

DeepSeek-V4.1-Flash：軽量・低価格マルチモーダルAIの衝撃

2026年9月発表：画期的な技術仕様、圧倒的な低コスト、および実務上の利用リスク



技術的革新とベンチマーク性能

新アーキテクチャ「CED」と「CSA2」



1Mトークンのコンテキストとネイティブ画像理解



外付けではない「真のマルチモーダル」により、テキストと画像を同時に事前学習。

エージェント/コーディングで米トップ級に肉薄



導入時の留意点

導入時に確認すべき「2つの制約」

⚠️ 中国サーバーによるデータリスク



⚠️ 画像API入力時の解像度制限
(実質800x800相当)



キャッシュヒット
入力価格を
約60%引き下げ

100万トークンあたり約
0.45円～ (90.003) という
桁違いの低価格を実現。

API価格の大幅な改定
(100万トークン、オフピーク)

	旧 V4 Flash	V4.1-Flash (新)
キャッシュヒット 入力	\$0.007	↓ \$0.003 (約60%減)
キャッシュミス 入力	\$0.22	\$0.15
出力	\$0.66	\$0.60

DeepSeek-V4.1-Flashは、独自のアーキテクチャによりKVキャッシュを劇的に圧縮し、ネイティブ・マルチモーダル対応を果たした最新のMoEモデルです。