Artificial Intelligence

Generative AI, Philosophy and Ethics of AI and the Future of AI

1131AI09
MBA, IM, NTPU (M5276) (Fall 2024)
Tue 2, 3, 4 (9:10-12:00) (B3F17)

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

Min-Yuh Day, Ph.D.
Professor

Institute of Information Management, National Taipei University

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

2024-12-10

Syllabus

Week	Date	Subject/Topics
1	2024/09/10	Introduction to Artificial Intelligence
2	2024/09/17	Mid-Autumn Festival (Day off)
3	2024/09/24	Artificial Intelligence and Intelligent Agents; Problem Solving
4	2024/10/01	Knowledge, Reasoning and Knowledge Representation; Uncertain Knowledge and Reasoning
5	2024/10/08	Case Study on Artificial Intelligence I
6	2024/10/15	Machine Learning: Supervised and Unsupervised Learning

Syllabus

Week	Date	Subject/Topics
7	2024/10/22	The Theory of Learning and Ensemble Learning
8	2024/10/29	Midterm Project Report
9	2024/11/05	Self-Learning
10	2024/11/12	Deep Learning and Reinforcement Learning
11	2024/11/19	Case Study on Artificial Intelligence II
12	2024/11/26	Deep Learning for Natural Language Processing

Syllabus

Week	Date	Subject/Topics
13	2024/12/03	Computer Vision and Robotics
14	2024/12/10	Generative AI, Philosophy and Ethics of AI and the Future of AI
15	2024/12/17	Final Project Report I
16	2024/12/24	Final Project Report II

Generative AI, Philosophy and Ethics of AI and the Future of AI

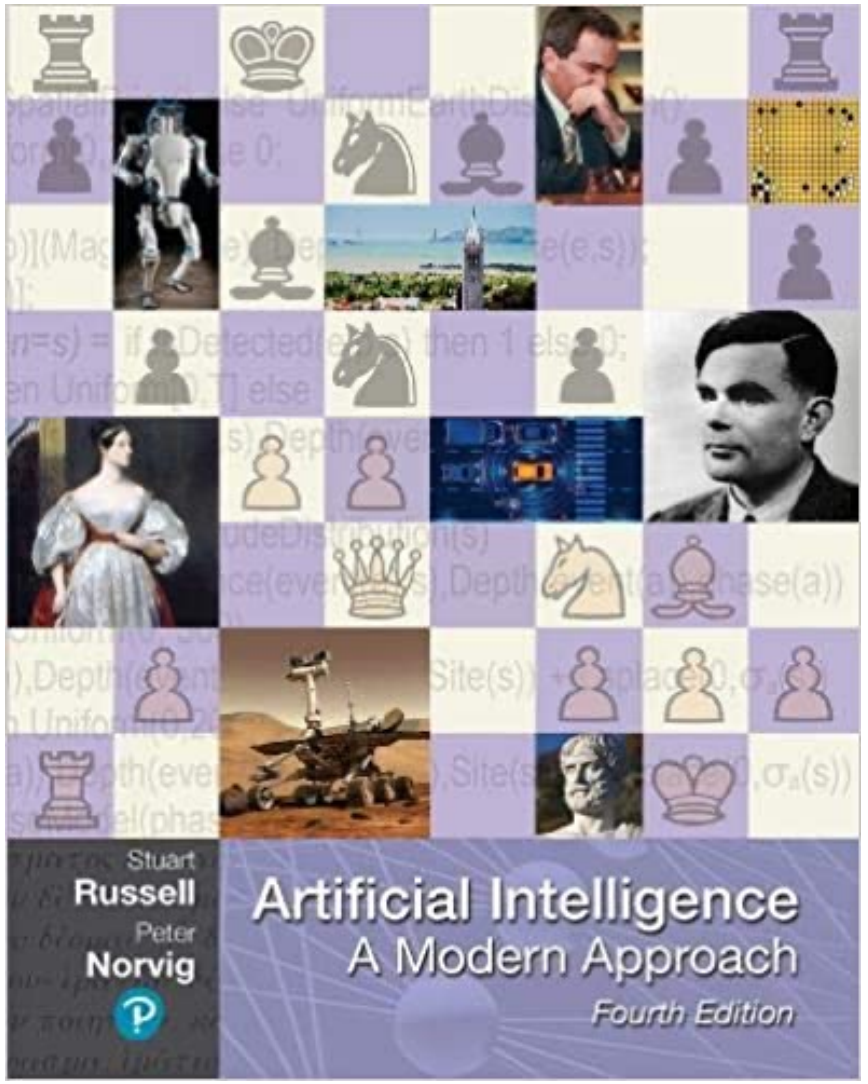
Outline

- Generative AI
- Philosophy, Ethics, and Safety of AI
  - The Limits of AI
  - Can Machines Really Think?
  - The Ethics of AI
- The Future of AI
  - AI Components
  - AI Architectures

Stuart Russell and Peter Norvig (2020),

Artificial Intelligence: A Modern Approach,

4th Edition, Pearson



Source: Stuart Russell and Peter Norvig (2020), Artificial Intelligence: A Modern Approach, 4th Edition, Pearson

https://www.amazon.com/Artificial-Intelligence-A-Modern-Approach/dp/0134610997/

Artificial Intelligence: A Modern Approach

1. Artificial Intelligence
2. Problem Solving
3. Knowledge and Reasoning
4. Uncertain Knowledge and Reasoning
5. Machine Learning
6. Communicating, Perceiving, and Acting
7. Philosophy and Ethics of AI

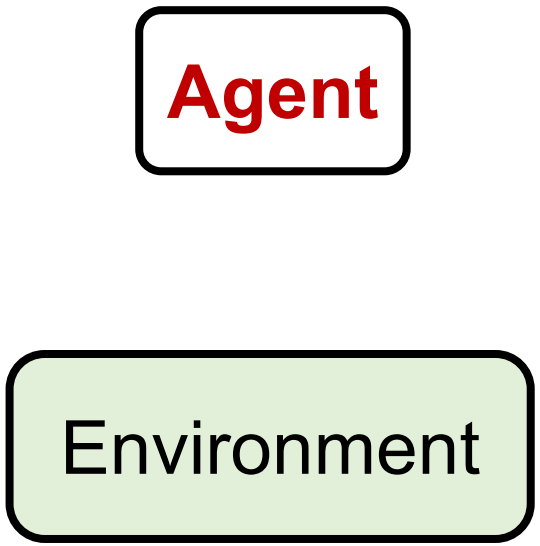
Source: Stuart Russell and Peter Norvig (2020), Artificial Intelligence: A Modern Approach, 4th Edition, Pearson

Artificial Intelligence:
7. Philosophy and Ethics of AI

- Philosophy, Ethics, and Safety of AI
  - The Limits of AI
  - Can Machines Really Think?
  - The Ethics of AI
- The Future of AI
  - AI Components
  - AI Architectures

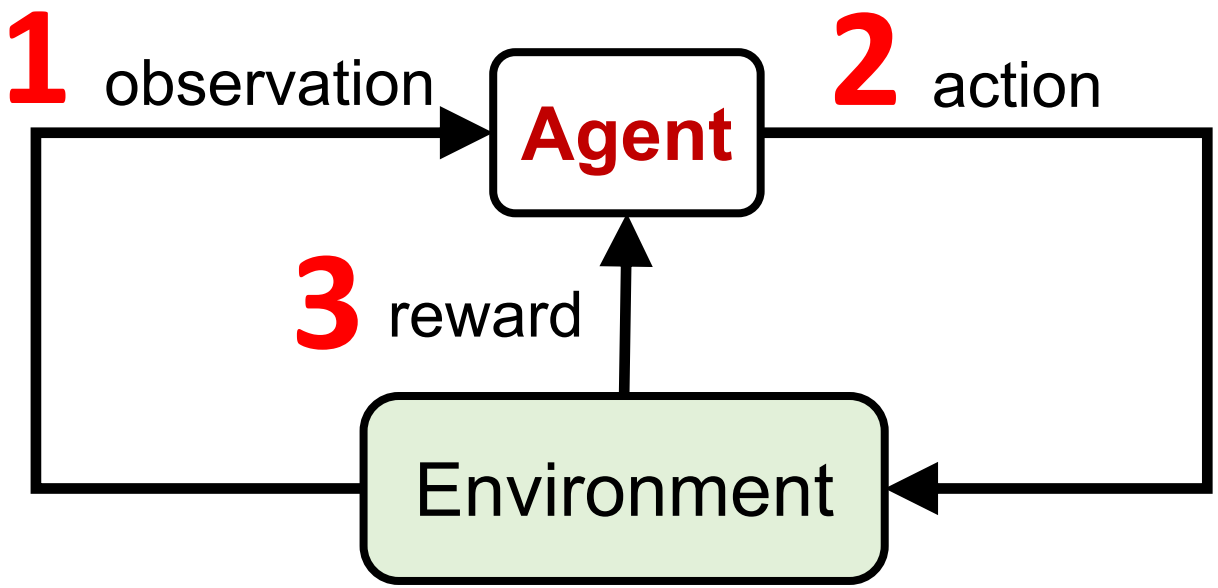
Source: Stuart Russell and Peter Norvig (2020), Artificial Intelligence: A Modern Approach, 4th Edition, Pearson

Reinforcement Learning (DL)



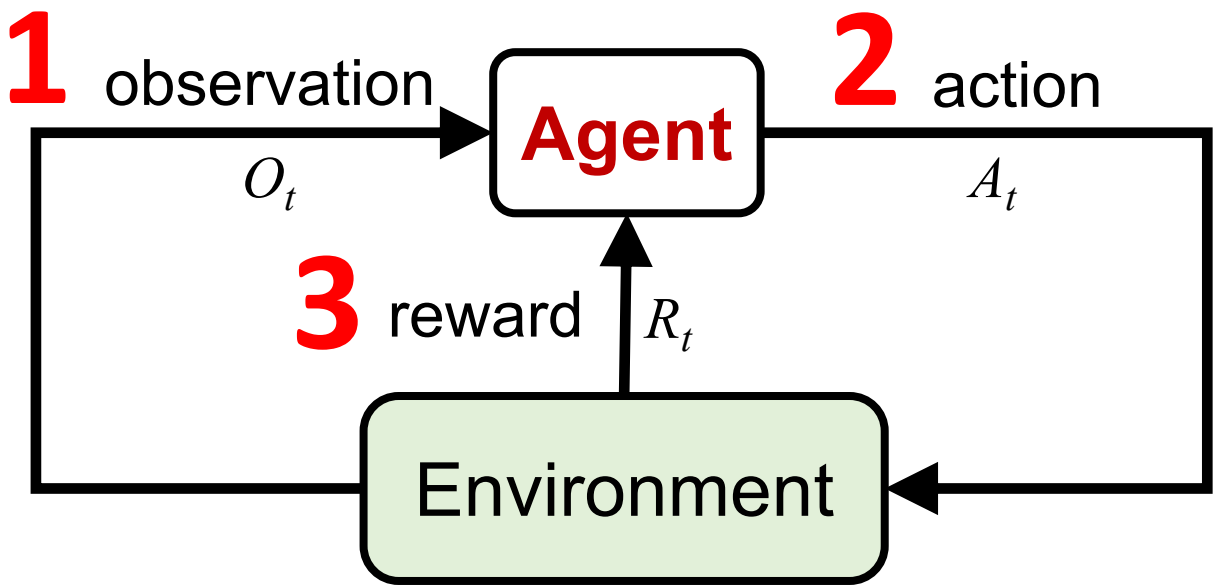
Source: Richard S. Sutton & Andrew G. Barto (2018), Reinforcement Learning: An Introduction, 2nd Edition, A Bradford Book.

Reinforcement Learning (DL)



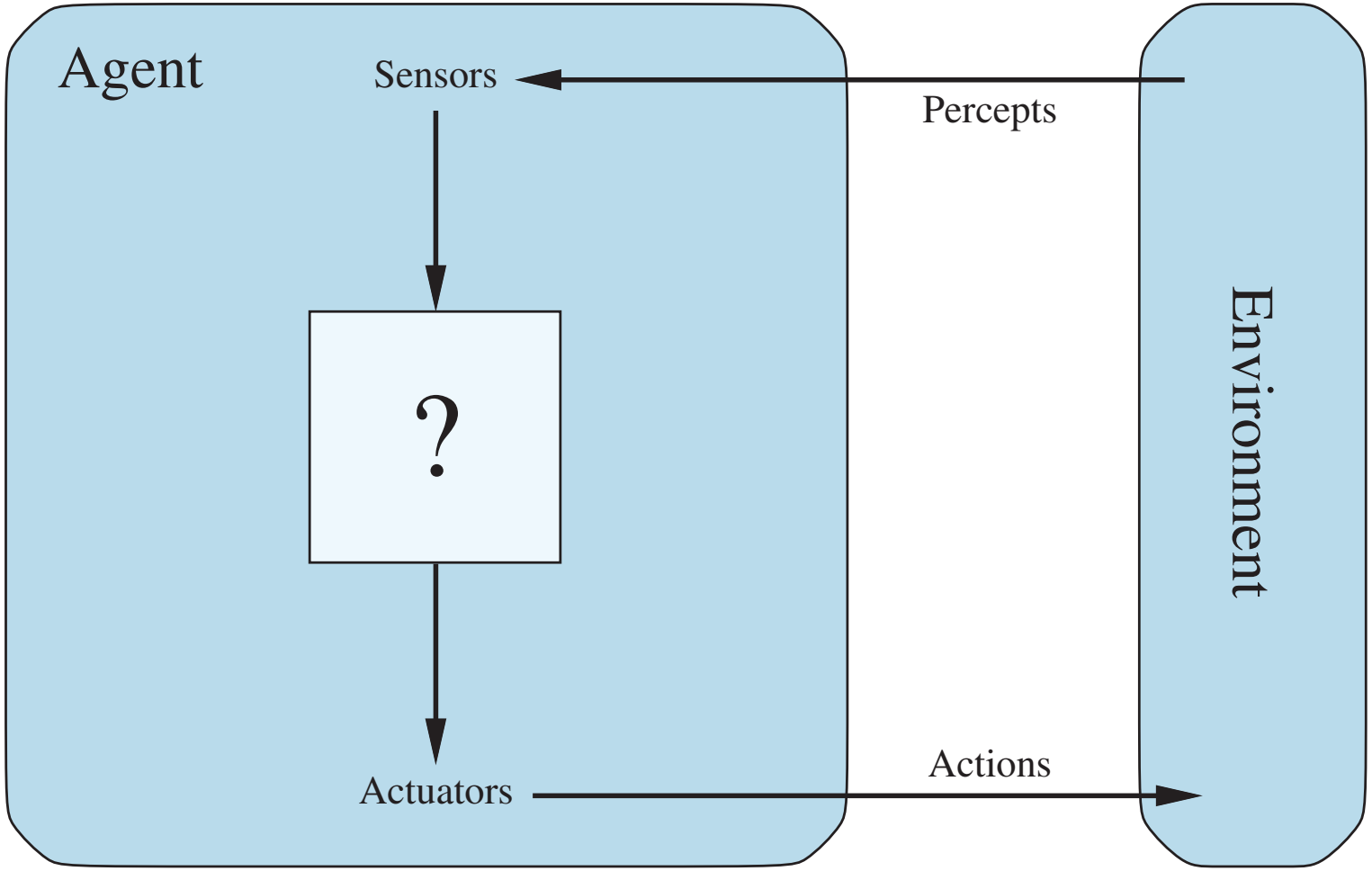
Source: Richard S. Sutton & Andrew G. Barto (2018), Reinforcement Learning: An Introduction, 2nd Edition, A Bradford Book.

Reinforcement Learning (DL)



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

Agents interact with environments through sensors and actuators



Source: Stuart Russell and Peter Norvig (2020), Artificial Intelligence: A Modern Approach, 4th Edition, Pearson

AI Acting Humanly:
The Turing Test Approach
(Alan Turing, 1950)

- Knowledge Representation
- Automated Reasoning
- Machine Learning (ML)
  - Deep Learning (DL)
- Computer Vision (Image, Video)
- Natural Language Processing (NLP)
- Robotics

Source: Stuart Russell and Peter Norvig (2020), Artificial Intelligence: A Modern Approach, 4th Edition, Pearson

Generative AI
Large Language Models (LLMs)
Foundation Models

NVIDIA Developer Program

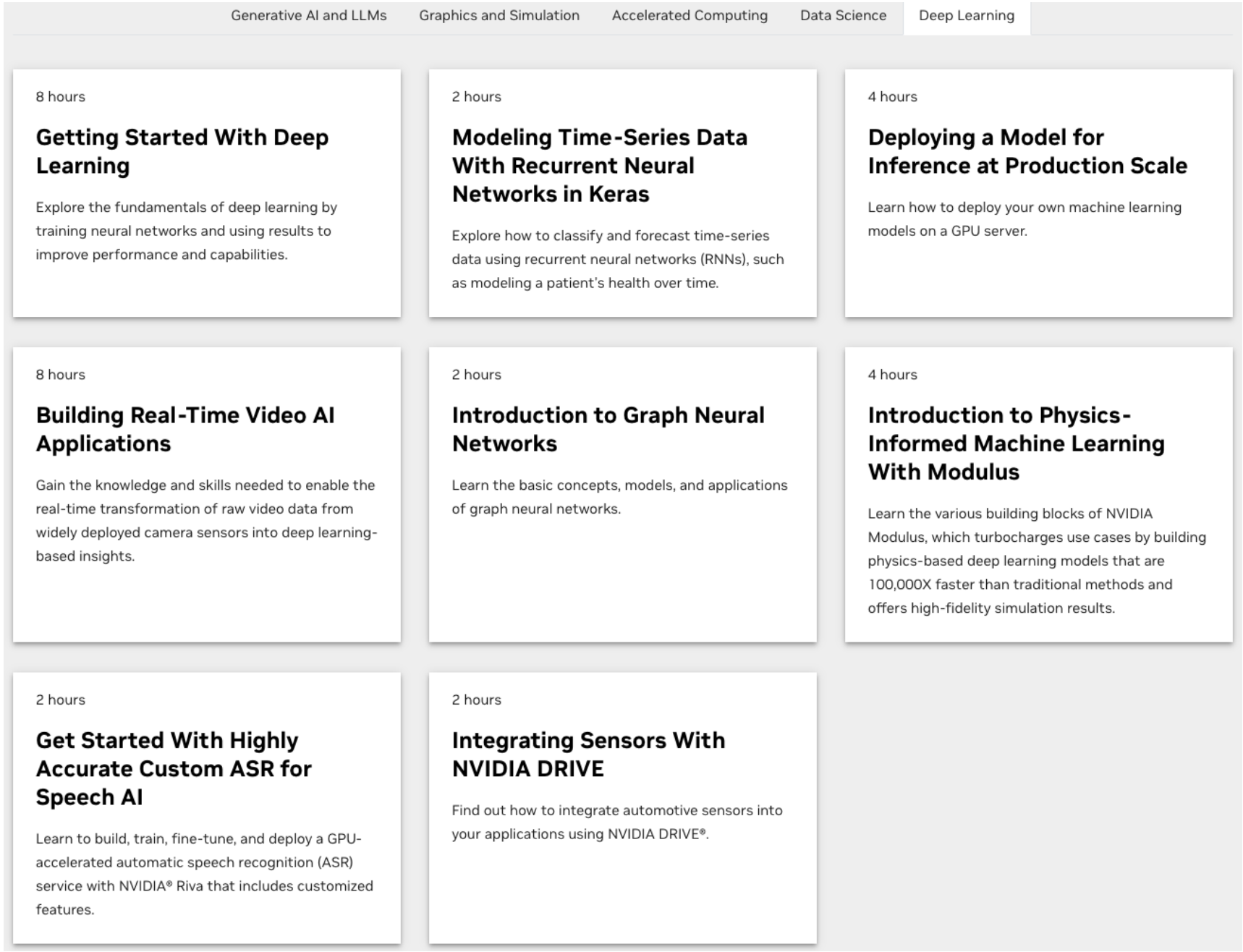
https://developer.nvidia.com/join-nvidia-developer-program

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)



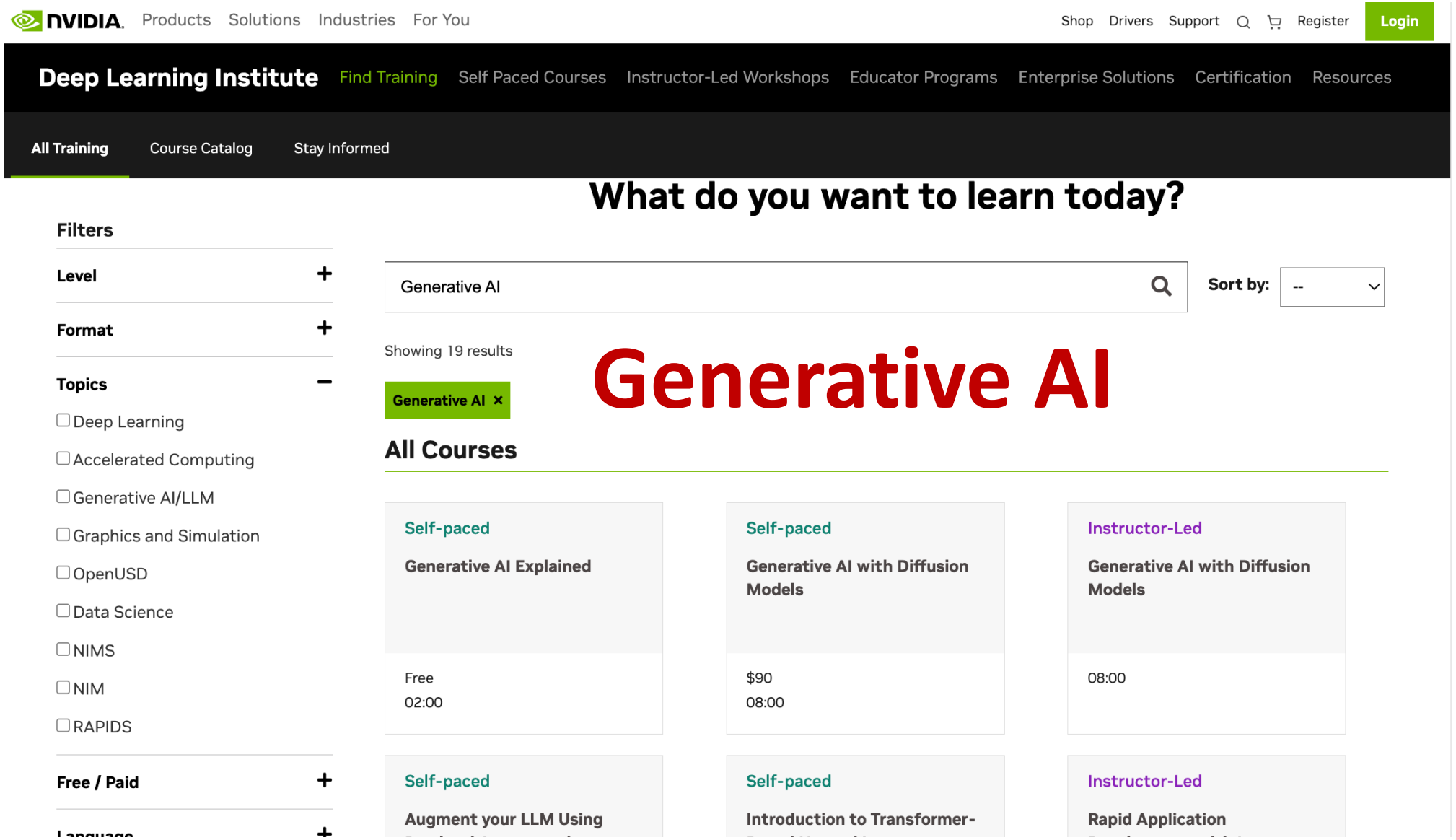
https://developer.nvidia.com/join-nvidia-developer-program

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

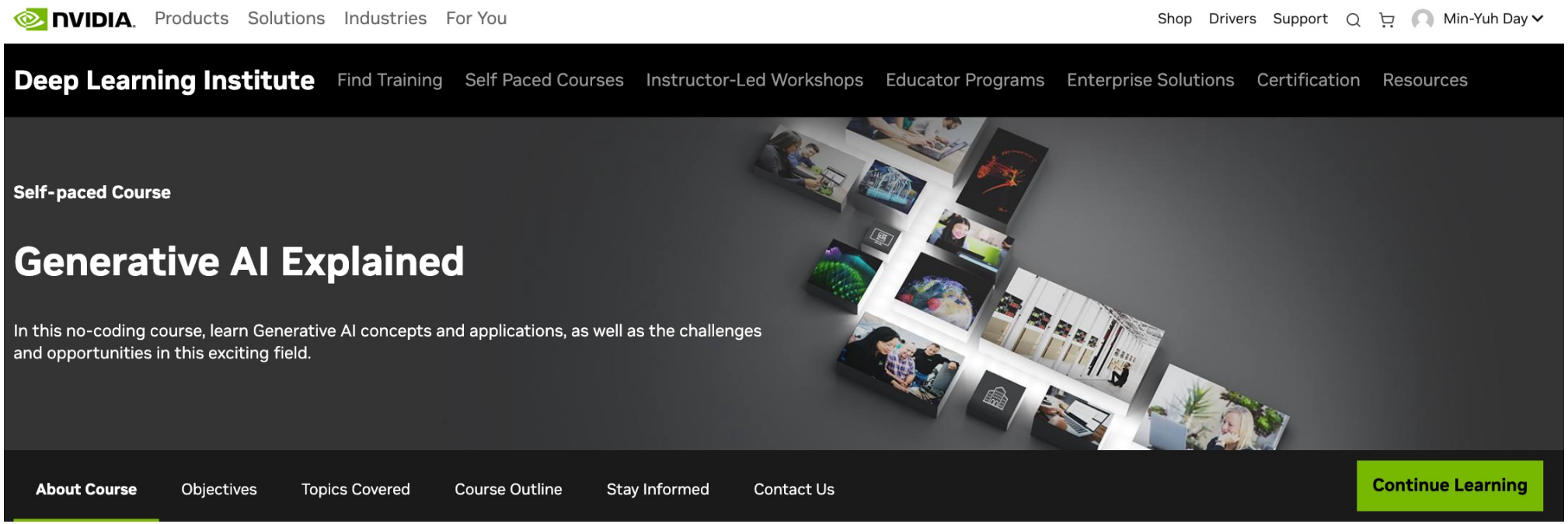
https://www.nvidia.com/en-us/learn/learning-path/generative-ai-llm/

