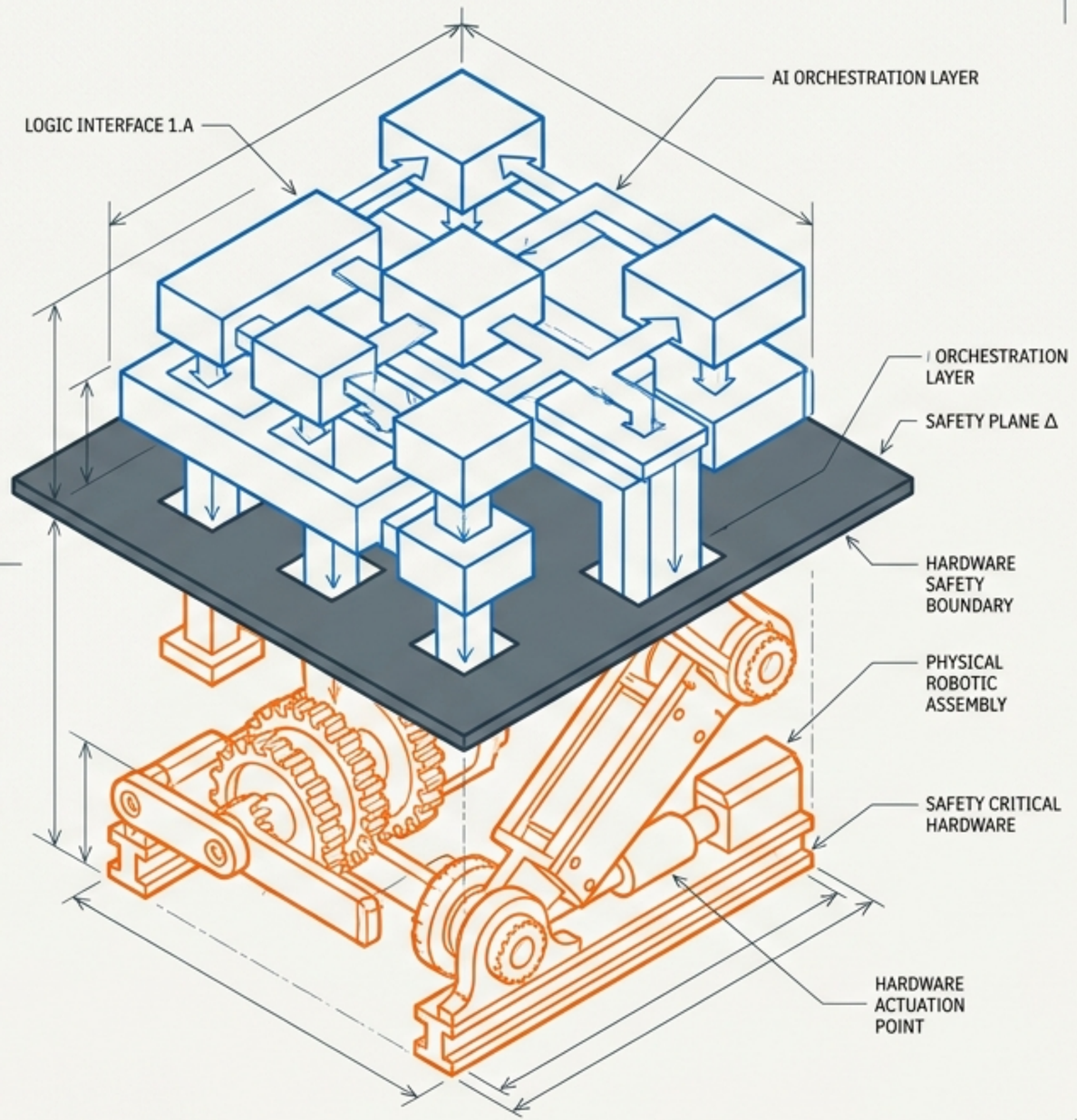
Anthropic「Model Hardware Standard（MHS）」解析解析報告

