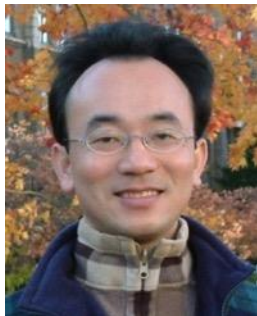


國立臺北大學 商學院碩士在職專班
企業倫理與永續發展
(Business Ethics and Sustainable Development)



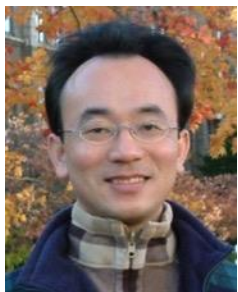
Generative AI for ESG and Sustainable Development (生成式 AI 於 ESG 與永續發展)

Time: 18:50-21:30, Tuesday, March, 11, 2025
企業倫理與永續發展 (Business Ethics and Sustainable Development)
任課教師：陳宥杉，戴敏育



戴敏育 教授 (Prof. Min-Yuh Day)
國立臺北大學 資訊管理研究所 教授
金融科技暨綠色金融研究中心 主任
永續辦公室 永續發展組 組長





戴敏育 教授

Prof. Min-Yuh Day



Professor, Information Management, NTPU

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

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

Division Director, Sustainable Development, Sustainability Office, NTPU

Visiting Scholar, IIS, Academia Sinica

Ph.D., Information Management, NTU

Publications Co-Chairs, International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013-)

Program Co-Chair, IEEE International Workshop on Empirical Methods for Recognizing Inference in Text (IEEE EM-RITE 2012-)

Publications Chair, The IEEE International Conference on Information Reuse and Integration for Data Science (IEEE IRI 2007-)



Outline

- 1. Generative AI**
- 2. ESG**
- 3. Sustainable Development**

衡量企業永續關鍵指標

臺北大學獨創ESG永續評鑑系統

社會(S)	經濟(E)	環境(E)	揭露(D)
<ul style="list-style-type: none"> 1.人權 2.員工溝通與福利 3.人力資本發展 4.多元組成與包容性 5.供應鏈社會面向控管 6.客戶關係管理 7.產品安全 8.企業公民與慈善 	<ul style="list-style-type: none"> 1.股東權益 2.董事會結構與運作 3.行為準則與內控 4.風險及危機管理 5.永續金融 6.ESG創新 	<ul style="list-style-type: none"> 1.環境系統與治理 2.空氣管理 3.能源與氣候變遷 4.水管理 5.原物料與廢棄物管理/ 資源與廢棄物管理 6.生物多樣性 7.供應商及產品生命週期管理/ 供應鏈環境面向管理 	<ul style="list-style-type: none"> 1.ESG 揭露



