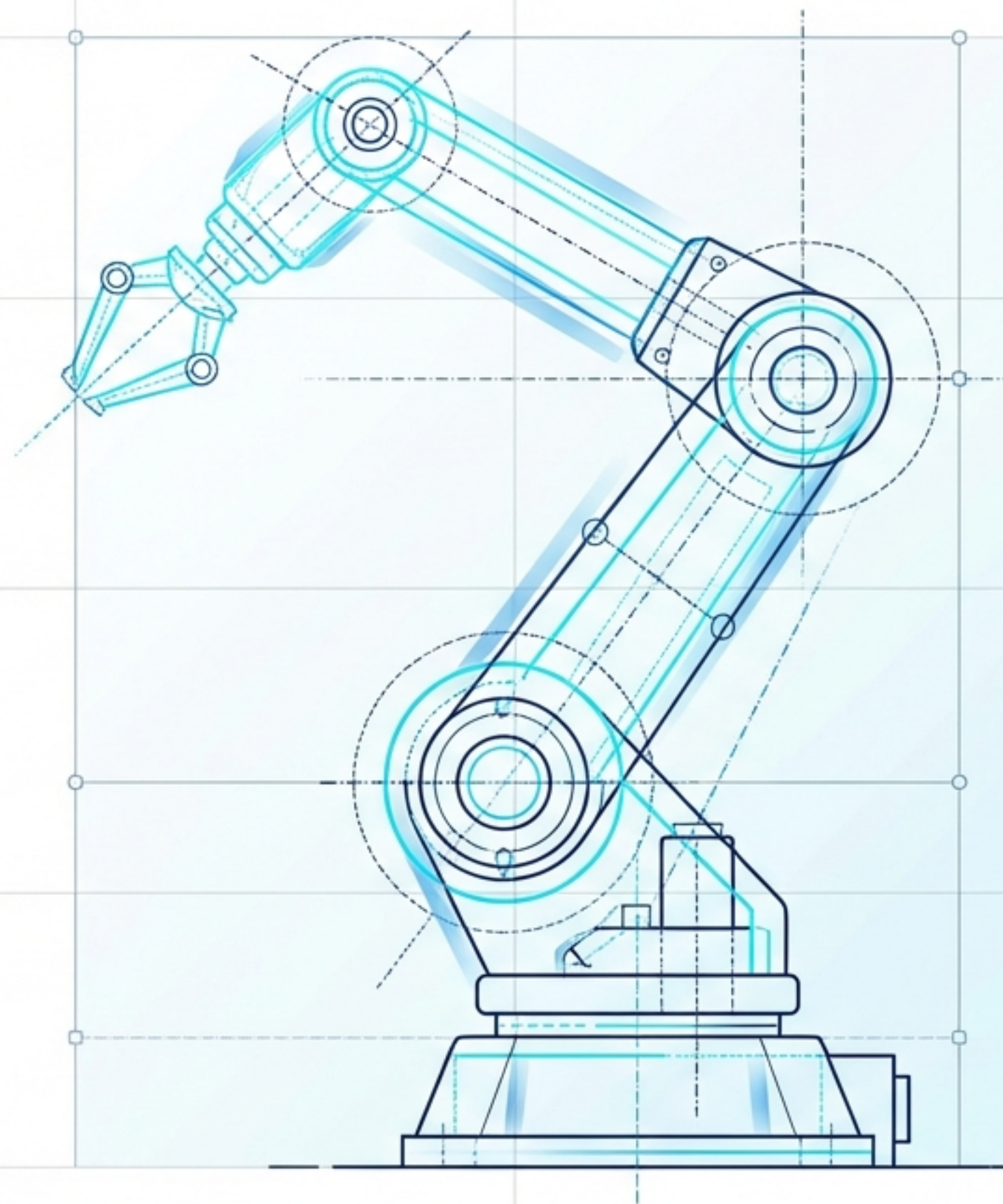


# GEN-1 Architecture & Claims Tear-Down

物理AI基盤モデル「GEN-1」の技術解剖と  
実装に向けたPoC評価フレームワーク

## TARGET:

- Robotics Dev Teams
- Engineering Leaders
- R&D Executives

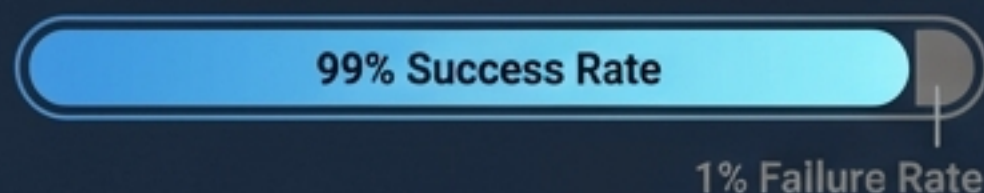


# 「単純な物理タスクにおける熟達 (Mastery)」の提示と実態



Reliability (信頼性)

平均成功率99%

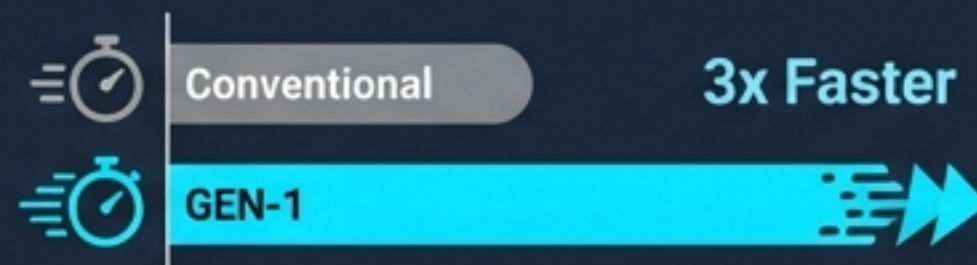


一貫した高いパフォーマンスと安定性を強調。



Speed (速度)

従来比約3倍の実行速度



大幅な効率化による迅速なタスク完了。



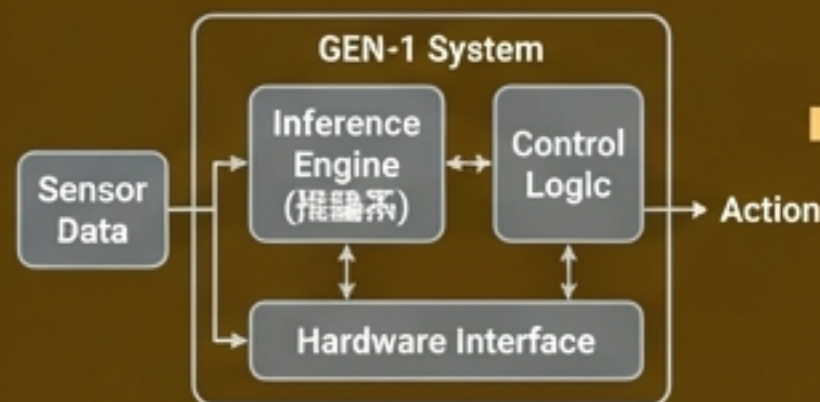
Improvisation (即興性)

各タスク約1時間の実機データで適応



限られたデータからの迅速な学習と柔軟な対応。

## Reality Check: 提示と実態のギャップ



### System vs. Model

GEN-1は単一モデルではなく、推論系を含めた「システム最適化」の産物。  
モデル単体の能力だけでなく、全体システムの統合と最適化が重要。



### Black Box Warning

・パラメータ数・周波数(Hz)・レイテンシなどの重要仕様は未開示。  
外部チームによる厳密な「介入率」「失敗モード」の検証が不可欠。



# Self-Reported Telemetry Dashboard: 連続稼働と適応能力の主張

## Continuous Execution (Stress Tests)

[ 1,800+ ]

回連続: ブロック梱包

[ 200+ ]

回連続: ロボット掃除機整備 / 箱折り

[ 86 ]

回連続: Tシャツ折り

## Speed / Cycle Time (タクトタイム)

12.1 sec



箱組み立て (従来約34秒から大幅短縮)

15.5 sec



スマホ梱包

## Data Engine (学習基盤)

[ 500,000 hrs ]



