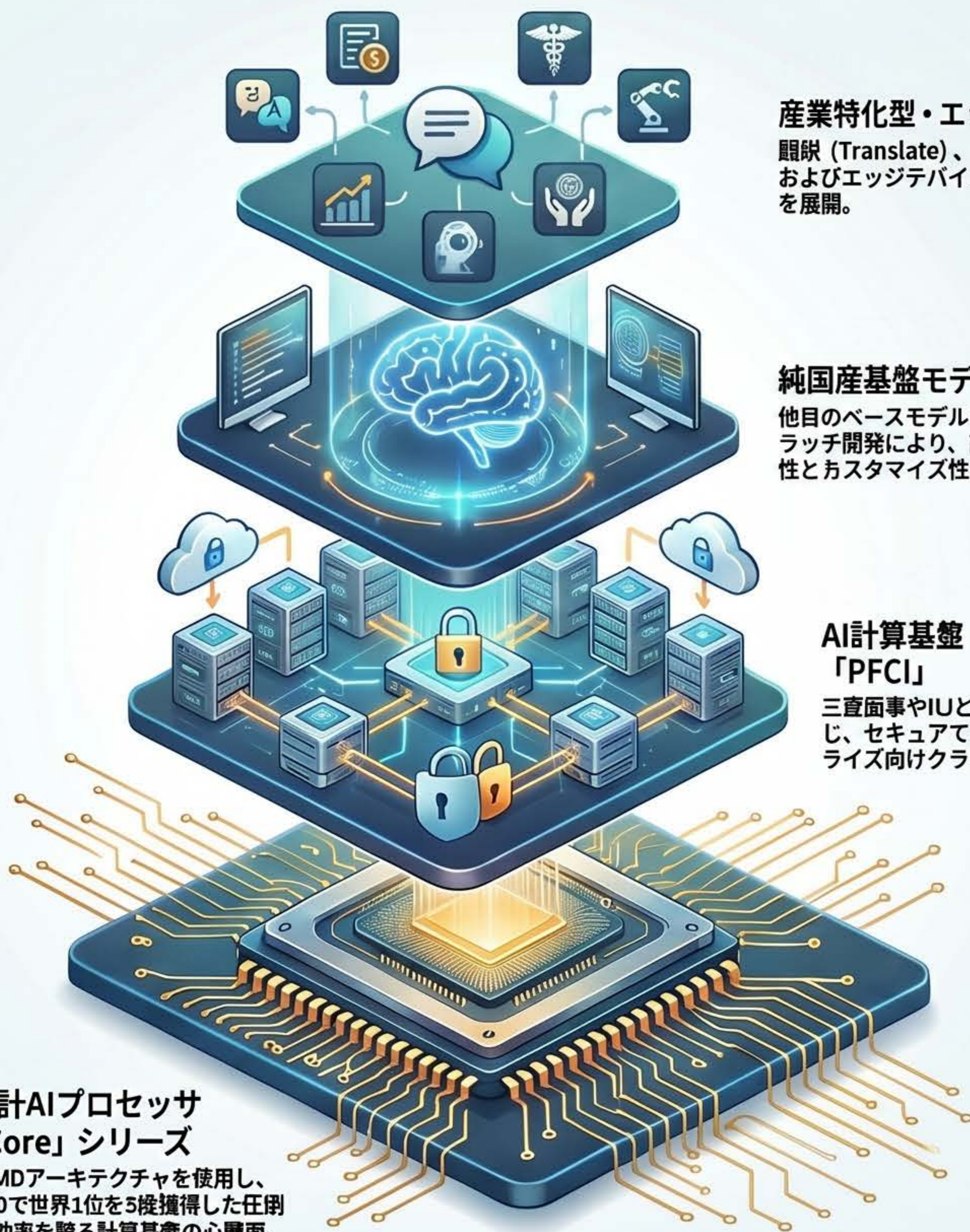


PFNの垂直統合型AI経営戦略：半導体からLLMまで、日本発「AIインフラ」の全貌

フルスタックAIアーキテクチャ



産業特化型・エッジAIソリューション
 翻訳 (Translate)、金融 (fin-Prime)、医療、およびエッジデバイス向けの軽量モデル (SLM) を展開。

