

株式会社Preferred Networks (PFN) の全貌 日本発「垂直統合型AIインフラ」の経営戦略と進化



ソブリンAIの必然性：データ主権と産業競争力の死守

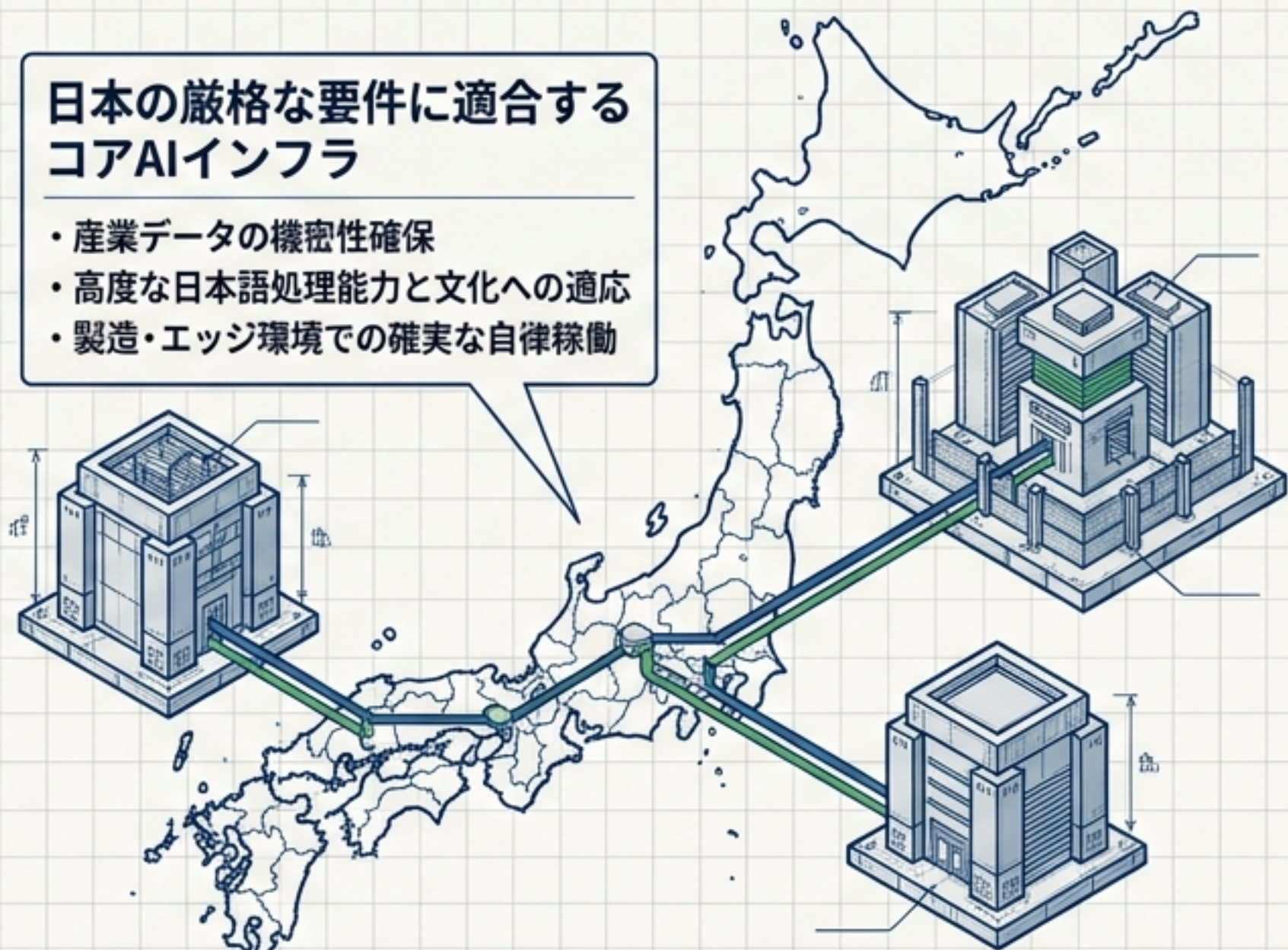
グローバル巨大LLMへの依存リスク

- データ主権の喪失
- 海外ブラックボックスモデルへの過度な依存

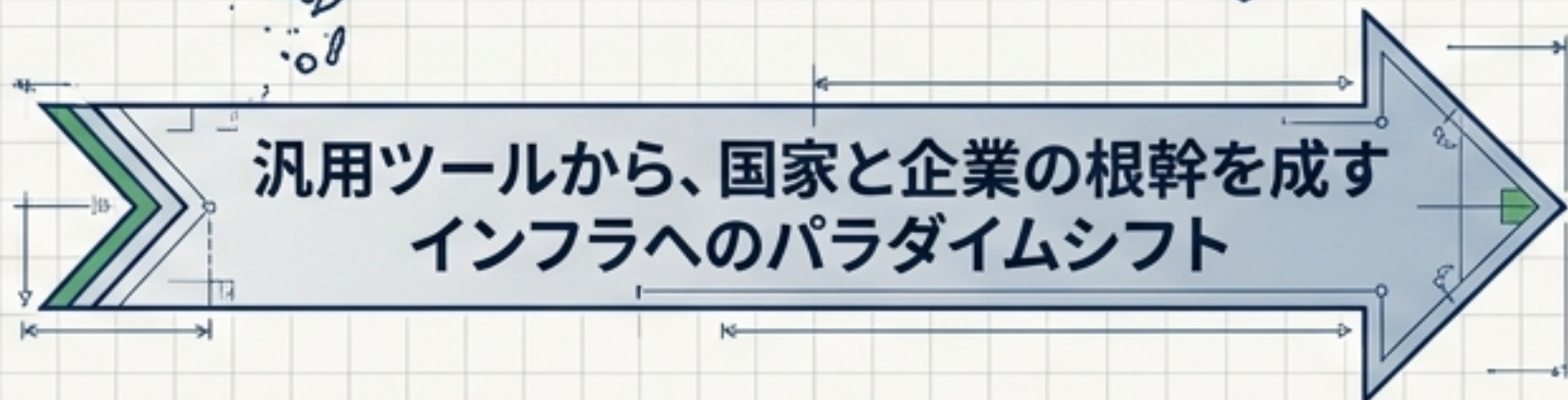


日本の厳格な要件に適合するコアAIインフラ

- 産業データの機密性確保
- 高度な日本語処理能力と文化への適応
- 製造・エッジ環境での確実な自律稼働

