

日本の主要AIモデルにおける高度推論・数理工学的適応性に関する包括的研究報告書： 2026年版

Gemini 3 pro

1. 序論：生成AIにおける「推論」のパラダイムシフトと本報告書の目的

1.1 背景：System 1からSystem 2への進化的転回

2025年から2026年にかけて、人工知能(AI)研究、特に大規模言語モデル(LLM)の開発領域において、かつてない規模のパラダイムシフトが発生しました。それは、従来の「確率的なトークン予測による流暢な言語生成(System 1: 直感的思考)」から、「論理的整合性と多段階の検証を伴う深い推論(System 2: 熟考的思考)」への移行です。OpenAIの「o1」およびその後継である「o3」、あるいはDeepSeekの「R1」といったモデル群は、ユーザーからの問いに対して即座に回答を出力するのではなく、内部的に「思考の連鎖(Chain of Thought: CoT)」と呼ばれる中間推論プロセスを生成し、自己検証と自己修正を繰り返すことで、従来は不可能とされた難解な数学的証明、物理法則の適用、そして複雑なコーディングタスクにおいて、人間エキスパートを凌駕する性能を示し始めました¹。

この技術的潮流は、単なるチャットボットの性能向上にとどまらず、AIが「科学的発見(Scientific Discovery)」の主体となり得る可能性を示唆しています。数学における未解決問題の証明や、新材料の物理シミュレーションといった領域において、AIは補助ツールから自律的な探求者へと役割を変えつつあります⁴。

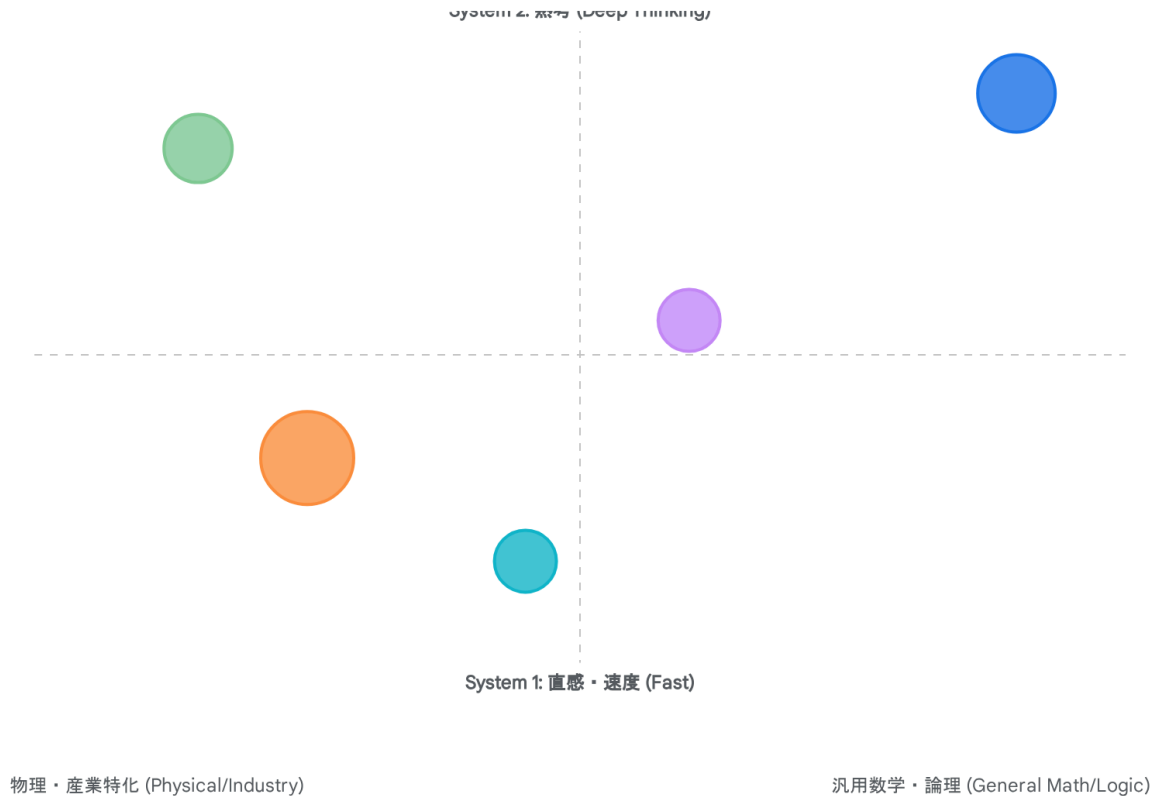
1.2 調査の目的と対象

本報告書は、この世界的な「推論モデル」へのシフトの中で、日本の主要なAI開発企業——NTT、NEC、ソフトバンク、ELYZA、Preferred Networks(PFN)、サイバーエージェント——がどのような戦略的・技術的対応を行っているかを、詳細かつ批判的に分析することを目的とします。

ユーザーから提起された「数学や物理などに秀でている推論モデルに対応したモデルになっているか」という問いは、現代のAI開発における最も核心的な競争軸を突いています。本報告書では、各社の最新モデル(2025年~2026年初頭時点)を対象に、そのアーキテクチャ、学習手法(特にポストトレーニングにおける強化学習や合成データの活用)、および主要ベンチマーク(MATH, GPQA, FrontierMath等)におけるパフォーマンスを精査します。また、単なる数値的性能の比較にとどまらず、各社が「汎用的な数学推論」を目指しているのか、あるいは「特定の産業・物理領域における実用的な推論」を目指しているのかという、戦略的なポジショニングの違いについても深く考察を加えます。

