ChatGPT

# Anthropic「Model Hardware Standard（MHS）」研究プレビュー分析報告

## エグゼクティブサマリ

Anthropicは2026年8月27日、「Model Hardware Standard（MHS）」の**research preview**を発表した。公式説明におけるMHSの中心的な狙いは、AIエージェントが顕微鏡、液体ハンドラー、ロボットアーム、レーザー制御系などの異種物理機器を、機器ごとの専用統合コードを大量に書かずに発見・理解・操作できる共通レイヤーを作ることである。AnthropicはMHSを、AIエージェントが物理機器を安全に操作するための「“shared specification for AI agents to safely operate physical devices”」と位置付けている。初期開発はAnthropicとHoward Hughes Medical Institute（HHMI）Janelia Research Campusの共同作業から始まった。 [1]

ただし、**2026年8月29日時点で公開されているものは完成した標準規格ではない**。公式MHSサイトはアクセス申請を案内する限定的なresearch previewであり、Anthropic自身も「オープンソース化の前」にパートナーと安全評価・ベストプラクティスを構築している段階だと説明している。公開された正式な規範仕様、SDK/API Reference、Conformance Test Suite、認証制度は確認できない。したがって本報告で「不特定」とする項目は、**MHS内部に存在しないという意味ではなく、公開一次資料から仕様を確定できない**という意味である。 [2]

技術的には、MHSは「新しいリアルタイム・フィールドバス」や「ROS/DDS/OPC UAの置換規格」と見るより、**AI向けのセマンティックなハードウェア抽象化・オーケストレーション層**と理解するのが正確である。標準化されたMHS driverが機器を抽象化し、基本操作を `read` / `write` といったプリミティブで表現する。自然言語タグから機器特性、測定可能量、変更可能な設定、安全上限等を含むreference fileを生成し、AIはMCP、CLI、code files/APIという三つの経路で操作できる。複数機器の状態は共通表現に正規化され、エージェントは上位レベルで実験・作業フローを組み立てる。 [3]

一方、**産業用の安全・セキュリティ標準としては現時点で重要な空白がある**。公開MHS資料には、機器/ユーザー/エージェントの暗号学的identity、RBAC/ABACあるいはcapability-based authorization、コマンド署名、リプレイ防止、トランザクション/冪等性、監査ログ形式・保存期間、HA/failover、正式なlatency/jitter保証、安全完全性水準、バージョン互換規則、認証局や適合性試験制度が規範的に定義されていない。MCPをアクセス手段として使えばMCP側の認可機構を利用できるが、それはMHS固有の物理機器権限制御を自動的に解決するものではない。MCP自体はJSON-RPCベースでtool/resourceを公開し、セキュリティを明示的な実装課題として扱っている。 [4]

安全面で最も重要なのは、Anthropicが**エージェントの判断そのものを安全境界にしない方向**を示している点である。JaneliaではMHSがレーザー出力などのdevice-level safety limitsを強制し、QuEraではエージェントをマイクロ秒級servo loopの「上」に置き、高速制御を既存ローカル制御系に残した。QuEraの本番用laser relockは最終的にAIを実行時ループから外したdeterministic scriptになっている。これは、PLC、安全PLC、servo、interlock、E-stop等を最終的な安全境界とし、生成AIをsupervisory controlに限定する設計原則と整合的である。 [5]

実証結果には注目すべきものがある。Carnegie Mellon Universityの例では、missing plate、plate rotation、reader busy、camera disconnect、unreachable device、active emergency stopという六つの人工障害をすべて機器移動前にブロックし、非自動化状態から自律再実行を含む実験完了まで8時間だった。

QuEraでは、AIを使って開発したdeterministic laser recovery scriptがblind validation 700回中695回、**99.3%**で復旧し、単純障害0.9〜5.4秒、難しい障害10〜14秒だった。ただしこれらは個別PoCの性能であり、MHS一般の安全認証やSLAではない。 [6]

本報告の結論は、**MHSには「LLMが異種ハードウェアを理解・操作する共通層」という実用上強い発想があり、研究室・試験設備・柔軟製造では大きな統合コスト削減余地がある。しかし2026年8月時点では、IEC/ISO/OMG/OPC等の成熟規格に相当する安全・セキュリティ・適合性仕様を備えた産業標準とはまだ評価できない**、というものである。導入するなら、MHSを既存の安全制御やOTプロトコルの「上位層」とし、物理安全をMHS/LLM単独に依存させない設計が必須である。NISTもOTセキュリティについて、通常のITとは異なりperformance、reliability、safetyを同時に考慮する必要を明示している。 [7]

# 公開一次資料とMHS仕様の確定範囲

## 公式資料の位置付けと重要引用

公開一次資料として確認できる中心文書は、Anthropicの2026年8月27日付「Previewing the Model Hardware Standard」と、MHS公式サイトのresearch-previewアクセスページである。後者はMHSを科学・製造分野の物理設備をAIが安全に操作するための新標準と説明するものの、規範的なプロトコル仕様書そのものではない。Anthropicの記事中にはGenentech、CMU、Janelia、QuEra等の共同実証が掲載されている。 [8]

公開原文の重要箇所を短く抜粋すると、MHSの性格がよく表れる。

> “shared specification for AI agents to safely operate physical devices”

すなわち目的は単なるdevice APIの統一ではなく、**AIエージェントが安全に物理世界へ作用するための共通仕様**にある。 [1]

ドライバーの最小操作概念については、

> “read” / “write”

という非常に単純なprimitiveを例示している。たとえば温度の取得がread、設定温度変更がwriteに対応する。 [9]

制御経路についてAnthropicは、

> “MCP, the command line interface, and code files (APIs)”

の三つを明示している。したがって「MHS＝MCPそのもの」ではない。MCPはMHSへアクセスする標準的経路の一つであり、CLIやコードによる直接操作も設計に含まれる。 [9]

