

# Generative AI Innovative Applications

# Introduction to Generative AI Innovative Applications

1132GAIIA01

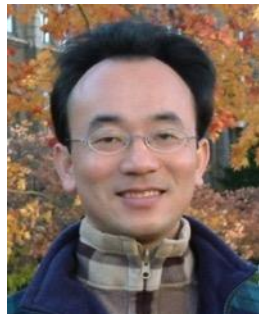
MBA, IM, NTPU (M6031) (Spring 2025)

Tue 2, 3, 4 (9:10-12:00) (B3F17)



<https://meet.google.com/paj-zhhj-mya>

aws  
educate | Cloud  
Ambassador  
2020 Cohort



Min-Yuh Day, Ph.D,  
Professor

[Institute of Information Management, National Taipei University](https://web.ntpu.edu.tw/~myday)

<https://web.ntpu.edu.tw/~myday>





國立臺北大學  
National Taipei University



# Min-Yuh Day, Ph.D.

aws educate | Cloud Ambassador

2020 Cohort

**Professor, Information Management, NTPU**

**Visiting Scholar, IIS, Academia Sinica**

**Ph.D., Information Management, NTU**

**Director, Intelligent Financial Innovation Technology, IFIT Lab, IM, NTPU**

**Director, Fintech and Green Finance Research Center, NTPU**

**Division Director, Sustainable Development, Sustainability Office, NTPU**

**Artificial Intelligence, Generative AI, ESG and Green Financial Technology,  
Big Data Analytics, Electronic Commerce, Biomedical Informatics**

aws educate | Cloud Ambassador

2020 Cohort

aws academy

Accredited  
Educator

aws certified

Solutions  
Architect

Associate

aws certified

Cloud  
Practitioner



國立臺北大學  
National Taipei University



Spring 2025

**Generative AI**  
**Innovative Applications**

# Course Syllabus

## National Taipei University

### Academic Year 113, 2<sup>nd</sup> Semester (Spring 2025)

- **Course Title: Generative AI Innovative Applications**
- **Instructor: Min-Yuh Day**
- **Course Class: MBA, IM, NTPU (3 Credits, Elective)**
- **Details**
  - **In-Class and Distance Learning EMI Course (3 Credits, Elective, One Semester) (M6031)**
- **Time & Place: Tue, 2, 3, 4, (9:10-12:00) (B3F17)**
- **Google Meet: <https://meet.google.com/paj-zhhj-mya>**



<https://meet.google.com/paj-zhhj-mya>



# Course Objectives

1. Understand the **fundamental concepts and research issues of Generative AI Innovative Applications**..
2. Equip with Hands-on practices of **Generative AI Innovative Applications**..
3. Conduct information systems research in the context of **Generative AI Innovative Applications**..

# Course Outline

- This course introduces the **fundamental concepts, research issues, and hands-on practices** of **Generative AI Innovative Applications**.
- Topics include:
  1. Introduction to Generative AI Innovative Applications
  2. Transformers for Natural Language Processing and Computer Vision
  3. Large Language Models (LLMs)
  4. **NVIDIA Building RAG Agents with LLMs**
  5. Generative AI for Multimodal Information Generation
  6. **NVIDIA Generative AI with Diffusion Models**
  7. AI Agents and Large Multimodal Agents (LMAs)
  8. **Case Study on Generative AI Innovative Applications**

# Core Competence

- **Exploring new knowledge in information technology, system development and application 80 %**
- **Internet marketing planning ability 10 %**
- **Thesis writing and independent research skills 10 %**

# Four Fundamental Qualities

- **Professionalism**
  - **Creative thinking and Problem-solving 40 %**
  - **Comprehensive Integration 40 %**
- **Interpersonal Relationship**
  - **Communication and Coordination 10 %**
  - **Teamwork 5 %**
- **Ethics**
  - **Honesty and Integrity 0 %**
  - **Self-Esteem and Self-reflection 0 %**
- **International Vision**
  - **Caring for Diversity 0 %**
  - **Interdisciplinary Vision 5 %**

# College Learning Goals

- **Ethics/Corporate Social Responsibility**
- **Global Knowledge/Awareness**
- **Communication**
- **Analytical and Critical Thinking**

# Department Learning Goals

- **Information Technologies and System Development Capabilities**
- **Internet Marketing Management Capabilities**
- **Research capabilities**

# Syllabus

**Week Date Subject/Topics**

**1 2025/02/18 Introduction to Generative AI Innovative Applications**

**2 2025/02/25 Transformers for Natural Language Processing and  
Computer Vision**

**3 2025/03/04 Large Language Models (LLMs),  
NVIDIA Building RAG Agents with LLMs Part I**

**4 2025/03/11 Case Study on Generative AI Innovative Applications I**

**5 2025/03/18 NVIDIA Building RAG Agents with LLMs Part II**

**6 2025/03/25 NVIDIA Building RAG Agents with LLMs Part III**

# Syllabus

**Week Date Subject/Topics**

7 2025/04/01 Self-Learning

**8 2025/04/08 Midterm Project Report**

**9 2025/04/15 Generative AI for Multimodal Information Generation**

**10 2025/04/22 NVIDIA Generative AI with Diffusion Models Part I**

**11 2025/04/29 NVIDIA Generative AI with Diffusion Models Part II**

**12 2025/05/06 Case Study on Generative AI Innovative Applications II**

# Syllabus

**Week Date Subject/Topics**

**13 2025/05/13 NVIDIA Generative AI with Diffusion Models Part III**

**14 2025/05/20 AI Agents and Large Multimodal Agents (LMAs)**

**15 2025/05/27 Final Project Report I**

**16 2025/06/03 Final Project Report II**

# Teaching Methods and Activities

- **Lecture**
- **Discussion**
- **Practicum**

# Evaluation Methods

- **Individual Presentation 60 %**
- **Group Presentation 10 %**
- **Case Report 10 %**
- **Class Participation 10 %**
- **Assignment 10 %**

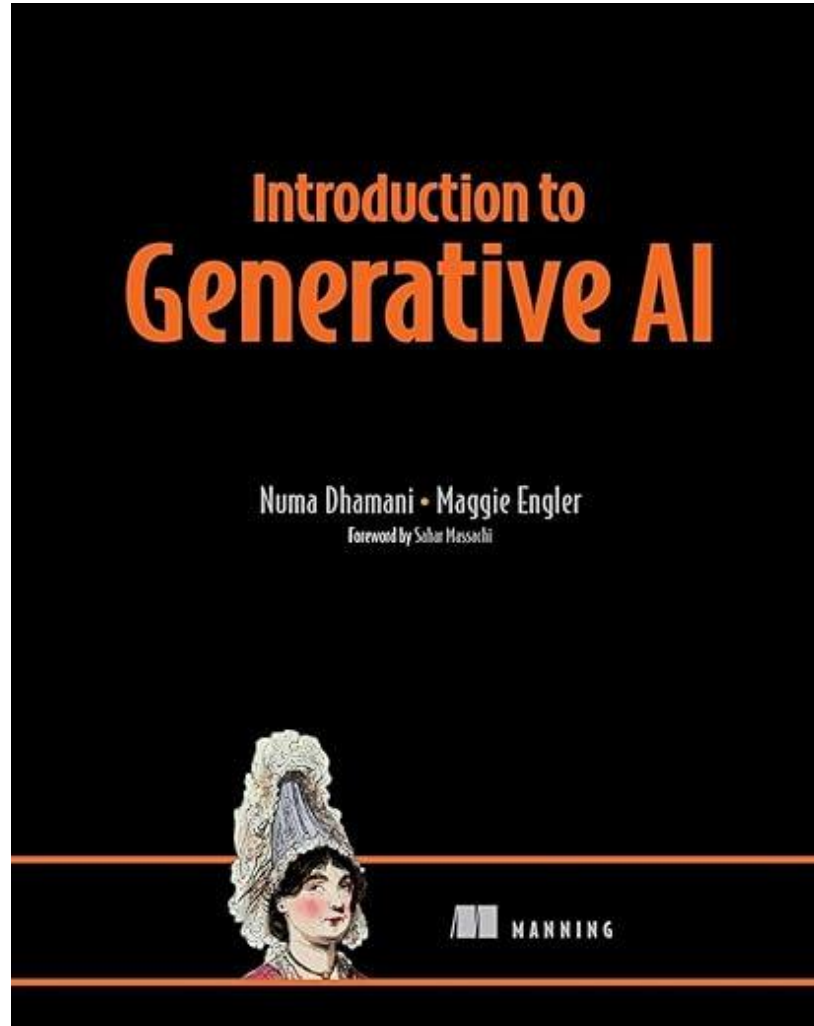
# Required Texts

- 1. Numa Dhamani and Maggie Engler (2024), Introduction to Generative AI, Manning**
- 2. Denis Rothman (2024), Transformers for Natural Language Processing and Computer Vision: Explore Generative AI and Large Language Models with Hugging Face, ChatGPT, GPT-4V, and DALL-E 3, 3<sup>rd</sup> Edition, Packt Publishing**

# References

- NVIDIA DLI (2024), Building RAG Agents with LLMs, [https://learn.nvidia.com/courses/course-detail?course\\_id=course-v1:DLI+S-FX-15+V1](https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-15+V1)
- NVIDIA DLI (2024), Generative AI with Diffusion Models, [https://learn.nvidia.com/courses/course-detail?course\\_id=course-v1:DLI+S-FX-14+V1](https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1)
- Denis Rothman (2024), RAG-Driven Generative AI: Build custom retrieval augmented generation pipelines with LlamaIndex, Deep Lake, and Pinecone, Packt Publishing
- Alammam and Maarten Grootendorst (2024), Hands-On Large Language Models: Language Understanding and Generation, O'Reilly Media
- Ben Auffarth (2023), Generative AI with LangChain: Build large language model (LLM) apps with Python, ChatGPT and other LLMs, Packt Publishing
- Chris Fregly, Antje Barth, and Shelbee Eigenbrode (2023), Generative AI on AWS: Building Context-Aware Multimodal Reasoning Applications, O'Reilly Media
- David Foster (2023), Generative Deep Learning: Teaching Machines to Paint, Write, Compose, and Play, 2nd Edition, Oreilly & Associates Inc

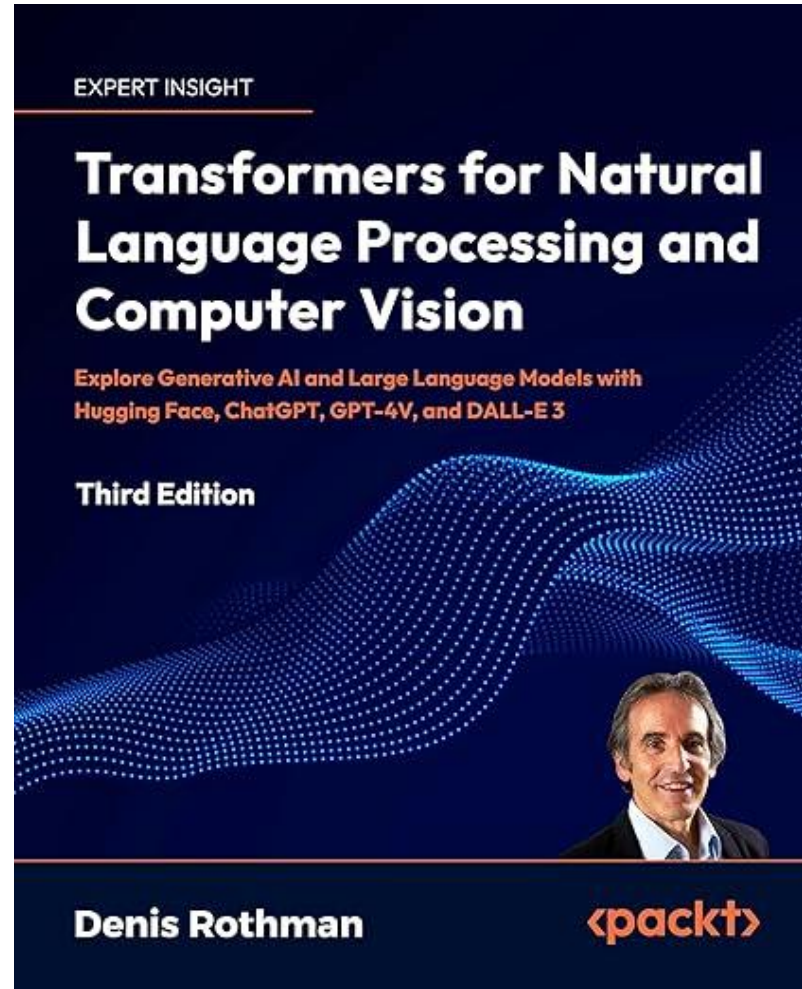
Numa Dhamani and Maggie Engler (2024),  
**Introduction to Generative AI,**  
Manning



Source: Numa Dhamani and Maggie Engler (2024), Introduction to Generative AI, Manning  
<https://www.amazon.com/Introduction-Generative-AI-Numa-Dhamani/dp/1633437191/>

Denis Rothman (2024),

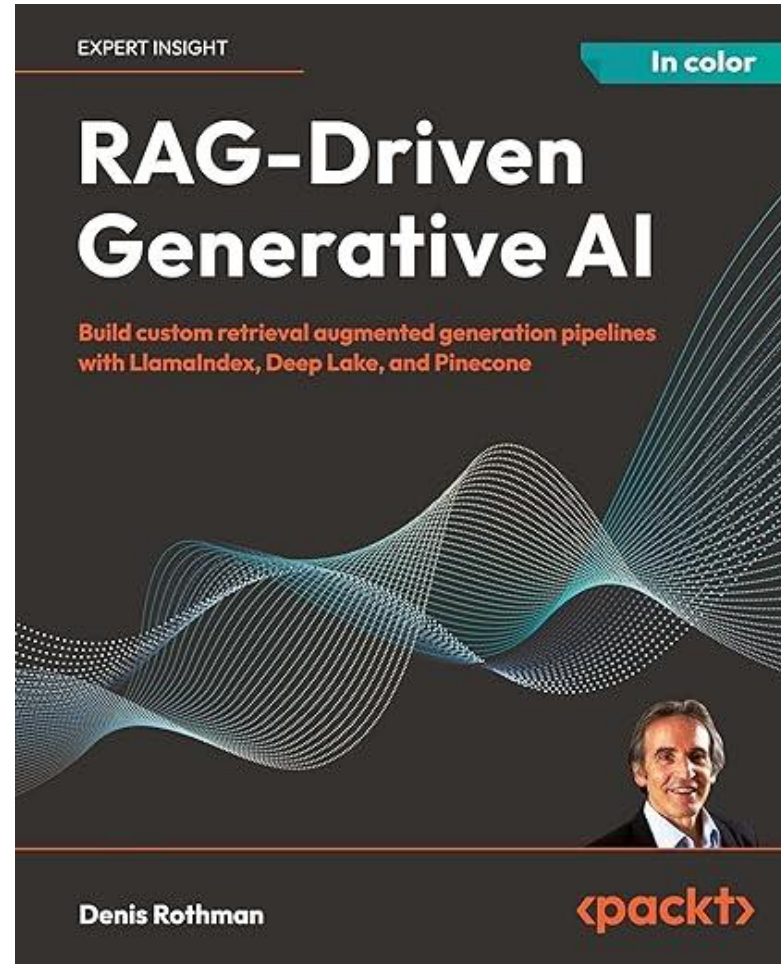
# Transformers for Natural Language Processing and Computer Vision: Explore Generative AI and Large Language Models with Hugging Face, ChatGPT, GPT-4V, and DALL-E 3, 3rd Edition, Packt Publishing



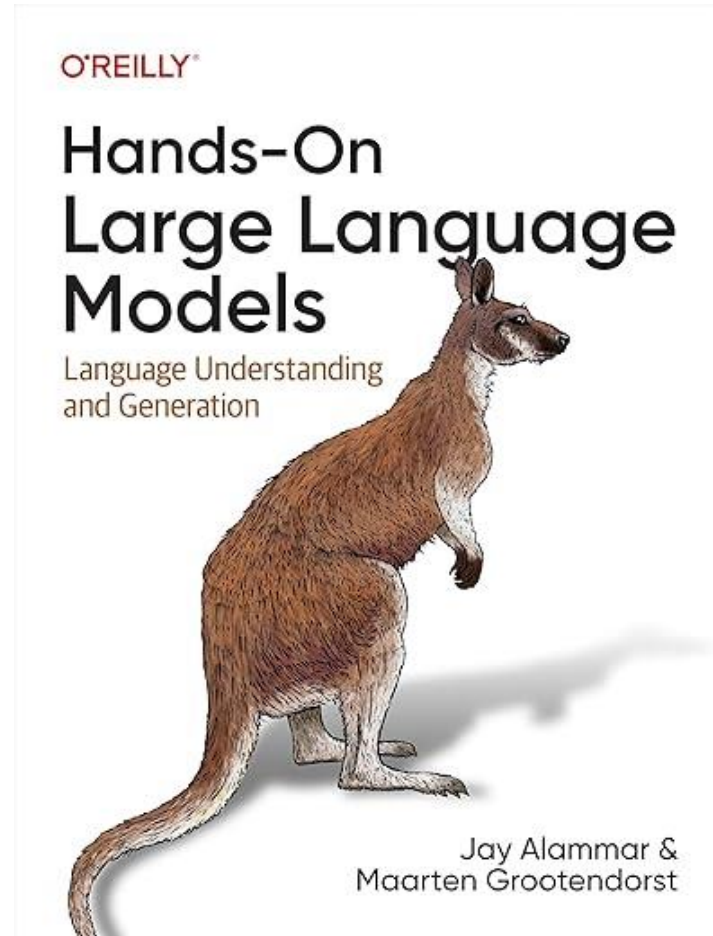
Denis Rothman (2024),

# RAG-Driven Generative AI:

Build custom retrieval augmented generation pipelines with LlamaIndex, Deep Lake, and Pinecone,  
Packt Publishing



Jay Alammar and Maarten Grootendorst (2024),  
**Hands-On Large Language Models:  
Language Understanding and Generation,**  
O'Reilly Media



Jay Alammar and Maarten Grootendorst (2024),  
**Hands-On Large Language Models:  
Language Understanding and Generation,**  
O'Reilly Media

Chapter 1: Introduction to Language Models

Chapter 2: Tokens and Embeddings

Chapter 3: Looking Inside Transformer LLMs

Chapter 4: Text Classification

Chapter 5: Text Clustering and Topic Modeling

Chapter 6: Prompt Engineering

Chapter 7: Advanced Text Generation Techniques and Tools

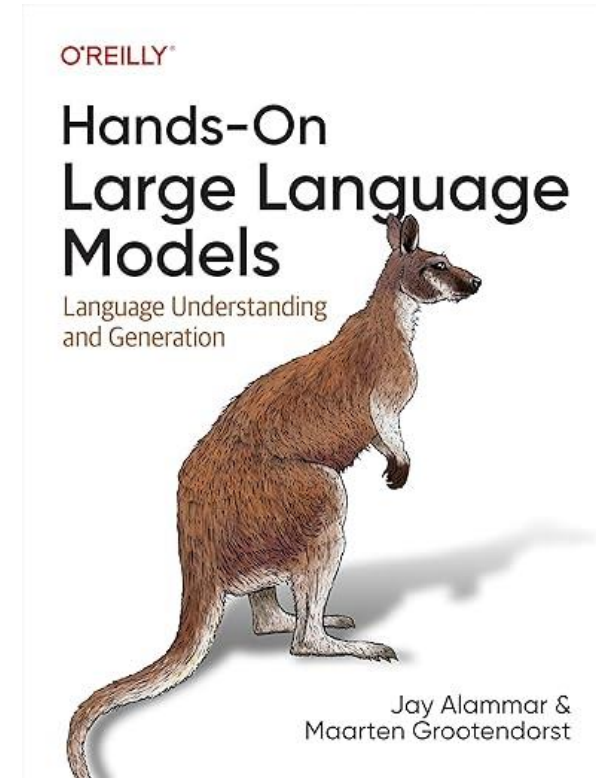
Chapter 8: Semantic Search and Retrieval-Augmented Generation

Chapter 9: Multimodal Large Language Models

Chapter 10: Creating Text Embedding Models

Chapter 11: Fine-tuning Representation Models for Classification

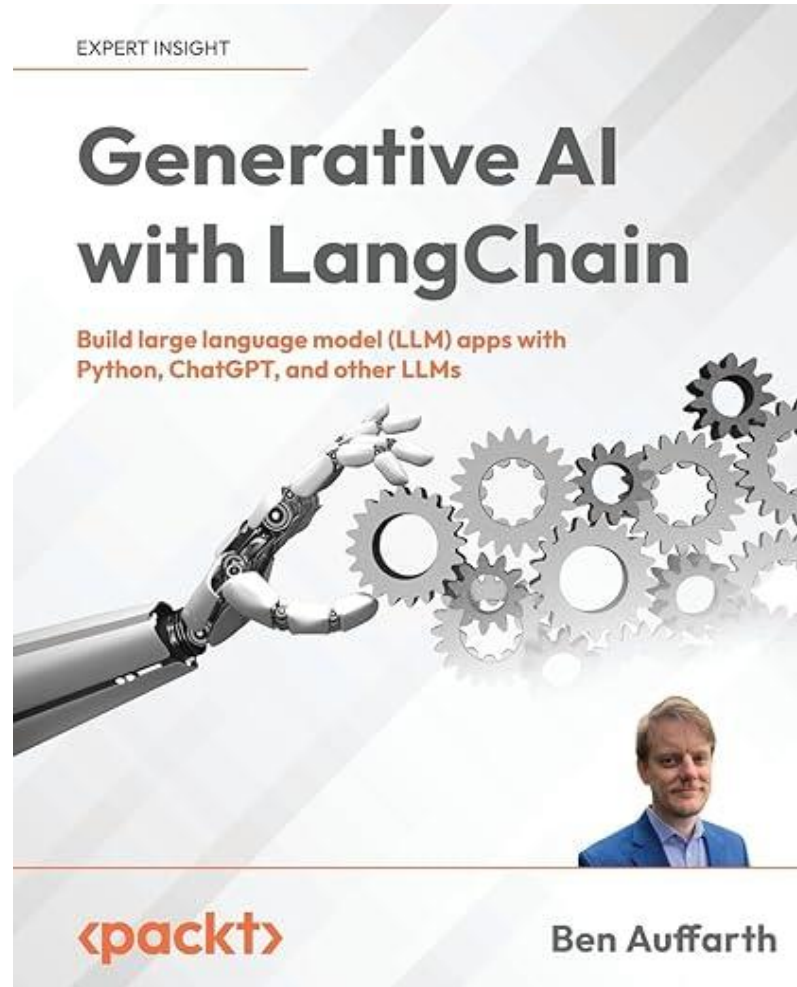
Chapter 12: Fine-tuning Generation Models



Ben Auffarth (2023),

# Generative AI with LangChain:

Build large language model (LLM) apps with Python, ChatGPT and other LLMs,  
Packt Publishing.

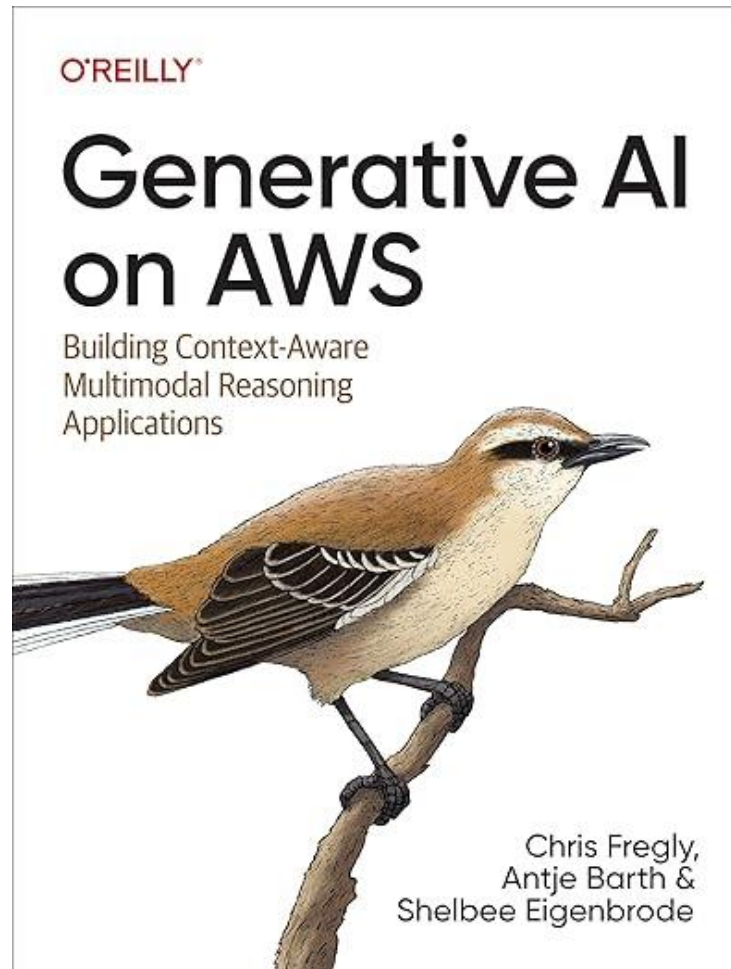


Chris Fregly, Antje Barth, and Shelbee Eigenbrode (2023),

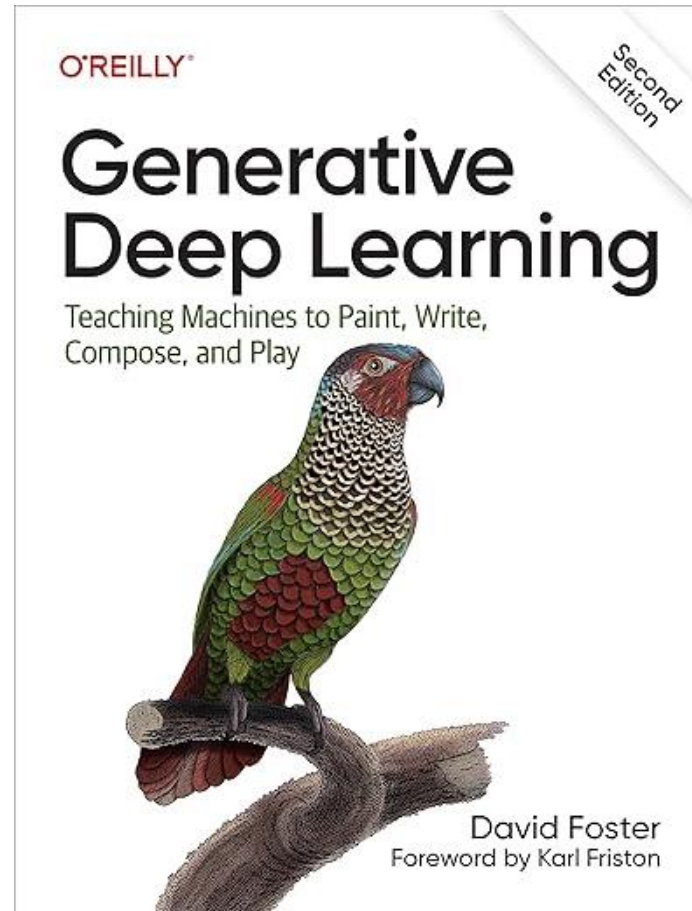
# Generative AI on AWS:

Building Context-Aware Multimodal Reasoning Applications,

O'Reilly Media



David Foster (2023),  
**Generative Deep Learning:**  
**Teaching Machines to Paint, Write, Compose, and Play,**  
2nd Edition, O'Reilly & Associates Inc



# **Generative AI**

# **Large Language Models**

# **(LLMs)**

# **Foundation Models**

# NVIDIA Developer Program

<https://developer.nvidia.com/join-nvidia-developer-program>

## NVIDIA

## Deep Learning Institute (DLI)

<https://learn.nvidia.com/>

# Join the NVIDIA Developer Program

take one of the  
complimentary  
technical self-  
paced courses  
(worth up to \$90)

8 hours

## Getting Started With Deep Learning

Explore the fundamentals of deep learning by training neural networks and using results to improve performance and capabilities.

