A Beginner's Guide to AI Research

Insights from a Multimodal AI Perspective

Hao Fei
https://haofei.vip/

Senior Research Fellow
NUS CTIC

Sep 4, 2025

CONTENTS

CONTENTS

What is Research?

Systematic Exploration, Discovery, and Understanding of New Knowledge

Core Definition

Research is the systematic process of exploring, discovering, and understanding new knowledge. It goes beyond mere information gathering, involving critical examination and innovative expansion of existing knowledge.

The Essence of Research

Curiosity-Driven

Driven by curiosity for the unknown and a desire to solve problems.

Systematic Approach

Follows rigorous scientific methodology: problem formulation, experimental design, data collection, analysis, and conclusion drawing.

Knowledge Creation

Aims to generate new theories, methods, technologies, or deepen existing understanding.

Why Do Research?

Exploring the Multiple Motivations and Value of Research

Advance Knowledge Frontiers

Expand human understanding, solve unsolved mysteries, and drive disciplinary and cognitive development.

Solve Real-World Problems

Apply theoretical knowledge to real-world scenarios, creating practical value (e.g., new medical diagnostic tools, optimized traffic systems, improved education quality).

Personal Growth & Development

Cultivate critical thinking, problem-solving skills, innovation, and lifelong learning habits, enhancing overall personal qualities.

Career Advancement

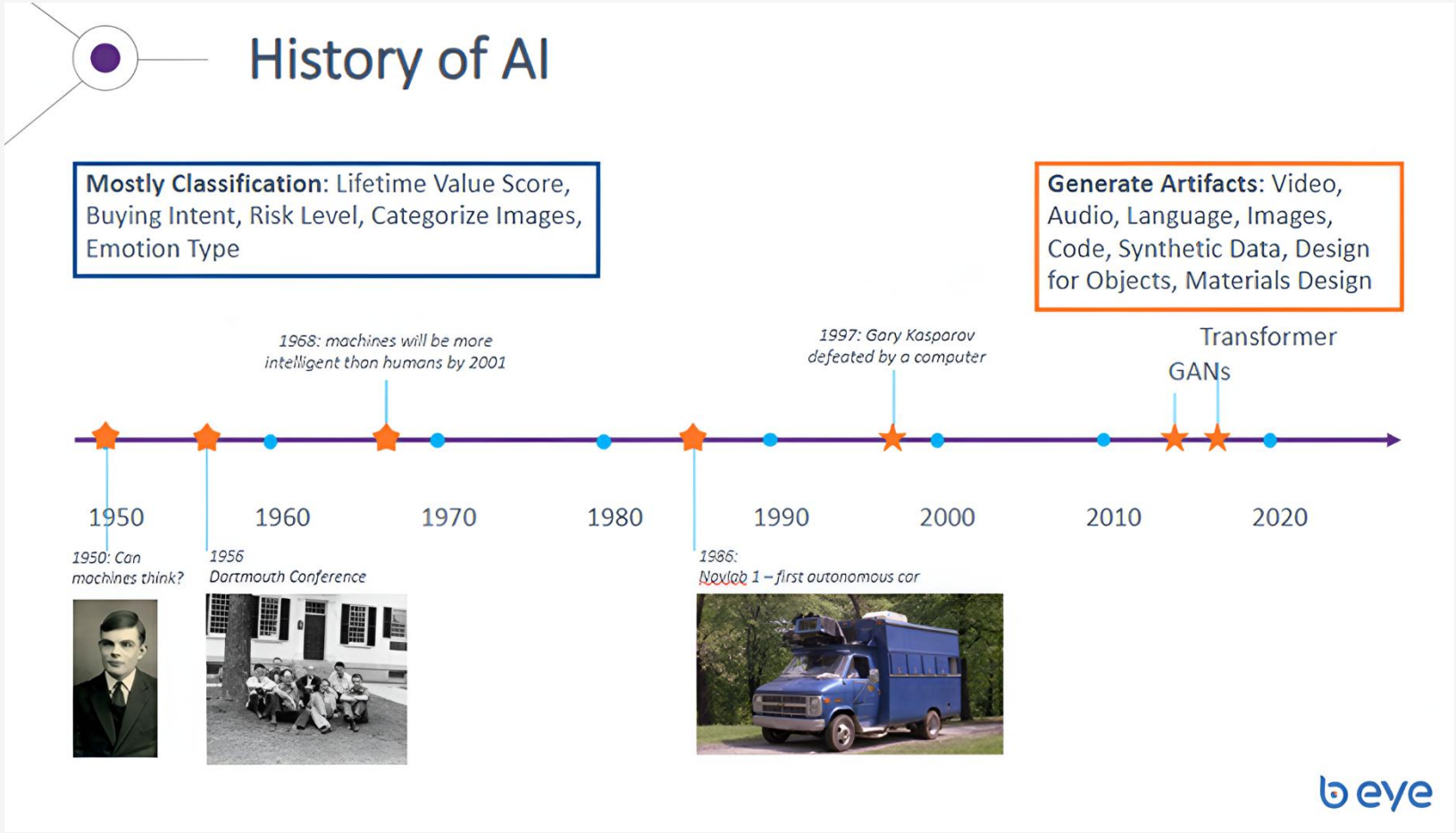
Provide a solid foundation for academia, industrial R&D, or entrepreneurship, boosting career competitiveness and future prospects.

AI Research is Special

Unique Characteristics and Challenges Compared to Traditional Disciplines

Rapidly Evolving Field

- **Rapid Transformation:** AI has undergone revolutionary changes in the past decade, especially with the rise of deep learning and large models.
- **Technological Iteration:** From traditional machine learning to deep learning, and now generative AI and multimodal AI, the tech stack and research paradigms are constantly updating.

