

# Gemini Robotics On-Device: 次世代オンデバイスロボットAIの革新と競合分析

Google DeepMindが2025年6月24日に発表したGemini Robotics On-Deviceは、ロボティクス分野において画期的な技術革新をもたらすVision-Language-Action (VLA) モデルです<sup>[1]</sup><sup>[2]</sup><sup>[3]</sup>。この軽量化されたAIモデルは、インターネット接続を必要とせずにローカルのGPU上で実行可能であり、既存のオンデバイスモデルを大幅に上回る性能を実現しています<sup>[2]</sup><sup>[3]</sup><sup>[4]</sup>。本レポートでは、この革新的技術の詳細な分析と競合他社との包括的な比較を通じて、ロボティクスAI市場における影響と将来性を探求します。

## Gemini Robotics On-Deviceの技術革新

### 基本概要と発表背景

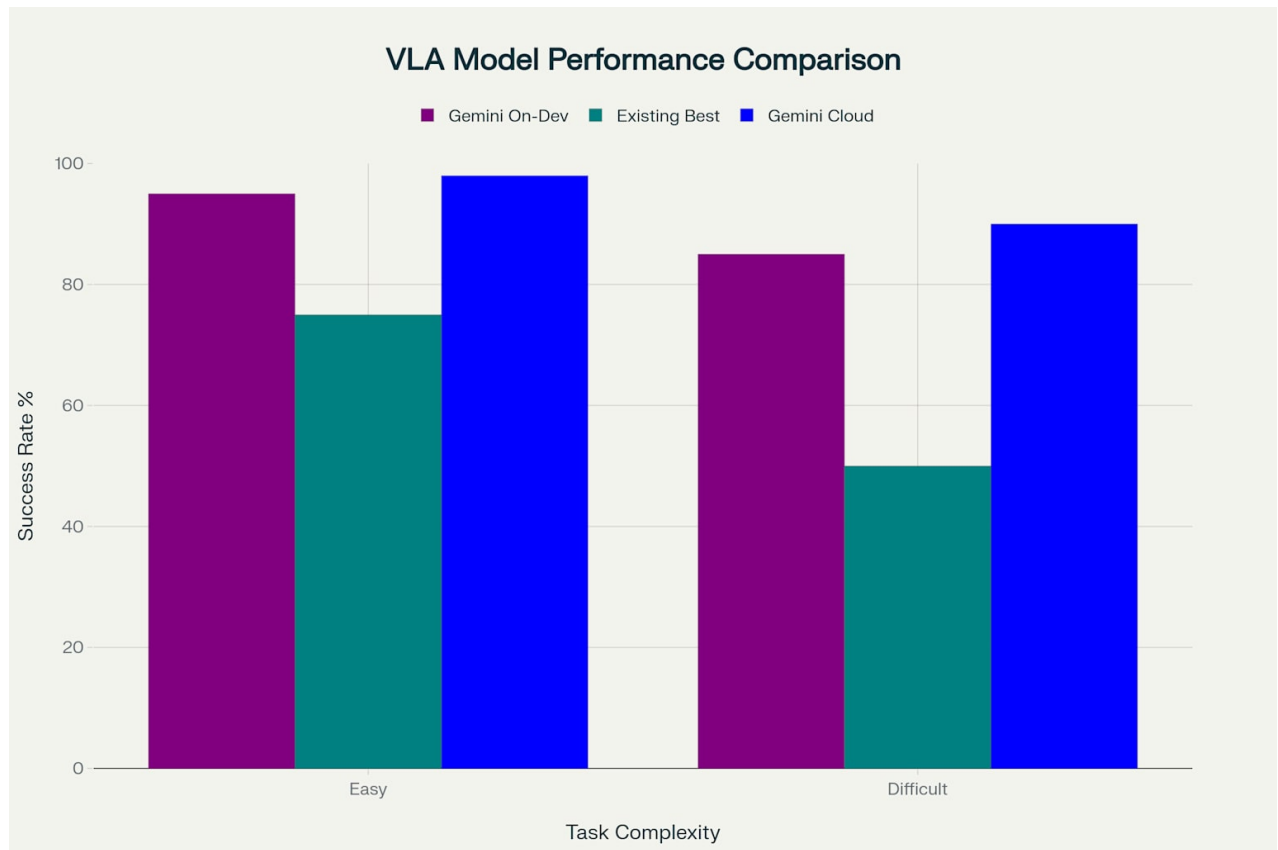
Gemini Robotics On-DeviceはGemini 2.0をベースとして開発された軽量版VLAモデルであり、2025年3月に発表されたクラウドベースのGemini Roboticsの後継技術として位置づけられています<sup>[1]</sup><sup>[5]</sup>。Google DeepMindのロボティクス部門責任者であるCarolina Paradaによると、このモデルは「一般的な用途の器用さと強力なタスク汎化能力」を示しており、特に2本アームロボット向けに最適化されています<sup>[3]</sup><sup>[6]</sup>。

