Generative AI Innovative Applications

Transformers for Natural Language Processing and Computer Vision; Large Language Models (LLMs)

1142GAIIA02
MBA, IM, NTPU (M6031) (Spring 2026)
Tue 2, 3, 4 (9:10-12:00) (B3F17)

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

Min-Yuh Day, Ph.D,
Professor and Director

Institute of Information Management, National Taipei University

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

2026-03-03

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

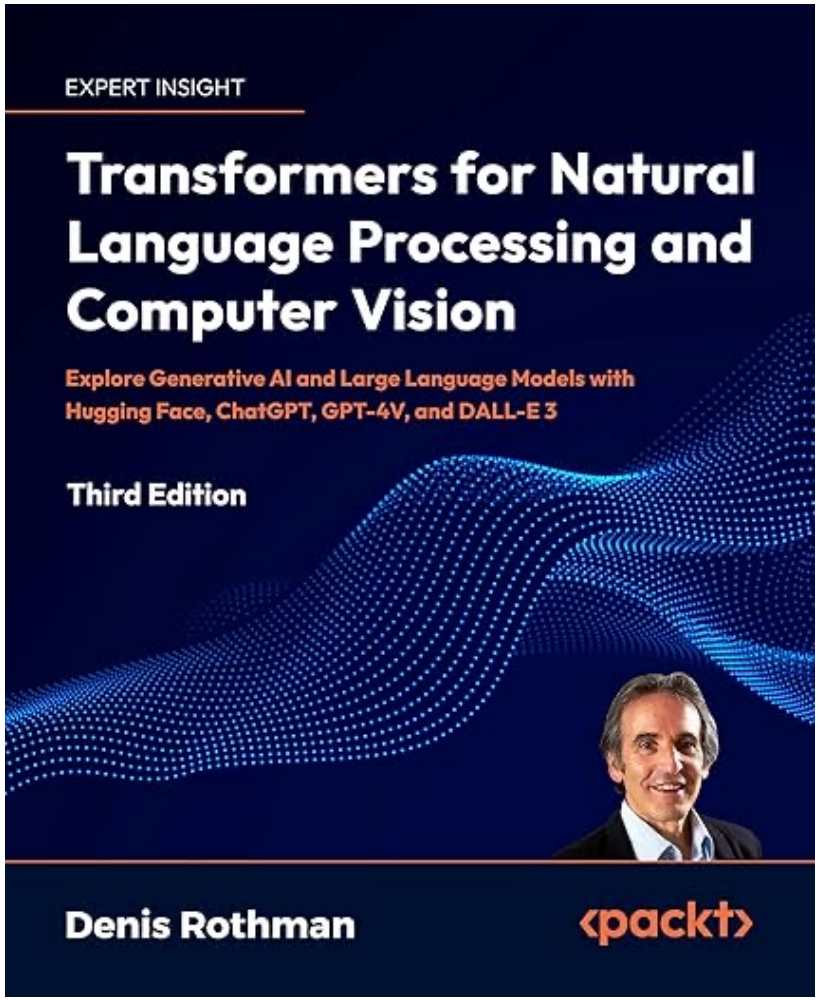
Syllabus

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

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



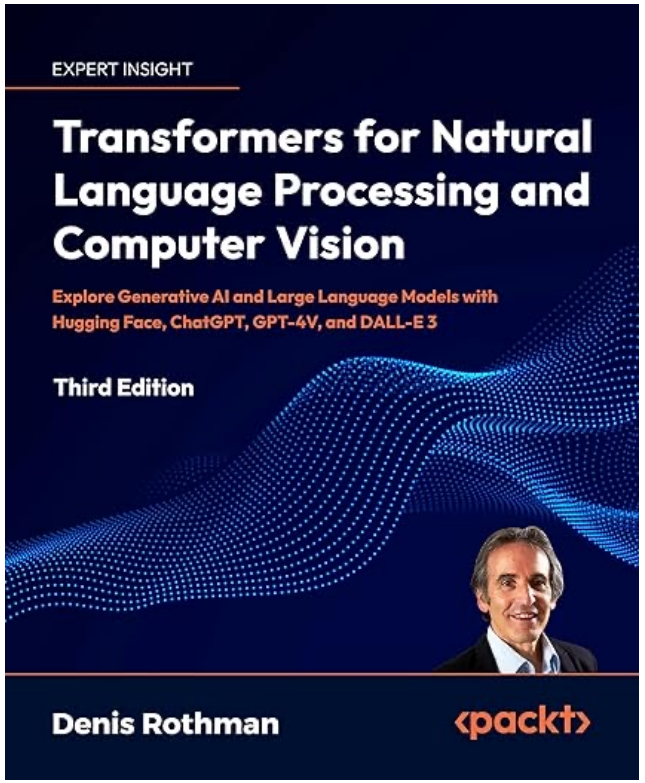
Source: https://www.amazon.com/Transformers-Natural-Language-Processing-Computer/dp/1805128728

Denis Rothman (2024),

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

1.What Are Transformers?
2.Getting Started with the Architecture of the Transformer Model
3.Emergent vs Downstream Tasks: The Unseen Depths of Transformers
4.Advancements in Translations with Google Trax, Google Translate, and Gemini
5.Diving into Fine-Tuning through BERT
6.Pretraining a Transformer from Scratch through RoBERTa
7.The Generative AI Revolution with ChatGPT
8.Fine-Tuning OpenAI GPT Models
9.Shattering the Black Box with Interpretable Tools
10.Investigating the Role of Tokenizers in Shaping Transformer Models
11.Leveraging LLM Embeddings as an Alternative to Fine-Tuning
12.Toward Syntax-Free Semantic Role Labeling with ChatGPT and GPT-4
13.Summarization with T5 and ChatGPT
14.Exploring Cutting-Edge LLMs with Vertex AI and PaLM 2
15.Guarding the Giants: Mitigating Risks in Large Language Models
16.Beyond Text: Vision Transformers in the Dawn of Revolutionary AI
17.Transcending the Image-Text Boundary with Stable Diffusion
18.Hugging Face AutoTrain: Training Vision Models without Coding
19.On the Road to Functional AGI with HuggingGPT and its Peers
20.Beyond Human-Designed Prompts with Generative Ideation



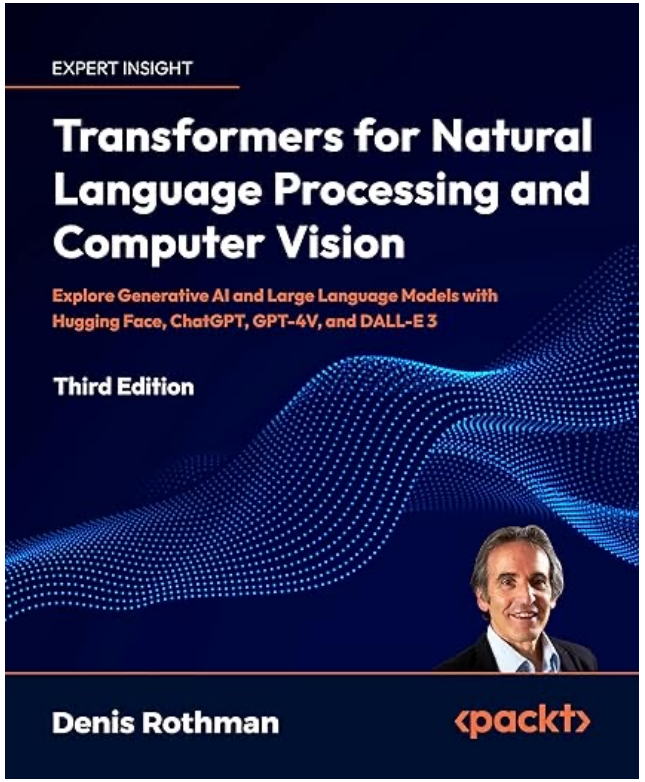
Source: https://github.com/Denis2054/Transformers-for-NLP-and-Computer-Vision-3rd-Edition

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

Chapter	Colab	Kaggle	Gradient	StudioLab
Part I The Foundations of Transformer Models				
Chapter 1: What are Transformers?				
• 🛠️O_1_and_Accelerators.ipynb • ChatGPT_Plus_writes_and_explains_AI.ipynb	Open in Colab Open in Colab	Open in Kaggle Open in Kaggle	Run on Gradient Run on Gradient	Open Studio Lab Open Studio Lab
Getting started with DeepSeek-R1 Reasoning models. Integrated into HuggingFace Hub and Together.				
• 🐬DeepSeek_Hugging_Face.ipynb	Open in Colab	Open in Kaggle	Run on Gradient	Open Studio Lab
Chapter 2: Getting Started with the Architecture of the Transformer Model				
• 🛠️Multi_Head_Attention_Sub_Layer.ipynb • positional_encoding.ipynb	Open in Colab Open in Colab	Open in Kaggle Open in Kaggle	Run on Gradient Run on Gradient	Open Studio Lab Open Studio Lab
Chapter 3: Emergent vs Downstream Tasks: the Unseen Depths of Transformers				
• From_training_to_emergence.ipynb • Transformer_tasks_with_Hugging_Face.ipynb	Open in Colab Open in Colab	Open in Kaggle Open in Kaggle	Run on Gradient Run on Gradient	Open Studio Lab Open Studio Lab

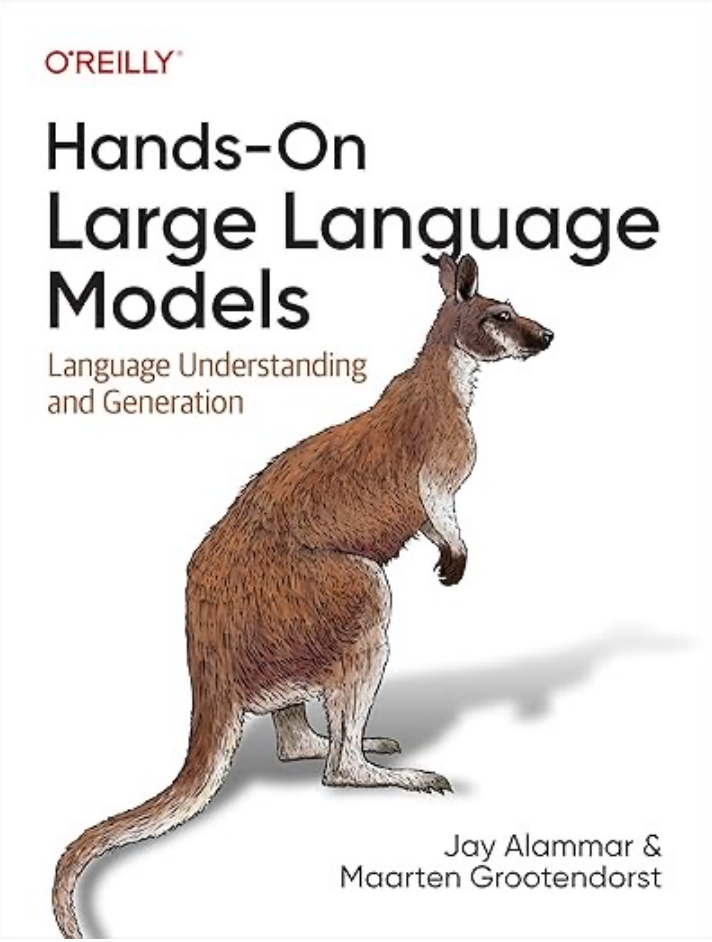


Source: https://github.com/Denis2054/Transformers-for-NLP-and-Computer-Vision-3rd-Edition

Jay Alammar and Maarten Grootendorst (2024),

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

O'Reilly Media



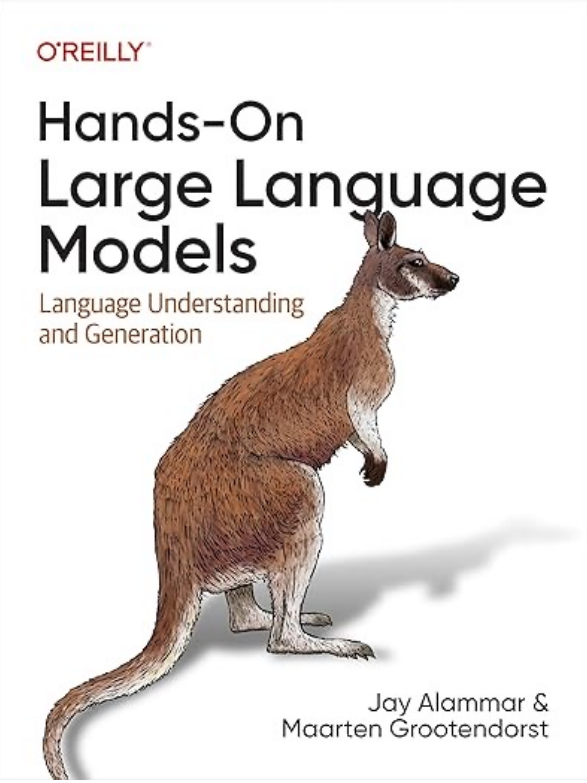
Source: https://www.amazon.com/Hands-Large-Language-Models-Understanding/dp/1098150961/

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

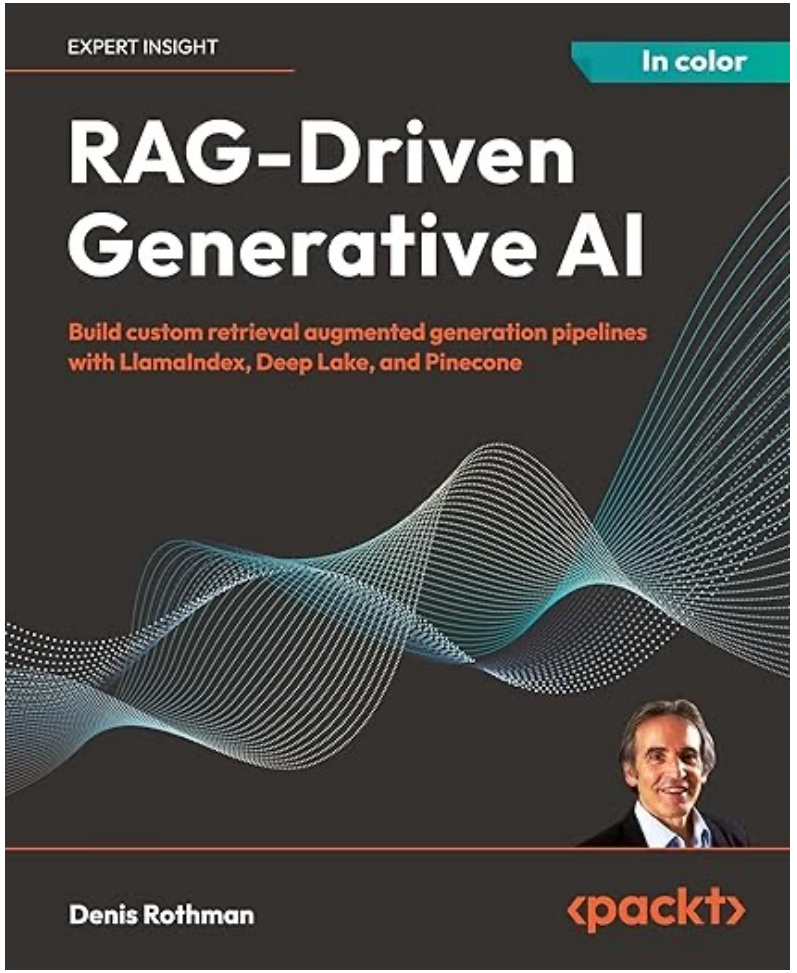


Source: https://github.com/HandsOnLLM/Hands-On-Large-Language-Models

Denis Rothman (2024),

RAG-Driven Generative AI:

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



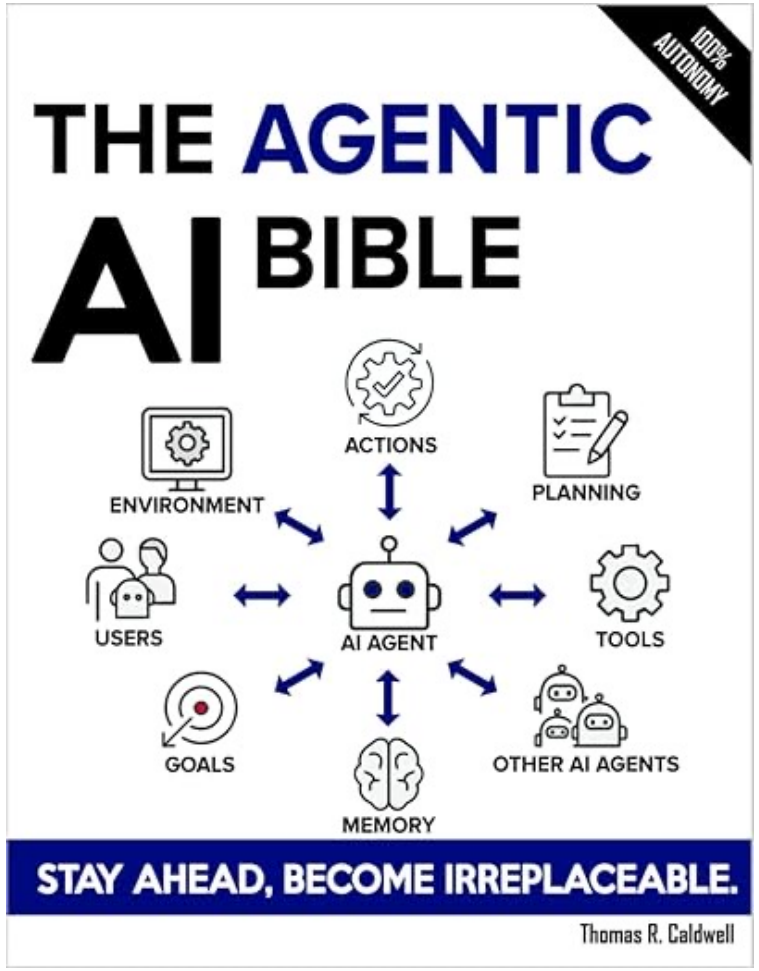
Source: https://www.amazon.com/RAG-Driven-Generative-retrieval-generation-LlamaIndex/dp/1836200919/

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

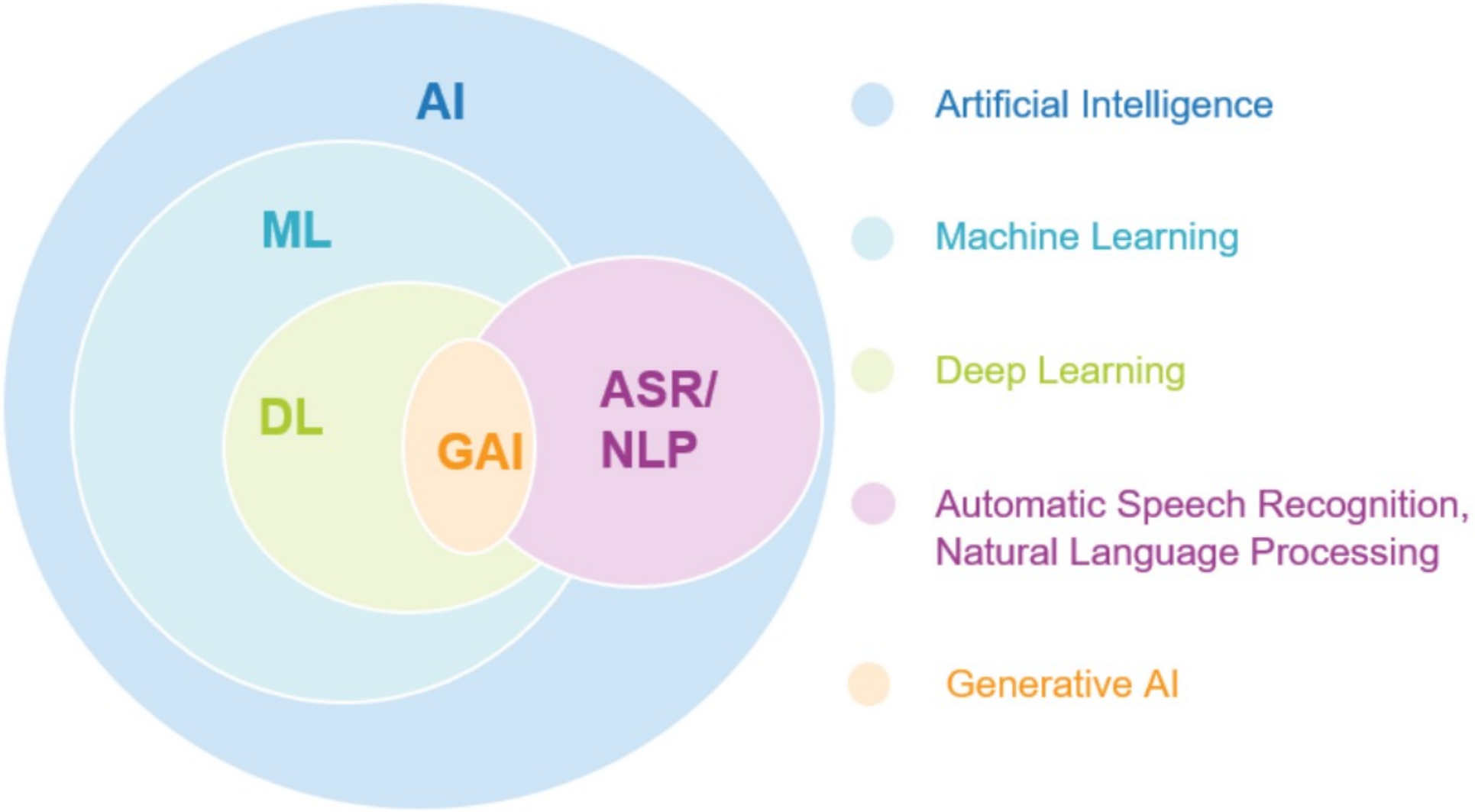
Generative AI
Large Language Models (LLMs)
Foundation Models

Generative AI (Gen AI)

AI Generated Content (AIGC)

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

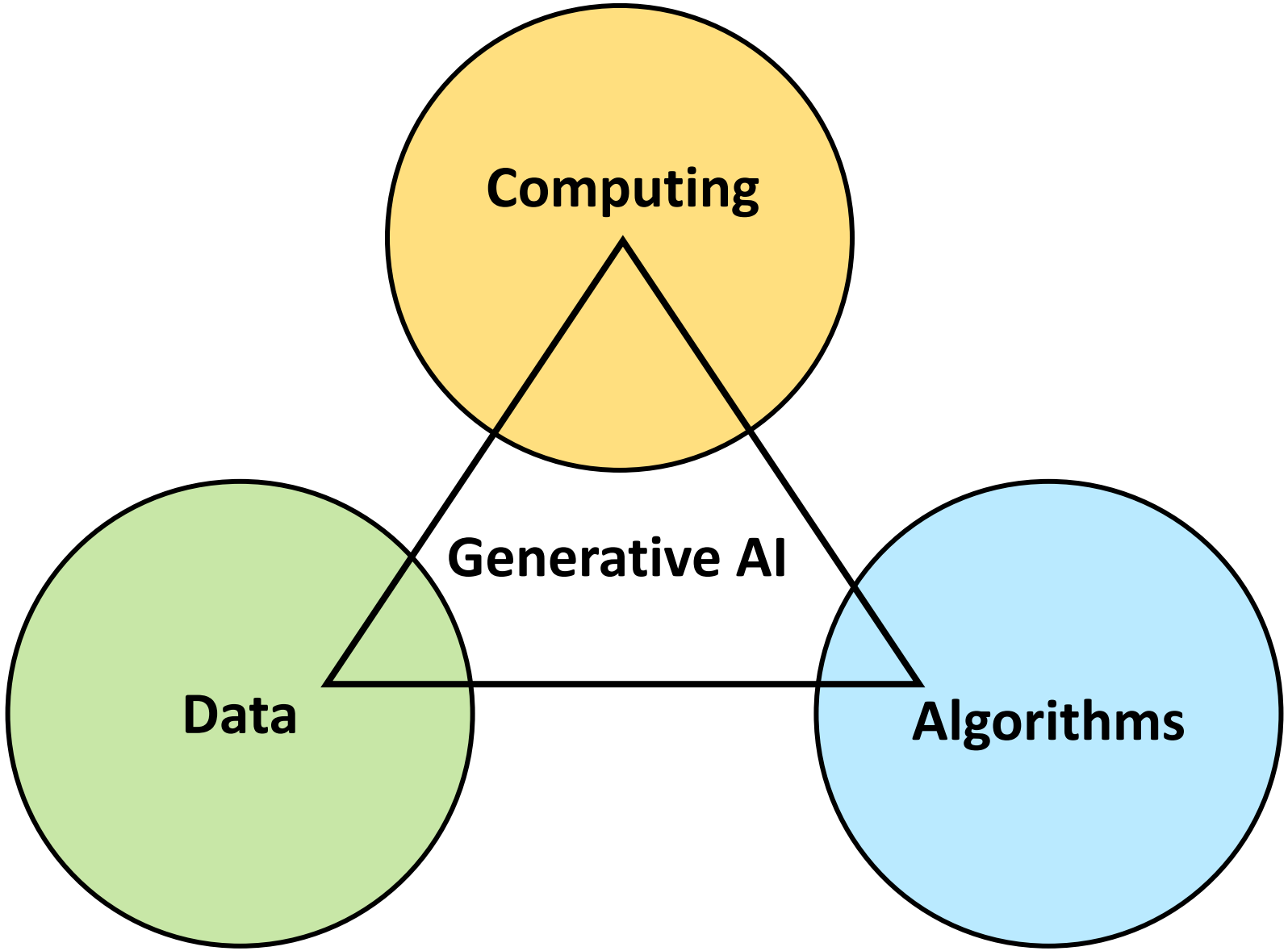
Generative AI
Digital marketing
Content creation

Agentic AI
Coding assistants
Customer service
Patient care

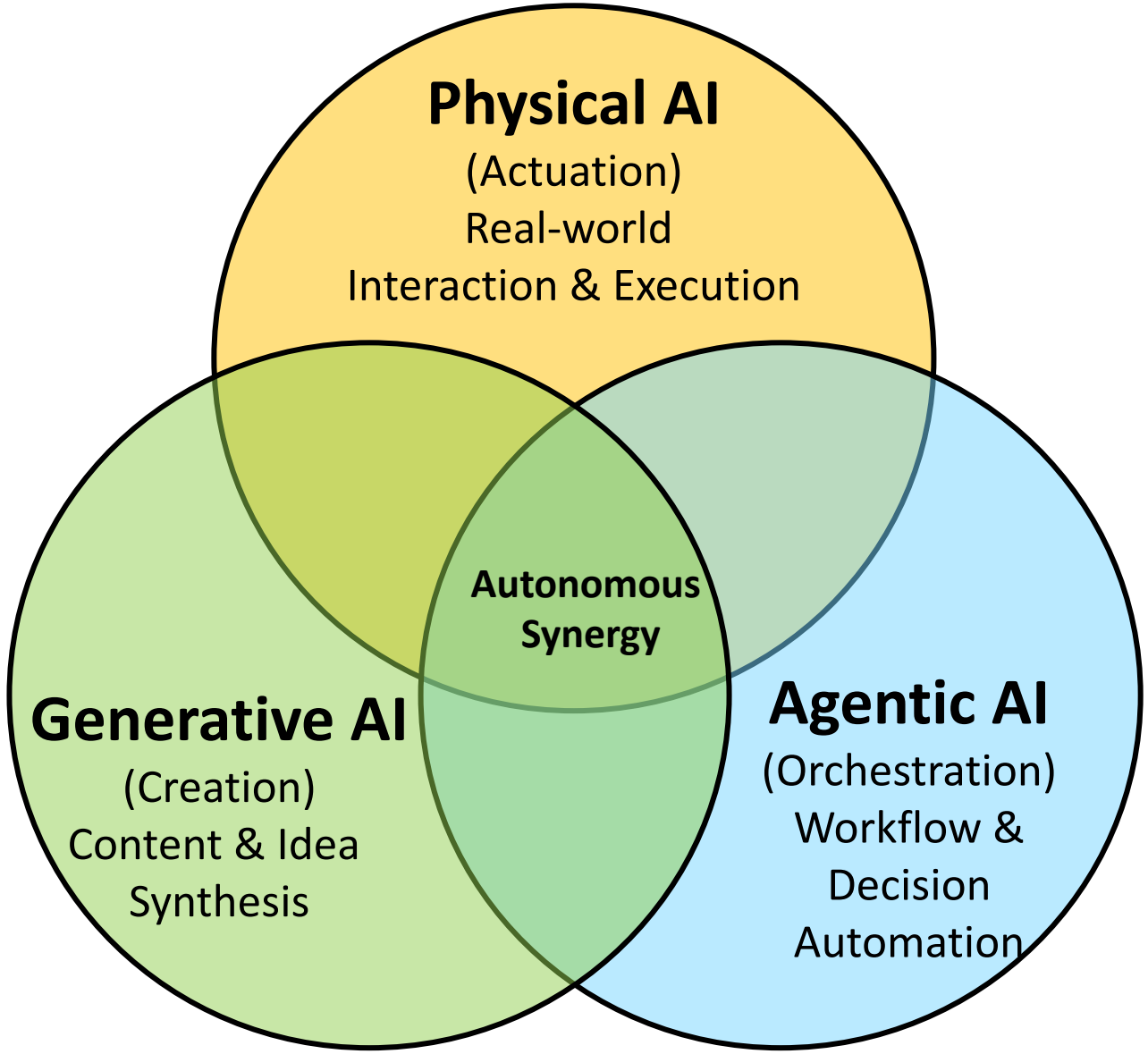
Physical AI
Self-driving cars
General robotics

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

Generative AI

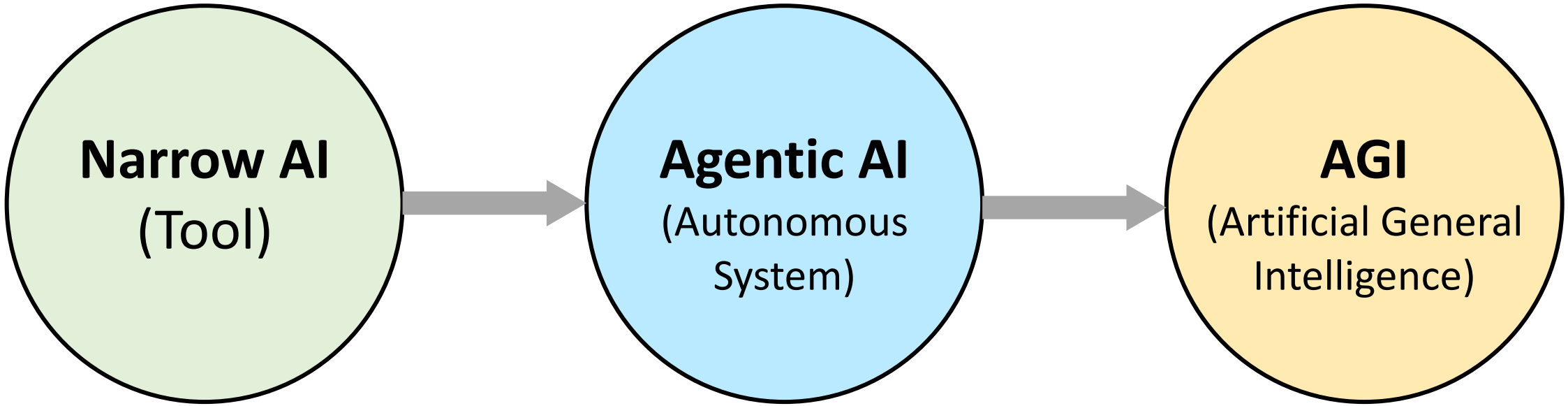


Generative AI, Agentic AI, Physical AI



New Economic Paradigm Shift: From Creation to Execution

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



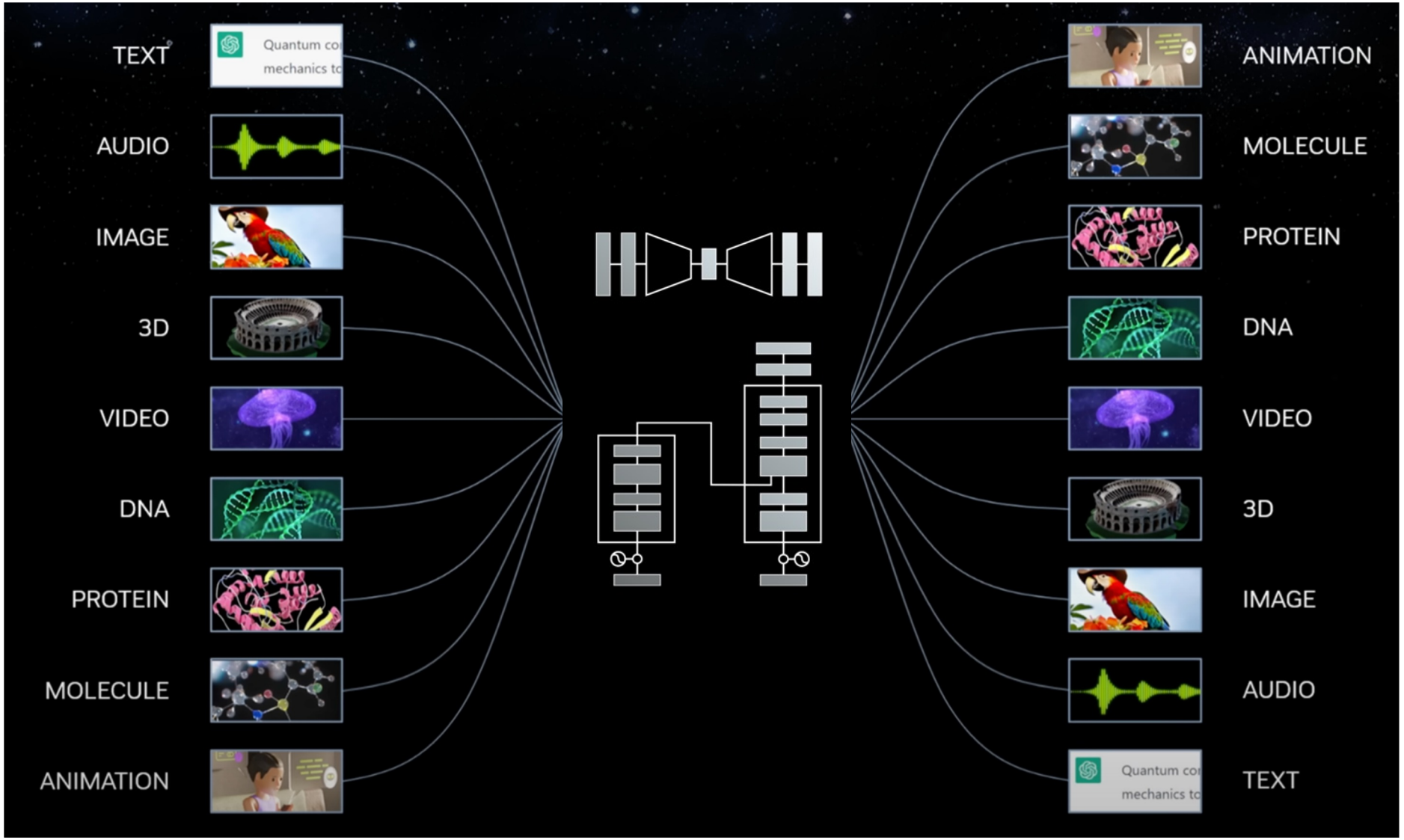
Executes Specific Tasks,
Content Generation

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

Connecting Multimodal GenAI
to Future AGI

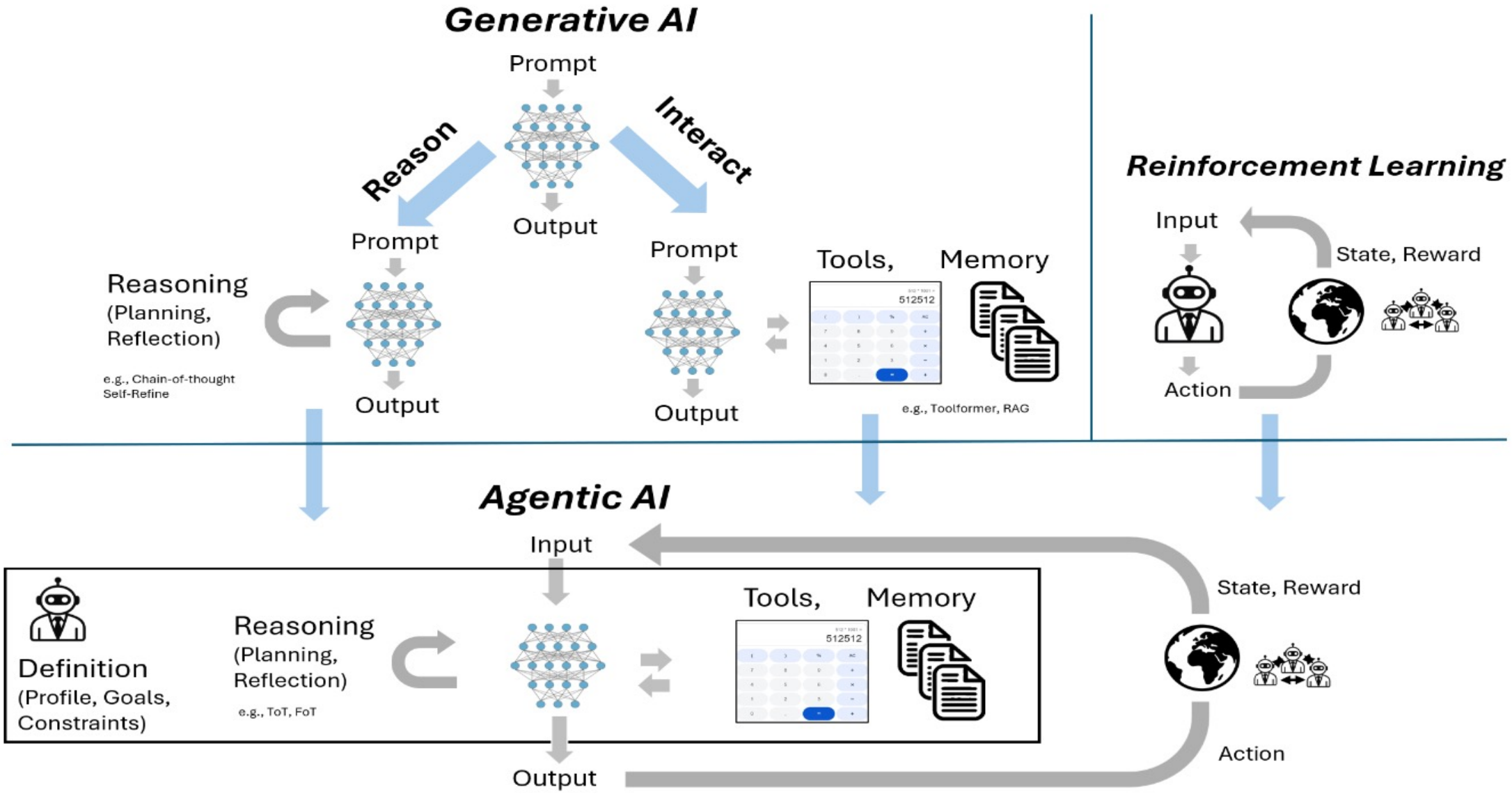
Modular Modalities

Where Can The Transformer Fit?



