

# 非対称な上昇: DeepSeek-V3.2およびV3.2-Specialeにおける技術的・戦略的変曲点の包括的分析

## Gemini 3 pro

### 1. 序論: 巨鯨の帰還とAI開発競争のパラダイムシフト

2025年12月1日、中国のAI研究機関であるDeepSeek(深度求索)は、同社の最新モデル群である「DeepSeek-V3.2」およびその高演算バリアント「DeepSeek-V3.2-Speciale」を公開しました<sup>1</sup>。このリリースは、単なるバージョンアップにとどまらず、生成AI業界における勢力図を塗り替える可能性を秘めた、極めて戦略的な一手として位置づけられます。

#### 1.1 背景: 2025年後半のAI情勢と「沈黙」の意味

2025年のAI業界は、Googleの「Gemini 3」シリーズの展開と、OpenAIによる「GPT-5」ラインナップの本格稼働によって特徴づけられていました<sup>2</sup>。これらの米国勢によるフロンティアモデルは、数万基規模のNVIDIA H100/Blackwellクラスタを用いた大規模学習によって実現されており、「スケーリング則(Scaling Laws)」の正当性を証明するかのように、圧倒的な計算資源の優位性を見せつけていました。

一方、DeepSeekは2025年初頭に「R1」や「V3」といったモデルで世界的な注目を集めた後、数ヶ月間の沈黙を保っていました<sup>2</sup>。業界の一部では、米国の輸出規制強化によるハードウェア制約が限界に達したのではないか、あるいは「DeepSeek旋風」は一過性の現象に過ぎなかつたのではないかという懐疑的な見方も浮上していました。しかし、12月1日の発表は、この沈黙が停滞ではなく、次世代アーキテクチャへの移行に向けた助走期間であったことを証明しました。

#### 1.2 リリースの核心: オープンウェイトによるフロンティア性能の民主化

今回のリリースの最大の特徴は、GPT-5やGemini 3 Proといった世界最高峰のクローズドモデルと同等の性能を持つモデルを、オープンモデル(Open Weights)として無料で公開した点にあります<sup>3</sup>。

- **DeepSeek-V3.2(Base):** 日常的なタスクやエージェントワークフローに最適化された「バランス型」モデル。推論コストと性能の均衡を保つつつ、GPT-5レベルの汎用性能を提供します<sup>1</sup>。
- **DeepSeek-V3.2-Speciale:** 推論能力を極限まで高めた「特化型」モデル。数学や競技プログラミングにおいて、Gemini 3.0 Proに匹敵、あるいは凌駕する性能を発揮し、国際数学オリンピック(IMO)で金メダル相当のスコアを記録しました<sup>5</sup>。

この二極化されたモデル戦略は、計算資源に制約のある環境下で、いかにしてSOTA(State-of-the-Art)を達成するかという問い合わせに対するDeepSeekの回答であり、同時に西側のAI霸権

に対する強烈なアンチテーゼもあります。本レポートでは、これら両モデルの技術的特異性、ベンチマークにおける詳細な比較分析、そしてこれらが示唆する経済的・地政学的な含意について、徹底的に深掘りします。

## 2. 技術的アーキテクチャ: 制約が生んだ革新

DeepSeekの強みは、単にパラメータ数を増やすことではなく、アルゴリズムレベルでの効率化にあります。特に、V3.2シリーズで導入された「DeepSeek Sparse Attention(DSA)」と、洗練された強化学習フレームワークは、ハードウェアの制約を克服するための重要な技術的ブレイクスルーです。

### 2.1 DeepSeek Sparse Attention (DSA) の機構と優位性

大規模言語モデル(LLM)のコンテキストウインドウが拡大するにつれ、従来の「密(Dense)」なアテンションメカニズムは計算コストの壁に直面していました。標準的なアテンションの計算量は、シーケンス長  $L$  に対して二次関数的 ( $O(L^2)$ ) に増加するため、128kトークンを超えるような長文脈処理においては、計算リソースとメモリ帯域を圧迫します。

