを構築しようとしています²⁰。

課題と批判

- **独立した検証不足:** 多くのベンチマーク結果はBaidu自身の発表に依存しており、第三者評価の必要性が指摘されています⁷²¹。
- **バグ報告:** SVG生成時に意図せずツールを呼び出すバグが報告されており⁷、開発者からのフィードバックに基づき修正が進められています²²。
- **言語バイアス:** Chineseデータへの最適化が強く、英語や他言語での性能は未知数との指摘もあります²³。
- **トレーニング環境:** パラメータ数が極めて大きく、推論時のコストやデプロイメント要件が不透明です。独立評価や長文処理への対応（コンテキストウィンドウは128K～131Kトークンと報じられる²⁴²⁵）が企業導入の鍵となります。

ERNIE-4.5-VL-28B-A3B-Thinking – 軽量オープンソースモデル

概要

- **発表日:** 2025年11月11日。
- **ライセンス:** Apache 2.0 ライセンスで公開され、商用利用を含め制限がほとんどない²。
- **基本アーキテクチャ:** ERNIE 4.5 ファミリーの一員で、28億（28B）の総パラメータを持つ Mixture-of-Experts (MoE) アーキテクチャを採用。入力ごとに約3億（3B）のパラメータのみをアクティブにするため、大規模モデルの性能を維持しながら計算コストを劇的に削減している²⁶²⁷。
- **ハードウェア要件:** 単一の 80GB GPU で動作可能²⁸²⁹。大規模GPUクラスタを必要とする競合モデルに比べ、導入コストが大幅に低減する。
- **文脈長:** Baiduのドキュメントでは 128Kトークン前後のコンテキストウィンドウを備えると記載されている²⁴²⁵。

技術的特徴

1. MoEアーキテクチャによる高効率推論

2. ERNIE-4.5-VL-28B-A3B-Thinkingは複数の“専門家ネットワーク”を備え、入力に応じて適切な専門家を選択的に活性化する仕組みを持つ²⁶。これにより **総パラメータ280億のうち約30億のみが推論時に利用** され、計算負荷が大幅に削減される³⁰。

3. VentureBeatも同様の構造を説明し、選択的ルーティングにより少ない計算資源で大きなモデルに匹敵する性能を発揮すると報じている³¹。

4. “Thinking with Images” 機能

5. このモデルには、人間のように画像を自由にズームイン／ズームアウトしながら詳細を分析する **Thinking with Images** 機能が搭載されている³²³³。静的解像度で画像を処理する従来の視覚言語モデルと異なり、複雑な技術図面の解析や製造業における微細な欠陥検出など多様な応用が可能となる³²。

6. Hugging Faceのモデルカードでも、画像ズームや画像検索ツールと組み合わせることで細粒度情報へのアクセスを高めると説明している³⁴。

7. 強化学習と安定化技術

8. Baiduは大規模な視覚と言語データで**中間トレーニング**を行い、GSPOやIcePopといった戦略を組み合わせたマルチモーダル強化学習を導入している³⁵³⁶。これによりMoEトレーニングを安定化し、効率的な学習を実現した。

9. 主要能力

能力	内容	出典
視覚推論	多段階の推論やチャート解析、因果推論など複雑な視覚タスクで卓越した性能を示す ³⁷ 。	37
STEM推論	画像から問題を解き、橋回路の等価抵抗計算などの科学・工学タスクを処理できる ³⁸ 。	38
視覚的グラウンディング	言語指示に基づいて画像内の対象物を検出し、バウンディングボックスの座標をJSON形式で返す ³⁹ 。	39
ツール利用	画像検索やズームツールなど外部ツールを適切に呼び出して長尾知識を取得する能力を備える ⁴⁰ 。	40
動画理解	時間的認識とイベント局在化能力があり、動画の異なる時間セグメントでの内容変化を識別する ⁴¹ 。	41