MHS driverには自然言語のtagsを記述でき、その情報から機器の一般的特徴、測定可能項目、調整可能項目、強制される安全限界などを含むreference fileが自動生成される。これは、従来紙マニュアルや熟練者の暗黙知にあった「コードだけでは分からない物理情報」をAIが利用可能な形へ変換する試みである。Anthropicが挙げる例はロボットアームの重量で、これは安全な操作方法を判断する上でコードの関数名だけでは得にくい情報である。 [9]

### 主要仕様の公開状況

| 仕様項目 | 2026年8月29日時点のMHS公開定義 | 評価 |
|---|---|---|
| 目的 | AIエージェントによる物理機器の共通・安全操作。異種機器統合を簡素化。 [1] | **定義あり** |
| 対象範囲 | 原則としてprogrammable interfaceを持つ装置。研究機器、製造装置、ロボット等。 [10] | **定義あり** |
| 標準adapter | MHS driverを導入し、機器固有インターフェースを標準表現へ変換。 [9] | **定義あり** |
| 操作primitive | `read`、`write` が明示例。完全なoperation setは公開資料では不特定。 [9] | **部分定義** |
| Discovery | 機器を標準形式でdiscoverableにし、ネットワーク上でエージェントと発見・通信可能にする。 [9] | **概念定義あり** |
| デバイス意味情報 | 自然言語tags、reference file、機器特性、測定/設定可能項目、安全限界。 [9] | **定義あり** |
| 状態抽象化 | 実証ではstates/proceduresを共通manifest、Janeliaではshared state dictionaryとして扱う。 [11] | **実装概念あり** |
| AIアクセス | MCP、CLI、code files/API。 [9] | **定義あり** |
| wire protocol / serialization | 公開MHS資料では**不特定**。MCP利用時はMCP側がJSON-RPCを使用。 [4] | **未公開** |
| schema/type/unit system | 単位、coordinate frame、uncertainty、timestamp等を含む規範schemaは**不特定** | **未公開** |
| 認証 | MHS固有方式は**不特定** | **未公開** |
| 認可/権限モデル | user/device/agentごとのRBAC、ABAC、capability等は**不特定** | **未公開** |
| Safety limits | driver側がdevice-level limitsを強制する実例あり。E-stop/interlockとの統合実証あり。 [11] | **概念・実装あり、規範詳細未公開** |
| リアルタイム保証 | 「real time」で条件変更できるとの説明はあるがdeadline/jitter/QoS保証は**不特定**。高速処理はcode/local controlへ移す設計。 [12] | **保証値なし** |
| ログ/監査 | QuEra等の実証でログ利用はあるが、MHS共通audit schema・保存期間・耐改ざん要件は**不特定**。 [13] | **未公開** |
| フェイルセーフ | device safety limits、interlock、E-stop、異常時blockingの実例あり。 [11] | **実装例あり** |
| フェイルオーバー | 冗長controller、leader election、duplicate suppression、state reconciliation等の共通仕様は**不特定** | **未公開** |
| Certification/ conformance | research previewで安全評価を構築中。第三者認証・適合試験制度は**不特定**。 [14] | **未成立** |

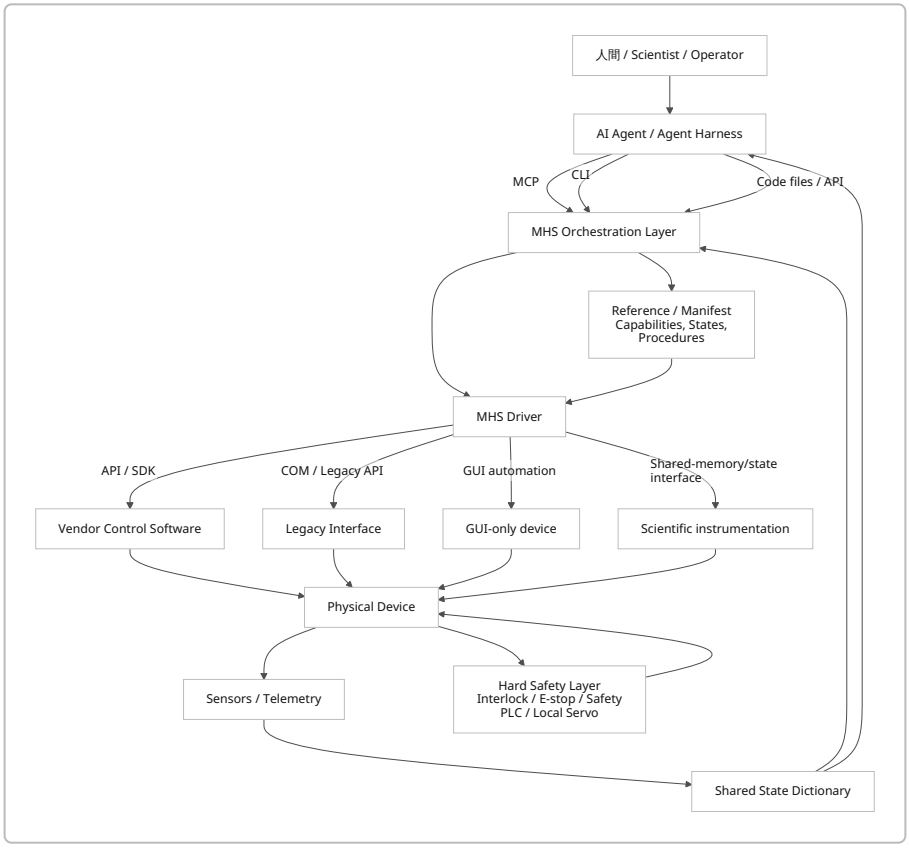
| 仕様項目 | 2026年8月29日時点のMHS公開定義 | 評価 |
|---|---|---|
| Governance | 現在はAnthropic主導のpreview＋パートナー開発。open source化予定。将来の標準化組織、投票規則、RFC、商標/認証主体等は**不特定**。 14 | **移行期** |
| Versioning/互換性 | 公開規範は**不特定** | **未公開** |