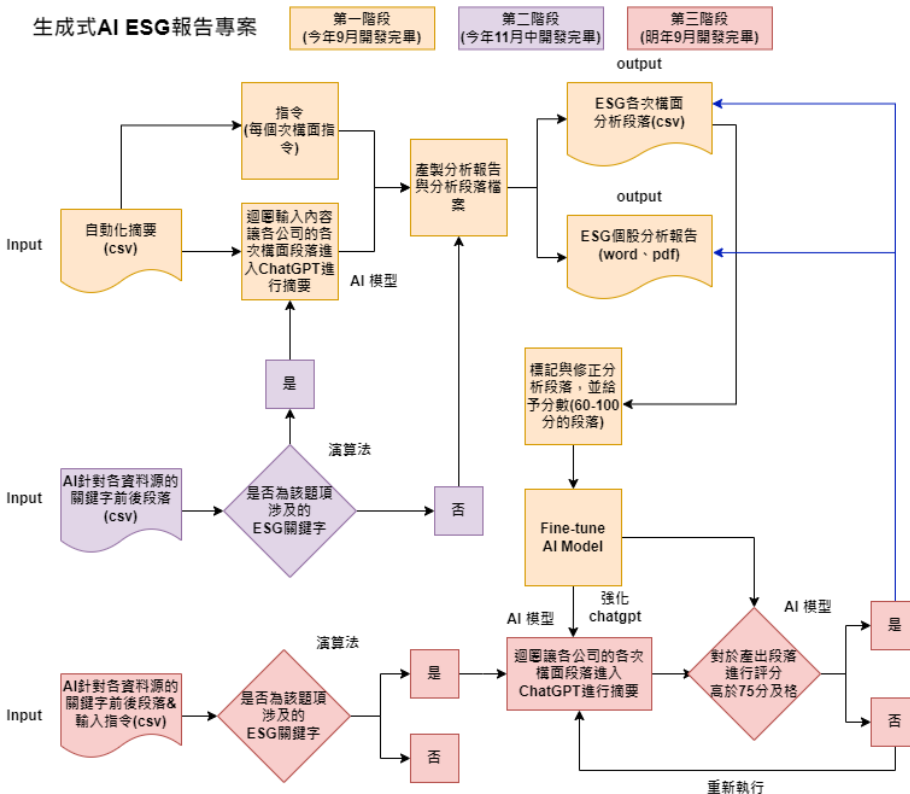
台灣永續評鑑

國立臺北大學商學院企業永續發展研究團隊

透過 AI SEED 提升評鑑效率

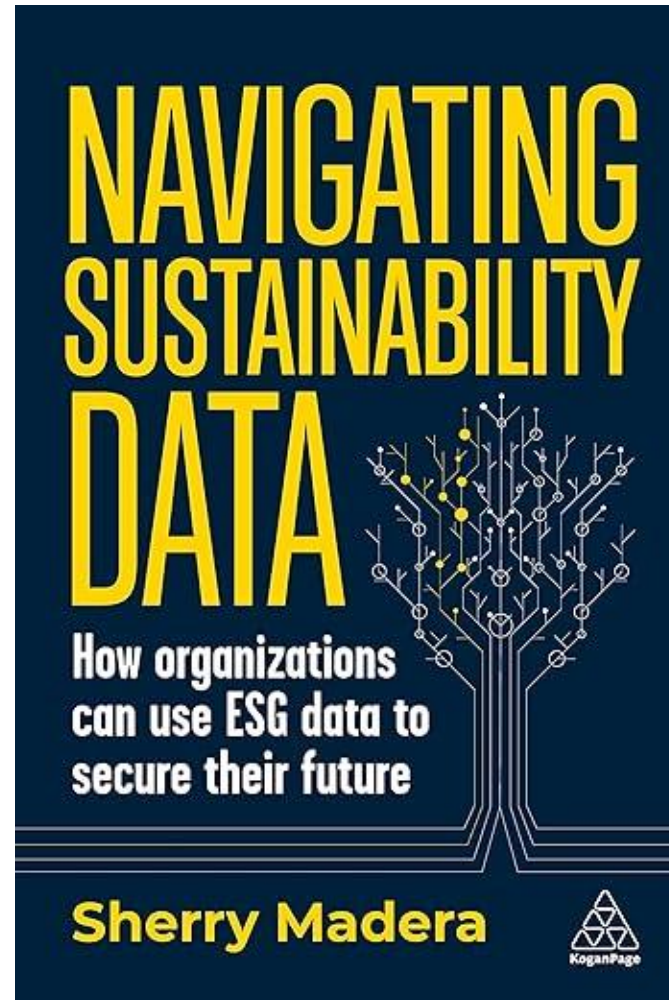


為使評鑑效率提升，與國立臺北大學資管所及資工所合作，開發相關程式，已有25%題項自動或半自動化，大幅提升評鑑效率，並持續開發機械學習，持續透過AI 輔助評鑑進行。另也透過AI SEED團隊持續將部分流程自動化，提升評鑑正確性，減少人力出錯可能。

