フィジカルAIの新たな地平：Generalist AIによるロボット基盤モデル「GEN-1」の技術的・戦略的包括分析

Gemini 3.1 pro

1. 序論：ロボティクスにおける「ChatGPTモーメント」の到来とフィジカルAIの台頭

人工知能（AI）の進化の歴史は、長らくデジタル空間におけるデータ処理能力の向上と軌を一にしてきた。チェスの世界王者を打ち負かし、タンパク質の立体構造を予測し、人間と遜色のない自然な文章や極めて写実的な映像を生成するに至ったAIの目覚ましい発展は、主にサイバースペース内に留まるものであった。しかし、AIが物理世界に干渉し、人間と同様の身体性を持って現実環境のタスクをこなす「フィジカルAI（Embodied AI：身体性AI）」の領域においては、モラベックのパラドックスが示す通り、極めて高度な推論を要するチェスよりも、シャツを折りたたんだりテーブルを片付けたりする単純な物理的作業の方が遥かに困難なエンジニアリングの壁として立ちはだかってきた[1]。

この長年の技術的障壁に対し、2026年4月2日、米国シリコンバレーに拠点を置く新興企業Generalist AIは、ロボット向けのリアルタイム動作用大規模マルチモーダル基盤モデル「GEN-1」を発表し、ロボティクス分野における極めて重要なマイルストーンを打ち立てた[2]。同モデルは、単に研究室のデモ環境で限定的なタスクをこなすという従来の概念を打ち破り、実際の運用環境において不可欠な「信頼性（Reliability）」「速度（Speed）」「即興性（Improvisation）」を兼ね備えた、汎用物理AIとしての「熟達（Mastery）」の閾値を超えた初のシステムであると定義づけられている[2]。

本稿では、Generalist AIの組織的背景から始まり、前世代モデル「GEN-0」で実証されたロボティクスにおけるスケーリング則の力学、そして「GEN-1」を支える「Harmonic Reasoning（調和的推論）」や「Paged Attention」といった推論アーキテクチャの革新、さらには「Data Hands」と呼ばれる独自のデータ収集パラダイムに至るまで、その技術的優位性と産業的意義を多角的な視点から詳細に分析・考察する。

2. Generalist AIの組織的基盤：フロンティアAIのトップタレント結集と目標駆動型アプローチ

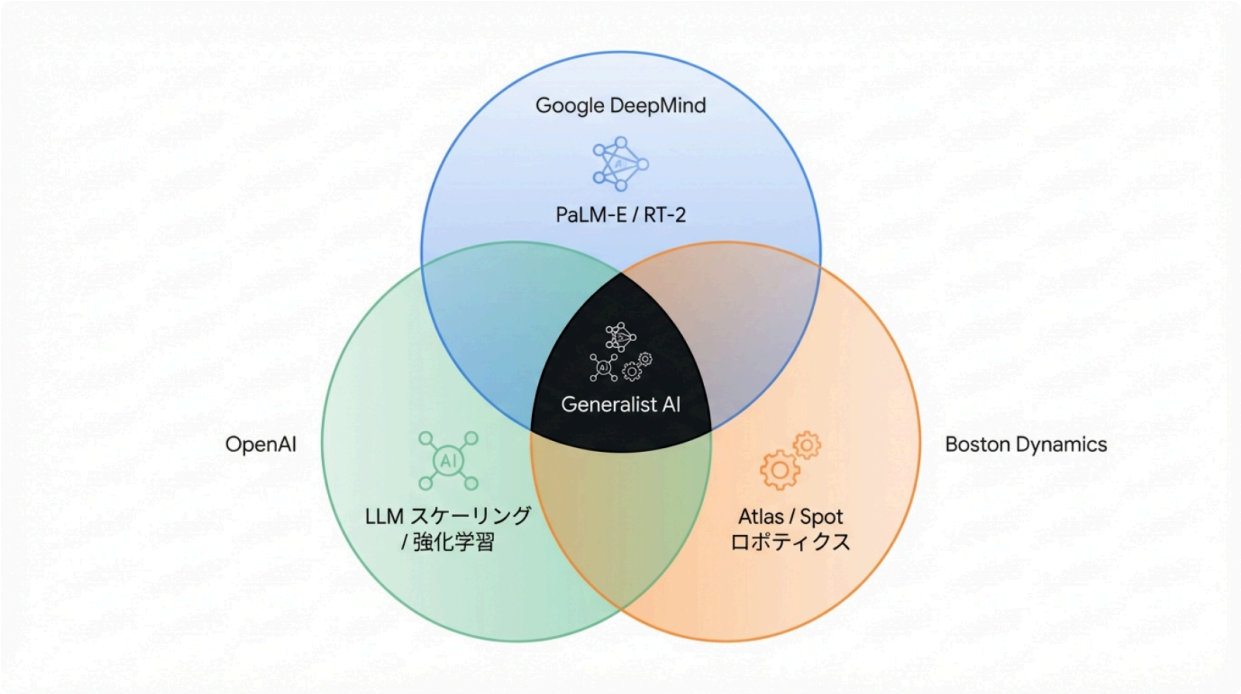
Generalist AIが設立からわずかな期間で「GEN-1」という画期的なブレイクスルーを達成できた背景には、世界のトップAI研究機関から結集した卓越したエンジニアリングチームの存在がある。2024年に設立された同社は、OpenAI、Google DeepMind、Boston Dynamicsといった最先端のフロンティアAIラボの出身者によって構成されており、AIとロボティクスの双方を自社の「DNAの核」として位置づけている[4]。同社は2025年時点で1億4000万ドルの資金調達を行い、評価額は4億4000万ドルに達しており、Spark Capital、NVIDIA、Bezos Expeditionsといった世界的なトップ投資家からの支援