このモデルの最大の革新点は、従来のクラウドベースVLAモデルが持つネットワーク遅延の問題を解決し、リアルタイムでの高精度なロボット制御を実現している点にあります<sup>[3]</sup><sup>[7]</sup><sup>[8]</sup>。インターネット接続が不安定な環境や、セキュリティ要件の厳しい産業環境での運用も可能となっています<sup>[8]</sup>。

### 技術仕様と性能指標

Gemini Robotics On-Deviceの技術仕様は、軽量性と高性能を両立させた設計が特徴的です<sup>[1]</sup><sup>[2]</sup>。モデルは50~100回のデモンストレーションのみで新しいタスクに適応できる学習能力を持ち、これは従来のロボット学習手法と比較して大幅な効率化を実現しています<sup>[2]</sup><sup>[4]</sup>。

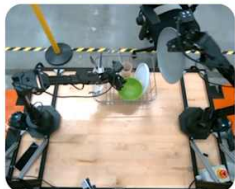
性能面では、簡単なタスクで95%、複雑なタスクで85%の成功率を達成しており、これは既存の最高性能オンデバイスモデルを大きく上回る結果です<sup>[2]</sup>。推論速度も従来モデルより30%高速化されており、リアルタイム制御に必要な低遅延要件を満たしています<sup>[4]</sup>。



VLAモデルの性能比較 - タスク複雑度別成功率

## 対応ロボットと実行可能タスク

Gemini Robotics On-Deviceは多様なロボットプラットフォームに対応しており、主要なものとしてGoogleのALOHAロボット、Franka Research 3産業用ロボットアーム、そしてApptронik社のApolloヒューマノイドロボットがあります<sup>[1]</sup> <sup>[2]</sup> <sup>[3]</sup>。



overhead camera



worms-eye camera



follower gripper



leader gripper

Google ALOHA 2 robotic workstations with bimanual manipulation arms and detailed component overview.

実行可能なタスクは高度な器用性を要求されるものが多く、バッグのジッパーを開ける、物体の選別と配置、衣類の折り畳み、組み立て作業、工具の使用など、従来のロボットでは困難とされていた作業を自然言語指示により実行できます [2] [3] [4]。



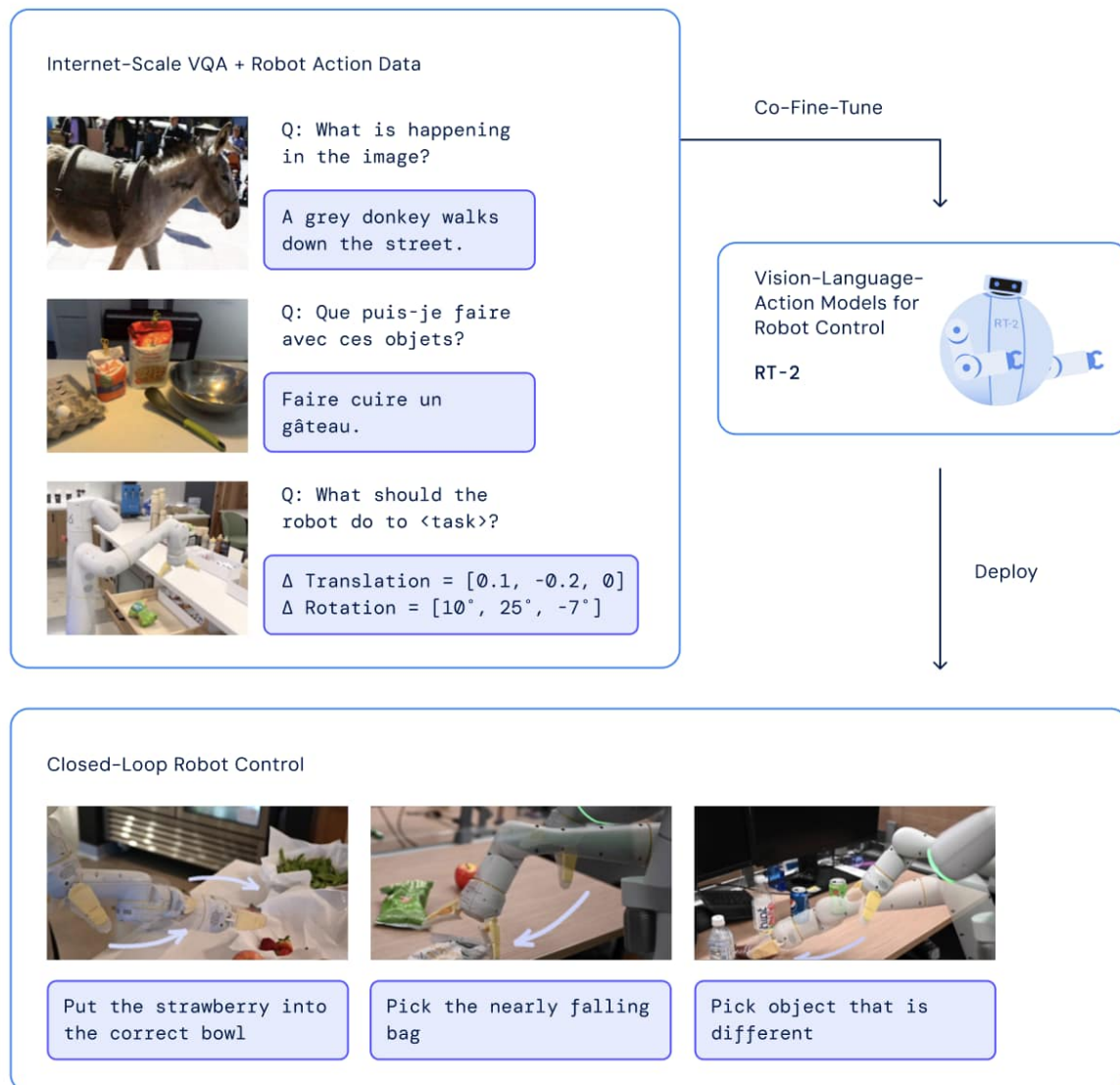
Apollo humanoid robot by Apptronik designed for manipulation tasks and human interaction.