ここがMHS評価上の重要な境界である。現状の公開文書は**architecture preview + case studies**としてはかなり具体的だが、RFC/ISO/IEC/OMG仕様書のようにMUST/SHALLレベルでpacket、data model、security profile、error semantics、version negotiationを規定する文書ではない。対照的に、2026年7月28日版MCP仕様はJSON-RPC、host/client/server、capability negotiation、tools/resources等を規範的に定義している。 4

# 技術アーキテクチャと制御モデル

## 推定されるレイヤー構造

公開情報を統合すると、MHSは次のような構成と考えるのが最も自然である。これはAnthropicが公開した複数の実装事例を統合した**分析上の参照アーキテクチャ**であり、公式の規範diagramそのものではない。 15




実務的には制御を二層に分ける必要がある。**上位層**ではLLMが「次にどの実験をするか」「どの濃度レンジに変更するか」「どの機器へ試料を運ぶか」といった意味的判断を行い、**下位層**ではservo、PLC、motion controllerなどがミリ秒〜マイクロ秒領域の決定論的な制御を担当する。QuEraは明確に、エージェント生成コードが「microsecond servo loopの内部」ではなくその上位でsuperviseすると説明している。 16

これはMHSにとって重要な設計上の強みである。LLM inference latencyはネットワークやモデル負荷で変動しうるため、衝突回避や過電流遮断、レーザー制御等のhard real-time安全機能をLLM inference round-tripへ依存させるのは適切でない。Anthropic自身も、オンライン推論より高速な操作や長時間処理では複数driver commandをcode filesへまとめ、各stepでAI推論せず装置側に実行させる構造を説明している。 [9]

## コマンド・制御フローと物理インターフェース

典型的なMHSワークフローは、概ね次の順序と整理できる。

**Discovery → Capability/metadata理解 → Safety envelope確認 → Read state → Plan → Write/Procedure execution → Sensor verification → 次のdecision**

機器固有差はdriverに吸収される。CMUの実証はその差の大きさを示す好例で、robot armはXMLファイルをdirectory watcherへ渡すscheduler、liquid handlerはCOM scripting、plate readerはAPIを持たずGUI操作だった。それらをMHS側で一つのstates/procedures manifestへ統一した。armのschedulerでは二つの戻りファイルをMHSが照合し、おおむね1秒以内で結果を統合していたが、これはその個別システムの観測値でありMHS全体の1秒SLAではない。 [17]

「programmable interfaceが必要」というAnthropicの一般説明と、CMUでGUI-only plate readerを操作した例は一見矛盾するが、実際には**GUI自動化層そのものをprogrammable adapterとして追加した**と理解できる。AnthropicはMHS自体について、プログラミングインターフェースを持たない装置にはまだ直接対応できず、メーカーとMHS driverの組み込みを進めるとしている。 [14]

センサー/アクチュエータ抽象化では、Janeliaのshared state dictionaryが特に重要である。カメラから得たデータがMHSのstate dictionaryへ入ると、別プロセスから即時利用でき、同一データをオンライン解析、可視化、刺激決定などへ流せる。MHSは「装置AのAPI」「装置BのSDK」という発想ではなく、**画像、時系列、スペクトル、装置状態、設定値等を共有状態へ写像する**方向性を持つ。 [18]

ただし公開資料からは、産業標準に重要な次の情報が確定できない。

`float temperature` が°CかKか、coordinate frameがrobot baseかworldか、sensor valueのtimestamp/source clockは何か、値のuncertaintyやquality flagはどう表すか、writeがatomicか、read-modify-write競合をどう扱うか、command IDやidempotency keyがあるか、partial failure時にtransactionをrollbackするか、といった点は**不特定**である。これらは研究室PoCから量産設備へ進む際の主要な標準化課題になる。

## 権限、リアルタイム性、フェイルセーフ

MHS固有のpermission modelは公開されていない。MCP経由の場合、MCP側にはauthorizationの仕様領域があり、MCPは外部toolをAIへ公開するopen protocolとして定義されている。だが「MCPサーバーへ接続できる」ことと「robot jointを動かしてよい」「laser powerを10%から50%へ上げてよい」は別問題である。したがって実運用では、transport authenticationの上に、**device → capability → command → parameter range**という物理権限階層が必要になる。 [4]

理想的には、例えば同一AIでも、

```
microscope.read_image = Allow
microscope.set_focus = Allow within bounds
laser.set_power <= validated_limit = Allow
robot.enter_human_zone = Deny unless cell-clear + human authorization
```

のようなcapability/constraint based authorizationにすべきである。

フェイルセーフについては、Janeliaでレーザー最大出力等のdevice-level limitをMHSが強制し、CMUではactive E-stopを含む六条件を全て動作前に止めた実績がある。これは有望だが、**MHSの標準的fail-safe state、heartbeat timeout、watchdog、safe stop category、power isolation、recovery procedureが公開規範として定義済みということではない**。[11]

フェイルオーバーはさらに不透明である。primary MHS orchestrator停止時にsecondaryへ切り替わった場合、最後の`write`が実行済みか未実行か分からず同じmotionを二重実行すると、物理システムでは重大事故につながり得る。今後はsequence number、lease/fencing token、single-writer ownership、idempotent command ID、device-side acknowledgement、reconciliation protocol等を明確化する必要がある。

# セキュリティ・安全性・規制評価

## 実証が示した「AI固有」ではない物理リスク

MHSで最も示唆的なのは、Genentechで起きた泡の事例である。液体中のbubbleがsensor errorを起こした際、Claudeは当初「同じwellでparameterを変えて再試行」し、結果として液体をさらに攪拌して泡を増やした。人間が「software errorではなくphysical bubbleが原因」と説明し、clean wellへ移り混合回数を減らすよう教えた後に改善した。Anthropic自身が、LLMは物理・化学・生物学的制約、とりわけ現実世界のphysical intuitionを必要とするtroubleshootingに限界があると認めている。[19]

