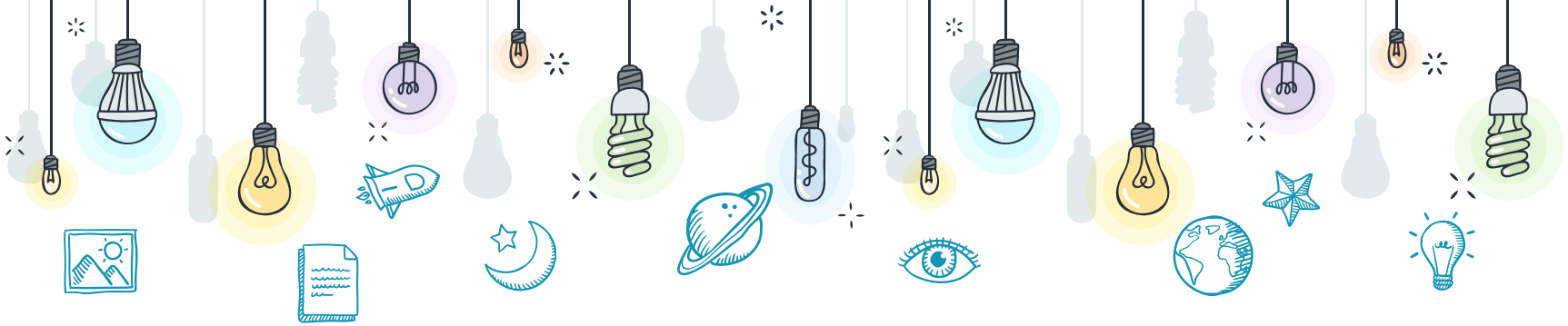


Evaluations and *Benchmarks* in Context of Multimodal LLM

https://mllm2024.github.io/CVPR2025/

Hao Fei

National University of Singapore

Xiang Yue

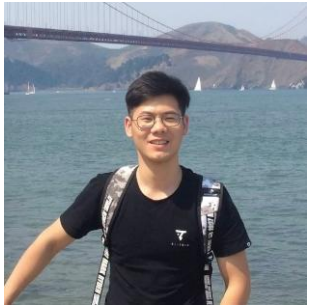
Carnegie Mellon University

Kaipeng Zhang

Shanghai AI Lab

Long Chen

HKUST



Jian Li

Tencent YoutuLab

Xinya Du

University of Texas at Dallas

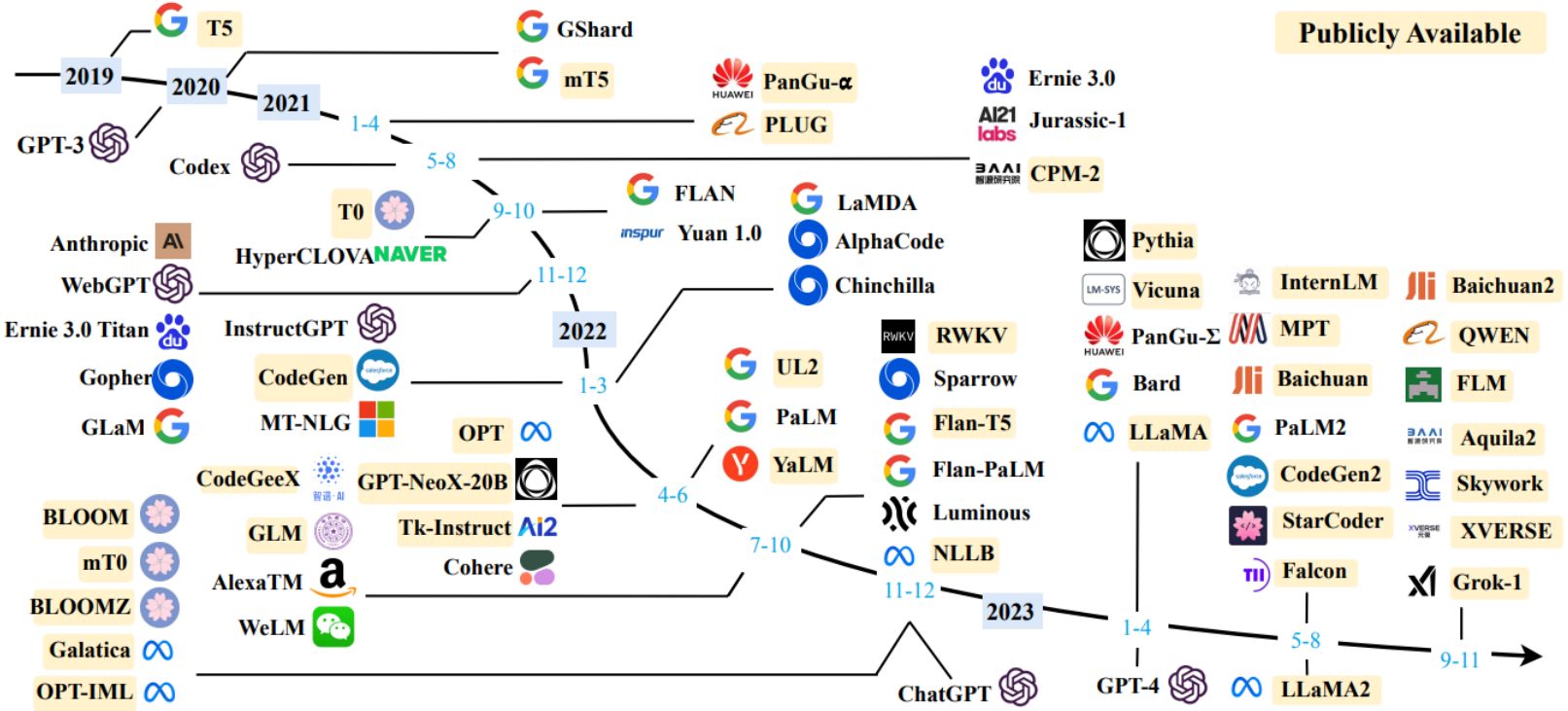
Part-I

Background and Introduction

Xiang Yue
Postdoc Researcher
Carnegie Mellon University
https://xiangyue9607.github.io/

