

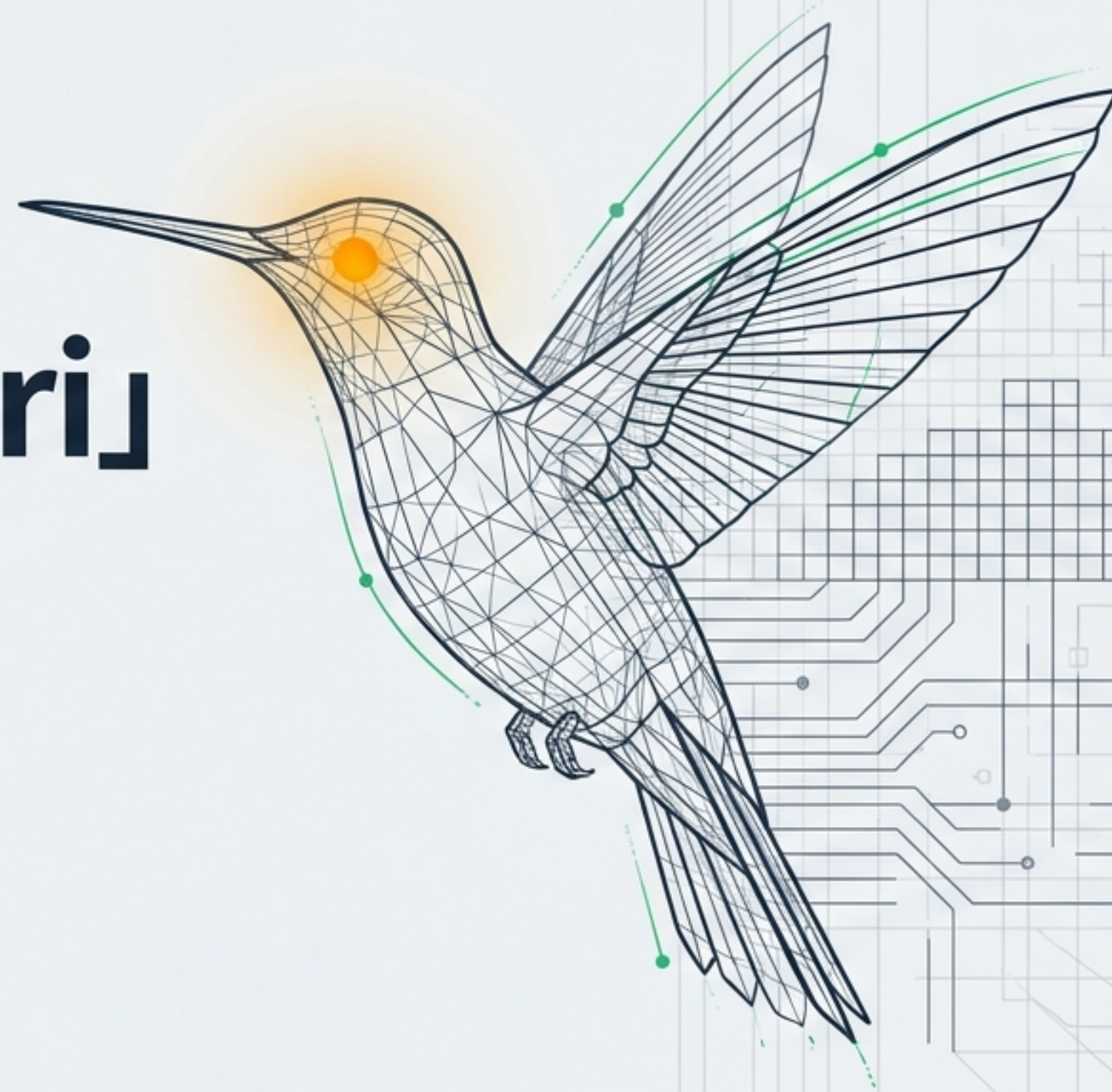
# Aleph Alpha「Kolibri」 戦略分析レポート

欧州発ソブリンAIの現実と  
日本市場へのインプリケーション

**Document Classification:** Strategic Intelligence Brief

**Date:** October 2026

**Target:** Policymakers & Enterprise Strategists



# エグゼクティブ・サマリー：なぜ日本が「Kolibri」を注視すべきか



## Model Specs - 高効率アーキテクチャ

総パラメータ78.1B / アクティブ3.46BのMoEモデル。特定言語（独・英）に特化し、GPT-5を凌ぐトークン圧縮効率を実現。



