

DeepSeek-V3.2およびDeepSeek-V3.2-Specialeの詳細調査

モデルアーキテクチャ

DeepSeek-V3.2シリーズは、大規模Mixture-of-Experts (MoE)構造を採用したTransformerモデルです。総パラメータ数は約6,850億に及びますが、各トークンの予測にはその一部（約370億規模のエキスパート）だけが活性化するため、効率的に計算できます¹。このMoEアーキテクチャにより、モデル全体の容量は巨大でありながら推論コストを抑えることに成功しています。さらにDeepSeek Sparse Attention (DSA)と呼ばれる長文最適化のスパース・アテンション機構を導入しており、最大128,000トークンもの長いコンテキストを扱えるのが特徴です²³。DSAでは低精度の軽量な「インデクサ」ネットワークが全文脈を走査し、各クエリに関連する上位 k 個のトークンだけに注意を集中させます。このハイブリッド戦略により、効果的に計算量を $O(L \times k)$ に削減しつつ、フルアテンション並みの性能を維持しています⁴。例えば推論時のコストは従来比で約70%削減されており、128kトークン（300ページ相当）の文脈を処理する場合でも現実的な計算資源で対応可能です⁵。またメモリ効率化のためヘッド間でキー・バリュー表現を共有する設計になっており、超長文でもメモリ使用量が抑制されています⁶。

DeepSeek-V3.2の基本構造はTransformerですが、上述のMoEによる専門化とDSAによる長文対応といった独自拡張が施されています。モデルはまず既存モデル（V3.1系）のチェックポイントから開始し、密な注意機構でウォームアップした後、段階的にスパース注意に切り替える2段階訓練で安定性を確保しています⁷。このときKLアライメント損失を併用してスパース注意が密な注意の振る舞いに近づくよう調整し、性能劣化を防いでいます⁸。以上の工夫により、従来のモデルと同等の「記憶力」や性能を保ったまま長文処理を実現しています⁷。

トレーニングデータとプロセス

DeepSeek-V3.2の事前学習（ベースモデル）は、前世代（V3シリーズ）の手法を踏襲し、大規模なWebページや電子書籍コーパスから次単語予測訓練を行っています⁹。ベースモデル訓練時には意図的な人工データの混入を避け、インターネット上から収集した自然文を用いています（ただしクロールデータ中に含まれるOpenAIモデル生成文は除去し切れない可能性があることが報告されています）⁹。このようにして基本的な言語知識を身につけた後、DeepSeek-V3.2は大規模な強化学習パイプラインによる追加訓練が行われています。

DeepSeek社は1,800以上の仮想環境と85,000件以上の複雑なタスクからなる膨大な合成データを生成し、モデルに「思考しながらツールを使う」能力を後付けしました¹⁰。具体的には、検索エンジンAPIやコード実行環境、ファイル操作など実際のツールを用いたシナリオをシミュレートし、旅行プラン最適化やソフトウェアのバグ修正、複数ステップの情報検索といった難解だが検証可能な課題を多数作成しています¹⁰。例えば「予算制約下での3日間の旅行計画」や「8言語にまたがるバグ修正」「数十回のウェブ検索を要する調査」など、多岐にわたる指令が含まれます¹⁰。これらのタスクには正解検証手段（単体テストや解答チェック）が組み込まれており、モデルが試行錯誤しながら解を導き出せるよう設計されています¹¹¹²。

まず領域別（数学・コーディング・推論）のスペシャリストモデルを訓練し、それらから知識を蒸留して汎用モデルに統合する専門家Distillationも実施されました¹³。特に数学領域ではDeepSeek独自のDeepSeek-Math-V2のデータセットと報酬モデルを組み込み、定理証明の能力向上が図られています¹⁴。最終的なDeepSeek-V3.2モデルは、人間による指示追従強化（RLHF）やポリシー最適化型のRLを数万ステッ

ブにわたり適用され、高度な推論力とエージェント行動を獲得しました¹⁵。この強化学習後の総追加学習トークン数は**数百億～数千億規模**に上り、事前学習計算量の**10%以上に相当する計算予算**がポストトレーニングに投入されています¹⁶。また**長い思考チェーン（思考プロンプト）**を扱うため、RL時に出力長のペナルティを下げて「長考」を促す工夫も行われています¹⁷。

以上のように、DeepSeek-V3.2は巨大コーパスによる基礎学習に加えて、大規模な合成タスク群での専門強化学習を経ることで、単なる言語モデルを超えた**高度な問題解決エージェント**へと仕上げられています。