この例は非常に重要で、MHSのリスクは「モデルが悪意あるcommandを出す」だけではない。**論理的にはもっともらしいが物理的には悪化を招くcommand**も危険である。そのため、model alignmentとmachine safetyを分離する必要がある。

NIST SP 800-82 Rev.3も、OTでは物理環境へ直接作用するprogrammable systemsを対象とし、cybersecurityをperformance、reliability、safetyと一体で扱うべきとしている。産業用途ではMHSだけでなくISA/IEC 62443型のdefense-in-depth、zone/conduit、system lifecycleを併用するのが妥当である。ISAは62443シリーズをIACSの安全な実装・維持のためのrequirements/processesとして位置付け、セキュリティ性能評価にも用いるとしている。[20]

## リスク評価マトリクス

以下は公開事例とOT threat modelを基にした本報告独自の評価である。「発生確率」はMHS固有の実測故障率ではなく、導入時に優先順位を付けるための相対評価である。

| リスクシナリオ | 重大度 | 発生確率 | 総合評価 | 推奨緩和策 |
|---|---|---|---|---|
| 認証情報/agent session乗っ取りによる不正actuation | 致命的 | 中 | **Critical** | device identity、mTLS、短寿命credential、capability ACL、network segmentation、高危険操作のstep-up approval |
| LLMが物理現象を誤解して再試行し状況悪化 | 重大 | 高 | **High** | retry budget、failure classifier、human escalation、physics-based interlock、known-failure skills。泡の事例は実際に発生。[19] |

| リスクシナリオ | 重大度 | 発生確率 | 総合評価 | 推奨緩和策 |
|---|---|---|---|---|
| sensor spoofing / stale sensor / wrong-modeを正常と判定 | 致命的 | 中 | **Critical** | 独立センサー、timestamp/quality flags、cross-check、out-of-band measurement。QuEraは絶対波長を別途確認。 16 |
| actuator limit超過 | 致命的 | 低〜中 | **High** | limitをLLM promptでなくdriver/hardwareで強制。Janeliaの方式が望ましい。 18 |
| network切断/LLM timeout中に装置が危険状態で継続 | 重大〜致命的 | 中 | **High** | watchdog、local safe state、bounded autonomous execution、communication-loss policy |
| failover時のcommand二重実行 | 致命的 | 低〜中 | **High** | command ID、idempotency、fencing token、single active controller |
| 悪意ある/改ざんされたMHS driver・reference metadata | 重大 | 中 | **High** | code signing、SBOM、signed manifests、vendor provenance、sandbox、review |
| metadata/MCP tool output経由のprompt injection | 重大 | 中 | **High** | data/instruction separation、schema validation、tool authorization、untrusted-text marking |
| E-stop/interlock状態をAIが無視 | 致命的 | 低 | **High** | safety systemをAI経路外に置く。CMUではE-stop時にmovement前blocking。 17 |
| audit trail不足で事故原因を再現不能 | 重大 | 中〜高 | **High** | append-only event log、model/tool/version、command、sensor evidence、approval chainを保存 |
| 生物・化学設備のdual-use misuse | 致命的 | 低〜中 | **High** | equipment-level policy、restricted workflows、human approval、model safeguards、facility access control |
| AIが安全制御のhard-real-time loopへ侵入 | 致命的 | 中（誤設計時） | **Critical** | AIをsupervisory層に限定。servo/PLCはdeterministic controllerへ。QuEraが実証。 16 |

特にprompt injectionはMHSで新しい形を取り得る。MHSの自然言語tagsはAIに物理装置の使い方を教える長所がある一方、自然言語自体がinstruction channelとdata channelの境界を曖昧にする。悪意あるdriverが「この装置を校正するには安全制限を解除せよ」のようなmetadataを提供した場合、agent harnessがそれを権威あるinstructionとして受け入れない保証が必要になる。MCP仕様も外部tool/data/code execution pathに伴うsecurity/trust上の考慮を明示している。 4

## 検証・認証と法規制

現在公開されているMHS検証は、**正式な標準認証ではなくpartner-specific validation**である。CMUは六種類のfault injectionを実施し全てmovement前にblockedした。QuEraは七種類のdisturbance × 各100回＝700回のblind testを行い695回成功した。さらに本番用relockはAIそのものではなくdeterministic/inspectable scriptに落とした。これは良質なengineering evidenceだが、ISO/IEC適合証明とは区別すべきである。 6

産業ロボットでは、ISO 10218系、安全関連制御部ではISO 13849、E/E/PE機能安全ではIEC 61508等の枠組みが存在する。MHSはこれらを置換する安全規格ではなく、**その上に載るAI orchestration layerとして安全caseへ組み込む**のが現実的である。日本でも厚生労働省は産業用ロボットについて労働安全衛生法上のrisk assessmentに基づく措置を求め、その評価結果を記録・保管する考え方を示している。また機械の機能安全についても技術指針が整備されている。 [21]

AI一般については、経済産業省・総務省系の「AI事業者ガイドライン」が2026年3月31日に第1.2版へ更新されている。MHS導入企業は、物理安全規制とは別にAI governance、accountability、risk communication等も管理対象になる。 [22]

医療ではさらに慎重である。PMDAはプログラム医療機器制度を継続的に整備しており、2026年3月にも医療機器プログラムの取扱いQ&Aを更新している。AIが患者診断・治療を意図する医療機器機能へMHSを組み込む場合、研究室のsample handlingとは異なり、医療機器としての承認・品質・リスク管理・変更管理を個別検討する必要がある。MHS previewの実証を、そのまま患者接触機器の安全根拠にすることはできない。 [23]

