

株式会社Preferred Networks (PFN) の「PLaMo」シリーズ進化と垂直統合型AI経営戦略の全貌

Gemini 3.1 pro

1. イントロダクション: 日本発「AIインフラ」の確立を目指すPFNの独自性と戦略的優位性

現代の産業競争力において、大規模言語モデル(LLM)をはじめとする生成AI技術は、単なるソフトウェアツールを超え、国家や企業の根幹を成す最重要インフラストラクチャーとして位置づけられている。グローバルなテクノロジー覇権競争が激化する中、日本国内においてもデータの主権(ソブリンAI)を確保し、独自の文脈や産業制約に適合する基盤モデルの構築が急務となっている。このようなマクロ環境下において、株式会社Preferred Networks(以下、PFN)は、「現実世界を計算可能にする(Make the real world computable)」というミッションを掲げ、類を見ないアプローチで生成AI市場に挑んでいる¹。

PFNの最大の特異性は、単にAIアプリケーションやソフトウェア層を開発する企業にとどまらず、深層学習用プロセッサ(ハードウェア)から、クラウドコンピューティング基盤、独自の基盤モデル(LLM)、そして特定産業向けのソリューションに至るまで、AIバリューチェーン全体を包括する「垂直統合型(フルスタック)」のアプローチを採用している点にある¹。本レポートは、PFNがフルスクラッチで独自開発を続ける純国産大規模言語モデル「PLaMo(プラモ)」シリーズの技術的進化の系譜を詳細に紐解くとともに、ハードウェア(MN-Coreシリーズ)の開発、そして国内の主要企業との戦略的アライアンスから見てくる同社の経営戦略について、多角的な視点から徹底的に分析を行う。特に、2026年6月現在における最新の動向を踏まえ、PFNが日本の産業界において「コアAIインフラ」としてどのような役割を担おうとしているのか、その深層を明らかにする。

2. 純国産LLM「PLaMo」シリーズの技術的進化とアーキテクチャの系譜

PLaMoは、PFNが独自のアーキテクチャと学習データを用いて、ゼロからの事前学習・事後学習を通して開発した純国産の基盤モデルである⁴。近年、多くの企業がMetaのLLaMAなどをベースにしたファインチューニングモデルを発表しているが、PLaMoは他社のオープンモデルを一切ベースにしていない。そのため、社外ライセンスの縛りや、開発過程における不明瞭なブラックボックスが存在せず、商用利用や独自カスタマイズにおいて極めて高い透明性と自由度を誇る純国産基盤モデルとなっている⁴。

2.1. 黎明期と基盤構築: PLaMo-13BからPLaMo-100Bへの飛躍

PFNは、基盤モデルが多様な産業におけるイノベーションを劇的に加速させると見込み、2023年春より本格的なLLM開発に着手した⁶。同年秋には、マルチモーダル基盤モデルの開発を加速させるため、新子会社「株式会社Preferred Elements(PFE)」を設立し、PFNの最高研究責任者(Chief

Executive Researcher)である岡野原大輔氏が社長に就任する体制を敷いた⁶。この強固な研究開発体制のもとで生み出された初期の代表的な成果が「PLaMo-13B」である。このモデルは、約1.4兆トークンの日本語および英語の公開データを収集・処理したデータセットで学習され、単一の標準的なGPUサーバーで推論可能な130億パラメータというコンパクトなサイズでありながら、当時の同規模モデルの中で世界トップクラスの性能を発揮した⁶。PFNは、このPLaMo-13BをApache License v2.0の下で無償公開し、商用および研究目的での広範な利用を促進することで、国内のエコシステム形成に寄与した⁶。

続いて、経済産業省と国立研究開発法人新エネルギー・産業技術総合開発機構(NEDO)が主導する「GENIAC(Generative AI Accelerator Challenge)」プロジェクトの強力な支援を受け、2024年に開発されたのが、1000億パラメータを誇る「PLaMo-100B」である⁵。このモデルは、計2兆トークン(英語1.3兆トークン、日本語0.7兆トークン)という膨大なコーパスを用いてフルスクラッチで学習された⁹。

PLaMo-100Bの事前学習(Pre-training)は、学習の効率性と精度の最適化を図るため、1.5兆トークンを用いた「フェーズ1」と、0.5兆トークンを用いた「フェーズ2(Continued Pre-training)」の二段階に分けて実行された⁵。以下の表は、各フェーズにおけるデータセットの構成割合を示している。

データセットソース	フェーズ1 (1.5兆トークン)	フェーズ2 (0.5兆トークン)
RefinedWeb (英語)	42%	17%
その他の英語データセット	28%	33%
独自のCommonCrawl-JP (日本語)	18%	46%
その他の日本語データセット	12%	4%

このデータ構成の変遷から、PFNの明確な学習戦略が読み取れる。フェーズ1では、英語の高品質データ(RefinedWebなど)を全体の70%投入することで、モデルの基礎的な論理推論能力や世界知識を構築している。一方、フェーズ2では、PFNが独自に精製した日本語データセット(CommonCrawl-JP)の比率を18%から46%へと劇的に引き上げている。この戦略的なカリキュラム学習により、高度な英語の推論能力を維持したまま、極めて自然で高精度な日本語処理能力をモデルに定着させることに成功しているのである⁹。