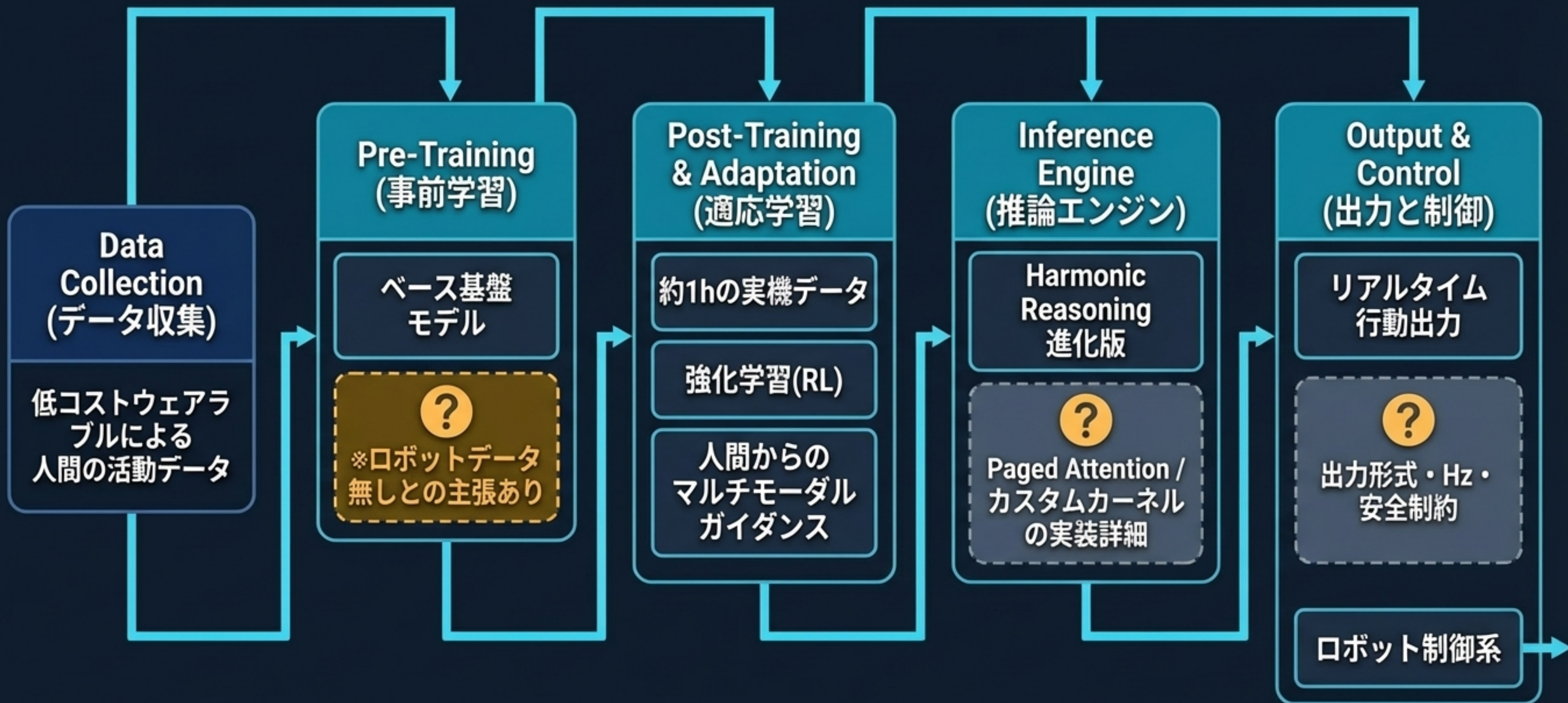
高忠実度な物理相互作用データ

[ 1 hr ]



各タスク適応に必要な実機データ

# GEN-1 System Architecture Blueprint



# Core Engine & The Data Paradox (中核技術とデータ定義の矛盾)

## Harmonic Reasoning (推論パラダイム)

センシングとアクションの非同期・連続時間のトークン流を「同時に考えながら動く」ように処理。巨大モデル化を推論時ガイダンス等に依存せず実現するGEN-0からの独自思想。