## 競合他社との詳細比較分析

### 主要VLAモデルとの性能比較

ロボティクスAI市場における競合状況を分析すると、Gemini Robotics On-Deviceは明確な差別化要因を持っています<sup>[9] [10] [11]</sup>。HuggingFaceのSmolVLAは450Mパラメータで軽量性を重視していますが、タスク成功率ではGemini Robotics On-Deviceが大幅に優位に立っています<sup>[9]</sup>。

OpenVLAやChatVLA-2などの大規模モデルは汎用性が高い一方で、クラウド接続が必須であり、レイテンシーの問題を抱えています<sup>[10] [11]</sup>。Google DeepMindの以前のモデルであるRT-2は55Bパラメータの大規模モデルですが、オンデバイス実行には適していません<sup>[1] [12]</sup>。



RT-2 vision-language-action model workflow showing training on diverse data and deployment for closed-loop robot control tasks.

## 産業特化型ロボットAIとの比較

Tesla Bot AIは製造業に特化した設計となっており、Tesla独自のFSDチップを活用した高速処理能力を持っています<sup>[13] [14]</sup>。2025年第1四半期の業績発表によると、Teslaは年末までにOptimus Gen 3の生産ラインを完成させる計画を発表しており、数千台のOptimusユニットを北米工場に配備予定です<sup>[13]</sup>。

Boston DynamicsのAtlas AIは強化学習ベースのアプローチを採用しており、2025年2月にRobotics & AI Instituteとの提携により、動的で汎化可能なモバイル操作行動の構築を目指しています<sup>[15] [16]</sup>。このパートナーシップでは、シミュレーションから実世界への転移技術の向上に焦点を当てています<sup>[16]</sup>。

## 市場動向と投資状況の分析

### エンボディドAI市場の成長予測

エンボディドAI市場は急速な成長を示しており、2023年の35億ドルから2030年には231億ドルに達すると予測されています<sup>[17]</sup>。年平均成長率（CAGR）15.2%という高い成長率は、AI技術の進歩とロボティクス応用の拡大によるものです。