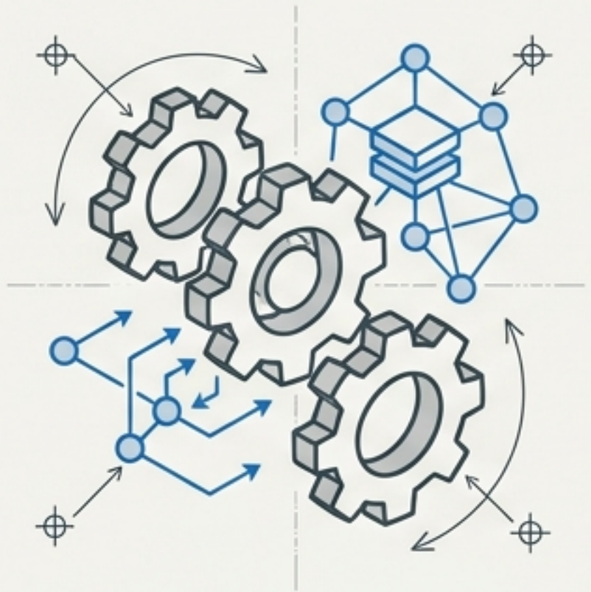
AIエージェントを物理世界に安全に着陸させるためのアーキテクチャと実装の現実

August 2026 Research Preview Analysis

Target: CTO / OT Engineers / Safety Management



Executive Summary: MHSの現在地と戦略的方針



❶ 革新性: 統合コストの劇的削減

MHSは、AIが異種物理機器を**専用コードなしで発見・操作**するための共通レイヤー（Agent-facing facade）を提供する。

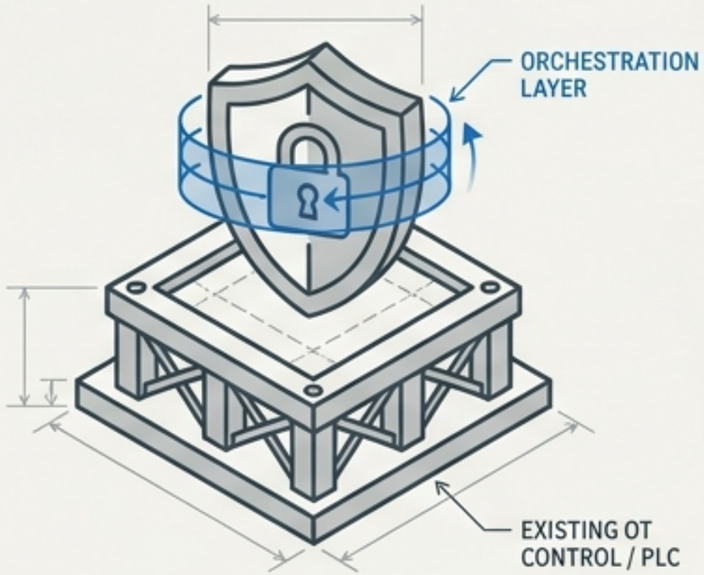
NO CODE REQUIRED

❷ 現状: あくまで「研究プレビュー」

2026年8月現在、産業用の安全・セキュリティ仕様は**未定義**。IECやOPC UAに代替する完成した標準規格ではない。

SPECIFICATION: UNDEFINED

IEC/OPC UA REPLACEMENT



❸ 結論: 既存OT制御の「上位層」

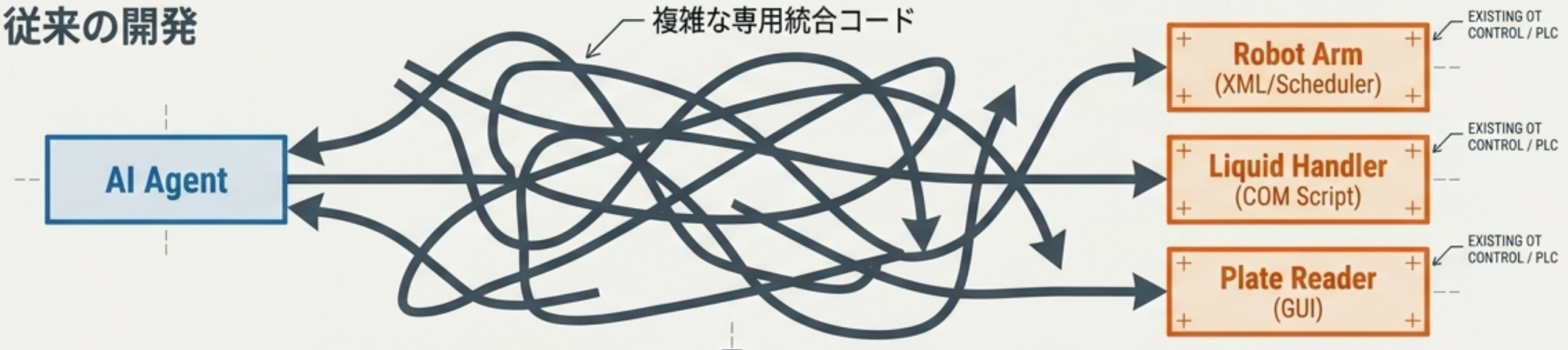
魔法の杖として単独使用せず、既存の安全PLCや通信スタックの上にオーケストレーション層として追加する実装が必須。