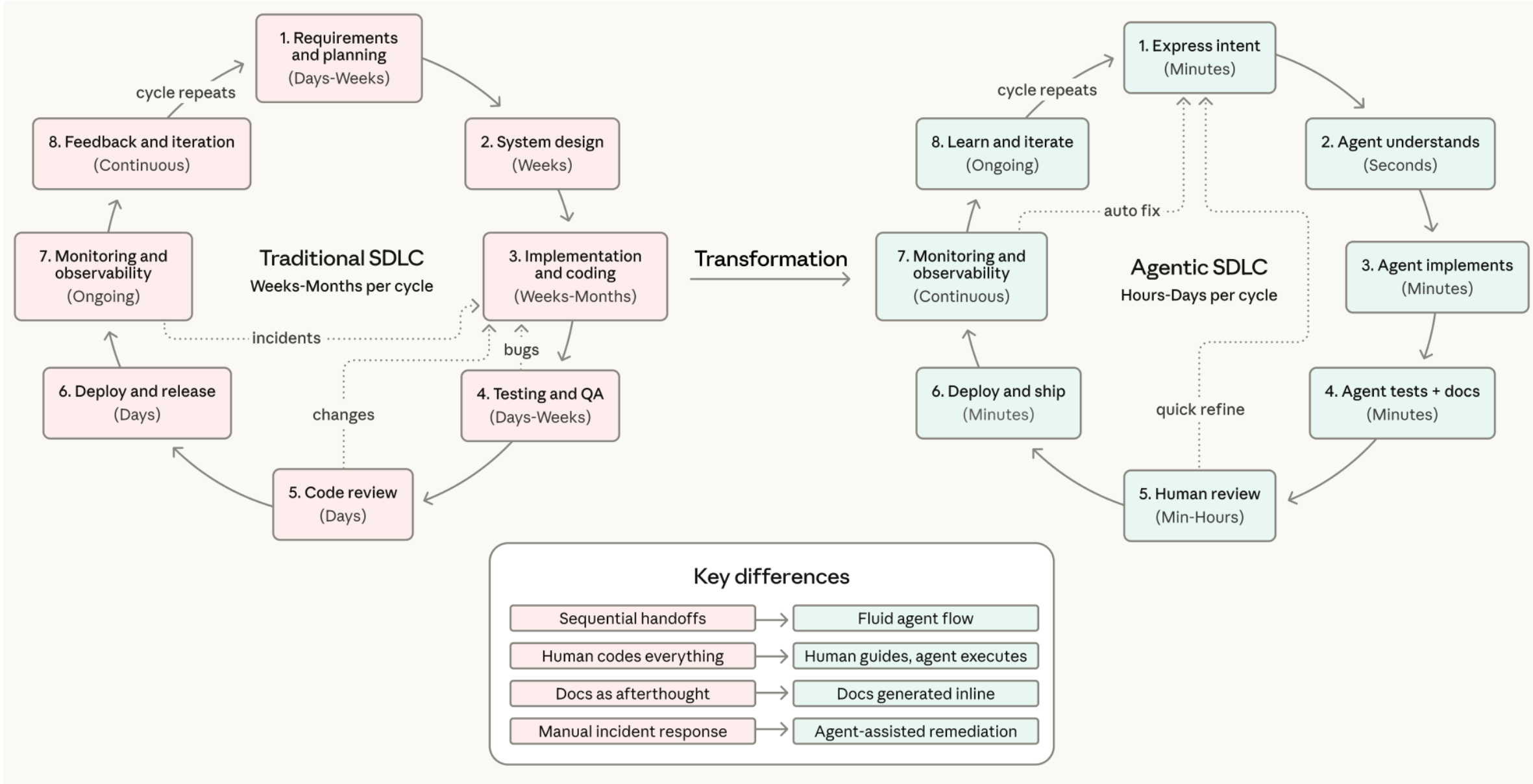
Source: NVIDIA DLI (2025), Rapid Application Development with Large Language Models (LLMs), https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-26+V1

From Generative AI to Agentic AI



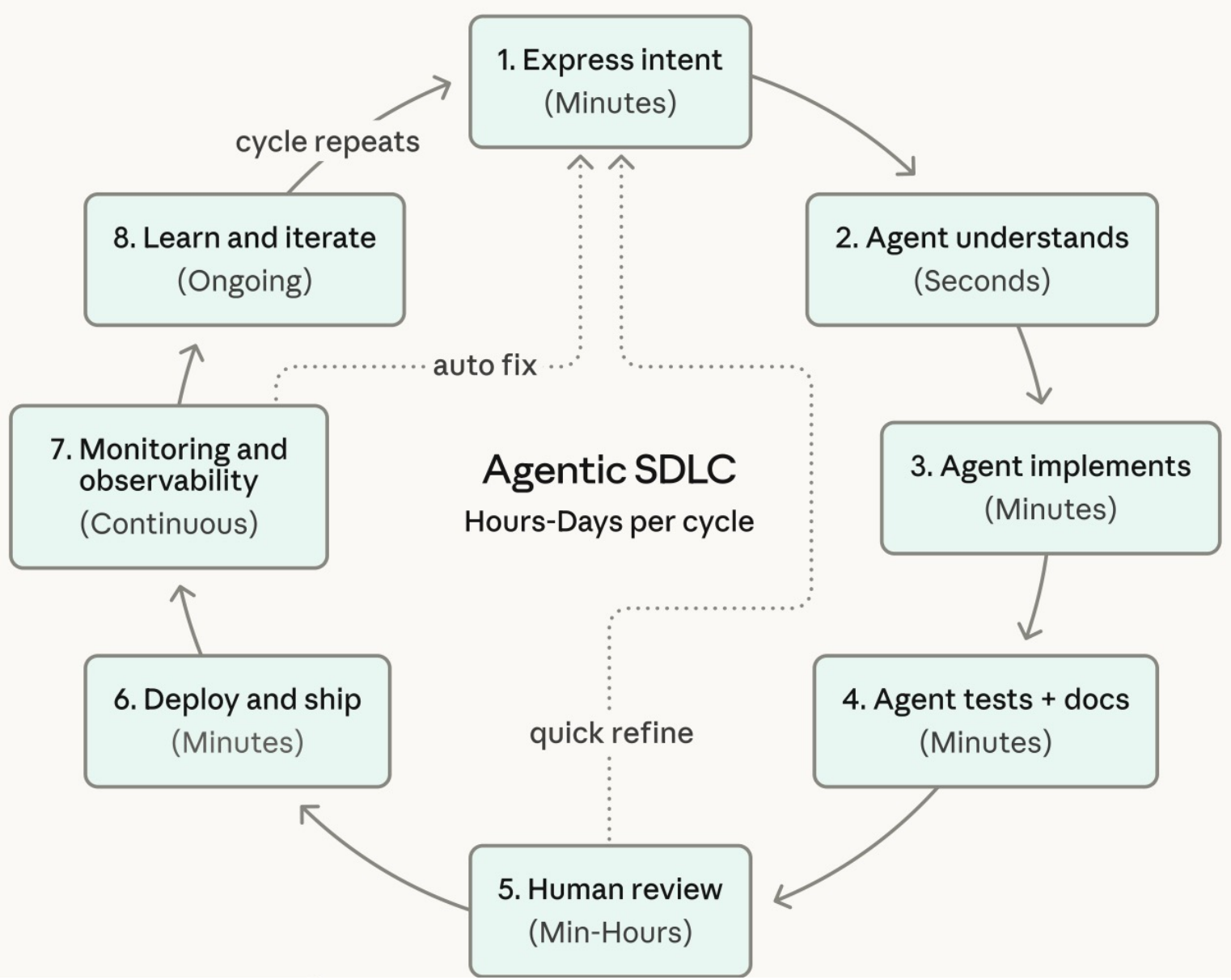
Source: Schneider, Johannes. "Generative to Agentic AI: Survey, Conceptualization, and Challenges." arXiv preprint arXiv:2504.18875 (2025).

Agentic Coding Software Development Lifecycle



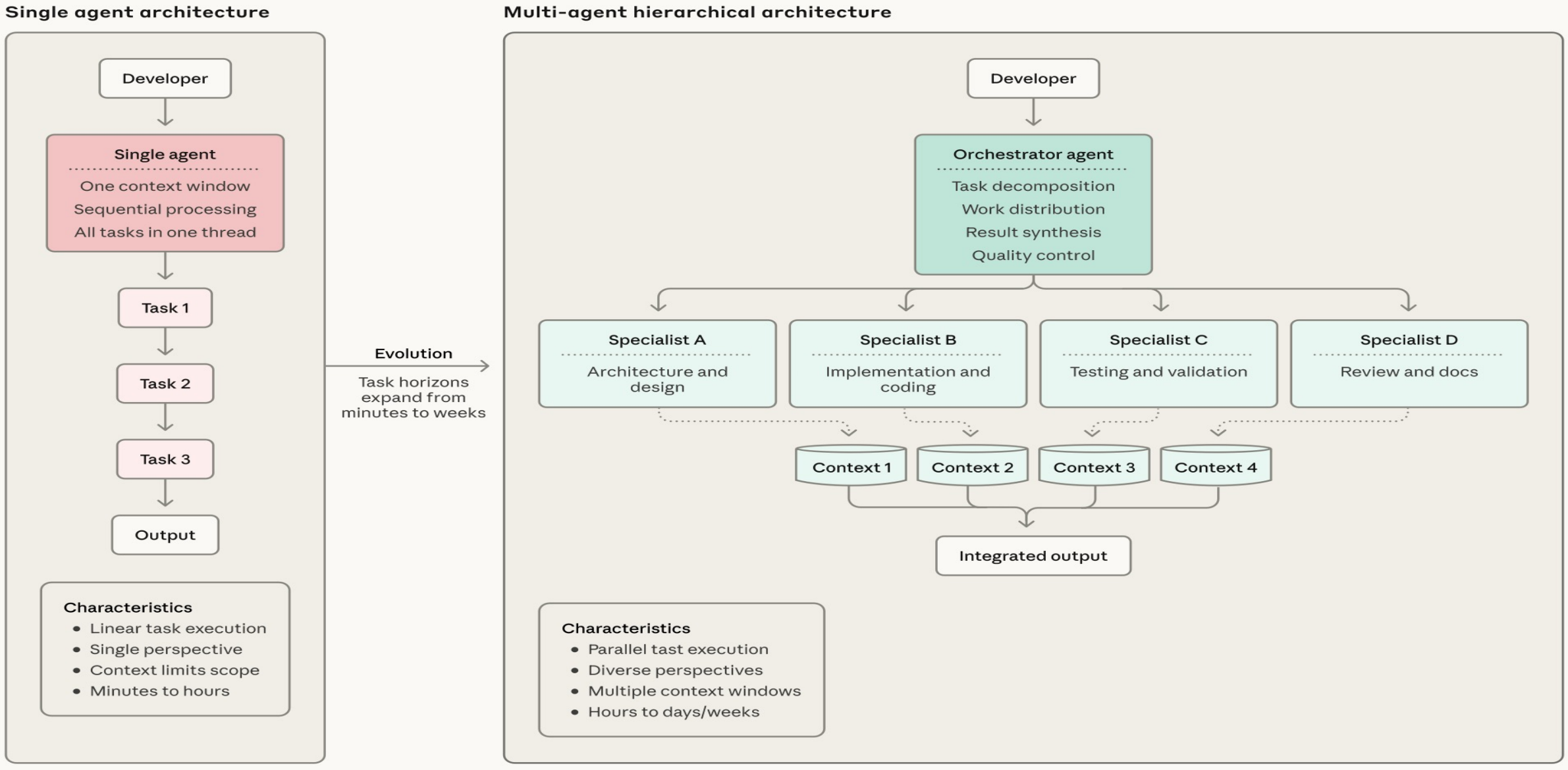
Source: Anthropic (2026), 2026 Agentic CodingTrends Report: How coding agents are reshaping software development, https://resources.anthropic.com/hubfs/2026%20Agentic%20Coding%20Trends%20Report.pdf

Agentic Coding Software Development Lifecycle



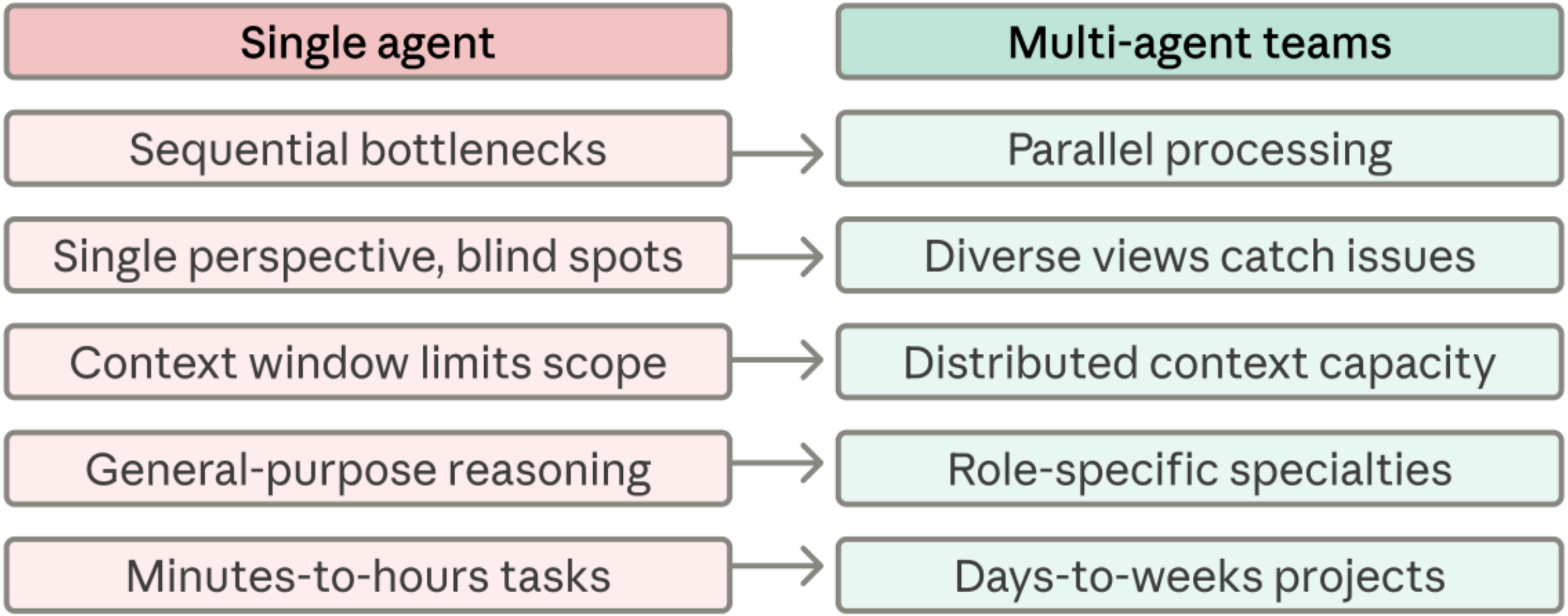
Source: Anthropic (2026), 2026 Agentic CodingTrends Report: How coding agents are reshaping software development, https://resources.anthropic.com/hubfs/2026%20Agentic%20Coding%20Trends%20Report.pdf

Coding Agent Architectures: From Single Agents to Coordinated Teams



Source: Anthropic (2026), 2026 Agentic CodingTrends Report: How coding agents are reshaping software development, https://resources.anthropic.com/hubfs/2026%20Agentic%20Coding%20Trends%20Report.pdf

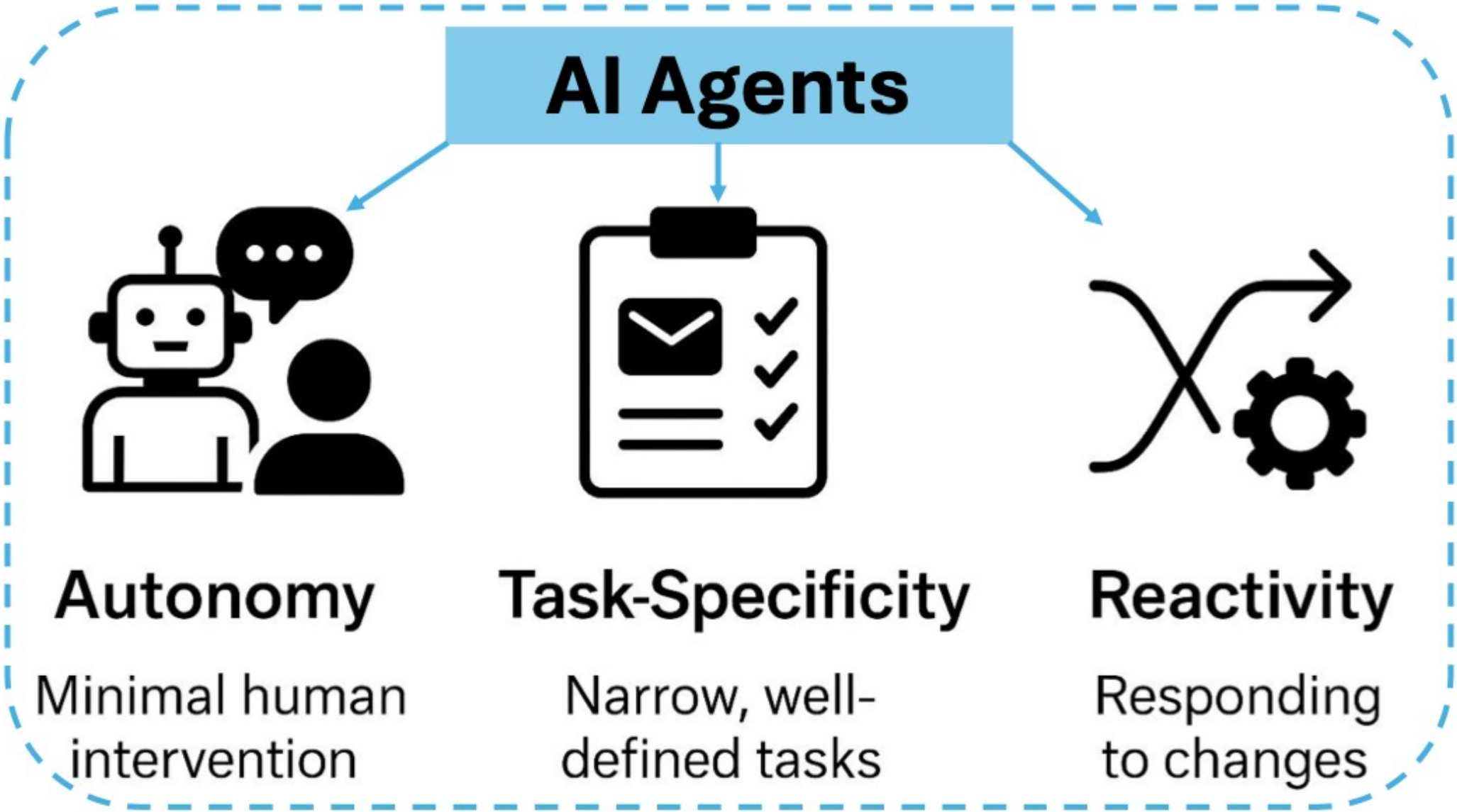
Coding Agent Architectures: Single Agent to Multi-Agent Teams Performance Impact



Source: Anthropic (2026), 2026 Agentic CodingTrends Report: How coding agents are reshaping software development, https://resources.anthropic.com/hubfs/2026%20Agentic%20Coding%20Trends%20Report.pdf

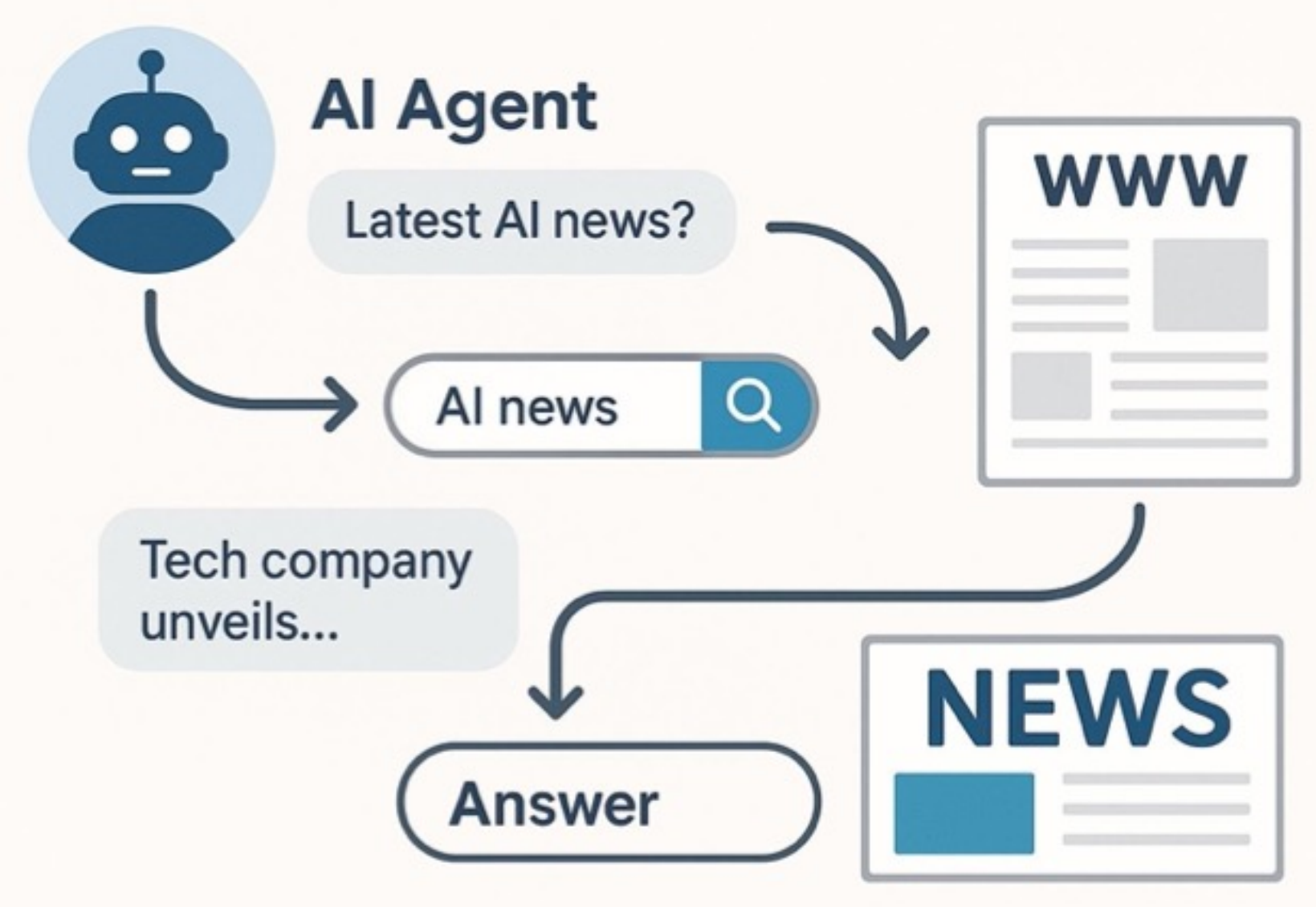
Agentic AI

AI Agents



Source: Ranjan Sapkota, Konstantinos I. Roumeliotis, and Manoj Karkee. (2025) "Ai agents vs. agentic ai: A conceptual taxonomy, applications and challenges." arXiv preprint arXiv:2505.10468 (2025).

AI Agents



Source: Ranjan Sapkota, Konstantinos I. Roumeliotis, and Manoj Karkee. (2025) "Ai agents vs. agentic ai: A conceptual taxonomy, applications and challenges." arXiv preprint arXiv:2505.10468 (2025).

Comparison of Generative AI and Traditional AI

Feature	Generative AI	Traditional AI
Output type	New content	Classification/Prediction
Creativity	High	Low
Interactivity	Usually more natural	Limited

AI Agent / Agentic AI, Generative AI, Traditional AI

Feature	AI Agent / Agentic AI	Generative AI	Traditional AI
Core Concept	To autonomously perceive its environment, make decisions, and take actions to achieve specific goals.	To create new, original content (text, images, code, etc.) that resembles its training data.	To execute specific tasks based on pre-programmed rules or statistical patterns.
Primary Function	Action & Goal Achievement. Executes a series of tasks to complete an objective (e.g., "Book me a flight to Taipei next Tuesday.").	Creation & Synthesis. Creates novel outputs in response to a prompt (e.g., "Write a poem about rain.").	Classification & Prediction. Answers questions with a known range of outcomes (e.g., "Is this spam?").
Decision Making	Based on a continuous loop: Perceive -> Plan -> Act. It reasons about its goal, breaks it down, and executes steps.	Based on probabilistic patterns learned from massive, unstructured datasets. It predicts the next most likely word, pixel, or note.	Based on explicitly programmed logic (if-then rules) or learned patterns from structured data.
Key Characteristic	Autonomous & Goal-Oriented. Proactively takes steps and can adapt its plan based on new information.	Creative & Probabilistic. Can produce a wide variety of unique outputs from the same prompt.	Deterministic & Logic-Based. Given the same input, it will almost always produce the same output.

AI Agent / Agentic AI, Generative AI, Traditional AI

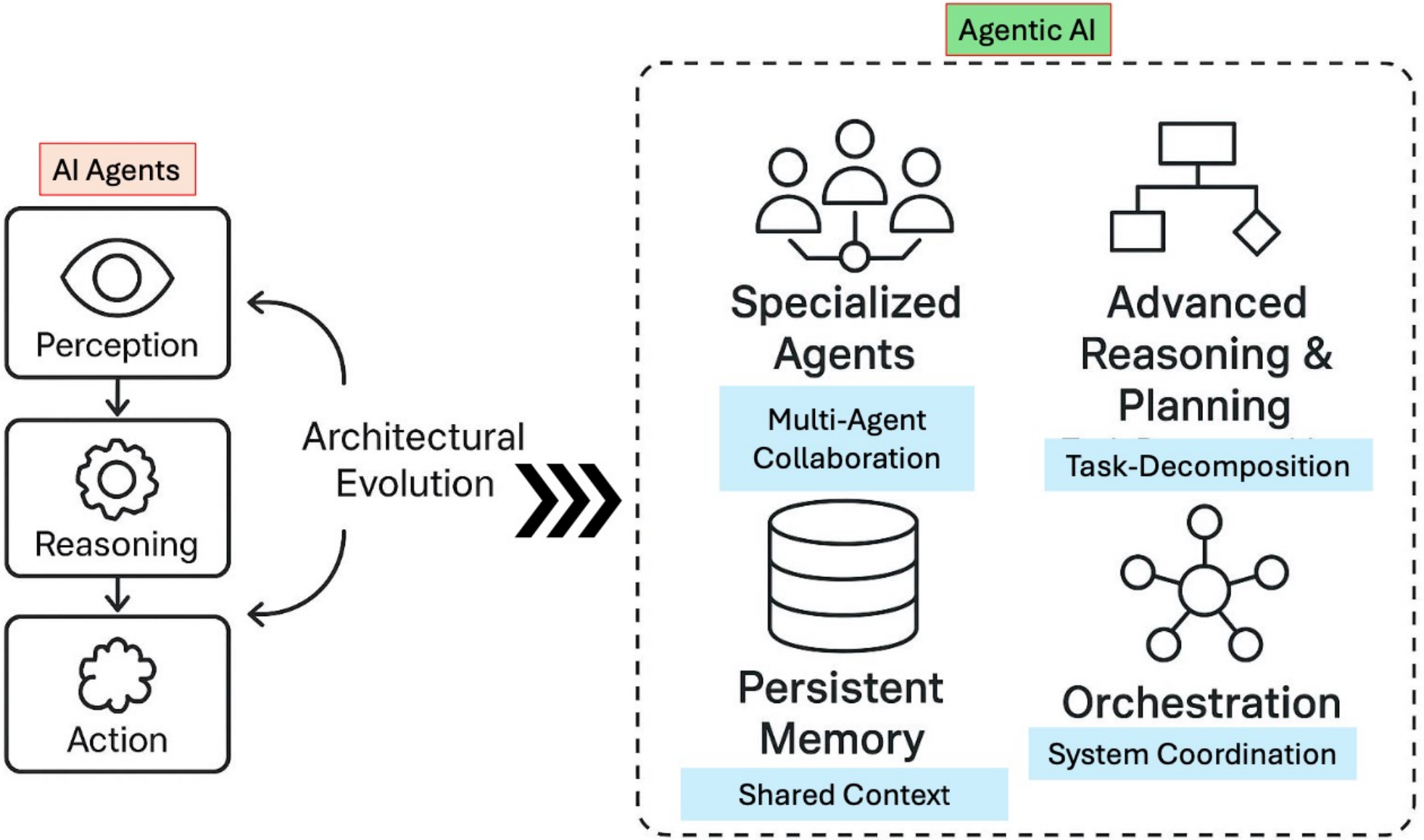
Feature	AI Agent / Agentic AI	Generative AI	Traditional AI
Interaction Model	Proactive & Interactive. Actively observes its environment (digital or physical) and takes actions to change it.	Responsive. Engages in a dialogue or responds to a user's prompt to generate content.	Reactive. Responds to a direct input or query. It doesn't act on its own.
Example Technologies	Architectural frameworks like ReAct (Reason + Act), and systems that combine LLMs with tools and memory.	Large Language Models (LLMs) like GPT-4, Diffusion Models (for images), Generative Adversarial Networks (GANs).	Expert systems, decision trees, linear regression, traditional machine learning (ML) models.
Common Use Cases	Self-driving cars, autonomous trading bots, smart assistants that manage calendars, customer service agents that process refunds.	ChatGPT, Google Gemini, Midjourney (image generation), Copilot (code generation), music composition.	Spam filters, chess engines, recommendation systems (e.g., Netflix), credit scoring, medical diagnosis from scans.
Relationship to Others	An architecture or system that often uses Generative AI to reason and Traditional AI for specific sub-tasks to accomplish a goal.	Can serve as the "brain" or reasoning engine for an AI Agent, enabling it to understand, plan, and generate actions.	The foundation for modern AI. Its techniques can be components within larger AI systems.

AI Agents vs Agentic AI

Feature	AI Agents	Agentic AI
Definition	Autonomous software programs that perform specific tasks.	Systems of multiple AI agents collaborating to achieve complex goals.
Autonomy Level	High autonomy within specific tasks.	Broad level of autonomy with the ability to manage multi-step, complex tasks and systems.
Task Complexity	Typically handle single, specific tasks.	Handle complex, multi-step tasks requiring coordination.
Collaboration	Operate independently.	Involve multi-agent information sharing, collaboration and cooperation.
Learning and Adaptation	Learn and adapt within their specific domain.	Learn and adapt across a wider range of tasks and environments.
Applications	Customer service chatbots, virtual assistants, automated workflows.	Supply chain management, business process optimization, virtual project managers.

Source: Ranjan Sapkota, Konstantinos I. Roumeliotis, and Manoj Karkee. (2025) "Ai agents vs. agentic ai: A conceptual taxonomy, applications and challenges." arXiv preprint arXiv:2505.10468 (2025).

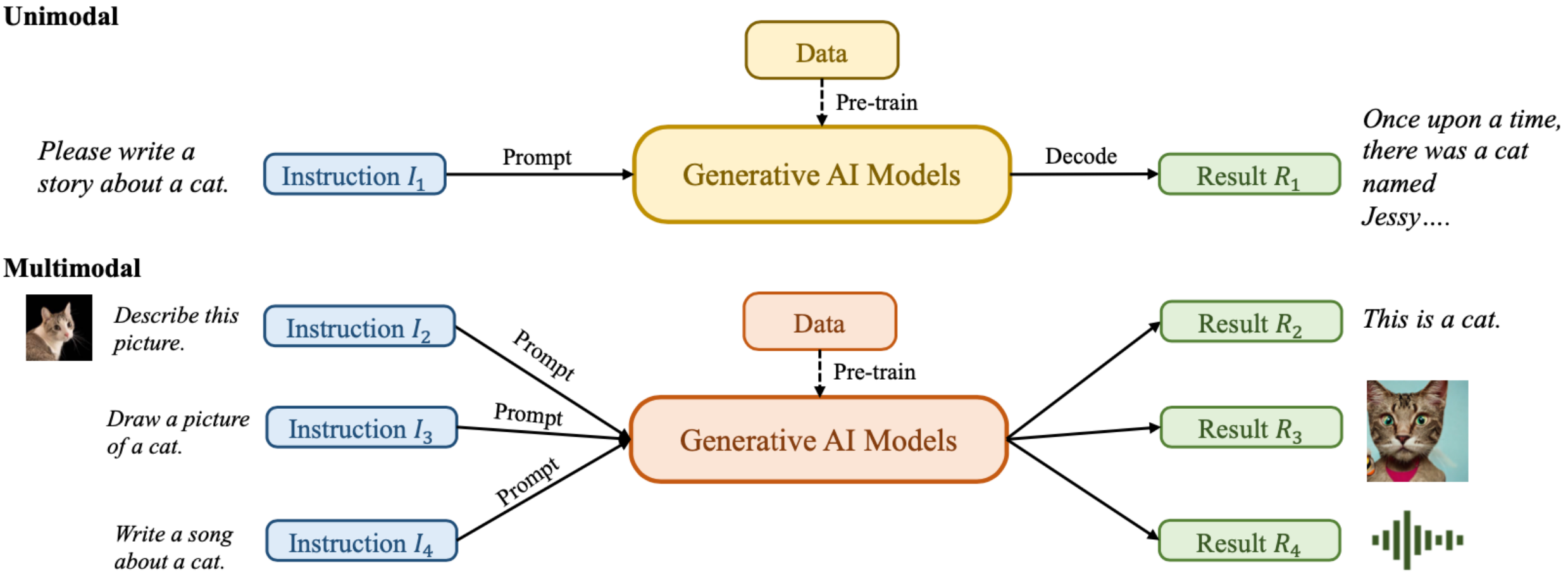
AI Agents vs Agentic AI



Source: Ranjan Sapkota, Konstantinos I. Roumeliotis, and Manoj Karkee. (2025) "Ai agents vs. agentic ai: A conceptual taxonomy, applications and challenges." arXiv preprint arXiv:2505.10468 (2025).

Generative AI (Gen AI)

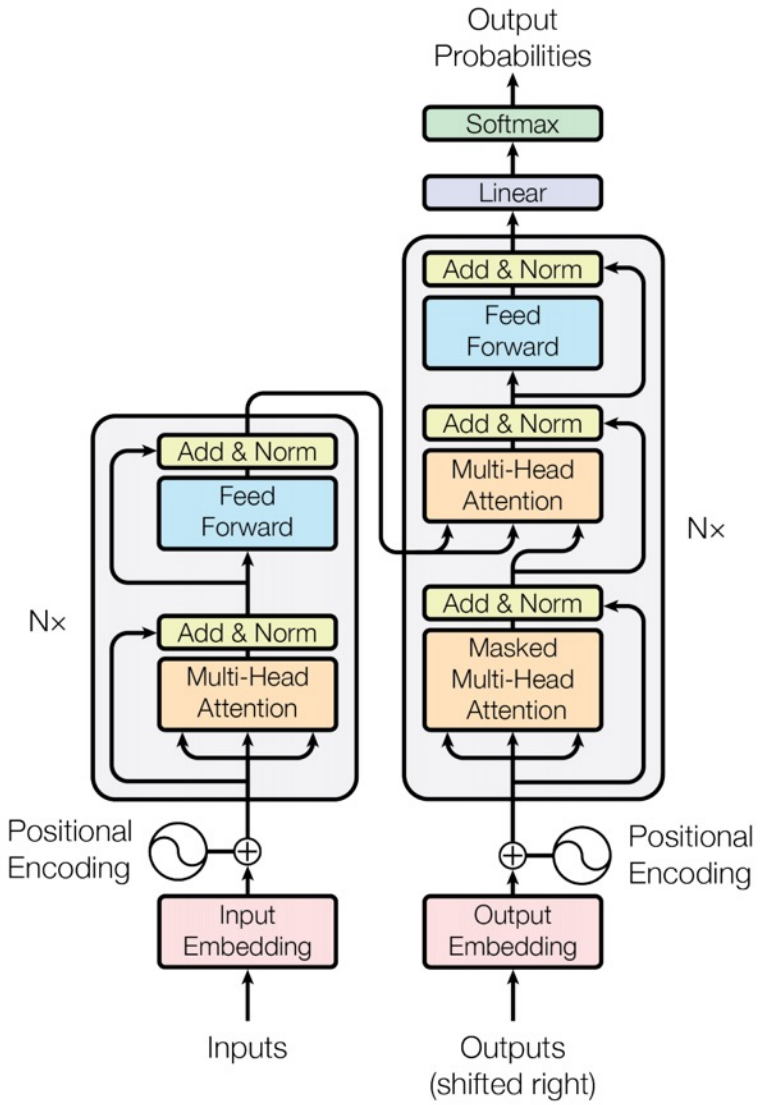
AI Generated Content (AIGC)



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.

Transformer (Attention is All You Need)

(Vaswani et al., 2017)



Source: Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need." In *Advances in neural information processing systems*, pp. 5998-6008. 2017.

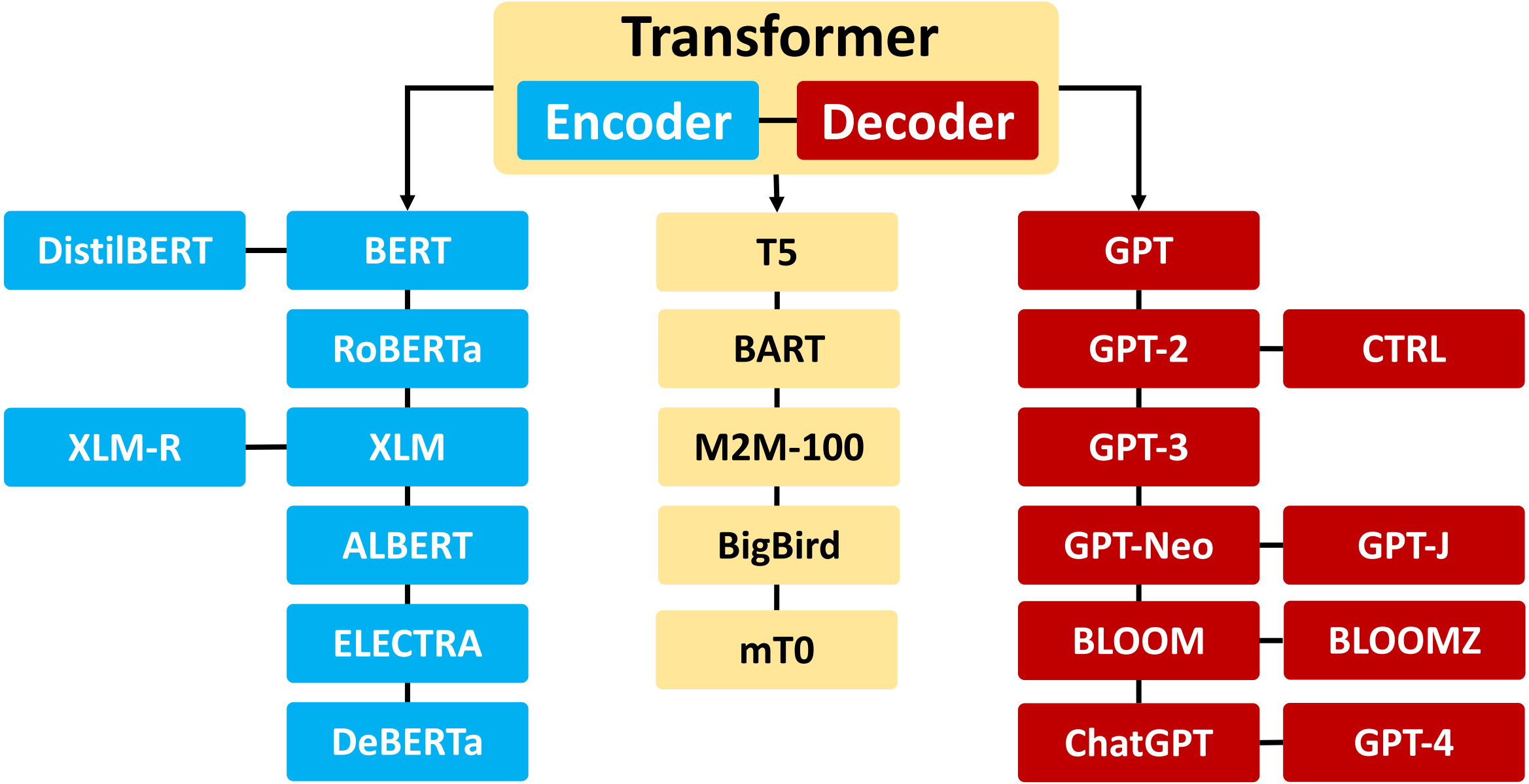
Transformer (Attention is All You Need)

(Vaswani et al., 2017)

- A **Transformer** is a type of **deep learning model** introduced in the paper "**Attention Is All You Need**" (Vaswani et al., 2017).
- It revolutionized **Natural Language Processing (NLP)** by replacing traditional **sequence models** like **RNNs and LSTMs** with a **self-attention mechanism** that enables highly parallelizable training.