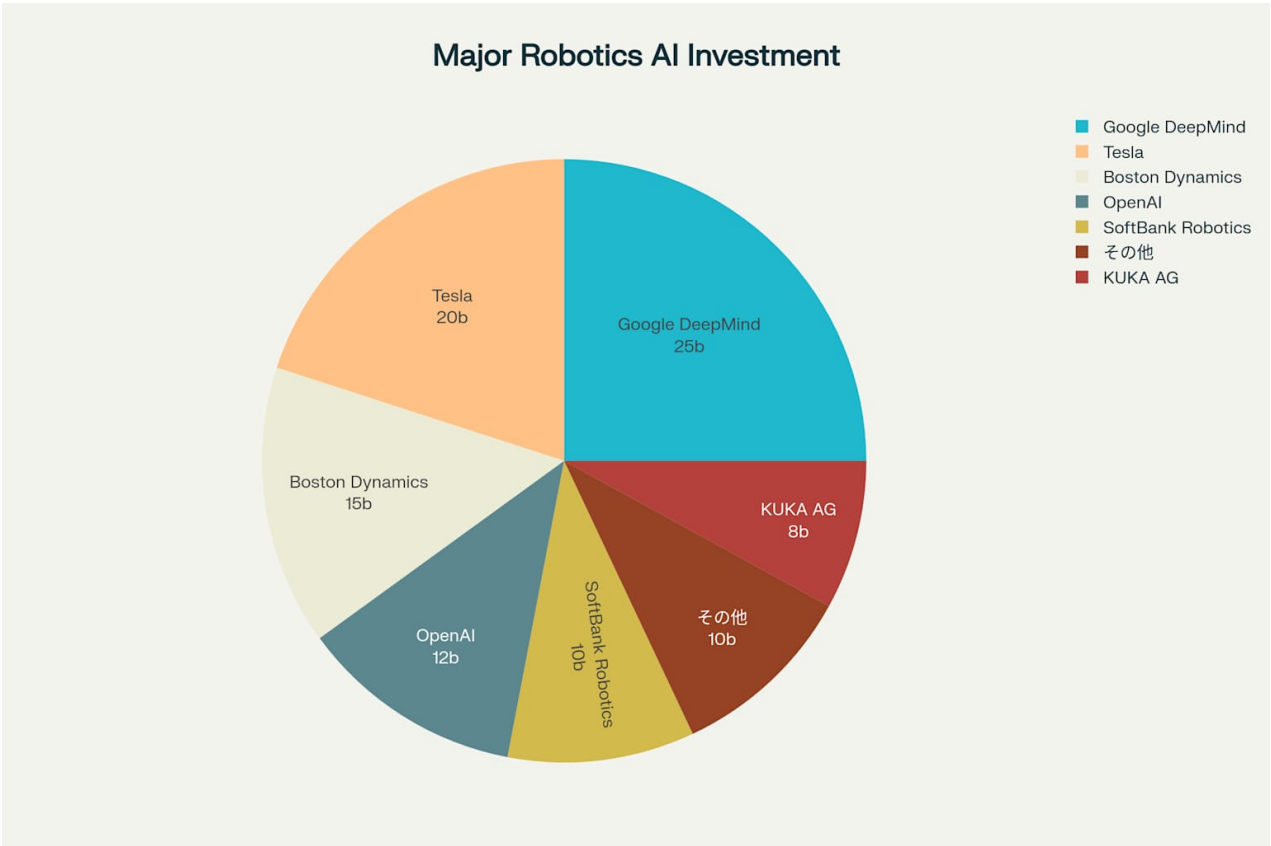


### エンボディドAI市場規模の成長予測（2023-2030年）

この成長を牽引する要因として、高齢化社会における労働力不足、製造業での自動化需要の増加、そして人工知能技術の向上が挙げられています<sup>[18] [17]</sup>。特に日本やヨーロッパなどの先進国では、構造的な労働力不足がロボティクス導入の大きな推進力となっています<sup>[18]</sup>。

## 主要企業への投資動向

ロボティクスAI分野への投資は2024年から2025年にかけて急激に増加しており、2024年10月だけで70億ドル以上の投資が実行されました<sup>[18]</sup>。Google DeepMindは25%のシェアで最大の投資を受けており、Teslaの20%、Boston Dynamicsの15%が続いています<sup>[18]</sup><sup>[17]</sup>。