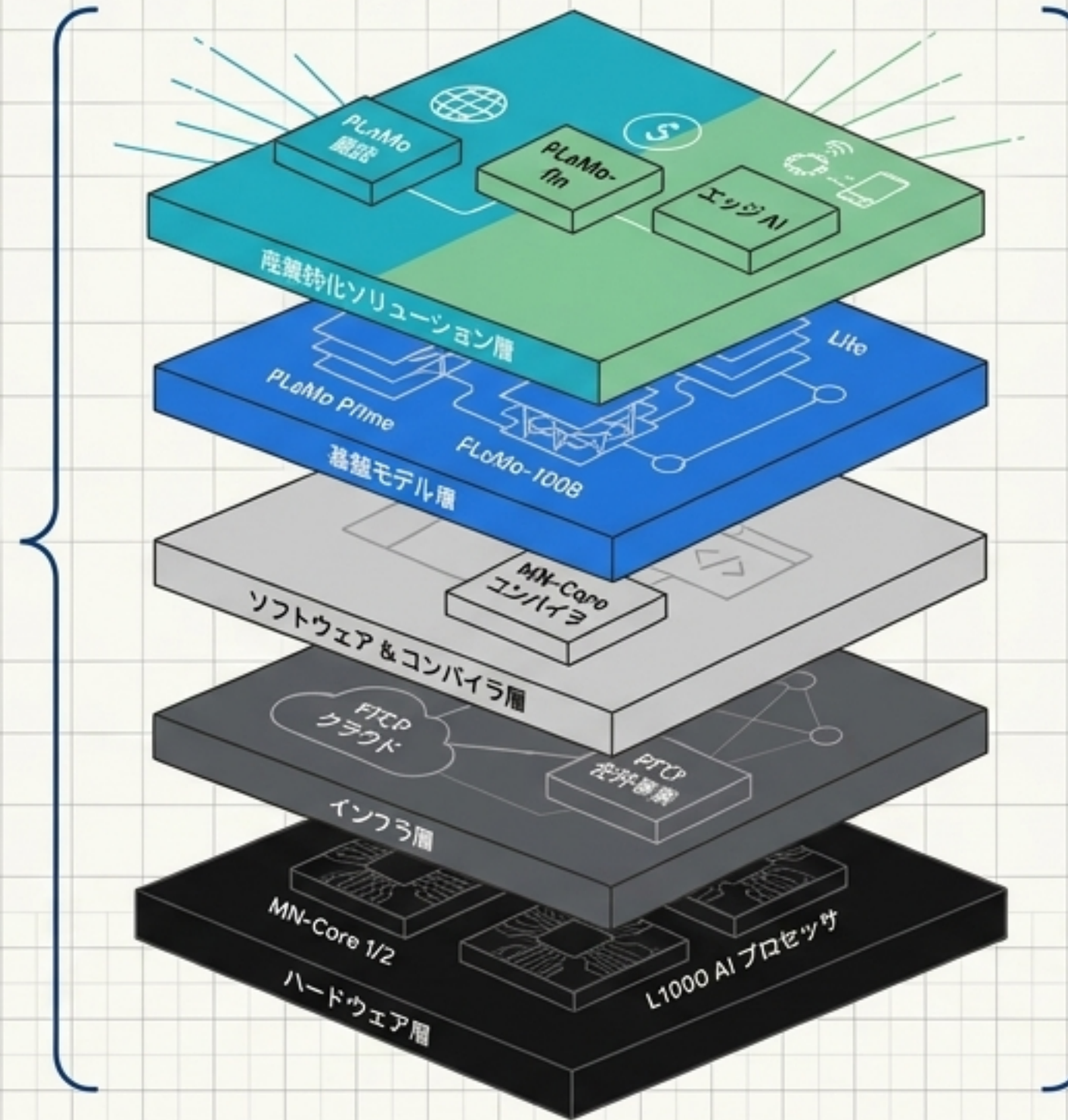


汎用ツールから、国家と企業の根幹を成す
インフラへのパラダイムシフト



類を見ない技術的堀：他社を圧倒する「垂直統合型（フルスタック）」アプローチ

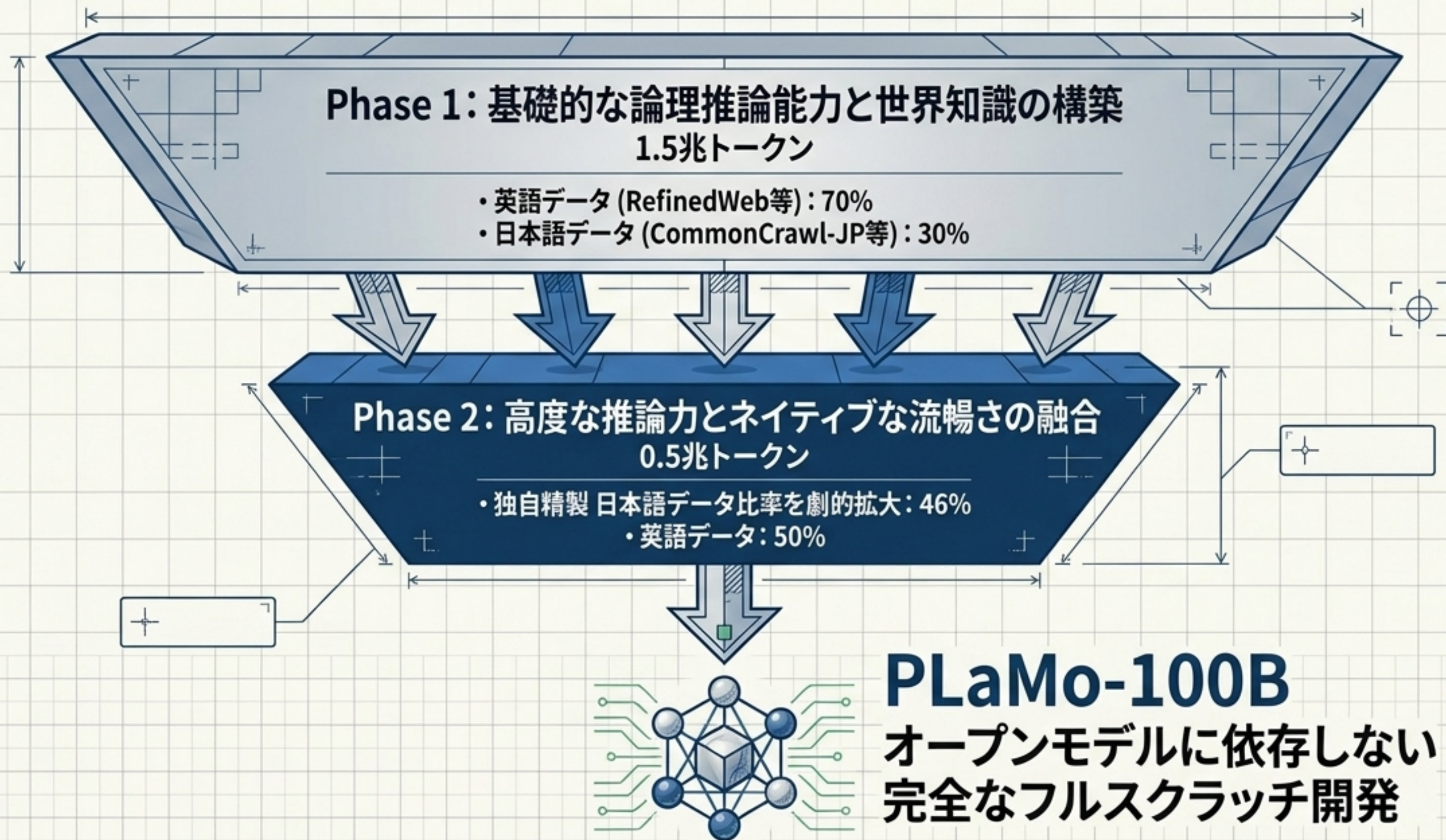
「現実世界を計算可能にする
(Make the real world
computable)」



他社には模倣不可能な
性能とコスト優位性

最下層のAI用プロセッサチップから最上位の特定産業向けLLMに至るまで、全階層を自社で設計・最適化。APIラッパーに留まる他社とは一線を画す、圧倒的な物理・デジタル統合基盤。

