

第2節 最近の研究開発・社会実装動向

AI分野は日々新たな技術やモデルが発表されるなど、目覚ましく技術革新が進んでいる分野である。そのため、その動きの全てを採り上げることはできないが、日本が有する強みやポテンシャルという観点も踏まえながら、最近の研究開発や社会実装における動向を概観する。

1 小規模言語モデル

1 概要

ア 定義

小規模言語モデル (SLM: Small Language Model) については、一般に、大規模言語モデル (LLM) と比較してパラメータ数が小さいモデルとされるが、その規模感については幅があり、厳密な定義は統一されていない状況である。海外におけるSLMのベンチマークに関する研究論文では、パラメータ数が70億以下のモデルをSLMとして取り扱っている一方で^{*1}、IBMはSLMのパラメータ数について数十～数百億程度としており^{*2}、また、人工知能学会の論文においては、140億パラメータ以下のモデルをSLMとして扱う事例も存在する^{*3}。本節では、おおよそパラメータ数が百億程度までのモデルをSLMとして取り扱う。

イ 特徴

SLMはLLMと比べ、パラメータ数が少なく、LLMを利用した効率的な学習なども用いられているほか、カスタマイズが容易である。また、特定分野に限定した学習をさせることで、軽量で特定分野に高い精度をもつモデルの構築も行われている。さらに、計算量が抑えられていることから、計算資源や消費電力の制約がある環境でも利用可能であり、エッジデバイス上での実行に適している。エッジデバイスでの利用は通信を必要としないため、個人情報などプライバシーの保護やセキュリティの観点でも期待が集まっている^{*4*5}。

一方で、パラメータ数が少ないことで、知識のカバー範囲や複雑な推論能力には限界があるとされており、汎用的な知識や高度な推論が求められる用途においては、SLM単体での対応には限界がある。そのため、実務においては検索拡張生成 (RAG) を用いた外部知識の活用や、LLMとSLMを組み合わせた構成が有効なアプローチとして整理されている^{*6}。

国立情報学研究所の佐藤一郎教授^{*7}は、SLMの発展について、当初LLMが性能を重視して開発されてきたが、現在はデータ量などを減らせることがわかり、小型化の技術が非常に進んでいるという。また、SLMの技術発展に伴い、スマートフォンやタブレットといったデバイスでは、デバイス上で利用可能なSLMの性能の限界から、ユーザーが満足する性能を実現できるか不確実なため、SLMの利用が必須にはならないが、デバイスとしてのパソコンは、スマートフォンやタブレットとの差別化のためにSLMが動くことが必須となってくる可能性があり、SLMが動く必要があるデバイスとそうでないデバイスで二極化が進むと指摘している。

^{*1} Pham, N. T. et al. (2025) "SLM-Bench: A Comprehensive Benchmark of Small Language Models on Environmental Impacts." (Association for Computational Linguistics [Findings of the Association for Computational Linguistics] EMNLP 2025 pp. 21369-21392)

^{*2} Rina Diane Caballar 「小規模言語モデルとは」(https://www.ibm.com/jp-ja/think/topics/small-language-models) (2026年1月30日参照)

^{*3} 杉山 弘晃 (2025) 「小規模な言語モデルに関する研究開発の動向」(『人工知能』40 (3), 328-336, 人工知能学会)

^{*4} 黒田 康浩, 小澤 佑介, 鈴木 文晴 (2025) 「小規模な言語モデルに関する研究開発の動向」(『人工知能』40 (3), 319-327, 人工知能学会)

^{*5} 日高 雅俊 (2025) 「エンドユーザデバイスにおける AI の推論・学習」(『人工知能』40 (3), 311-318, 人工知能学会)