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

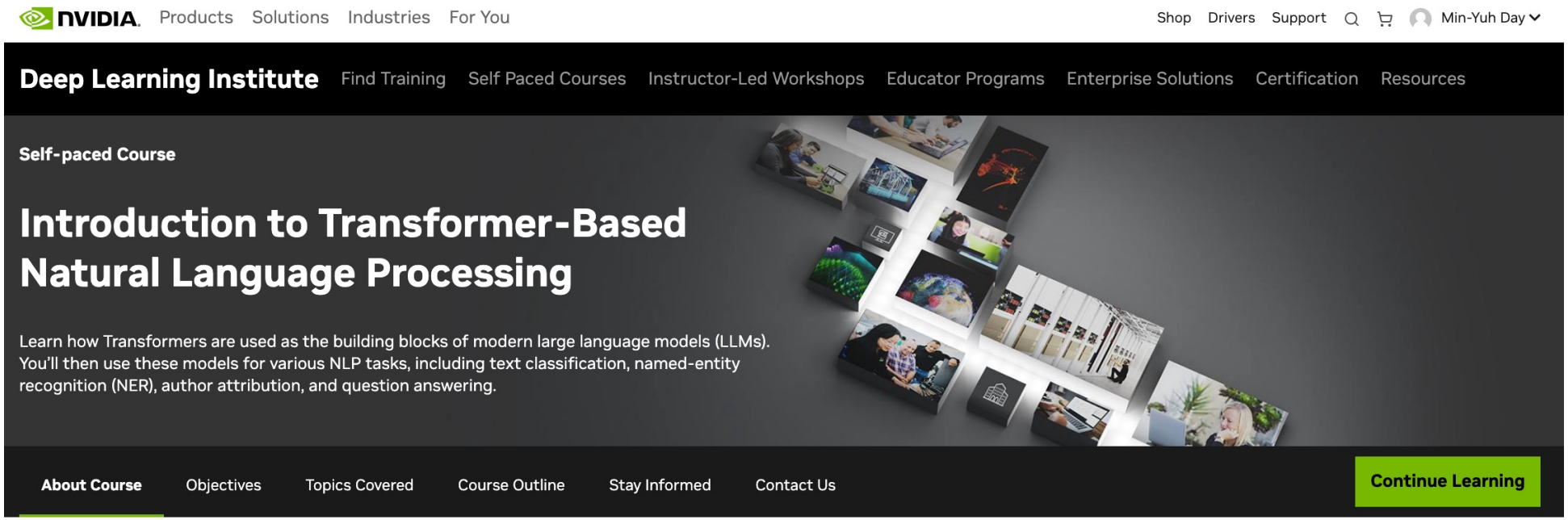
Level: Technical - Beginner

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

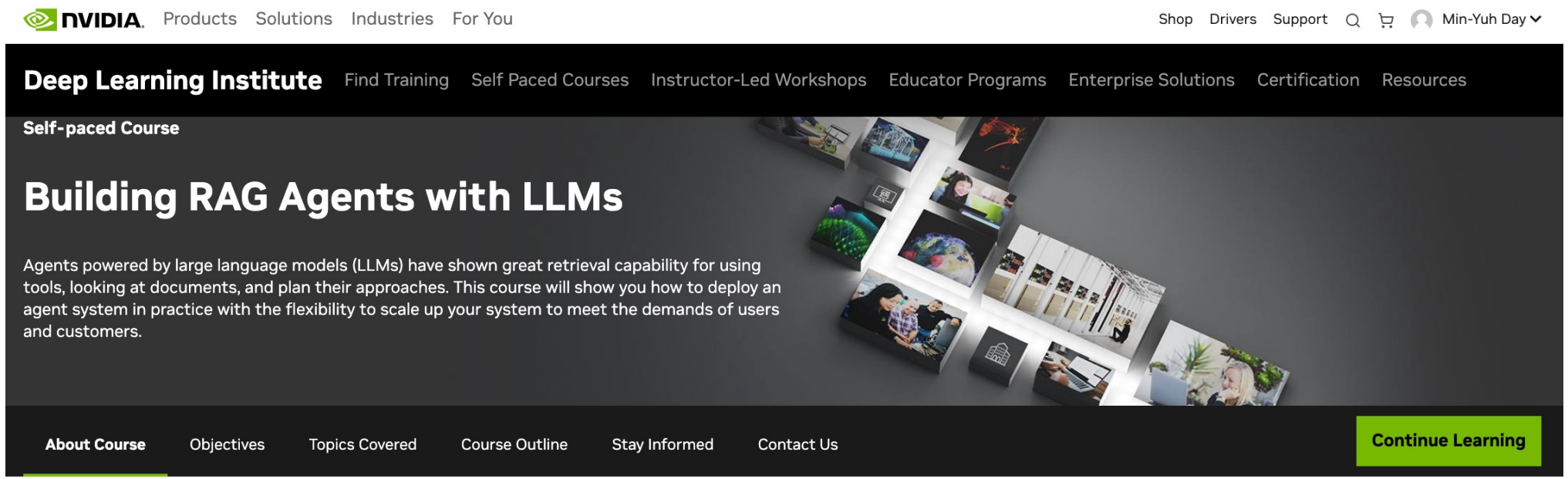
Level: Technical - Beginner

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

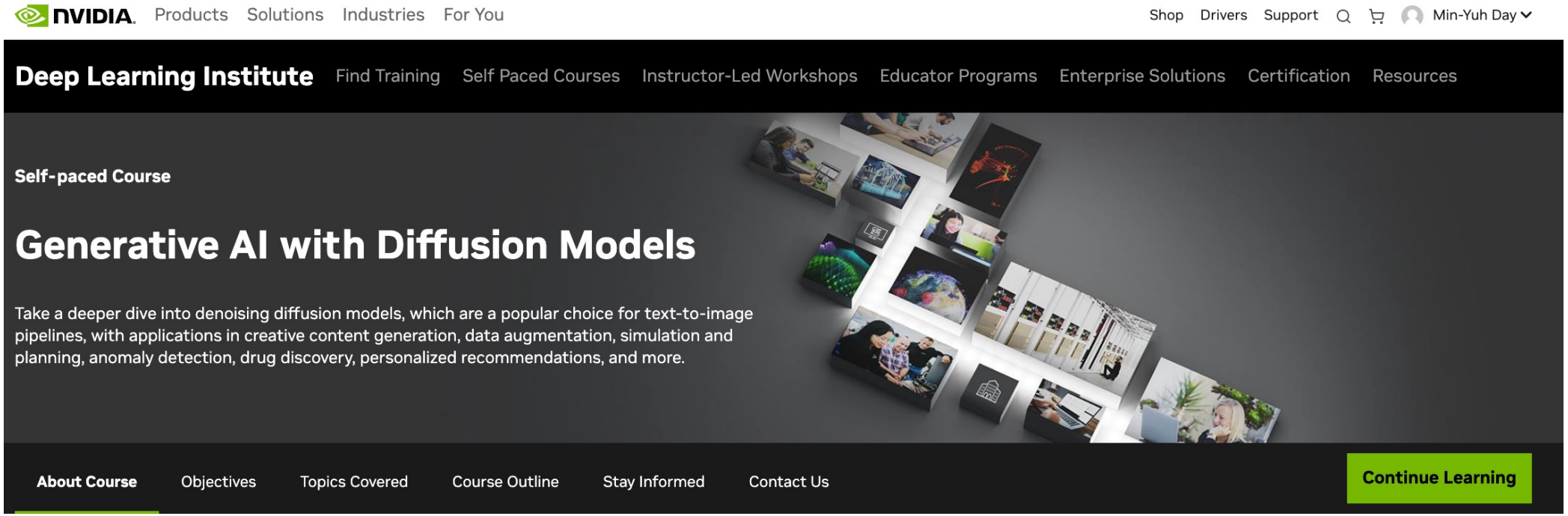
Subject: Generative AI/LLM

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

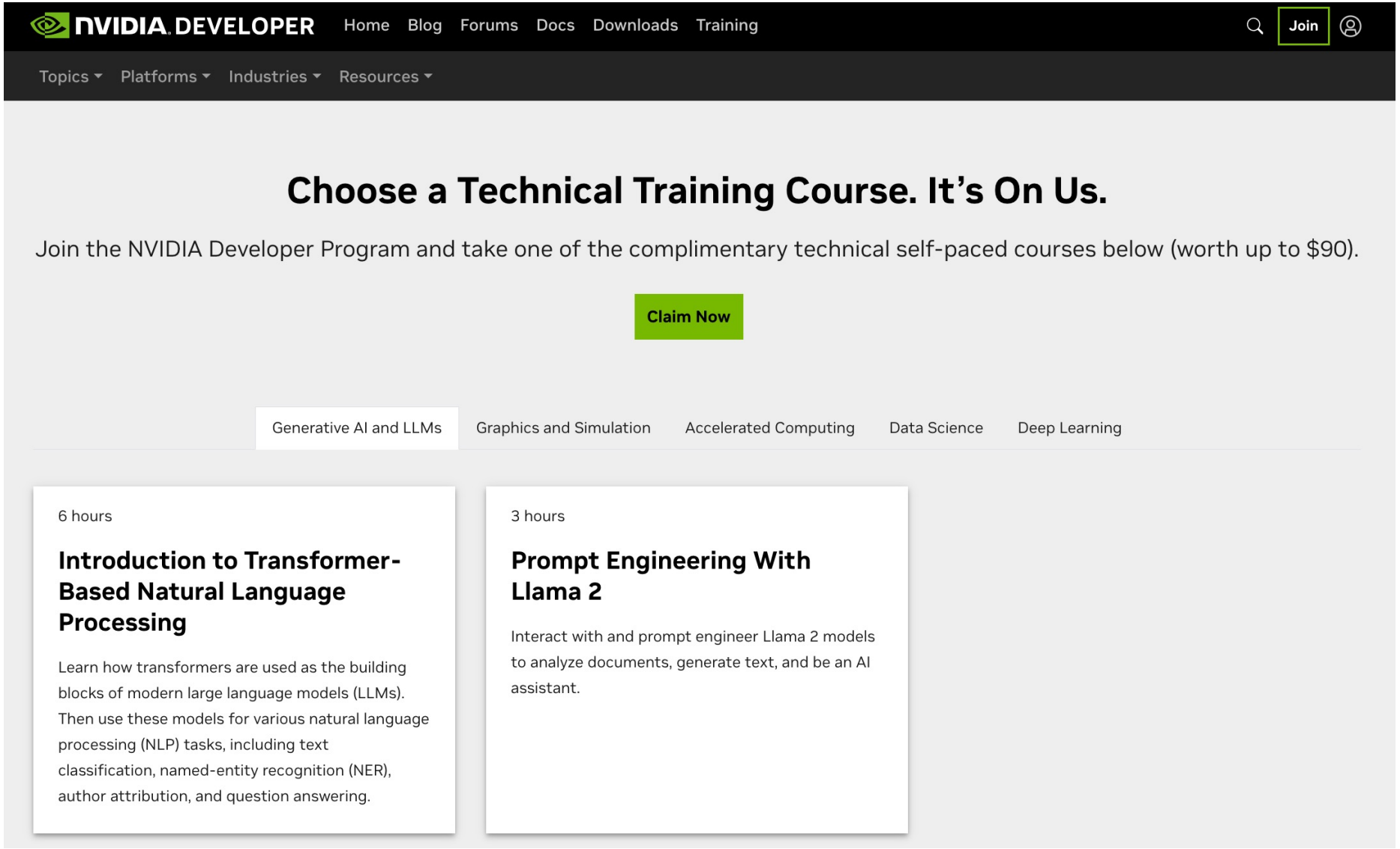
Course Prerequisites:
A basic understanding of Deep Learning Concepts.

Learning Objectives

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

Join the NVIDIA Developer Program

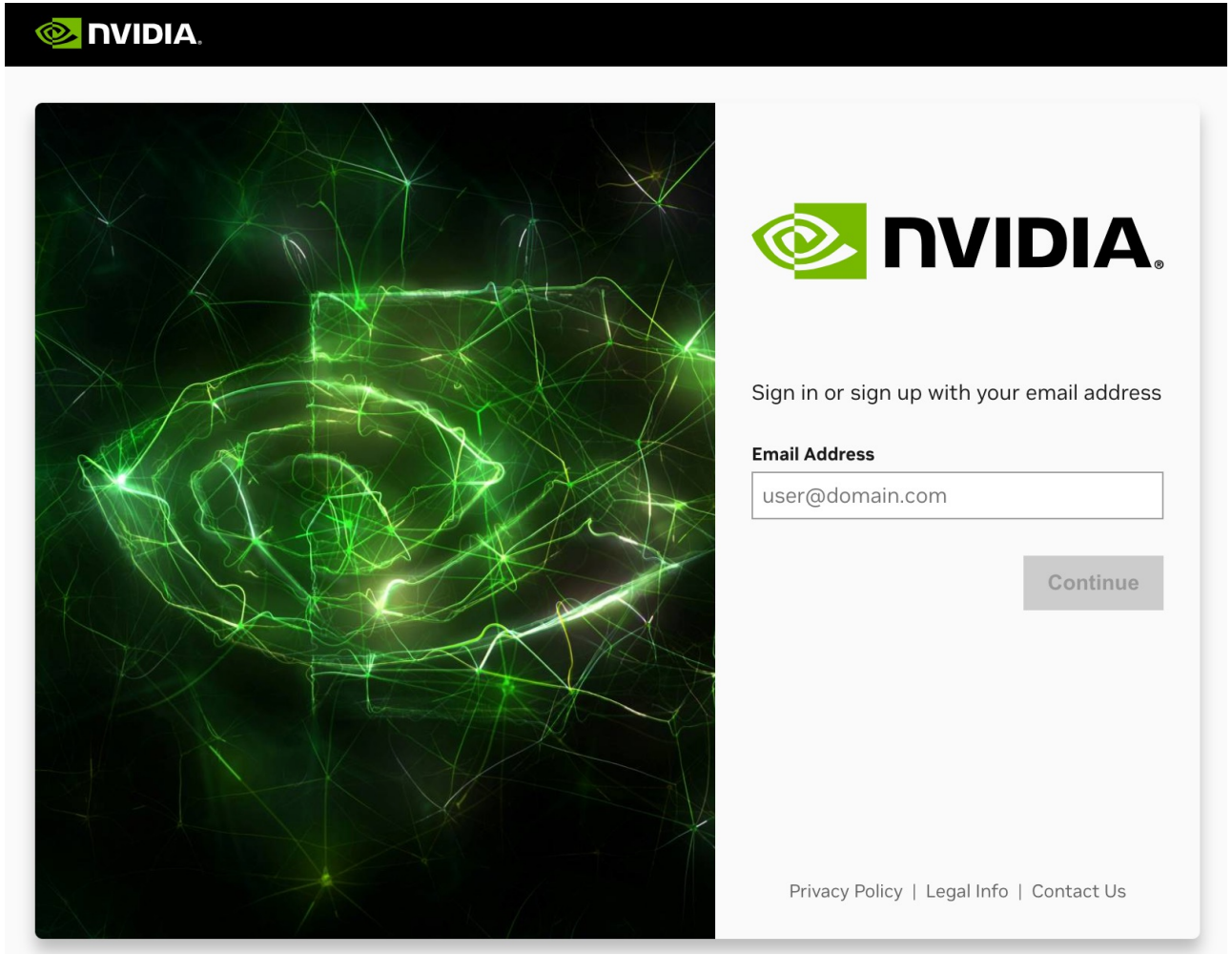
take one of the complimentary technical self-paced courses (worth up to $90)



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

Join the NVIDIA Developer Program

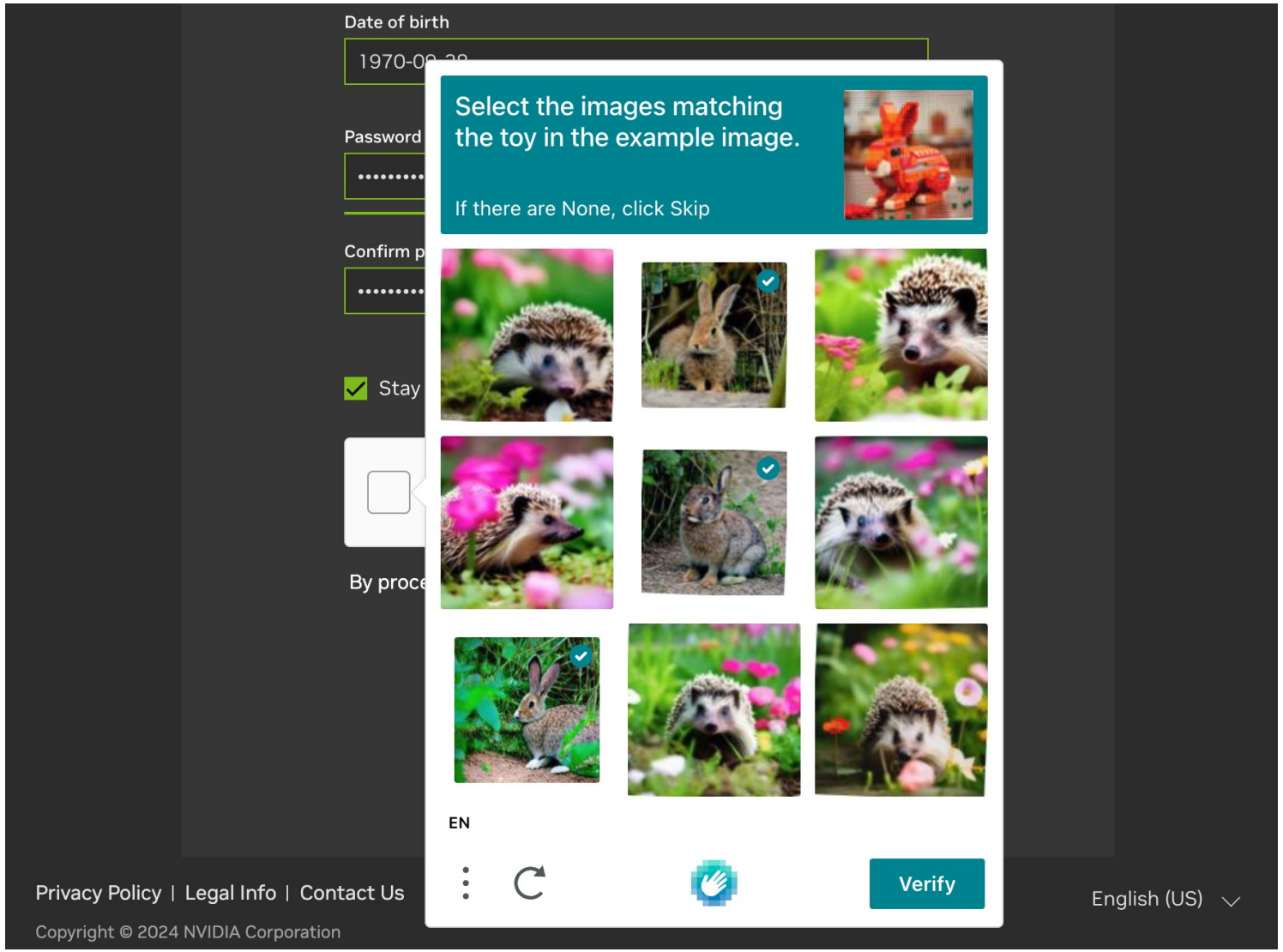
take one of the complimentary technical self-paced courses (worth up to $90)



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

Join the NVIDIA Developer Program

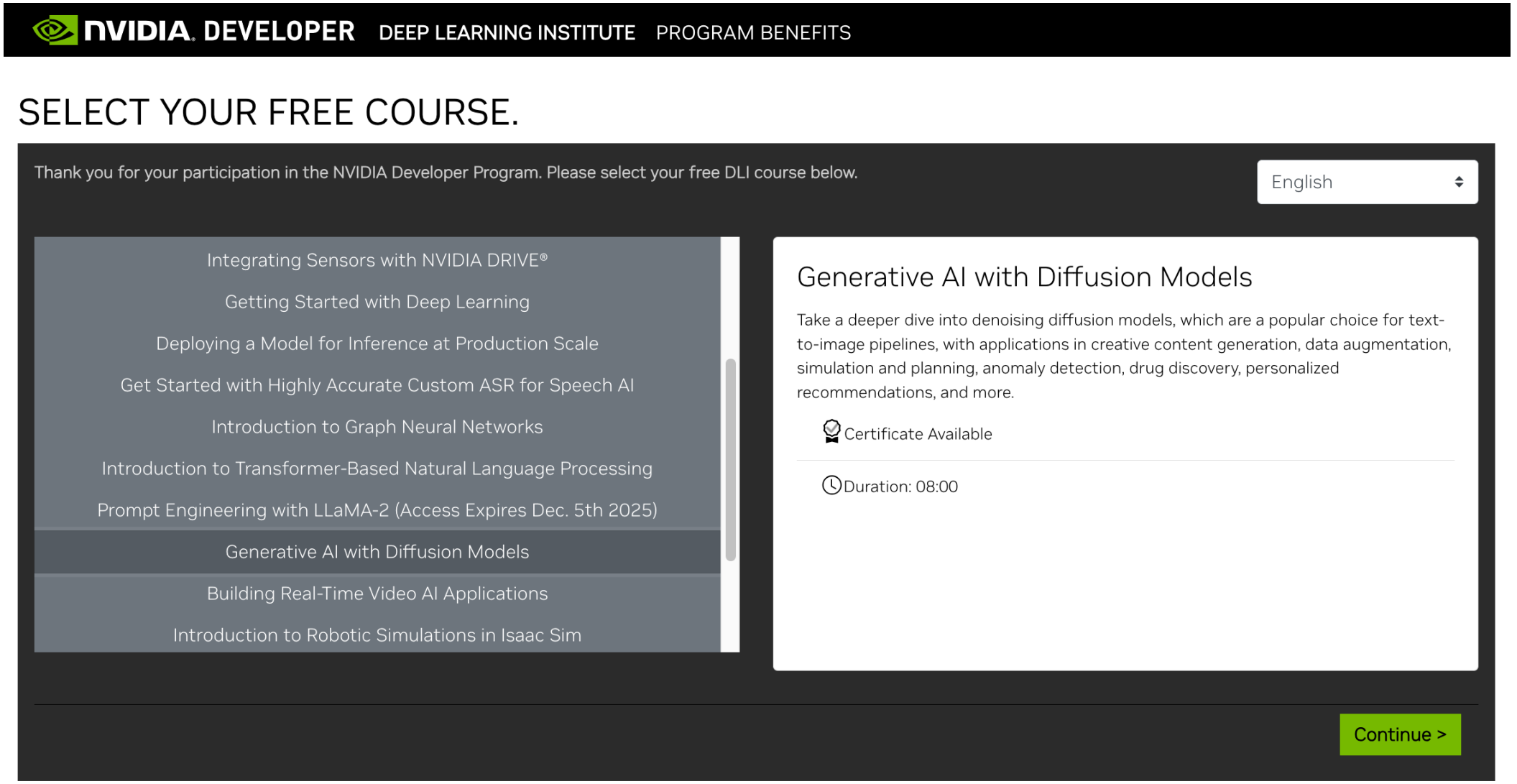
take one of the complimentary technical self-paced courses (worth up to $90)