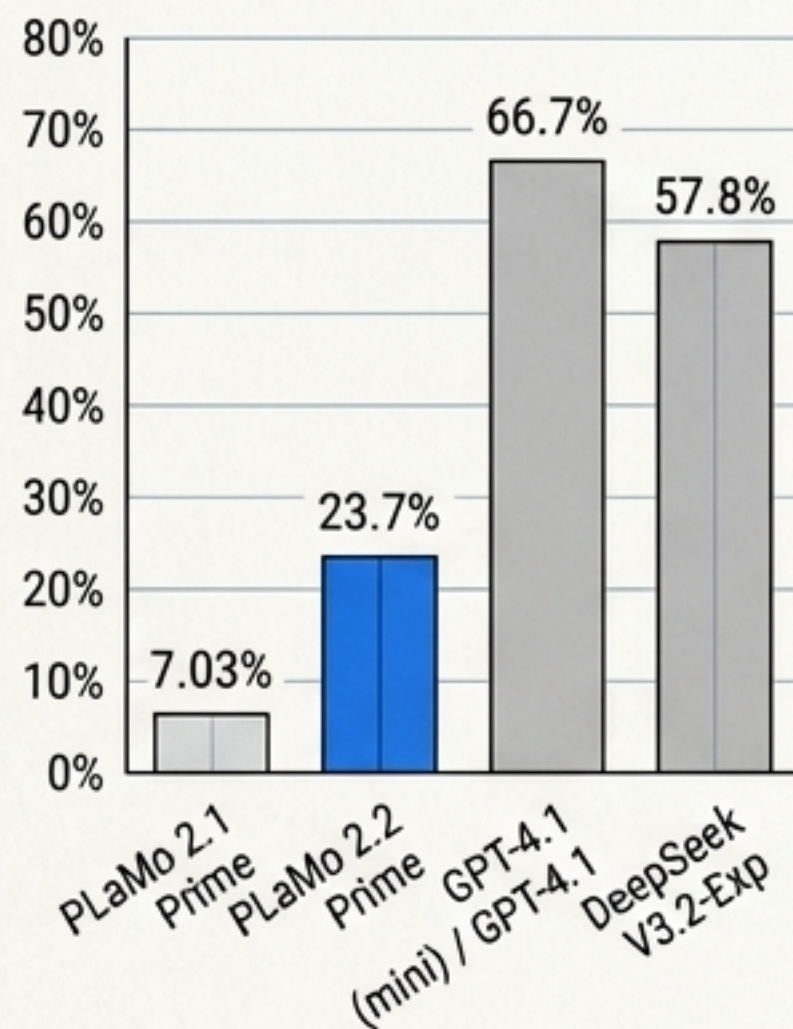
PLaMo-100Bの独自構築：戦略的「カリキュラム学習」の全容



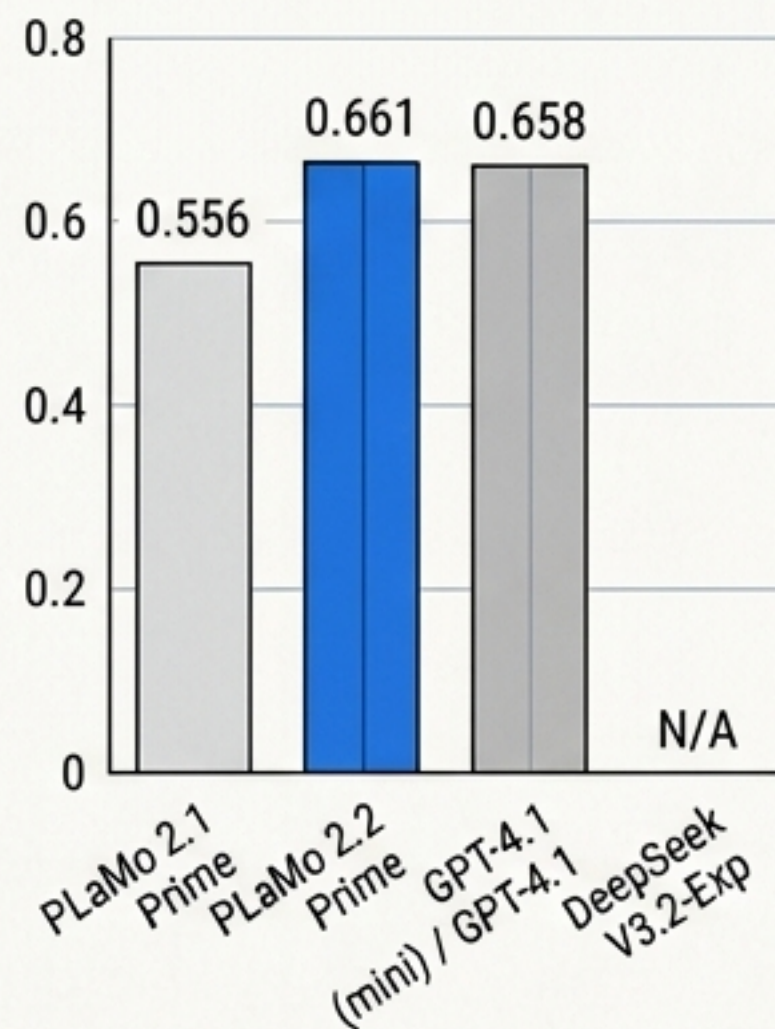
規模から「指示追従性」と「論理的推論」への進化： PLaMo 2.2 Prime & 3.0 Prime

■ 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

Multi-turnロールプレイ
指示追従率 (%)



MedRECT-ja 臨床文書
誤り検出 (F1スコア)



指示追従性の劇的改善

Multi-turnロールプレイ指示追従率が、PLaMo 2.1 Primeの7.03%から、2.2 Primeでは23.7%へと大幅に跳ね上がり、複雑なプロンプトや制約への厳格な対応能力を証明。

医療・専門ドメインにおける推論のブレイクスルー

- MedRECT-ja (誤り検出 F1スコア) : 0.661 (海外フロンティアモデルを凌駕)
- JMLE (日本医師国家試験) 正解率 : 55.1% から 70.7% (394/557問) へ向上。

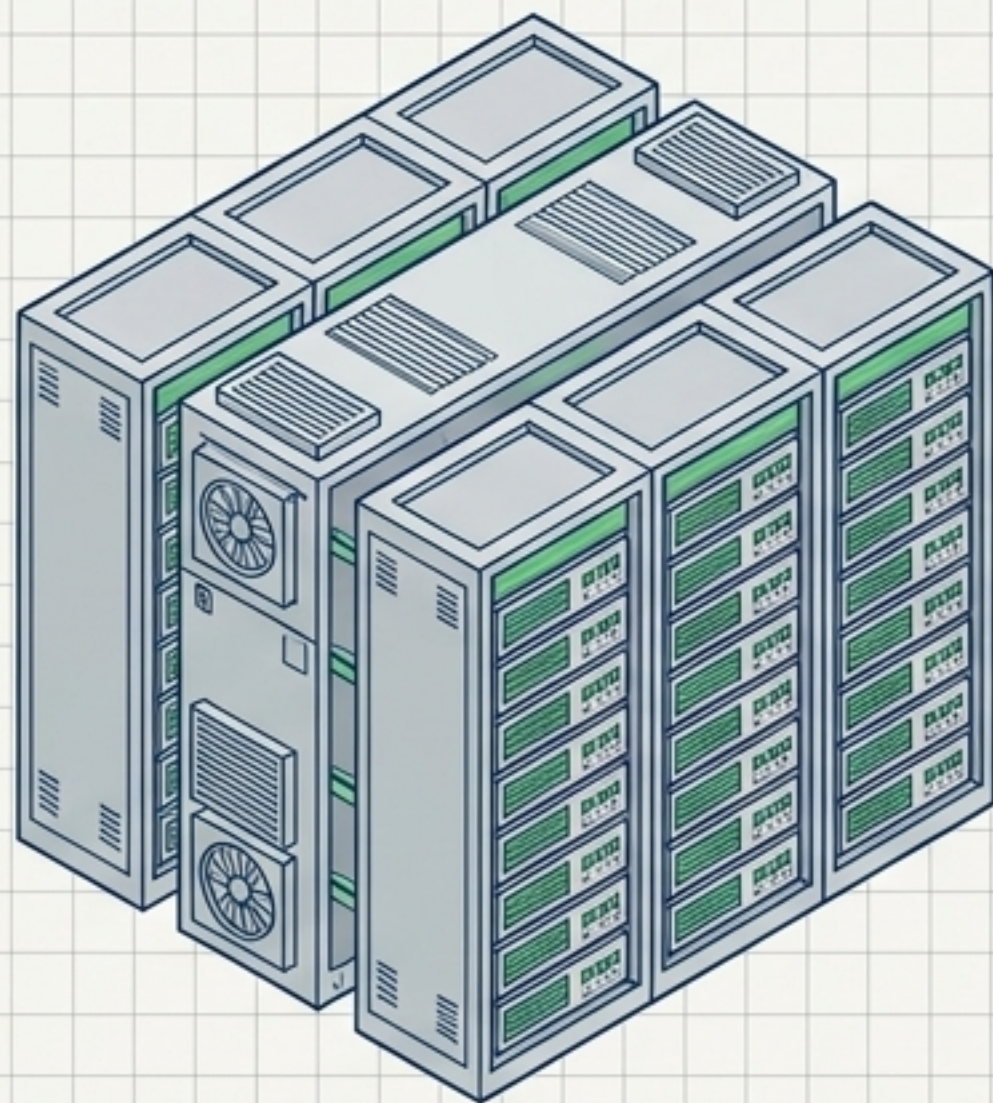
産学官連携の結実

PLaMo 3.0 Prime : NICT (情報通信研究機構) との共同研究成果をベースに、PFN独自の最適化を施した最新フラッグシップモデル。

産業課題をピンポイントで解決する「特化型モデル」マトリクス

PLaMo Translate (翻訳特化)	PLaMo-fin-Prime (金融特化)	PLaMo Lite / 8B (エッジ特化)
【特性】 コンパクトサイズでの高度なフォーマット維持機能。	【特性】 金融実務の専門知識を追加学習。機密データの安全なファインチューニング。	【特性】 1000億クラスに匹敵する極小80億パラメータのモール言語モデル(SLM)。
【展開環境】 セキュアなオンプレミス / ローカルPC。	【展開環境】 オンプレミス展開 / Snowflake内での直接実行 (PLaMo App for Snowflake)。	【展開環境】 通信・電力制約の厳しいエッジデバイス。
【導入実績】 デジタル庁ガバメントAI「源内」に正式導入。	【導入実績】 厳格なコンプライアンス要件を持つ金融機関向け。	【導入実績】 製造機器、自動運転、ロボティクスへの直接組み込み。

