Generative AI Innovative Applications

國立臺北大學
National Taipei University

Introduction to Generative AI Innovative Applications

1142GAIIA01
MBA, IM, NTPU (M6031) (Spring 2026)
Tue 2, 3, 4 (9:10-12:00) (B3F17)

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

NVIDIA University Ambassador Certified Instructor

aws educate Cloud Ambassador 2020 Cohort

aws academy Accredited Educator

aws certified Cloud Practitioner

aws certified Solutions Architect Associate

Min-Yuh Day, Ph.D,
Professor

Institute of Information Management, National Taipei University

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

2026-02-24

NTPU
國立臺北大學
National Taipei University

AACSB ACCREDITED

aws educate | Cloud Ambassador
2020 Cohort

aws academy
Accredited Educator

aws certified
Solutions Architect
Associate

aws certified
Cloud Practitioner

Prof. Min-Yuh Day

NVIDIA
University Ambassador
Certified Instructor

Director, Information Management, NTPU

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

Director, Fintech and Green Finance Center (FGFC), NTPU

Division Director, Sustainable Development, Sustainability Office, NTPU

Visiting Scholar, IIS, Academia Sinica

Ph.D., Information Management, NTU

Artificial Intelligence, Agentic AI, ESG and Green Financial Technology, Big Data Analytics, Electronic Commerce

NTPU
國立臺北大學
National Taipei University

中央研究院 ACADEMIA SINICA 1928

國立臺灣大學

Spring 2026

Generative AI Innovative Applications

Course Syllabus
National Taipei University
Academic Year 114, 2nd Semester (Spring 2026)

- **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. Agentic AI 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	2026/02/24	Introduction to Generative AI Innovative Applications
2	2026/03/03	Transformers for Natural Language Processing and Computer Vision; Large Language Models (LLMs)
3	2026/03/10	NVIDIA Building RAG Agents with LLMs Part I
4	2026/03/17	Case Study on Generative AI Innovative Applications I
5	2026/03/24	NVIDIA Building RAG Agents with LLMs Part II
6	2026/03/31	NVIDIA Building RAG Agents with LLMs Part III
7	2026/04/07	Make-up holiday for NTPU Sports Day (No Classes)
8	2026/04/14	Midterm Project Report

Syllabus

Week	Date	Subject/Topics
9	2026/04/21	Generative AI for Multimodal Information Generation
10	2026/04/28	NVIDIA Generative AI with Diffusion Models Part I
11	2026/05/05	NVIDIA Generative AI with Diffusion Models Part II
12	2026/05/12	Case Study on Generative AI Innovative Applications II
13	2026/05/19	NVIDIA Generative AI with Diffusion Models Part III
14	2026/05/26	Agentic AI and Large Multimodal Agents (LMAs)
15	2026/06/02	Final Project Report I
16	2026/06/09	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, 3rd Edition, Packt Publishing
3. Thomas R. Caldwell (2025), The Agentic AI Bible: The Complete and Up-to-Date Guide to Design, Build, and Scale Goal-Driven, LLM-Powered Agents that Think, Execute and Evolve, Independently published

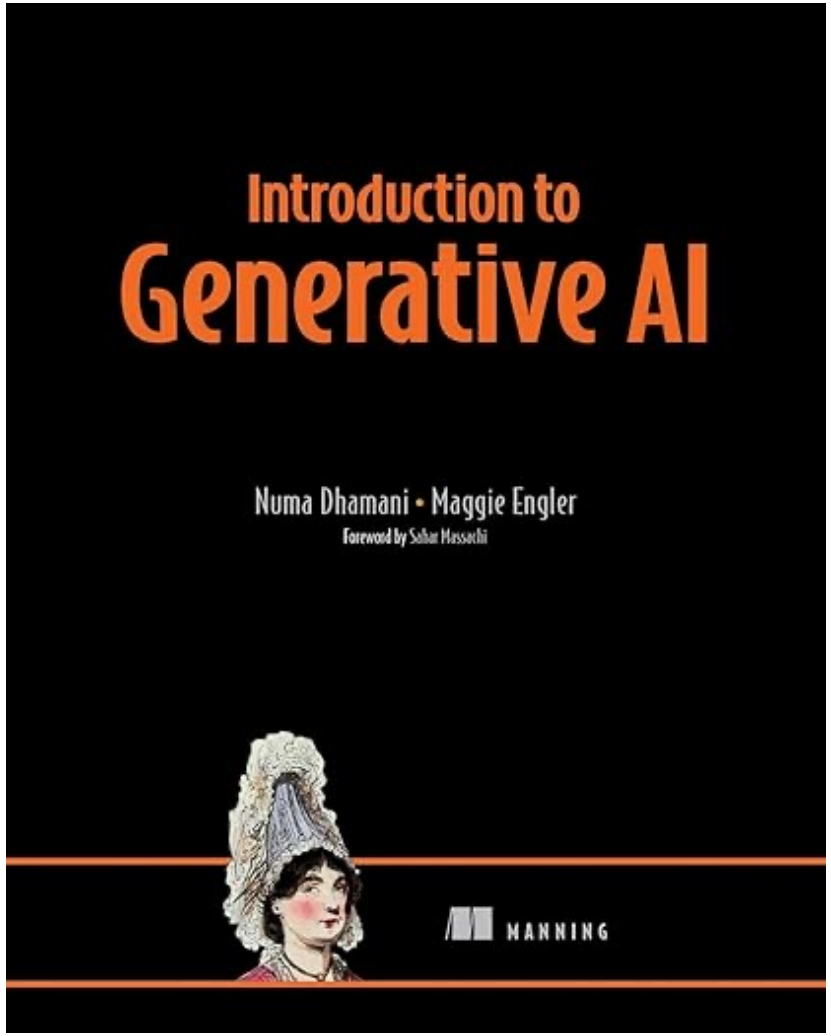
References

- NVIDIA DLI (2024), Building RAG Agents with LLMs, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-15+V1
- NVIDIA DLI (2024), Generative AI with Diffusion Models, 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
- Alammar 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



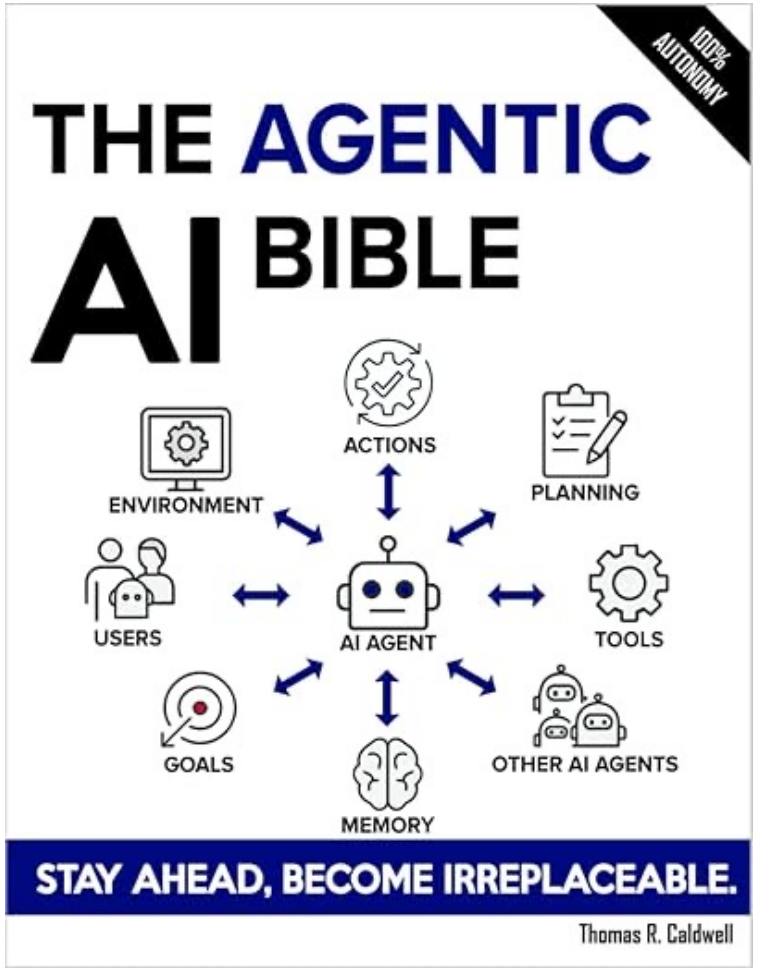
Source: Numa Dhamani and Maggie Engler (2024), Introduction to Generative AI, Manning
https://www.amazon.com/Introduction-Generative-AI-Numa-Dhamani/dp/1633437191/

Thomas R. Caldwell (2025),

The Agentic AI Bible:

The Complete and Up-to-Date Guide to Design, Build, and Scale Goal-Driven, LLM-Powered Agents that Think, Execute and Evolve,

Independently published

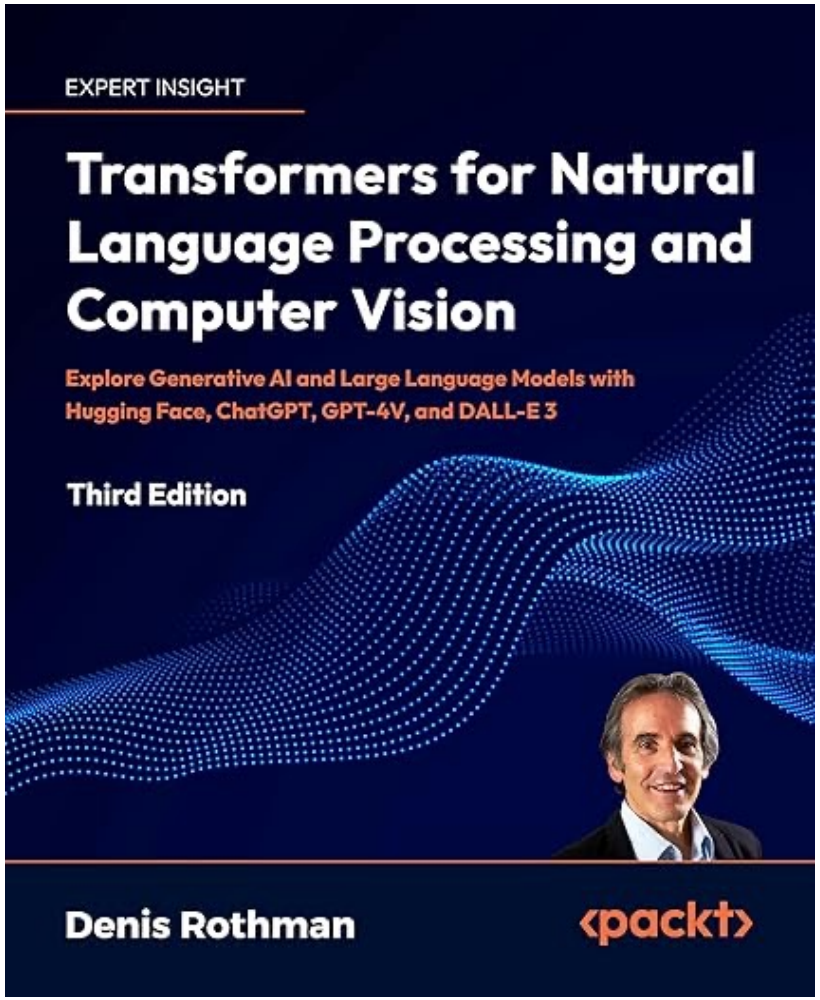


Source: https://www.amazon.com/Agentic-Bible-Up-Date-Goal-Driven/dp/B0FL21R86Q

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

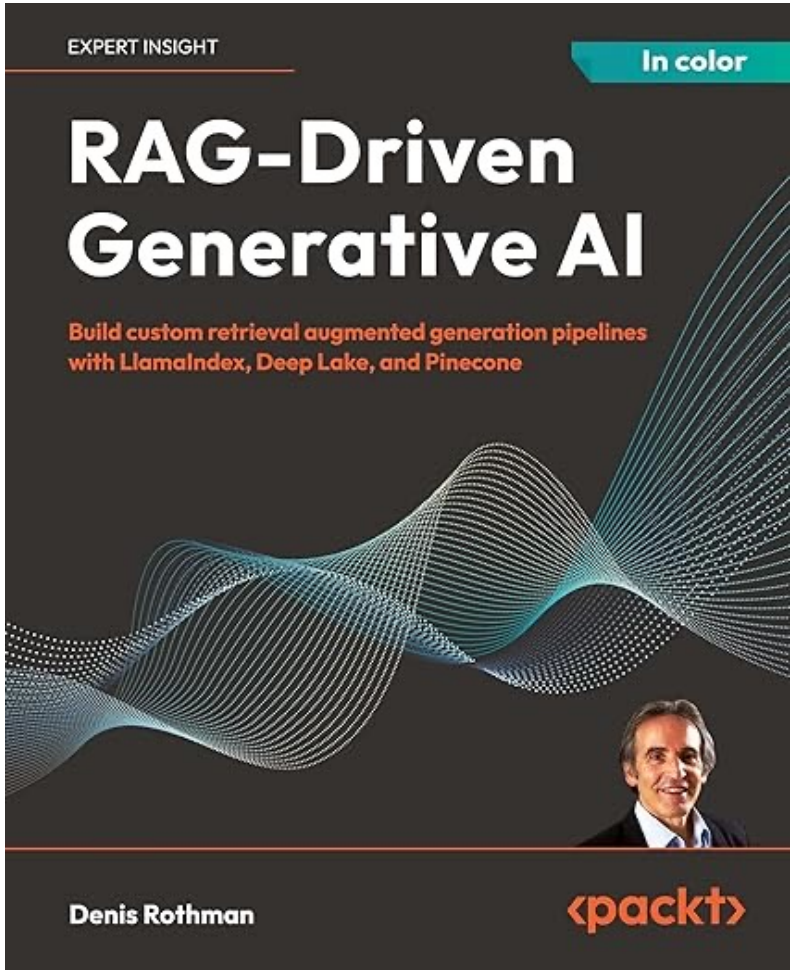


Source: https://www.amazon.com/Transformers-Natural-Language-Processing-Computer/dp/1805128728

Denis Rothman (2024),

RAG-Driven Generative AI:

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

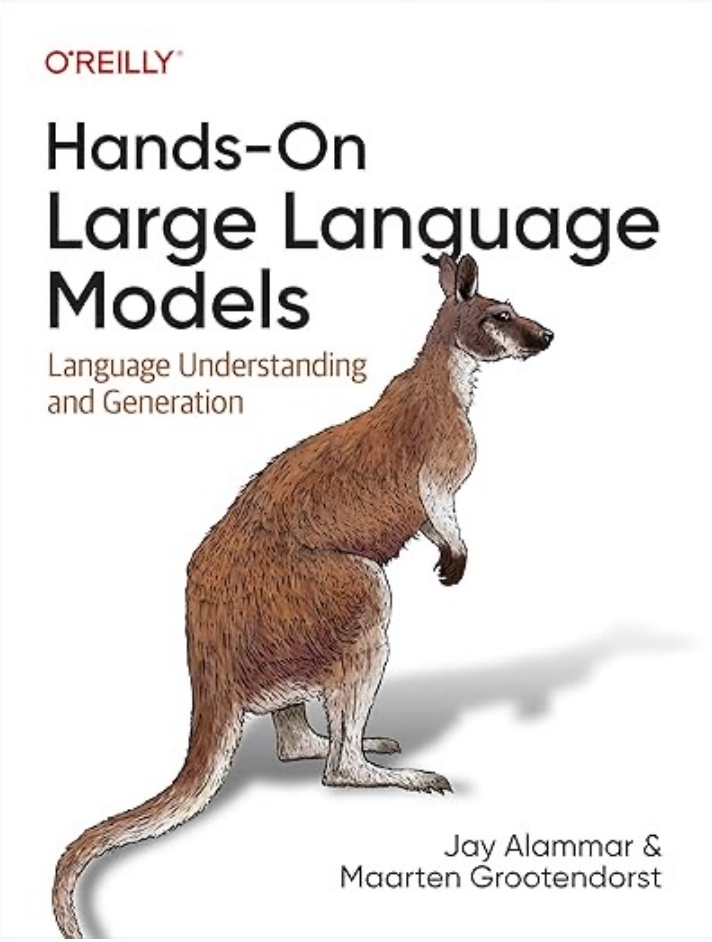


Source: https://www.amazon.com/RAG-Driven-Generative-retrieval-generation-LlamaIndex/dp/1836200919/

Jay Alammar and Maarten Grootendorst (2024),

Hands-On Large Language Models: Language Understanding and Generation,

O'Reilly Media



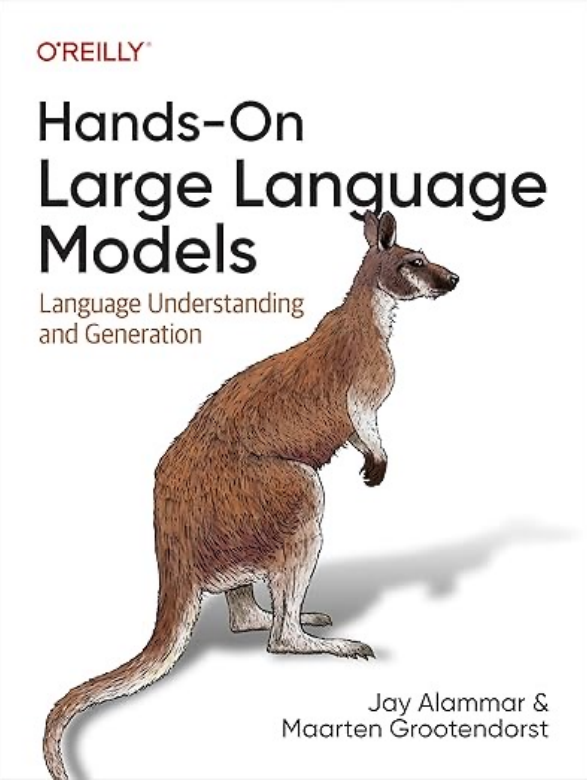
Source: https://www.amazon.com/Hands-Large-Language-Models-Understanding/dp/1098150961/

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

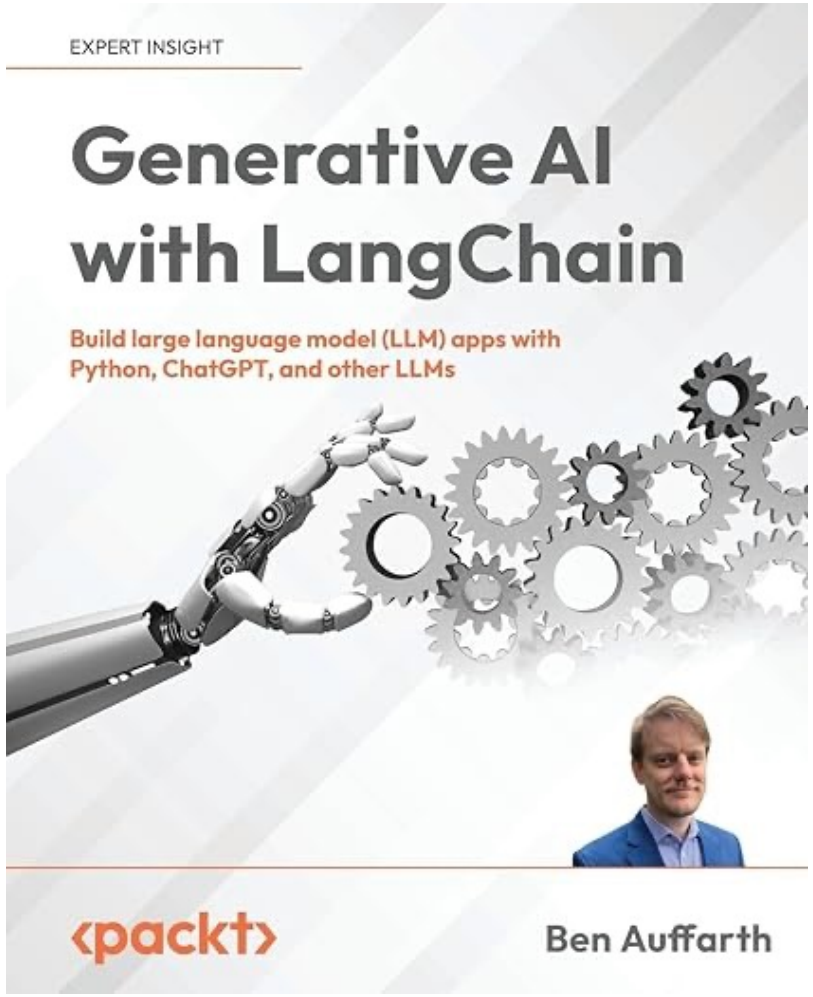


Source: https://github.com/HandsOnLLM/Hands-On-Large-Language-Models

Ben Auffarth (2023),

Generative AI with LangChain:

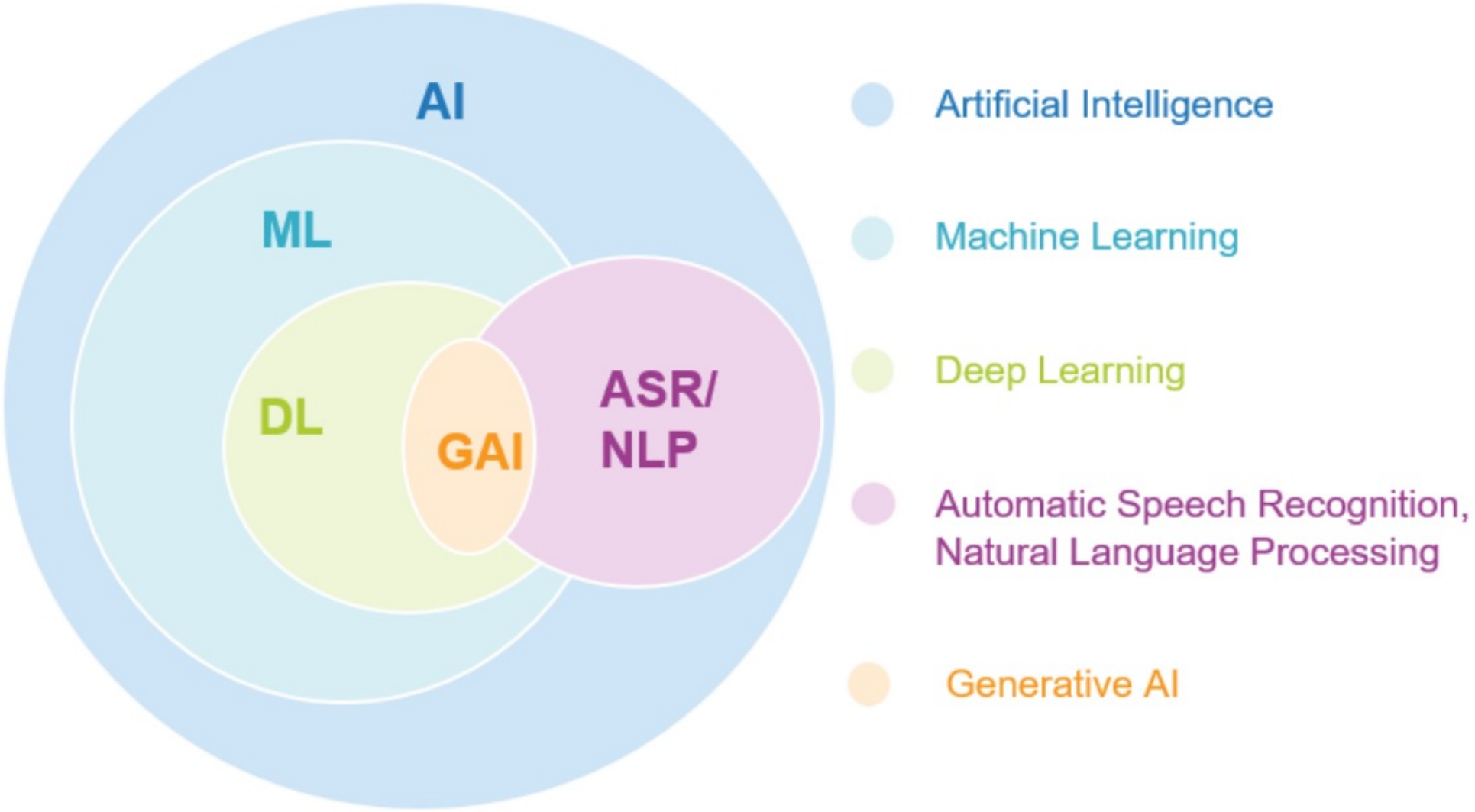
Build large language model (LLM) apps with Python, ChatGPT and other LLMs, Packt Publishing.



Source: https://www.amazon.com/Generative-AI-LangChain-language-ChatGPT/dp/1835083463

Artificial Intelligence (AI)

AI, ML, DL, Generative AI



Source: Jeong, Cheonsu. "A Study on the Implementation of Generative AI Services Using an Enterprise Data-Based LLM Application Architecture." arXiv preprint arXiv:2309.01105 (2023).