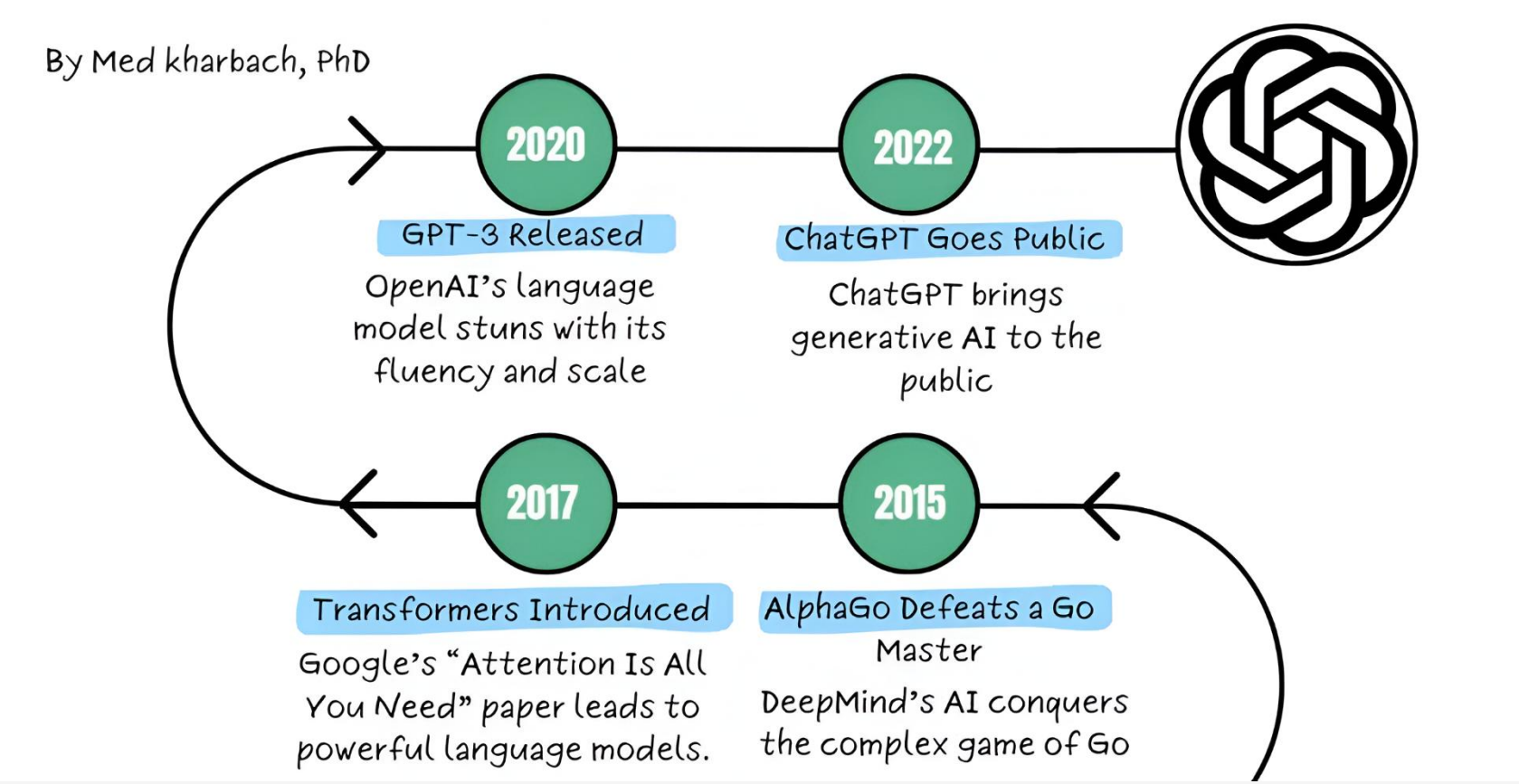


AI Research is Special

Unique Characteristics and Challenges Compared to Traditional Disciplines

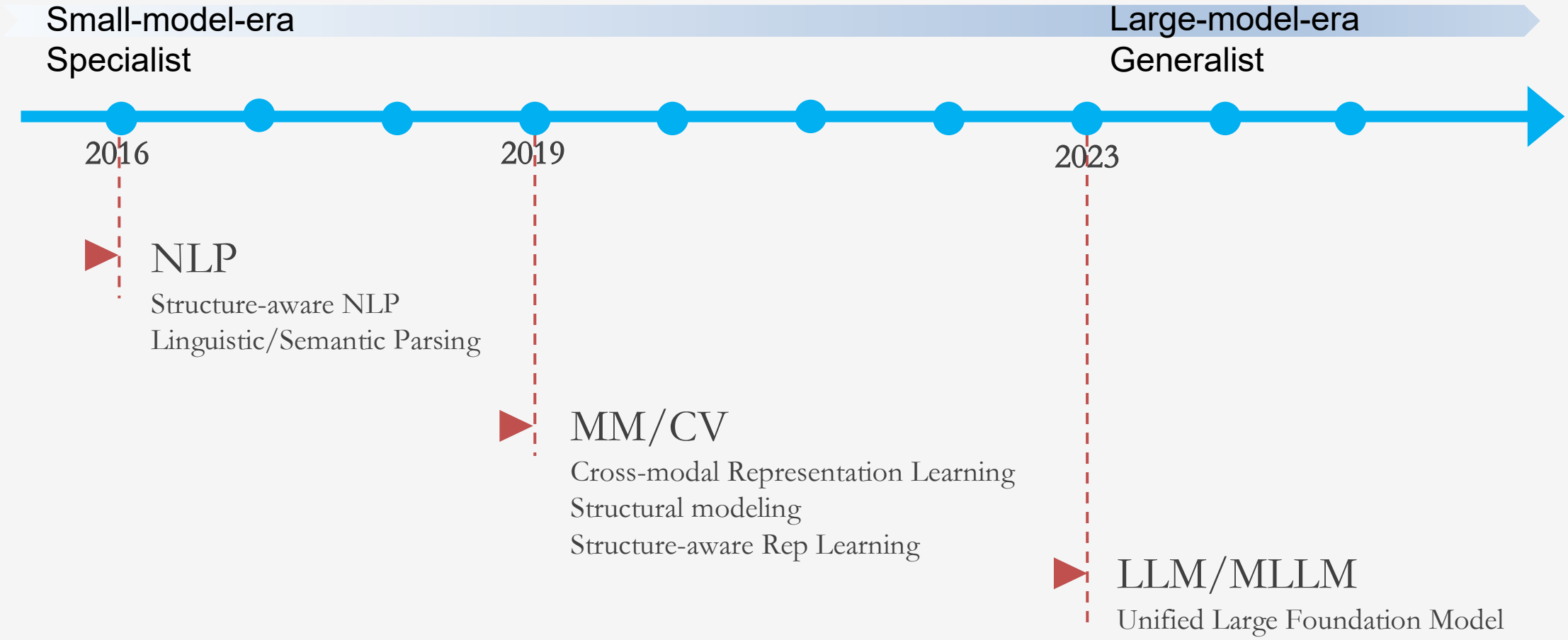
Rapidly Evolving Field

- **Rapid Transformation:** AI has undergone revolutionary changes in the past decade, especially with the rise of deep learning and large models.
- **Technological Iteration:** From traditional machine learning to deep learning, and now generative AI and multimodal AI, the tech stack and research paradigms are constantly updating.



AI Research is Special

My research journey: Stage, Interest and Focus



AI Research is Special

Unique Characteristics and Challenges Compared to Traditional Disciplines

Rapidly Evolving Field

- **Rapid Transformation:** AI has undergone revolutionary changes in the past decade, especially with the rise of deep learning and large models.
- **Technological Iteration:** From traditional machine learning to deep learning, and now generative AI and multimodal AI, the tech stack and research paradigms are constantly updating.

Strong Real-World Connection

- **Application-Driven:** AI research often directly addresses real-world applications, with results quickly transforming into products and services that impact daily life.
- **Data-Driven:** Much AI research relies on large-scale real-world data, making data acquisition, processing, and ethical considerations crucial.

Opportunities for Early Student Contribution

- **Open Ecosystem:** AI boasts active open-source communities (e.g., GitHub), rich competition platforms (e.g., Kaggle), and public datasets.
- **Highly Practical:** Students can quickly gain practical experience and even produce impactful work by participating in open-source projects, replicating papers, and joining competitions.

Example Case: Multimodal AI

From Idea to Impact: A Complete Research Flow

Case Study: From Idea to Impact

Idea Phase

Observing how humans understand the world through multiple senses, posing whether AI can achieve similar multimodal understanding.