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

Join the NVIDIA Developer Program

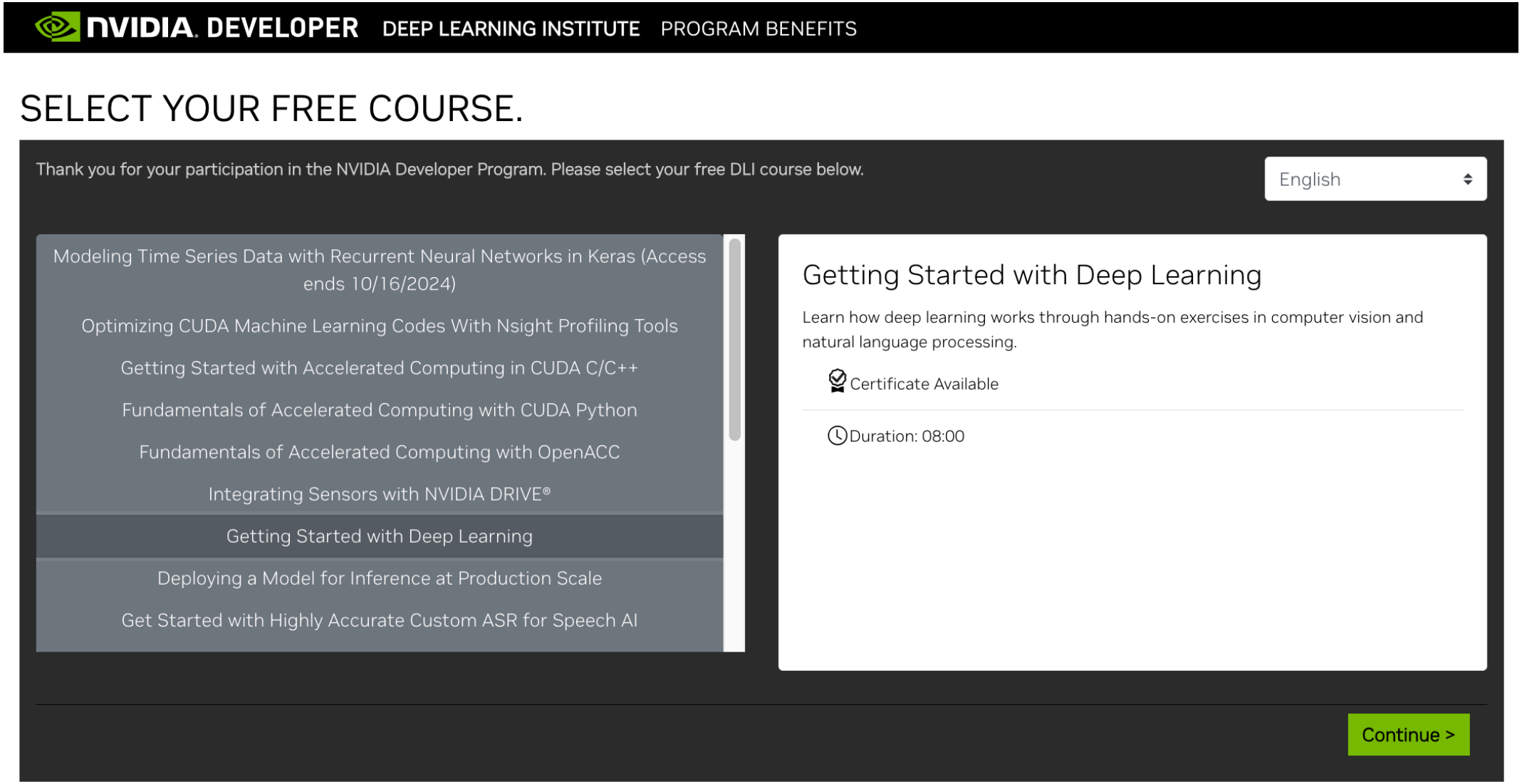
take one of the complimentary technical self-paced courses (worth up to $90)



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

Join the NVIDIA Developer Program

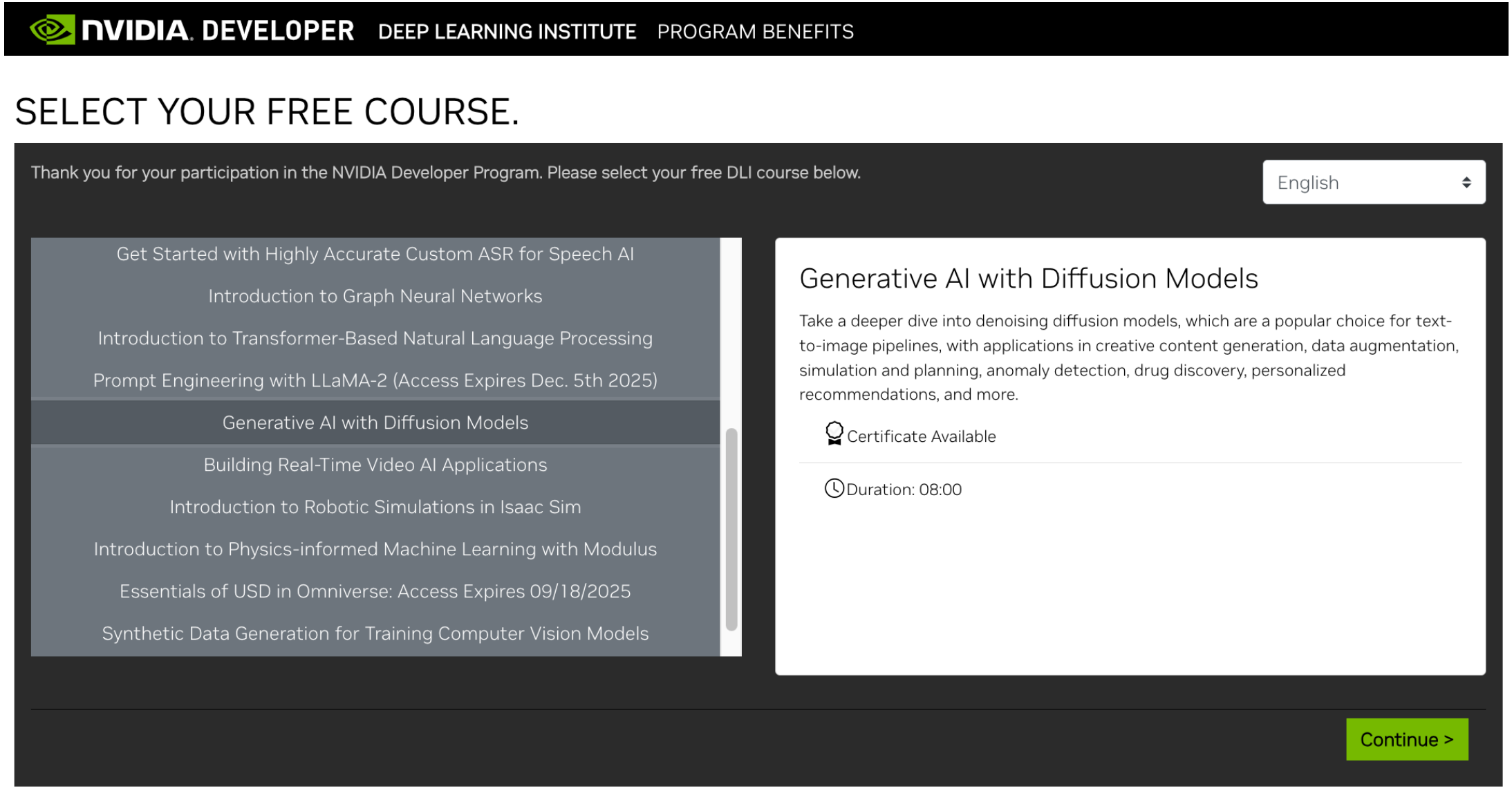
take one of the complimentary technical self-paced courses (worth up to $90)



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

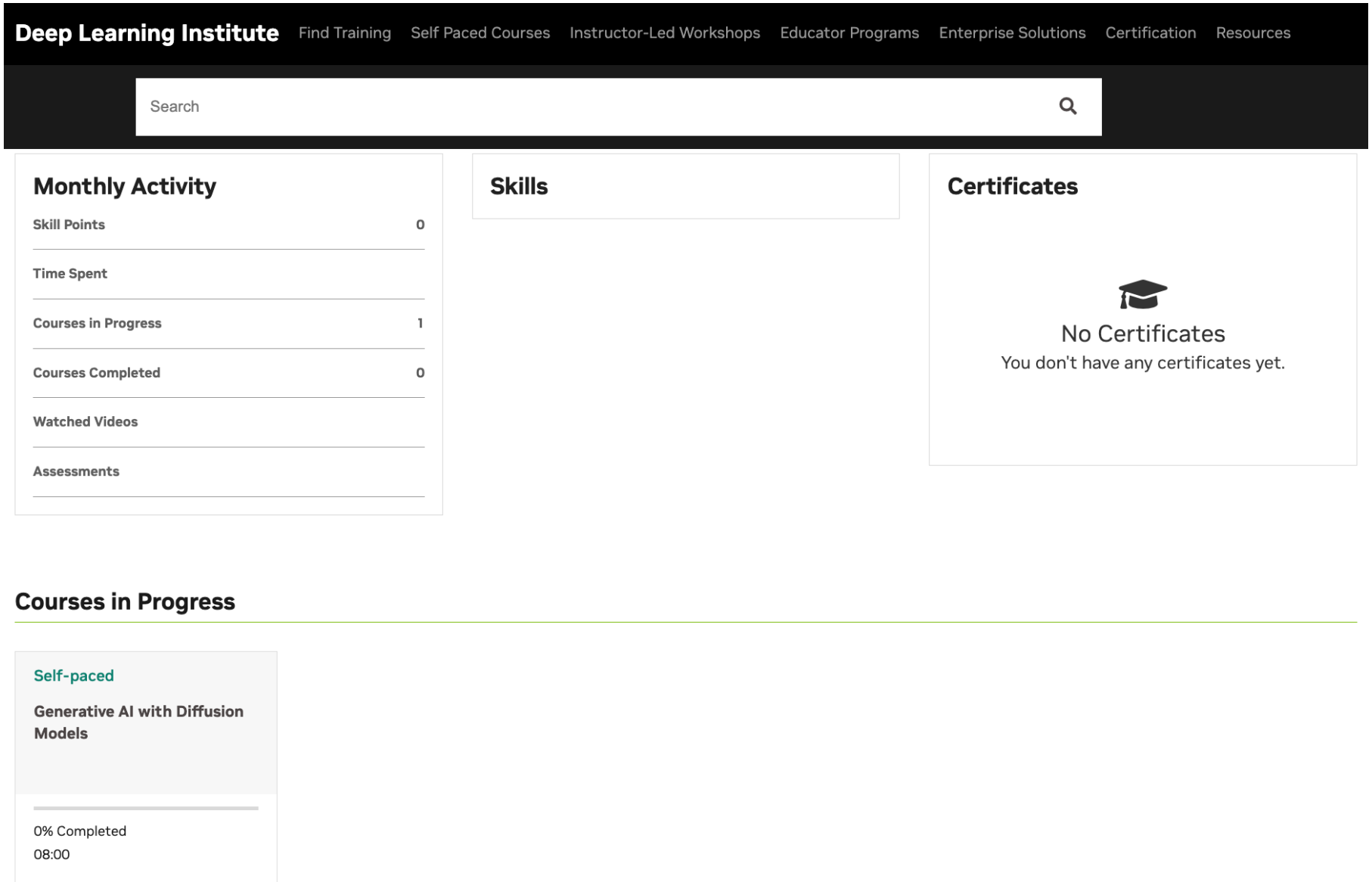
Join the NVIDIA Developer Program

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



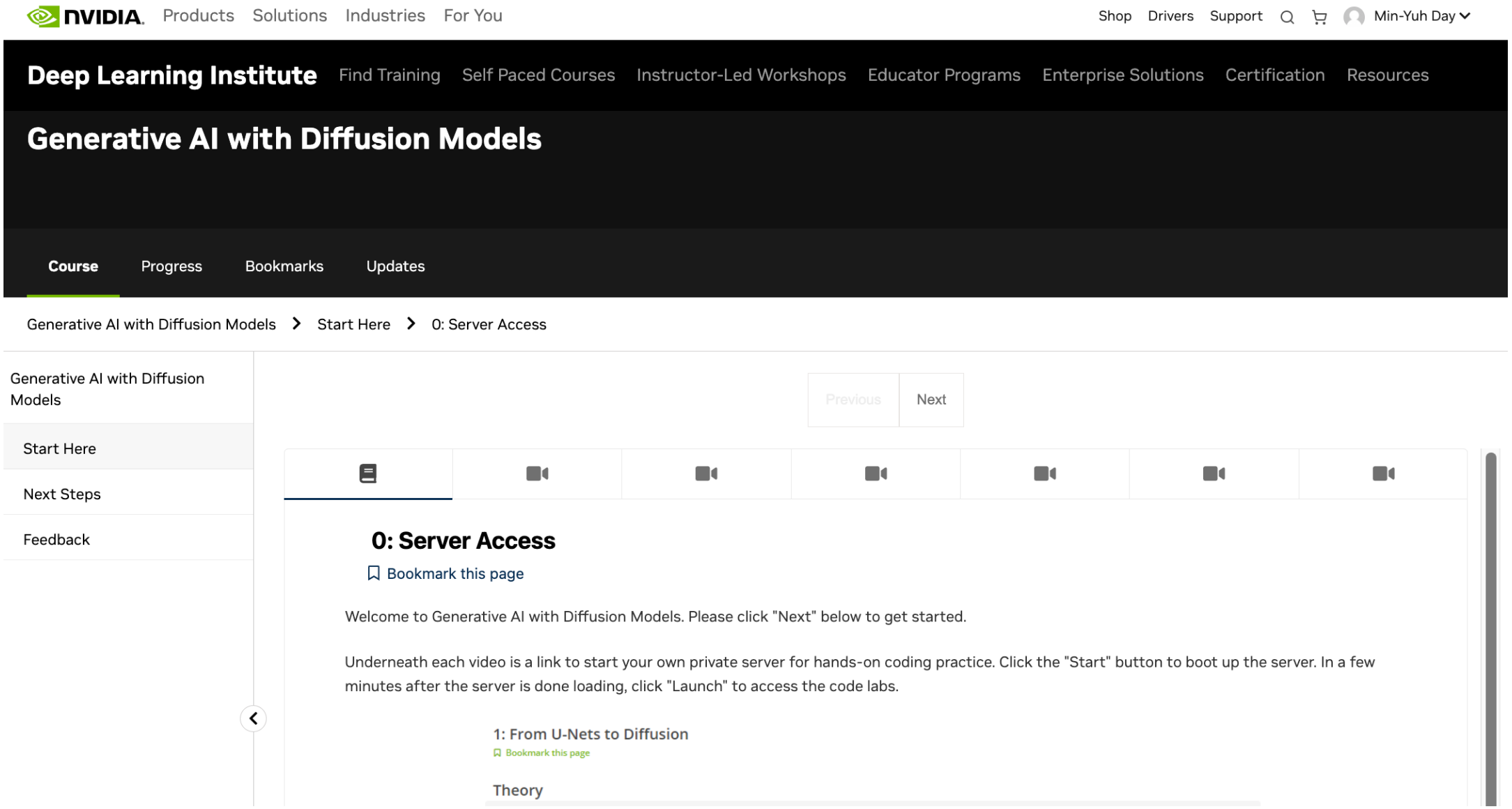
https://developer.nvidia.com/join-nvidia-developer-program

NVIDIA Deep Learning Institute (DLI)



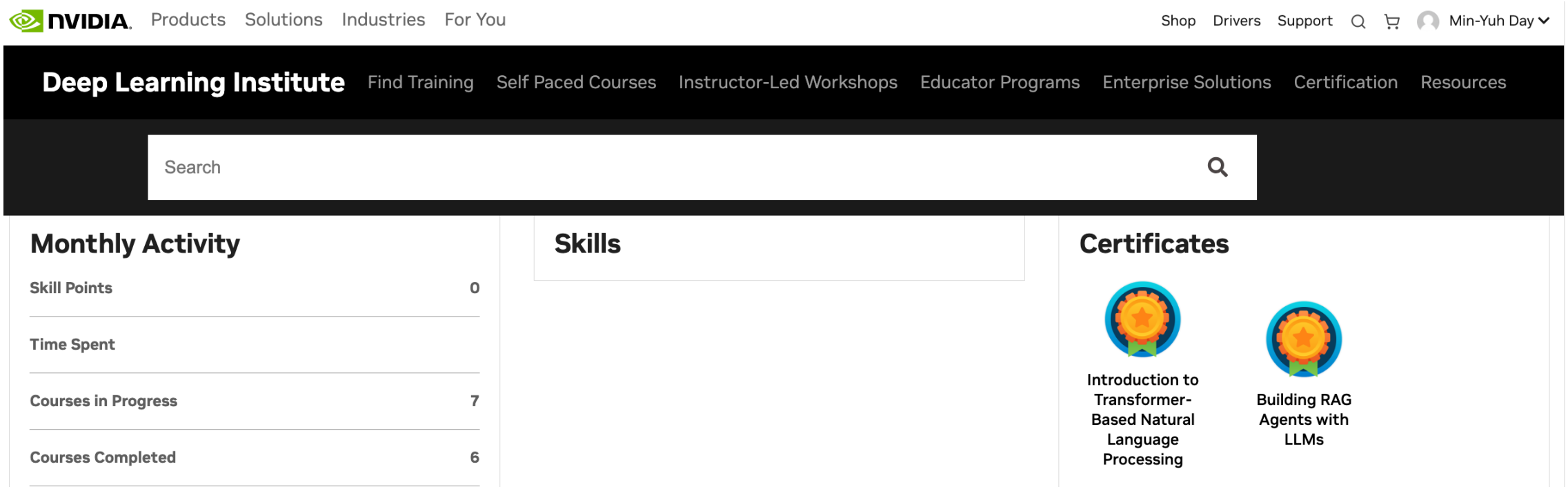
https://learn.nvidia.com/my-learning

NVIDIA Deep Learning Institute (DLI)



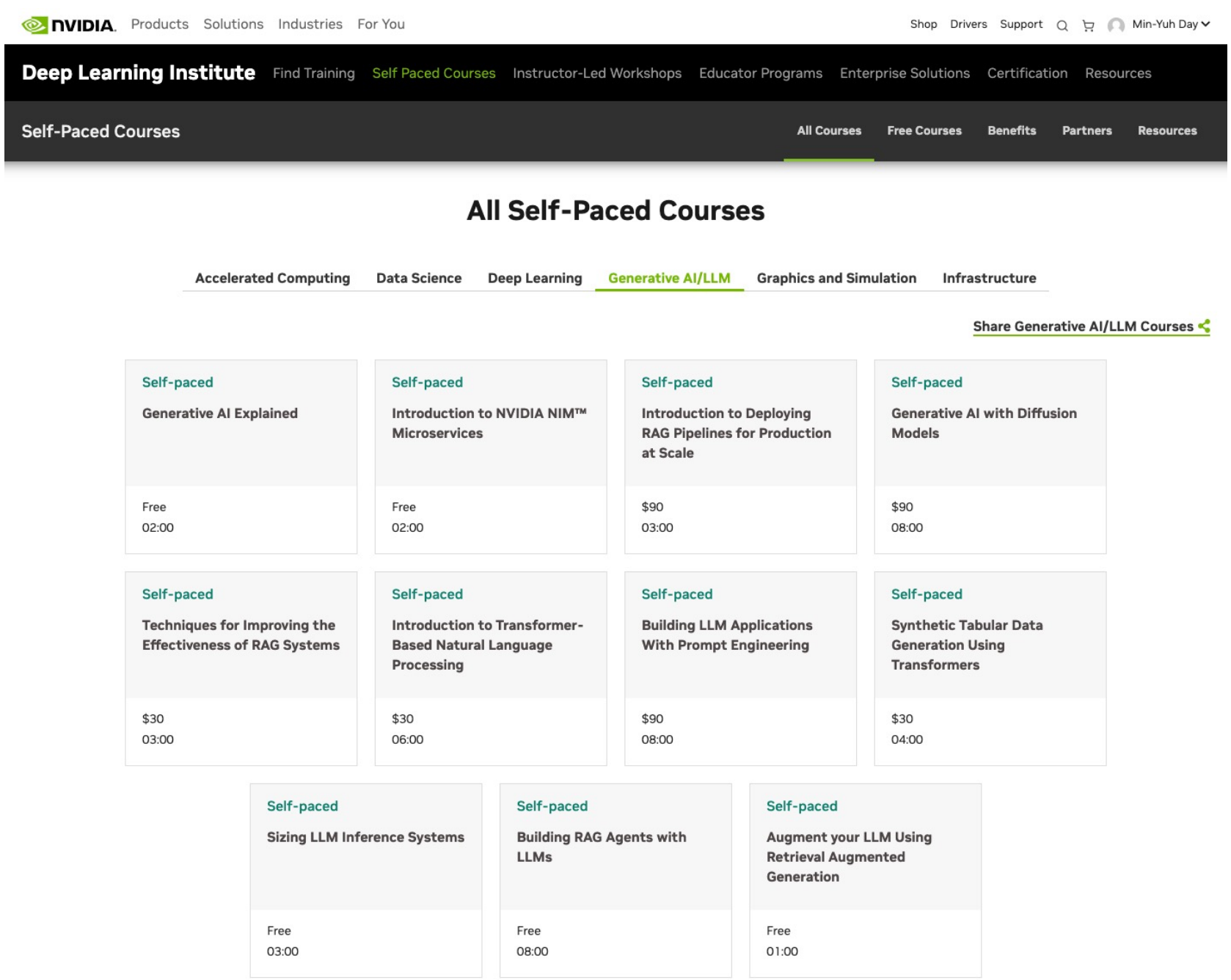
https://learn.nvidia.com/my-learning

Deep Learning Institute (DLI)



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

NVIDIA Deep Learning Institute (DLI)



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

NVIDIA

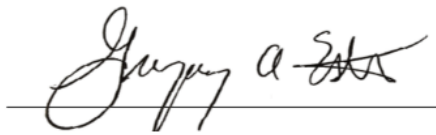
Certificate of Competency

This certificate is awarded to

Min-Yuh Day

for demonstrating competence in the completion of

Introduction to Transformer-Based Natural Language Processing



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

NVIDIA

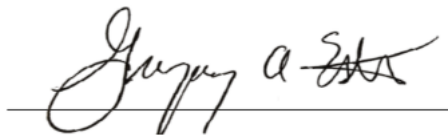
Certificate of Completion

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

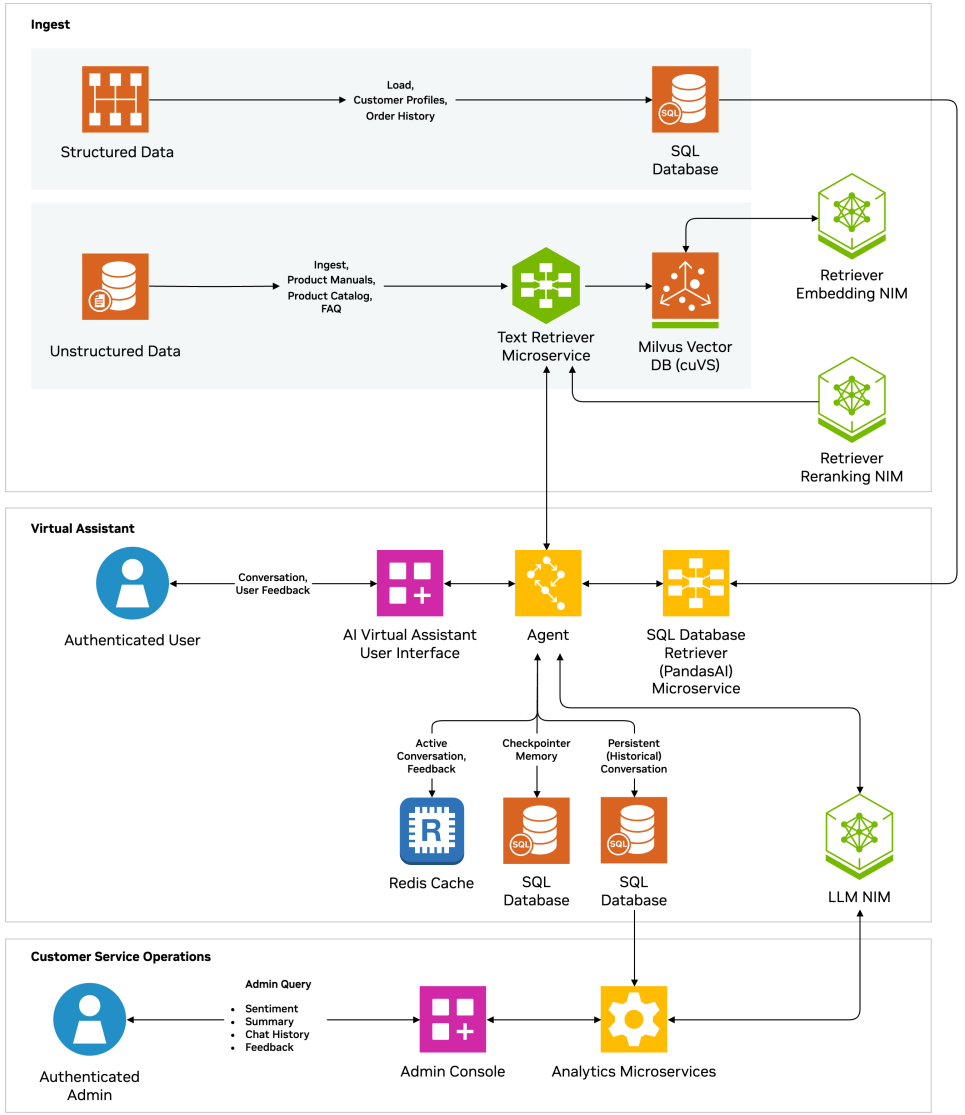
https://learn.nvidia.com/certificates?id=ed-qOCIMQaatzU8SNUNxgw