Generative AI, Agentic AI, Physical AI

2012 AlexNet
Deep learning breakthrough

Perception AI
Speech recognition
Deep recommender systems
Medical imaging

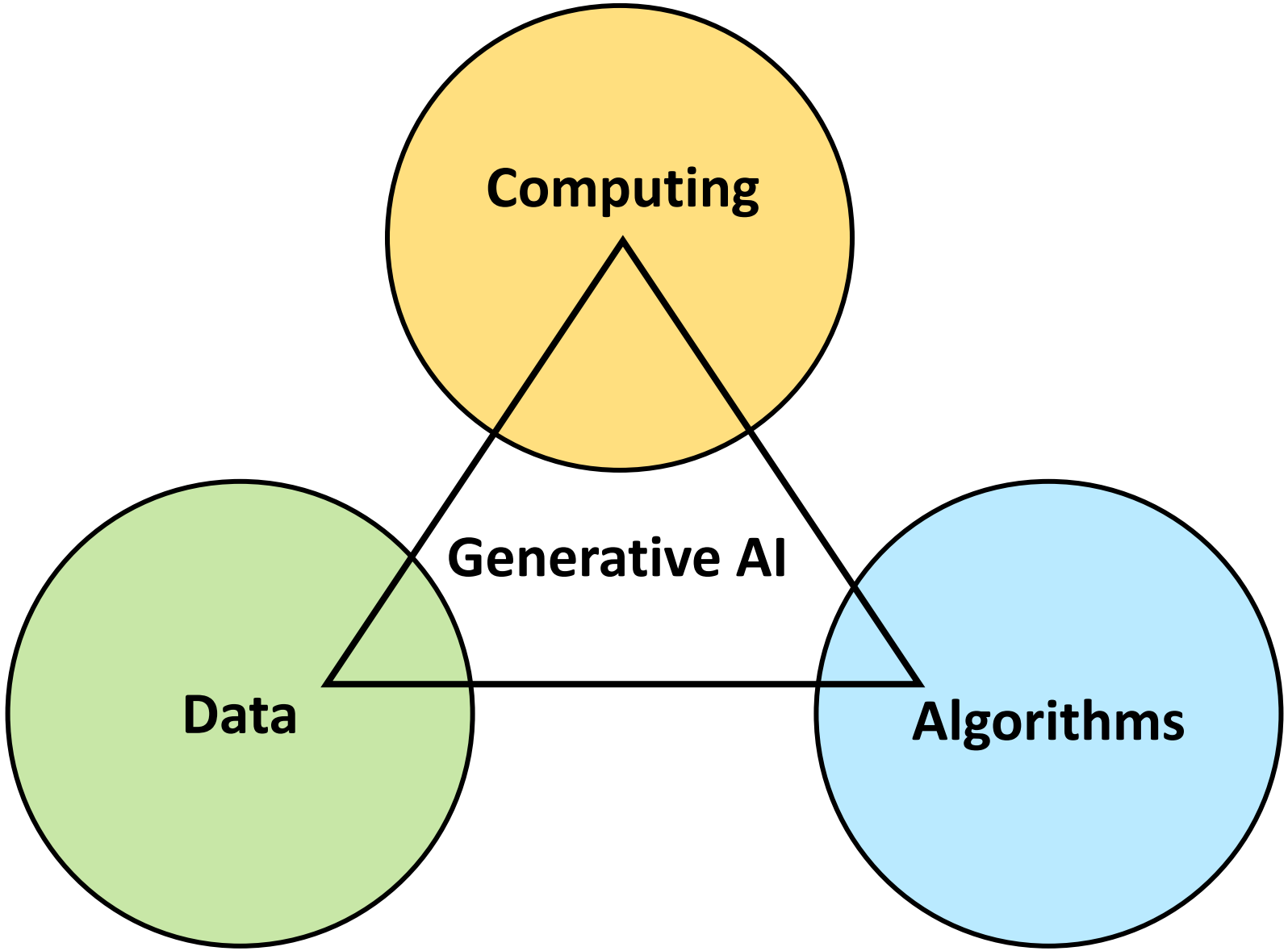
Generative AI
Digital marketing
Content creation

Agentic AI
Coding assistants
Customer service
Patient care

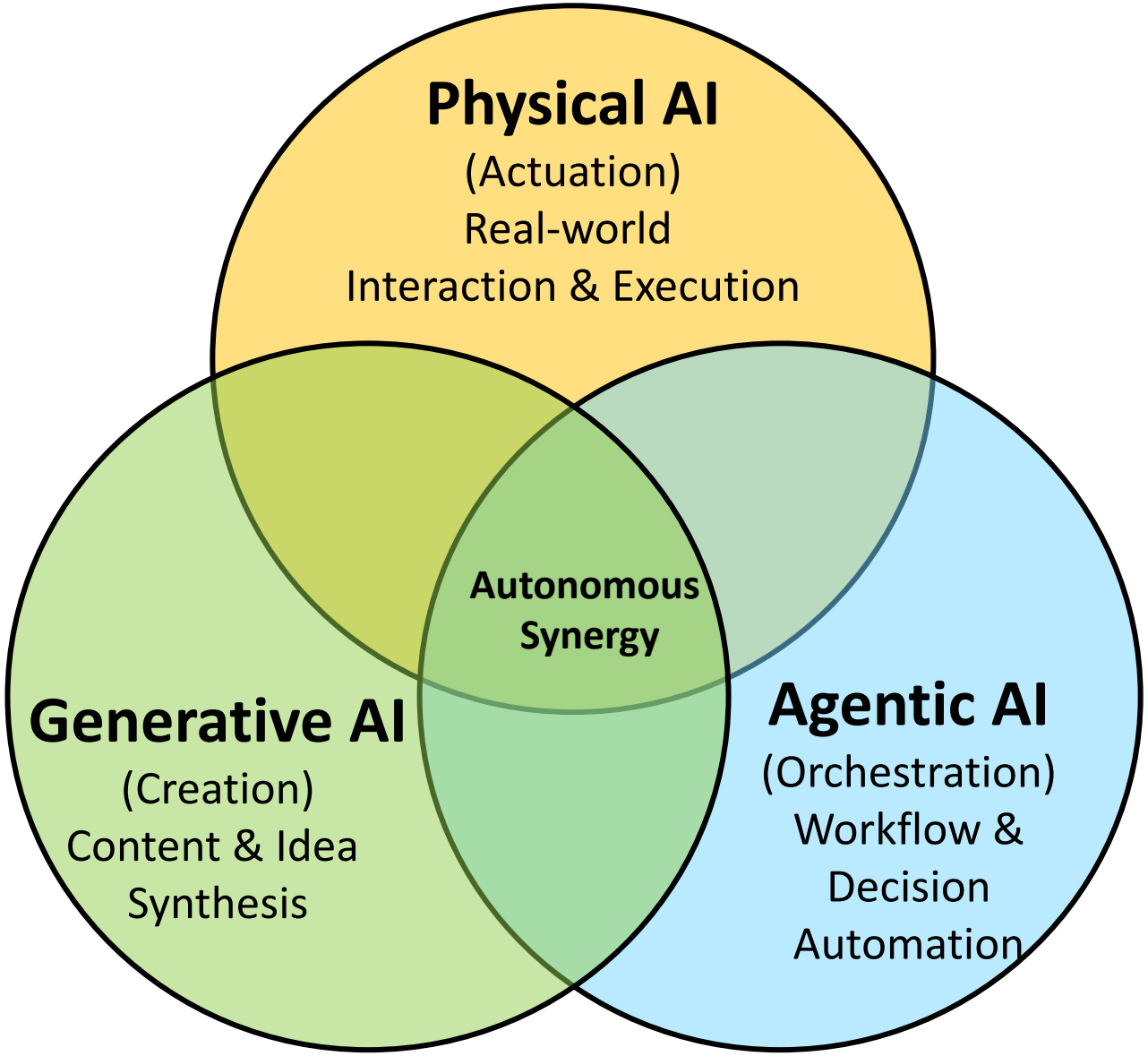
Physical AI
Self-driving cars
General robotics

Source: NVIDIA (2025), GTC March 2025 Keynote with NVIDIA CEO Jensen Huang, https://www.youtube.com/watch?v=_waPvOwL9Z8

Generative AI

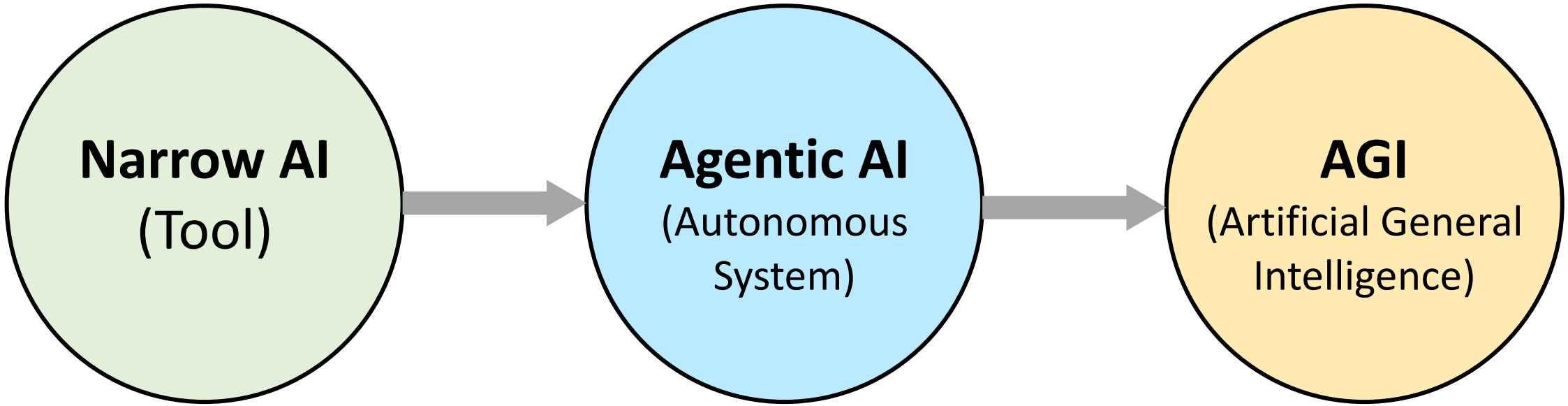


Generative AI, Agentic AI, Physical AI



New Economic Paradigm Shift: From Creation to Execution

The Future of AI
From Tools to Agents:
The Rise and Autonomy of Agentic AI



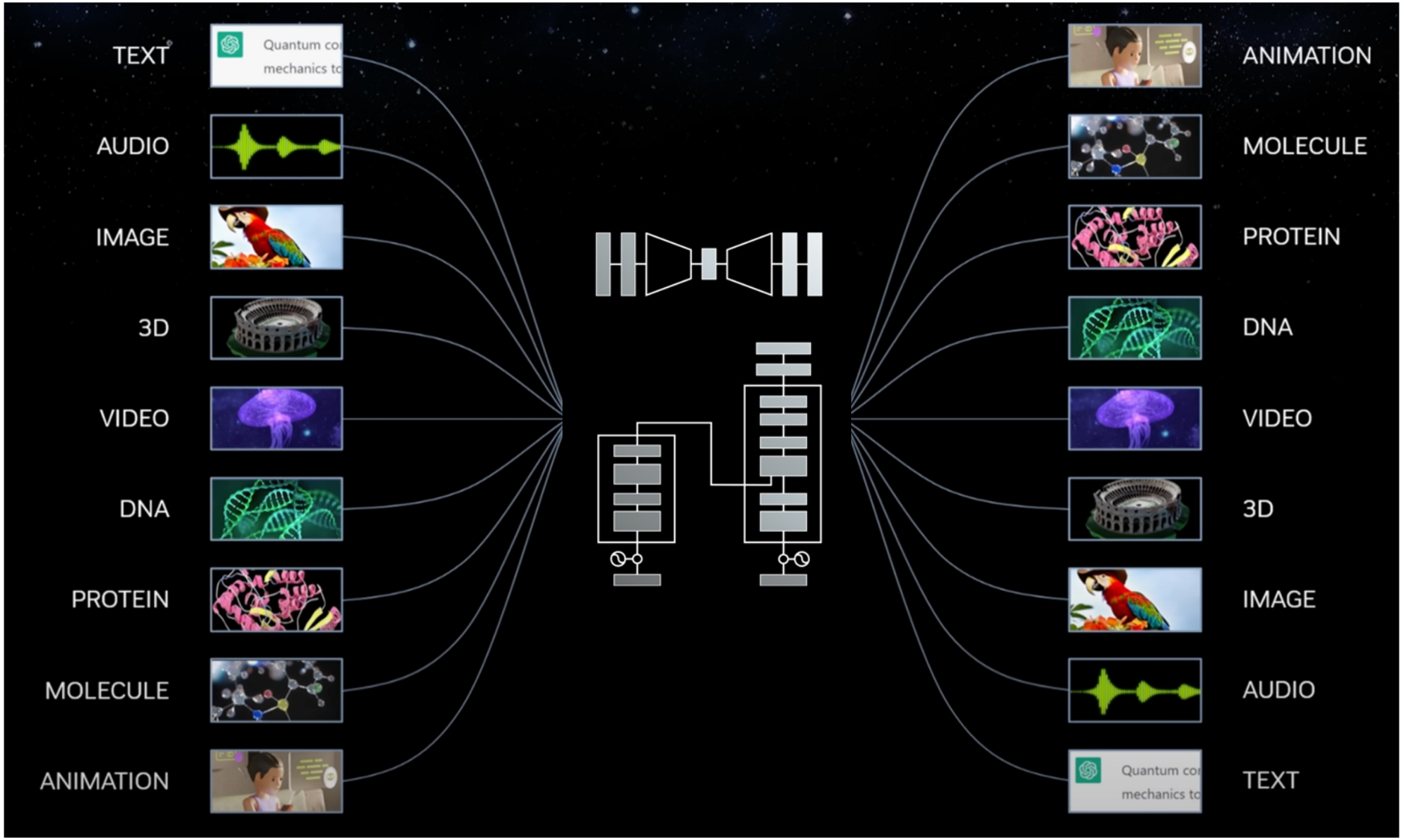
Executes Specific Tasks,
Content Generation

Goal-Oriented Behavior,
Autonomous Decision-Making,
Multi-Step Workflows,
Self-Improvement within Parameters

Connecting Multimodal GenAI
to Future AGI

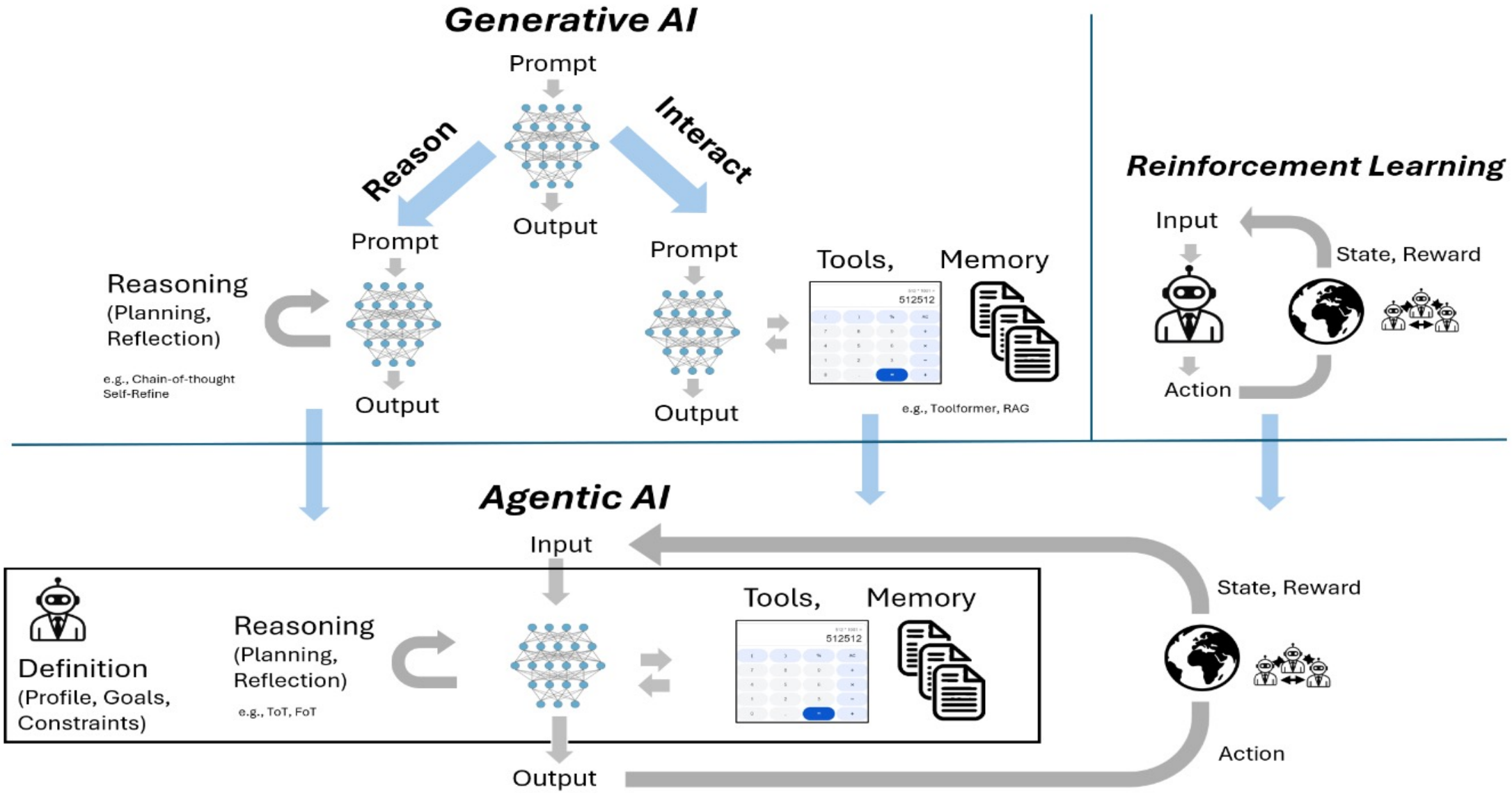
Modular Modalities

Where Can The Transformer Fit?



Source: NVIDIA DLI (2025), Rapid Application Development with Large Language Models (LLMs), https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-26+V1

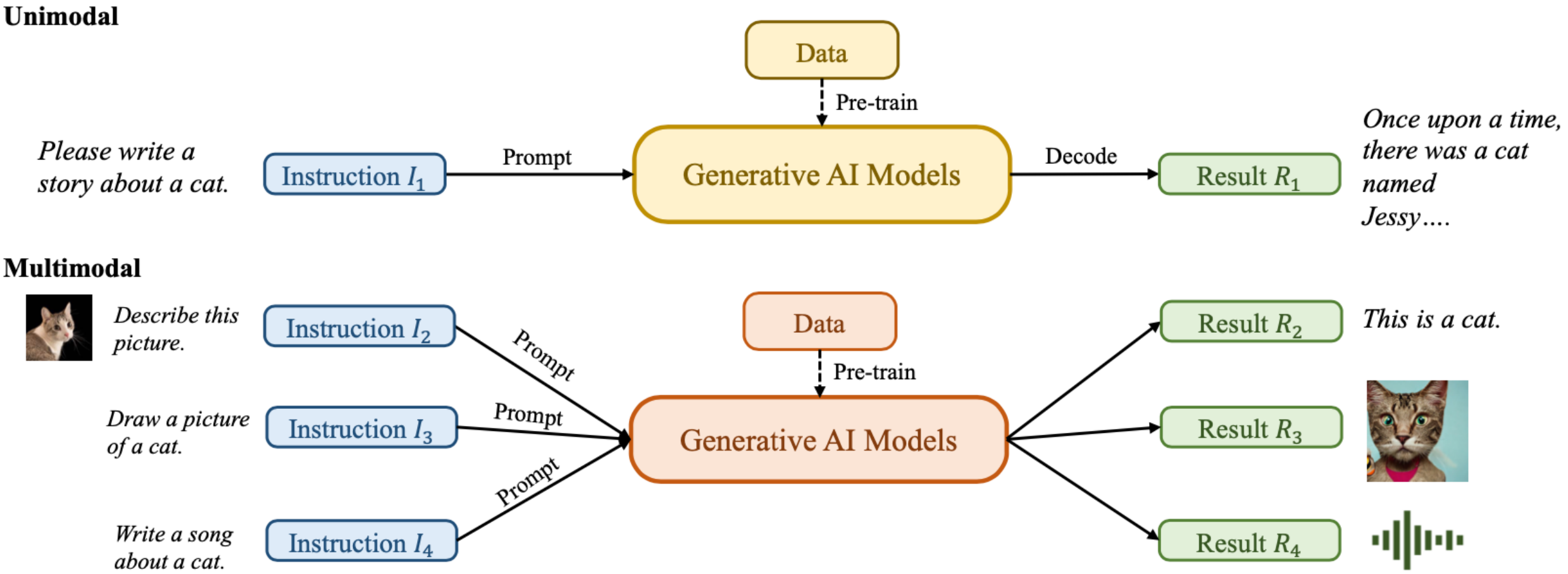
From Generative AI to Agentic AI



Source: Schneider, Johannes. "Generative to Agentic AI: Survey, Conceptualization, and Challenges." arXiv preprint arXiv:2504.18875 (2025).

Generative AI (Gen AI)

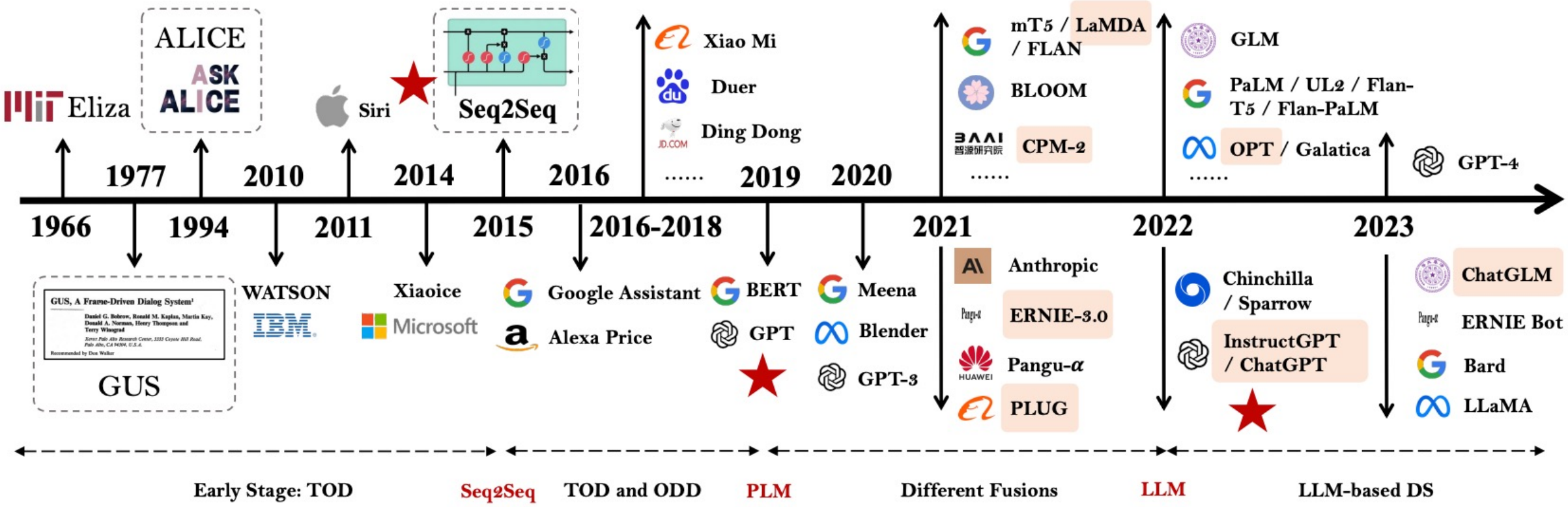
AI Generated Content (AIGC)



Source: Yihan Cao, Siyu Li, Yixin Liu, Zhiling Yan, Yutong Dai, Philip S. Yu, and Lichao Sun (2023). "A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT." arXiv preprint arXiv:2303.04226.

The Development of LM-based Dialogue Systems

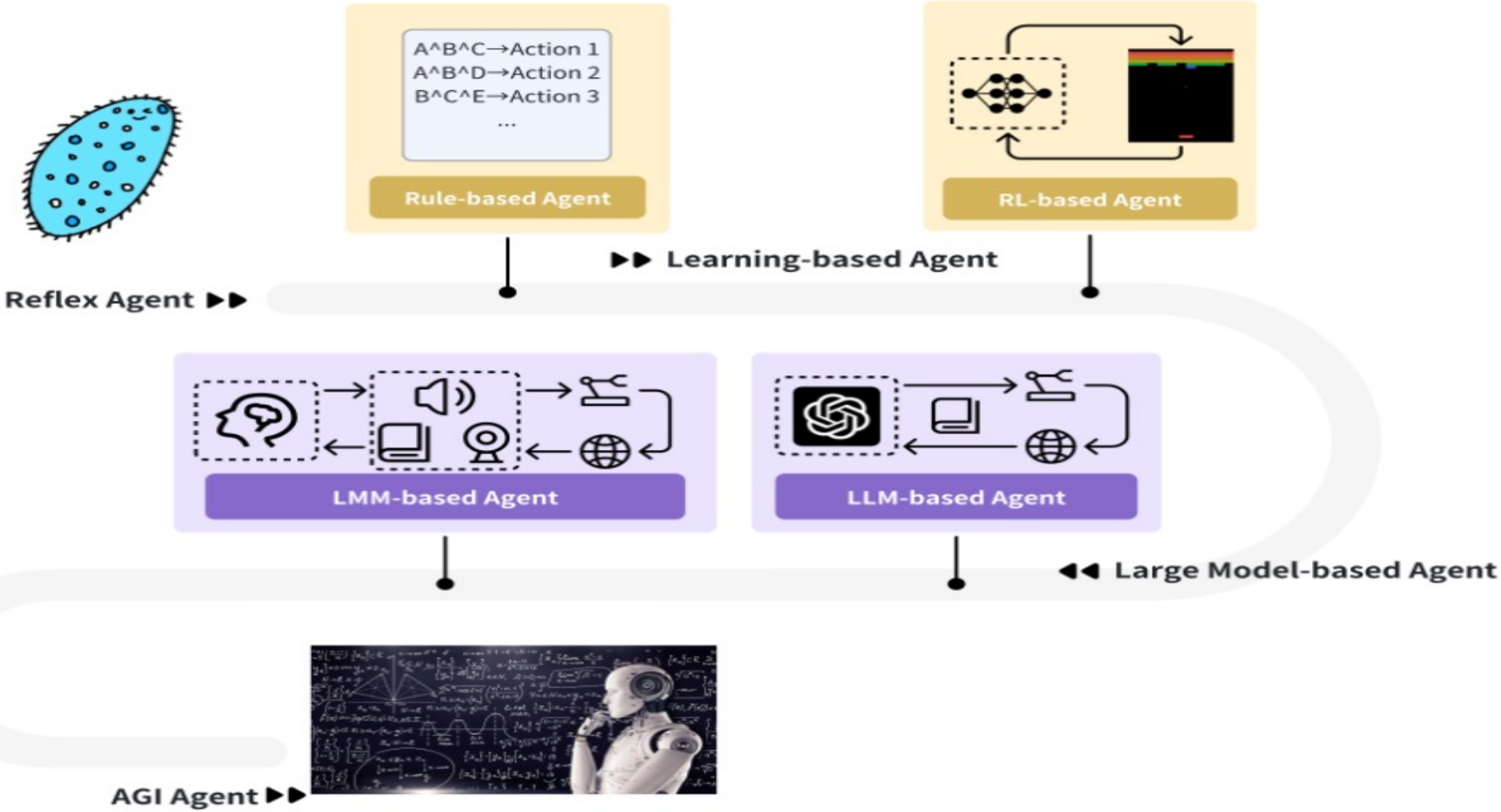
1) Early Stage (1966 - 2015)
2) The Independent Development of TOD and ODD (2015 - 2019)
3) Fusions of Dialogue Systems (2019 - 2022)
4) LLM-based DS (2022 - Now)



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

Source: Wang, Hongru, Lingzhi Wang, Yiming Du, Liang Chen, Jingyan Zhou, Yufei Wang, and Kam-Fai Wong. "A Survey of the Evolution of Language Model-Based Dialogue Systems." arXiv preprint arXiv:2311.16789 (2023).

Intelligent Agents Roadmap



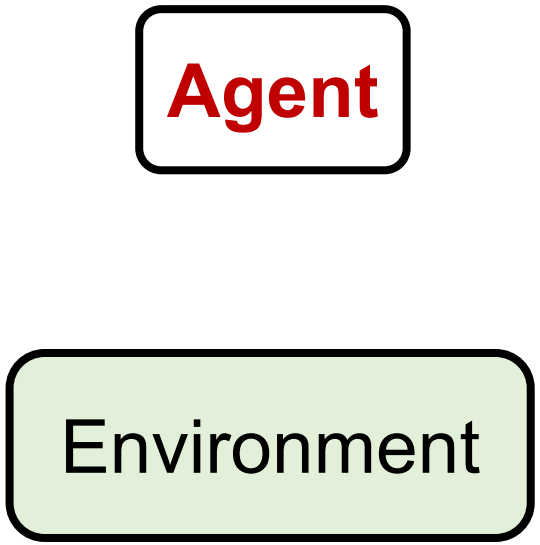
Source: Cheng, Yuheng, Ceyao Zhang, Zhengwen Zhang, Xiangrui Meng, Sirui Hong, Wenhao Li, Zihao Wang et al. "Exploring large language model based intelligent agents: Definitions, methods, and prospects." arXiv preprint arXiv:2401.03428 (2024).

