

## ロボットの AI モデル開発で 数十テラバイト級データを受け渡すデータハブ ネットワーク性能、データ耐久性、価格で 選ばれた Wasabi Hot Cloud Storage

### 概要

一般社団法人 AI ロボット協会 (AI Robot Association、以下、AIRoA) は、ロボットと AI の融合を通じてロボット開発技術を革新し、社会や暮らしの中でロボットが活躍できるよう「ロボットデータエコシステム」の構築をめざす非営利の団体です。大学や産業界の R&D 機関など、ロボット・AI 研究に関わる多くの組織が運営に参加しています。

近年、AI は大規模言語モデル (LLM) や視覚言語モデル (VLM) などの基盤モデルを軸に急速に進化し、大量のデータと計算資源を背景に大企業主導で開発が進み、データ・計算資源・ユーザーフィードバックが一部の企業に集中し、新規参入が難しくなっています。潮流はロボティクスにも及び、動作データを学習するロボット基盤モデルの研究が進展する一方、汎用ロボットではデータ不足と共有の遅れが課題となっています。AI を活用したロボティクス分野の競争力強化のためには、開かれたデータ基盤の整備が急務であるとの強い思いから協会が設立されました。

### 課題

- › 産業界で共有可能なデータハブが必要
- › 数十テラバイト級のデータを受け渡し
- › データ欠損のないことが必須
- › 非営利団体のため予算には上限

「ロボットデータエコシステム」の構築に当たって、AIRoA では具体的に 2 つのプロジェクトを計画しました。Phase 1 では、VLM・VLA 構築に向けて、同協会に提供された生活支援ロボットを利用し、そこから生成されたデータを活用する PoC を推進。一方、Phase 2 では、小売・製造業・物流などの分野でユースケースデータを取得し、ベースモデルを構築、事後学習を行って社会実装を行います。ここで必要なのがベースモデル構築のためのデータ基盤です。アーキテクチャは次のように設計されました。

ロボットデータエコシステムを支えるデータハブ基盤の確立に成功

データ容量が増え続ける中でも安定稼働

信頼性の高いデータ耐久性を発揮

決められた予算の中で継続活用が可能

#### お客様名

- › 一般社団法人 AI ロボット協会

#### 業種

- › 研究機関

#### ユースケース

- › AI データ活用



使える予算には上限があるため、不確実性の高い費用がかかるものはあまり入れたくありません。その点、データ転送料と API 費用のかからない Wasabi は向いているなと思いました。



東京大学大学院 工学系研究科  
技術経営戦略学専攻  
松尾・岩澤研究室  
学術専門職員  
(AIRoA より委託を受けたプロジェクトに従事)  
野海 芳博 氏

まず、ロボットから動作データを収集し、これをデータハブにアップロードします。次に、そのデータを取り出して学習用データとして変換、パッケージ化して再びデータハブに戻します。そして、今度はこれを GPU クラスタへ分配して学習に用います。さらに、学習で確立したモデルに関して汎用性という観点で評価を行います。ここで中核となるのがデータハブで、Raw data の保存と配布、学習データの保存、GPU クラスタへの分配、実験環境の保存と配布、評価用のデータセット保存と配布と、まさに扇の要となる役割を果たします (図 1)。