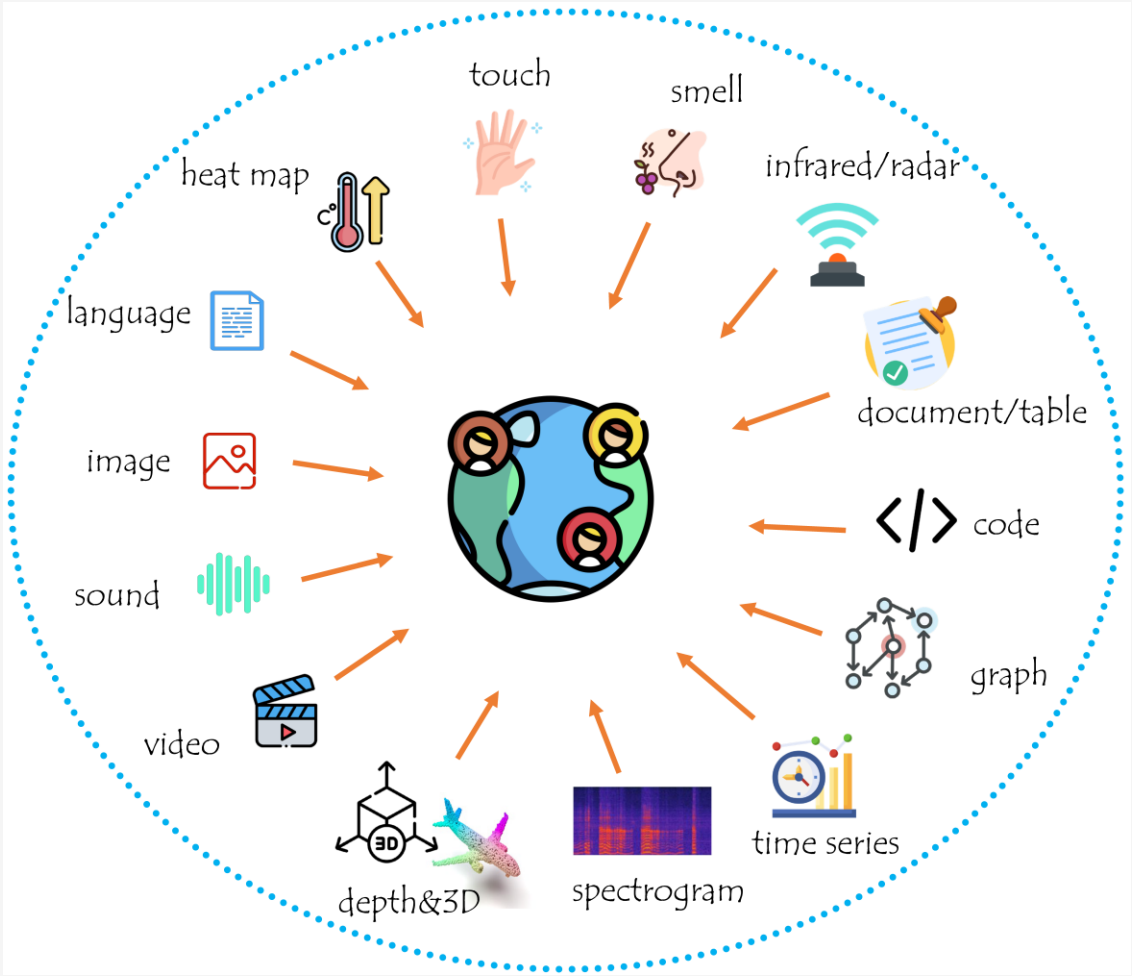
Definition of Multimodality

Multimodal AI aims to process and understand information from various modalities:

- Visual (image, video, 3D)
- Linguistic (text, speech)
- Auditory (audio)

The goal is to achieve more comprehensive, human-like intelligent perception and understanding.

The world functions with varied multimodal information and signals



Example Case: Multimodal AI

NExT-GPT

Foundation Models

NExT-GPT: A multimodal large language model capable of processing various inputs and outputs including text, image, audio, and video.

Research Challenges

Modality Alignment
How to effectively map information from different modalities into a common representation space, enabling mutual understanding and correlation.

Semantic Understanding
How to enable models to deeply understand the intrinsic semantics and contextual relationships across modalities.

Cross-Modal Generation
How to achieve generation from one modality to another (e.g., text-to-image, image-to-text description).