を受けている[4]。

同社のCEOであるPete Florenceは、かつてGoogle DeepMindにおいて上級研究員を務め、「PaLM-E」や「RT-2」といった視覚・言語・行動モデル（VLA: Vision-Language-Action Models）の開発を主導した人物である[7]。また、共同創業者でありChief Scientistを務めるAndy Zengは、プリンストン大学で博士号を取得後、Google DeepMindにおいてFlorenceと共に「PaLM-E」プロジェクトに携わったほか、「TossingBot」や「Code as Policies」など、残差物理学（Residual Physics）や言語モデルを用いたロボット制御の研究で多数の最高論文賞を受賞した実績を持つ[7]。さらに、暗黙的行動クローニング（Implicit Behavioral Cloning）や交差身体性逆強化学習（XIRL）といった、接触の多いタスクにおける政策学習や異種ロボット間でのデータ転移技術においても、彼らは先駆的な研究成果を残してきた[8]。

加えて、Generalist AIのチームには、ChatGPTやGPT-4を数億人のユーザー規模へとスケーリングさせたOpenAIの元エンジニアたちや、Atlas、Spot、Stretchといった次世代ロボットの極限的なハードウェア制御（パルクールから精密操作まで）を担ってきたBoston Dynamicsの出身者が多数参画している[4]。近年、Ilya Sutskever（Safe Superintelligence創業者）やJohn Schulman（Anthropicへ移籍後退社）、Mira Murati（Thinking Machines Lab創業者）など、OpenAIのコアメンバーがAIの安全性や新たなアプローチを求めて独立・移籍する動きが加速しており、Generalist AIもこうした高度な技術人材の流動化の恩恵を受け、ソフトウェアとハードウェアの最高峰の知見を融合させることに成功している[9]。

Generalist AIを支える技術的系譜と創業者DNA



Generalist AIの創業チームは、Google DeepMind（VLAモデルの先駆者）、OpenAI（LLMのスケーリング則と強化学習）、Boston Dynamics（極限のロボットハードウェア制御）の知見を統合し、汎用物理AIの開発を推進している。

このような強力な人材基盤を持つGeneralist AIの研究開発における最大の特徴は、徹底した「目標駆動型（Goal-driven）」の哲学にある。学術界や産業界においては、「世界モデル（World Models）」や「VLA」といった特定のアーキテクチャやパラダイムの流行（バンドワゴン）に乗る「アイデア駆動型（Idea-driven）」の研究が横行しやすい[13]。しかし同社は、自らがVLAの共同発明者であり、長年世界モデルの研究を行ってきたにもかかわらず、自社のモデルに意図的に特定のラベルを貼ることを避けている[13]。彼らの究極の目標は、タスク固有のデータなしに未知のタスクを高速実行する「完全なゼロショット・ロボティクス」の実現と、実用的な商業展開のための「1時間のロボットデータで99%以上の成功率を達成すること」に明確に設定されており、その目標達成の障害となる問題を解決するためであれば、どのような手段やアーキテクチャの変更も厭わないというアジャイルな姿勢を貫いている[13]。この哲学こそが、既存の視覚・言語モデルにロボットの動作を安易に後付けするのではなく、パラメータの約99%を物理相互作用データからゼロから（from scratch）学習させるという、極めて計算資源を要するが根本的な制御が可能なアプローチをGEN-1において選択させた最大の要因である[3]。

3. 熟達（Mastery）の再定義：GEN-1を特徴づける3つのパフォーマンス次元

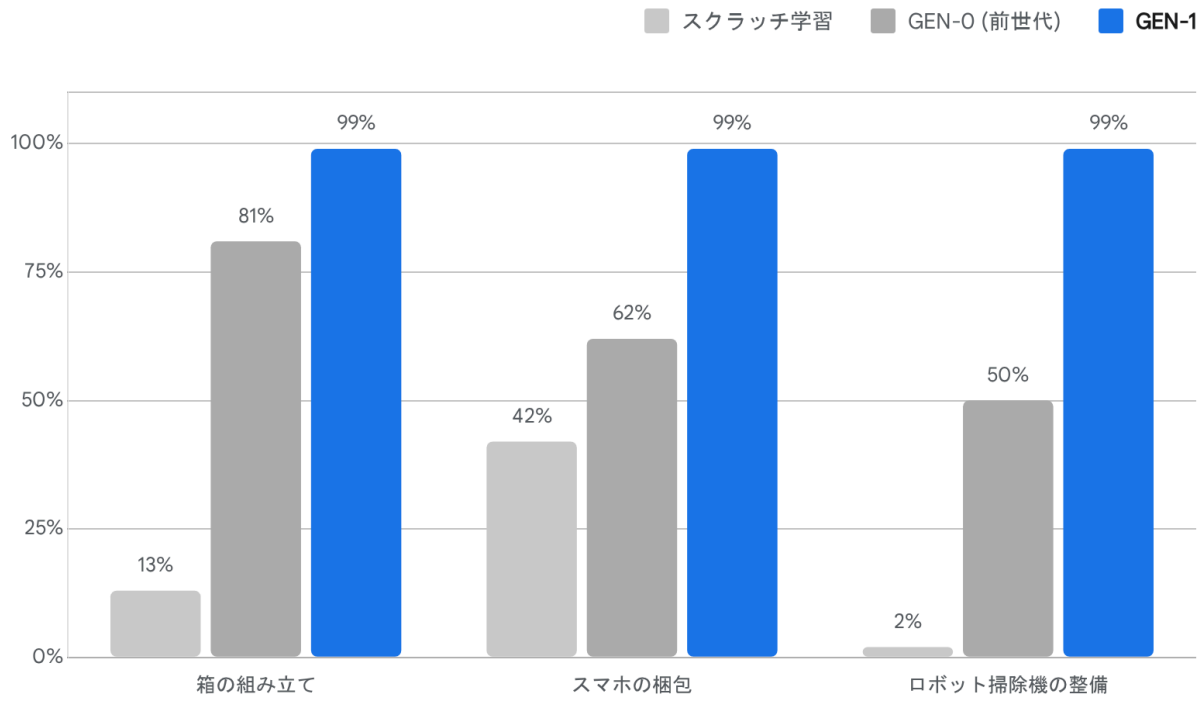
2026年4月に発表された「GEN-1」は、ロボットの評価基準を根本的に書き換えるものであった。従