AI Agents

- Traditional AI Agents
  - Simple reflex agents
  - Model-based reflex agents
  - Goal-based agents
  - Utility-based agents
  - Learning agents

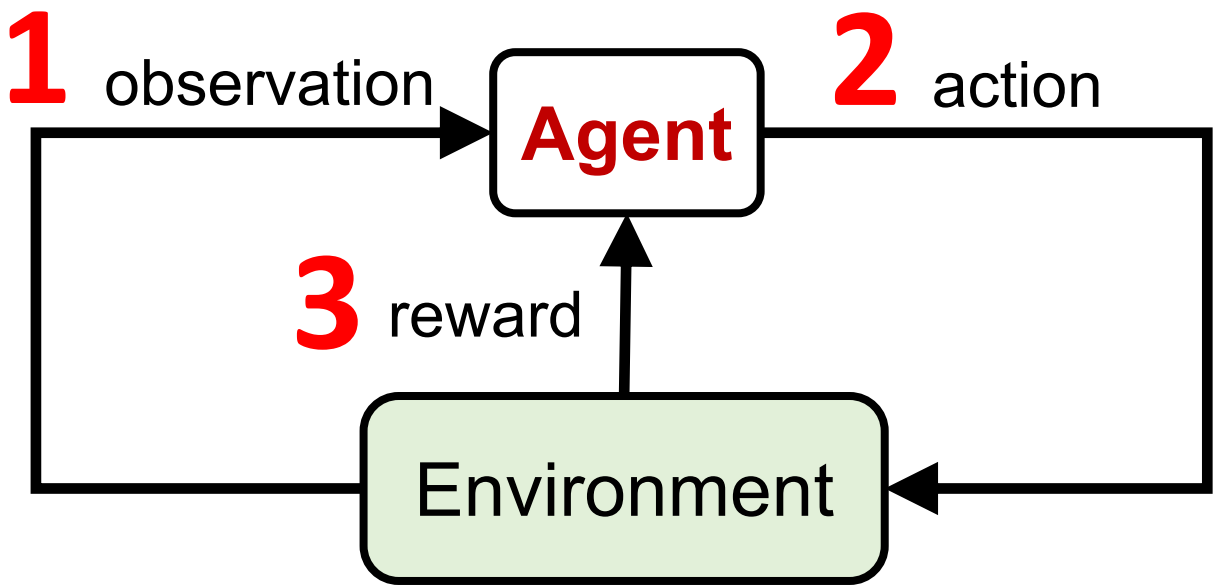
- Evolution of AI Agents
  - LLM-based Agents
  - Multi-modal agents
  - Embodied AI agents in virtual environments
  - Collaborative AI agents

Reinforcement Learning (DL)



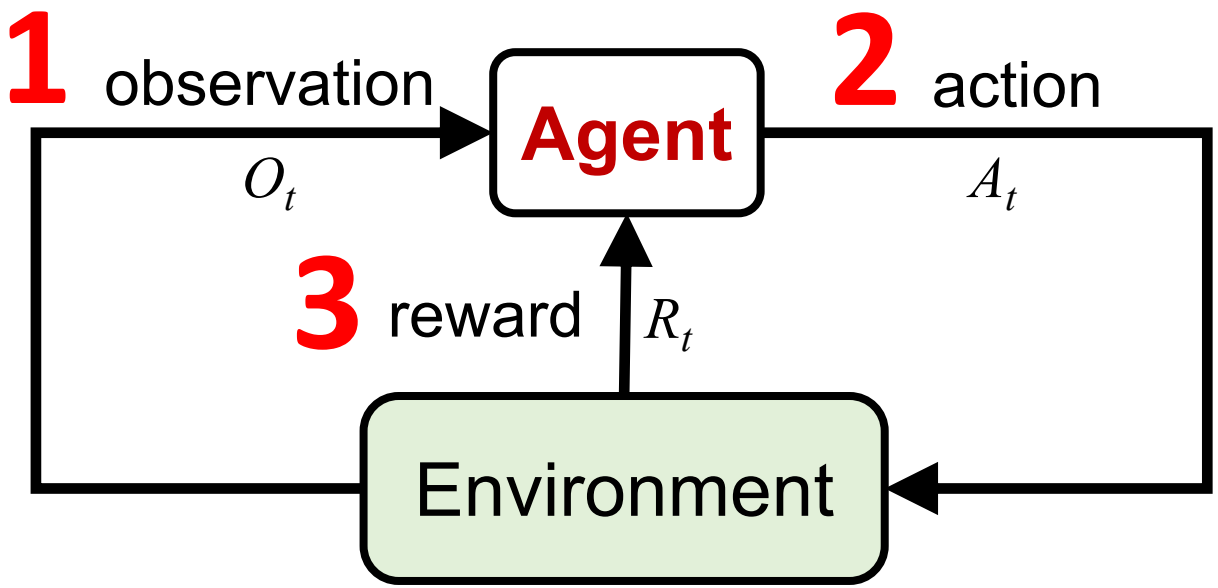
Source: Richard S. Sutton & Andrew G. Barto (2018), Reinforcement Learning: An Introduction, 2nd Edition, A Bradford Book.

Reinforcement Learning (DL)



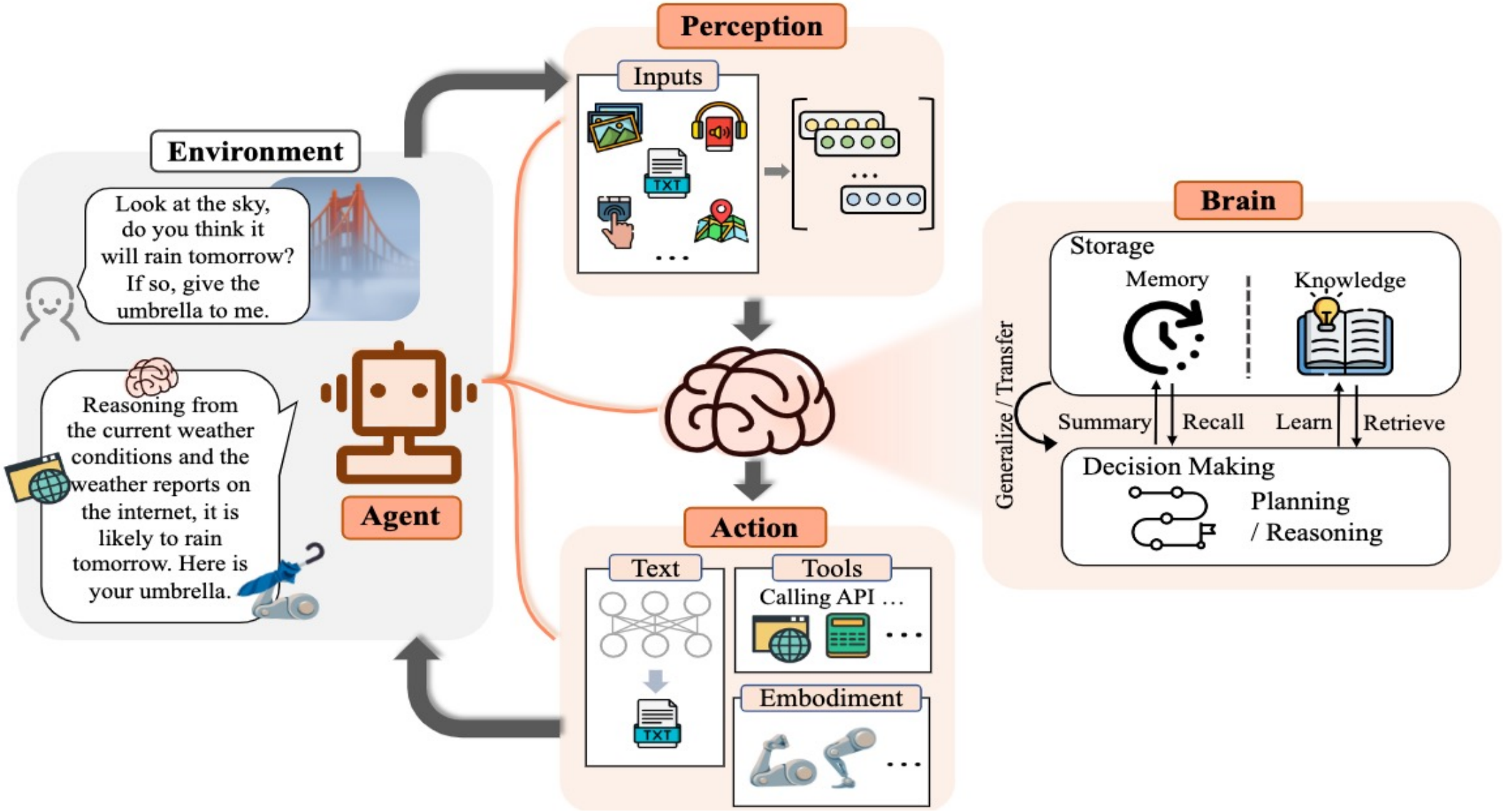
Source: Richard S. Sutton & Andrew G. Barto (2018), Reinforcement Learning: An Introduction, 2nd Edition, A Bradford Book.

Reinforcement Learning (DL)



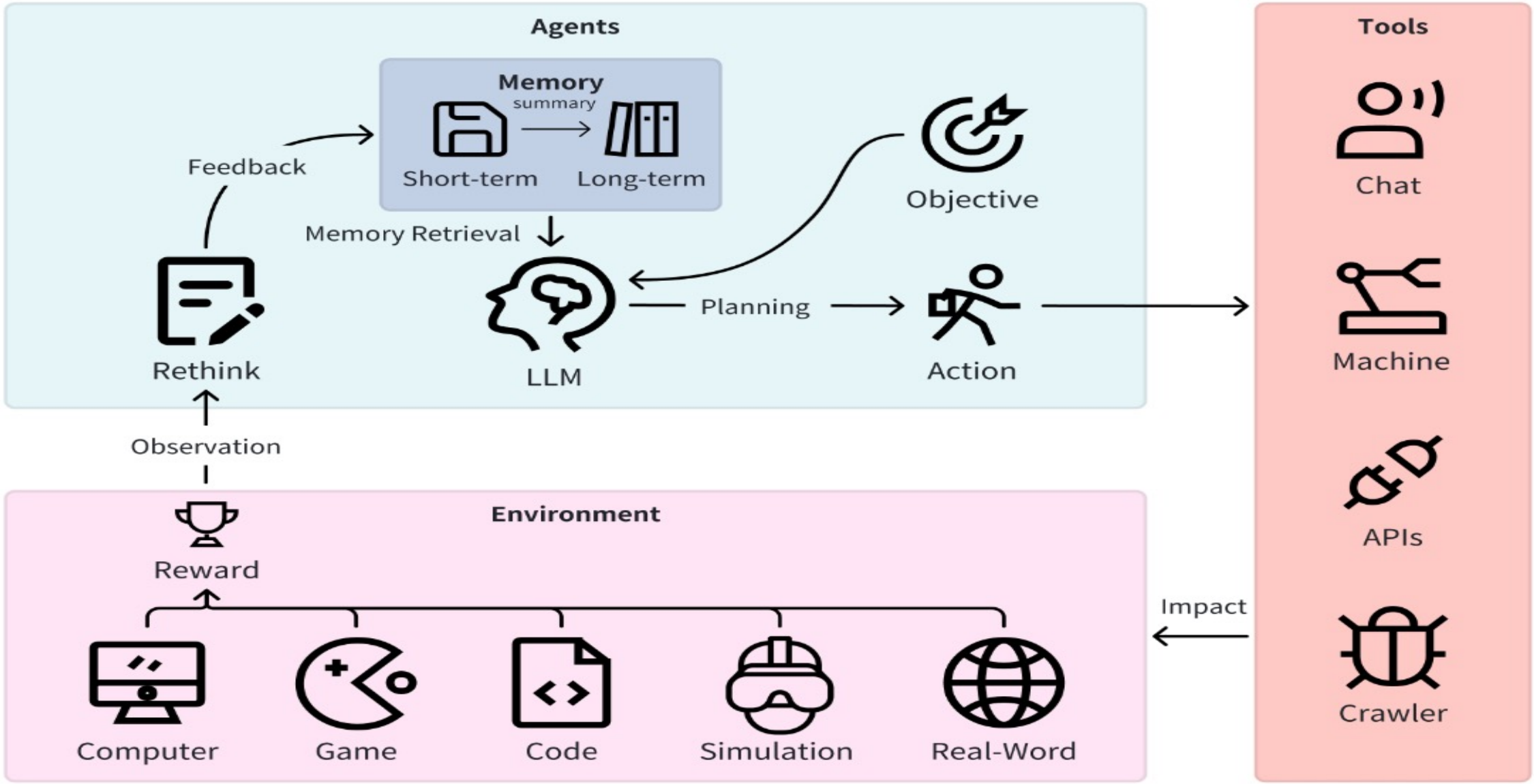
Source: Richard S. Sutton & Andrew G. Barto (2018), Reinforcement Learning: An Introduction, 2nd Edition, A Bradford Book.

Large Language Model (LLM) based Agents



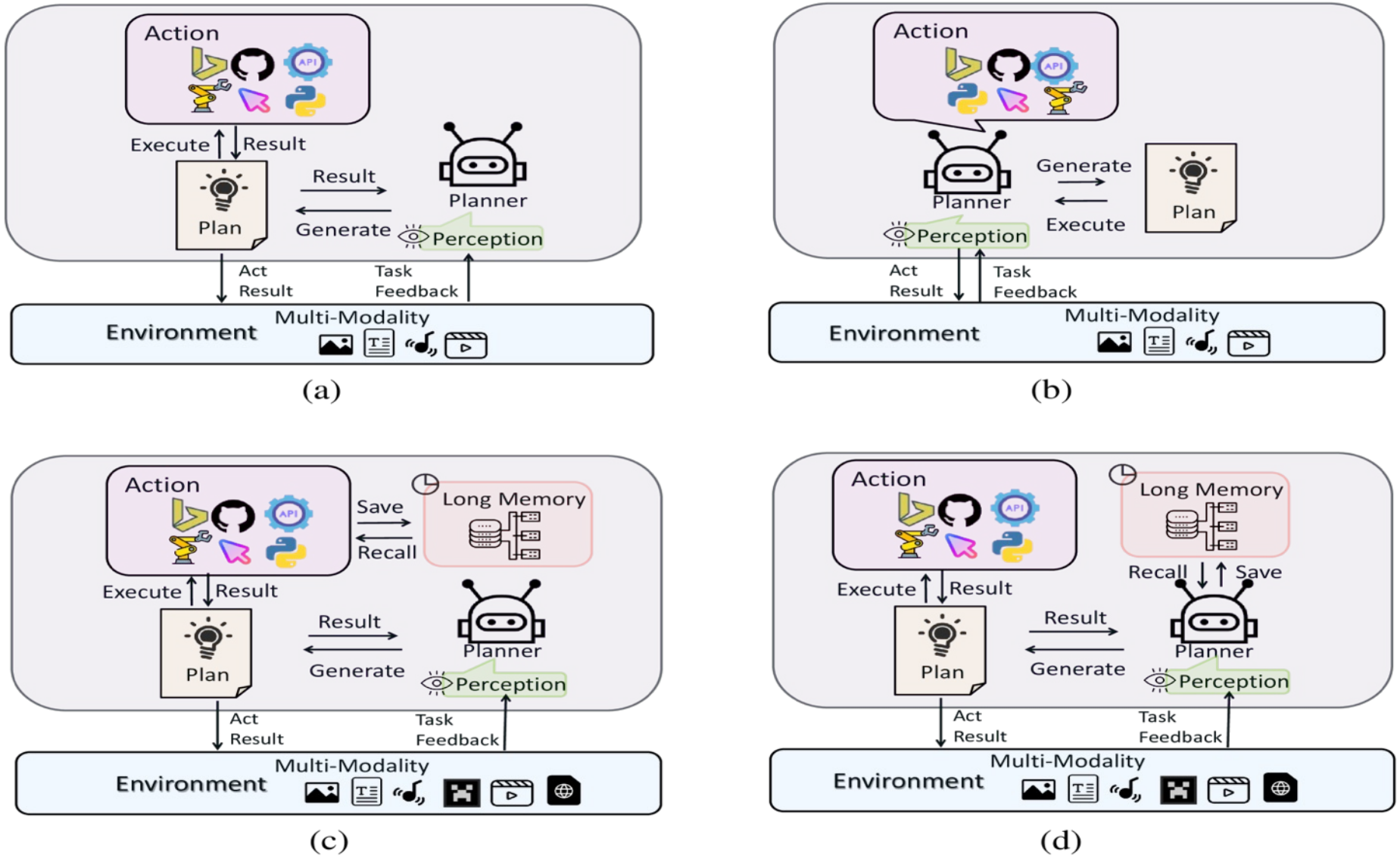
Source: Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., ... & Gui, T. (2023). The rise and potential of large language model based agents: A survey. arXiv preprint arXiv:2309.07864.

LLM-based Agents



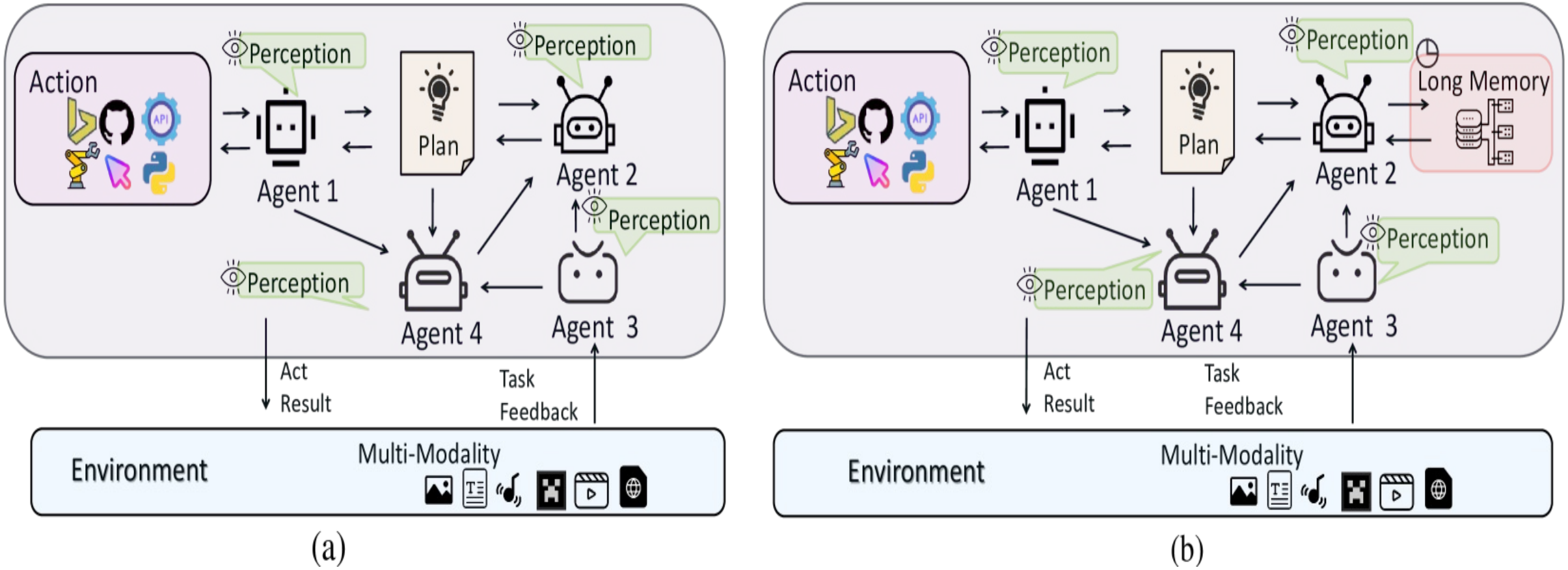
Source: Cheng, Yuheng, Ceyao Zhang, Zhengwen Zhang, Xiangrui Meng, Sirui Hong, Wenhao Li, Zihao Wang et al. "Exploring large language model based intelligent agents: Definitions, methods, and prospects." arXiv preprint arXiv:2401.03428 (2024).

Large Multimodal Agents (LMA)



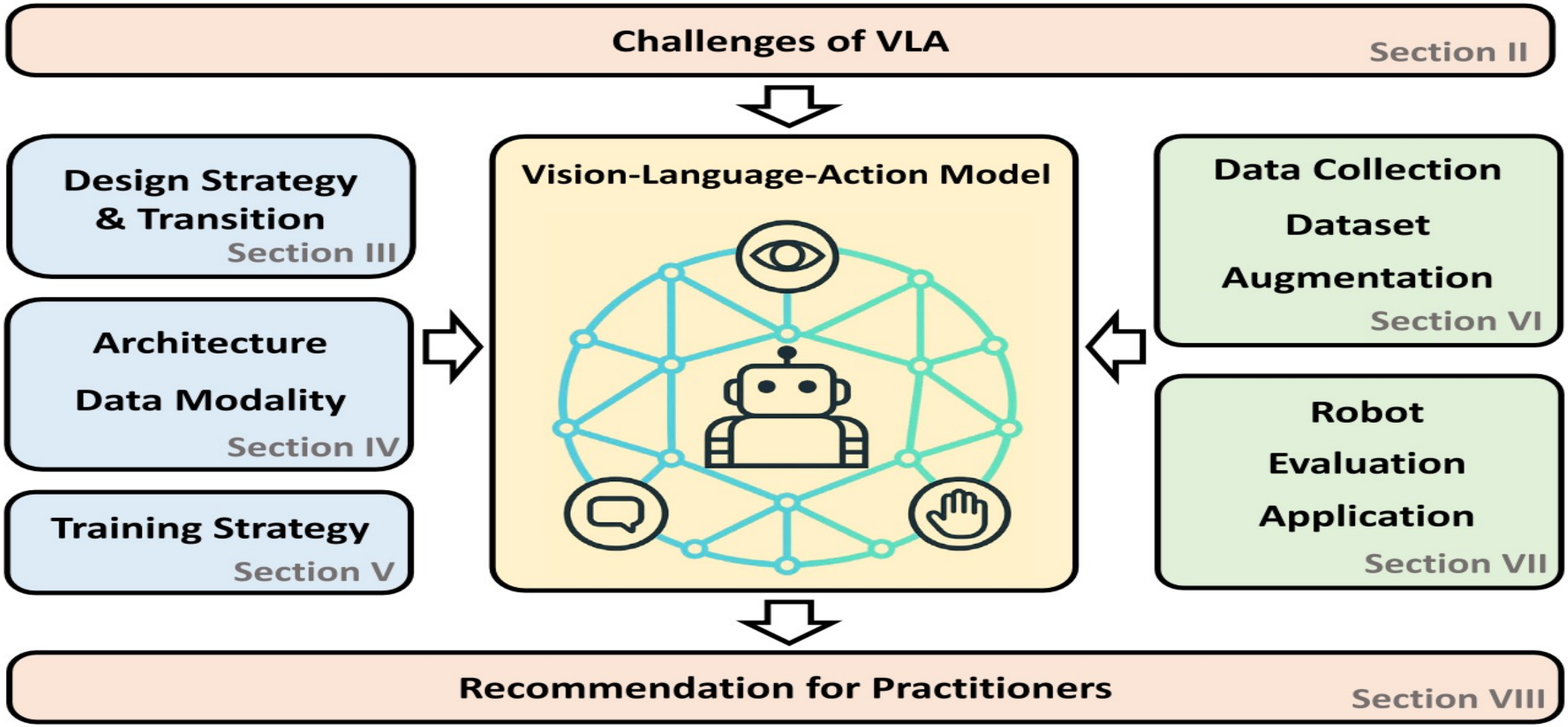
Source: Xie, J., Chen, Z., Zhang, R., Wan, X., & Li, G. (2024). Large Multimodal Agents: A Survey. ArXiv, abs/2402.15116.

Large Multimodal Agents (LMA)



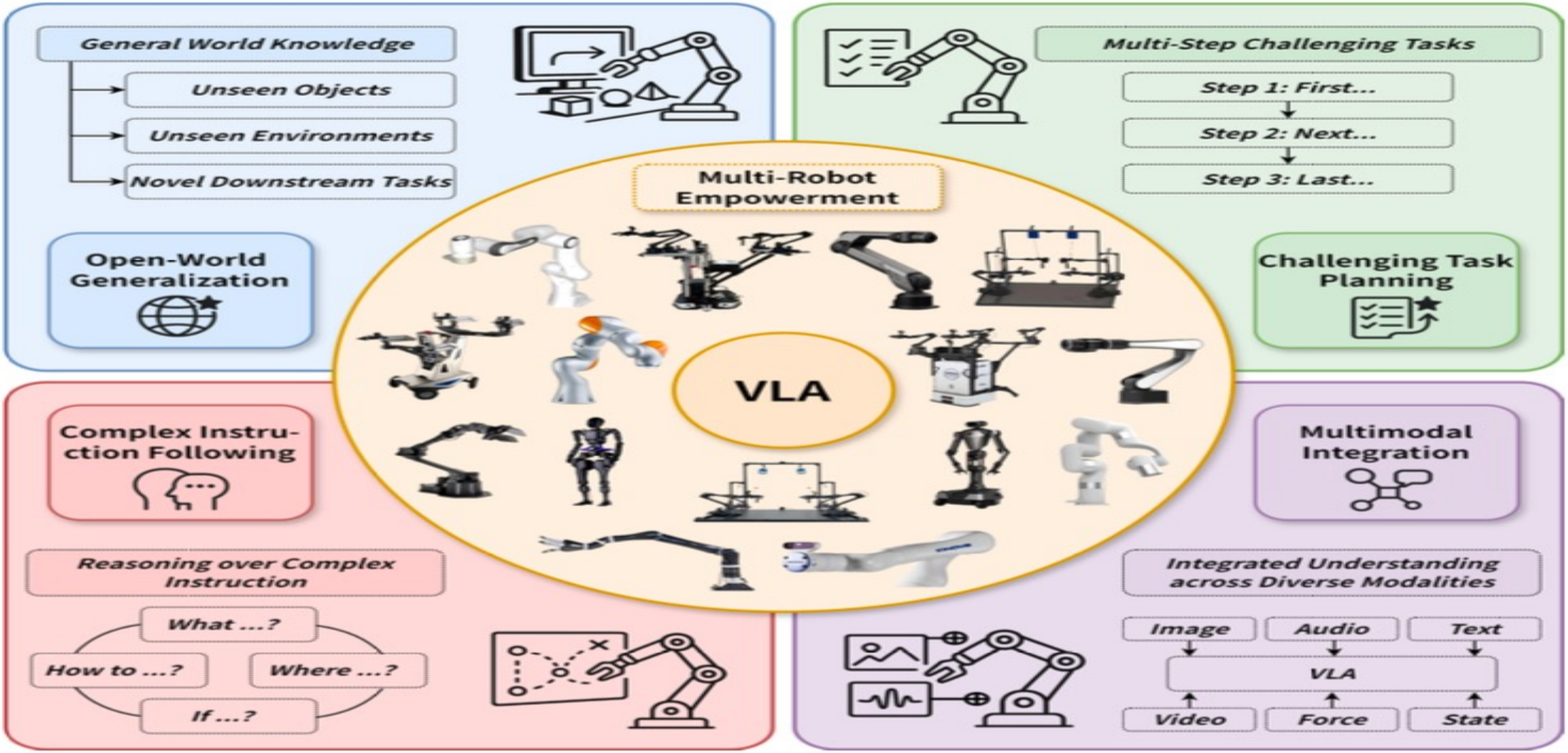
Source: Xie, J., Chen, Z., Zhang, R., Wan, X., & Li, G. (2024). Large Multimodal Agents: A Survey. ArXiv, abs/2402.15116.