DeepSeek-V3.2で採用されたDSAは、この課題に対する根源的な解決策です<sup>5</sup>。

#### 2.1.1 階層的疎性(Hierarchical Sparsity)の導入

DSAは、アテンションの計算対象を動的に、かつ階層的に選択することで、計算量を劇的に削減します。

- 粗粒度の圧縮(Coarse-Grained Compression): 文脈全体をブロック単位で評価し、関連性の低いセグメントを計算から除外します。
- 細粒度の選択(Fine-Grained Selection): 選択されたブロック内で、さらに重要なトークンのみにアテンションを集中させます。
- 計算量の削減: この機構により、主要なアテンション計算の複雑性は  $O(L^2)$  から  $O(L \cdot k)$  (ここで  $k$  は選択されたトークン数であり、 $k \ll L$ ) へと低減されます<sup>7</sup>。

#### 2.1.2 ネイティブな学習(Natively Trained Sparse Attention)

従来の疎なアテンション手法の多くは、学習済みのモデルに対して事後的に適用されることが多く、精度低下を招く傾向がありました。しかし、DeepSeekのDSA(論文中ではNSA: Native Sparse Attentionとも記述)は、プレトレーニング段階からこの疎性を前提として学習されています<sup>8</sup>。これにより、モデルは「どの情報を無視すべきか」自体を学習し、情報の損失を最小限に抑えつつ、推論効率を最大化することに成功しています。

表 1: DSA導入による効率化の定量的評価

指標	改善率(対Dense Attention)	備考
----	-----------------------	----

	)	
学習時の前方パス速度	最大 9.0倍	64kコンテキスト時 <sup>10</sup>
学習時の後方パス速度	最大 6.0倍	64kコンテキスト時 <sup>10</sup>
推論時のデコーディング速度	最大 11.6倍	メモリアクセス量の削減に起因 <sup>10</sup>

この圧倒的な効率化こそが、DeepSeekがGPT-5クラスのモデルを低成本で提供し、さらに「Speciale」のような超長考モデルを実現できた技術的基盤です。

## 2.2 スケーラブルな強化学習(RL)フレームワーク

V3.2シリーズのもう一つの核となる技術は、推論能力を飛躍的に高めるための強化学習プロセスです。DeepSeekは、従来のマルチステージRL(SFT → Reward Model → PPO)を進化させ、推論(Reasoning)、エージェント(Agent)、およびアライメント(Alignment)の学習を単一のステージに統合した「Mixed RL Training」を採用しました<sup>7</sup>。

- **Group Relative Policy Optimization (GRPO):** V3.2では、主要なRLアルゴリズムとしてGRPOを採用しています。これは、グループ内での相対的な報酬に基づいてポリシーを最適化する手法であり、多様なドメインにまたがるタスクにおいても学習の安定性を保つことができます<sup>11</sup>。
- **ポストトレーニング計算量のスケーリング:** DeepSeekの研究チームは、プレトレーニングだけでなく、ポストトレーニング(事後学習)段階における計算リソースを大幅に拡張しました。具体的には、全計算予算の10%以上をRLステージに配分しており<sup>11</sup>、これによりモデルは膨大な試行錯誤を通じて、複雑な論理的推論やエージェントとしての振る舞いを獲得しています。
- **報酬設計の二重化:**
  - 客観的タスク(数学・コード): 正解・不正解が明確なタスクに対しては、ルールベースの報酬と、出力長に対するペナルティ、言語的一貫性への報酬を組み合わせた厳密な評価を行います。
  - 主観的タスク(一般会話): 正解が一つではないタスクに対しては、プロンプトごとに固有の評価基準を持つ生成的報酬モデル(Generative Reward Model)を使用し、柔軟な評価を実現しています<sup>7</sup>。

## 2.3 Mixture-of-Experts (MoE) アーキテクチャの継承と進化

V3.2は、前世代であるV3のアーキテクチャを踏襲しつつ、さらなる最適化が施されています。

- 総パラメータ数: 6,850億(685B)<sup>12</sup>。
- アクティブパラメータ数: トークンあたり約370億(37B)<sup>13</sup>。
- 専門家の粒度: 微細な粒度の専門家(Experts)を多数配置し、各トークン処理時に少数の専門家のみを活性化することで、知識の広さと推論の速さを両立しています。
- 負荷分散戦略: 補助的な損失関数(Auxiliary Loss)を使用せずに負荷分散を行う独自の戦略