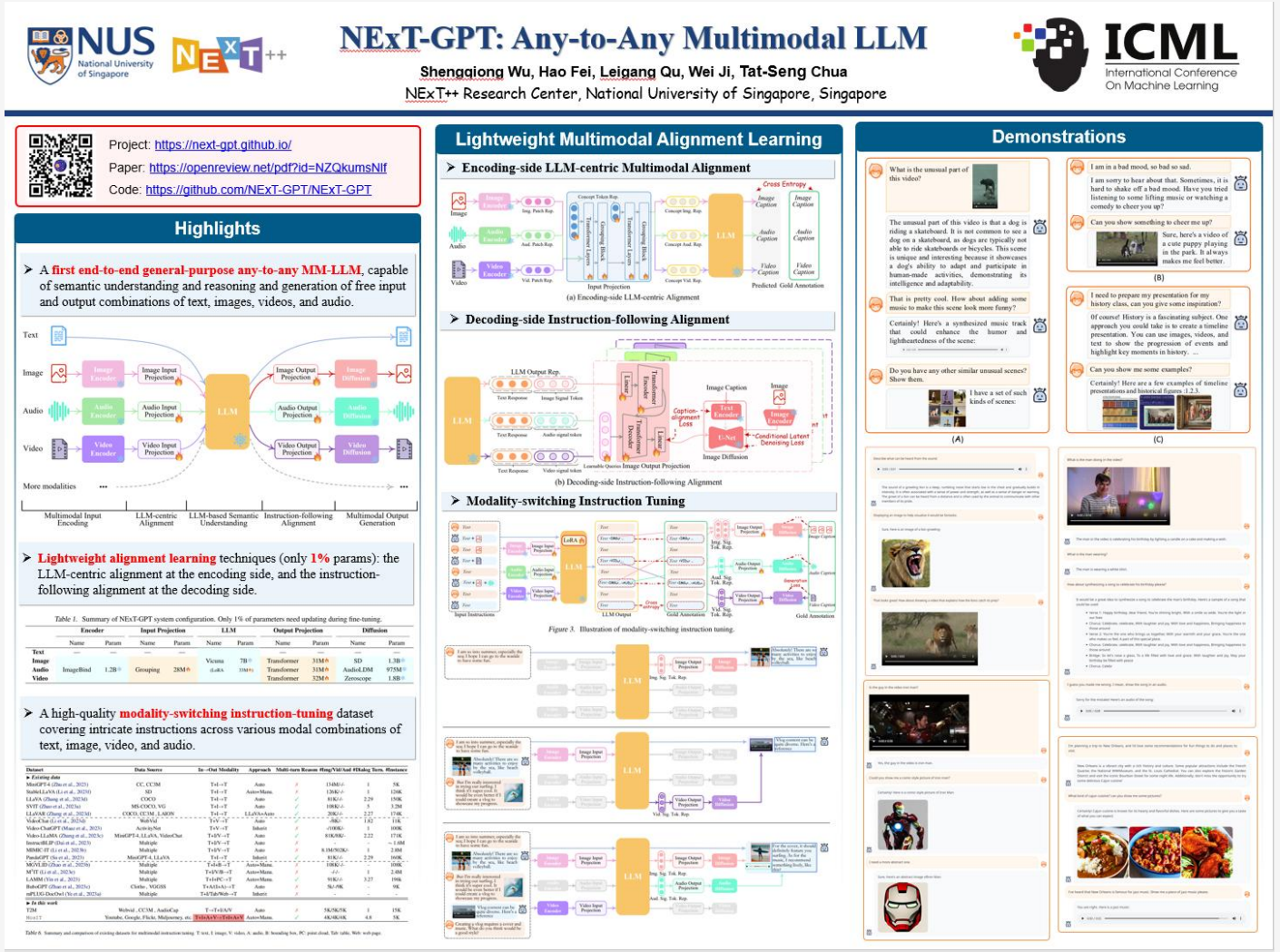
Example Case: Multimodal AI

NExT-GPT

Project: https://next-gpt.github.io

Paper: https://arxiv.org/pdf/2309.05519

Code: https://github.com/NExT-GPT/NExT-GPT

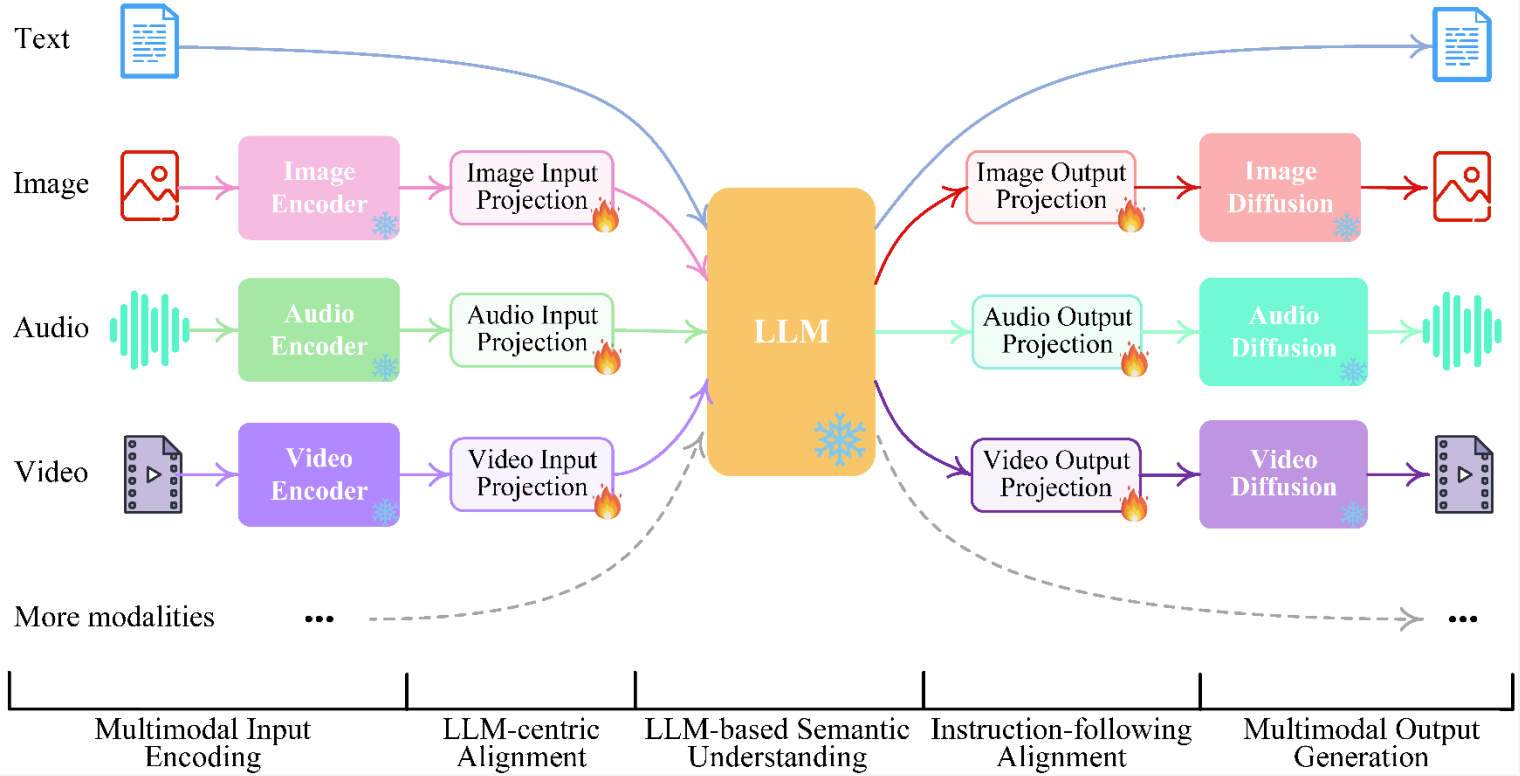


- Shengqiong Wu, **Hao Fei**, Leigang Qu, Wei Ji, Tat-Seng Chua. "NExT-GPT: Any-to-Any Multimodal LLM". **ICML**. 2024.

Example Case: Multimodal AI

NExT-GPT

- The first end-to-end MLLM that **perceives** input and **generates** output in arbitrary combinations (any-to-any) of text, image, video, and audio and beyond.
- Realizing Human-like Multimodal Interaction Mode



- Shengqiong Wu, **Hao Fei**, Leigang Qu, Wei Ji, Tat-Seng Chua. "NExT-GPT: Any-to-Any Multimodal LLM". **ICML**. 2024.