来、ロボティクスの論文やデモンストレーションにおいては「いかに多様なタスクができるか（What it can do）」という点が強調される傾向にあったが、GEN-1はこれを「実際の運用環境において、どれほど安定して、速く、柔軟に実行できるか」という基準へと引き上げた[2]。Generalist AIはこれを「熟達（Mastery）」と定義し、「信頼性（Reliability）」「速度（Speed）」「即興性（Improvisation）」という3つの主要な指標の交差点として、その性能を実証している[3]。

3.1. 信頼性（**Reliability**）：商業化の閾値を超える**99%**のタスク成功率

産業用オートメーションや物流の現場において、ロボットの導入を阻んできた最大の要因は「失敗率の高さ」である。100回に1回の失敗であっても、高速で稼働する製造ラインや物流拠点においてはシステム全体の致命的な遅延や停止を招くため、99%以上の成功率の達成は商業化に向けた絶対的な閾値とされてきた[14]。GEN-1は、シンプルな物理作業において平均成功率99%という、まさに「生産レベル（Production-level）」の信頼性を達成した[2]。

これは、事前学習を行わないスクラッチ学習モデル（平均成功率19%）や、わずか5ヶ月前に発表された前世代の基盤モデルGEN-0（平均成功率64%）からの劇的な飛躍である[3]。長時間の自律連続実行（人間の介入なし）を測定するストレステストにおいても、GEN-1は驚異的な耐久性と精度を示している。具体的には、ブロックの梱包タスクで1,800回の連続成功、ロボット掃除機の整備タスクで200回以上の連続成功（GEN-0の50%から99%へ向上）、箱の組み立てタスクで200回の連続成功（GEN-0の81%から99%へ向上）、スマートフォンの梱包タスクで100回以上の連続成功（GEN-0の62%から99%へ向上）、そしてTシャツの折りたたみタスクで86回の連続成功を記録している[3]。自動車部品のキッティング作業においても、1時間以上にわたって人間の介入なしに自律稼働を続けることが実証されている[3]。



GEN-1は、箱折りやロボット掃除機の整備といった複雑な物理タスクにおいて、前世代のGEN-0（平均64%）やスクラッチ学習モデルを大きく上回り、商業的実用性の閾値とされる99%の成功率を達成した。

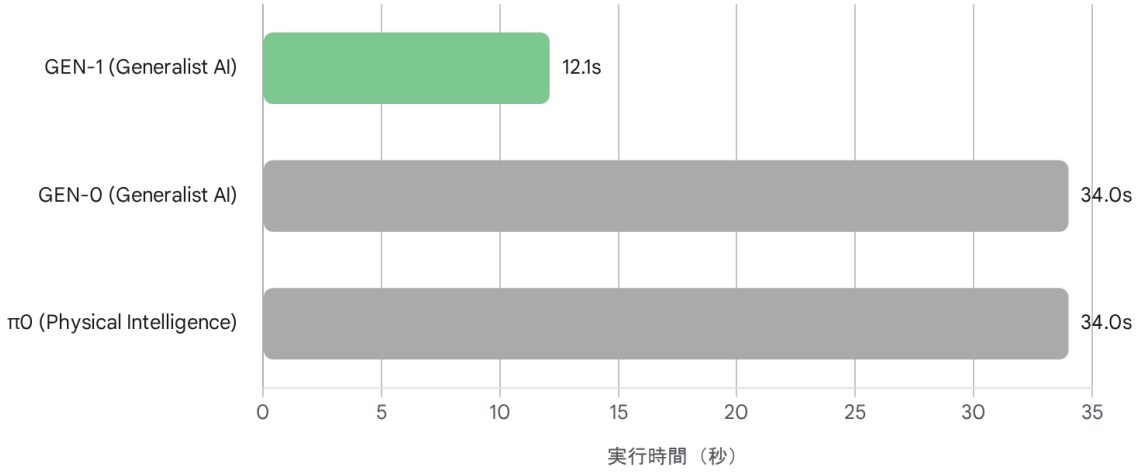
Data sources: Generalist AI

3.2. 速度（Speed）：SOTAを大幅に凌駕するタスク実行効率

ロボットが人間の代替として商業的に成立するためには、単にタスクを成功させるだけでなく、それを経済的に意味のある速度で実行できなければならない。GEN-1は、タスクの完了時間を当時の最高水準（State-of-the-Art: SOTA）と比較して約3倍（正確には約2.8倍）に高速化することに成功している[2]。

ベンチマークとして公開された段ボール箱の組み立て（箱折り）タスクにおいて、前世代のGEN-0や、競合であるPhysical Intelligence社の汎用モデル「π0（pi-zero）」がいずれも約34秒の時間を要していたのに対し、GEN-1はわずか12.1秒でこの作業を完了させる[3]。また、スマートフォンをケースに梱包するタスクも15.5秒という驚異的な速度で完了する[3]。この大幅な速度向上は、モーターの出力を物理的に引き上げた結果ではなく、後述する推論インフラの最適化（低レイテンシ化）と、事前学習による対象物の物理的挙動に対する深い理解がもたらした「経路計画における迷いのなさ」によるものである[14]。人間のデモンストレーション速度を上回るペースでタスクを実行し、その速度域においても未知の物理現象に対してリアルタイムに反応できる能力は、GEN-1のアーキテクチャの堅牢性を示している[3]。