を採用しており、これにより学習の安定性を損なうことなく、専門家間の利用率を均等化しています<sup>12</sup>。

---

### 3. モデルバリエーションと戦略的ポジショニング

DeepSeek-V3.2のリリース戦略における最大の特徴は、汎用性と専門性を明確に分離した「二刀流」のアプローチにあります。

#### 3.1 DeepSeek-V3.2 (Base): バランスの取れた万能選手

「DeepSeek-V3.2」は、V3.2-ExpおよびV3.1-Terminusの正当な後継として位置づけられています<sup>1</sup>。

- 目的: 日常的なタスク、チャットボット、および一般的なエージェントワークフローにおける「GPT-5レベル」のパフォーマンスの提供。
- 特徴: 思考(Thinking)とツール利用(Tool Use)の統合。従来のモデルでは、思考プロセスとツール呼び出しが分離していましたが、V3.2では思考の過程で動的にツールを呼び出し、その結果をフィードバックして思考を継続することが可能です<sup>1</sup>。
- 利用形態: アプリ、ウェブ、APIを通じて広範に提供されており、コストパフォーマンスと応答速度のバランスが重視されています。

#### 3.2 DeepSeek-V3.2-Special: 推論の怪物

一方、「DeepSeek-V3.2-Special」は、DeepSeekが「推論の限界(Boundaries of Reasoning)」に挑戦するために開発した野心的なモデルです<sup>1</sup>。

- 目的: 数学の定理証明、競技プログラミング、複雑な論理パズルなど、極めて高度な認知能力を要するタスクの解決。
- 特徴:
  - 超長考(Ultra-Long Thinking Chain): 最大128kトークンにも及ぶ思考連鎖(Chain of Thought)を生成することが可能で、人間が数時間～数日かけて考えるような問題を、深く、粘り強く検討します<sup>14</sup>。
  - ツール利用の制限: 初期リリース時点ではAPI経由でのツール呼び出しがサポートされておらず、純粋な論理推論能力のみで勝負する仕様となっています<sup>5</sup>。
  - 期間限定のAPI: 2025年12月15日までの期間限定でAPIエンドポイントが開放されており、研究コミュニティによるベンチマーク検証を目的としています<sup>1</sup>。これは、このモデルの運用コストが極めて高いことを示唆しており、DeepSeek側がリソース管理を慎重に行っていることが伺えます。

---

### 4. 「Thinking Mode」における認知パラダイムの転換

DeepSeek-V3.2シリーズ、特にSpecialにおける最大の差別化要因は、OpenAIのo1シリーズで普

及した「System 2」的な思考プロセスを、より透明性の高い形で実装した点にあります。

## 4.1 可視化された思考プロセス vs ブラックボックス

OpenAIのo1モデルは、推論プロセス(Chain of Thought)をユーザーから隠蔽しており、最終的な回答のみを提示します。これに対し、DeepSeekは思考プロセスをAPIの reasoning\_content フィールドとして明示的に返します<sup>15</sup>。

- 透明性: ユーザーや開発者は、モデルがどのように結論に至ったか、どこで躓いたか、どのように自己修正(Self-Correction)を行ったかを詳細に追跡できます。
- 教育的価値: 特に数学やコーディングの学習者にとって、正解だけでなく「解法に至る思考の道筋」自体が貴重な学習リソースとなります。
- UI/UXの課題: この仕様変更により、既存のチャットUI(Kilo CodeやOpen WebUIなど)の一部では、思考プロセスが表示されないという互換性の問題が発生しました<sup>16</sup>。これは、DeepSeekの仕様が業界標準の先を行っているがゆえの摩擦と言えます。

## 4.2 ツール利用と思考の融合 (Thinking in Tool-Use)

V3.2(Baseモデル)では、思考モード中にツールを利用する機能が実装されました。これはエージェントAIにとって革命的な進歩です<sup>1</sup>。

- 従来のReAct: 「思考 → 停止 → ツール実行 → 結果取得 → 思考再開」という断続的なプロセス。
- DeepSeekのアプローチ: 思考の流れ(Stream of Thought)の中にツール呼び出しが有機的に組み込まれています。モデルは「思考しながら」必要な情報を検索し、得られた情報を即座に検証し、仮説を修正します。これにより、試行錯誤の効率が飛躍的に向上し、幻覚(Hallucination)のリスクが低減されます<sup>18</sup>。