ベンチマークスコアと評価指標

DeepSeek-V3.2シリーズは様々な評価ベンチマークで**最先端に匹敵する成績**を収めています。知識・推論分野の指標である**MMLU (Massive Multi-Task Language Understanding)**では約**85%前後**の正答率を達成し、これは同時期のGPT-5モデルとほぼ同等の水準です¹⁸。特に最新の拡張版**MMLU-Pro**では、V3.2が**85.0%**、Specialeは**90%前後**に達し、Gemini 3.0 Pro（約90%）と肩を並べる結果が示されています¹⁹²⁰。日常的な文章問題データセット（例えば**Humanity's Last Exam**テキスト版）でも、Specialeは約**37.7%**のPass@1を記録し、従来オープンモデルを大きく上回っています²¹。

数学推論に関して、競技会形式の難問で卓越した性能が報告されています。アメリカの数学競技**AIME 2025**では、**DeepSeek-V3.2-Specialeが96.0%の正答率**を達成し、GPT-5（94.6%）やGoogleのGemini-3.0-Pro（95.0%）を僅差で上回りました²²。また**HMMT 2025（ハーバード・MIT数学大会）**でもSpecialeは**99.2%**という突出した成績を収め、Gemini-3.0-Proの97.5%を凌駕しています²²。通常版のDeepSeek-V3.2も**AIMEで93.1%、HMMTで92.5%**と、最先端に迫るスコアを記録しています²³。これらの成績は、同モデルが**数理推論においてGPT-5と同等以上、場合によってはGemini級の能力を持つ**ことを示すエビデンスです。

さらに注目すべきは実際の**国際コンペティション**での成果です。DeepSeek-V3.2-Specialeは**2025年国際数学オリンピック(IMO)で42点満点中35点**を獲得して金メダルに相当する成績を収め、また**国際情報オリンピック(IOI)でも600点中492点**をマークして金メダルを獲得（全参加者中10位相当）しました²⁴。プログラミング世界大会の**ICPC 2025**でも提示された12問中10問に正解し、総合2位に相当する成績を残しています²⁴。これらはいずれも**インターネット非接続・ツール非使用**の条件下で、公平に人間参加者と同様の制限時間内で達成されたものです²⁵。

コード生成・デバッグ分野でも最先端級の性能が確認されています。例えばオープンソースのソフトウェア不具合修正ベンチマーク**SWE-Verified**では、通常版DeepSeek-V3.2が**73.1%**のバグ修正成功率を記録し、これはGPT-5-Highの74.9%とほぼ同等でした²⁶。またプログラミングの複雑なマルチステップタスクを評価する**Terminal-Bench 2.0**では、DeepSeek-V3.2が**46.4%**のスコアで**GPT-5-Highの35.2%を大きく上回**りました²⁶。一方、ウェブ検索やツール利用を伴う総合的エージェント評価（例：**Tool Decathlon**等）でも、DeepSeek-V3.2は**既存のGPTベースのモデルに匹敵または上回る水準**を示し、最難関タスクでGeminiやClaudeをわずかに下回る程度まで肉薄しています²⁷。総じて、**数学、コーディング、常識推論からツール使用まで幅広いベンチマークで、DeepSeek-V3.2は従来のオープンモデルの性能格差を埋め、トップクラスのクローズドモデルとの差をごく僅かなものに縮めています**²⁸²⁶。

用途・応用分野

DeepSeek-V3.2は汎用的な大規模言語モデルとして設計されており、チャットボット対話、質問応答、コード生成、長文要約など多岐にわたる用途に利用できます。特に**マルチラウンドの対話**や**長い履歴を含むチャット**に強みがあり、公式のチャットアプリやAPI経由で**ChatGPTのような会話エージェント**として利用可能です²⁹。最大128kトークンのコンテキスト長を活かし、**長大なドキュメントや書籍の解析・要約、大規模なコードベースの理解**といったタスクにも適しています²³⁰。たとえば「長大なPDFレポートを読んでまと

める」あるいは「巨大なコードリポジトリからバグの原因を推論する」といった場面でも、DeepSeek-V3.2はコンテキストを途切れさせず一度に処理できる強みがあります。