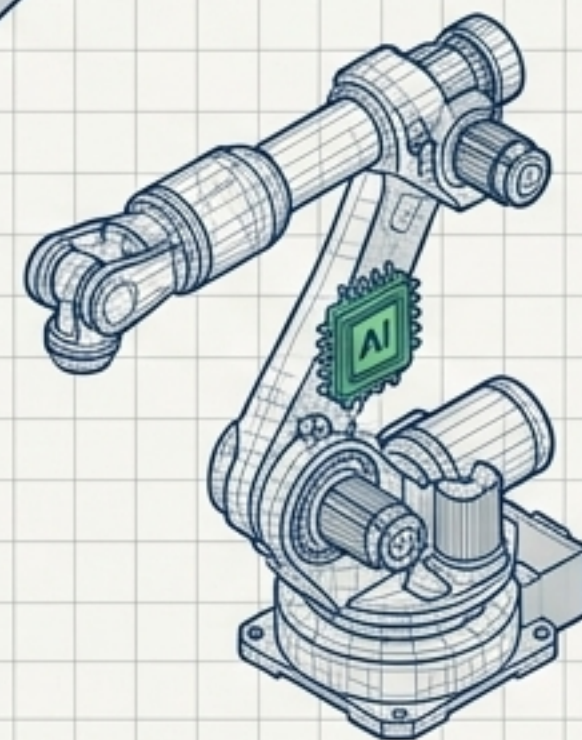
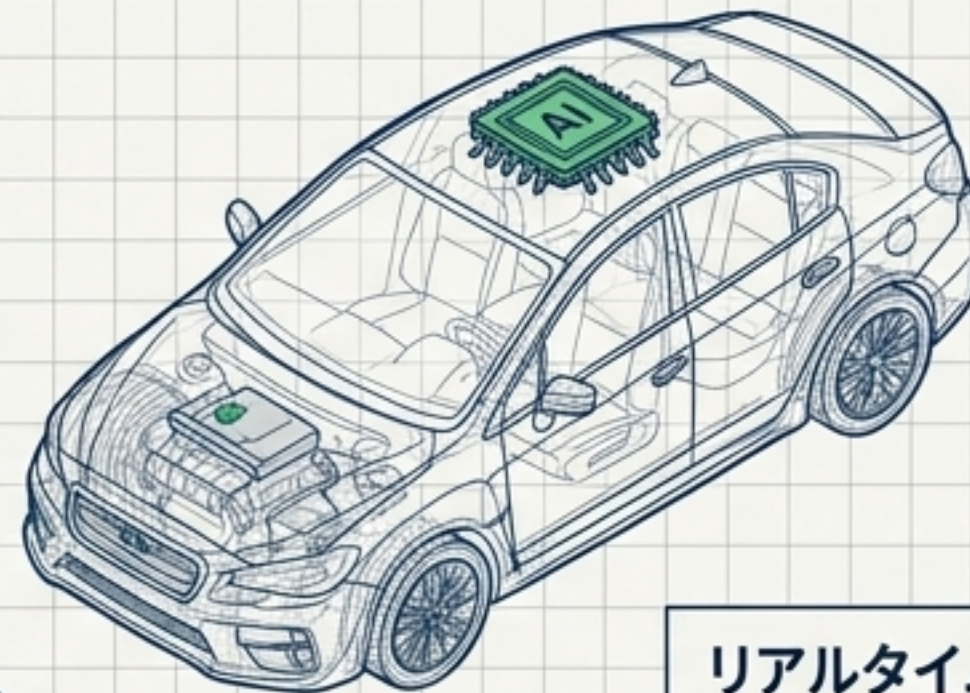
クラウド依存を脱却する「エッジAI」へのパラダイムシフト



従来の汎用巨大LLM
(クラウド依存・高遅延・高消費電力)