## 4.3 「Developer」ロールと利用制限

Specialeのチャットテンプレートには、検索エージェント専用の developer ロールが新たに追加されました<sup>5</sup>。しかし、公式APIではこのロールの使用が制限されています。

- これは、Specialeの強力な推論能力が悪用されるリスク(例: 高度なサイバー攻撃の自動化など)を懸念した安全措置である可能性が高いです。
- 一方で、オープンウェイト版(ローカル実行)ではこの制限を技術的に回避できる可能性があり、倫理的な利用に関する議論を呼ぶ可能性があります。

---

## 5. ベンチマーク分析: 世界最高峰との対峙

DeepSeek-V3.2-Specialeの性能は、特定のドメインにおいて、現在世界最強とされるGoogle Gemini 3.0 ProやOpenAI GPT-5 Highに匹敵、あるいは凌駕しています。

### 5.1 数学的推論: 金メダル級の衝撃

数学分野におけるDeepSeekの成果は、AI史に残るマイルストーンと言えます。

表 2: 主要な数学・競技プログラミングベンチマーク比較

ベンチマーク / コンテスト	DeepSeek V3.2-Speciale	GPT-5 High	Gemini 3.0 Pro	備考
IMO 2025 (国際数学オリンピック)	金メダル (35/42)	データなし	データなし	歴史的快挙 <sup>4</sup>
IOI 2025 (国際情報オリンピック)	金メダル (492/600)	データなし	データなし	アルゴリズム実装力の証明 <sup>4</sup>
AIME 2025 (Pass Rate)	96.0%	94.6%	95.0%	米国招待数学試験 <sup>4</sup>
HMMT 2025	99.2%	データなし	97.5%	ハーバード-MIT数学大会 <sup>4</sup>

- 分析: IMOでの金メダル獲得は、これまでGoogle DeepMindの「AlphaGeometry」のようなニューロシンボリックAI(推論特化型システム)の領域とされてきました。汎用LLMであるDeepSeekが単独でこの水準に達したことは、言語モデルが「パターンマッチング」を超えて、深い論理的推論と定理証明の能力を獲得しつつあることを示しています。

## 5.2 コーディングとソフトウェアエンジニアリング

コーディング領域では、純粋なアルゴリズム力と実務的なエンジニアリング能力で結果が分かれています。

- アルゴリズム (Codeforces/IOI): SpecialeはIOIで金メダルを獲得し、Codeforcesでも高いパーセンタイル(上位)を記録しています。これは、明確に定義された問題に対する解決能力が極めて高いことを示しています。
- 実務 (SWE-Verified): 実際のGitHubリポジトリのバグ修正を模したSWE-Verifiedベンチマークでは、Speciale(73.1%)はGemini 3.0 Pro(76.2%)やGPT-5(74.9%)にわずかに遅れをとっています<sup>4</sup>。
- 考察: この差は、Gemini 3が持つ超長文脈(1M+トークン)と、マルチモーダル理解、および広範なコードベースへのアクセス能力によるものと推測されます。Specialeは「深さ(論理)」では勝っていますが、「広さ(コンテキスト把握)」においては、まだ巨大なコンテキストウィンドウを持つGeminiに分がある可能性があります。

## 5.3 一般知識と限界

「Humanity's Last Exam (HLE)」のような、マルチモーダルかつ分野横断的な超難問ベンチマークにおいて、Specialeのスコアは30.6%にとどまり、Gemini 3 Proの37.7%と差が開いています<sup>19</sup>。これは、DeepSeekが(現時点では)テキストベースの推論に特化しており、Geminiのようなネイティブマルチモーダル能力(画像、動画、音声を同時に理解し推論する力)を持たないことに起因すると考えられます。

## 6. 経済性と地政学的含意:H800が生んだパラドックス

DeepSeekの躍進を語る上で避けて通れないのが、米中技術摩擦と、それに伴うハードウェア制約、そして圧倒的なコスト競争力です。

### 6.1 「コモドール・モーメント」:知能の価格破壊

かつてコモドール64がパーソナルコンピュータを大衆化させたように、DeepSeekは「フロンティア級の知能」をコモディティ化しようとしています。

表 3: 100万トークンあたりのAPI価格比較

モデル	入力 (Cache Hit)	入力 (Cache Miss)	出力
DeepSeek-V3.2	\$0.028	\$0.28	\$0.42
GPT-5.1	N/A	\$1.25	\$10.00
Gemini 3.0 Pro	N/A	\$2.00	\$12.00
Claude 3.5 Sonnet	N/A	\$3.00	\$15.00