NExT-GPT

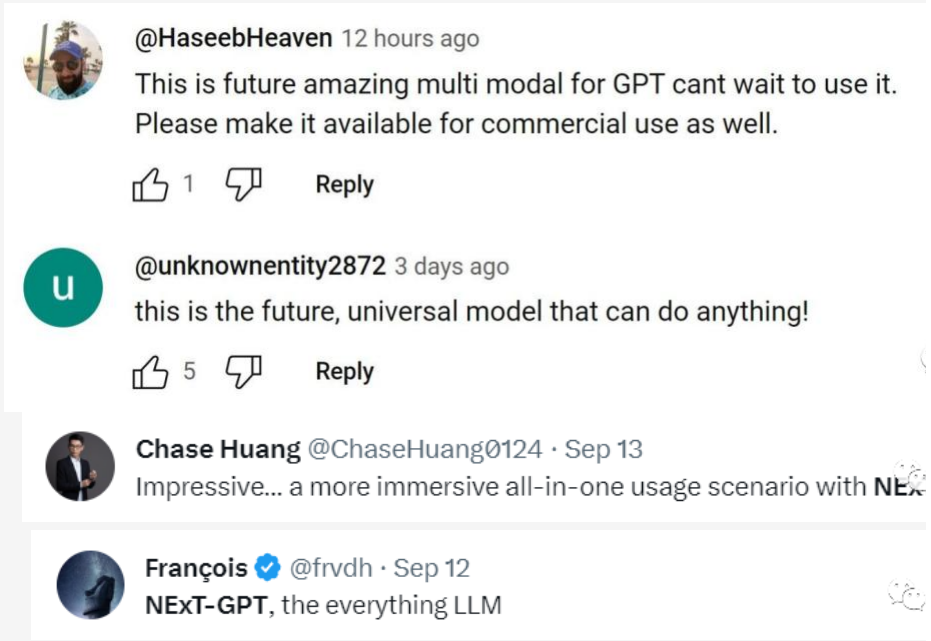
Text + Audio

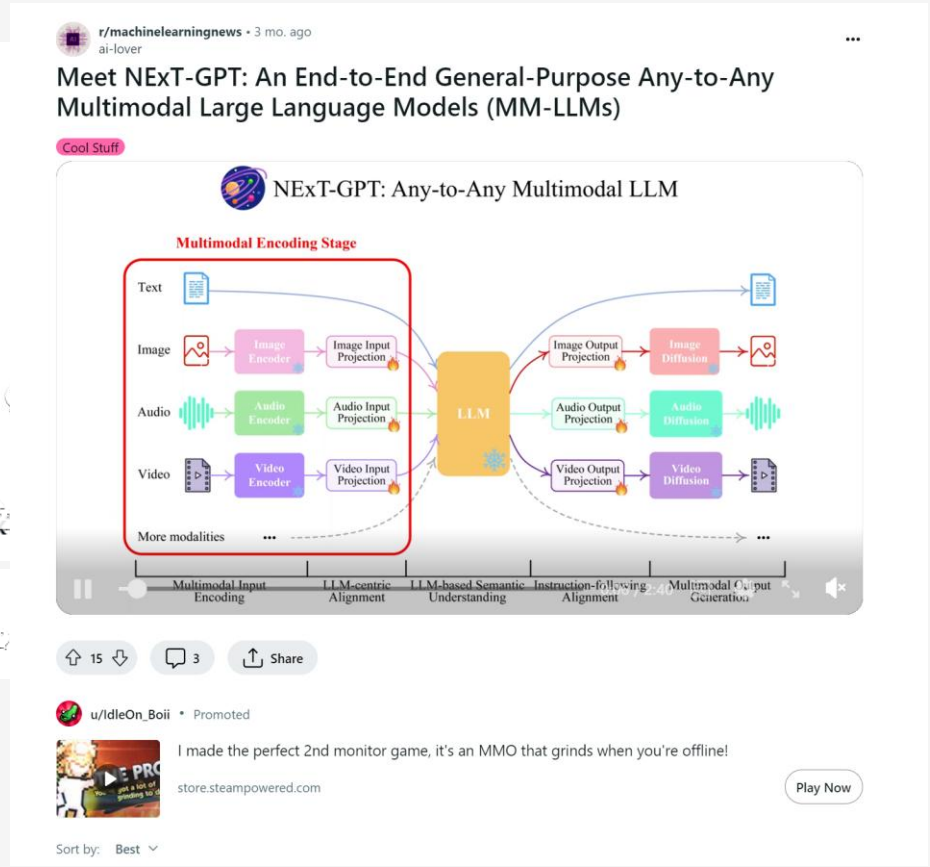
↓

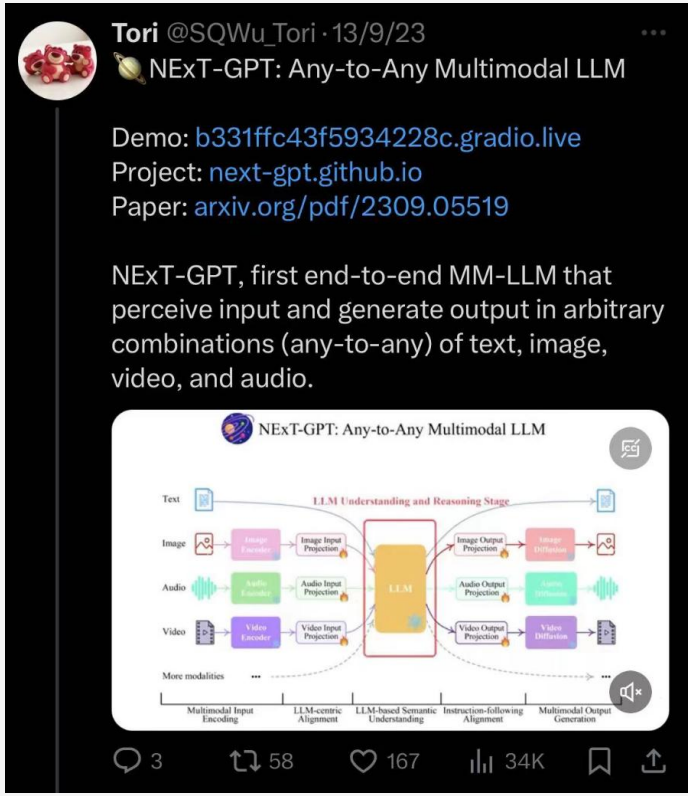
Text + Image + Video

Example Case: Multimodal AI

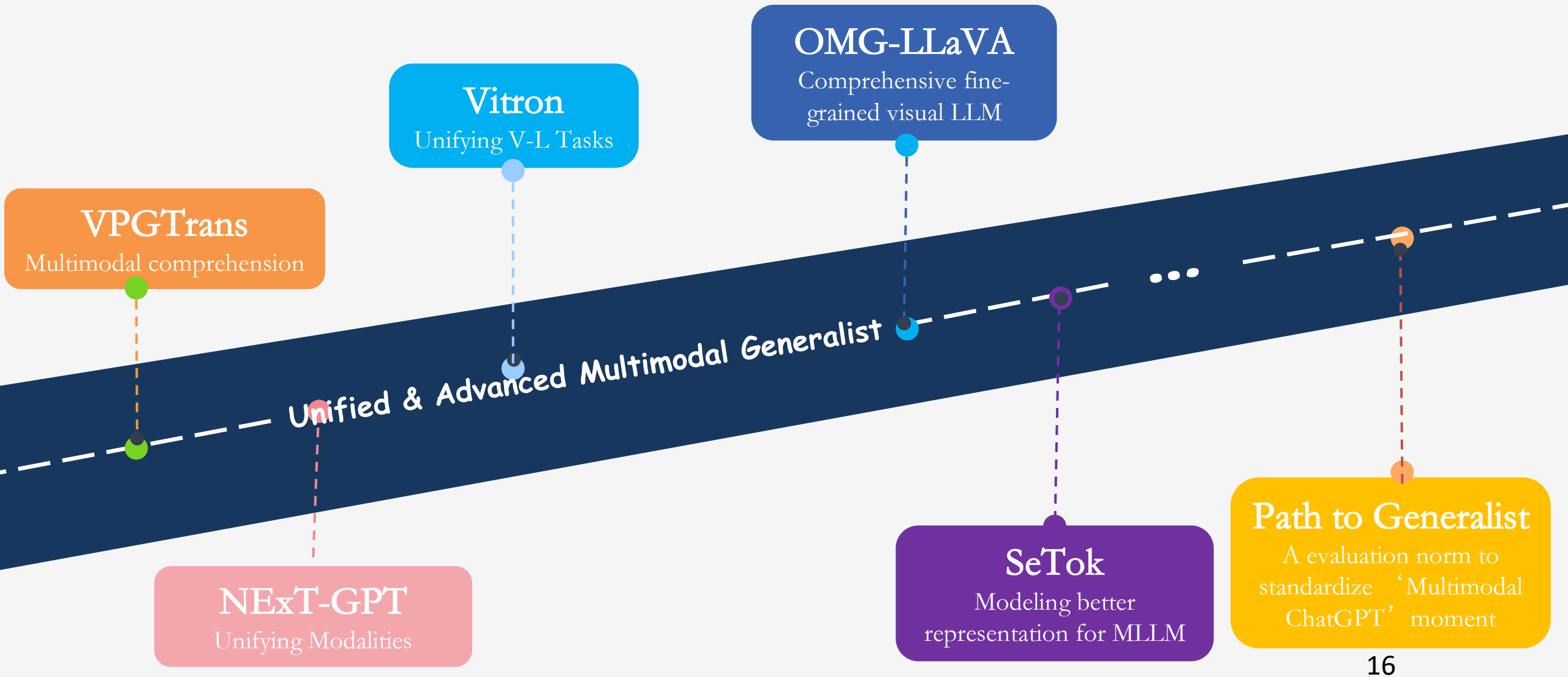
Win tons of high evaluation and praise





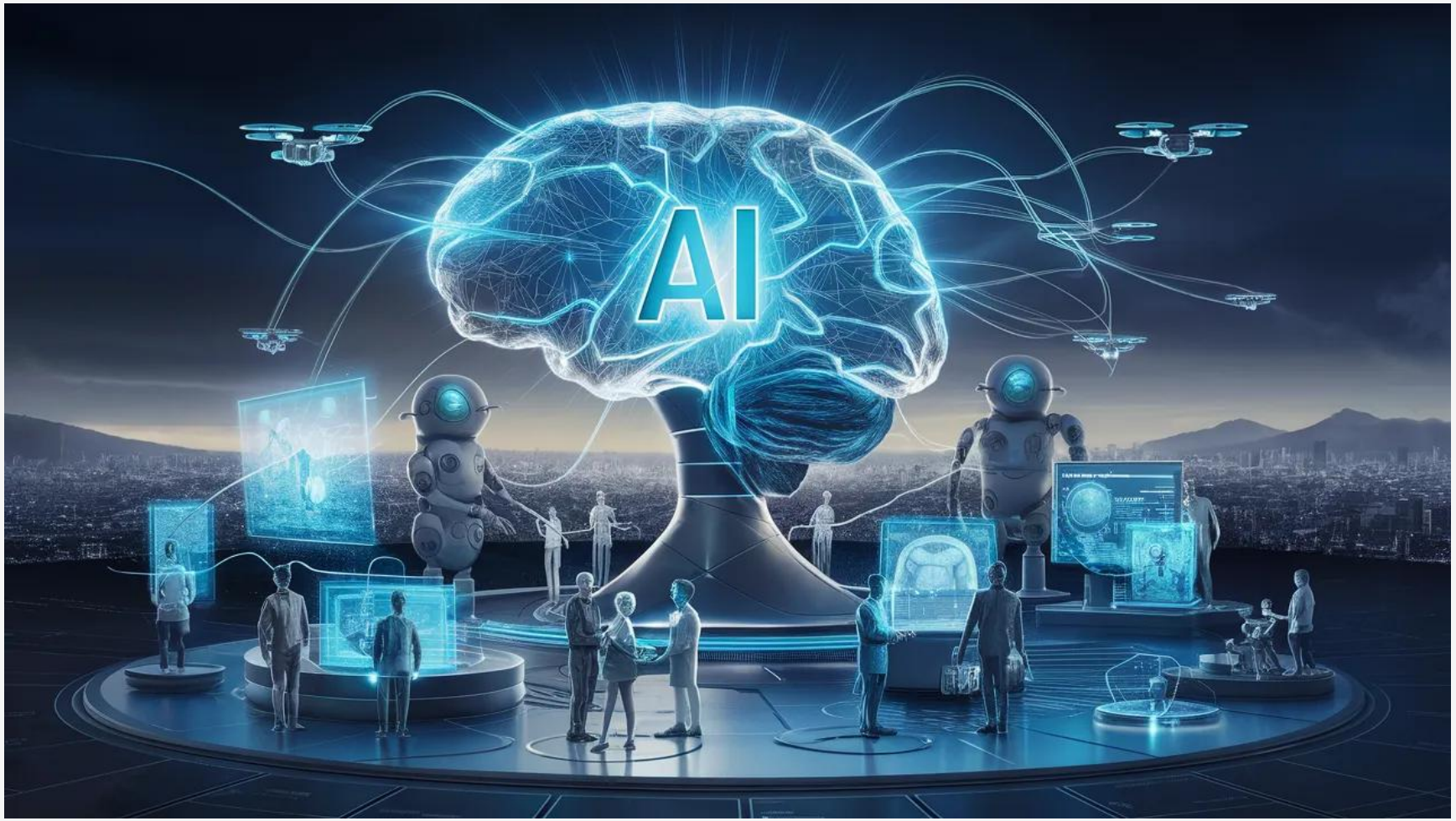