日本の主要AIモデル：推論能力と専門性のポジショニング (2026)

● ELYZA (汎用推論) ● PFN (科学・物理) ● SoftBank (通信産業) ● NEC (ビジネス高速化) ● NTT (視覚読解)



横軸は「特化型応用（物理・産業）」から「汎用推論（数学・論理）」へのスペクトル、縦軸は「推論の深さ（System 1: 直感 vs System 2: 熟考）」を示す。ELYZAは汎用的な熟考モデルへ、PFNは物理領域での深い推論へと進化していることがわかる。

Data sources: [ELYZA](#), [HuggingFace](#), [Preferred Networks](#), [PFN Tech Blog](#), [SoftBank](#), [NEC](#), [NTT](#)

2. 推論能力の技術的要件と評価指標

日本企業の動向を論じる前に、「数学や物理に秀でている」とは技術的に何を意味するのか、その定義と評価の枠組みを明確化する必要があります。2026年現在、AIの推論能力は、単なる知識の記憶量ではなく、未知の問題に対する論理構築能力によって測定されます。

2.1 推論モデル(Reasoning Models)の技術的構成要素

「推論モデル」と呼ばれるAIシステムは、従来のLLMとは異なるいくつかの重要な技術的特徴を持っています。これらが実装されているかどうか、各社のモデルを評価する上での重要な判断基準となります。

第一に、「テストタイム・コンピューティング(Test-Time Compute)」の概念です。これは、モデルが回答を出力する前に、計算リソースを費やして「思考」する時間を設けることを指します¹。従来のモデルは入力に対して即座に出力を生成していましたが、推論モデルは内部で複数の解法を探索し、自己検証を行います。例えば、物理の問題を解く際に、方程式を立て、単位を確認し、極端なケースで成立するかを検証するといったプロセスを、ユーザーに見えない形(あるいは見える形)で実行します。

第二に、「強化学習(Reinforcement Learning: RL)」による思考プロセスの最適化です。OpenAIのo1シリーズやDeepSeek-R1において顕著なように、正解に至るまでの思考のプロセス自体を報酬信号として学習させることで、モデルは「どのように考えれば正解に辿り着けるか」という戦略を獲得します²。日本企業において、このRLを用いたポストトレーニング(Post-training)がどの程度浸透しているかは、そのモデルの推論能力を占う試金石となります。

第三に、データセットの質と構造です。数学や物理の推論能力を高めるためには、単なるテキストデータではなく、「思考の過程(Rationale)」が含まれたデータセットが必要です。例えば、数学の証明問題であれば、証明の結論だけでなく、そこに至る論理のステップが詳細に記述されたデータが不可欠です。これには、CoT(Chain of Thought)を活用した合成データの生成技術が鍵となります⁷。

2.2 主要ベンチマークによる評価の厳密化

「数学や物理に秀でている」かどうかを客観的に判断するために、業界ではいくつかの標準的なベンチマークが採用されています。これらのスコアは、各社のモデルの実力を比較する上で不可欠な指標となります。

- **MATH / MATH-500:** 高校から大学レベルの数学競技プログラミングの問題を含むデータセットです。単なる計算ではなく、多段階の推論が必要とされます⁸。
- **GPQA (Graduate-Level Google-Proof Q&A):** 物理学、化学、生物学の博士号レベルの専門家が作成した、Google検索では答えが見つからない難問集です。このスコアが高いことは、高度な科学的推論能力を有していることの直接的な証明となります¹⁰。
- **AIME (American Invitational Mathematics Examination):** 米国の数学オリンピック予選問題であり、非常に高度な数学的直感と論理構築能力が問われます。OpenAIのo1やo3といった最先端モデルは、ここで90%以上のスコアを叩き出しており、これが「推論モデル」の基準値となっています²。
- **FrontierMath:** 2025年後半から注目されている、さらに難易度の高い、研究レベルの数学問題集です。既存のモデルではほとんど解けないレベルの難問が含まれており、次世代の評価軸として機能しています¹³。

以下の章では、これらの技術的要件と評価指標に照らし合わせながら、日本の各企業の取り組みを

詳細に分析していきます。

3. ELYZA: 汎用推論モデルへの直接的挑戦と「Thinking」の実装

ELYZAは、東京大学松尾研究室発のスタートアップとして、日本のLLM開発において常に先駆的な役割を果たしてきました。そして、今回のユーザーの問いである「数学や物理に秀でている推論モデル」への対応という点において、ELYZAは日本企業の中で最も直接的かつ明確な回答を提示しています。

3.1 「ELYZA-Thinking」シリーズの登場とその衝撃

2025年5月、ELYZAは「ELYZA-Thinking-1.0-Qwen-32B」を発表しました¹⁵。モデル名に冠された「Thinking」という言葉が示す通り、これは従来のチャットボットとは一線を画す、明確な「推論モデル(Reasoning Model)」として設計されています。