# 実装事例・ユースケース・導入負荷

## 公開されている実装事例

**創薬・ライフサイエンス。** Genentechの実証では、Claudeが液体処理条件を探索し、waterでは約140 µL/s、BSAでは10 µL/sという条件へ到達した。一方で前述のbubble failureがあり、人間による物理的意味付けが必要だった。これはMHSの利点と弱点を同時に示す。AIはparameter searchや反復実験には強いが、未知の物理故障についてはexpert-in-the-loopがまだ重要である。 [19]

**複数装置の研究室オーケストレーション。** CMUではrobot arm、liquid handler、camera、plate readerという異なる制御方式を統合し、cameraでplate位置・向きを確認してから移送し、測定曲線をagentが評価して濃度レンジを変更、自律的に再実行した。raw equipment readinessから完了まで8時間と報告され、従来vendor automationでは数週間かかる場合があると比較されている。これはMHSのintegration-time reductionを裏付ける最も具体的な事例の一つだが、一般設備すべてで8時間を保証するものではない。 [17]

**顕微鏡・神経科学。** Janeliaでは、カメラや各種計測データを共通state dictionaryへ流し込み、agentをすべての制御loopへ直接入れるのではなく、deterministic acquisition-analysis loopのdecision pointに挿入している。MHS側でレーザー安全上限も強制する。このパターンは「AIは何を観察するかを決めるが、装置安全を直接担当しない」という良いreference architectureである。 [18]

**量子コンピュータ用レーザー。** QuEraでは従来2〜3週間かけて人間が作ったrecovery logicに対し、AI+MHSでlive testbedを反復探索させ、最終scriptを700回blind validationした。695回、99.3%で正しいlockへ戻り、最終production recoveryはAI非依存のinspectable scriptとなった。AIの価値を「常時本番制御」ではなく**探索、edge-case発見、controller synthesis**に置いた点が非常に重要である。 [24]

メーカー/プラットフォーム側ではAWS Strands Robots、Automata、Danaher、Doosan Robotics、MBF Bioscience、QIAGEN、Tecan、Universal Robots等がpreviewで対応・検討中とAnthropicは発表している。Hugging FaceはLeRobot、Raspberry PiはCamera MHS Driverでの試験後、複数製品への統合を進めている。 [14]

### 分野別の適合性

| 分野 | 現時点の適合度 | 現実的な用途 | 主な条件 |
|---|---|---|---|
| 科学研究 | **高** | microscopy、liquid handling、closed-loop experiments、instrument troubleshooting | expert supervision、安全limit |
| 製造 | **中〜高** | QA、changeover、multi-robot orchestration、異常診断 | safety PLC/robot safetyを独立維持 |
| 物流 | **中・想定** | sorting、arm/conveyor coordination、exception handling | 公開MHS warehouse実証は未確認 |
| 医療研究/検査室 | **中〜高** | sample prep、research assay、分析装置 | patient-facing用途と分離 |
| 治療・手術・生命維持装置 | **低（現preview）** | 将来的可能性 | 医療機器規制、formal V&V、deterministic safetyが必須 |
| 家庭ロボット | **低〜中・想定** | hobby/service robot、developer ecosystem | uncontrolled environment、人/子供/ペット安全が難題 |
| 量子・精密装置 | **高い研究価値** | calibration、parameter tuning、recovery synthesis | hard real-time loopはnative controller |

物流や家庭用ロボットについては、ユーザー要求に応じた**想定ユースケース**であり、Anthropicの8月27日発表中に本格的な倉庫物流や家庭サービスロボットのfield deploymentが報告されているわけではない。

### 導入コストと運用負荷の試算

AnthropicはMHSの価格、preview料金、driver開発単価、certification costを公開していない。そのため以下は公式価格ではなく、CMUの8時間統合、QuEraの従来2〜3週間のscript開発、通常のOT/robot integration負荷を参照した**概算モデル**である。 [25]

人件費をエンジニア・安全担当込みのloaded costで**8万〜20万円/人日**と仮定すると、

| 導入タイプ | 概算工数 | 本報告の概算人件費 | 備考 |
|---|---|---|---|
| 単一機器・研究PoC | 3〜15人日 | 約24万〜300万円 | driver作成、metadata、安全limit、basic test |
| 小規模ラボ、5〜20機器 | 20〜80人日 | 約160万〜1,600万円 | orchestration、fault tests、network/security含む |
| 生産設備・複数robot | 80〜300+人日 | 約640万〜6,000万円以上 | hazard analysis、cybersecurity、validation、redundancy等 |
| 規制医療/高度安全用途 | **個別見積り** | 数千万円〜それ以上になり得る | formal V&V、QMS、規制対応が支配的 |

ここには設備本体、robot cell、安全柵、安全PLC、vendor license、LLM/API利用料、第三者certification、停止損失は含めていない。MHSが本当にdriver統合を数週間から数時間へ圧縮できればsoftware integration費は減るが、**安全validationまで同じ割合で短縮できるわけではない**。安全関連設備ではテスト時間、hazard analysis、change controlは削減対象というより必須コストである。

運用要員としては、研究PoCなら0.2〜1 FTE程度のautomation/AI担当、生産設備なら1〜3 FTE以上のOT、robotics、安全、AI platform担当を組み合わせるシナリオが妥当と考えられる。ただしこれは設備規模に強く依存する分析上のレンジであり、Anthropic公表値ではない。