Intelligence in Language

- Very Rapid Evolvement of Language-based LLMs

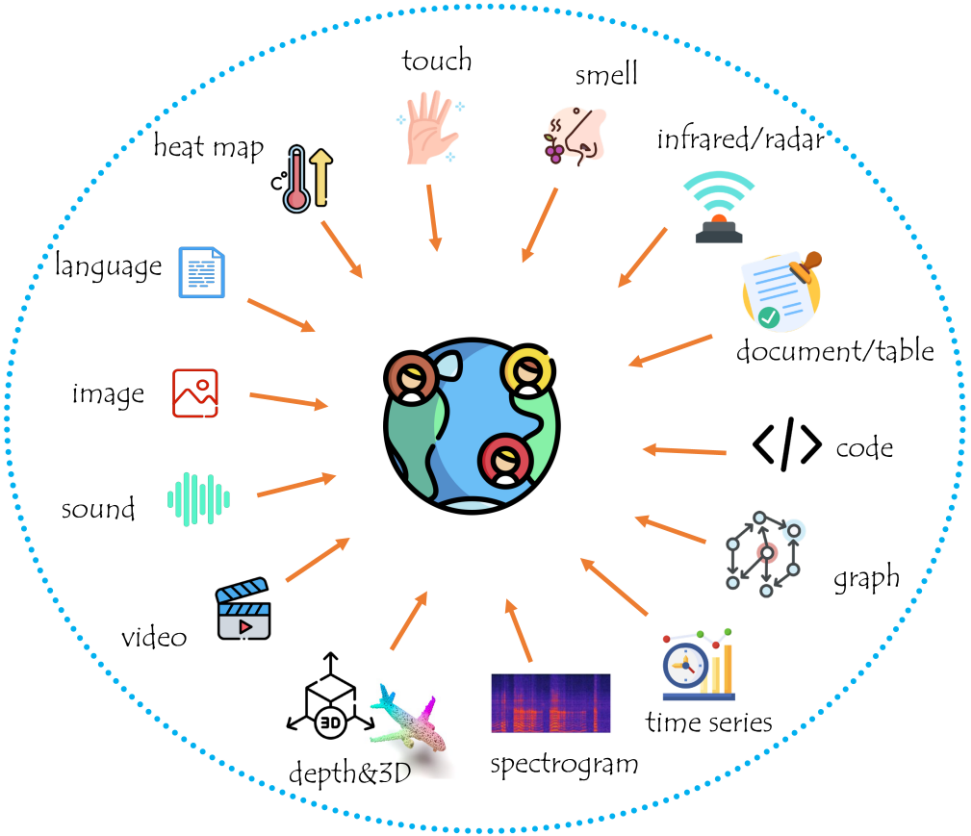


[1] A Survey of Large Language Models. https://github.com/RUCAIBox/LLMSurvey, 2023

Intelligence in Multi-Sensory Data

- ## Harnessing Multimodality

☞ This world we live in is replete with multimodal information & signals, not just language.



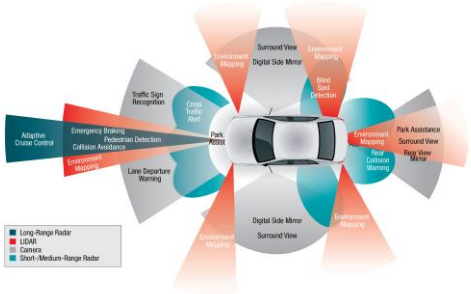
Intelligence in Multi-Sensory Data

Harnessing Multimodality

This world we live in is replete with multimodal information & signals, not just language.

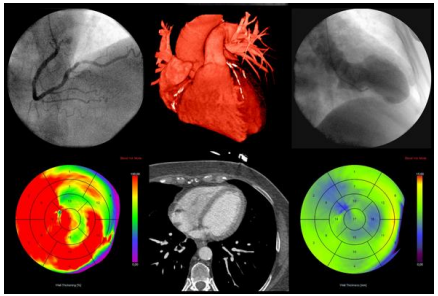
Autonomous Driving Systems

In this application, vehicles use a combination of visual data (cameras), spatial data (LiDAR), and auditory signals (sonar) to navigate safely.



Healthcare Diagnostics

*Medical **imaging** tools like MRI, CT scans, and X-rays, along with patient history and verbal symptoms, are used to diagnose diseases.*