AI Virtual Assistant for Customer Service

AI Virtual Assistant to reduce handling time, boost customer satisfaction

Benefits

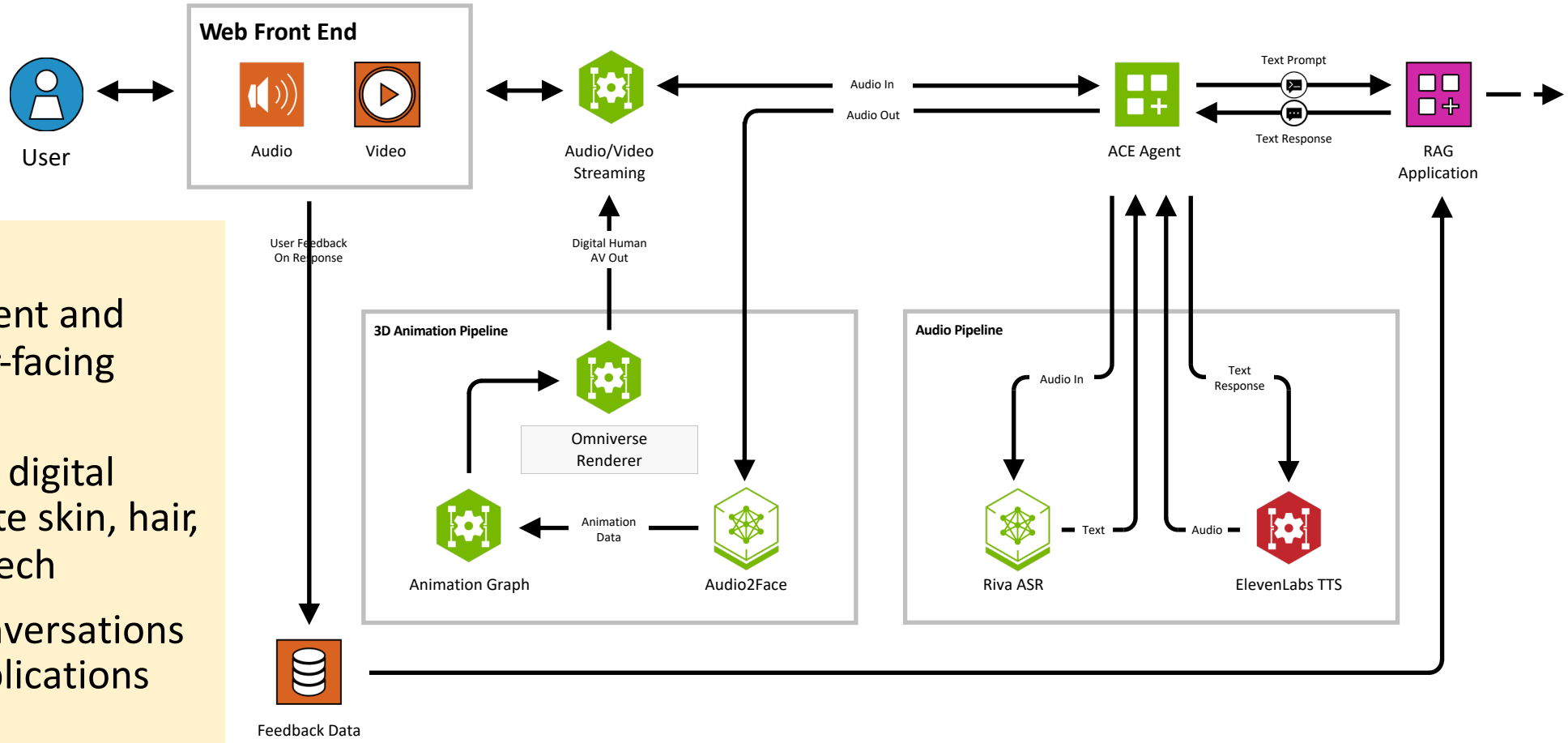
- Personalized Responses: Handles structured and unstructured customer queries (e.g., order details, spending history).
- Multi-Turn Dialogue: Offers context-aware, seamless interactions across multiple questions.
- Custom Conversation Style: Adapts text responses to reflect corporate branding and tone.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

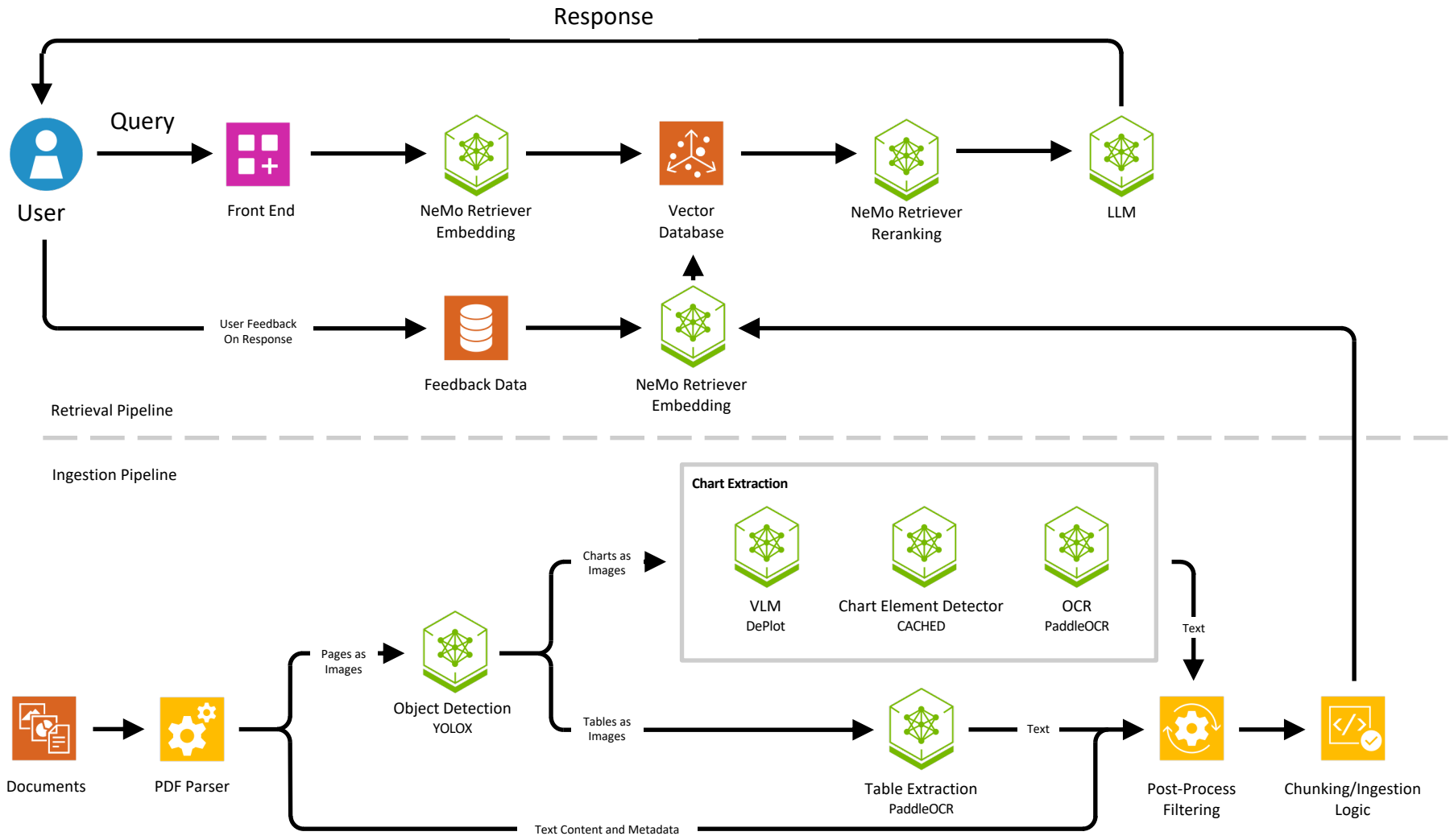
Source: NVIDIA DLI (2024), Introduction to NVIDIA NIM™ Microservices, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-23+V1

Multimodal PDF Data Extraction for Enterprise RAG

Unlocks Knowledge from trillions of PDFs

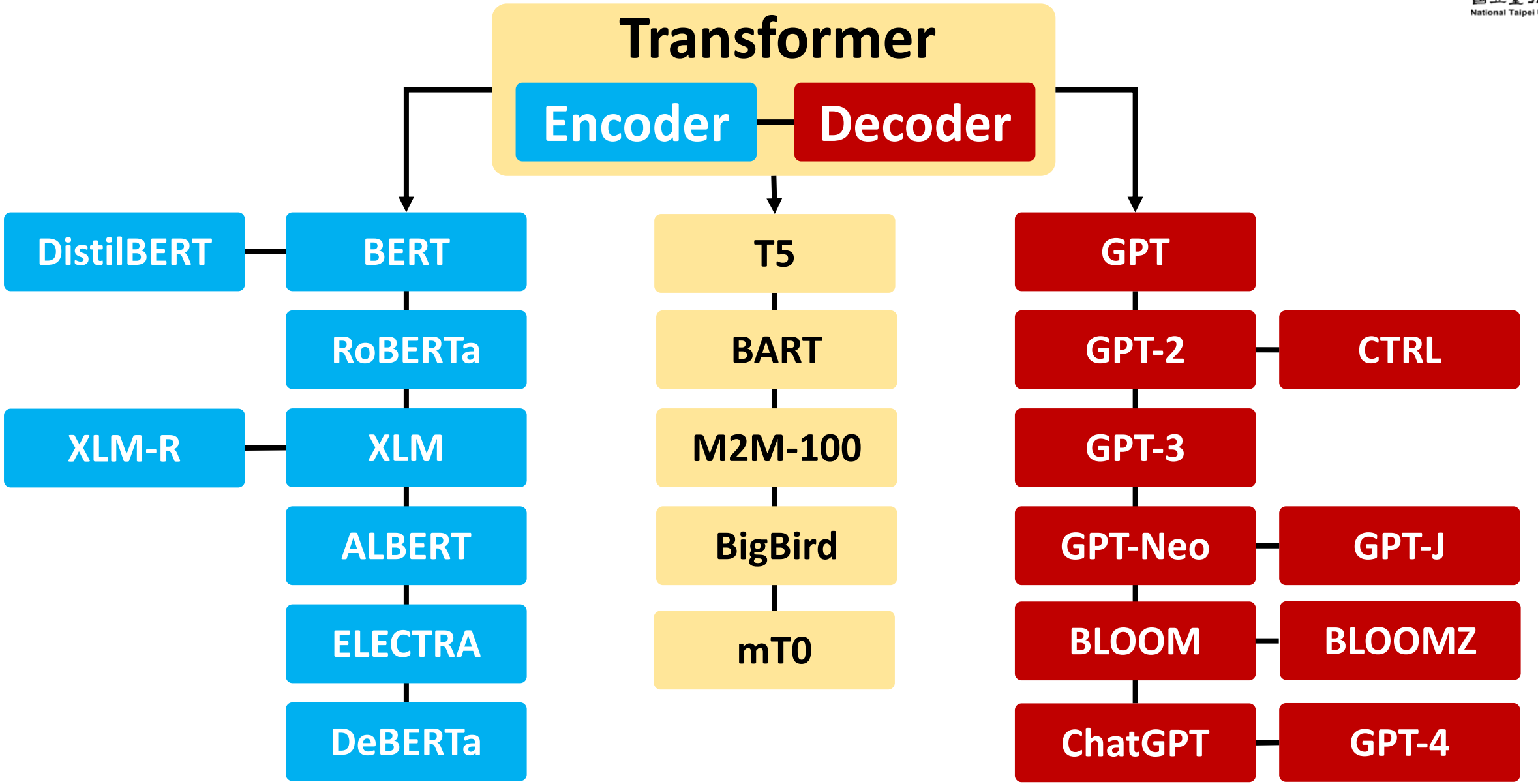
Benefits

- **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

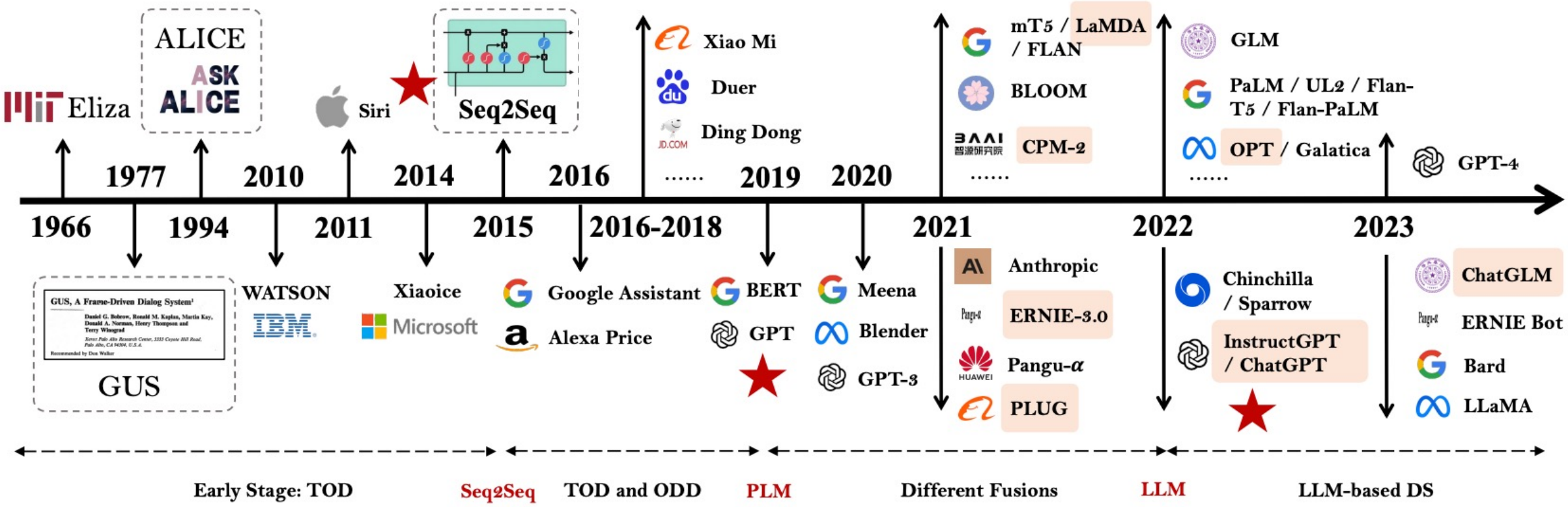
Transformer Models



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

The Development of LM-based Dialogue Systems

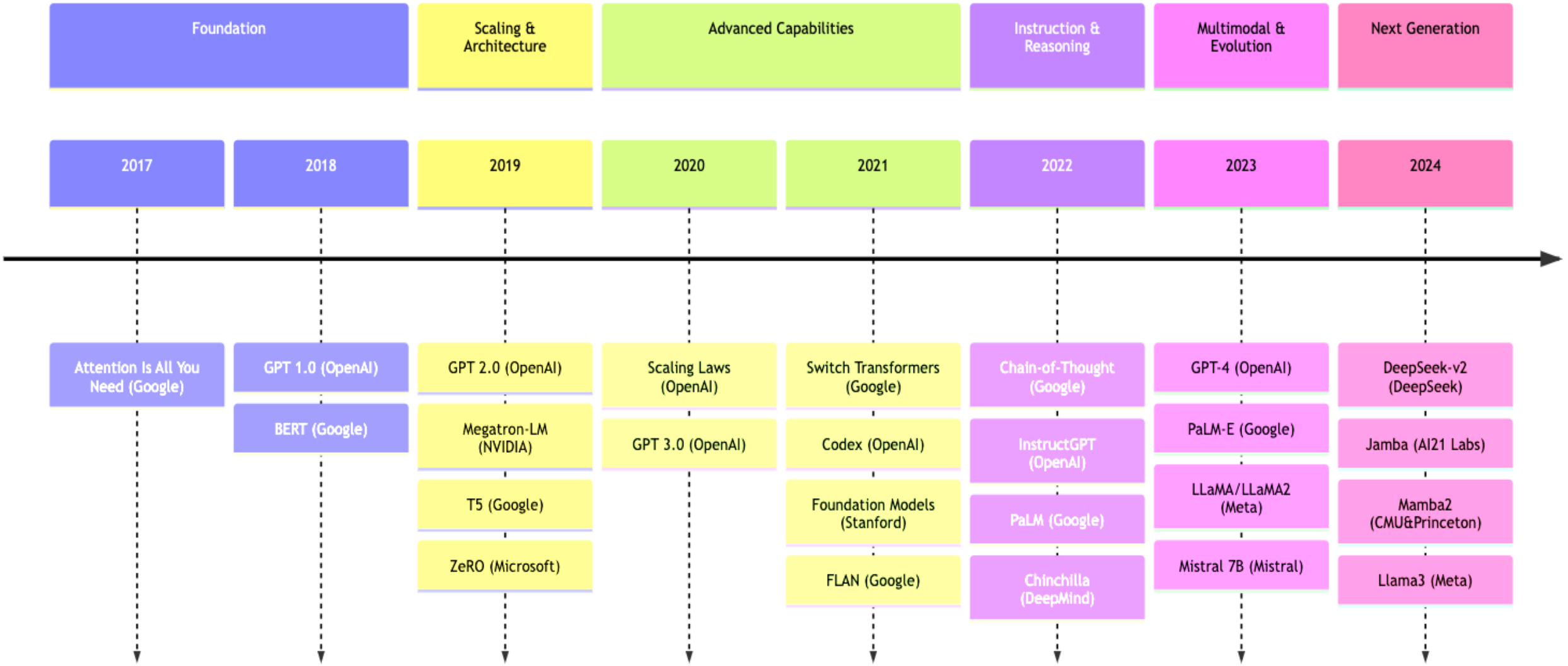
1) Early Stage (1966 - 2015)
2) The Independent Development of TOD and ODD (2015 - 2019)
3) Fusions of Dialogue Systems (2019 - 2022)
4) LLM-based DS (2022 - Now)



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

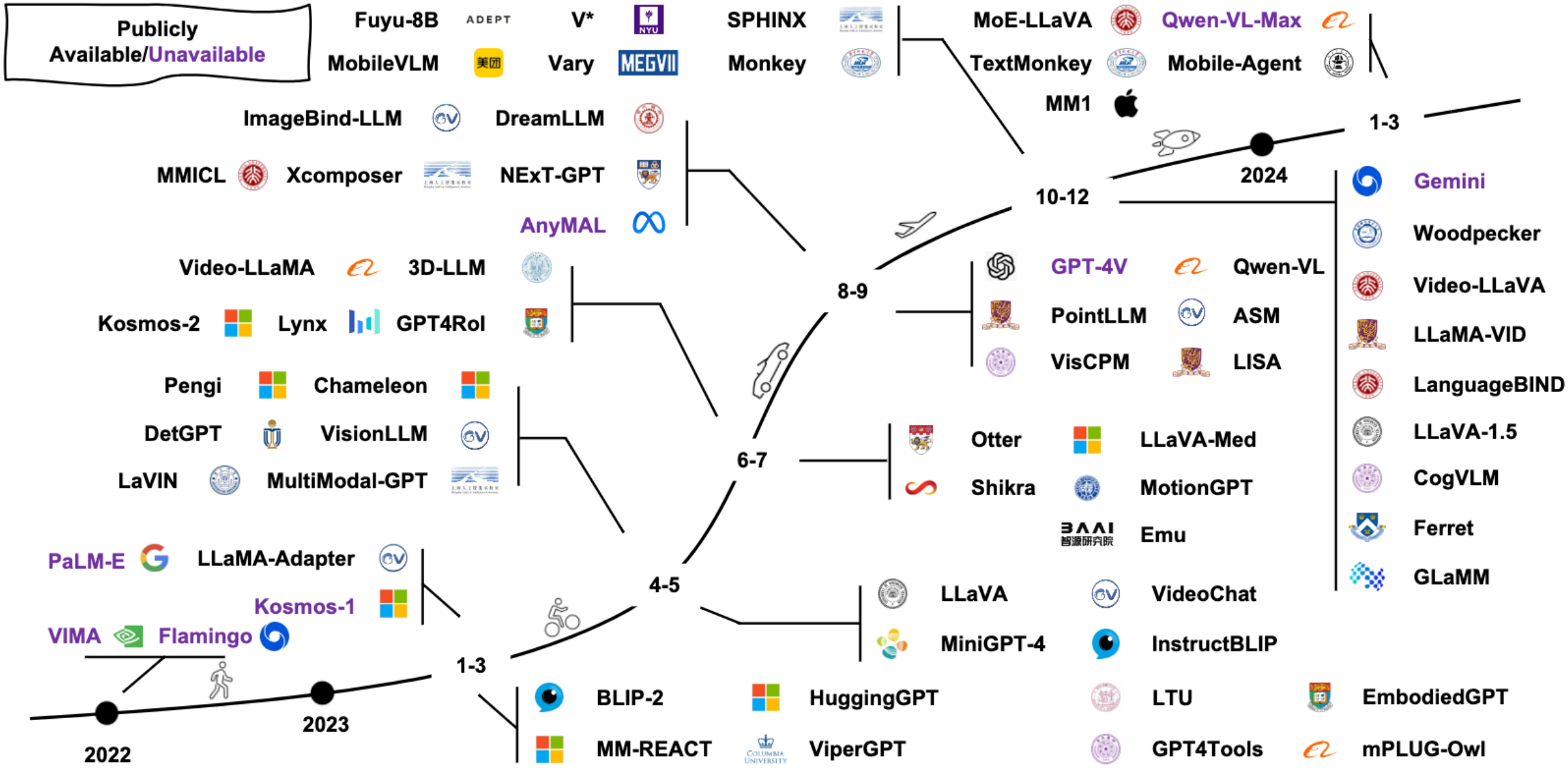
Source: Wang, Hongru, Lingzhi Wang, Yiming Du, Liang Chen, Jingyan Zhou, Yufei Wang, and Kam-Fai Wong. "A Survey of the Evolution of Language Model-Based Dialogue Systems." arXiv preprint arXiv:2311.16789 (2023).

Major GenAI LLMs Research Milestones (2017-2024)



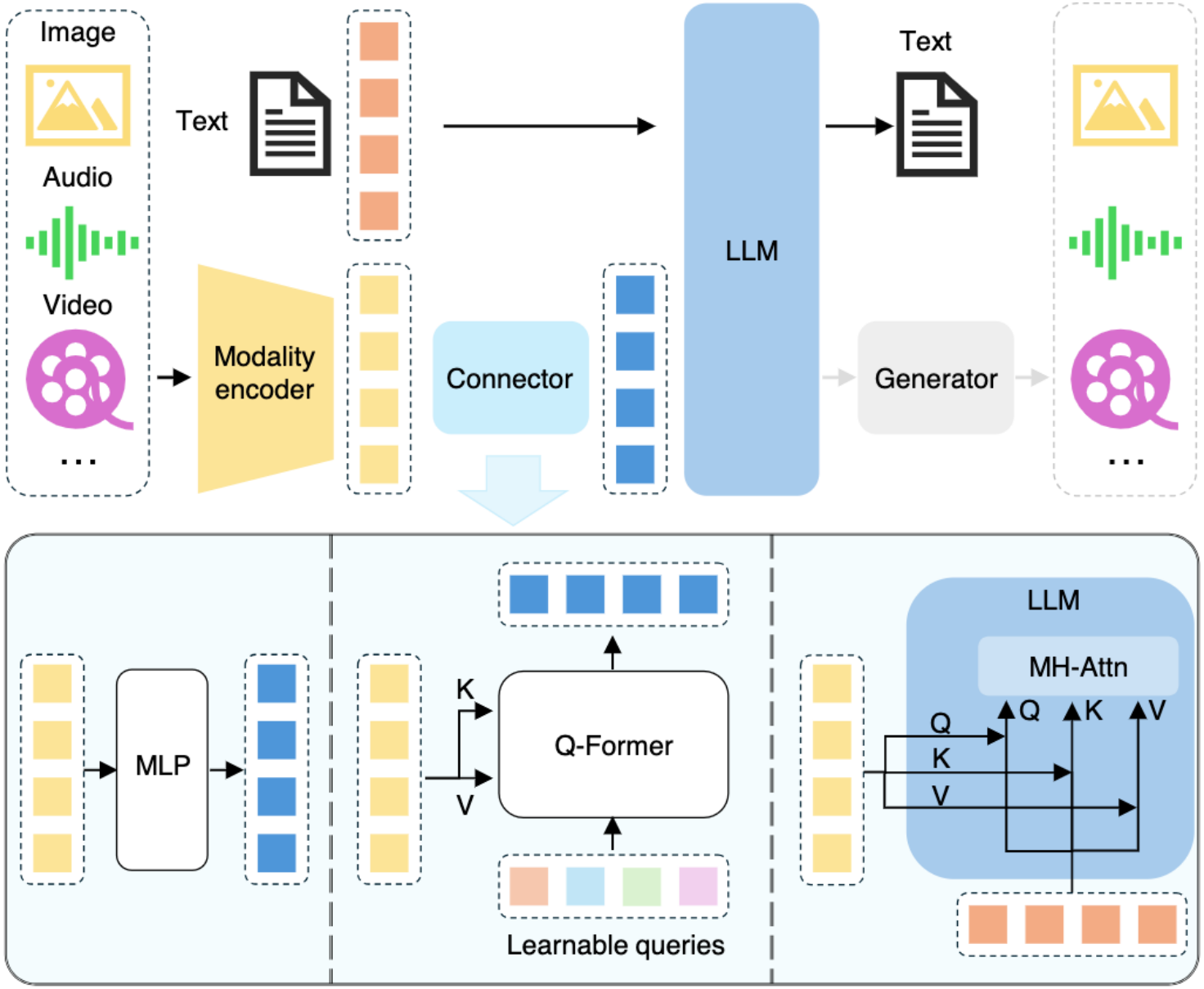
Source: https://github.com/Hannibal046/Awesome-LLM

Multimodal Large Language Models (MLLM)



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.