技術的なアーキテクチャの側面では、巨大モデルの学習過程における不安定性を排除するため、「QK Normalization」や「Z-Loss」といった先進的な安定化技術が採用されている⁵。また、事前学習完了後に続く事後学習(Post-training)フェーズにおいては、SFT(Supervised Fine-Tuning)およびDPO(Direct Preference Optimization)が適用され、モデルの出力品質が大幅に洗練された⁵。PFNの技術ブログによれば、数学的推論能力を向上させるためのSFTデータセット構築においては、LLMによる自動生成に頼るのではなく、手作業で数式のテンプレートを作成し、数値を変更することでデータを増強するという職人的なアプローチが採られた。これは、LLMが計算結果を生成する際、

トークンの分布が決定論的になる(一意の正解に収束する)という特性を深く理解した上での最適化手法である¹¹。

2.2. 実用化と市場展開の加速: PLaMo 2シリーズの躍進と「指示追従性」の極大化

PLaMo 1.0シリーズで培われた事前学習および事後学習の高度なノウハウは、続く「PLaMo 2」シリーズにおいてさらなる洗練と飛躍的な性能向上を見せた。2024年12月にリリースされた「PLaMo 2.0 Prime」は、同社の商用フラッグシップモデルとして位置づけられ、フルマネージドのPLaMoプラットフォームAPIやAmazon Bedrock Marketplace経由、さらにはオンプレミス環境など、企業の要件に応じた柔軟な導入形態で提供が開始された⁴。PLaMo 2.0 Primeは、miibo(国産AI構築プラットフォーム)、Tachyon Generative AI、さらには全国150以上の自治体で導入されているQommonsAIといった外部の主要サービスにデフォルトのエンジンとして組み込まれ、急速な社会実装を達成している⁷。この一連の社会実装の実績が高く評価され、PLaMo 2.0 Primeは2025年の「日経優秀製品・サービス賞」において最優秀賞を受賞するという栄誉に輝いた¹⁴。

さらに、PFNの開発スピードは止まることなく、2026年1月には事後学習データのさらなる拡充と最適化によって性能を飛躍的に向上させた「PLaMo 2.2 Prime」がリリースされた¹⁵。ここから読み取れるPFNの技術的な力点は、「単純なパラメータ規模の力任せな拡大」から、「指示追従性(Instruction Following)」と「特定ドメイン(特に医療や高度専門分野)における推論能力」の極大化への明確なシフトである¹⁵。

PLaMo 2.2 Primeにおける特筆すべき技術的進化は、ユーザーの複雑なプロンプトや制約に厳格に従う能力の向上である。PFNは、英語の指示追従性能を向上させるため、RLVR(Reinforcement Learning from Verbal Feedback/Reasoning)向けのデータセット「IFTrain」から安全で商用利用可能なプロンプトを厳選し、SFTおよびDPOのデータセットを構築した¹⁵。同時に、日本語特有の出力フォーマット制約を正確に評価・改善するため、独自の評価ベンチマーク「JFBench(Japanese instruction Following Benchmark)」および学習用の「JFTrain」を独自に開発した。JFTrainは、日本語の16の制約グループにわたる174種類の制約から厳選された72種類の制約で構成されており、「出力フォーマットに関する制約」や「複数の制約を同時に満たす能力」を重視した設計となっている¹⁵。

この徹底した事後学習の結果、PLaMo 2.1 Primeでは7.03%に留まっていたMulti-turn(20~40ターン)ロールプレイ能力の指示追従率が、PLaMo 2.2 Primeでは23.7%へと劇的に改善された¹⁵。また、高度な専門知識と論理的推論が要求される医療分野のベンチマークにおいても、目覚ましい成果を上げている。以下の表は、PLaMo 2.2 Primeと主要な海外フロンティアモデルにおける医療ベンチマークの比較を示している。

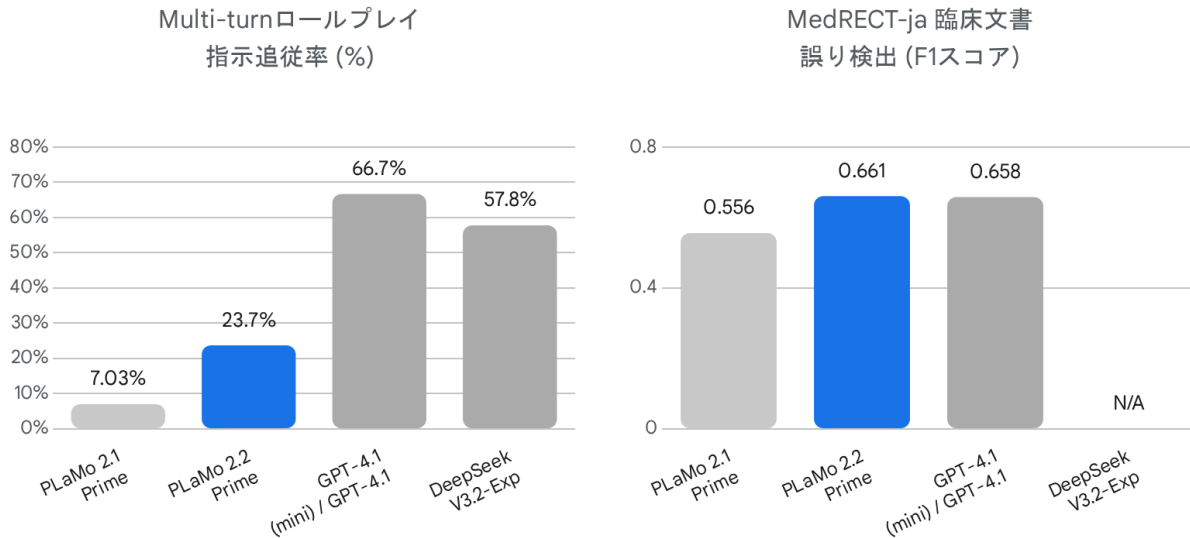
評価モデル	MedRECT-ja (誤り検出 F1スコア)	MedRECT-ja (誤り文抽出 正解率)	MedRECT-ja (誤り訂正 ROUGE-1)	JMLE(日本医師国家試験 正解率)
PLaMo 2.2 Prime	0.661	57.0%	0.492	70.7% (394/557問)