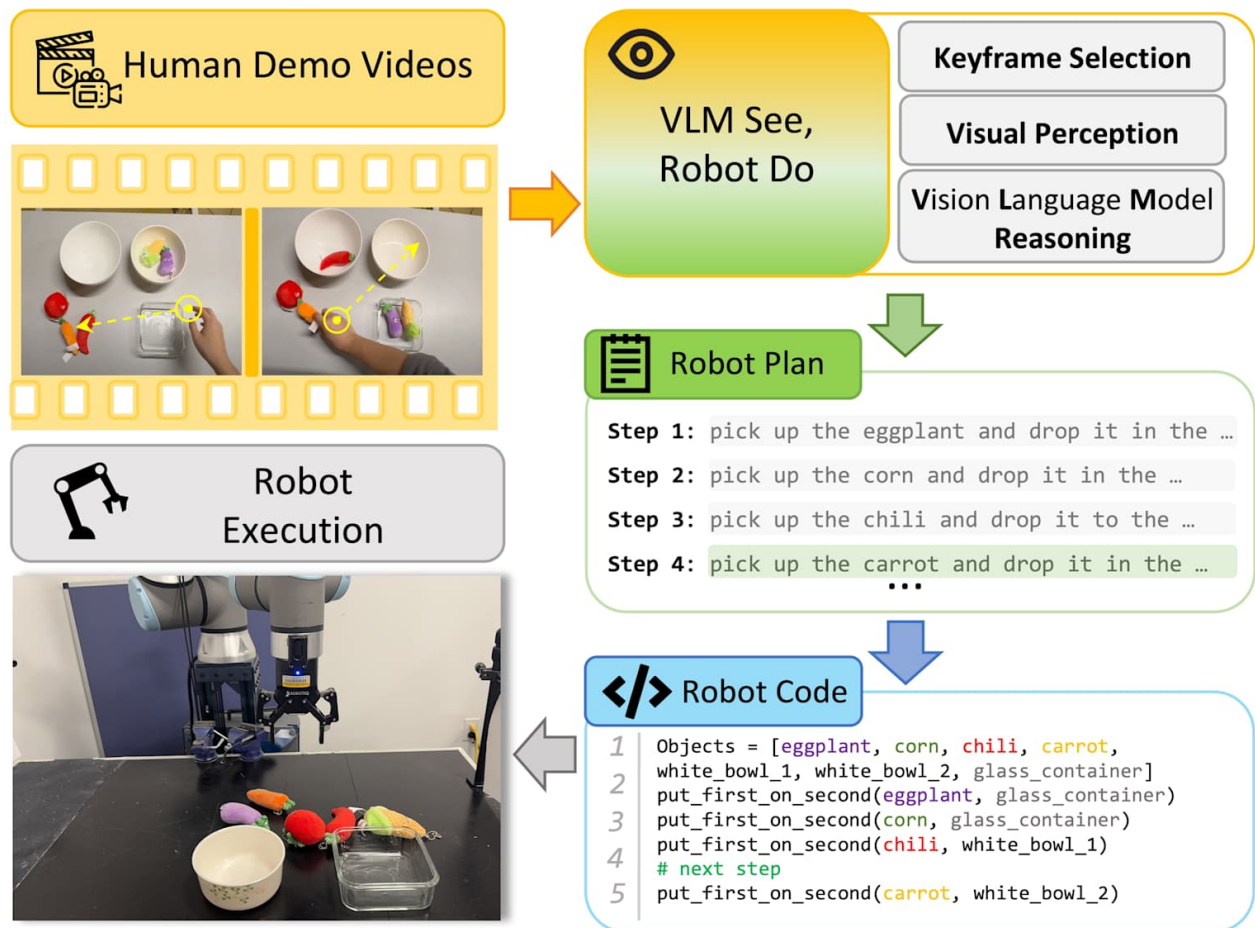
主要ロボティクスAI企業への投資分布（2024-2025年）

この投資増加は、ヒューマノイドロボット、農業自動化、外科用ロボティクスなどの新興カテゴリーに集中しており、長期的な技術開発と商業化を可能にしています<sup>[18]</sup>。防衛分野でも自律システムやAI駆動ロボティクスへの政府投資が増加しており、ドローン技術やロボット戦闘車両の開発が進んでいます<sup>[18]</sup>。

## 技術アーキテクチャと実装詳細

### Vision-Language-Actionモデルの仕組み

VLAモデルの基本アーキテクチャは、視覚入力、言語理解、行動出力を統合した単一のフレームワークで構成されています<sup>[4]</sup><sup>[11]</sup><sup>[19]</sup>。Gemini Robotics On-Deviceでは、この統合アプローチにより、人間のデモンストレーションから学習し、複雑な操作タスクを自律的に実行する能力を実現しています<sup>[4]</sup>。