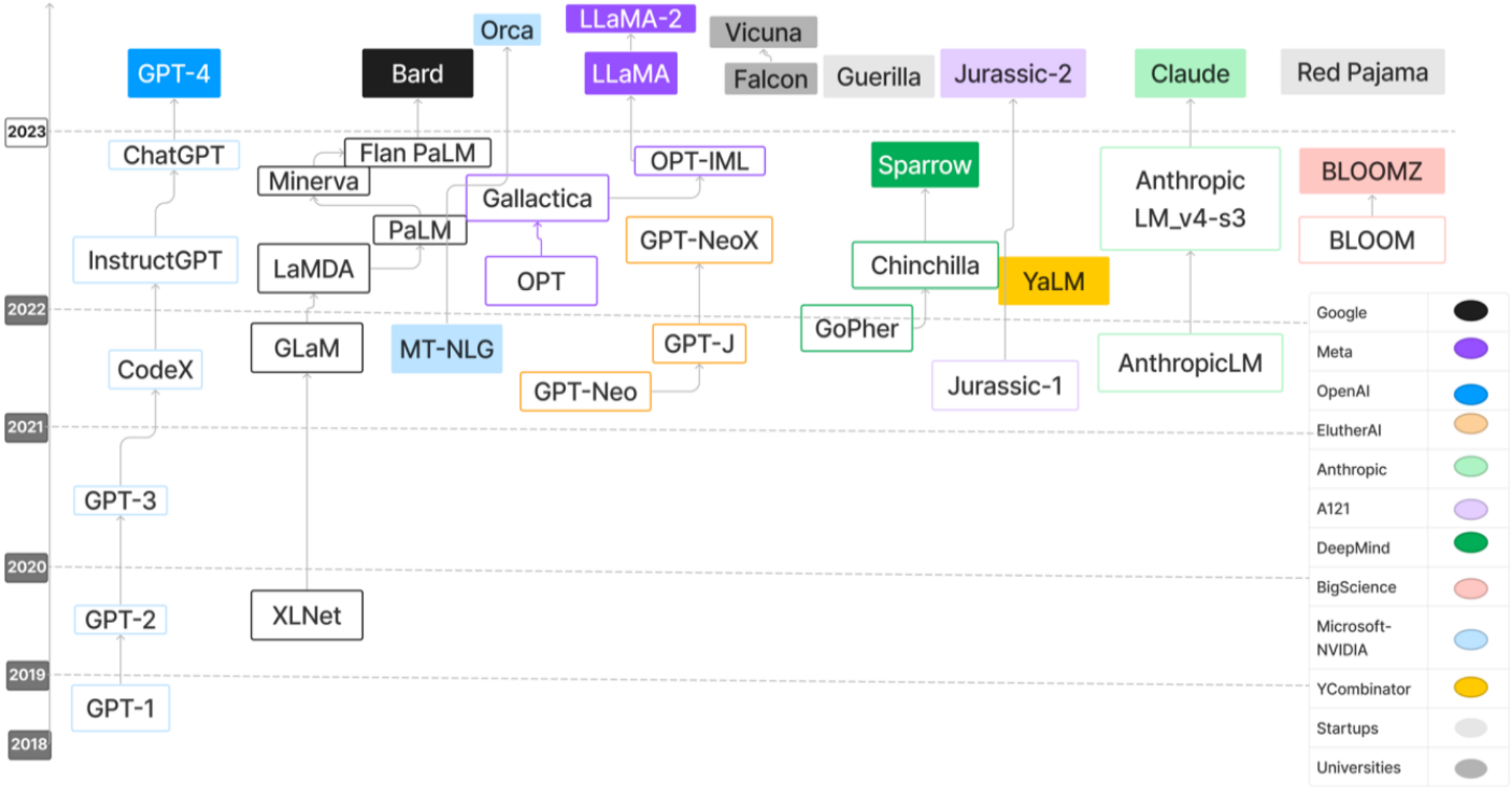
Source: Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need." In *Advances in neural information processing systems*, pp. 5998-6008. 2017.

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.

Large Language Models (LLMs)

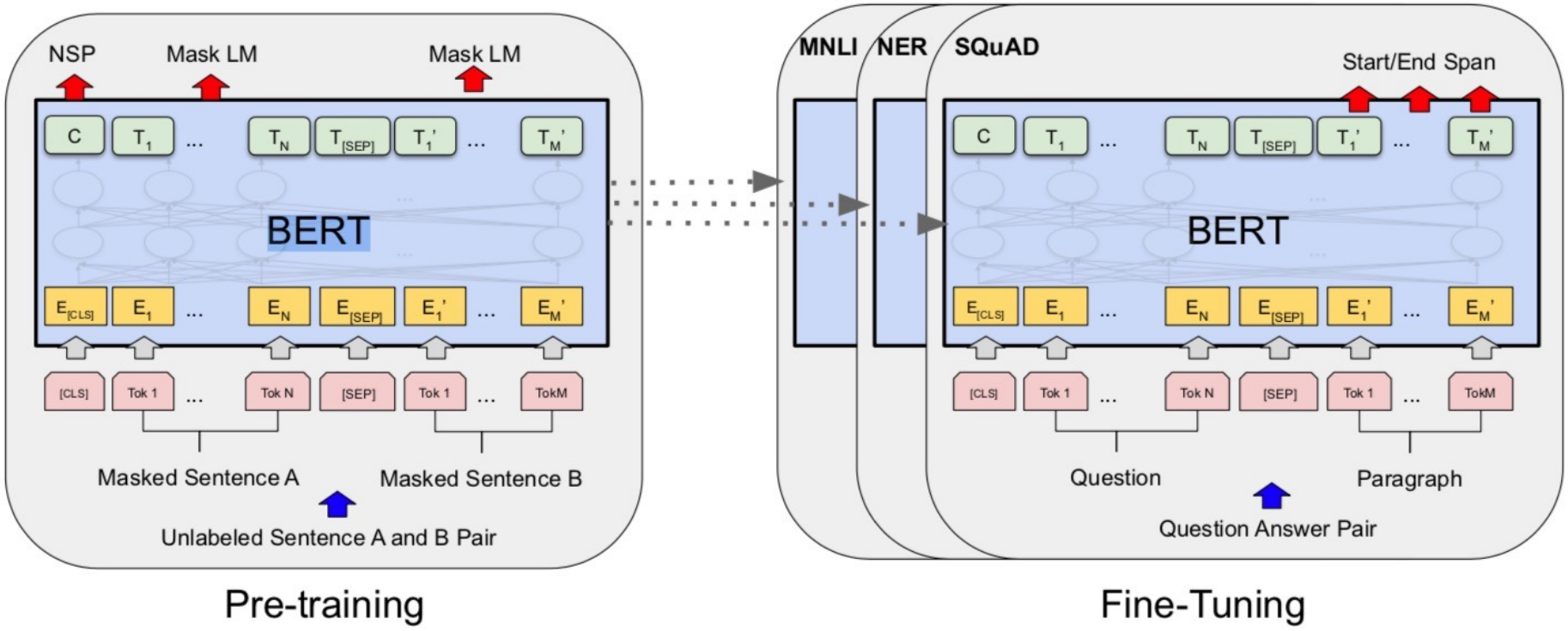


Source: Hadi, Muhammad Usman, Rizwan Qureshi, Abbas Shah, Muhammad Irfan, Anas Zafar, Muhammad Bilal Shaikh, Naveed Akhtar, Jia Wu, and Seyedali Mirjalili. "Large language models: a comprehensive survey of its applications, challenges, limitations, and future prospects." Authorea Preprints (2023).

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

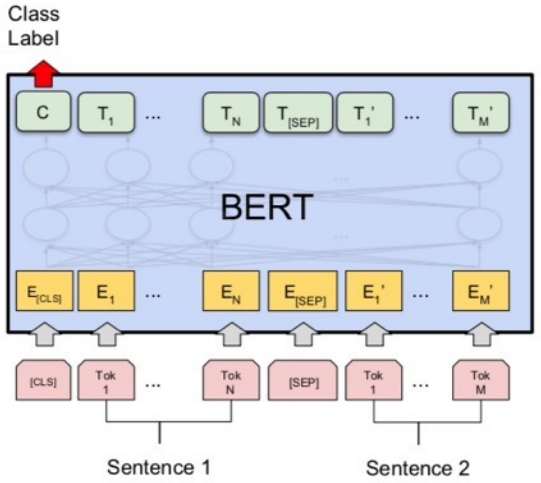
BERT (Bidirectional Encoder Representations from Transformers)

Overall pre-training and fine-tuning procedures for BERT

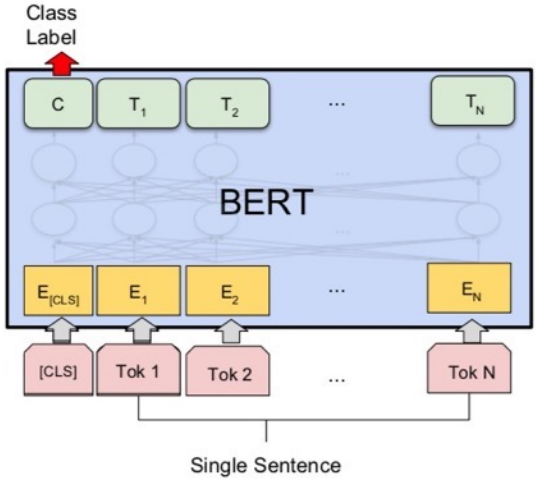


Source: Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2018).
"Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805.

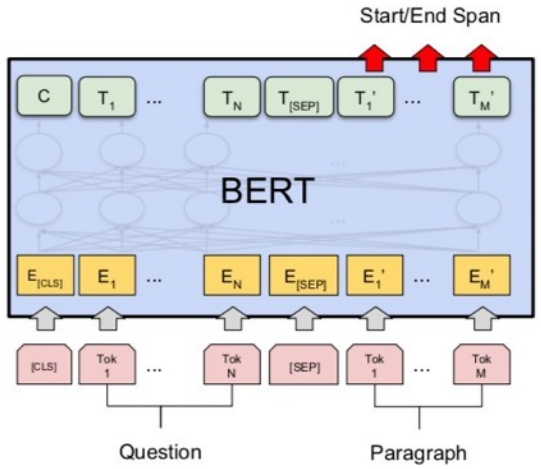
Fine-tuning BERT on Different Tasks



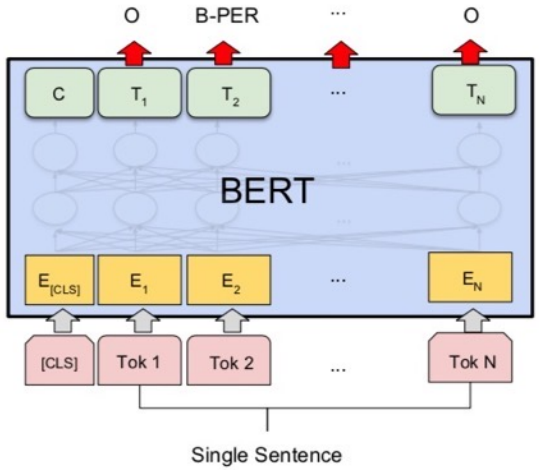
(a) Sentence Pair Classification Tasks: MNLI, QQP, QNLI, STS-B, MRPC, RTE, SWAG



(b) Single Sentence Classification Tasks: SST-2, CoLA



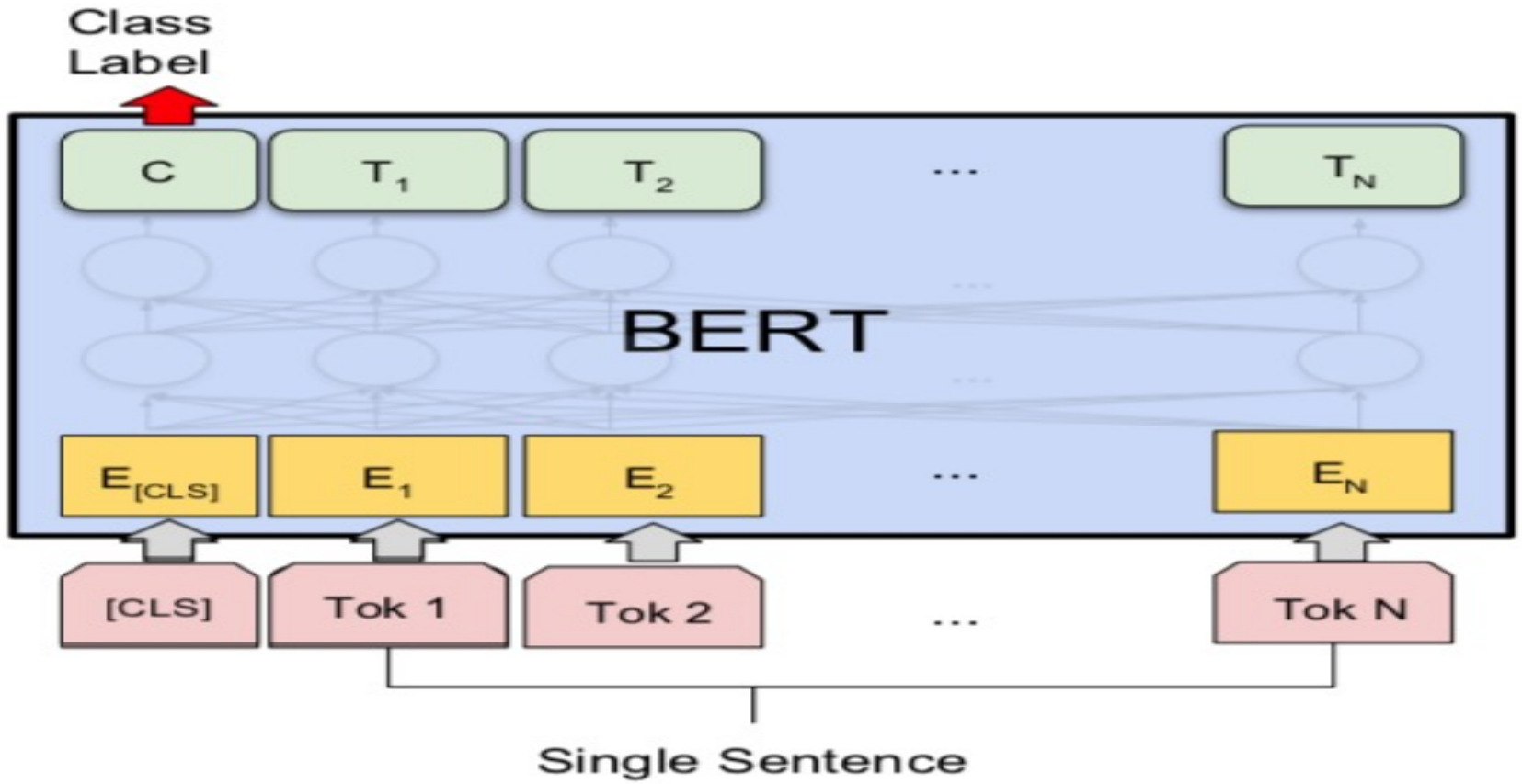
(c) Question Answering Tasks: SQuAD v1.1



(d) Single Sentence Tagging Tasks: CoNLL-2003 NER

Source: Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2018). "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805.

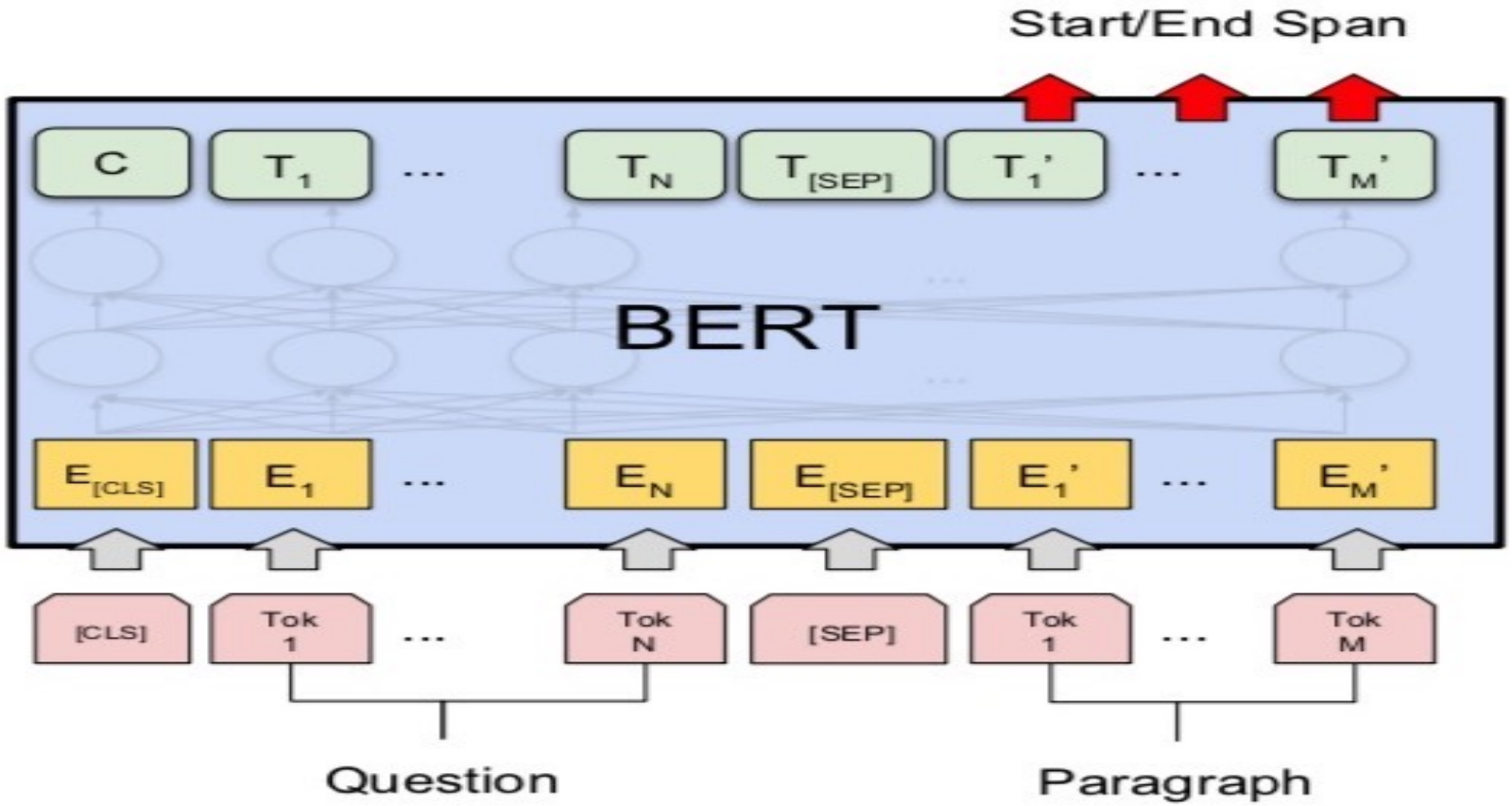
Sentiment Analysis: Single Sentence Classification



(b) Single Sentence Classification Tasks: SST-2, CoLA

Source: Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2018). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." arXiv preprint arXiv:1810.04805

Fine-tuning BERT on Question Answering (QA)

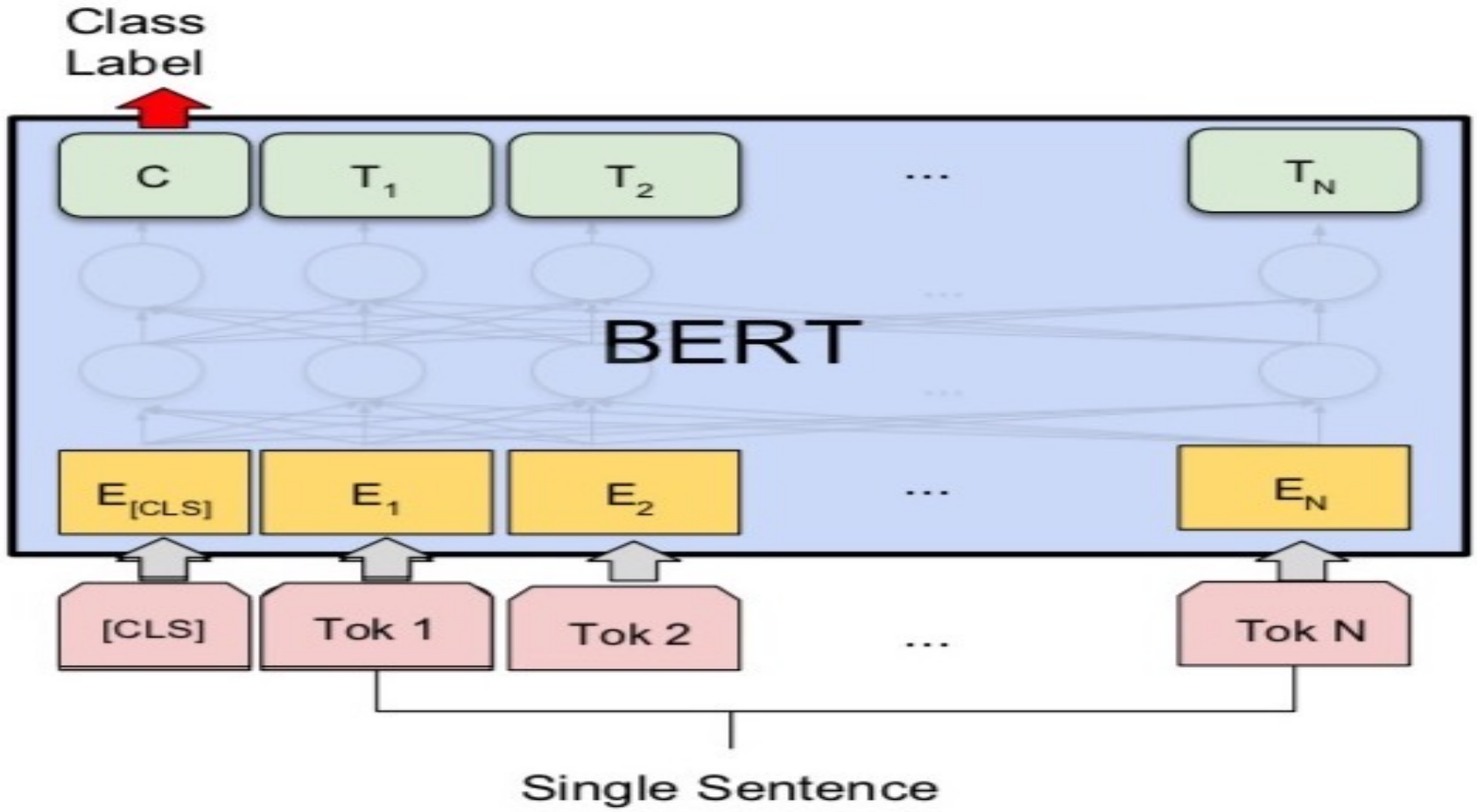


(c) Question Answering Tasks: SQuAD v1.1

Source: Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2018). "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805.

Fine-tuning BERT on Dialogue

Intent Detection (ID; Classification)

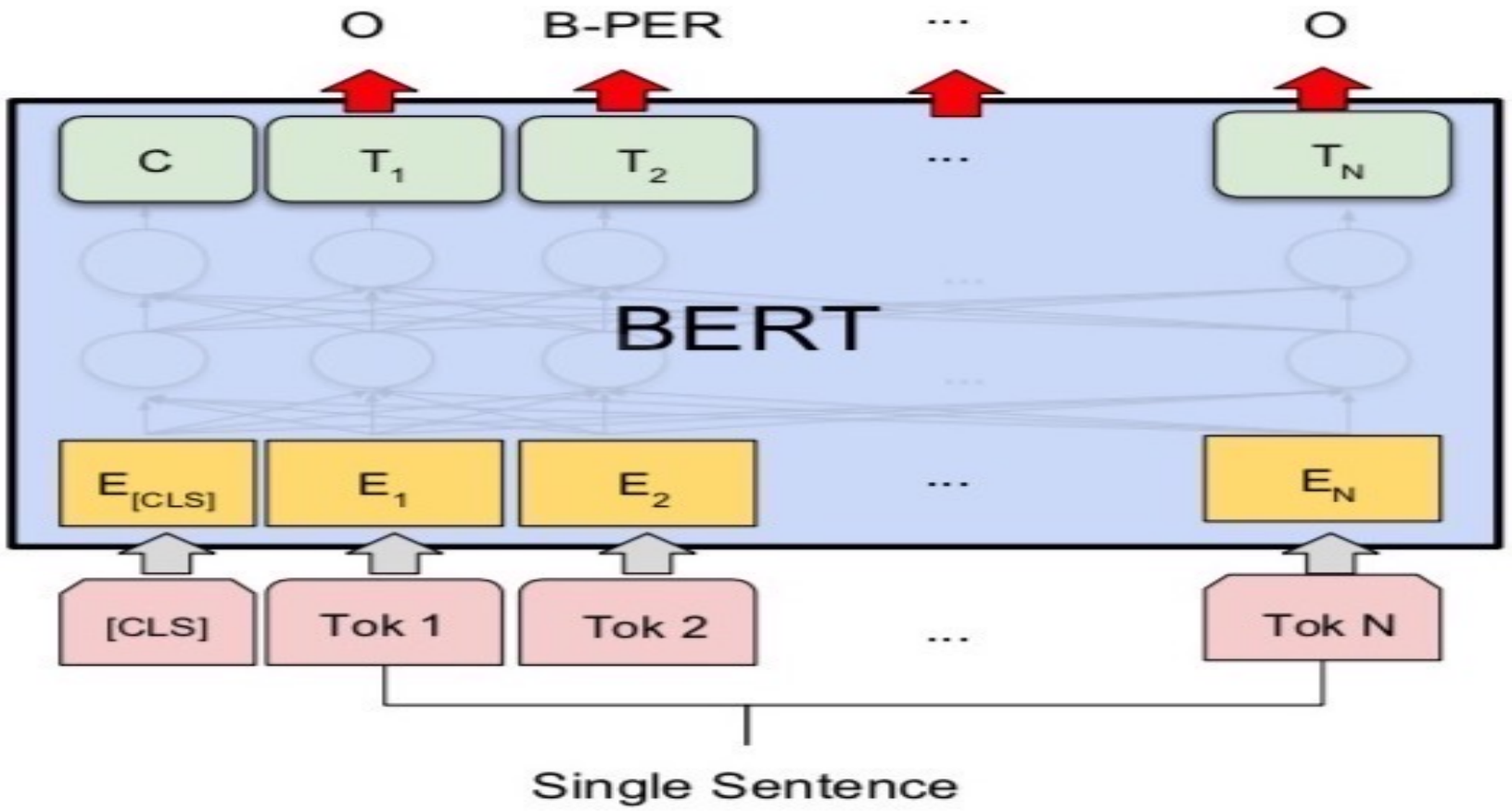


(b) Single Sentence Classification Tasks: SST-2, CoLA

Source: Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2018). "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805.

Fine-tuning BERT on Dialogue

Slot Filling (SF)



(d) Single Sentence Tagging Tasks:
CoNLL-2003 NER

Source: Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2018).
"Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805.

Key Features of the Transformer Model

- **Self-Attention Mechanism**: Allows the model to weigh the importance of different words in a sentence, regardless of their position.
- **Positional Encoding**: Since Transformers don't use recurrence (like RNNs), positional encodings are added to input embeddings to retain word order information.
- **Multi-Head Attention**: The model attends to different words simultaneously in multiple ways, capturing various relationships.
- **Feed-Forward Layers**: After attention, the output passes through dense layers for further transformation.
- **Layer Normalization & Residual Connections**: Improve gradient flow and training stability.
- **Encoder-Decoder Architecture**:
  - **Encoder**: Processes input text and converts it into contextual embeddings.
  - **Decoder**: Generates output text, often used in translation or text generation tasks.

Source: 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

Popular Transformer-Based Models

- **BERT (Bidirectional Encoder Representations from Transformers)**
  - Used for tasks like classification and question answering.
- **GPT (Generative Pre-trained Transformer)**
  - Generates text based on input prompts.
- **T5 (Text-To-Text Transfer Transformer)**
  - Converts all NLP tasks into a text-to-text format.
- **ViT (Vision Transformer)**
  - Applies Transformer architecture to computer vision.

Source: 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

Tokens in NLP (Text Processing)

Aspect	Description
Tokenization	Splitting text into meaningful units (words, subwords, or characters).
Subword Tokens	Methods like Byte Pair Encoding (BPE) and WordPiece break words into reusable subunits (e.g., "unhappiness" → ["un", "happiness"]).
Embeddings	Converts tokens into numerical vectors for processing.
Semantic Role Labeling (SRL)	Identifies sentence structure by assigning roles to tokens (e.g., subject, object).
Special Tokens	[CLS], [SEP]

Source: 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

Tokens in CV (Image Processing)

Aspect	Description
Patch Tokens	Vision Transformers (ViT) split an image into small patches (e.g., 16x16 pixels), treating them as tokens.
Positional Encoding	Since images lack inherent sequence order like text, positional embeddings help ViTs understand spatial structure.
Midjourney API Tokens	Midjourney's AI processes text prompts into image tokens, converting descriptions into AI-generated art.
CLIP Tokens	OpenAI's CLIP model tokenizes both text and images, allowing cross-modal understanding (e.g., "dog" matches a picture of a dog).

Source: 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

Tokens in NLP vs CV

Feature	NLP Tokens	CV Tokens
Basic Unit	Words, subwords, or characters	Image patches (e.g., 16×16 pixel grids)
Processing	Tokenized using BPE, WordPiece, SentencePiece	Tokenized as patch embeddings
Positional Encoding?	Needed to retain word order	Needed to retain spatial information
Example Models	BERT, GPT, T5, RAG	ViT, DINOv2, CLIP
Use Case	Text-based tasks (chatbots, summarization, RAG)	Vision-based tasks (image classification, AI art generation)

Source: 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

Attention in NLP (Text Processing)

Aspect	Description
Self-Attention	Each token attends to every other token in a sentence, capturing dependencies across long text sequences.
Scaled Dot-Product Attention	Computes attention scores using query (Q), key (K), and value (V) vectors.
Multi-Head Attention (MHA)	Improves attention by using multiple attention heads that learn different relationships.
Causal Attention (Decoder-Only Models)	Restricts attention to past tokens only, enabling text generation without looking ahead.
Cross-Attention	The decoder attends to encoder outputs in seq-to-seq tasks like translation (e.g., T5, BART).
Retrieval-Augmented Attention	Fetches external knowledge before generating a response (RAG models).

Source: 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

Attention in CV (Vision Processing)

Aspect	Description
Self-Attention in ViTs	Treats images as a sequence of patches (like words in text) and applies attention to learn spatial relationships.
Multi-Head Attention in ViTs	Similar to NLP, multiple attention heads capture different visual features.
Positional Encoding in Vision	Since images lack inherent order, positional embeddings help maintain spatial structure.
Cross-Attention in Multimodal AI	Used in models like CLIP and Midjourney, where text descriptions attend to visual features.
Attention Maps in Vision	Heatmaps showing which image regions the model focuses on (e.g., for explainability).

Source: 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

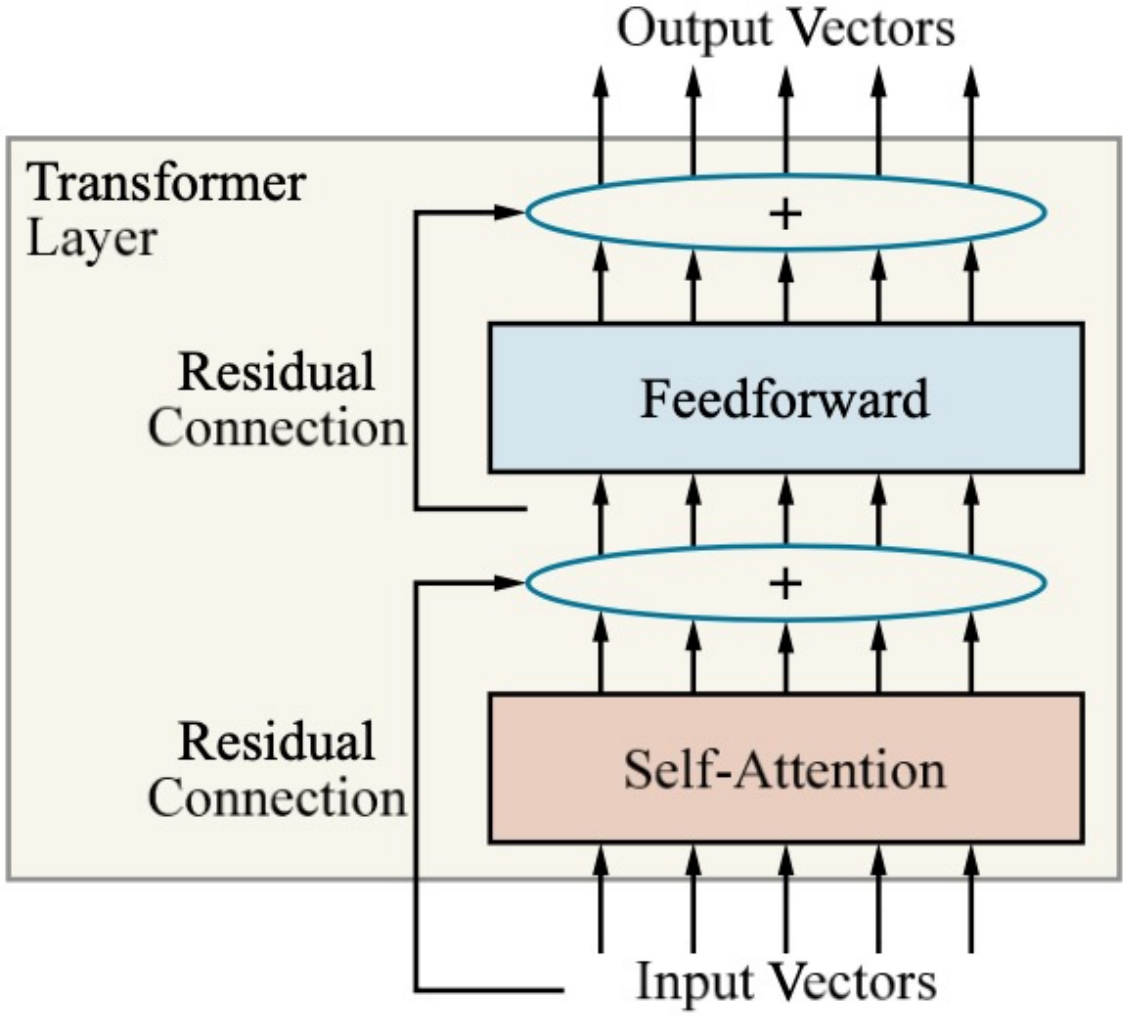
Attention in NLP vs CV

Feature	NLP Attention (Text)	CV Attention (Images)
Input Format	Tokenized text sequences	Image patches
Basic Unit	**Words/subwords**	**Pixels/patches**
Role of Attention	Captures long-range dependencies	Learns spatial & contextual relationships
Sequential Processing?	No, operates on full input (parallelizable)	No, processes patches like text
Model Examples	BERT, GPT, T5, RAG	ViT, CLIP, Midjourney