PLaMo 2.1 Prime	0.556	15.8%	0.243	55.1%
GPT-4.1 mini / 4.1	0.658	52.6%	-	94.3%
DeepSeek-V3-0324	-	42.2%	-	88.3%

このデータが示す通り、記録された臨床文書内の矛盾や誤りを検出・訂正する「MedRECT-ja」ベンチマークにおいて、PLaMo 2.2 PrimeはGPT-4.1などのフロンティアモデルを上回るF1スコア(0.661)および誤り文抽出正解率(57.0%)を達成し、極めて高い実用レベルの性能を証明した¹⁵。JMLE(医師国家試験)の正解率においても、前バージョンの55.1%から70.7%へと大幅なスコアアップを果たしており、LLMが単なる「流暢な文章生成ツール」から、高度な専門知識を要求される「論理的推論エンジン」へと昇華しつつあることを如実に示している¹⁵。

PLaMo 2.2 Primeの性能向上：指示追従率と医療分野での評価

■ PLaMo 2.2 Prime ■ 前バージョン (PLaMo 2.1) ■ 海外フロンティアモデル等



PLaMo 2.2 Primeは、前バージョン（2.1 Prime）から大幅な性能改善を達成。特に医療臨床文書の誤り検出（MedRECT-ja F1スコア）では、海外の主要な非推論（non-reasoning）フロンティアモデルを上回る精度を記録した。

Data sources: [Preferred Networks](#)

2.3. 最新フラッグシップ：PLaMo 3.0 Primeの登場と産学官連携の結実

そして、本日2026年6月22日、PFNはさらなる進化を遂げた最新のフラッグシップモデル「PLaMo 3.0 Prime」を正式にリリースした¹⁶。このモデルは、単なる自社単独の開発成果ではなく、国立研究開発法人情報通信研究機構（NICT）との共同研究成果として得られた事前学習モデルをベースに、PFNが独自の最適化と調整を施して開発したものである¹⁷。この事実は、PFNの実装力・計算インフラと、国内の公的研究機関が持つ膨大な学術的知見やデータ資産が融合することで、研究開発のサイクルが極めて高速に回り、産学官連携の枠組みが純国産LLMの基盤技術をいかに加速させているかを示す象徴的なマイルストーンと言える。

3. 産業ニーズに直結する「ドメイン特化型モデル」と「エッジ環境」への適応戦略

PFNのLLM戦略において、汎用モデル（Primeシリーズ）と双璧をなす重要な柱が、特定産業のペインポイントをピンポイントで解消するための「ドメイン特化型モデル（Specialized Models）」と、計算リソースの物理的制約を克服する「軽量化モデル（Small Language Models: SLM）」の多角的な市場展開である⁸。数千億パラメータを持つ汎用巨大モデルをクラウドAPI経由で一律に提供する海外

ビッグテックの画一的な手法とは一線を画し、PFNは日本のエンタープライズ企業が抱える厳格なセキュリティ要件や、製造・物流現場のオンプレミス環境に柔軟に適応する戦略を志向している¹²。

3.1. デジタル庁も採用する軽量・高精度な翻訳特化型「PLaMo Translate」

グローバルビジネスを展開する日本企業にとって、高精度かつセキュアな言語翻訳機能は不可欠である。このニーズに対し、PFNは2025年秋よりサブスクリプションサービスとして「PLaMo Translate」の提供を開始した⁷。これは、日本語と英語のテキスト翻訳タスクに特化して、PFNがスクラッチから独自データセット(日本語と英語の高い比率を含む)を用いて学習させた特化型LLMである⁷。PLaMo Translateの最大の強みは、その「コンパクトなサイズ」と「フォーマット維持能力」にある。一般的な汎用LLMがその巨大さゆえに常時接続のクラウド環境を必須とするのに対し、PLaMo Translateは翻訳に特化したアーキテクチャ設計を採用することでモデルサイズを極めて小さく抑えつつ、海外の巨大モデルを凌駕する翻訳品質と自然な日本語の流暢さを実現している⁷。これにより、一定のスペックを満たすローカルPCやオンプレミス環境での単独稼働が可能となり、企業は機密情報を含む社内文書や研究データを外部ネットワークに一切出すことなく、安全に翻訳できるという決定的な利点を得る⁷。また、会話文、ニュース記事、学術論文といった元のテキストのスタイル(文体)を検出し、忠実に再現する機能も備えている⁷。その実用性の高さやセキュリティ水準は、デジタル庁が推進するガバメントAI「源内」に正式導入され、行政文書などの高度な日本語翻訳に実践投入されていることから強みに裏付けられている¹⁴。