出典:<sup>1</sup>

DeepSeekの価格設定は破壊的です。特に出力コスト(\$0.42)は、競合他社の約30分の1です。推論モデルは内部で大量の思考トークンを生成するため、出力コストがボトルネックになりますが、DeepSeekはこの障壁を取り払いました。これは、DSAによる計算効率化と、FP8(8ビット浮動小数点)精度での学習・推論最適化が、実用レベルでコスト削減に寄与していることを証明しています<sup>22</sup>。

### 6.2 H800 vs H100: 制約は発明の母か?

DeepSeekは、V3の学習に2,048基のNVIDIA H800 GPUを使用したと公表しています<sup>13</sup>。H800は、

米国の対中輸出規制に準拠するために帯域幅が大幅に制限されたチップです。

- 効率の逆説: H100を潤沢に使える米国のラボが「力技」のスケーリングに走る一方で、帯域幅に制限のあるDeepSeekは、通信ボトルネックを回避するためのアルゴリズム(MoEの通信オーバーラップ最適化など)を極限まで洗練させる必要がありました。
- 密輸疑惑: 西側の一部のアナリストは、「H800だけでこれほどの性能が出せるはずがない」として、H100の密輸を疑っています<sup>24</sup>。しかし、技術レポートは、彼らがハードウェアの制約をソフトウェア(DSAやアーキテクチャ設計)で克服したことを詳細に記述しており<sup>12</sup>、単なるハードウェアの隠匿ではなく、技術的ブレイクスルーの結果である可能性が高いと見る専門家も多いです<sup>25</sup>。
- コスト: 学習コストは約560万ドル(約8億円)と見積もられており<sup>13</sup>、これはGPT-4等の学習コスト(1億ドル規模と言われる)と比較して桁違いに安価です。

---

## 7. 実運用とデプロイメントの課題

技術的には輝かしいV3.2ですが、実際に導入・運用する上ではいくつかの課題も存在します。

### 7.1 APIの期間限定性と安定性

V3.2-SpecialeのAPI提供は2025年12月15日までと極めて短期間に設定されています<sup>1</sup>。これは、Specialeの超長考プロセスが膨大な計算リソースを消費するため、恒久的なサービスとして維持するにはまだコストやインフラの調整が必要であることを示唆しています。企業が本番環境でSpecialeを採用するには、ローカルデプロイか、将来の安定版APIを待つ必要があります。

### 7.2 ローカル運用のハードル

「オープンモデル」とはいえ、685Bパラメータのモデルをローカルで動かすには、エンタープライズ級のハードウェアが必要です。

- **VRAM要件:** FP8量子化を行っても、数百GB～TBクラスのVRAMが必要です。H100/H800の8枚～16枚構成のクラスタが必要となり、個人の愛好家や中小企業が手軽に動かせるものではありません<sup>12</sup>。
- **蒸留モデルへの期待:** コミュニティでは、この巨大モデルの能力を蒸留(Distillation)した小型モデル(7B～70Bクラス)の登場が強く待望されています。

---

## 8. 結論: AI霸権の行方

DeepSeek-V3.2およびV3.2-Specialeの登場は、2025年のAI業界における「ブラックスワン」イベントと言えるでしょう。

1. オープンソースの勝利: クローズドなフロンティアモデル(GPT-5, Gemini 3)と同等の性能を持つモデルが、MITライセンスで公開されたことは、AI技術の独占構造を根本から搖るがします<sup>27</sup>。

2. 中国AIの底力: 輸出規制という逆風の中で、ハードウェアの劣勢をアルゴリズムの革新で跳ね返した事実は、中国のAI研究開発能力の高さ(および弾力性)を世界に知らしめました。
3. 推論コストの崩壊: 圧倒的な低価格は、今後AIアプリケーションのビジネスモデルを再定義するでしょう。「考えるAI」はもはや高嶺の花ではなく、あらゆるアプリに組み込める標準機能となります。