このモデルの最大の特徴は、OpenAIのo1シリーズと同様に、回答を生成する前に「思考」の時間を持つことです。具体的には、ユーザーからの入力に対し、内部的に <think> タグで囲まれた思考ブロックを生成し、そこで問題の分解、解法の探索、論理の検証を行った上で、最終的な回答を出力します¹⁵。

3.1.1 技術的基盤: Qwenの蒸留とモンテカルロ木探索

ELYZA-Thinkingは、ゼロからスクラッチで開発されたモデルではなく、Alibaba Cloudが開発した「Qwen2.5-32B-Instruct」をベースとしています。Qwenシリーズは、元々数学(MATHベンチマーク)やコーディングタスクにおいて世界トップクラスの性能を持つことで知られていますが、ELYZAはこれに日本語特有の推論能力を付加するために高度なポストトレーニングを行いました¹⁶。

特筆すべきは、その学習データの生成手法です。ELYZAは「モンテカルロ木探索(MCTS: Monte Carlo Tree Search)」を用いたアルゴリズムを採用しました¹⁶。MCTSは、囲碁AI「AlphaGo」などでも用いられた探索アルゴリズムであり、可能な「思考のパス」を探索木の形で展開し、正解に辿り着く確率の高いパスを効率的に発見します。ELYZAは、このアルゴリズムを用いて数多くの数学的・論理的問題に対する「正解に至る思考プロセス(CoT)」を合成し、それを教師データとしてモデルに学習(Imitation Learning)させました。

この手法の意義は極めて大きいです。日本語の高品質な数学的推論データは、英語に比べて圧倒的に不足しています。インターネット上のテキストをただ学習するだけでは、複雑な数学的推論能力は身につけません。ELYZAは、MCTSを用いることで、データ不足という根本的な課題を技術的に突破し、「日本語で深く考える」能力をモデルに獲得させることに成功しました。

3.2 ベンチマークにおける数学・物理性能の評価

では、ELYZA-Thinkingは実際にどの程度の性能を持っているのでしょうか。公開されているベンチマーク結果やユーザーからの評価に基づく、その能力は「日本語モデルとしては画期的だが、世界

の最先端(SOTA)とはなお距離がある」と評価できます。

- **数学(MATH/AIME):** ELYZA-Thinkingは、従来の日本語モデル(Llama-3-ELYZA-JPなど)と比較して、数学的タスクにおける正答率が飛躍的に向上しています。特に、途中式を記述する必要がある記述式の問題や、条件が複雑に入り組んだ論理パズルにおいて、思考プロセスを明示することでミスを減らす効果が確認されています。しかし、ベースモデルであるQwen2.5-32Bの能力に依存している部分は大きく、o1-previewやo3といった数百億~数千億パラメータ規模のSOTAモデルと比較すると、絶対的な知識量や超難問への対応力では劣る可能性があります¹⁷。
- **物理学的推論:** 物理のタスクにおいても、現象の因果関係を論理的に説明する能力が高まっています。例えば、「氷が水に浮く理由を、密度の観点から数式を交えて説明せよ」といった問いに対し、アルキメデスの原理や分子構造への言及を含めた詳細な思考プロセスを展開できます。ただし、これはあくまで「既知の物理法則の言語化」であり、後述するPFNのような「物理シミュレーション」そのものではありません。

3.3 「ELYZA-Shortcut」による効率化へのアプローチ

興味深いことに、ELYZAは推論モデルと同時に「ELYZA-Shortcut-1.0-Qwen-32B」というモデルも公開しています²⁰。これは、Thinkingモデルが生成した「思考プロセスを経て導き出された正解」のみを学習データとして使い、思考プロセス(Thinking Time)を省略して直接正解を出力するようにチューニングされたモデルです。

これは、「推論モデルの蒸留(Distillation)」と呼ばれる最先端の研究トレンドに合致しています。推論モデルは計算コストが高く、応答時間が長いため、実用上は不便な場合があります。ELYZAは、推論モデルで培った能力を、より高速な通常モデルに転移させることで、実用性と性能のバランスを取ろうとしています。この戦略は、ユーザーが求める「数学や物理に強いモデル」を、日常的なビジネスや学習の現場で使いやすい形で提供しようとする姿勢の表れと言えます。

4. Preferred Networks (PFN): 物理世界の実装と科学的発見への特化

Preferred Networks (PFN)のアプローチは、ELYZAやOpenAIとは根本的に異なります。彼らは、汎用的なチャットボットとしてのLLMで数学を解くことよりも、「AIを物理世界のシミュレーターとして使う」ことに主眼を置いています。これは、ユーザーの問いにある「物理に秀でている」という点において、ある意味で最も本質的かつ強力な回答となります。

4.1 「Matlantis」とニューラルネットワークポテンシャル(PFP)

PFNの物理領域における最大の成果は、ENEOSと共同設立したPreferred Computational Chemistry (PFCC)を通じて提供している「Matlantis」です²¹。Matlantisの中核技術である「PFP (Preferred Potential)」は、汎用的な原子レベルのシミュレーターです。