## Corporate Reality - 資本とエコシステムの現実

カナダCohereとの事業統合（2026年9月）。「欧州単独のソブリンAI」から「大西洋横断連合」への現実的なピッチシフト。



## Regulatory Standard - コンプライアンスの製品化

EU AI ActのGPAI行動規範に準拠。海賊版データ排除と著作権クリアランスを「製品の強み（モート）」として再定義。



## Impact on Japan - 日本市場への間接的波及

Cohereを通じた富士通「Takane」への技術流入と、デジタル庁「ガバメントAI（源内）」調達基準におけるEU基準の事実上の標準化。

# ソブリンAIの「理想」と「現実」



**2026年10月3日（ドイツ統一の日）**

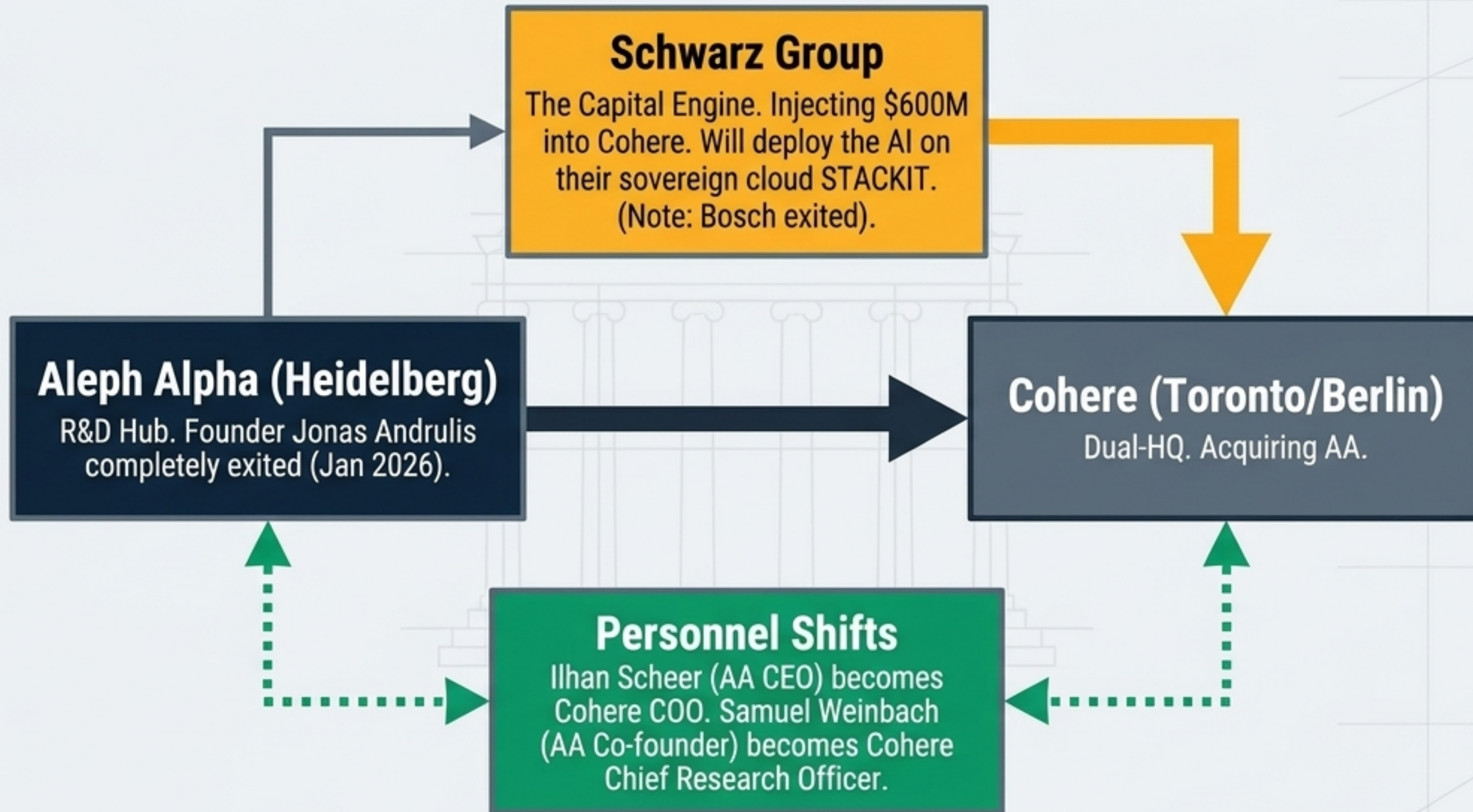
象徴的なリリース。ドイツのチーム、欧州内のインフラ（ドイツ・フィンランド）のみを使用し、外国の支配を受けない「技術的主権」の誇示。