タスク実行速度におけるSOTAの凌駕（箱の組み立て時間）



GEN-1は、箱の組み立てタスクにおいて約12.1秒という記録を達成し、競合モデル（π0）や前世代モデル（GEN-0）の約34秒と比較して、約2.8倍の実行速度の向上を実現した。

データソース: Generalist AI, Let's Data Science

3.4. 即興性（Improvisation）と物理的常識の創発

GEN-1が従来のプログラムによって厳密に制御されたロボットや、単純な模倣学習（Imitation Learning）モデルと決定的に異なるのは、「物理的常識（Physical Commonsense）」に基づく高度な即興的なエラー回復（リカバリ）能力を備えている点である[3]。

工場や家庭といった現実空間は非構造的であり、常に対象物の位置がミリ単位でずれたり、布やビニールのような変形しやすい物体が想定外の形状になったりする不確実性（ロングテールイベント）に満ちている。従来のロボティクスでは、これらすべてのエラーパターンを想定し、事前にリカバリ動作をスクリプトとしてコーディングしようと試みてきたが、無限のバリエーションに対処することは不可能であった[14]。しかし、GEN-1は台本から外れた事態に直面した際、自律的かつ創造的な解決策をその場で編み出すことができる[3]。

テスト環境において観察された具体的なリカバリ行動の例は以下の通りである。自動車用の小さなワッシャーを適切な角度で把持できなかった場合、ロボットは一度それを台の上に置き、もう片方の手を使って正しい角度で持ち直す「両手を用いた持ち直し（Bimanual in-hand regrasping）」を自発的に行った[14]。また、物体を特定の隙間に挿入する際、対象物の向きが合わない場合には、周囲のスリットや環境の物理的形状をテコのように利用して姿勢を正す「外部器用さ（Extrinsic dexterity）」

を披露した[3]。さらに、ぬいぐるみがビニール袋の入り口に引っかかった際には、人間が日常的に行うように「袋を軽く振って（Shaking a bag）重力で底に落とし込む」という、直感的で高度な物理ダイナミクスの理解を伴う動作を実行している[3]。

重要なのは、エンジニアがこれらの回復動作を個別の条件分岐として事前にプログラムしたわけではないという点である[14]。膨大な物理相互作用データの事前学習を通じて、モデル自身が「物体がどう振る舞うか」という物理的な直感を獲得し、その結果として有用な行動が「自発的に創発（Emergent）」したのである[14]。この物理的常識の獲得こそが、GEN-1を真のフィジカルAIたらしめている核心である。

4. データ戦略のパラダイムシフト：「Data Hands」による物理データの獲得

LLMがインターネット上に存在する数兆ものテキストデータを貪欲に学習することで飛躍的な進化を遂げたのに対し、ロボティクス分野は深刻な「データのボトルネック」に苦しんできた[5]。物理空間の複雑な操作を伴う「筋肉の記憶（Muscle memory）」に相当するようなデータは、ウェブ上には存在しないためである[14]。

従来のロボット開発において、この問題を解決するための主流なアプローチは、人間がVRゴーグルや専用のジョイスティックを用いてロボットアームを遠隔操作（テレオペレーション）し、その軌道を記録して学習させることであった[15]。しかし、テレオペレーションによるデータ収集は、ロボットの動作速度の限界やシステム遅延（レイテンシ）、そして何より人間が物体を触った際の微細な力覚（フォースフィードバック）の欠如により、生成されるデータ自体が「不自然で、遅く、不器用」なものになりがちであるという構造的な欠陥を抱えていた[15]。

4.1. ウェアラブルデバイス「Data Hands」の導入と人間の活動データの収集

このボトルネックを根本的に打破するため、Generalist AIは「Data Hands（データ・ハンズ）」と呼ばれる独自の低コスト・ウェアラブルデバイスを開発し、データ収集のパラダイムを完全に転換した[5]。これは、人間の手首や手に装着するピンサー（ハサミ）型のデバイスであり、装着者の手を事実上「ロボットのグリッパー」へと変換するものである[5]。

同社はこの「Data Hands」をグローバルに配布し、人間の作業員が日常的なタスク（掃除、調理、部品の組み立て、洗濯など）を行う際の視覚的変化、微妙な力加減、滑りの感覚、そして失敗から立ち直る修正動作に至るまで、すべての微細な物理相互作用データを収集した[5]。この手法により、ロボットのテレオペレーションというボトルネックを迂回し、人間本来の滑らかで高速なダイナミクスをそのままデータ化することに成功したのである。

4.2. ロボットデータを用いない事前学習と驚異的な適応効率

このデータ収集戦略の結果、Generalist AIは、家庭、ランドリー、倉庫、工場など様々な環境から収集された**50**万時間（ペタバイト級）を超える高忠実度な物理相互作用データの巨大なコーパスを構築した[3]。

ここで特筆すべきは、GEN-1のベースとなる基盤モデルの事前学習（Pretraining）において、実際の

ロボット由来のデータが一切使用されていないという点である[3]。モデルは、人間が「Data Hands」を通じて生成した膨大な活動データのみから、重力、摩擦、変形、慣性といった物理世界のダイナミクスの根底にある法則と、マルチタスクの処理能力を純粋に学習する[3]。いわば、特定の身体（ロボットハードウェア）に宿る前に、物理法則に対する直感的な理解（物理的常識）を構築してしまうのである。

そして、この強固な事前学習の土台があるため、GEN-1を特定のロボットハードウェア（新しいエンボディメント）や全く未知のタスクに適応（Adaptation）させるプロセスは劇的に効率化される。新たなタスクの習得に必要なロボット固有のファインチューニングデータは、わずか約**1**時間分で十分である[2]。学習後の強化学習（Reinforcement Learning from Experience）やマルチモーダルな人間のガイダンスを組み合わせることで、前世代のGEN-0と比較して10分の1以下のタスク固有データとファインチューニングステップで、99%の成功率という熟達の域に到達することが可能となった[3]。