3.2. 高度な機密性を担保する金融特化型「PLaMo-fin-Prime」

金融業界は、厳格なコンプライアンス、個人情報保護、そして情報セキュリティ要件の高さから、パブリッククラウド上の生成AI利用に最も慎重な姿勢を崩していない産業の一つである。この課題をブレイクスルーするため、PFNは金融実務に必要な専門知識を追加学習させた特化型モデル「PLaMo-fin-Prime」を開発した¹³。

このモデルは、高いセキュリティ要件を満たすために各金融機関のセキュアなオンプレミス環境に直接展開することが可能である⁸。さらにPFNは、各社固有の機密データ、商品情報、社内ワークフロー、コンプライアンス規程などをPLaMo-fin-Primeにファインチューニング(追加学習)させるカスタマイズサービスを提供している⁸。このアプローチにより、金融機関は外部へのデータ漏洩リスクを完全に遮断した上で、顧客対応の高度化や専門業務の効率化に生成AIを活用できる。特筆すべきは、グローバルなデータクラウドプラットフォームであるSnowflakeのマーケットプレイス上でも「PLaMo App for Snowflake」として提供されている点である¹。これにより、企業はSnowflake内に蓄積された自社の膨大なデータを移動させることなく、セキュアな環境内でPLaMoの高度な日本語生成・分析能力をシームレスに実行することが可能となっている¹。

3.3. 軽量モデル(SLM)によるエッジAIへのパラダイムシフト

「パラメータ数の増大こそが性能の向上をもたらす」というLLM黎明期のスケーリング則に盲従するのではなく、PFNは学習データの「質」の向上と学習プロセスの最適化によって、小規模パラメータでも巨大モデルを凌駕する性能を引き出す研究開発に注力している¹⁹。現在、PFNが開発を進めている「PLaMo 8B(仮称)」は、わずか80億パラメータという極めて小規模なサイズでありながら、かつて主流であった1000億パラメータ級のモデル(PLaMo-100Bなど)に匹敵、あるいはそれを上回る性能を叩き出していると報告されている¹⁹。

この軽量化技術の結晶である「PLaMo Lite」シリーズなどのスモール言語モデル (SLM) は、自動車 (自動運転車載システム)、製造現場の制御機器、ロボティクスといった、通信帯域や消費電力に厳しい制約があるエッジデバイスへのLLMの直接組み込みを視野に入れたものである⁷。PFNの祖業であり強みであるIoTや産業用ロボットの制御技術と、この軽量かつ高性能な生成AIモデルが融合することで、クラウドに依存せずリアルタイムで自律的な判断を下す「エッジAI」の領域において、PFNは他社の追随を許さない強力なシナジーと技術的優位性を確立しつつある。

4. PFNの垂直統合型エコシステム: 究極のハードウェアとインフラの融合

AIの真の性能と競争力は、「アルゴリズム (基盤モデル)」、「データ」、そしてそれらを駆動する「計算機 (ハードウェア)」の三位一体によって決定される。PFNが国内外の数多あるAIスタートアップと決定的に、そして絶望的なまでの技術的堀 (Moat) を築いて異なるのは、最下層のAI用プロセッサチップから独自のクラウドインフラ、そして最上位のLLMに至るまでの全スタックを自社で設計・保有・最適化している点にある¹。

4.1. MN-Coreシリーズ: 深層学習における圧倒的な電力効率とアーキテクチャ

PFNは、AIモデルが巨大化するはるか前の2016年から、独自のAIプロセッサ「MN-Core」シリーズの開発に莫大なりソースを投じ続けてきた²¹。第一世代のMN-Coreを搭載した深層学習用スーパーコンピュータ「MN-3」は、その驚異的な省電力性能により、スーパーコンピュータの電力効率を競う世界ランキング「Green500」において、2020年6月から2021年11月にかけて実に3度の世界第1位を獲得するという歴史的快挙を成し遂げている²。