Vision-Language-Action (VLA) Models for Robotics



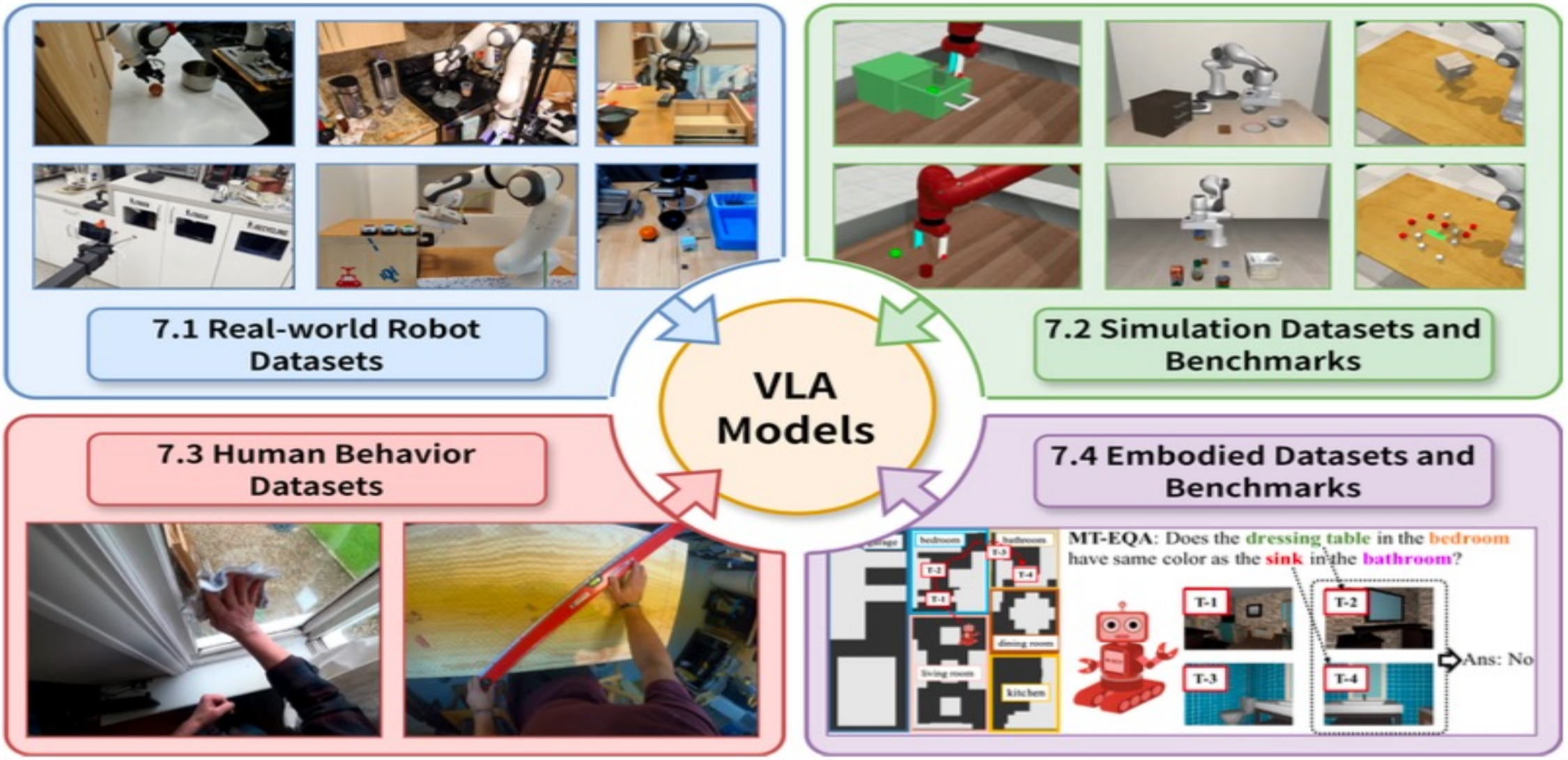
Source: Kento Kawaharazuka, Jihoon Oh, Jun Yamada, Ingmar Posner, and Yuke Zhu. (2025) "Vision-language-action models for robotics: A review towards real-world applications." IEEE Access (2025).

Large VLM-based Vision-Language-Action Models for Robotic Manipulation



Source: Rui Shao, Wei Li, Lingsen Zhang, Renshan Zhang, Zhiyang Liu, Ran Chen, and Liqiang Nie. (2025) "Large vlm-based vision-language-action models for robotic manipulation: A survey." arXiv preprint arXiv:2508.13073 (2025).

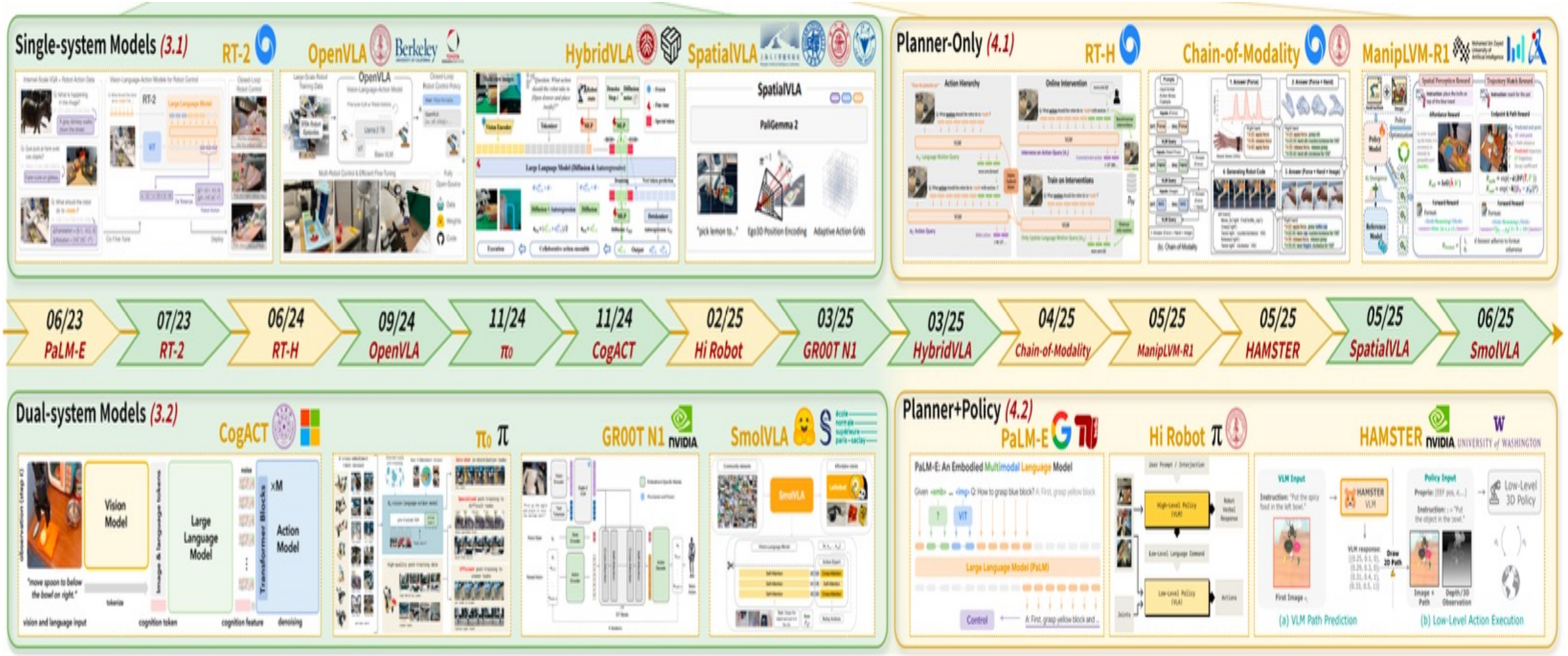
Large VLM-based Vision-Language-Action Models for Robotic Manipulation



Source: Rui Shao, Wei Li, Lingsen Zhang, Renshan Zhang, Zhiyang Liu, Ran Chen, and Liqiang Nie. (2025) "Large vlm-based vision-language-action models for robotic manipulation: A survey." arXiv preprint arXiv:2508.13073 (2025).

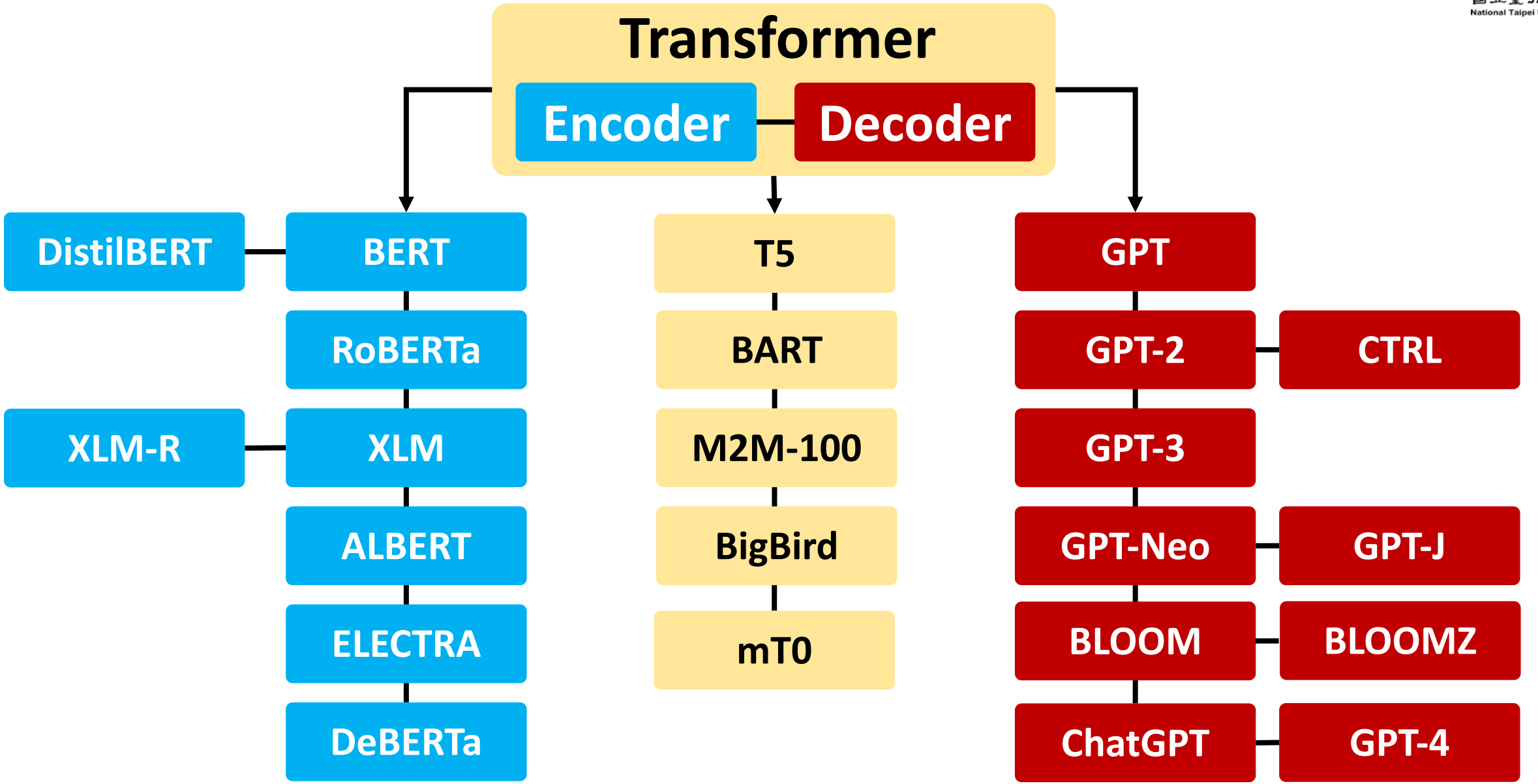
Large VLM-based Vision-Language-Action Models for Robotic Manipulation (Timeline)

Monolithic models and Hierarchical Models



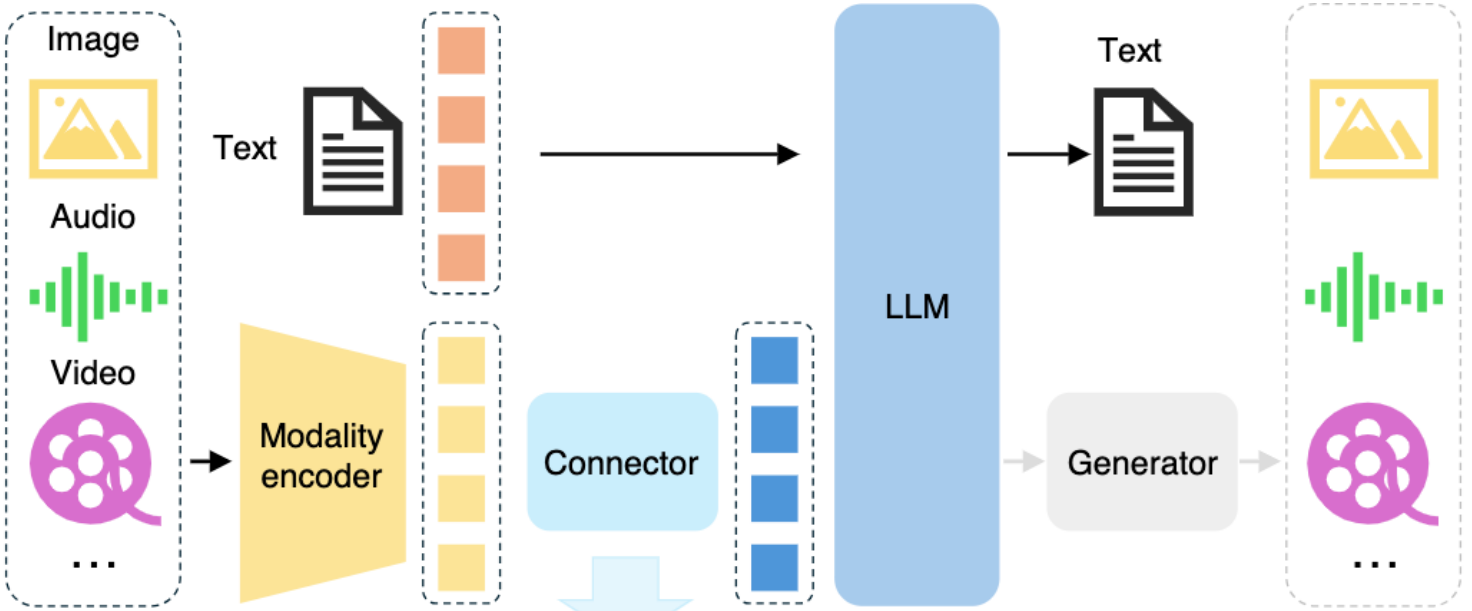
Source: Rui Shao, Wei Li, Lingsen Zhang, Renshan Zhang, Zhiyang Liu, Ran Chen, and Liqiang Nie. (2025) "Large vlm-based vision-language-action models for robotic manipulation: A survey." arXiv preprint arXiv:2508.13073 (2025).

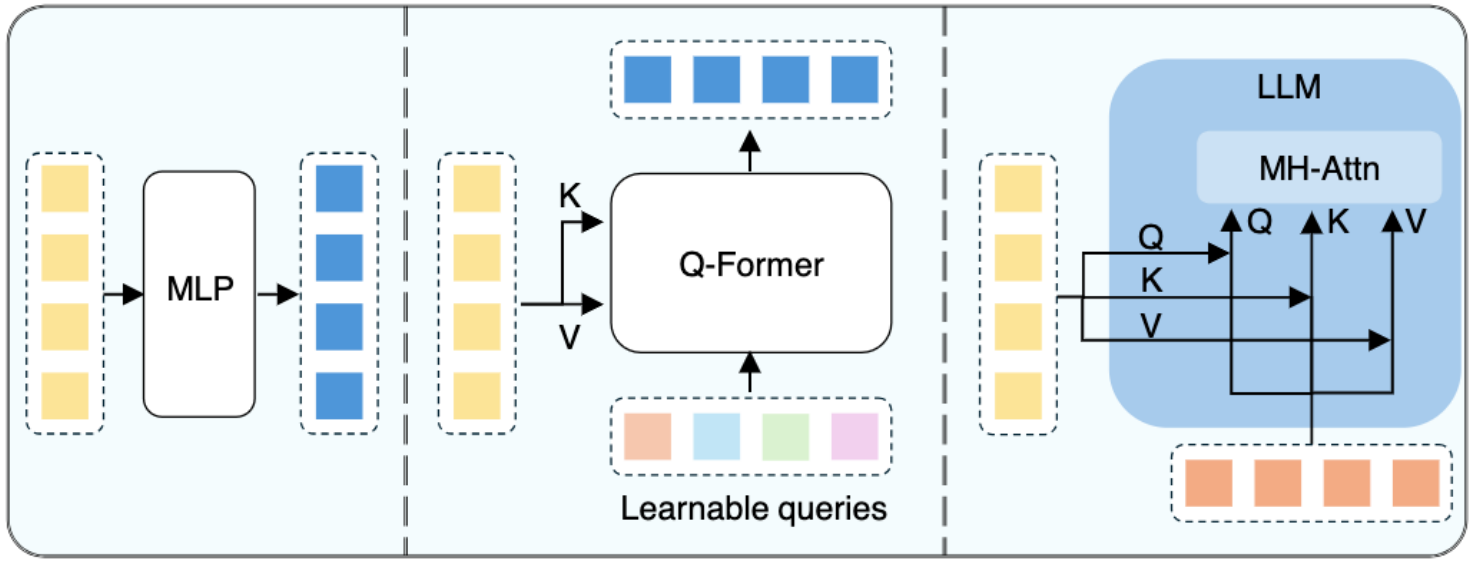
Transformer Models



Source: Lewis Tunstall, Leandro von Werra, and Thomas Wolf (2022), Natural Language Processing with Transformers: Building Language Applications with Hugging Face, O'Reilly Media.

Multimodal Large Language Models (MLLM)

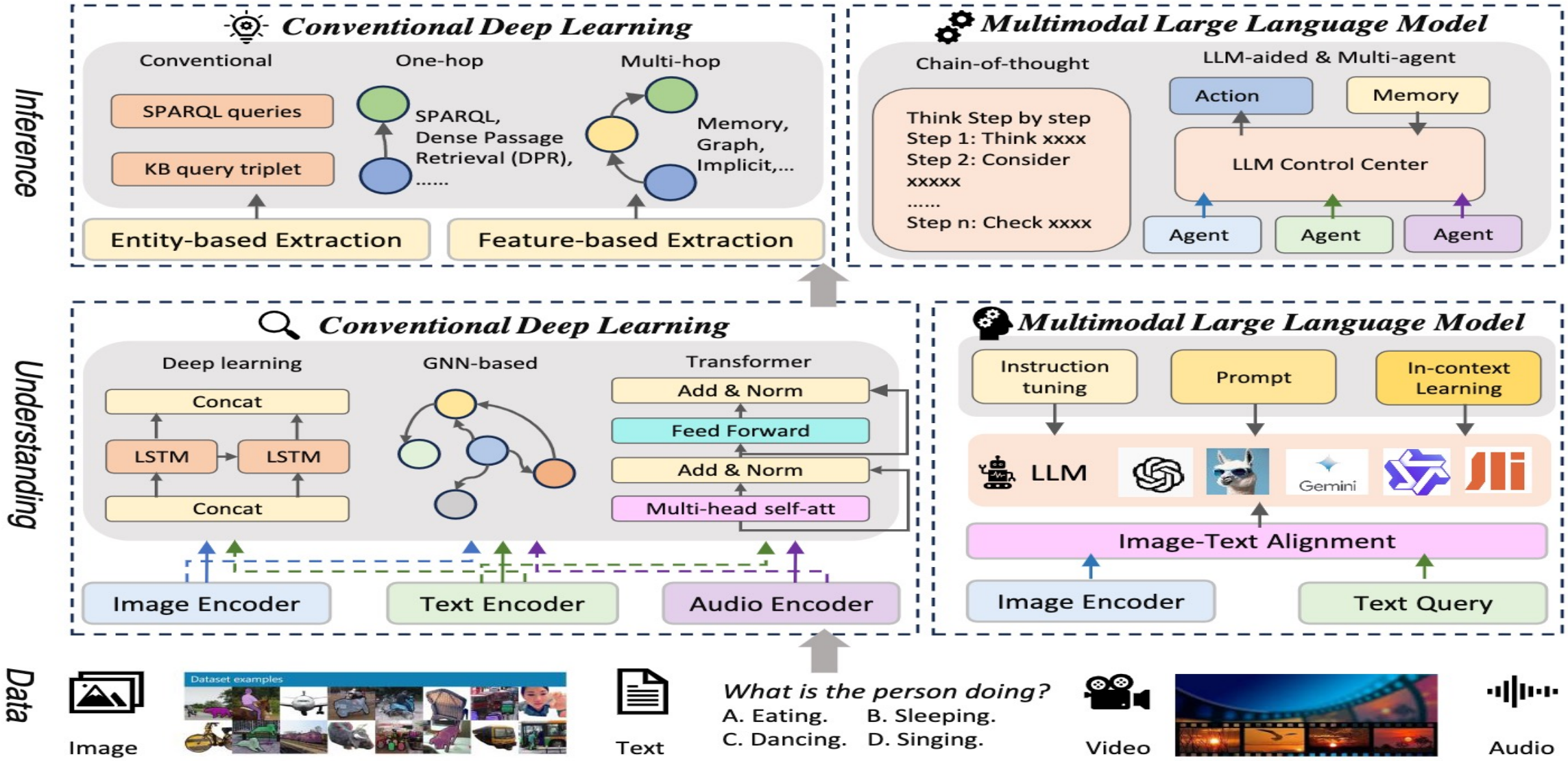




Multimodall LLM
Three types of connectors:
1. projection-based
2. query-based
3. fusion-based connectors

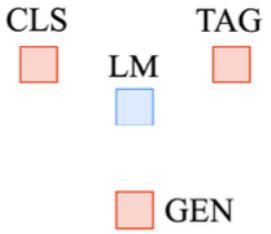
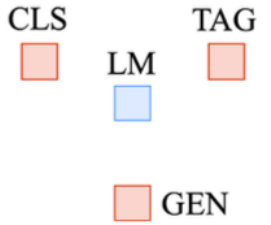
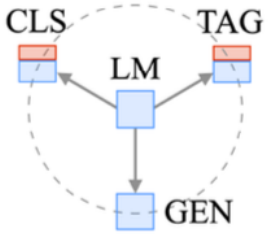
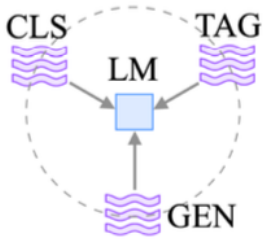
Source: Shukang Yin, Chaoyou Fu, Sirui Zhao, Ke Li, Xing Sun, Tong Xu, and Enhong Chen. (2024) "A survey on multimodal large language models." National Science Review (2024): nwae403.

Multimodal Large Language Model (MLLM) for Vision Question Answering



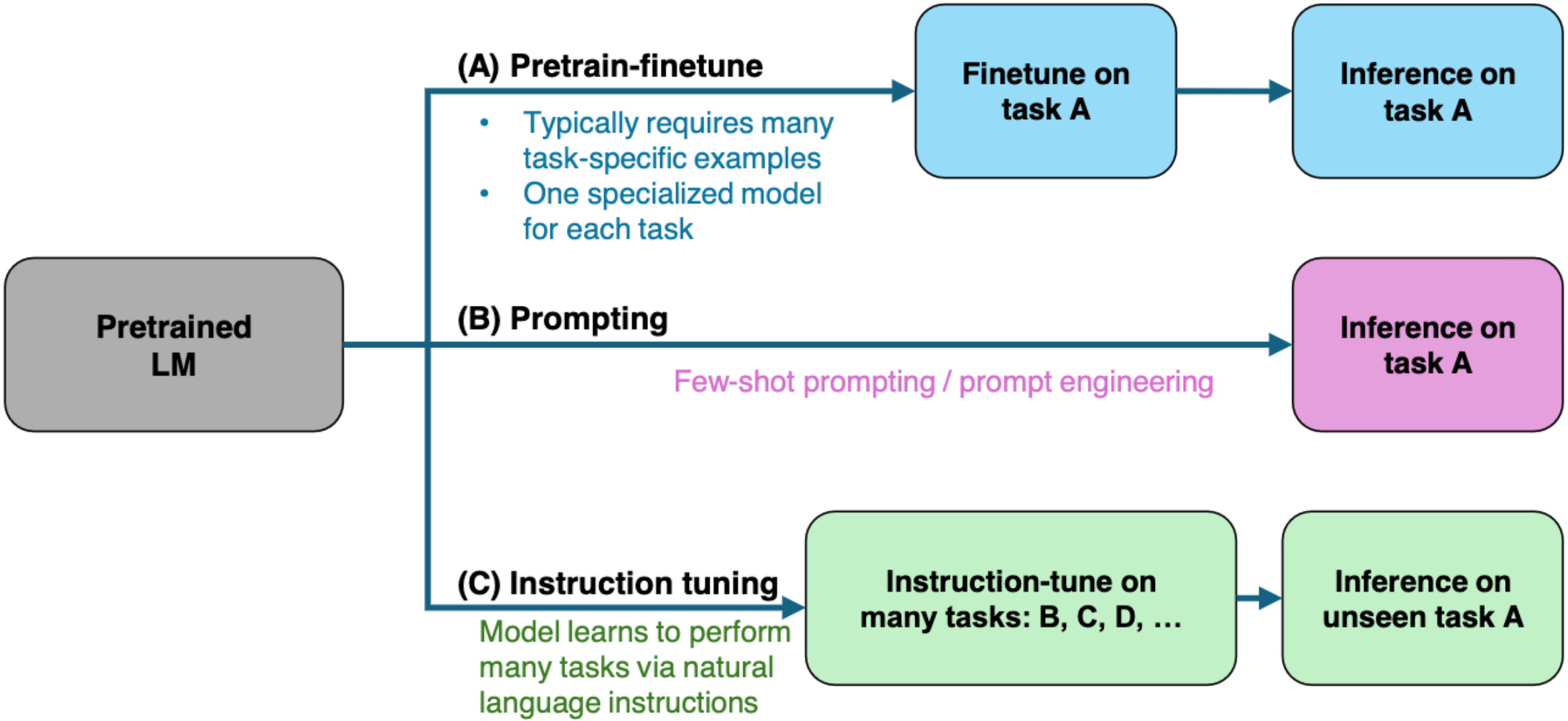
Source: Jiayi Kuang, Jingyou Xie, Haohao Luo, Ronghao Li, Zhe Xu, Xianfeng Cheng, Yinghui Li, Xika Lin, and Ying Shen. (2024) "Natural Language Understanding and Inference with MLLM in Visual Question Answering: A Survey." arXiv preprint arXiv:2411.17558.