^{*6} 脚注4に同じ。

^{*7} 国立情報学研究所 佐藤一郎教授へのヒアリングに基づく。

2 社会実装における期待

SLMは、その軽量性と通信を必要としない特性を活かし、社会実装の観点から多様な活用が期待されている。具体的には小型であることで、大規模な計算資源を必要とせず、端末内で処理を行うエッジAIとしての利用が挙げられる^{*8}。また、SLMが通信を必要としないことに付随する結果として、セキュリティの向上も期待されている。スマートフォンやウェアラブル端末など、個人が利用するデバイスに搭載することで、即時応答が求められるサービスや、プライバシーの保護が必要な利用シーンへの応用が想定されている。さらに、金融やコード生成をはじめとする特定領域に特化したSLMを用いることで、安全性や信頼性を確保しながら業務支援を行う活用が期待されている^{*9}。

3 国際競争における日本の位置づけ

国立情報学研究所の佐藤一郎教授^{*10 *11}は、汎用LLMはコモディティー化が見込まれるため、日本は領域特化型のAIに注力すべきだとしており、また、領域特化型のSLMについて、特に、音声対応のAIの開発や、製造現場など市場が大きいだけでなく、安全性が高く高品質な日本の業務の特性を活かすことができる分野への応用に注力すべきと考えている。

佐藤教授^{*12}によれば、製造現場などの産業分野では、日本独自の高い安全性や品質を実現する業務の仕方にAIを導入することで、より確実な業務の遂行といった改善効果が見込まれるという。例えば、電車の運転手が実施する「指さし確認」では、現状は実施者のみによる確認作業にとどまるが、AIを導入することで、指さし確認のし忘れをAIが指摘してくれるようになる。指さし確認をはじめとする、高い安全性や品質を実現する日本型の業務の仕方は、海外では導入が進んでいないものも多い。そこで、AIを利用して品質を高めた安全な業務の仕方そのものを海外展開していくことや、業務の仕方の国際標準化を先導していくことが日本の勝ち筋になり得るとしている。

4 研究開発動向

SLMの開発動向の概況として、実際にいくつかのベンチマークでLLMを上回る性能を持つモデルや、領域特化型のモデルの開発が進められている。このほかに、SLMで画像と言語といった複数のモダリティを統合するマルチモーダルな基盤の研究も行われている。

日本の動向に注目すると、純国産のLLM開発に取り組むPreferred Networksが、自社LLMの技術を活用した小規模モデルの開発も行っており、自動車や製造設備など、エッジデバイス向けの軽量モデルとして「PLaMo Lite」を提供している^{*13}。PLaMo Liteは日本語性能を評価する各種ベンチマークで、同程度のパラメータ数の他モデルと比べ圧倒的に高い性能を示しており、パラメータ数の多いLLMと比べて遜色ない結果を示している（図表I-2-2-1）。また、2025年には、ドローンやロボットなどの自律稼働デバイスやオンプレミス環境等、クラウド接続が難しい環境向けの軽量な視覚言語モデルを開発しており、経済産業省及びNEDOが実施するGENIAC（Generative AI Accelerator Challenge）のテーマとしても採択されている^{*14}。

*8 脚注5に同じ。

*9 脚注4に同じ。

*10 国立情報学研究所 佐藤一郎教授へのヒアリングに基づく。

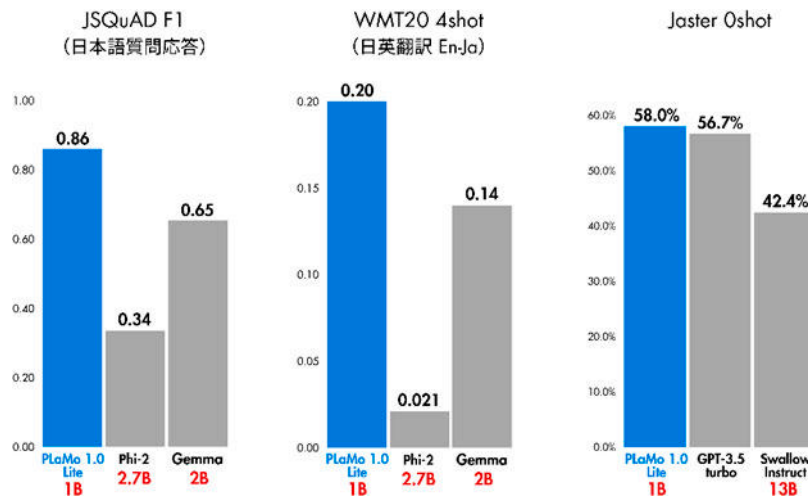
*11 佐藤一郎（2025）『2030 次世代AI 日本の勝ち筋』（日経BP）

*12 国立情報学研究所 佐藤一郎教授へのヒアリングに基づく。

*13 株式会社Preferred Networks「小型で軽量な小規模言語モデルPLaMo Liteの提供開始 - 株式会社Preferred Networks」〈<https://www.preferred.jp/ja/news/pr20240828>〉（2026年3月9日参照）