以下の表は、MN-3がGreen500で記録した主要な測定結果である。

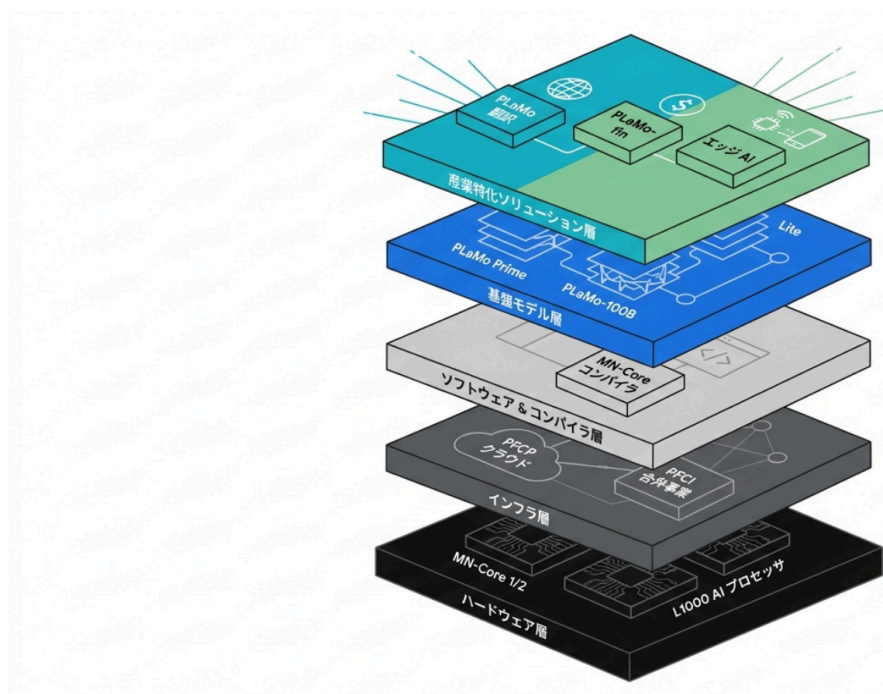
測定時期	ノード数	MN-Core数	CPUコア数	ピーク性能 (理論値)	備考
2021年11月	32ノード	128個	1,536個	3.390 Pflops	3度目の Green500 世界1位
2021年6月	-	-	-	3.138 Pflops	
2020年11月	-	-	-	3.138 Pflops	

2023年からは、性能をさらに向上させた第二世代プロセッサ「MN-Core 2」の本格稼働が開始された²。MN-Coreシリーズの最大の特徴は、一般的なGPUが持つ「データキャッシュ」をあえて持たず、ソフトウェア (コンパイラ) 側で明示的にデータ移動を制御する独自の大規模SIMD (Single Instruction, Multiple Data) アーキテクチャを採用している点である²。これにより、ワット当たりの演算性能だけでなく、シリコン面積当たりの性能も劇的に向上させている²。

2024年9月には、このMN-Core 2チップを8基搭載した「MN-Server 2」や、1チップを搭載したワーク

ステーション「MN-Core 2 Devkit」の提供を開始した²¹。これらの革新的なハードウェア群は、PFN独自のクラウドコンピューティング基盤「PFCP (Preferred Computing Platform)」の中核として稼働しており、自社のPLaMoの膨大な事前学習や、顧客企業向けのセキュアなカスタマイズ事後学習プロセスを、低コストかつ極めて高速に処理する心臓部として機能している¹。

PFNが構築するフルスタックAIアーキテクチャ



PFNのエコシステムは、半導体設計（MN-Core）を土台とし、専用のコンパイラとクラウドインフラ（PFCP）、その上で学習される独自の基盤モデル（PLaMo）群によって構築される完全な垂直統合型モデルである。

4.2. 推論専用プロセッサ「MN-Core L1000」によるゲームチェンジ

PFNのインフラ戦略において、業界構造を根底から覆す可能性を秘めた次なる決定打が、2027年の市場投入に向けて開発が発表された生成AI推論専用プロセッサ「MN-Core L1000」である²¹。

生成AIの運用フェーズ（推論：Inference）において、現在のシステムアーキテクチャが抱える最大のボトルネックは、チップの「演算性能（FLOPS）」ではなく、「メモリ帯域幅（Memory Bandwidth）」にある。LLMがテキストを1トークン生成するごとに、プロセッサは数百ギガバイトに及ぶモデルのウェイト（重み）データをメモリから演算器へと繰り返し読み込む必要があり、この膨大なデータ転送速度の物理的限界が、AIの応答速度（Inference-time scaling）の壁となっている²¹。

PFNはこの構造的課題を打破するため、最新の「3D積層技術（3D-Stacked DRAM）」を大胆に採用し、ロジック（演算器）の真上に大容量メモリを直接積層するという革新的なアプローチを選択した²¹。これにより、現在ハイエンドGPUに標準搭載されているHBM（High Bandwidth Memory）をも凌駕する圧倒的な広帯域メモリと、MN-Coreが元来持つ高密度な演算器・分散メモリのアーキテクチャが

完璧に融合。消費電力と排熱を劇的に抑えつつ、データ転送速度を極限まで引き上げることに成功した²¹。

L1000の登場がもたらす最も破壊的なインパクトは、その提供形態にある。L1000は、大規模なデータセンターや特殊な冷却設備を備えたラックスペースを一切必要とせず、標準的なPCIeスロットに挿入できるワークステーションのフォームファクタで提供される²⁵。これにより、単一のサーバー上で700億(70B)パラメータ以上の長文脈を扱う巨大LLMの超高速推論(従来比で10倍の性能向上)が可能になる²⁵。これは、企業がクラウドの制約から解放され、自社のデスクトップ環境で巨大な生成AIをセキュアかつ低遅延で稼働させる未来(エッジ・オンプレミスへのAI計算資源の劇的な分散化)を現実のものにする革新的な技術である。

