

ROS 2 を使用した汎用自律エージェント プラットフォームの開発

株式会社サイバーエージェント AI Lab
Agent Development Team
Research Engineer
兵頭 亮哉



CyberAgent **AI Lab**

自己紹介

兵頭 亮哉 Hyodo Katsuya

株式会社サイバーエージェント
AI事業本部 AI Lab

Agent Development Team
Research Engineer

マルチモーダル音声対話プラットフォームの開発



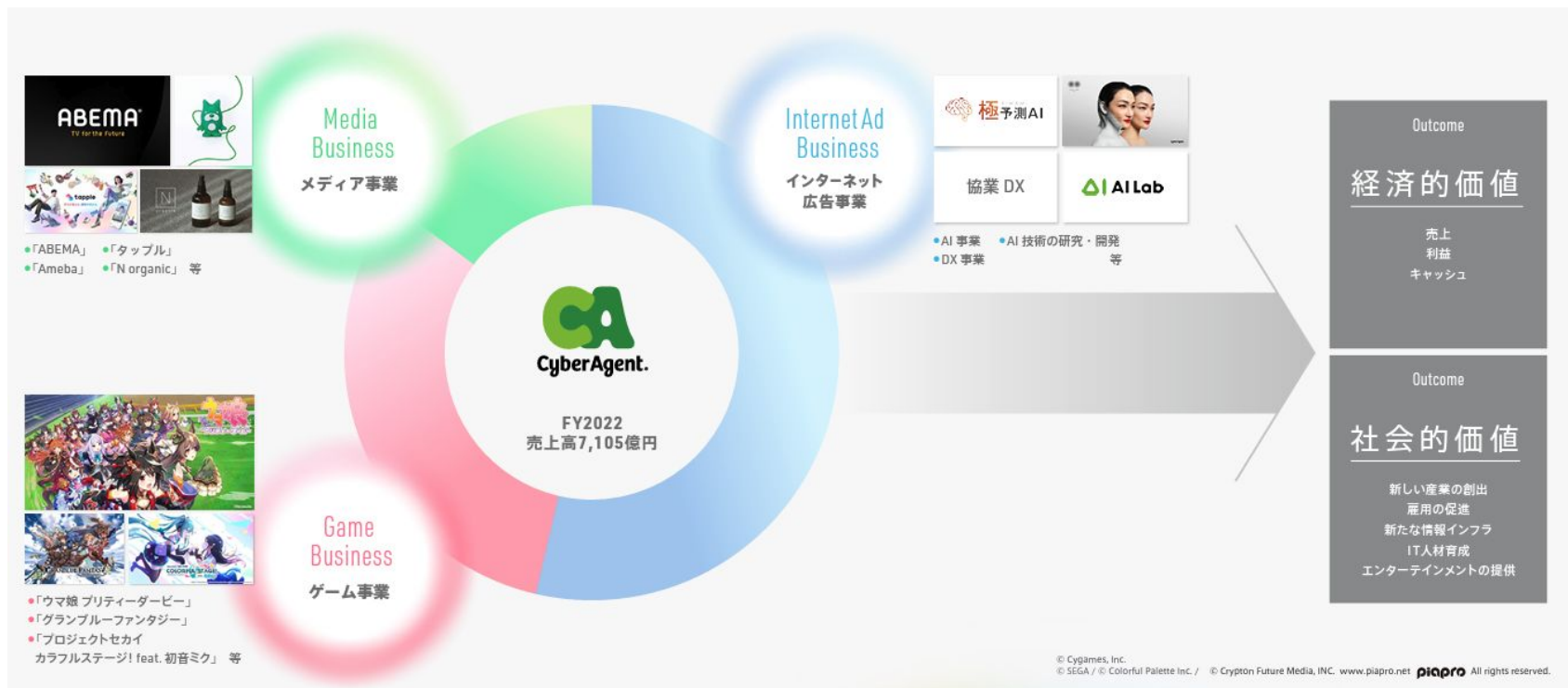