最大の特徴は、**エージェントとして外部ツールを駆使するタスク**に対するネイティブな対応です。DeepSeek-V3.2は**チェイン・オブ・ソート（思考の連鎖）**を内部で保持しつつ、途中で**コード実行・ウェブ検索・計算ツール**などを呼び出すことができます³¹³²。従来のモデルは外部ツールを呼ぶたびに思考過程がリセットされる問題がありましたが、DeepSeek-V3.2では**ツール使用中も思考の文脈（考えた過程）を保持**する独自アーキテクチャになっており、ツール呼び出し前後で一貫した推論が続けられます³³。これにより、「**考える（think）と実行する（act）を交互に行う複雑なマルチステップ問題**において、従来よりもスムーズで効率的な解決が可能です³⁴³⁵。典型的な例として、**対話型のコーディング支援**（コードを書いてコンパイル→エラーを踏まえてデバッグ）、**調査エージェント**（ユーザ質問→検索ツールで情報収集→まとめて回答）、**計画立案**（ユーザ要件→複数APIを順次呼び出し→最適プラン提案）など、**思考とツール利用を行き来するシナリオ**で威力を発揮します³²³⁶。マルチモーダル（画像・音声）入力への対応は公式には言及されていませんが、テキストベースでの**コードや構造化データの処理**、および上述の外部システム連携により、実質的に幅広い形式の情報を扱えます。現在、DeepSeek社の提供するウェブインターフェース「chat.deepseek.com」やAPIで、このモデルを用いたチャット対話やエージェント機能を試すことができます²⁹。

一方、**DeepSeek-V3.2-Speciale**は主に**研究用途や高度な問題解決専用**に位置づけられています³⁷。Specialeは後述するようにツール使用機能を持たないため、インタラクティブなエージェントというより**オフラインで与えられた難問に長時間熟考して答えを出す用途**に適しています。例えば数学競技の問題や高度な論証問題を解かせる、といったケースで威力を発揮します。通常のV3.2モデルよりも出力が冗長になる傾向があるため、**日常的なチャットボットや要約用途には向きません**³⁷。そのためDeepSeek社もSpecialeを「研究目的限定」と位置づけており、一般利用にはまず標準版V3.2を推奨しています³⁸。

ライセンスと利用条件

DeepSeek社はV3.2およびV3.2-Specialeの**モデル重みとコードをオープンソースで公開**しています。ライセンスはMIT Licenseであり、研究・商用を問わず**誰でも自由に利用・改変・再配布可能**な非常に寛容な条件となっています³⁹。実際、Hugging Face上で両モデルの重みがダウンロード可能であり、企業や開発者はこれを取得して自社インフラ上でホストしたり、独自にファインチューニングしてサービス展開することも許されています³⁹。利用に当たっての特別な制限（例えば商用利用禁止やクレジット表記義務）はMITライセンス上は存在せず、DeepSeek社も「**オープンな研究と商用利用を歓迎**」するとしています⁴⁰³⁹。

このオープンモデル戦略は競合他社と対照的です。OpenAIやAnthropicが最先端モデルを非公開・クラウドAPI限定としている中で、DeepSeekはフロンティア級のモデルを無償公開することで**AI業界のビジネスモデルに一石を投じている**との評価もあります⁴¹。誰もがGPT-5相当のモデルを入手できるため、専有モデルに高額なAPI料金を支払う必要がなくなり、開発コストの劇的な低下につながります⁴¹。その一方で、中国発の企業であるDeepSeekに対してはデータ管理や安全保障上の懸念から、欧米で規制の動きも報じられています（例：EU一般データ保護規則に絡みドイツ当局が懸念を表明⁴²）。しかしモデルそのものは公開されているため、技術的には規制による進展阻止は難しいとも言われています⁴³。

総じて、**DeepSeek-V3.2シリーズは完全オープンで商用利用可能なモデル**であり、必要な計算資源さえあれば誰でもこの強力なLLMを活用できる状況です。GitHub上で学習コードやドキュメントも公開されているため、実装の透明性も確保されています³⁹。このような開かれた姿勢はコミュニティからも歓迎されており、公開直後からHugging Faceで多数のダウンロードと高評価を獲得しています。

GPT-5・Gemini 3との性能比較

DeepSeek社は「V3.2はGPT-5と同等の性能、SpecialeはGPT-5を超えGemini-3.0-Proに匹敵する」と主張しています⁴⁴。この主張は前述のベンチマーク結果によって裏付けられています。例えば、数学競技のAIMEやHMMTでSpecialeがGPT-5やGeminiを上回る正答率を示したこと、コードバグ修正や対話型ターミナル操作でV3.2がGPT-5と肩を並べたことなどが具体的な根拠です²²²⁶。総合知識テストのMMLUでも、DeepSeek-V3.2はGPT-5とほぼ同等の正答率を記録し、Specialeは一部領域でGPT-5を上回りました¹⁹。また**推論の深さが要求される競技的タスク**（IMOやIOI）で金メダル級の成績を収めている点は、**Gemini 3.0 Pro級の推論力**を持つことの傍証と言えます⁴⁵。