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)	
Transfer Learning: Pre-training, Fine-Tuning (FT) c. Pre-train, Fine-tune	Objective (e.g. masked language modeling, next sentence prediction)	
GAI: Pre-train, Prompt, and Predict (Prompting) d. Pre-train, Prompt, Predict	Prompt (e.g. cloze, prefix)	

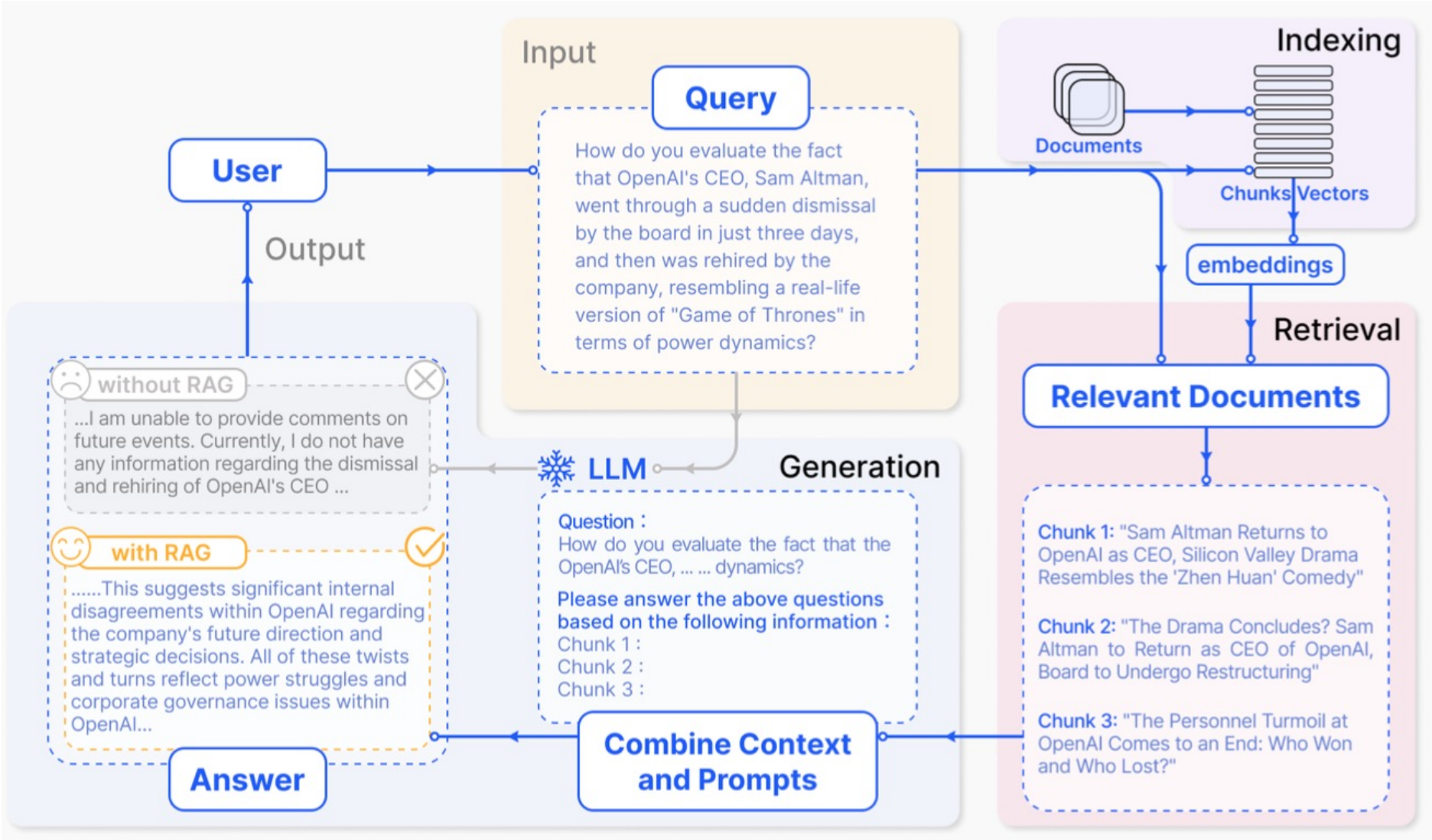
Source: Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. (2023) "Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing." ACM Computing Surveys 55, no. 9 (2023): 1-35.

Large Language Models (LLM)
Three typical learning paradigms



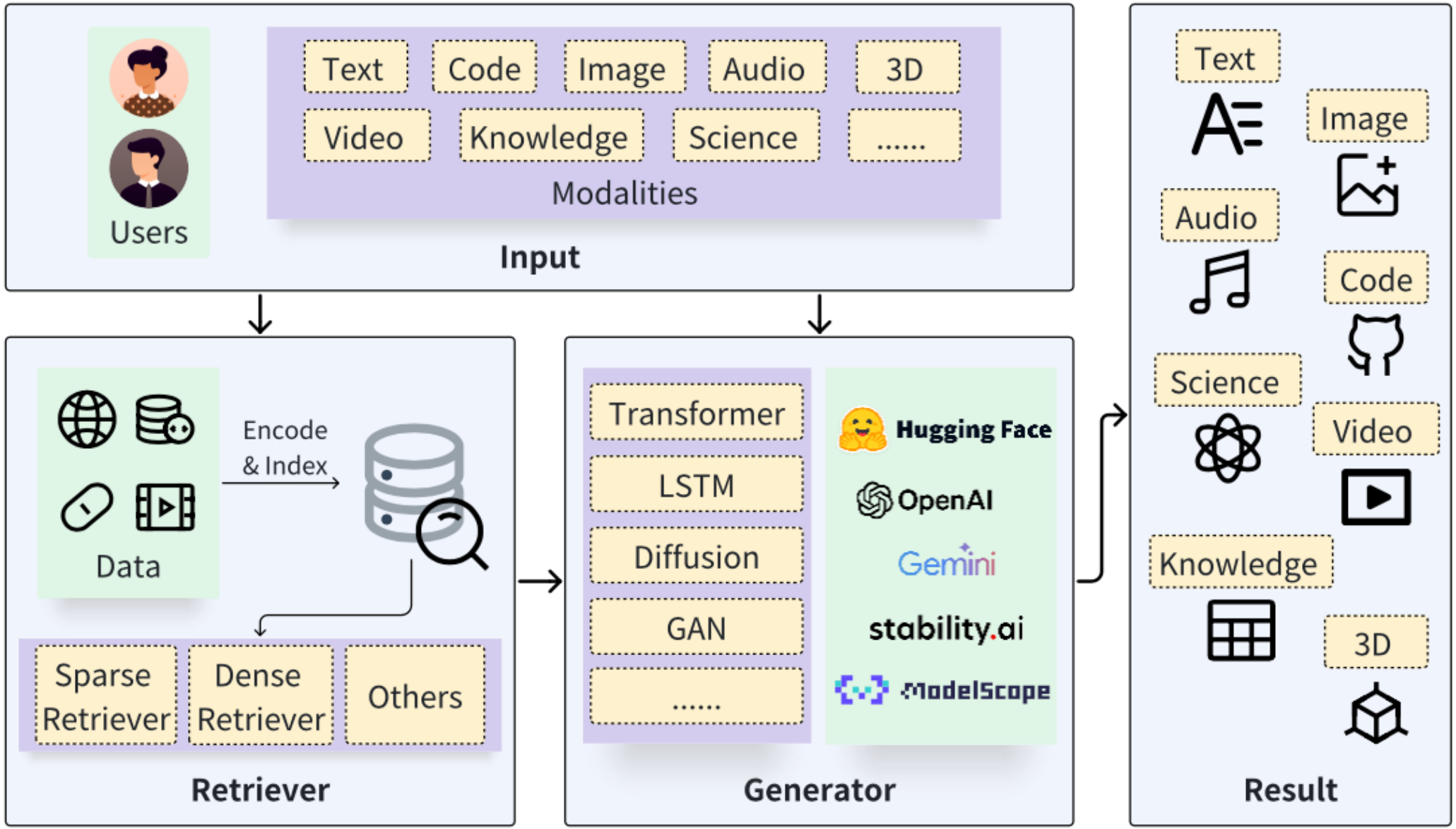
Source: Shukang Yin, Chaoyou Fu, Sirui Zhao, Ke Li, Xing Sun, Tong Xu, and Enhong Chen. (2024) "A survey on multimodal large language models." National Science Review (2024): nwae403.

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



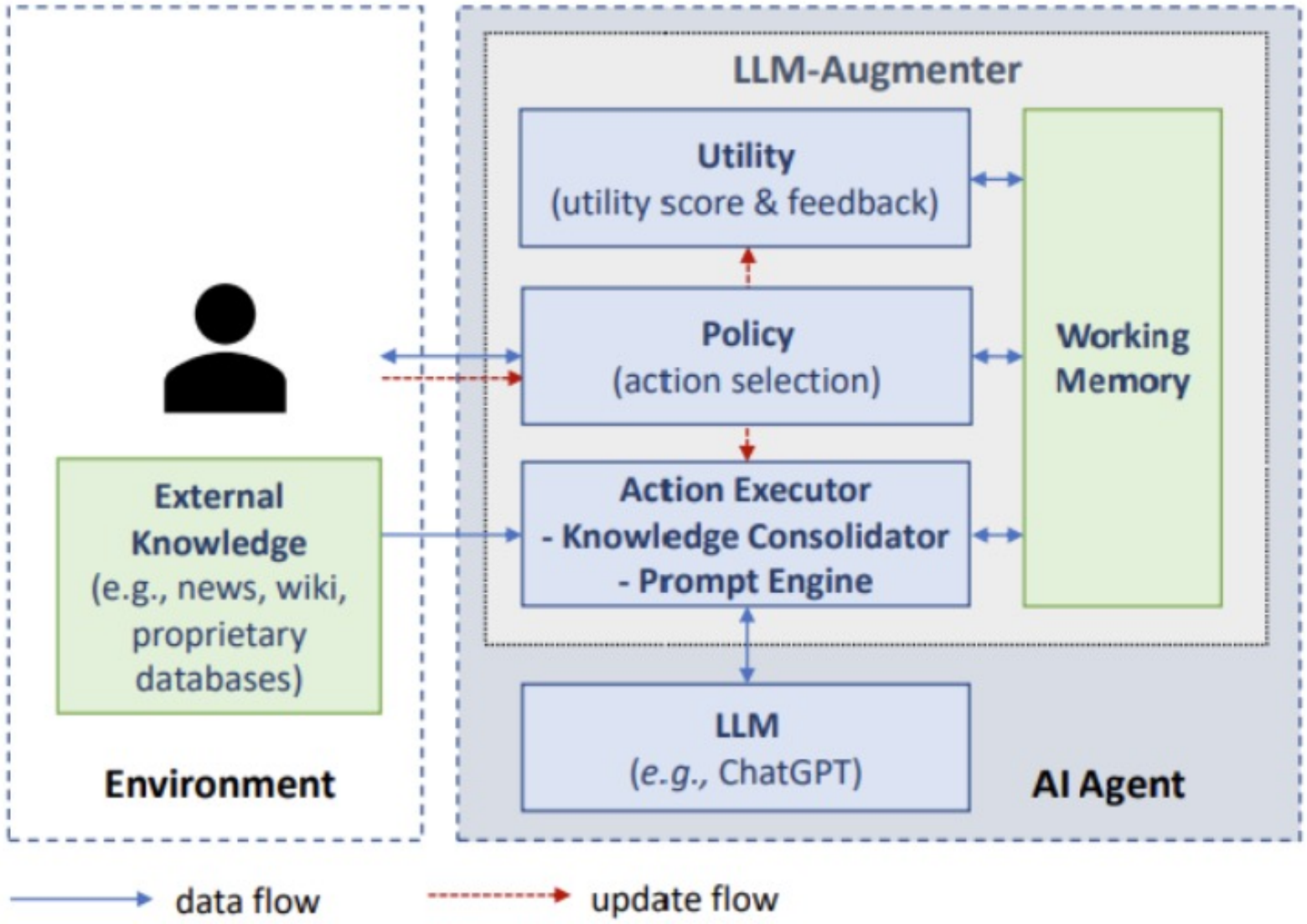
Source: Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., ... & Wang, H. (2023). Retrieval-augmented generation for large language models: A survey. arXiv preprint arXiv:2312.10997.

Retrieval-Augmented Generation (RAG) Architecture



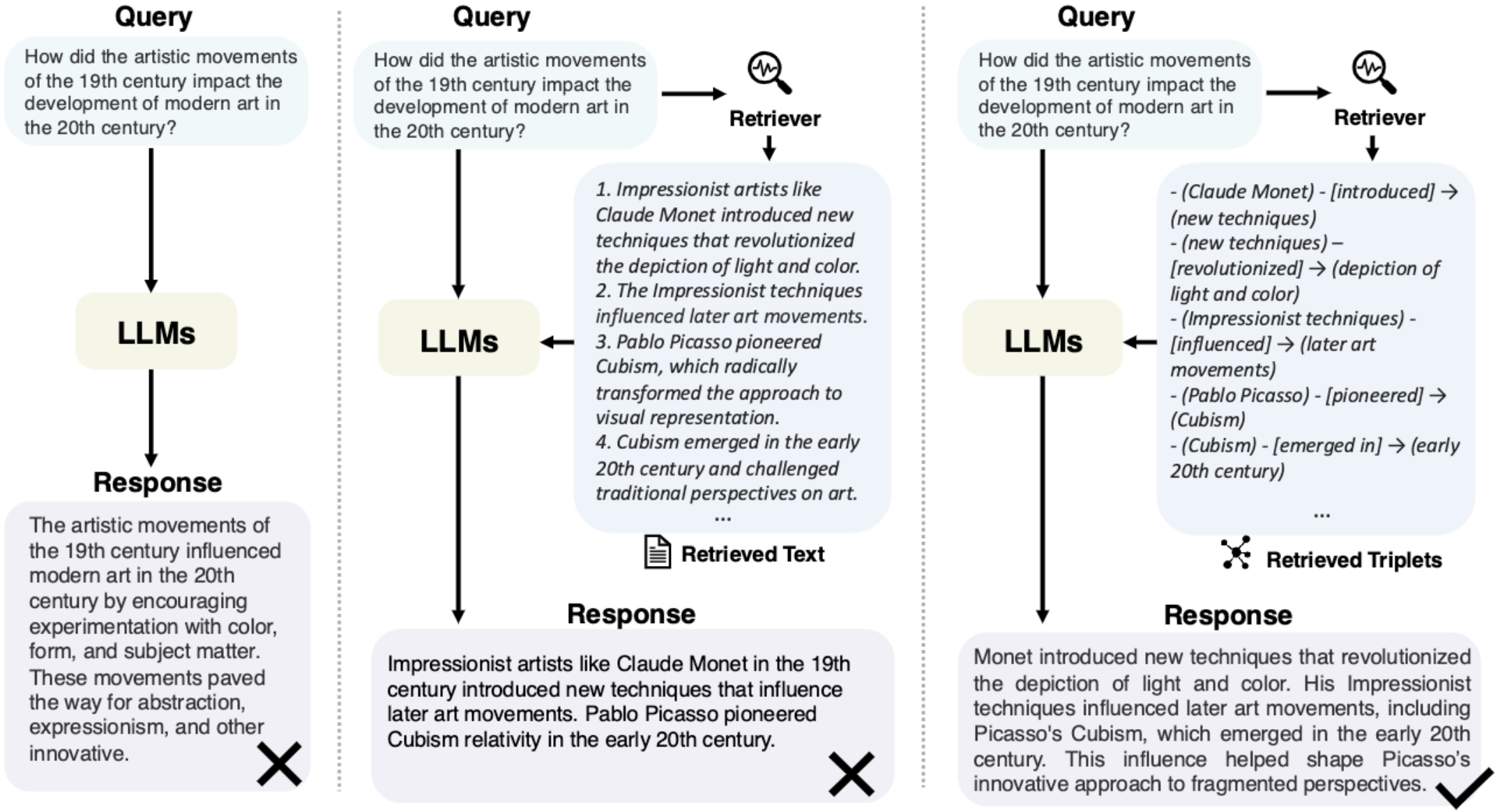
Source: Zhao, P., Zhang, H., Yu, Q., Wang, Z., Geng, Y., Fu, F., ... & Cui, B. (2024). Retrieval-augmented generation for ai-generated content: A survey. arXiv preprint arXiv:2402.19473.

A LLM-based Agent for Conversational Information Seeking



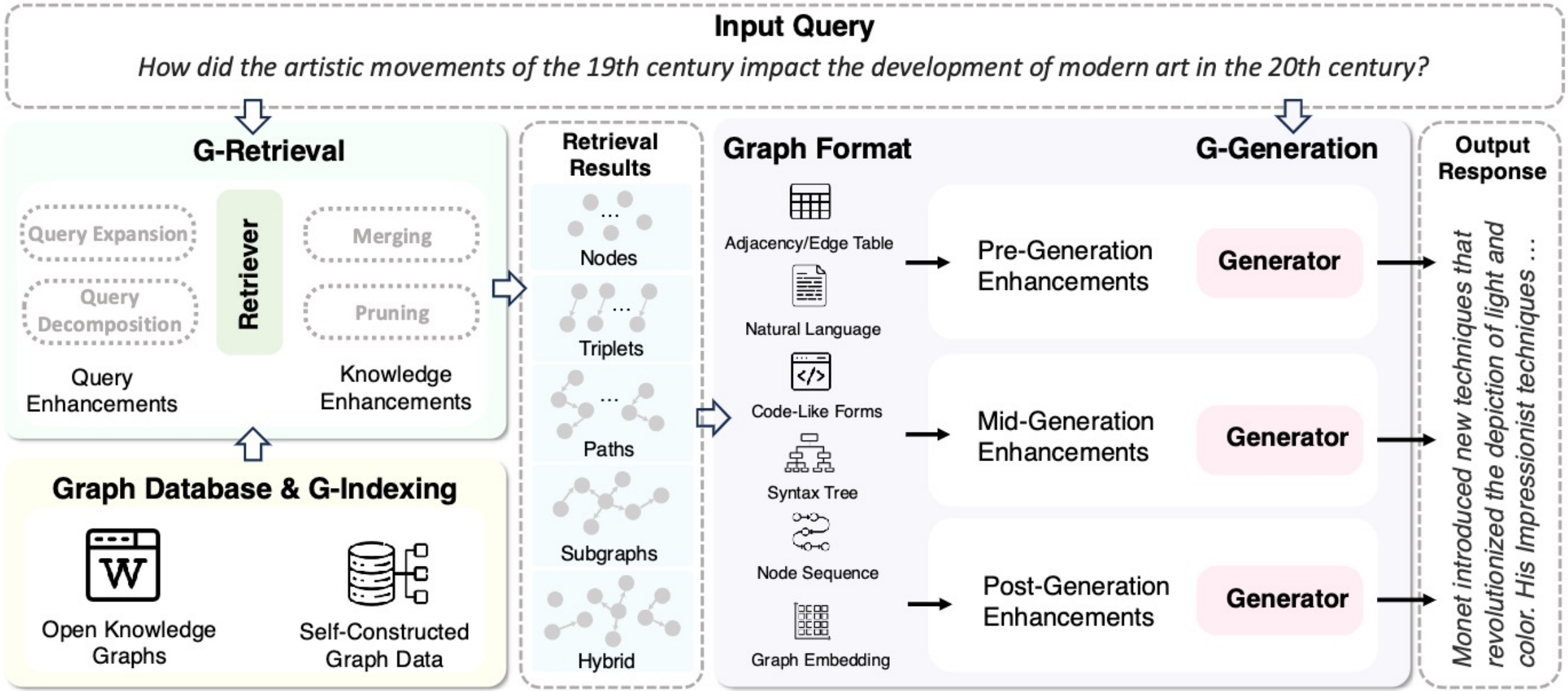
Source: Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X., & Gao, J. (2024). Large language models: A survey. arXiv preprint arXiv:2402.06196.

Direct LLM, RAG, and GraphRAG



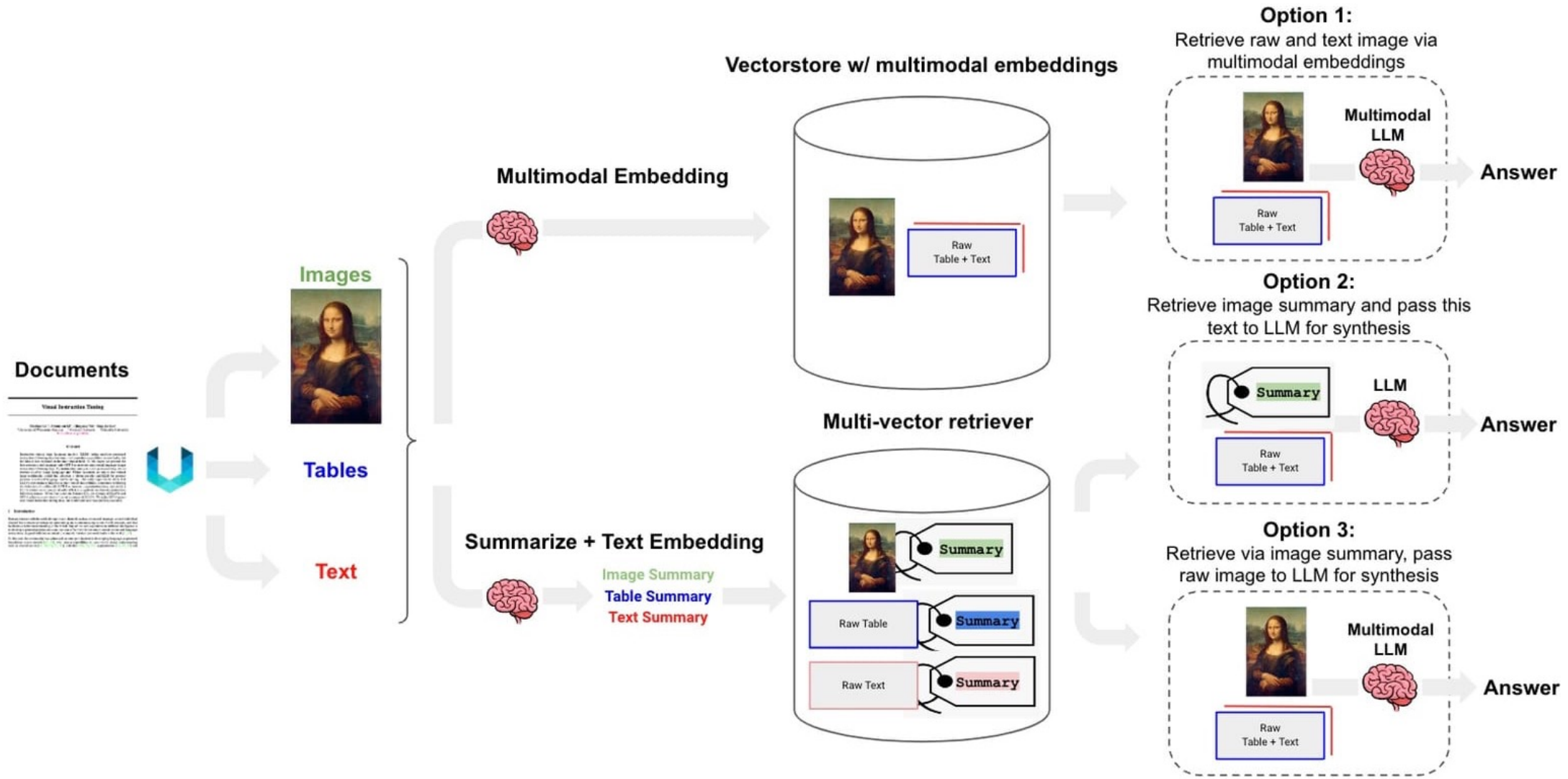
Source: Peng, B., Zhu, Y., Liu, Y., Bo, X., Shi, H., Hong, C., ... & Tang, S. (2024). Graph retrieval-augmented generation: A survey. arXiv preprint arXiv:2408.08921.

GraphRAG Framework for Question Answering



Source: Peng, B., Zhu, Y., Liu, Y., Bo, X., Shi, H., Hong, C., ... & Tang, S. (2024). Graph retrieval-augmented generation: A survey. arXiv preprint arXiv:2408.08921.