Source: 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

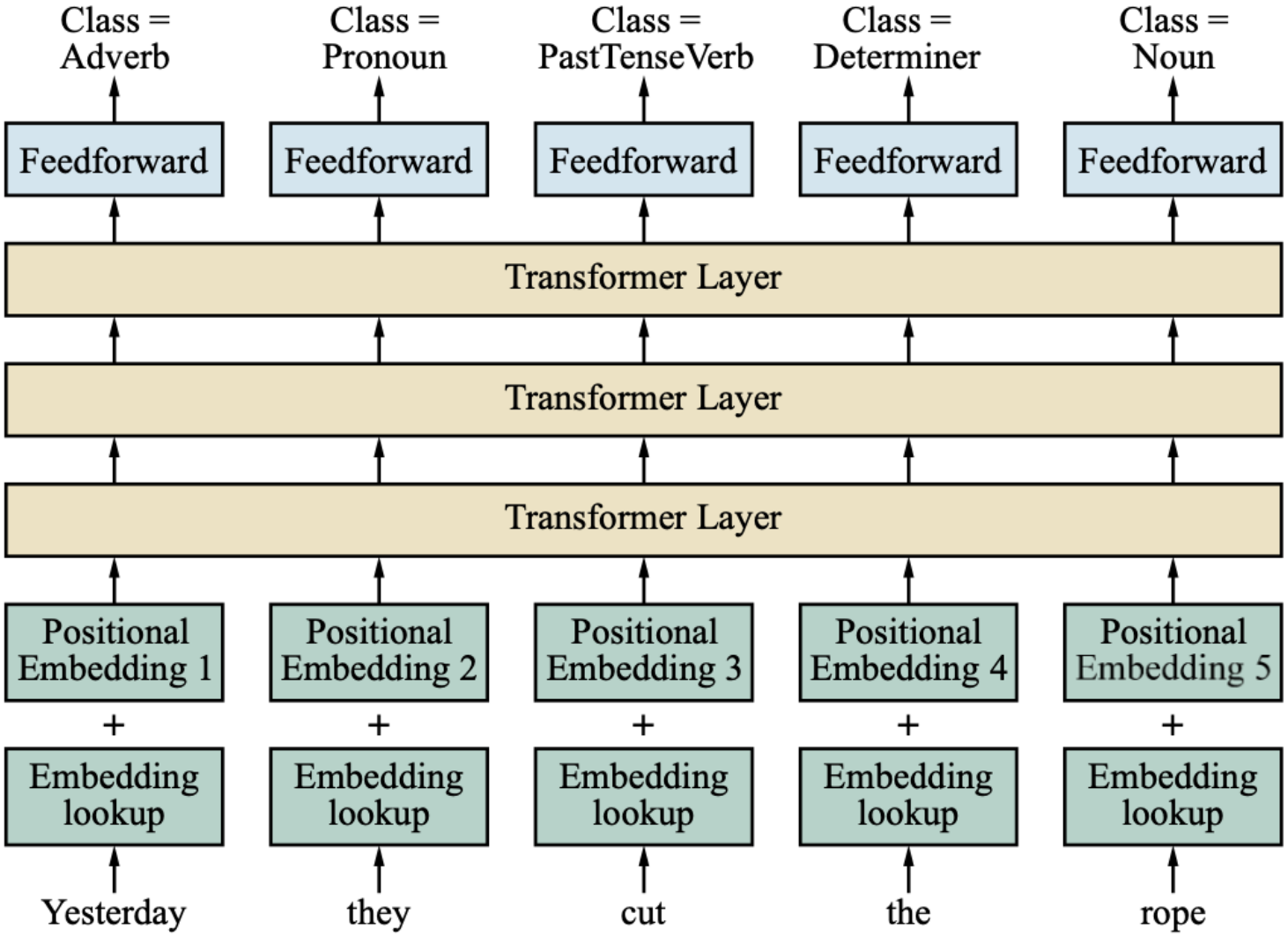
Single-layer Transformer

consists of self-attention,
a feedforward network, and residual connection



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

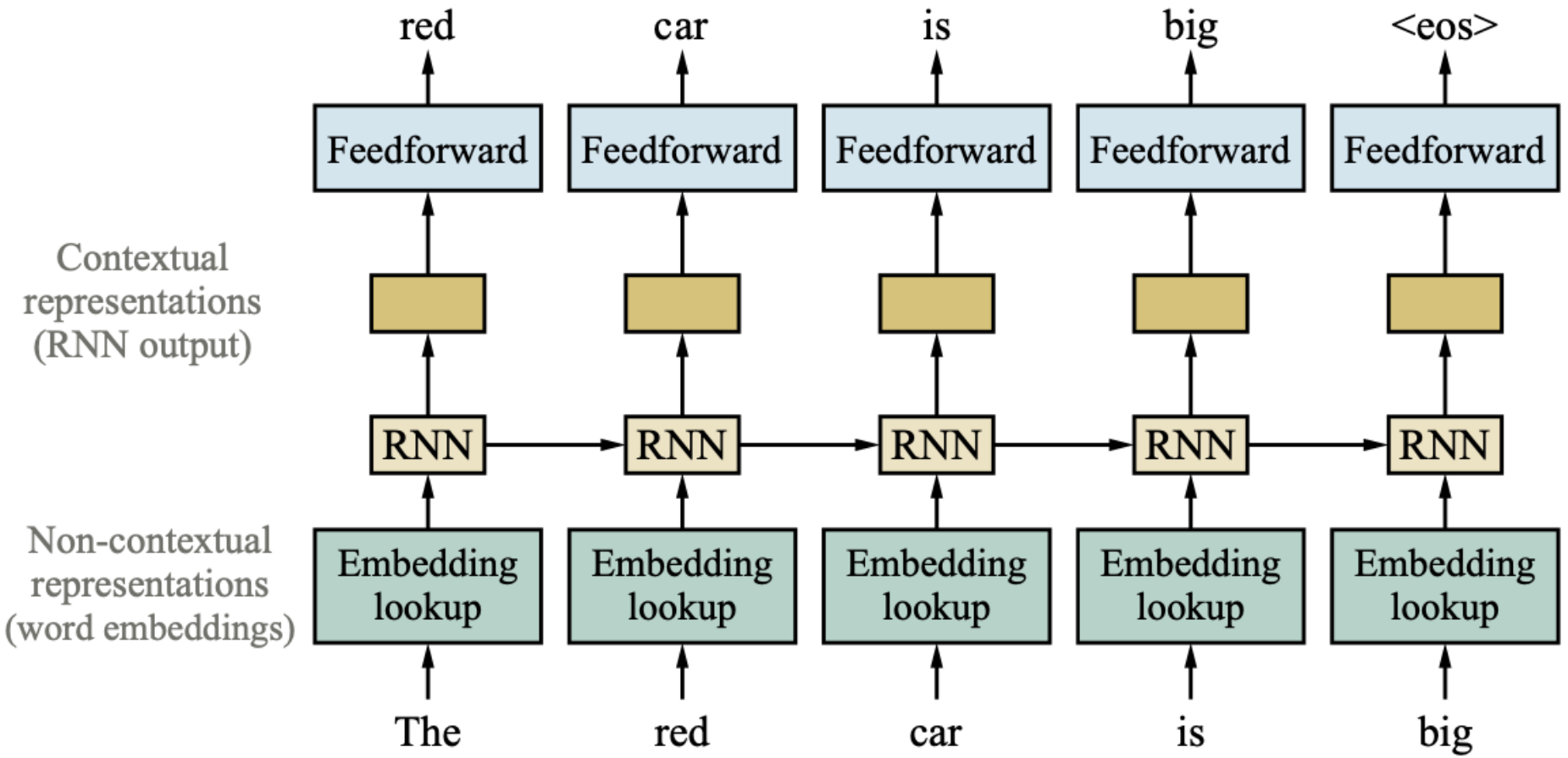
Transformer Architecture for POS Tagging



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

Training Contextual Representations

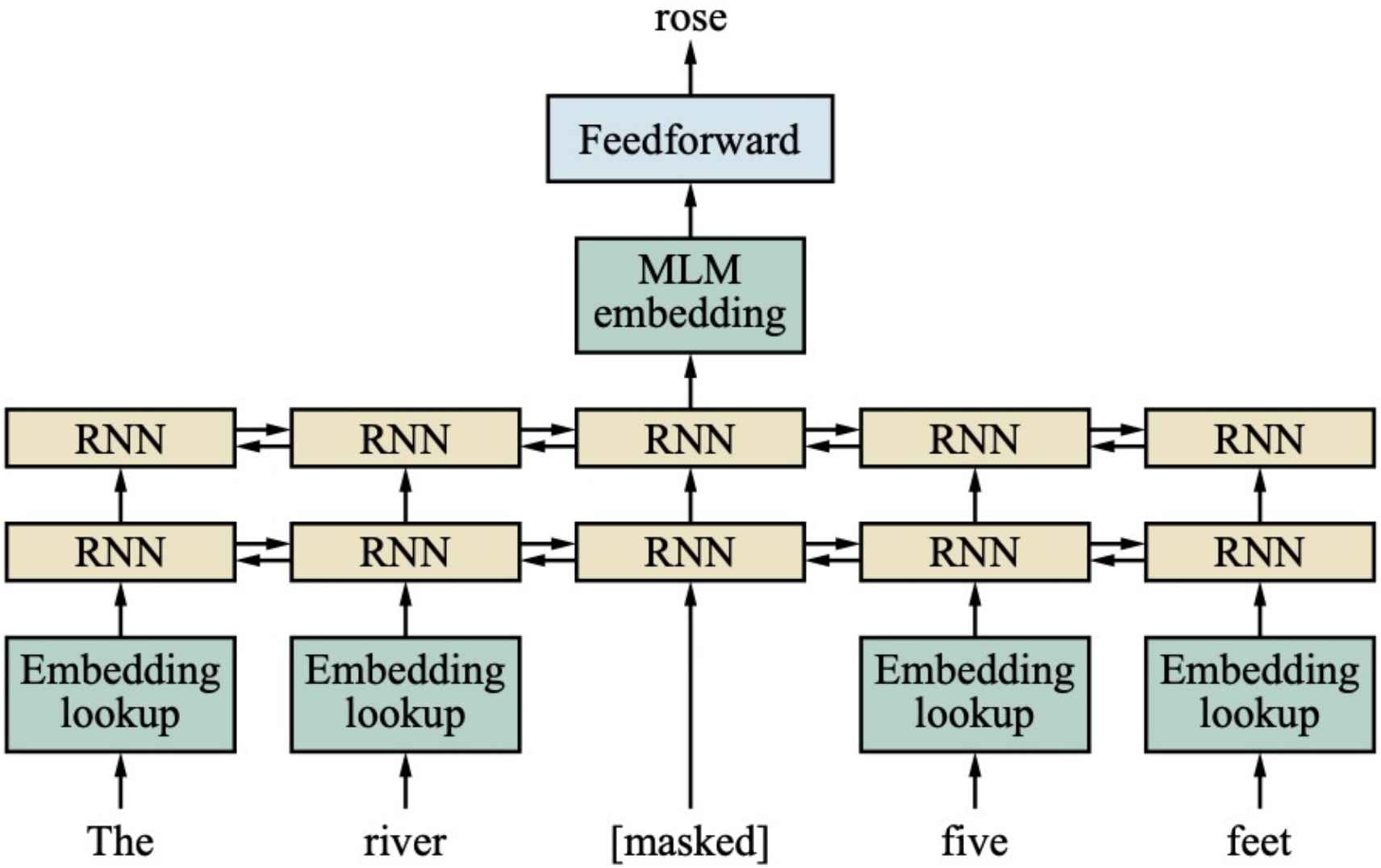
using a left-to-right Language Model



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

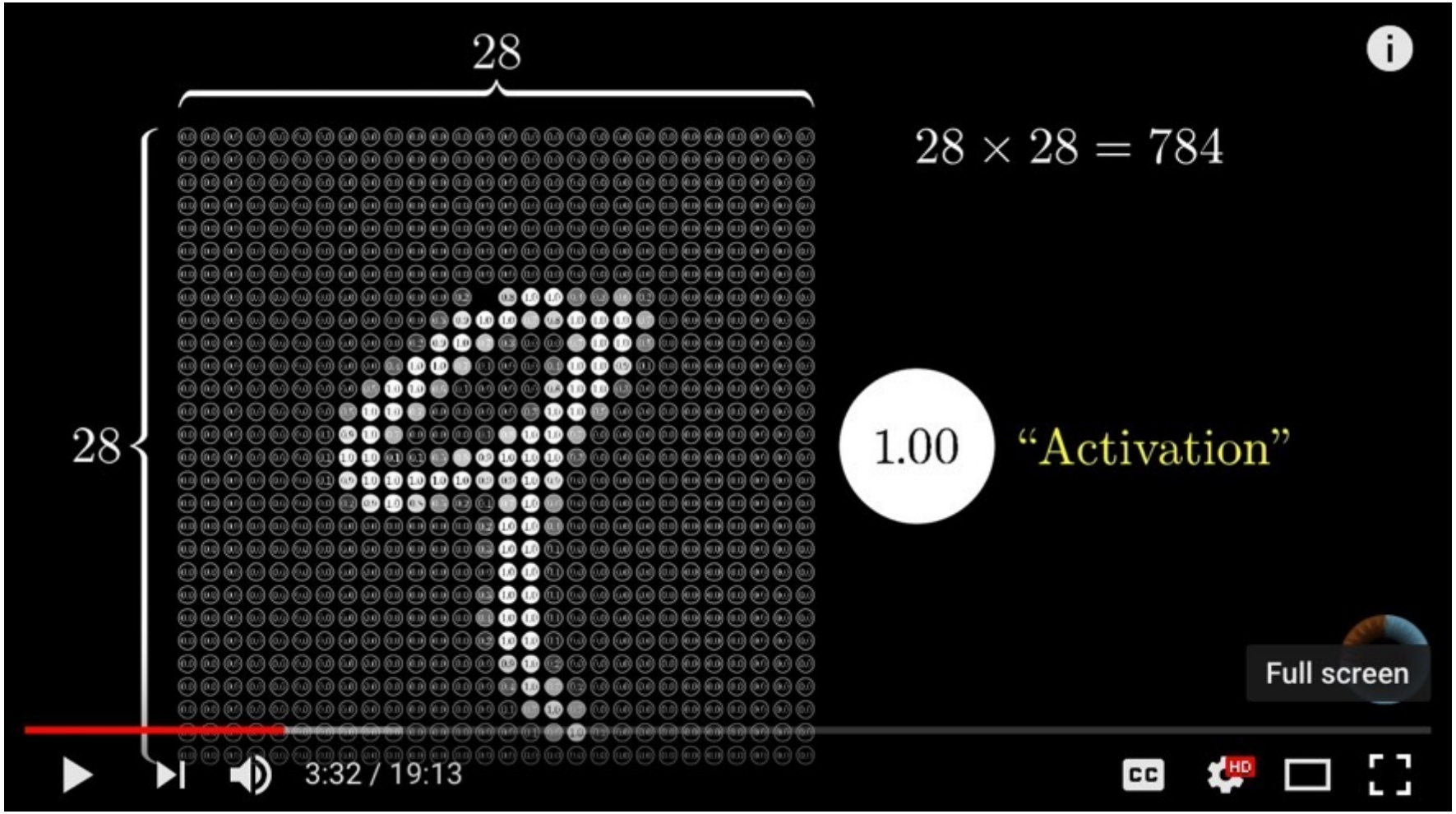
Masked Language Modeling:

Pretrain a Bidirectional Model



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

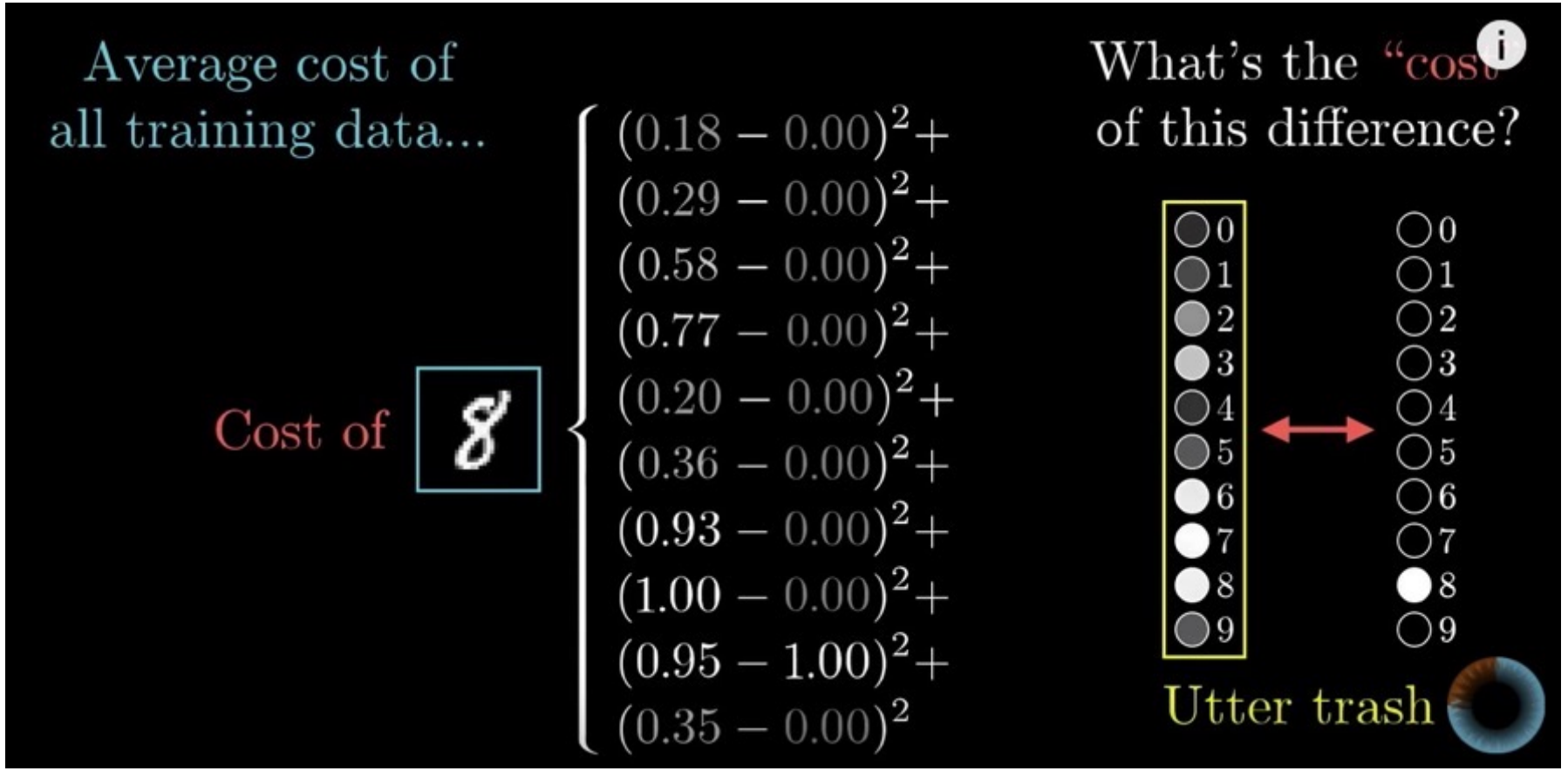
Neural Network and Deep Learning



Source: 3Blue1Brown (2017), But what *is* a Neural Network? | Chapter 1, deep learning, https://www.youtube.com/watch?v=aircAruvnKk

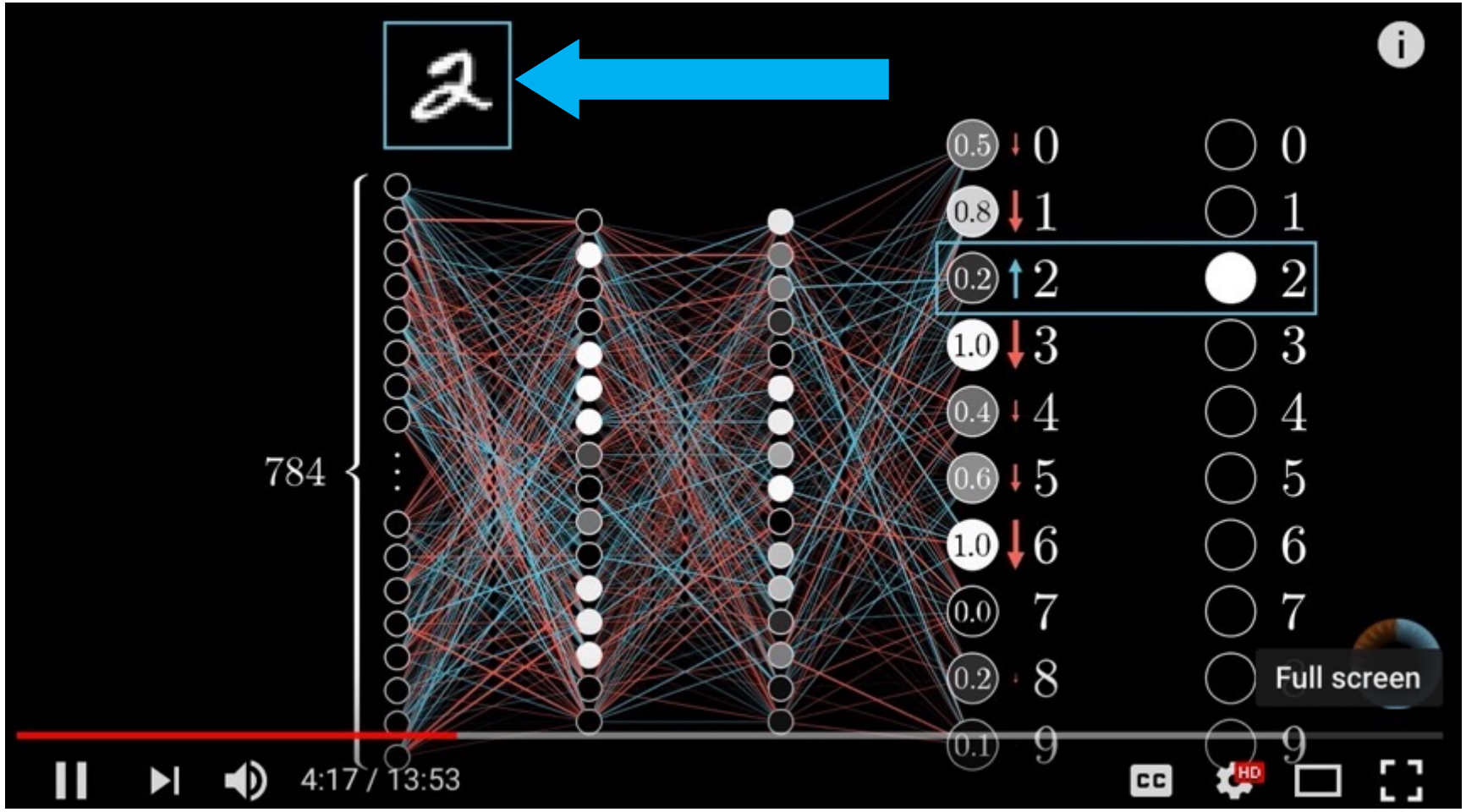
Gradient Descent

how neural networks learn



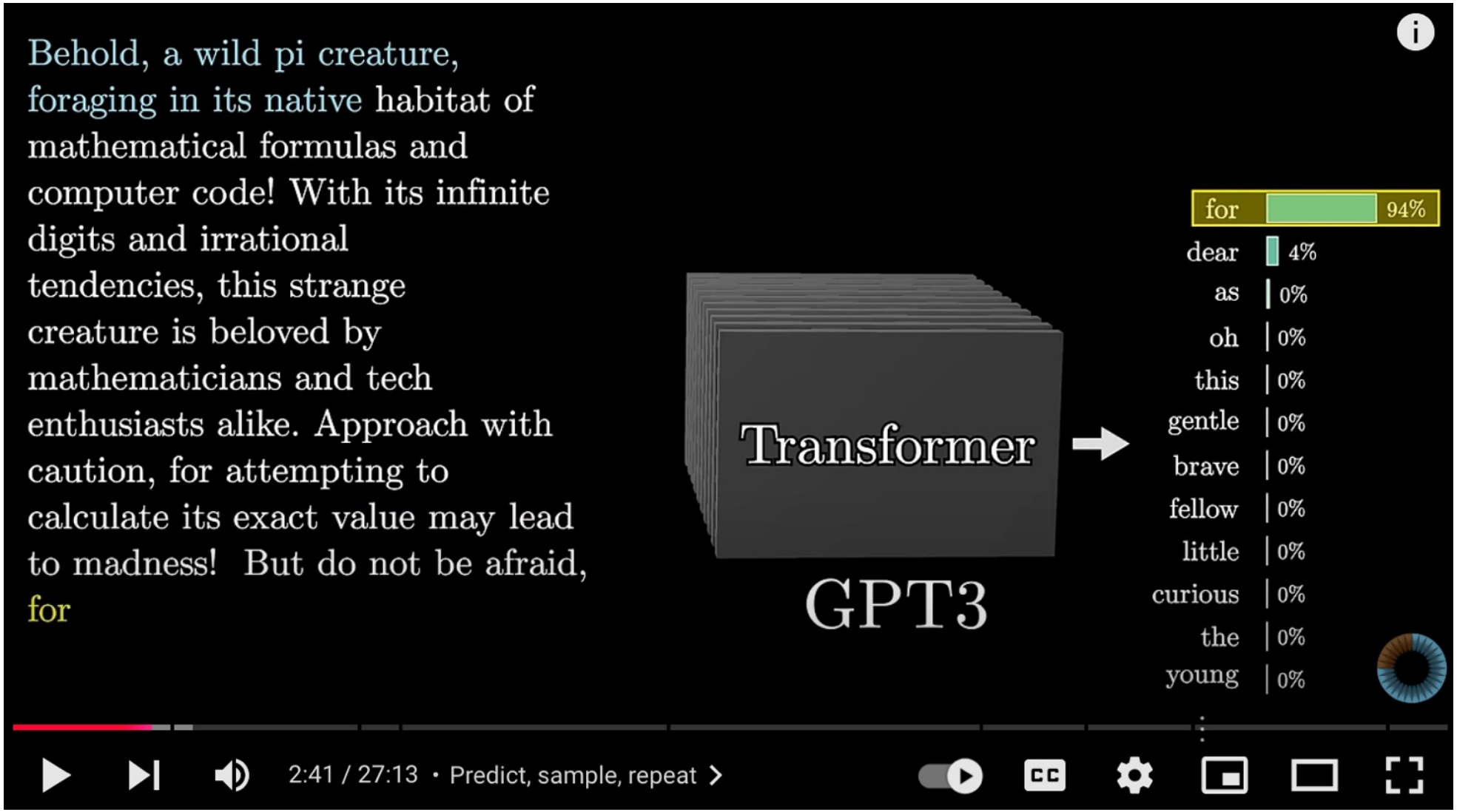
Source: 3Blue1Brown (2017), Gradient descent, how neural networks learn | Chapter 2, deep learning, https://www.youtube.com/watch?v=IHZwWFHWa-w

Backpropagation



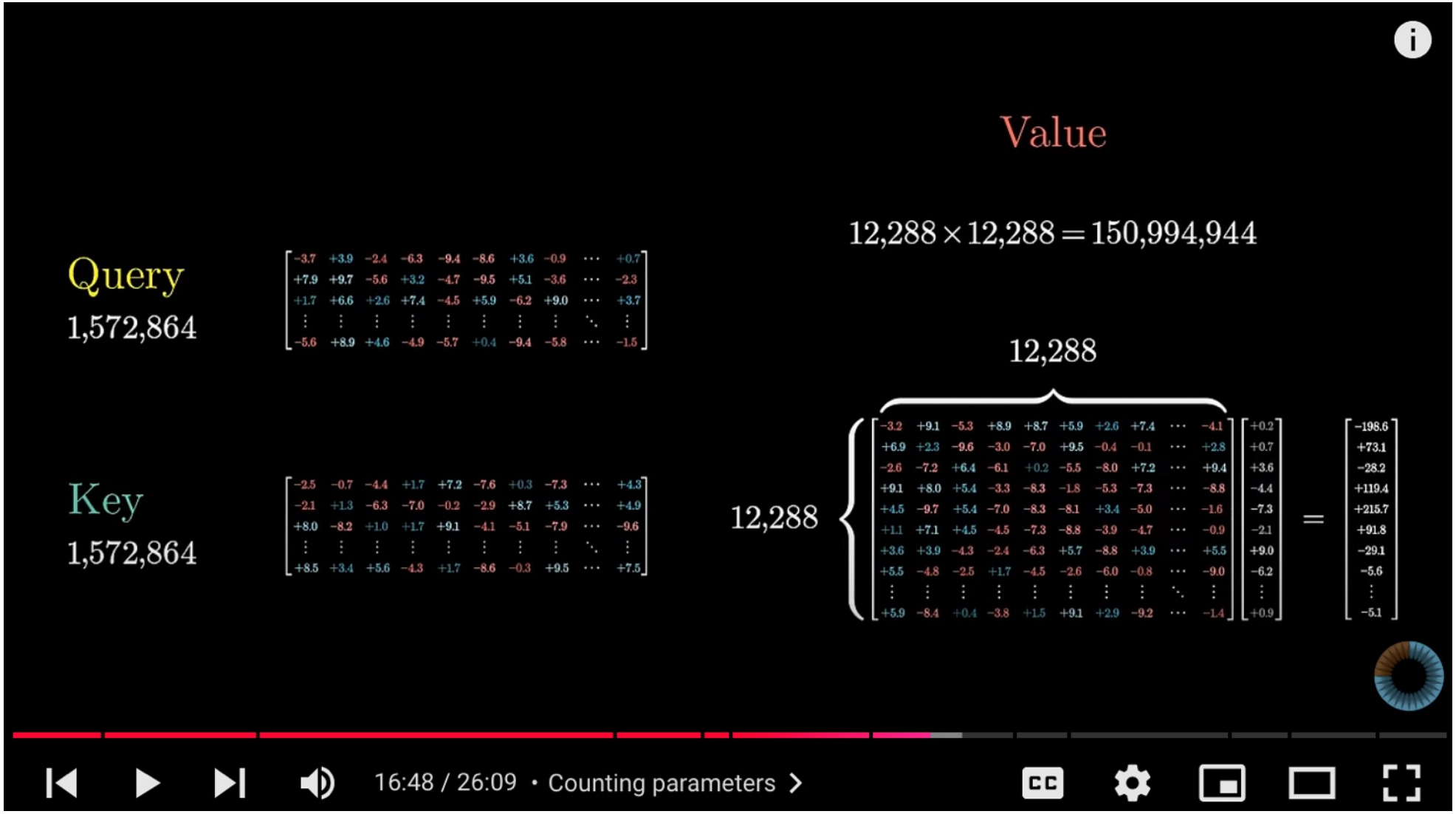
Source: 3Blue1Brown (2017), What is backpropagation really doing? | Chapter 3, deep learning, https://www.youtube.com/watch?v=Ilg3gGewQ5U

Transformers (how LLMs work)



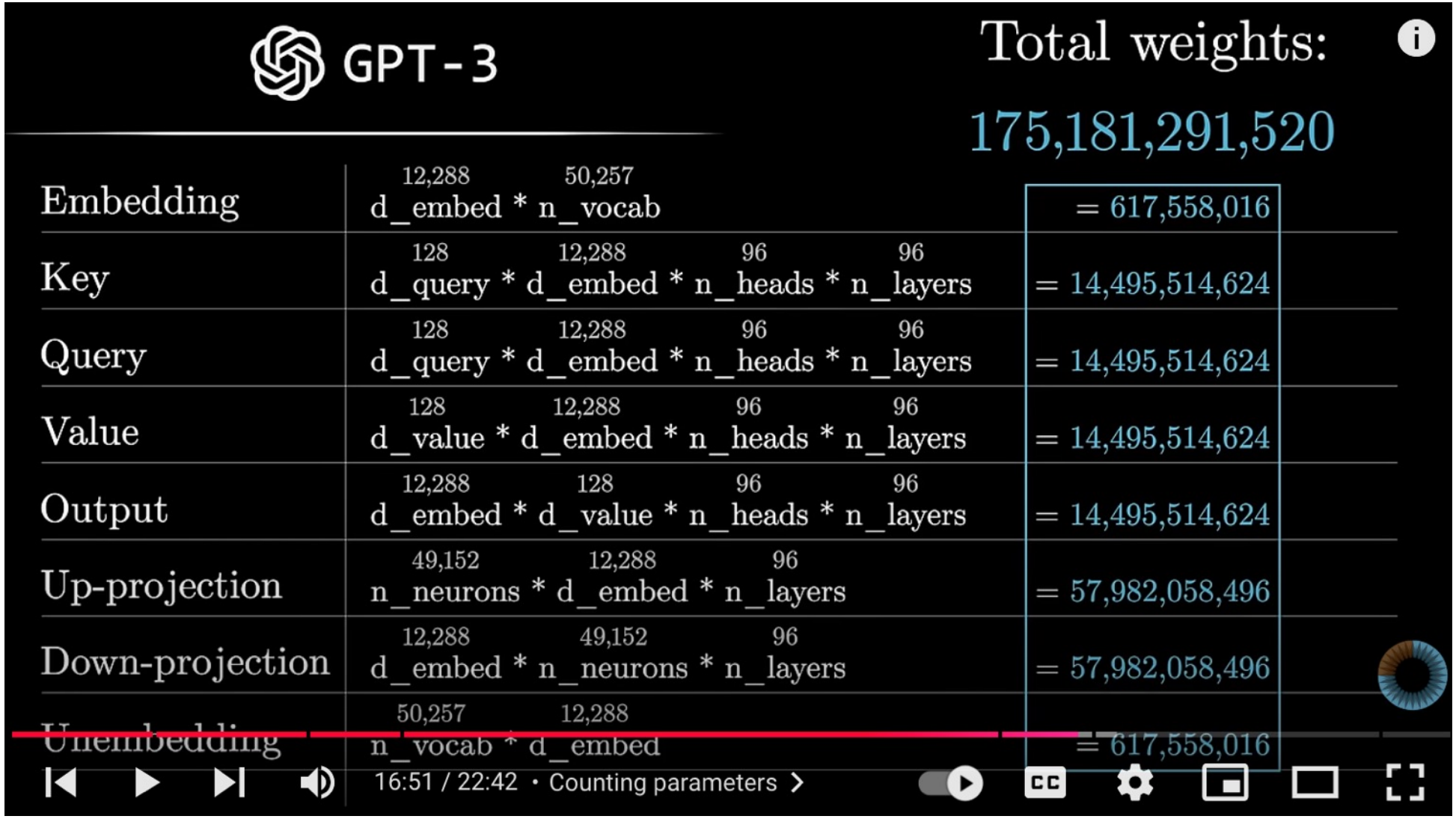
Source: 3Blue1Brown (2024), Transformers (how LLMs work) explained visually | DL5, https://www.youtube.com/watch?v=wjZofJX0v4M

Attention in Transformers



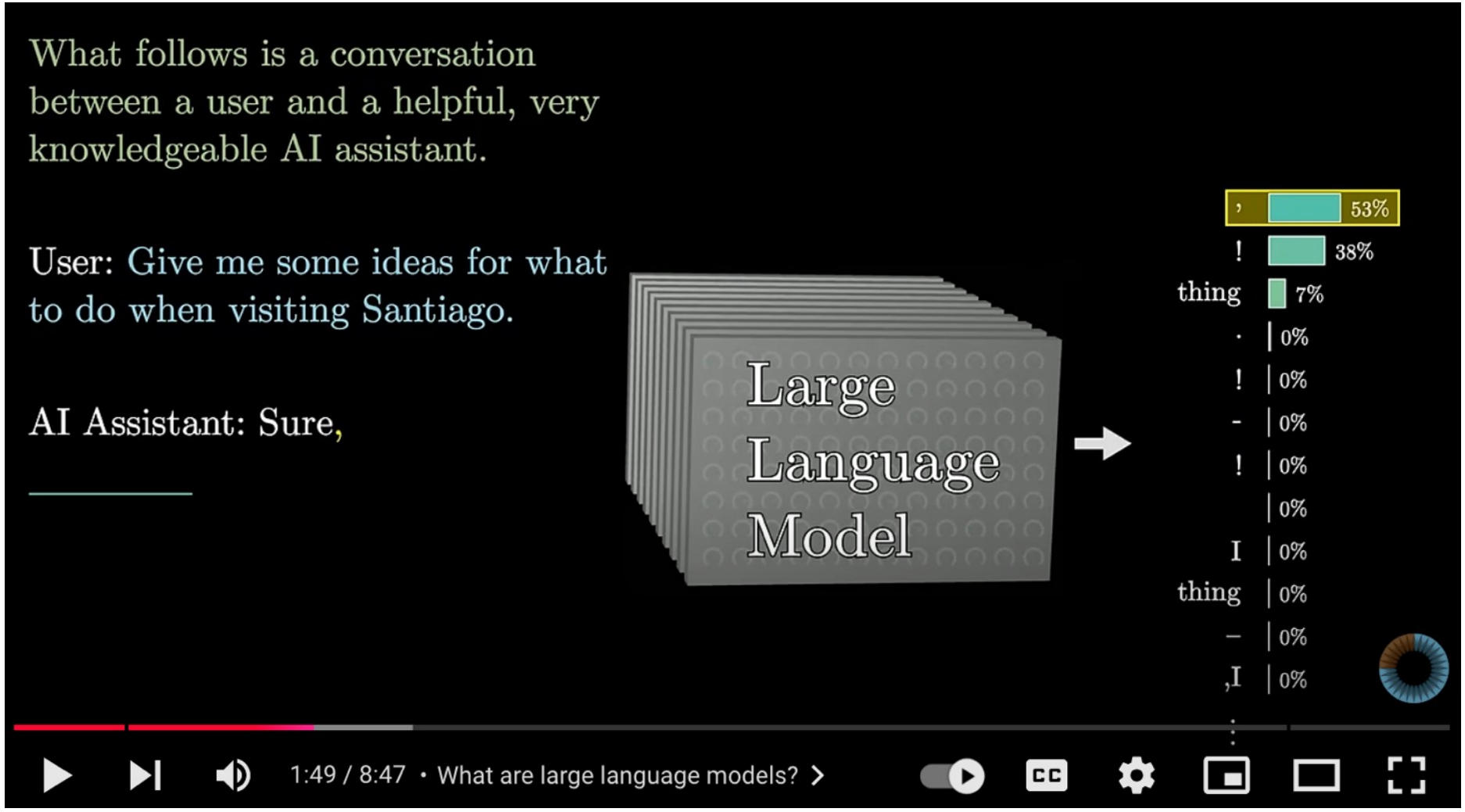
Source: 3Blue1Brown (2024), Attention in transformers, visually explained | DL6, https://www.youtube.com/watch?v=eMlx5fFNoYc

How might LLMs store facts



Source: 3Blue1Brown (2024), How might LLMs store facts | DL7,
https://www.youtube.com/watch?v=9-Jl0dxWQs8

Large Language Models explained briefly



Source: 3Blue1Brown (2024), Large Language Models explained briefly, https://www.youtube.com/watch?v=LPZh9BOjkQs

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

Chatbot

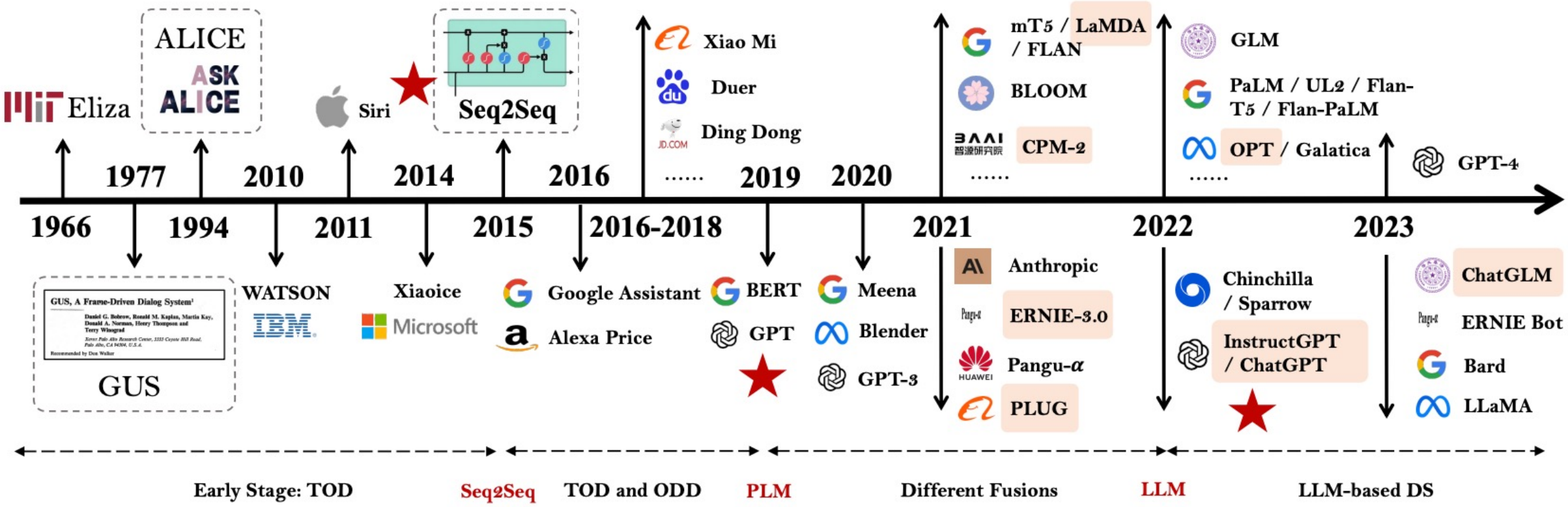
Dialogue System

Intelligent Agent

Conversational AI

The Development of LM-based Dialogue Systems

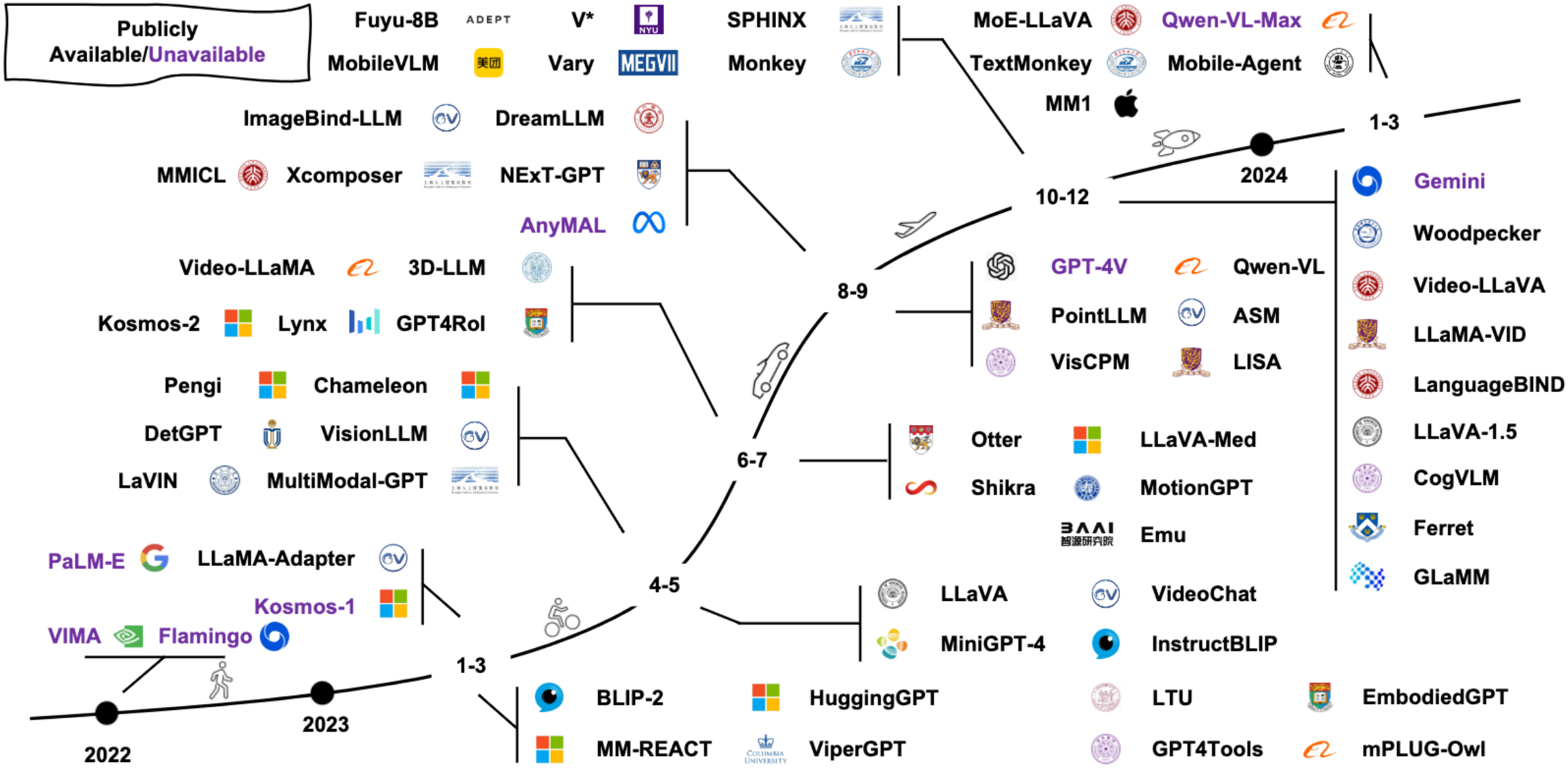
1) Early Stage (1966 - 2015)
2) The Independent Development of TOD and ODD (2015 - 2019)
3) Fusions of Dialogue Systems (2019 - 2022)
4) LLM-based DS (2022 - Now)



