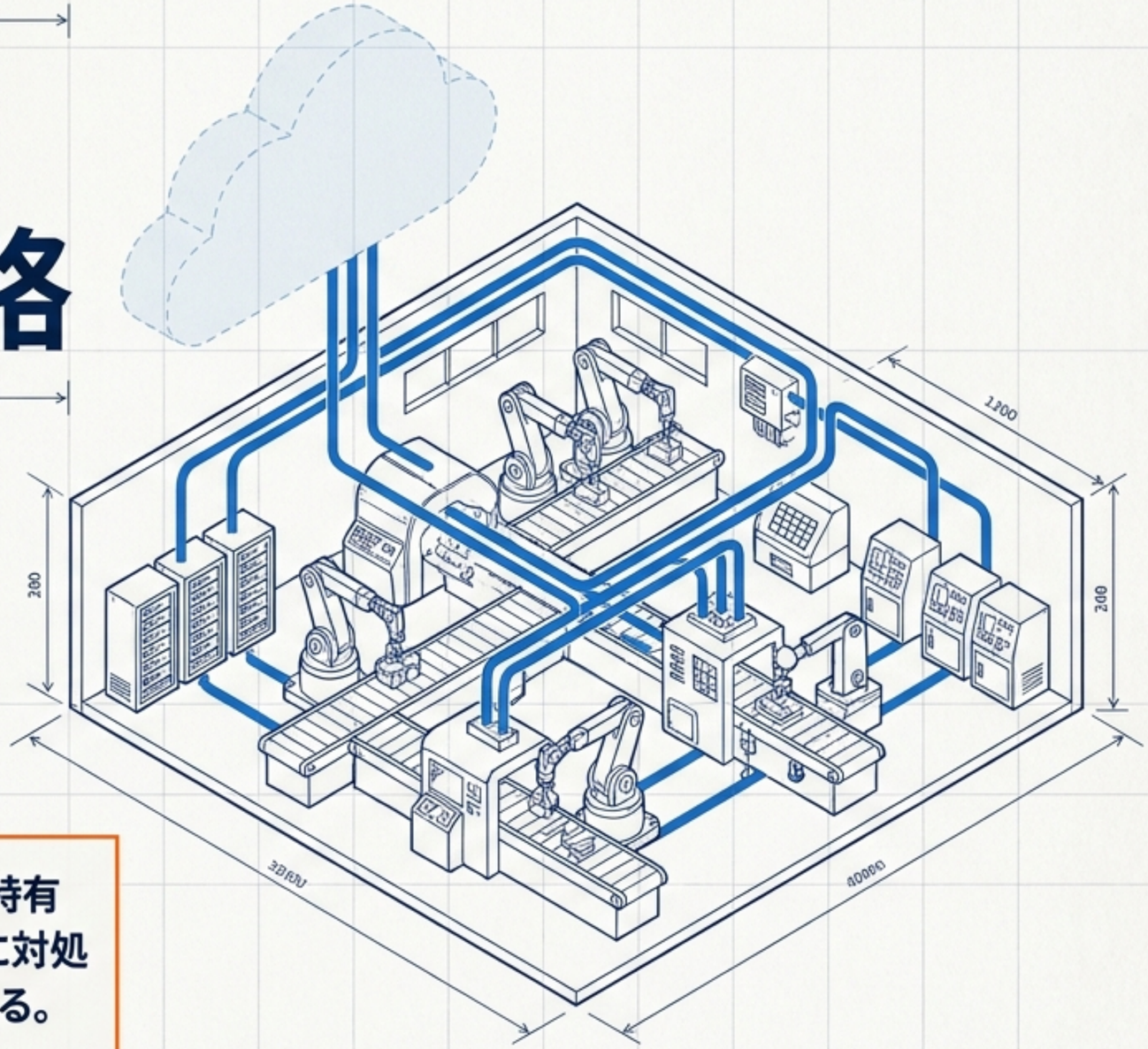
製造業における「ローカルAI」戦略

クラウドの限界を超える、
ハイブリッドアーキテクチャと
知財ガバナンス



ローカルAIはクラウドの単なる代替ではない。製造業特有の機密データ保護と、予測不能なトークンコスト膨張に対処するための必須の「ハイブリッド・コンポーネント」である。

Gemini Notebook

クラウド完結型アプローチが製造現場で破綻する理由

製造現場の重力



1. 機密性の壁: 門外不出の設計図面、生産実績、稼働ログ。
2. オフラインの壁: ネットワーク切断が許されないライン連動と即時応答（低レイテンシ）の要求。
3. コストの壁: 常時稼働によるトークン課金の青天井。