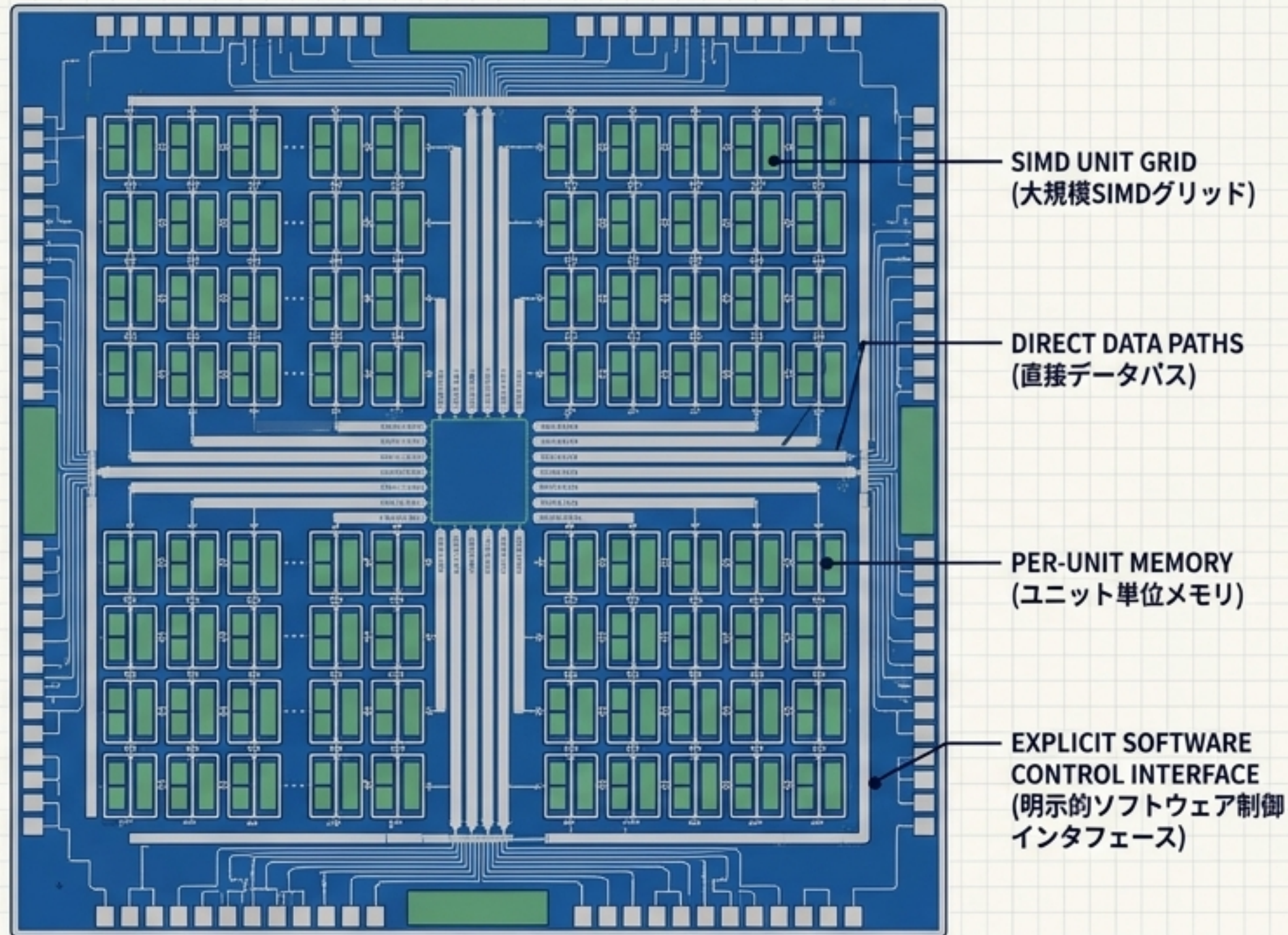
PLaMo 8B:
わずか80億パラメータで
1000億クラスの性能を発揮



リアルタイム自律判断の実現

通信帯域や消費電力の制約が極めて厳しい自動車(自動運転車載システム)、製造現場の制御機器、ロボティクスへのLLM直接組み込みを可能にする。

究極の物理レイヤー：圧倒的な電力効率を誇る独自AIプロセッサ「MN-Core」



独自の大規模SIMDアーキテクチャ。ハードウェアキャッシュを持たず、ソフトウェアでデータ移動を明示的に制御。

世界No.1の電力効率実績 MN-Core搭載スーパーコンピュータ「MN-3」



2020年6月: Green500 第1位



2020年11月: Green500 第1位

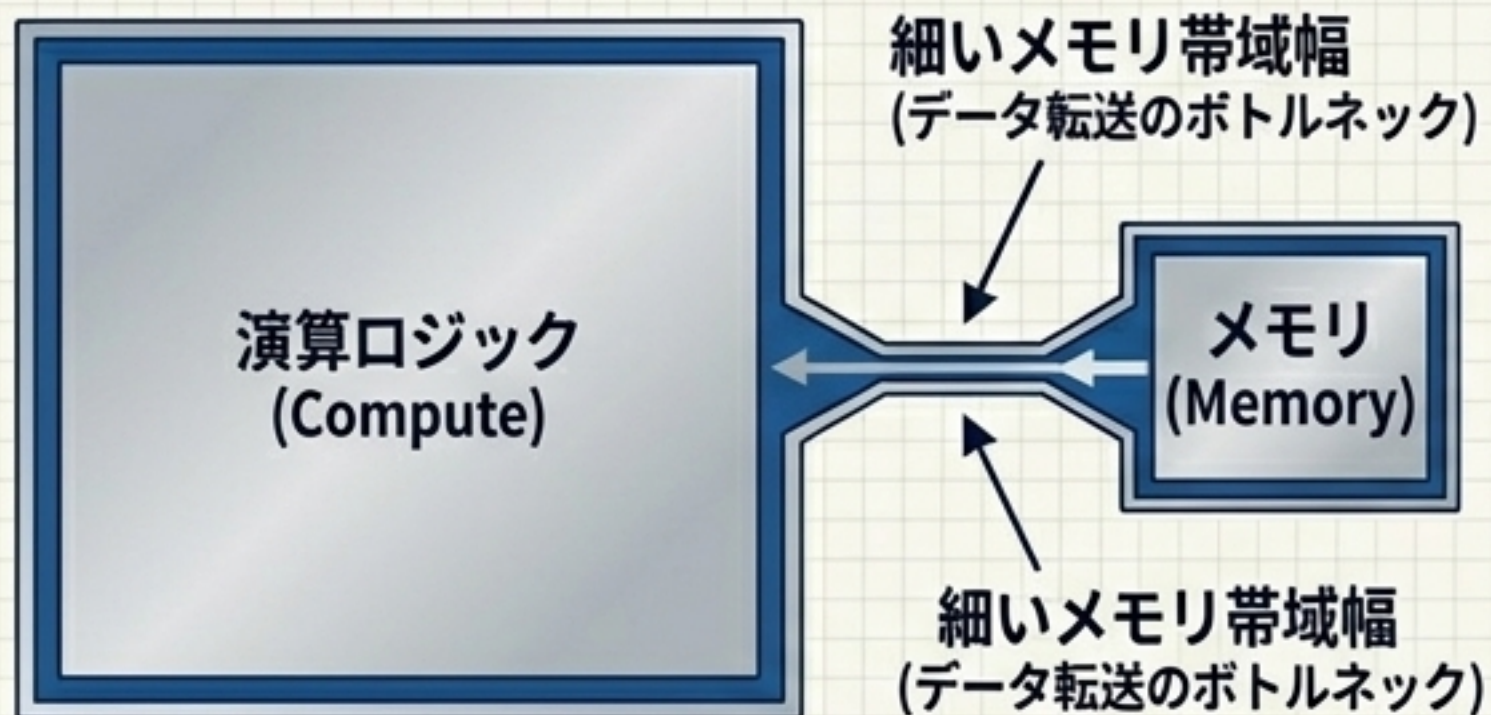


2021年11月: Green500 第1位
(39.38 GFlops/W を記録)

現在：第2世代「MN-Core 2」がクラウドインフラPFCPの中核として稼働中。

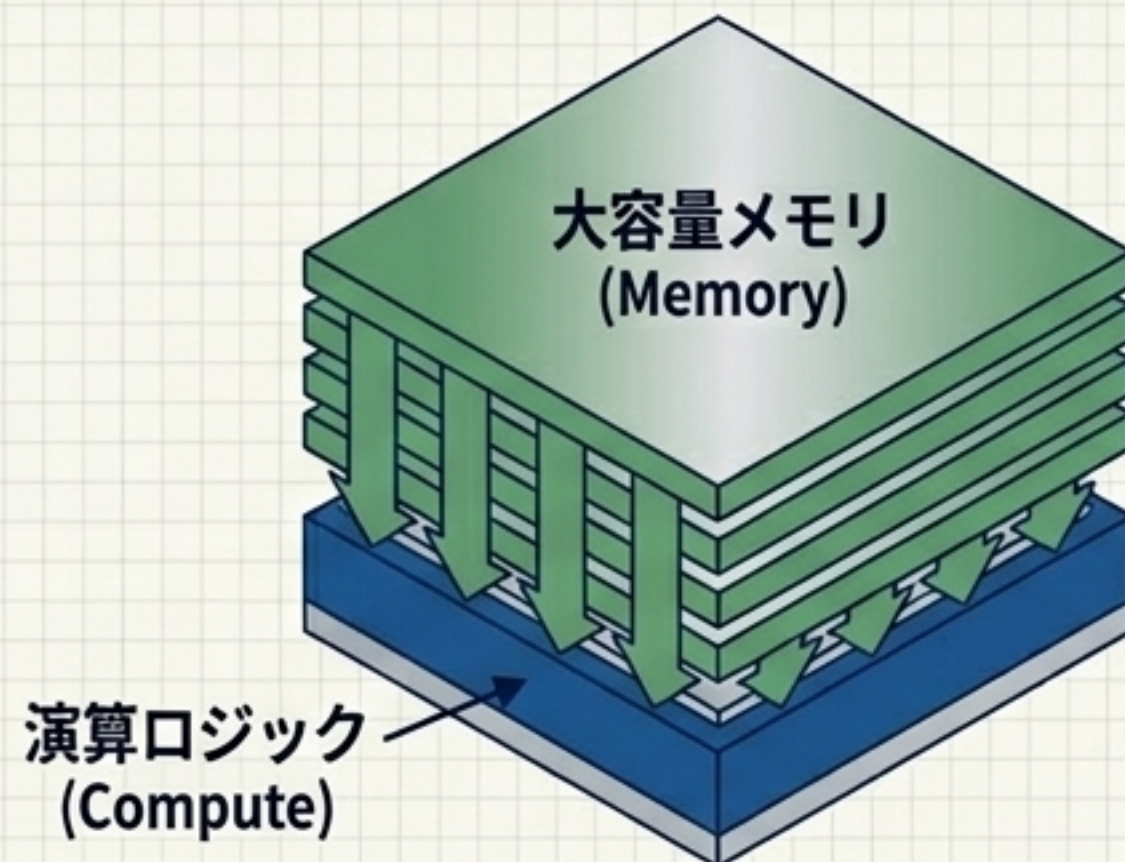
推論時代の「メモリの壁」を打破する：生成AI推論専用プロセッサ「MN-Core L1000」

標準的なGPU (Standard GPU)



従来の推論: 演算性能(FLOPS)は高くても、データ転送速度が律速となる「メモリの壁」が発生。

MN-Core L1000 (3D-Stacked DRAM)