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

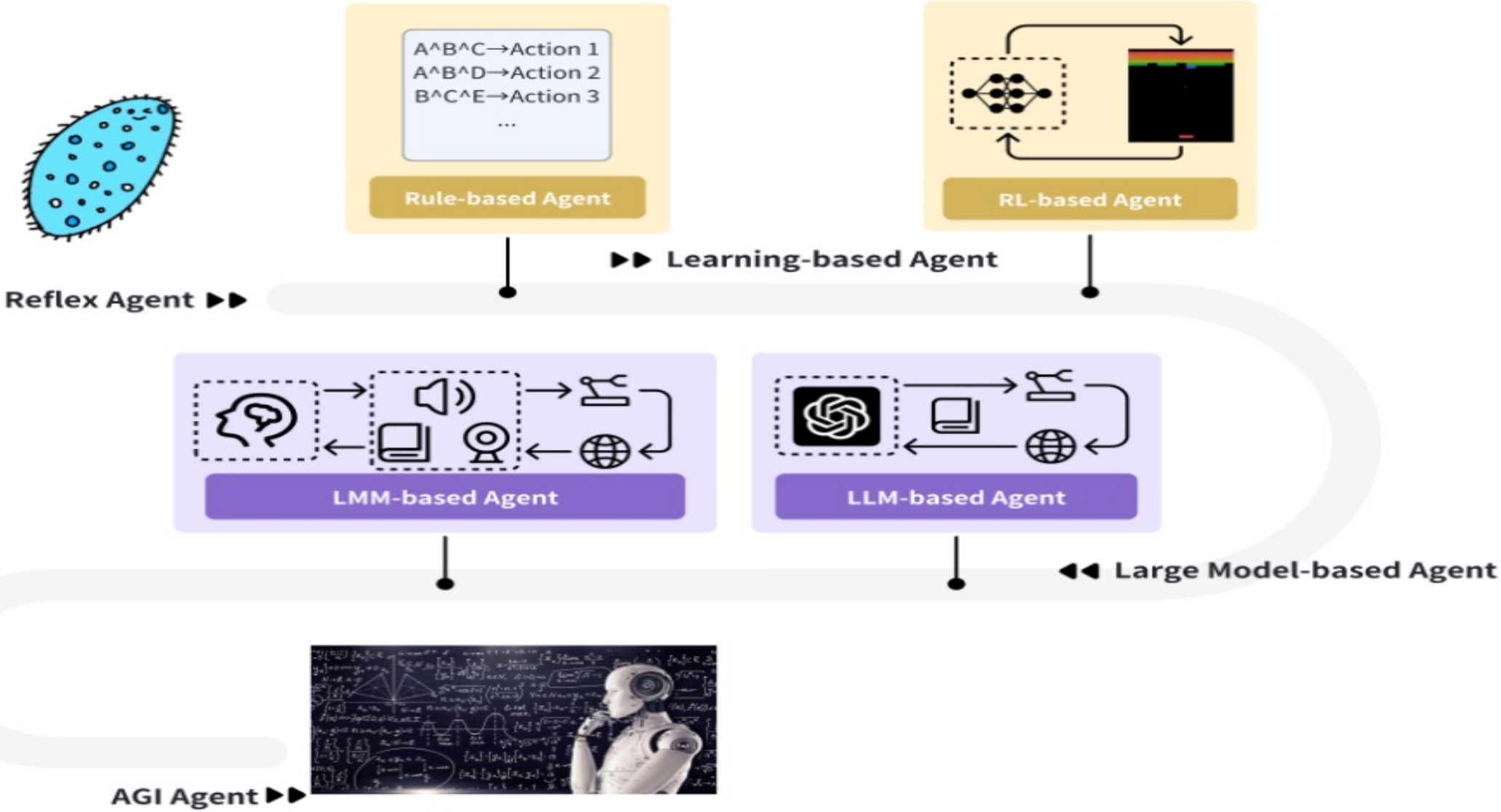
Source: Wang, Hongru, Lingzhi Wang, Yiming Du, Liang Chen, Jingyan Zhou, Yufei Wang, and Kam-Fai Wong. "A Survey of the Evolution of Language Model-Based Dialogue Systems." arXiv preprint arXiv:2311.16789 (2023).

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.

Intelligent Agents Roadmap



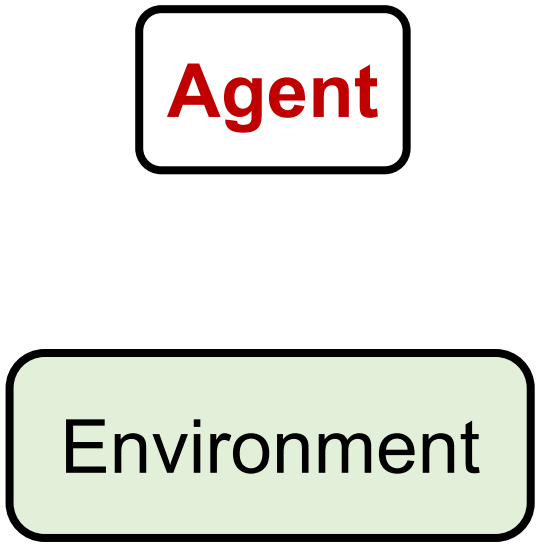
Source: Cheng, Yuheng, Ceyao Zhang, Zhengwen Zhang, Xiangrui Meng, Sirui Hong, Wenhao Li, Zihao Wang et al. "Exploring large language model based intelligent agents: Definitions, methods, and prospects." arXiv preprint arXiv:2401.03428 (2024).

AI Agents

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

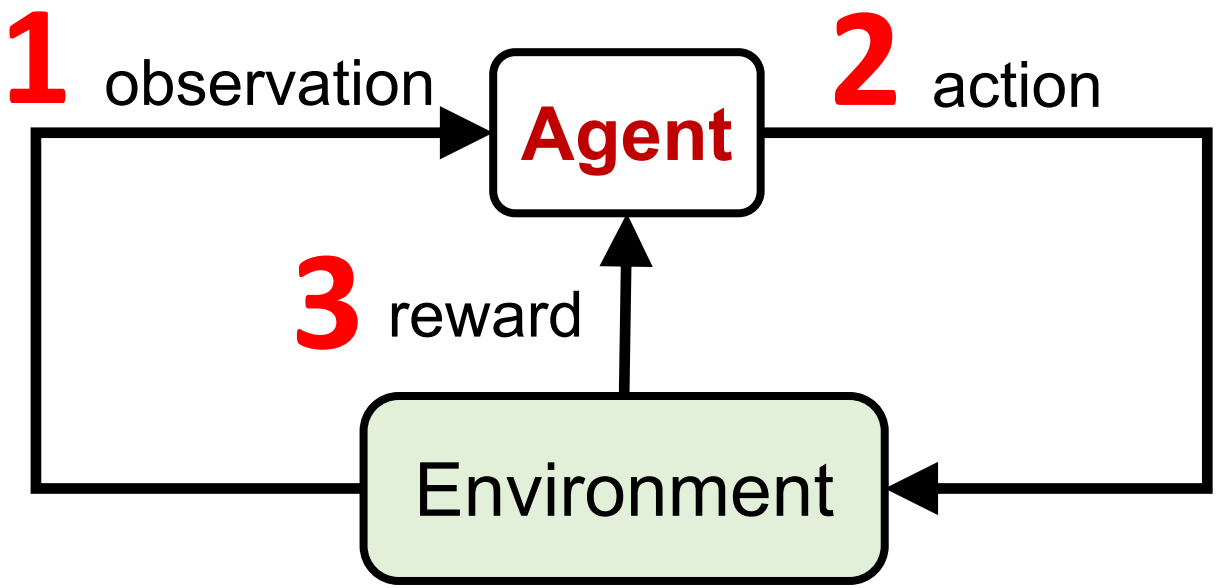
- Evolution of AI Agents
  - LLM-based Agents
  - Multi-modal agents
  - Embodied AI agents in virtual environments
  - Collaborative AI agents

Reinforcement Learning (DL)



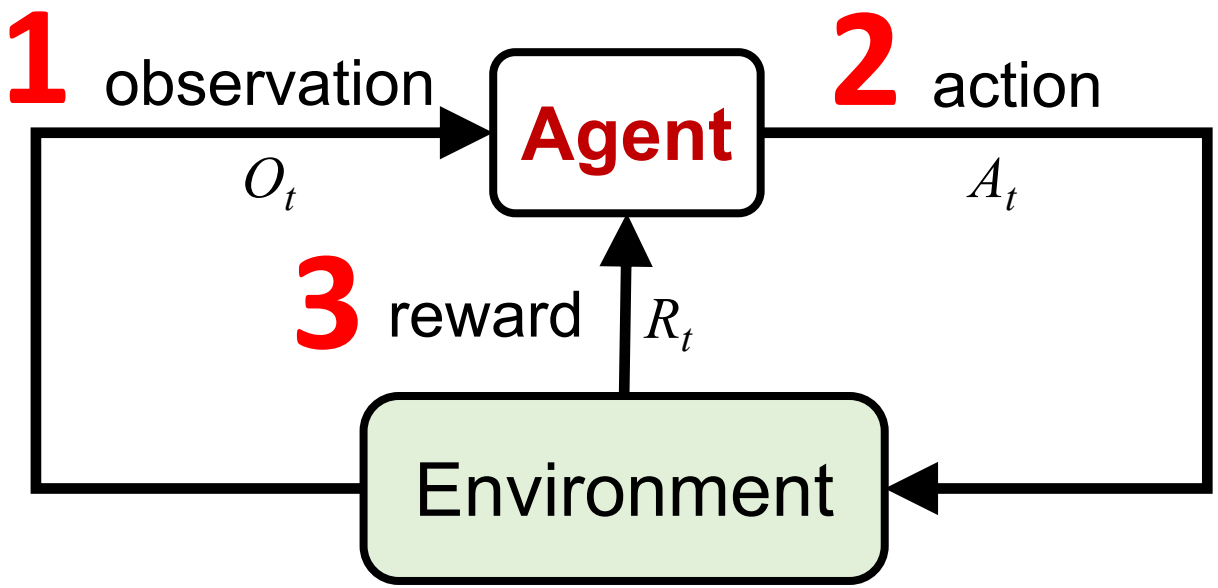
Source: Richard S. Sutton & Andrew G. Barto (2018), Reinforcement Learning: An Introduction, 2nd Edition, A Bradford Book.

Reinforcement Learning (DL)



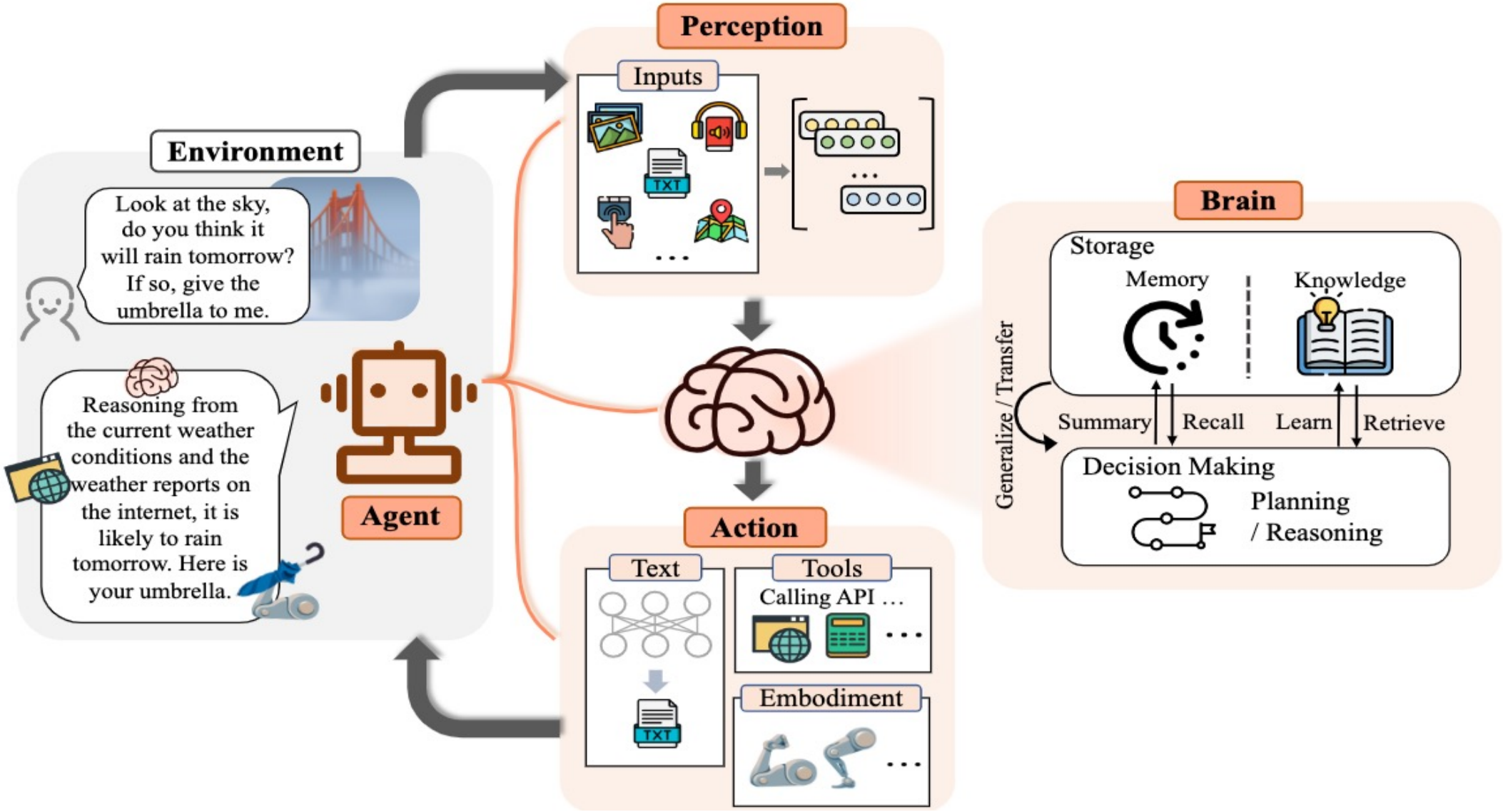
Source: Richard S. Sutton & Andrew G. Barto (2018), Reinforcement Learning: An Introduction, 2nd Edition, A Bradford Book.

Reinforcement Learning (DL)



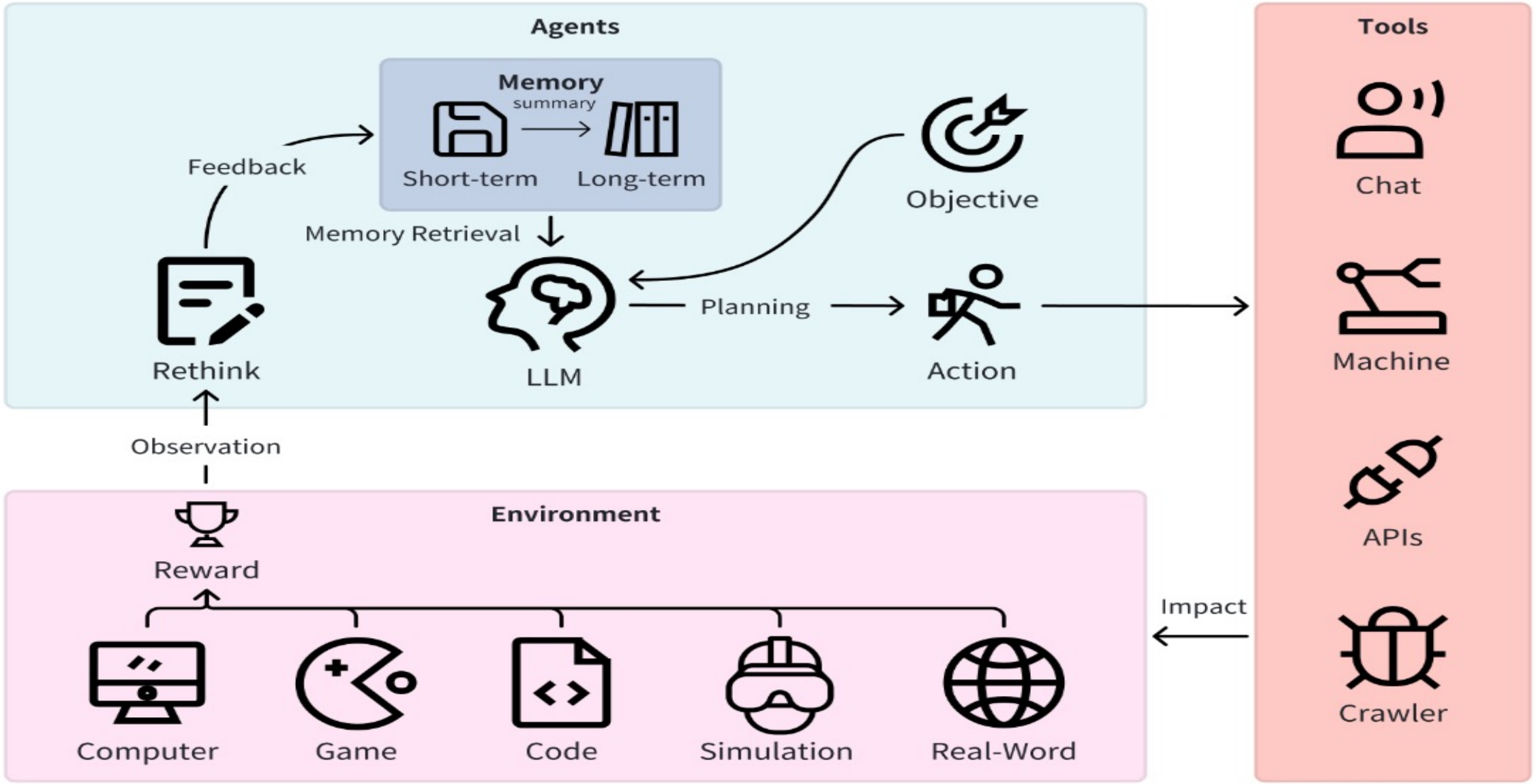
Source: Richard S. Sutton & Andrew G. Barto (2018), Reinforcement Learning: An Introduction, 2nd Edition, A Bradford Book.

Large Language Model (LLM) based Agents



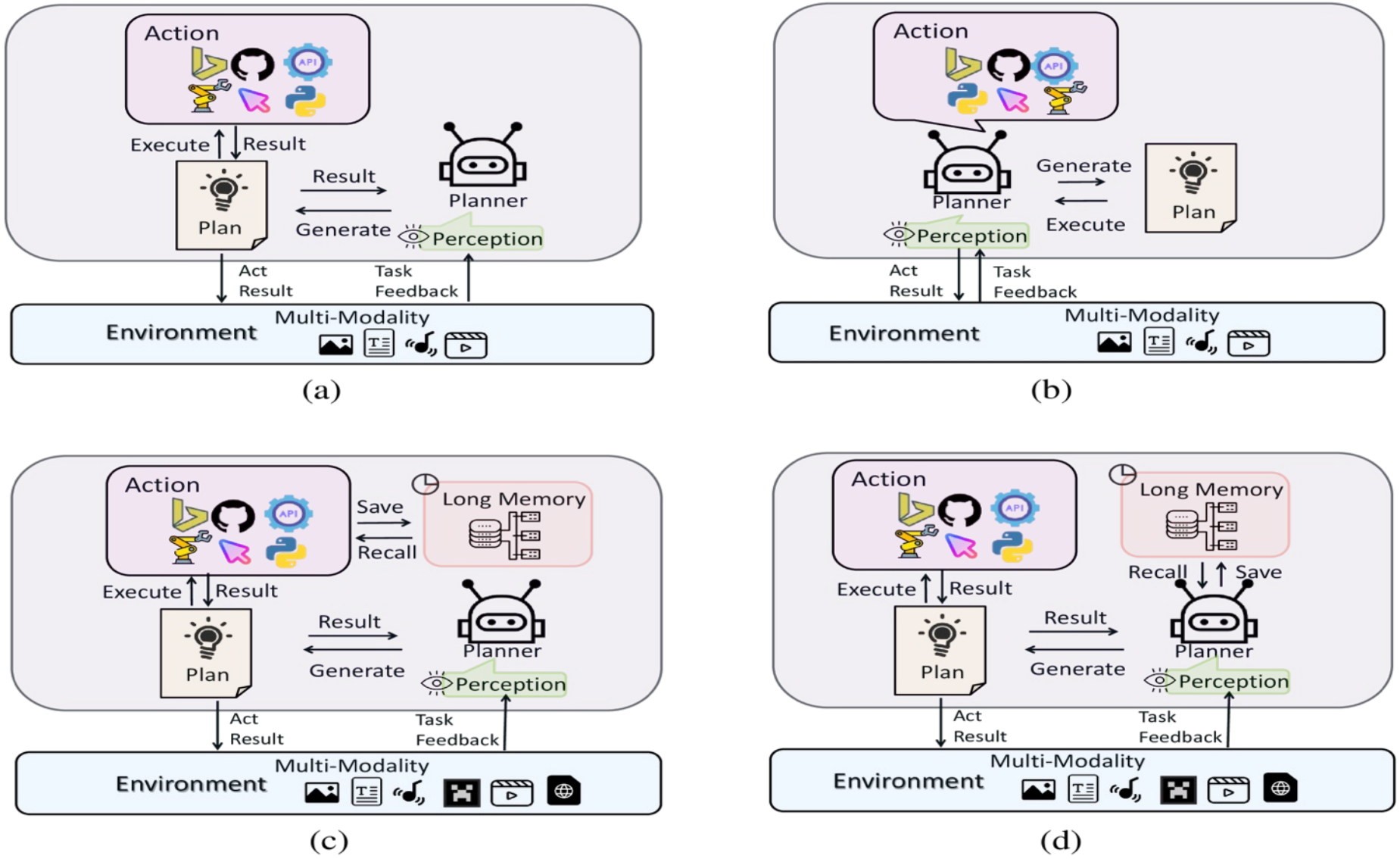
Source: Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., ... & Gui, T. (2023). The rise and potential of large language model based agents: A survey. arXiv preprint arXiv:2309.07864.

LLM-based Agents



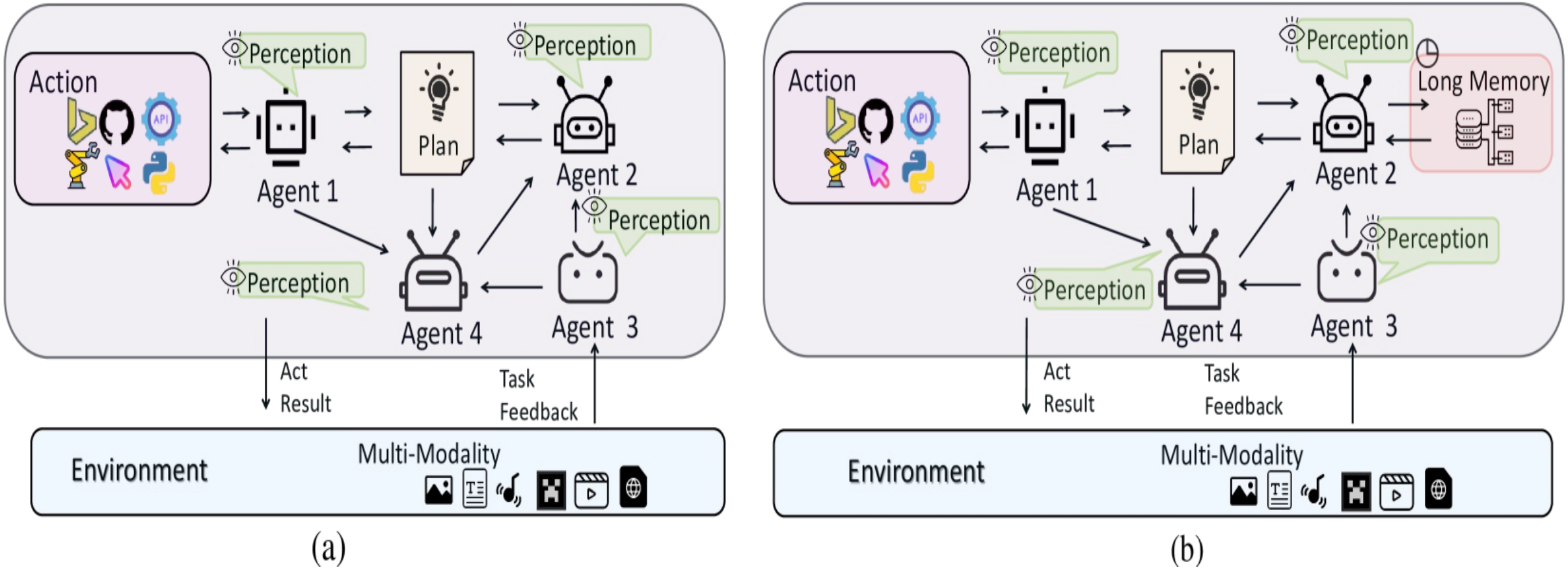
Source: Cheng, Yuheng, Ceyao Zhang, Zhengwen Zhang, Xiangrui Meng, Sirui Hong, Wenhao Li, Zihao Wang et al. "Exploring large language model based intelligent agents: Definitions, methods, and prospects." arXiv preprint arXiv:2401.03428 (2024).

Large Multimodal Agents (LMA)



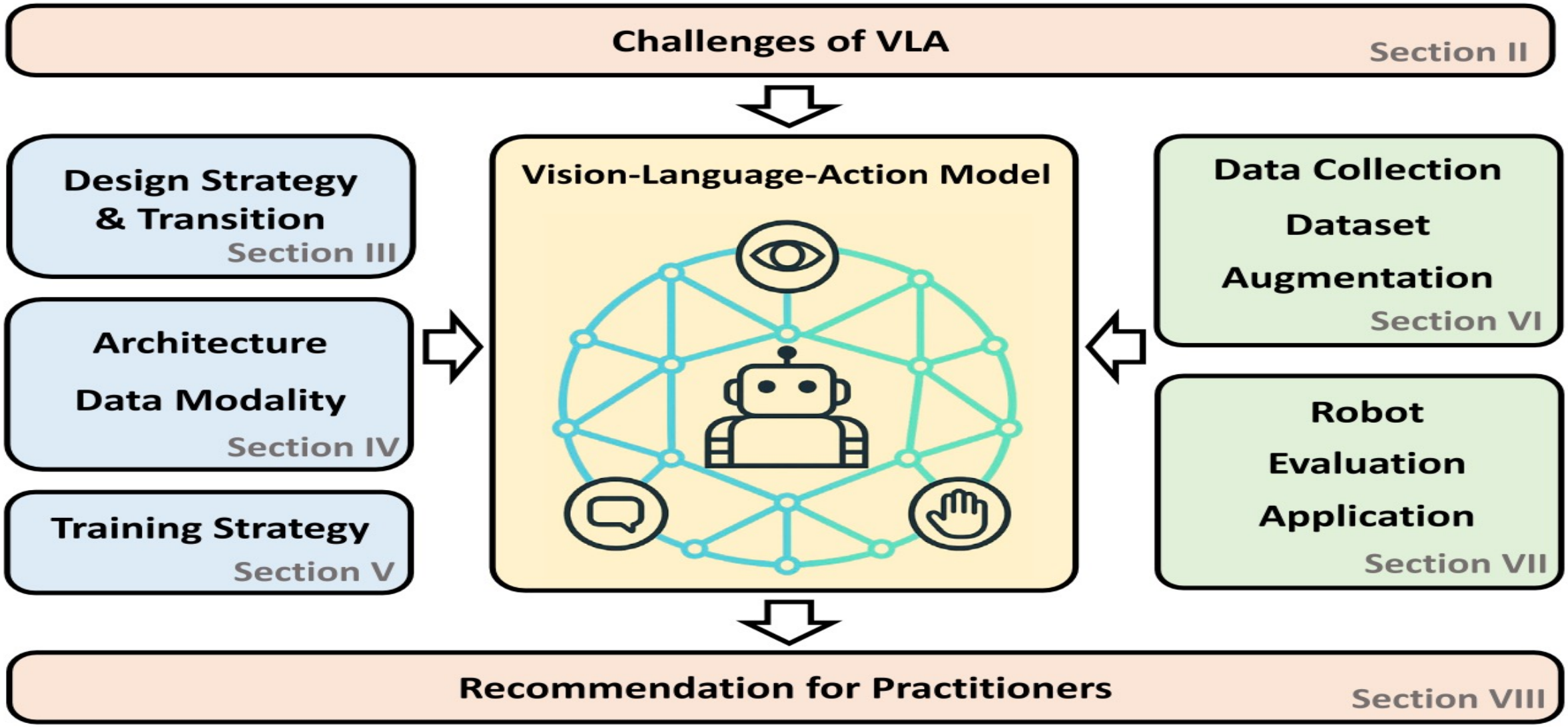
Source: Xie, J., Chen, Z., Zhang, R., Wan, X., & Li, G. (2024). Large Multimodal Agents: A Survey. ArXiv, abs/2402.15116.

Large Multimodal Agents (LMA)



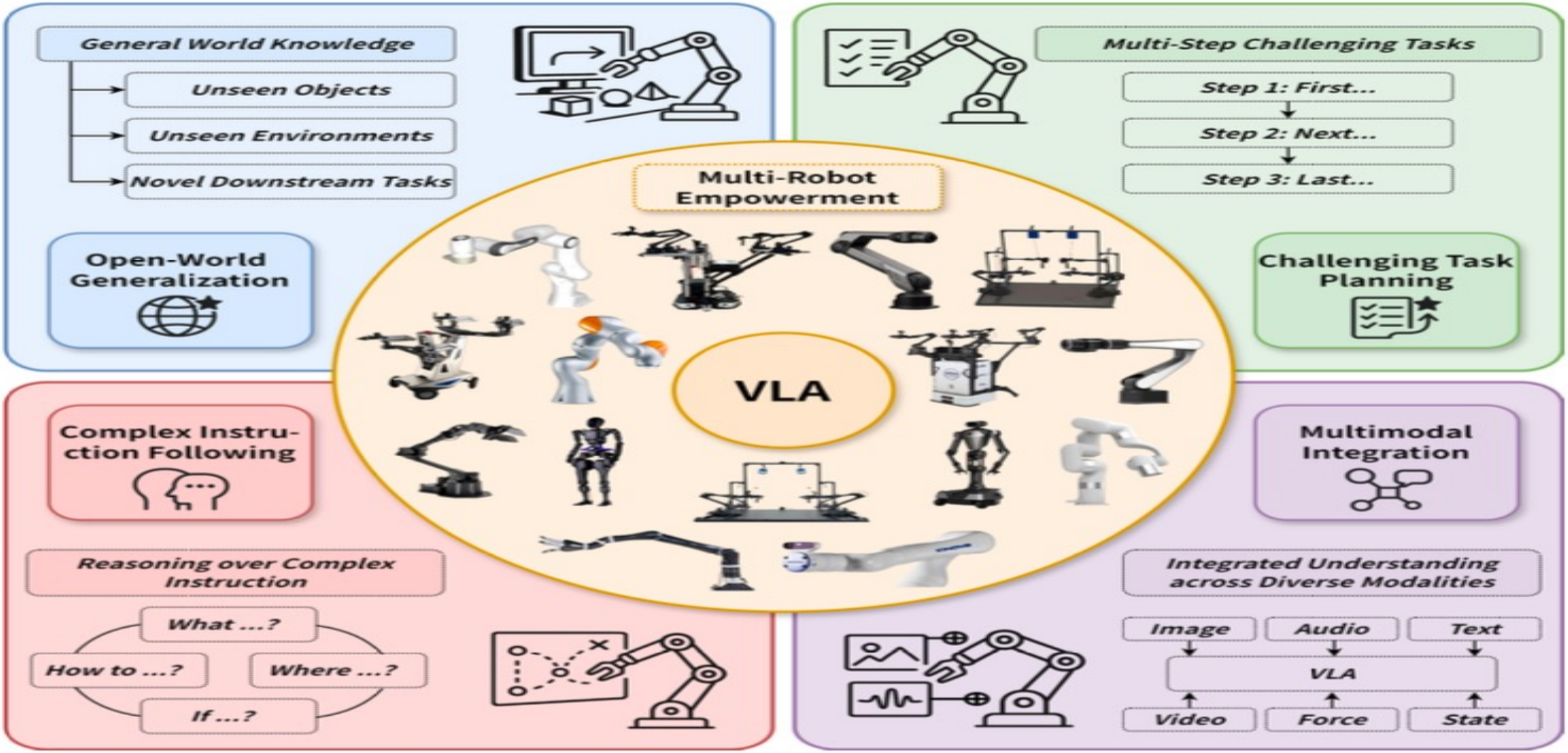
Source: Xie, J., Chen, Z., Zhang, R., Wan, X., & Li, G. (2024). Large Multimodal Agents: A Survey. ArXiv, abs/2402.15116.

Vision-Language-Action (VLA) Models for Robotics



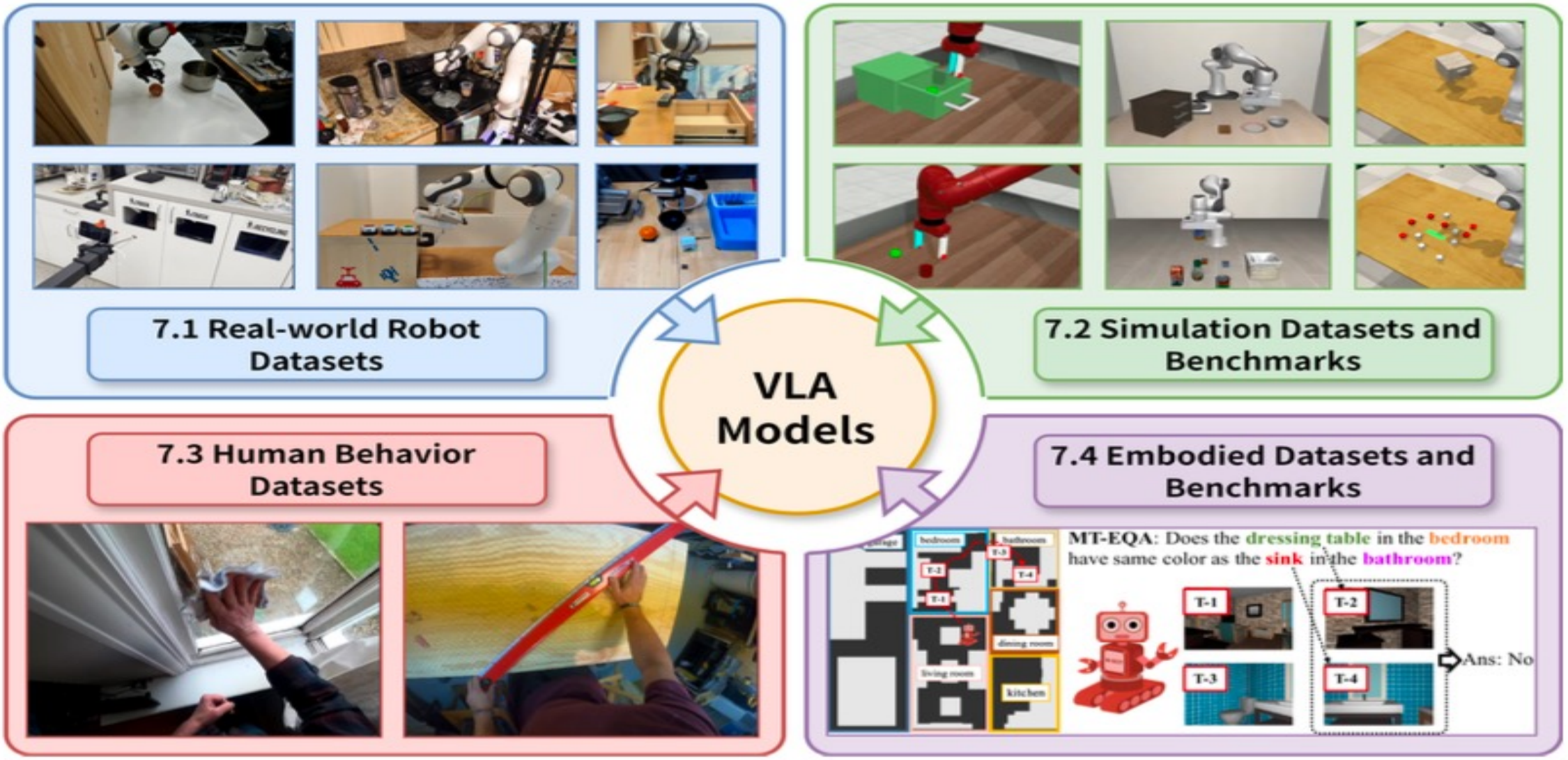
Source: Kento Kawaharazuka, Jihoon Oh, Jun Yamada, Ingmar Posner, and Yuke Zhu. (2025) "Vision-language-action models for robotics: A review towards real-world applications." IEEE Access (2025).