Towards Unified & Advanced Multimodal LLMs/Generalists



Ultimate Goal

☞ What will the next-generation of **multimodal foundation models/agents** look like?



MLLM Tutorial Series

- **Homepage:**

  COLING'24: https://mllm2024.github.io/COLING2024/

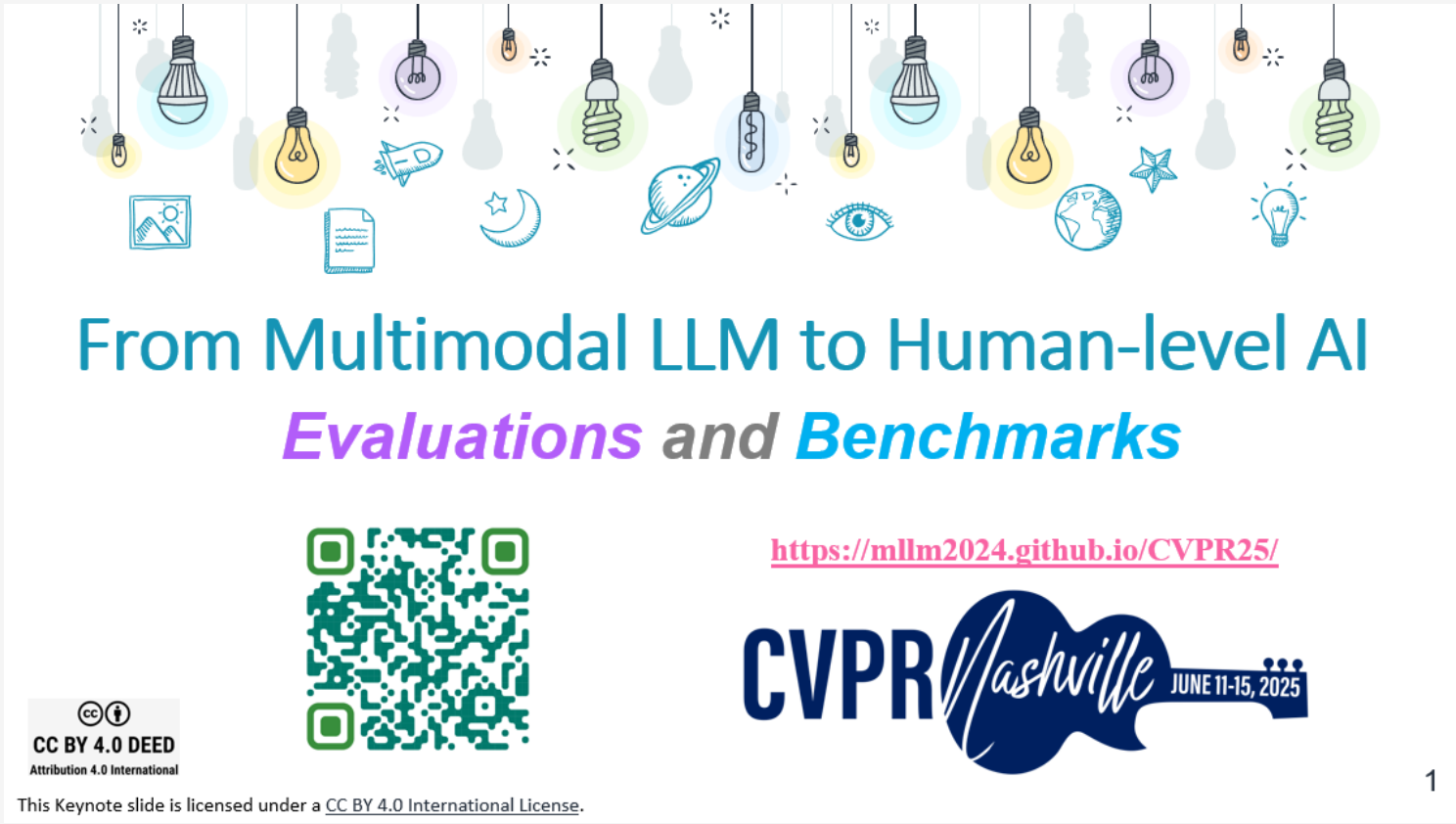
  CVPR'24: https://mllm2024.github.io/CVPR2024/