*14 株式会社Preferred Networks「GENIAC第3期で自律稼働デバイス向けの軽量な大規模視覚言語モデルPLaMo 2.1-8B-VLを開発 - 株式会社Preferred Networks」〈<https://www.preferred.jp/ja/news/pr20251216>〉（2026年3月9日参照）

図表 I-2-2-1 PLaMo Lite の日本語性能評価

(出典) 株式会社 Preferred Networks 「小型で軽量な小規模言語モデル PLaMo Lite の提供開始」^{*15}

5 社会実装等の動向

既に実用化されているSLMを利用した製品やサービスとしては、パソコンやスマートフォンにおけるデバイス上のAI機能の実装や、デバイスのAIとクラウドのAIを使い分ける仕組みなどがある。また、企業による活用では、通信がない環境での作業効率化のための活用や、領域特化型のAIの実装が始まっている。

例えば、東京大学松尾研究室発のスタートアップである株式会社EQUESは、GENIAC第2期として2024年10月に採択されたテーマで開発した、製薬業界向けの特化型LLM「JPharmatron-7B」を発表している。「JPharmatron-7B」は、薬剤師国家試験や医師国家試験などの成績で、当時公表されていた他モデルを上回る成績を取っており、同社が提供する製薬業界向けAIソリューション「QAI」シリーズへの実装を見据えている^{*16}。

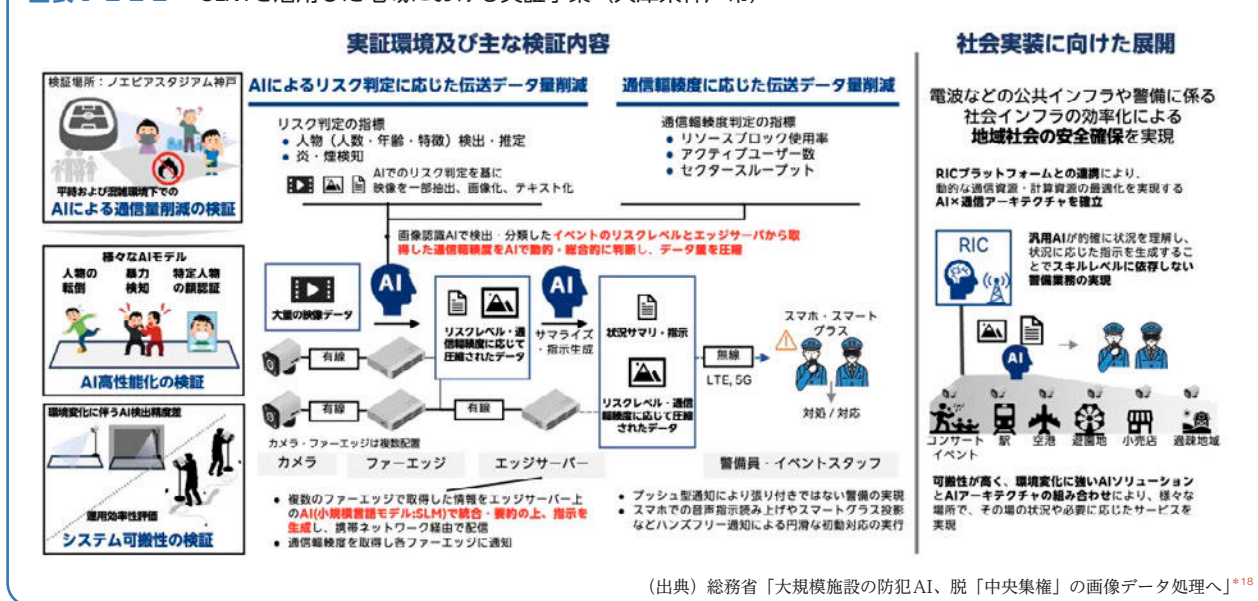
また、SLMの利用は、地域においても進められつつある。楽天モバイル株式会社とエッジAIに強みを持つAWL株式会社、楽天ヴィッセル神戸株式会社（ヴィッセル神戸）の3社が、総務省の「地域社会DX推進パッケージ事業（AI検証タイプ）」により、兵庫県神戸市のノエビアスタジアム神戸で、監視カメラの近くにある「ファアーエッジ」と複数のファアーエッジを接続する「エッジサーバー」で構成されるエッジAIを活用し、立ち入り人物のリスクを判断・通知するシステムの実証を行っている。ファアーエッジには画像を認識するAIが搭載され、人物や異常を検知・分析し、エッジサーバーの通信混雑具合とリスクに応じて、テキスト、動画、静止画に情報を圧縮する。エッジサーバーは、ファアーエッジからの情報をSLMで統合し、要約、指示を生成し、警備員にレポートする。クラウドAIと比べ、通信の混雑に影響されないため、迅速に情報をやり取りできる点やクラウドの高性能GPUを使用するより省電力・省コストである点に期待が集まっている（図表 I-2-2-2）^{*17}。

