Exploring Multimodal AI *beyond* Vision and Language

Multimodal AI Community UK

Introduction

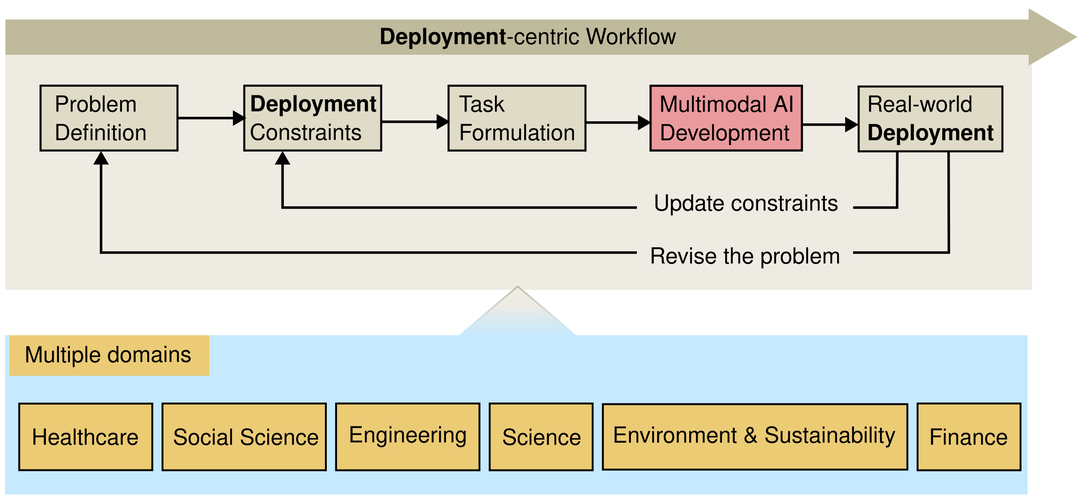
Motivation

- Human developed methods to gather and integrate information to understand our world.
- Multimodal AI improves machine understanding and prediction by integrating and processing diverse data sources and domain knowledge.
- Current research mainly focuses on vision and language within specific domains.