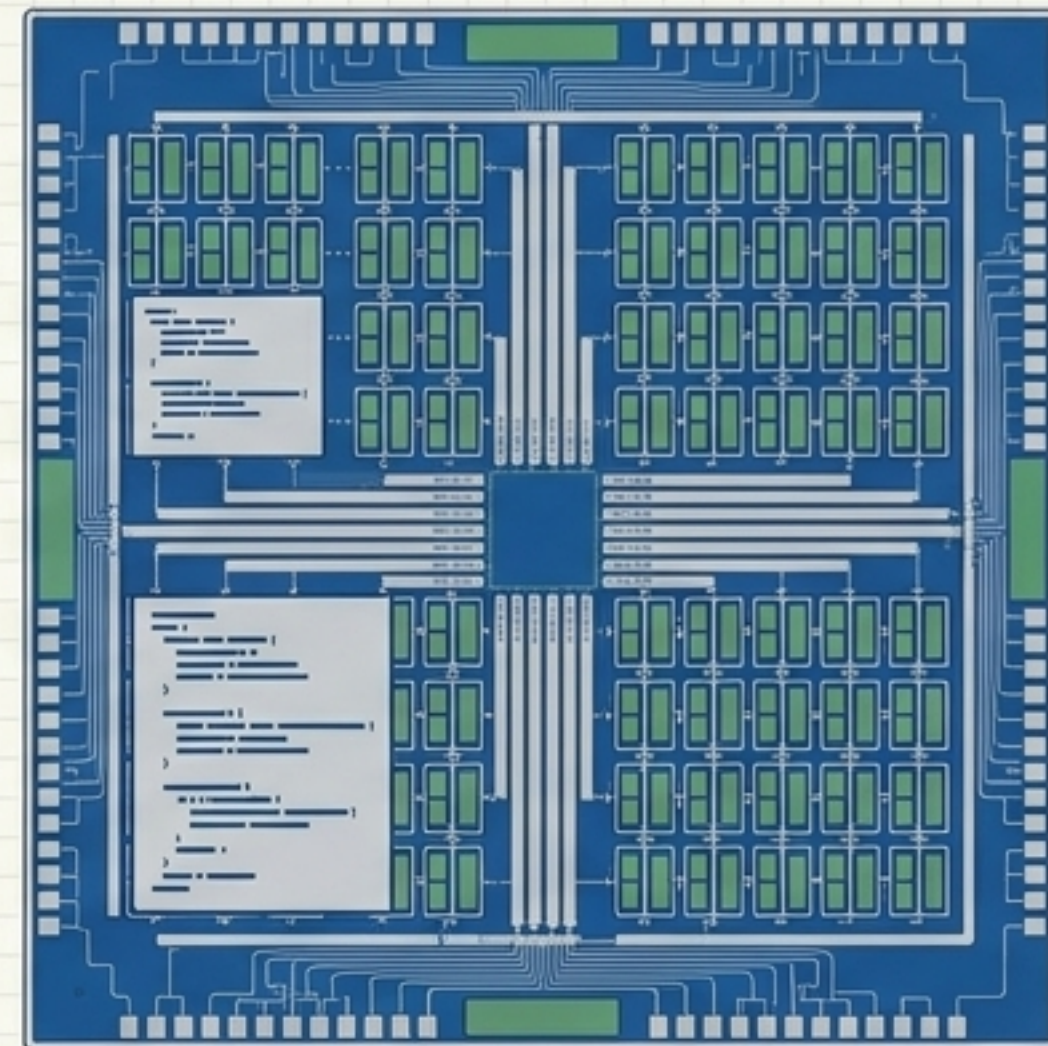
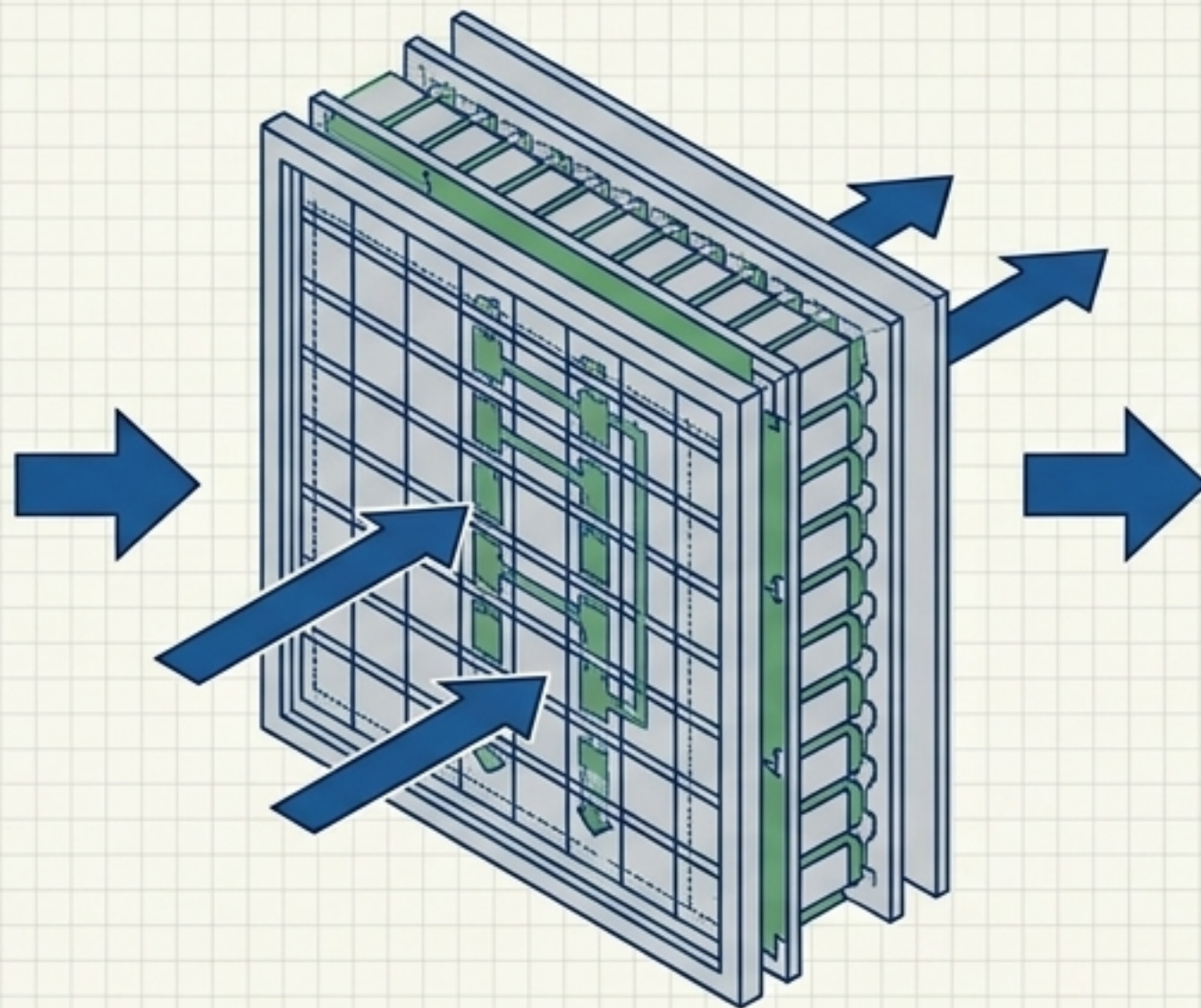
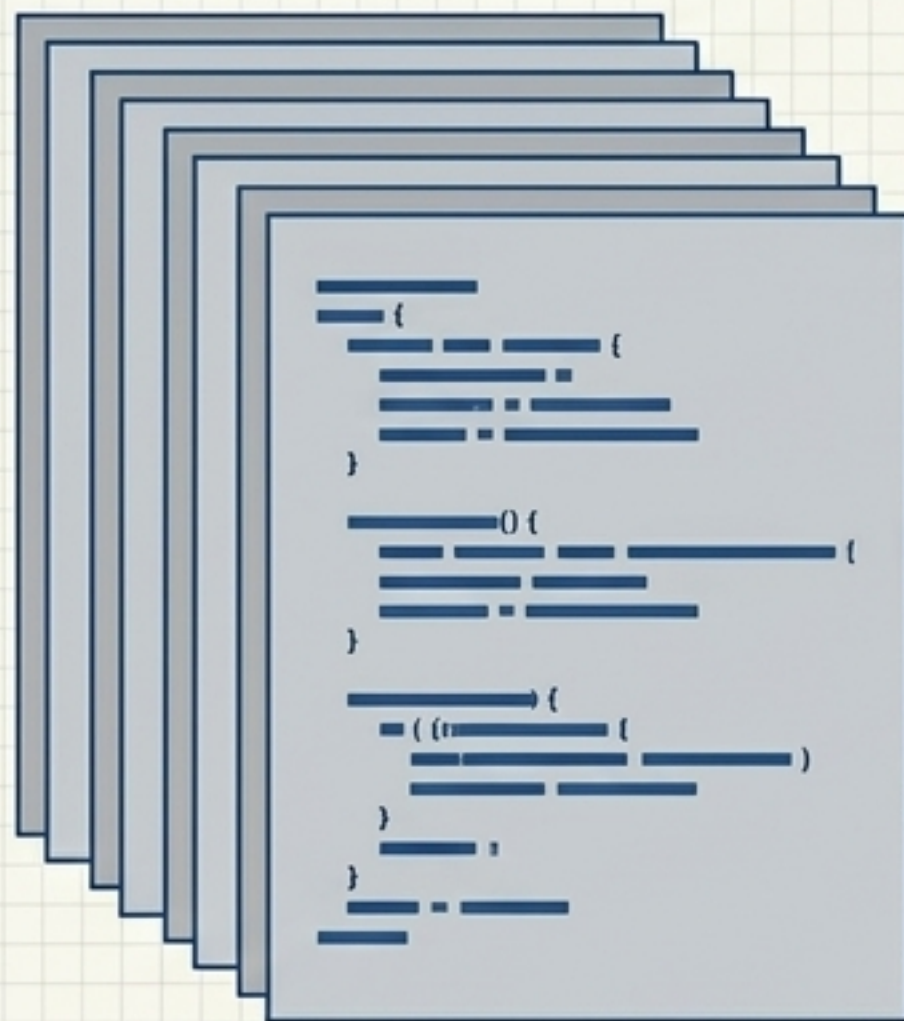
MN-Core L1000: 3D積層技術によりロジックの真上にメモリを直接配置。超広帯域通信を実現。