**2026年9月16日（最終統合契約締結）**

資本の限界。単独での基盤モデル開発競争から撤退し、カナダCohereとの事業統合（評価額約200億ドル規模）による生き残りを選択。

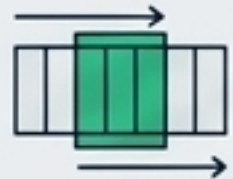
# 資本とエコシステムのアーキテクチャ



# Kolibri 1 アーキテクチャ：広さより「深さ」の戦略



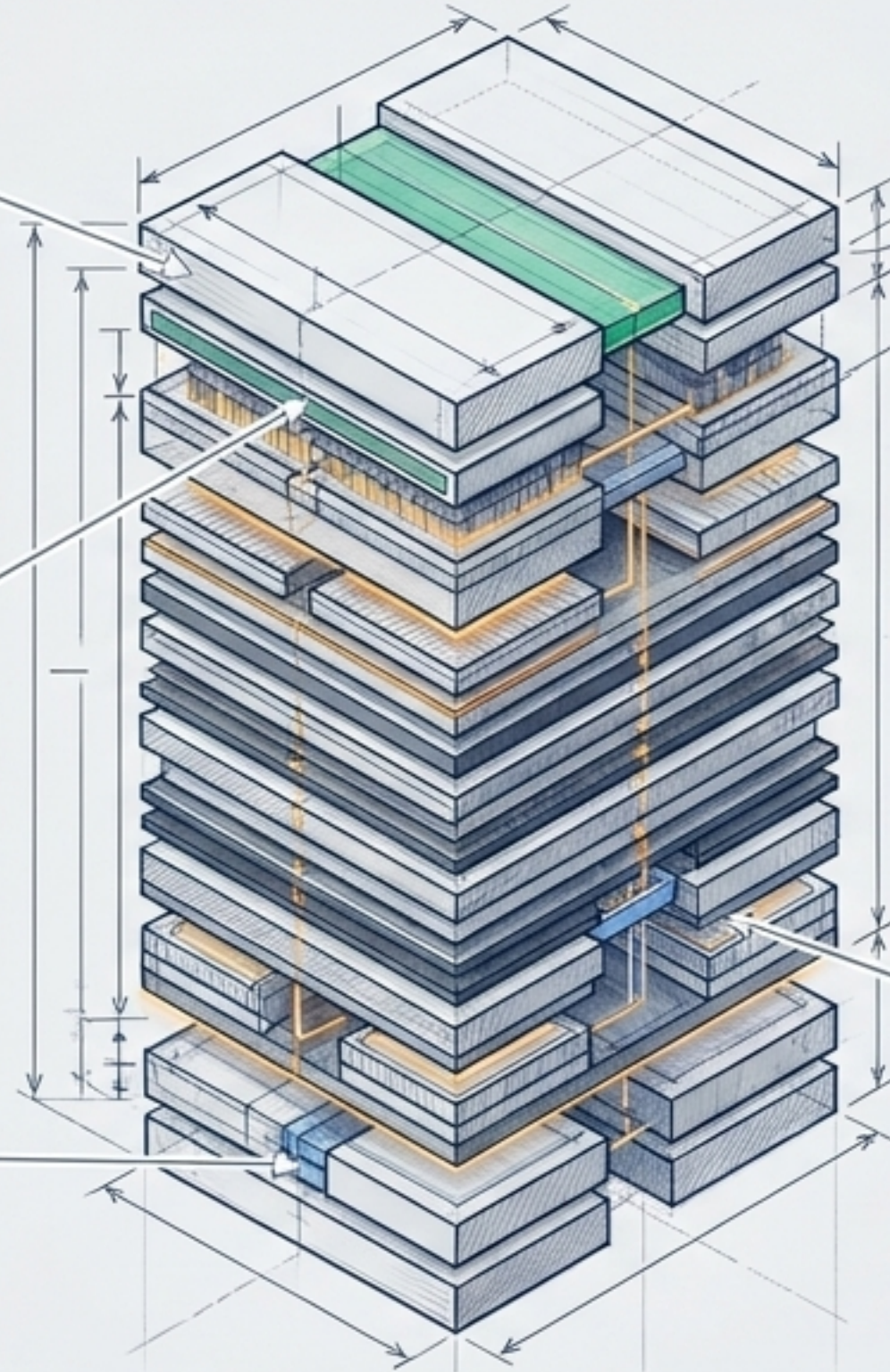
**Scale:** 総78.1B / アクティブ3.46B (4.4%)。50層MoE。推論時の軽量化に特化。