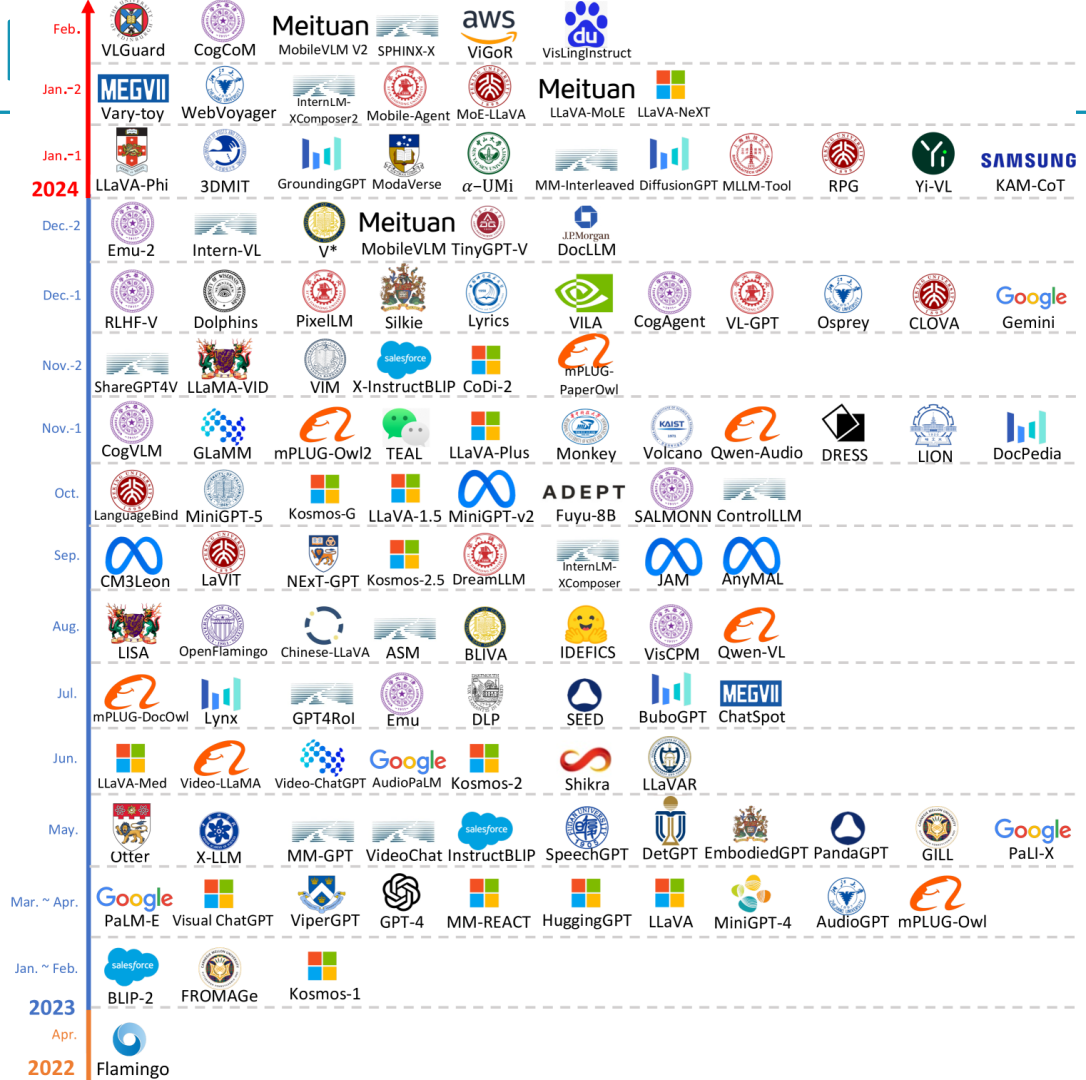


Smart Home Assistants

*Devices like Amazon Alexa and Google Home use voice commands (**audio**), physical interaction (**touch**), and sometimes **visual** cues to operate.*