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).

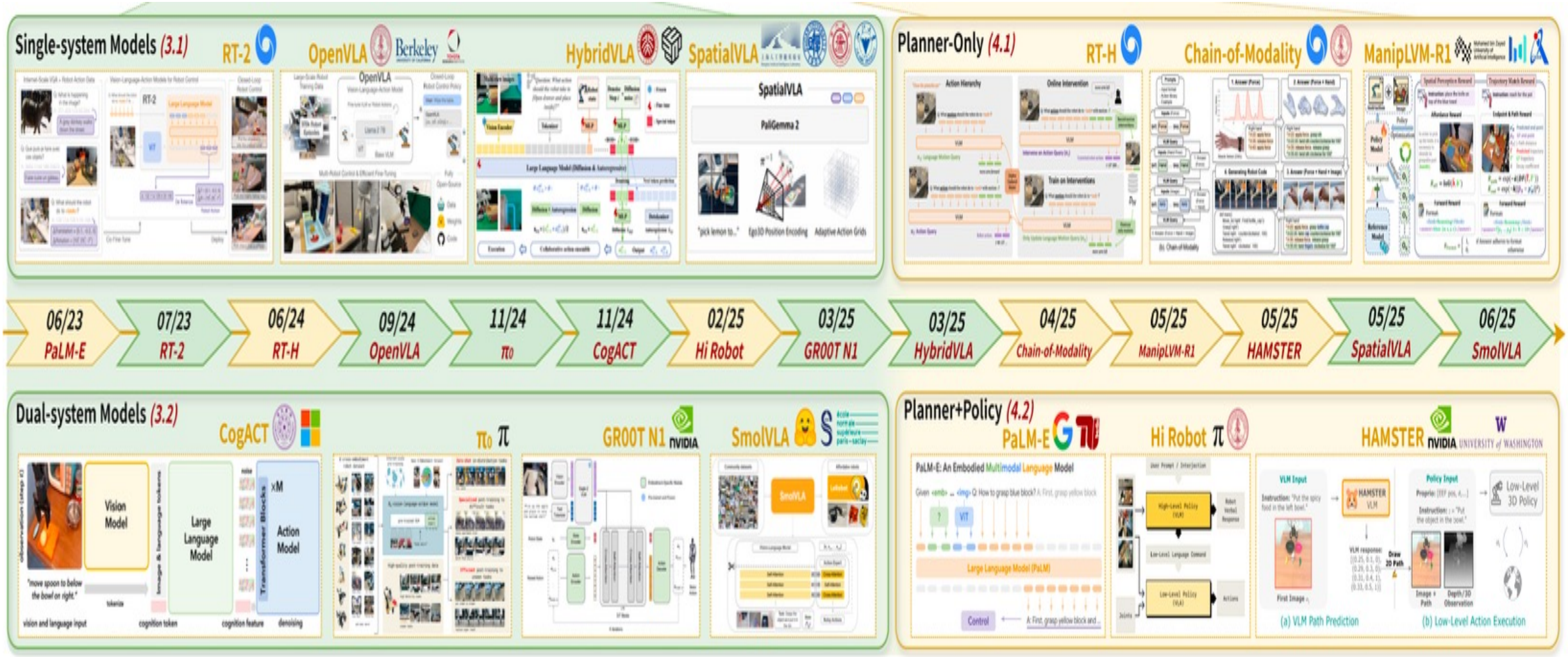
Large VLM-based Vision-Language-Action Models for Robotic Manipulation



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).

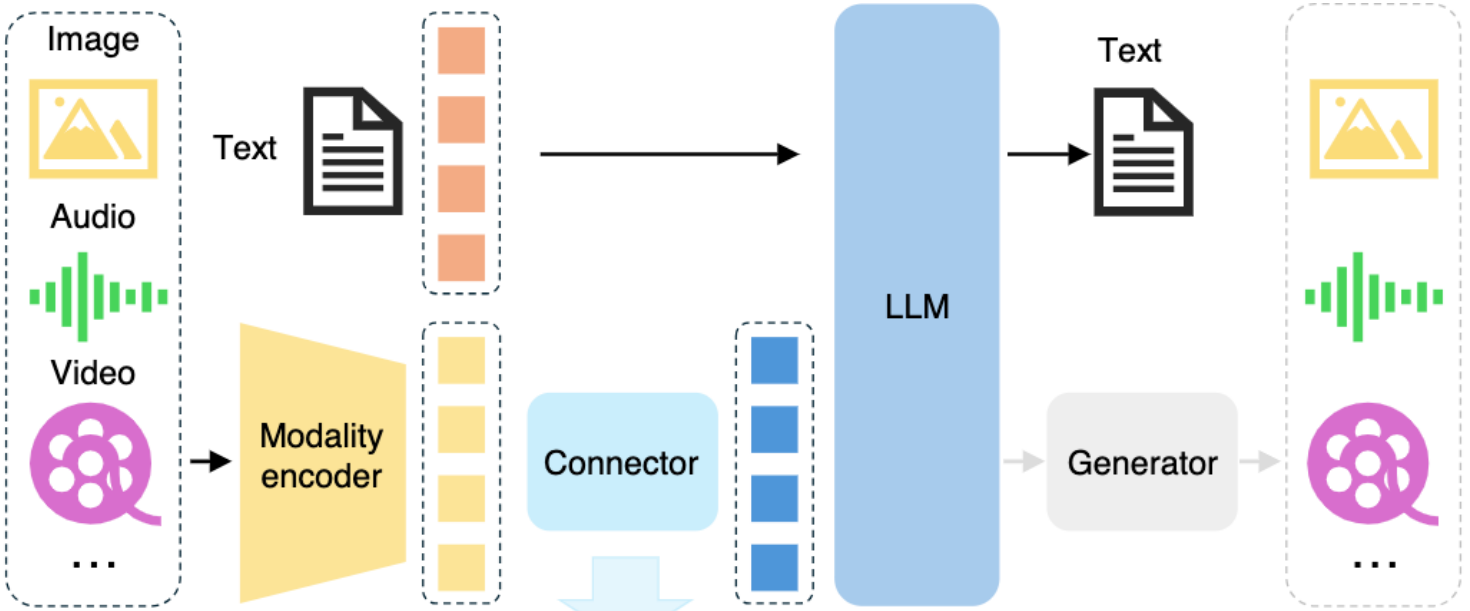
Large VLM-based Vision-Language-Action Models for Robotic Manipulation (Timeline)

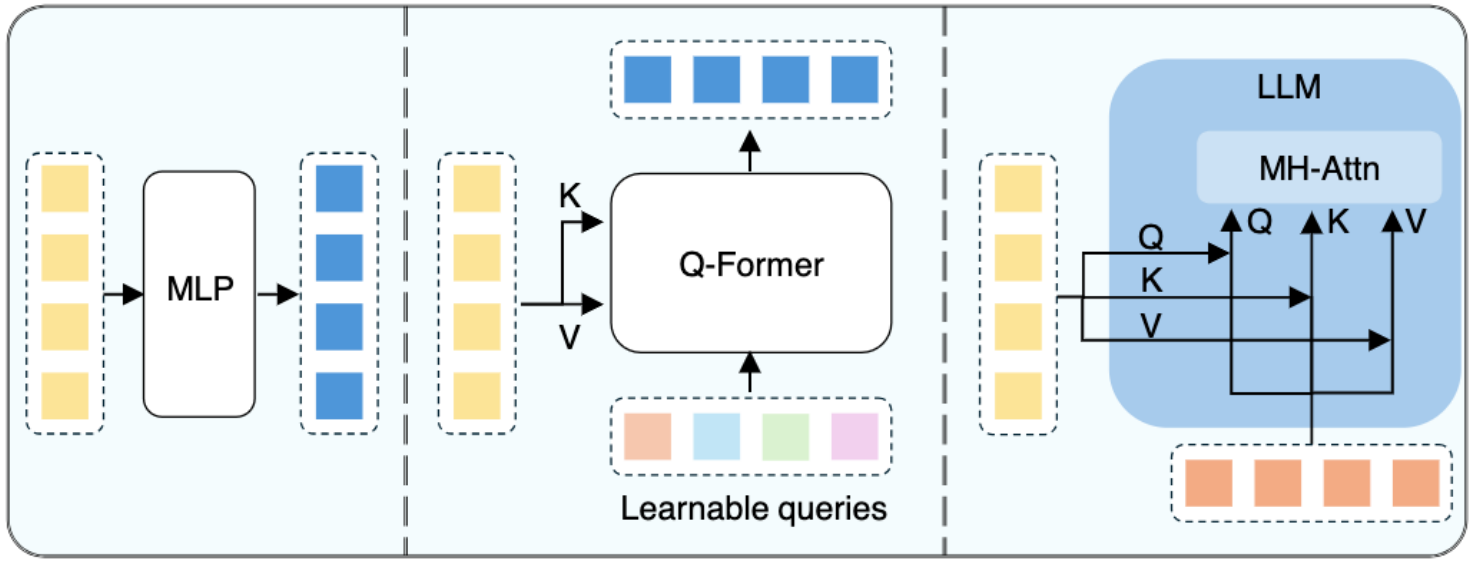
Monolithic models and Hierarchical Models



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).

Multimodal Large Language Models (MLLM)

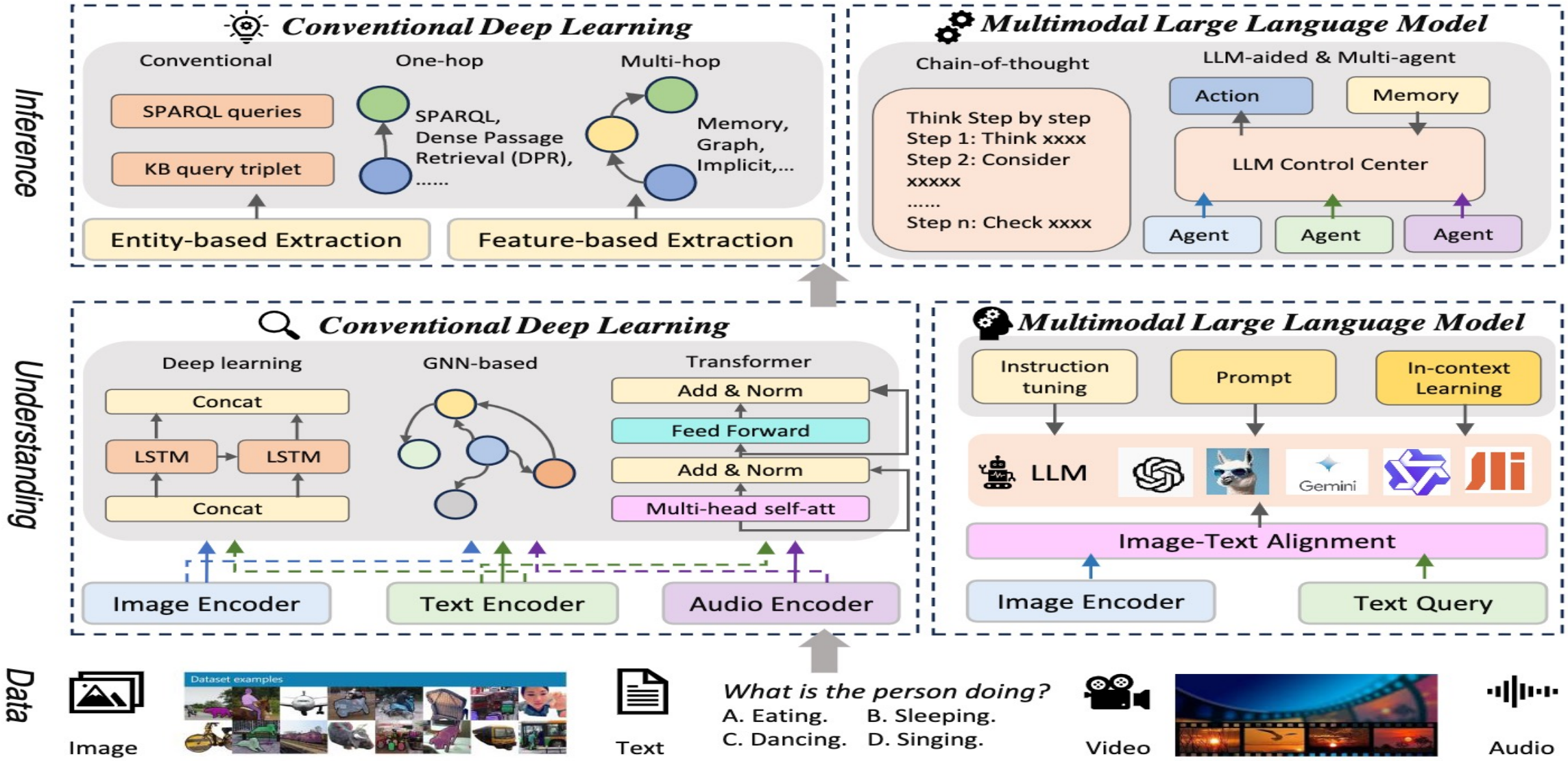




Multimodall LLM
Three types of connectors:
1. projection-based
2. query-based
3. fusion-based connectors

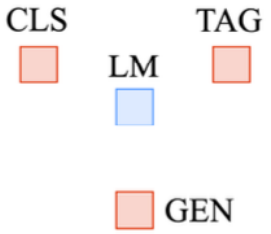
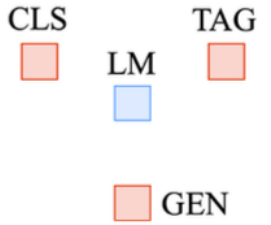
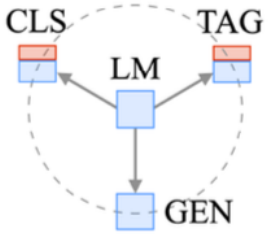
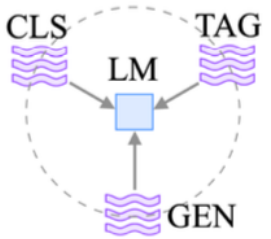
Source: Shukang Yin, Chaoyou Fu, Sirui Zhao, Ke Li, Xing Sun, Tong Xu, and Enhong Chen. (2024) "A survey on multimodal large language models." National Science Review (2024): nwae403.

Multimodal Large Language Model (MLLM) for Vision Question Answering



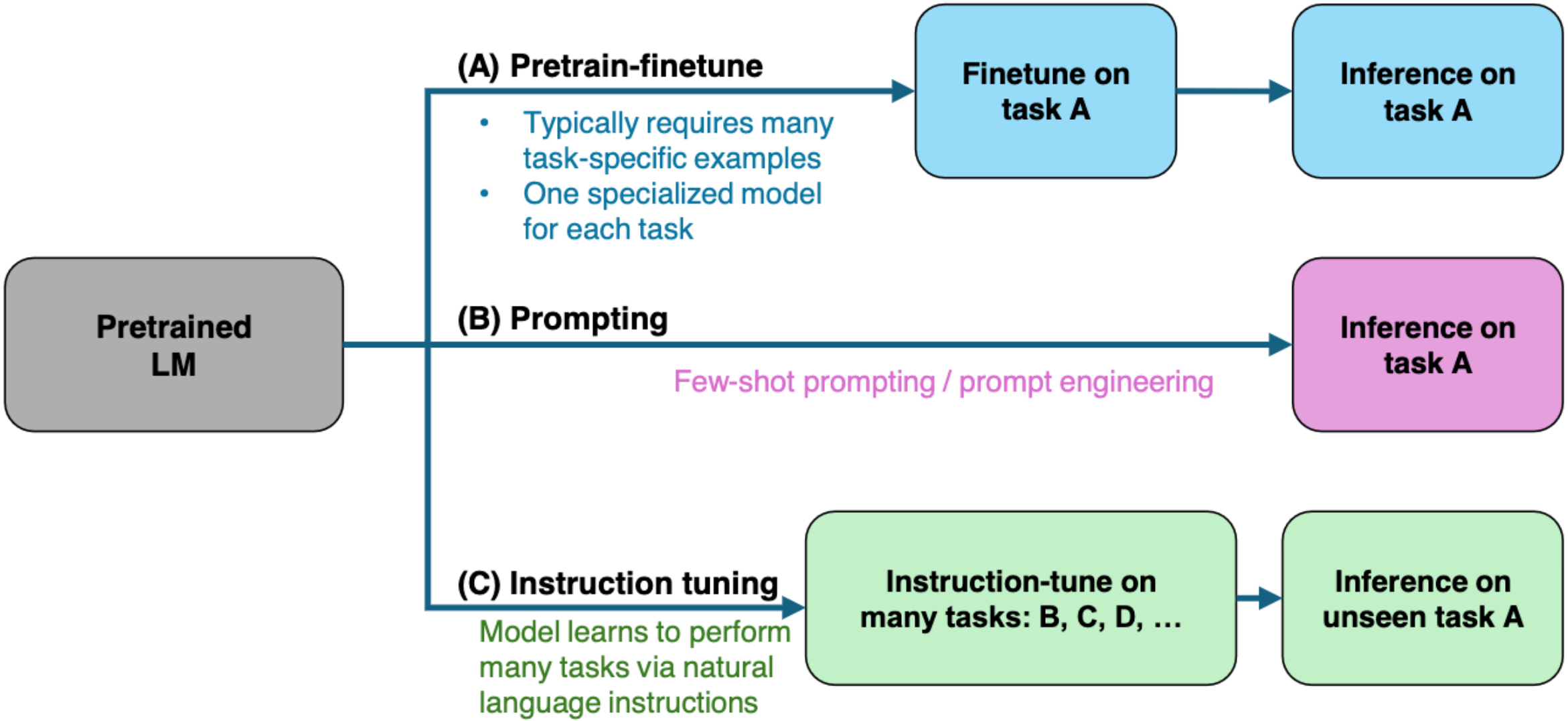
Source: Jiayi Kuang, Jingyou Xie, Haohao Luo, Ronghao Li, Zhe Xu, Xianfeng Cheng, Yinghui Li, Xika Lin, and Ying Shen. (2024) "Natural Language Understanding and Inference with MLLM in Visual Question Answering: A Survey." arXiv preprint arXiv:2411.17558.

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.

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

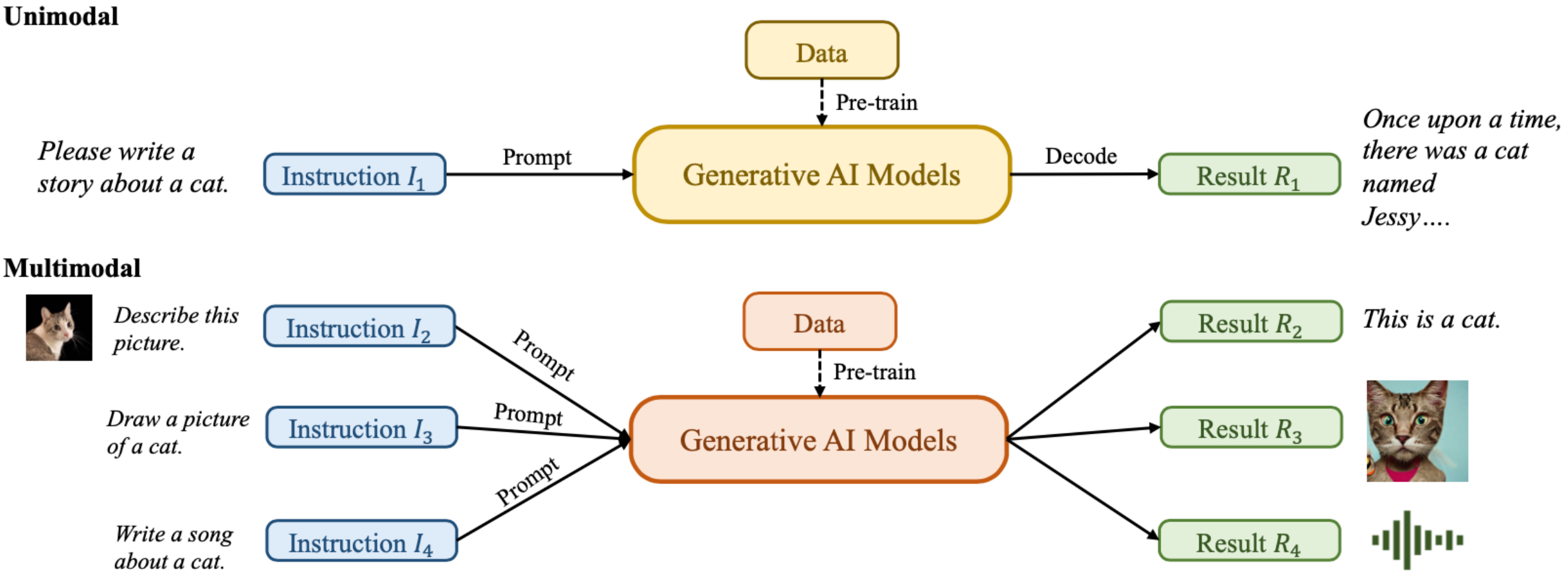
Figure 2

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.

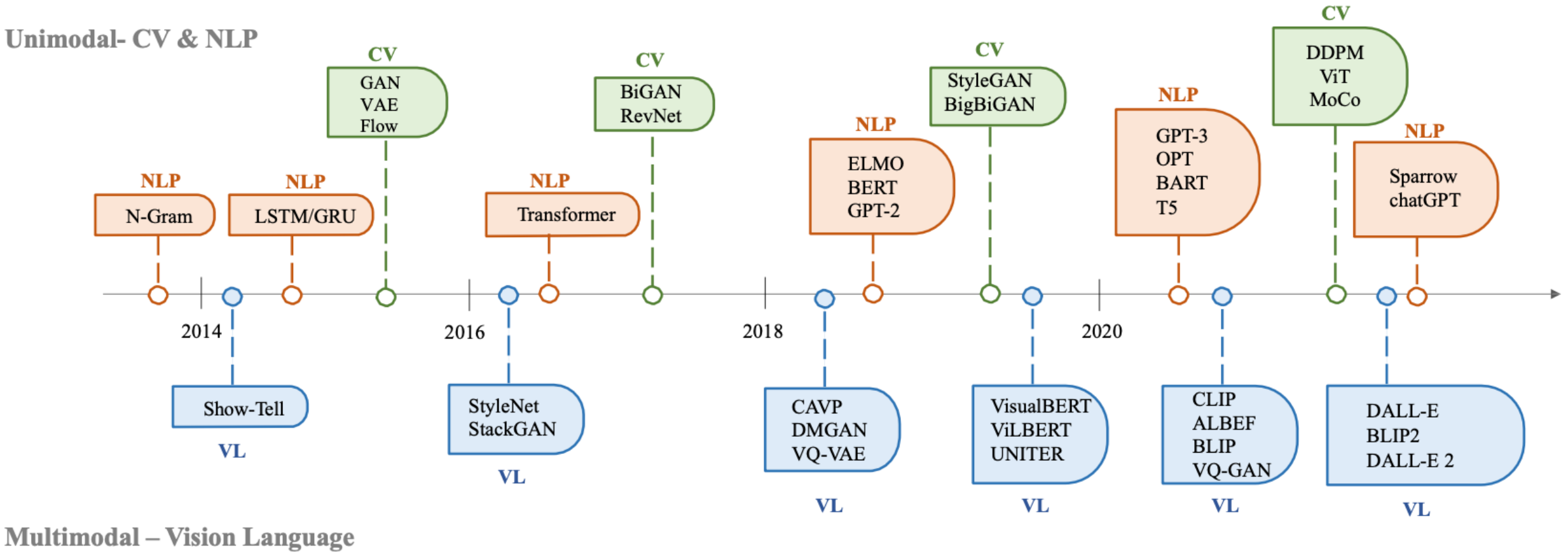
Generative AI (Gen AI)

AI Generated Content (AIGC)



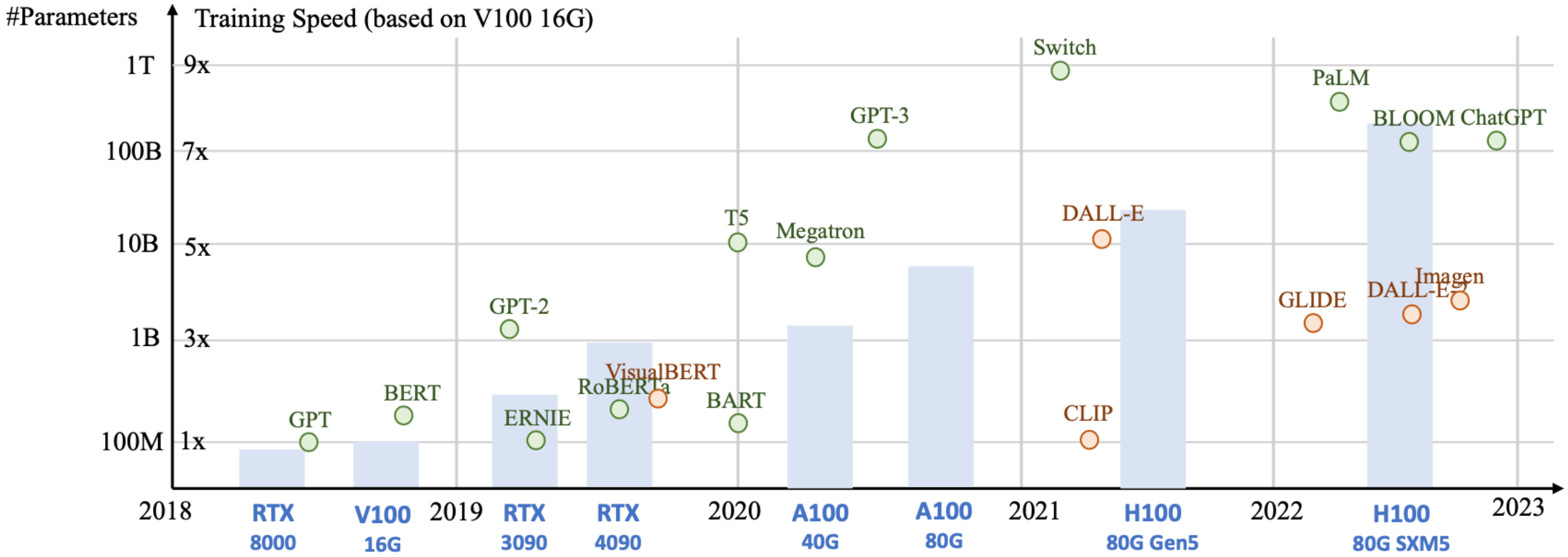
Source: Yihan Cao, Siyu Li, Yixin Liu, Zhiling Yan, Yutong Dai, Philip S. Yu, and Lichao Sun (2023). "A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT." arXiv preprint arXiv:2303.04226.

The history of Generative AI in CV, NLP and VL



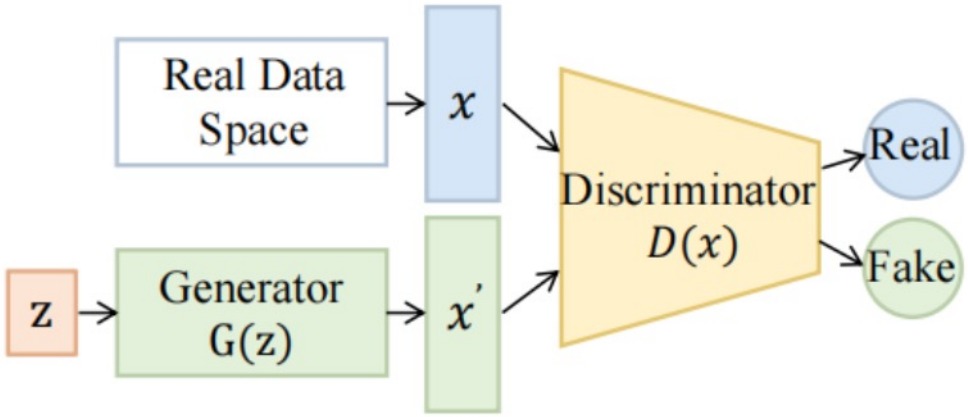
Source: Yihan Cao, Siyu Li, Yixin Liu, Zhiling Yan, Yutong Dai, Philip S. Yu, and Lichao Sun (2023). "A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT." arXiv preprint arXiv:2303.04226.

Generative AI Foundation Models

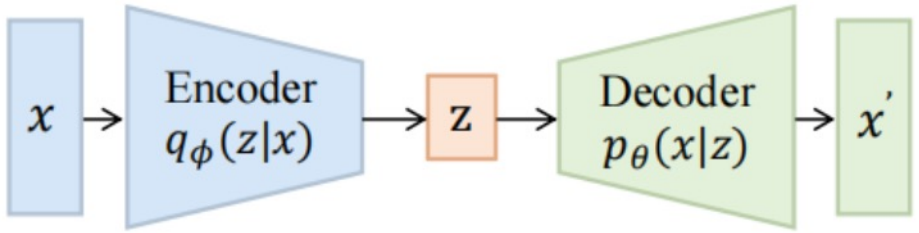


Source: Yihan Cao, Siyu Li, Yixin Liu, Zhiling Yan, Yutong Dai, Philip S. Yu, and Lichao Sun (2023). "A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT." arXiv preprint arXiv:2303.04226.

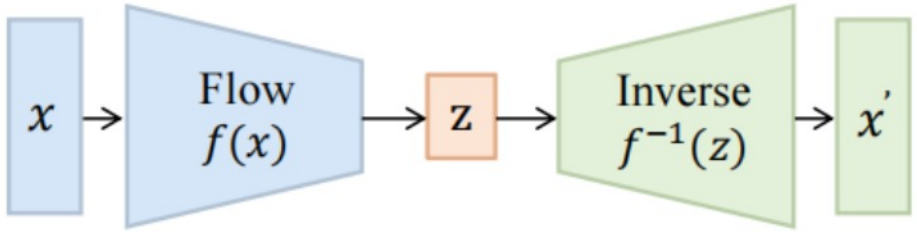
Categories of Vision Generative Models



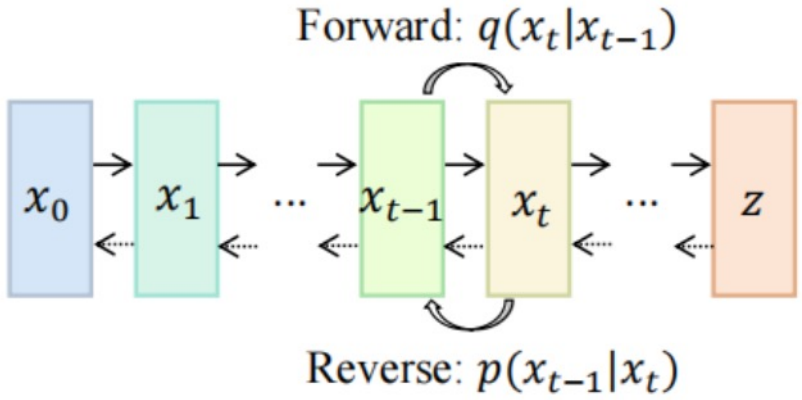
(1) Generative adversarial networks



(2) Variational autoencoders



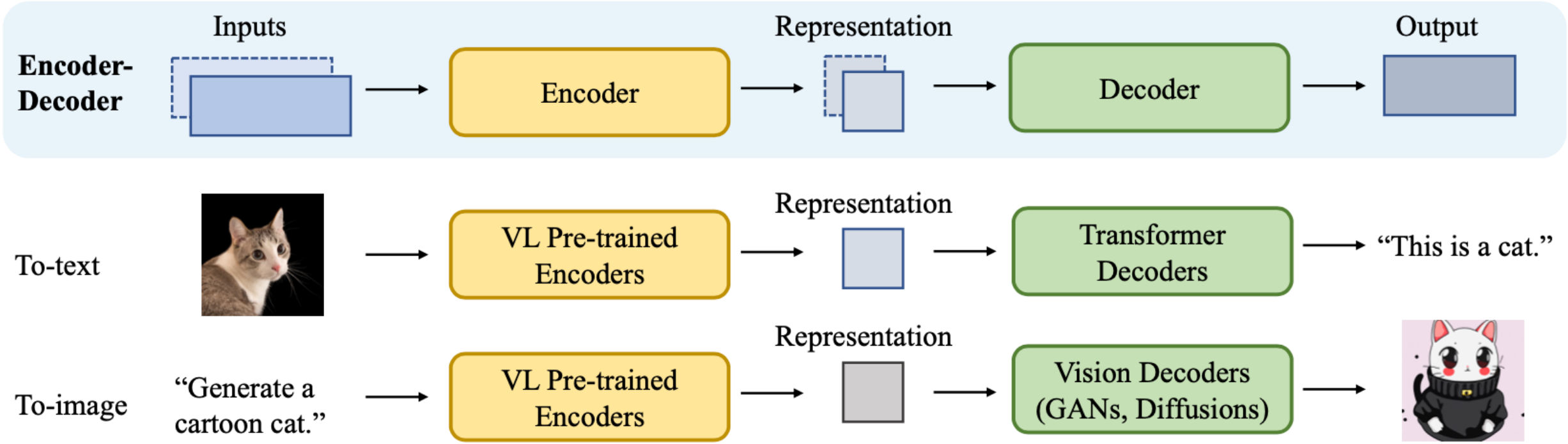
(3) Normalizing flows



(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

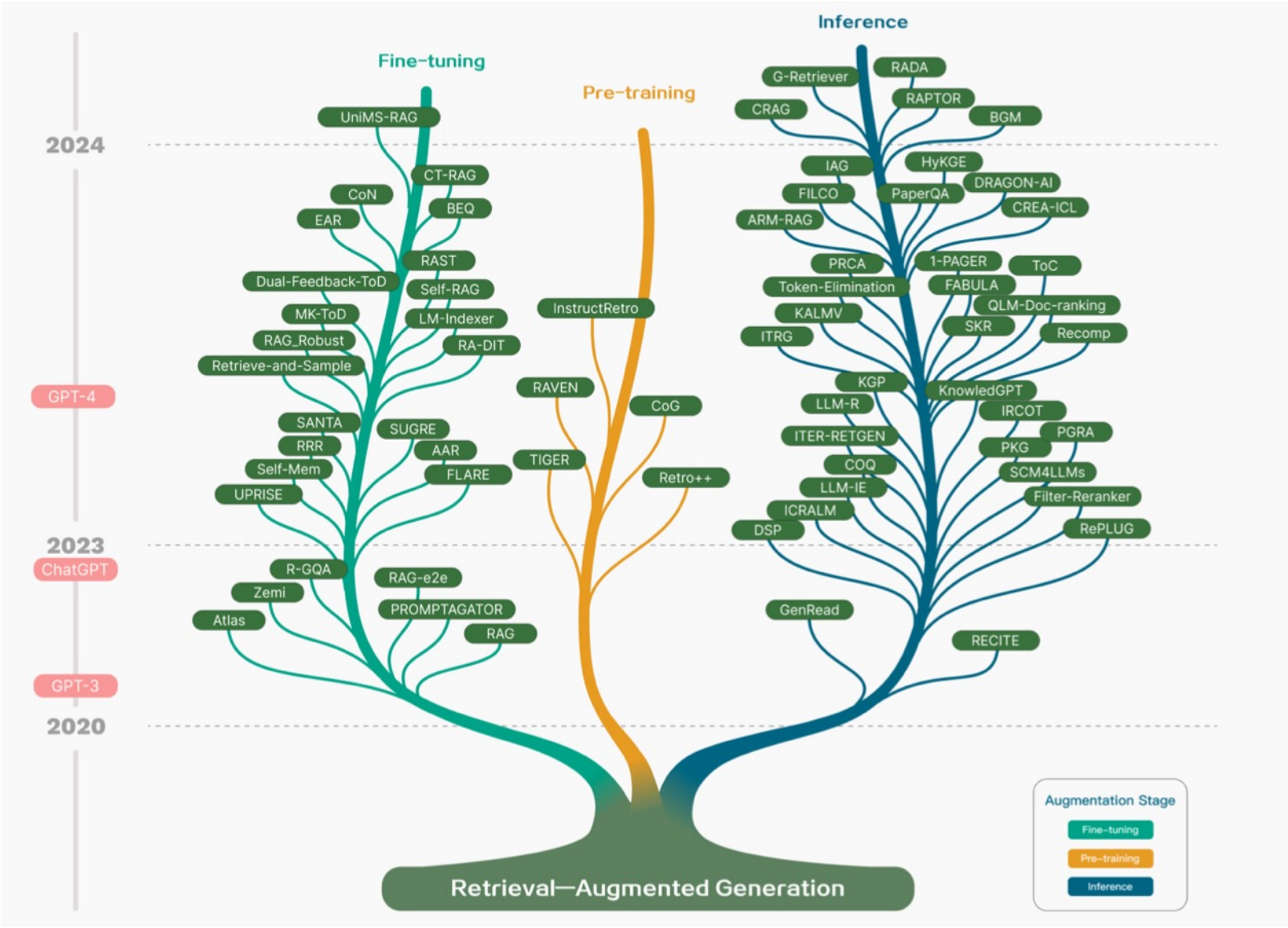


Source: Yihan Cao, Siyu Li, Yixin Liu, Zhiling Yan, Yutong Dai, Philip S. Yu, and Lichao Sun (2023). "A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT." arXiv preprint arXiv:2303.04226.