- AI計算資源の分散化：巨大データセンター不要。
- PCIeスロット搭載の標準ワークステーション単体で、70B(700億)パラメータの巨大LLMを超高速推論可能に。

ハードとソフトの完璧な密結合：専用コンパイラによる限界突破

汎用的な高水準言語
(PyTorch / JAX)

PFN専用コンパイラ

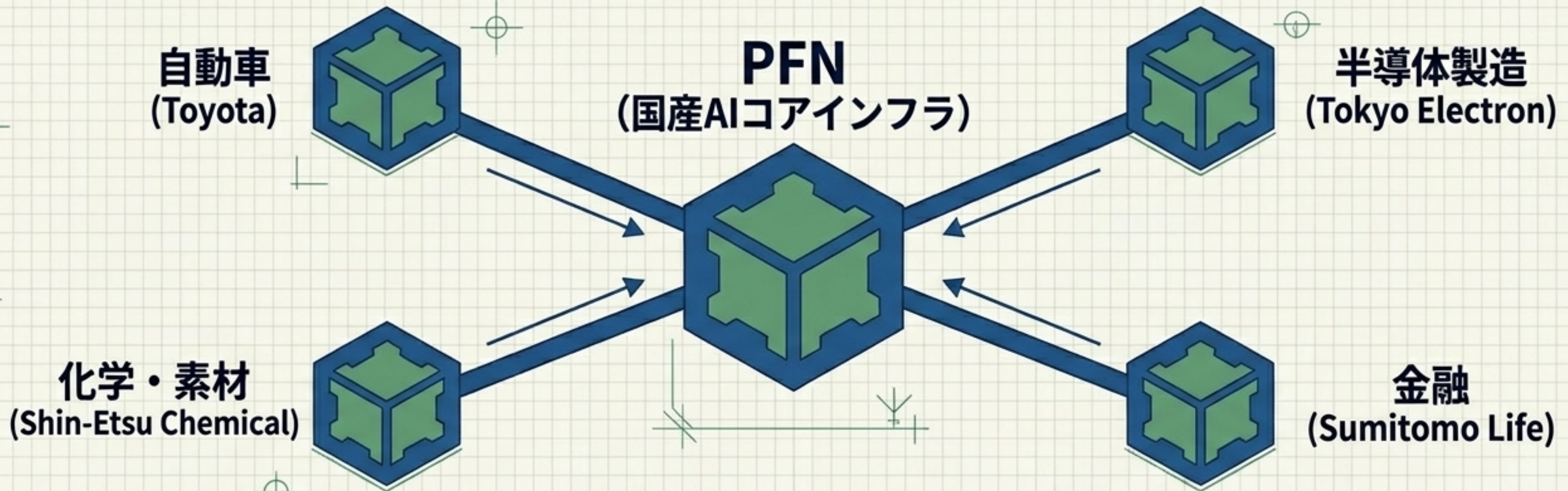


コードを変更することなく
そのまま入力

- データ移動の効率的なスケジューリング
- 抽象度に応じた階層的なコンパイル処理
- 最適化された低レベル機械語命令の自動生成

プロセッサの潜在能力を
極限まで引き出す完璧な密結合

日本産業界の総意：「コアAIインフラ」への戦略的合流



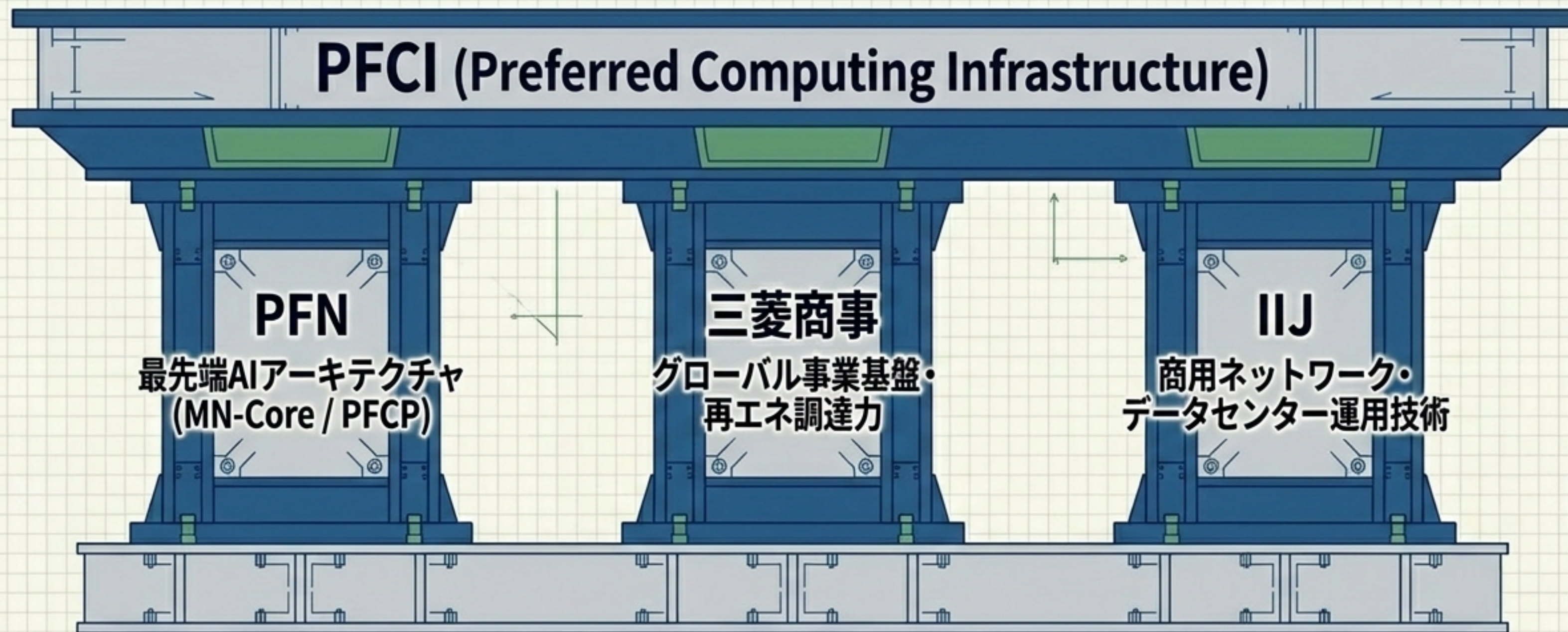
異例の大型調達

総額240億円の資金調達を完了。
スタートアップ単独のリソース限界を突破。

強固なコンセンサス

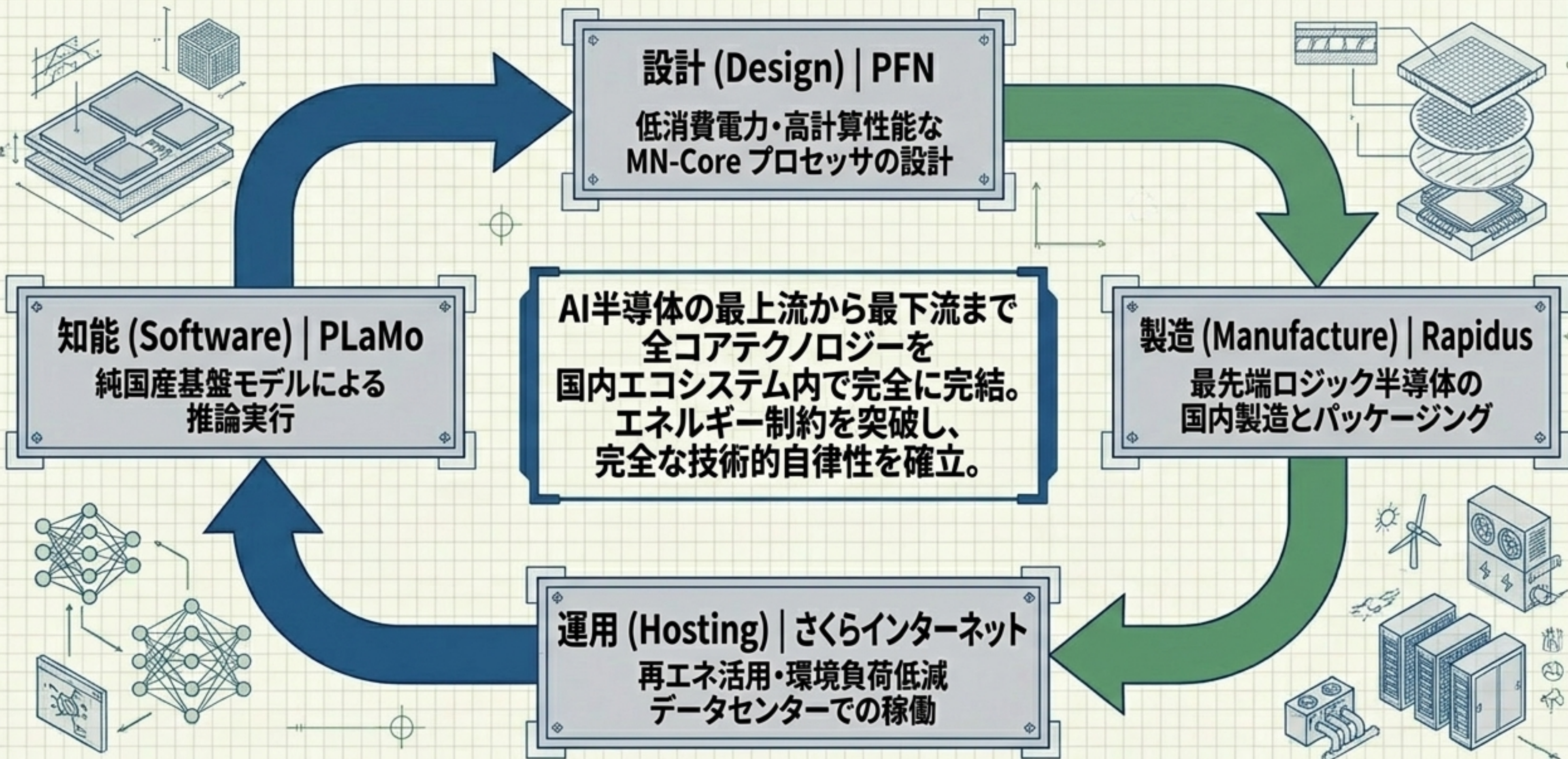
