

2026年 国産LLM戦略マップ：巨大化か、効率化か

2026年、日本のAI開発は海外モデルの単純な追跡から脱却。データ主権、日本語の高度な処理能力、コスト効率といった国内特有のニーズに応え、独自の強みを打ち出す「差別化と実証」の時代へ。

国産LLM、2つの進化論：規模追求型 (MoE)

巨大な知識量を持ち、汎用性能を追求。
クラウドでの大規模サービスに向いています。(SoftBank / Rakuten)

2025年末時点の日本語性能ベンチマーク(MT-Bench)スコア

Rakuten AI 3.0 (700B)	8.88	🏆
GPT-4o (Global Baseline)	8.67	
ABEJA Qwen2.5 32B	8.04	

国産モデルがグローバルトップレベルに匹敵することを示します。

国産LLM、2つの進化論：効率特化型 (Adapter/Edge)

軽量モデルと追加モジュールで専門性を実現。
オンプレミスや特定業務に最適です。(NTT / NEC)

ソフトバンク「Sarashina」：圧倒的な
計算資源でAGI(汎用人工知能)を目指す
通信インフラとAIを融合させ、グローバル
水準の超巨大モデル開発に挑みます。

楽天「Rakuten AI 3.0」：国内最大級モ
デルをオープン化しエコシステムを構築
7000億パラメータモデルを公開し、
日本のAI技術全体の底上げに貢献します。

NTT「tsuzumi」：
軽量・省電力でオンプレミス市場を深耕
1GPUで動作。機密情報を扱う金融や自治体
でのセキュアな利用を推進します。

NEC「cotomi」：
高速推論と業種特化で現場実装力を発揮
GPT-4oの2.2倍以上の速度。製造・金融など
特定業務の自動化に注力します。