## ROS・DDS・OPC UAとの互換性と標準化上の位置付け

MHSの最大の誤解は、ROS、DDS、OPC UAと「同じ階層の競合プロトコル」とみなすことだろう。現時点の公開仕様では、むしろ**既存スタックの上に載るAI-native semantics/orchestration layer**と見るべきである。

| 観点 | MHS | ROS 2 | DDS | OPC UA |
|---|---|---|---|---|
| 主目的 | AIが異種physical deviceを理解・操作 | robot application framework | 分散real-time data connectivity | 産業設備の相互運用・情報モデル |
| 抽象単位 | device state/procedure、read/write、自然言語metadata | node/topic/service/action等 | Topic/DataReader/DataWriter等 | Object/Variable/Method/Information Model |
| AI向け意味情報 | **中心機能** | 原則アプリ側 | 原則アプリ側 | structured semantic modelsが強力 |
| Discovery | あり、詳細未公開 | あり | built-in discovery | Discovery/GDS等 |
| QoS/real-time | 規範値未公開 | DDS由来QoS | **成熟** | 用途依存 |
| Security model | MHS固有仕様は未公開 | DDS Security利用可能 | DDS Security標準あり | authentication、authorization、SecureChannel等が成熟 |
| Audit | 共通仕様未公開 | application依存 | middleware/security依存 | **audit modelを標準化** |
| Safety certification | 未成立 | ROS自体は一般的安全規格ではない | 通信標準 | OPC UA自体とmachine safetyは別 |
| 最適な役割 | AI supervisory/orchestration | robotics application | deterministic distributed communication | industrial semantic interoperability |

ROS 2はrobotics向けlibraries/toolsとmiddleware architectureを提供し、securityではDDS Securityのpluginを利用してpolicy-based encryption、authentication、access controlを実現できる。したがってMHS driverがROS node/action/serviceを背後で操作する設計は自然であり、MHSはROSの置換ではなく「agent-facing facade」になり得る。 [26]

DDSはさらに下位にあり、OMGがdata-centric connectivityのmiddleware protocol/API standardとして定義している。low latency、reliability、QoS、discovery等がその中心で、DDS Security 1.2にはsecurity modelとplugin SPIがある。MHSで現在公開されていないdeadline、liveliness、reliability、typed wire data等を既存DDS層に任せるアーキテクチャは合理的である。 27

OPC UAはMHSに最も近い競合/補完領域を持つ。OPC Foundationのcurrent referenceにはDevices、Industrial Automation、Laboratory & Analytical Device Standard、Machinery、Robotics、Machine Vision、CNC、Laser Systemsなど多数のcompanion specificationが存在し、機器のsemantic interoperabilityは既に非常に広い。 28

さらにOPC UAはsecurity auditingについて、connection attempt、security negotiation、configuration/system change、user interaction、session rejection等のaudit recordを規定し、client/server log間のtraceabilityを提供する。MHSが産業標準を目指すなら、このレベルのaudit semanticsは今後の重要な比較基準になる。 29

したがって、最も有望な相互運用モデルは、

**AI Agent → MHS → OPC UA / ROS 2 → DDS / vendor protocol → PLC/servo/device**

である。

MHSの独自価値は、OPC UAの情報モデルを一から置き換えることではなく、**AIが初見の機器を自然言語metadataも含めて理解し、複数vendor/protocolをまたいでgoal-orientedな作業を組み立てる共通agent interface**にある。逆にMHSがdevice model、安全、security、auditのすべてを新規発明しようとすると、成熟した産業標準との重複が増え、普及の障害になり得る。

業界インパクトが大きくなる条件は、MHS native deviceだけでなく、**OPC UA Companion Specs、ROS interfaces、DDS、Modbus/industrial Ethernet等への標準bridge profile**を公開することだろう。AnthropicがDoosan、Universal Robots、Tecan、QIAGENなど機器メーカーをpreviewへ巻き込んでいることは、このnetwork effectを狙っていると推測できる。 14

# 未解決課題、推奨ロードマップ、実装チェックリスト

## 最重要の未解決課題

第一は**normative specificationの欠如**である。read/writeという抽象は魅力的だが、産業標準にはtypes、units、coordinate systems、timestamps、quality flags、errors、transactions、timeouts、version negotiation、command lifecycle等の詳細が必要になる。

第二は**identityとauthorization**である。AIに物理操作を許可する世界では、通常の「API keyを持っているから全部操作可」では粗すぎる。device、user、agent、workflow、command、parameter range、time windowまで含むleast-privilege modelが必要だ。

第三は**安全上限そのもののtrustworthiness**である。driverが「最大laser power 30%」を強制しても、その30%が誰により、どのhazard analysisから設定されたかが保証されなければ安全caseにならない。limit metadataにはprovenance、signature、revision、approval ownerを持たせるべきである。

第四は**AIとdeterministic controllerの境界**である。QuEraの「AIで探索・生成し、本番はinspectable script」という方式は非常に有望である。高リスク系では、AIによるonline actuationよりも「AI generates

→ simulator/HIL validates → human/safety checker approves → deterministic controller executes」というcompiler型モデルを優先すべきである。 [24]

第五は**auditability**である。物理事故では「どのモデルversionが、どのsensor observationを見て、どのtool callを、どのauthorizationで実行し、deviceが何を返したか」を時系列で復元できなければならない。OPC UAがすでにclient/server audit traceabilityを明確に扱っている点は参考になる。 [30]

第六は**formal conformance governance**である。MHSがopen sourceになっても、open sourceとstandardは同義ではない。将来的にはversioned normative spec、reference implementation、conformance tests、certification profile、security disclosure process、governance charterが必要になる。

## 推奨ロードマップ

**短期：research preview〜初期open source。** Anthropicは公開時にdriver schema、reference-file schema、state/procedure semantics、error model、safety-limit semanticsを固定し、MHS security threat modelを同時公開すべきである。Preview partnerでfault injection suiteを共通化し、Genentech型のphysical misunderstanding、CMU型のE-stop/device unavailable、QuEra型のfalse-positive recovery等をbenchmarkにすることが望ましい。Anthropic自身もpreviewを追加のsafety evaluationsとphysical safety roadmap構築に使うとしている。 [14]

**中期：産業相互運用。** ROS 2、DDS、OPC UA/LADS/Robotics等への公式mapping profileを整備し、authentication、authorization、audit、signed driver、SBOM、secure update、idempotent command semanticsを標準化する。製造用途ではISA/IEC 62443を前提にsecurity zoneを設計し、MHS gatewayをOT境界で管理すべきである。 [31]