2 hours

## Modeling Time-Series Data With Recurrent Neural Networks in Keras

Explore how to classify and forecast time-series data using recurrent neural networks (RNNs), such as modeling a patient's health over time.

4 hours

## Deploying a Model for Inference at Production Scale

Learn how to deploy your own machine learning models on a GPU server.

8 hours

## Building Real-Time Video AI Applications

Gain the knowledge and skills needed to enable the real-time transformation of raw video data from widely deployed camera sensors into deep learning-based insights.

2 hours

## Introduction to Graph Neural Networks

Learn the basic concepts, models, and applications of graph neural networks.

4 hours

## Introduction to Physics-Informed Machine Learning With Modulus

Learn the various building blocks of NVIDIA Modulus, which turbocharges use cases by building physics-based deep learning models that are 100,000X faster than traditional methods and offers high-fidelity simulation results.

2 hours

## Get Started With Highly Accurate Custom ASR for Speech AI

Learn to build, train, fine-tune, and deploy a GPU-accelerated automatic speech recognition (ASR) service with NVIDIA® Riva that includes customized features.

2 hours

## Integrating Sensors With NVIDIA DRIVE

Find out how to integrate automotive sensors into your applications using NVIDIA DRIVE®.

# NVIDIA Deep Learning Institute (DLI)

Self-Paced Course

**Generative AI Explained**

Free  
2 hours

Self-Paced Course

**Getting Started With Deep Learning**

Certificate available  
\$90  
8 hours

Instructor-Led Workshop

**Fundamentals of Deep Learning**

Certificate available  
\$500  
8 hours

Self-Paced Course

**Introduction to Transformer-Based Natural Language Processing**

Certificate available  
\$30  
6 hours

Self-Paced Course

**Building RAG Agents With LLMs**

Certificate available  
Free  
8 hours

Instructor-Led Workshop

**Building RAG Agents With LLMs**

Certificate available  
\$500  
8 hours

Self-Paced Course

**Generative AI with Diffusion Models**

Certificate available  
\$90  
8 hours

Instructor-Led Workshop

**Generative AI with Diffusion Models**

Certificate available  
\$500  
8 hours

## What do you want to learn today?

### Filters

Level +

Format +

Topics -

- Deep Learning
- Accelerated Computing
- Generative AI/LLM
- Graphics and Simulation
- OpenUSD
- Data Science
- NIMS
- NIM
- RAPIDS

Free / Paid +

Language +

Generative AI



Sort by: --

Showing 19 results

Generative AI x

# Generative AI

## All Courses

Self-paced

Generative AI Explained

Free  
02:00

Self-paced

Generative AI with Diffusion Models

\$90  
08:00

Instructor-Led

Generative AI with Diffusion Models

08:00

Self-paced

Augment your LLM Using

Self-paced

Introduction to Transformer-

Instructor-Led

Rapid Application

# Generative AI Explained

Self-paced Course

## Generative AI Explained

In this no-coding course, learn Generative AI concepts and applications, as well as the challenges and opportunities in this exciting field.

[About Course](#) [Objectives](#) [Topics Covered](#) [Course Outline](#) [Stay Informed](#) [Contact Us](#)[Continue Learning](#)

## About this Course

Generative AI describes technologies that are used to generate new content based on a variety of inputs. In recent time, Generative AI involves the use of neural networks to identify patterns and structures within existing data to generate new content. In this course, you will learn Generative AI concepts, applications, as well as the challenges and opportunities in this exciting field.

## Learning Objectives

Upon completion, you will have a basic understanding of Generative AI and be able to more effectively use the various tools built on this

## Course Details

**Duration:** 02:00**Price:** Free**Level:** Technical - Beginner**Subject:** Generative AI/LLM**Language:** English

[https://learn.nvidia.com/courses/course-detail?course\\_id=course-v1:DLI+S-FX-15+V1](https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-15+V1)

# Introduction to Transformer-Based Natural Language Processing

## Self-paced Course

## Introduction to Transformer-Based Natural Language Processing

Learn how Transformers are used as the building blocks of modern large language models (LLMs). You'll then use these models for various NLP tasks, including text classification, named-entity recognition (NER), author attribution, and question answering.

[About Course](#)[Objectives](#)[Topics Covered](#)[Course Outline](#)[Stay Informed](#)[Contact Us](#)[Continue Learning](#)

### About this Course

Large Language Models (LLMs), or Transformers, have revolutionized the field of natural language processing (NLP). Driven by recent advancements, applications of NLP and generative AI have exploded in the past decade. With the proliferation of applications like chatbots and intelligent virtual assistants, organizations are infusing their businesses with more interactive human-machine experiences. Understanding how Transformer-based large language models (LLMs) can be used to manipulate, analyze, and generate text-based data is essential. Modern pre-trained LLMs can encapsulate the nuance, context, and sophistication of language, just as humans do. When fine-tuned and deployed correctly, developers can use these LLMs to build powerful NLP applications that provide natural and seamless human-computer interactions within chatbots, AI voice agents, and more. In this course, you'll learn how Transformers are used as the building blocks of modern large language models (LLMs). You'll then use these models for various NLP

### Course Details

**Duration:** 06:00**Price:** \$30**Level:** Technical - Beginner**Subject:** Generative AI/LLM**Language:** English

[https://learn.nvidia.com/courses/course-detail?course\\_id=course-v1:DLI+S-FX-08+V1](https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-08+V1)

# Building RAG Agents with LLMs

Deep Learning Institute Find Training Self Paced Courses Instructor-Led Workshops Educator Programs Enterprise Solutions Certification Resources

Self-paced Course

## Building RAG Agents with LLMs

Agents powered by large language models (LLMs) have shown great retrieval capability for using tools, looking at documents, and plan their approaches. This course will show you how to deploy an agent system in practice with the flexibility to scale up your system to meet the demands of users and customers.



About Course Objectives Topics Covered Course Outline Stay Informed Contact Us

Continue Learning

## About this Course

This course is free for a limited time.

The evolution and adoption of large language models (LLMs) have been nothing short of revolutionary, with retrieval-based systems at the forefront of this technological leap. These models are not just tools for automation; they are partners in enhancing productivity, capable of holding informed conversations by interacting with a vast array of tools and documents. This course is designed for those eager to explore the potential of these systems, focusing on practical deployment and the efficient implementation required to manage the considerable demands of both users and deep learning models. As we delve into the intricacies of LLMs, participants will gain insights into advanced orchestration techniques that include internal reasoning, dialog management, and effective tooling strategies.

## Course Details

**Duration:** 08:00

**Price:** Free

**Level:** Technical - Intermediate

**Subject:** Generative AI/LLM

**Language:** English

**Course Prerequisites:**

Introductory deep learning knowledge, with comfort

[https://learn.nvidia.com/courses/course-detail?course\\_id=course-v1:DLI+S-FX-15+V1](https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-15+V1)

# Generative AI with Diffusion Models

Deep Learning Institute Find Training Self Paced Courses Instructor-Led Workshops Educator Programs Enterprise Solutions Certification Resources

Self-paced Course

## Generative AI with Diffusion Models

Take a deeper dive into denoising diffusion models, which are a popular choice for text-to-image pipelines, with applications in creative content generation, data augmentation, simulation and planning, anomaly detection, drug discovery, personalized recommendations, and more.

About Course Objectives Topics Covered Course Outline Stay Informed Contact Us

Continue Learning

## About this Course

Thanks to improvements in computing power and scientific theory, generative AI is more accessible than ever before. Generative AI plays a significant role across industries due to its numerous applications, such as creative content generation, data augmentation, simulation and planning, anomaly detection, drug discovery, personalized recommendations, and more. In this course, learners will take a deeper dive into denoising diffusion models, which are a popular choice for text-to-image pipelines.

## Learning Objectives

[https://learn.nvidia.com/courses/course-detail?course\\_id=course-v1:DLI+S-FX-14+V1](https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1)

## Course Details

**Duration:** 08:00

**Price:** \$90

**Subject:** Generative AI/LLM

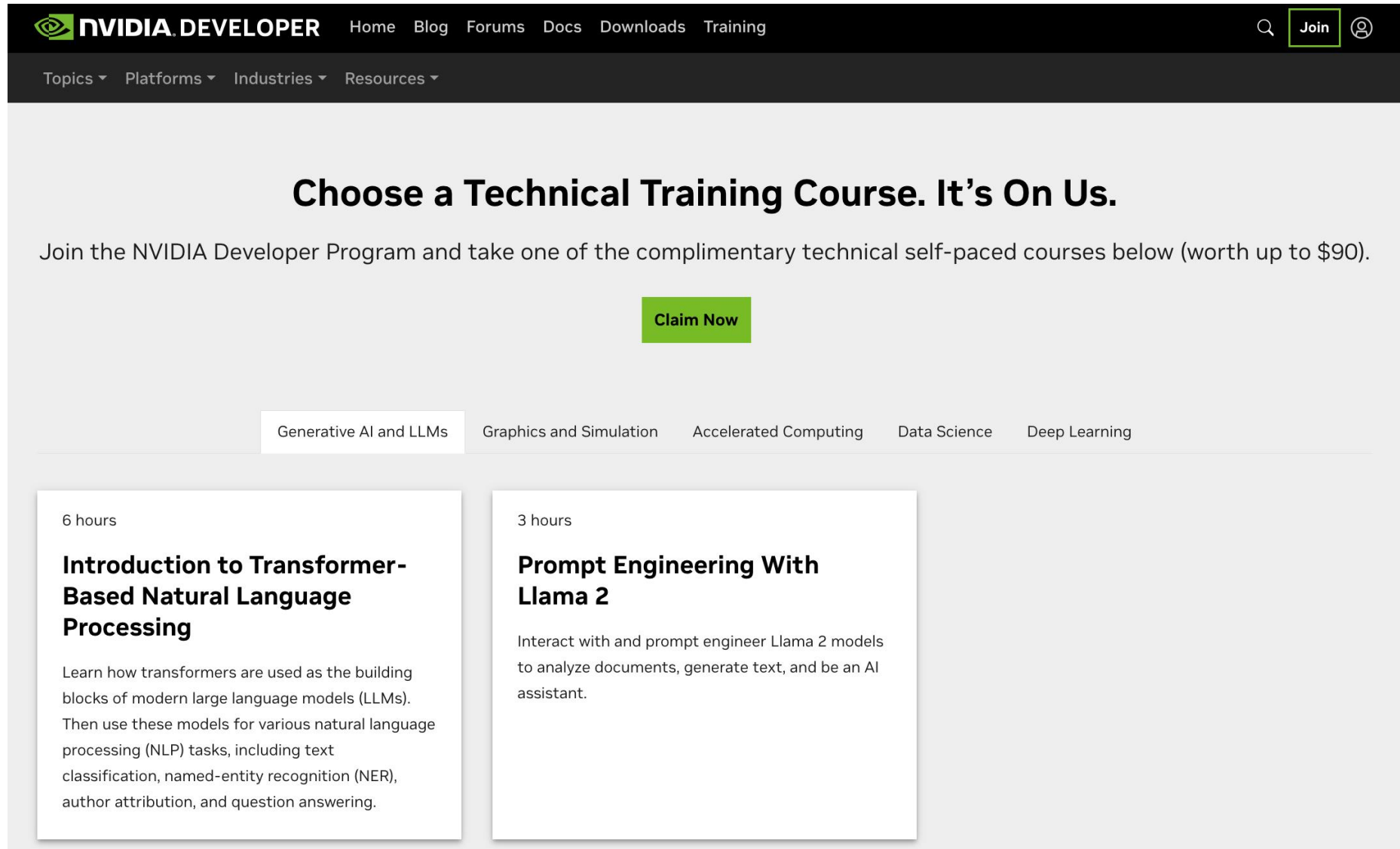
**Language:** English

**Course Prerequisites:**

A basic understanding of [Deep Learning Concepts](#).

# Join the NVIDIA Developer Program

take one of the complimentary technical self-paced courses (worth up to \$90)

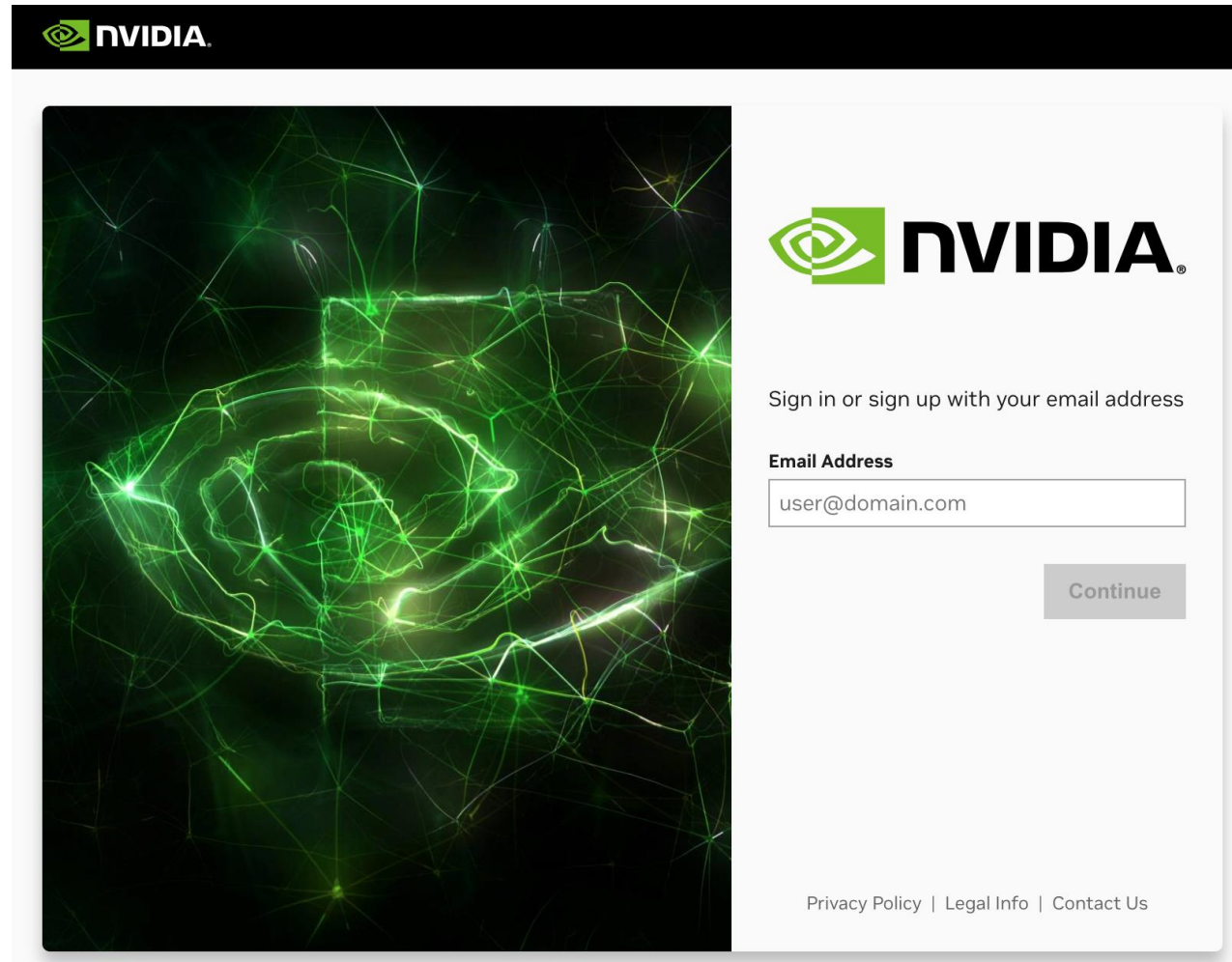


The screenshot shows the NVIDIA Developer Program website. At the top, there is a navigation bar with the NVIDIA logo, the text "NVIDIA DEVELOPER", and links for Home, Blog, Forums, Docs, Downloads, and Training. A search icon and a "Join" button are also present. Below the navigation bar, there are dropdown menus for Topics, Platforms, Industries, and Resources. The main content area features a large heading "Choose a Technical Training Course. It's On Us." followed by a sub-heading "Join the NVIDIA Developer Program and take one of the complimentary technical self-paced courses below (worth up to \$90)." and a prominent green "Claim Now" button. Below this, there are five category tabs: "Generative AI and LLMs" (which is selected), "Graphics and Simulation", "Accelerated Computing", "Data Science", and "Deep Learning". Two course cards are displayed under the "Generative AI and LLMs" tab. The first card is titled "Introduction to Transformer-Based Natural Language Processing" and is 6 hours long. The second card is titled "Prompt Engineering With Llama 2" and is 3 hours long. The description for the second card reads: "Interact with and prompt engineer Llama 2 models to analyze documents, generate text, and be an AI assistant."

<https://developer.nvidia.com/join-nvidia-developer-program>

# Join the NVIDIA Developer Program

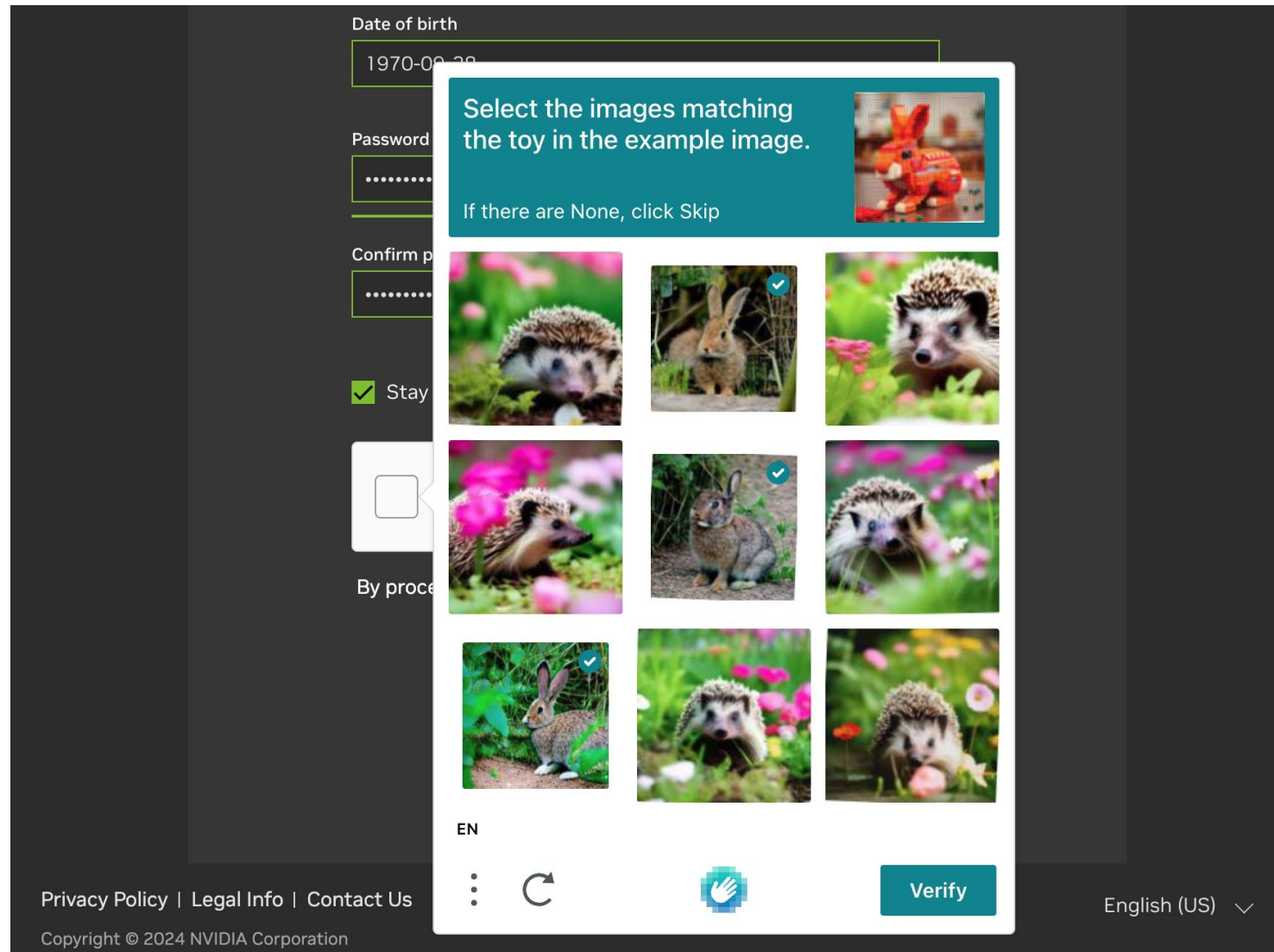
take one of the complimentary technical self-paced courses (worth up to \$90)



The screenshot shows the NVIDIA Developer Program sign-up page. It features a black header with the NVIDIA logo. The main content area is split into two columns. The left column contains a large, abstract image of a glowing green neural network or data visualization. The right column contains the NVIDIA logo, the text "Sign in or sign up with your email address", an "Email Address" label, a text input field containing "user@domain.com", and a "Continue" button. At the bottom of the right column, there are links for "Privacy Policy", "Legal Info", and "Contact Us".

# Join the NVIDIA Developer Program

take one of the complimentary technical self-paced courses (worth up to \$90)



The screenshot displays a registration form with fields for "Date of birth" (1970-00-20), "Password", and "Confirm password". A "Stay" checkbox is checked. A CAPTCHA challenge is overlaid, asking the user to "Select the images matching the toy in the example image." The example image shows a red toy rabbit. The challenge grid contains nine images: three rabbits and six hedgehogs. Three rabbits are selected with blue checkmarks. At the bottom of the CAPTCHA window, there is a "Verify" button and a language dropdown set to "English (US)".

Date of birth  
1970-00-20

Password  
.....

Confirm password  
.....

Stay

By proceeding, I agree to the [Terms of Service](#) and [Privacy Policy](#).

EN

Privacy Policy | Legal Info | Contact Us

Copyright © 2024 NVIDIA Corporation

English (US) ▾

# Join the NVIDIA Developer Program

take one of the complimentary technical self-paced courses (worth up to \$90)

 NVIDIA DEVELOPER DEEP LEARNING INSTITUTE PROGRAM BENEFITS

## SELECT YOUR FREE COURSE.

Thank you for your participation in the NVIDIA Developer Program. Please select your free DLI course below.

English

Integrating Sensors with NVIDIA DRIVE®

Getting Started with Deep Learning

Deploying a Model for Inference at Production Scale

Get Started with Highly Accurate Custom ASR for Speech AI

Introduction to Graph Neural Networks

Introduction to Transformer-Based Natural Language Processing

Prompt Engineering with LLaMA-2 (Access Expires Dec. 5th 2025)

Generative AI with Diffusion Models

Building Real-Time Video AI Applications

Introduction to Robotic Simulations in Isaac Sim

### Generative AI with Diffusion Models

Take a deeper dive into denoising diffusion models, which are a popular choice for text-to-image pipelines, with applications in creative content generation, data augmentation, simulation and planning, anomaly detection, drug discovery, personalized recommendations, and more.

 Certificate Available

 Duration: 08:00

Continue >

# Join the NVIDIA Developer Program

take one of the complimentary technical self-paced courses (worth up to \$90)

 NVIDIA DEVELOPER DEEP LEARNING INSTITUTE PROGRAM BENEFITS

## SELECT YOUR FREE COURSE.


Thank you for your participation in the NVIDIA Developer Program. Please select your free DLI course below.

English ▾


- Modeling Time Series Data with Recurrent Neural Networks in Keras (Access ends 10/16/2024)
- Optimizing CUDA Machine Learning Codes With Nsight Profiling Tools
- Getting Started with Accelerated Computing in CUDA C/C++
- Fundamentals of Accelerated Computing with CUDA Python
- Fundamentals of Accelerated Computing with OpenACC
- Integrating Sensors with NVIDIA DRIVE®
- Getting Started with Deep Learning**
- Deploying a Model for Inference at Production Scale
- Get Started with Highly Accurate Custom ASR for Speech AI

### Getting Started with Deep Learning

Learn how deep learning works through hands-on exercises in computer vision and natural language processing.

 Certificate Available

---

 Duration: 08:00

[Continue >](#)

# Join the NVIDIA Developer Program

take one of the complimentary technical self-paced courses (worth up to \$90)

 NVIDIA DEVELOPER DEEP LEARNING INSTITUTE PROGRAM BENEFITS

## SELECT YOUR FREE COURSE.

Thank you for your participation in the NVIDIA Developer Program. Please select your free DLI course below.