純国産基盤モデル「PLaMo Prime」
 他目的のベースモデルを一切使用しないフルスタック開発により、前用効用における高い透明性とカスタマイズ性を実現。

AI計算基盤「PFCP」および「PFCI」
 三直面事やIUとの合併会社 (PFCI) を通じ、セキュアでスケーラブルなエンタープライズ向けクラウド環境を提供。

自社設計AIプロセッサ「MN-Core」シリーズ
 独自のSIMDアーキテクチャを使用し、Green500で世界1位を5度獲得した圧倒的な電力効率を誇る計算基盤の心臓面。

国産LLM「PLaMo」進化の系譜



2023年：
 PLaMo-13Bのリリース
 日本発・英産の1.4兆トークンを学習。コンパクトながら毎時の世界トップクラスの性能を達成し、無償公開。

2024年：
 PLaMo-100Bへの飛躍
 GENIACプロジェクトの支援を受け、2兆トークンを学習。英語での種別推論能力と素精緻な日本語能力を両立する戦略的カリキュラム学習を採用。

2026年：
 PLaMo 2.2 Primeによる「指示追従性」の極大化
 独自の評価ベンチマーク「JFBench」を創設し、指示追従率を7.03%から23.7%へと大幅に改善。

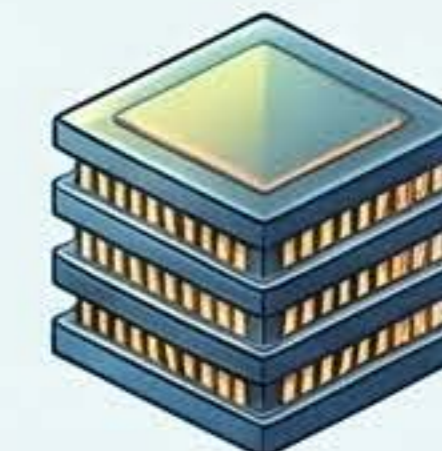
最新フラッグシップ：
 PLaMo 3.0 Prime
 NICTとの共同研究成果をベースに開発され、液状普通集による幾連な研究開発サイクルを象徴する最新モデル (2026年6月12日リリース)。

実社会を計算可能にする「専門性」と「効率性」

PLaMo Translate (翻訳特化型)
 デジタル庁のカバメントAI「通内」に採用、オンプレミス環境での稼働が可能で、高度な機密性を確保。



医療分野における卓越した推論能力
 臨床文書の誤り検出 (MedRECT-je) において、GPT-4.1 mini等の海外トップモデルを上回る精度を記録。



次世代推論プロセッサ「MN-Core L1000」
 3D構造ORAM型により、メモリ構造のボトルネックを軽減、ワークステーションサイズで巨大LLMの高速推論を実現し、エッジAIを加速。

PLaMo 2.2 Primeが海外フロンティアモデルを上回る医療分野の性能を示し、実用性の高さを示す。

評価項目 (ベンチマーク)	PLaMo 2.2 Prime	GPT-4.1 mini / 4.1
MedRECT-ja (誤り検出 F1スコア)	0.661	0.658
MedRECT-ja (誤り検出 正検率)	57.0%	52.6%
JMLE (日本医療用語訳 正検率)	70.7%	94.3%

日本産業界の総意としての「ソブリンAI」戦略

国内主要企業による
 240億円の出資



トヨタ、三井物産、東京エレクトロン、SRI、信越化学など、日本を代表する多様な企業がPFNを支援。

国産「グリーンAI」
 サプライチェーンの構築

さくらインターネット

PFN (設計)、Rapidus (製造)、さくらインターネット (データセンター) の連携により、国内産製型の持続可能なAIインフラを構築。