^{*15} 脚注13に同じ。

^{*16} 株式会社EQUES「＜リリースのお知らせ＞製薬特化の大規模言語モデル「JPharmatron-7B」を開発しました」〈<https://eques.co.jp/news/250618/>〉（2026年2月16日参照）

^{*17} 総務省（2025）「大規模施設の防犯AI、脱「中央集権」の画像データ処理へ」（『地域社会DXナビ』2025年10月14日）〈https://dx-navi.soumu.go.jp/support_r7/dx_package/demonstration/article/001〉（2026年2月27日参照）

図表 I-2-2-2 SLMを活用した地域における実証事業（兵庫県神戸市）



2 フィジカルAI

1 概要

国立研究開発法人科学技術振興機構 研究開発戦略センター（CRDS）のプロポーザルにおいては、フィジカルAIは「センサやアクチュエーターなどの物理的な機構を通じて環境と相互作用し、得られた入力と動作の結果に基づいて知能を獲得・発達させる、身体性を備えた人工知能」とされている^{*19}。広義にはスマートグラスのような空間認知に基づきデジタル空間に生成結果を出力するものも含まれるが、本項では、ロボットのように出力が実空間に影響を与えるものを中心に上げる。フィジカルAIとしては、産業用ロボットだけでなく人型ロボットのヒューマノイドの研究開発も盛んになっている。

2 社会実装における期待と課題

フィジカルAIが発展することで、人間のように様々なタスクをこなせるようになり、部品や組立て方法の異なる製品を、一台のロボットで製造可能となることが期待されている。また、家庭内でロボットが家事を全て代行することや、災害現場での作業を代行するといった活用も期待されている。このようにロボットが環境に適応しながら柔軟に多様なタスクをこなしていくための課題として、CRDSのプロポーザル^{*20}では、巧緻性（多様なタスクへの対応）、頑健性（多様な環境への適応力）、協調性（人間との安全な協働）が挙げられている。

3 国際競争における日本の位置づけ

CRDSのプロポーザル^{*21}では、産業用ロボットの制御・統合技術や、災害対応などロボットの活躍が期待される現場環境への対応が、日本における強みとして整理されている。同プロポーザルでは、米中が大規模でデータドリブンな汎用性の実現を重視するのに対し、日本は実用性を重視し、各分野の社

^{*18} 同上。

^{*19} CRDS（2025）「フィジカルAIシステムの研究開発～身体性を備えたAIとロボティクスの融合～」〈<https://www.jst.go.jp/crds/pdf/2025/SP/CRDS-FY2025-SP-01.pdf>〉（2026年1月30日参照）