English

Get Started with Highly Accurate Custom ASR for Speech AI

Introduction to Graph Neural Networks

Introduction to Transformer-Based Natural Language Processing

Prompt Engineering with LLaMA-2 (Access Expires Dec. 5th 2025)

Generative AI with Diffusion Models

Building Real-Time Video AI Applications

Introduction to Robotic Simulations in Isaac Sim

Introduction to Physics-informed Machine Learning with Modulus

Essentials of USD in Omniverse: Access Expires 09/18/2025

Synthetic Data Generation for Training Computer Vision Models

### Generative AI with Diffusion Models

Take a deeper dive into denoising diffusion models, which are a popular choice for text-to-image pipelines, with applications in creative content generation, data augmentation, simulation and planning, anomaly detection, drug discovery, personalized recommendations, and more.


 Certificate Available

 Duration: 08:00

Continue >

# NVIDIA Deep Learning Institute (DLI)

**Deep Learning Institute** Find Training Self Paced Courses Instructor-Led Workshops Educator Programs Enterprise Solutions Certification Resources


Search 

### Monthly Activity

Skill Points	0
Time Spent	
Courses in Progress	1
Courses Completed	0
Watched Videos	
Assessments	

### Skills

### Certificates

  
No Certificates  
You don't have any certificates yet.

## Courses in Progress

Self-paced

**Generative AI with Diffusion Models**

---

0% Completed  
08:00

## Generative AI with Diffusion Models

[Course](#) [Progress](#) [Bookmarks](#) [Updates](#)

Generative AI with Diffusion Models &gt; Start Here &gt; 0: Server Access

Generative AI with Diffusion Models

[Start Here](#)[Next Steps](#)[Feedback](#)[Previous](#)[Next](#)

### 0: Server Access

[Bookmark this page](#)

Welcome to Generative AI with Diffusion Models. Please click "Next" below to get started.

Underneath each video is a link to start your own private server for hands-on coding practice. Click the "Start" button to boot up the server. In a few minutes after the server is done loading, click "Launch" to access the code labs.

#### 1: From U-Nets to Diffusion

[Bookmark this page](#)

Theory

<https://learn.nvidia.com/my-learning>

**Deep Learning Institute** Find Training Self Paced Courses Instructor-Led Workshops Educator Programs Enterprise Solutions Certification Resources

### Monthly Activity

Skill Points	0
Time Spent	
Courses in Progress	9
Courses Completed	8

### Skills

### Certificates



Introduction to Transformer-Based Natural Language Processing



Building RAG Agents with LLMs



Building RAG Agents with LLMs

### Completed Courses

View more

<p>Self-paced</p> <p><b>Sizing LLM Inference Systems</b></p> <hr/> <p>100% Completed 03:00</p>	<p>Self-paced</p> <p><b>Augment your LLM Using Retrieval Augmented Generation</b></p> <hr/> <p>100% Completed 01:00</p>	<p>Self-paced</p> <p><b>Building RAG Agents with LLMs</b></p> <hr/> <p>100% Completed 08:00</p>	<p>Self-paced</p> <p><b>Generative AI Explained</b></p> <hr/> <p>100% Completed 02:00</p>	<p>Self-paced</p> <p><b>Introduction to Transform Based Natural Language Processing</b></p> <hr/> <p>100% Completed 06:00</p>
--	---	---	---	---

## All Self-Paced Courses

Accelerated Computing Data Science Deep Learning **Generative AI/LLM** Graphics and Simulation Infrastructure

[Share Generative AI/LLM Courses](#)

<p>Self-paced</p> <p><b>Generative AI Explained</b></p> <p>Free 02:00</p>	<p>Self-paced</p> <p><b>Introduction to NVIDIA NIM™ Microservices</b></p> <p>Free 02:00</p>	<p>Self-paced</p> <p><b>Introduction to Deploying RAG Pipelines for Production at Scale</b></p> <p>\$90 03:00</p>	<p>Self-paced</p> <p><b>Generative AI with Diffusion Models</b></p> <p>\$90 08:00</p>
<p>Self-paced</p> <p><b>Techniques for Improving the Effectiveness of RAG Systems</b></p> <p>\$30 03:00</p>	<p>Self-paced</p> <p><b>Introduction to Transformer- Based Natural Language Processing</b></p> <p>\$30 06:00</p>	<p>Self-paced</p> <p><b>Building LLM Applications With Prompt Engineering</b></p> <p>\$90 08:00</p>	<p>Self-paced</p> <p><b>Synthetic Tabular Data Generation Using Transformers</b></p> <p>\$30 04:00</p>
<p>Self-paced</p> <p><b>Sizing LLM Inference Systems</b></p> <p>Free 03:00</p>	<p>Self-paced</p> <p><b>Building RAG Agents with LLMs</b></p> <p>Free 08:00</p>	<p>Self-paced</p> <p><b>Augment your LLM Using Retrieval Augmented Generation</b></p> <p>Free 01:00</p>	

# NVIDIA Deep Learning Institute (DLI)



## Certificate of Competency

This certificate is awarded to

**Min-Yuh Day**

for demonstrating competence in the completion of

**Introduction to Transformer-Based Natural Language Processing**

A handwritten signature in black ink, appearing to read "Greg Estes", written over a horizontal line.

**Greg Estes**

Vice President, NVIDIA

Issue Date: : December 5, 2024

Certification ID: twpWsrB4TDCQ0ErgAoEt6w |

[https://learn.nvidia.com/certificates?id=twpWsrB4TDCQ0ErgAoEt6w/courses/course?course\\_id=course-v1:DLI+S-FX-08+V1](https://learn.nvidia.com/certificates?id=twpWsrB4TDCQ0ErgAoEt6w/courses/course?course_id=course-v1:DLI+S-FX-08+V1)

<https://learn.nvidia.com/certificates?id=twpWsrB4TDCQ0ErgAoEt6w>



## Certificate of Completion

This certificate is awarded to

**Min-Yuh Day**

for successfully completing

**Building RAG Agents with LLMs**

A handwritten signature in black ink, appearing to read "Greg Estes", written over a horizontal line.

**Greg Estes**

Vice President, NVIDIA

Issue Date: : December 8, 2024

Certification ID: ed-qOCIMQatzU8SNUNxgw |

[https://learn.nvidia.com/certificates?id=ed-qOCIMQatzU8SNUNxgw/courses/course?course\\_id=course-v1:DLI+S-FX-15+V1](https://learn.nvidia.com/certificates?id=ed-qOCIMQatzU8SNUNxgw/courses/course?course_id=course-v1:DLI+S-FX-15+V1)

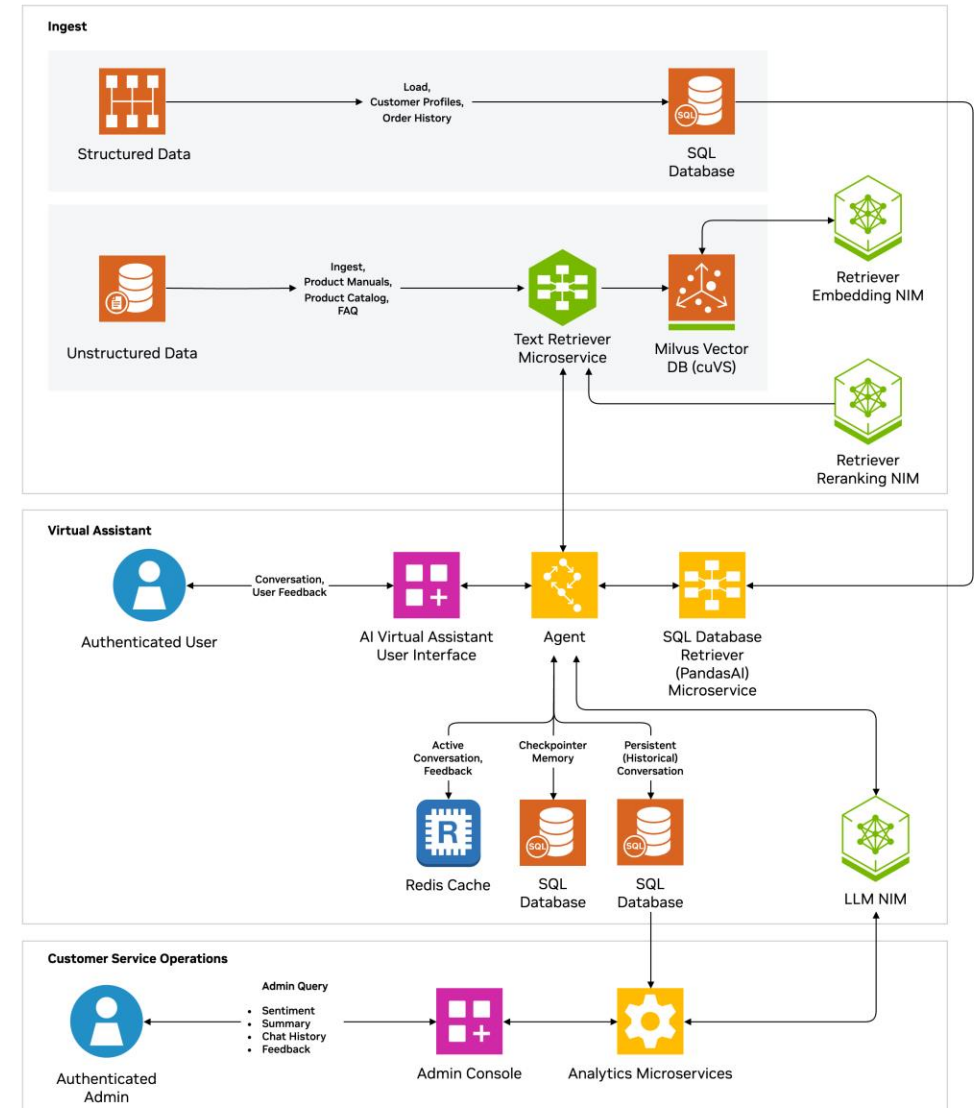
<https://learn.nvidia.com/certificates?id=ed-qOCIMQatzU8SNUNxgw>

# AI Virtual Assistant for Customer Service

AI Virtual Assistant to reduce handling time, boost customer satisfaction

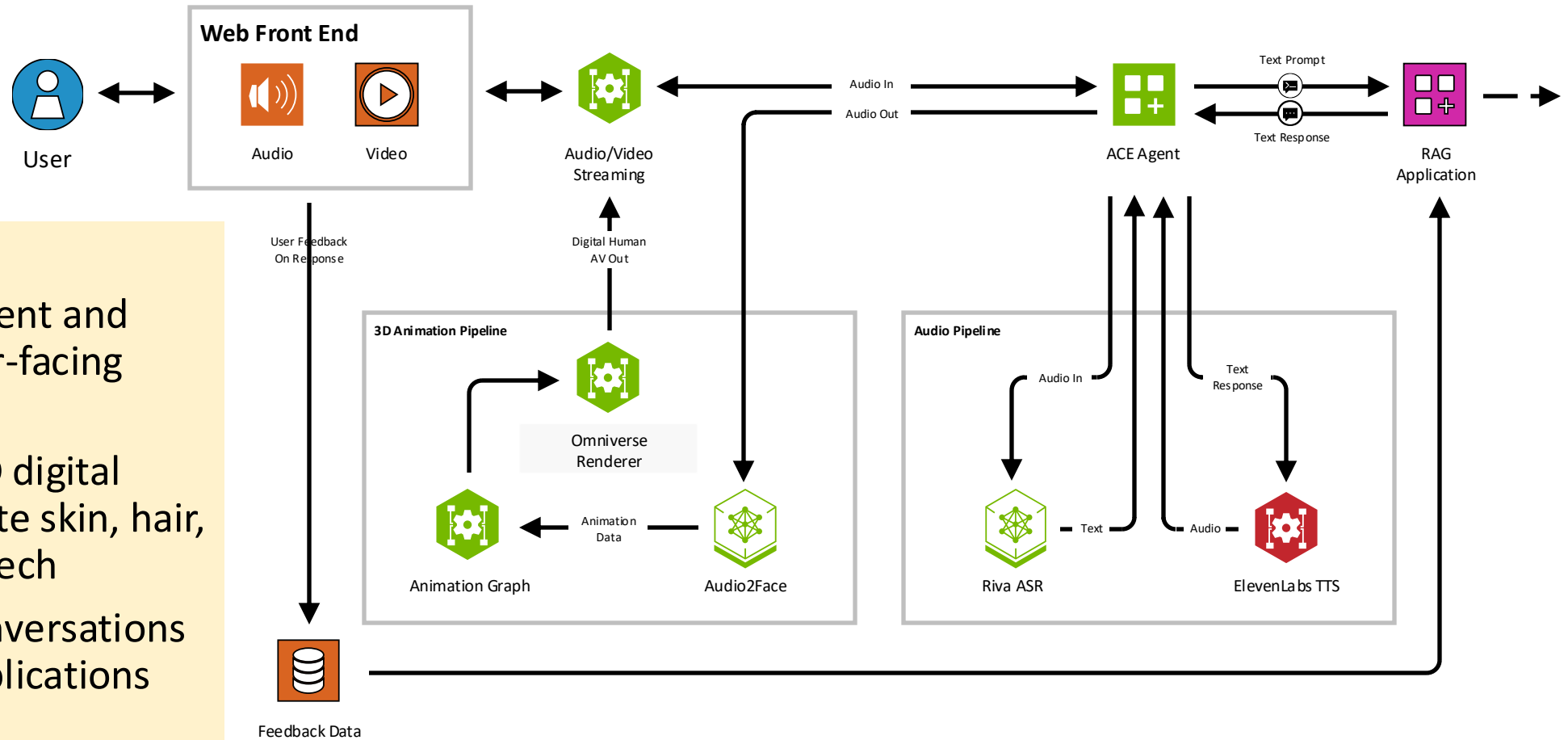
## Benefits

- **Personalized Responses:** Handles structured and unstructured customer queries (e.g., order details, spending history).
- **Multi-Turn Dialogue:** Offers context-aware, seamless interactions across multiple questions.
- **Custom Conversation Style:** Adapts text responses to reflect corporate branding and tone.
- **Sentiment Analysis:** Analyzes real-time customer interactions to gauge sentiment and adjust responses.
- **Multi-Session Support:** Allows for multiple user sessions with conversation history and summaries.
- **Data Privacy:** Integrates with on-premises or cloud-hosted knowledge bases to protect sensitive data.



# Digital Humans for Customer Service

\$125B market for digital human economy by 2035

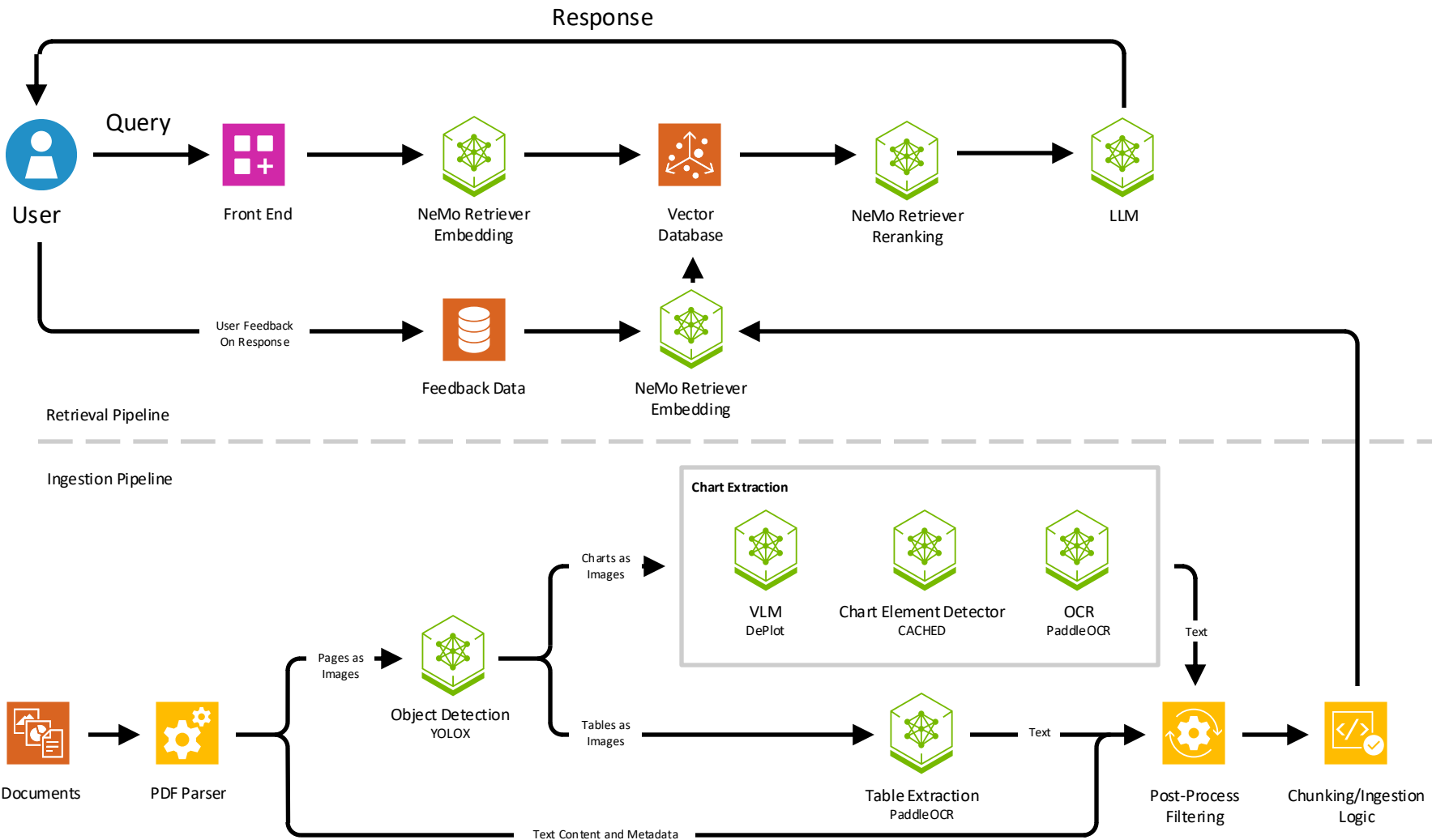


## Benefits

- Increases engagement and satisfaction for user-facing applications
- Creates a lifelike 3D digital human with accurate skin, hair, animation, and speech
- Enables natural conversations with enterprise applications and data

# Multimodal PDF Data Extraction for Enterprise RAG

## Unlocks Knowledge from trillions of PDFs

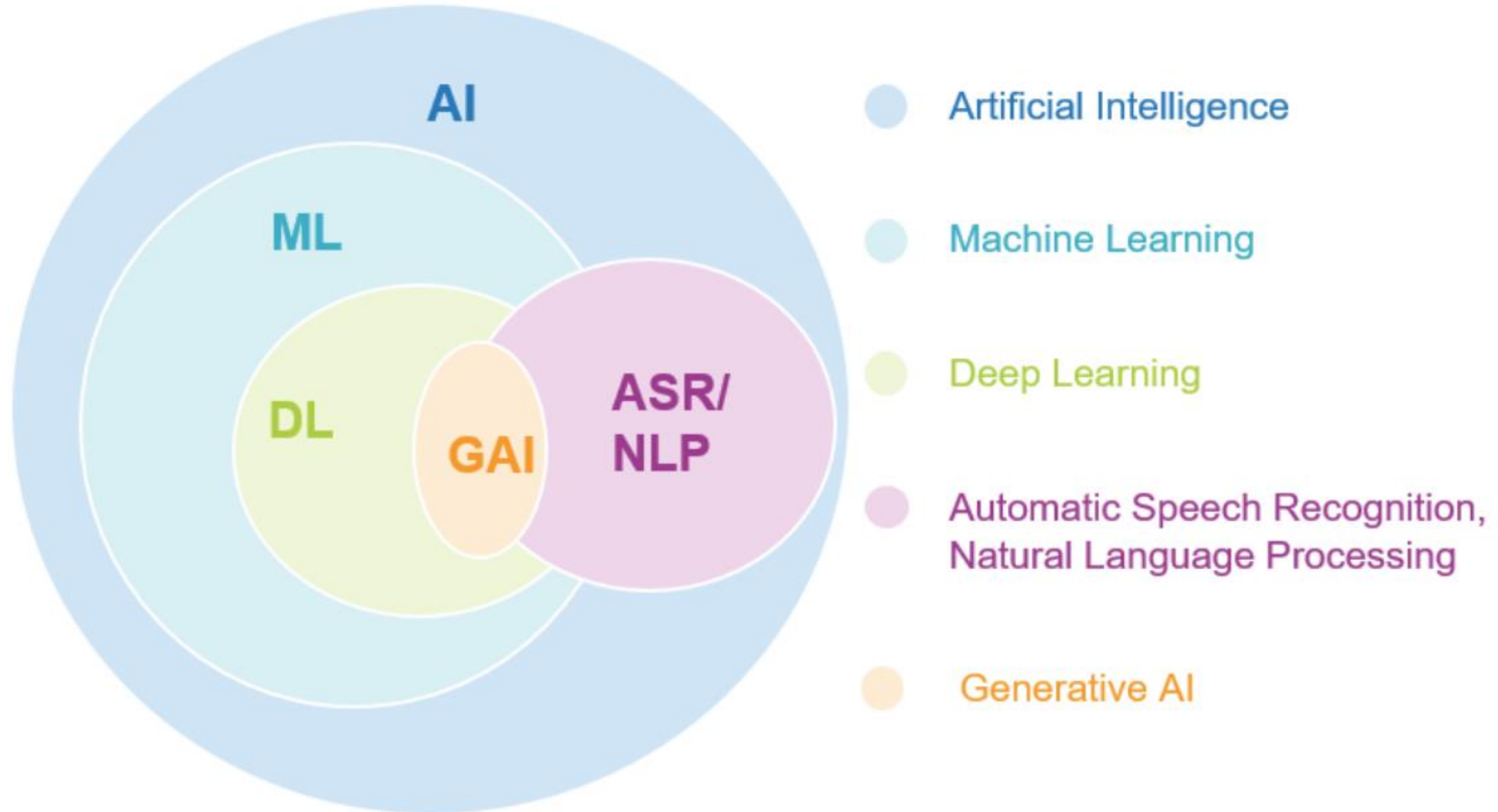


### Benefits

- Unlocks the next level of indexable enterprise data from text to images and charts
- High-accuracy extraction and responses
- Enterprise-scale PDF ingestion

# Artificial Intelligence (AI)

# AI, ML, DL, Generative AI



**Generative AI,  
Agentic AI,  
AI Agent,  
RAG LLM  
for  
QA and Dialogue Systems**

**Chatbot**

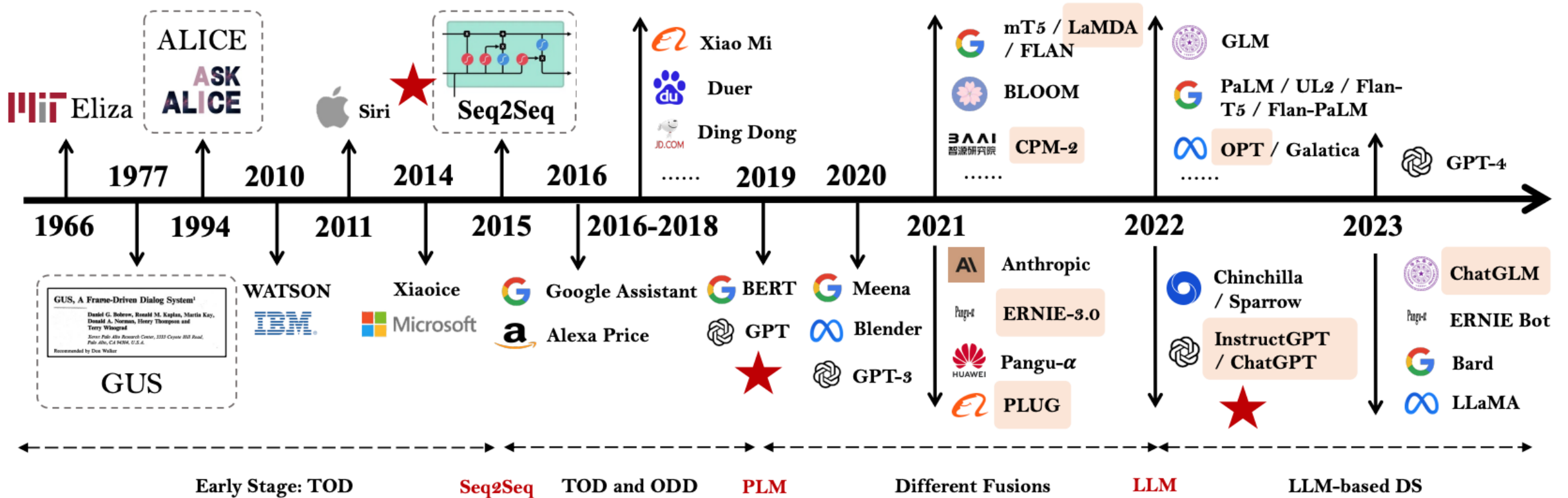
**Dialogue System**

**Intelligent Agent**

**Conversational AI**

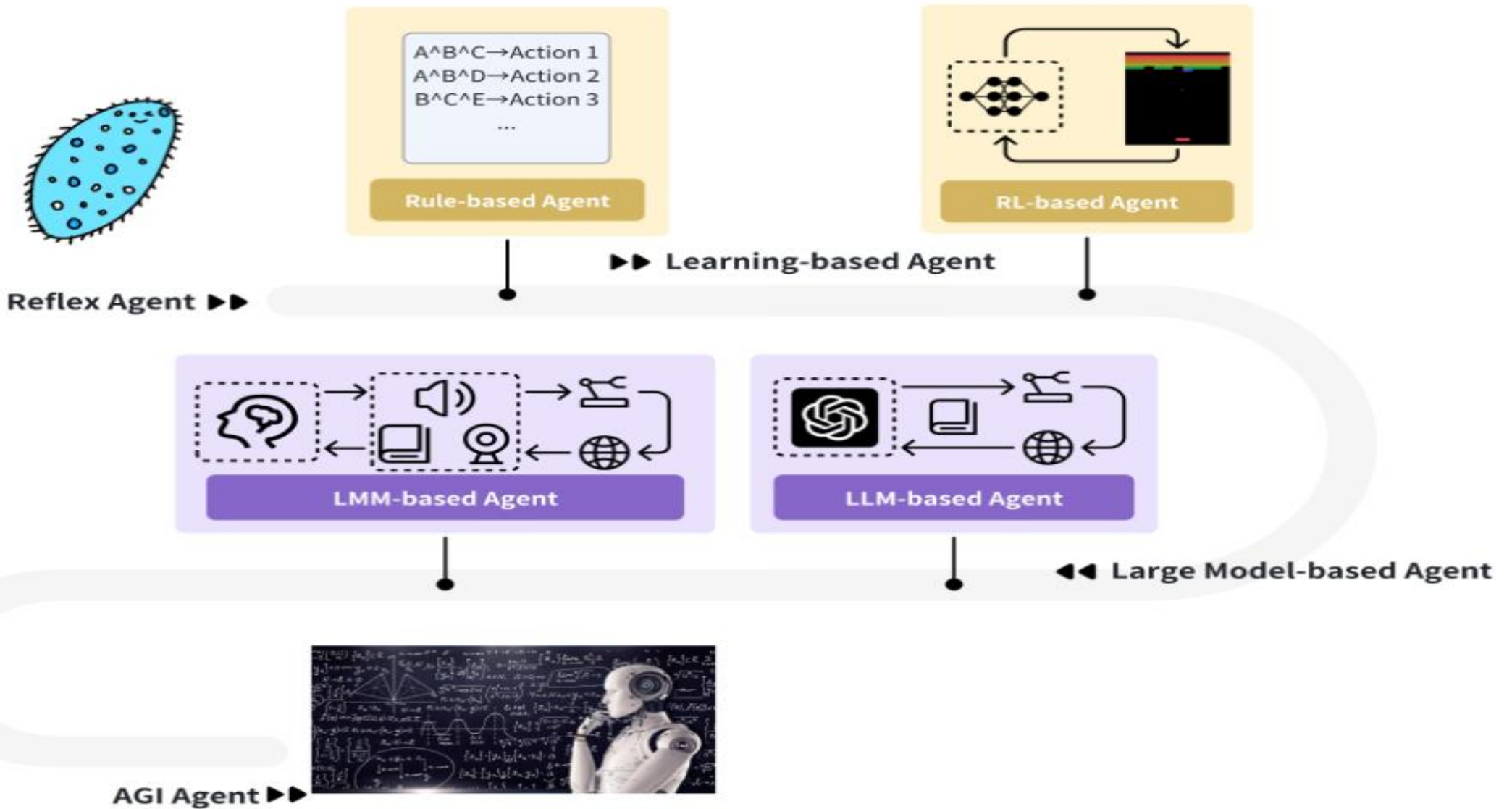
# The Development of LM-based Dialogue Systems