**Context Window:** ネイティブ262kトークン（最大1Mまで検証済）。



**Training:** NVIDIA B200 768基で21日間（推計950MWh）。フロンティア級の物量戦を回避した高効率学習。



## 独自のUniBPEトークナイザー

Kolibri

4.90

GPT-5

4.35

bytes/token (ドイツ語)



ドイツ語の圧縮率「4.90 bytes/token」。

GPT-5 (4.35) と比較して約11% 少ないトークン数で処理可能。日本語等の多言語を切り捨て、独・英（とコード）にリソースを集中させた結果の優位性。

# パフォーマンスと「回答を控える技術」

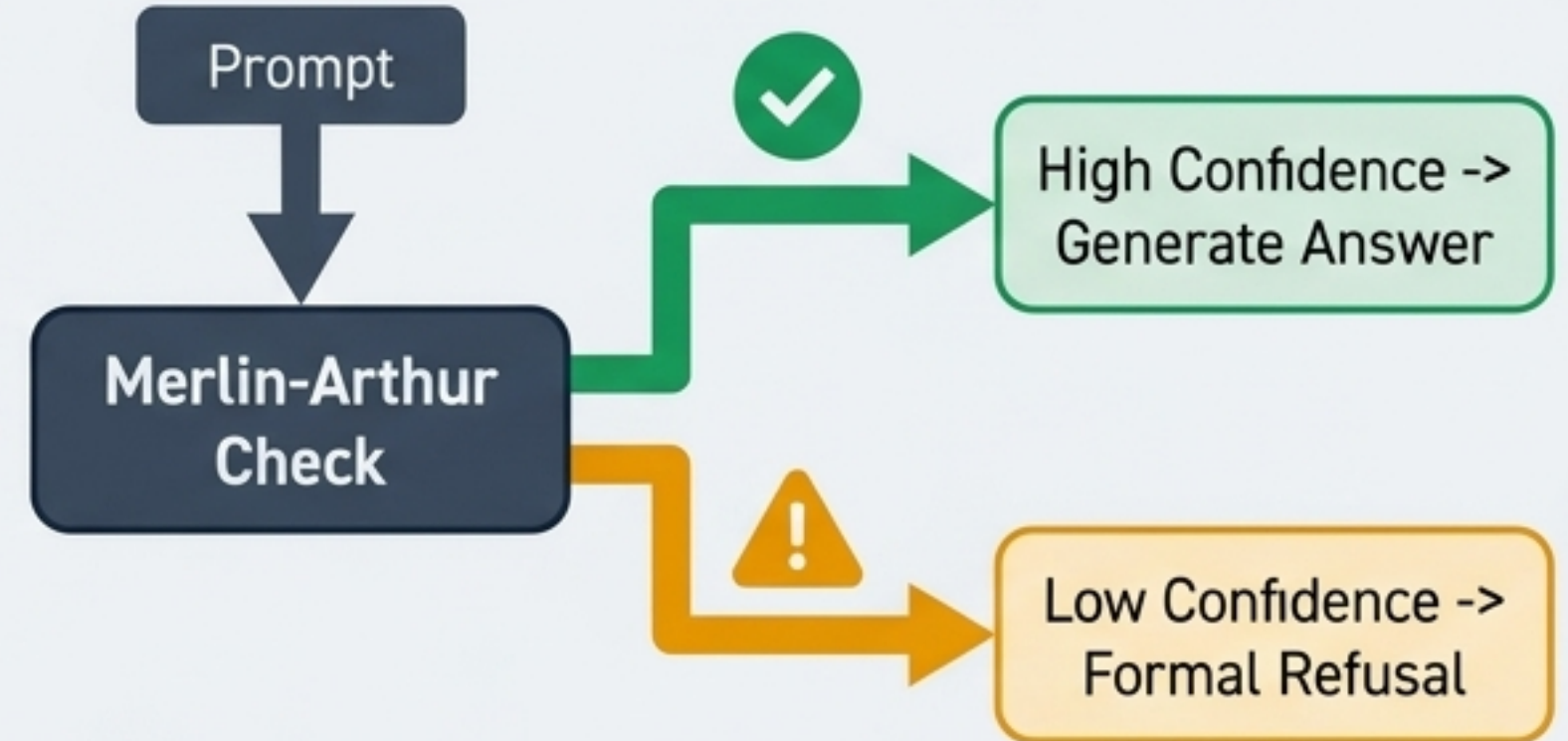
## Niche Dominance



**Strengths:** コーディング (LiveCodeBench 85.9) と金融 ( $\tau^3$ -bench banking 38.1) で他を圧倒。

**Weaknesses:** 一般知識系のベンチマークでは Qwen3.8等に劣る。(※全て自社測定値に基づく)

## The "Merlin-Arthur" Protocol



44%

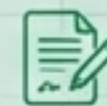
AA-Omniscienceベンチマークにて、誤答せずに「回答を控えた」割合が44%に到達（旧モデルの15%から大幅改善）。ミッションクリティカル用途におけるハルシネーション抑制の最適解。

# 規制対応の製品化：EU AI Actレディネス

## Compliance Stack

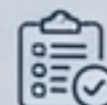


### GPAI行動規範への署名



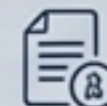
欧州委員会のGPAI（General Purpose AI）行動規範に署名。モデルの自律的評価と透明性を担保。

### 学習データのホワイトボックス化



EUテンプレートに沿った「学習データ公開要約」の発表。権利者向けの直接連絡窓口（オプトアウト権の尊重）をモデルカードに明記。

### 徹底した著作権クリアランス



欧州委員会のPiracy Watch Listに基づく450万件超のURLブロックリストを適用。第三者データセットのライセンスを個別審査。

# 日本市場への波及ルート：技術と政策のインダレクト・インパクト

## Direct Use

Target: 欧州に拠点を持つ日系企業  
(製造・自動車部品・半導体)。

Impact: EUの厳格なコンプライアンス環境下で稼働する、安全なオープンプリミス型Apache 2.0モデルとしての実利用（日本語非対応でも独・英文書処理で価値）。