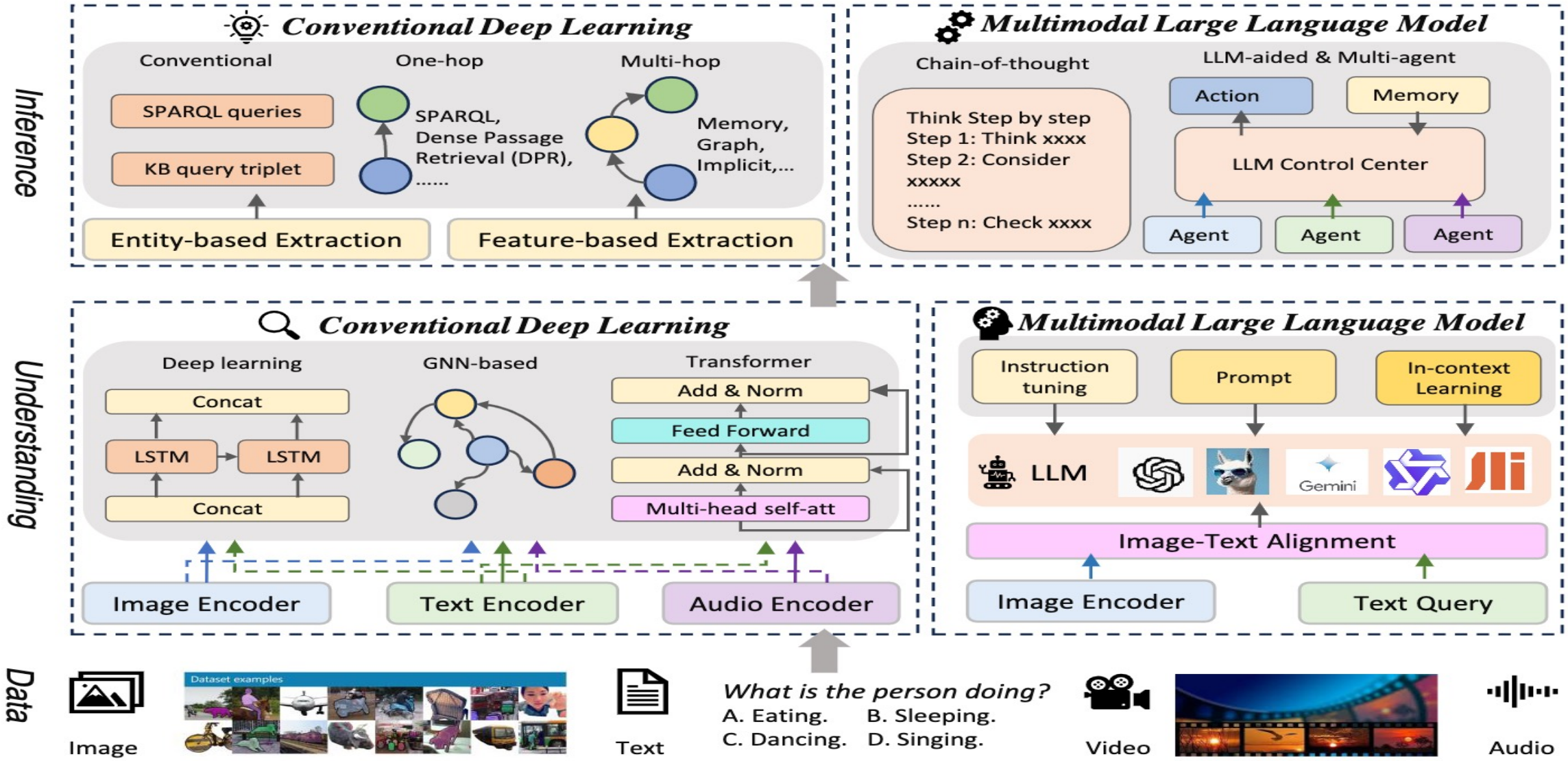
Multimodal Large Language Models (MLLM)



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.

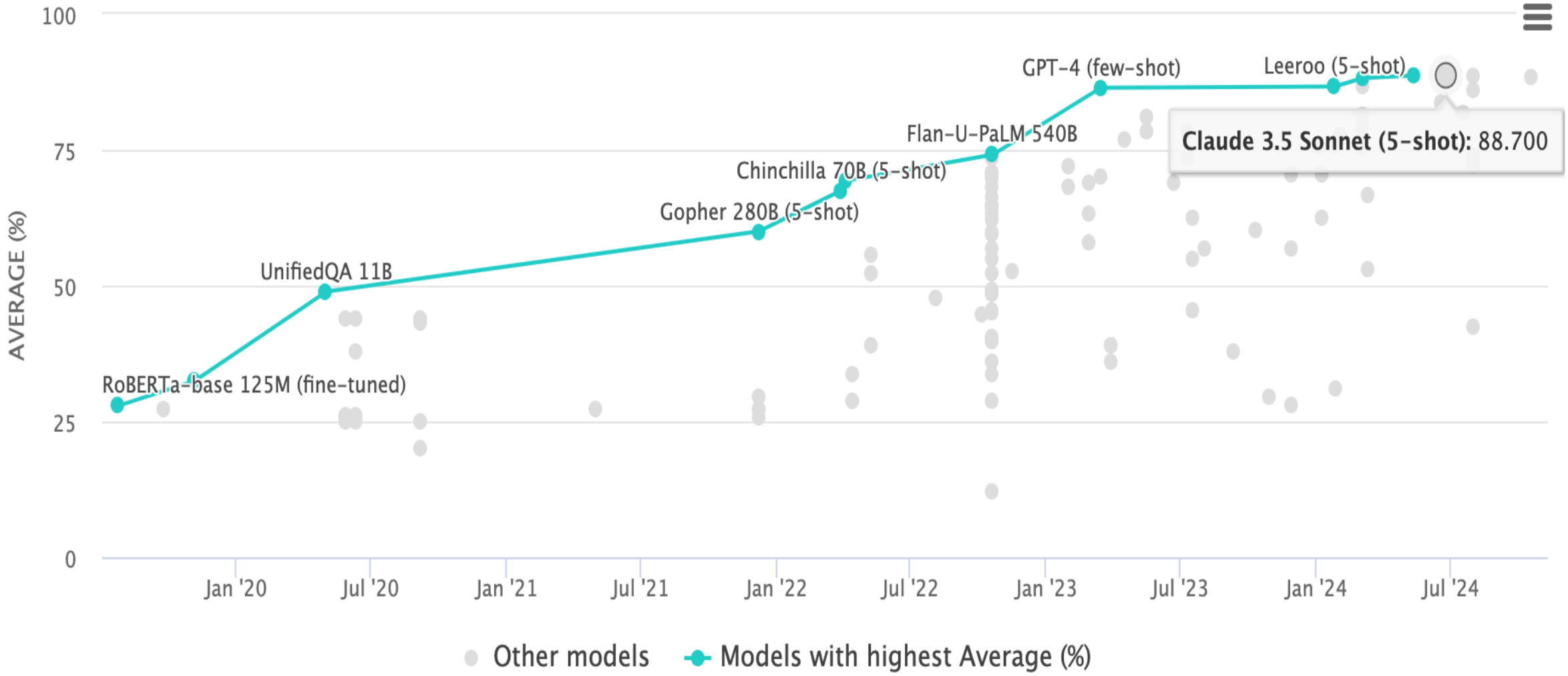
Multimodal Large Language Model (MLLM) for Vision Question Answering



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.

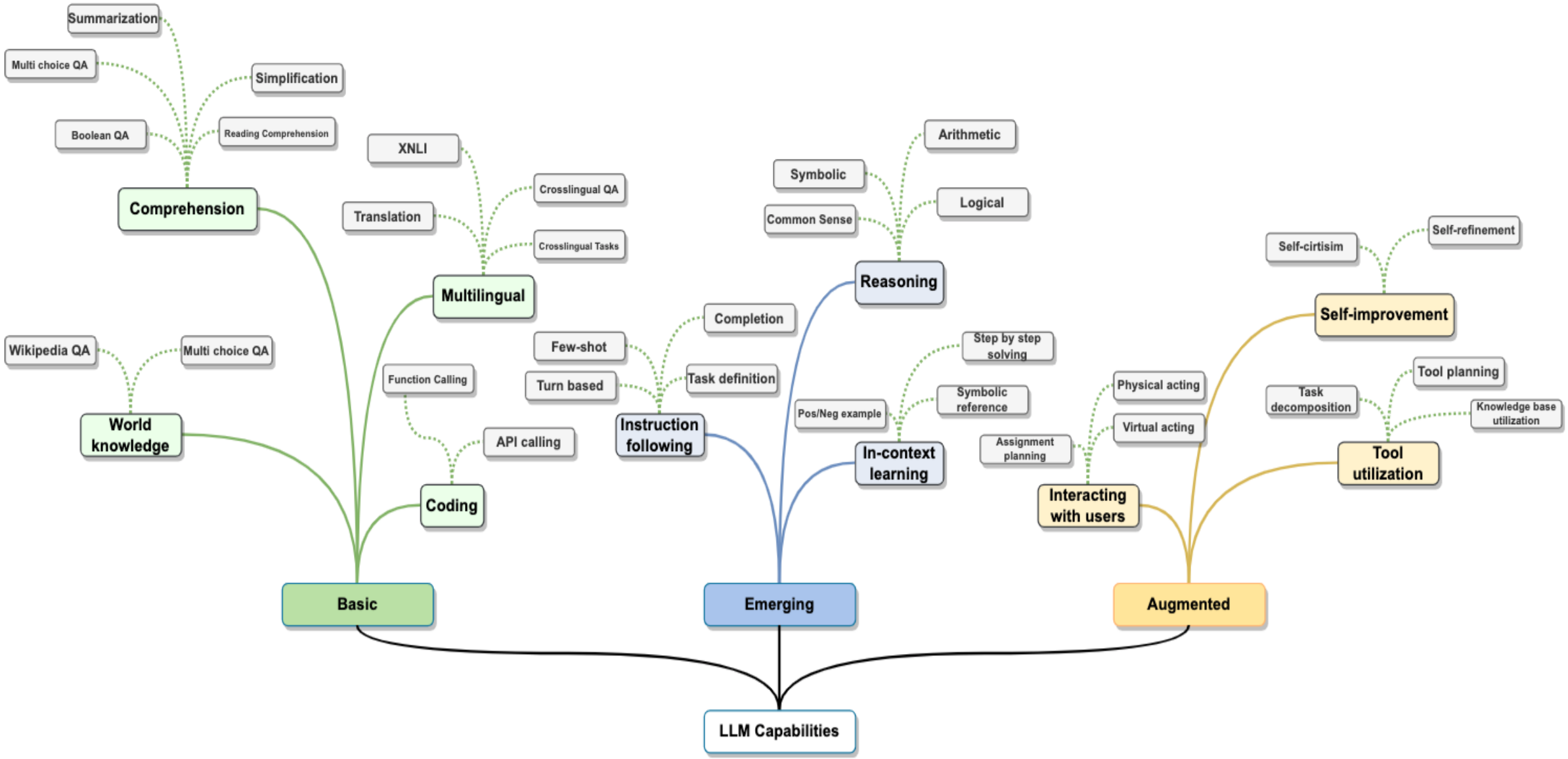
Multi-task Language Understanding on MMLU
GPT-4, Claude 3.5 Sonnet

Massive Multitask Language Understanding (MMLU)



Source: https://paperswithcode.com/sota/multi-task-language-understanding-on-mmlu

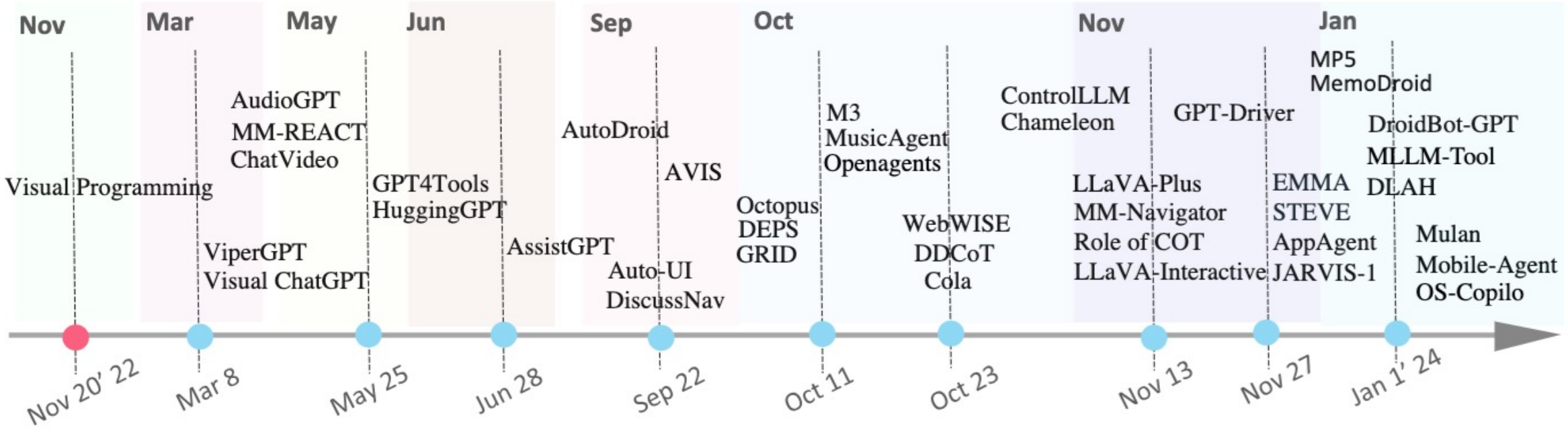
LLM Capabilities



Source: Minaee, Shervin, Tomas Mikolov, Narjes Nikzad, Meysam Chenaghlu, Richard Socher, Xavier Amatriain, and Jianfeng Gao. (2024) "Large language models: A survey." arXiv preprint arXiv:2402.06196 (2024).

LLM-powered Multimodal Agents

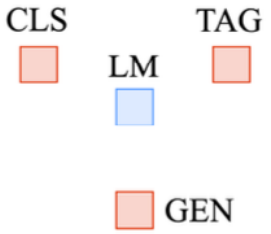
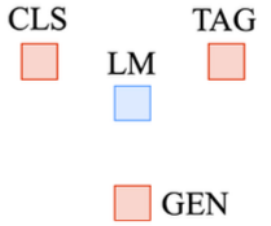
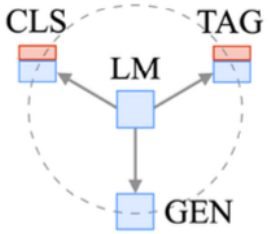
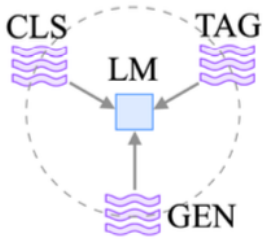
Large Multimodal Agents (LMAs)



2022 2023 2024

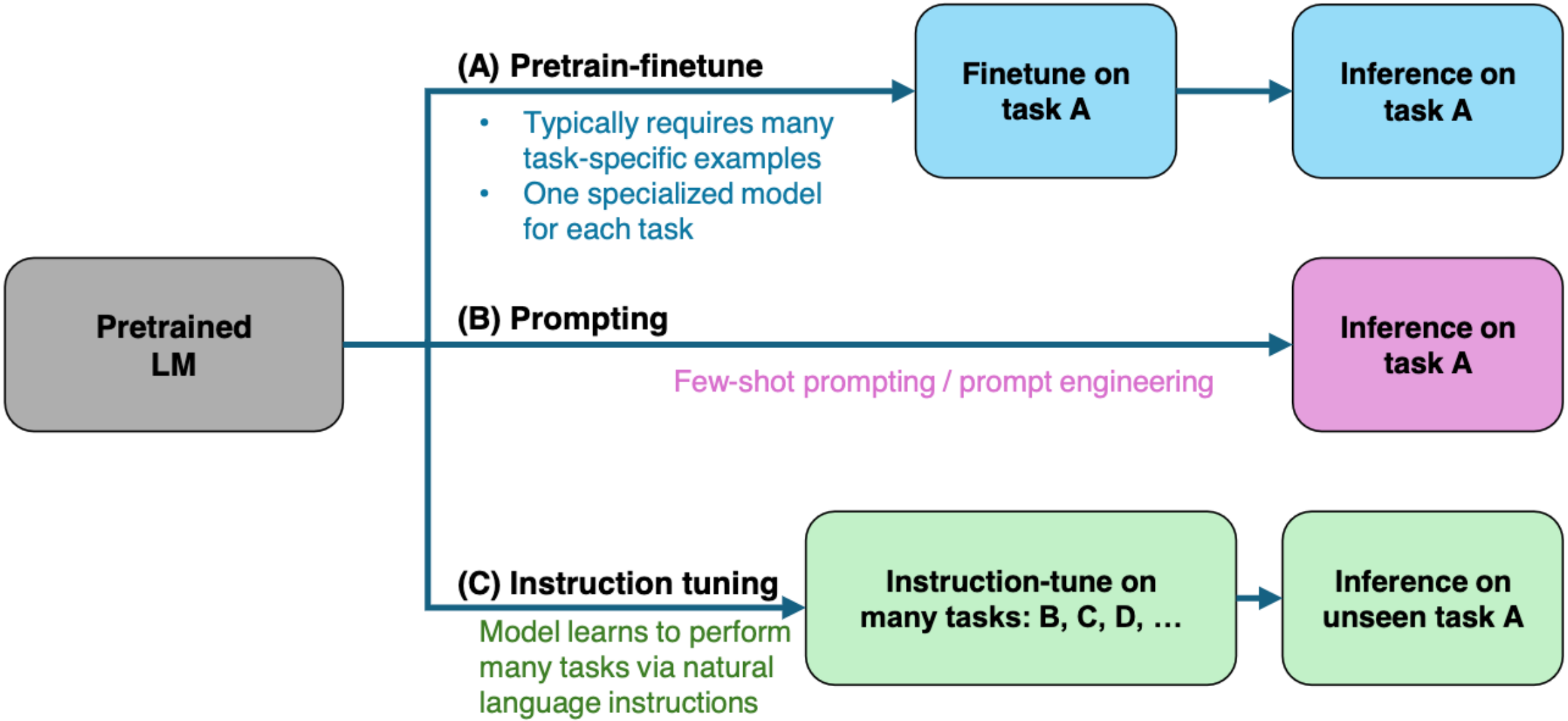
Source: Xie, Junlin, Zhihong Chen, Ruifei Zhang, Xiang Wan, and Guanbin Li. "Large Multimodal Agents: A Survey." arXiv preprint arXiv:2402.15116 (2024).

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.

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)
- Chat.LMSys.org (lmarena.ai)
- Perplexity.ai
- Stable Diffusion
- Video: D-ID, Synthesia
- Audio: Speechify

LMSYS Chatbot Arena Leaderboard

GPT-4o

Claude 3.5

Rank* (UB)	Rank (StyleCtrl)	Model	Arena Score	95% CI	Votes	Organization	License
1	1	Gemini-Exp-1206	1379	+10/-5	5052	Google	Proprietary
2	1	ChatGPT-4o-latest (2024-11-20)	1366	+4/-5	21929	OpenAI	Proprietary
2	4	Gemini-Exp-1121	1364	+4/-5	15004	Google	Proprietary
4	3	o1-preview	1334	+5/-4	30448	OpenAI	Proprietary
5	7	o1-mini	1307	+4/-3	37176	OpenAI	Proprietary
5	6	Gemini-1.5-Pro-002	1302	+5/-3	32758	Google	Proprietary
7	10	Grok-2-08-13	1289	+3/-3	55616	xAI	Proprietary
7	11	Yi-Lightning	1287	+4/-4	29193	01 AI	Proprietary
7	6	GPT-4o-2024-05-13	1285	+3/-3	116858	OpenAI	Proprietary
7	4	Claude 3.5 Sonnet (20241022)	1283	+3/-3	34846	Anthropic	Proprietary
8	17	Athene-v2-Chat-72B	1278	+6/-6	8700	NexusFlow	NexusFlow
11	16	GLM-4-Plus	1274	+4/-5	28006	Zhipu AI	Proprietary
11	18	GPT-4o-mini-2024-07-18	1273	+3/-3	54539	OpenAI	Proprietary
11	21	Gemini-1.5-Flash-002	1271	+4/-4	26276	Google	Proprietary
11	28	Llama-3.1-Nemotron-70B-Instruct	1269	+6/-6	7676	Nvidia	Llama 3.1

https://lmarena.ai/

Claude 3.5 Sonnet State-of-the-art vision

	Claude 3.5 Sonnet (new)	Claude 3.5 Haiku	Claude 3.5 Sonnet	GPT-4o*	GPT-4o mini*	Gemini 1.5 Pro	Gemini 1.5 Flash
Graduate level reasoning *GPQA (Diamond)*	65.0% 0-shot CoT	41.6% 0-shot CoT	59.4% 0-shot CoT	53.6% 0-shot CoT	40.2% 0-shot CoT	59.1% 0-shot CoT	51.0% 0-shot CoT
Undergraduate level knowledge *MMLU Pro*	78.0% 0-shot CoT	65.0% 0-shot CoT	75.1% 0-shot CoT	—	—	75.8% 0-shot CoT	67.3% 0-shot CoT
Code *HumanEval*	93.7% 0-shot	88.1% 0-shot	92.0% 0-shot	90.2% 0-shot	87.2% 0-shot	—	—
Math problem-solving *MATH*	78.3% 0-shot CoT	69.2% 0-shot CoT	71.1% 0-shot CoT	76.6% 0-shot CoT	70.2% 0-shot CoT	86.5% 4-shot CoT	77.9% 4-shot CoT
High school math competition *AIME 2024*	16.0% 0-shot CoT	5.3% 0-shot CoT	9.6% 0-shot CoT	9.3% 0-shot CoT	—	—	—
Visual Q/A *MMMU*	70.4% 0-shot CoT	—	68.3% 0-shot CoT	69.1% 0-shot CoT	59.4% 0-shot CoT	65.9% 0-shot CoT	62.3% 0-shot CoT
Agentic coding *SWE-bench Verified*	49.0%	40.6%	33.4%	—	—	—	—
Agentic tool use *TAU-bench*	Retail 69.2% Airline 46.0%	Retail 51.0% Airline 22.8%	Retail 62.6% Airline 36.0%	—	—	—	—

* Our evaluation tables exclude OpenAI's o1 model family as they depend on extensive pre-response computation time, unlike typical models. This fundamental difference makes performance comparisons difficult.

Source: https://www.anthropic.com/news/3-5-models-and-computer-use

Llama 3.2 90B vision LLMs

Modality	Category / Benchmark	Llama 3.2 11B	Llama 3.2 90B	Claude 3 – Haiku	GPT-4o-mini
Image	College-level Problems and Mathematical Reasoning MMMU (val, 0-shot CoT, micro avg accuracy)	50.7	**60.3**	50.2	59.4
	MMMU-Pro, Standard (10 opts, test)	33.0	**45.2**	27.3	42.3
	MMMU-Pro, Vision (test)	23.7	33.8	20.1	**36.5**
	MathVista (testmini)	51.5	**57.3**	46.4	56.7
	Charts and Diagram Understanding ChartQA (test, 0-shot CoT relaxed accuracy)*	83.4	85.5	81.7	—
	AI2 Diagram (test)*	91.1	92.3	86.7	—
	DocVQA (test, ANLS)*	88.4	90.1	88.8	—
	General Visual Question Answering VQAv2 (test)	75.2	78.1	—	—
Text	General MMLU (0-shot, CoT)	73.0	**86.0**	75.2 (5-shot)	82.0
	Math MATH (0-shot, CoT)	51.9	68.0	38.9	**70.2**
	Reasoning GPQA (0-shot, CoT)	32.8	**46.7**	33.3	40.2
	Multilingual MGSM (0-shot, CoT)	68.9	86.9	75.1	**87.0**

Source: https://ai.meta.com/blog/llama-3-2-connect-2024-vision-edge-mobile-devices/

Llama 3.3 70B instruction-tuned

Category Benchmark	Llama 3.1 70B	Llama 3.3 70B	Amazon Nova Pro	Llama 3.1 405B	Gemini Pro 1.5	GPT-4o	Claude 3.5 Sonnet
General MMLU Chat (0-shot, CoT)	86.0	**86.0**	85.9	**88.6**	87.1	87.5	**88.9**
MMLU PRO (5-shot, CoT)	66.4	**68.9**	-	**73.4**	76.1	73.8	**77.8**
Instruction Following IFEval	87.5	**92.1**	**92.1**	**88.6**	81.9	84.6	**89.3**
Code HumanEval (0-shot)	80.5	**88.4**	**89.0**	**89.0**	89.0	86.0	**93.7**
MBPP EvalPlus (base) (0-shot)	86.0	**87.6**	-	**88.6**	87.8	83.9	86.8
Math MATH (0-sho, CoT)	67.8	**77.0**	76.6	**73.9**	**82.9**	76.9	78.3
Reasoning GPQA Diamond (0-shot, CoT)	48.0	**50.5**	-	**49.0**	53.5	47.5	**65.0**
Tool use BFCL v2 (0-shot)	77.5	**77.3**	-	**81.1**	80.3	74.0	79.3
Long context NIH/Multi-needle	97.5	**97.5**	-	**98.1**	94.7	-	**99.4**
Multilingual Multilingual MGSM (0-shot)	86.9	**91.1**	-	**91.6**	89.6	90.6	**92.8**
Pricing* 1M Input tokens (Cheapest among providers)*	$0.1	**$0.1**	$0.80	**$1.0**	$1.30	2.5$	$3.0
1M Output tokens (Cheapest among providers)*	$0.4	**$0.4**	$3.20	**$1.8**	$5.0	10.0$	$15.0

https://www.llama.com/

Mistral Pixtral Large (124B)