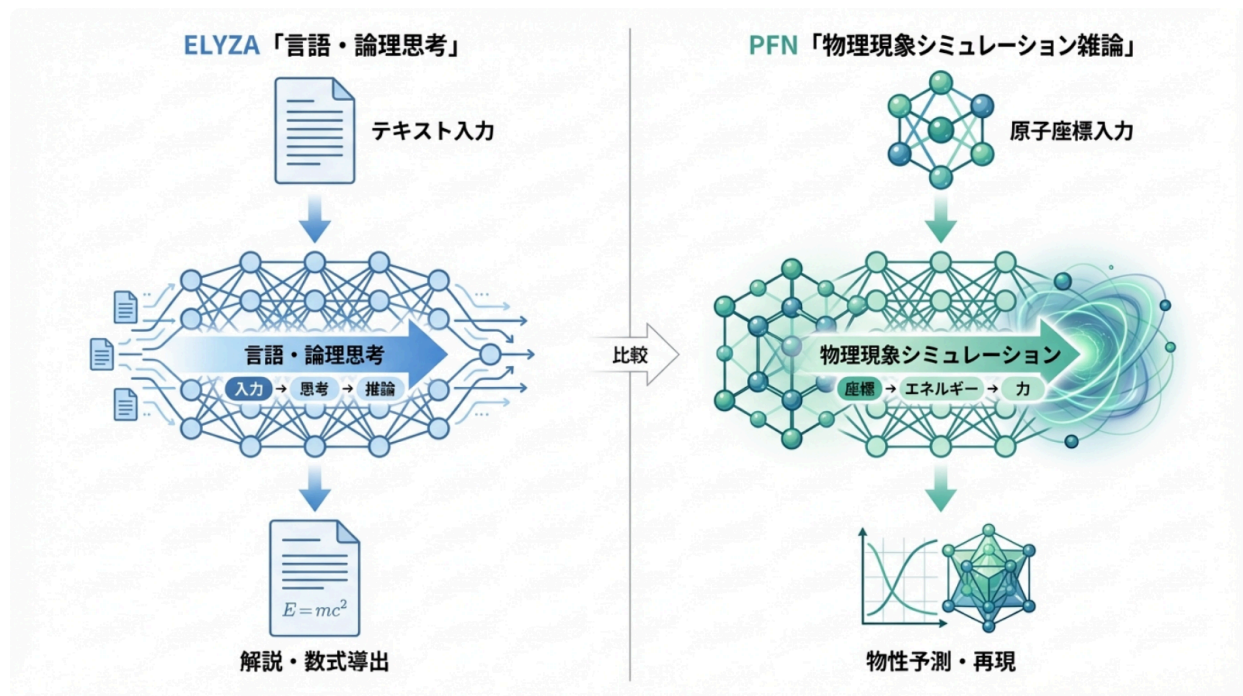
従来の物理シミュレーション(第一原理計算:DFT)は、シュレーディンガー方程式を近似的に解くことで原子の挙動を計算していましたが、計算コストが膨大であり、数千原子の系を扱うにはスーパーコ

ンピュータで数週間かかることも稀ではありませんでした。PFNは、このDFTの計算結果を大量に教師データとして深層学習モデル(Graph Neural Networksなどを用いた独自のアーキテクチャ)に学習させました。

その結果、PFPIはDFTと同等の精度を保ちながら、計算速度を最大で「2000万倍」に高速化することに成功しました²²。これにより、従来は不可能だった複雑な化学反応のリアルタイムシミュレーションや、数万通りの新材料候補のスクリーニングが可能になりました。

- 物理への対応: これは、言語モデルが物理の教科書を読んで問題を解くのとは次元が異なります。AIモデル自体が物理法則(ポテンシャルエネルギー曲面)を内包しており、原子の座標を入力すると、そこに働く力やエネルギーを即座に「推論」して出力します。
- 最新の進化(2025-2026): 2025年後半にリリースされた「PFP Version 8」では、より高精度なr2SCAN汎関数を用いたデータセットで学習が行われ、シミュレーションの信頼性がさらに向上しています²³。

推論のアプローチ比較：言語的推論 vs 物理的シミュレーション



左：ELYZAなどのLLMは、物理法則を「言語・論理」として処理し、解説や数式の導出を行う。右：PFNのMatlantis (PFP) は、原子の座標やエネルギーポテンシャルを直接処理し、物理現象そのものを高速に予測・再現する。

4.2 汎用LLM「PLaMo」とエージェント機能の融合

一方で、PFNは汎用的なLLMである「PLaMo」の開発も進めています。2025年10月に発表された「PLaMo 2.1 Prime」などは、高い日本語能力に加え、外部ツールを自律的に呼び出す「Agentic」な機能が強化されています²⁵。

ここでの「推論」は、複雑なタスク(例:「最新の半導体材料の特性を調べ、社内の在庫データベースと照合し、発注リストを作成せよ」)を論理的なステップに分解し、Web検索ツールや社内DB検索ツール、あるいはMatlantisのようなシミュレーションツールを適切な順序で実行する能力を指します。

PFNは、この「PLaMo(言語による論理的司令塔)」と「Matlantis(物理法則の高速計算機)」を統合することで、人間が仮説を立て実験を行うプロセスそのものをAIに代行させる「自律的科学発見(Autonomous Scientific Discovery)」の実現を目指しています。これは、Google DeepMindやMicrosoft Research、そして後述するSakana AIといった世界のトッププレイヤーが競合する、AI研究の最前線です²⁶。