## Technology Flow

Target: 国産LLMエコシステム。

Impact: AAの技術と人材  
(Weinbach氏等) がCohereへ統合。  
-> Cohereを基盤とする富士通  
「Takane」の将来モデルへ、欧州の  
トークナイザー技術やハルシネーシ  
ョン抑制技術が流入する可能性。

## Policy Flow

Target: デジタル庁 / 政府調達。

Impact: Kolibriの「学習データ要約」「海賊版排除プロセス」が、次期ガバメントAI調達の事実上のグローバルスタンダードとなる。

# ソブリンAIの共通課題：欧州と日本のパラレル・リアリティ

## Europe (Kolibri)

ドイツ語データの不足を、自前クロールと「言い換え (rephrasing)」で補填（翻訳データへの依存を回避）。

UniBPEによるドイツ語の極限圧縮（4.90 bytes/token）で推論コストを削減。

計算資源（B200×768）を確保するも、継続的な単独生存を断念し北米資本と統合。

ローカルデータの枯渇

推論コストとトークナイザー

資本の壁  
(サステナビリティ)

## Japan (GENIAC/LLM-jp)

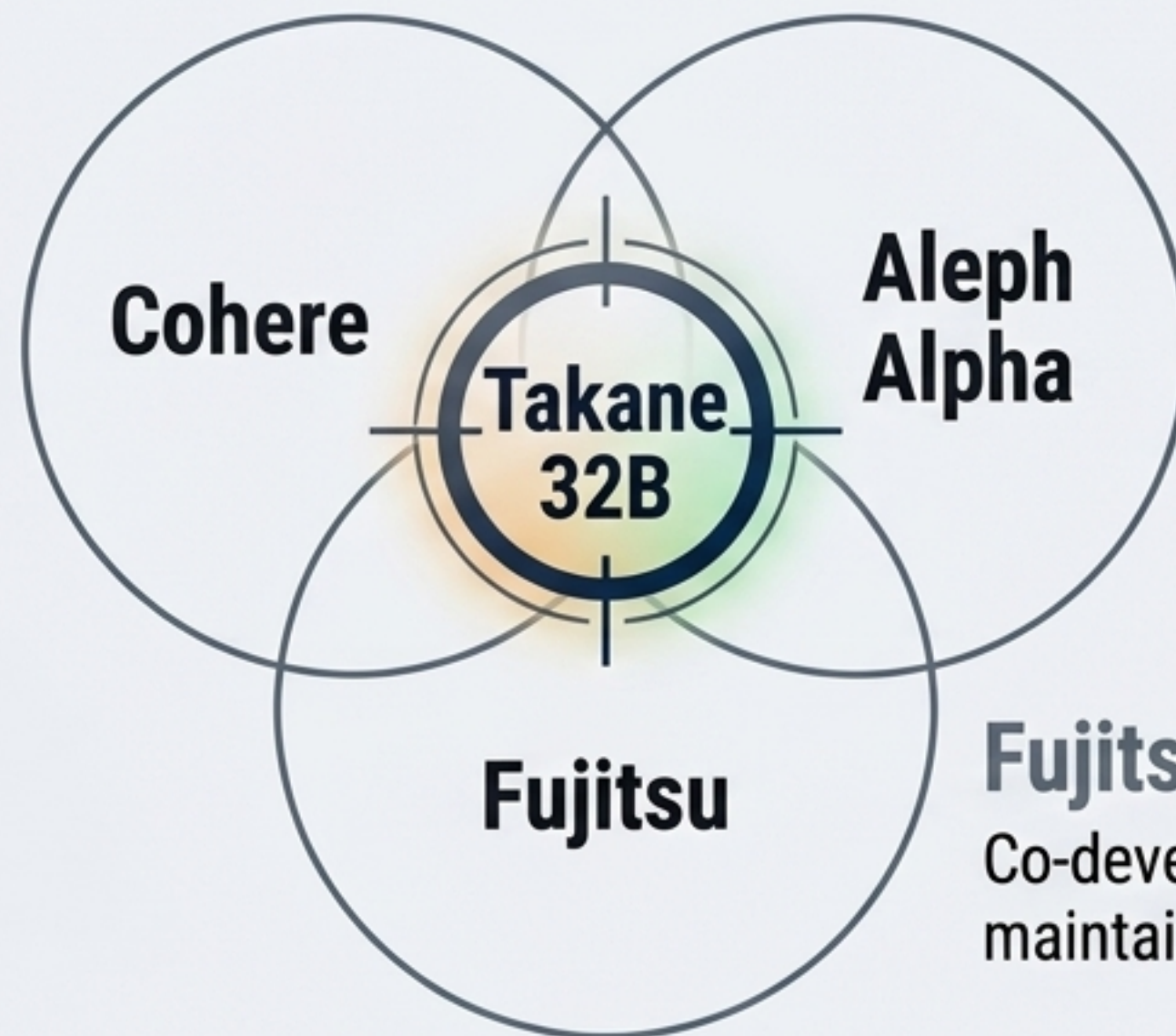
高度な日本語学習データの枯渇と、合成データ/翻訳データ活用への依存。

日本語特化ボキャブラリによるトークン効率化の追求。

国の計算資源支援（GENIAC）終了後の、国内スタートアップの自立的資金調達とハイパースケーラーへの対抗戦略という共通のジレンマ。

# 「国産LLM」の再定義：Cohere・富士通・Aleph Alphaの交差点

**Cohere**  
The Global Platform  
(Dual HQ Berlin/Toronto).



**Aleph Alpha**  
Injecting European Regulatory  
IP & Research Leadership  
(Samuel Weinbach as CRO).

**Fujitsu**  
Co-developing Takane 32B with Cohere,  
maintaining exclusive global rights.