- 1) Early Stage (1966 - 2015)
- 2) The Independent Development of TOD and ODD (2015 - 2019)
- 3) Fusions of Dialogue Systems (2019 - 2022)
- 4) LLM-based DS (2022 - Now)



Task-oriented DS (TOD), Open-domain DS (ODD)

# Intelligent Agents Roadmap



# AI Agents

- **Traditional AI Agents**

- **Simple reflex agents**
- **Model-based reflex agents**
- **Goal-based agents**
- **Utility-based agents**
- **Learning agents**

- **Evolution of AI Agents**

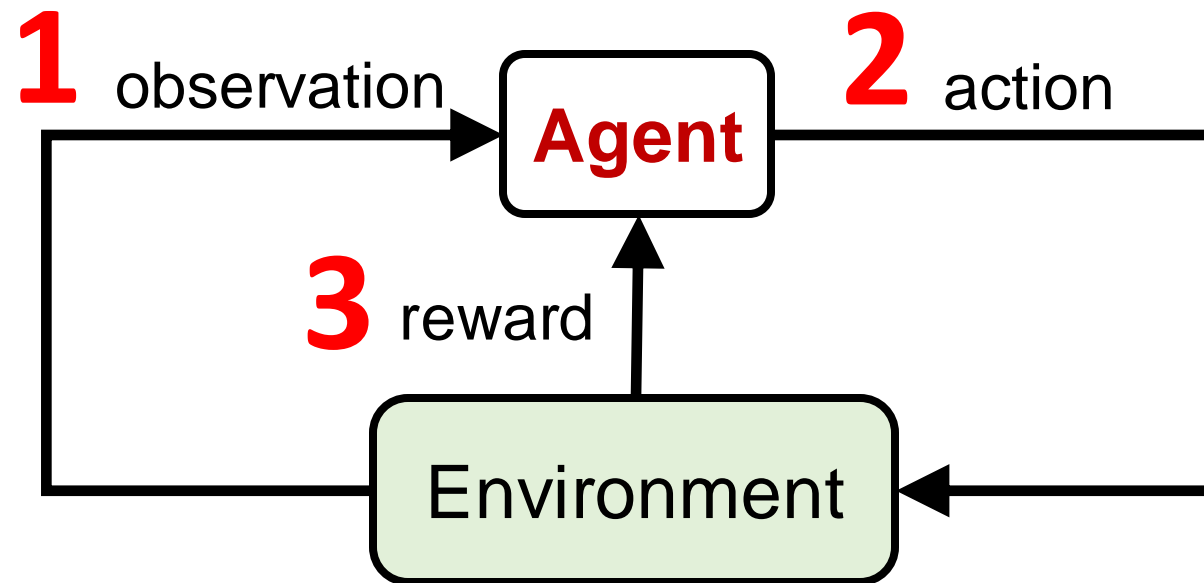
- **LLM-based Agents**
- **Multi-modal agents**
- **Embodied AI agents in virtual environments**
- **Collaborative AI agents**

# Reinforcement Learning (DL)

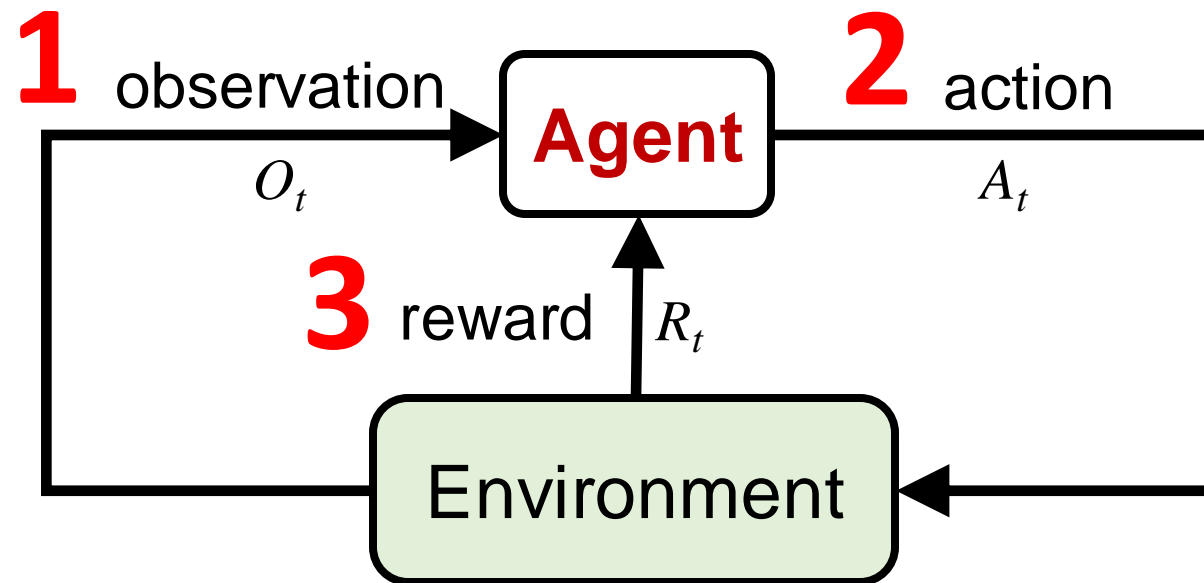
**Agent**

Environment

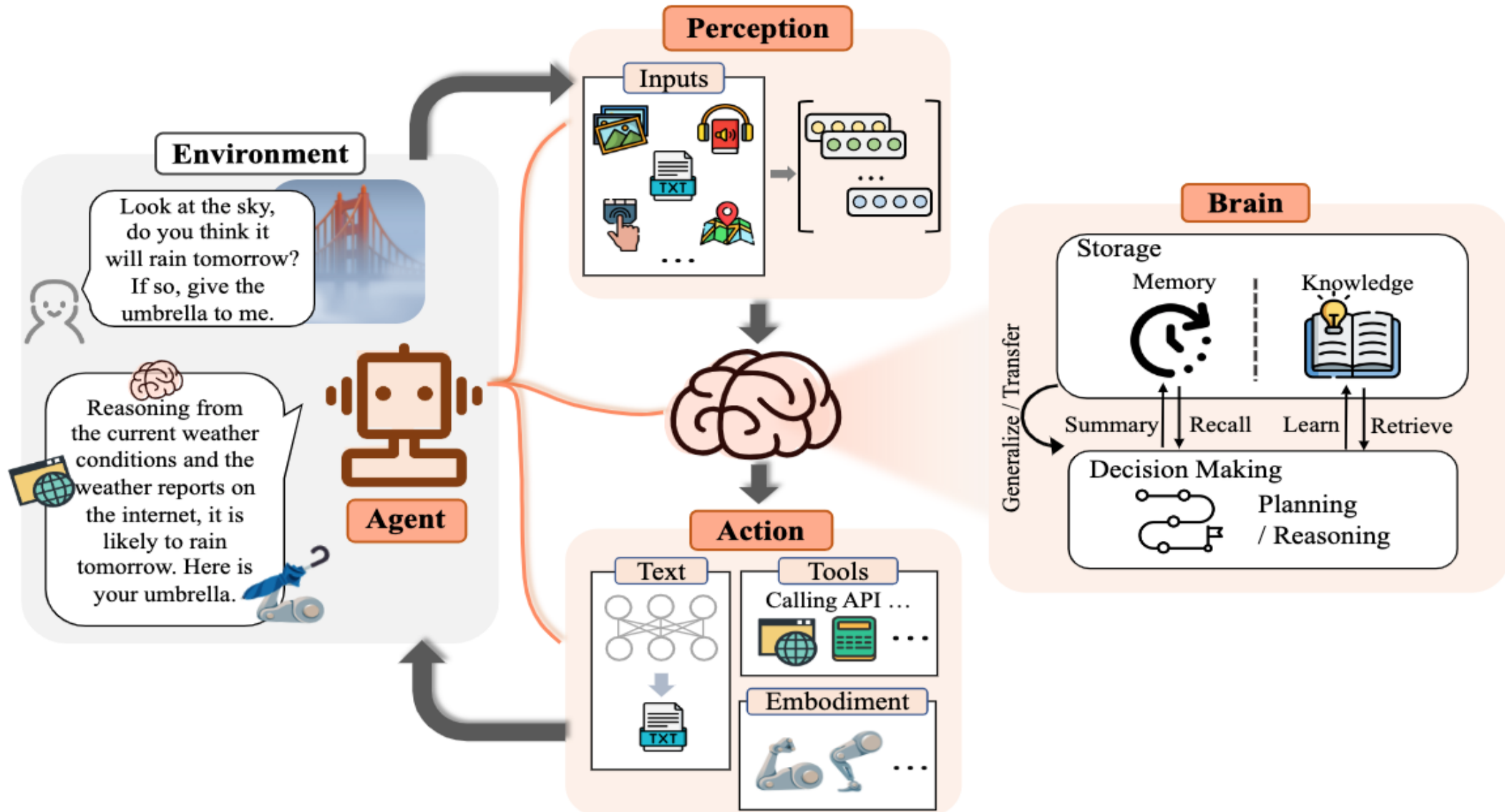
# Reinforcement Learning (DL)



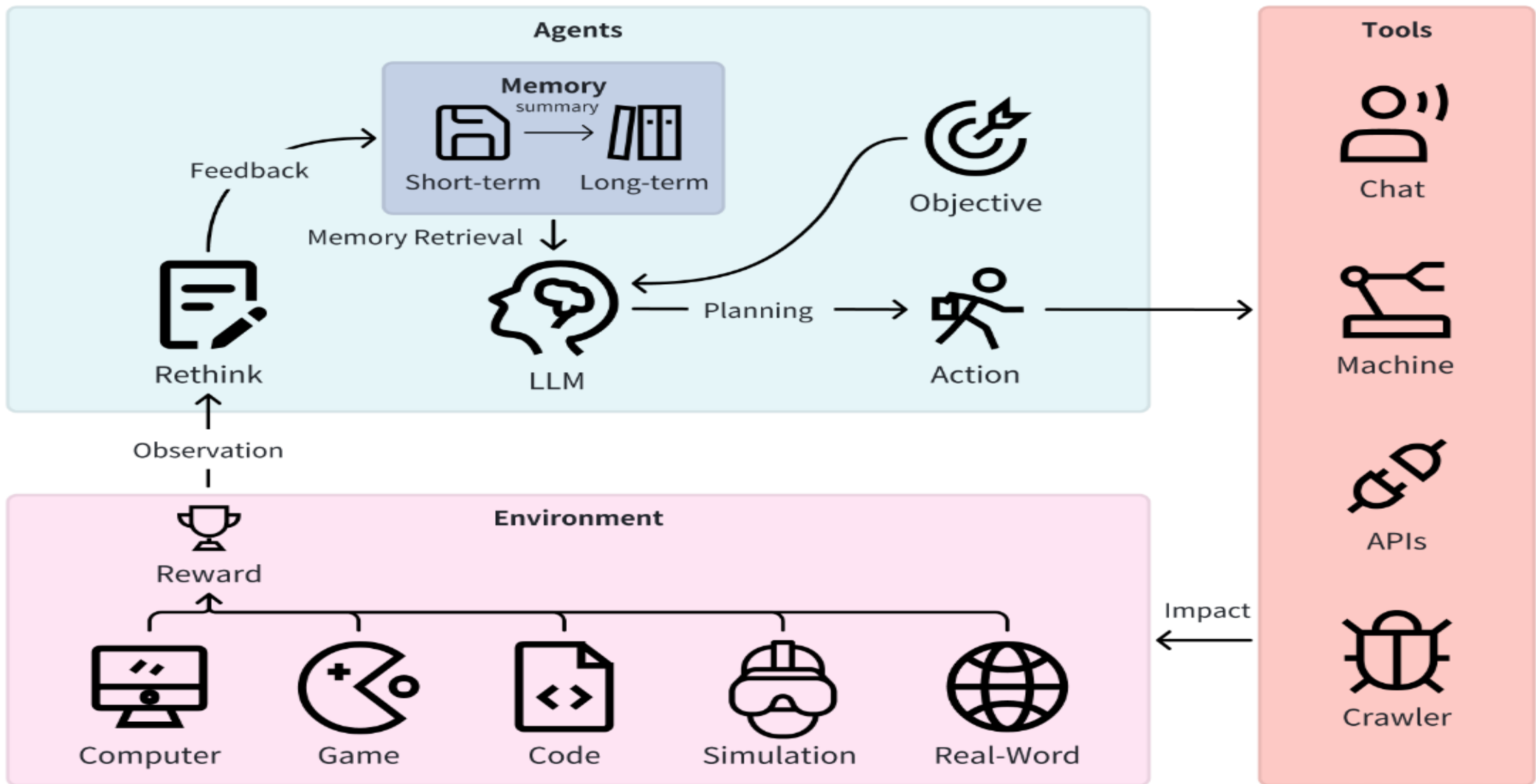
# Reinforcement Learning (DL)



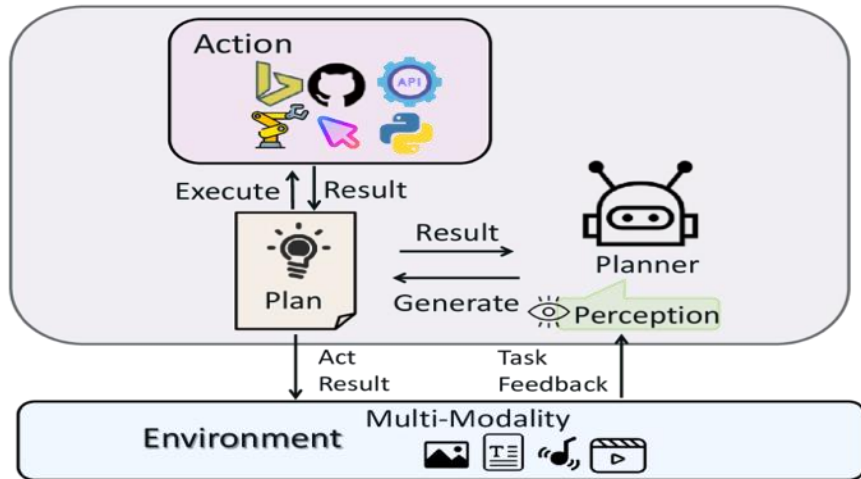
# Large Language Model (LLM) based Agents



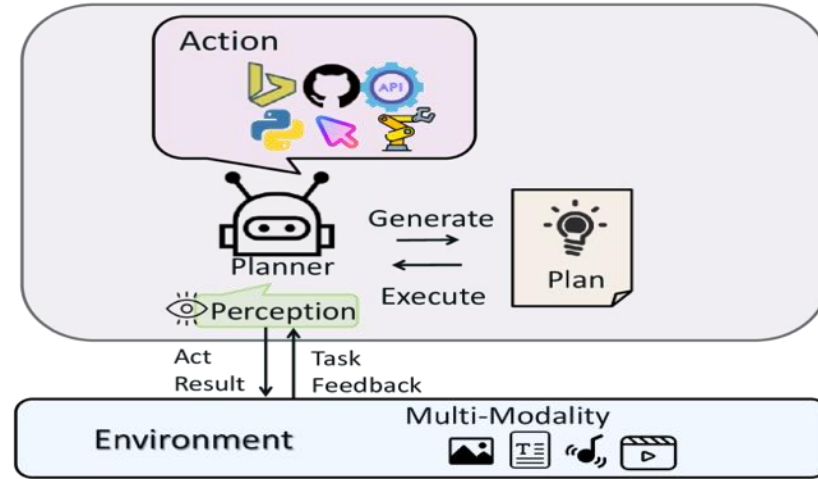
# LLM-based Agents



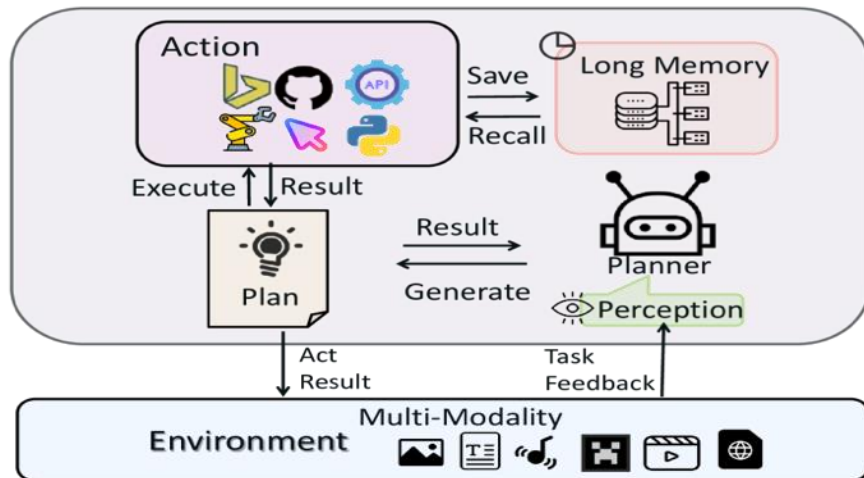
# Large Multimodal Agents (LMA)



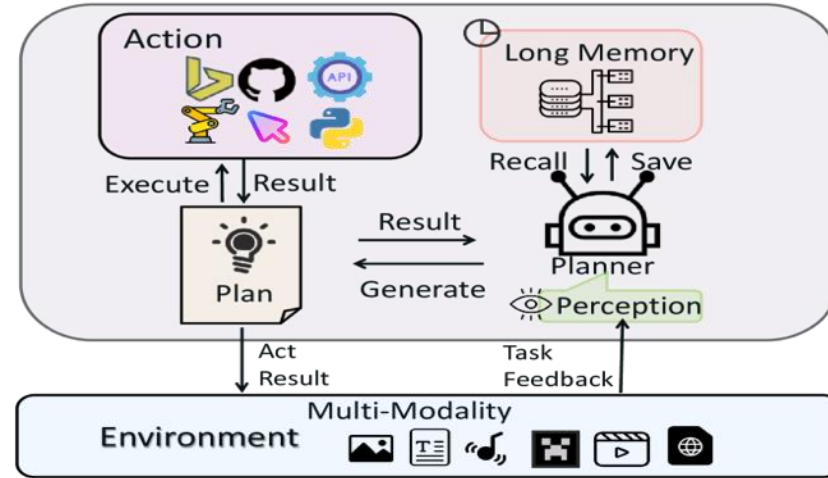
(a)



(b)

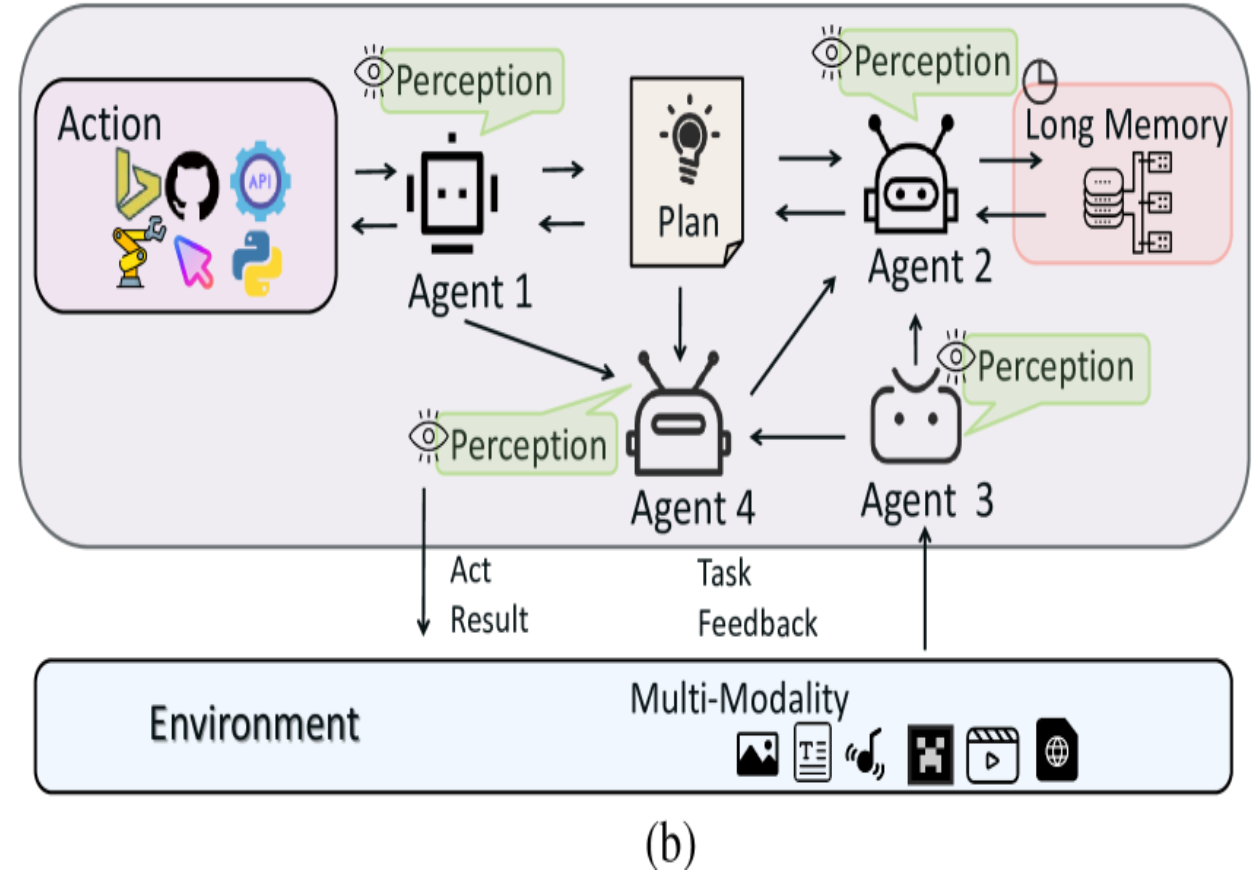
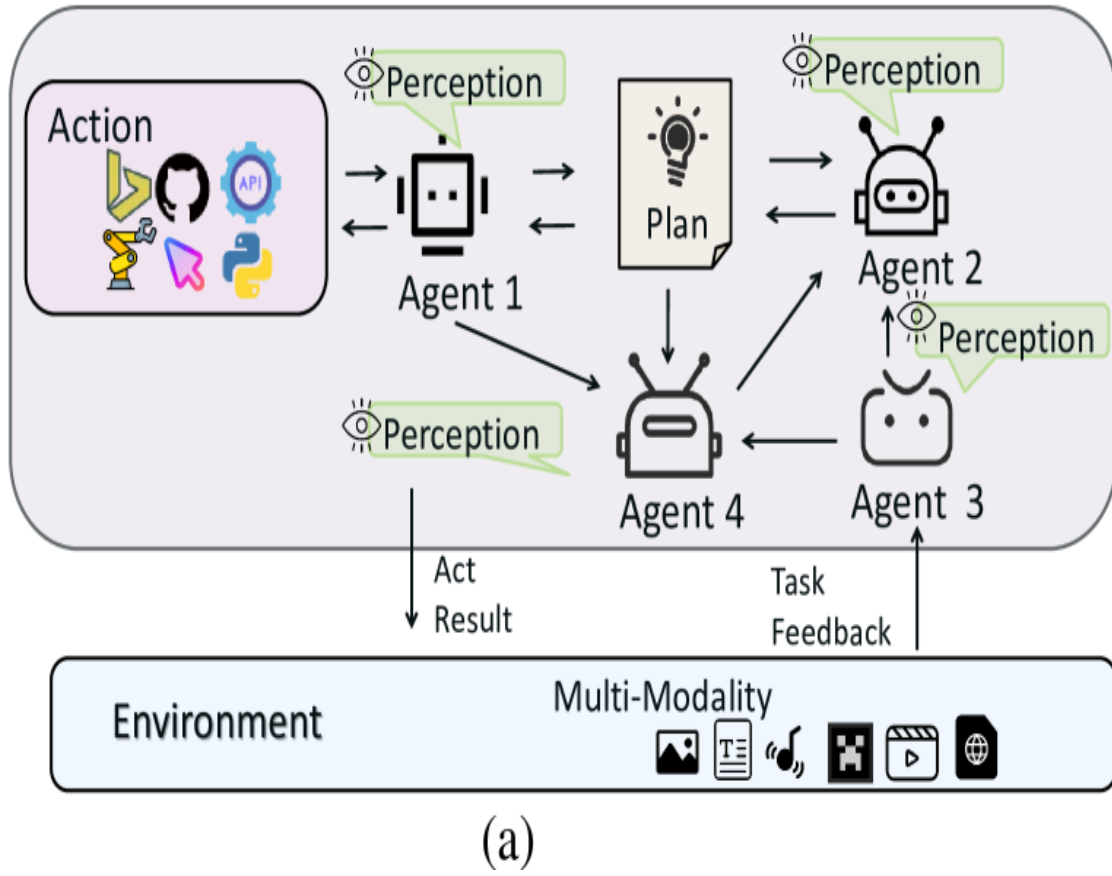