Multimodal LLM RAG
Multi-Vector Retriever for RAG



Source: https://blog.langchain.dev/deconstructing-rag/

Evaluating RAG with Ragas Metrics

ragas score

generation

faithfulness

how factually acurate 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

Source: https://blog.langchain.dev/evaluating-rag-pipelines-with-ragas-langsmith/

NVIDIA

University Ambassador

This certificate acknowledges that

Min-Yuh Day

has been certified to deliver NVIDIA instructor-led workshop for academia

Greg Estes

Vice President, NVIDIA

Issue Date: : March 7, 2025

Ambassador Certification ID: cCFh1ZWWTvqKTq7dcKkEWw

https://learn.nvidia.com/certificates?id=cCFh1ZWWTvqKTq7dcKkEWw

NVIDIA

Certified Instructor

This certificate acknowledges that

Min-Yuh Day

has been certified to deliver the instructor-led workshop

Building RAG Agents with LLMs

Greg Estes

Vice President, NVIDIA

Issue Date: : March 7, 2025

Certification ID: OVmqY4cSSya0BdMQBWHxzw

https://learn.nvidia.com/certificates?id=OVmqY4cSSya0BdMQBWHxzw

NVIDIA

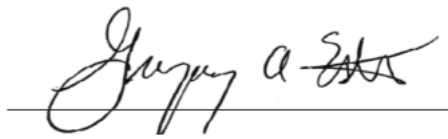
Certificate of Completion

This certificate is awarded to

Min-Yuh Day

for successfully completing

Building RAG Agents with LLMs



Greg Estes

Vice President, NVIDIA

Issue Date: : December 8, 2024

Certification ID: ed-qOCIMQaatzU8SNUNxgw |

https://learn.nvidia.com/certificates?id=ed-qOCIMQaatzU8SNUNxgw/courses/course?course_id=course-v1:DLI+S-FX-15+V1

https://learn.nvidia.com/certificates?id=ed-qOCIMQaatzU8SNUNxgw

NVIDIA

Certificate of Competency

This certificate is awarded to

Min-Yuh Day

for demonstrating competence in the completion of

Generative AI with Diffusion Models

Greg Estes

Vice President, NVIDIA

Issue Date: : February 28, 2025

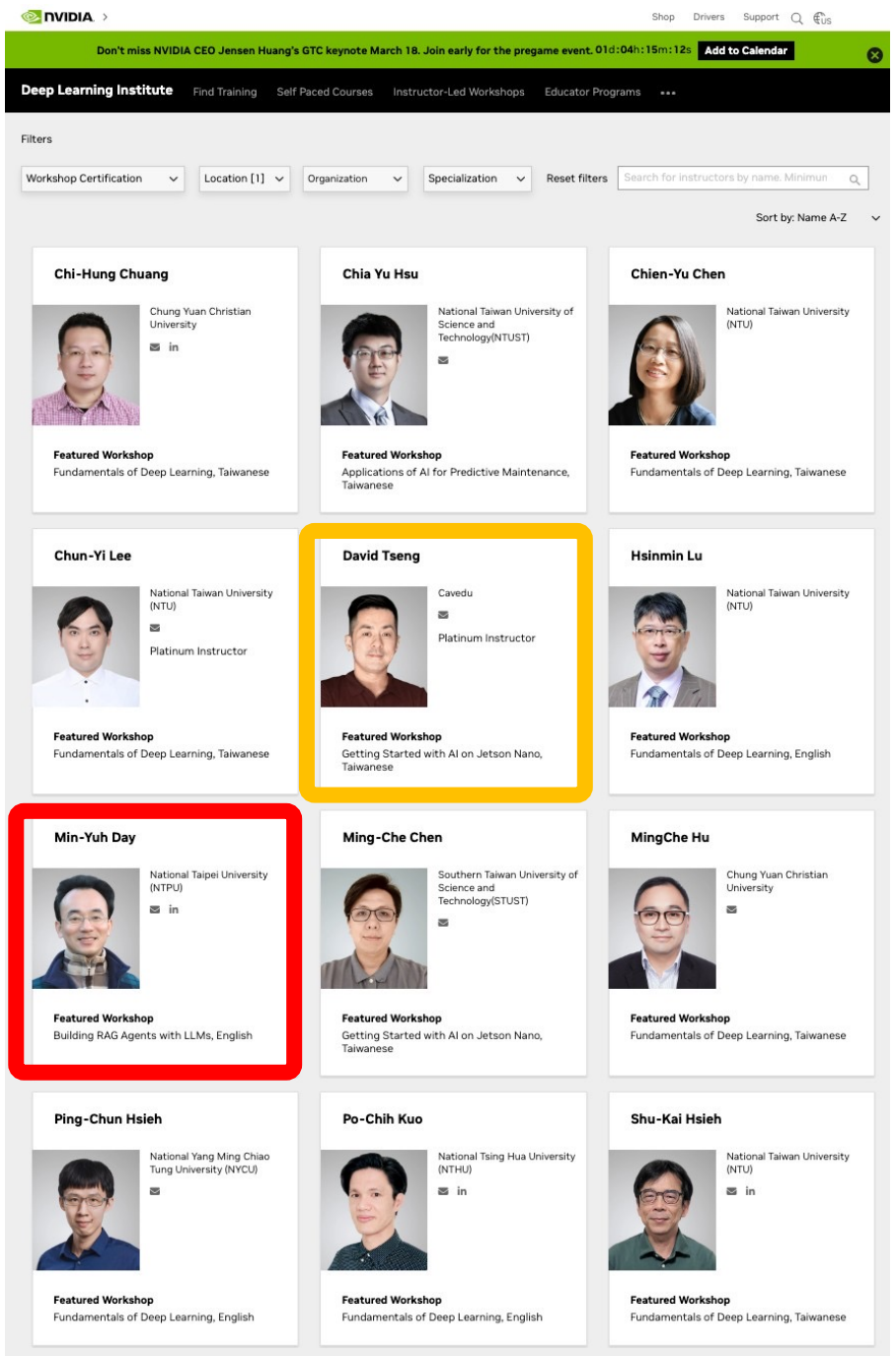
Certification ID: q3oo-oBhTQKtyCCote2E-Q

https://learn.nvidia.com/certificates?id=q3oo-oBhTQKtyCCote2E-Q

NVIDIA Certified Instructors

(12 from Taiwan)

(March 2025)

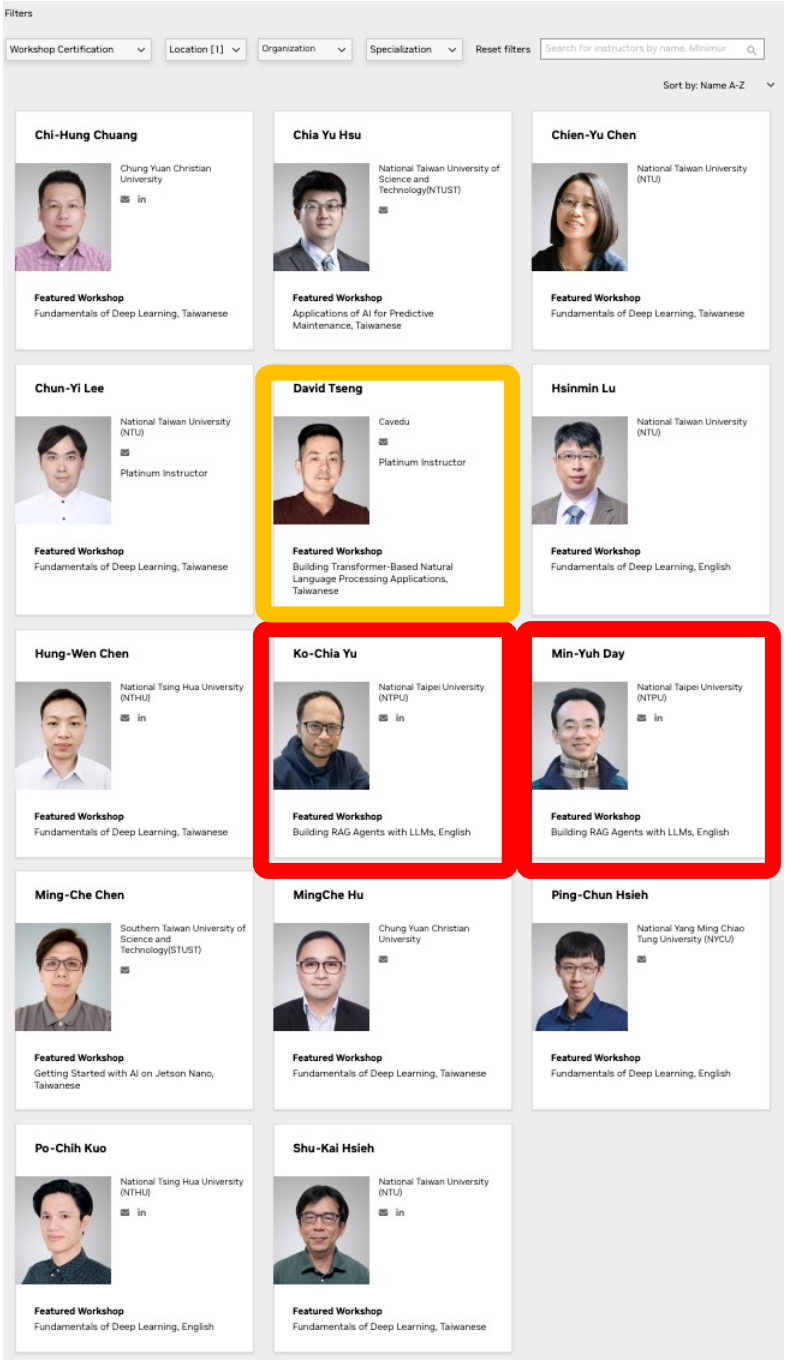


https://www.nvidia.com/en-us/training/instructor-directory/search/

NVIDIA Certified Instructors

(14 from Taiwan)

(2 from NTPU)(April 2025)



https://www.nvidia.com/en-us/training/instructor-directory/search/

NVIDIA Certified Instructor

Min-Yuh Day

Dr. Min-Yuh Day is a Professor in the Graduate Institute of Information Management at National Taipei University, Taiwan. He holds a Ph.D. degree in Information Management from National Taiwan University, Taiwan. His research focuses on AI, generative AI, ESG green fintech, big data, e-commerce, and biomedical informatics, with publications in top journals and conferences.

Organization

National Taipei University (NTPU)

Location

Taiwan

Instructor Tier

Silver

Workshop Certifications

- Building RAG Agents with LLMs, English
- Generative AI with Diffusion Models, English
- Fundamentals of Deep Learning, English
- Building Agentic AI Applications with LLMs, English

https://www.nvidia.com/en-us/training/instructor-directory/search/

NVIDIA Developer Program

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

NVIDIA Deep Learning Institute (DLI)

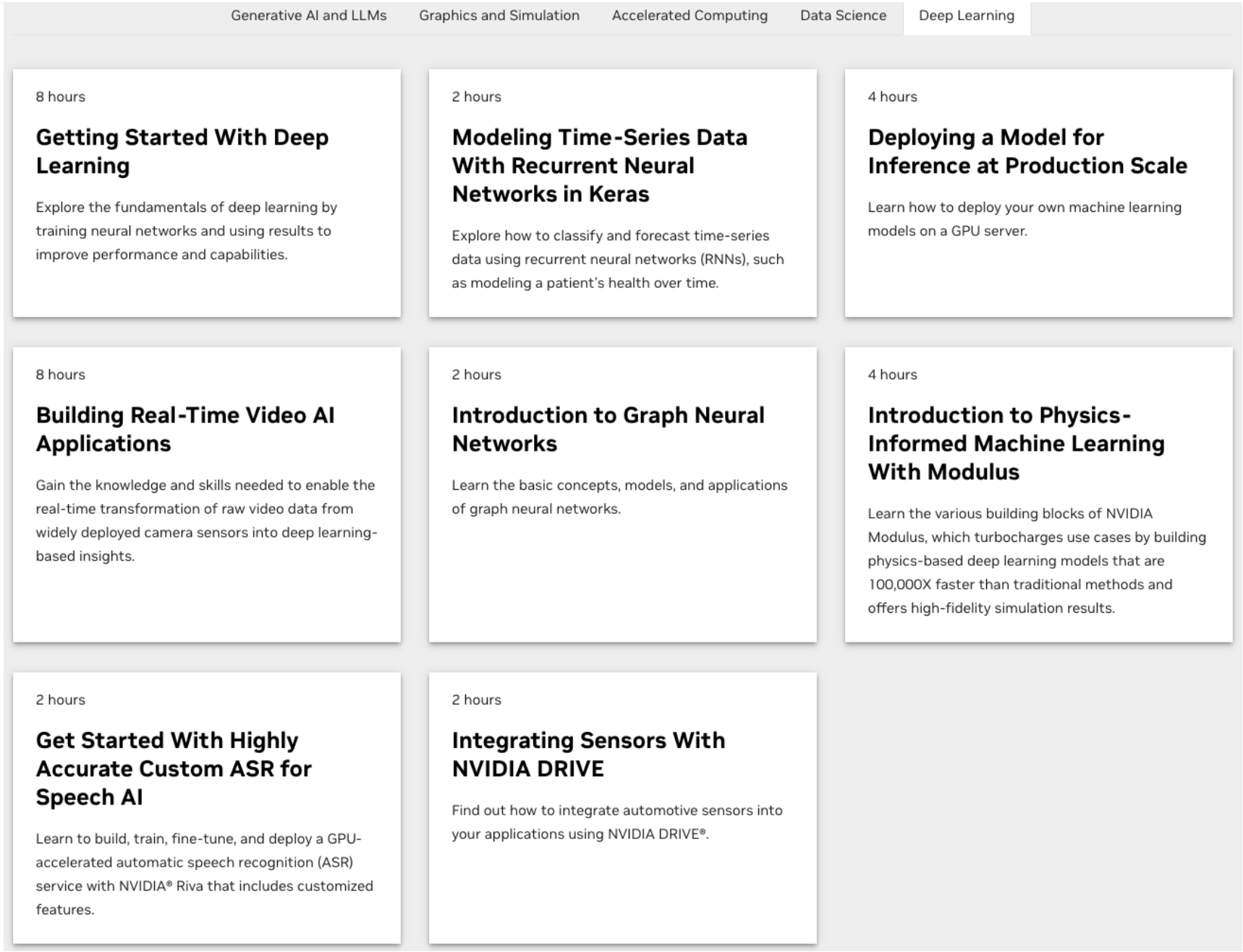
https://learn.nvidia.com/

Get NVIDIA DLI Certificate Before the NVIDIA Workshop

- **Step 1. Join NVIDIA Developer Program (Free)**
  https://developer.nvidia.com/join-nvidia-developer-program
- **Step 2. Visit NVIDIA Deep Learning Institute (DLI)**
  https://learn.nvidia.com/
- **Step 3. Enroll "Building RAG Agents with LLMs" Self-Paced Course (Free)**
  https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-15+V1

Join the NVIDIA Developer Program

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



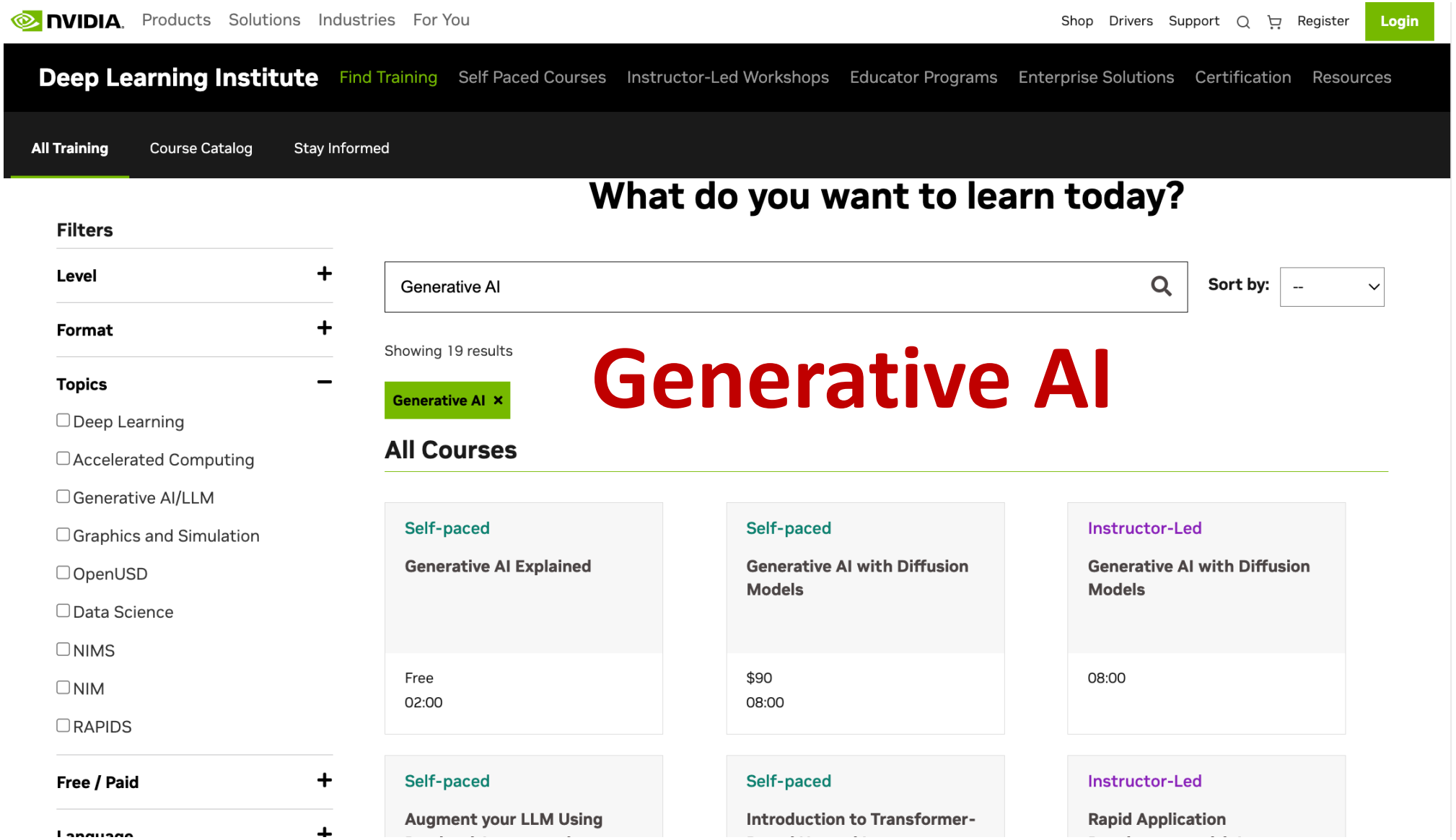
https://developer.nvidia.com/join-nvidia-developer-program

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

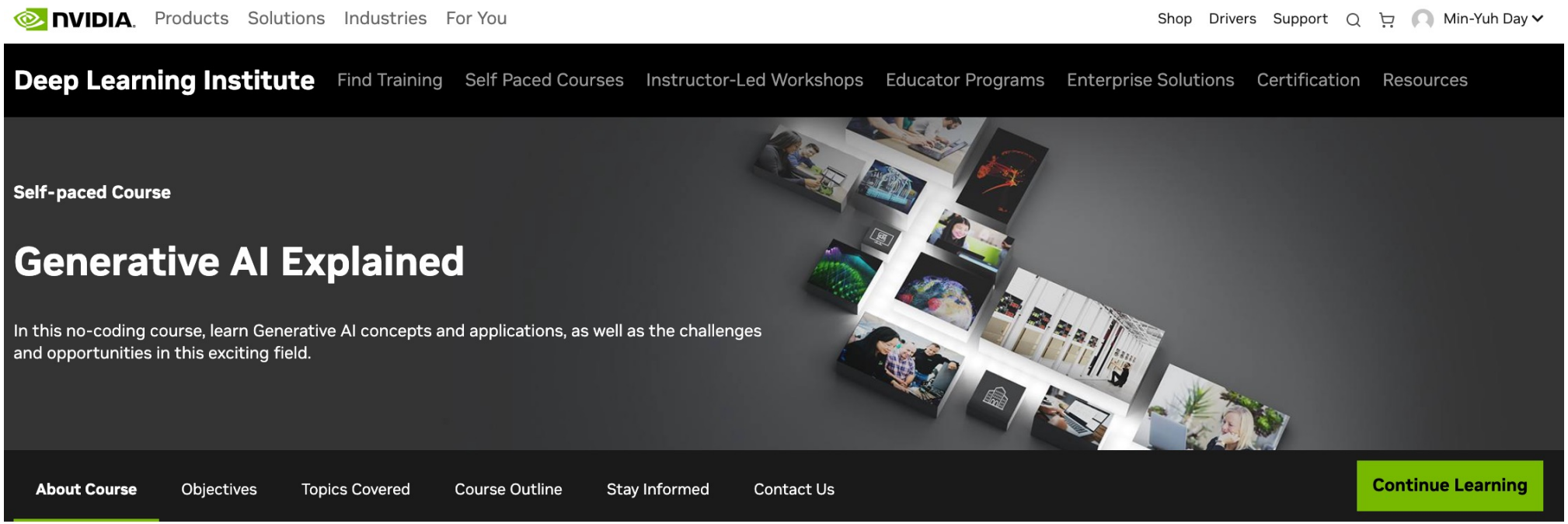
https://www.nvidia.com/en-us/learn/learning-path/generative-ai-llm/

NVIDIA Deep Learning Institute (DLI)



https://learn.nvidia.com/en-us/training/find-training?q=Generative+AI

Generative AI Explained



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

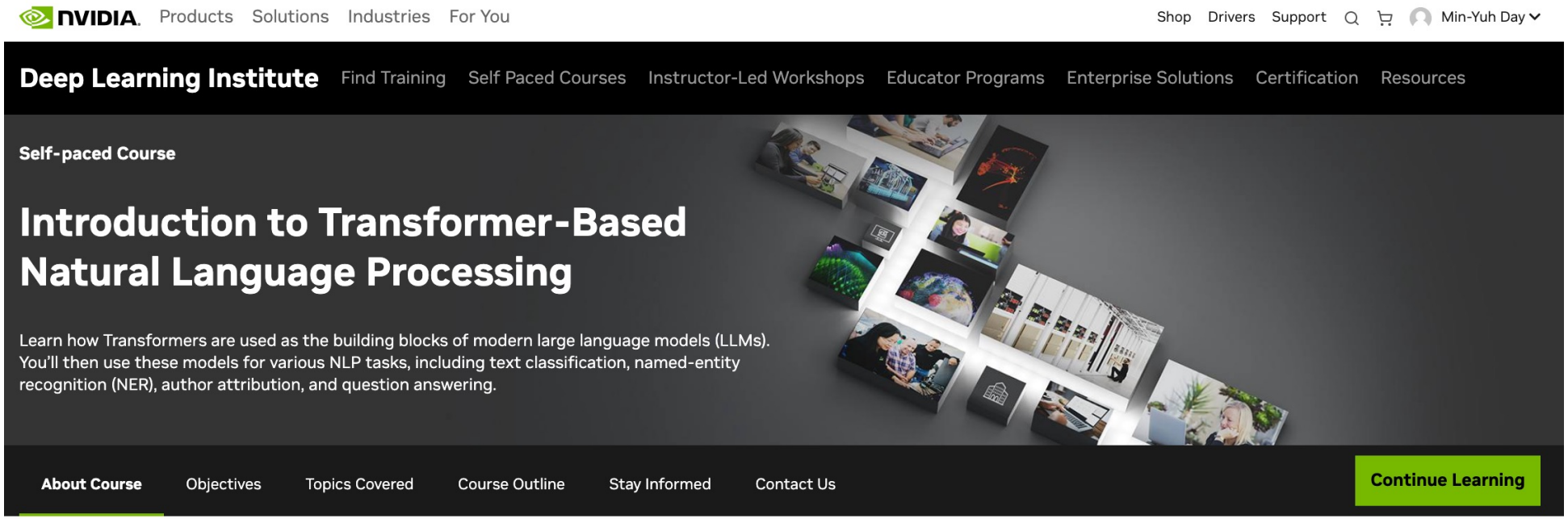
Level: Technical - Beginner

Subject: Generative AI/LLM

Language: English

https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-15+V1

Introduction to Transformer-Based Natural Language Processing



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

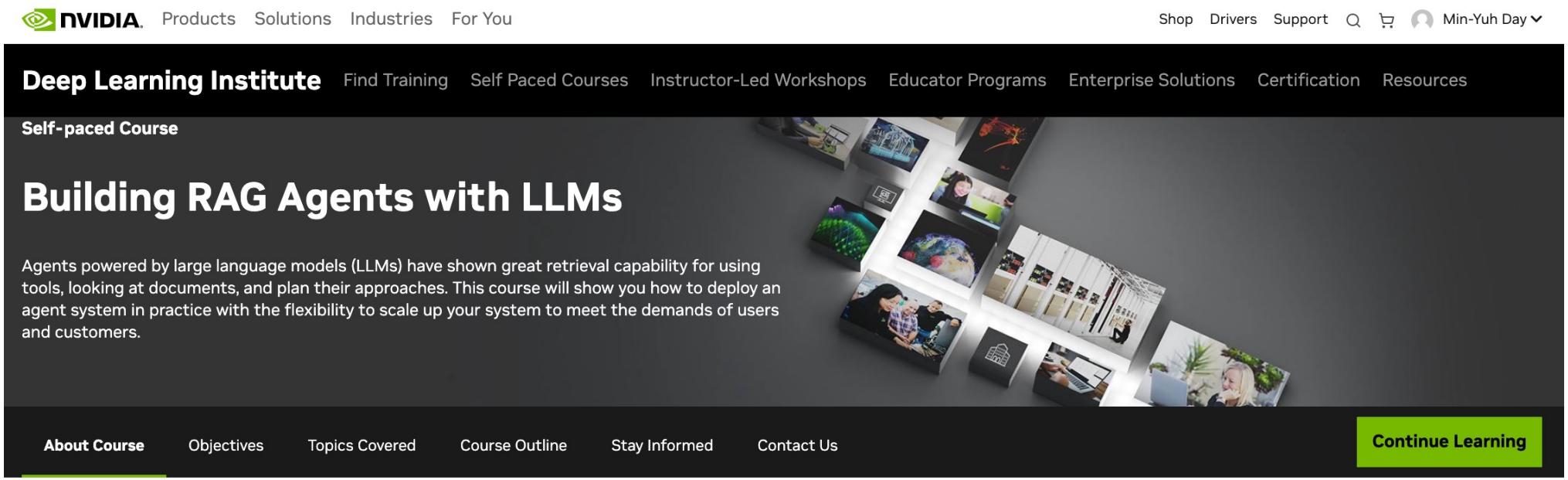
Level: Technical - Beginner

Subject: Generative AI/LLM

Language: English

https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-08+V1

Building RAG Agents with LLMs



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

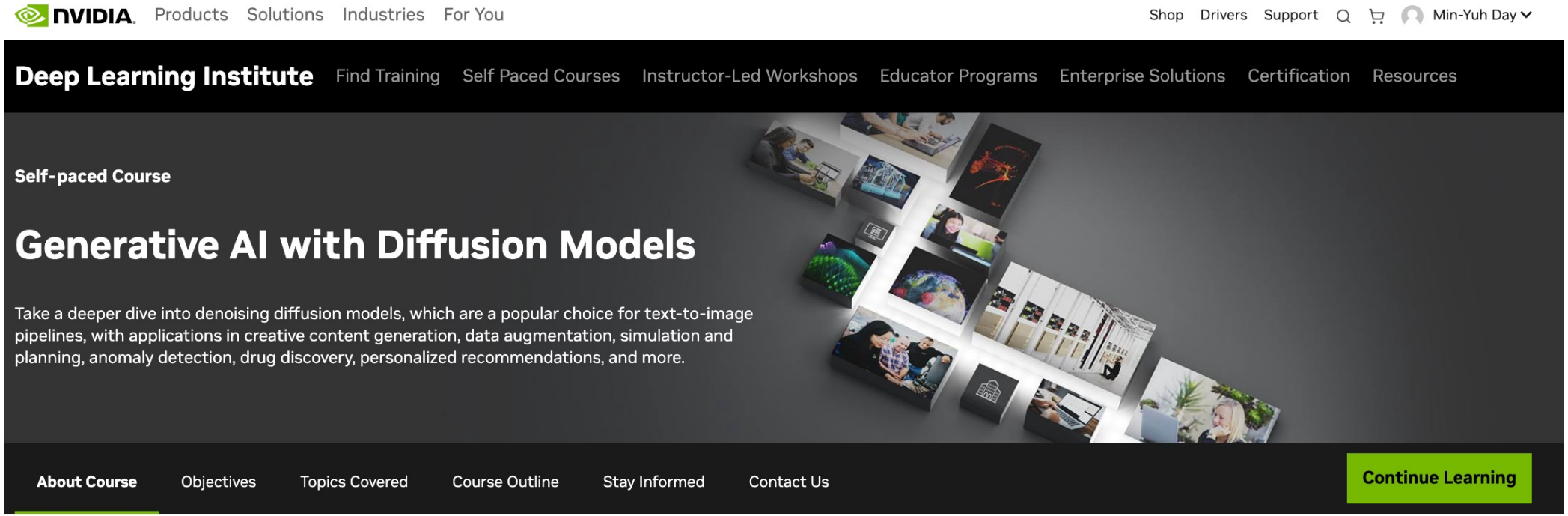
Subject: Generative AI/LLM

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

Generative AI with Diffusion Models



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.