オーケストレーション層 ← ADDITIONAL LAYER

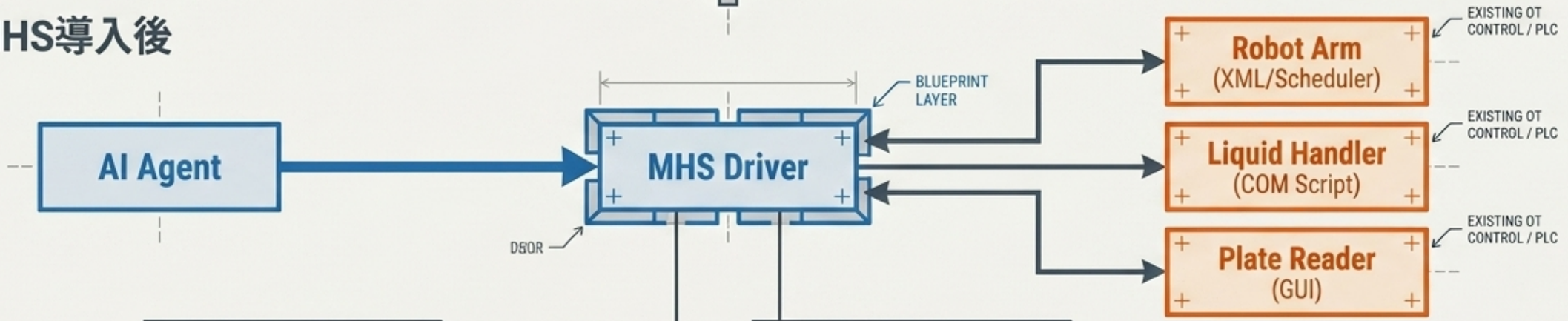
安全PLC/通信スタック ← IMPLEMENTATION: MANDATORY