## The Pre-training Data Paradox (事前学習データの緊張関係)

### GEN-1 Claims:

「ロボットデータ無しで、低コストウェアラブルを装着した人間の数百万の活動データで事前学習」

### GEN-0 Precedent:

「270,000時間超の実世界マニピュレーション軌跡を含む“ロボティクス事前学習データセット”で事前学習」

**Insight:** 同社の「robot data」の定義が曖昧。比較評価において「何が性能に寄与したか」の特定を困難にする要因。

# The Dark Matter: 未開示のエンジニアリング・ブラックボックス

## Model Scale (モデル規模)

GEN-0では「**10B+までスケール**」と明記されたが、GEN-1のパラメータ数は完全未記載。何B相当かは不明。

## Real-Time Telemetry (リアルタイム仕様)

Paged attentionやカスタムカーネルを強調する一方、推論周波数(Hz)やエンドツーエンドのレイテンシ(遅延)の定量開示がない。

## Modalities (入出力内訳)

「大規模マルチモーダル」と称するが、入力センサー(深度、力覚、触覚など)やアクション出力形式(関節角、トルク、レート等)の内訳が不明。

## Safety & Constraints (安全機構)

行動許可領域、緊急停止、力制限などの具体的なセーフティ設計が未公開。  
「創発的挙動は負債にもなる」と自認しつつもフェイルセーフ仕様は不明。

# The Battlefield: ロボットAI基盤モデル 競合ランドスケープ

Model	Modality	Real-Time Control	Openness	Hardware
<b>GEN-1</b> (Generalist AI)	大規模マルチモーダル -> 行動	Hz/遅延未記載	クローズド (Early Access)	詳細未開示
<b>Isaac GROOT N1.6</b> (NVIDIA)	言語+画像 -> 連続アクション	推論周波数実測提示 (22.8-27.3Hz)	オープン (Apache-2.0)	RTX 4090/5090, H100
<b>π0 / openpi</b> (Physical Intelligence)	画像+テキスト -> モータコマンド	最大50Hz, RTCで 遅延耐性明記	オープン (コード/重み)	複数プラット フォーム
<b>Gemini Robotics</b> (Google DeepMind)	VLM基盤 (Web+ロボ) -> 制御	VLAとしてのE2E制御 (詳細未記載)	限定プレビュー (SDKあり)	オンデバイス~ ハイブリッド
<b>Helix / Helix 02</b> (Figure AI)	Pixels -> 全身単一NN制御	オンボード・ リアルタイム主張	クローズド	未記載

**Key Insight:** オープンソース系 (NVIDIA, Physical Intelligence) がHzや遅延などの「工学仕様」を定量開示し標準化を牽引する一方、GEN-1は「タスク成功率」という結果主導で差別化を図っている。

# Acceleration Timeline: 基盤モデルの急速な進化

Oct 2024

Feb 2025

Nov 2025

Dec 2025

Apr 2026

**π0 (Physical Intelligence)**  
公開

50Hz / RTC /  
Open Source

**Helix (Figure AI)**  
公開

Pixels-to-Actions

**GEN-0 (Generalist AI)**  
公開

10B+ / 270k  
hours data

**Isaac GROOT N1.6 (NVIDIA)**  
公開

3B / Hz  
specifications  
explicitly tested

**GEN-1 (Generalist AI)**  
発表

99% Success /  
500k hours /  
~3x Speed

**Strategic Takeaway:** わずか半年間で「シミュレーション主導」から「実世界でのスケールデータ×高速適応」へのパラダイムシフトが加速。GEN-1は「高成功率×高速×少量適応」を最も強く押し出したマイルストーン。

# The Evaluation Playbook Phase 1: Measure & Fill Metrics

仕様が未開示である以上、公開デモの「成功/失敗」だけでは実運用ギャップは埋まらない。PoCでは以下の指標を自社環境で計測しブラックボックスを埋める。

## Reliability Test (信頼性の解像度)

- 完全達成と部分達成 (Partial credit) の分離
- 連続無介入実行の最長回数と平均/分散
- 失敗モードの分類 (落下、衝突、破損、停止)