クラウドAIの引力



結論：現場の制約（重力）はクラウドの進化では解決できない。
AI処理を物理空間に引き戻す「ローカルAI」への回帰が必然となる。

Gemini Notebook

「ローカルAI」を構成する3つの物理レイヤー



Tier 1: オンプレミスLLM (全社・データセンター)

自社サーバー室のGPUサーバー。
用途：全社の文書検索(RAG)、横断的AIエージェント。

Tier 2: デスクサイド・ワークステーション (部門・個人)

DGX Spark等、机上に置ける強力なAI端末。
用途：機密業務のPoC、映像解析、小規模な追加学習。

Tier 3: エッジSLM/VLM (装置・ゲートウェイ)

産業用PC(IPC)や設備組込み。
用途：リアルタイム外観検査、エラーの即時言語化（設備保全）。

【注意】クラウド上の「専有テナント（プライベートクラウド）」はデータ隔離されていても、推論環境が社外にあるため本稿の定義する「ローカルAI」には含まれない。

Gemini Notebook

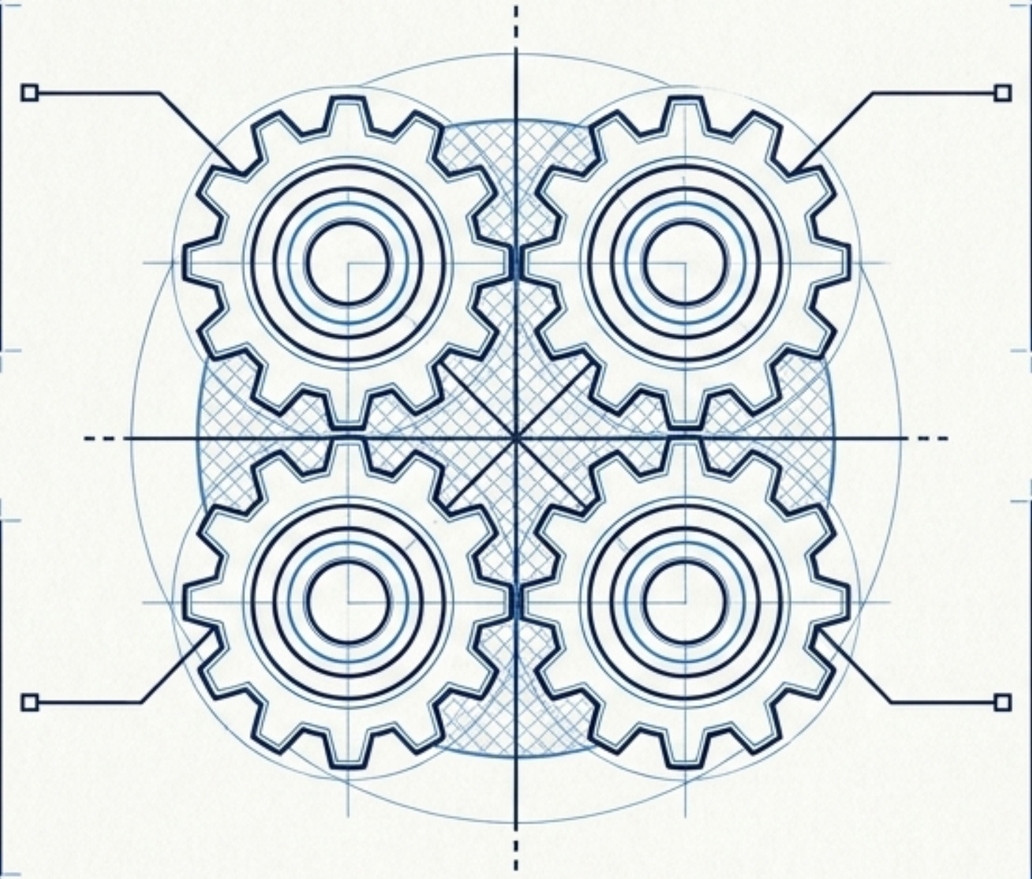
ローカルAI実装を加速させる4つの駆動要因

1. 機密性 (Confidentiality)

設計図、加工条件、未公開発明情報など、製造業のコアバリューは社外送信が不可能。

2. コスト (Cost)

LLMの常時利用・エージェント化に伴うトークン消費量の指数関数的な膨張（予算の予測不能性）。



3. 技術進化 (Technology)