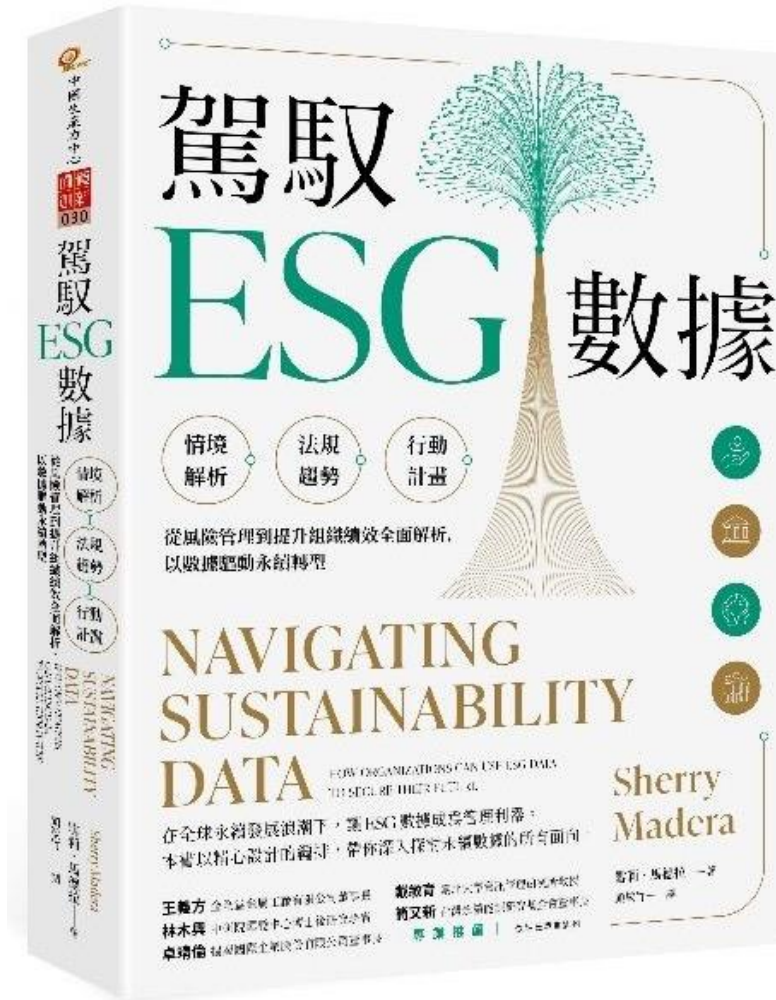
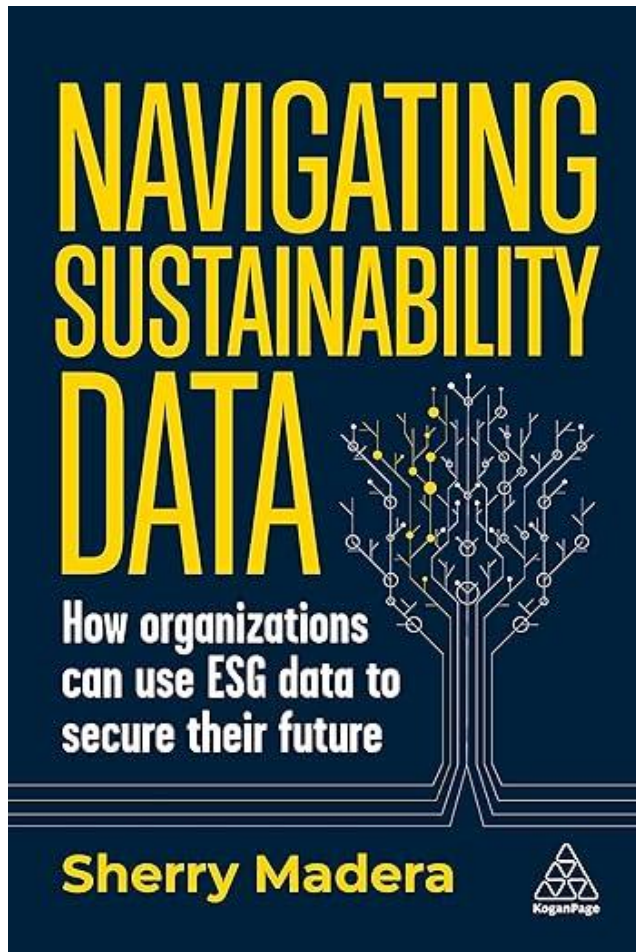


題號	題目關鍵字	完成度
2-1-15	僅分派董監酬勞未分派股利	100%
2-2-3	獨董達董事席次1/2以上	100%
2-2-4	至少兩名獨董任期不超過9年	100%
2-2-14	設提名委員會且半數以上為獨董	100%
2-2-30	董事長兼任總經理	100%
2-2-31	1/3以上董事任期超過15年	100%
2-3-8	破產 / 面臨下市	100%
2-3-15	資安長或資訊安全委員會	100%
2-4-7	無保留意見	100%