Intelligence in

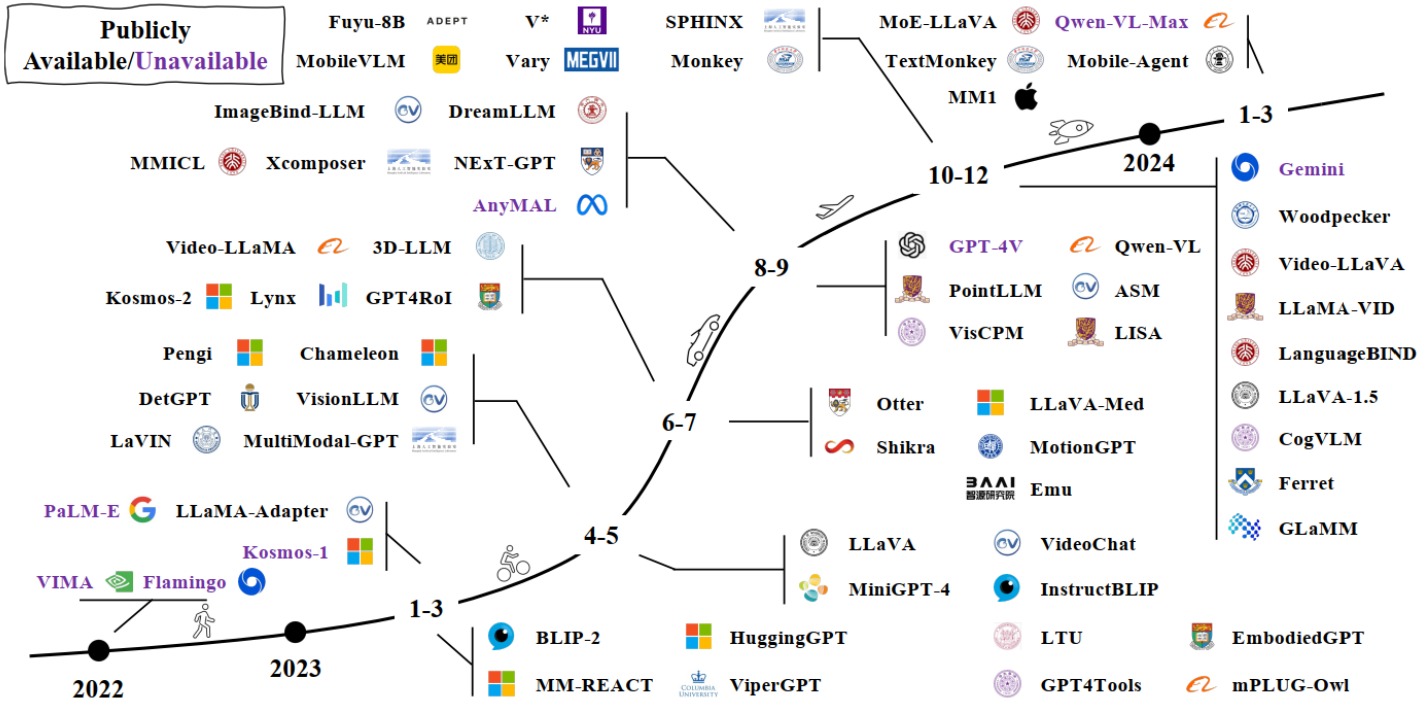
- Trends of MLLMs



[1] MM-LLMs: Recent Advances in MultiModal Large Language Models, 2023.

Intelligence in Multimodal

- Trends of MLLMs



[1] A Survey on Multimodal Large Language Models. https://github.com/BradyFU/Awesome-Multimodal-Large-Language-Models, 2023.

Why should we focus on evaluation?

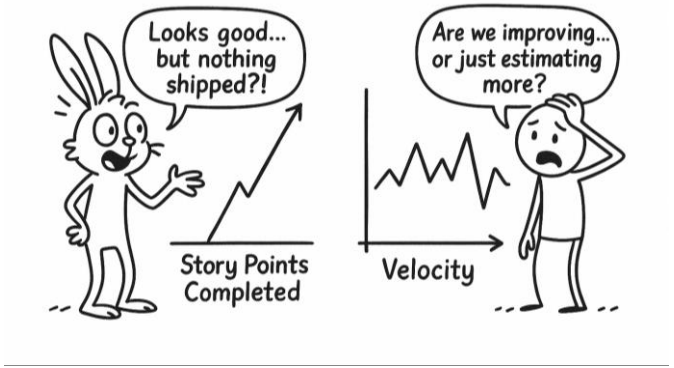
"*If you can't measure it, you can't manage it*"

--Peter Drucker (Father of Management)

Why should we focus on evaluation?

We Optimize What We Measure

Evaluation Reflects Real-World Use



Misleading Metrics Can Harm Trust

Goal of This Tutorial

- What are now?
  - *Walking through the recent key techniques on building MLLM evaluations and benchmarks*
  - *Taxonomies of existing research.*
- Where to go next?
  - *Key insights, current challenges & open problems.*
  - *How to build next generation MLLM benchmarks?*

Evaluations of Multimodal LLMs

Schedule Overview

- **Wed, 11 June, 2024**, **13:00-17:00** Nashville Local Time

Time	Section	Presenter
13:00-13:05	Part 1: Background and Introduction	Xiang Yue
13:05-13:40	Part 2: Existing MLLM Benchmark Overall Survey	Jian Li
13:40-14:15	Part 3: Vision-Language Capability Evaluation	Kaipeng Zhang
14:15-14:50	Part 4: Video Capability Evaluation	Long Chen
14:50-15:10	**Coffee Break**	
15:10-15:45	Part 5: Expert-level Discipline Capability Evaluation	Xiang Yue
15:45-16:20	Part 6: Beyond Evaluation: Path to Multimodal Generalist	Hao Fei
16:20-17:00	Part 7: Multimodal Reasoning & Agent	Xinya Du

From MLLMs to Human-level AI

- ## Contact & QA & Discussions

  - All slides and reading list are available at tutorial homepage:

    https://mllm2024.github.io/CVPR2025/

  - We welcome all Q&A and discussions via Google Group:

    - *Post your questions on Google Group:*

      https://groups.google.com/g/mllm24

    - *Email us:*

      mllm24@googlegroups.com