Exploring Multimodal Al beyond Vision and Language

Multimodal AI Community UK

Introduction

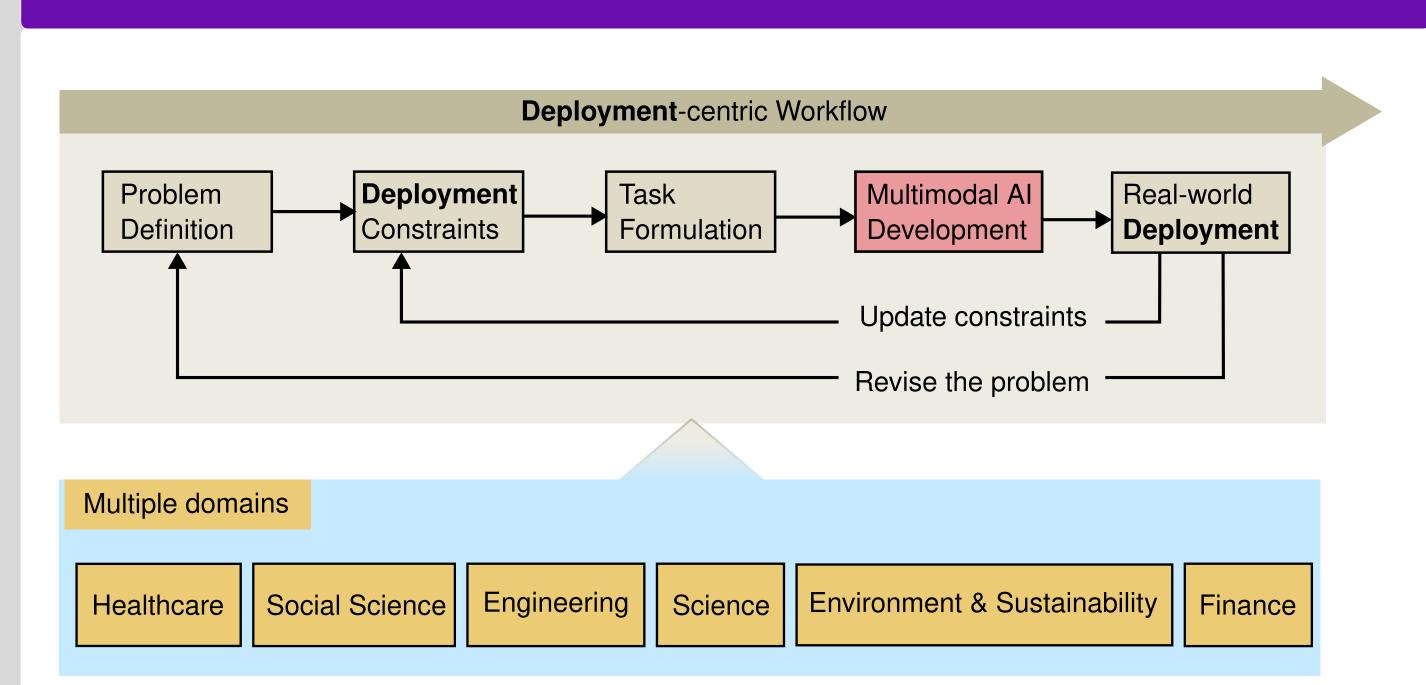
Motivation

- 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.