**Sherry Madera (2024),
Navigating Sustainability Data: How Organizations can use ESG
Data to Secure Their Future, Kogan Page**



雪莉·馬德拉 (Sherry Madera) (顏敏竹 譯) (2024), 駕馭ESG數據 (Navigating Sustainability Data), 財團法人中國生產力中心



專業推薦：

王義方

(金全益金屬工廠有限公司董事長)

林木興

(中研院環變中心博士後研究學者)

卓靖倫

(揚秦國際企業股份有限公司董事長)

戴敏育

(國立臺北大學資訊管理研究所教授)

簡又新

(台灣永續能源研究基金會董事長)

Generative AI-Driven ESG Report Generation Technology

Industrial Technology Research Institute (ITRI),
Fintech and Green Finance Center (FGFC, NTPU),
NTPU-113A513E01, 2024/03/01~2024/12/31

Generative AI

Powering

Digital Sustainability

Transformation

Generative AI

(Gen AI)

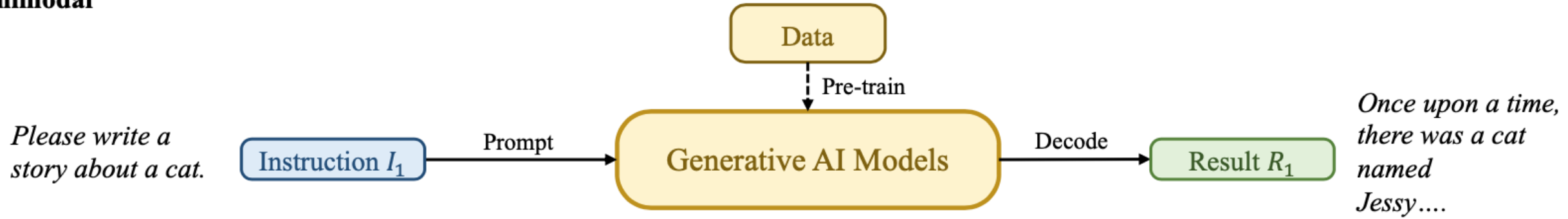
AI Generated Content

(AIGC)

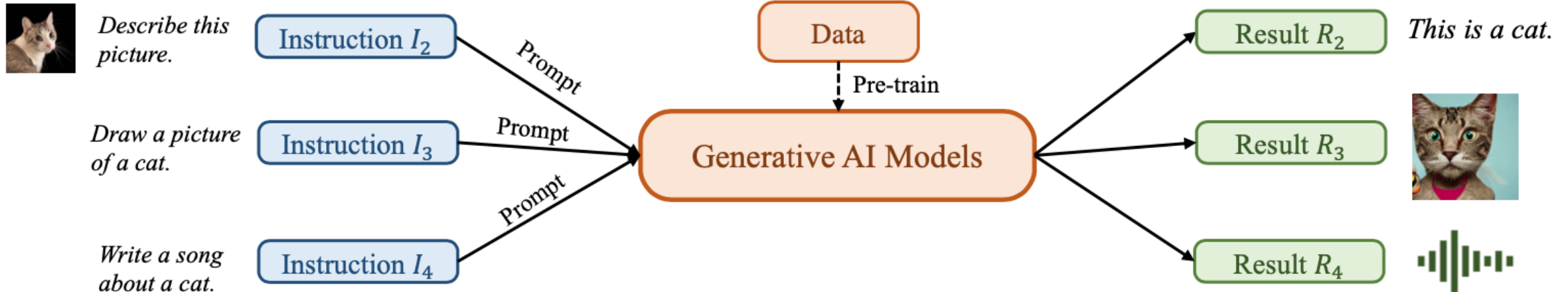
Generative AI (Gen AI)

AI Generated Content (AIGC)