Course Details

Duration: 08:00

Price: $90

Subject: Generative AI/LLM

Language: English

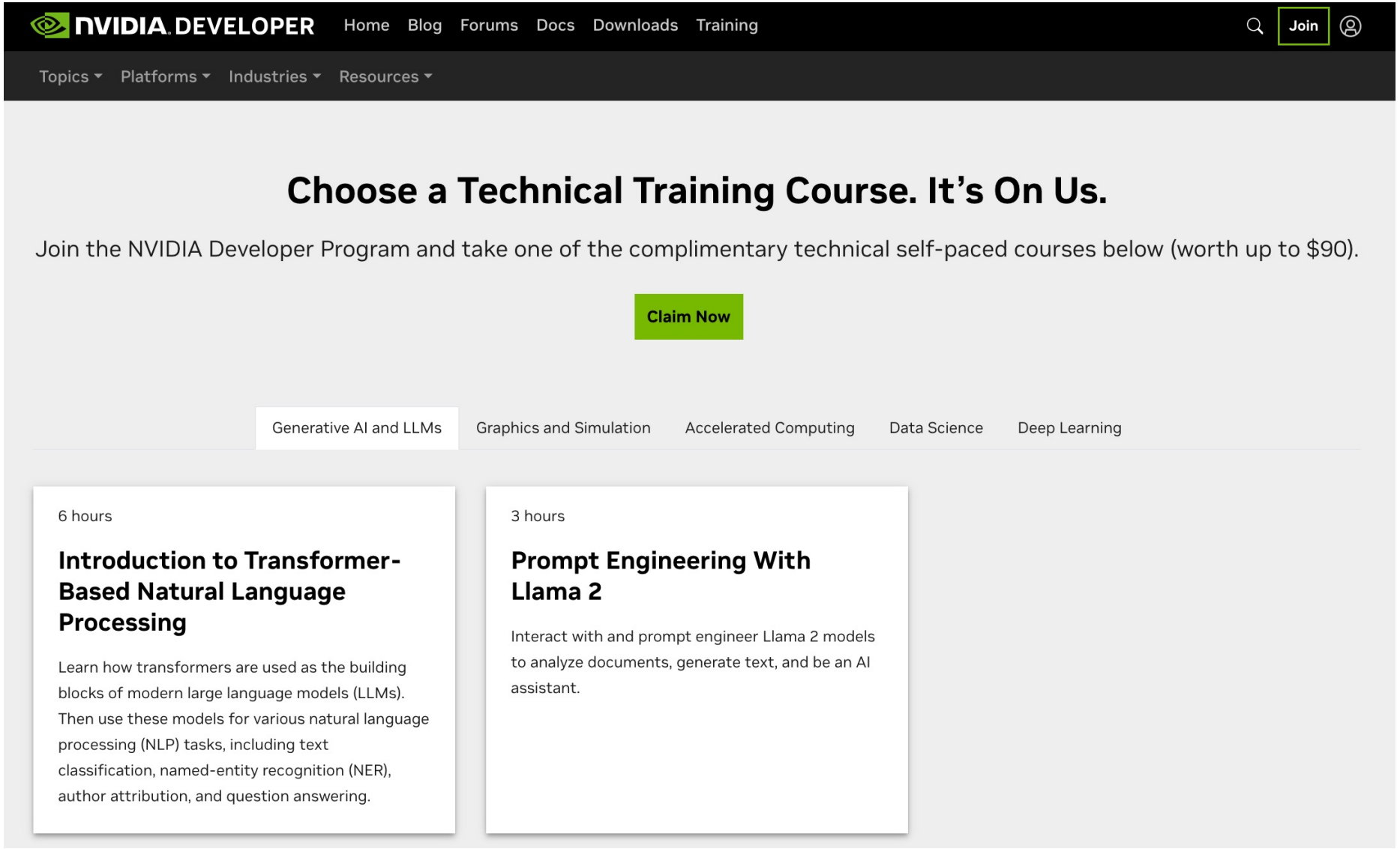
Course Prerequisites:
A basic understanding of Deep Learning Concepts.

Learning Objectives

https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Join the NVIDIA Developer Program

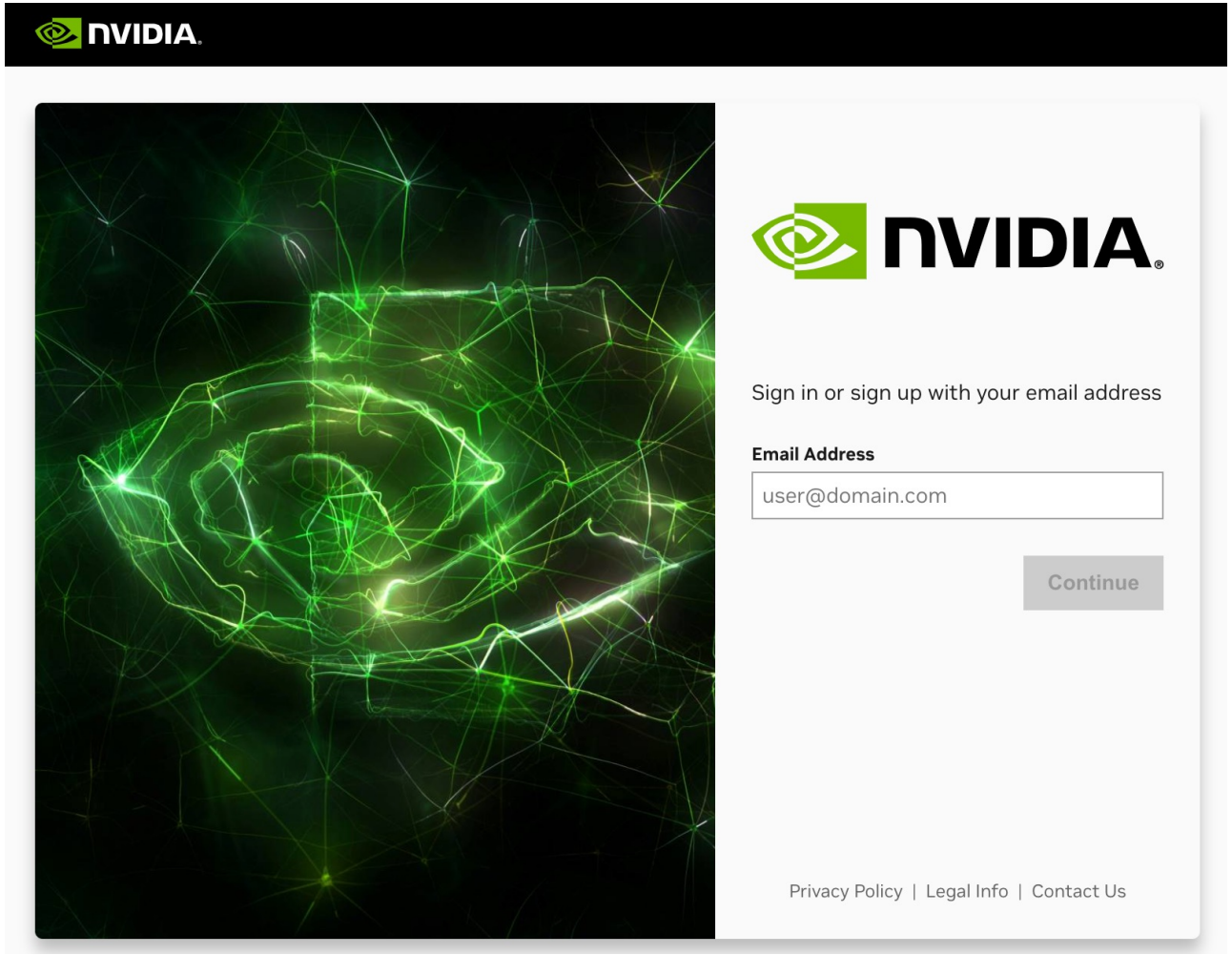
take one of the complimentary technical self-paced courses (worth up to $90)



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

Join the NVIDIA Developer Program

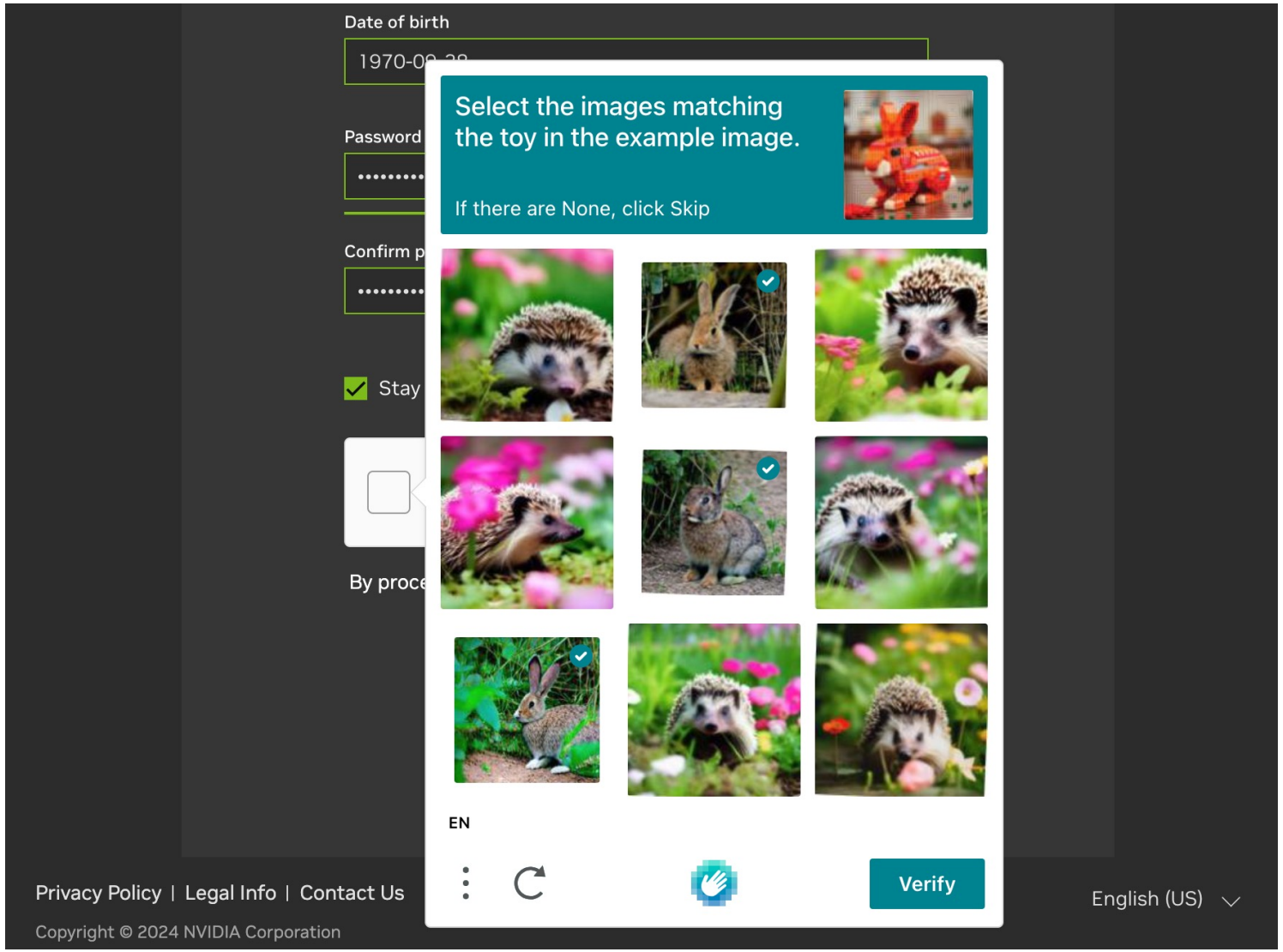
take one of the complimentary technical self-paced courses (worth up to $90)



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

Join the NVIDIA Developer Program

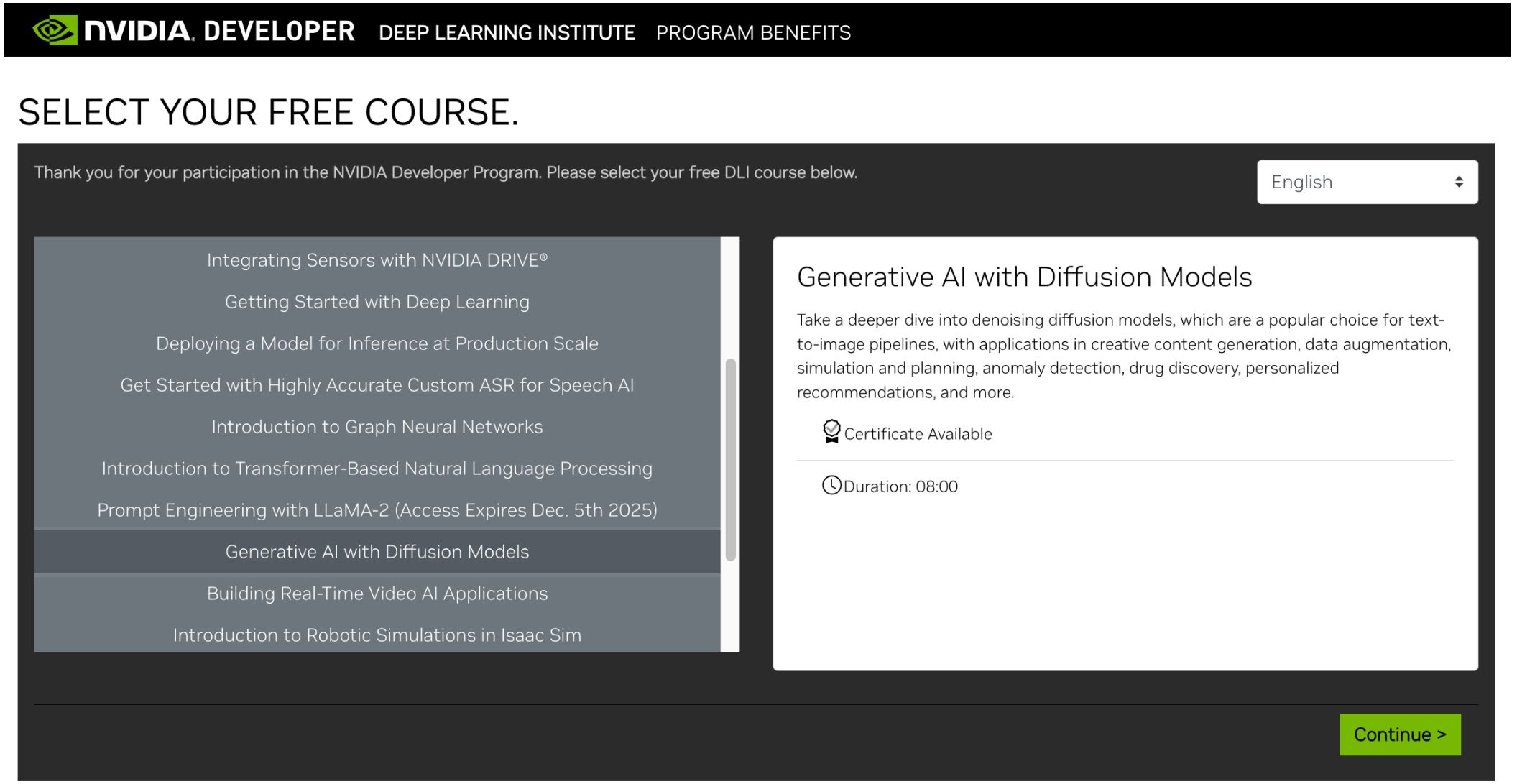
take one of the complimentary technical self-paced courses (worth up to $90)



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

Join the NVIDIA Developer Program

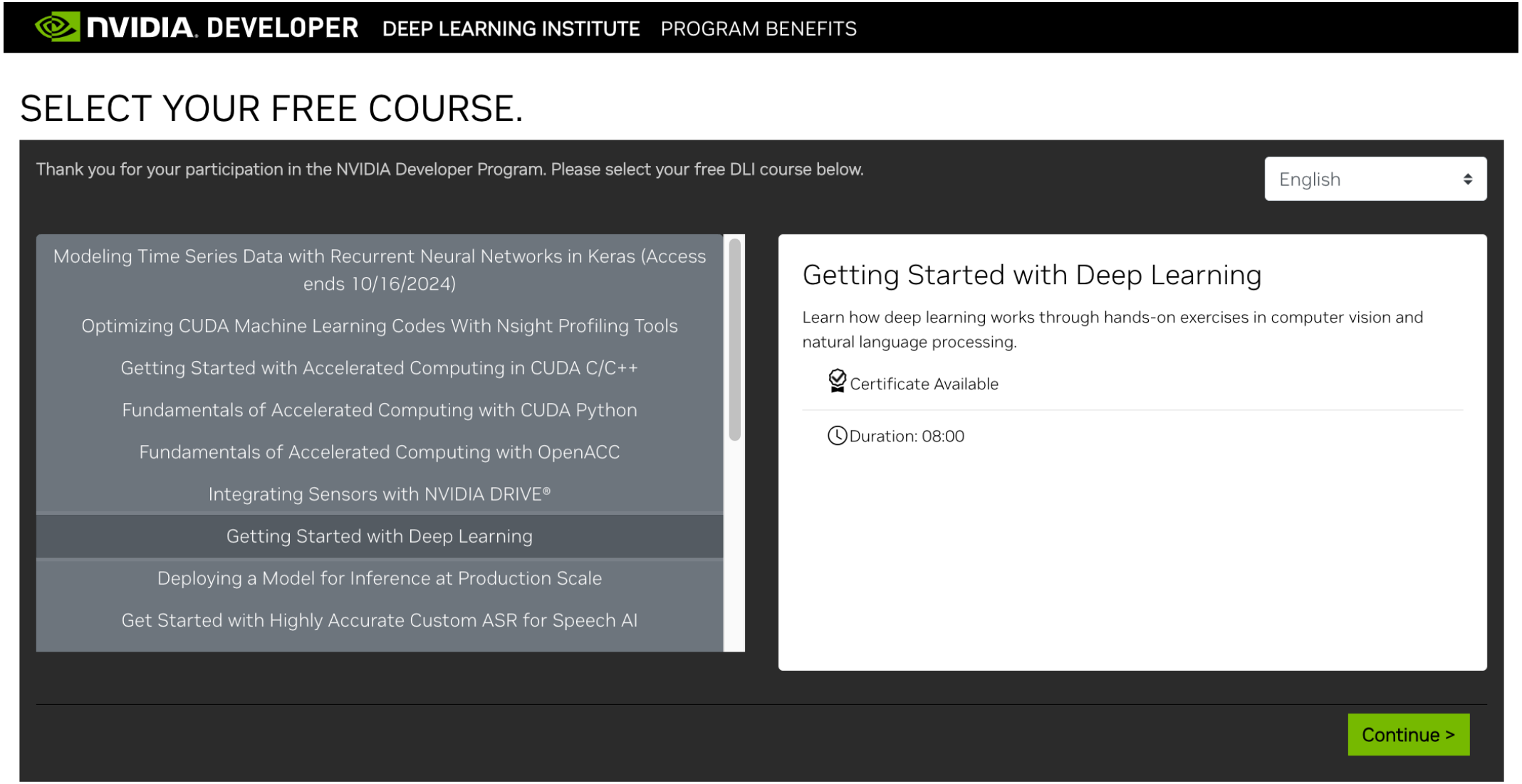
take one of the complimentary technical self-paced courses (worth up to $90)



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

Join the NVIDIA Developer Program

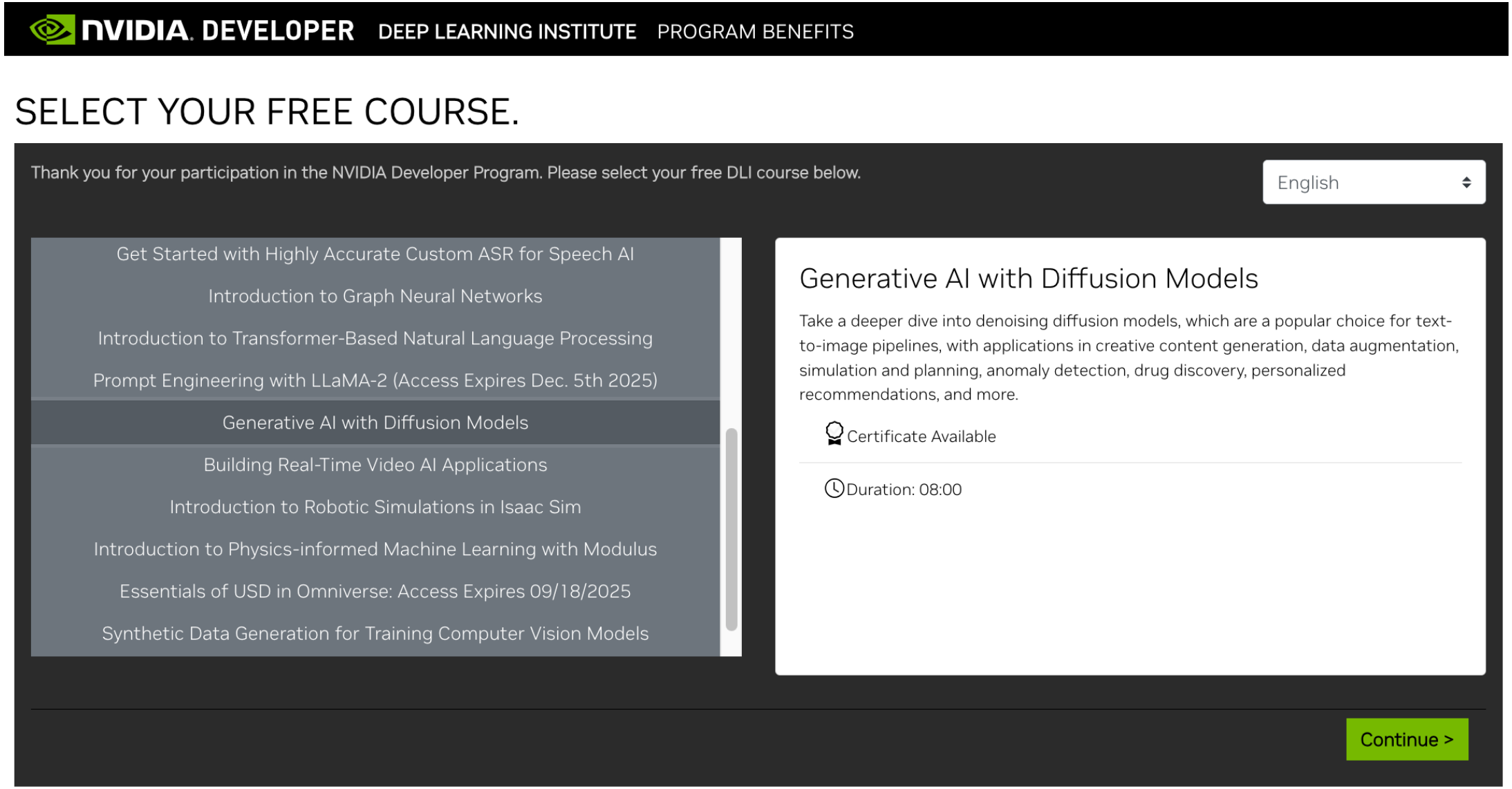
take one of the complimentary technical self-paced courses (worth up to $90)



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

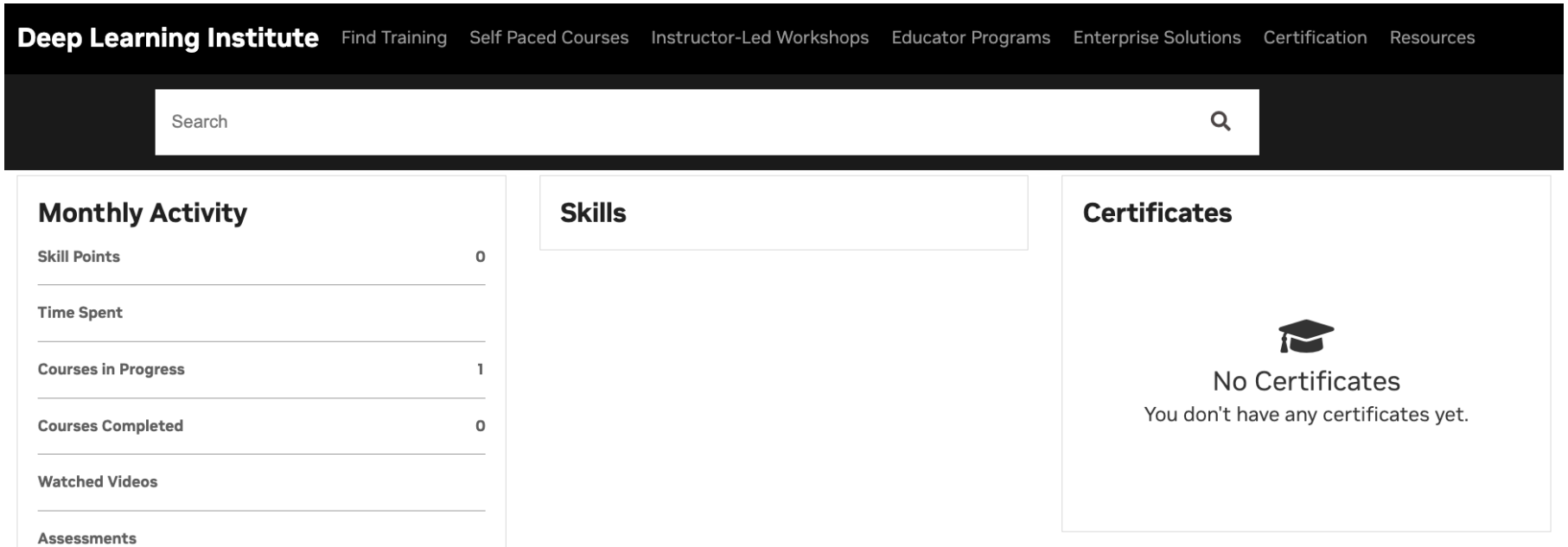
Join the NVIDIA Developer Program

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



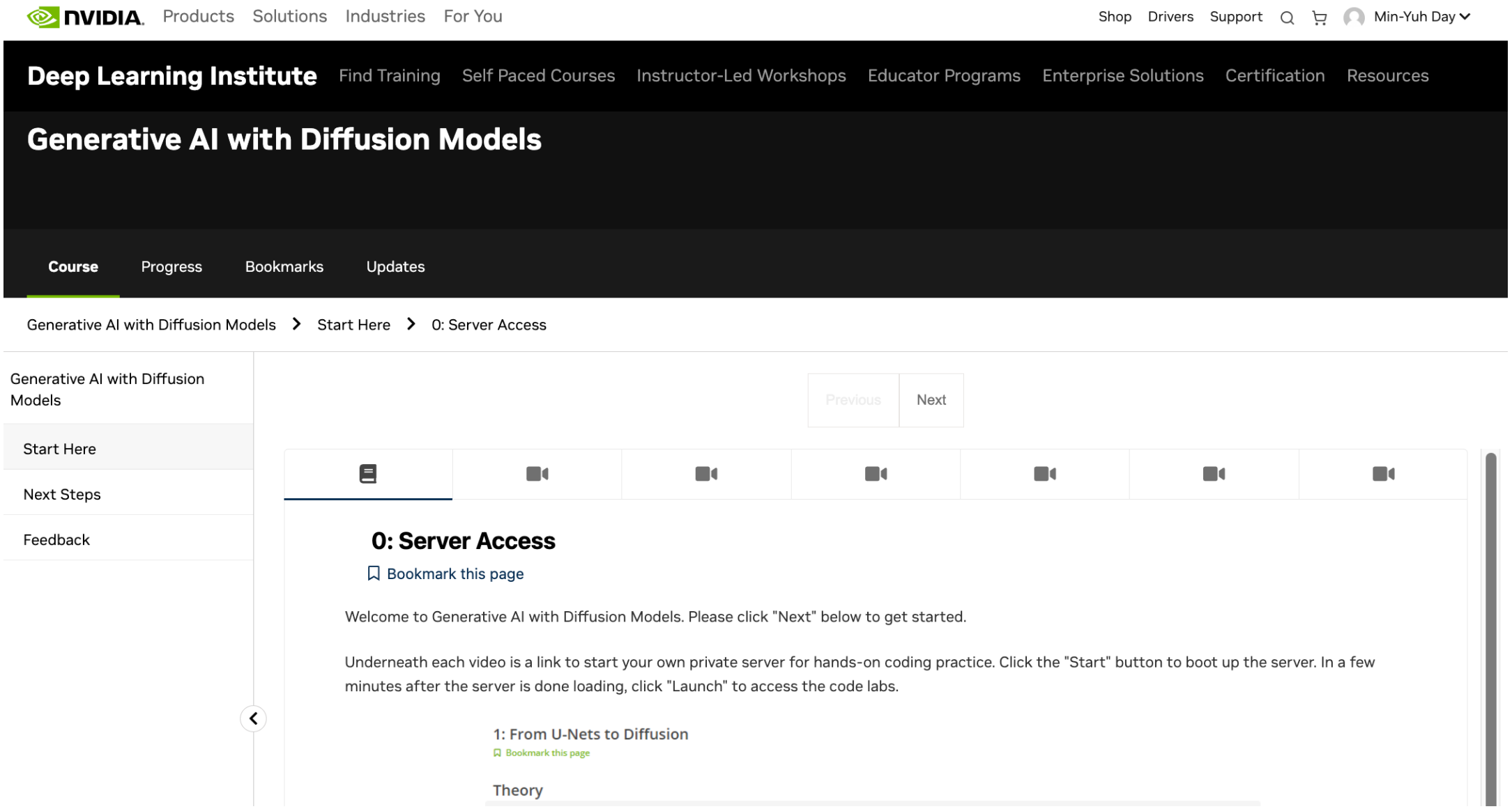
https://developer.nvidia.com/join-nvidia-developer-program