@PINTO003091



@PINTO00309



サイバーエージェントのビジネスモデル



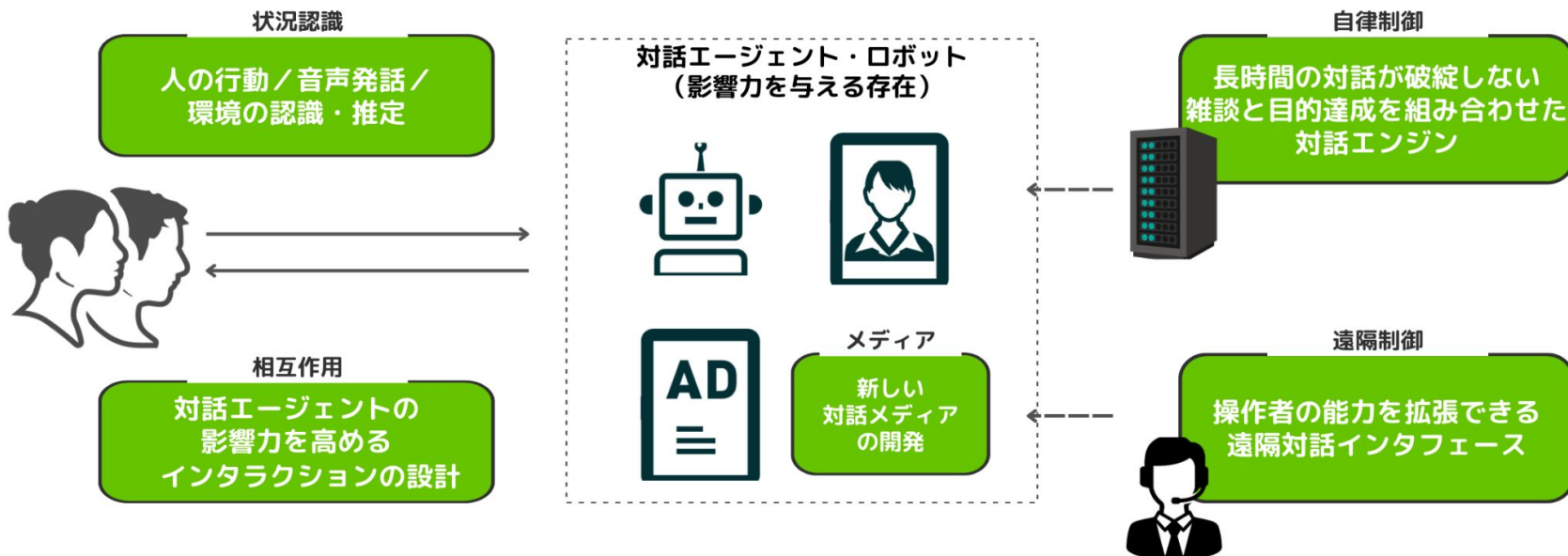
AI Lab

- 2016年設立された研究組織
 - 論文投稿・事業貢献・社会貢献をポリシーとした研究組織
- 現在13チームによって構成
 - 広告画像生成、自然言語処理、CG、音声、接客対話、強化学習、経済、等々
- 正社員67名（Research Scientist & Research Engineer）2023年8月現在
 - 博士取得者63%、社会人博士を含むと72%