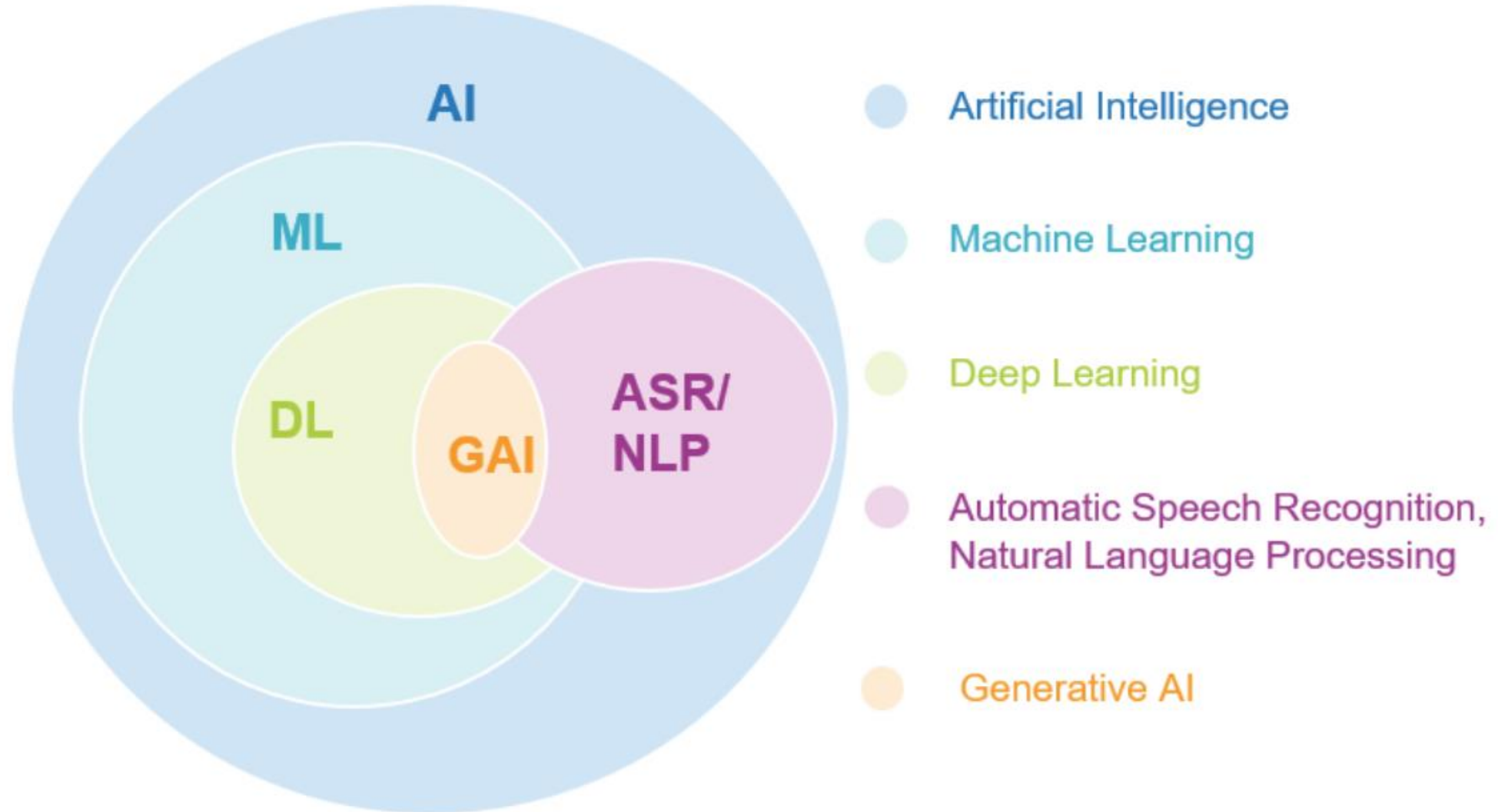
Unimodal



Multimodal



AI, ML, DL, Generative AI



Definition of Artificial Intelligence (A.I.)