  ACM MM'24: https://mllm2024.github.io/ACM-MM2024/

  CVPR'25: https://mllm2024.github.io/CVPR2025/

  …

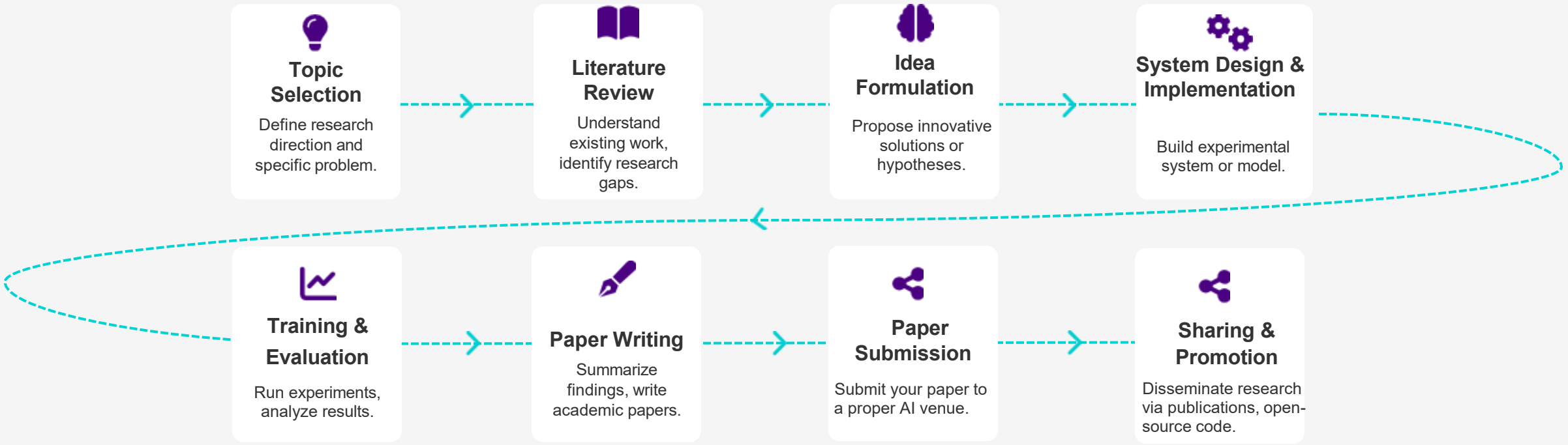
Video: https://www.youtube.com/watch?v=pHBT3zXxQX8



CONTENTS

The AI Research Pipeline

A Standardized Research Path from Topic Selection to Result Sharing



Note: AI research is an iterative optimization process. Each stage is interconnected, influencing subsequent stages, often requiring multiple iterations for validation.

Choosing a Research Topic

Three Key Criteria for Selecting a Valuable Research Direction

Impactful

Addresses real and valuable problems, offering practical application and solving current pain points or bottlenecks. For instance, in multimodal AI, resolving cross-modal information gaps or enhancing multimodal content understanding.

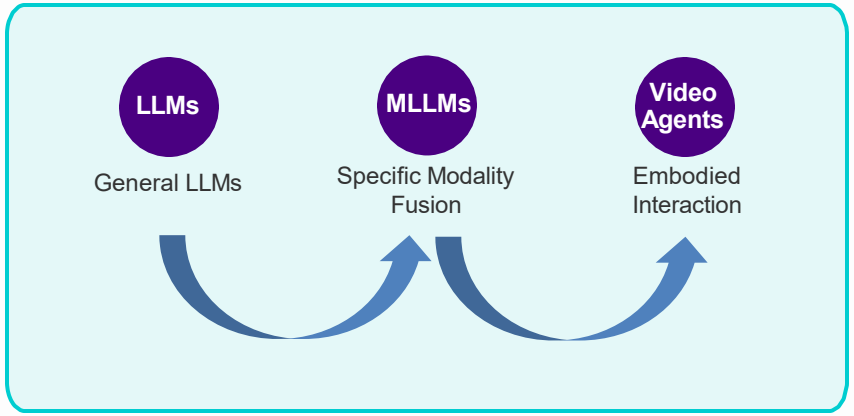
Trending

Focuses on emerging directions, keeping pace with AI frontiers like LLMs, Multimodal LLMs, Embodied AI, and Video Agents. Avoids overly researched or low-potential areas.

Practical

Considers research feasibility, availability of datasets, tools, and open-source resources. Aligns with personal interests and existing skills to ensure productive engagement.

Research Path Example



This evolutionary path from general large models to specific modality fusion and then embodied interaction demonstrates the natural expansion of research directions.

Reading Papers Effectively

Strategies for Efficiently Acquiring and Understanding Frontier Knowledge

Reading Content

Survey Papers

Provide comprehensive overviews, development history, and main methods of a field.

SOTA Papers

Represent the current State-of-the-Art methods and results.

Conference Papers

Papers from top AI conferences (e.g., ICML, CVPR, ACL, NeurIPS).

'Three-Pass' Reading Method

1. **Abstract & Introduction**
   Quickly grasp the problem, motivation, main contributions, and core conclusions.
2. **Figures & Tables**
   Visually understand the core ideas of the method and trends of experimental results.
3. **Conclusion & Experiments**
   Deeply understand the paper's contributions, experimental setup, result analysis, and limitations.

Key Resources

arXiv

Preprint server for accessing the latest unpublished papers.

PapersWithCode

Combines papers with open-source code for easy reproduction and understanding.

Reading Tip

Skim and Deep Read: Quickly browse to filter relevant papers, then deep read.

From Idea to System

Transforming Abstract Ideas into Functional Systems

Idea Refinement

Start with a Concrete Idea

e.g., how to make multimodal models better understand emotions in video.

Refine into a Hypothesis

Transform the idea into a testable scientific hypothesis, e.g., "Introducing temporal attention mechanisms can improve multimodal model performance in video emotion recognition."

Methodology Design

Model/Algorithm Design

Design new model architectures or improve existing algorithms based on the hypothesis.

Phased Implementation

Idea → Prototype → Full System

Data & Baseline

Identify Datasets

Find suitable datasets for the research problem, e.g., multimodal video emotion datasets.

Determine Baseline Models

Select existing SOTA or classic models as baselines for comparison and evaluation.

Training & Evaluation

Validation and Performance Measurement of AI Models

Training Process

Loss Function Optimization

The core objective of training is to minimize a carefully designed loss function by iteratively adjusting model parameters through gradient descent.

Backpropagation

Efficiently calculates gradients through the computational graph to update model weights in the direction that reduces loss.

Iterative Training

Training requires multiple epochs (passes through the dataset) to gradually improve model performance and generalization.

Multi-dimensional Evaluation Metrics

BLEU
Machine translation and text generation.

F1 Score
Classification tasks, especially with imbalanced classes.

Accuracy
Classification tasks.

CIDEr
Image caption generation.

Other Metrics
Recall@K for cross-modal retrieval, FID (image generation quality), etc.

Importance of Reproducibility & Open- Source

Reproducibility

Ensuring others can replicate your experimental results using your code.

Open-Source

Sharing code and models to foster community collaboration and evolution.