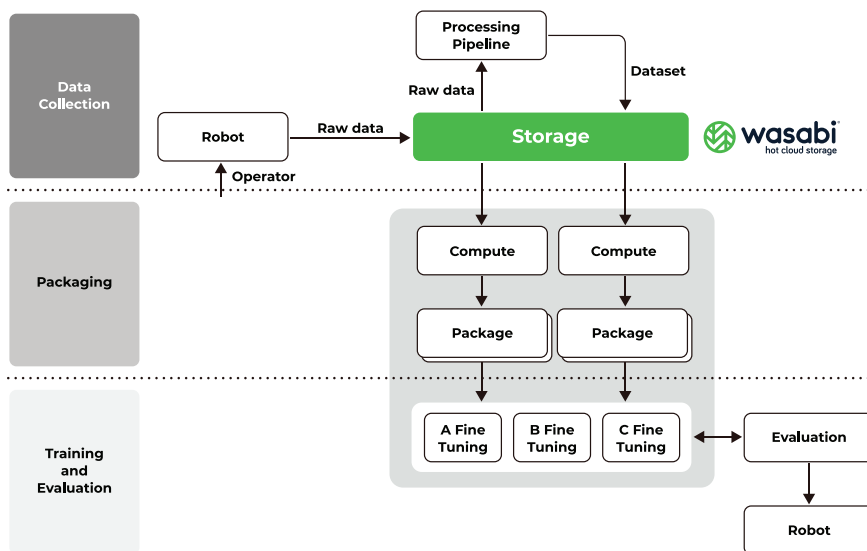


図 1 開発・学習・検証するための全体アーキテクチャ

そのため、データハブには高度な要件が設けられました。AIRoA にインフラ構築支援で携わる、東京大学大学院 工学系研究科 技術経営戦略学専攻 松尾・岩澤研究室 学術専門職員 野海 芳博 氏は、次のように語ります。

「要件は大きく 3 つ挙げました。1 つめは、数十テラバイト級のデータを常にアップロードし続けるため、安定したネットワークでつながれたストレージであることでした。2 つめは、その大量データがアップロード、保存、ダウンロード時に欠損のリスクがないことです。欠損すると、学習用データとして使えなくなってしまいます。ですから、高いデータ耐久性が保証されていることも重要でした。3 つめが、コストパフォーマンスが高いことです。AIRoA は非営利団体ですから、決められた予算の中で工面する必要がありました」

“

要件は大きく 3 つ挙げました。1 つめは、数十テラバイト級のデータを常にアップロードし続けるため、安定したネットワークでつながれたストレージであることでした。2 つめは、その大量データがアップロード、保存、ダウンロード時に欠損のリスクがないことです。欠損すると、学習用データとして使えなくなってしまいます。ですから、データ耐久性が高いことも重要でした。3 つめが、コストパフォーマンスが高いことです。AIRoA は非営利団体ですから、決められた予算の中で工面する必要がありました。

“

ここまで何も問題が起きていません。それが最大の成果だと考えています。この AI ロボット基盤モデルの構築というのは、誰も試みたことのない先進的な取り組みで、何が起こるかわからない側面がありますから、少なくともインフラ部分には神経を使いたくありません。そうした中で Wasabi は、規模を拡大し続けるという動的な使い方をするにも関わらず、何事もなく動作しており、データの正しさにも信頼が置き、AIRoA の活動に貢献しているといえます。

## ソリューション

そうした中で、最有力候補として挙げたのが Wasabi Hot Cloud Storage でした。松尾研で総合商社から紹介されたことを野海氏が記憶していました。「なぜ Wasabi という名前？」と聞いて、「Hot なデータの格納庫だから」と返ってきた答えが印象的だったそうです。実は、当初にはオンプレミスのストレージや別のオンラインストレージも検討されました。Phase 1 では AIRoA 内で活動がほぼ完結するため、オンプレミスストレージを GPU クラスタに隣り合わせて設置することも考えられたのですが、将来的にロボットデータエコシステムの構築をめざす AIRoA にとって、どこからでもデータがアップロード／ダウンロードできるクラウドにデータハブを置く方がよりよい方法でした。学習用データへの変換には AWS を利用することから、Amazon S3 も選択肢の一つでしたが、構築するデータハブはデータの受け渡しが頻繁に行われ、しかもデータ容量は膨大で、正確に予測する事が難しい状況でした。データ転送料のかかるクラウドストレージでは、億円単位でコストが増大する試算も出ました。

野海氏はこう語ります。

「使える予算には上限があるため、不確実性の高い費用がかかるものはあまり入れたくありません。その点、データ転送料と API 費用のかからない Wasabi は向いているなと思いました」

今回のプロジェクトでは、複数クラウド上の GPU クラスタへのデータ転送に加え、データのアップロード／ダウンロード、バックアップ時のデータの差分チェックやオブジェクトのリストアップなどで API 利用も非常に多く発生します。ここに課金しない Wasabi Hot Cloud Storage を利用すれば、コストマネジメントは格段に容易になります。

これらをふまえて、Amazon S3 互換であるという Wasabi が“Amazon S3 のように使えるか”という観点で検証を行いました。それはもう試し使いというレベルではなく、実際にロボットから収集された Raw data をアップロードするなど、すでに Phase 1 の活動の中に組みこむことを想定した利用だったといえます。