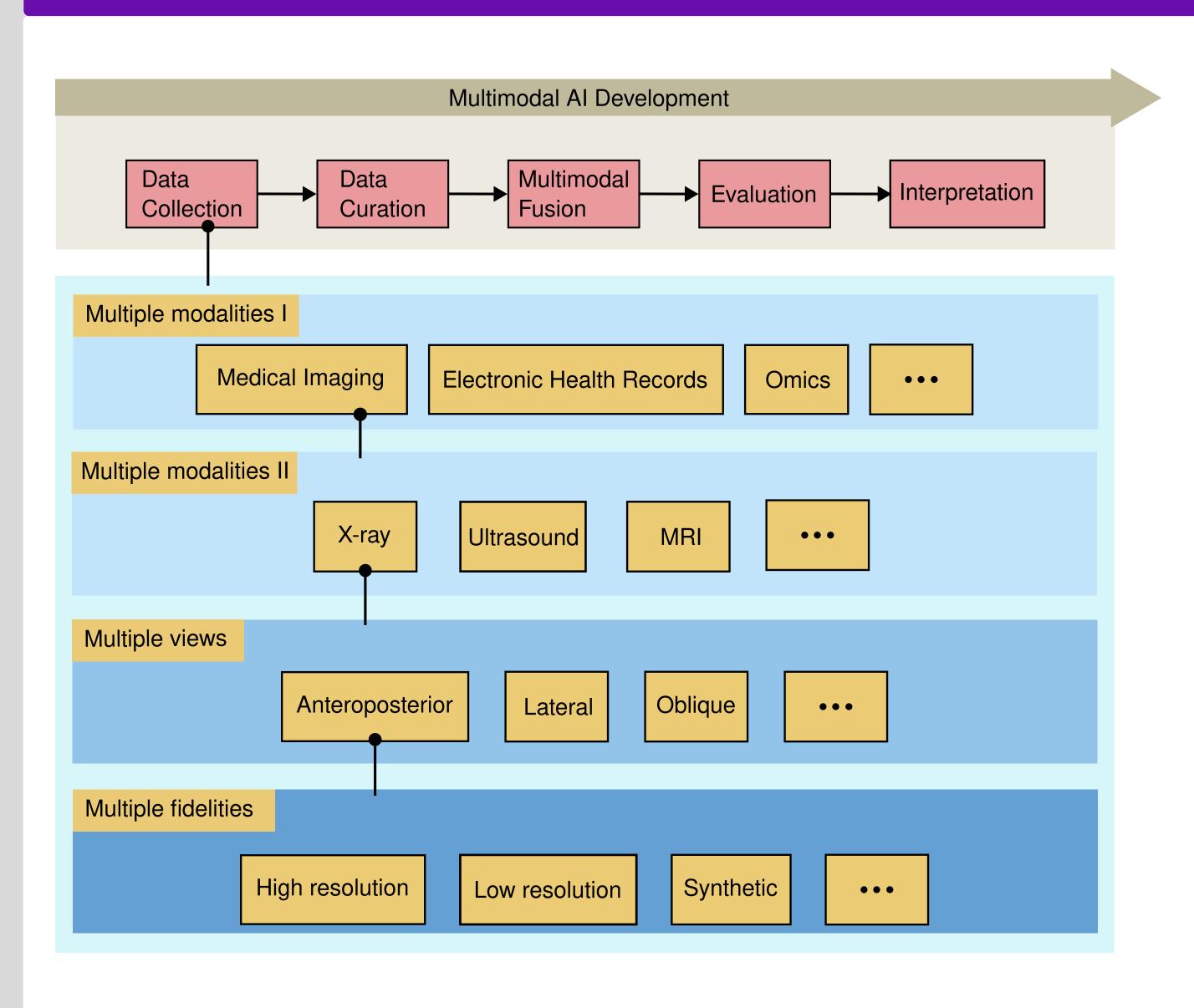
Contributions

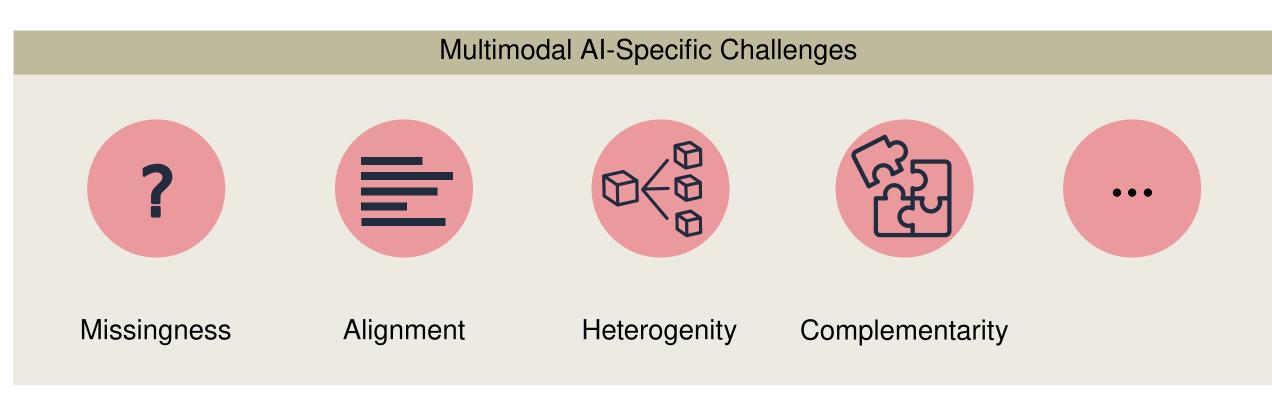
- Multimodal Al-specific Problems: Missing modality, cross-modality alignment, heterogeneity, complementarity, ...
- Beyond vision and language, and beyond specific domain.
- **Deployment-centric** Perspective: Early consideration of deployment constraints.