具体的な比較として、DeepSeek-V3.2-Specialeは**Gemini-3.0-Proとほぼ同等の「考える力」**を持つと評価されています⁴⁶。Specialeは複雑な数学・論理問題でGeminiと互角に渡り合い、多くの場合**Claude 4.5やGPT-5など他の開源モデルを凌駕する結果**を示しました²⁷²²。一方、通常版のDeepSeek-V3.2（Thinkingモード有効）は**GPT-5-Highに匹敵する総合性能**を達成しており、日常的なタスクではGPT-4.5相当のClaude 2系やOpenAIのGPT-5に肉薄、あるいは同等との評価です⁴⁷⁴⁸。例えば先述の**ソフトウェアバグ修正**ではGPT-5と僅差、**対話型エージェントタスク**ではGPT系に勝るケースも報告されました²⁶。

ただし、DeepSeek自身も認めているように**出力効率（答案の簡潔さ）**ではまだ課題が残ります⁴⁹。Gemini-3.0-Proなどは短い解答で高品質な結果を出すのに対し、DeepSeek-V3.2は**同等の回答品質を得るためにより長い思考過程（トークン）を要する傾向**があります⁴⁹。つまり「**知識の密度**」で若干劣る可能性が指摘されています⁵⁰。とはいえ、これら差分は「もはや僅かな漸進的差異であり、かつてのような絶対的な性能格差ではない」ことが強調されています⁵¹。オープンモデルでありながらGPT-5やGeminiと同**リーグで肩を並べる性能**を示したDeepSeek-V3.2/Specialeは、2025年現在のLLM研究における一大マイルストーンと位置付けられています。

「V3.2」と「V3.2-Speciale」の主な違い

同時公開された**DeepSeek-V3.2（標準版）**と**DeepSeek-V3.2-Speciale**は、基本的なアーキテクチャやパラメータ数は共通ですが、**設計目的と調整方針が異なるバリエーション**です⁵²⁵³。以下に主な違いを整理します。

- **目的・位置付け**: V3.2はV3.2-Expの正式後継として位置づけられた**汎用モデル**で、日常利用に適した**バランス型**です²⁹⁵⁴。対してV3.2-Specialeは**高計算コスト版（high-compute variant）**と銘打たれ、**研究用途で推論力を極限まで追求した特化モデル**となっています⁵⁵。
- **出力の長さ・思考様式**: V3.2は**推論力と出力効率のバランス**を重視し、回答が冗長になりすぎないように**適度に最適化**されています⁵⁴。一方Specialeは「**考えること**」に**最大重きを置いて長めの思考チェーン**を許容しており、**長い推論過程（チェーン・オブ・ソート）**によって難問解決能力を高めています³⁷。つまりSpecialeでは**出力長に関する制約を緩和（ペナルティを低減）し、必要に応じいくらかでも深く推論させる設計**です³⁷。その結果、Specialeは回答までに要するトークン数が多くなりがちですが、その分**難解な問題で最高レベルの正答率**を狙えるようになっています⁴⁹。
- **外部ツール使用**: V3.2は**ツール呼び出しに対応**しており、内部で思考を保持しつつ検索やコード実行を行えるのが大きな特徴です⁵⁶。一方**Specialeはツール使用機能をサポートしていません**⁵⁷。Specialeは**深い内省的な推論（自己完結型の考察）**に特化するため、ツール呼び出しのようなインタラクションは訓練されておらず無効化されています⁵⁷。そのためSpecialeは与えられたテキスト入力だけで解答するスタイルで、たとえばWeb検索を伴う質問には適しません。
- **強化学習の調整**: 両者とも大規模RLを経ているですが、Specialeでは**数学定理証明データセットの追加導入や長考モードでの最終調整**が行われています¹⁴¹⁷。具体的には、DeepSeek-Math-V2由来の数学問題データと報酬設計を組み込み、Specialeに**高度な定理証明スキル**を付与しました¹⁴。また、RL最終段階でSpecialeは**思考プロセスを途中で打ち切らないよう設定**されており、そのぶん推論に時

間（トークン）を割くよう最適化されています¹⁷。標準版V3.2もRLによって高い推論力を得ていますが、Specialeはさらに計算資源を費やし理詰めで解を探す能力を尖らせた形です⁵⁵³⁷。