(c)

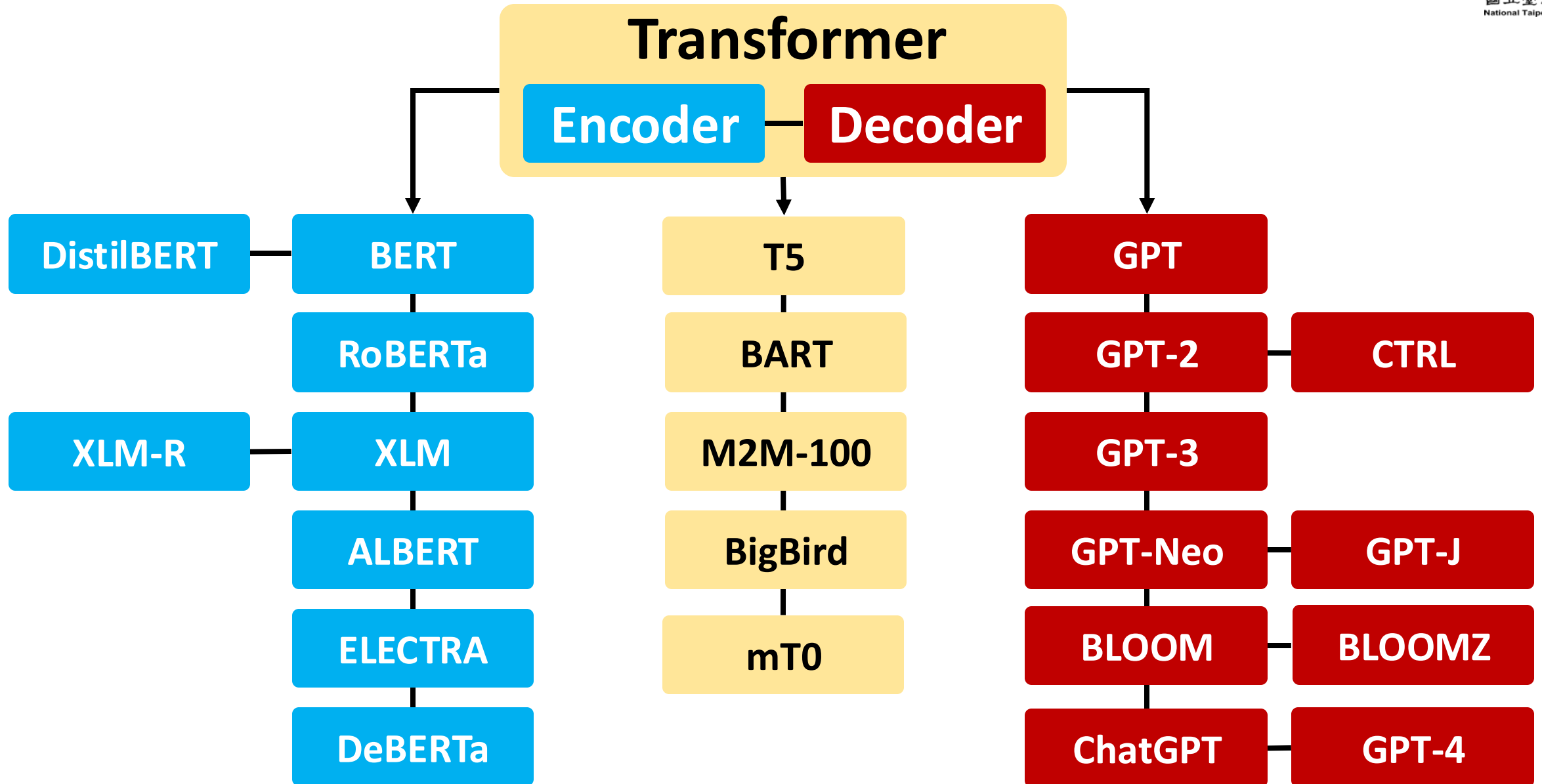


(d)

# Large Multimodal Agents (LMA)

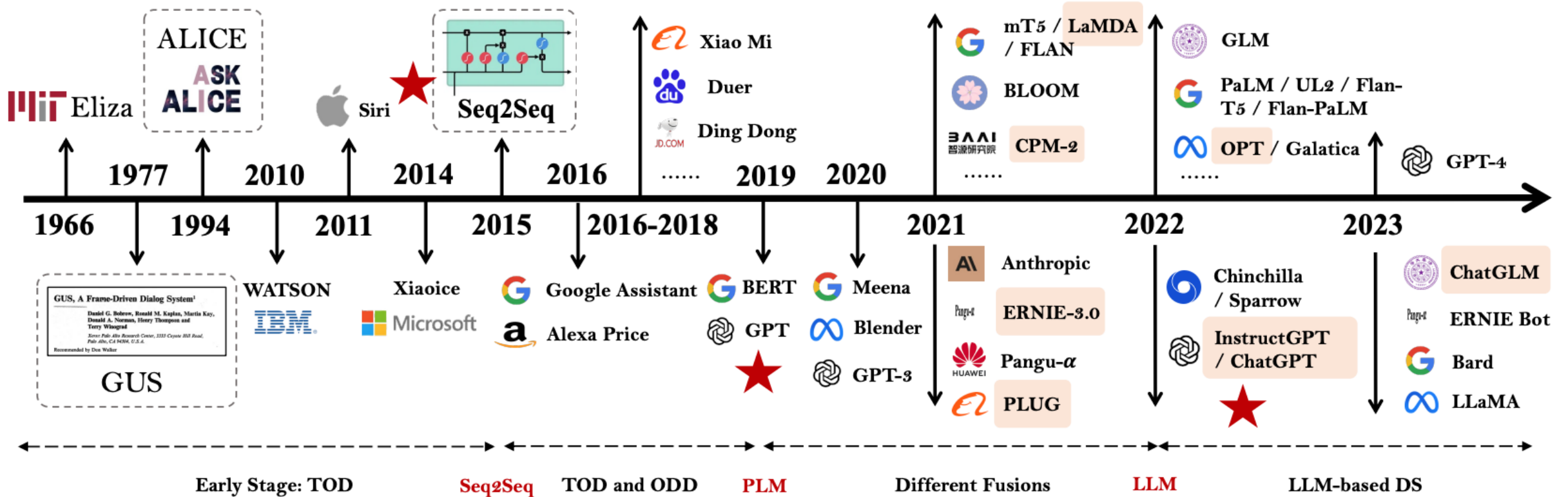


# Transformer Models



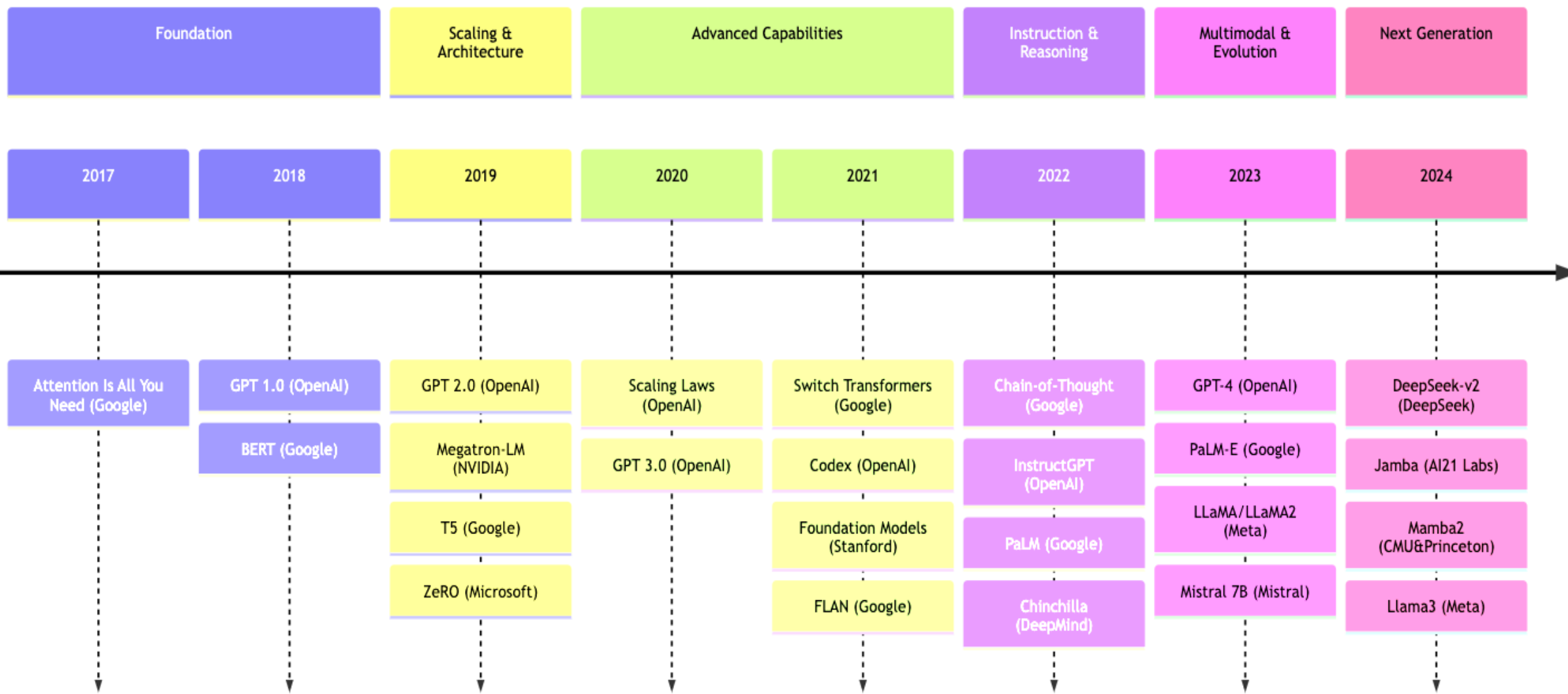
# The Development of LM-based Dialogue Systems

- 1) Early Stage (1966 - 2015)
- 2) The Independent Development of TOD and ODD (2015 - 2019)
- 3) Fusions of Dialogue Systems (2019 - 2022)
- 4) LLM-based DS (2022 - Now)

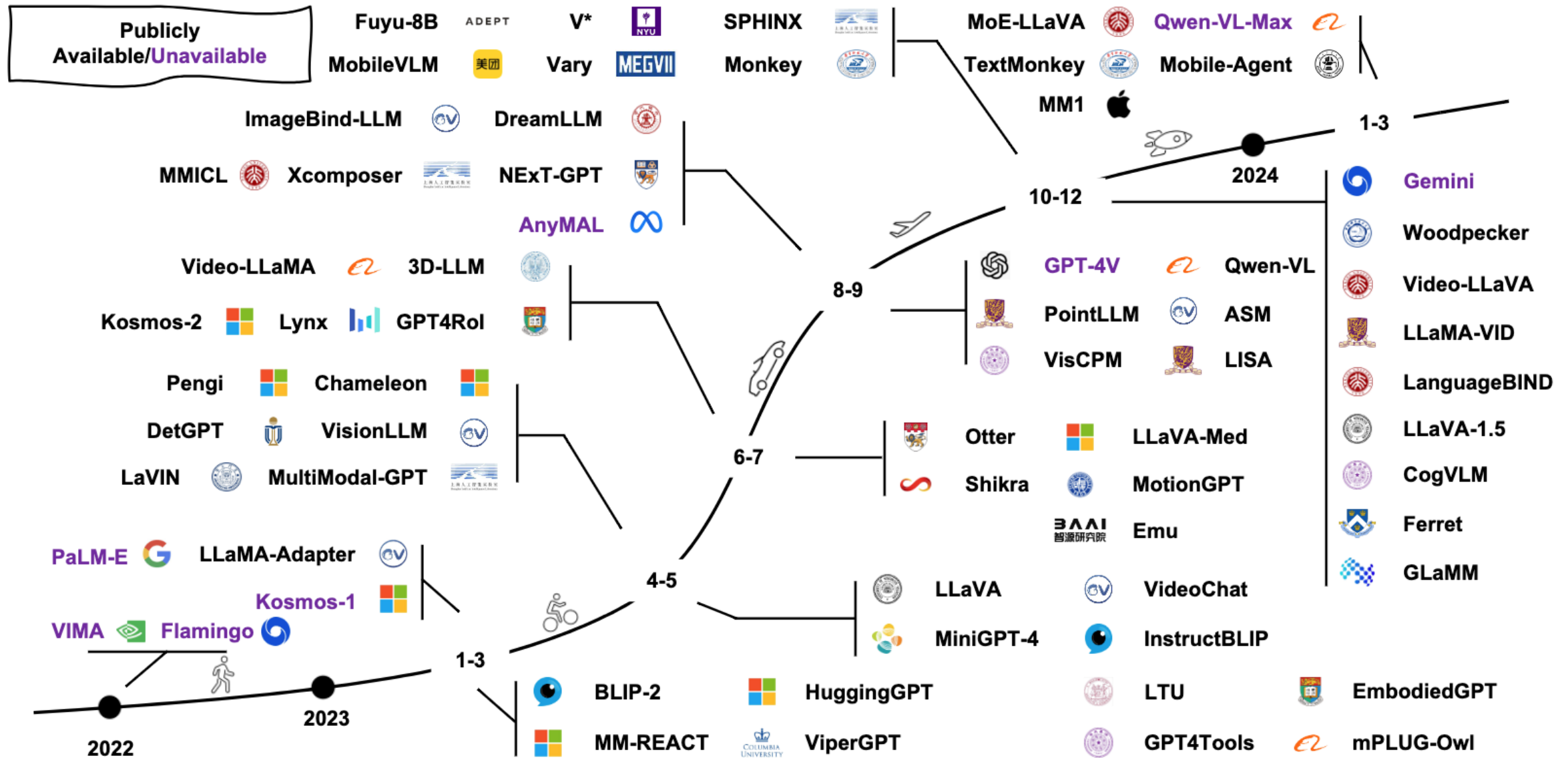


Task-oriented DS (TOD), Open-domain DS (ODD)

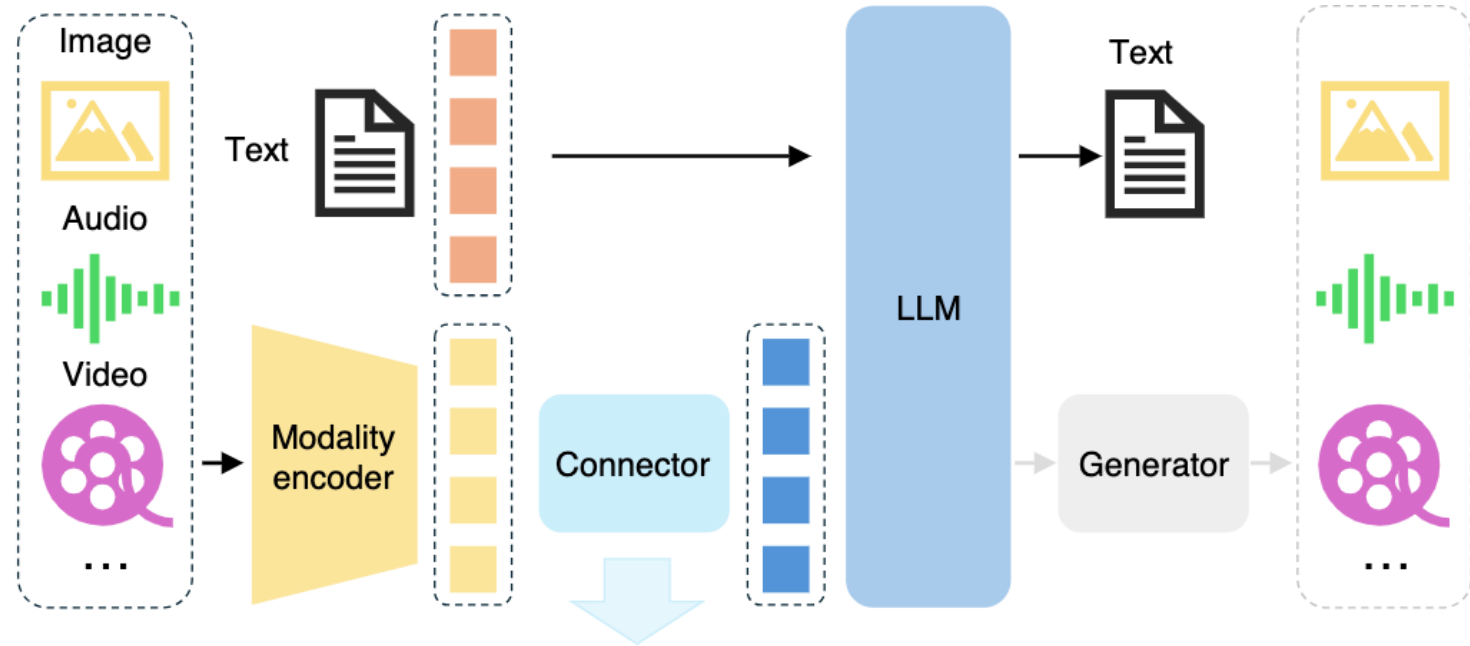
# Major GenAI LLMs Research Milestones (2017-2024)



# Multimodal Large Language Models (MLLM)



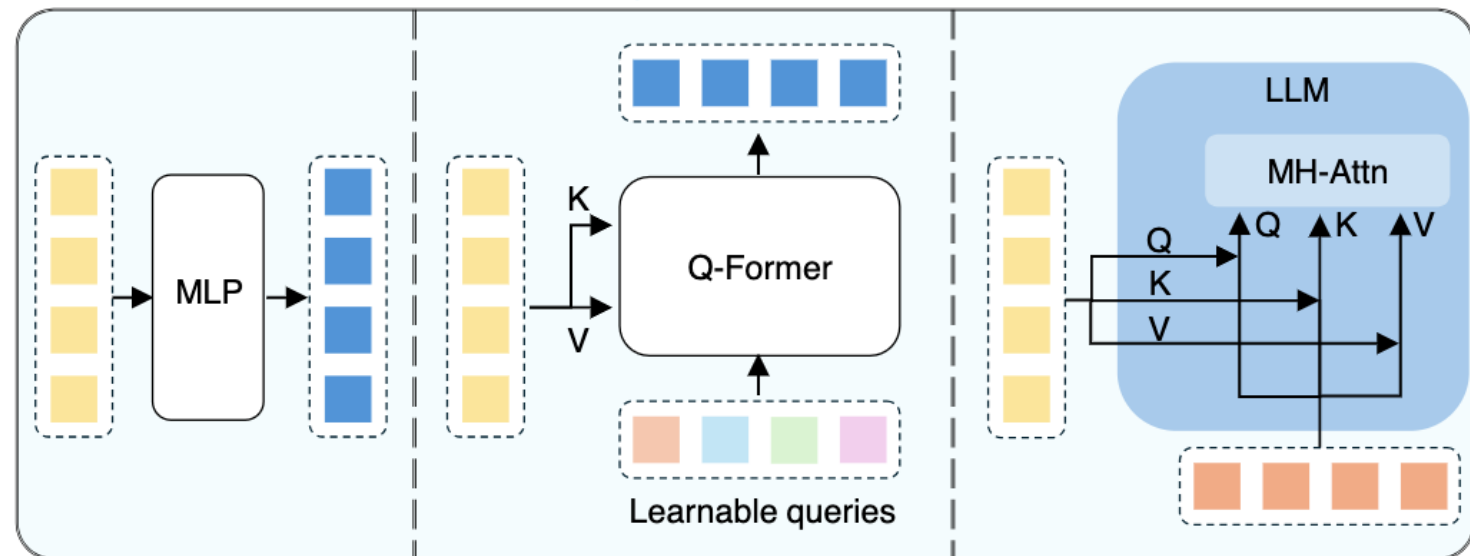
# Multimodal Large Language Models (MLLM)



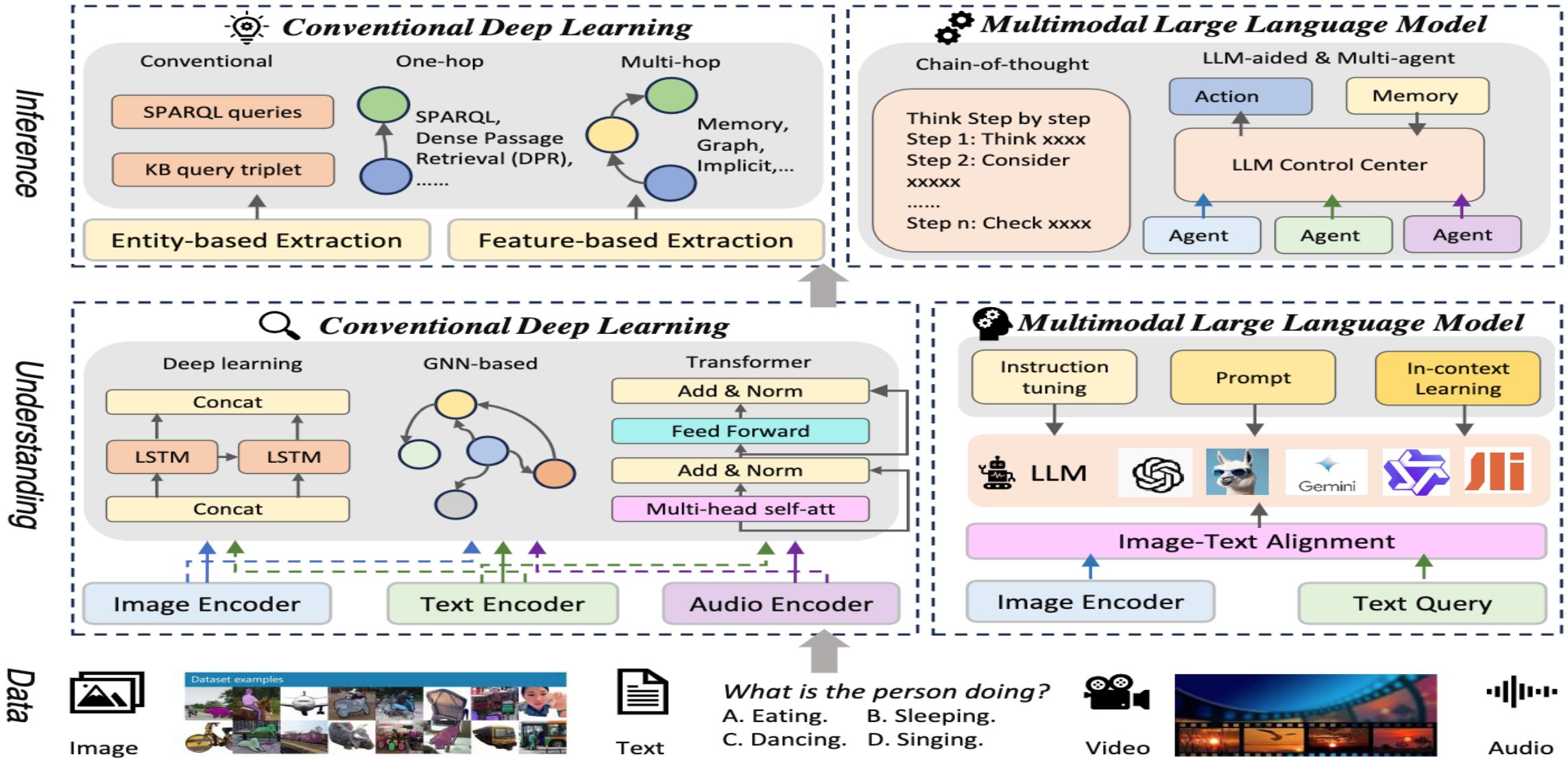
## Multimodal LLM

Three types of connectors:

1. projection-based
2. query-based
3. fusion-based connectors



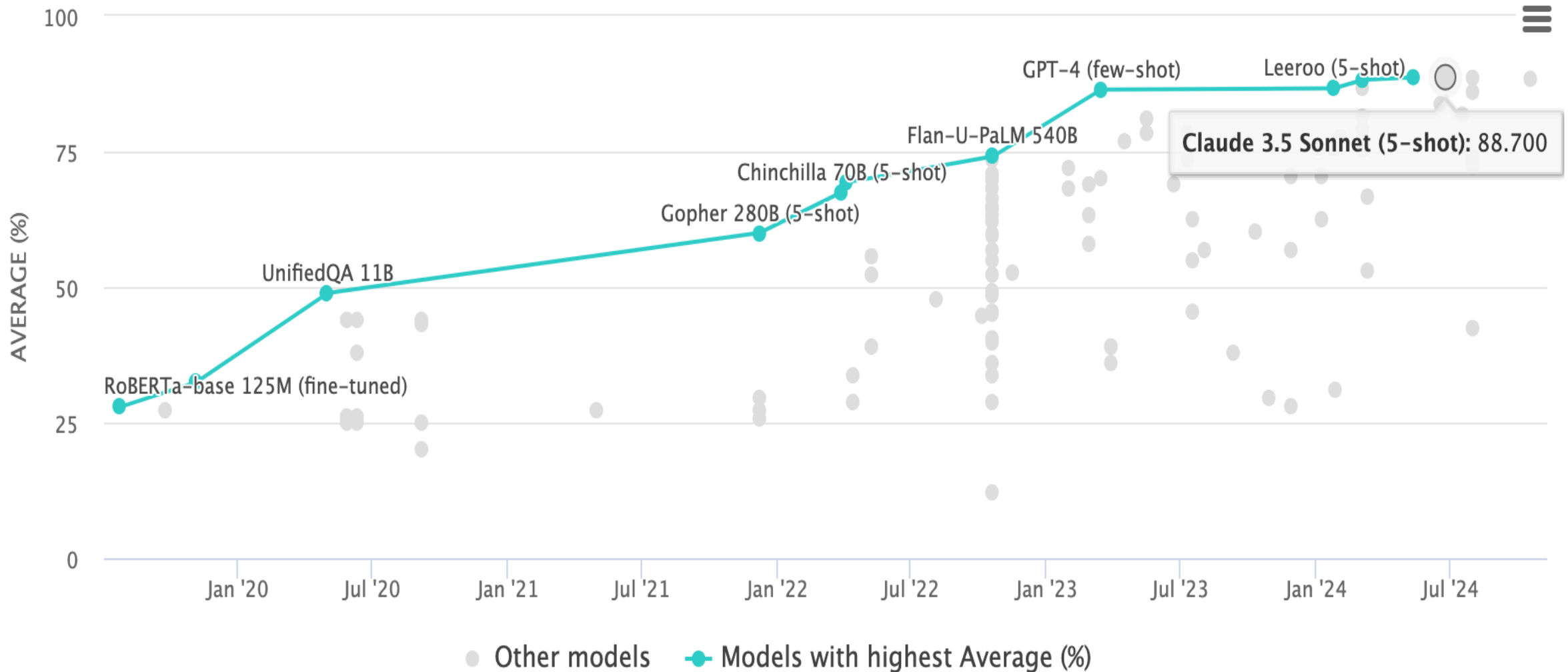
# Multimodal Large Language Model (MLLM) for Vision Question Answering