これは単なる投資ではなく、海外特定ビッグテックへの
過度な依存を回避し、日本独自の「国産AI基盤」を
自社DXの中核に据えるという産業界の明確な決意。

エンタープライズ提供のスケールアップ：合併会社「PFCI」の設立

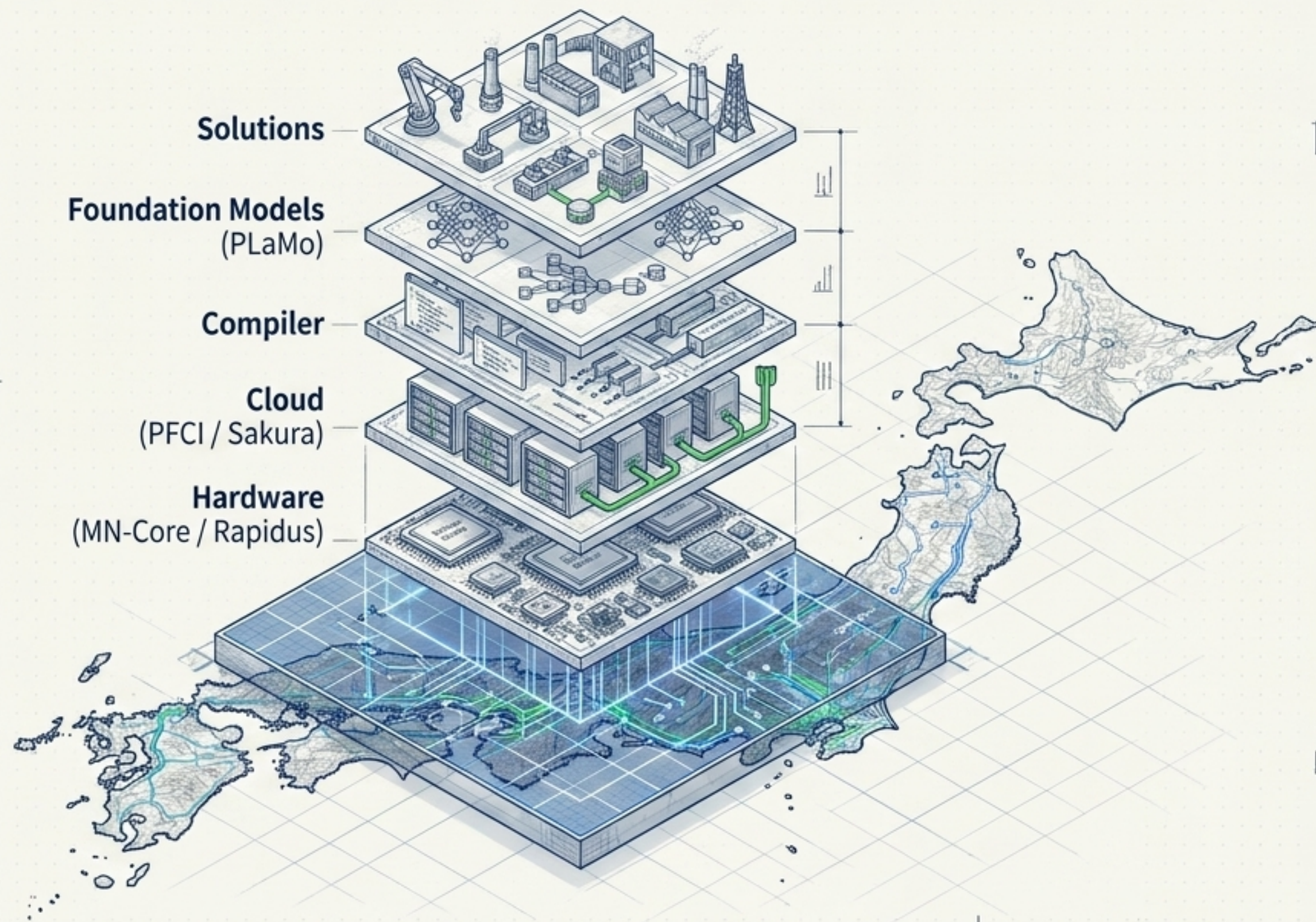


スタートアップ単独の物理的な設備投資限界を、国内トップ企業とのジョイントベンチャーで突破。極めて強固でスケラブルなAI計算インフラをエンタープライズ市場へ安定供給する体制の完成。

究極のクローズドループ:「グリーン・ソブリンAI」サプライチェーンの完成



現実世界を計算可能にする、究極の産業用AIエンジン



- PFNの戦場は「汎用チャットボット」の競争ではない。
- チップ（物理法則の制御）から、クラウド（インフラ）、LLM（高度な論理）に至る全階層を掌握。
- 日本初、世界最高峰のAIインフラストラクチャーとして、実社会のミッションクリティカルなワークフローを駆動し続ける。