- **パラメータ数・モデルサイズ:** いずれも約6850億パラメータで、基本的なネットワーク構成（層数やMoEエキスパート数など）は同一です⁵²⁵⁸。Specialeだからといってパラメータが増えているわけではなく、差異は主に訓練内容と推論モードにあります。
- **性能と適応タスク:** V3.2（標準版）は「日常のGPT-5」と言える位置づけで、汎用的なQAやチャット、コーディング支援など幅広く高性能を発揮します⁵⁴。一方Specialeは「オープン界のGemini-Pro」とも称され、数学・アルゴリズム・論理パズルなど極めて困難な課題でこそ真価を発揮します⁵⁹。言い換えると、Specialeは難解な競技問題を解くことに特化し、日常的な雑談や簡単なタスクではその長考ぶりがむしろ冗長になる可能性があります³⁷。
- **提供形態:** オープンソースとしては両モデルともHugging Faceで重みが公開されています⁶⁰。ただ公式サービスとしては、V3.2標準版はDeepSeekのアプリ・Web・APIで常時利用可能なのに対し、Specialeは2025年12月15日までの期間限定APIエンドポイントとして提供されました⁶¹。Specialeは将来的に標準モデルへフィードバック統合される計画らしく、DeepSeek社は期間終了後にSpecialeの能力を標準V3.2に取り込む意向を示唆しています⁶²。

以上より、DeepSeek-V3.2とV3.2-Specialeは「万能型」と「究極特化型」という関係です。共通の巨大基盤モデルを土台としつつ、一方は実用上のバランスを取り、他方は考え得る限りの推論力を追求しています⁵³³⁷。ユーザーは用途に応じて使い分けことが想定されており、日常対話や一般的なタスクには軽量のV3.2を、研究的な難問にはSpecialeを、といった選択が可能です。両モデルの登場により、オープンソースコミュニティでもGPT-5やGeminiに匹敵する技術を自由に検証・応用できる時代が到来したと言えるでしょう³⁹⁴¹。

参考文献・情報源: DeepSeek公式サイトのニュースリリース⁶³⁶⁴、Hugging Faceモデルカード⁶⁵⁶⁶、DeepSeek技術報告書⁶⁷¹⁹、信頼できる報道（VentureBeat）²²³⁹ および有志による分析記事³⁴⁴⁷ など。これらにより、モデルの設計思想から評価結果まで確認しました。以上がDeepSeek-V3.2およびV3.2-Specialeに関する詳細調査結果です。

¹ ⁹ ¹⁴ ¹⁶ ³⁷ ³⁸ ⁴⁷ ⁵³ ⁵⁴ ⁵⁵ The Complete Guide to DeepSeek Models: V3, R1, V3.1, V3.2 and Beyond

<https://www.bentoml.com/blog/the-complete-guide-to-deepseek-models-from-v3-to-r1-and-beyond>

² ⁴ ⁶ ⁷ ⁸ ¹³ ¹⁵ ²⁷ ²⁸ ³⁰ ³¹ ³² ³⁴ ³⁵ ³⁶ ⁴⁸ ⁵¹ DeepSeek V3.2 : Bye Bye Gemini 3.0, Claude & GPT-5 | by Mehul Gupta | Data Science in Your Pocket | Dec, 2025 | Medium

<https://medium.com/data-science-in-your-pocket/deepseek-v3-2-bye-bye-gemini-3-0-claude-gpt-5-78b17bee76d8>

³ ⁵ ¹⁰ ²² ²³ ²⁴ ²⁵ ²⁶ ³³ ³⁹ ⁴¹ ⁴² ⁴³ ⁴⁹ ⁶² DeepSeek just dropped two insanely powerful AI models that rival GPT-5 and they're totally free | VentureBeat

<https://venturebeat.com/ai/deepseek-just-dropped-two-insanely-powerful-ai-models-that-rival-gpt-5-and>

¹¹ ¹² ¹⁷ ¹⁸ ¹⁹ ²⁰ ²¹ ⁵⁰ ⁶⁷ huggingface.co

<https://huggingface.co/deepseek-ai/DeepSeek-V3.2/resolve/main/assets/paper.pdf>

²⁹ ⁵⁶ ⁵⁷ ⁶⁰ ⁶¹ ⁶³ ⁶⁴ DeepSeek-V3.2 Release | DeepSeek API Docs

<https://api-docs.deepseek.com/news/news251201>

⁴⁰ deepseek-ai/DeepSeek-V3 - Hugging Face

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

⁴⁴ ⁴⁵ ⁴⁶ ⁵⁹ DeepSeek's New AI Models: V3.2 and V3.2-Speciale

<https://semiengineering.com/deepseeks-new-ai-models-v3-2-and-v3-2-speciale/>

52 58 65 66 deepseek-ai/DeepSeek-V3.2 · Hugging Face
<https://huggingface.co/deepseek-ai/DeepSeek-V3.2>