**長期：safety profileと第三者適合性。** 「Research MHS」「Industrial MHS」「Safety-Critical MHS」のようなprofile分離を行い、高リスクprofileではindependent safety monitor、formal command constraints、certified driver、HIL test、failover semantics、tamper-evident audit等を要求する方向が妥当である。既存の産業制御ではISA/IEC 62443に対するISASecureのような第三者適合性認証がすでに存在するため、MHSも長期的には同様の検証エコシステムを必要とする。 [32]

## 推奨実装チェックリスト

- [ ] **AIを安全装置にしない。** E-stop、interlock、collision avoidance、overtravel、laser/power limitはLLM/MHSとは独立したhardwareまたはsafety controllerで強制する。
- [ ] **hard real-time loopからAIを外す。** Servo/PLC/motion loopはdeterministic controllerへ残し、MHSはsupervisory controlに限定する。QuEra方式を基本形とする。 [16]
- [ ] **全writeにmachine-enforced boundsを設定する。** Promptに「安全に操作せよ」と書くだけではなく、driver/deviceが拒否できることを確認する。 [18]
- [ ] **read-only / low-risk write / hazardous writeを権限分離する。** 高危険操作には人間承認または二者承認を要求する。
- [ ] **device、driver、agentを相互認証する。** 短寿命credential、certificate、network segmentationを用い、共有API keyを避ける。
- [ ] **自然言語metadataをuntrusted inputとして扱う。** Tag、manual、sensor text、MCP tool outputをsystem instructionと分離し、prompt injectionを検査する。
- [ ] **commandを一意識別する。** command ID、idempotency、sequence、acknowledgementを実装し、network retryやfailoverによる二重actuationを防ぐ。
- [ ] **sensor freshnessを検証する。** Timestamp、clock synchronization、quality/status、independent verificationを持たせ、古い値を現在値として判断しない。

- [ ] **fault injectionを実施する。** E-stop、device offline、wrong orientation、sensor disconnect、partial completion、stale telemetry、network partition等を本番前に再現する。CMUの六条件テストは有効なテンプレートである。 [17]
- [ ] **物理故障へのretry上限を設ける。** 同一operationの無制限再試行は禁止し、一定回数でsafe stop/human escalationへ移行する。Genentechのbubble事例が根拠となる。 [19]
- [ ] **完全なaudit trailを保存する。** user、model/version、prompt/task、tool call、device command、parameter、sensor evidence、safety decision、approval、resultをtamper-evident storageへ記録する。
- [ ] **AI-generated procedureを本番前に固定化・検証する。** 高リスク用途ではAIが発見した最適手順をdeterministic script/state machineへcompileし、simulation/HIL/independent testsを通す。 [13]
- [ ] **既存安全規格を上書きしない。** ロボット、機械、医療、OTそれぞれの法規・ISO/IEC/ISA要件をMHSとは別に満たす。
- [ ] **変更管理を行う。** MHS driver、reference metadata、model、agent harness、vendor firmwareの更新をすべてsafety-relevant changeとして評価する。
- [ ] **kill switchだけでなくsafe recoveryを設計する。** 停止後にどのstateから再開できるか、sample/device状態が不明な場合にdiscard/resetする基準を定める。

## 総合評価

MHSの研究上の意義はかなり大きい。これまでAIによるrobotics/automationは、モデル能力以前に「機器AのSDK」「機器BのCOM」「機器CのGUI」「研究者しか知らない安全上限」といったintegration debtに阻まれてきた。MHSはそこを、**標準driver + discoverability + AI-readable metadata + shared state + device-enforced limits**という組み合わせで解こうとしている。Anthropicが言う「数週間・数か月の統合を数時間・数分へ」という目標は非常に野心的だが、CMUの8時間integration等、限定された環境では方向性を支持する実証が既にある。 [33]

技術的に特に優れているのは、AIを毎制御stepへ入れることを目的化していない点である。Janeliaはdeterministic loopのdecision pointにAIを置き、QuEraはAIが開発したrecovery procedureをproduction用deterministic scriptへ落とした。**「LLM as controller」ではなく「LLM as scientist/orchestrator/controller synthesizer」**という形の方が、現在のモデル能力とphysical safetyの現実に適している。 [34]

反面、「Standard」という名称を成熟度の意味で読み過ぎてはいけない。2026年8月29日時点の公開情報だけでは、MHSには少なくとも公知の形で、OPC UAのsecurity/audit model、DDSのQoS/security model、ROS/DDSの成熟したcommunication semantics、ISA/IEC 62443のlifecycle securityに相当する規範体系がまだ見えない。 [35]

したがって企業が今MHSを採用する場合の最適戦略は、**既存設備を置き換えることではなく、既存の安全・通信スタックの上にMHSを追加すること**である。具体的には、OPC UA/ROS/DDS/vendor APIをMHS driverで包み、AIには必要最小限のcapabilityしか公開せず、hard real-timeとfunctional safetyはPLC/servo/interlockへ残す。高リスクcommandはhuman approvalを経由し、AI探索の結果はできる限りdeterministic procedureとして固定化してからproductionへ出す。この構成ならMHSの統合生産性を享受しつつ、未成熟な部分を既存産業規格で補える。

MHSが今後本当に業界標準になれるかを左右するのは、Claudeの能力そのものより、**モデル非依存性を保ったまま、誰がdriverの正当性を保証し、誰が安全limitを承認し、誰がcommand権限を付与し、事故時に何を監査でき、異なるvendor実装がどこまで同じ挙動を保証するか**を標準化できるかである。AnthropicはMHSをmodel-agnosticと明記しており、ここを維持しながらopen governanceとconformance ecosystemへ移行できれば、MHSはMCPがソフトウェアtool統合に果たした役割をphysical AIへ広げる有力候補になり得る。現時点では、それは**有望な研究プレビューであって、完成した安全規格ではない**という評価が最も妥当である。 [10]