What is MHS? : サイロ化からの脱却と意味的抽象化

Before: 従来の開発



After: MHS導入後



Read / Writeの抽象化

全基本操作を2つのプリミティブに集約 (例：温度取得＝Read、設定変更＝Write)。

Natural Language Tags

自然言語から機器特性や安全限界を含むReference fileを自動生成し、AIに物理制約を理解させる。

The Reality Check: 公開仕様の到達度と空白領域

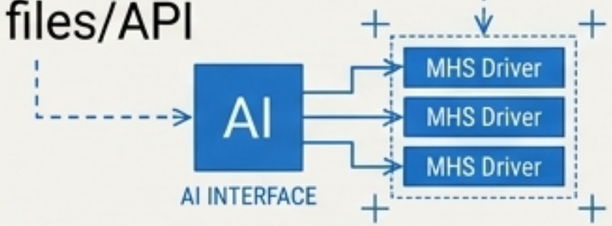
Status Dashboard

定義あり

対象範囲: Programmable interfaceを持つ装置

標準Adapter: 機器固有I/Fを変換するMHS Driver

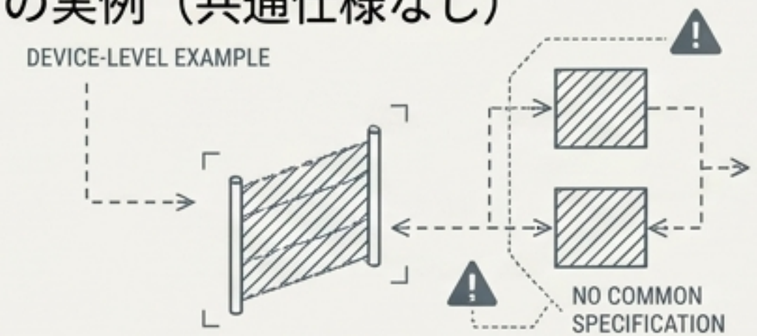
AIアクセス: MCP, CLI, Code files/API



概念定義のみ

安全限界: Device-level limitsの実装例（規範は未公開）

フェイルセーフ: 異常時ブロックの実例（共通仕様なし）



未公開 / ブランク

リアルタイム保証: Deadline / Jitter / QoSの保証値なし