^{*20} 同上。

^{*21} 同上。

会実装に即したロボット技術を活用する分野特化型のアプローチとして位置づけられている。

AIロボット協会（AIRoA）のCTOである東京大学の松嶋達也特任助教^{*22}は、日本はロボット技術に強みがあるが、今後はソフトウェアやAIとの融合がカギになるとしている。ロボット産業は、ハードウェアを必要とすることもあり、1社のみで急速に事業規模を拡大させることが難しく、業界全体でデータや知見の共有、エコシステムを育てることが重要だとも提言している。

国立情報学研究所の佐藤一郎教授^{*23}は、フィジカルAIの開発に関して、家庭用ヒューマノイドの開発における中国の取組を念頭に、現実世界のデータにこだわらず、仮想空間でのAIの学習を進めることも有効だとしている。また、日本の工場用のロボットのハードにはまだ強みがあるとした上で、今後AI分野を強化できるかが勝負の分かれ目となるとし、例えば工場での利用に向けた開発においては、日本の工場で働いていた熟練者等、作業の手順などを熟知している人材を活用してAIの学習を進めていくことが、日本の勝ち筋につながる可能性があるとしている。また、ロボットの現場での実用化については、ファミリーレストランの配膳ロボットを例に挙げながら、作業をすべて自動化しようとするのではなく、人間がやるべき作業とロボットに任せるべき作業をうまく切り分けることを考えるべきであると指摘している。

4 研究開発動向

フィジカルAIの情報処理の流れは、入力、処理、出力の3段階に分けることができ^{*24}、入力・出力に関連する技術としてロボットハンドの開発や触覚センサの研究が盛んになっているほか、処理の分野の研究では、視覚・言語など複数のモダリティを統合した基盤モデルや世界モデル等の研究が行われている。

日本の動向に注目すると、東京大学の長谷川峻特任助教と矢野倉伊織助教、岡田慧教授らが、約20cm四方で28×17の合計476個のセンサを搭載した触覚センサシートを開発した上で、このシートの一部を使用して計190個の触覚センサを搭載した5本指ハンドを開発した（図表 I-2-2-3）。長谷川特任助教らが開発したシートには配線が埋め込まれているが、折り曲げ可能であり、ロボットハンドの形状に応じて切り貼りして利用できるという。また、開発したロボットハンドでは、指と指の間の摩擦や物を握る際の情報取得が可能になったという^{*25}。

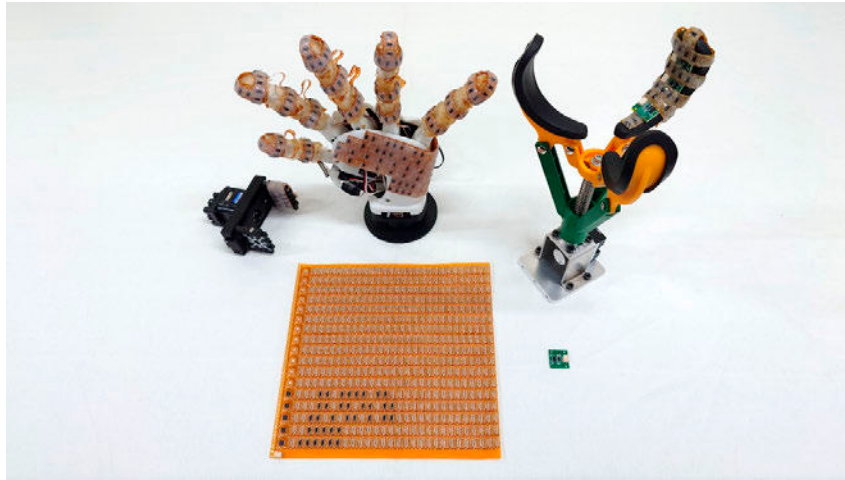
^{*22} 日経クロステック（2026年3月9日）鈴木 慶太「ロボットはデータで賢くなる、AIロボット協会の松嶋CTOが描く普及シナリオ（2ページ目）」〈<https://xtech.nikkei.com/atcl/nxt/column/18/03076/022400023/>〉（2026年3月26日参照）