NVIDIA Deep Learning Institute (DLI)



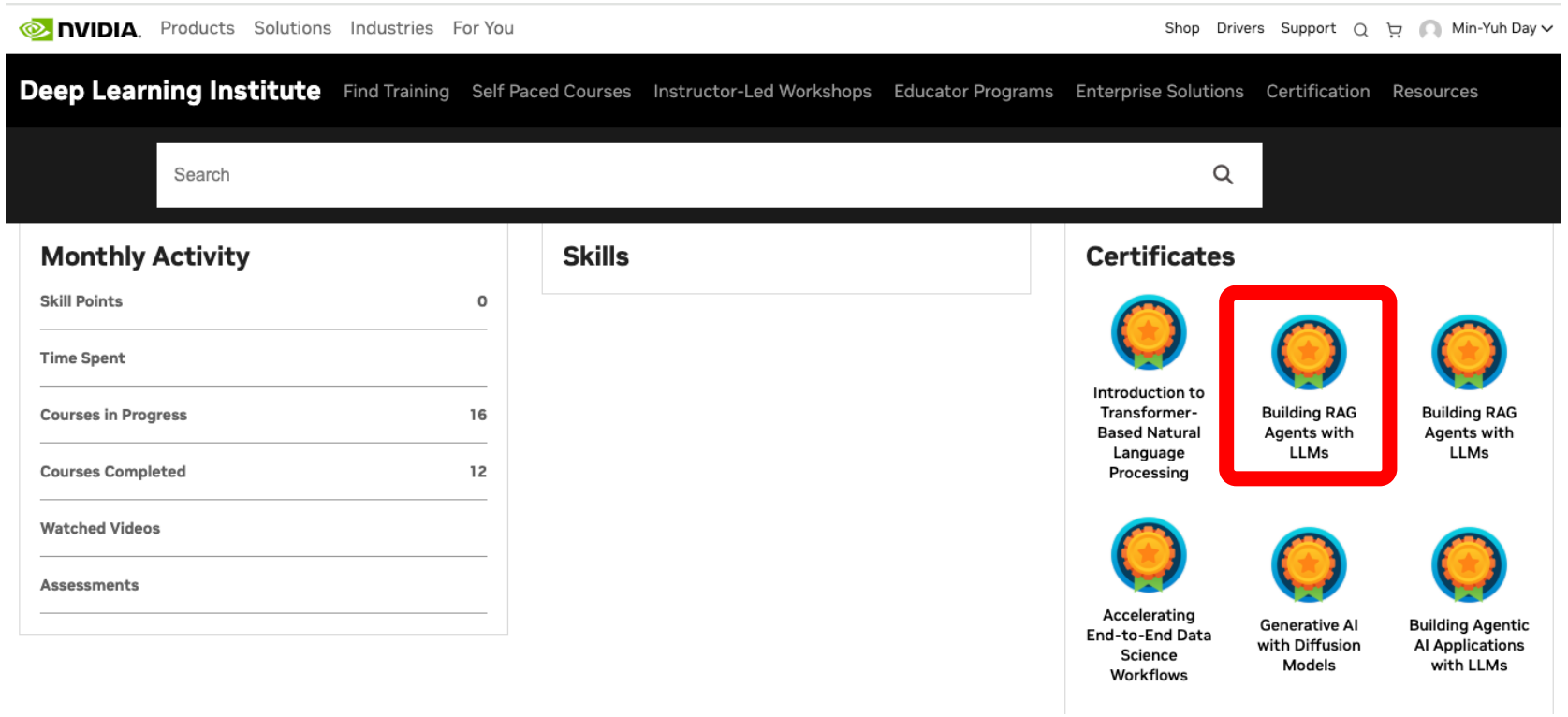
https://learn.nvidia.com/my-learning

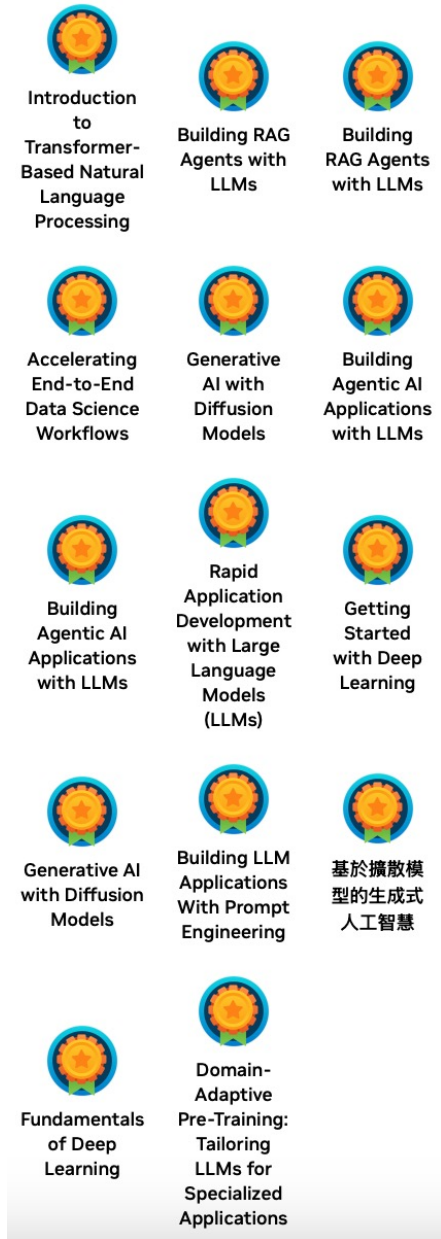
NVIDIA Deep Learning Institute (DLI)



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

Deep Learning Institute (DLI)





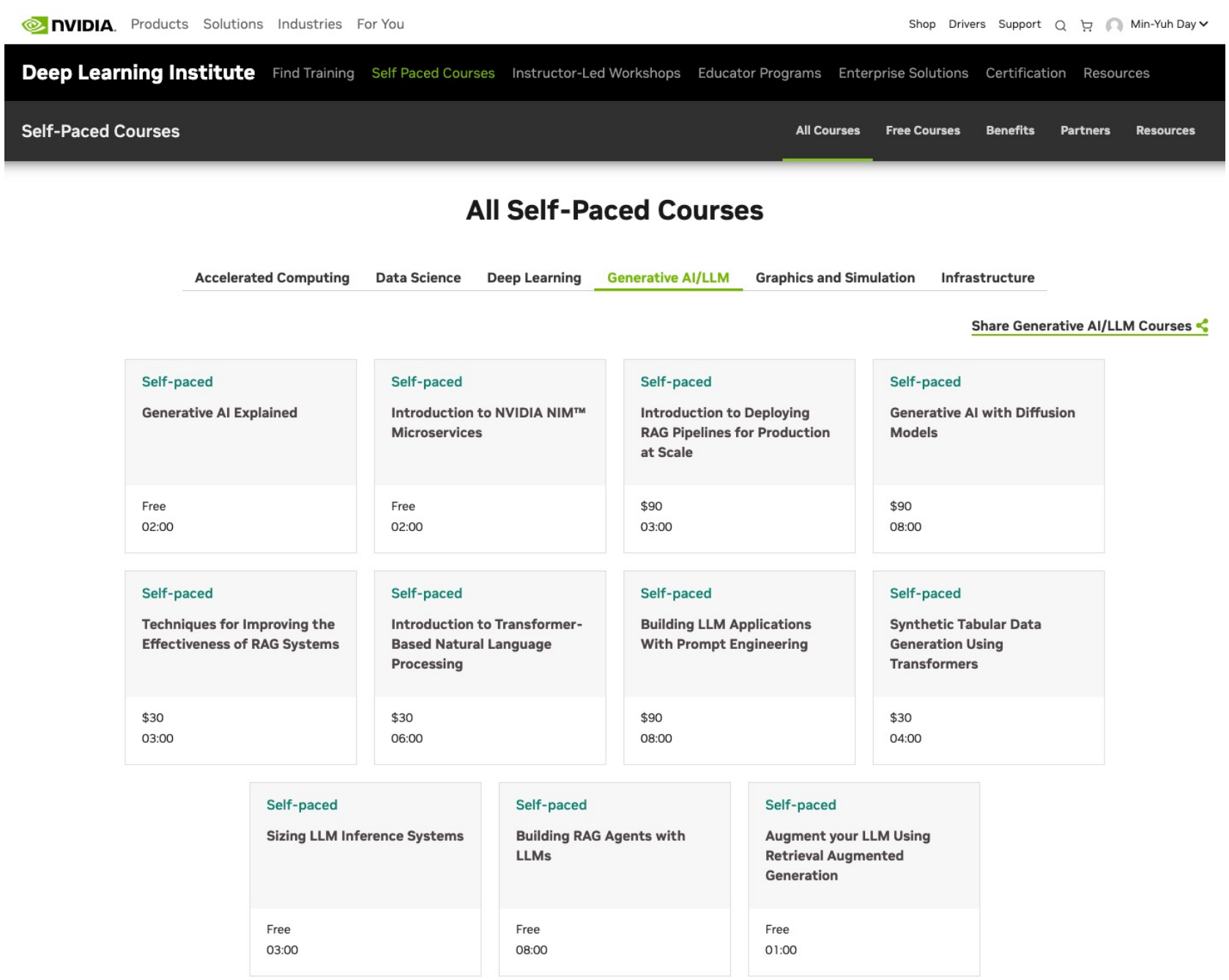
Completed Courses

View more

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

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

NVIDIA Deep Learning Institute (DLI)



https://learn.nvidia.com/en-us/training/self-paced-courses

NVIDIA

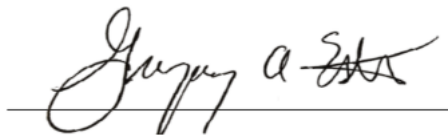
Certificate of Completion

This certificate is awarded to

Min-Yuh Day

for successfully completing

Building RAG Agents with LLMs



Greg Estes

Vice President, NVIDIA

Issue Date: : December 8, 2024

Certification ID: ed-qOCIMQaatzU8SNUNxgw |

https://learn.nvidia.com/certificates?id=ed-qOCIMQaatzU8SNUNxgw/courses/course?course_id=course-v1:DLI+S-FX-15+V1

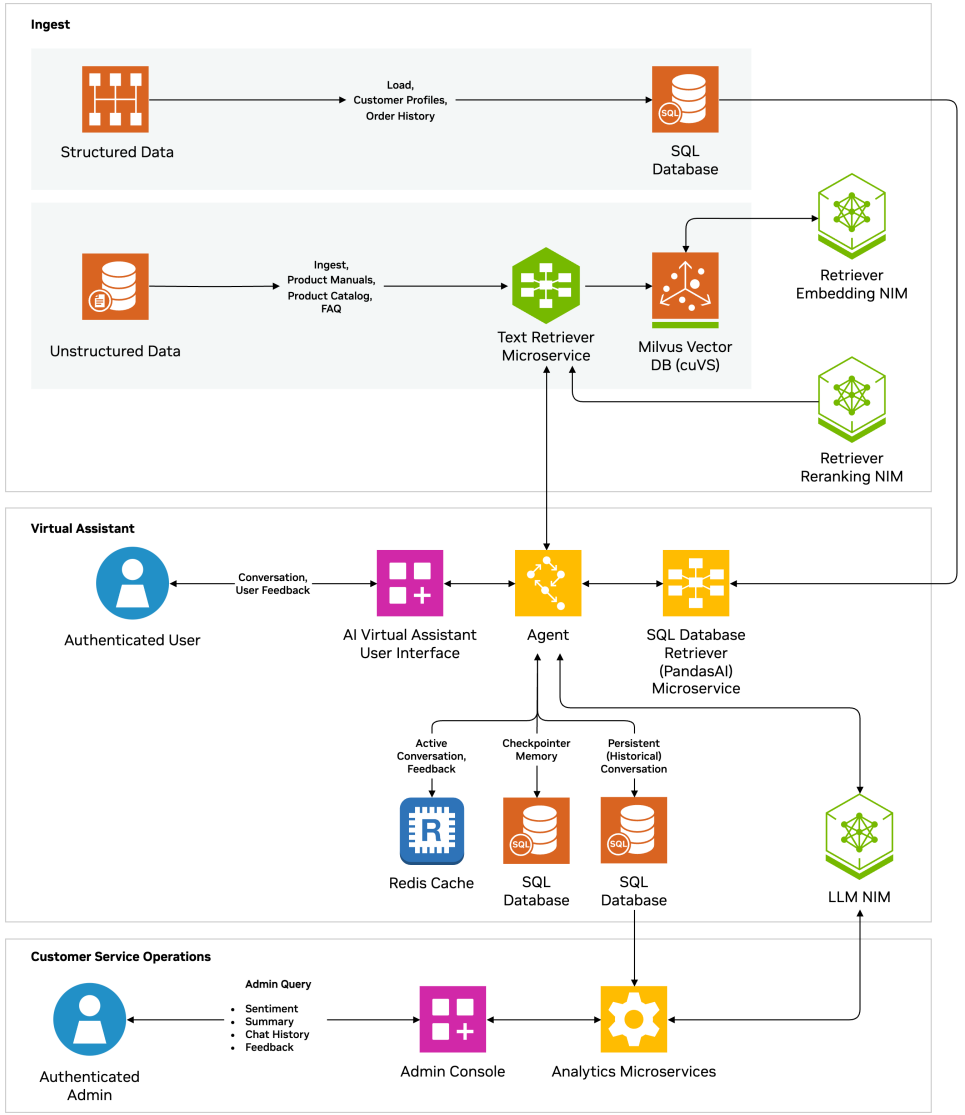
https://learn.nvidia.com/certificates?id=ed-qOCIMQaatzU8SNUNxgw

AI Virtual Assistant for Customer Service

AI Virtual Assistant to reduce handling time, boost customer satisfaction

Benefits

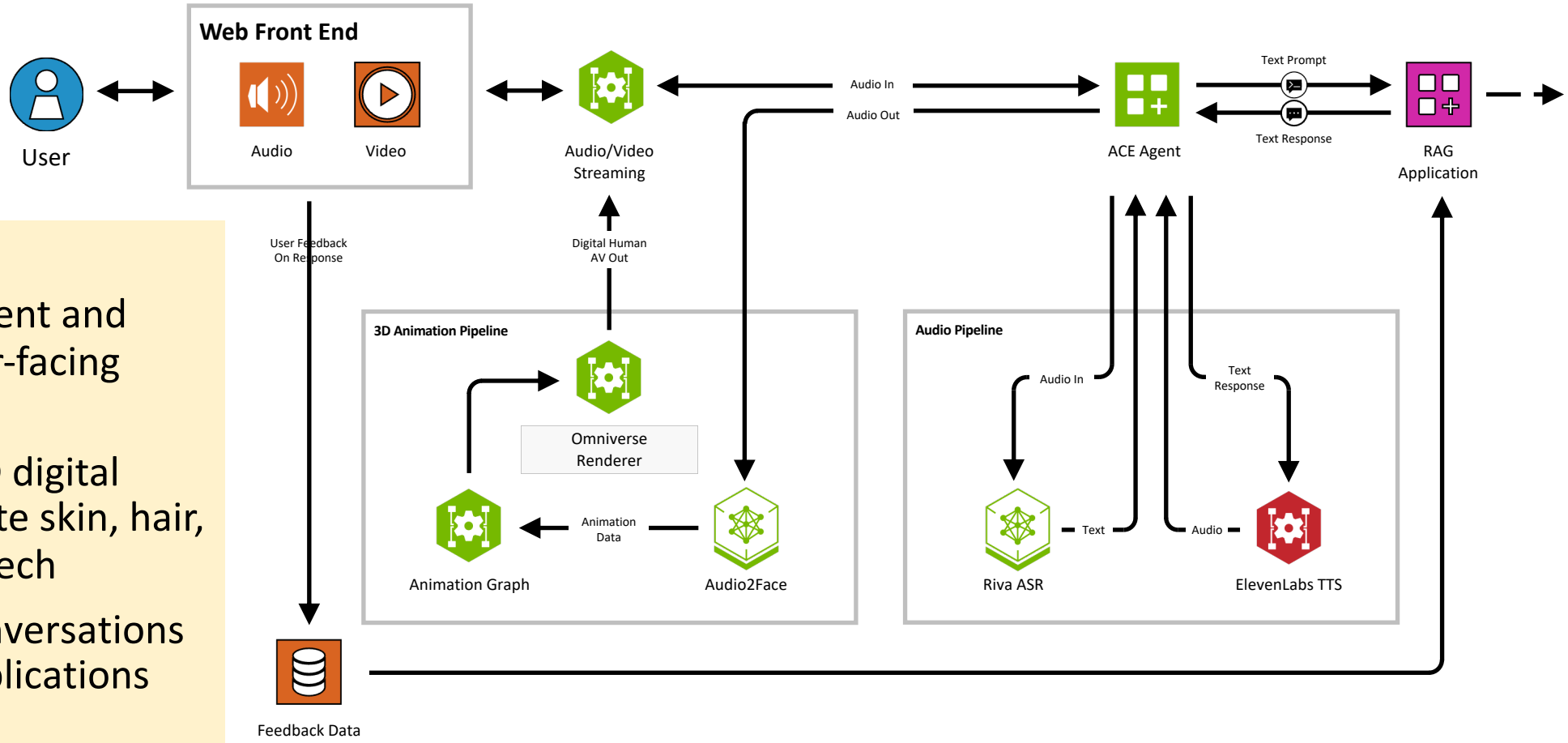
- Personalized Responses: Handles structured and unstructured customer queries (e.g., order details, spending history).
- Multi-Turn Dialogue: Offers context-aware, seamless interactions across multiple questions.
- Custom Conversation Style: Adapts text responses to reflect corporate branding and tone.
- 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.



Source: NVIDIA DLI (2024), Introduction to NVIDIA NIM™ Microservices, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-23+V1

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

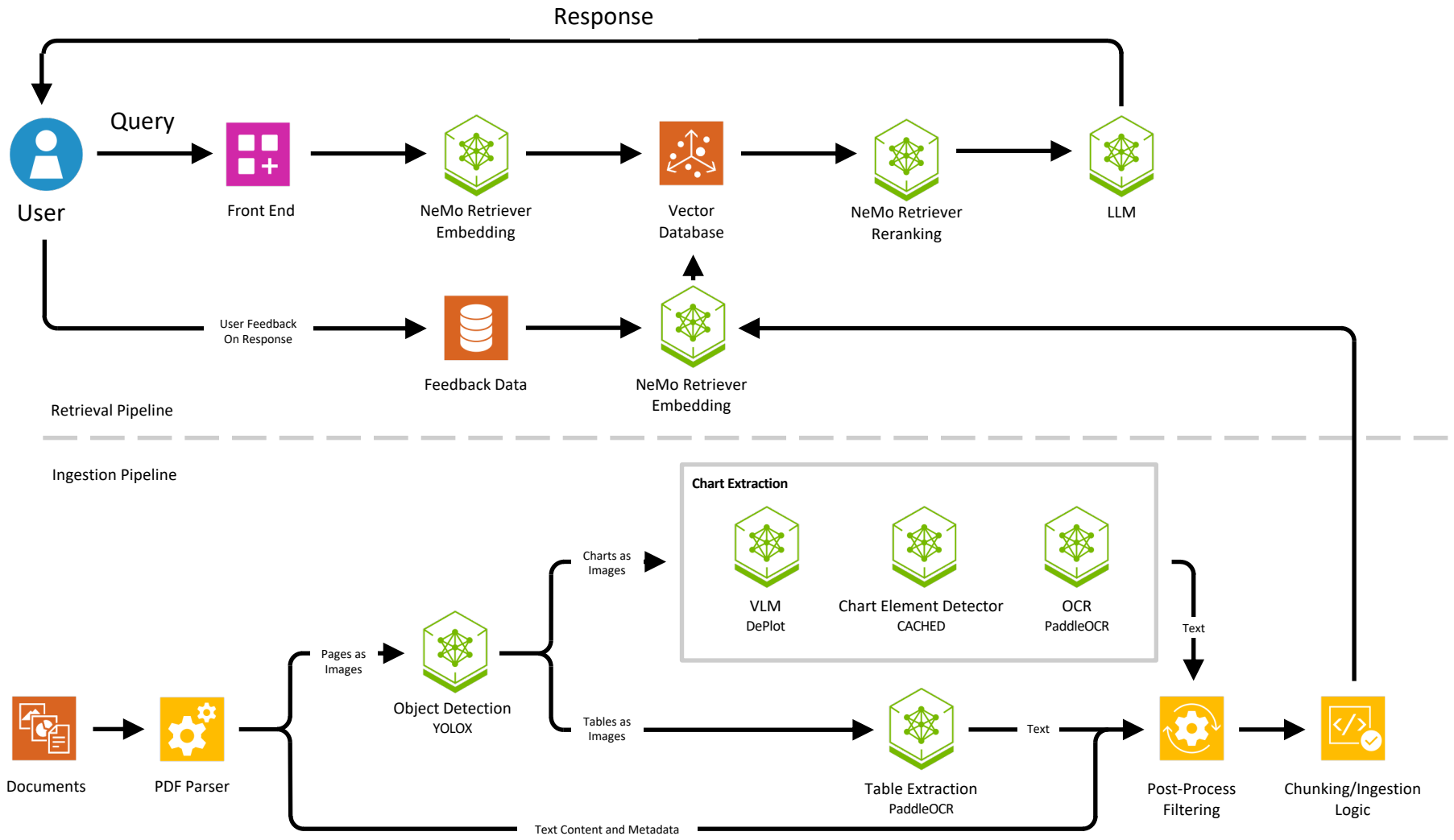
Source: NVIDIA DLI (2024), Introduction to NVIDIA NIM™ Microservices, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-23+V1

Multimodal PDF Data Extraction for Enterprise RAG

Unlocks Knowledge from trillions of PDFs

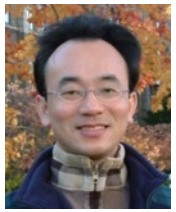
Benefits

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



Source: NVIDIA DLI (2024), Introduction to NVIDIA NIM™ Microservices, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-23+V1

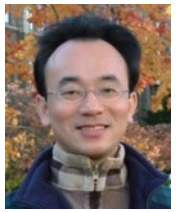
Teaching



- **Generative AI Innovative Applications**
  - Spring 2025, Spring 2026
- **Artificial Intelligence in Finance and Quantitative**
  - Fall 2021, Fall 2022, Fall 2023, Spring 2025, Spring 2026
- **Software Engineering**
  - Fall 2020, Fall, 2021, Spring 2022, Spring 2023, Spring 2024, Spring 2025, Spring 2026
- **Artificial Intelligence**
  - Spring 2021, Fall 2022, Fall 2024, Fall 2025
- **Sustainability and ESG Data Analytics**
  - Spring 2024, Fall 2024, Fall 2025
- **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, ,Fall 2025
- **Foundation of Business Cloud Computing**
  - Spring 2021, Spring 2022, Spring 2023, Spring 2024

https://web.ntpu.edu.tw/~myday/teaching.htm

Research Projects



1. **PI, Generative AI Multi-Agent Systems with LLM-Based RAG for ESG Reporting Automation**
   - NSTC (E4104), NSTC 114-2221-E-305-002-, 2025/08/01~2026/07/31
2. **PI, ESG Sustainability Commitment Verification Competition and Annotation Data Collection Project (AI Cup 2026 VeriPromiseESG)**
   - Ministry of Education (MoE), AI CUP 2026 MoE National Collegiate Artificial Intelligence Professional Domain Thematic Competition Project, 2025/10/01~2026/09/30
3. **Co-PI, Physical AI and Physical AI Robotics Curriculum Development Project, Forward-Looking AI Talent Cultivation Program (NAPAI)**
   - Ministry of Education (MoE), 2025/11/01~2027/03/31
4. **Co-PI, Industry-Academia Alliance Program for Smart Healthcare: Smart Precision Next-Generation Preventive Medicine**
   - NSTC 114-2622-8-371-001-IE, 2025/11/01~2026/10/31
5. **Co-PI, Digital Support, Unimpeded Communication: The Development, Support and Promotion of AI-assisted Communication Assistive Devices for Speech Impairment(3/3), Sub-project 3: Multimodal Cross-lingual Task-Oriented Dialogue System for Inclusive Communication Support**
   - NSTC (HZZ22), NSTC 114-2425-H-305-003-, 3 Years (2023/05/01-2026/04/30) Year 3: 2025/05/01~2026/04/30

https://web.ntpu.edu.tw/~myday/cindex.htm#projects

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. Agentic AI and Large Multimodal Agents (LMAs)
  8. Case Study on Generative AI Innovative Applications

NVIDIA University Ambassador Certified Instructor

aws educate Cloud Ambassador 2020 Cohort

aws academy Accredited Educator

aws certified Solutions Architect Associate

aws certified Cloud Practitioner

Generative AI Innovative Applications

NTPU 國立臺北大學 National Taipei University

AACSB ACCREDITED

Contact Information

Min-Yuh Day, Ph.D.

Professor and Director

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

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