1. 開発者向けエコシステム

2. モデルはHugging Faceで公開され、`transformers` ライブラリやvLLM、高速デプロイツールFastDeployに対応している⁴²。

3. BaiduはERNIEKitという産業向けトレーニング／圧縮ツールキットも提供し、企業が独自の最適化や量子化を行えるようにしている ⁴³。

ベンチマークと採用面での議論

- VentureBeatによると、文書理解やチャート分析においてGemini 2.5 ProやGPT-5-Highを上回る性能を示すとBaiduは主張しているが、独立検証はまだ行われていない ⁴⁴。性能は特定タスクに限られる可能性があり、創造的な視覚生成やリアルタイム動画解析では優位性が不明との指摘もある ²¹。
- モデルのコンテキストウィンドウは128Kトークンと大きいものの、非常に長い技術マニュアルや長時間の動画には不足する可能性があり、敵対的入力や分布外データへの耐性についても情報が限られている ⁴⁵ ²⁵。
- 80GB GPUは中規模企業でも導入可能だが、一般的なクラウドGPUより高価であり、特にGPU基盤を持たない組織ではクラウドサービスの利用が必要となり継続的なコストが発生する ⁴⁶。
- Apache 2.0ライセンスでの提供により、競合他社のクローズドモデルと比べて自由度が高い。ライセンス費用や収益分配が不要なため、商用利用の障壁が低くなる ⁴⁷ ⁴⁸。

政策的・市場的背景

- 2025年初頭にDeepSeekが高性能オープンソースモデルを発表した“DeepSeekショック”をきっかけに、中国企業はオープンソース戦略へ舵を切り始めました ⁴⁹。BaiduのCEO Robin Li氏も従来のクローズドソース中心の姿勢を転換し、ERNIE 4.5シリーズを6月からオープンソース化すると宣言 ⁵⁰。
- アメリカの大手AI企業もオープンソース戦略に追従し、OpenAIは2025年8月に5年ぶりのオープンソースモデルを公開した ⁵¹。こうした競争環境が、ERNIE-4.5-VL-28B-A3B-Thinkingの公開につながっています。

ERNIE 5.0とERNIE-4.5-VL-28B-A3B-Thinkingの比較

項目	ERNIE 5.0	ERNIE-4.5-VL-28B-A3B-Thinking
発表日	2025年11月13日 ⁵²	2025年11月11日 ⁵³
ライセンス	プロプライエタリ（商用）；Qianfan経由のAPI提供 ¹²	Apache 2.0オープンソース; Hugging Faceで公開 ⁵⁴ ⁴⁷
パラメータ数	約2.4兆パラメータで、全モダリティを共同モデリング ³	総パラメータ280億のうち約30億のみが推論時にアクティブ（MoE構造） ²⁶ ²⁷
アーキテクチャ	ネイティブ・オムニモーダル; モダリティ統合型の単一モデル ¹	Mixture-of-Experts; 入力毎に最適な「専門家」サブネットを選択し計算を削減 ²⁶
主な特徴	オムニモーダル統合、2.4兆パラメータ、長文対応（～131Kトークン）、高性能エージェント機能	Thinking with Images、強化学習による高精度推論、128Kコンテキスト、シングル80GB GPU動作
提供形態	API経由 (ERNIE Bot / Qianfan); 料金はGPT-5.1より低め ¹⁴	モデルとコードを公開; Transformers、vLLM、FastDeployに対応 ⁴²
対象ユーザー	大企業や特定用途で最高性能を求める顧客; 総合的なAIエコシステムの中核	開発者コミュニティや中規模企業; 自社環境でコスト効率良く活用したい組織

項目	ERNIE 5.0	ERNIE-4.5-VL-28B-A3B-Thinking
ベンチ マーク 評価	文書理解やチャート解析で GPT-5-High/Gemini 2.5 Proと同等か 優位と主張 ⁵⁵ ⁵ ; 独立検証不足	同様に文書理解やチャート分析で上回ると主張 ⁵⁶ ⁴⁴ ; 独立検証不足