野海氏自身、データをアップロードしてスピードが出ていることを確認し、活動メンバーからも「普通に使えている」というフィードバックが得られたことから、導入を正式に決定。ネットワーク性能とデータ耐久性、コストパフォーマンスの3点で十分に要件を満たしていたことが決め手となりました。

“

いろいろな拠点からデータをアップロードするとなると、最初に思い浮かぶのは Amazon S3 です。しかし、S3 はデータの取り出し（データ転送量）に応じた費用が発生するため、大量データを頻繁にダウンロードする用途ではコストが大きくなります。そこにコストがかからないものにしないと、AIRoA が設立時に掲げたデータ公開が非現実的なものになってしまいます。

“

活動では、アップロードしたデータをまたダウンロードして同じ状況を再現する場面が結構あります。そのため、データを動かしたときに欠損がないという再現性が非常に重要です。データがきちんと保存され、それを戻してもきちんと動くということも研究の一部なのです。「こういう成果が出ました」と発表したのに、再現しようとしたらできなかった、それでは話になりません。

“

今後、開発予定の AIRoA のプラットフォーム経由で、ロボット動作に関する情報や方法やノウハウ、ロボットのアクションなどさまざまな交換が行われ、そこでは“掛け合わせ”が起きてくるでしょう。ロボットが多分アクションを勝手に生成し始めるとはいますが、単に自動化だけではない、次のひらめきといえるような何かが起きるはず。このプラットフォームは、そのような偶発的な何かが起きる場所だと考えています。

## 結果

2025年4月、1年のプロジェクトとしてPhase 1が始まりました。ここでは計画が提出された中で2チームのモデルが採用され、それぞれに冒頭で言及した流れで活動を行いました。2026年2月時点ですべてのチームのモデル作成と提出は完成しており、現在は評価を待っている段階です。一方、Phase 2も2025年10月にスタート、こちらは4年間のプロジェクトで、両腕を備えた移動型マニピュレータ・ヒューマノイドを使って、研究室および工場・店舗など実際の事業拠点でデータを収集し始めています。

Phase 1はデータ収集にかかる時間が数万時間ですが、Phase 2では数十万時間になる見込みです。現状、Wasabiに保管されているデータボリュームは2.2PBですが、容量は活動が進むにつれて今後も増え続けていきます。

野海氏は、Wasabiを導入した効果について、次のように述べています。

「ここまで計画に影響があるような問題は何も起きていません。それが最大の成果だと考えています。このロボット基盤モデルの構築というのは、大規模なロボットデータ基盤としては、まだ世界的にも事例の少ない取り組みで、何が起こるかかわからない側面がありますから、少なくともインフラ部分には神経を使いたくありません。そうした中でWasabiは、規模を拡大し続けるという動的な使い方をしているにも関わらず、何事もなく動作しており、データの正しさにも信頼が置き、AIRoAの活動に貢献しているといえます」

AIRoAの描く理想の未来像は、確立したモデルがさまざまな産業界を横断して活発に利用され、それを搭載したAIロボットが飛躍的に実用性を高めていくことです。一例を挙げると、物流の運搬動作で蓄積した知見が、店舗の陳列動作で活かされるといった具合です。

「ロボットはAIと組み合わせることで、文脈を理解した動きができるようになります。たとえば、コップを置くという動作一つを取っても、状況に合わせて『やさしく置く』『人の腕の外ではなく、人の前に置く』など、さまざまなパターンが考えられます。AIロボットはロボットの可能性を大きく広げ、日本の国際競争力を高めることにもつながります」(野海氏)

そのような配慮ができる賢いロボットは、きっと世界中で歓迎されることでしょう。そうしたロボットの“核心”部分が、今Wasabi上で生み出され続けています。

“

AIは、データを収集し、整備し、学習させ、活用するというテクノロジーです。企業でAIモデルを開発するのであれば、企業内で生まれたデータをすべて取っておいた方がいいと思います。それは企業内でしか生めないものであり、捨てたデータは戻ってきません。データの使い道はいくらでもあります。AIで何かするのなら、データはすべて保存することをお勧めします。そういう意味では、データの保存体制や基盤作りを最初から十分考えておくことは重要です。

“

個人的には、いつかロボットを相棒として人類が今まで行ったことのない場所に行ってみたいですね。