CRITICAL GAP

セキュリティ: 権限モデル(RBAC等)・監査ログ仕様なし

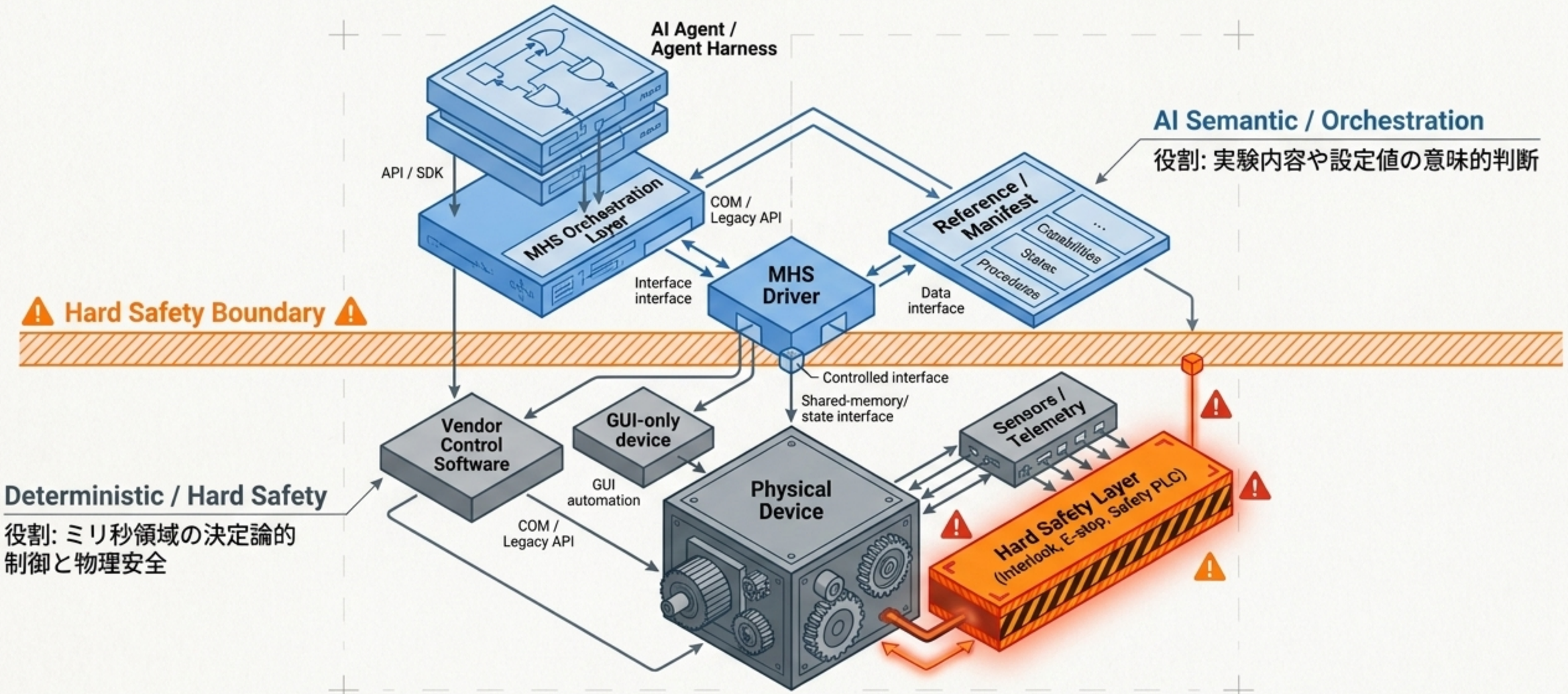
NO SECURITY MODEL

冗長性: フェイルオーバー・冪等性の規定なし

NO REDUNDANCY SPEC

Insight: MCPがJSON-RPCでツールを公開するのに対し、MHSは物理操作のパケット、データモデル、エラーセマンティクスをMUST/SHALLレベルで規定する規範文書を未だ欠いている。

Reference Architecture: 2層制御モデルの実装設計



Key Takeaway: AIの推論レイテンシに衝突回避等のリアルタイム制御を依存させず、高速処理は下位層のLocal Controlに一任する。

The MHS Workflow: 機器発見から実行までのプロセス

1. Discovery 発見
ネットワーク上の機器を標準形式で発見

3. Safety Envelope 確認
操作が安全限界内かチェック

5. Plan
LLMによる計画立案

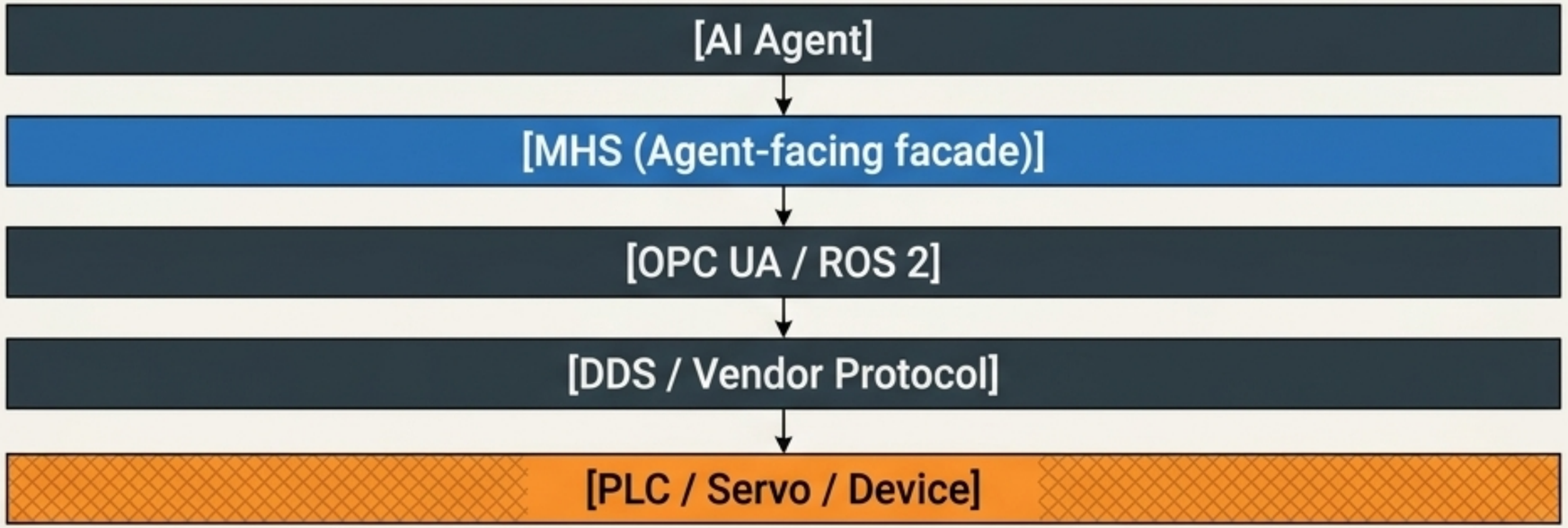
Deep-Dive: GUI操作の統合 (CMU事例)