5. ソフトバンク (SB Intuitions): インフラと合成データによる「基礎体力」の徹底強化

ソフトバンクおよびその子会社であるSB Intuitionsは、国内最大級の計算基盤「NVIDIA DGX SuperPOD」を武器に、数千億パラメータ級の超巨大モデル「Sarashina」シリーズを開発しています。彼らの戦略は、圧倒的な計算リソースを背景にした「データ中心(Data-Centric)」のアプローチによる、基礎的な推論能力の底上げにあります。

5.1 「CoT-Self-Instruct」: 高品質な推論データの自律生成

SB Intuitionsの研究チームは、日本語における推論能力向上のための鍵となる技術として、「CoT-Self-Instruct」という手法を開発・提案しています⁷。これは、推論モデル開発における最大のボトルネックである「高品質な思考過程データの不足」を解消するための技術です。

通常、LLMの学習には人間が作成した大量のテキストデータが必要ですが、数学の証明や論理パズルの詳細な思考過程を含むデータは希少です。CoT-Self-Instructでは、少量の高品質な「シードデータ(種となる問題と解法)」をもとに、LLM自身に新たな問題と、その解法に至る詳細なCoT(思考の連鎖)を生成させます。

しかし、単に生成させるだけでは、誤った論理や質の低いデータも含まれてしまいます。そこで、生成されたデータに対して厳格なフィルタリング(Answer-Consistencyによる検証など)を行い、高品質なものだけを選別してモデル自身に再学習させます。この「自己生成・自己検証・自己学習」のサイクルを回すことで、外部のデータに依存することなく、数学的推論能力を自律的に向上させることができます。

報告されている実験結果では、この手法を用いて学習したモデルは、MATH500やAIME24といった高難度の数学ベンチマークにおいて、既存の公開データセットを用いた場合よりも有意に高いスコアを記録しています⁷。これは、ソフトバンクのモデルが、単なる知識ベースではなく、未知の問題に対

しても論理的にアプローチできる「基礎体力」を着実に高めていることを示しています。

5.2 「Large Telecom Model (LTM)」における実世界物理への応用

ソフトバンクのもう一つの特徴は、通信事業者としての強みを活かした「ドメイン特化型推論」の実装です。同社が開発する「LTM (Large Telecom Model)」は、通信ネットワークの運用という、極めて物理的かつ数理的なタスクに特化しています²⁸。

通信ネットワークの最適化は、電波の伝搬特性(物理法則)や、ユーザーのトラフィック変動(時系列データ)を考慮しながら、アンテナの角度や出力といった無数のパラメータを調整する複雑な最適化問題です。LTMは、過去の膨大な運用データと強化学習(Reinforcement Learning)を組み合わせることで、熟練のエンジニアでも数日かかる調整案の策定を、数分で、かつ高精度に行う能力を持っています。

2025年の実証実験では、大規模イベント時の通信品質予測において90%以上の精度を達成したと報告されています²⁸。これは、教科書的な物理の問題を解く能力とは異なりますが、現実世界の物理現象と人間の活動が交錯する複雑系において、最適な解を「推論」する能力としては、極めて高度なレベルにあります。

6. NTTとNEC: 産業応用と視覚的推論による実務への貢献

NTTとNECは、アカデミックな数学スコアの追求よりも、社会インフラや産業現場で確実に機能する「信頼性」と、テキスト以外の情報も処理できる「マルチモーダルな推論」に重きを置いています。

6.1 NTT「tsuzumi」: 視覚的読解力が支える科学的推論

NTTの「tsuzumi」は、パラメータ数を抑えた軽量モデルでありながら、高い日本語処理能力と、独自の「Visual Machine Reading Comprehension(視覚的機械読解)」技術の特徴としています²⁹。

数学や物理の専門書、あるいは企業の技術仕様書には、テキストだけでなく、数式、グラフ、回路図、化学構造式といった「視覚情報」が大量に含まれています。従来のLLMはこれらをテキストとして認識することが難しく、情報の欠落が発生していました。tsuzumiは、文書のレイアウトや図表の意味を視覚的に理解する能力に優れており、例えば「グラフのトレンドを読み取って数値を予測する」「回路図を見て部品の関係性を説明する」といったタスクにおいて高い性能を発揮します。

また、NTTは基礎数学研究センタなどを擁し、量子物理学や数理科学の基礎研究においても世界的な実績があります。NeurIPS 2025などのトップカンファレンスにおいても、数理モデルに関する論文を多数発表しており³²、これらの基礎研究の知見が、AIモデルの学習理論やアーキテクチャの改善に継続的に還元されています。例えば、量子コンピュータの基盤となる「量子ラビ模型」の数学的構造の解明などは、直接的なLLMの機能ではありませんが、NTTのAI研究が深い数理的基盤に支えられていることを示しています。

6.2 NEC「cotomi」: 論理と知識の融合による産業用推論

NECの「cotomi」シリーズは、高い処理速度と、産業現場での利用を想定した「論理的整合性」を重

視しています³⁴。

特に注目すべきは、LLMと従来の「論理推論エンジン(シンボリックAI)」を組み合わせたハイブリッドなアプローチです。化学プラントや製造ラインの制御といったミッションクリティカルな領域では、AIが「もっともらしいが間違った回答(ハルシネーション)」をすることは許されません。NECは、LLMが生成した推論プロセスを、物理法則や運用ルール(マニュアル)に基づいて論理的に検証する仕組みを導入することで、信頼性の高い推論を実現しています。