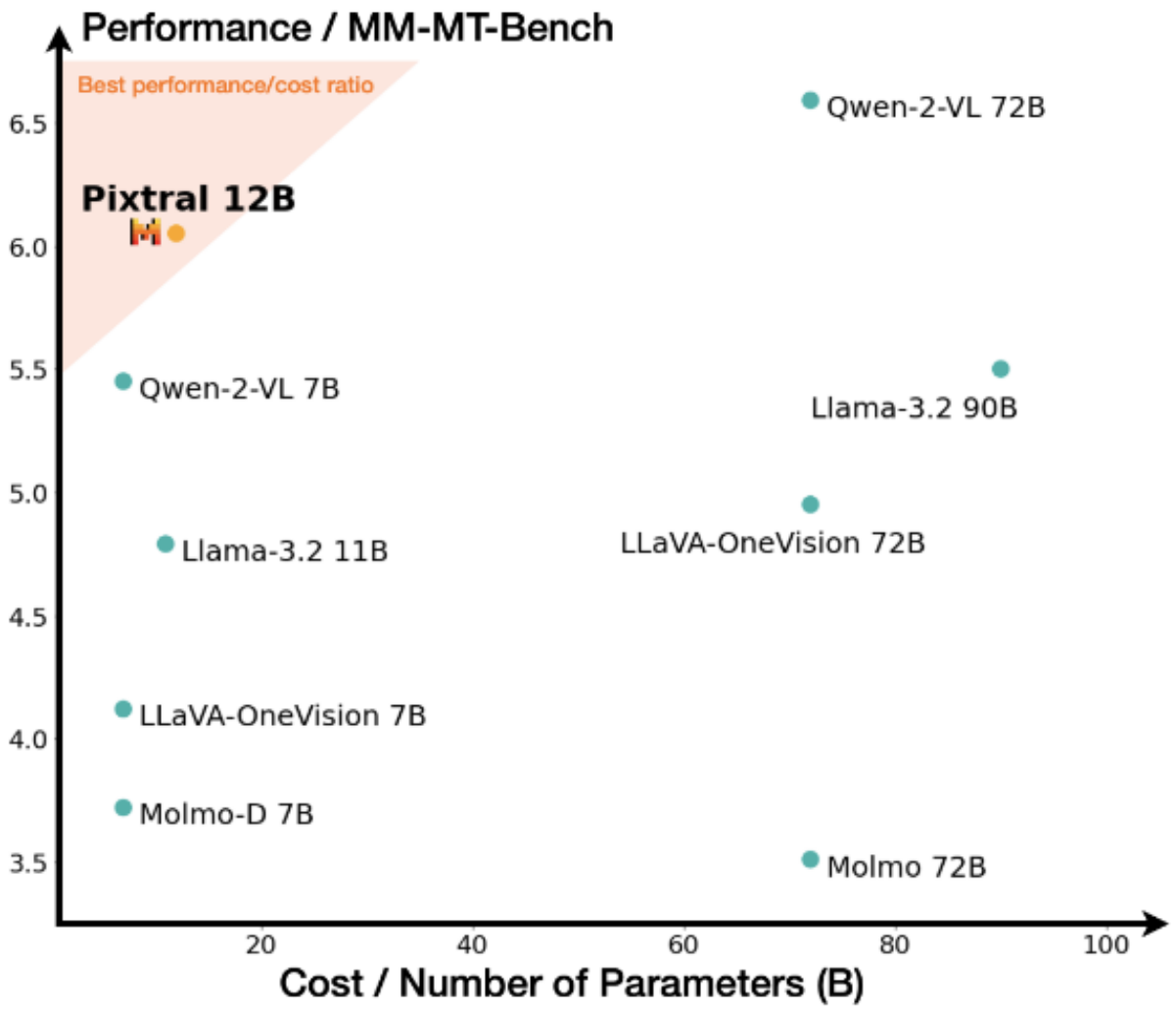
Frontier-class multimodal performance

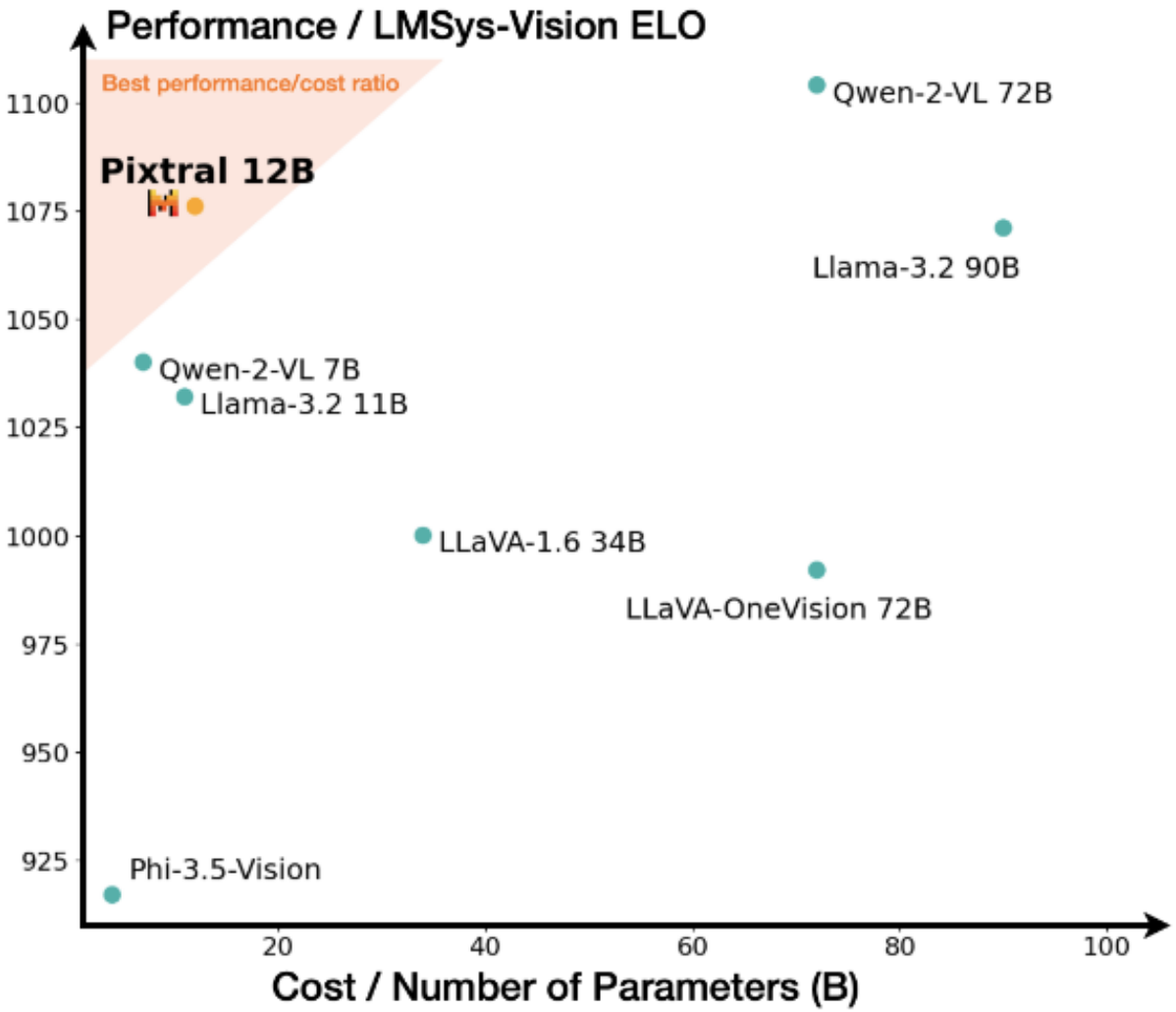
	Model	Mathvista (CoT)	MMMU (CoT)	ChartQA (CoT)	DocVQA (ANLS)	VQAv2 (VQA Match)	AI2D (BBox)	MM MT-Bench
Open Weights	Pixtral Large (124B)	**69.4**	64.0	88.1	**93.3**	**80.9**	93.8	**7.4**
	Llama-3.2 90B (measured)	49.1	53.7	70.8	85.7	67.0	-	5.5
	Llama-3.2 90B (reported)	57.3	60.3	85.8	90.1	80.2	92.3	-
Closed	Gemini-1.5 Pro (measured)	67.8	66.3	83.8	92.3	70.6	**94.6**	6.8
	Gemini-1.5 Pro (reported)	68.1	65.9	-	-	-	-	-
	GPT-4o (measured)	65.4	**68.6**	85.2	88.5	76.4	93.2	6.7
	GPT-4o (reported)	63.8	69.1	85.7	92.8	-	-	-
	Claude-3.5 Sonnet (measured)	67.1	68.4	**89.1**	88.6	69.5	76.9	7.3
	Claude-3.5 Sonnet (reported)	70.7	70.4	90.8	94.2	-	95.3	-
Unreleased	Llama-3.1 505B (reported)	-	64.5	85.8	92.6	80.2	94.1	-
	Grok-2 (reported)	69.0	66.1	-	93.6	-	-	-

Specific model versions evaluated: Claude-3.5 Sonnet (new) [Oct 24], Gemini-1.5 Pro (002) [Sep 24], GPT-4o (2024-08-06) [Aug 24]

Source: https://mistral.ai/news/pixtral-large/

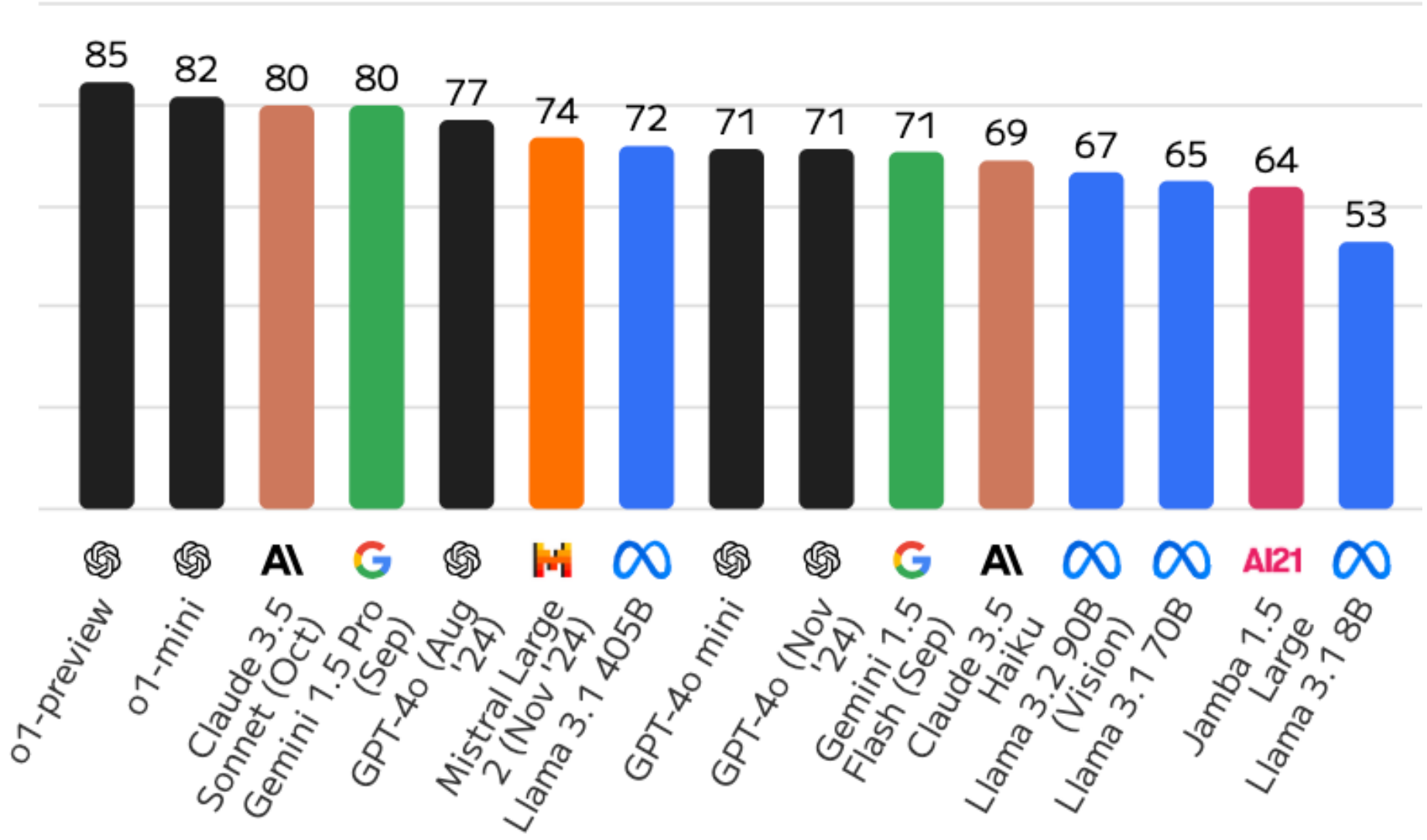
Mistral Pixtral 12B





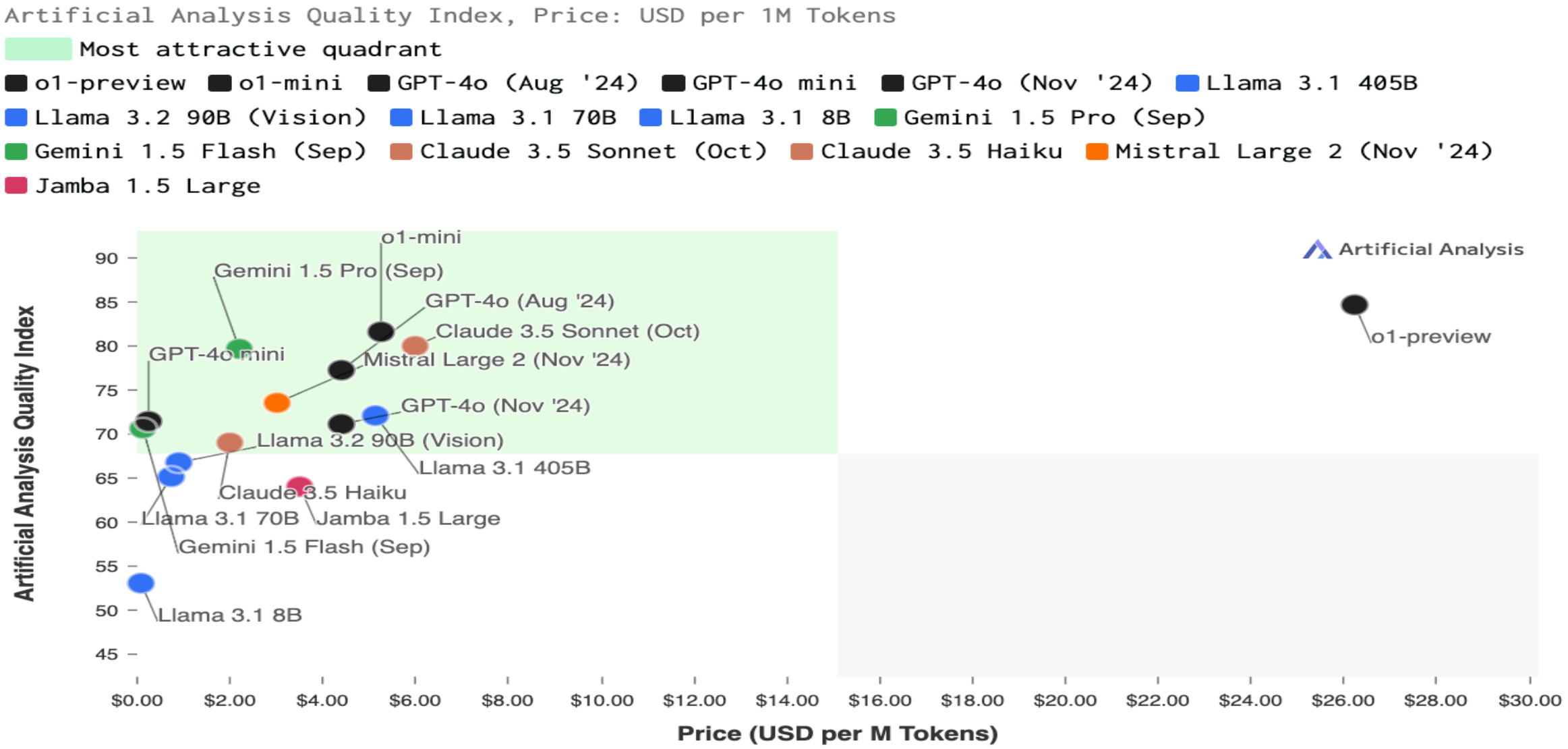
Source: Agrawal, Pravesh, Szymon Antoniak, Emma Bou Hanna, Baptiste Bout, Devendra Chaplot, Jessica Chudnovsky, Diogo Costa et al. (2024) "Pixtral 12B." arXiv preprint arXiv:2410.07073.

Large Language Models (LLMs)
Artificial Analysis Quality Index



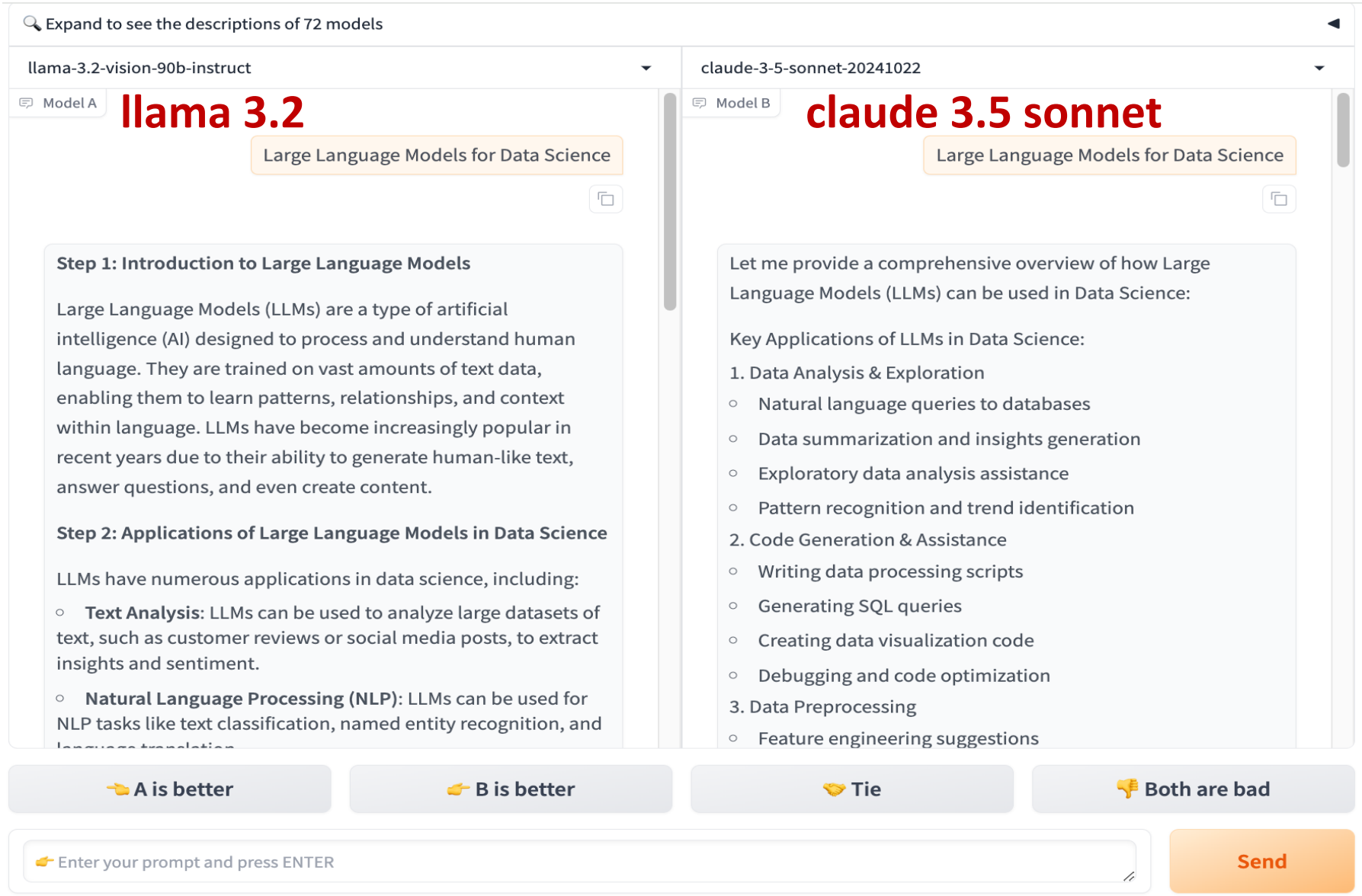
Source: https://artificialanalysis.ai/

Large Language Models (LLMs)
Quality vs. Price



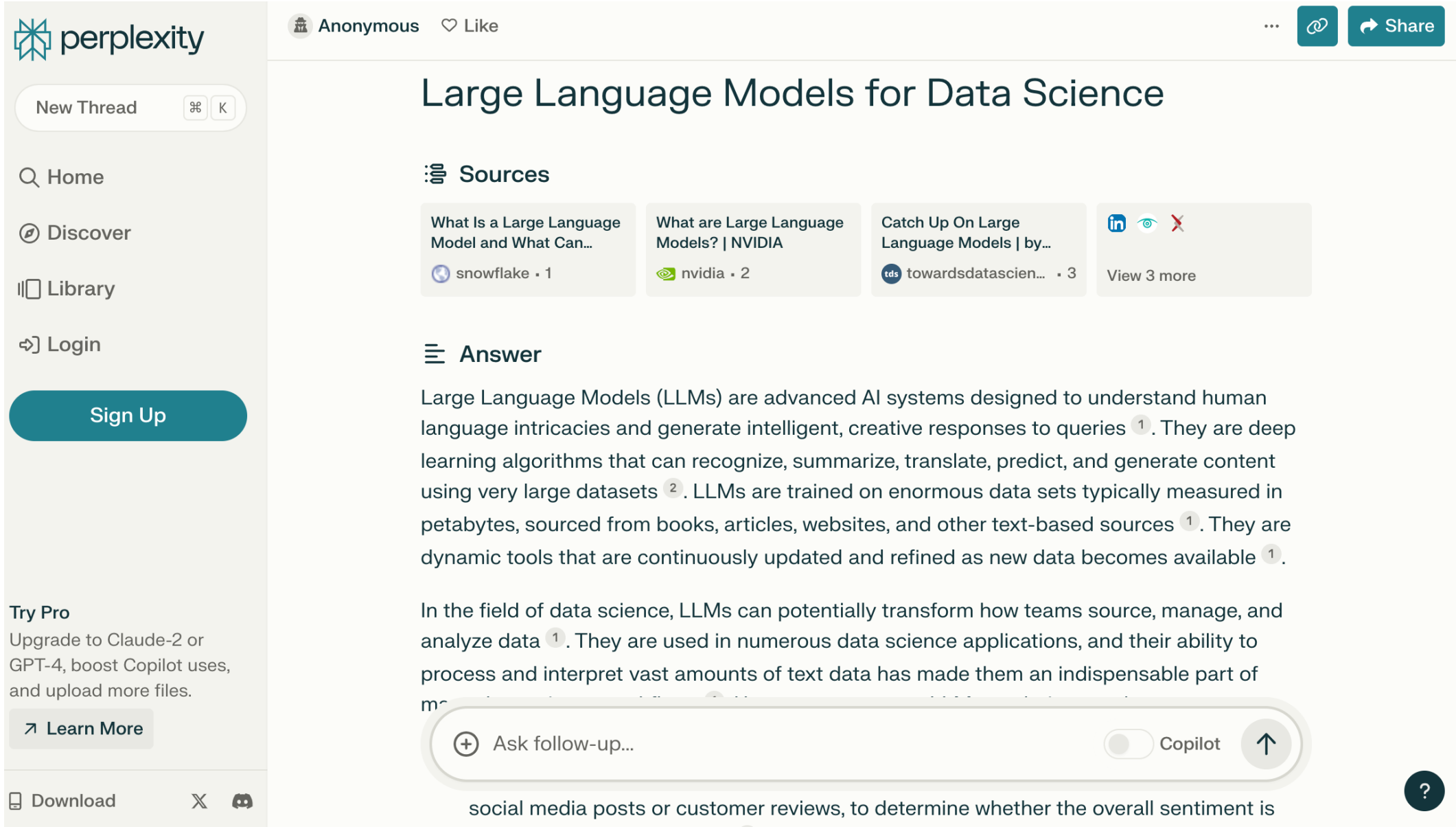
Source: https://artificialanalysis.ai/

Chat with Open Large Language Models: Chatbot Arena



https://lmarena.ai/

Perplexity.ai



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

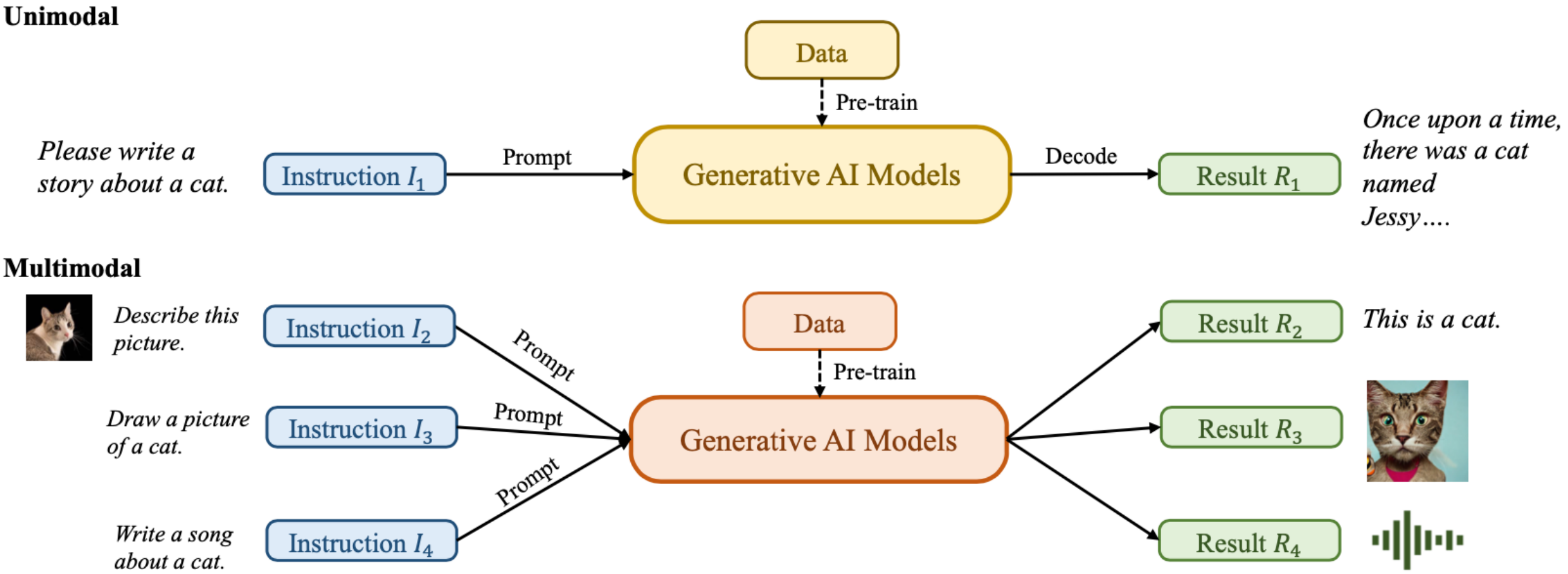
Figure 2

OpenAI DALL·E 2

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.