## Speed & Latency (速度とリアルタイム性)

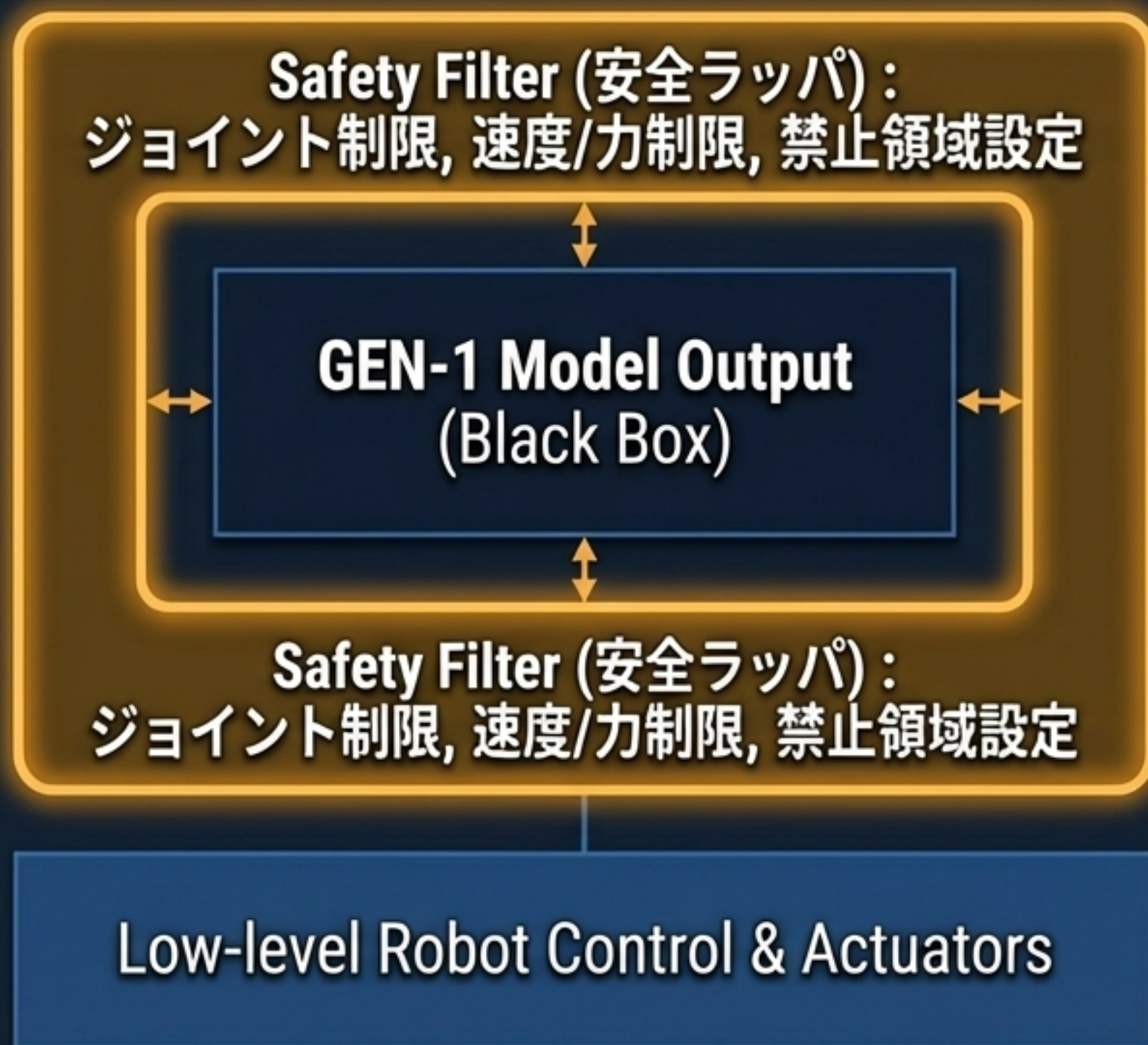
- 厳密なタクトタイム計測 (開始/終了条件の固定)
- 推論のエンドツーエンド (E2E) 遅延
- 制御更新レートとネットワーク遅延感度