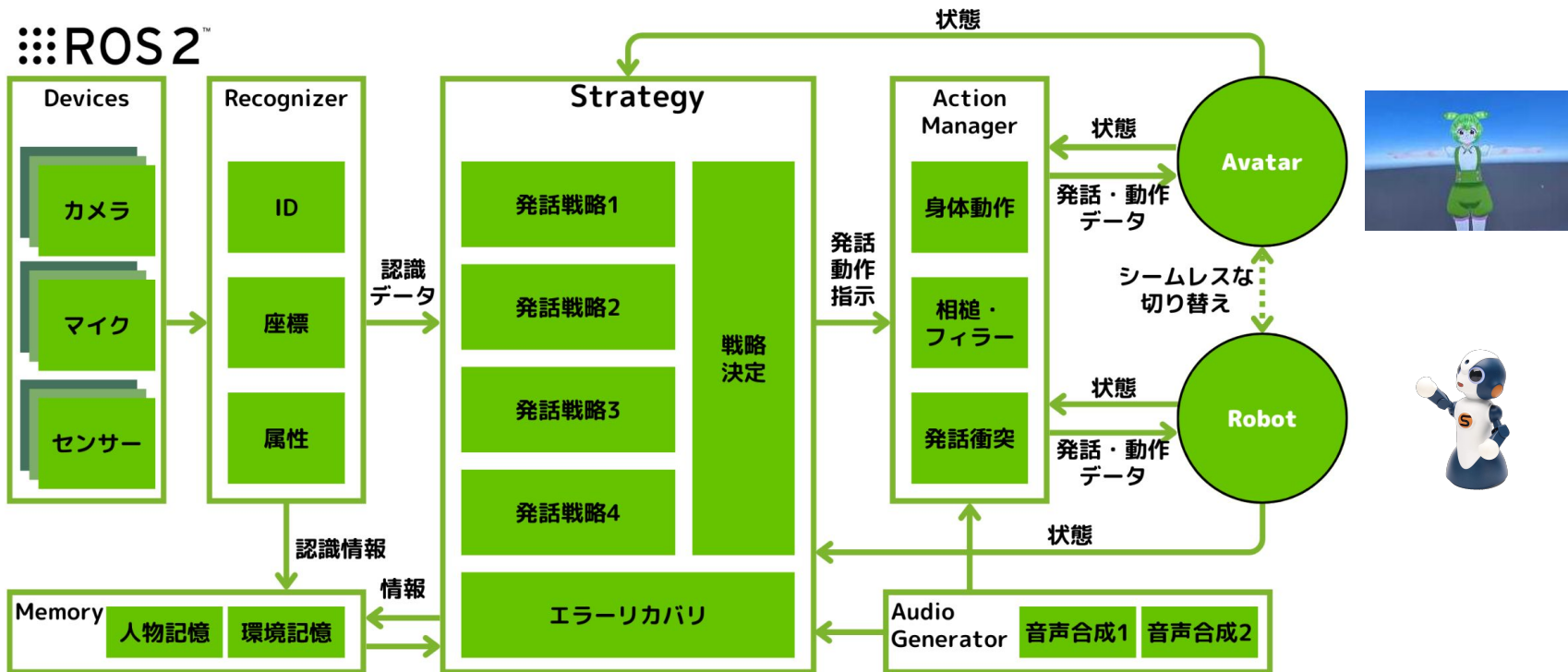


研究・開発チームのテーマ

ヒトが信頼したくなる対話エージェント・ロボットの研究開発

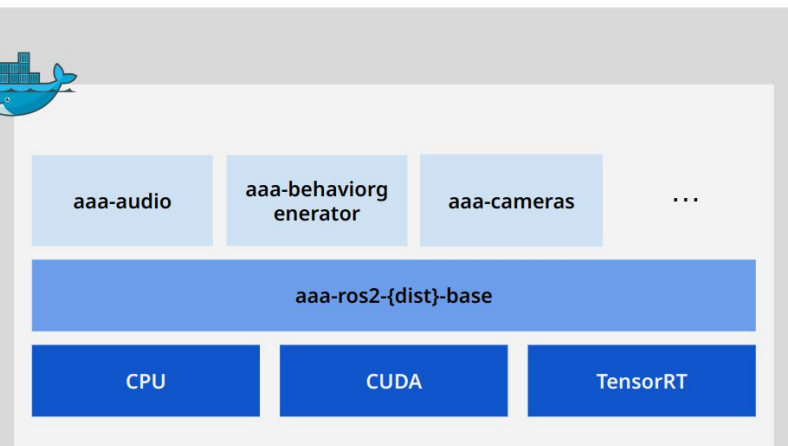
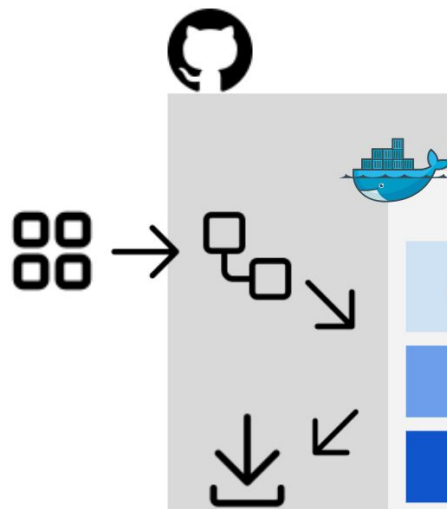


ROS 2 によるプラットフォーム開発



HW / SW 構成

- ROS 2
- 単眼RGBカメラ
- RealSense, ZED2
- スマートスピーカー
- 指向性マイク
- 環境センサー
- GPU
- CUDA
- TensorRT
- 2Dエージェント
- ロボット

