国産初、“長考”するAI：PLaMo 3.0 Prime が切り拓く推論の未来

「長考 (Reasoning)」を可能にする技術的特徴

思考過程への学習 (SFT/DPO/RL)

最終回答だけでなく、結論に至るまでの「思考のプロセス」にも学習を適用しています。

最大12,288トークンの推論予算

12,288

複雑な課題に対し、内部で膨大な試行錯誤や推敲を行うための十分な「思考スペース」を確保しています。

日本語に特化した推論の最適化

思考を日本語で行う再訓練を検討中で、トークン節約と表現の構度向上を目指しています。

ベンチマーク性能とビジネス活用

日本語対話・指示追従で世界水準

PLaMo 3.0 Prime β とグローバル主要モデルとの性能比較

日本語対話 (J-MTBench)	PLaMo 3.0 Prime β	9.01
	Qwen3-235B	8.99
	gpt-oss-120b	8.95
日本語指示追従 (JFBench)	PLaMo 3.0 Prime β	0.371
	Qwen3-235B	0.301
	gpt-oss-120b	0.323
数学推論 (AIME 2024)	PLaMo 3.0 Prime β: 0.579	Qwen3-235B: 0.857
	gpt-oss-120b: 0.782	

実務に強い3つの重点領域

- 医療・法令の専門QA
- 長文契約書の整合性チェック
- 複雑な条件の意思決定支援

商用化へのロードマップ

2026年3月19日：β版公開

無償利用を前提としたモニター企業の募集を開始し、実運用での検証をスタート。

2026年6月中旬：商用版提供（予定）

モニターからのフィードバックに基づき、安定性と処理速度を改善した正式版をリリース。