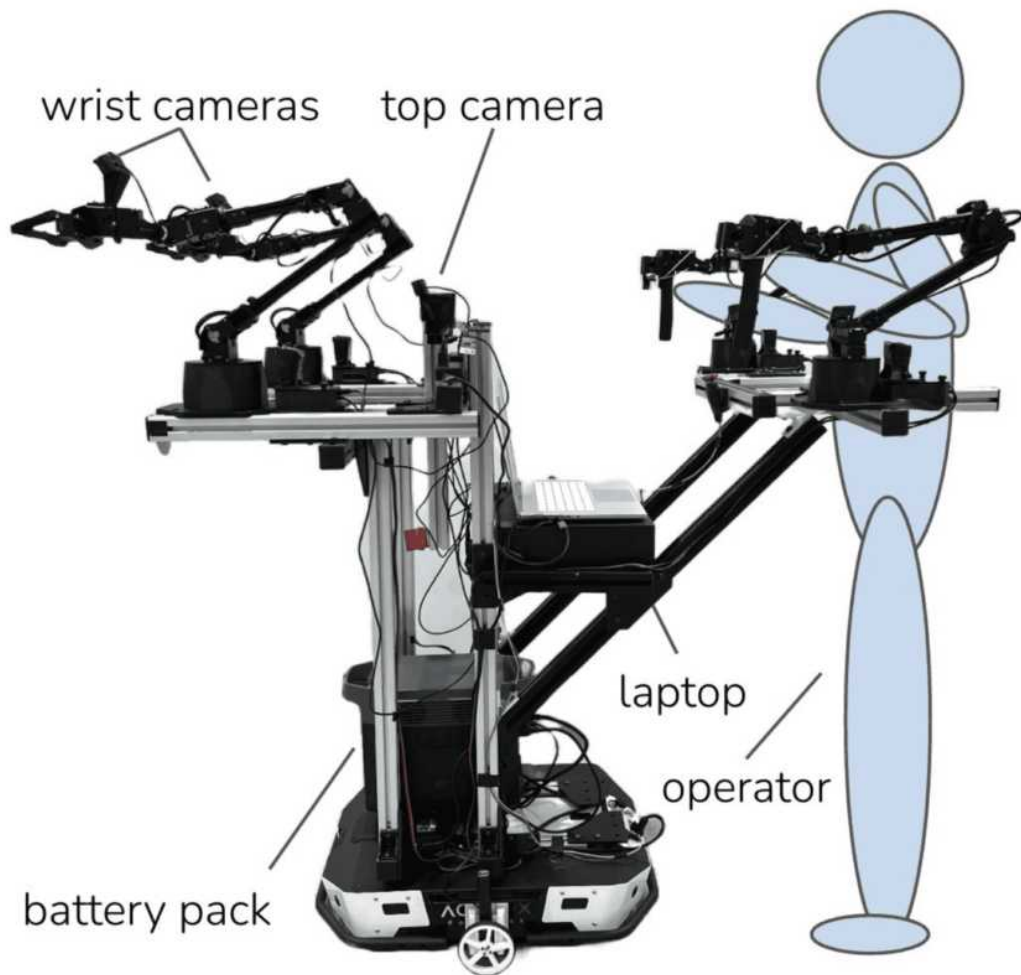


Vision-language model pipeline enabling a robot to learn from human demo videos and execute object manipulation tasks autonomously.

モデルの学習プロセスでは、視覚データ、言語指示、行動ログの3つの要素を組み合わせたマルチモーダルデータを使用します<sup>[1] [4]</sup>。この統合的アプローチにより、ロボットは新しい状況に対しても既存の知識を適用できる汎化能力を獲得しています<sup>[4] [8]</sup>。

## 開発環境とSDK

Gemini Robotics SDKはMuJoCo物理シミュレーター統合により、開発者がモデルを実際のハードウェアに展開する前にテストと最適化を行えるプラットフォームを提供しています<sup>[20]</sup>。このSDKは信頼できるテスター向けに限定公開されており、学術研究者や開発者のアクセス拡大が計画されています<sup>[20]</sup>。



A robotic system with articulated arms and wrist cameras for manipulation tasks.

## 産業応用と社会的影響

### 製造業および物流分野への影響

Gemini Robotics On-Deviceの産業応用は多岐にわたり、特に製造業での組み立て作業や品質管理において大きな変革をもたらす可能性があります<sup>[3] [6]</sup>。ローカル処理能力により、ネットワーク環境に依存しない安定した動作が可能となり、工場の生産ラインでの実用性が大幅に向上しています<sup>[3] [7]</sup>。

物流分野では、複雑な形状の物体の仕分けや梱包作業において、従来のロボットでは困難だった柔軟性と適応性を実現しています<sup>[4]</sup>。Amazon、FedEx、UPSなどの大手物流企業が類似技術の導入を検討しており、人間とロボットの協働による効率向上が期待されています<sup>[18]</sup>。

### ヘルスケアおよびサービス業への展開

ヘルスケア分野では、高齢者ケアや患者モニタリング、リハビリテーション支援において、Gemini Robotics On-Deviceの自然言語理解能力と精密な操作能力が活用される見込みです<sup>[17]</sup>。特に日本のような超高齢社会では、介護ロボットの需要が急速に拡大しており、このような技術の社会実装が急務となっています<sup>[17]</sup>。