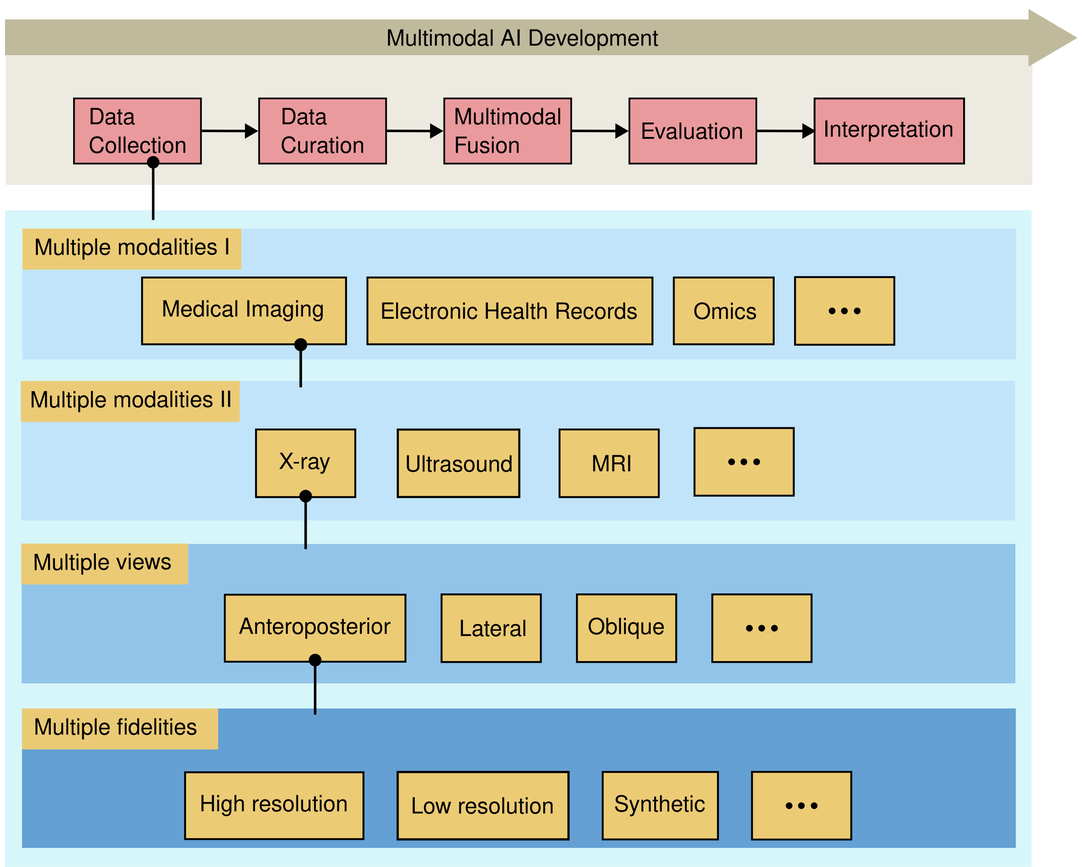
Contributions

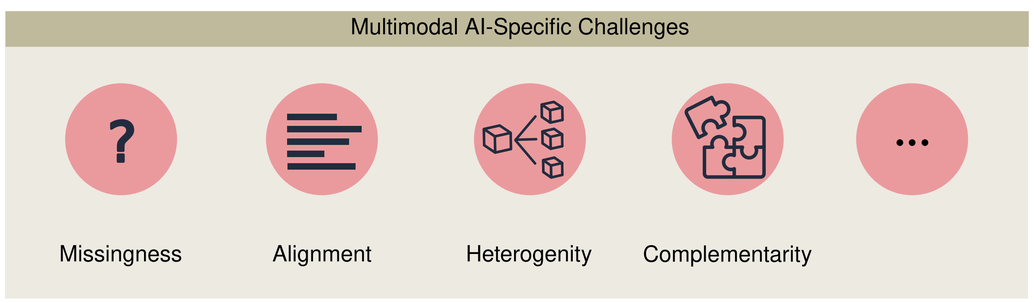
- **Multimodal AI-specific** Problems: Missing modality, cross-modality alignment, heterogeneity, complementarity, ...
- **Beyond** vision and language, and beyond specific domain.
- **Deployment-centric** Perspective: Early consideration of deployment constraints.