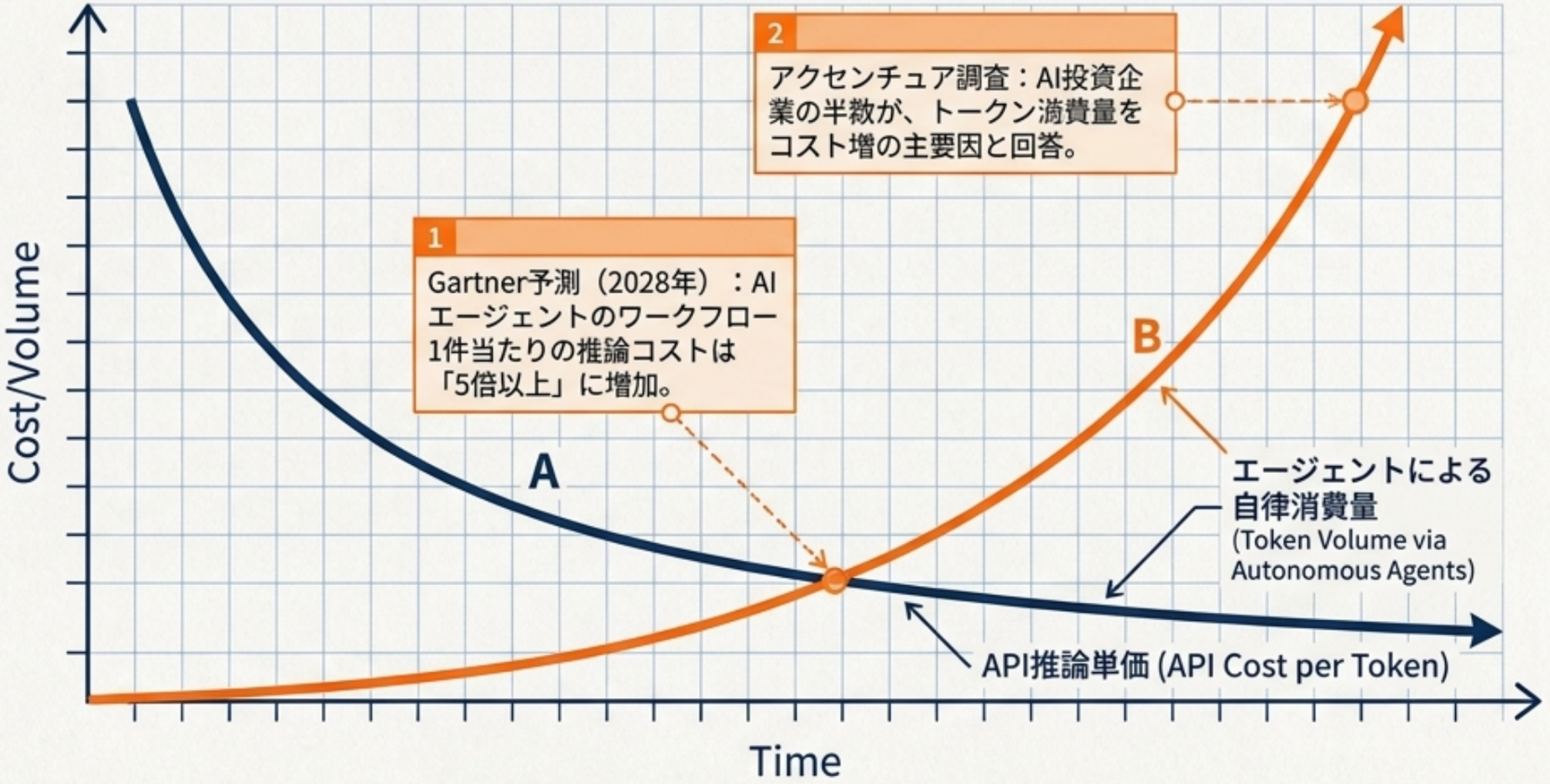
モデルの小型化（SLM）と、強力な小型ハードウェアの成熟 (128GBメモリ搭載のDGX Spark等）。

4. 経済安全保障 (Economic Security)

データ・モデル・基盤を国内・自社内で完結させる「ソブリンAI」への志向（NTT、富士通等の戦略）。

Gemini Notebook

トークン経済のパラドックス：AIエージェント化によるコスト爆発



モデル自体の単価は下落しているが、自律的にループ処理を行う「エージェント型」の普及により総消費量は爆発する。
製造業特有の膨大な設備ログや常時映像解析を扱う場合、「固定費型」のローカル推論ハードウェアが強力なコスト防波堤となる。

Gemini Notebook

急速に整備されたローカルAIの供給エコシステム

ハードウェア基盤 (Hardware)