DeepSeekは、「巨鯨」のロゴが示す通り、深く静かに潜行し、浮上するたびに大きな波を起こしています。V3.2-Specialeが示した「金メダル級の知性」が、今後どのように世界中の開発者の手によって応用され、新たなイノベーションを生み出していくのか。その動向から目を離すことはできません。

---

## 出典インデックス

- リリース・ニュース:<sup>1</sup>
- 技術・アーキテクチャ:<sup>5</sup>
- ベンチマーク:<sup>4</sup>
- 経済性・価格:<sup>1</sup>
- **Hugging Face / ウェイト:**<sup>5</sup>
- 思考モード・API:<sup>15</sup>
- 地政学・ハードウェア:<sup>24</sup>

## 引用文献

1. DeepSeek-V3.2 Release, 12月 3, 2025にアクセス、  
<https://api-docs.deepseek.com/news/news251201>
2. DeepSeek releases updated AI that matches Gemini 3 and ChatGPT 5, internet says Whale is back, 12月 3, 2025にアクセス、  
<https://www.indiatoday.in/technology/news/story/deepseek-releases-updated-ai-that-matches-gemini-3-and-chatgpt-5-internet-says-whale-is-back-2829475-2025-12-02>
3. DeepSeek just dropped a free GPT-5.1 rival, 12月 3, 2025にアクセス、  
<https://www.techradar.com/ai-platforms-assistants/gemini/deepseek-just-gave-a-way-an-ai-model-that-rivals-gpt-5-and-it-could-change-everything>
4. DeepSeek Launches V3.2 AI Models to Challenge GPT-5 and Gemini - ASO World, 12月 3, 2025にアクセス、  
<https://marketingtrending.asoworld.com/en/discover/deepseek-launches-v3-2-ai-models-to-challenge-gpt-5-and-gemini/>
5. deepseek-ai/DeepSeek-V3.2-Speciale - Hugging Face, 12月 3, 2025にアクセス、  
<https://huggingface.co/deepseek-ai/DeepSeek-V3.2-Speciale>
6. DeepSeek Releases New Reasoning Models to Take On ChatGPT and Gemini, 12月 3, 2025にアクセス、  
<https://www.cnet.com/tech/services-and-software/deepseek-releases-new-reasoning-models-to-take-on-chatgpt-and-gemini/>
7. DeepSeek-V3.2: Pushing the Frontier of Open Large Language Models - arXiv, 12