Deployment-centric Workflow

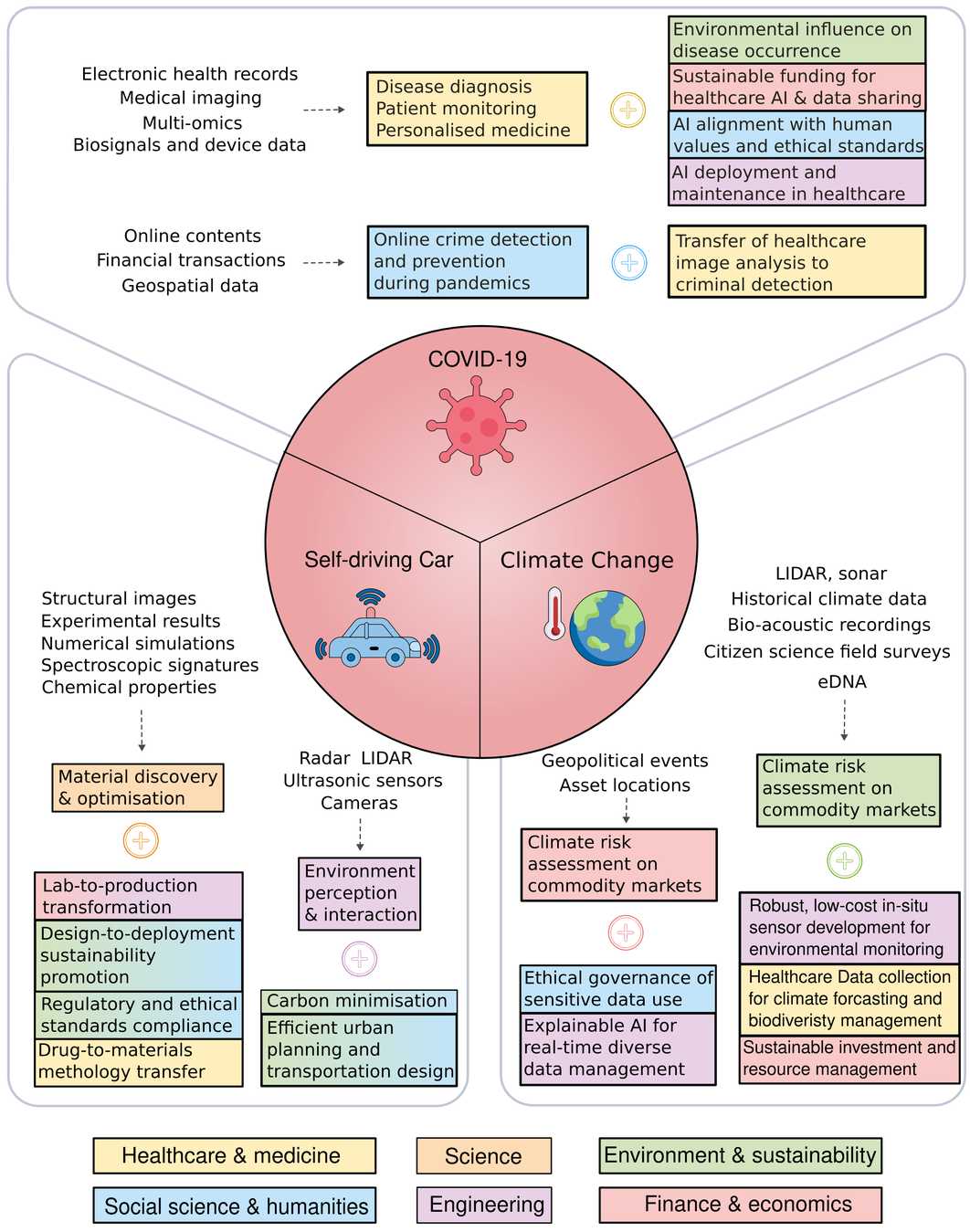


Multimodal AI Development





Use Cases



Benchmark Datasets

Datasets	Year	Country	Data Modalities	Access	Sample Size	Applications
Quilt-1M	2022-2023	U.S.	Text, image	Open	1,000,000	Histopathology tasks
TCGA	2006-2018	U.S.	Omics	Open	>20,000	Cancer diagnosis Patient stratification
Houston Human Brain Activity	2023	U.S.	Wearable device data, EEG	Open	30	Regulation of brain cognitive states
IGDD Project	2020-2021	U.S.	Text, image, numerical metadata	Restricted	4,181,970	Concersation classification
EASE	2016-2017	U.S.	Text, image, audio, numerical metadata	Restricted	>800,000	Emotion recognition, mental health classification
Trafficking-10k	2016-2017	Canada, U.S.	Text, image	Restricted	10,000	Human trafficking detection
NuScenes	2020-2024	U.S.	LiDAR, radar, scenes, image, video, graph	Open	>1,000,000	Autonomous vehicle perception
DAIR-V2X	2021	China	LiDAR, image, pointcloud	Open	>150,000	Autonomous vehicle perception
Rank2Tell	2024	U.S.	Text, image, video, GPS data	Open	116	Urban traffic importance understanding
Materials Project	2011	Multiple	Text, image, graph, sensor data	Open	150,000	Battery property search, Stability prediction
ESPRIT	2020	U.S.	Text, video	Open	2441	Physics reasoning explaination
BubbleML	2023	Multiple	Symbolic, numerical data	Open	79	Boiling processes modelling
iNaturalist	2012-present	U.S.	Species taxa, images, timestamps, coordinates	Open	85,727,262	Biodiversity modelling Green Finance
BioAcoustica	2016-2019	U.K., U.S., Spain	Timestamps, locations, air movement, etc.	Open	315,888	Systematics biogeography Automated species identification
International Soil Moisture Network	1952-2022	Multiple	Numerical metadata (soil moisture level, temperature, precipitation level etc.)	Open	73	Weather prediction Flood forecasting etc.
Earnings conference calls	2019	U.S.	Text, audio	Open	500	Stock volatility prediction
Monetary Policy Calls (MPC)	2022	Multiple	Text, images, audio	Open	464	Stock market indices, gold price, currency exchange rates and bond prices prediction
Markets Data	?	?	Text, audio, video	Open	17	Daily charts with market movement

Open access Restricted access

Outlook