5. 推論アーキテクチャの革新：Harmonic Reasoningと Paged Attentionの統合

高い成功率とSOTAの3倍という圧倒的な実行速度を支えているのは、質の高いデータだけでなく、ソフトウェアインフラおよび推論アーキテクチャにおけるシステムレベルの大規模な再設計である[3]。ロボットの制御においては、生成AIの常識とは異なる極めて厳しい制約が存在する。

5.1. Harmonic Reasoning（調和的推論）による認知と行動のシームレスな統合

テキストベースの大規模言語モデル（LLM）やチャットボットにおいては、モデルが複雑な問題に対して「考える（Thinking）」ための時間を長く取る（例えばChain-of-Thoughtプロンプティングによって中間推論ステップを生成する）ことは比較的容易である[18]。しかし、現実世界の物理システムにおいては、モデルが思考している間も物体は落下し続け、環境は変化し続ける。すなわち「物理法則は止まってくれない（physics doesn't stop）」のである[18]。

従来のロボット制御アーキテクチャの多くは、高次の認知や計画を担う「System 2（遅い思考・熟慮的）」と、低次の反射的なモーター制御を担う「System 1（速い思考・直感的）」を分離して設計し、それらを統合しようとしてきた[7]。しかし、この階層化は推論時のオーバーヘッドを生み、複雑な環境での高速な反応を阻害する要因となっていた。

これに対し、GEN-1（およびその前身であるGEN-0）の中核技術として導入された**「Harmonic Reasoning（調和的推論）」**は、この構造的な分離を排除する全く新しいアプローチである[18]。Harmonic Reasoningは、センサーからの入力（知覚トークン）ストリームと、アクションを出力する（行動トークン）ストリームを、明示的な同期ポイントを設けることなく、非同期の連続時間ドメインで「調和的」に相互作用させる訓練手法である[7]。

このアーキテクチャにより、モデルは「考えること」と「行動すること」を同時にシームレスに並行処理できるようになる[18]。結果として、System 1とSystem 2の分離や、推論時の重いガイダンスコントローラーへの依存から脱却し、モデルサイズを極大化（10Bパラメータ以上）させても、低レイテンシ（遅延）でのリアルタイム制御（例えば100Hzの制御ループ）を維持することが可能となった[7]。例えば「カ

メラのキットを組み立てる」といった、段ボールを折り、ビニール袋からカメラを取り出し、小さなフラップを差し込むという複雑で長期的なタスク（Long-horizon tasks）であっても、サブタスクの明示的な概念を保持することなく、単一のHarmonic Reasoningのストリーム内で完結させることができる [18]。

5.2. ロボティクス推論における**Paged Attention**の実装

さらに、GEN-1のリアルタイム推論をハードウェア・システムレベルで決定的に支えているのが、ロボティクスの制御ループ向けに新しく発明された**「Paged Attention」**の実装である [3]。

Paged Attentionは、元来vLLMなどの大規模言語モデルの推論フレームワークにおいて、GPUメモリ上のKVキャッシュ（Key-Value Cache）の断片化という深刻な問題を解決するために考案された技術である [21]。従来の自己注意力機構では、シーケンスが長くなるにつれて連続した広大なメモリ領域が必要となり、メモリの浪費とパフォーマンスの低下を招いていた。Paged Attentionは、OSの仮想メモリページングの概念を応用し、KVキャッシュを非連続なメモリブロック（ページ）に分割して管理することで、メモリの利用効率とスループットを飛躍的に向上させる [21]。

Generalist AIのエンジニアリングチームは、このPaged Attentionのメカニズムをカスタマイズし、ロボットの自己回帰的な行動生成ループに組み込んだ [15]。連続的なセンサー入力と行動出力が繰り返されるロボット制御において、過去の知覚トークンや行動トークンのKVを毎ステップ再計算することなく効率的にキャッシュして再利用することで、推論レイテンシを極小化している [16]。この革新的なアテンション機構と、学習の安定性を高めるカスタムコンピュートカーネルの開発により、複雑で長期にわたる身体的推論（Long-horizon embodied reasoning）を、極めて低い遅延でロボットの制御ループ内に収めることが可能となり、SOTAの3倍というタスク実行速度が実現されたのである [16]。

6. ロボティクスにおけるスケーリング則の証明（**GEN-0**から**GEN-1**への進化）

AIの領域、特にLLMの発展を牽引してきたのは「データ量と計算資源（パラメータサイズ）を指数関数的に増やせば、モデルの性能は予測可能な形で向上する」という強力な**スケーリング則（Scaling Laws）**の存在である。Generalist AIの開発史において特筆すべきは、このスケーリング則が、テキスト空間だけでなくロボティクスという物理空間においても明確に成立することを証明した点である [18]。

6.1. パラメータサイズの拡大とインテリジェンスの相転移

前世代モデルであるGEN-0の開発過程において、研究チームはモデルのパラメータサイズを拡張していく中で、能力が飛躍的に向上する明確な「相転移（Phase transition）」を観測している [7]。

- **1B（10億）パラメータ規模**：このサイズのモデルは、多様な環境からの複雑な感覚運動データを吸収しきれず、パフォーマンスの向上が頭打ちになる「硬直化（Ossification）」という現象を示す [7]。
- **6B（60億）パラメータ規模**：この規模に達すると、単一のタスクだけでなく、複数のタスクを同時にこなす強力なマルチタスク能力が明確に発現し始める [7]。
- **7B（70億）パラメータ規模以上（インテリジェンスの閾値）**：7Bを超えるとモデルは質的な変化を遂げ、大規模なロボット事前学習データを極めて効果的に内面化する。これにより、全く新しい