二刀流戦略の意義

Baiduは、**プロプライエタリモデル** と **オープンソースモデル** の双方を展開することで、異なるユーザー層を取り込もうとしています。XenoSpectrumはこの戦略を「二刀流」と評し、最先端性能を求める大企業にはプレミアムなERNIE 5.0を、コスト効率とカスタマイズ性を求める開発者や中小企業にはオープンソースモデルを提供することで、市場を広く網羅しようとしていると分析します ⁵⁷。このアプローチにより、Baiduは国内外のAIエコシステムにおける影響力を拡大し、独自の半導体・アプリケーション群と組み合わせた垂直統合型の競争力を高めています。

結論

- **ERNIE 5.0** はBaiduのフラッグシップモデルとして、統合型のオムニモーダル処理能力と大規模パラメータを武器に、企業向け文書処理や視覚解析、AIエージェントなど幅広い用途での活用を目指しています。2.4兆パラメータによる性能向上とエージェント機能は魅力的ですが、独立したベンチマーク検証やバグ修正、言語的偏りへの対処が今後の課題です。
- **ERNIE-4.5-VL-28B-A3B-Thinking** は、MoEアーキテクチャで効率的な推論を実現した軽量モデルです。80GB GPUで動作し、Open Source licenceにより自由度が高く、開発者や中規模組織にとって魅力的な選択肢となります。Thinking with Imagesや強化学習による複雑な視覚推論など、高度な機能を備える一方、長文・大容量データに対する制約や独立検証の不足が課題です。
- Baiduの二刀流戦略は、**高性能とオープン性**の両立を通じてグローバル市場での存在感を高める試みであり、中国勢と米国勢が競い合う生成AI競争の中で重要な位置を占めます。今後は、透明性の高い第三者評価とユーザーからのフィードバックに基づく改善が、両モデルの信頼性と採用拡大の鍵となるでしょう。



1 3 5 8 9 12 15 17 18 19 52 57 Baidu、GPT-5超えを謳う「ERNIE 5.0」発表。真のオムニモーダルAIが拓く未来 | XenoSpectrum

<https://xenospectrum.com/baidu-ernie-5-omnimodal-ai-beats-gpt5/>

2 21 24 26 28 30 32 45 49 50 51 53 54 56 Baidu ERNIE-4.5新モデル公開：280億パラメータで Gemini 2.5 Proを上回る性能

<https://innovatopia.jp/ai/ai-news/71466/>

4 16 Baidu Unveils ERNIE 5.0 and a Series of AI Applications at Baidu World 2025, Ramps Up Global Push

<https://www.prnewswire.com/news-releases/baidu-unveils-ernie-5-0-and-a-series-of-ai-applications-at-baidu-world-2025--ramps-up-global-push-302614531.html>

6 7 10 11 55 「Baidu ERNIE 5.0」登場 | 新時代マルチモーダルAIの最新ベンチマーク

<https://innovatopia.jp/ai/ai-news/71772/>

13 14 22 Baidu unveils proprietary ERNIE 5 beating GPT-5 performance on charts, document understanding and more | VentureBeat

<https://venturebeat.com/ai/baidu-unveils-proprietary-ernie-5-beating-gpt-5-performance-on-charts>

20 23 Baidu launches new generation of Ernie AI | InfoWorld

<https://www.infoworld.com/article/4089778/baidu-launches-new-generation-of-ernie-ai.html>

25 27 29 31 33 36 42 43 44 46 47 48 Baidu just dropped an open-source multimodal AI that it claims beats GPT-5 and Gemini | VentureBeat

<https://venturebeat.com/ai/baidu-just-dropped-an-open-source-multimodal-ai-that-it-claims-beats-gpt-5>

34 35 37 38 39 40 41 ERNIE-4.5-VL-28B-A3B-Thinking: A Breakthrough in Multimodal AI | ERNIE Blog

<https://yiyan.baidu.com/blog/posts/ernie-4.5-vl-28b-a3b-thinking/>