- 月 3, 2025にアクセス、<https://arxiv.org/html/2512.02556v1>
- 8. [2502.11089] Native Sparse Attention: Hardware-Aligned and Natively Trainable Sparse Attention - arXiv, 12月 3, 2025にアクセス、<https://arxiv.org/abs/2502.11089>
  - 9. Hardware-Aligned and Natively Trainable Sparse Attention - ACL Anthology, 12月 3, 2025にアクセス、<https://aclanthology.org/2025.acl-long.1126/>
  - 10. DeepSeek Native Sparse Attention: Advanced Attention Mechanism for LLMs | by Mehul Gupta | Data Science in Your Pocket | Medium, 12月 3, 2025にアクセス。  
<https://medium.com/data-science-in-your-pocket/deepseek-native-sparse-attention-advanced-attention-mechanism-for-langs-6ac68fc014ff>
  - 11. DeepSeek Researchers Introduce DeepSeek-V3.2 and DeepSeek-V3.2-Speciale for Long Context Reasoning and Agentic Workloads - MarkTechPost, 12月 3, 2025にアクセス。  
<https://www.marktechpost.com/2025/12/01/deepseek-researchers-introduce-deepseek-v3-2-and-deepseek-v3-2-speciale-for-long-context-reasoning-and-agentic-workloads/?amp>
  - 12. deepseek-ai/DeepSeek-V3 - GitHub, 12月 3, 2025にアクセス。  
<https://github.com/deepseek-ai/DeepSeek-V3>
  - 13. DeepSeek-V3 Technical Report - arXiv, 12月 3, 2025にアクセス。  
<https://arxiv.org/pdf/2412.19437.pdf>
  - 14. What is DeepSeek V3.2— and what official version's changed - CometAPI - All AI Models in One API, 12月 3, 2025にアクセス。  
<https://www.cometapi.com/what-is-deepseek-v3-2-and-what-s-changed/>
  - 15. DeepSeek Prompting Techniques: strategies, limits, best practices, etc - Data Studios, 12月 3, 2025にアクセス。  
<https://www.datastudios.org/post/deepseek-prompting-techniques-strategies-limits-best-practices-etc>
  - 16. reasoning/thinking content not displayed when using deepseek R1 through Litellm proxy · Issue #2770 · Kilo-Org/kilocode - GitHub, 12月 3, 2025にアクセス。  
<https://github.com/Kilo-Org/kilocode/issues/2770>
  - 17. Open WebUI Does Not Display Deepseek API Reasoning Content Due to Unsupported reasoning\_content Field #9431 - GitHub, 12月 3, 2025にアクセス。  
<https://github.com/open-webui/open-webui/discussions/9431>
  - 18. How DeepSeek V3.2's New Attention Makes Long Contexts Cheap | by Sai Dheeraj Gummadi | Dec, 2025 | Medium, 12月 3, 2025にアクセス。  
<https://medium.com/@gsaidheeraj/how-deepseek-v3-2s-new-attention-makes-long-contexts-cheap-46b3747d98a7>
  - 19. DeepSeek's New AI Model Achieves Gold-Level Results, Rivals GPT-5 and Gemini 3 Pro, 12月 3, 2025にアクセス。  
<https://beebom.com/deepseek-v3-2-achieves-gold-level-results-rivals-gpt-5-mini-3-pro/>
  - 20. DeepSeek surfaces again with open-source V3.2 release - The Rundown AI, 12月 3, 2025にアクセス、<https://www.therundown.ai/p/deepseek-strikes-again>
  - 21. AI Leaderboards 2025 - Compare LLM, TTS, STT, Video, Image & Embedding Models, 12月 3, 2025にアクセス、<https://llm-stats.com/>
  - 22. DeepSeek-V3.2-Exp Complete Analysis: 2025 AI Model Breakthrough and

In-Depth Analysis of Sparse Attention Technology - DEV Community, 12月 3, 2025にアクセス、

<https://dev.to/czmilo/deepseek-v32-exp-complete-analysis-2025-ai-model-break-through-and-in-depth-analysis-of-sparse-3gcl>

23. [2412.19437] DeepSeek-V3 Technical Report - arXiv, 12月 3, 2025にアクセス、

<https://arxiv.org/abs/2412.19437>

24. DeepSeek: A Deep Dive - CSIS, 12月 3, 2025にアクセス、

<https://www.csis.org/analysis/deepseek-deep-dive>

25. On DeepSeek and Export Controls - Dario Amodei, 12月 3, 2025にアクセス、

<https://www.darioamodei.com/post/on-deepseek-and-export-controls>

26. How to install DeepSeek V3.2 Speciale Setup: vLLM + Transformers -

Sonusahani.com, 12月 3, 2025にアクセス、

<https://sonusahani.com/blogs/deepseek-speciale>

27. deepseek-ai/DeepSeek-V3.2-Exp - Hugging Face, 12月 3, 2025にアクセス、

<https://huggingface.co/deepseek-ai/DeepSeek-V3.2-Exp>

28. Hardware-Aligned and Natively Trainable Sparse Attention - arXiv, 12月 3, 2025に

アクセス、<https://arxiv.org/pdf/2502.11089.pdf>

29. DeepSeek R1 vs OpenAI o1: Which One is Faster, Cheaper and Smarter? -

Analytics Vidhya, 12月 3, 2025にアクセス、

<https://www.analyticsvidhya.com/blog/2025/01/deepseek-r1-vs-openai-o1/>

30. Develop reasoning apps with DeepSeek models on Azure AI Foundry using the OpenAI SDK | Microsoft Learn, 12月 3, 2025にアクセス、

<https://learn.microsoft.com/en-us/azure/developer/ai/how-to/use-reasoning-model-inference>

31. DeepSeek's Lesson: America Needs Smarter Export Controls | RAND, 12月 3, 2025にアクセス、

<https://www.rand.org/pubs/commentary/2025/02/deepseeks-lesson-america-needs-smarter-export-controls.html>