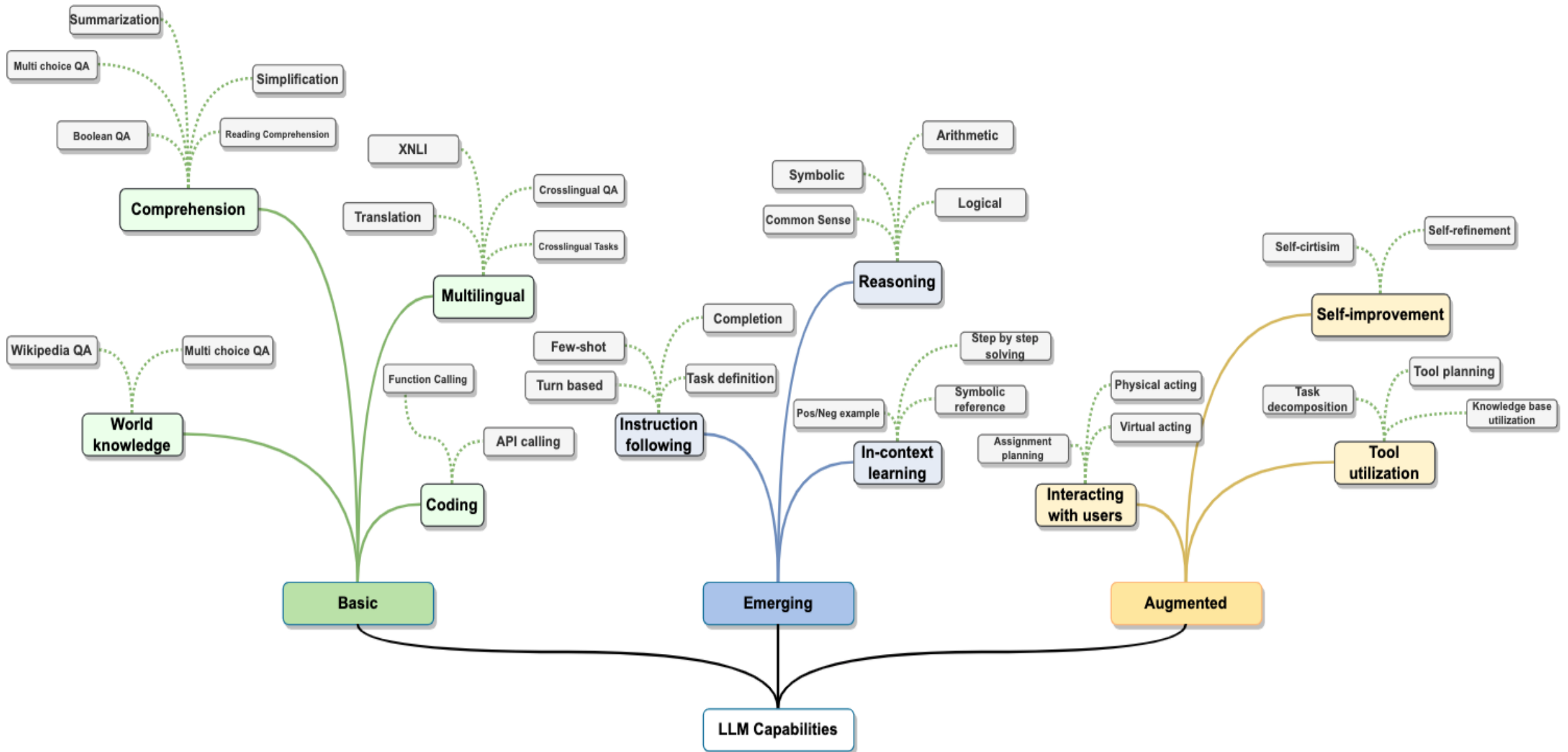
# Multi-task Language Understanding on MMLU

## GPT-4, Claude 3.5 Sonnet

### Massive Multitask Language Understanding (MMLU)

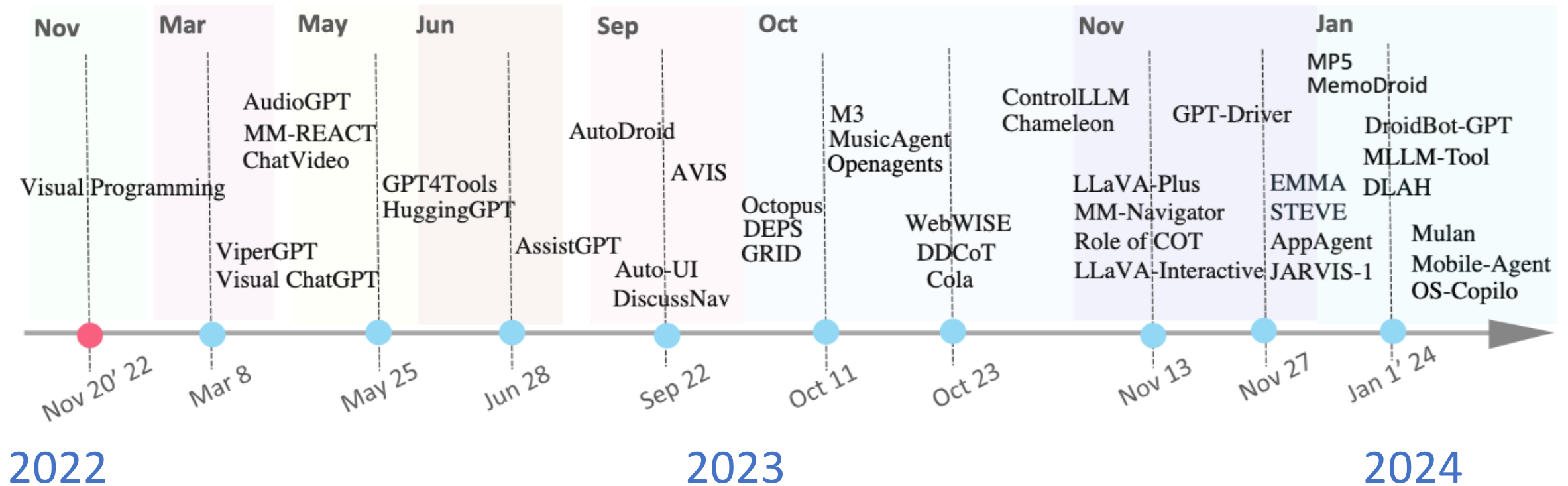


# LLM Capabilities


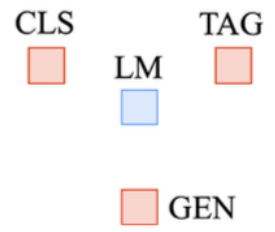
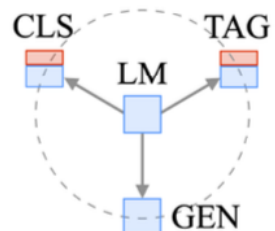
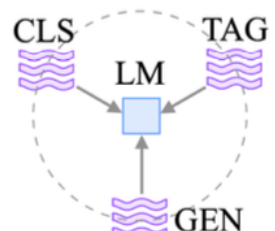


# LLM-powered Multimodal Agents

## Large Multimodal Agents (LMAs)

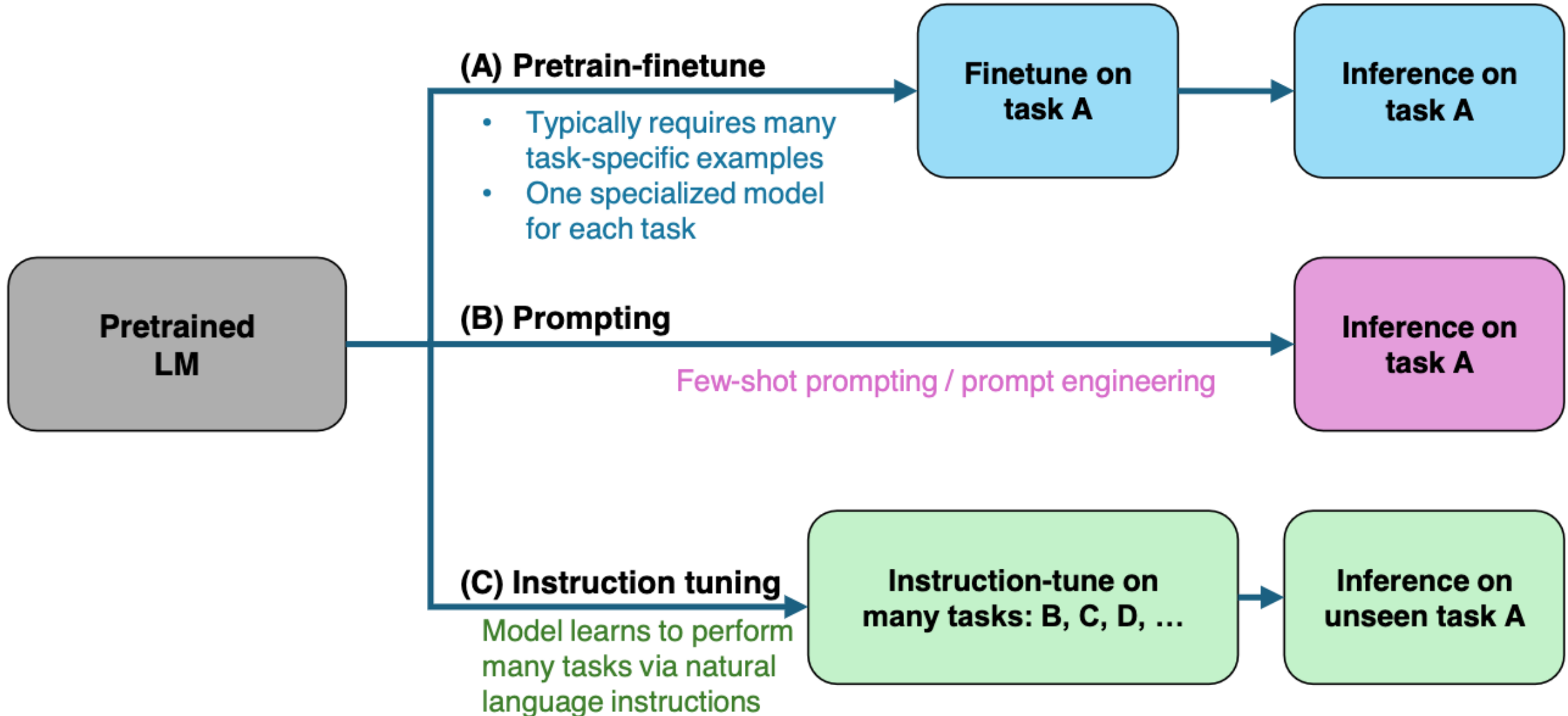


# Four Paradigms in NLP (LM)

Paradigm	Engineering	Task Relation
a. Fully Supervised Learning (Non-Neural Network)	Feature (e.g. word identity, part-of-speech, sentence length)	
b. Fully Supervised Learning (Neural Network)	Architecture (e.g. convolutional, recurrent, self-attentional)	
<b>Transfer Learning: Pre-training, Fine-Tuning (FT)</b>		
c. Pre-train, Fine-tune	Objective (e.g. masked language modeling, next sentence prediction)	
<b>GAI: Pre-train, Prompt, and Predict (Prompting)</b>		
d. Pre-train, Prompt, Predict	Prompt (e.g. cloze, prefix)	

# Large Language Models (LLM)

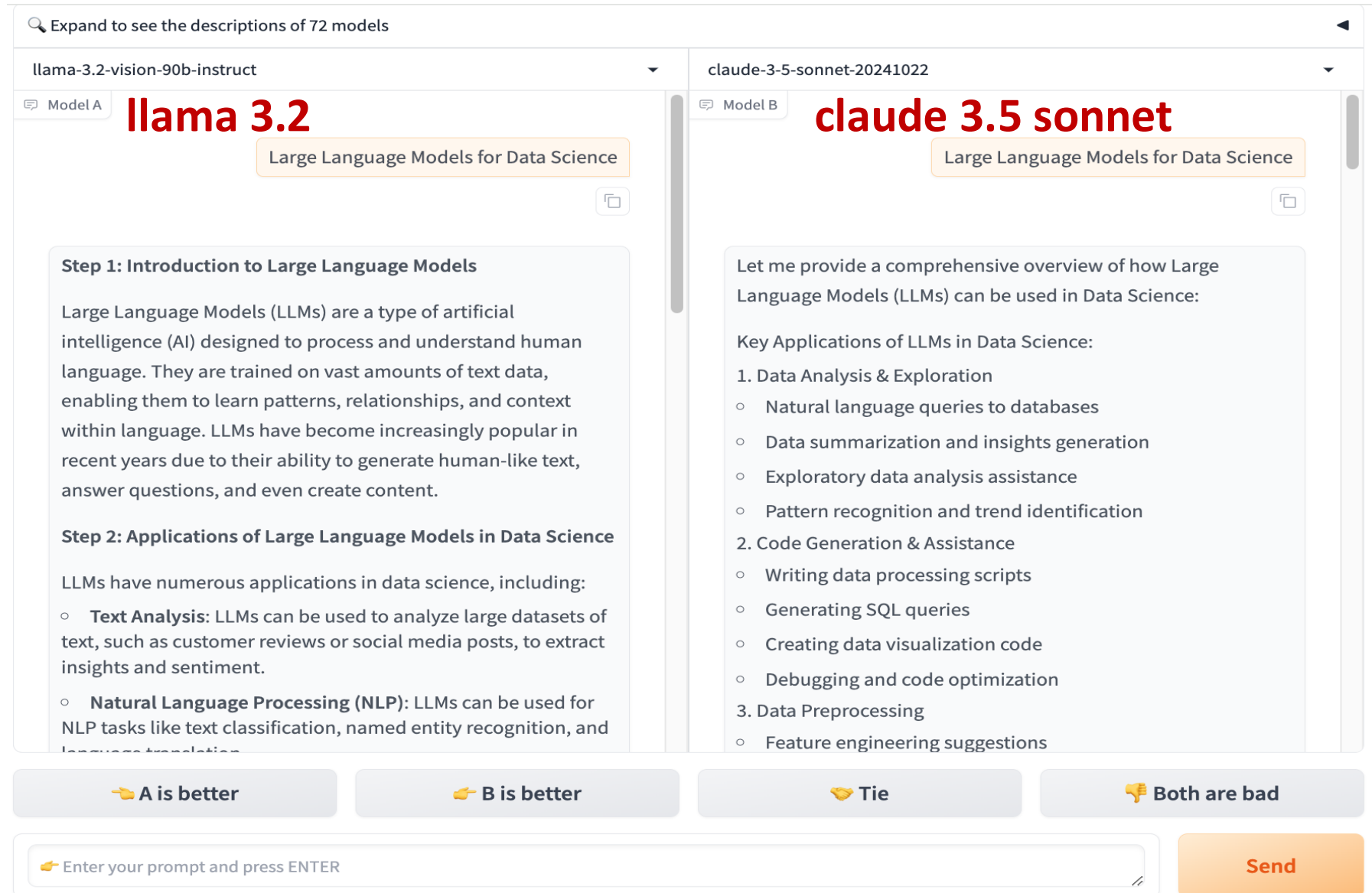
## Three typical learning paradigms



# Popular Generative AI

- **OpenAI ChatGPT (GPT-o1, GPT-4o, GPT-4)**
- **Claude.ai (Claude 3.5)**
- **Google Gemini**
- **Meta Llama 3.3, Llama 3.2 Vision**
- **Mixtral Pixtral (mistral.ai)**
- **DeepSeek**
- **Chat.LMSys.org (lmarena.ai)**
- **Perplexity.ai**
- **Stable Diffusion**
- **Video: D-ID, Synthesia**
- **Audio: Speechify**

# Chat with Open Large Language Models: Chatbot Arena



Expand to see the descriptions of 72 models

llama-3.2-vision-90b-instruct      claude-3-5-sonnet-20241022

Model A **llama 3.2**      Model B **claude 3.5 sonnet**

Large Language Models for Data Science

**Step 1: Introduction to Large Language Models**

Large Language Models (LLMs) are a type of artificial intelligence (AI) designed to process and understand human language. They are trained on vast amounts of text data, enabling them to learn patterns, relationships, and context within language. LLMs have become increasingly popular in recent years due to their ability to generate human-like text, answer questions, and even create content.

**Step 2: Applications of Large Language Models in Data Science**

LLMs have numerous applications in data science, including:

- **Text Analysis:** LLMs can be used to analyze large datasets of text, such as customer reviews or social media posts, to extract insights and sentiment.
- **Natural Language Processing (NLP):** LLMs can be used for NLP tasks like text classification, named entity recognition, and language translation.

Let me provide a comprehensive overview of how Large Language Models (LLMs) can be used in Data Science:

Key Applications of LLMs in Data Science:

1. Data Analysis & Exploration
  - Natural language queries to databases
  - Data summarization and insights generation
  - Exploratory data analysis assistance
  - Pattern recognition and trend identification
2. Code Generation & Assistance
  - Writing data processing scripts
  - Generating SQL queries
  - Creating data visualization code
  - Debugging and code optimization
3. Data Preprocessing
  - Feature engineering suggestions

👉 A is better      👉 B is better      🤝 Tie      🗑️ Both are bad

👉 Enter your prompt and press ENTER      Send

# Perplexity.ai



Anonymous Like



New Thread



Home

Discover

Library

Login

Sign Up

Try Pro

Upgrade to Claude-2 or GPT-4, boost Copilot uses, and upload more files.

Learn More

Download



## Large Language Models for Data Science

### Sources

What Is a Large Language Model and What Can...

snowflake · 1

What are Large Language Models? | NVIDIA

nvidia · 2

Catch Up On Large Language Models | by...

towardsdatascien... · 3



View 3 more

### Answer

Large Language Models (LLMs) are advanced AI systems designed to understand human language intricacies and generate intelligent, creative responses to queries <sup>1</sup>. They are deep learning algorithms that can recognize, summarize, translate, predict, and generate content using very large datasets <sup>2</sup>. LLMs are trained on enormous data sets typically measured in petabytes, sourced from books, articles, websites, and other text-based sources <sup>1</sup>. They are dynamic tools that are continuously updated and refined as new data becomes available <sup>1</sup>.

In the field of data science, LLMs can potentially transform how teams source, manage, and analyze data <sup>1</sup>. They are used in numerous data science applications, and their ability to process and interpret vast amounts of text data has made them an indispensable part of m

Ask follow-up...

Copilot



social media posts or customer reviews, to determine whether the overall sentiment is

<https://www.perplexity.ai/>



# Generative AI (Gen AI)

## AI Generated Content (AIGC)

### Image Generation

**Instruction 1:**

*An astronaut riding a horse in a photorealistic style.*

**Instruction 2:**

*Teddy bears working on new AI research on the moon in the 1980s.*

Figure 1



Figure 2

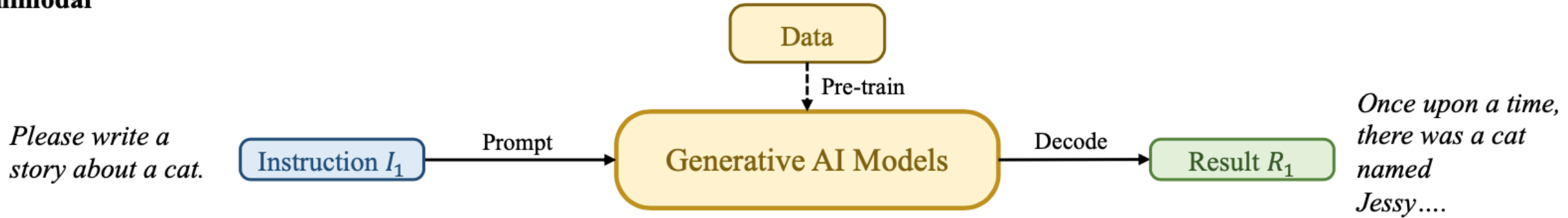


 **OpenAI DALL·E 2**

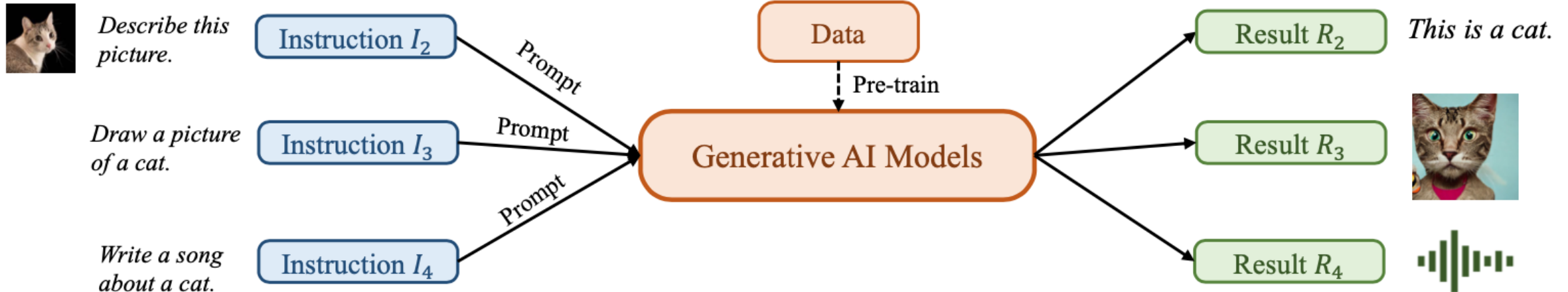
# Generative AI (Gen AI)

## AI Generated Content (AIGC)

### Unimodal

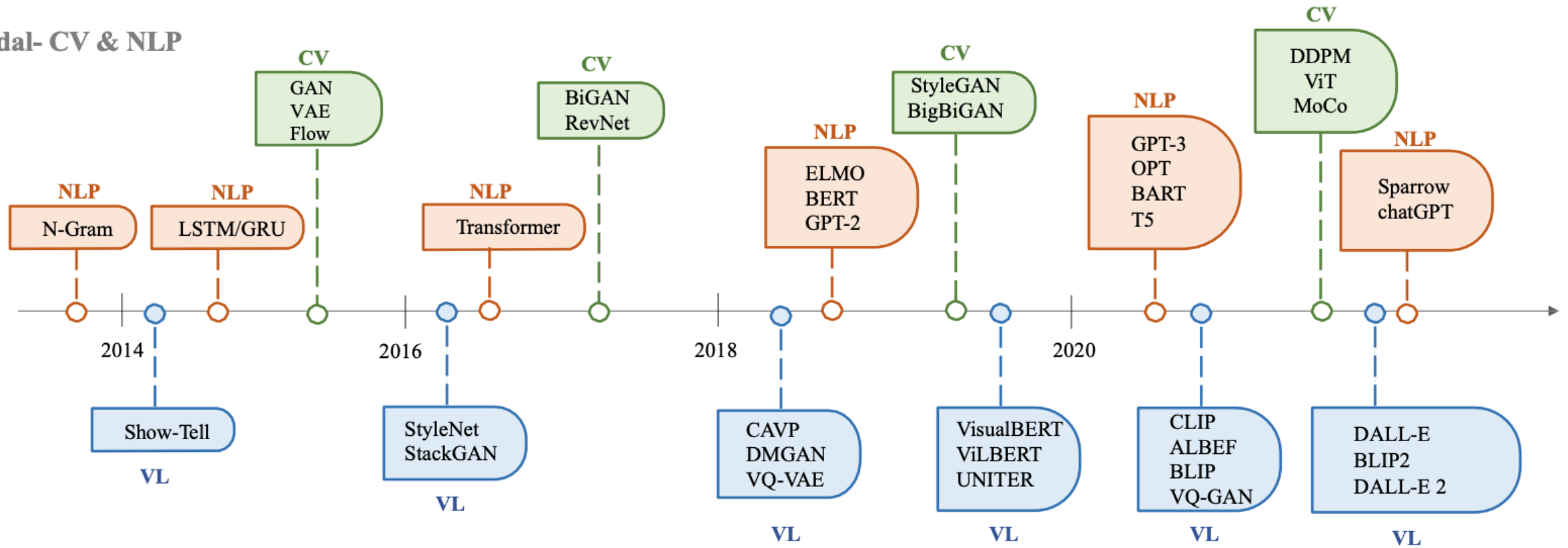


### Multimodal



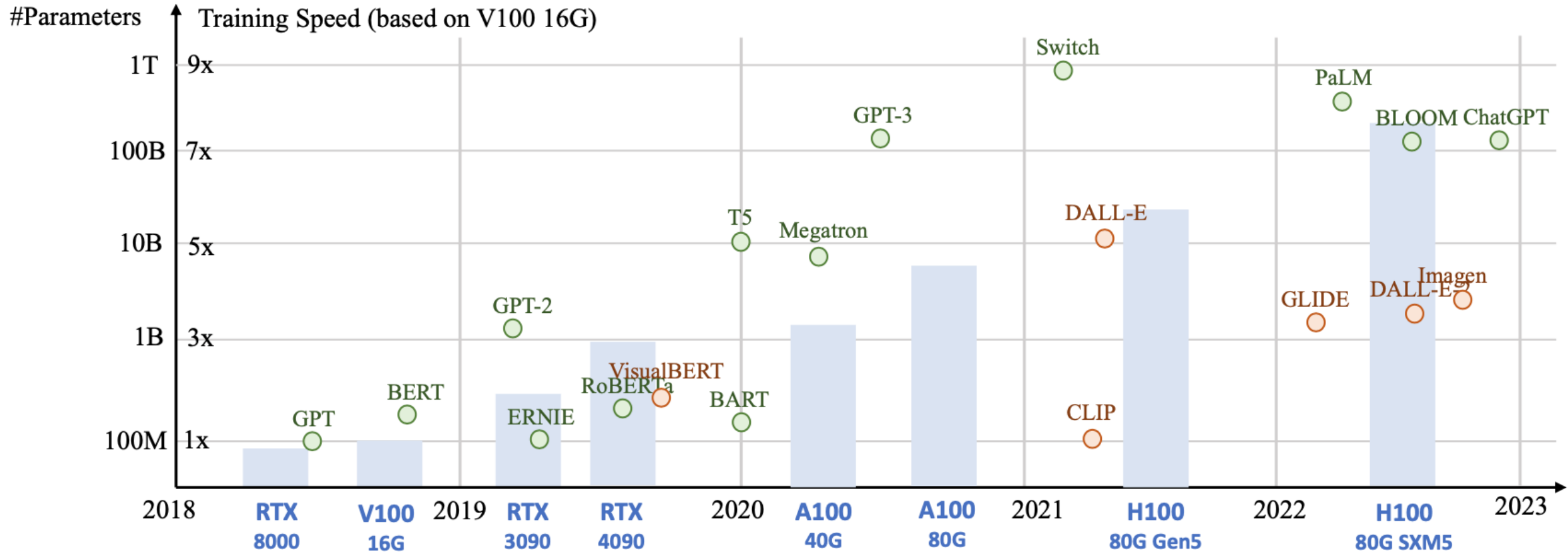
# The history of Generative AI in CV, NLP and VL