- **NVIDIA / 菱洋エレクトロ**
  - DGXシリーズ
  - 工場運営AI「FOX」ブループリント展開

エンタープライズ・オンプレミス

- **NTT**
  - tsuzumi 2（300億パラメータ、利用の約6割がオンプレミス）
- **富士通**
  - Kozuchi Enterprise AI Factory等（1,000万円台から）
- **インテック**
  - ローカルLLM導入支援（約500万円から）

デスクサイド・部門導入

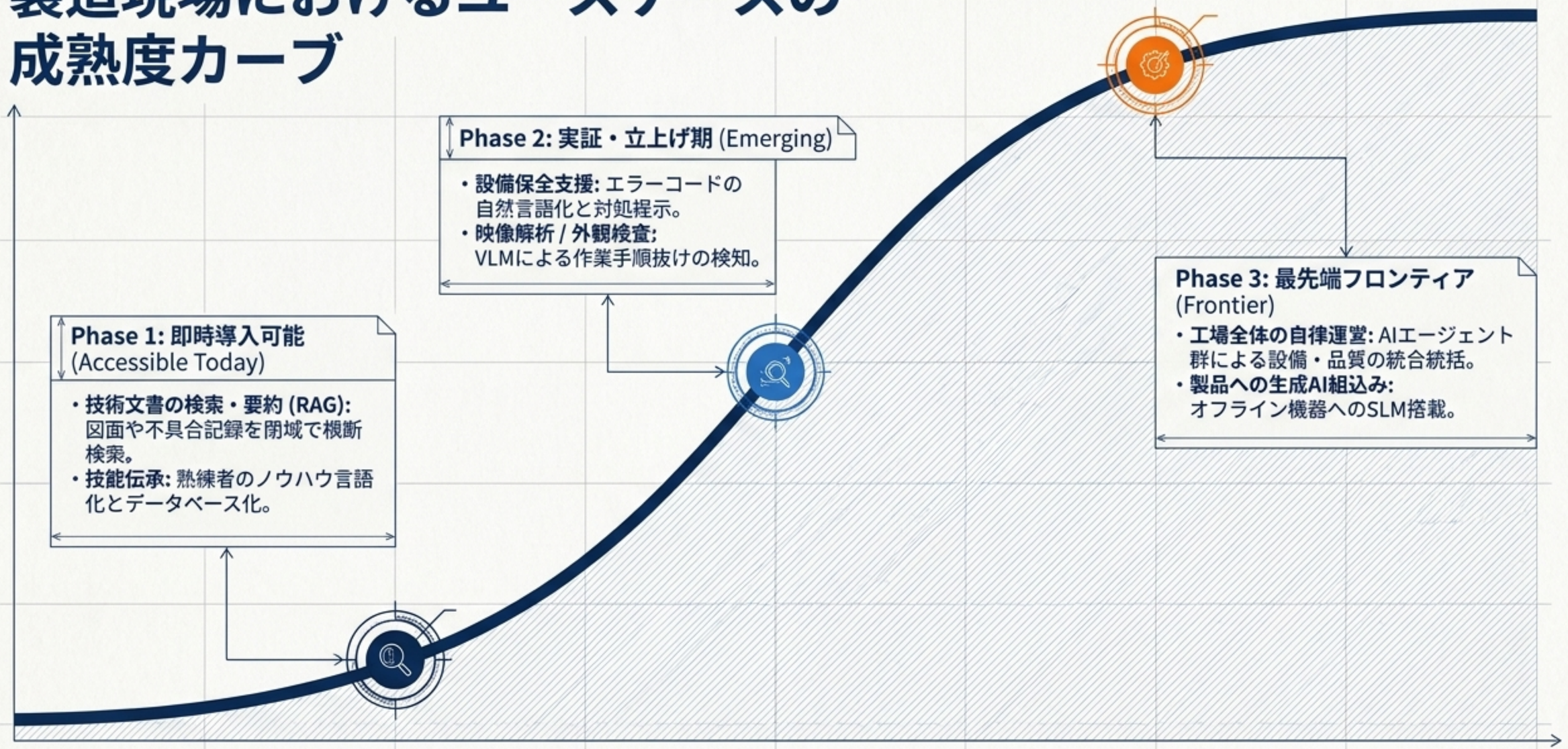
- **リコー**
  - オンプレLLMスターターキット
  - 従量課金なし

エッジ・FA組込み

- **三菱電機**
  - FA特化の18億パラメータSLM開発
- **東京エレクトロンデバイス**
  - エッジSLM評価支援
- **PFN**
  - 産業向け視覚言語モデル PLaMo-VL