4.3. 専用コンパイラによるソフトウェアスタックの高度な最適化

どれほど優れたハードウェアが存在しても、それを制御するソフトウェアが最適化されていなければ、その真価は発揮されない。PFNはこの点を熟知しており、MN-Coreシリーズ専用のコンパイラを自社で開発している²⁷。ユーザーがPyTorchやJAXといった汎用的な高水準言語で定義したAIワークロード(計算グラフ)に変更を加えることなく、コンパイラが自動的に最適化された命令を生成し、データ移動を効率的にスケジューリングする。計算グラフレベルの制御から低レベルな機械語命令の生成まで、抽象度に応じた階層的なコンパイル処理を行うことで、プロセッサの潜在能力を極限まで引き出している²⁷。このハードウェアとソフトウェアの完璧な密結合による擦り合わせこそが、PFNの高速・省電力なAI計算基盤の源泉となっている。

5. 経営戦略とアライアンス: 日本産業界の総意による「AIインフラ」の確立へ

PFNの経営戦略と資本政策を俯瞰すると、同社がもはや一介の有望なAIスタートアップという次元を超え、日本の産業界全体を根底から支える「ナショナル・インフラストラクチャー(国家基盤)」を構築する中核企業として認識され、それに向けた巨大な布石を打っていることが明白となる³。産業用ロボットの制御やデジタルツイン(Field OS)に強みを持つMUJINが、物流や製造の「現場(物理レイヤー)」の自動化に特化しているのに対し、PFNは半導体からLLMまでの「デジタル頭脳(情報インフラ)」そのものをゼロから設計・構築している点で、その事業の射程と資本集約度は根本的に異なる³。

5.1. 巨額の資金調達と日本産業界からの強力なバックアップ

ハードウェア(半導体)の設計・製造、巨大なサーバー群とデータセンターの構築、そして数兆トークンを処理する基盤モデル(LLM)の開発という「フルスタック」の戦略を実行するためには、ソフトウェア専門企業とは比較にならないほど莫大な資本が必要となる³。

PFNは、2024年12月にSBIグループ等をリード投資家とするエクイティファイナンスや金融機関からの借入により、第1クローズとして総額190億円を調達した³。さらに、2025年4月のエクステンションラウンドで50億円を追加調達し、直近のラウンド合計で実に240億円という、国内スタートアップとしては異例の大型資金調達を完了させている³。

ここで最も注目すべきは、調達された金額の大きさ以上に、その株主構成の多様性と重厚さである。積水ハウス系ファンド、信越化学工業、住友生命系ファンド、TBS系ファンド、東京エレクトロン(TEL)、トヨタ自動車、さらには東映アニメーションやワコムといった企業群がPFNを支援している³。製造、

半導体、化学、素材、自動車、通信、金融、メディア、エンターテインメントに至るまで、日本を代表するあらゆる産業のリーディングカンパニーがPFNに出資しているという事実は、日本の産業界全体がPFNを「自社のデジタルトランスフォーメーションとAI活用を担保するコア候補」として見なし、海外の特定ビッグテックへの過度な依存を回避するための「国産AI基盤の確立」に戦略的投資を行っていることの確固たる証左である³。

5.2. 戦略的合併会社の設立：PFCIIによるインフラ提供のスケールアップ

PFNのインフラ拡充戦略は、自社単独での設備投資に留まらない。2024年12月、PFNは三菱商事および株式会社インターネットイニシアティブ（IIJ）と共同で、エンタープライズ向けのAIクラウドサービスを提供する合併会社「Preferred Computing Infrastructure（PFCI）」の設立に合意した²⁴。

この合併事業の枠組みは極めて合理的かつ強力である。PFNが持つ最先端のAI計算アーキテクチャ（MN-CoreプロセッサやPFCPのソフトウェア技術）に、三菱商事が誇るグローバルな事業基盤、データセンター構築力、再生可能エネルギーの調達力を掛け合わせる。さらに、国内ISPの雄であるIIJが持つ商用ネットワークインフラの運用ノウハウと堅牢なデータセンター運用技術を統合する²⁴。これにより、PFN単独では限界があった大規模なAI計算インフラを、極めて安定的かつスケラブルに国内のエンタープライズ市場へ提供する強固な事業体制が完成したのである。

5.3. グリーンAI基盤に向けた協業：Rapidus、さくらインターネットとの「ソブリンAI」サプライチェーン

さらに2025年1月、PFNは日本の半導体復権を担い最先端ロジック半導体の製造を目指すRapidus（ラピダス）および、国内最大手のクラウド・データセンター事業者であるさくらインターネットと、日本発の「グリーンなAIインフラ」の構築に向けた基本合意書を締結した²⁴。

この三社提携は、経済安全保障とテクノロジー戦略の観点から極めて重要な意味を持つ。PFNが「低消費電力かつ高計算性能なAIプロセッサ（MN-Core）」の設計・開発を行い、Rapidusが自社の「最先端の製造プロセスとパッケージング技術」を用いてそのプロセッサを国内で製造する。そして、完成したチップを搭載したサーバー群を、さくらインターネットが「再生可能エネルギーと外気冷却を活用した環境負荷の低い国内データセンター」で運用し、サービスとして提供する³¹。

このサイクルが実現すれば、AI半導体の最上流である「設計」から「製造」、データの保管・処理を行う「データセンターの稼働」、そしてその上で推論を実行する「基盤モデル（PLaMo）」に至るまで、生成AIに関するすべてのコアテクノロジーを国内のサプライチェーン内で完全に完結させることが可能になる。これは、激化する米中AI覇権競争の中において、日本が独自の計算資源と技術的自律性（ソブリンAI）を確保し、エネルギー制約の壁を突破するための最も強力かつ現実的なエコシステムの形成を意味している。