RAG LLM
Dialogue Systems

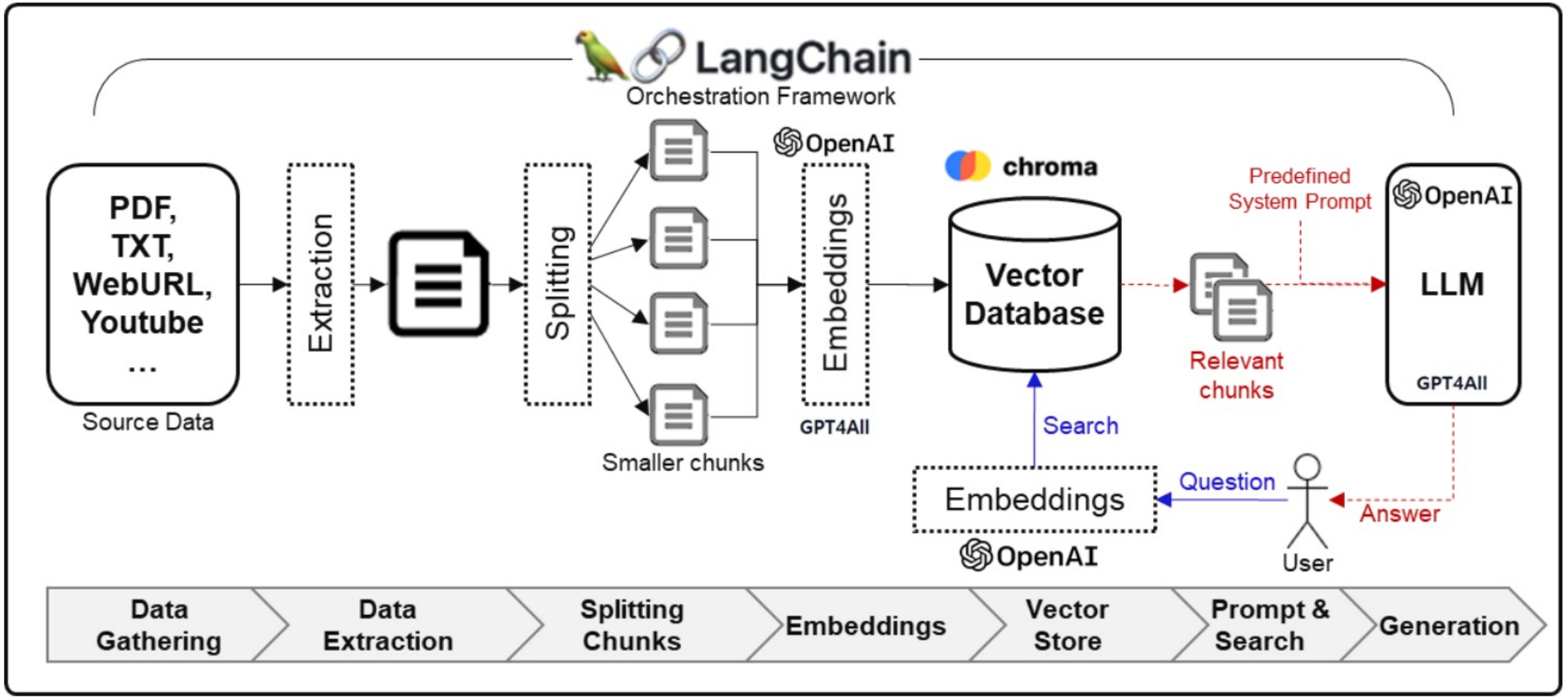
Technology Tree of RAG Research

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



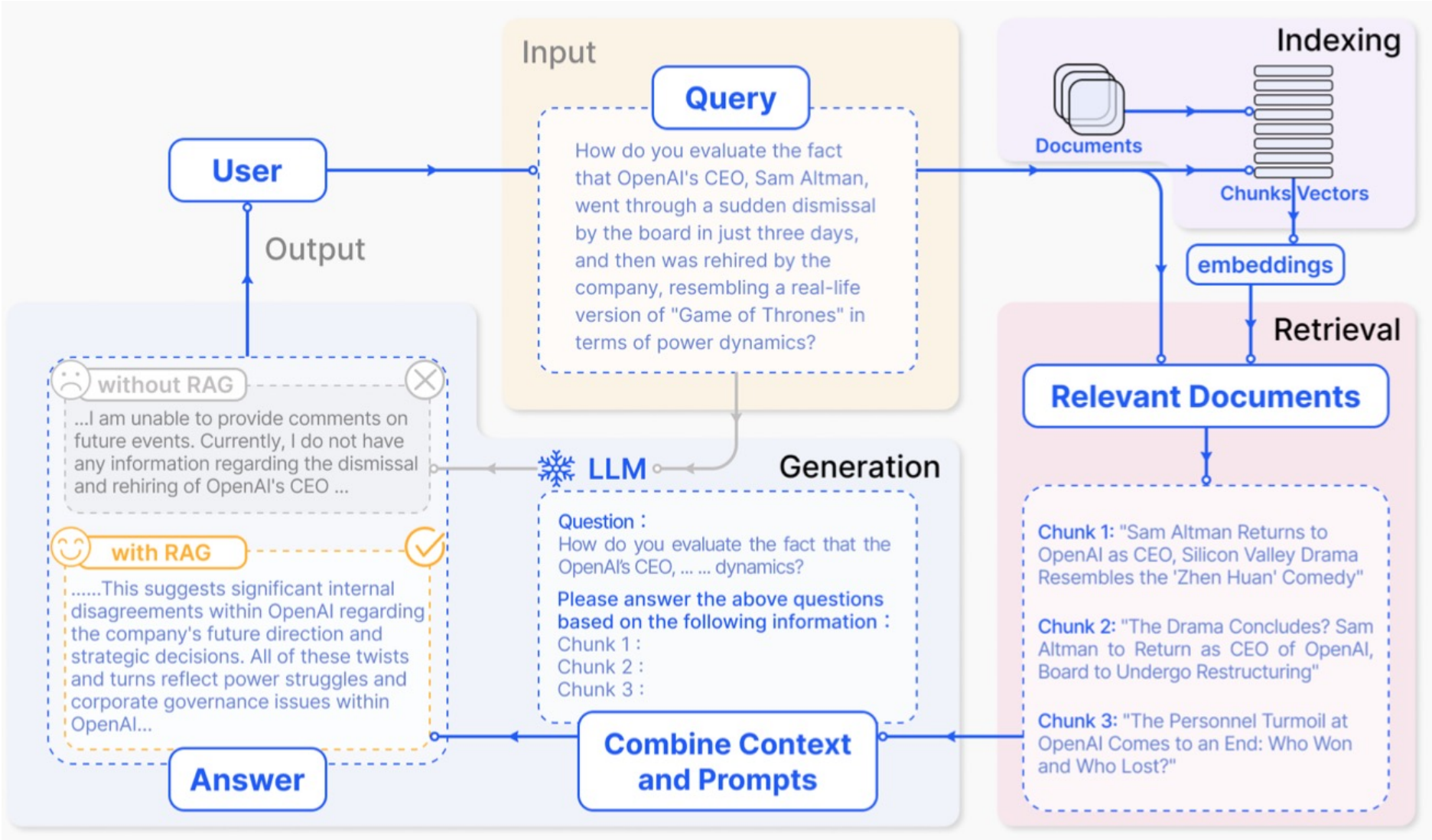
Source: Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., ... & Wang, H. (2023). Retrieval-augmented generation for large language models: A survey. arXiv preprint arXiv:2312.10997.

Framework for Implementing Generative AI Services using RAG Model



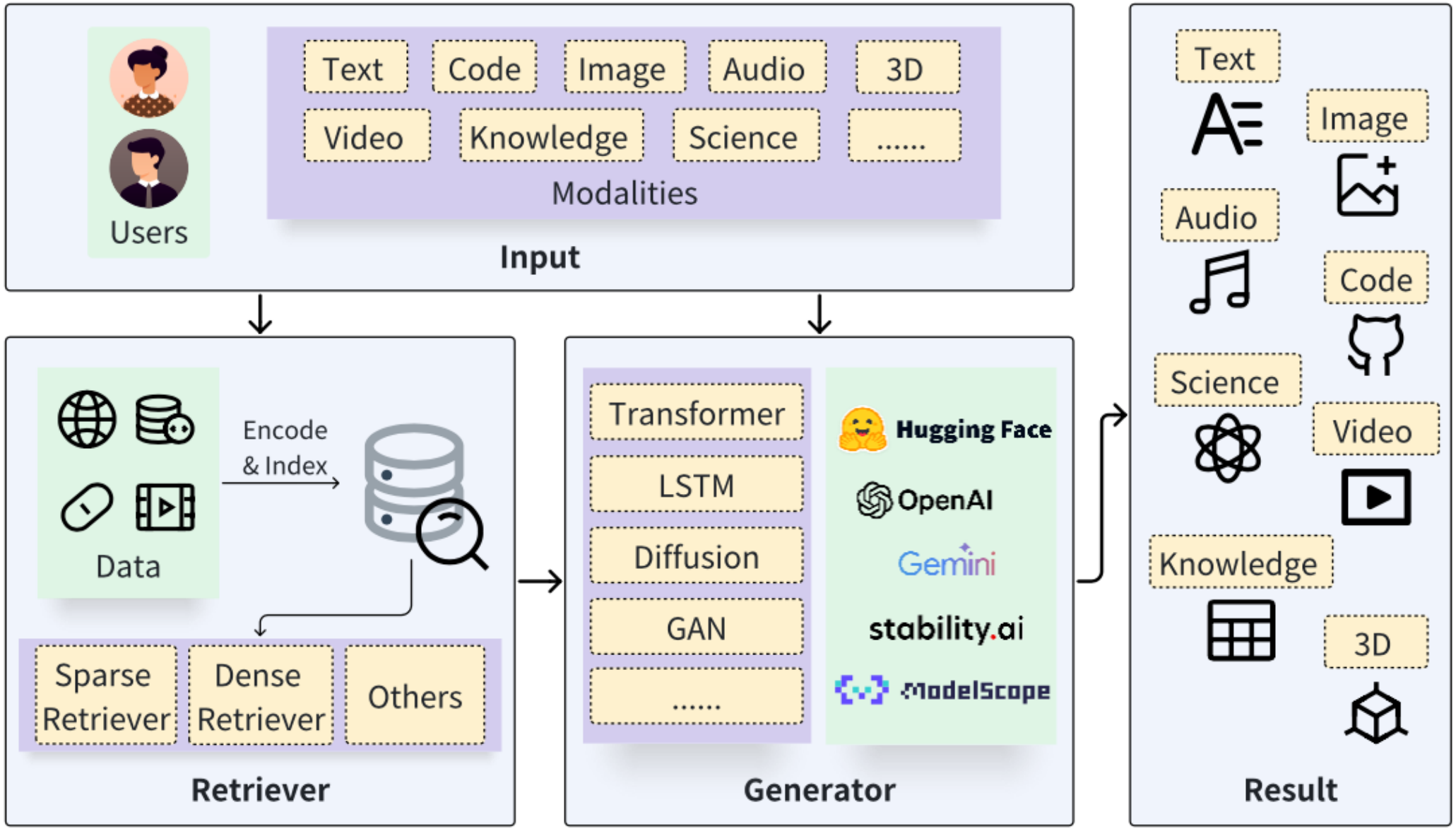
Source: Jeong, Cheonsu. "A Study on the Implementation of Generative AI Services Using an Enterprise Data-Based LLM Application Architecture." arXiv preprint arXiv:2309.01105 (2023).

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



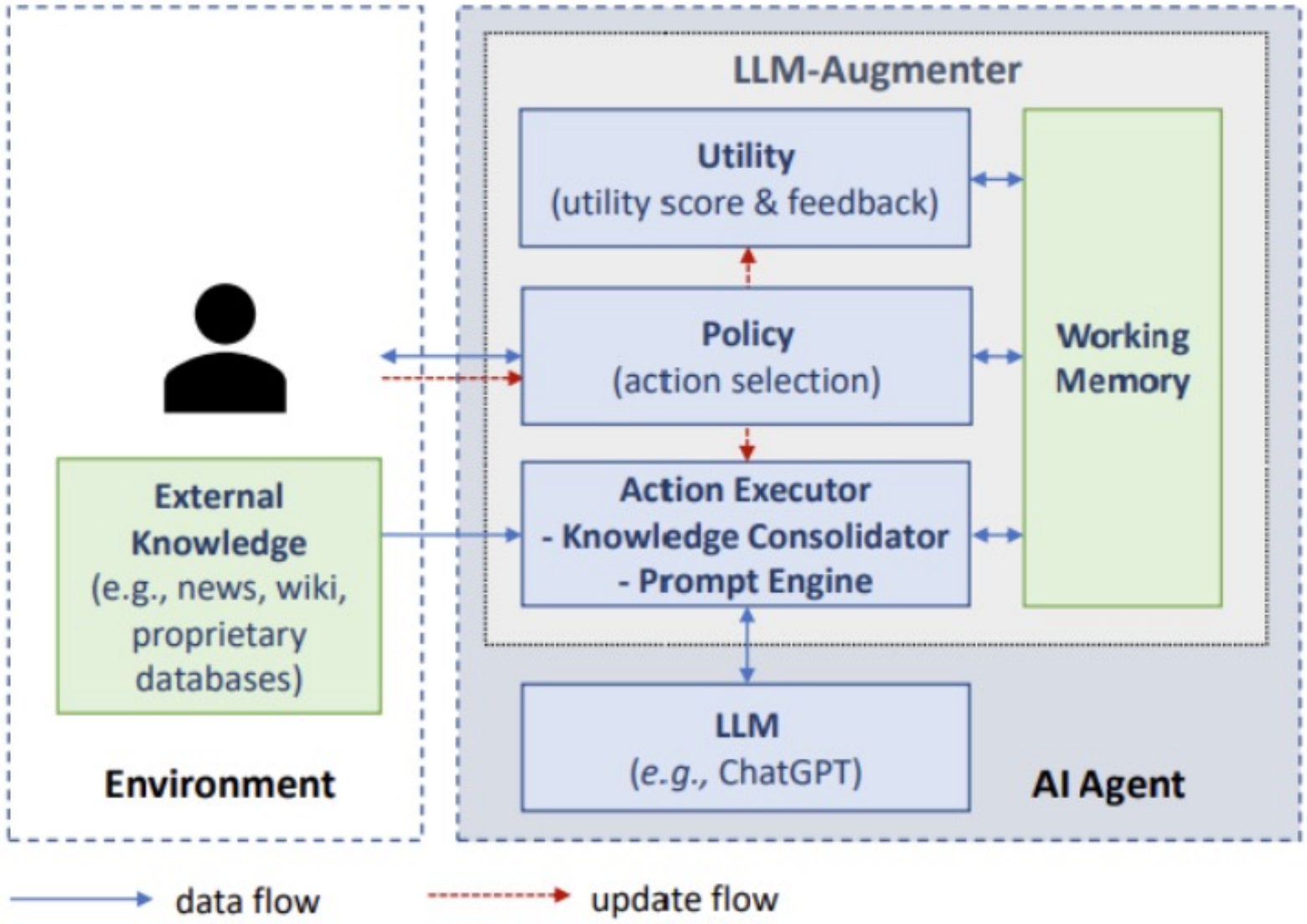
Source: Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., ... & Wang, H. (2023). Retrieval-augmented generation for large language models: A survey. arXiv preprint arXiv:2312.10997.

Retrieval-Augmented Generation (RAG) Architecture



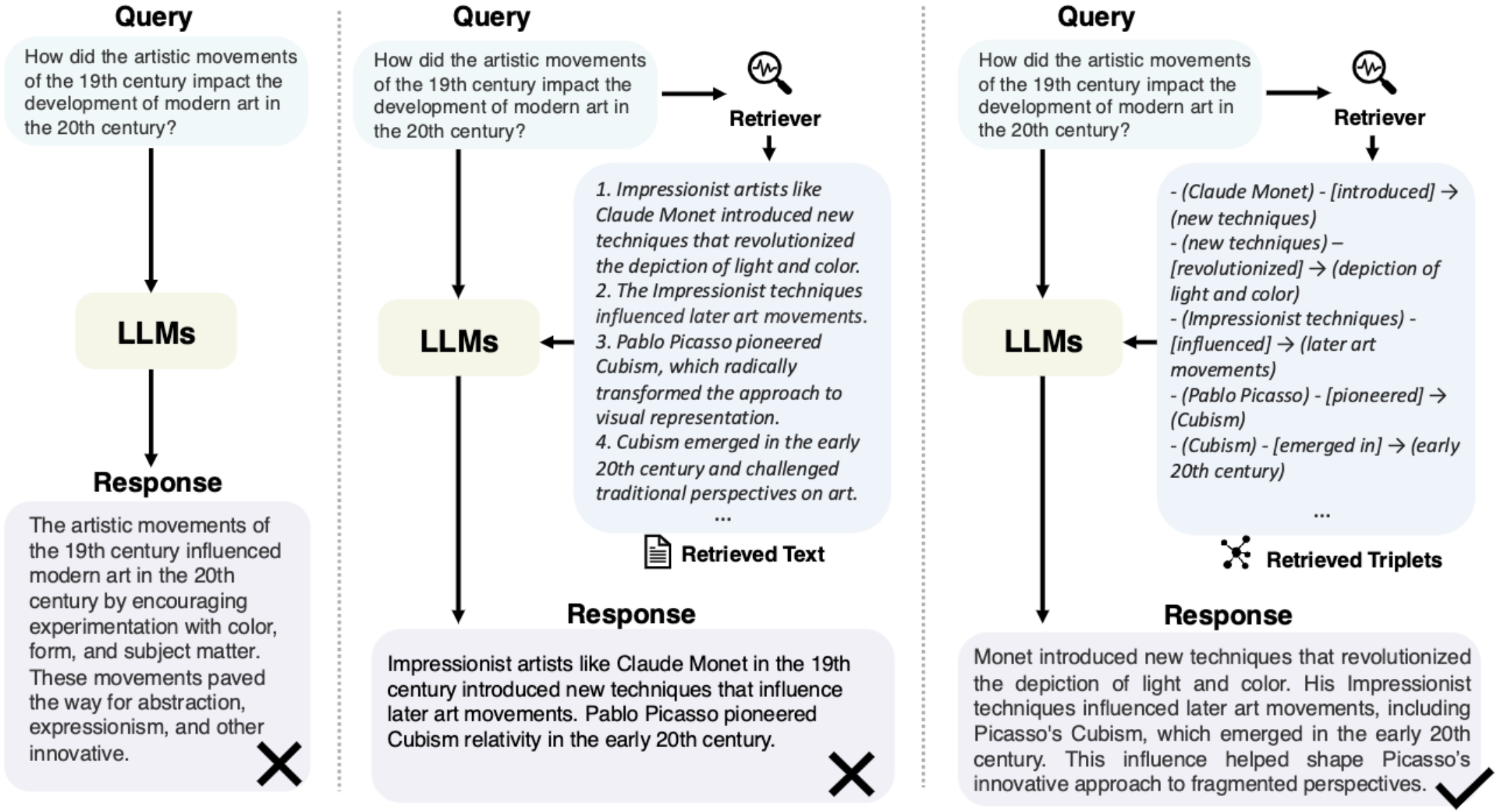
Source: Zhao, P., Zhang, H., Yu, Q., Wang, Z., Geng, Y., Fu, F., ... & Cui, B. (2024). Retrieval-augmented generation for ai-generated content: A survey. arXiv preprint arXiv:2402.19473.

A LLM-based Agent for Conversational Information Seeking



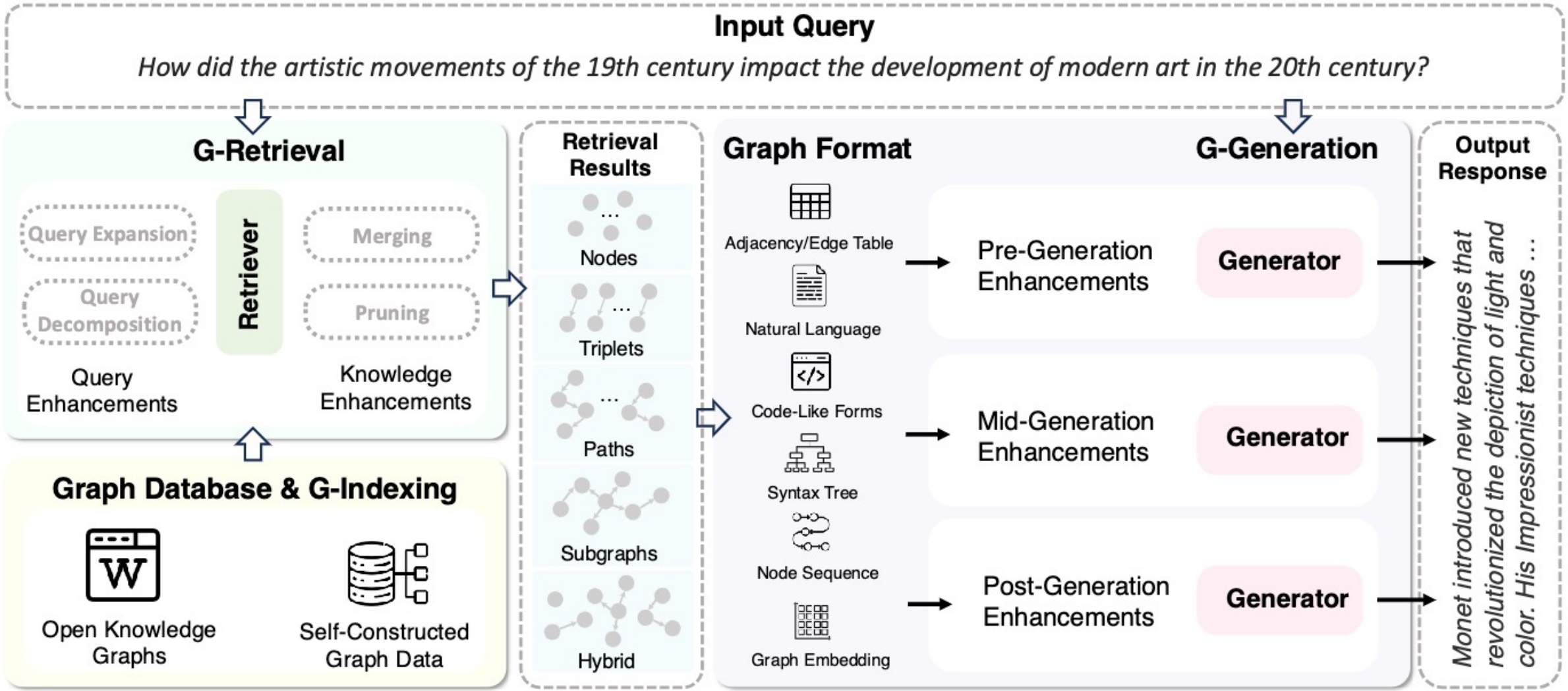
Source: Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X., & Gao, J. (2024). Large language models: A survey. arXiv preprint arXiv:2402.06196.

Direct LLM, RAG, and GraphRAG



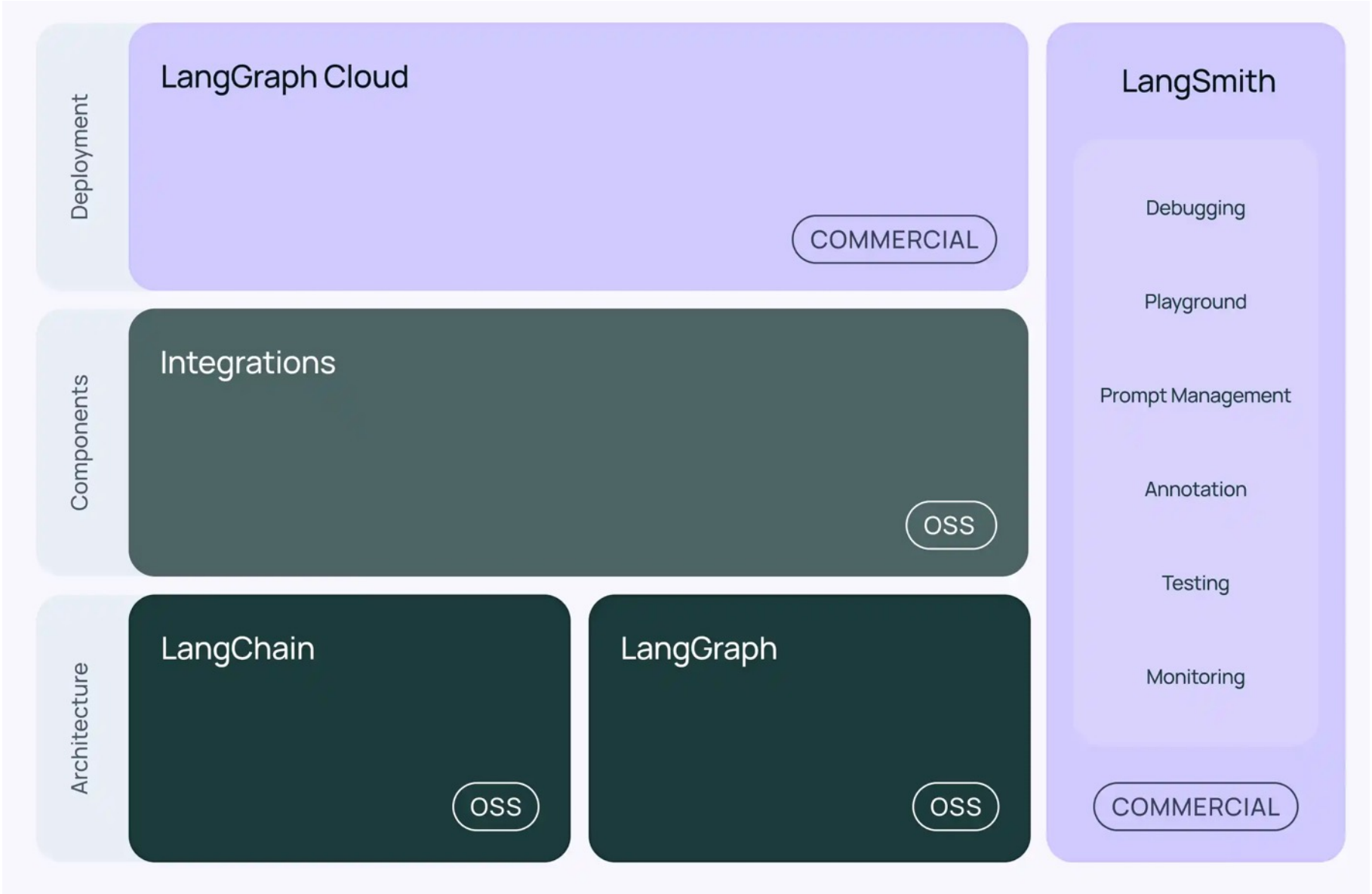
Source: Peng, B., Zhu, Y., Liu, Y., Bo, X., Shi, H., Hong, C., ... & Tang, S. (2024). Graph retrieval-augmented generation: A survey. arXiv preprint arXiv:2408.08921.

GraphRAG Framework for Question Answering



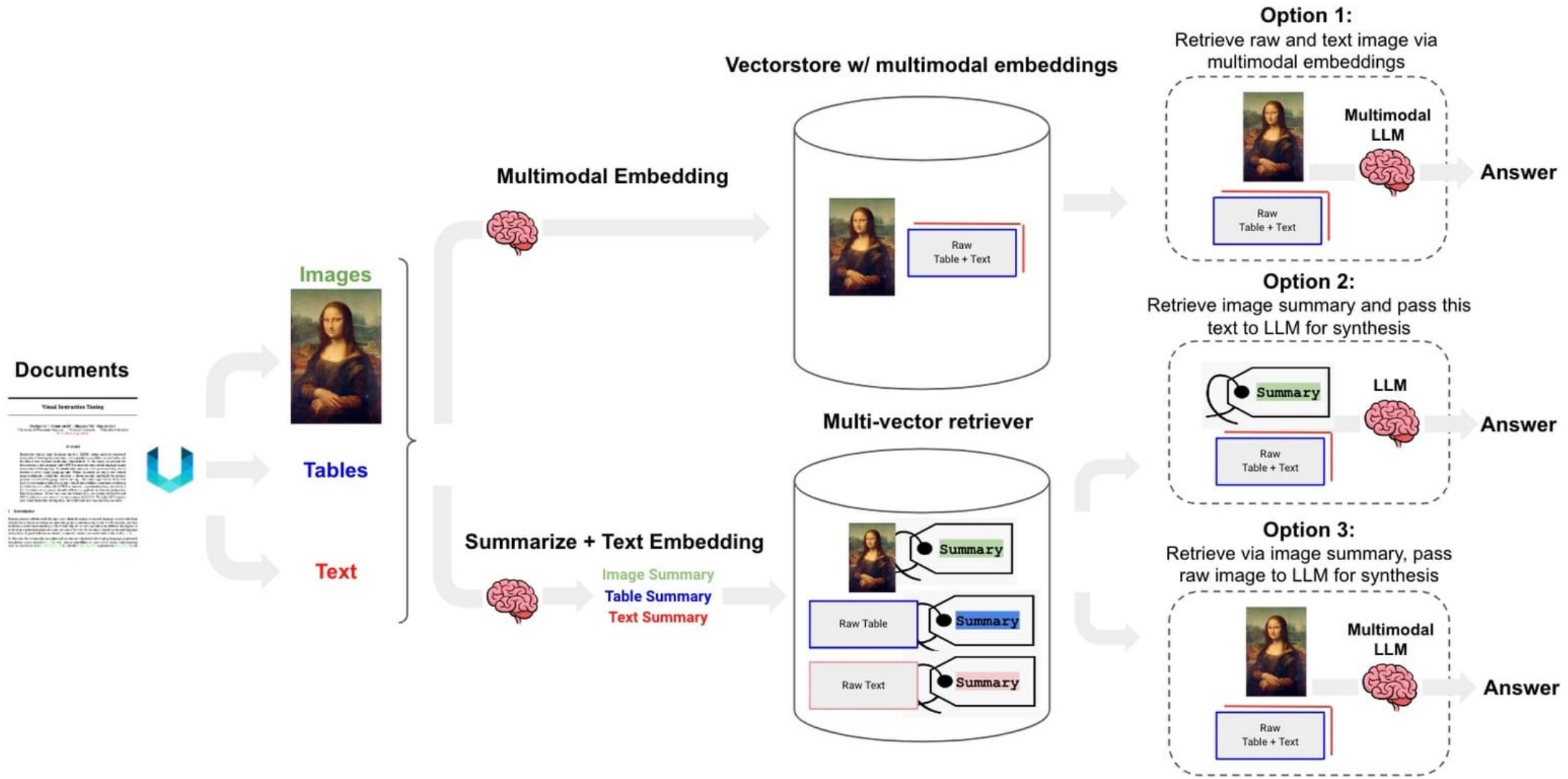
Source: Peng, B., Zhu, Y., Liu, Y., Bo, X., Shi, H., Hong, C., ... & Tang, S. (2024). Graph retrieval-augmented generation: A survey. arXiv preprint arXiv:2408.08921.

LangChain Architecture



Source: https://www.langchain.com/

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

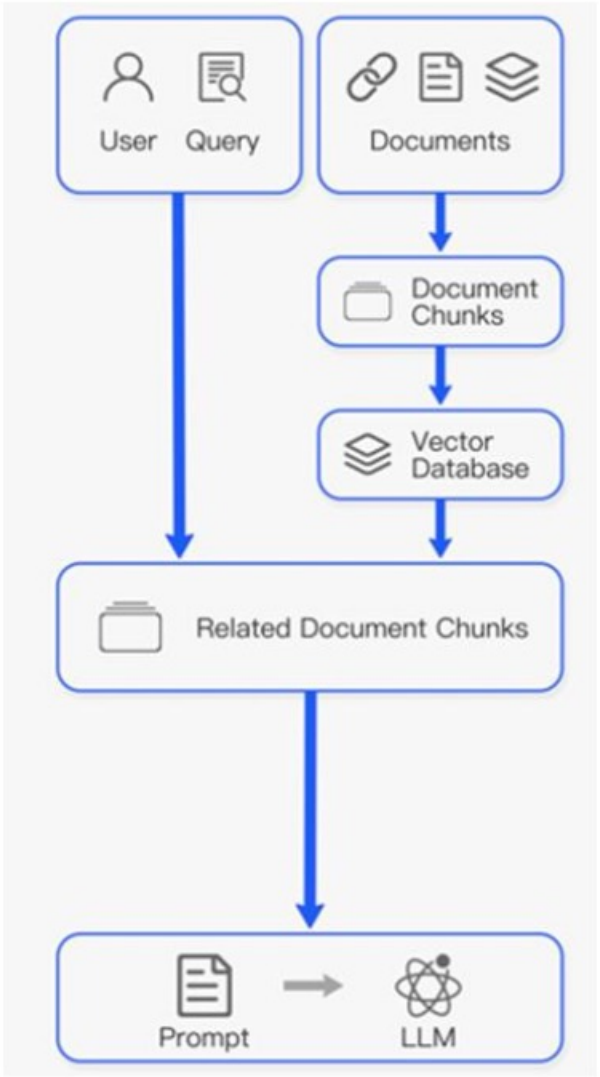
the signal to noise ratio of retrieved
context

context recall

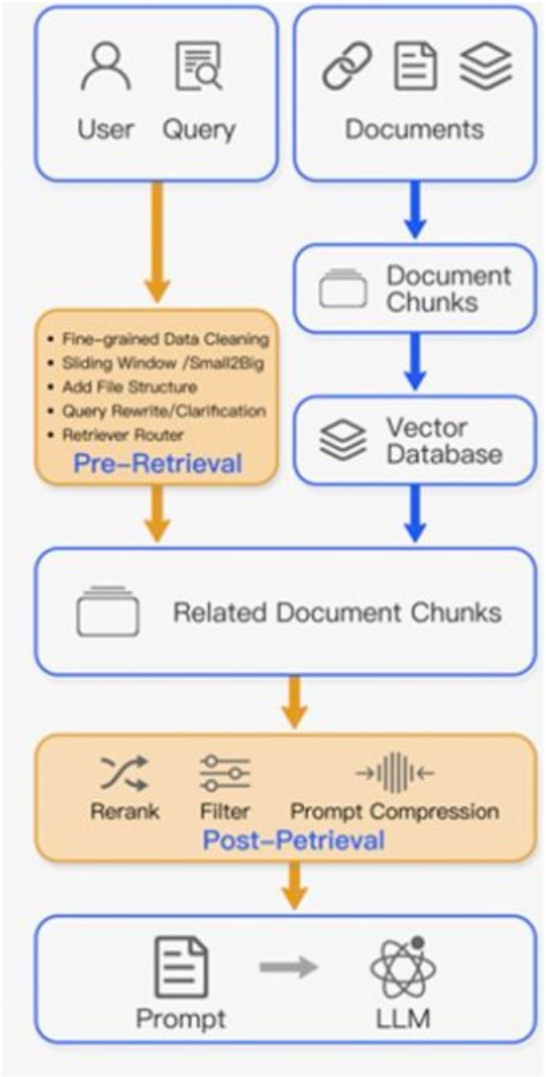
can it retrieve all the relevant information
required to answer the question

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

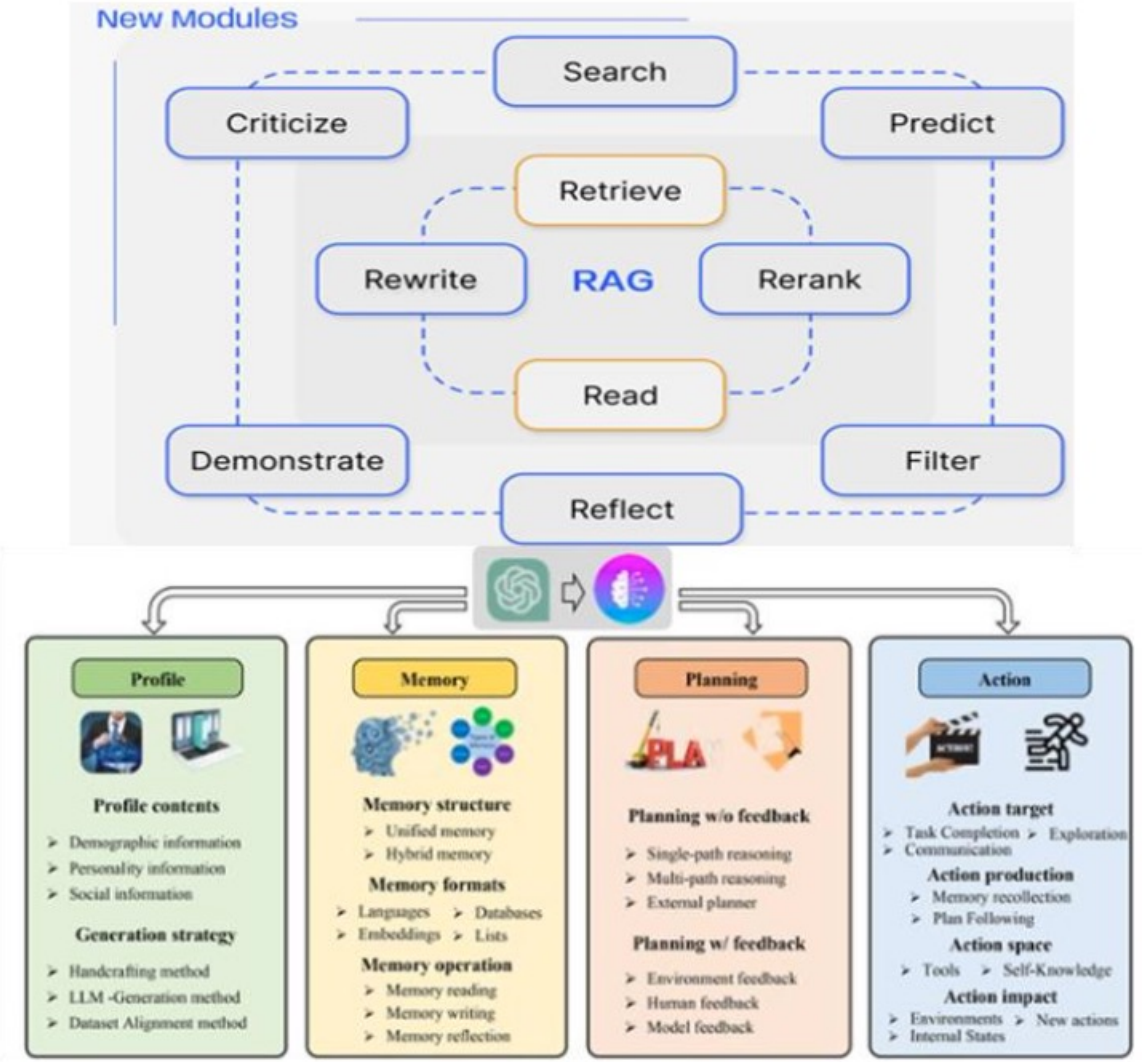
Native RAG, Advanced RAG, Agentic RAG



Native RAG



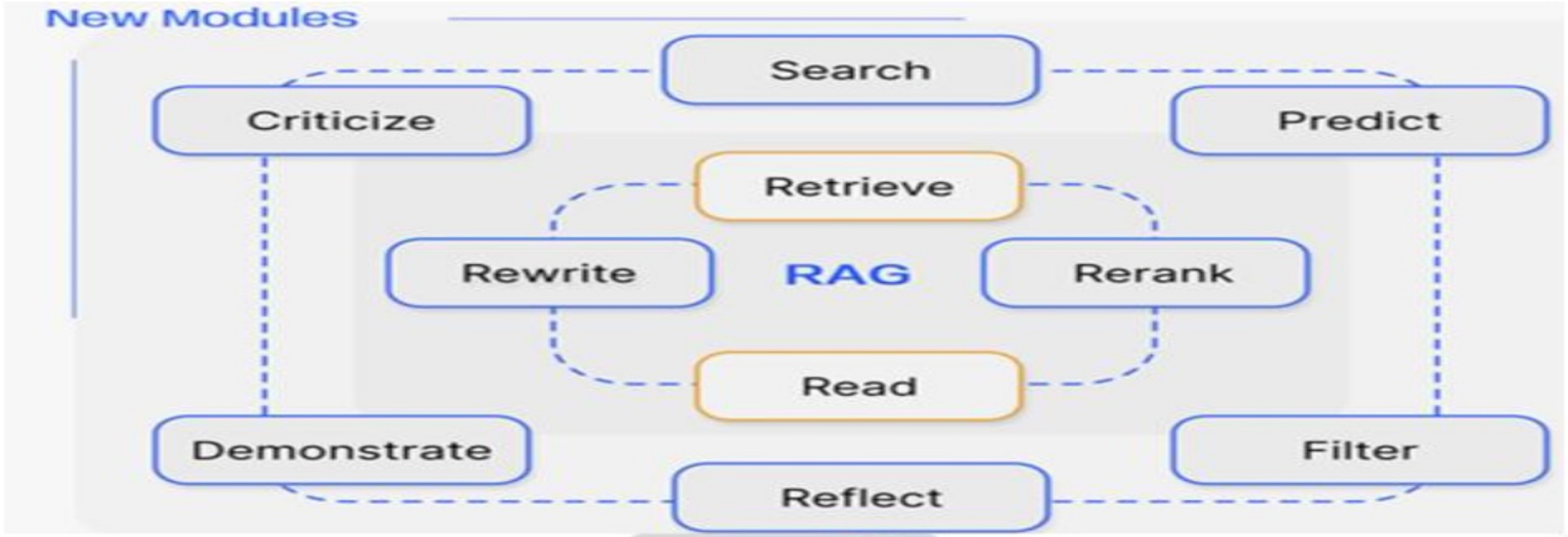
Advanced RAG

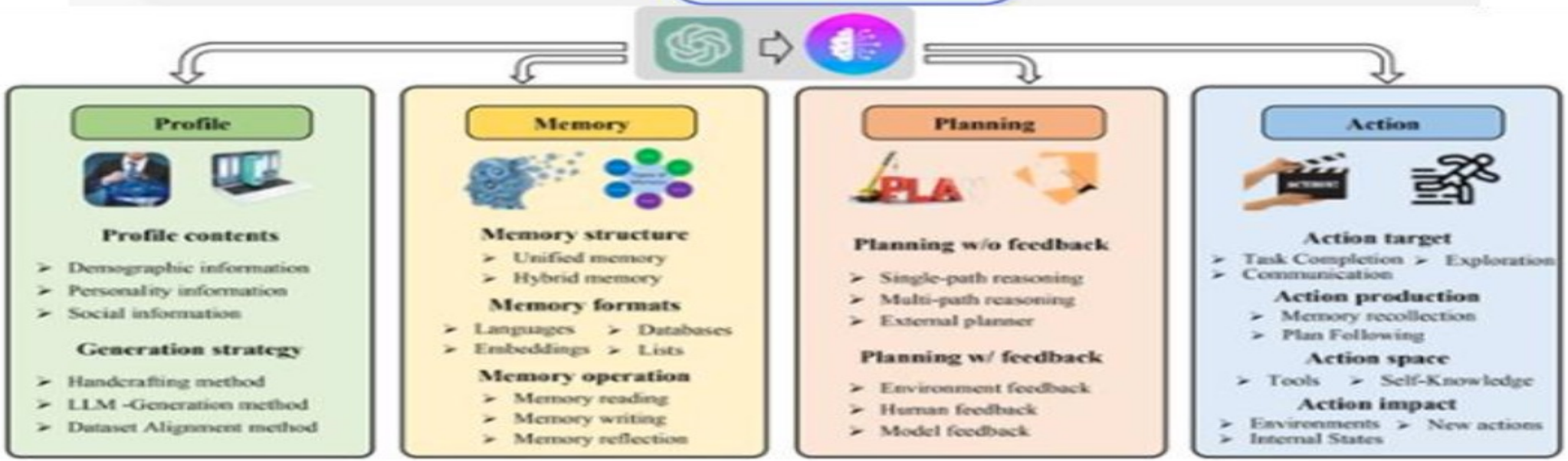


Modular + Agentic RAG

Source: Jun Xu (2024). "GenAI and LLM for Financial Institutions: A Corporate Strategic Survey." Available at SSRN 4988118.