Gemini Notebook

製造現場におけるユースケースの成熟度カーブ

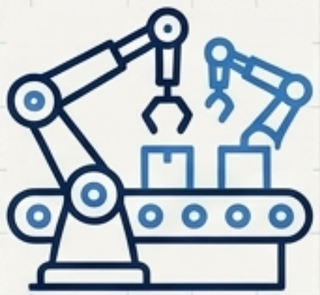


Gemini Notebook

グローバル最前線と国内製造業の実態（現在地）

Global Leaders: 工場の自律化 (Factory Autonomy)

- 【Siemens（独）】
  アンベルク工場などで「Industrial Copilot」を産業用PC(IPC)に展開し、はんだ付け装置のエラー対応をオフライン・ローカルで実現。
- 【Foxconn（台湾）】
  700億パラメータの自社LLM「FoxBrain」を開発。複数AIが連携する製造オペレーションシステム「MoMClaw」を構築中。



国内の実態: 研究開発とバックオフィス業務

- 【国内大手（Panasonic HD, 三菱電機）】
  1,000億パラメータ級モデルの独自開発やFA特化SLMの実証・構想段階。
- 【現場の実像】
  国内事例の多くは依然としてバックオフィスの「社内文書RAG」に集中しており、生産ライン自律化の公表事例は乏しい。



Gemini Notebook

ローカルAI実装を阻む4つの「現実的課題」



1. 性能のギャップ (Performance Gap)

ローカルの小型モデルは、最上位のクラウドLLMと同等の汎用推論精度は出ない（用途の絞り込みが必須）。



2. 運用負荷 (Operational Burden)

サーバー、GPU、ネットワーク、モデル更新、精度監視を自社で運用・保守する重いITインフラ負担。



3. データ整備の壁 (Data Readiness)

国内製造業で製造プロセスデータを「活用できている」 企業はわずか51.4%。AI以前のデータサイロ問題。



4. セキュリティの錯覚 (Security Paradox)