5.4. GENIAC等を通じた官民連携の深化

PFNの技術開発は、民間資金や企業間アライアンスだけでなく、国家的なプロジェクトとも密接に連動して進行している。前述した経済産業省とNEDOが主導する生成AI開発力強化プロジェクト「GENIAC」において、PFNは第1サイクルでのPLaMo-100B開発に加え、翻訳特化型モデル（PLaMo Translate）の開発支援を受け、さらに2024年10月の第2サイクルへも継続採択されるなど、政府から多大な支援と継続的な期待を集めている⁷。

また、2026年6月22日に発表された最新モデル「PLaMo 3.0 Prime」が、国立研究開発法人情報通

信研究機構 (NICT) との共同研究成果をベースに開発されたことから分かるように¹⁶、国内の研究機関が長年蓄積してきた学術的知見・高品質なコーパスデータと、PFNの圧倒的なエンジニアリング実装力・計算インフラが融合することで、日本国内におけるAI研究開発のサイクルはこれまでにない速度で回転し始めている。

6. 結論：産業インフラとしての「AIエコシステム」完成への道程

株式会社Preferred Networksが展開する「PLaMo」シリーズの技術的進化と、それを支える一連の事業戦略は、単にパラメータ数の多さを競うグローバルなソフトウェア開発競争の枠組みに留まるものではない。それは、日本特有の産業構造、高い言語的壁、そしてエンタープライズ企業が求める厳格な機密性・セキュリティ要件という独自の課題に対する、極めて緻密かつ戦略的な解答である。

1. 実用性と専門性の追求：PLaMo 2.2 Primeで証明された、ユーザーの複雑な制約に従う高い指示追従性と、医療 (MedRECT-ja, JMLE) など高度な推論が求められる専門領域における傑出したスコア。さらに金融 (PLaMo-fin-Prime) や行政・ビジネス翻訳 (PLaMo Translate) への特化モデル展開は、PFNがLLMを「汎用的な対話のおもちゃ」から、「実社会のミッションクリティカルなワークフローに組み込まれる産業用エンジン」へと明確に昇華させたことを示している。
2. 推論コストと物理的制約の打破：来たるInference-time scaling (推論スケーリング) の時代において、3D積層DRAM技術を用いた推論専用プロセッサ「MN-Core L1000」の開発は、巨大なデータセンターへの過度な依存を断ち切り、LLMの計算リソースをエッジや各企業のオンプレミス環境 (ワークステーション) へと解放する画期的なアプローチである。これにより、導入コストとレイテンシの壁が劇的に下がり、あらゆる産業機器へのAI実装が加速する。
3. 国産サプライチェーンの強靱化とソブリンAIの確立：製造、素材、金融、メディアに及ぶ幅広い重厚長大な株主構成と、三菱商事、IIJ、Rapidus、さくらインターネットら国内トップ企業との戦略的アライアンスは、PFNが一企業としての利益追求を超え、日本産業界の総意としての「コアAIインフラ」の確立を背負っていることを意味する。

PFNの経営戦略の真髄は、「アルゴリズム (PLaMo)」、「ハードウェア (MN-Core)」、「計算基盤 (PFCP/PFCI)」の全レイヤーを垂直統合することで、AI時代の新たな頭脳から神経網に至るまでの全てを自らの手でデザインし、それをオープンかつセキュアな形で産業界へ提供することにある。2026年6月22日のPLaMo 3.0 Primeのリリースは、この壮大なエコシステム構築における一つの通過点に過ぎない。PFNは今後も、「物理世界 (現実世界) を真の意味で計算可能にする」という創業からのビジョンのもと、日本初にして世界最高峰の産業用AIインフラストラクチャーとしての地位を不動のものにし、グローバル市場において極めてユニークかつ不可欠なプレイヤーとして君臨していくと考えられる。

引用文献

1. Preferred Networks, Inc. | Snowflake Marketplace, 6月 22, 2026にアクセス、
<https://app.snowflake.com/marketplace/providers/GZT2ZX8M74/Preferred%20Networks%2C%20Inc.>
2. Second-generation processor of MN-Core architecture for AI and general-purpose HPC application - Hot Chips 2024 -, 6月 22, 2026にアクセス、
https://hc2024.hotchips.org/assets/program/conference/day2/15_HC2024.Preferr ed.Makino.final.pdf
3. Who is Preferred Networks? | 香川友志 - note, 6月 22, 2026にアクセス、