Artificial Intelligence

**“... the science and
engineering
of
making
intelligent machines”
(John McCarthy, 1955)**

Artificial Intelligence

**“... technology that
thinks and acts
like humans”**

Artificial Intelligence

**“... intelligence
exhibited by machines
or software”**

4 Approaches of AI

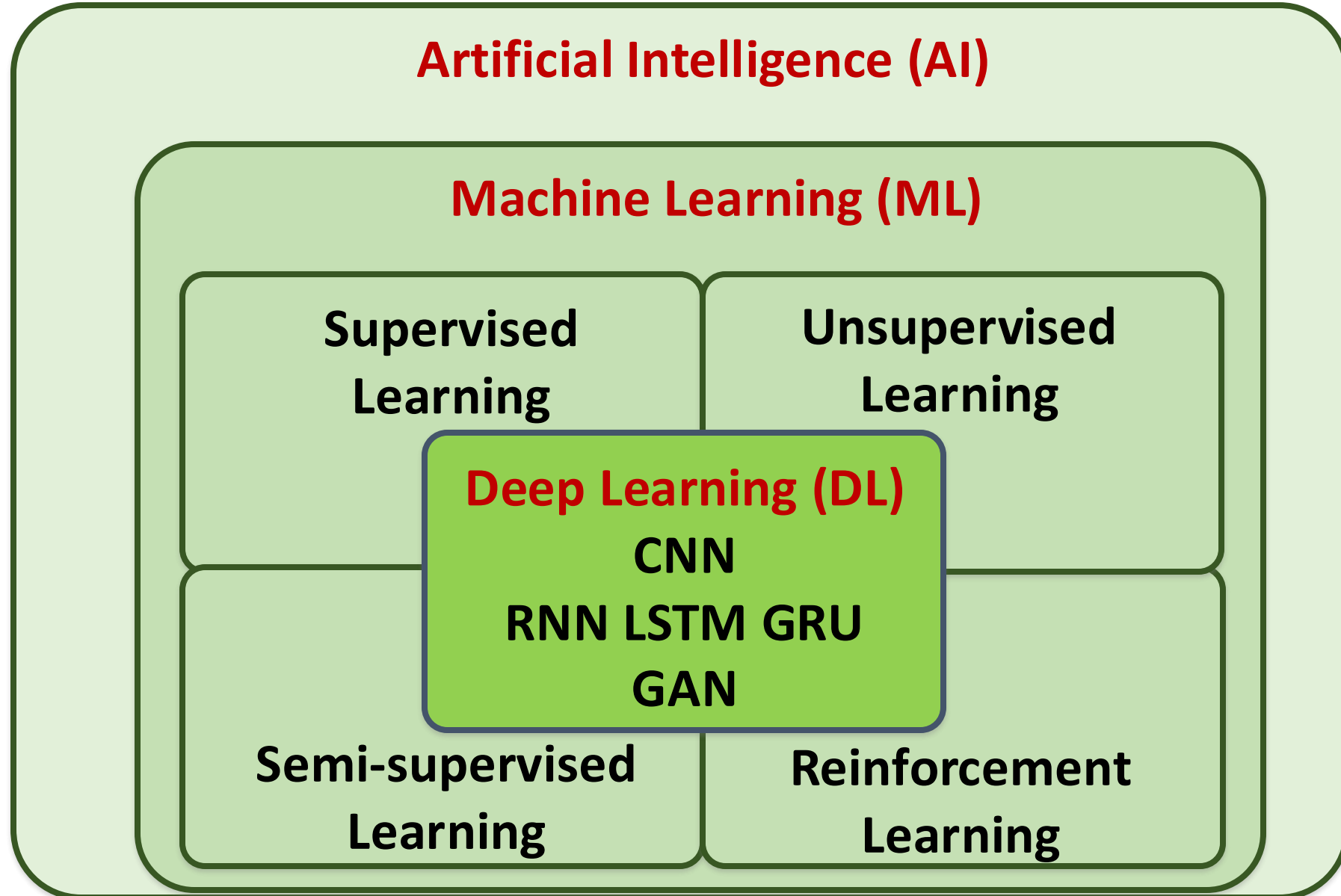
<p>2. Thinking Humanly: The Cognitive Modeling Approach</p>	<p>3. Thinking Rationally: The “Laws of Thought” Approach</p>
<p>1. Acting Humanly: The Turing Test Approach (1950)</p>	<p>4. Acting Rationally: The Rational Agent Approach</p>