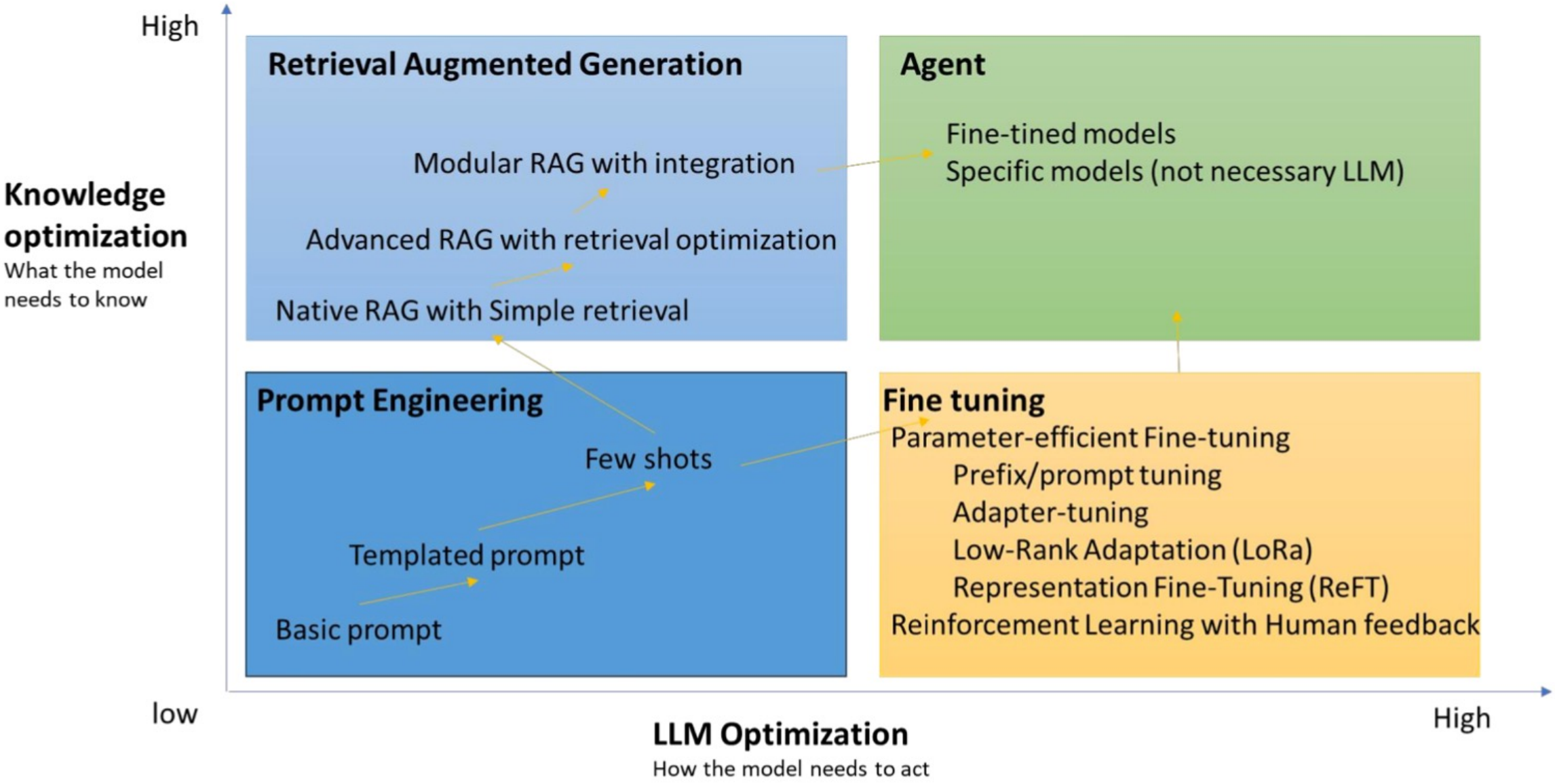
Modular + Agentic RAG





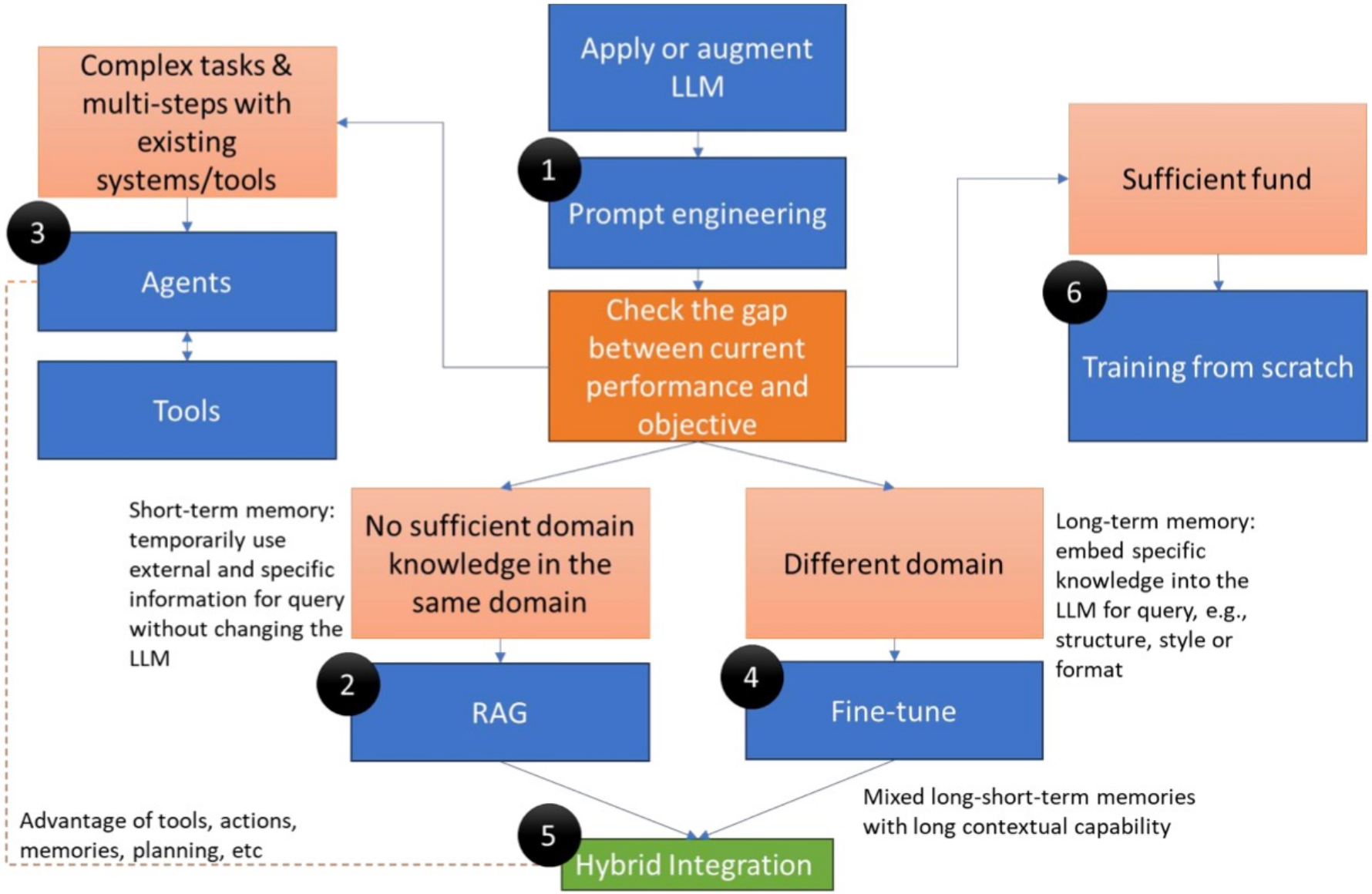
Source: Jun Xu (2024). "GenAI and LLM for Financial Institutions: A Corporate Strategic Survey." Available at SSRN 4988118.

LLM Prompt, RAG, Agent, Fine tuning



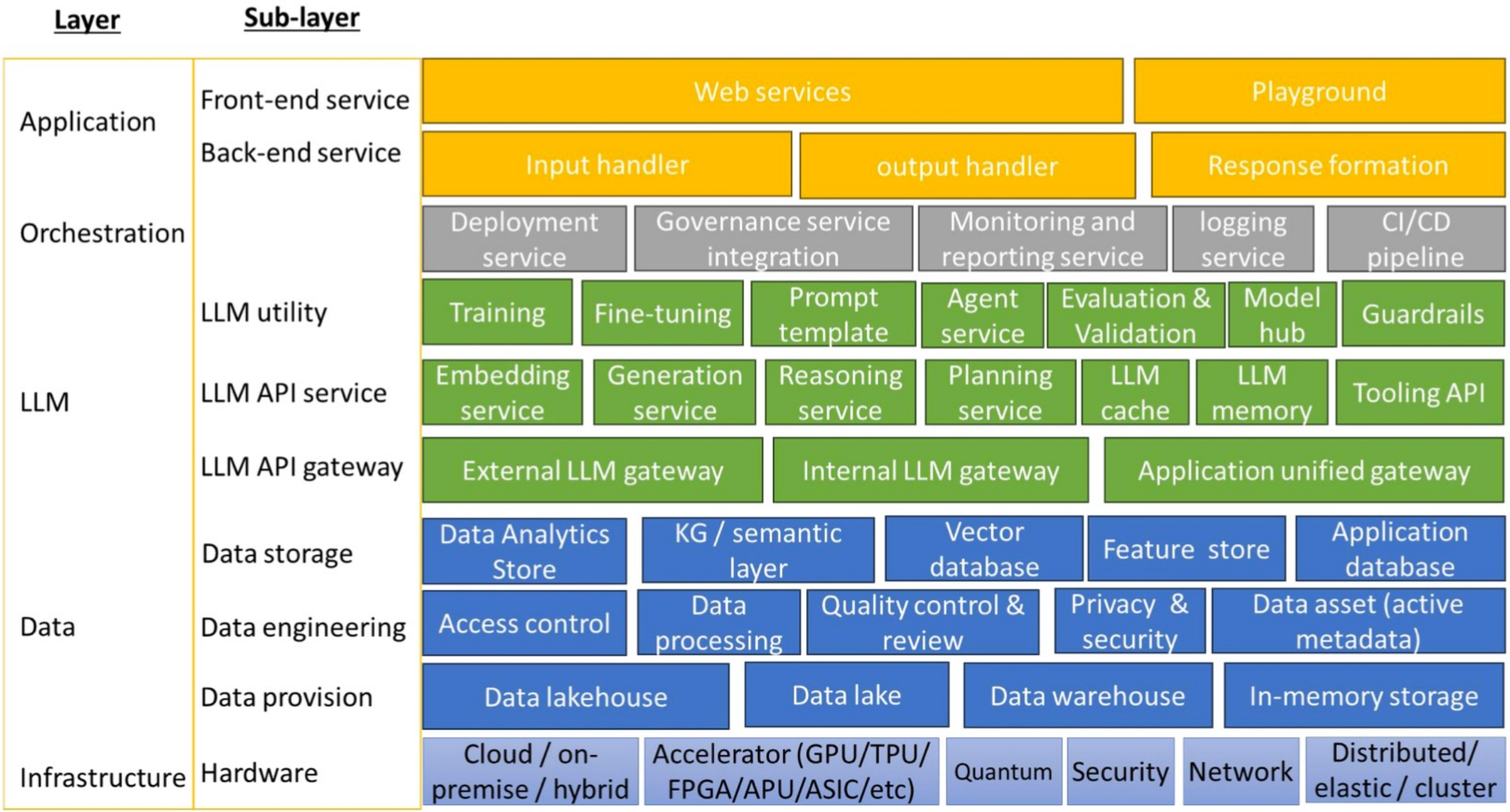
Source: Jun Xu (2024). "GenAI and LLM for Financial Institutions: A Corporate Strategic Survey." Available at SSRN 4988118.

LLM Decision Path for Suitable Techniques



Source: Jun Xu (2024). "GenAI and LLM for Financial Institutions: A Corporate Strategic Survey." Available at SSRN 4988118.

LLM Stack



Source: Jun Xu (2024). "GenAI and LLM for Financial Institutions: A Corporate Strategic Survey." Available at SSRN 4988118.

LLM Functionalities

Service management

Scenario management

Configuration	Task creation
Knowledge management	Application management
pipeline	Automated test

Model/data management

Metadata / data annotation	Collaboration	Content moderation
Data mining & synthesis	Security and privacy: PII detection & masking	

Agent/prompt management

Prompt studio / recipe	Agent studio / template	Tool management
NLP recipe	Safety and compliance	

Online services

Security/persona Auth	Monitoring /Logging / cost reporting	audit trail
guardrails	Validation evaluation	
QC	API gateway	

Engineering

Prompt engineering

Zero-short/ few-short	Chain/tree/graph-of-thought
Pattern-based template: co-star/automate	self-criticism / role-playing

RAG engineering

Query rewriting	Routing	Rerank
Auto-merging	Recursive	Hybrid fusion
compression		Dense x

Agent engineering

ReAct	self-criticism	memory
tooling	planning	decision
workflow	collaboration	

Fine-tuning engineering

LoRa / QLoRa	quantization	SFT / RLHF
distillation	ReFT	Adaptor
PEFT	ZeRo /DeepSpeed	

Technology foundation

LLM (generation)

OpenAI	Gemini / gemma	Claudra
ChatGLM	Llama	Mistral / Mixtral
Phi-MoE	Yi / Qwen	WizardLM
deepseek	Kimi	Blossom

Embedding

BGE	FinGPT
Bert/ FinBert	M3E / ERNIE

Core algorithms

Context fusion	memory
multimodal	Long context

Framework

Langchain	Lammaindex
GPTflow	graphGPT
AutoAgent	HF Agents
Langroid	Haystack

Vector database

pinecore	Chroma	pgvector
Faiss	Milvus	Qdrant

Feature store

Feathr	Databricks	Feast
Vertex AI FS	Hopsworks	AWS FS

Source: Jun Xu (2024). "GenAI and LLM for Financial Institutions: A Corporate Strategic Survey." Available at SSRN 4988118.

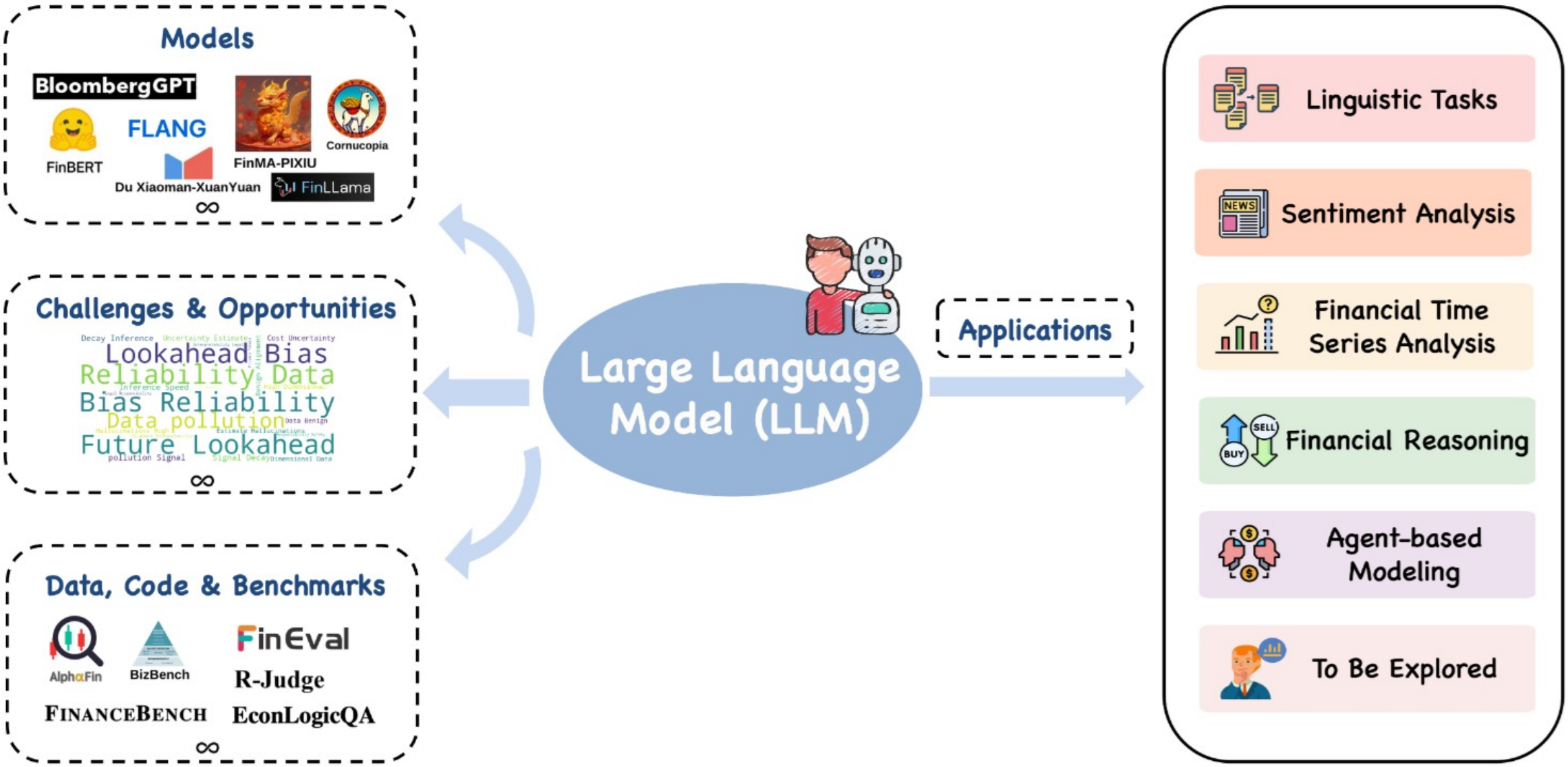
Domain-Specific Language Model (DSLM) for ESG

Model Architecture	Base Model	Parameter Count	Training Focus	Primary Use Case	Key Performance Metric
ClimateBERT	DistilRoBERTa	~82M	Climate news, science, & corporate reports	Fact-checking, Climate Specific Classification	48% improvement in domain MLM tasks (Webersinke et al., 2021)
ESG-BERT	BERT-Base	~110M	Sustainable investing corpora	Sentiment Analysis, Category Classification	F1-score of 0.90 on ESG text mining (Mehra et al., 2022)
SusGen-GPT	Llama/Mistral	7B - 8B	SusGen-30K (Financial + ESG tasks)	TCFD Report Generation, Relation Extraction	~2% performance gap to GPT-4 on specialized tasks (Wu et al., 2024)
ClimateChat	Llama 2 (tuned)	7B+	ClimateChat-Corpus (Q&A pairs)	Climate Science Q&A, Policy Analysis	Improved accuracy on scientific discovery tasks (ClimateChat Team, 2025)

GenAI in ESG Datasets

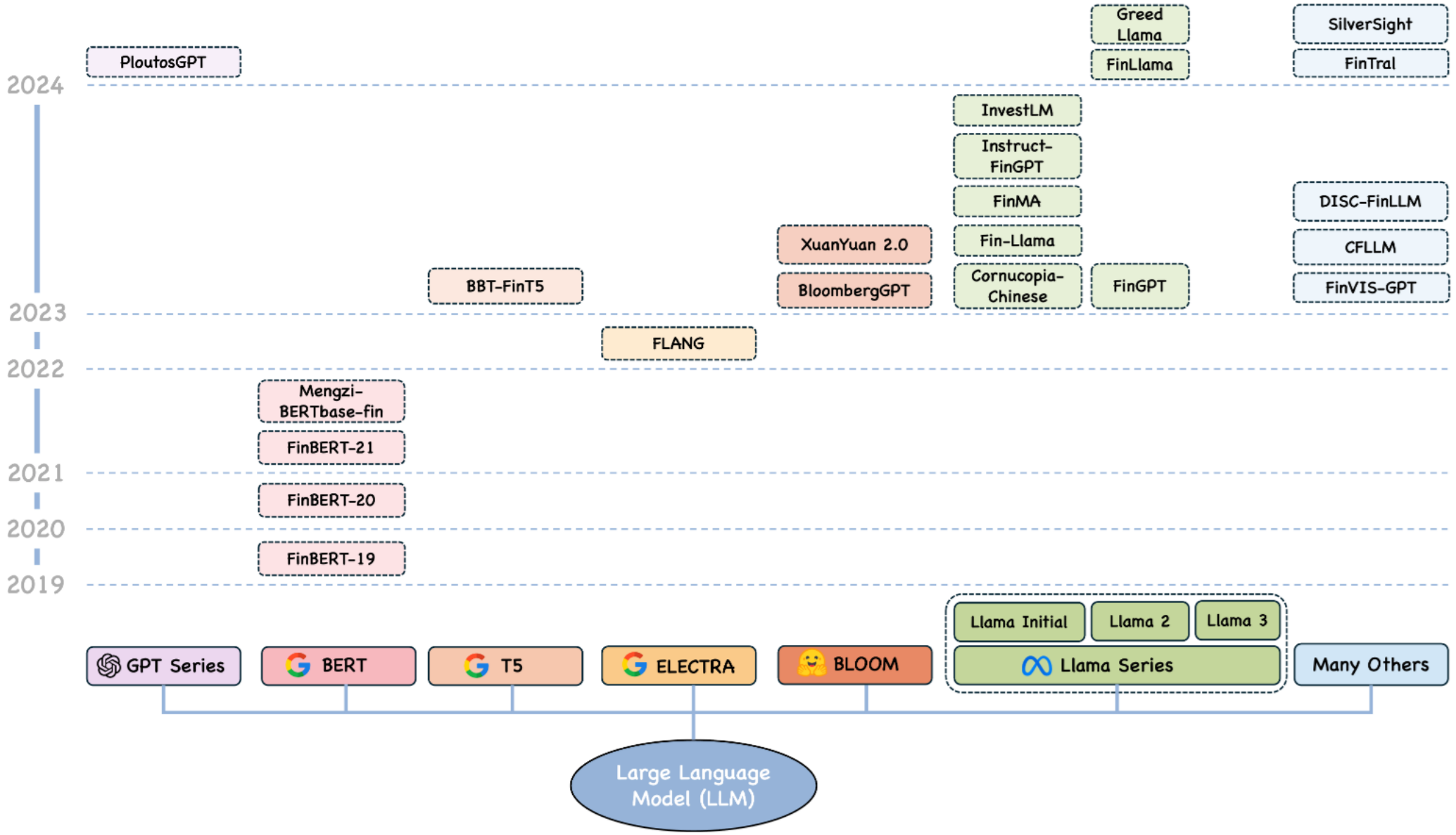
Dataset Name	Size / Composition	Primary Application	Source
SusGen-30K	30,000 samples; 7 financial/ESG tasks. Sourced from TCFDHub & Hugging Face.	Training SusGen-GPT; Balancing financial NLP with ESG reporting tasks.	Wu et al. (2024)
ClimateChat-Corpus	Instruction-tuned Q&A pairs derived from scientific docs and web scraping.	Fine-tuning conversational agents for climate science.	ClimateChat Team (2025)
A3CG	1,679 sustainability reports (SGX companies); Annotated Aspect-Action pairs.	Greenwashing detection; Cross-category generalization testing.	Ong et al. (2025)
ESG-Activities	1,325 labeled text segments classified by EU ESG taxonomy.	Benchmarking LLM performance on granular ESG activity identification.	Intelligent ESG Evaluation Team (2025)
FinRpt	Chinese & English Equity Research Reports.	Evaluating automated generation of comprehensive financial reports.	Li et al. (2025)

Large Language Models for Financial Applications



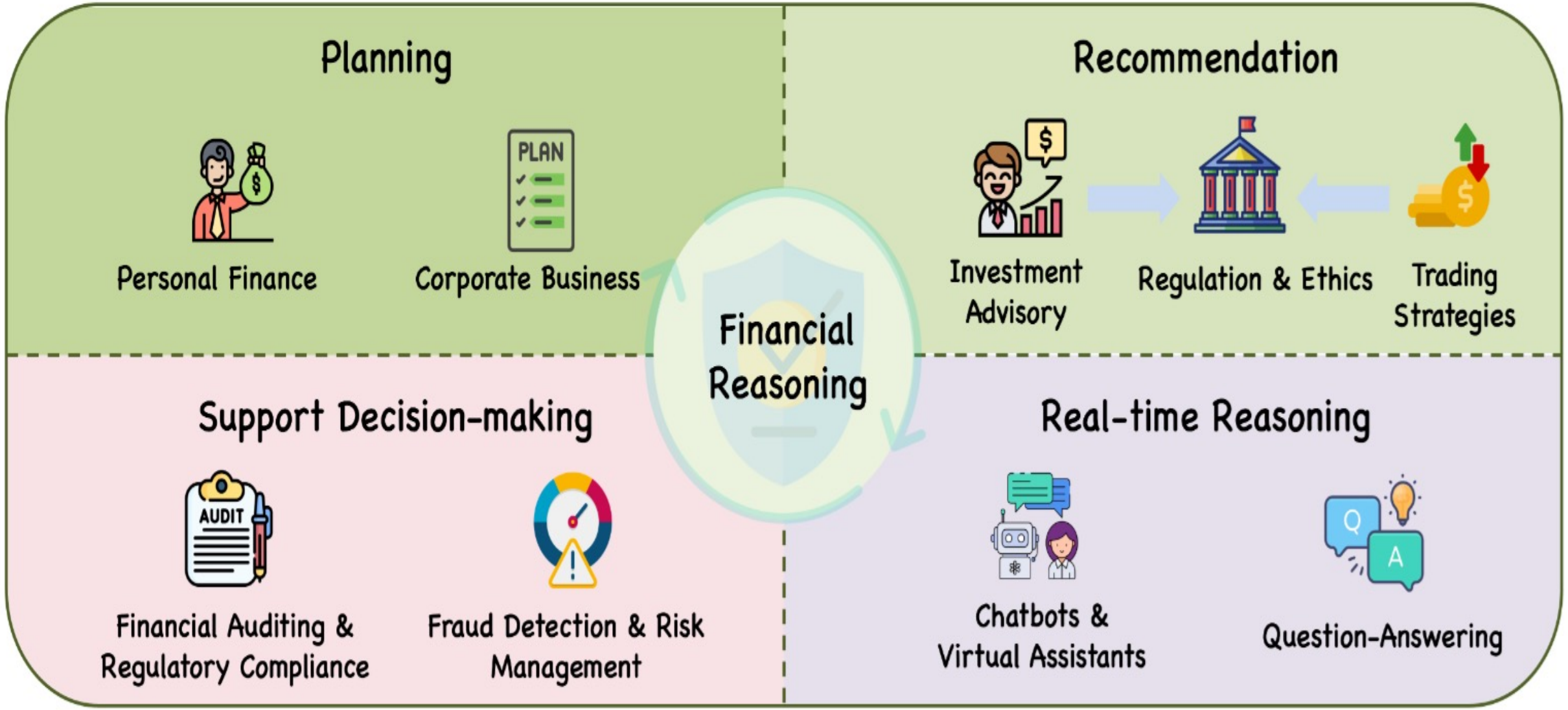
Source: Yuqi Nie, Yaxuan Kong, Xiaowen Dong, John M. Mulvey, H. Vincent Poor, Qingsong Wen, and Stefan Zohren (2024).
"A Survey of Large Language Models for Financial Applications: Progress, Prospects and Challenges." arXiv preprint arXiv:2406.11903.

Financial Large Language Models (Fin LLMs)



Source: Yuqi Nie, Yaxuan Kong, Xiaowen Dong, John M. Mulvey, H. Vincent Poor, Qingsong Wen, and Stefan Zohren (2024).
"A Survey of Large Language Models for Financial Applications: Progress, Prospects and Challenges." arXiv preprint arXiv:2406.11903.

Financial Reasoning Tasks



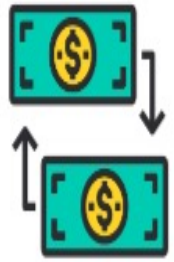
Source: Yuqi Nie, Yaxuan Kong, Xiaowen Dong, John M. Mulvey, H. Vincent Poor, Qingsong Wen, and Stefan Zohren (2024).
"A Survey of Large Language Models for Financial Applications: Progress, Prospects and Challenges." arXiv preprint arXiv:2406.11903.

Financial LLM Agent

Interaction

Downstream Applications

Investment and Trading



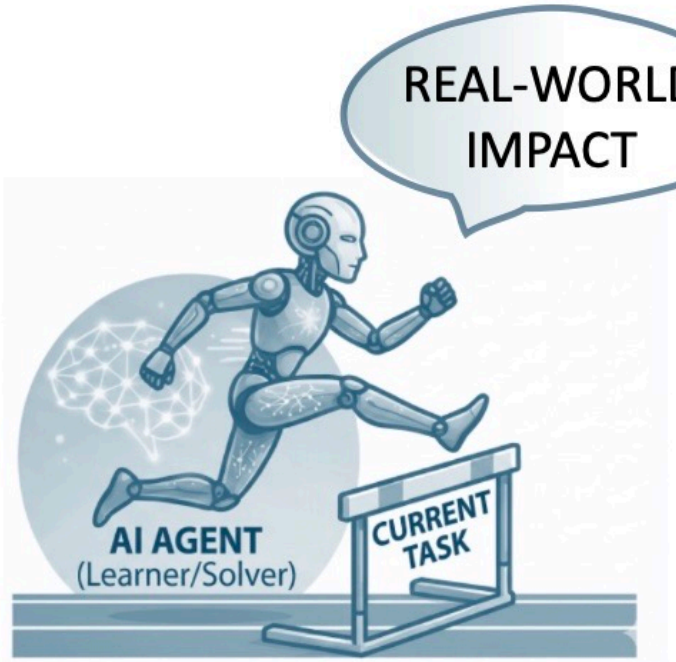
Market Simulator

Automated Process

Multi-agent System

Source: Yuqi Nie, Yaxuan Kong, Xiaowen Dong, John M. Mulvey, H. Vincent Poor, Qingsong Wen, and Stefan Zohren (2024).
"A Survey of Large Language Models for Financial Applications: Progress, Prospects and Challenges." arXiv preprint arXiv:2406.11903.

ESG Agent and ESG Benchmark



ESG Agent

Converter
Python Interpreter
Deep Researcher
Deep Analyzer
Retriever
Plotter
Report tool

ESG Benchmark

Designed Benchmark
Challenge/Standard

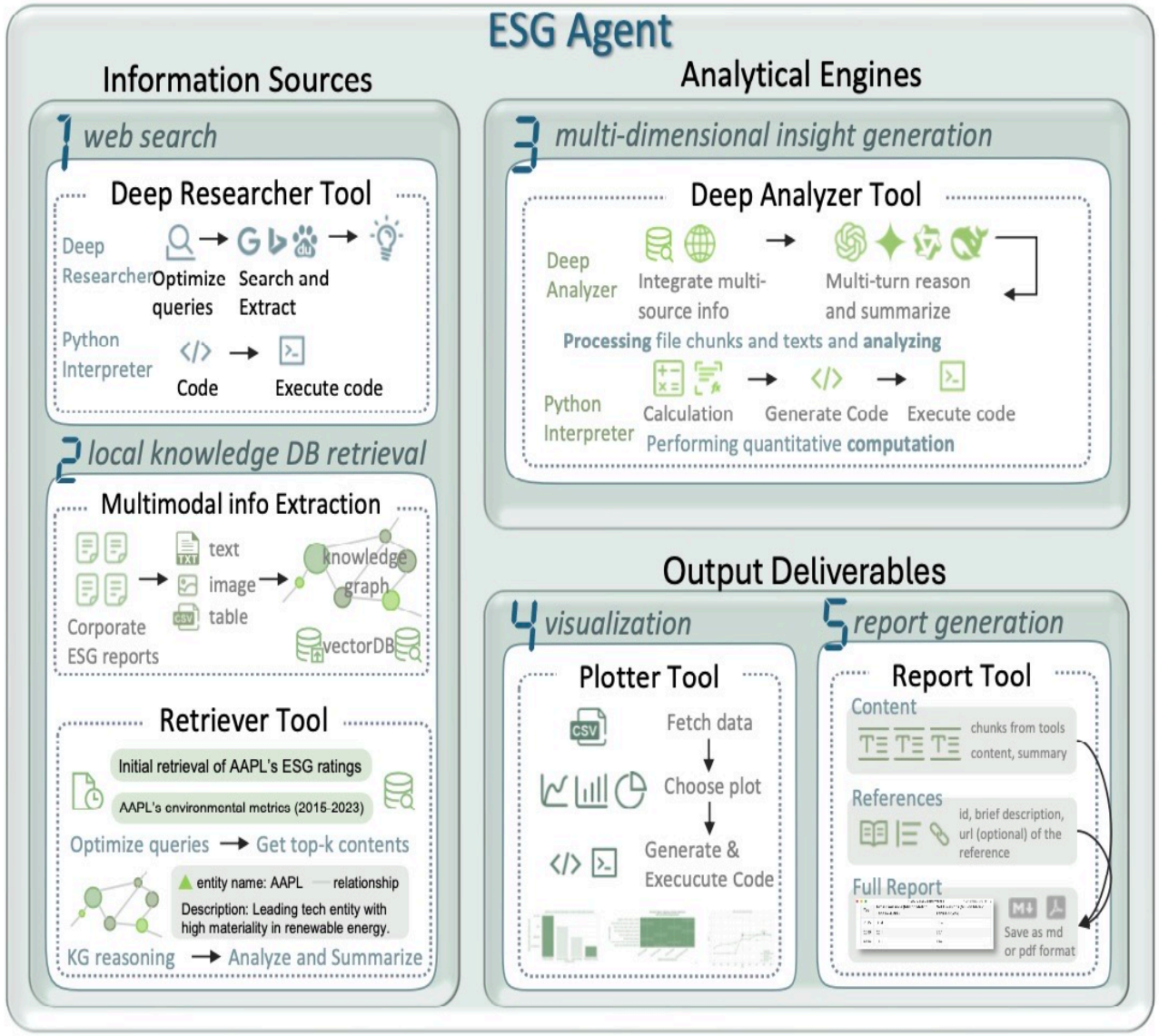
Evaluation & Feedback

Source: Yilei Zhao, Wentao Zhang, Xiao Lei, Yandan Zheng, Mengpu Liu, and Wei Yang Bryan Lim. (2026)
"Advancing ESG Intelligence: An Expert-level Agent and Comprehensive Benchmark for Sustainable Finance." arXiv preprint arXiv:2601.08676 (2026)

ESG Agent Overall Architecture and Workflow

Query / Task

Evaluate **AAPL's** environmental sustainability performance from 2015 to 2023 based on the provided corpus. Your response should include a quantitative trend of **ESG rating analysis** and at least one comparative insight regarding their **carbon neutrality goals**. Formulate your findings into a professional report structure.



AAPL's ESG performance — 4561 Characters

Year	Gross Emissions (Million Metric Tons CO_2e)	Net Emissions (Million Metric Tons CO_2e)
2015	38.4	38.4
2019	25.1	25.1
2023	16.1	15.6

Apple 2030 vs. Industry Peers

Apple's 'Apple 2030' goal is to be carbon neutral across its entire value chain by 2030, which involves a 75% reduction in gross emissions from 2015 levels.

Comparative Insight: Compared to the IPCC recommendation of reaching net zero by 2050, Apple's target is 20 years ahead of global benchmarks. While peers like Microsoft have set ambitious 'carbon negative' targets by 2030, Apple's focus on a 75% absolute reduction in *gross* emissions highlights a rigorous approach to physical decarbonization rather than relying solely on offsets. [30, 31, 36]

1.4 Visualizing the ESG Evolution

The following visualizations track Apple's ESG performance metrics and the strategic pivot in 2019 where environmental performance became a primary driver of the overall score.

Note: The chart highlights the significant improvement in Environmental and Social scores over the nine-year period, alongside the stability of Governance ratings. [36]

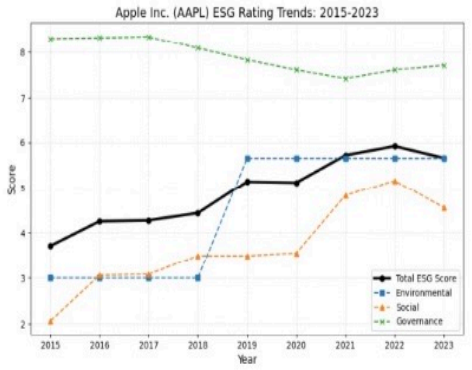


Chart: AAPL ESG Score Breakdown (2015-2023). Note the 2019 inflection point in Environmental scoring corresponding to 100% renewable energy achievement for global

Source: Yilei Zhao, Wentao Zhang, Xiao Lei, Yandan Zheng, Mengpu Liu, and Wei Yang Bryan Lim. (2026) "Advancing ESG Intelligence: An Expert-level Agent and Comprehensive Benchmark for Sustainable Finance." arXiv preprint arXiv:2601.08676 (2026)

References

- Numa Dhamani and Maggie Engler (2024), Introduction to Generative AI, Manning
- 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
- 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
- Yilei Zhao, Wentao Zhang, Xiao Lei, Yandan Zheng, Mengpu Liu, and Wei Yang Bryan Lim. (2026) "Advancing ESG Intelligence: An Expert-level Agent and Comprehensive Benchmark for Sustainable Finance." arXiv preprint arXiv:2601.08676 (2026).
- Anthropic (2026), 2026 Agentic CodingTrends Report: How coding agents are reshaping software development, https://resources.anthropic.com/hubfs/2026%20Agentic%20Coding%20Trends%20Report.pdf
- NVIDIA DLI (2026), Building RAG Agents with LLMs, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-15+V1
- NVIDIA DLI (2026), 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
- Jay 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