APIを持たない機器（Plate reader等）でも、GUI自動化層をProgrammable adapterとして追加することで、MHS側の一つのStates manifestへと統合可能。

Context & Interoperability: 既存の産業規格との共存

Protocol Comparison Matrix (Diagnostic Table)

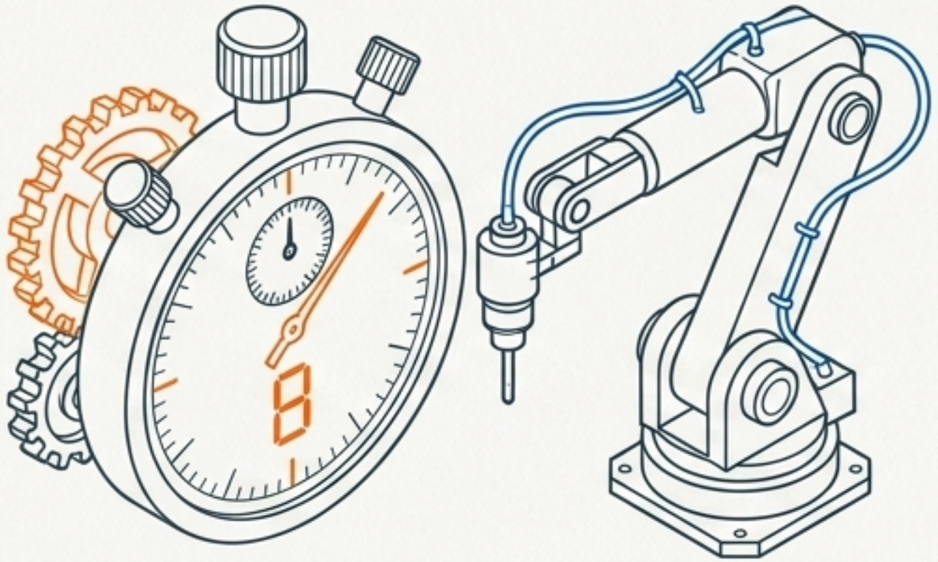
Protocol	主目的	抽象単位	QoS / リアルタイム性	セキュリティ
MHS	AI向け意味情報層	自然言語metadata	未公開	未公開
ROS 2	ロボティクスアプリ	Topic/Service	DDS由来QoS	DDS Security
DDS	分散リアルタイム通信	DataWriter等	低遅延・高信頼	成熟
OPC UA	設備の相互運用・情報	Object/Method	用途依存	極めて成熟 (Audit等)



Insight: MHSはOPC UAを代替するものではなく、複数ベンダーを跨ぐAI向け意味情報層（オーケストレーション）として機能する。

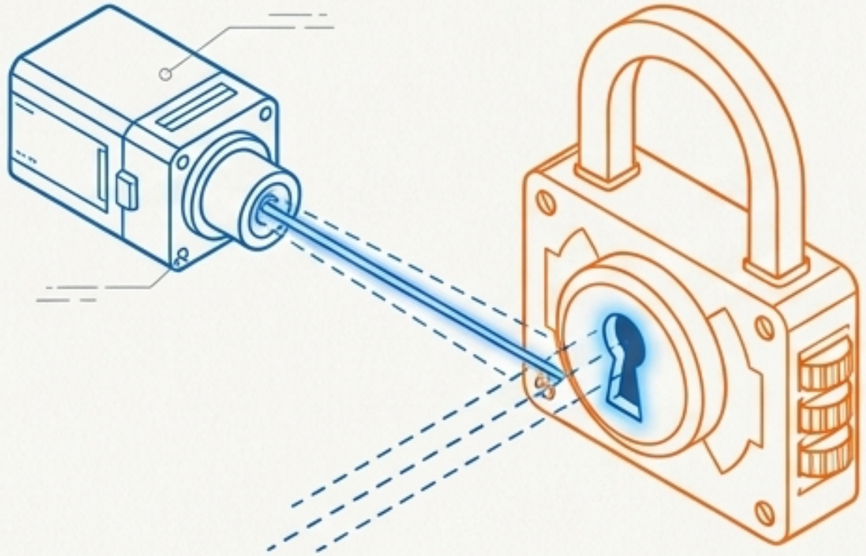
Success Cases: 実証結果とベストプラクティス

Case 1: CMU (複数ラボ機器の統合)