「ローカル＝安全」ではない。アクセス権限設計の不備や、モデルファイル（重み）自体の物理的持ち出しリスクが存在する。

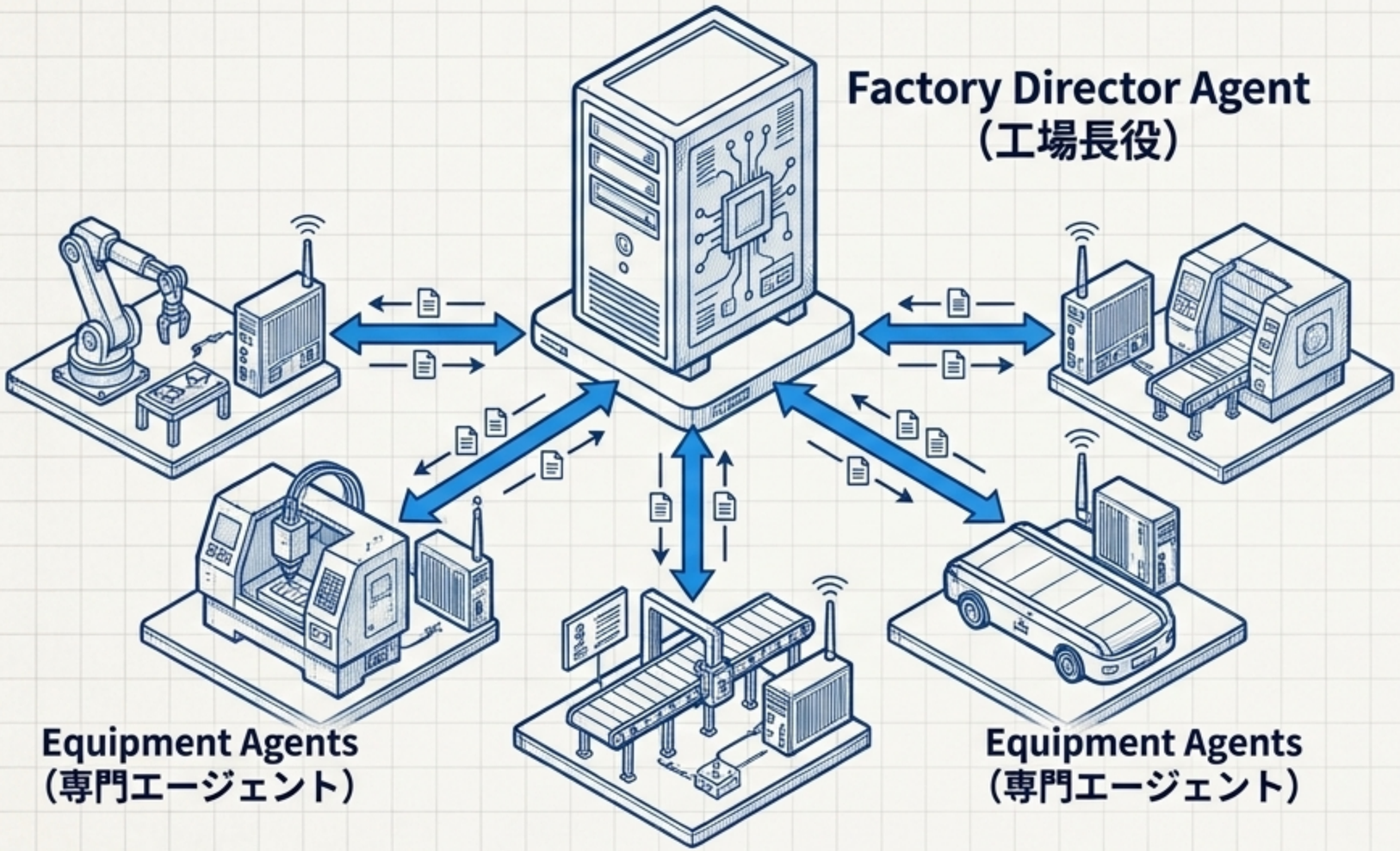
Gemini Notebook

【SYNTHESIS】クラウド vs. ローカル ワークロード振り分けマトリクス

	Local（ローカル）	Cloud（クラウド）
データの機密度	未公開発明、詳細図面、加工条件、原価情報。	公開情報の調査、一般的な文章ドラフト作成。
処理量と定型性	大量・常時稼働（ログの要約・分類・映像抽出）。	少量・非定型・複雑な推論が必要なタスク。
応答速度と接続性	閉域・オフライン環境、ライン連動のミリ秒単位の即時応答。	常時接続が前提で、数秒の遅延が許容される業務。
運用体制	AIインフラ運用人材、または専任の保守委託先を確保可能。	運用インフラを持ちたくない、リーンなIT体制。

Gemini Notebook

次世代のフロンティア：工場を自律化する「マルチエージェント」



質問応答から「自律制御」へ

ローカルAIは単なるチャットボットから、機械を直接制御する頭脳へと進化している。

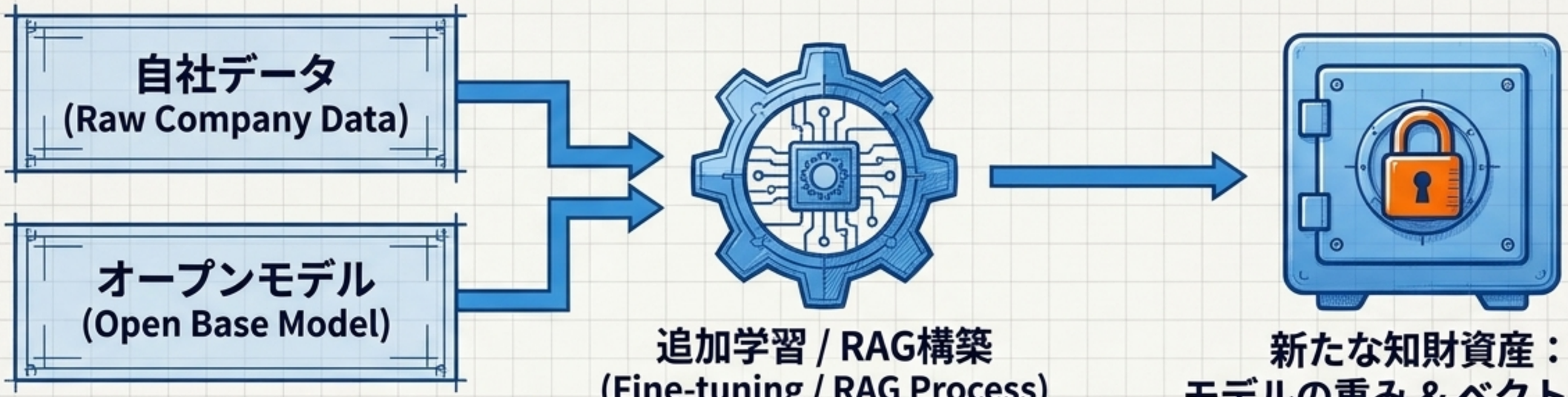
NVIDIA FOX ブループリント

「工場長役」のメインエージェントが、設備・品質・アラートのデータを集約。各ラインの「専門エージェント」に対して指示を出し、工場全体のエネルギー最適化や自律運営を行う構想（台湾EMS各社が先行導入）。