Deployment-centric Workflow



Multimodal Al Development





Use Cases Environmental influence on disease occurrence Electronic health records Sustainable funding for Disease diagnosis Medical imaging healthcare AI & data sharing Patient monitoring Multi-omics Al alignment with human Personalised medicine Biosignals and device data values and ethical standards Al deployment and maintenance in healthcare Online contents Online crime detection Transfer of healthcare Financial transactions ----> and prevention image analysis to during pandemics criminal detection Geospatial data COVID-19 Self-driving Car Climate Change LIDAR, sonar Structural images Historical climate data Experimental results Bio-acoustic recordings Numerical simulations Citizen science field surveys Spectroscopic signatures eDNA Chemical properties Radar LIDAR Geopolitical events Climate risk Material discovery Ultrasonic sensors **Asset locations** assessment on & optimisation Cameras commodity markets Climate risk assessment on Environment Lab-to-production commodity markets perception Robust, low-cost in-situ transformation & interaction sensor development for Design-to-deployment environmental monitoring sustainability Healthcare Data collection Ethical governance of promotion for climate forcasting and sensitive data use **Carbon** minimisation Regulatory and ethical biodiveristy management Explainable AI for standards compliance Efficient urban real-time diverse Sustainable investment and Drug-to-materials planning and data management transportation design resource management methology transfer Healthcare & medicine Environment & sustainability Science Engineering Finance & economics Social science & humanities

Benchmark Datasets

Datasets	Year	Country	Data Modalities	Access	Sample Size	Applications
Quilt-1M	2022-2023	U.S.	Text, image	0	1,000,000	Histopathology tasks
TCGA	2006-2018	U.S.	Omics	0	>20,000	Cancer diagnosis Patient stratification
Houston Human Brain Activity	2023	U.S.	Wearable device data, EEG	0	30	Regulation of brain cognitive states
IGDD Project	2020-2021	U.S.	Text, image, numerical metadata	0	4,181,970	Concersation classification
EASE	2016-2017	U.S.	Text, image, audio, numerical metadata	0	>800,000	Emotion recognition, mental health classification
Trafficking-10k	2016-2017	Canada, U.S.	Text, image	0	10,000	Human trafficking detection
NuScenes	2020-2024	U.S.	LiDAR, radar, scenes, image, video, graph	0	>1,000,000	Autonomous vehicle perception
DAIR-V2X	2021	China	LiDAR, image, pointcloud	0	>150,000	Autonomous vehicle perception
Rank2Tell	2024	U.S.	Text, image, video, GPS data	0	116	Urban traffic importance understanding
Materials Project	2011	Multiple	Text, image, graph, sensor data	0	150,000	Battery property search, Stability prediction
ESPRIT	2020	U.S.	Text, video	0	2441	Physics reasoning explaination
BubbleML	2023	Multiple	Symbolic, numerical data	0	79	Boiling processes modelling
iNaturalist	2012-present	U.S.	Species taxa, images, timestamps, coordinates	0	85,727,262	Biodiversity modelling Green Finance
BioAcoustica	2016-2019	U.K., U.S., Spain	Timestamps, locations, air movement, etc.	0	315,888	Systematics biogeography Automated species identification
International Soil Moisture Network	1952-2022	Multiple	Numerical metadata (soil moisture level, temperature, precipitation level etc.)	0	73	Weather prediction Flood forecasting etc.
Earnings conference calls	2019	U.S.	Text, audio	0	500	Stock volatility prediction
Monetary Policy Calls (MPC)	2022	Multiple	Text, images, audio	0	464	Stock market indices, gold price, currer exchange rates and bond prices predic
Markets Data	?	?	Text, audio, video	0	17	Daily charts with market movement

Outlook

Restricted access

Open access

