完全自動運転に向けた身体性を持つマルチモーダル基盤モデルの開発

実施者	Turing株式会社
事業概要	完全自動運転の実現に向けて、視覚・言語・行動を統合した「身体性」を持つマルチモーダルAIモデルを開発する。

本事業では、視覚・言語・行動を統合して車両の運転行動を生成できる「**身体性を持つAIモデル**」の開発を目指した。日本語に特化した視覚-言語モデルの構築と、自社走行車両によるセンサデータを活用した大規模三次元行動データセットを整備する。これらを組み合わせ、**空間理解や運転判断が可能なマルチモーダルモデル**を開発、シミュレータ上で運転タスクの評価を実施する。学習済みモデルやデータセットは順次公開する他、将来の自動車メーカーとの技術連携やADAS領域での社会実装を見据えた基盤技術を構築する。

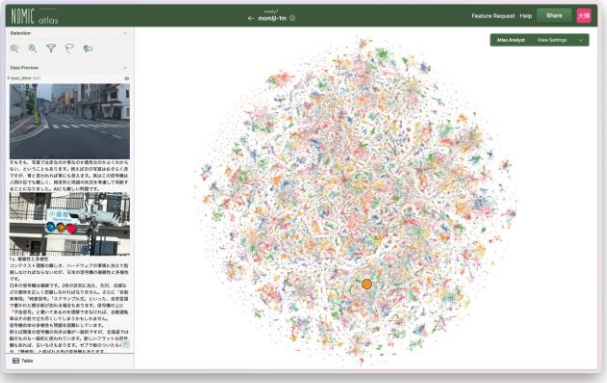
社会実装イメージ

カメラ＋AIによる自動運転機能の実現に向け、自動車メーカーおよびサプライヤーと連携する。東京都内をはじめとする複雑な公道において、言語理解を含むEnd-to-Endな運転モデルの走行を実現する。将来的にはあらゆる場所・状況で人間の運転を代替する**完全自動運転**を、市販される乗用車に搭載する。

事業成果

日本語視覚・言語モーダルに特化した大規模データセット「MOMIJI」を構築し、WebスケールのHTML文書から収集・整形した2.49億枚の画像と日本語テキストを、インターリーブ形式で整理した。このデータを用い、NVILAアーキテクチャによる**視覚-言語モデル「Heron-NVILA」**シリーズ（1B～33B）を学習し15BモデルでHeron VLM Leaderboardで4.84を達成。70B以下のオープンモデルとしては国内外最高の性能を記録した。さらに、東京都内で3,500時間以上の走行データを取得し、**100時間超の視覚-言語-行動データセット「STRIDE-QA」**を整備。交通オブジェクトへ三次元空間情報を付与し、自動記述アルゴリズムにより自然言語への変換を実現した。これらのデータを用いて運転状況を予測可能なマルチモーダルモデルを構築した。また、シミュレータ上において自律走行タスクを言語モデルで行動までを一貫して学習させ、最終的にCARLA AD Leaderboardで6.53を記録、**マルチモーダルモデルの自動運転への応用を定量的に実証**した。構築したモデルやデータは、Hugging Faceやテックブログを通じてすべて公開し、実験用のiOSアプリも一般公開した。なお、今回の成果の基盤モデルを用いた事業化に向けたPoCも開始している。

STRIDE-QA



MOMIJI