## Key Insight Box

デジタル庁「源内」プロジェクトにおいて、Takane 32Bは「国産LLM」枠として契約済み（2026年6月）。Cohereが「独・加連合のソブリンAI企業」へと変質する中、知財の源泉やデータガバナンスの観点から、「国産」の定義と調達要件の境界線が極めて曖昧になりつつある。

# デジタル庁「源内」2027年調達に向けたベンチマーク

## Digital Agency Requirements (源内公募要件)

- 収集方法の説明
- 海賊版等が含まれていないことの担保方法
- 著作権法・個人情報保護法等の遵守

MATCH

MATCH

MATCH

## Kolibri's Gold Standard Implementation

- EUテンプレートに基づく詳細な「学習データ公開要約」の提出。
- Piracy Watch List等の450万件ブロックリストによる機械的な海賊版排除プロセス。
- データセットごとの個別ライセンス審査と、オプトアウト受付窓口の常設。

**Takeaway:** 日本の著作権法（30条の4）は学習に寛容だが、実際の政府調達プロセスにおいては、Kolibriが体現する「**EU型の厳格な説明責任**」が**デファクトスタンダード**として求められ始めている。

# アクションプランと今後のウォッチリスト

## 政策担当者 (デジタル庁・経産省) へ

Kolibriの「技術報告書」と「学習データ要約」を、2027年度の「源内」有償調達に向けた国産LLMの開示テンプレート作成の直接的なリファレンスとして活用せよ。

## エンタープライズ (製造・金融) へ

日本国内向け業務でのKolibri採用検討は不要。ただし、EU拠点における技術・法務文書処理のPoC候補（安全なオンプレミス環境）としては最有力。自社データでの評価を推奨。

## 富士通「Takane」 ユーザーへ

CohereとAleph Alphaの統合クロージング後、ロードマップ、データの所在（データレジデンシ）、モデル所有構造の変更が、自社の国内調達要件やコンプライアンスに影響しないか継続確認を。

### 監視すべき先行指標

ベルリンSCCON（公共部門見本市）での導入実績 / Kolibri対Commandシリーズの統合ロードマップ / 第三者によるベンチマーク検証結果。