下流タスクに対しても、わずか数千ステップの事後学習（Post-training）で適応できる強力な汎化能力を獲得する[7]。

Generalist AIはその後、GEN-0のアーキテクチャを**10B**（**100**億）パラメータ以上のサイズへとスケーリングさせることに成功している。サイズが大きくなるほど、新たなタスクへの適応に必要な事後学習のデータ量とステップ数が減少するという、LLMと同様の現象が物理AIにおいても確認された[18]。

6.2. 予測可能な性能向上と交差身体性

スケーリング則のもう一つの重要な側面は、性能の予測可能性である。GEN-0の研究において、事前学習のデータセットサイズ（D）に対する、下流タスクの検証予測誤差（L）は、べき乗則 $L(D) =$ に正確に従うことが示された[18]。これは、ロボットの開発において「商業的に要求される性能（例：特定のピッキング作業での99.9%の成功率）に到達するためには、あとどれだけのデータと計算資源を投下すればよいか」を、勘や経験ではなく、科学的・数学的に予測・コントロールできるようになったことを意味する[18]。GEN-1の開発は、まさにこの予測可能なスケーリングの軌道上に位置しており、ペタバイト級の物理データを処理するために再設計された分散学習インフラがそれを支えている[3]。

さらに、これらの基盤モデルは設計段階から**「交差身体性（Cross-Embodiment）」**を前提としている。すなわち、単一の事前学習を済ませた同一のアーキテクチャが、6自由度（6DoF）の単純なアーム、7DoFの柔軟なアーム、そして16DoF以上の複雑な半ヒューマノイドシステムなど、全く構造の異なる異種ロボットのフリート全体を制御できることが実証されている[7]。これにより、ハードウェアの変更ごとにソフトウェアをゼロから構築し直す必要がなくなり、汎用ロボットの普及に向けたコストと時間が劇的に削減される。

7. 競合ランドスケープ：他社基盤モデルとの技術的比較分析

フィジカルAIおよび汎用ロボット基盤モデルの開発領域では、世界のトップ企業が異なるアプローチで熾烈な競争を繰り広げている。Generalist AIのGEN-1の立ち位置と技術的優位性をより明確にするため、主要な競合モデルであるGoogle DeepMindの「RT-2 / Gemini Robotics」、Physical Intelligenceの「pi0（π0）」、そしてNVIDIAの「Isaac GR00T N1」などとの比較を行う。

7.1. 各アプローチの思想的差異

現在のロボット基盤モデルの開発アプローチは、大きく分けて二つの潮流が存在する。一つは、インターネット上の膨大なテキストや画像で学習済みの既存のAIモデル（LLMやVLM）の強力な意味理解能力を基盤とし、そこにロボットの行動出力を接ぎ木（ファインチューニング）していくアプローチである。Google DeepMindの**RT-2**や最新の**Gemini Robotics**がこの代表例である[14]。この手法は、ロボットが「青い恐竜のぬいぐるみを拾って」といった複雑な自然言語の指示や抽象的な概念を理解し、ゼロショットで実行計画を立てる能力において極めて高いパフォーマンスを発揮する[14]。

これに対し、Generalist AIの**GEN-1**は、全く逆のアプローチをとっている。既存の言語モデルに依存せず、モデルパラメータの約99%を、純粋な物理相互作用データからゼロから（From scratch）学習させるのである[13]。これは、デジタル世界での言語や画像の学習だけでは、重力や摩擦といった「物

理的常識」を本質的に獲得することはできないという強力な思想に基づいている[13]。

また、Physical Intelligence社の**pi0（π0）**は、事前学習済みのVLM（PaliGemmaなど）をベースにしつつも、出力層にフロー・マッチング（拡散モデルに類似した技術）を採用し、滑らかで連続的な低レベルのモーターコマンドを直接出力することに特化している[1]。pi0は洗濯物の折りたたみやコーヒーの抽出など、複雑な長尺タスクにおいて優れたゼロショット汎化性能を示している[27]。しかし、pi0のデータ収集基盤は依然として大規模なロボットテレオペレーションに強く依存しており、データ生成のコストとスケーラビリティにおいて課題を残している[5]。この点において、ウェアラブルデバイス「Data Hands」による人間の活動データの直接収集という手法を確立したGEN-1は、データエンジンの拡張性で優位に立っている[5]。

7.2. アーキテクチャとパフォーマンスの比較

以下の表は、GEN-1と主要な競合モデルの技術的特性および公開されているパフォーマンス指標の比較である。

比較項目	**Generalist AI GEN-1**	**Google RT-2 / Gemini Robotics**	**Physical Intelligence pi0**	**NVIDIA Isaac GR00T N1**
アーキテクチャの起点	物理相互作用特化（ゼロから学習）	大規模視覚言語モデル（VLM）ベース	VLM + フロー・マッチング	ヒューマノイド特化基盤モデル
推論機構の革新	Harmonic Reasoning, Paged Attention	言語的・空間的推論（Chain-of-Thought）	拡散ベースの連続アクション生成	統合的なシミュレーションと強化学習
データ収集・学習基盤	Data Hands (人間ウェアラブルデータ)	インターネットデータ + ロボット実機データ	大規模ロボットテレオペレーション	Isaac Simによる大規模合成データと検証
実行速度（例：箱組立）	約**12.1**秒	（同等タスクの直接比較データなし）	約34.0秒	（同等タスクの直接比較データなし）

タスク成功率の指標	平均**99%**	タスクによるが一般に汎化能力を重視	未公開（汎用性と滑らかさを重視）	エコシステム全体での適応力を重視
主な技術的強み	生産レベルの信頼性、超高速実行、即興的な物理的リカバリ	圧倒的な意味理解、コード生成による即座の能力拡張	スムーズな連続軌道、強力なゼロショット多段タスク汎化	デジタルツインでの検証、NVIDIAハードウェアとの親和性