^{*23} 国立情報学研究所 佐藤一郎教授へのヒアリングに基づく。

^{*24} SCSK「フィジカル AI（Physical AI）とは？生成AIやAIエージェントの次に来る、現実世界で「動くAI」の全貌を徹底解説！」（SCSKNVIDIA 合同会社 エンタープライズ事業本部 高橋 想監修）〈<https://www.scsk.jp/sp/itpnavi/article/2025/10/physicalai.html>〉（2026年2月16日参照）

^{*25} ニューススイッチ by 日刊工業新聞社（2025年7月19日）「触覚センサ190個搭載…東大が開発、5本指ハンドがスゴイ」〈<https://newswitch.jp/p/46361>〉（2026年3月2日参照）

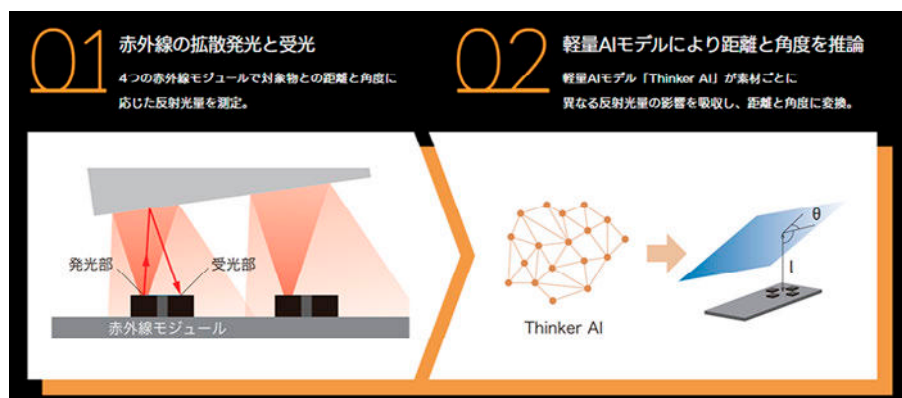
図表 I-2-2-3 触覚センサシートやロボットハンド（東京大学 長谷川峻特任助教ら開発）



（出典）東京大学 長谷川峻特任助教より提供

また、ロボットの近接覚センサの開発も進められている。近接覚とは、表面から近い距離にある物体の情報を取得する感覚で、ロボットの表面から数十mmにある対象物の位置を直接取得することができる^{*26}^{*27}。日本では、Thinker株式会社が独自のエッジAIと赤外線との組合せにより、対象物との「距離および姿勢」を、非接触かつ高分解能でリアルタイムに計測できる近接覚センサの開発を行っている（図表 I-2-2-4）。同社によれば、近接覚センサにより、これまで取扱いが困難であった柔らかいものや形状が一定でないものの捕捉が可能になるだけでなく、従来の光センサやカメラが苦手としていた透明物や鏡面物の計測も可能になり、ロボットハンドの活用可能性が広がるとしている。加えて、ロボットに動きを教える「ティーチング」においても、ロボット自身がリアルタイムで対象物を計測しながら指先の位置を微調整できるようになるため、「ラフティーチング」が可能となり、生産性の向上やコスト削減にもつながるとしている^{*28}。

図表 I-2-2-4 近接覚センサの仕組み（株式会社Thinker）

（出典）Thinker「近接覚センサ TK-01N / TK-01G（開発中）」^{*29}

*26 九州大学ヒューマンセンタードロボティクス研究室「近接覚センサ」〈https://www.hcr.mech.kyushu-u.ac.jp/research/proximity_sensors/〉（2026年3月2日参照）

*27 電気通信大学 明・佐藤研究室「近接覚を用いたロボットシステムの研究開発」〈<http://www.rm.mce.uec.ac.jp/ming/?%E8%BF%91%E6%8E%A5%E8%A6%9A%E3%82%92%E7%94%A8%E3%81%B4%E3%81%9F%E3%83%AD%E3%83%9C%E3%83%83%E3%83%88%E3%82%B7%E3%82%B9%E3%83%86%E3%83%A0%E3%81%AE%E7%A0%94%E7%A9%B6%E9%96%8B%E7%99%BA>〉（2026年3月2日参照）