Paper Writing

Effectively Communicating Your Research Findings

Anatomy of a Research Paper

- **Abstract**
  Concise summary of the problem, methods, results, and conclusions.
- **Introduction**
  Sets the stage, outlines the problem, states contributions, and previews the paper structure.
- **Related Work**
  Discusses previous research, highlighting gaps your work addresses.
- **Methodology**
  Details your approach, experimental setup, and model architecture.
- **Experiments & Results**
  Presents findings, data analysis, and comparisons with baselines.
- **Discussion (optional)**
  Interprets results, discusses implications, limitations, and future directions.
- **Conclusion**
  Summarizes key findings and reiterates contributions.

Tips for Effective Writing

- **Clarity & Conciseness**
  Use precise language, avoid jargon, and get straight to the point.
- **Logical Flow**
  Ensure smooth transitions between sections and paragraphs.
- **Strong Visuals**
  Use clear figures, tables, and diagrams to convey complex information.
- **Iterative Process**
  Writing is rewriting. Draft, revise, get feedback, and refine.
- **Target Audience**
  Write for researchers in your field, assuming some background but explaining novel concepts clearly.
- **Proofread**
  Check for grammar, spelling, and formatting errors meticulously.

> *"Good writing is not a natural gift. It is a skill you can learn and improve."*

Paper Submission

Key Research Areas and Prestigious Venues in AI

Key AI Research Areas

Computer Vision (CV)
Image recognition, object detection, scene understanding, video analysis, and visual reasoning.

Natural Language Processing (NLP)
Text understanding, generation, translation, sentiment analysis, and large language models.

Multimodal AI
Integration of multiple modalities (text, image, audio, video) for comprehensive understanding and generation.

Machine Learning
Deep learning, reinforcement learning, unsupervised learning, and foundational algorithms.

Recommender Systems & Information Retrieval
Personalized recommendations, search algorithms, ranking systems, and information filtering.

Important AI Venues

Important AI Journals

TPAMI TNNLS TKDE TOIS AIJ

Important Top-tier AI Conferences

- CV
  - CVPR ICCV ECCV
- NLP
  - ACL EMNLP NAACL
- Multimodal
  - ACM MM ICME
- ML
  - ICML NeurIPS ICLR AAAI
- IR
  - SIGIR WWW RecSys

Sharing & Promotion

Maximizing the Impact and Visibility of Your Research

Publication Channels

Strategically share your research through multiple channels to reach diverse audiences:

GitHub
Share code, models, and implementation details

arXiv
Rapid dissemination of research findings

Demo Videos
Visual demonstrations of research outcomes

Twitter/X
Accessible explanations for broader audiences

Research Promotion

Strategic promotion helps your research reach the right audience and maximize impact:

1. **Craft a Compelling Narrative**
   - Highlight the problem's significance and your unique solution
   - Create accessible explanations for different audience levels

2. **Leverage Social Media**
   - Share on Twitter/X, LinkedIn, and specialized communities
   - Create visual summaries and thread explanations

3. **Network Strategically**
   - Connect with researchers in your field and adjacent areas
   - Participate in workshops, discussions, and research forums

CONTENTS

Build Your Foundations

Laying a Solid Groundwork for a Successful AI Research Journey

Theoretical Foundations

Mathematics
Linear Algebra, Probability & Statistics, Optimization Theory

AI Fundamentals
Machine Learning, Deep Learning, NLP, CV, Transformer, Attention…

> *Theory is the cornerstone of research.*

Programming Skills

Python
The most commonly used programming language in AI.

Deep Learning Frameworks
PyTorch/Keras/Tensorflow

Version Control
Git (code management and collaboration)

Operating Systems
Linux (commonly used in server environments)

> *Programming is the tool to turn ideas into reality.*

Continuous Learning Mindset

Lifelong Learning
The AI field evolves rapidly, requiring continuous acquisition of new knowledge and skills.

Curiosity & Exploration
Maintain enthusiasm and a desire to explore new things.

Critical Thinking
Cultivate problem-solving thinking, not blindly following existing conclusions.

> *Mindset determines research success.*

Start Small, Grow Big

A Gradual Growth Path in Research

Replicate Existing Work

- Run Open-Source Code
- Understand Implementation Details
- Debug and Troubleshoot

Small Projects

- Micro-Innovations
- Modify Existing Models
- Experiment with New Datasets

Conference Paper

- Document Innovative Work
- Academic Paper Writing
- Submit to Academic Conferences

Research Direction

- Formulate Long-Term Research Direction
- Deepen Domain Expertise
- Continuously Explore Innovation

> *"Research is a long journey, not a sprint. Start with small steps, gradually accumulate experience, and you will eventually reach the pinnacle of academia."*

Find a Team & Mentor

Collaboration is Key to Research Success

The Value of Collaboration

Research is Not a Solo Endeavor

Team members leverage diverse strengths to solve complex problems; varied perspectives and ideas spark innovation.

Mentors Provide Guidance

Share valuable research experience and methodologies, help clarify research directions, and provide computing resources, datasets, and academic network support.

Teams Offer Feedback & Motivation

Provide mutual feedback to improve work; allocate tasks efficiently; face challenges together, maintaining research enthusiasm.

Ways to Find

Labs

Join university or research institution labs, participate in ongoing projects, and gain research experience.

Professors

Proactively contact professors of interest, express research intent, and participate in their supervised projects.

Internships

Engage in industry or academic internships to encounter real-world problems and gain practical experience.

> *"Excellent mentors and teams not only accelerate your research growth but also help you avoid common pitfalls and collectively explore the frontiers of AI."*

Final Words & Encouragement

Key Attitudes for Starting Your AI Research Journey

Key Research Attitudes

Learning by Doing

Practice is the best way to learn; don't be afraid to get hands-on.

Be Curious, Patient, Persistent

These are essential qualities for a researcher.

Long Journey, Not a Sprint; Get well motivated

Research requires patience and perseverance; results often take time to accumulate.

Embrace Failures

Learn from failures, continuously adjust and improve.

Quote

> ***"You don't need to be extraordinary to start, but you need to start to become extraordinary."***
>
> **— Fei-Fei Li**

Useful Links & Resource

About AI, Cognition and Thought

- Cognitive Neuroscience of Language (David Kemmerer)
  https://www.routledge.com/Cognitive-Neuroscience-of-Language/Kemmerer/p/book/9781138318403
- AGI and World Model (Yann LeCun)
  https://ai.meta.com/blog/yann-lecun-advances-in-ai-research/
- Commonsense and Moral Reasoning (Yejin Chio)
  https://www.youtube.com/watch?v=KGRGrMUJd8U
- Intelligence is about compression (Ilya Sutskever)
  https://the-decoder.com/openai-co-founder-explains-the-secret-sauce-behind-unsupervised-learning/?utm_source=chatgpt.com

About Paper Writing for Beginner

- A Survival Guide to a PhD (by Andrej Karpathy)
  http://karpathy.github.io/2016/09/07/phd/
- Good Research and Good Papers (by Fei-Fei Li)
  http://qinhongwei.com/2015/03/25/Fei-fei-Li-s-advice/
- How to write a good paper/review/rebuttal (CVPR Tutorial)
  https://sites.google.com/view/making-reviews-great-again/
- How to give a great research talk (Simon Peyton Jones)
  https://www.microsoft.com/en-us/research/uploads/prod/2016/07/How-to-give-a-great-research-talk.pdf
- How to Properly Review AI Papers? (Jiebo Luo)
  https://valser.org/webinar/slide/slides/20200710/howtoproperlyreviewaipapers-200710022751.pdf

Thank you!

https://haofei.vip/

haofei37@nus.edu.sg