- **実績:** 異種4機器の統合から自律実行までをわずか「**8時間**」で完了。
- **安全実証:** E-stop等、**6種類の人工障害（Faults）**をすべて機器移動前にブロック成功。

Case 2: QuEra (量子コンピュータ用レーザー制御)



- **実績:** AI開発のレーザー復旧ロジックが700回のブラインドテスト中695回 (**99.3%**) 成功。
- **Best Practice:** 本番環境ではAIを常時ループに置かず、生成された決定論的スクリプト (**Deterministic script**) として利用。

AIの最適な活用法: AIを直接の『コントローラー』としてではなく、手順の『探索・合成器 (Synthesizer)』として活用せよ。

The Physical Risk: 論理的安全性と物理的安全性のギャップ

1. 液体中の泡（Bubble）が
センサーエラーを誘発

2. AIが論理的エラー
（パラメータ設定ミス）と誤認

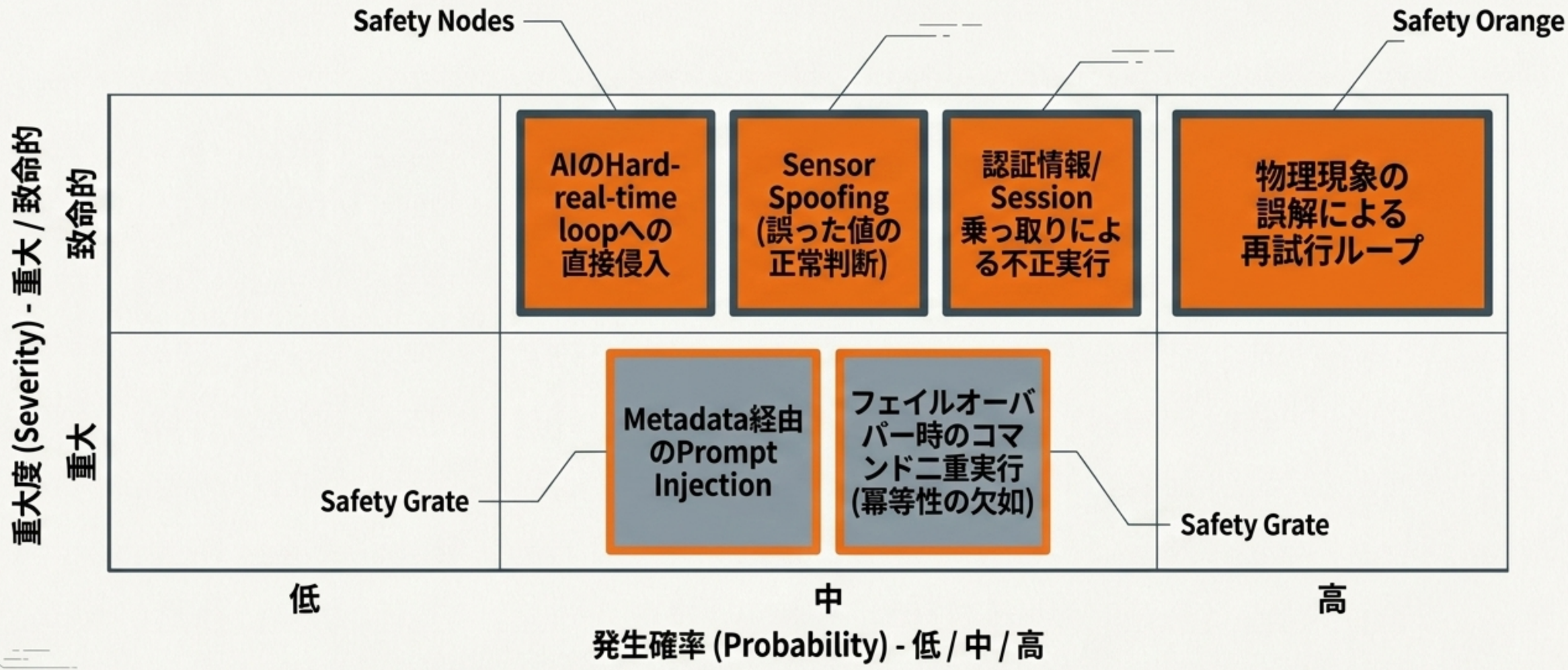
3. AIがパラメータを変更し、
同じウェルで再試行（Retry）

4. 液体が攪拌され、さらに
泡が増加（事態の悪化）

5. 人間が介入し、物理的
要因を教えるまでループ

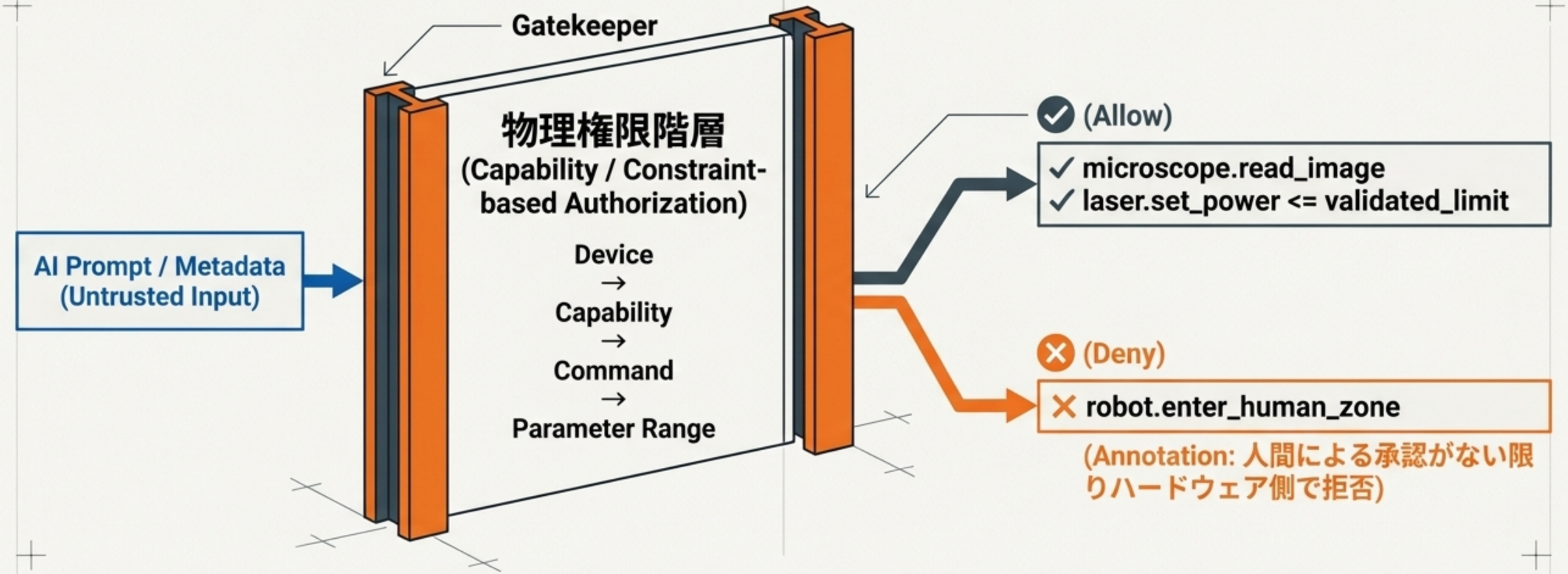
The Genentech Lesson: Model AlignmentとMachine Safetyは別物である。AIには物理的直感（Physical intuition）が欠如しており、物理現象の誤解が事態を致命的に悪化させる。

