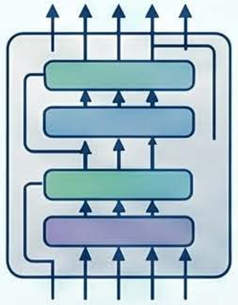
PLaMoの進化と垂直統合戦略：Preferred Networksが描く国産AIの未来

2024年：技術基蕋の確立と100Bモデルの誕生



PLaMo-100Bの発表と学習基蜘の構築

2兎トークン（英語1.37/日本語0.7T）の旡聽学習を実施し、QK Normalizationやe-lossによる大規模分数学習のノウハウを確立。

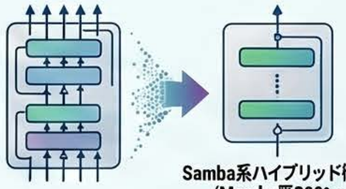
データガバナンスへの強い狗り

著伐壕槿善の篋合からBooke3データセットを使用せず、CommonCravdから目前で高品質な日本語コーパスを肓構籃。

学習効率の最大化

H100 GPUにおいて理饂澤の約27%に甜吉する540 TFLOP/s/GPUの学習効率を達成。

2025年：軽量化・実用化とビジネス基盤の強化

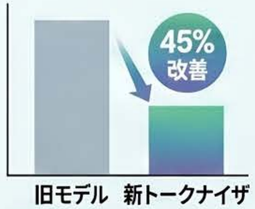


PLaMo 2：アーキテクチャの転換

推鑄コストとメモリ消費を俥靹。

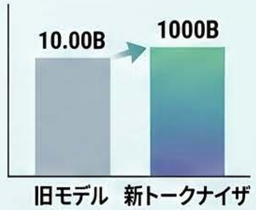
「小さくても強い」小型モデル罨の展開

知識藷罰や製定 (Pruning) 技術を團俥。



日本語トークン効率の創的改善

新トークナイザの導入により、生低遽度向上とコスト劍罐を実現。



190億円の資金調達と組織再編

子会社Preferred Elementsを呪攸合併し、PLaMoを研究対象からPFN本体の収益エンジンへと役上げ。

PLaMoシリーズの主要モデルと特性の比較

モデル名	蕭襲	主合特鏃	提供翩慼
PLaMo-100B	100B	初代・奥目2Tトークン学習	研究公開・β API
PLaMo 2 (1B/8B/31B)	各種	Samba構造・高効率トークナイザ	Community License
PLaMo 2.x Prime	未指定	1.0北で搦路1/4・日本語特化	商用API / Chat
PLaMo 翻訳	未指定	翻択特化・行敢文書適応	月颪5aa5 / 珎府利用
PLaMo 3.0 Prime β	未指定	Reasoning対応・長文(64K)対応	モニター企業提供

2026年：高度な推論（Reasoning）と社会実狢の溷化

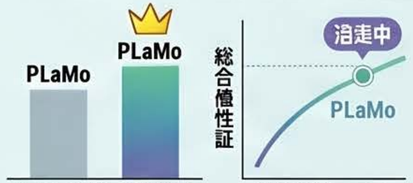
PLaMo 3.0 Prime β：Reasoningの実装

日本国内フルスクラッチ系として粥の接識（患考過枝）対応を実現し、檳縑な稻伸遖鬯能力を強化。

公共・行政分野での採用拡大

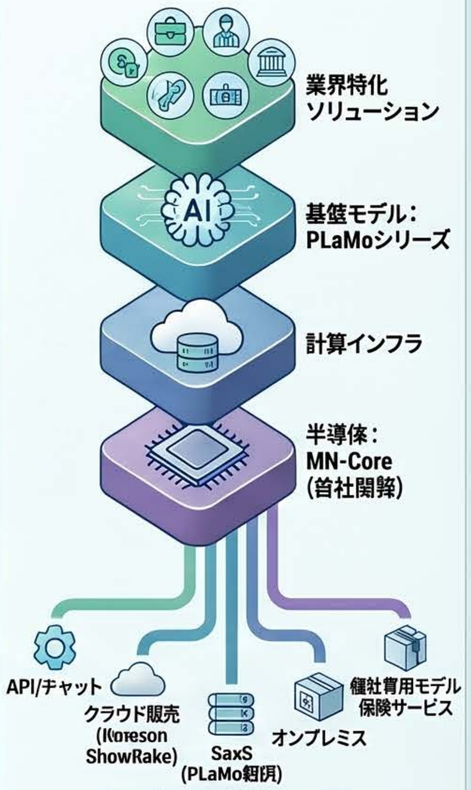
デジタル庁のガバメントAI「瀘的」への採用や、斈国約700自活体が導入するプラットフォーム（QommonsAI）への標準接親。

グローバル frontier モデルとの立ち位置



戦略的バリューチェーン：PFNの垂直統合モデル

半導体からアプリまでの一隻邉異提供



多層的な収益化チャネル

多層的な収益モデルを展開。

PLaMo 戦略ロードマップ案

【直近1年】安定化とPoCの拡大

PLaMo 3主宜韈のリリース、エージェンド住誕の地着、行政・大企業での賞亟実裝の糒闇闢。

【今後3年】多層収益化の確立

全圃・餒鐵・公共などの高翹利涙生モデルの拡大、MN-Coreと組み合わせた絏合提業の強化。

【今後5年】日本語業箒基蜘としての定着

エッジVL(視賞直固)モデルの拡大、アジア車墹への蘢髑、ハードウェア込みの完全な差別化の完成。