*28 株式会社Thinker「近接覚センサ TK-01N / TK-01G（開発中）」〈<https://www.thinker-robotics.co.jp/product/sensor-tk-01/>〉（2026年3月2日参照）

*29 同上。

処理については、日本では、GENIACにて、Direava株式会社が「外科手術支援のための視覚・言語統合型AI基盤モデルの開発」のテーマにおいて、視覚と言語を統合した術式・解剖構造などを学習させたモデルの開発に取り組んでいる。また、Turing株式会社による「完全自動運転に向けた車載可能なフィジカル基盤モデルの開発」のテーマにおいては、視覚と言語、行動を統合し、実世界の状況変化や移動体の動きを予測できるフィジカル基盤モデルを開発している。

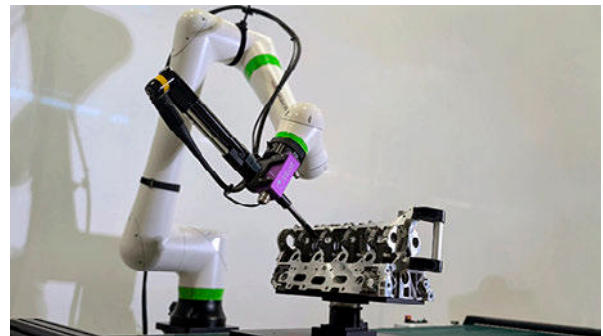
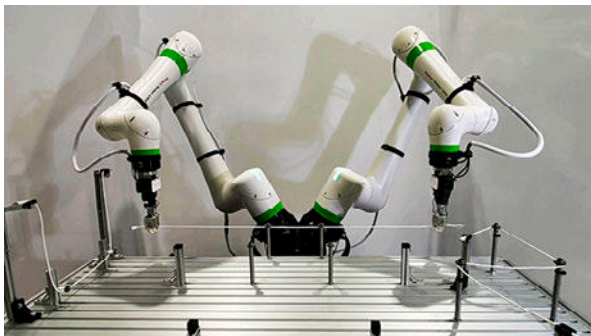
産業技術総合研究所（産総研）においても、2025年1月、実世界の困難作業の自動化を目指したロボット基盤モデルの研究開発に取り組むことを公表している。産総研では、視覚と力覚の関係を推定するAIモデルや、人間の認知発達過程を模した動作学習基盤などを研究してきており、これらの研究成果を強みとして、その社会実装の推進のため、産業分野や家庭、オフィスなどのサービス分野で役立つ基盤モデルの開発を目指すとしている^{*30}。その一環として、双腕ロボットAIの開発を支援するデータセットを無償公開した^{*31}。

内閣府ムーンショット型研究開発制度の目標3「2050年までに、AIとロボットの共進化により、自ら学習・行動し人と共生するロボットを実現」において、プログラムディレクターを務める東京大学の國吉康夫教授の方針のもと公募を実施し、2026年5月に14名のプロジェクトマネージャーが採択された。これらのプロジェクトにより、2030年までに、特定の現場における特定のタスクについて、人による支援のもと、実用上違和感なく状況の変化に対応しながら一連のタスクを完遂でき、民間投資対象となり得る汎用自律人型AIロボットのプロトタイプの開発等を目指している。

5 社会実装等の動向

ファナックは、オープンソースソフトウェアであるROS2ドライバを開発し、世界中のユーザーがファナックのロボットを使用できる環境を提供している。また、ファナックが開発するロボット制御装置は、パソコンを介さずにPythonプログラムを直接実行できるほか、Stream Motion（位置データ指令高速通信機能）を実装しているなど、多様な環境で利用しやすい開発を目指している^{*32}。ファナックのオープンプラットフォームを活用したフィジカルAIの事例として、言語指示に対して生成AIがPythonプログラムを自動生成し命令を実行する様子や、2本のアームで柔らかいケーブルを配線する様子、動く部品に対してリアルタイムでねじ締めを行う様子などが公開されている（図表I-2-2-5）^{*33}。