サービス業においても、ホスピタリティ分野での顧客対応ロボットや、清掃・メンテナンス作業の自動化など、幅広い応用が期待されています<sup>[17]</sup>。これらの応用は、労働力不足の解決と同時に、サービス品質の向上にも寄与する可能性があります<sup>[17]</sup>。

## 技術的課題と今後の展望

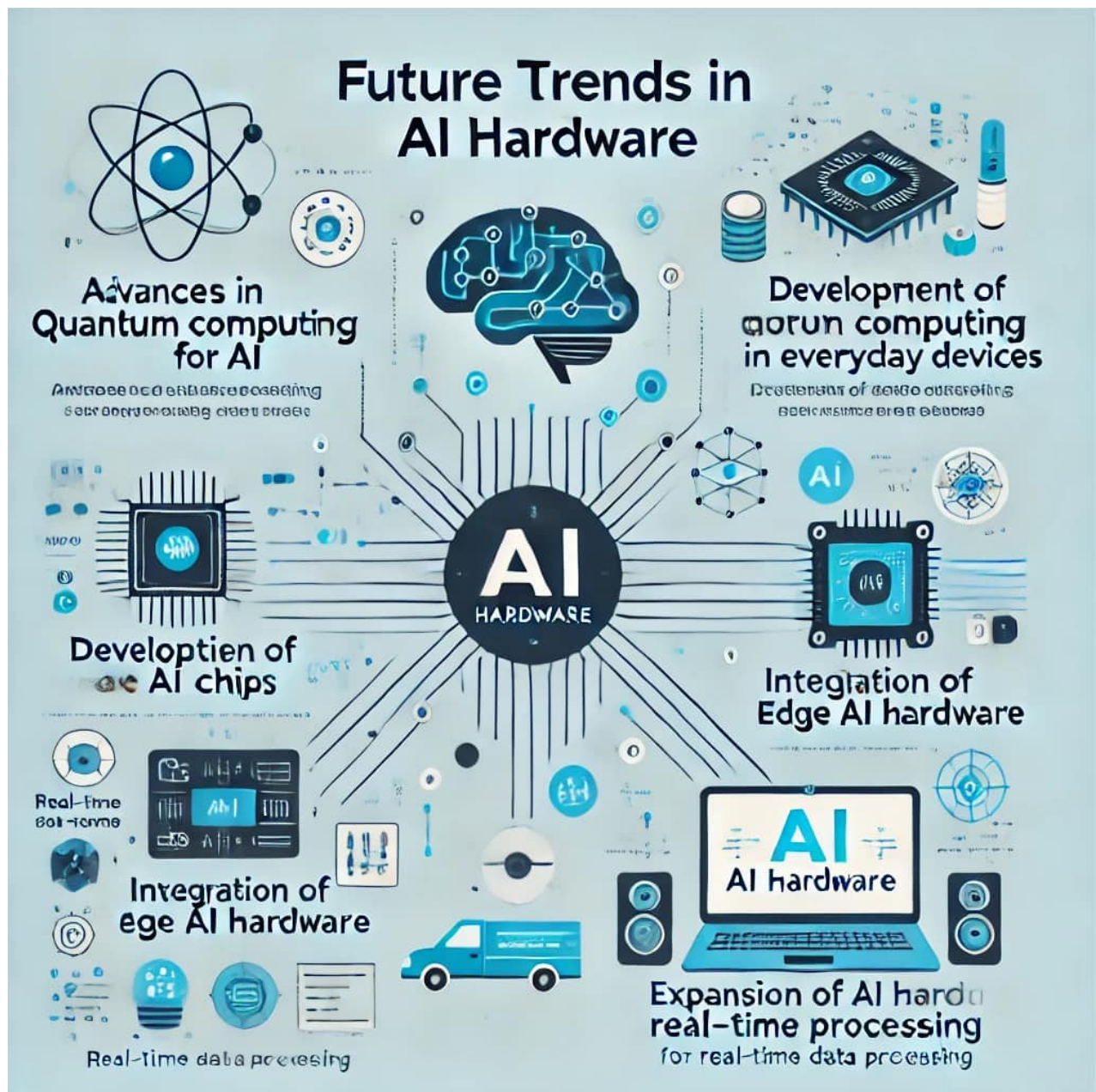
### 現在の技術的境界と改善課題

VLAモデルの評価において、現在のモデルは実用展開に必要な堅牢性が不足していることが指摘されています<sup>[21]</sup>。VLATestフレームワークによる評価では、障害物の数、照明条件、カメラポーズ、未知の物体などの要因がモデル性能に大きく影響することが明らかになっています<sup>[21]</sup>。