## Improvisation & Safety (即興性と安全限界)

- 外乱シナリオ (位置ずれ、変形、滑り) への回復率と時間
- 最大力 / 最大速度 / 接触回数のモニタリング
- 衝突・挟み込みのヒヤリハット率

# The Evaluation Playbook Phase 2: Infrastructure & Risk Mitigation

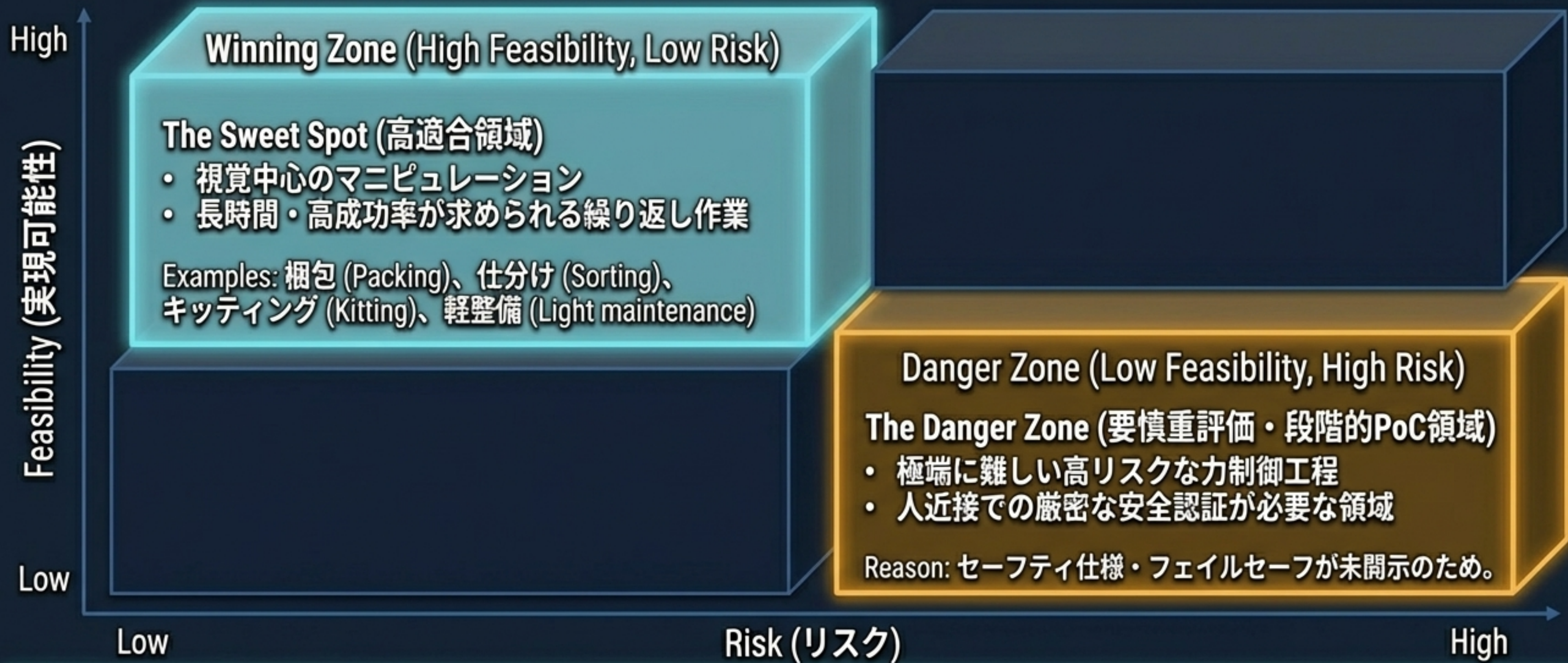
## The Safety Wrapper (フェイルセーフ階層)



## Required Infrastructure

- 計測用GPUサーバ: 推論・ログ収集・再生評価用の計測環境 (GROOT等の公開モデルと同一環境での比較を推奨)
- データ基盤: 生ログ+アノテーション+再生環境 (失敗の切り分け設計)
- 停止・復帰手順: 保護停止条件と監査ログの保存メカニズム
- データガバナンス: E2E学習における現場映像・機密情報の扱いルール設計

# The Verdict: 導入の「勝ち筋」とタイムライン見立て



## The Industry Outlook:

「一般公開・重重み公開」が見えない現状では、早期アクセスパートナー経由での産業PoCが先行する。学術・汎用展開においては、オープンなπ0やGROOTのエコシステムと比較検討することが現実的な戦略となる。