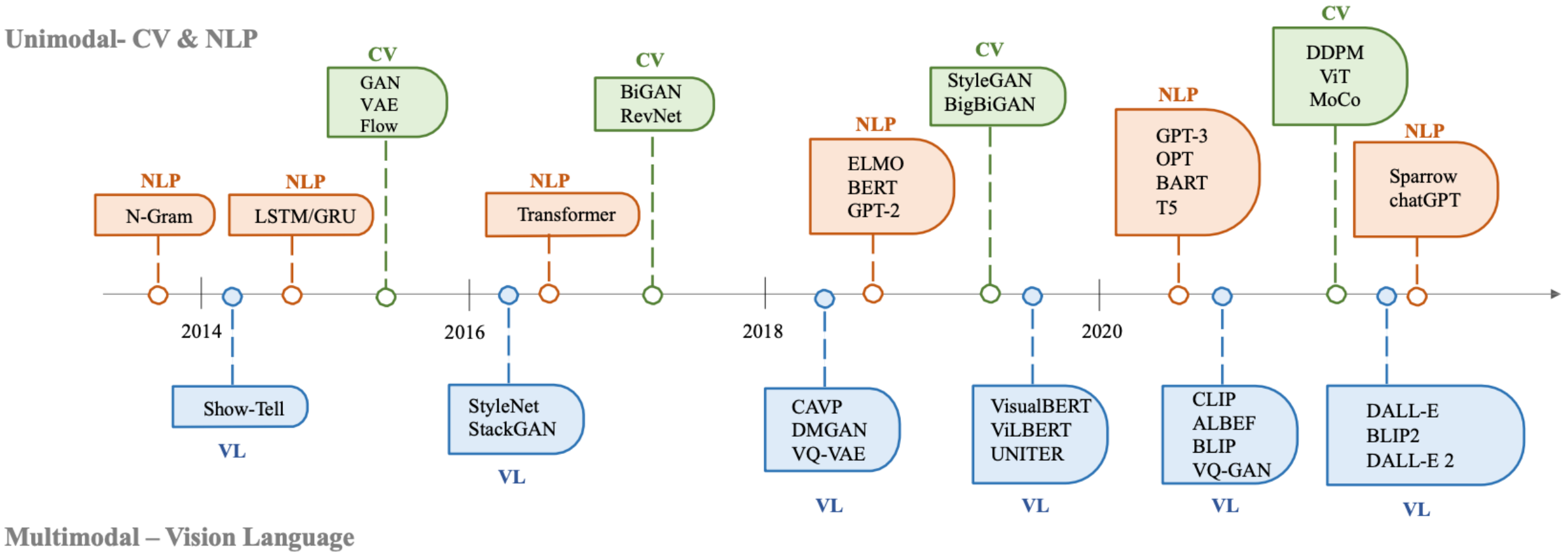
Generative AI (Gen AI)

AI Generated Content (AIGC)



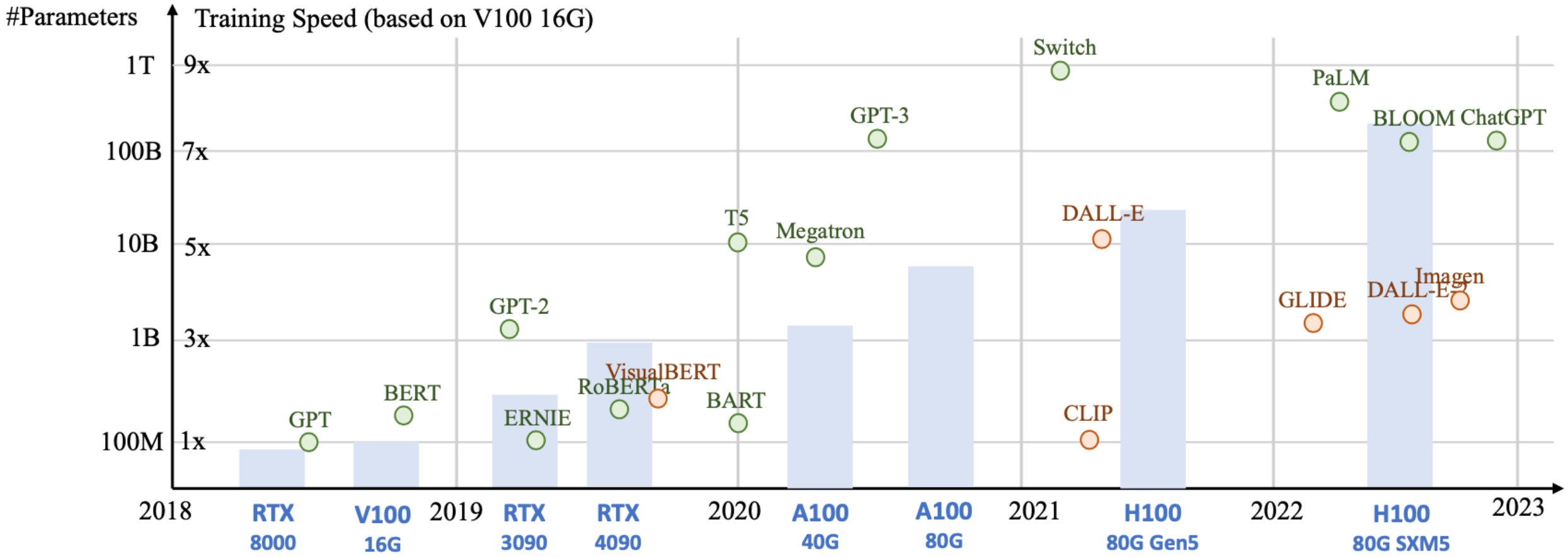
Source: Yihan Cao, Siyu Li, Yixin Liu, Zhiling Yan, Yutong Dai, Philip S. Yu, and Lichao Sun (2023). "A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT." arXiv preprint arXiv:2303.04226.

The history of Generative AI in CV, NLP and VL



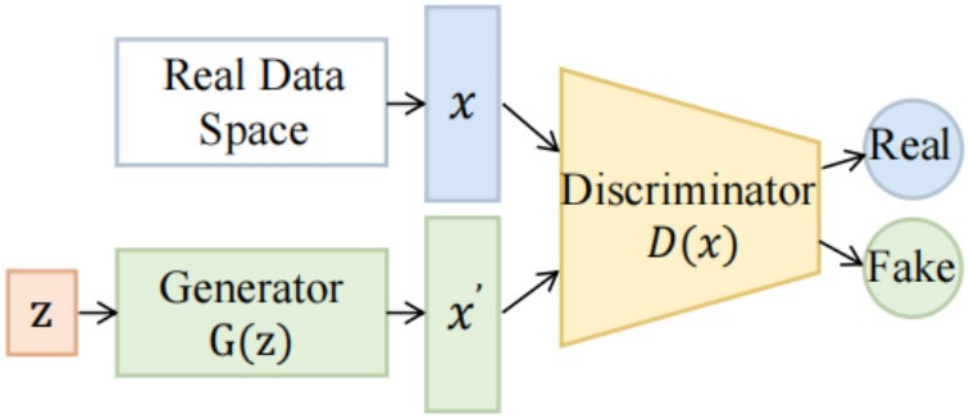
Source: Yihan Cao, Siyu Li, Yixin Liu, Zhiling Yan, Yutong Dai, Philip S. Yu, and Lichao Sun (2023). "A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT." arXiv preprint arXiv:2303.04226.

Generative AI Foundation Models

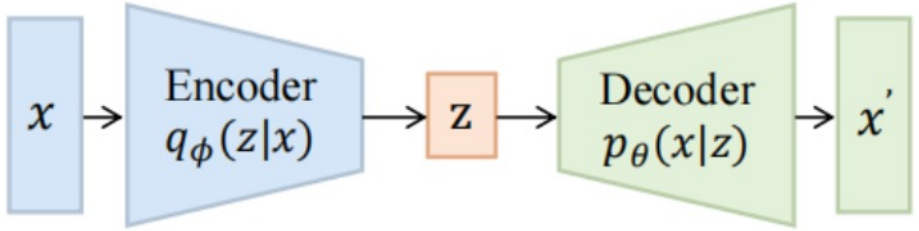


Source: Yihan Cao, Siyu Li, Yixin Liu, Zhiling Yan, Yutong Dai, Philip S. Yu, and Lichao Sun (2023). "A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT." arXiv preprint arXiv:2303.04226.

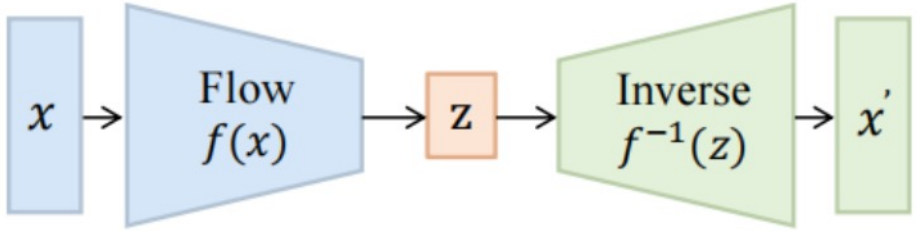
Categories of Vision Generative Models



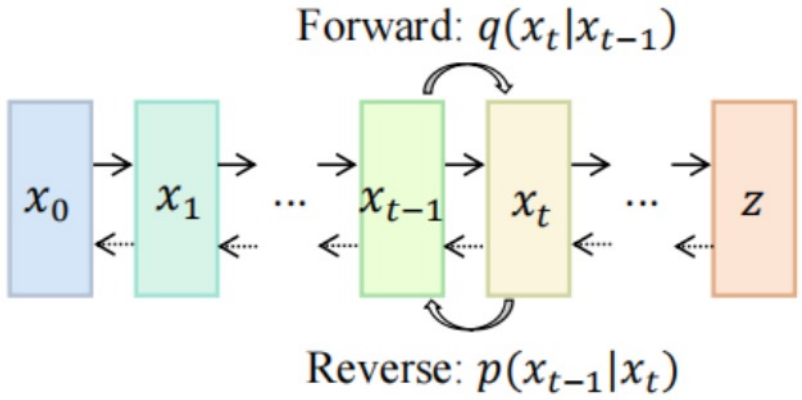
(1) Generative adversarial networks



(2) Variational autoencoders



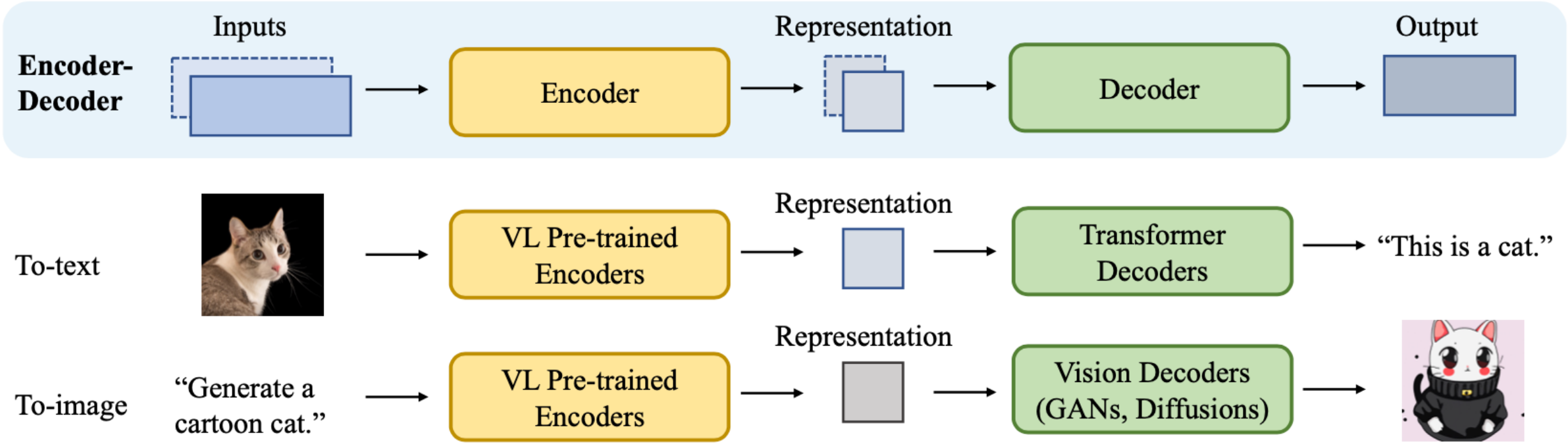
(3) Normalizing flows



(4) Diffusion models

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 General Structure of Generative Vision Language



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.

Philosophy and Ethics of AI

Philosophy, Ethics, and Safety of AI

- The Limits of AI
- Can Machines Really Think?
- The Ethics of AI

Source: Stuart Russell and Peter Norvig (2020), Artificial Intelligence: A Modern Approach, 4th Edition, Pearson

Philosophy of AI

- Philosophers use the term
  - **weak AI** for the hypothesis that machines could possibly behave intelligently
  - **strong AI** for the hypothesis that such machines would count as having actual minds (as opposed to simulated minds)

Source: Stuart Russell and Peter Norvig (2020), Artificial Intelligence: A Modern Approach, 4th Edition, Pearson

4 Approaches of AI

Thinking Humanly	**Thinking Rationally**
Acting Humanly	**Acting Rationally**

Source: Stuart Russell and Peter Norvig (2020), Artificial Intelligence: A Modern Approach, 4th Edition, Pearson

4 Approaches of AI

2. **Thinking Humanly:** **The Cognitive Modeling Approach**	**3.** **Thinking Rationally:** **The "Laws of Thought" Approach**
1. **Acting Humanly:** **The Turing Test Approach (1950)**	**4.** **Acting Rationally:** **The Rational Agent Approach**

Source: Stuart Russell and Peter Norvig (2020), Artificial Intelligence: A Modern Approach, 4th Edition, Pearson

AI Acting Humanly:
The Turing Test Approach
(Alan Turing, 1950)

- Knowledge Representation
- Automated Reasoning
- Machine Learning (ML)
  - Deep Learning (DL)
- Computer Vision (Image, Video)
- Natural Language Processing (NLP)
- Robotics

Source: Stuart Russell and Peter Norvig (2020), Artificial Intelligence: A Modern Approach, 4th Edition, Pearson

Can machines think?

- Alan Turing rejected the question “Can machines think?” and replaced it with a behavioral test.
  - Alan Turing anticipated many objections to the possibility of thinking machines.
- Concentrate on their systems’ performance on practical tasks
  - rather than the ability to imitate humans.
- Consciousness remains a mystery.

Source: Stuart Russell and Peter Norvig (2020), Artificial Intelligence: A Modern Approach, 4th Edition, Pearson

The Ethics of AI

- **Given that AI is a powerful technology, we have a moral obligation to use it well, to promote the positive aspects and avoid or mitigate the negative ones.**

Source: Stuart Russell and Peter Norvig (2020), Artificial Intelligence: A Modern Approach, 4th Edition, Pearson

Principles of Robotics and AI

- Ensure safety
- Ensure fairness
- Respect privacy
- Promote collaboration
- Provide transparency
- Limit harmful uses of AI

Source: Stuart Russell and Peter Norvig (2020), Artificial Intelligence: A Modern Approach, 4th Edition, Pearson

Principles of Robotics and AI

- Establish accountability
- Uphold human rights and values
- Reflect diversity/inclusion
- Avoid concentration of power
- Acknowledge legal/policy implications
- Contemplate implications for employment

Source: Stuart Russell and Peter Norvig (2020), Artificial Intelligence: A Modern Approach, 4th Edition, Pearson

Safety of AI

- AI is a powerful technology, and as such it poses potential dangers, through lethal autonomous weapons, security and privacy breaches, unintended side effects, unintentional errors, and malignant misuse.
- Those who work with AI technology have an ethical imperative to responsibly reduce those dangers.

Source: Stuart Russell and Peter Norvig (2020), Artificial Intelligence: A Modern Approach, 4th Edition, Pearson

Robot Ethics

Laws of Robotics (Isaac Asimov, 1942, 1950)

1. **A robot may not injure a human being or, through inaction, allow a human being to come to harm.**
2. **A robot must obey orders given to it by human beings, except where such orders would conflict with the First Law.**
3. **A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.**

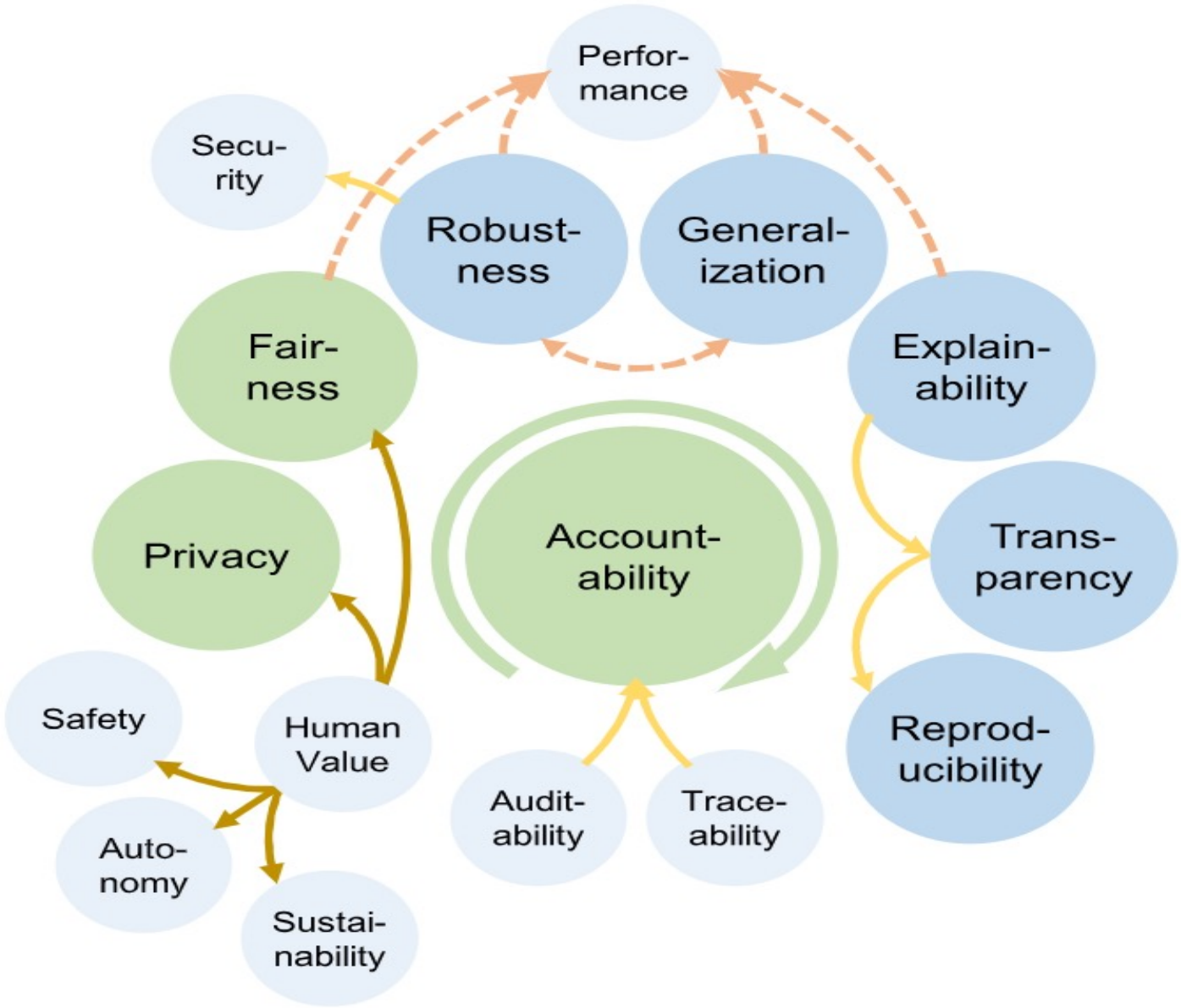
Source: Stuart Russell and Peter Norvig (2020), Artificial Intelligence: A Modern Approach, 4th Edition, Pearson

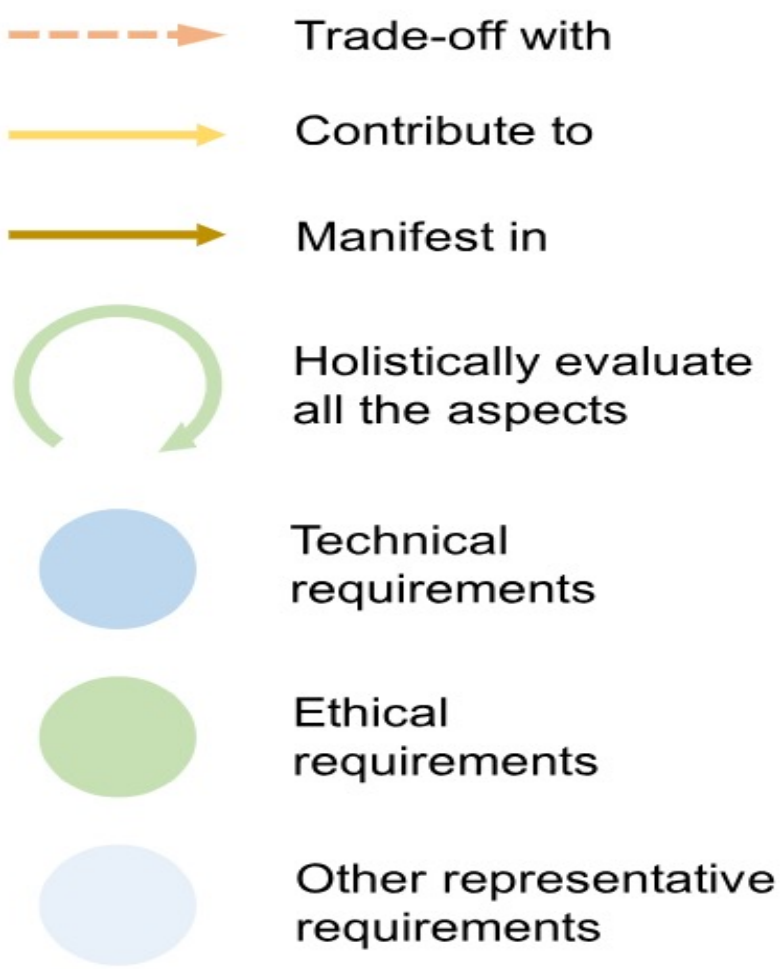
Fair, trustworthy, and transparent of AI

- **AI systems must be able to demonstrate they are fair, trustworthy, and transparent.**
- **There are multiple aspects of fairness, and it is impossible to maximize all of them at once.**
- **So a first step is to decide what counts as fair.**

Source: Stuart Russell and Peter Norvig (2020), Artificial Intelligence: A Modern Approach, 4th Edition, Pearson

Trustworthy AI





Source: Bo Li, Peng Qi, Bo Liu, Shuai Di, Jingen Liu, Jiquan Pei, Jinfeng Yi, and Bowen Zhou. (2023) "Trustworthy AI: From principles to practices." ACM Computing Surveys 55, no. 9 (2023): 1-46.

Explainable AI (XAI)

- When an AI system turns you down for a loan, you deserve an explanation.
  - In Europe, the GDPR enforces this for you.
  - An AI system that can explain itself is called **explainable AI (XAI)**.