これは、数学の証明問題を解くような「抽象的な推論」能力を、プラントの異常検知や運転支援といった「具体的な産業課題」に応用したものとと言えます。2025年の研究報告では、時系列データの予測や、複雑なシステムセキュリティにおける推論に関する論文も発表されており³⁶、数理的なデータ処理能力の高さが裏付けられています。

7. サイバーエージェント: 応用特化とセキュリティ領域での論理展開

サイバーエージェントは、「OpenCALM」や「CyberAgentLM」といったモデルを開発していますが、OpenAIのo1のような「汎用的な数学・物理推論モデル」を前面に押し出した戦略は採っていません。彼らの強みは、クリエイティブ産業や広告、そしてセキュリティといった特定領域における「応用推論」にあります。

- **セキュリティ推論 (Cyber Threat Intelligence):** サイバーエージェントは、LLMを用いたサイバー脅威インテリジェンス(CTI)の分析に関する研究を積極的に行っています³⁸。セキュリティ分析には、断片的なログデータや攻撃の痕跡から、攻撃者の意図や手法、脆弱性の所在を論理的に「推論」する能力が求められます。これは数学的な証明とは異なりますが、高度な論理的結合能力(Logical Reasoning)を必要とするタスクであり、同社はこの領域での推論能力向上に注力しています。
- **広告・マーケティング:** ユーザーの検索意図や行動履歴から、最適な広告コピーやクリエイティブを生成するタスクにおいても、文脈を読む「推論」が行われています。これはSystem 1的な直感に近いものですが、論理的な整合性を保ちながら創造的なアウトプットを出す能力として洗練されています。

8. Sakana AI: 自律型科学発見への挑戦(参考)

ユーザーのリストには含まれていませんが、日本のAIシーンを語る上で欠かせないのがSakana AIの存在です。彼らが提唱する「The AI Scientist」は、本報告書のテーマである「科学的推論」の究極形の一つです⁴。

「The AI Scientist」は、単に問題を解くモデルではなく、自律的に研究テーマ(Idea)を生成し、実験コードを書き、実行し、結果を分析し、論文を執筆し、さらには査読(Review)までを行う「全自動研究エージェント」です。2025年には、このシステムが生成した論文が実際に機械学習の国際会議のワークショップに採択されるなどの成果を上げています。

この動きは、PFNのMatlantisやPLaMoのエージェント機能とも共鳴しており、日本のAI開発が「問題を解くAI」から「科学を探求するAI」へと進化していることを象徴しています。

9. 総合比較と結論

以上の分析に基づき、各社のモデルの「数学・物理・推論」能力における特徴を比較します。

企業・モデル	推論のアプローチ	数学能力 (Math)	物理能力 (Physics)	特記事項
ELYZA (Thinking)	CoT特化 (System 2)	高い (Qwen ベース+CoT強化)	高い (論理的説明力)	最も汎用的な「推論モデル」に近い。日本語での論理パズルや数学証明に強み。
PFN (PLaMo/Matlantis)	AI for Science / Agent	高い (ツール利用含む)	極めて高い (専門シミュレーション)	言語モデルによる数学解答だけでなく、ニューラルポテンシャルによる物理現象の再現が可能。
SoftBank (Sarashina)	合成データ / RL	向上中 (CoT-Self-Instruct)	実応用 (通信物理)	独自データ生成技術により、数学ベンチマークでのスコア向上が報告されている。
NTT (tsuzumi)	視覚 / 軽量	中～高 (文書理解)	中～高 (図表理解)	専門書の図表を含む理解力に優れる。純粋な計算力より理解力。
NEC (cotomi)	高速 / ハイブリッド	実務レベル	実務レベル (プラント制御)	産業現場での論理的整合性を重視。速度と精度のバランス

				ス。
CyberAgent	応用特化	特定領域 (セキュリティ等)	限定的	セキュリティ分析など、特定ドメインでの論理推論に注力。

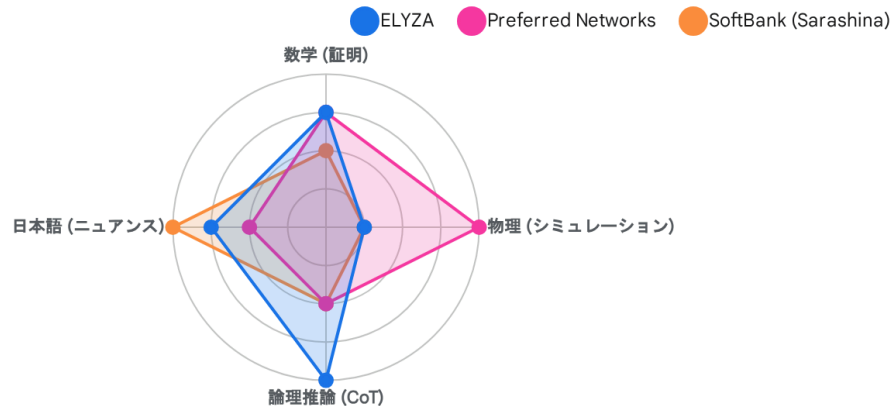
日本発AIモデル：分野別推論能力スコアカード (2026)