- <https://note.com/kagawatomo/n/ndd12ded20946?hl=en>
4. PLaMo, 6月 22, 2026にアクセス、<https://plamo.preferredai.jp/>
 5. PLaMo-100B: A Ground-Up Language Model Designed for Japanese Proficiency - arXiv, 6月 22, 2026にアクセス、<https://arxiv.org/html/2410.07563v1>
 6. PFN Releases PLaMo-13B Open-Source Large Language Model in Japanese and English for Research and Commercial Use, 6月 22, 2026にアクセス、
<https://www.preferred.jp/en/news/pr20230928>
 7. PFN Launches PLaMo Translate Large Language Model Specialized for Japanese Text Translation - Preferred Networks, Inc., 6月 22, 2026にアクセス、
<https://www.preferred.jp/en/news/pr20250527>
 8. Generative AI foundation models - Business - Preferred Networks, Inc., 6月 22, 2026にアクセス、<https://www.preferred.jp/en/business/genai>
 9. pfnnet/plamo-100b - Hugging Face, 6月 22, 2026にアクセス、
<https://huggingface.co/pfnnet/plamo-100b>
 10. PLaMo-100B: A Ground-Up Language Model Designed for Japanese Proficiency - arXiv, 6月 22, 2026にアクセス、<https://arxiv.org/abs/2410.07563>
 11. Post-training completed for PLaMo-100B, a proprietary LLM with 100 billion parameters, 6月 22, 2026にアクセス、
<https://tech.preferred.jp/en/blog/plamo-100b-post-training/>
 12. PFN Launches Flagship Japan-Made Large Language Model PLaMo Prime, 6月 22, 2026にアクセス、<https://www.preferred.jp/en/news/pr20241202>
 13. PFN's PLaMo Translate to Be Adopted by Japan's Government AI Project "Gennai", 6月 22, 2026にアクセス、<https://www.preferred.jp/en/news/657>
 14. PFNの国産大規模言語モデルPLaMo 2.0 Prime、2025年日経優秀製品・サービス賞にて最優秀賞を受賞 | 株式会社Preferred Networksのプレスリリース - PR TIMES, 6月 22, 2026にアクセス、
<https://prtimes.jp/main/html/rd/p/000000020.000156310.html>
 15. PLaMo 2.2 Primeをリリースしました - Preferred Networks Tech Blog, 6月 22, 2026にアクセス、<https://tech.preferred.jp/ja/blog/plamo-2-2-prime-release/>
 16. 国産生成AI基盤モデルPLaMo 3.0 Primeを正式リリース - 株式会社 ..., 6月 22, 2026にアクセス、<https://www.preferred.jp/ja/news/pr20260622>
 17. Preferred Networksが国産生成AI基盤モデル「PLaMo 3.0 Prime」をリリース | 2026年 - NICT, 6月 22, 2026にアクセス、
<https://www.nict.go.jp/publicity/topics/2026/06/22-1.html>
 18. Preferred Networks, Inc.: PLaMo Fin Prime 1.0 for Snowflake, 6月 22, 2026にアクセス、
<https://app.snowflake.com/marketplace/listing/GZT2ZX8M7G/preferred-networks-inc-plamo-fin-prime-1-0-for-snowflake>
 19. PLaMo 2 Technical Report - arXiv, 6月 22, 2026にアクセス、
<https://arxiv.org/html/2509.04897v1>
 20. PFN、新LLM「PLaMo 2 8B」(仮称) 開発中 / パラメータ数1000億→80億に、でも性能は維持, 6月 22, 2026にアクセス、
<https://www.itmedia.co.jp/aipplus/article/2502/18/1250218125/>
 21. PFN Begins Development of Generative AI Processor MN-Core L1000, 6月 22, 2026にアクセス、<https://www.preferred.jp/en/news/pr20241115>

22. PFNの深層学習用スーパーコンピュータMN-3、39.38GFlops/Wの電力効率を記録しGreen500ランキングで3度目の世界1位を獲得 | 神戸大学ニュースサイト - Kobe University, 6月 22, 2026にアクセス、
https://www.kobe-u.ac.jp/ja/news/article/2021_11_16_01/
23. ノーコードで言語モデルの「学習」を体験できるMN-Core Playground / SLM Customizeの遊び方, 6月 22, 2026にアクセス、
<https://tech.preferred.jp/ja/blog/mncore-playground-slm-customize/>
24. News - Preferred Networks, Inc., 6月 22, 2026にアクセス、
<https://www.preferred.jp/en/news/page/3>
25. MN-Core L1000 - A Chip for the Inference-Time Scaling Era, 6月 22, 2026にアクセス、
<https://mn-core.com/>
26. News - Preferred Networks, Inc., 6月 22, 2026にアクセス、
<https://www.preferred.jp/en/news/tag/88>
27. AI Chips - Business - Preferred Networks, Inc., 6月 22, 2026にアクセス、
<https://www.preferred.jp/en/business/chips>
28. 総額190億円の資金調達を実施 - 株式会社Preferred Networks, 6月 22, 2026にアクセス、
<https://www.preferred.jp/news/pr20241223>
29. エクステンションラウンドで総額50億円の資金調達を実施 - 株式会社Preferred Networks, 6月 22, 2026にアクセス、
<https://www.preferred.jp/news/pr20250430>
30. PFN、三菱商事、IJがAI向けクラウドサービスを提供する合弁会社 Preferred Computing Infrastructureを設立 - Mitsubishi Corporation, 6月 22, 2026にアクセス、
<https://www.mitsubishicorp.com/jp/ja/news/release/2024/20241223002.html>
31. Preferred Networks、Rapidus、櫻花互联网, 达成基本协议, 为提供对绿色社会贡献的国产AI基础设施而努力。 - Moomoo, 6月 22, 2026にアクセス、
<https://www.moomoo.com/hans/news/post/47868763/preferred-networks-rapidus-and-sakura-internet-have-signed-a-basic>
32. GENIAC第2サイクルに継続採択 - 株式会社Preferred Networks, 6月 22, 2026にアクセス、
<https://www.preferred.jp/news/pr20241010>