Source: Stuart Russell and Peter Norvig (2020), Artificial Intelligence: A Modern Approach, 4th Edition, Pearson

Explainable AI (XAI)

- A good explanation properties of XAI
  - it should be understandable and convincing to the user
  - it should accurately reflect the reasoning of the system
  - it should be complete
  - it should be specific in that different users with different conditions or different outcomes should get different explanations

Source: Stuart Russell and Peter Norvig (2020), Artificial Intelligence: A Modern Approach, 4th Edition, Pearson

Automation

- **Automation is already changing the way people work.**
- **As a society, we will have to deal with these changes.**

Source: Stuart Russell and Peter Norvig (2020), Artificial Intelligence: A Modern Approach, 4th Edition, Pearson

The Future of AI

The Future of AI

- AI Components
- AI Architectures

AI Components

- Sensors and actuators
- Representing the state of the world
- Selecting actions
- Deciding what we want
- Learning
- Resources
  - Shared data
  - Shared model

Source: Stuart Russell and Peter Norvig (2020), Artificial Intelligence: A Modern Approach, 4th Edition, Pearson

Learning

- Deep learning
- Data science
- Big data
- Transfer learning
- Apprenticeship learning
- Differentiable programming
- Weakly supervised learning
- Predictive learning

Source: Stuart Russell and Peter Norvig (2020), Artificial Intelligence: A Modern Approach, 4th Edition, Pearson

AI Architectures

- Which of the agent architectures should an agent use?
  - All of them!
- Real-time AI
- Anytime algorithm
- Decision-theoretic metareasoning
- Reflective architecture
- Agent = Architecture + Program
- Bounded optimality

Real-time AI

- **As AI systems move into more complex domains, all problems will become real-time, because the agent will never have long enough to solve the decision problem exactly.**

Source: Stuart Russell and Peter Norvig (2020), Artificial Intelligence: A Modern Approach, 4th Edition, Pearson

General AI

- Narrow tasks AI
  - DARPA Grand Challenge for autonomous cars
  - ImageNet object recognition competition
  - For each separate task, we build a separate AI system
  - A separate machine learning model trained from scratch with data collected specifically for this task.
- Human-level AI (HLAI)

Source: Stuart Russell and Peter Norvig (2020), Artificial Intelligence: A Modern Approach, 4th Edition, Pearson

AI Engineering

- Powerful tools and frameworks
  - TensorFlow, Keras, PyTorch, CAFFE, Scikit-Learn and SCIPY.
- Promising approaches
  - GANs
  - Deep reinforcement learning
  - Train properly in a new domain

Generative AI: Proprietary & Confidential

- **Valuable data needs to be private.**
- **Need to maintain confidentiality with private data.**
- **Need to consider industry-specific data norms.**

Source: NVIDIA DLI (2024), Generative AI Explained, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-07+V1

Challenges of Generative AI

- IP ownership
  - Does the model have the right to reproduce and extrapolate from this data
- IP attribution
  - Can the output of the model be attributed to the proper sources
  - Important for many businesses
- Harmful application
  - Is the output of the model used to manipulate or mislead
- Harmful outputs
  - Is the output of the model offensive

Source: NVIDIA DLI (2024), Generative AI Explained, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-07+V1

Training Generative AI Efficiently at Scale

- **NVIDIA NeMo enables efficient training of large language models at scale.**
- **Full-stack optimization from networking to compute to frameworks and libraries.**
- **Extraordinary throughput and flexibility.**
- **NVIDIA invests heavily in the most important frameworks: JAX, PyTorch, TensorFlow and more.**

Source: NVIDIA DLI (2024), Generative AI Explained, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-07+V1

The Future of AI

- AI has made great progress in its short history.
- We can see only a short distance ahead, but we can see that much remains to be done.
  (Alan Turing, 1950)
  [Computing Machinery and Intelligence]

Source: Stuart Russell and Peter Norvig (2020), Artificial Intelligence: A Modern Approach, 4th Edition, Pearson

The Future of AI

- **Past: Build each new system from scratch**
- **Future: Start with a single huge system**
  - **For each new task, extract from it the parts that are relevant to the task.**
- **Transformer language models (e.g., BERT, GPT-2, GPT-3, ChatGPT) with billions of parameters**
  - **An "outrageously large" ensemble neural network architecture that scales up to 175 billion parameters in one experiment.**
  - **GPT-3 (175B), BLOOM (176B), PaLM (540B)**

Source: Stuart Russell and Peter Norvig (2020), Artificial Intelligence: A Modern Approach, 4th Edition, Pearson

AI for Social Good (AI4SG)

Source: Nenad Tomašev, Julien Cornebise, Frank Hutter, Shakir Mohamed, Angela Picciariello, Bec Connelly, Danielle Belgrave et al. (2020) "AI for social good: unlocking the opportunity for positive impact." Nature Communications 11, no. 1: 1-6.

Sustainable Development Goals (SDGs)

Source: https://sdgs.un.org/goals

Sustainable Development Goals (SDGs) and 5P

Partnership

Peace

Prosperity

People

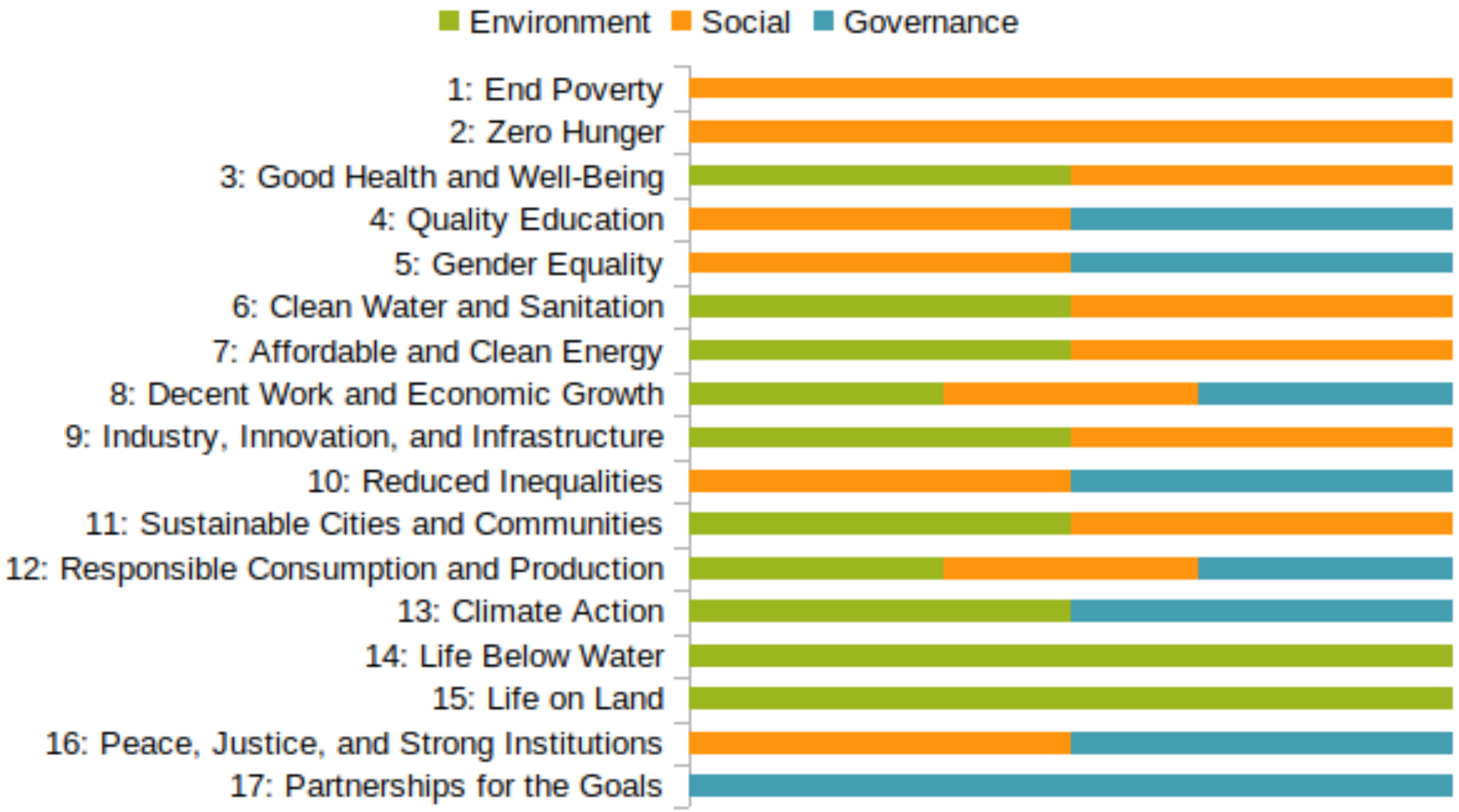
Planet

Source: Folke, Carl, Reinette Biggs, Albert V. Norström, Belinda Reyers, and Johan Rockström. "Social-ecological resilience and biosphere-based sustainability science."Ecology and Society 21, no. 3 (2016).

ESG to 17 SDGs

Source: Henrik Skaug Sætra (2021) "A Framework for Evaluating and Disclosing the ESG Related Impacts of AI with the SDGs." Sustainability 13, no. 15 (2021): 8503.

ESG to 17 SDGs



Source: https://sustainometric.com/esg-to-sdgs-connected-paths-to-a-sustainable-future/

AI for Social Good (AI4SG)
AI for Sustainable Development
AI4SG 10 Guidelines

- **AI Technology (G1, G2, G3)**
- **Applications (G4, G5, G6, G7, G8)**
- **Data Handling (G9, G10)**

Source: Nenad Tomašev, Julien Cornebise, Frank Hutter, Shakir Mohamed, Angela Picciariello, Bec Connelly, Danielle Belgrave et al. (2020) "AI for social good: unlocking the opportunity for positive impact." Nature Communications 11, no. 1: 1-6.

AI4SG 10 Guidelines
AI Technology (G1, G2, G3)

- **G1: Expectations of what is possible with AI need to be well-grounded.**
- **G2: There is value in simple solutions.**
- **G3: Applications of AI need to be inclusive and accessible, and reviewed at every stage for ethics and human rights compliance.**

Source: Nenad Tomašev, Julien Cornebise, Frank Hutter, Shakir Mohamed, Angela Picciariello, Bec Connelly, Danielle Belgrave et al. (2020) "AI for social good: unlocking the opportunity for positive impact." Nature Communications 11, no. 1: 1-6.

AI4SG 10 Guidelines
Applications (G4, G5, G6, G7, G8)

- **G4: Goals and use cases should be clear and well-defined.**
- **G5: Deep, long-term partnerships are required to solve large problems successfully.**
- **G6: Planning needs to align incentives, and factor in the limitations of both communities.**
- **G7: Establishing and maintaining trust is key to overcoming organisational barriers.**
- **G8: Options for reducing the development cost of AI solutions should be explored.**

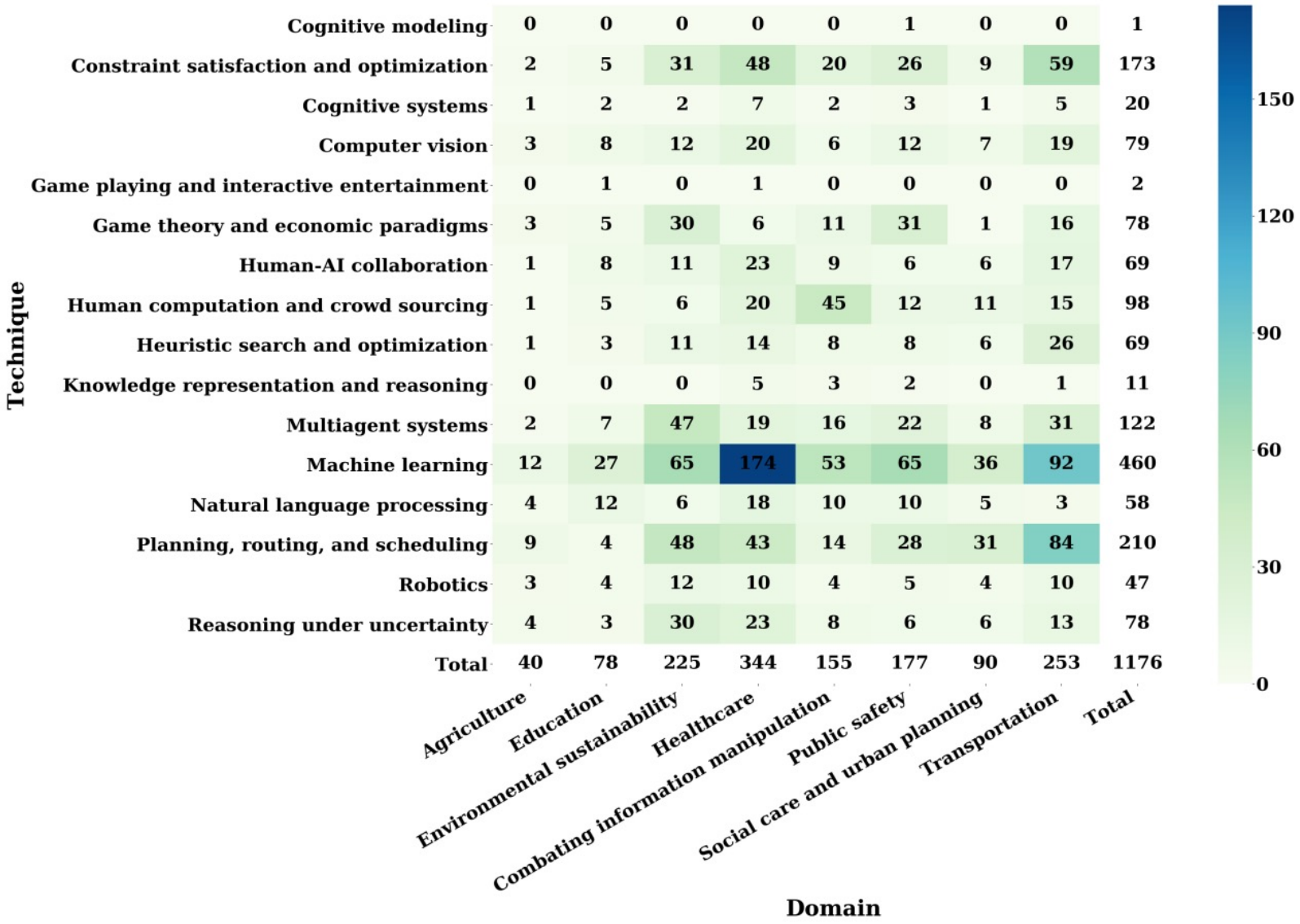
Source: Nenad Tomašev, Julien Cornebise, Frank Hutter, Shakir Mohamed, Angela Picciariello, Bec Connelly, Danielle Belgrave et al. (2020) "AI for social good: unlocking the opportunity for positive impact." Nature Communications 11, no. 1: 1-6.

AI4SG 10 Guidelines
Data Handling (G9, G10)

- **G9: Improving data readiness is key.**
- **G10: Data must be processed securely, with utmost respect for human rights and privacy.**

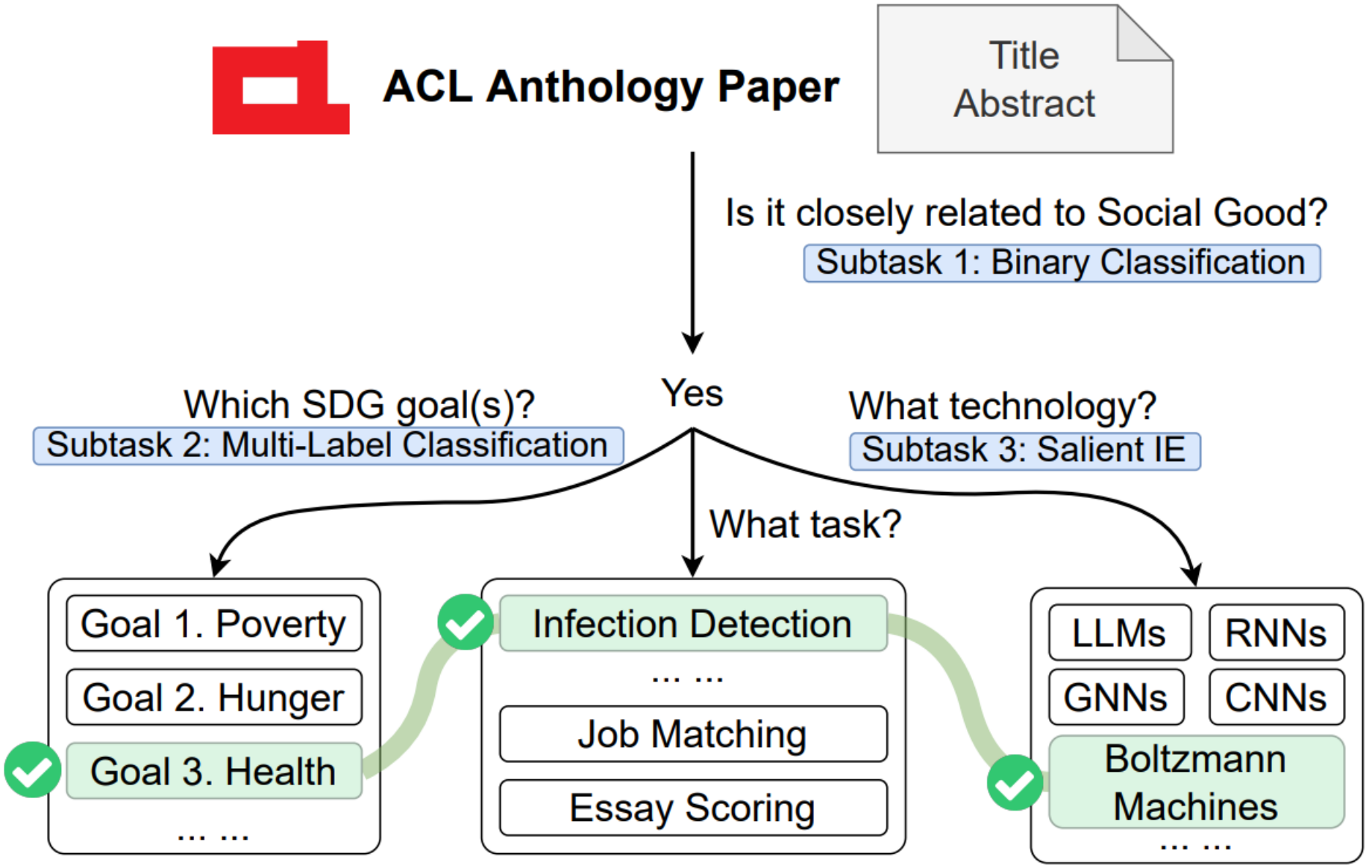
Source: Nenad Tomašev, Julien Cornebise, Frank Hutter, Shakir Mohamed, Angela Picciariello, Bec Connelly, Danielle Belgrave et al. (2020) "AI for social good: unlocking the opportunity for positive impact." Nature Communications 11, no. 1: 1-6.

AI for Social Good (AI4SG)
Domains and Techniques



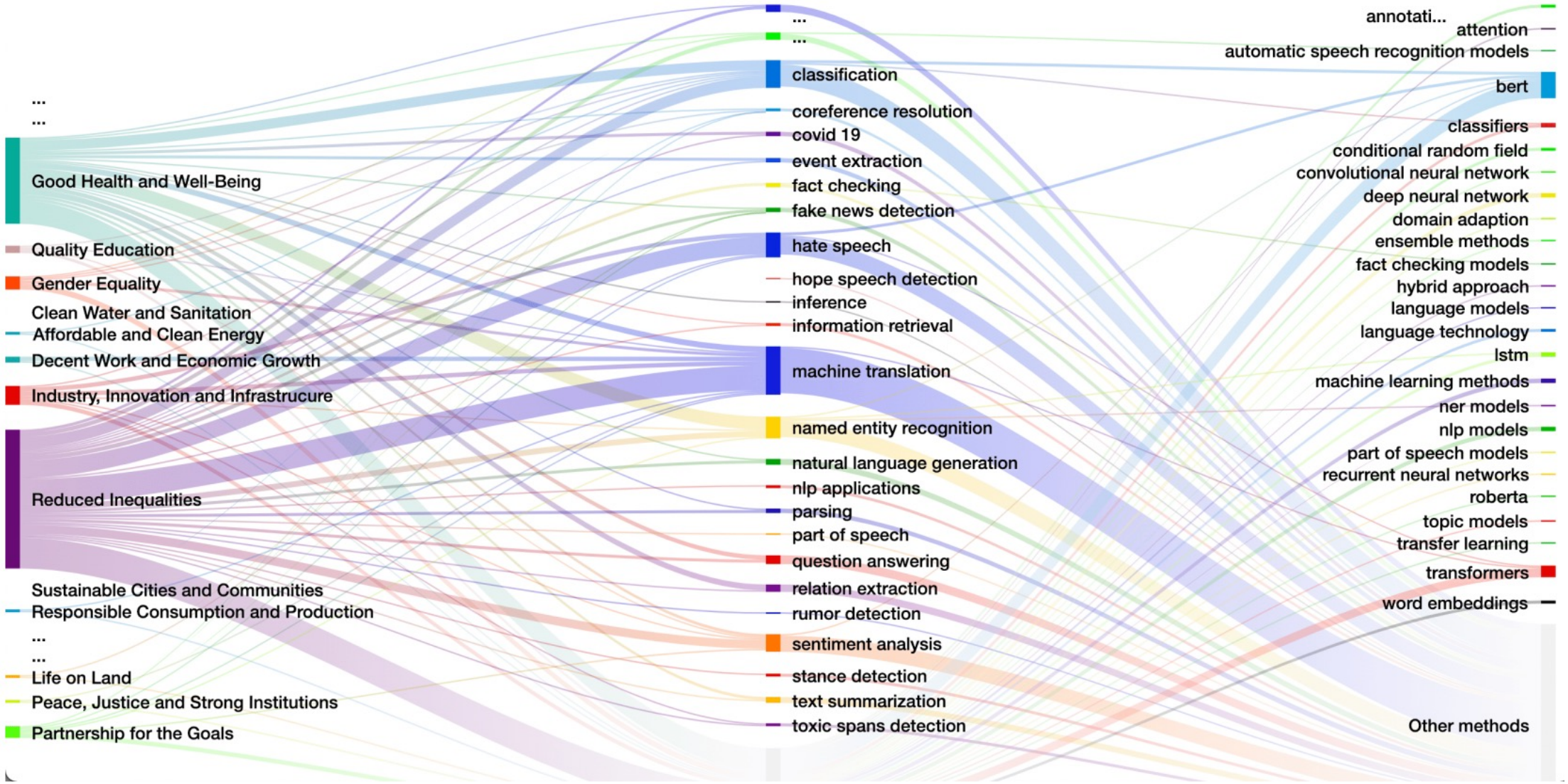
Source: Zheyuan Shi, Ryan, Claire Wang, and Fei Fang (2020). "Artificial intelligence for social good: A survey." arXiv preprint arXiv:2001.01818.

NLP for Social Good (NLP4SG)



Source: Fernando Gonzalez, Zhijing Jin, Jad Beydoun, Bernhard Schölkopf, Tom Hope, Rada Mihalcea, and Mrinmaya Sachan (2022).
"How Is NLP Addressing the 17 UN Sustainability Goals? A Challenge Set of Social Good Paper Classification and Information Extraction."

NLP for Social Good (NLP4SG) Visualization



Source: Fernando Gonzalez, Zhijing Jin, Jad Beydoun, Bernhard Schölkopf, Tom Hope, Rada Mihalcea, and Mrinmaya Sachan (2022).
"How Is NLP Addressing the 17 UN Sustainability Goals? A Challenge Set of Social Good Paper Classification and Information Extraction."

Papers with Code
State-of-the-Art (SOTA)

Computer Vision

Semantic Segmentation

185 benchmarks

3397 papers with code

Image Classification

390 benchmarks

2778 papers with code

Object Detection

269 benchmarks

2559 papers with code

Contrastive Learning

2 benchmarks

1119 papers with code

Image Generation

208 benchmarks

1097 papers with code

▸ See all 1415 tasks

Natural Language Processing

Language Modelling

458 benchmarks

2248 papers with code

Question Answering

181 benchmarks

1818 papers with code

Machine Translation

78 benchmarks

1721 papers with code

Sentiment Analysis

87 benchmarks

1040 papers with code

Text Generation

242 benchmarks

931 papers with code

▸ See all 664 tasks

https://paperswithcode.com/sota

Spring 2025

Generative AI Innovative Applications

Summary

- Generative AI
- Philosophy, Ethics, and Safety of AI
  - The Limits of AI
  - Can Machines Really Think?
  - The Ethics of AI
- The Future of AI
  - AI Components
  - AI Architectures

References

- Stuart Russell and Peter Norvig (2020), Artificial Intelligence: A Modern Approach, 4th Edition, Pearson.
- Denis Rothman (2024), Transformers for Natural Language Processing and Computer Vision - Third Edition: Explore Generative AI and Large Language Models with Hugging Face, ChatGPT, GPT-4V, and DALL-E 3, 3rd ed. Edition, Packt Publishing
- Aurélien Géron (2022), Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, 3rd Edition, O'Reilly Media.
- Steven D'Ascoli (2022), Artificial Intelligence and Deep Learning with Python: Every Line of Code Explained For Readers New to AI and New to Python, Independently published.
- Nenad Tomašev, Julien Cornebise, Frank Hutter, Shakir Mohamed, Angela Picciariello, Bec Connelly, Danielle Belgrave et al. (2020) "AI for social good: unlocking the opportunity for positive impact." Nature Communications 11, no. 1: 1-6.
- Fernando Gonzalez, Zhijing Jin, Jad Beydoun, Bernhard Schölkopf, Tom Hope, Rada Mihalcea, and Mrinmaya Sachan (2022). "How Is NLP Addressing the 17 UN Sustainability Goals? A Challenge Set of Social Good Paper Classification and Information Extraction."
- 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.
- 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.
- Bo Li, Peng Qi, Bo Liu, Shuai Di, Jingen Liu, Jiquan Pei, Jinfeng Yi, and Bowen Zhou. (2023) "Trustworthy AI: From principles to practices." ACM Computing Surveys 55, no. 9 (2023): 1-46.
- NVIDIA DLI (2024), Generative AI Explained, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-07+V1
- NVIDIA DLI (2024), Introduction to Transformer-Based Natural Language Processing, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-08+V1
- 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
- NVIDIA DLI (2024), Introduction to NVIDIA NIM™ Microservices, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-23+V1
- Min-Yuh Day (2024), Python 101, https://tinyurl.com/aintpupython101