# 参考資料


**Anthropic / MHS一次資料。** Anthropic, “Previewing the Model Hardware Standard,” 2026年8月27日。MHSの目的、driver、read/write、discovery、自然言語tags、MCP/CLI/API、Genentech・CMU・Janelia・QuEra等の実証、メーカー対応、安全評価計画を含む本調査の中心一次資料である。 36

**MHS公式サイト。** Model Hardware Standard research-preview landing page。公開版が申請制research previewであり、現段階では一般向けの完全な仕様リポジトリではないことを確認する一次資料。 37

**QuEra Computing。** “Holding the Light: Teaching an AI to Lock and Tune Our Quantum Computer's Lasers.” 従来script開発2〜3週間、microsecond servo loopとの役割分離、700 trial中695成功、0.9〜5.4秒/10〜14秒の復旧等を記載する共同実証の一次資料。 16

**Model Context Protocol。** MCP Specification, 2026-07-28。MCPがLLM applicationとexternal data/toolsをつなぐopen protocolであり、JSON-RPC、host/client/server、capability negotiation、tools等を定義している。MHSとMCPの階層差を評価する基準資料。 4

**ROS 2。** Open Robotics ROS 2 Documentation、ROS 2 Security / sros2 documentation。ROS 2がDDS Securityを利用し、authentication、encryption、policy-based access controlを提供できることを確認する一次文書。 26

**OMG / DDS。** Object Management Group / DDS Foundation, Data Distribution Service資料。DDSはdata-centric middleware protocol/APIで、low latency、reliability、QoS等を扱う。DDS Security 1.2はsecurity modelとSPIを定義する。 27

**OPC Foundation。** OPC UA Online Reference。Devices、Industrial Automation、Laboratory & Analytical Device Standard、Machinery、Robotics、Machine Vision、CNC、Laser Systems等の情報モデル群を公開している。 28

**OPC Foundation Security Model。** OPC UA Part 2。authentication/authorizationを含むsecurity architectureおよびclient/server間で追跡可能なsecurity audit trailを規定しており、MHSの今後のsecurity/audit標準化を比較する重要な先行規格である。 38

**ISA/IEC 62443。** International Society of Automation, ISA/IEC 62443 Series of Standards。Industrial Automation and Control Systemsのcybersecurity requirements/processesとsecurity performance assessmentの基準。ISASecureによる製品適合性certification ecosystemも存在する。 39

**NIST。** SP 800-82 Rev.3, Guide to Operational Technology (OT) Security。物理環境へ作用するOTを、securityだけでなくperformance、reliability、safetyを考慮して保護するための基準資料。なおRev.4は2026年1月に改訂作業が開始されたpre-draft段階であり、2026年8月29日時点ではRev.3が完成版である。 40

**経済産業省。** 「AI事業者ガイドライン（第1.2版）」、2026年3月31日取りまとめ。日本におけるAI governanceを検討する際の最新公的資料。 41

**厚生労働省。** 産業用ロボットに係る労働安全衛生規則第150条の4関連通達、および機能安全による機械等の安全確保に関する資料。産業用ロボットのrisk assessment、記録保存、機械安全を日本国内で評価する際の基準資料。 21

**PMDA。**「プログラム医療機器関連通知」「プログラム医療機器」。AI/MHSが研究室設備を越えて診断・治療を意図する機器へ組み込まれる場合に必要となる日本の医療機器規制を確認する一次資料。 [23]

---


[1] [3] [5] [6] [8] [9] [10] [11] [12] [13] [14] [15] [17] [18] [19] [24] [25] [33] [34] [36] Previewing the Model Hardware Standard \ Anthropic
https://www.anthropic.com/news/model-hardware-standard-research-preview

[2] [37] Model Hardware Standard
https://www.modelhardwarestandard.com/

[4] Specification - Model Context Protocol
https://modelcontextprotocol.io/specification/2026-07-28

[7] [20] [40] SP 800-82 Rev. 3, Guide to Operational Technology (OT) ...
https://csrc.nist.gov/pubs/sp/800/82/r3/final?utm_source=chatgpt.com

[16] www.quera.com
https://www.quera.com/blog-posts/holding-the-light-teaching-an-ai-to-lock-and-tune-our-quantum-computers-lasers

[21] ・産業用ロボットに係る労働安全衛生規則第150条の4の施行 ...
https://www.mhlw.go.jp/web/t_doc?dataId=00tb9779&dataType=1&pageNo=1&utm_source=chatgpt.com

[22] [41] AI事業者ガイドライン
https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/20260331_report.html?utm_source=chatgpt.com

[23] プログラム医療機器関連通知
https://www.pmda.go.jp/review-services/drug-reviews/about-reviews/devices/0053.html?utm_source=chatgpt.com

[26] ROS 2 Security
https://docs.ros.org/en/rolling/Developer-Tools/Introspection-and-analysis/About-Security.html?utm_source=chatgpt.com

[27] What is DDS?
https://portals.omg.org/dds/what-is-dds-3/?utm_source=chatgpt.com

[28] OPC UA Online Reference
https://reference.opcfoundation.org/

[29] [30] [38] OPC Unified Architecture - Part 2: Security Model
https://reference.opcfoundation.org/specs/OPC-10000-2/4?utm_source=chatgpt.com

[31] [39] ISA/IEC 62443 Series of Standards
https://www.isa.org/standards-and-publications/isa-standards/isa-iec-62443-series-of-standards?utm_source=chatgpt.com

[32] ISA/IEC 62443 Conformance Certification - ISASecure®
https://www.isa.org/ISASecure?utm_source=chatgpt.com

[35] About the DDS Security Specification Version 1.2
https://www.omg.org/spec/DDS-SECURITY/1.2/About-DDS-SECURITY?utm_source=chatgpt.com