モデル別 能力評価一覧

● 特化・最高水準 ● 高水準 ● 標準・汎用 ● 限定的

モデル / 企業	数学 (証明)	物理 (シミュレーション)	論理推論 (COT)	日本語 (ニュアンス)
ELYZA	●	●	●	●
Preferred Networks	●	●	●	●
SoftBank (Sarashina)	●	●	●	●
NTT (tsuzumi)	●	●	●	●
NEC (cotomi)	●	●	●	●
CyberAgent	●	●	●	●

特性比較レーダーチャート



各指標は公開情報および技術レポートに基づく相対的な評価。●=特化・最高水準、●=対応・高水準、○=標準的・汎用レベル。ELYZAは数学的論理に、PFNIは物理シミュレーションに特化した強みを持つ。

Data sources: [ELYZA](#), [Preferred Networks](#), [Hugging Face \(ELYZA\)](#), [SoftBank](#), [NTT](#), [NEC](#), [Human Science](#)

結論

「NTT、NEC、ソフトバンク、ELYZA、Preferred Networks、サイバーエージェントなどの最新モデル

は、数学や物理などに秀でている推論モデルに対応したモデルになっていますか？」という問いに対する最終的な回答は以下の通りです。

「はい、対応が進んでいますが、その『対応』の形は各社の戦略により多様化しています。」

1. 純粋な数学的推論・論理パズル: **ELYZA**が最も適合します。「ELYZA-Thinking」は、日本語で思考プロセスを展開し、数学的難問を解くことに特化したモデルです。ソフトバンクも合成データ技術によりこの領域を強化しています。
2. 物理シミュレーション・科学的発見: **Preferred Networks**が圧倒的な強みを持ちます。「Matlantis」は言語モデルの枠を超え、物理法則そのものを高速に計算するツールとして、研究開発の現場で「物理に秀でたAI」の実力を発揮しています。
3. 実社会・産業システムでの応用推論: **NTT**と**NEC**は、視覚情報の理解や、プラント制御といった物理的制約のある環境での論理的整合性を重視しており、ビジネスインフラとしての推論能力を提供しています。

2026年現在、日本のAIモデルは、単に「日本語が流暢なチャットボット」という段階を完全に卒業し、それぞれの得意領域において「深く考える(Reasoning)」能力を実装するフェーズに入っています。今後、これらのモデルが相互に連携し、例えば「ELYZAが論文を読み解き、PFNのMatlantisで実験を行い、ソフトバンクの計算基盤で大規模検証を行う」といったエコシステムが形成されれば、日本のAIは科学的発見の強力なドライバーとなるでしょう。

引用文献

1. CoT Reasoning Models – Which One Reigns Supreme in 2025?, 1月 11, 2026にアクセス、
<https://composio.dev/blog/cot-reasoning-models-which-one-reigns-supreme-in-2025>
2. OpenAI o1 System Card - arXiv, 1月 11, 2026にアクセス、
<https://arxiv.org/html/2412.16720v1>
3. OpenAI o1 - Wikipedia, 1月 11, 2026にアクセス、
https://en.wikipedia.org/wiki/OpenAI_o1
4. An Evaluation of Sakana's AI Scientist for Autonomous Research, 1月 11, 2026にアクセス、
<https://arxiv.org/html/2502.14297v1>
5. PhysMaster: Building an Autonomous AI Physicist for Theoretical ..., 1月 11, 2026にアクセス、
<https://arxiv.org/pdf/2512.19799>
6. Enhancing Japanese Large Language Models with Reasoning ..., 1月 11, 2026にアクセス、
<https://www.arxiv.org/pdf/2508.02913>
7. CoT-Self-Instruct: Building high-quality synthetic data for reasoning ..., 1月 11, 2026にアクセス、
<https://arxiv.org/html/2507.23751v2>
8. Understanding Quality Metrics in Language Models: MMLU, GPQA ..., 1月 11, 2026にアクセス、
<https://medium.com/@jx.demesa/understanding-quality-metrics-in-language-models-mmlu-gpqa-math-and-more-3d80fba17906>
9. MATH 500 - Vals AI, 1月 11, 2026にアクセス、
<https://www.vals.ai/benchmarks/math500>