## Unimodal- CV & NLP

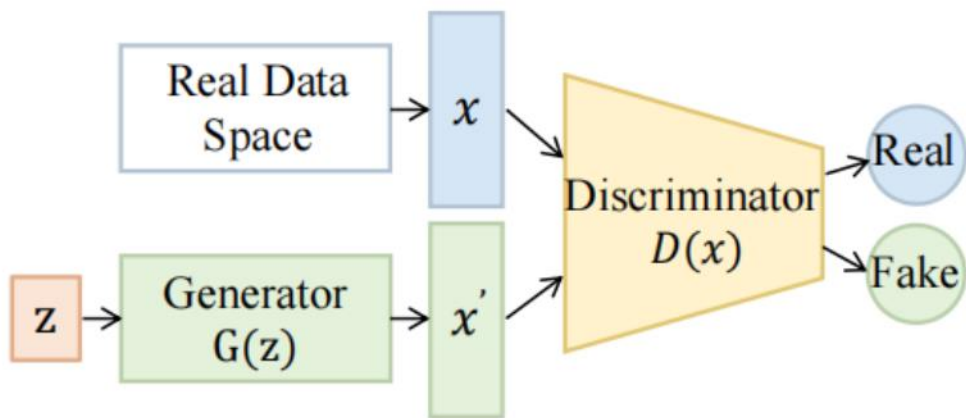


## Multimodal – Vision Language

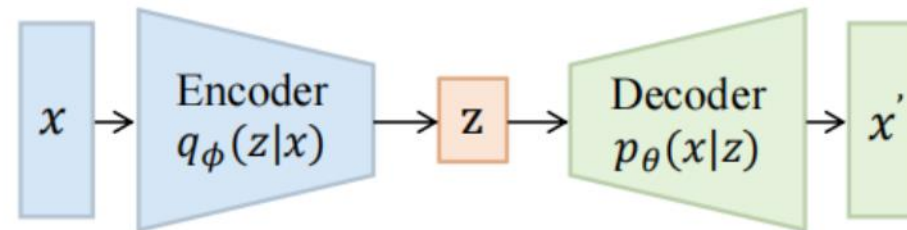
# Generative AI Foundation Models



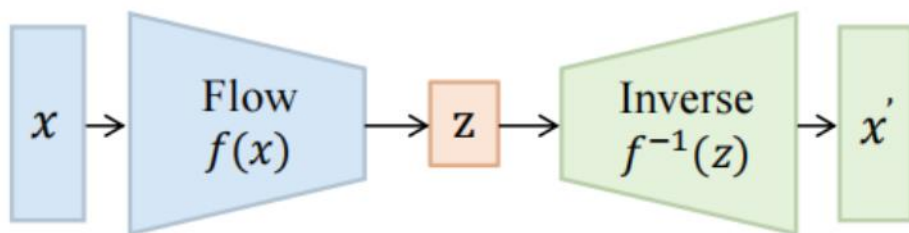
# Categories of Vision Generative Models



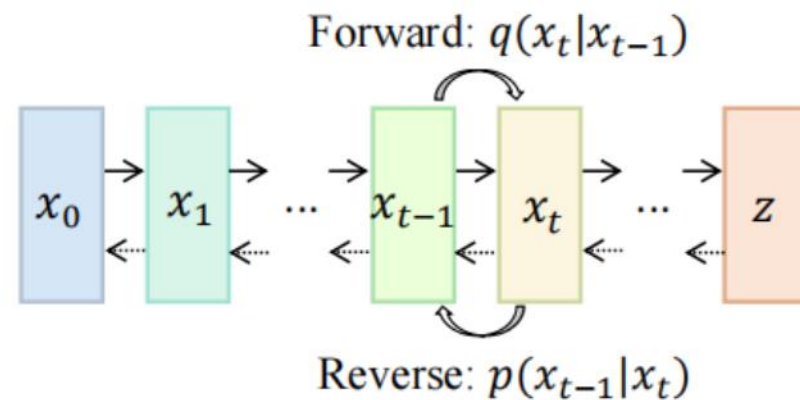
(1) Generative adversarial networks



(2) Variational autoencoders

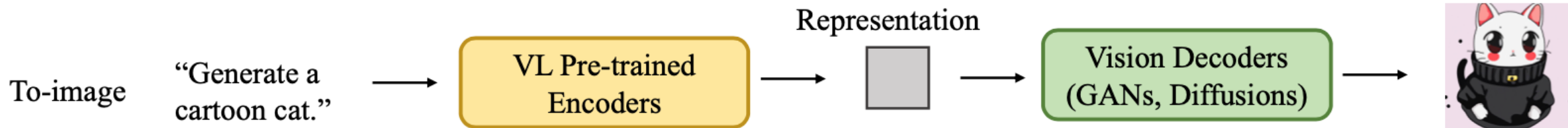
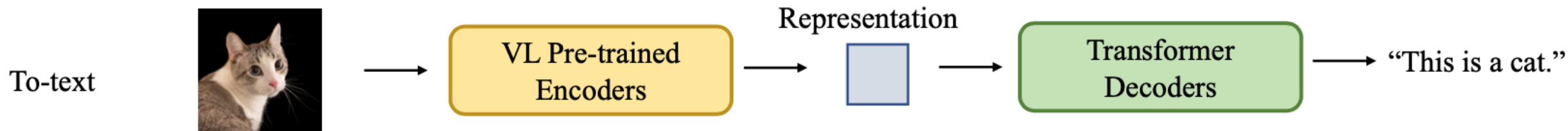
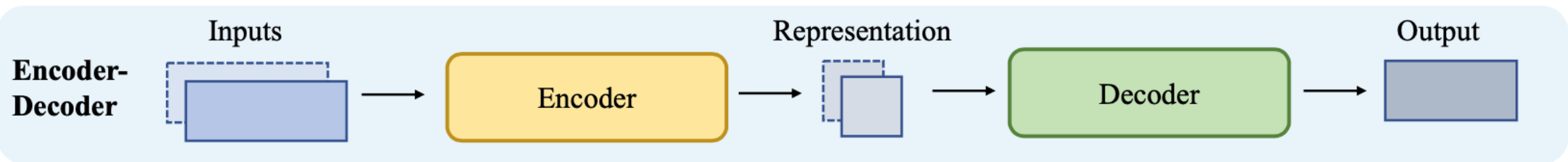


(3) Normalizing flows



(4) Diffusion models

# The General Structure of Generative Vision Language

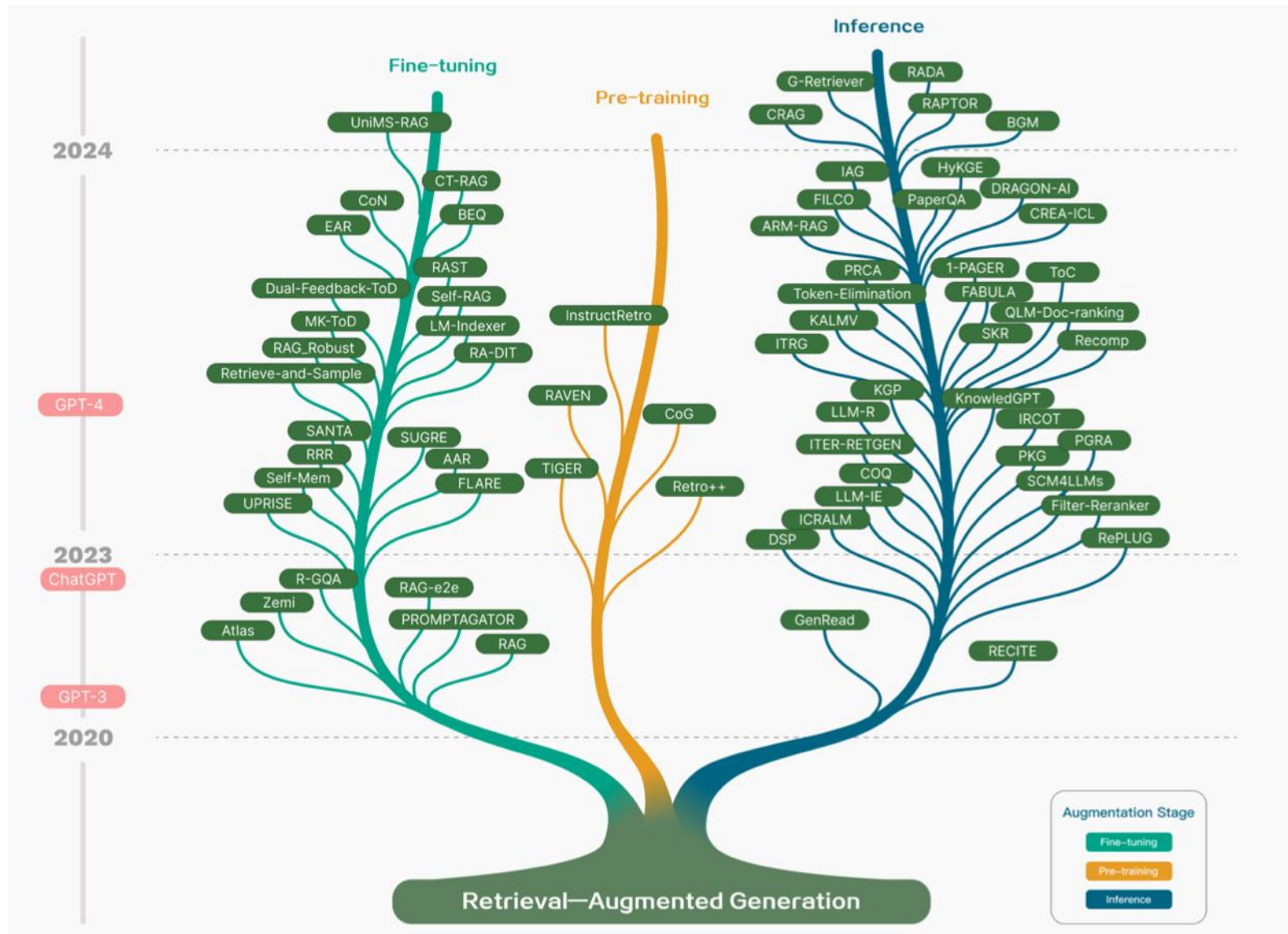


# **RAG LLM**

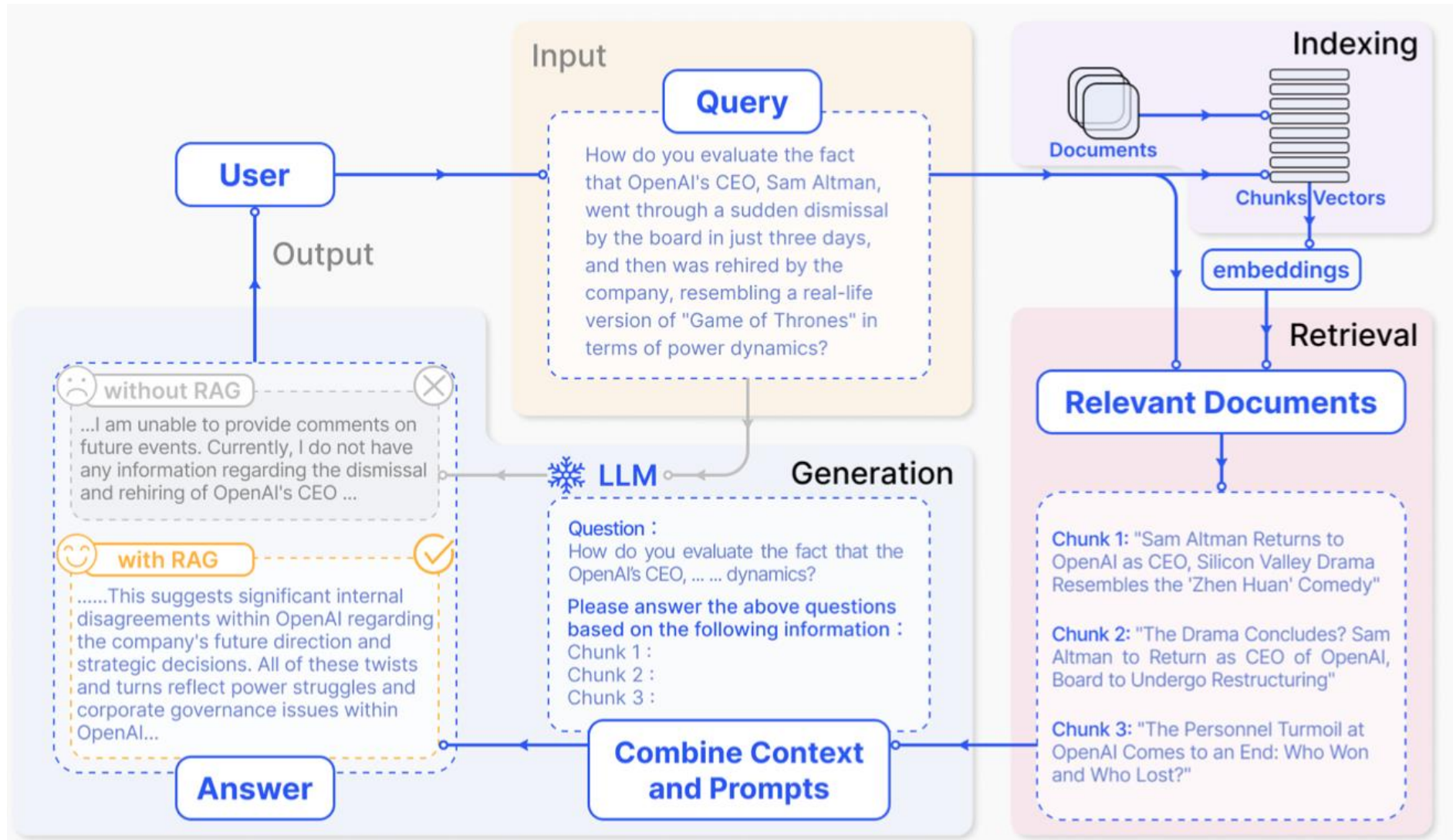
# **Dialogue Systems**

# Technology Tree of RAG Research

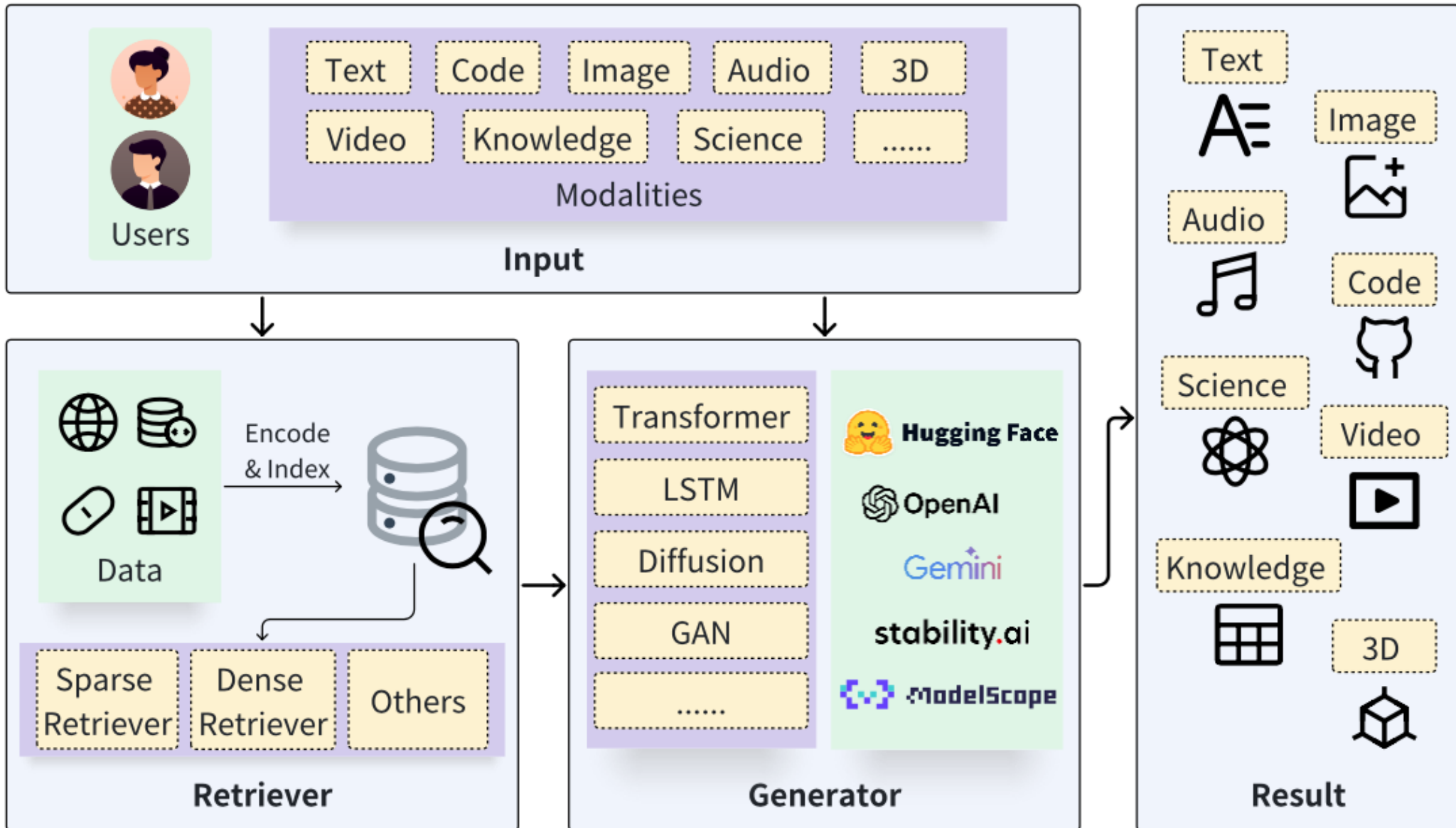
## Retrieval-Augmented Generation (RAG) for Large Language Models (LLMs)



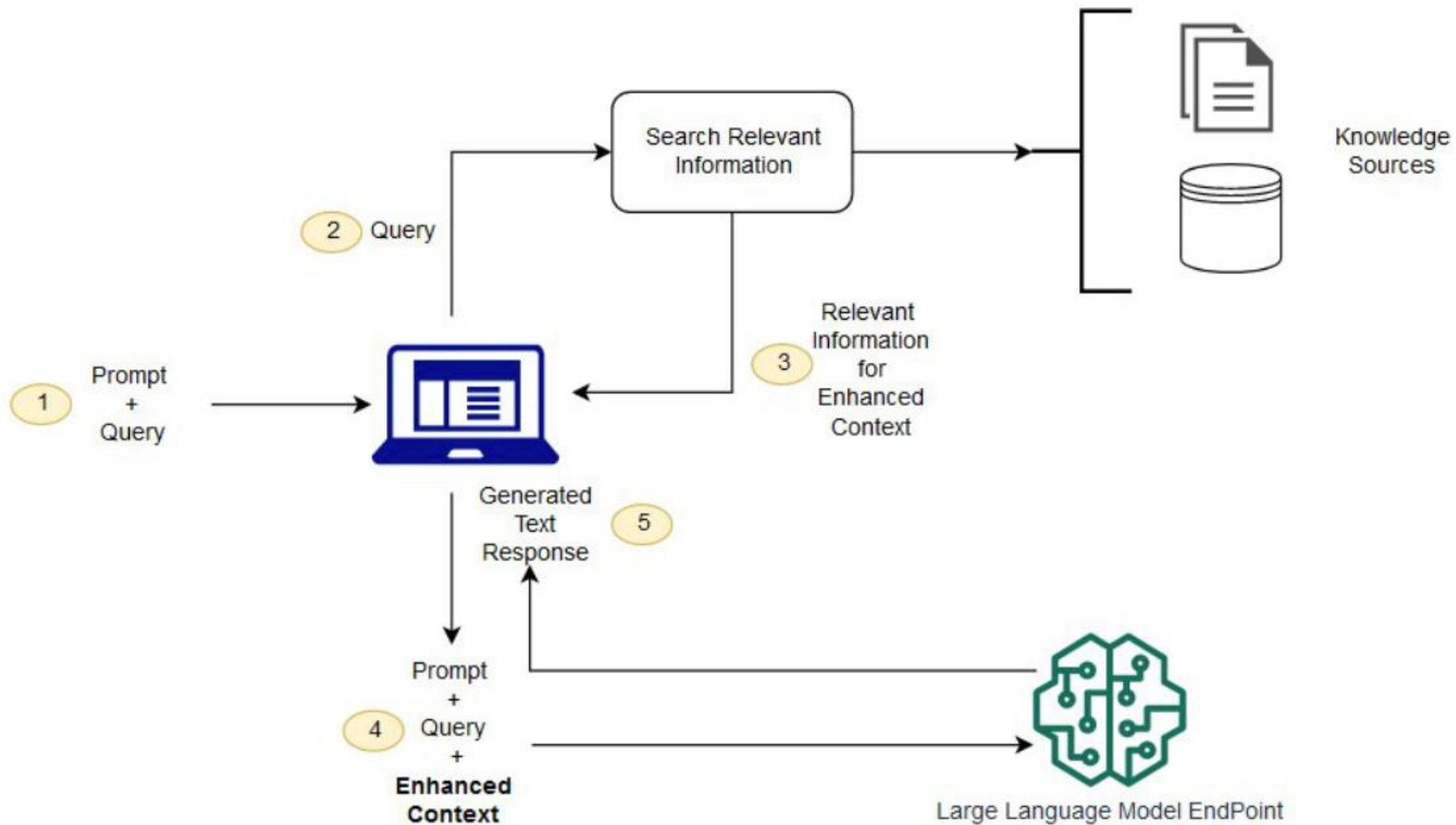
# Retrieval-Augmented Generation (RAG) for Large Language Models (LLMs)



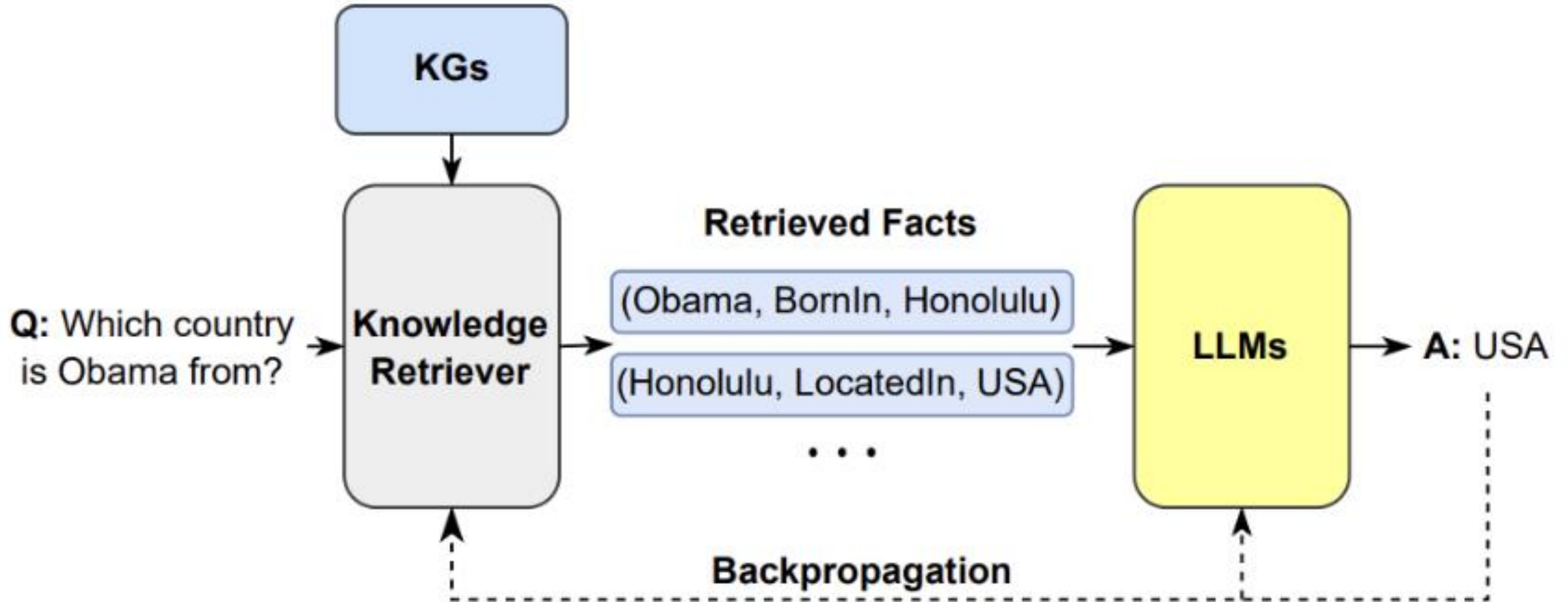
# Retrieval-Augmented Generation (RAG) Architecture