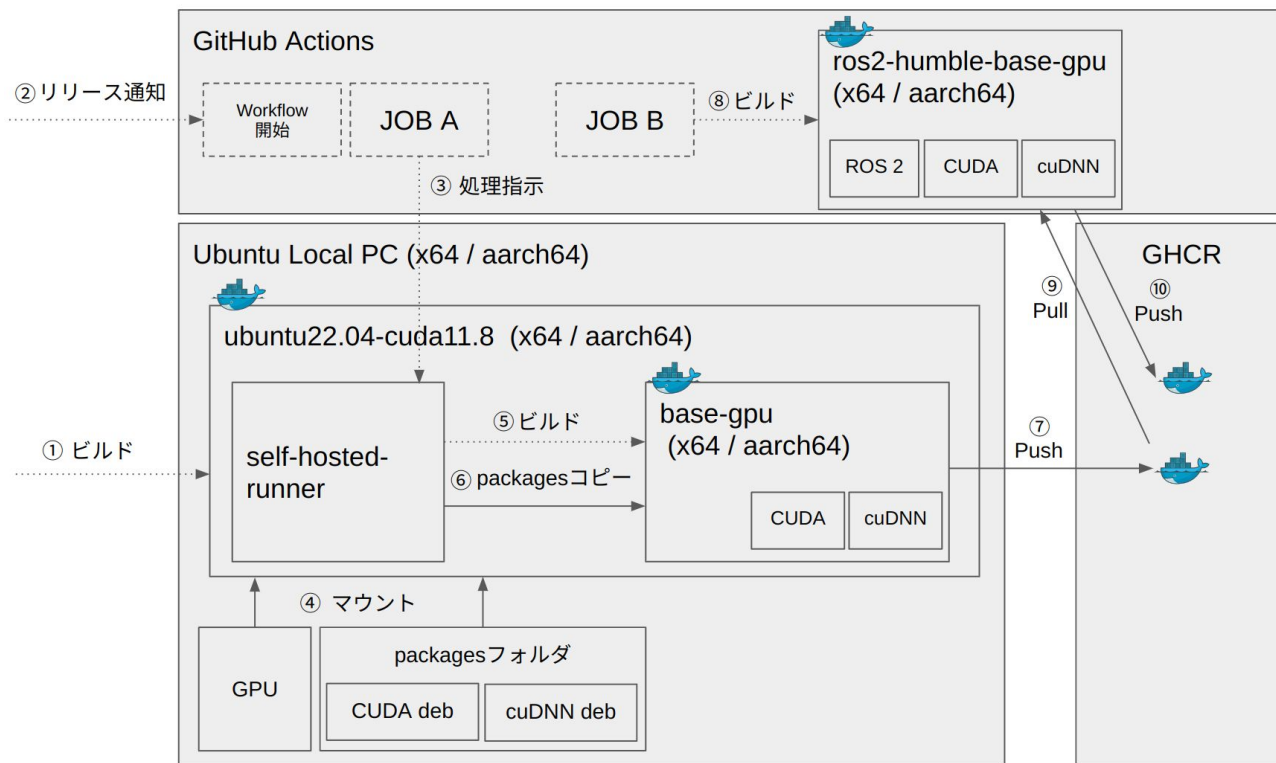


CI/CD環境の構築

- ・ ベースイメージの設計
- ・ CUDA / TensorRT / ONNX Runtime
 - From Scratch
 - GitHub Self-Hosted Runnerの活用
- ・ センサ固有の設定
 - カメラ、マイク、センサ
- ・ Multi-Arch
 - x64 / Arm



CI/CD環境の構築



- ・ テスト実行
- ・ ベースイメージのビルド
- ・ ROS 2 導入
- ・ CUDA 導入
- ・ TensorRT 導入
- ・ ONNX Runtime のビルド
- ・ Docker Image のストック
- ・ 開発用 devcontainer への利用
- ・ 配布用 so / wheel ビルド

CI/CD環境の構築



GHCR (GitHub Container Registry)

- Azure? の下りが不安定、エラーが頻発してセッションが切断される
- 10GB を超えると上り・下り共に明らかに遅い



Docker Hub

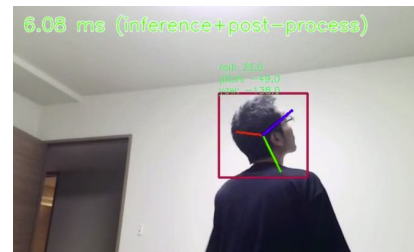
- とても安定している
- 10GB を超えても速い
- ベースイメージのみ Docker Hub を利用

状況認識・戦略決定・音声動作モジュール

- ・ オブジェクトディテクション
- ・ ReID
- ・ 骨格検出
- ・ 顔向き 6 D 推定
- ・ 注視方向推定
- ・ 立ち止まり検出
- ・ アクション検出
- ・ etc...