Insight: このレベルの自律オーケストレーションは、クラウドのレイテンシとセキュリティ 制約では実現できず、デスクサイド～エッジの強力なローカル推論基盤が不可欠となる。

Gemini Notebook

知財ガバナンス：ローカルAIが生み出す「新たな営業秘密」



経産省の指針改訂

「営業秘密管理指針（令和7年3月改訂見込）」への対応。ローカルAI環境で作られたデータは、従来存在しなかった新しい形の無形資産となる。

新たな管理対象の発生

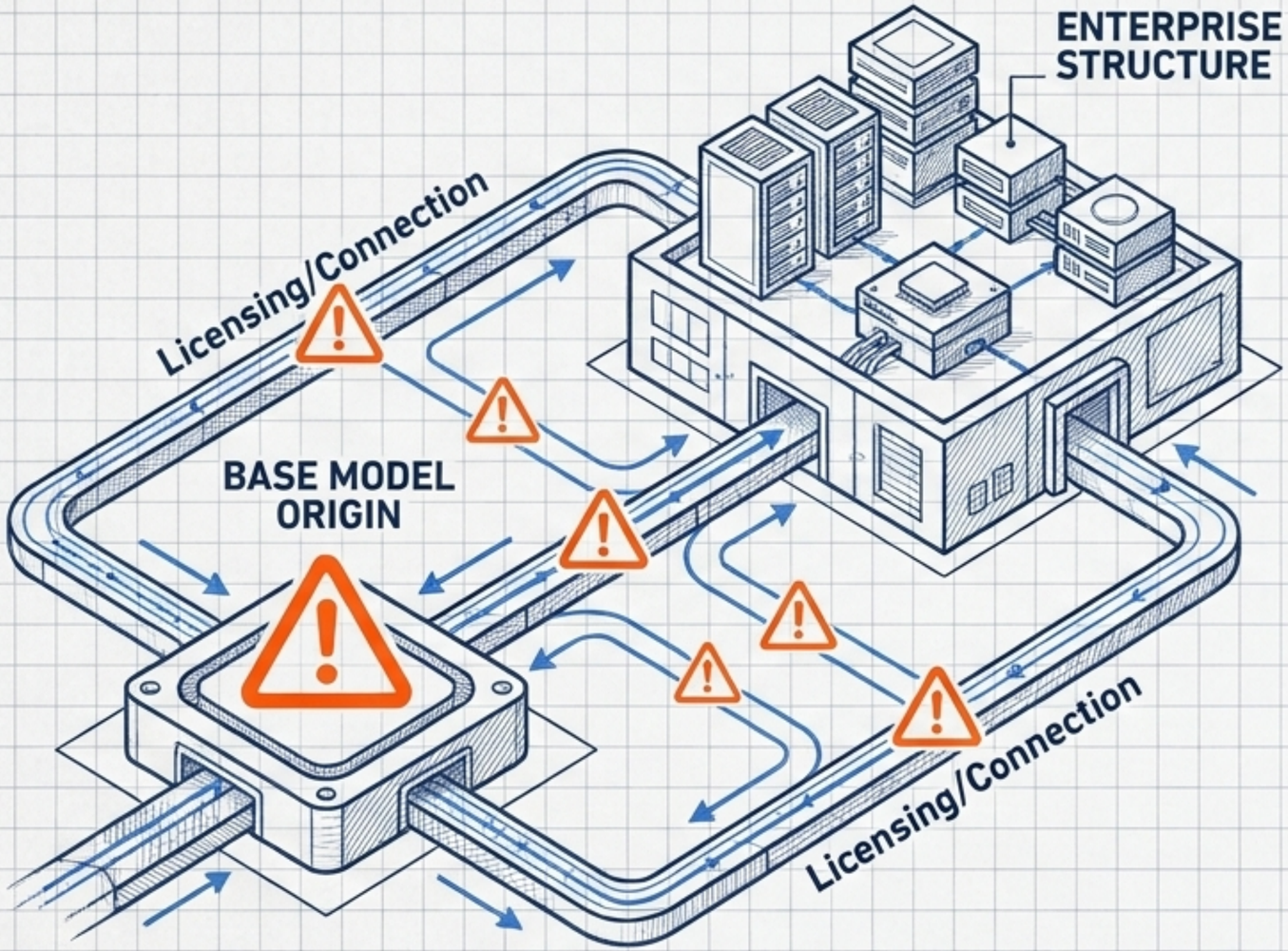
自社のノウハウで追加学習（ファインチューニング）した「モデルの重みデータ」や、RAG検索用の「ベクトルデータベース」は、企業秘密の結晶である。