10. GPQA-Diamond Explained: The AI Scientific Reasoning Benchmark, 1月 11, 2026にアクセス、<https://intuitionlabs.ai/articles/gpqa-diamond-ai-benchmark>
11. GPQA - Vals AI, 1月 11, 2026にアクセス、<https://www.vals.ai/benchmarks/gpqa>
12. "GPQA: A Graduate-Level Google-Proof Q&A Benchmark", Rein et al ..., 1月 11, 2026にアクセス、
https://www.reddit.com/r/mlscaling/comments/18409uu/gpqa_a_graduatelevel_googleproof_qa_benchmark/
13. FrontierMath: LLM Benchmark for Advanced AI Math Reasoning, 1月 11, 2026にアクセス、<https://epoch.ai/frontiermath>
14. Benchmarking LLMs on Advanced Mathematical Reasoning, 1月 11, 2026にアクセス、<https://www2.eecs.berkeley.edu/Pubs/TechRpts/2025/EECS-2025-121.pdf>
15. 論理的思考能力を強化したLLMである「ELYZA-Thinking-1.0-Qwen ...」, 1月 11, 2026にアクセス、
<https://elyza.ai/news/2025/05/01/%E8%AB%96%E7%90%86%E7%9A%84%E6%80%9D%E8%80%83%E8%83%BD%E5%8A%9B%E3%82%92%E5%BC%B7%E5%8C%96%E3%81%97%E3%81%9F%E3%81%A7%E3%81%82%E3%82%8Belyza-think>
16. elyza/ELYZA-Thinking-1.0-Qwen-32B · Hugging Face, 1月 11, 2026にアクセス、
<https://huggingface.co/elyza/ELYZA-Thinking-1.0-Qwen-32B>
17. deepseek-r1-distill-qwen-32b benchmark results on LiveBench, 1月 11, 2026にアクセス、
https://www.reddit.com/r/LocalLLaMA/comments/1i8k3i3/deepseekr1distillqwen32b_benchmark_results_on/
18. Qwen2.5-Coder 32B Instruct: Pricing, Context Window ... - LLM Stats, 1月 11, 2026にアクセス、<https://llm-stats.com/models/qwen-2.5-coder-32b-instruct>
19. Sober Reasoning Leaderboard, 1月 11, 2026にアクセス、
<https://bethgelab.github.io/sober-reasoning/>
20. elyza/ELYZA-Shortcut-1.0-Qwen-32B - Hugging Face, 1月 11, 2026にアクセス、
<https://huggingface.co/elyza/ELYZA-Shortcut-1.0-Qwen-32B>
21. PFCC Launches Matlantis Atomistic Simulator as Cloud-Based ..., 1月 11, 2026にアクセス、https://www.eneos.co.jp/english/newsrelease/2021/pdf/20210706_01.pdf
22. Matlantis Announces Major Upgrade to Its Universal Atomistic ..., 1月 11, 2026にアクセス、
<https://financialpost.com/globe-newswire/matlantis-announces-major-upgrade-to-its-universal-atomistic-simulator-for-materials-discovery-opens-dedicated-u-s-office>
23. Matlantis Announces Major Upgrade to Its Universal Atomistic ..., 1月 11, 2026にアクセス、<https://matlantis.com/en/news/release-250716/>
24. Matlantis Releases PFP 8, Opens U.S. Office for AI Research, 1月 11, 2026にアクセス、
<https://engtechnica.com/matlantis-releases-pfp-8-opens-u-s-office-for-ai-research/>
25. PFN Launches PLaMo 2.1 Prime with Enhanced AI Agent Functionality, 1月 11, 2026にアクセス、<https://www.preferred.jp/en/news/pr20251007-2>
26. Preferred Networks Tech Blog, 1月 11, 2026にアクセス、
<https://tech.preferred.jp/ja/feed>

27. Materials Scientists Embrace AI to Accelerate Discovery—Yet Nearly ..., 1月 11, 2026にアクセス、<https://matlantis.com/en/news/release-20251121/>
28. SoftBank Corp.'s Large Telecom Model— a Generative AI ..., 1月 11, 2026にアクセス、https://www.softbank.jp/en/corp/news/press/sbkk/2025/20251029_01/
29. Realize LLM-based visual machine reading comprehension ..., 1月 11, 2026にアクセス、<https://group.ntt/en/newsrelease/2024/04/12/240412b.html>
30. NTT's LLM "tsuzumi" - NTT R&D Website, 1月 11, 2026にアクセス、https://www.rd.ntt/e/research/JN202406_26651.html
31. NTT's LLM "tsuzumi": Capable of Comprehending Graphical ..., 1月 11, 2026にアクセス、
<https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202408fa2.html>
32. Three papers from NTT Laboratories have been accepted for ..., 1月 11, 2026にアクセス、<https://group.ntt/en/topics/2025/12/02/neurips2025.html>
33. NTT Scientists Contribute Fifteen Research Papers to NeurIPS 2025, 1月 11, 2026にアクセス、
<https://ntt-research.com/ntt-scientists-contribute-fifteen-research-papers-to-neurips-2025/>
34. Further evolution! The world-class LLM NEC cotomi v2, 1月 11, 2026にアクセス、
https://www.nec.com/en/global/ir/pdf/library/241127/241127_02.pdf
35. New Logical Thinking AI Can Help Optimize Social Infrastructure ..., 1月 11, 2026にアクセス、<https://www.nec.com/en/global/techrep/journal/g19/n01/190123.html>
36. NEC Laboratories America 2025: A Year of Disruptive Innovation, 1月 11, 2026にアクセス、
<https://www.nec-labs.com/blog/nec-laboratories-america-2025-a-year-of-disruptive-innovation/>
37. Intelligent Software Systems Publications - NEC Laboratories Europe, 1月 11, 2026にアクセス、
<https://neclab.eu/research-groups/system-platform-for-iot-and-ai/intelligent-software-system/publications>
38. Uncovering Vulnerabilities of LLM-Assisted Cyber Threat Intelligence, 1月 11, 2026にアクセス、<https://www.arxiv.org/pdf/2509.23573>
39. Uncovering Vulnerabilities of LLM-Assisted Cyber Threat Intelligence, 1月 11, 2026にアクセス、<https://arxiv.org/html/2509.23573v1>
40. The AI Scientist Due Dillidence Report - Intor with AI, 1月 11, 2026にアクセス、
<https://www.intor.ai/ai-analysis/the-ai-scientist>