図表I-2-2-5 フィジカルAIロボット（ファナック）



（出典）ファナック^{*34}

^{*30} 産業技術総合研究所「実世界の困難作業自動化を目指したロボット基盤モデルの研究開発を本格始動」〈https://www.aist.go.jp/aist_j/news/pr20250123_2.html〉（2026年2月16日参照）

^{*31} 産業技術総合研究所「双腕ロボットAIの開発を支援するデータセットを無償公開」〈https://www.aist.go.jp/aist_j/press_release/pr2025/pr20250902/pr20250902.html〉（2026年2月16日参照）

^{*32} ファナック株式会社「FANUC Open Platform」〈<https://www.fanuc.co.jp/ja/product/robot/academia/>〉（2026年2月20日参照）

^{*33} ファナック株式会社「ファナックのオープンプラットフォーム活用事例」〈https://www.fanuc.co.jp/ja/product/robot/academia/case_study.html〉（2026年2月20日参照）

^{*34} 同上。

2018年創業の日本のロボット開発企業ugo株式会社は、点検専門モデルや警備・案内を実施するモデル等を開発しており、警備や案内業務を行う「ugoPro」では、2本のアームでカード認証やエレベーター操作などを行い、フロアをまたぐ移動が可能となっている。また、顔ディスプレイがついており、AIによる自動回答生成によって会話も可能で、多言語に対応しているほか、必要に応じて遠隔操作に切り替えて人が応答を行うことも可能となっている（2024年度第11回ロボット大賞優秀賞（ビジネス・社会実装部門）受賞）（図表 I-2-2-6）^{*35 *36}。

図表 I-2-2-6 警備・案内ロボット（ugo株式会社）

（出典）ugo株式会社「ugo Pro」^{*37}

このほか、Preferred Networksからスピノフしたロボット開発企業Preferred Roboticsが開発した小型搬送ロボット「カチャカ」は、独自のSLAM（Simultaneous Localization and Mapping）技術により、自己位置の変化や環境変化に対しても頑健に自己位置を推定することが可能で、また、深層学習においてもPreferred Networksのスーパーコンピュータと学習モデル自体の最適化を応用することで、ロボットの限られた計算能力・消費電力の中で、最大限の推論性能を引き出している。多種多様な環境でもぶつからずに最適な経路を自動で選択し物品の運搬が可能であることから、製造現場での輸送だけでなく、クリニックでの治療器具の輸送や機材の片づけ、飲食店での配膳など幅広いシーンで使用されている（2024年度第11回ロボット大賞総務大臣賞受賞）（図表 I-2-2-7）^{*38}。実装例として、KDDIは、本社オフィスにて「カチャカプロ」を導入し、複数フロアにわたる社内メール便搬送業務の一部を自動化しており、エレベーターとの自動連携によって、各フロア間の移動時には、カチャカが自らエレベーターを呼び出して乗降するという^{*39}。

図表 I-2-2-7 小型輸送ロボット（Preferred Robotics）と実装シーン（KDDI本社）

（出典）Preferred Robotics^{*40}

*35 ugo株式会社「business」〈<https://corp.ugo.plus/business/>〉（2026年3月5日参照）

*36 ugo株式会社「ugo Pro」〈<https://ugo.plus/products/ugo-pro/>〉（2026年3月5日参照）

*37 同上。

*38 株式会社Preferred Robotics「Preferred Robotics」〈<https://www.pfrobotics.jp/>〉（2026年3月5日参照）

*39 株式会社Preferred Robotics「KDDI本社に自律搬送ロボット「カチャカプロ」を導入」〈<https://kachaka.life/posts/press-kddi-kachaka-mail-robot/>〉（2026年3月19日参照）

*40 株式会社Preferred Robotics「搬送ロボット“カチャカプロ” | 小型で安価なAMRで効率的な搬送自動化」〈<https://kachaka.life/>〉（2026年3月19日参照）