表が示す通り、GEN-1の最大の差別化要因は、実運用に直結する「実行速度の圧倒的な速さ」と「99%という高精度の反復信頼性」、そして「物理空間特有のノイズに対する即興的な回復力」の3点に集約される。高度な抽象的推論においてはGemini Roboticsが優位性を持つ可能性があるが、工場の組み立てラインや物流倉庫といった、ミリ秒単位の速度と絶対に失敗できない信頼性が求められる「現場の物理作業」においては、GEN-1のアーキテクチャが現状最も商業化の要求水準に近い位置にあると言える[15]。

8. 商業的展望と「身体性のアライメント」という新たな課題

GEN-1が実証した「99%の信頼性」と「1時間のデータでの適応力」は、これまで特注の自動化システム（Bespoke automation）を設計するか、人手に頼らざるを得なかった多品種少量生産の現場に、大きな変革をもたらす。複雑なプログラミングなしに、汎用ロボットが物流倉庫でのピッキング、軽工業での組み立て、アパレル産業での衣類の折りたたみ、さらには機器の保守点検といった多岐にわたるタスクを担うことが可能になりつつある[16]。

事実、Generalist AIはすでに製造業やロジスティクス企業を対象としたアーリーアクセス（早期アクセス）プログラムを通じてGEN-1の提供を開始しており、AWSが主導するPhysical AI Fellowshipなど、エコシステム全体を巻き込んだパートナーシップの構築を進めている[15]。

8.1. 創発的な振る舞いがもたらす「リスク」とアライメントの再定義

一方で、フィジカルAIの商業的・社会的な本格実装に向けては、デジタル空間のAIとは異なる特有の重大な課題が残されている。それは、ロボットの「アライメント（人間の価値観や意図との調整）」問題である[3]。

LLMにおけるアライメントは、主に「有害な文章や偏見を含む回答を出力させないこと」に主眼が置かれてきた。しかし、質量と運動エネルギーを持つロボットの場合、アライメントの失敗は直接的な物理的破壊や人命への危害（Liability）に直結する。GEN-1の最大の技術的強みである「即興的な物理的常識（台本にない回復行動を自発的に編み出す能力）」は、管理された工場や人間のいる協働環境においては、時に「予測不可能な動き（Emergent behaviors）」として極めて危険な要素となり得る[3]。例えば、ロボットが重い部品を取り落としそうになった際、即興的にそれを空中で掴み直そうと

する反射行動は、モデルとしては「優秀なリカバリ」であっても、その軌道上に人間の作業員の手があれば重大な事故を引き起こす。

Generalist AIのチームもこの問題を深く認識しており、「身体性知能のためのアライメント」の再定義を急務としている[3]。ロボットのタスクにおける「成功」は厳密にユーザーによって定義されるものであり、ロボットが「何をすべきか」だけでなく「何をしてはいけないか」という境界を明確に設定しなければならない。そのために、推論時のポリシー・ステアリング（Inference-time policy steering）や強化学習の報酬設計を通じて、モデルの創発的な即興性を活かしつつも、安全限界を決して超えないようにロボットの行動を厳密に制御（ハード化）する技術の確立が、今後の社会実装における最大の焦点となる[3]。

9. 結論

Generalist AIによる「GEN-1」の発表は、ロボティクスAIが長年停滞していた「研究所のデモンストレーション」という枠組みから脱却し、「予測可能なスケーリングと商業レベルの信頼性」という実運用の領域へと決定的に足を踏み入れた歴史的な転換点である。

本分析から明らかになったGEN-1の革新性は、大きく3点に集約される。第一に、「Data Hands」を用いたウェアラブルデバイスによるデータ収集パラダイムの転換であり、テレオペレーションのボトルネックを迂回して50万時間に及ぶ人間の物理相互作用データを基盤化したことである。第二に、「Harmonic Reasoning」とロボティクス向け「Paged Attention」の実装により、System 1/2の分離という旧来のアーキテクチャを打破し、思考と行動の連続的かつ極低レイテンシでの統合を実現したことである。第三に、これらのイノベーションとスケーリング則の証明により、産業界が求める「99%の信頼性」「従来の3倍の実行速度」、そして「物理的常識に基づく即興的なエラー回復」という、真の熟達（Mastery）を達成したことである。

Google DeepMindやOpenAI、Boston Dynamicsといったトップ機関でAIとロボティクスの最前線を牽引してきたチームが、「VLA」や「世界モデル」といった既存のラベルや流行に囚われず、「完全なゼロショット・ロボティクスの実現」という目標から逆算してシステム全体をゼロから再設計したことが、この劇的なブレイクスルーを生み出した。

GEN-1は現状のすべての物理作業を完璧にこなせるわけではなく、予測不可能な創発的行動をいかに安全に制御するかという身体性のアライメント課題も残されている。しかし、「データと計算資源を投下することで、物理空間におけるAIの性能が予測通りに向上する」という事実が証明されたことの意義は計り知れない。データ基盤がさらに拡大し、安全性のためのハードウェア制御技術が成熟するにつれて、製造や物流の現場、そして最終的には私たちの家庭環境において、人間と汎用ロボットがシームレスに協働する「フィジカルAIの時代」が本格的に幕を開けることになるだろう。

引用文献

1. π 0 : Our First Generalist Policy - Physical Intelligence, 4月 13, 2026にアクセス、https://www.pi.website/blog/pi0
2. GEN-1 の概要 | npaka - note, 4月 13, 2026にアクセス、https://note.com/npaka/n/na1f8ce42db4e
3. GEN-1: Scaling Embodied Foundation Models to Mastery - Generalist, 4月 13, 2026にアクセス、https://generalistai.com/blog/apr-02-2026-GEN-1