さらに、エンボディドAIシステムの包括的評価のため、EmbodiedEvalなどの新しいベンチマークが開発されており、ナビゲーション、物体操作、社会的相互作用、属性質問応答、空間質問応答の5つのカテゴリでMLLMsの評価が行われています<sup>[22]</sup>。これらの評価により、現在のモデルが人間レベルの性能に比べて大きな差があることが判明しています<sup>[22]</sup>。

### 将来技術の発展方向

AI技術の進歩により、今後のロボティクスシステムはより高度な推論能力と適応性を持つことが予想されます<sup>[5] [23]</sup>。量子コンピューティングとAIの統合、エッジAIハードウェアの発展、リアルタイムデータ処理能力の向上などが、次世代ロボティクスシステムの基盤技術となる見込みです<sup>[5]</sup>。



Infographic on future trends in AI hardware development and integration

特に、マルチモーダル基盤モデルとワールドモデルの統合により、より汎用的で適応性の高いエンボディドエージェントの実現が期待されています<sup>[23]</sup>。これらの技術は、サイバー空間と物理世界を橋渡しし、真の意味での人工汎用知能（AGI）の実現に向けた重要なステップとなる可能性があります<sup>[23]</sup>。

## 結論

Gemini Robotics On-Deviceの発表は、ロボティクスAI分野における重要な転換点を示しています<sup>[1]</sup><sup>[2]</sup><sup>[3]</sup>。オンデバイス実行による低遅延性、高い器用性、自然言語指示への対応能力を組み合わせることで、従来のロボット技術の限界を大幅に拡張しています<sup>[3]</sup><sup>[4]</sup><sup>[7]</sup>。

競合他社との比較においても、Google DeepMindは技術革新と市場投資の両面で優位性を保持しており、エンボディドAI市場の成長を牽引する立場にあります<sup>[18]</sup><sup>[17]</sup>。しかし、実用化に向けては技術的課題も存在し、特に複雑な実世界環境での堅牢性向上が重要な課題となっています<sup>[21]</sup>。

今後数年間で、ロボティクスAI技術は製造業、物流、ヘルスケア、サービス業など多様な分野で実装が進み、社会構造の変革をもたらす可能性があります<sup>[17]</sup>。Gemini Robotics On-Deviceのような革新的技術の継続的な発展により、人間とロボットが協働する新たな社会の実現が期待されています<sup>[23] [17]</sup>。



1. <https://developers.googleblog.com/en/gemini-25-for-robotics-and-embodied-intelligence/>
2. <https://gigazine.net/news/20250625-gemini-robotics-on-device/>
3. <https://techcrunch.com/2025/06/24/google-rolls-out-new-gemini-model-that-can-run-on-robots-locally/>
4. <https://wandb.ai/byyoung3/ml-news/reports/Google-DeepMind-releases-Gemini-Robotics-On-Device-Model--VmlldzoxMzMzMzMyNw>
5. <https://www.watch.impress.co.jp/docs/news/2025768.html>
6. <https://huggingface.co/blog/smolv1a>
7. <https://arstechnica.com/google/2025/06/google-releases-first-cloud-free-ai-robotics-model/>
8. <https://www.theverge.com/news/691882/google-deepmind-on-device-ai-robots-gemini>
9. <https://arxiv.org/abs/2505.21906>
10. <https://www.linkedin.com/pulse/ai-robotics-agents-april-2025s-key-developments-vicki-reyzelman-3dz5e>
11. <https://www.numberanalytics.com/blog/cognitive-robotics-performance-metrics>
12. <https://www.automate.org/industry-insights/google-deepmind-debuts-gemini-robotics-on-device-visual-language-model>
13. <https://autogpt.net/openai-plans-to-build-robots-for-real-world-applications/>
14. <https://arxiv.org/html/2506.05667v1>
15. [https://en.wikipedia.org/wiki/Optimus\\_\(robot\)](https://en.wikipedia.org/wiki/Optimus_(robot))
16. <https://arxiv.org/html/2505.04769v1>
17. [https://www.meegle.com/en\\_us/topics/robotics/robotic-investment-trends](https://www.meegle.com/en_us/topics/robotics/robotic-investment-trends)
18. <https://conf.researchr.org/details/fse-2025/fse-2025-research-papers/3/VLATEST-Testing-and-Evaluating-Vision-Language-Action-Models-for-Robotic-Manipulation>
19. <https://www.marketsandmarkets.com/ResearchInsight/embodied-ai-market.asp>
20. <https://www.youtube.com/watch?v=Wm5Yani53nM>
21. <https://virtuemarketresearch.com/report/embodied-ai-systems-market>
22. <https://arxiv.org/html/2407.06886v1>
23. <https://www.osmosisai.co/2025/03/19/the-rise-of-vision-language-action-models-in-autonomous-driving/>