Threat Model: MHS固有のOTリスク・ヒートマップ



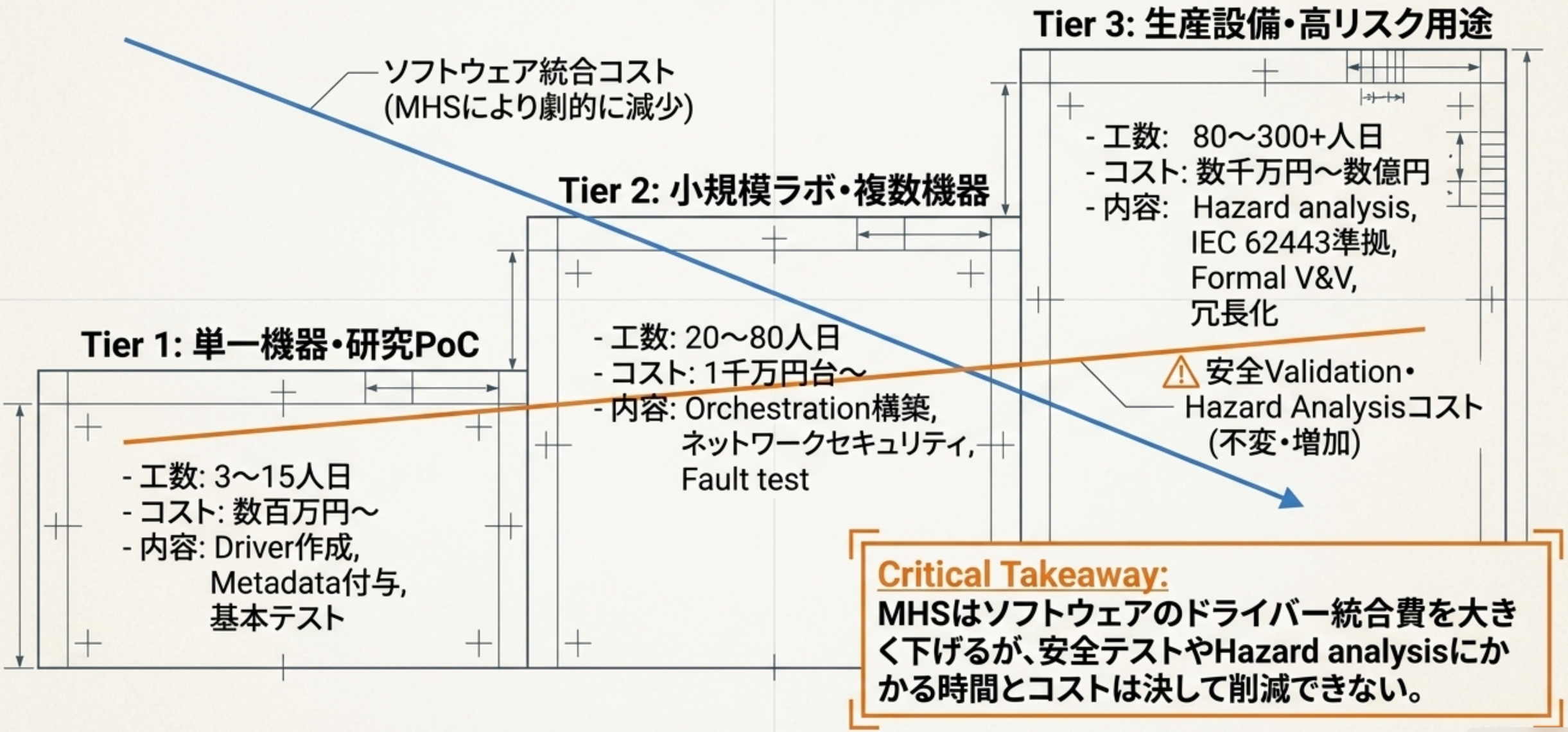
Reference: NIST SP 800-82 Rev.3に基づき、ITセキュリティだけでなくPerformance, Reliability, Safetyを一体で保護する設計が必須。

The Hard Safety Boundary: 権限階層と安全の境界線



Rule of Thumb: AIが生成する自然言語タグは信頼できない入力として扱い、安全限界は必ずドライバーやハードウェア側で強制（Machine-enforced）しなければならない。

Cost & Effort: 実装負荷と検証コストの現実



Critical Takeaway:

MHSはソフトウェアのドライバー統合費を大きく下げるが、安全テストやHazard analysisにかかる時間とコストは決して削減できない。

Implementation Checklist: 必須要件 (Must-haves)

Architecture & Safety	Security & Traceability
☐ **AIをハード・リアルタイム制御から外す** (AIはSupervisory層に留める)	☐ **高危険操作の権限分離** (危険なWrite操作には人間の承認を要求)
☐ **安全限界のハードウェア強制** (LimitはLLMではなくデバイス側で強制)	☐ **完全なAudit Trailの保存** (OPC UA水準の時系列ログを耐改ざんストレージへ)
☐ **AI生成手順の本番前固定化** (探索結果を決定論的スクリプトへコンパイル)	☐ **自然言語メタデータの隔離** (Prompt Injection対策としてUntrusted Input扱い)
☐ **物理故障へのRetry上限設定** (無制限な再試行を禁止しSafe stopへ移行)	☐ **既存安全規格の上書き禁止** (IEC等の法規要件を別途満たすこと)

Roadmap to Industrial Standard: 産業標準へのマイルストーン

Phase 1
Short-term
(Research Preview)

Focus
脅威モデルとベンチマーク

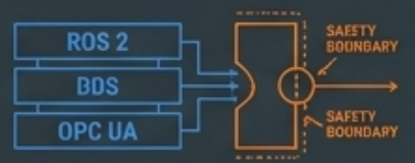
- MHS Threat modelの公開。
- CMU型のFault injection suiteの共通化。

GATE UNLOCKED

Phase 2
Mid-term
(Industrial Interoperability)

Focus
既存プロトコルとの統合

- ROS 2, DDS, OPC UAへの公式マッピング。
- 冪等性と監査（Audit）の標準化。
- ISA/IEC 62443に基づく境界管理。



GATE UNLOCKED

Phase 3
Long-term
(Certification)

Focus
第三者適合性とプロファイル分離

- Research用とSafety-Critical用のプロファイル分離。
- ISASecureのような第三者適合性認証エコシステムの確立。

ISASecure
CSTIF
SDTV

MHSは既存の制御スタックを破壊するものではない。AIを安全にオーケストレーションするための『最上位レイヤー』である。

1. Augment, Don’t Replace	2. AI as Synthesizer	3. Governance First
MHSは既存の安全・通信スタック（OPC UA, DDS, PLC）の上に「追加」せよ。	AIを直接のコントローラにせず、手順の「探索・合成器」として活用せよ。	真の価値はモデルの能力ではなく、誰が安全限界を承認し、事故時に何を監査できるかのガバナンスにかかっている。