# Synthesizing RAG with LLMs for Question Answering Application

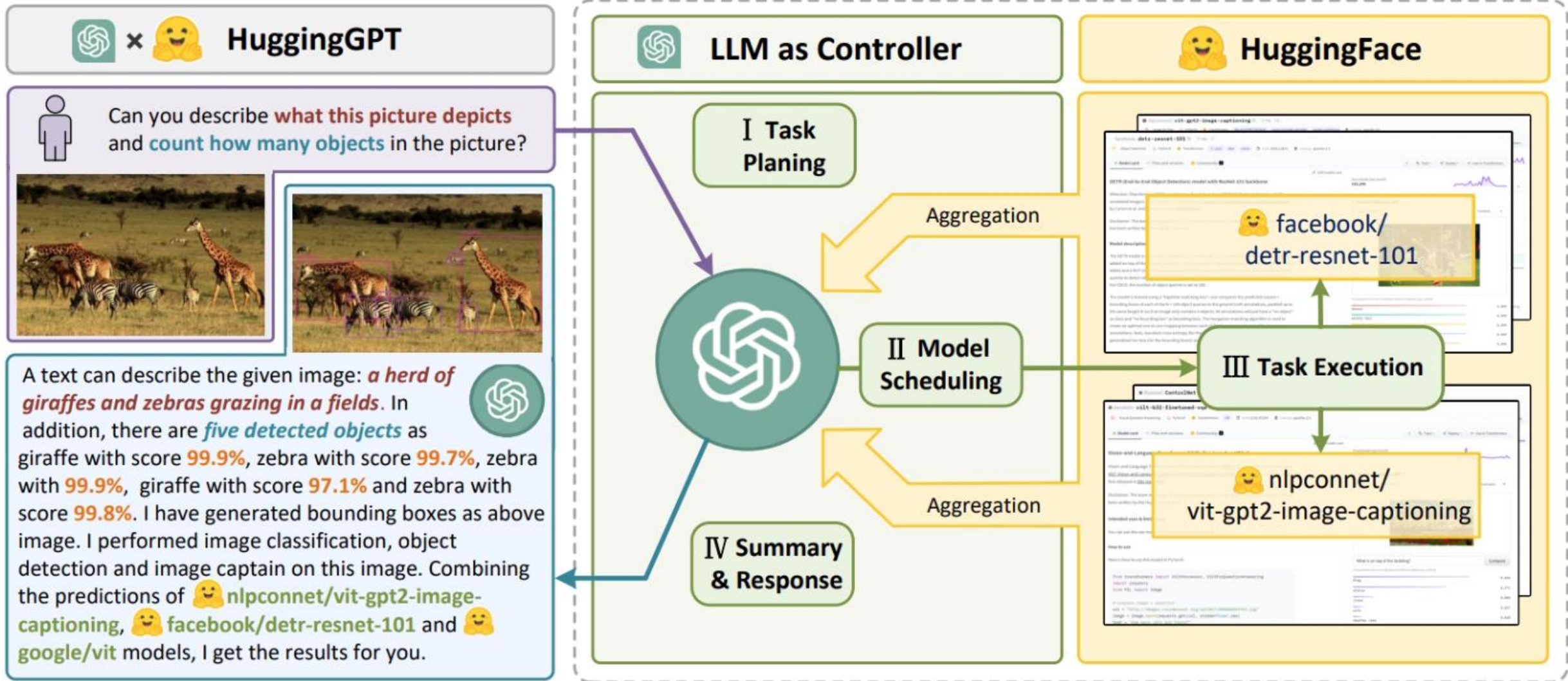


# Synthesizing the KG as a Retriever with LLMs

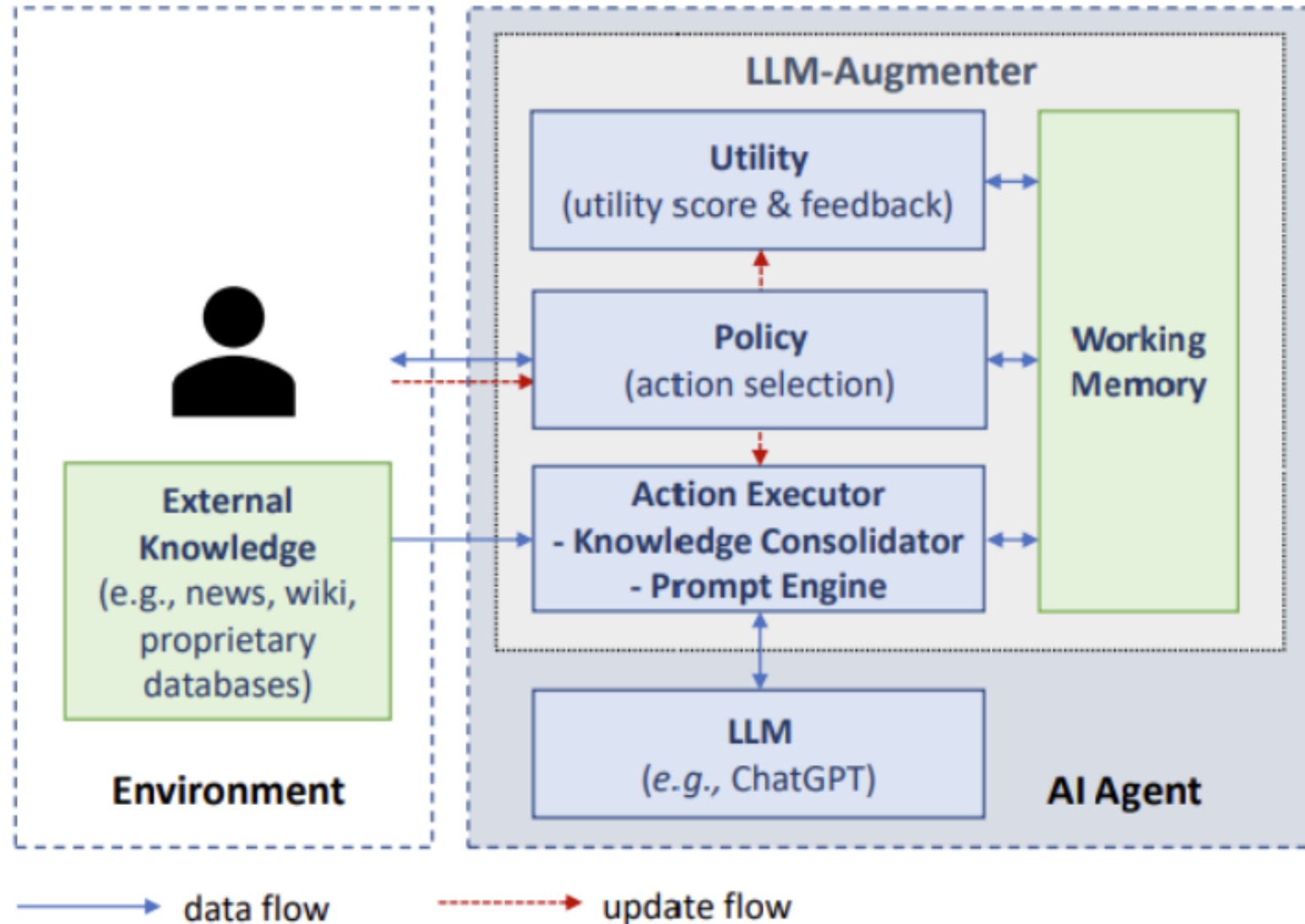


# HuggingGPT:

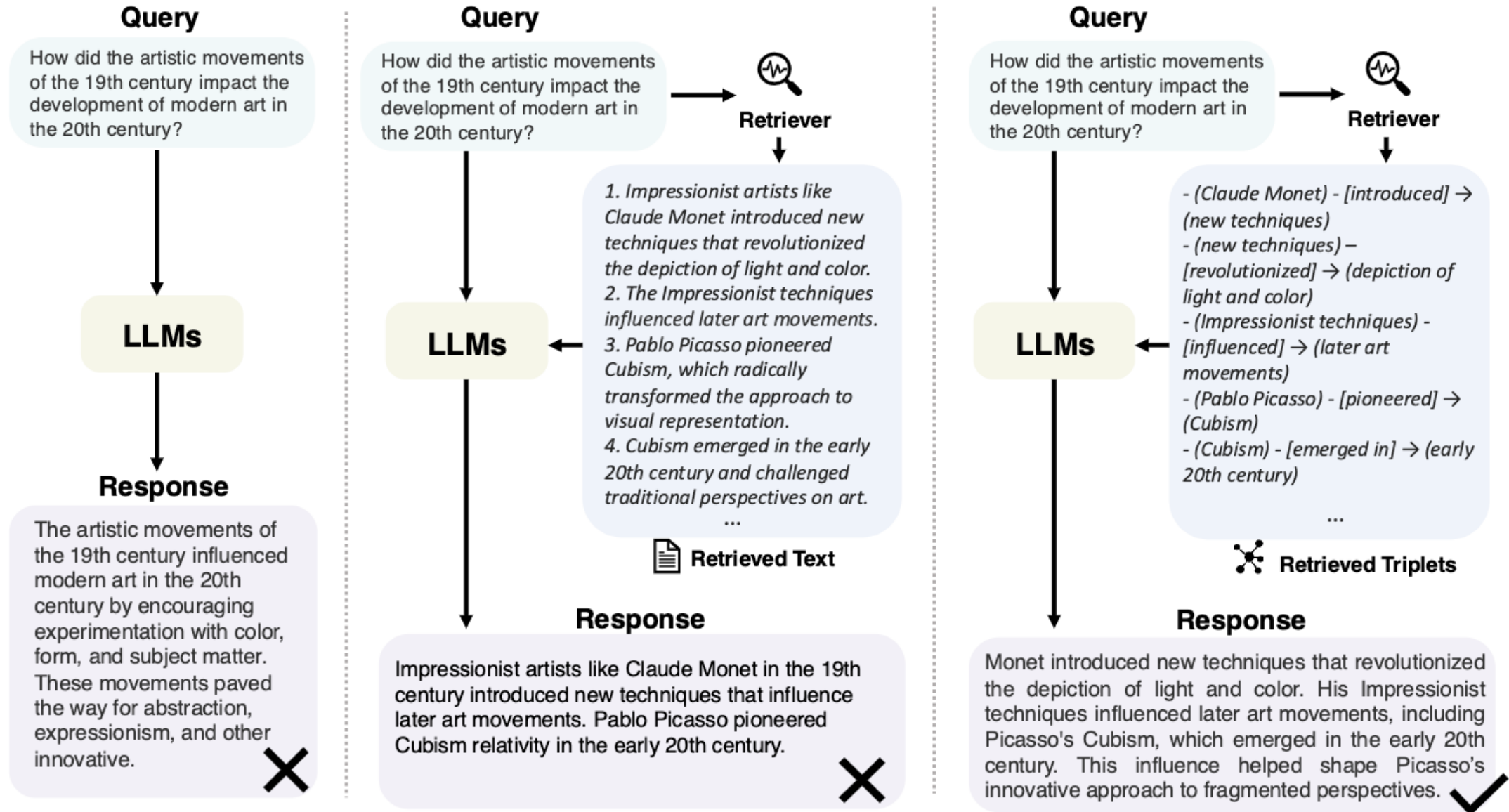
## An agent-based approach to use tools and planning



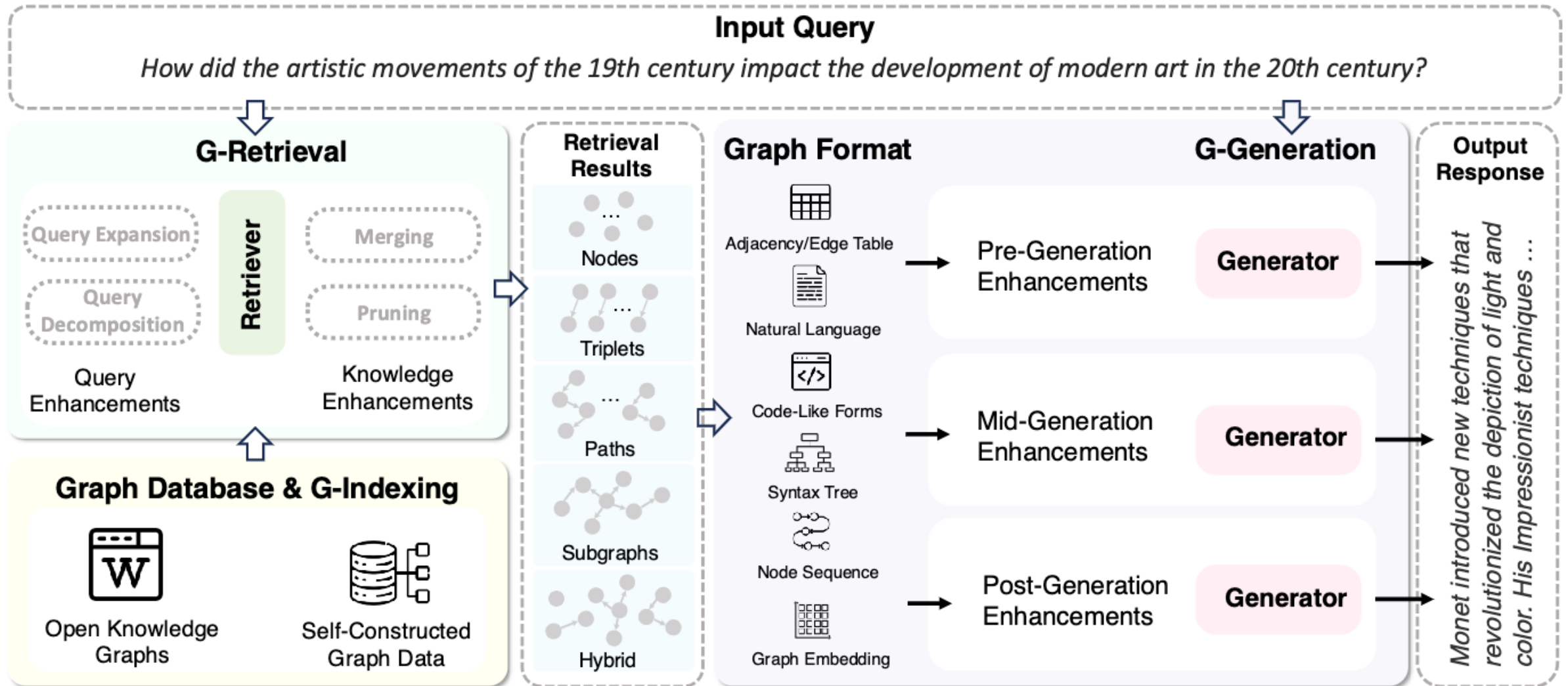
# A LLM-based Agent for Conversational Information Seeking



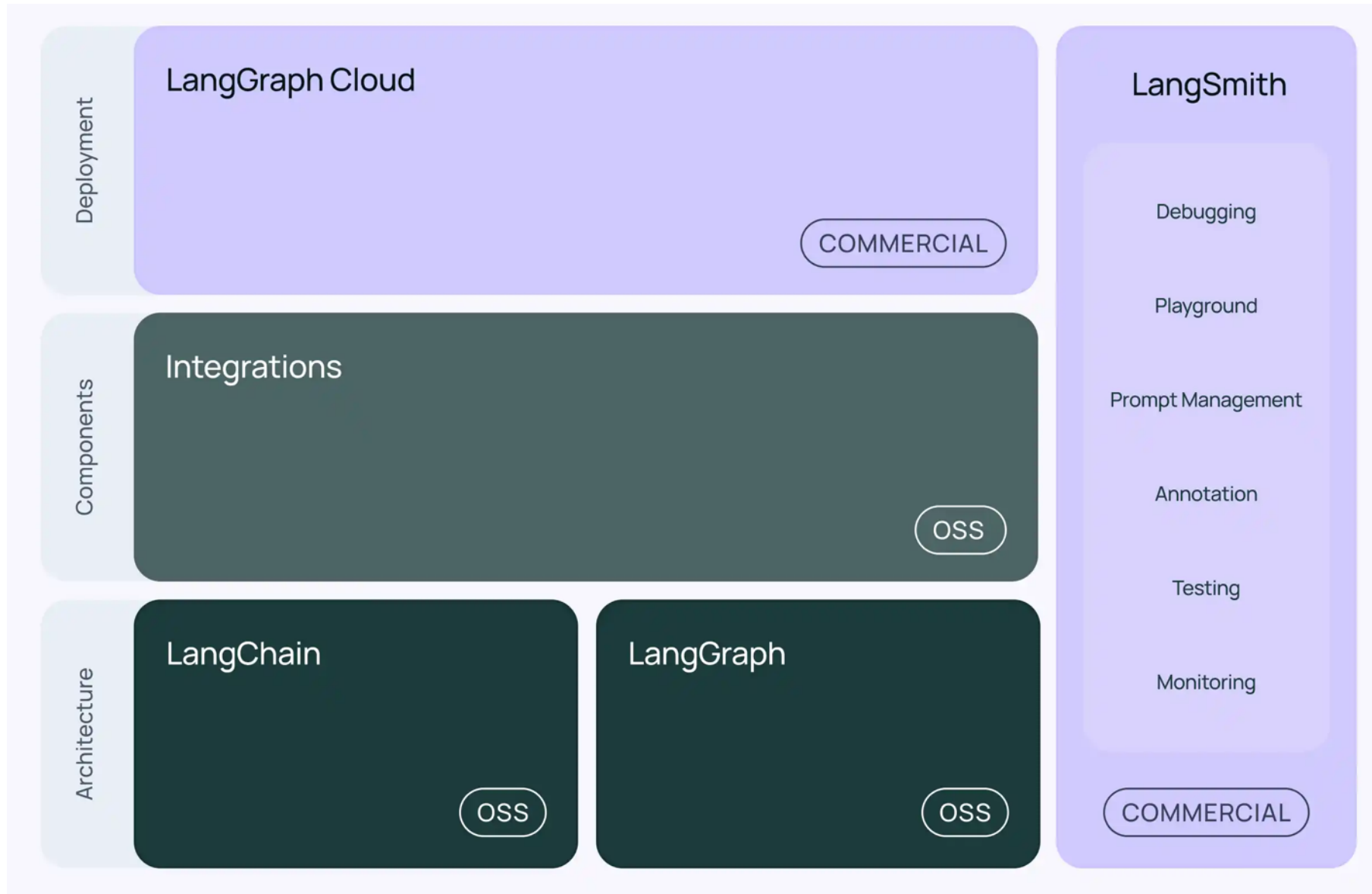
# Direct LLM, RAG, and GraphRAG



# GraphRAG Framework for Question Answering

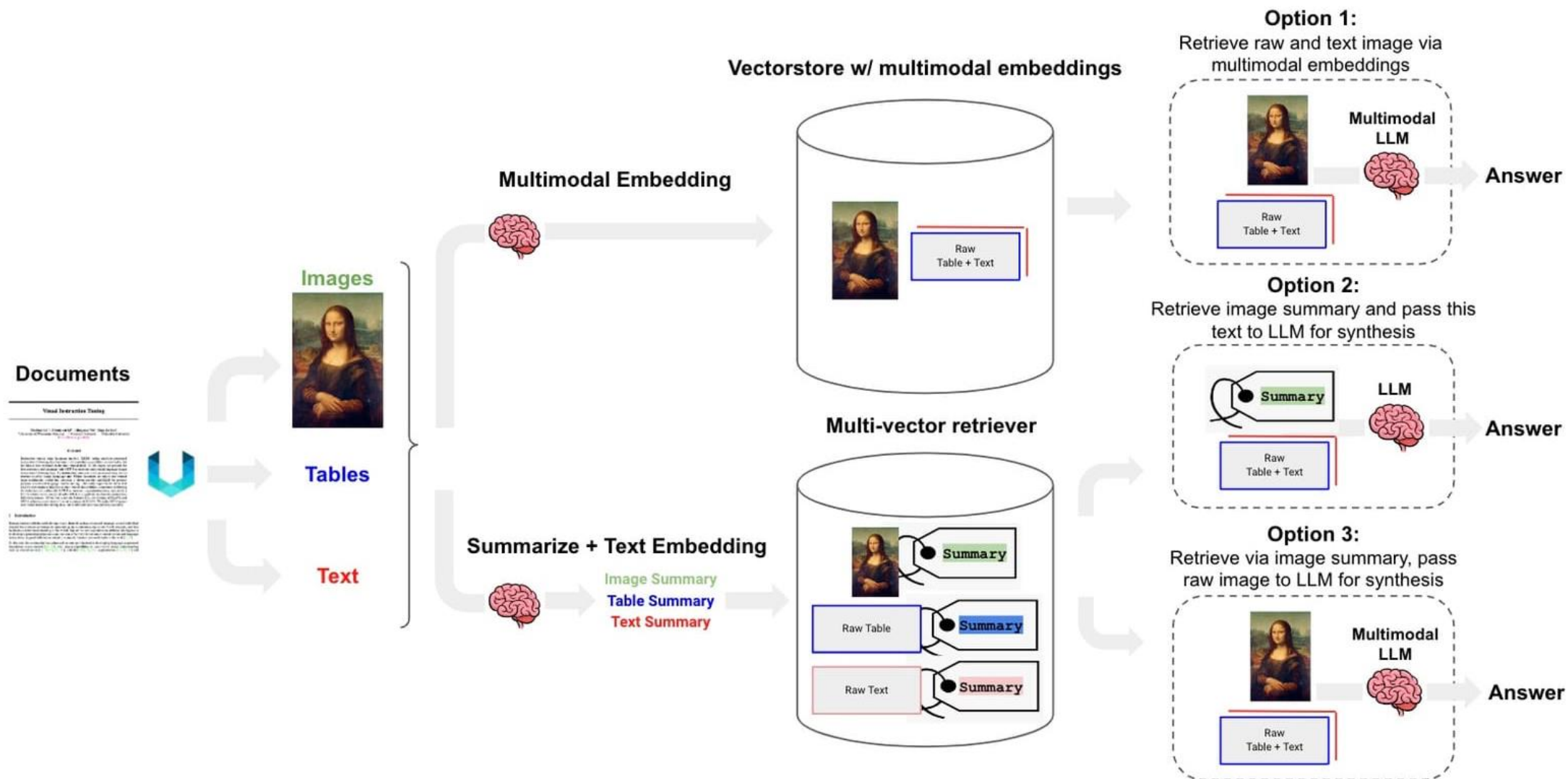


# LangChain Architecture



# Multimodal LLM RAG

## Multi-Vector Retriever for RAG



# Evaluating RAG with Ragas Metrics

## ragas score

generation

### faithfulness

how factually accurate is  
the generated answer

### answer relevancy

how relevant is the generated  
answer to the question

retrieval

### context precision

the signal to noise ratio of retrieved  
context

### context recall

can it retrieve all the relevant information  
required to answer the question

# Teaching



- **Generative AI Innovative Applications**
  - Spring 2025
- **Artificial Intelligence in Finance and Quantitative**
  - Fall 2021, Fall 2022, Fall 2023, Spring 2025
- **Software Engineering**
  - Fall 2020, Fall, 2021, Spring 2022, Spring 2023, Spring 2024, Spring 2025
- **Artificial Intelligence**
  - Spring 2021, Fall 2022, Fall 2024
- **Sustainability and ESG Data Analytics**
  - Spring 2024, Fall 2024
- **Big Data Analytics**
  - Fall 2020, Spring 2023, Spring 2024
- **Artificial Intelligence for Text Analytics**
  - Spring 2022, Fall 2023
- **Python for Accounting Applications**
  - Fall 2023, Fall 2024
- **Foundation of Business Cloud Computing**
  - Spring 2021, Spring 2022, Spring 2023, Spring 2024

# Research Projects



- 1. Digital Support, Unimpeded Communication: The Development, Support and Promotion of AI-assisted Communication Assistive Devices for Speech Impairment (2/3).**  
**Multimodal Cross-lingual Task-Oriented Dialogue System for Inclusive Communication Support**
  - NSTC 113-2425-H-305-002-, 3 Years (2023/05/01-2026/04/30) Year 1: 2024/05/01~2025/04/30
- 2. Research on speech processing, synthesis, recognition, and sentence construction of people with language disabilities. Multimodal Cross-lingual Task-Oriented Dialogue System**
  - NTPU, 114-NTPU\_ORDA-F-004, 2023/01/01~2025/12/31
- 3. Development of a Deep Learning for Dental Implant Detection in Panoramic Radiographs,**
  - USTP-NTPU-TMU-114-02, 2025/01/01~2025/12/31
- 4. Metaverse AI Multimodal Cross-Language Task-Oriented Dialogue System**
  - ATEC Group, Fintech and Green Finance Center (FGFC, NTPU), NTPU-112A413E01, 2 Years (2023/05/01~2025/04/30)
- 5. Establishment and Implement of Smart Assistive Technology for Dementia Care and Its Socio-Economic Impacts (3/3). Intelligent, individualized and precise care with smart AT and system integration**
  - NSTC, 113-2627-M-038-001-, 2024/08/01~2025/07/31

# Summary

- This course introduces the **fundamental concepts, research issues, and hands-on practices** of **Generative AI Innovative Applications**.
- Topics include:
  1. Introduction to Generative AI Innovative Applications
  2. Transformers for Natural Language Processing and Computer Vision
  3. Large Language Models (LLMs)
  4. **NVIDIA Building RAG Agents with LLMs**
  5. Generative AI for Multimodal Information Generation
  6. **NVIDIA Generative AI with Diffusion Models**
  7. AI Agents and Large Multimodal Agents (LMAs)
  8. **Case Study on Generative AI Innovative Applications**



# Generative AI Innovative Applications



2020 Cohort



Accredited  
Educator



Solutions  
Architect  
Associate



Cloud  
Practitioner

## Contact Information

**Min-Yuh Day, Ph.D.**

Professor

[Institute of Information Management, National Taipei University](#)

Tel: 02-86741111 ext. 66873

Office: B8F12

Address: 151, University Rd., San Shia District, New Taipei City, 23741 Taiwan

Email: [myday@gm.ntpu.edu.tw](mailto:myday@gm.ntpu.edu.tw)

Web: <http://web.ntpu.edu.tw/~myday/>