2023年09月26日現在
419種類

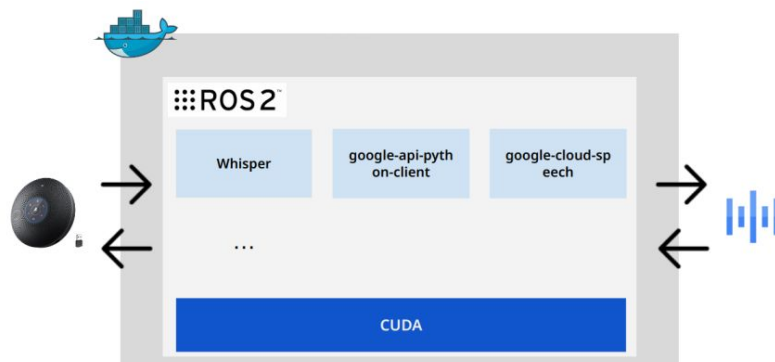


Recognizer

ROS 2

状況認識・戦略決定・音声動作モジュール

- LLM
- Faster-Whisper
- VAD
- etc...



Audio

ROS 2

マルチエージェント制御

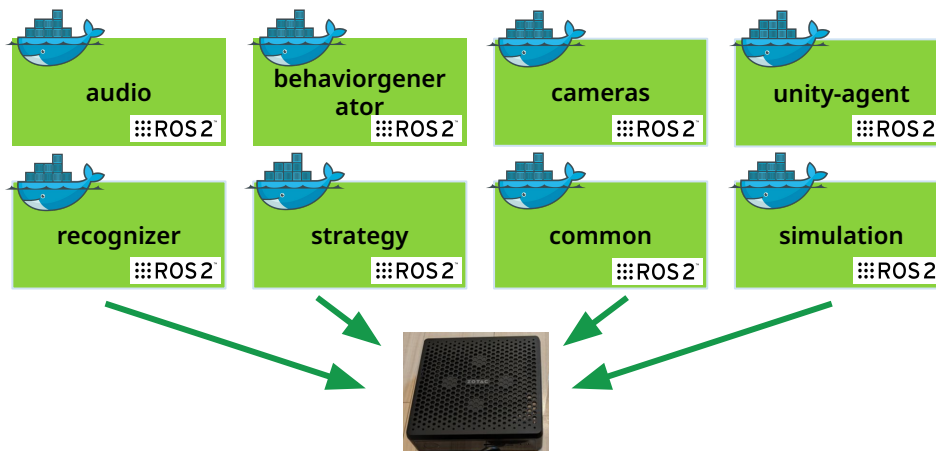
- ・ 複数体・複数種類のエージェントを同時制御



ROS 2 マルチエージェント環境構築の敷居を下げる

- ・ シェルスクリプト実行（有線LAN環境で約15分）

```
sudo chmod +x environment_build.sh \  
&& time ./environment_build.sh
```



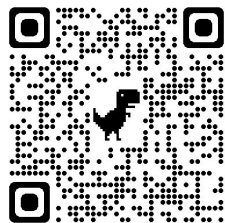
マルチモーダル認識・マルチエージェント制御の課題

1. ロボットの音声発話がそのままロボットのマイクに入ってしまう問題。
2. 単位時間あたりの命令数が多い場合、無線などの貧弱な通信環境でスタックする問題。
3. 2DアバターのようなCPU、GPUともに負荷の高いアプリケーションとROSを同一PCに同時に起動するとレンダリングの負荷でCPU処理がスタックする問題。
4. 異なるフレームワークを組み合わせる前提による座標系の統一問題。
5. ROS上に音声データをチャンクで流して再構成するのが難しい問題。
6. 音声処理系の遅延問題。
7. 人っぽさを再現するために必要な機能が多すぎてリソースが足りない問題。
8. リソース分散のために端末を増やすとコストが高過ぎる問題。
9. Docker Image 大きすぎ問題。



さいごに

- エンジニアも研究者も募集中です。



エンジニア向け



研究者向け