4. Generalist AI, 4月 13, 2026にアクセス、https://generalistai.com/
5. Generalist Bets Its Robot-Training Gloves Will Usher In Robotics' ChatGPT Moment - San Mateo Today - National Today, 4月 13, 2026にアクセス、https://nationaltoday.com/us/ca/san-mateo/news/2026/04/03/generalist-bets-its-robot-training-gloves-will-usher-in-robotics-chatgpt-moment/
6. Generalist AI unveils robot model it says masters simple physical tasks - Perplexity, 4月 13, 2026にアクセス、https://www.perplexity.ai/page/generalist-ai-unveils-robot-mo-Ya.bwalURhG6Rt8HhBVYgw
7. Generalist Uncovers Scaling Law of Embodied Intelligence, Enabling Model to Think and Act Simultaneously - 36氪, 4月 13, 2026にアクセス、https://eu.36kr.com/de/p/3561332469627784
8. Andy Zeng - Generalist, 4月 13, 2026にアクセス、https://andyzeng.github.io/
9. What is Thinking Machines building? we think it's a another ChatGPT like model, 4月 13, 2026にアクセス、https://www.themarketai.com/post/what-is-thinking-machines-building-we-think-it-s-a-another-chatgpt-like-model
10. Is there a complete list of open ai employees that have left due to safety issues? - Reddit, 4月 13, 2026にアクセス、https://www.reddit.com/r/ControlProblem/comments/1ituogu/is_there_a_complete_list_of_open_ai_employees/
11. Only 4 of OpenAI's 11 Founders Are Still With the Company—Where Are the Rest of Them?, 4月 13, 2026にアクセス、https://observer.com/2024/07/openai-founders-career/
12. Seven Unicorn Companies Founded by Former OpenAI Employees - 36氪, 4月 13, 2026にアクセス、https://eu.36kr.com/en/p/3412422888738184
13. Going Beyond World Models & VLAs - Generalist AI, 4月 13, 2026にアクセス、https://generalistai.com/blog/apr-07-2026-beyond-world-models
14. New AI Gives Robots Muscles Memory to Hit a 99% Success Rate on Delicate Manual Tasks - ZME Science, 4月 13, 2026にアクセス、https://www.zmescience.com/science/news-science/ai-muscle-memory-delicate-manual-taskl/
15. Generalist introduces GEN-1 general-purpose model for physical AI - The Robot Report, 4月 13, 2026にアクセス、https://www.therobotreport.com/generalist-introduces-gen-1-general-purpose-model-for-physical-ai/
16. Generalist Releases GEN-1 Robotic Intelligence Foundation Model - Let's Data Science, 4月 13, 2026にアクセス、https://letsdatascience.com/news/generalist-releases-gen-1-robotic-intelligence-foundation-mo-2fcfa7b8
17. Watch a Robot Stuff Cash Into a Wallet Just Like You Do - CNET, 4月 13, 2026にアクセス、https://www.cnet.com/tech/computing/gen-1-physical-ai-model-robot-dexterity/
18. GEN-0 / Embodied Foundation Models That Scale with Physical Interaction - Generalist, 4月 13, 2026にアクセス、

https://generalistai.com/blog/nov-04-2025-GEN-0
19. Embodied Foundation Model: GEN-0's Breakthrough in Robotics Intelligence - 高效码农, 4月 13, 2026にアクセス、https://www.xugj520.cn/en/archives/embodied-foundation-model-gen-0-robotics.html
20. A New Paradigm for General-Purpose Robotics, The Integration of, 4月 13, 2026にアクセス、https://jimsantana1.substack.com/p/a-new-paradigm-for-general-purpose?utm_source=youtube_shorts
21. How Far Are We From AGI: Are LLMs All We Need? - arXiv, 4月 13, 2026にアクセス、https://arxiv.org/html/2405.10313v2
22. From Language to Action: Can LLM-Based Agents Be Used for Embodied Robot Cognition?, 4月 13, 2026にアクセス、https://arxiv.org/html/2603.03148v1
23. Tencent Hunyuan Releases HPC-Ops: A High Performance LLM Inference Operator Library, 4月 13, 2026にアクセス、https://www.marktechpost.com/2026/01/27/tencent-hunyuan-releases-hpc-ops-a-high-performance-llm-inference-operator-library/
24. Pangu-Agent: A Fine-Tunable Generalist Agent with Structured Reasoning - arXiv, 4月 13, 2026にアクセス、https://arxiv.org/pdf/2312.14878
25. A paper review of RT-2 — What Happens When a Language Model Controls a Robot?, 4月 13, 2026にアクセス、https://medium.com/correll-lab/what-happens-when-a-language-model-controls-a-robot-a-review-of-rt-2-709e60ea9081
26. Generative Artificial Intelligence | PDF - Scribd, 4月 13, 2026にアクセス、https://www.scribd.com/document/728249841/Generative-artificial-intelligence
27. Comparing 5 Pioneering Robotics Foundation Models for ML-Based Control, 4月 13, 2026にアクセス、https://genki-sano.medium.com/a-practical-comparison-of-five-leading-ml-based-robotics-control-approaches-49e1977dd3ec
28. Meet Generalist at the Peak: How Did Qianxun Intelligence Secure $3 Billion in Just 30 Days?, 4月 13, 2026にアクセス、https://eu.36kr.com/en/p/3756066027209477
29. Generalist AI's GEN-1 Model Hits 99% Reliability in Production Robotics Tasks - MLQ.ai, 4月 13, 2026にアクセス、https://mlq.ai/news/generalist-ais-gen-1-model-hits-99-reliability-in-production-robotics-tasks/
30. Transforming the physical world with AI: the next frontier in intelligent automation - AWS, 4月 13, 2026にアクセス、https://aws.amazon.com/blogs/machine-learning/transforming-the-physical-world-with-ai-the-next-frontier-in-intelligent-automation/