現代の「図面持ち出し」

これらのファイル（数GB〜数十GB）を持ち出されることは、工場全体の設計図を盗まれることと同義。厳格なアクセス権限とログ管理が不可欠。

Gemini Notebook

オープンウェイト・モデルのジレンマと出自リスク



1. 偏在する「出自」リスク

高性能なオープンモデルは中国製が台頭。モデルルーター上のエンタープライズ利用において、中国発モデルの割合が63%に達したとの報告が存在。将来の提供継続性や安全保障上の評価が必要。

2. 複雑化するライセンス

商用利用の可否、派生モデルの取り扱い規定はモデルごとに異なる。ローカルで追加学習したモデルを自社製品（機器）に組み込んで外部販売する場合、ベースモデルのライセンス抵触リスクが跳ね上がる。

結論

「社内からデータを出さない」だけではコンプライアンスは担保されない。無形資産ガバナンスの確立が必須。

Gemini Notebook

ローカルAI実装への戦略ロードマップ（2025〜2027+）

短期 (〜2027年)

デスクサイドとRAGの普及

- Windows版DGX Station等の登場により、CAD/CAE利用者の机上で高度な推論が稼働。



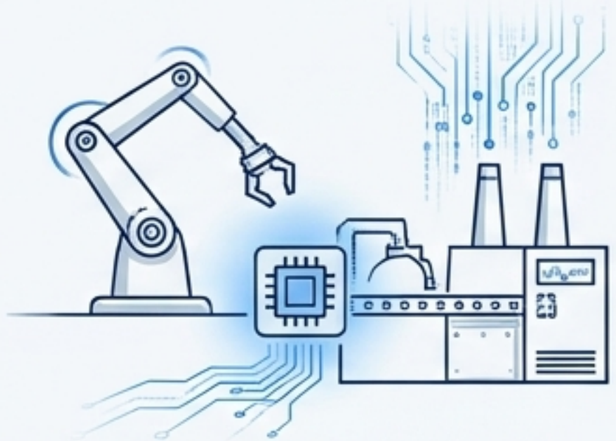
- 数百万〜1,000万円級のパッケージによる、部門単位の「文書検索・要約(RAG)」が一般化。



中期 (2027+)

エッジSLMの実運用への移行

- 国内FA機器メーカーの装置組込みSLMや、工場エージェント（FOX型）がPoC（概念実証）から本番稼働へシフト。



不確実性 (Wildcards) ⚠

想定されるゲームチェンジャー

⚠ ・クラウドAIの熾烈な価格競争によるコスト優位性の変動。



⚠ ・オープンウェイトモデルの商用ライセンス条項の急な変更リスク。



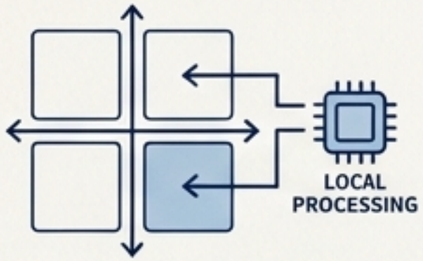
Gemini Notebook

アーキテクトのための次の一手（Actionable Checklist）

1

アーキテクチャ (Architecture): データの棚卸しと分離

全社一律のクラウド化を停止する。4象限マトリクスを用い、「外部に出せないデータ」と「常時発生するログ」を特定し、ローカル処理へルーティングせよ。



2

インフラストラクチャ (Infrastructure): デスクサイドからの小さく強いスタート

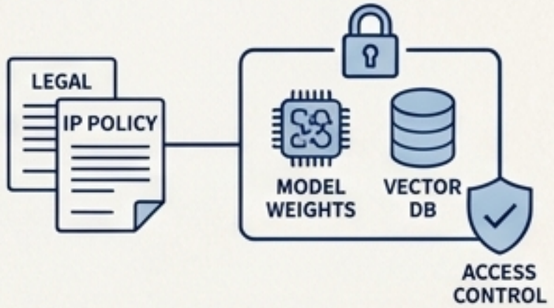
巨額のオンプレミスGPUサーバー投資の前に、まずは部門単位のデスクサイド機（ワークステーション）とスターターキットで「自社データでの推論」を検証せよ。



3

ガバナンス (Governance): 新たな知財ポリシーの即時策定

法務・知財部門と連携し、自社ファインチューニングモデルの「重みデータ」と「ベクトルDB」を営業秘密として再定義し、アクセス制御を設計せよ。



Gemini Notebook