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

AI, ML, DL



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

Generative AI

- **Generative AI: The Art of Creation**
- **Definition: AI systems capable of creating new content**
- **Characteristics: Creativity, interactivity**

Generative AI and Large Language Models (LLMs): Popular Generative AI Applications

Generative AI

Large Language Models

(LLMs)

Foundation Models

Spring 2025

Generative AI
Innovative Applications



University Ambassador



This certificate acknowledges that

Min-Yuh Day

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

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

Greg Estes

Vice President, NVIDIA

Issue Date: : March 7, 2025

Ambassador Certification ID: cCFh1ZWWTvqKTq7dcKkEWw



Certified Instructor



This certificate acknowledges that

Min-Yuh Day

has been certified to deliver the instructor-led workshop

Building RAG Agents with LLMs

A handwritten signature in black ink, appearing to read "Greg Estes".

Greg Estes

Vice President, NVIDIA

Issue Date: : March 7, 2025

Certification ID: OVmqY4cSSya0BdMQBWHxzw

NVIDIA Developer Program

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

NVIDIA

Deep Learning Institute (DLI)

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

Get NVIDIA DLI Certificate

- Welcome to NVIDIA DLI "**Building RAG Agents with LLMs**" **workshop** scheduled for **March 11th-25th, 2025** at NTPU.
- Step 1. Visit <https://learn.nvidia.com/dli-event>
- Step 2. Enter the event code: **NTPU_RAG_AMBASSADOR_MA25**
- Step 3. Complete the NVIDIA DLI course and review the course datasheet (on March 25, 2025) at <https://developer.nvidia.com/dli/getready>

Get NVIDIA DLI Certificate Before the NVIDIA Workshop

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

NVIDIA Deep Learning Institute (DLI)

Building RAG Agents with LLMs workshop

March 11th-25th, 2025 at NTPU

Self-Paced Course	Instructor-Led Workshop
Building RAG Agents With LLMs	Building RAG Agents With LLMs
Certificate available Free 8 hours	Certificate available \$500 8 hours

Step 1.

Visit

<https://learn.nvidia.com/dli-event>

Step 2.

Enter the event code:

NTPU_RAG_AMBASSADOR_MA25



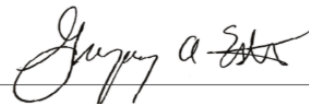
Certificate of Competency

This certificate is awarded to

Min-Yuh Day

for demonstrating competence in the completion of

Building RAG Agents with LLMs

[Download Certificate](#)[Add to Profile](#)

Greg Estes

Vice President, NVIDIA

Issue Date: : December 8, 2024

Certification ID: ed-qOCIMQaatzU8SNUNxgw

Join the NVIDIA Developer Program

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

Generative AI and LLMs Graphics and Simulation Accelerated Computing Data Science Deep Learning

<p>8 hours</p> <h3>Getting Started With Deep Learning</h3> <p>Explore the fundamentals of deep learning by training neural networks and using results to improve performance and capabilities.</p>	<p>2 hours</p> <h3>Modeling Time-Series Data With Recurrent Neural Networks in Keras</h3> <p>Explore how to classify and forecast time-series data using recurrent neural networks (RNNs), such as modeling a patient's health over time.</p>	<p>4 hours</p> <h3>Deploying a Model for Inference at Production Scale</h3> <p>Learn how to deploy your own machine learning models on a GPU server.</p>
<p>8 hours</p> <h3>Building Real-Time Video AI Applications</h3> <p>Gain the knowledge and skills needed to enable the real-time transformation of raw video data from widely deployed camera sensors into deep learning-based insights.</p>	<p>2 hours</p> <h3>Introduction to Graph Neural Networks</h3> <p>Learn the basic concepts, models, and applications of graph neural networks.</p>	<p>4 hours</p> <h3>Introduction to Physics-Informed Machine Learning With Modulus</h3> <p>Learn the various building blocks of NVIDIA Modulus, which turbocharges use cases by building physics-based deep learning models that are 100,000X faster than traditional methods and offers high-fidelity simulation results.</p>
<p>2 hours</p> <h3>Get Started With Highly Accurate Custom ASR for Speech AI</h3> <p>Learn to build, train, fine-tune, and deploy a GPU-accelerated automatic speech recognition (ASR) service with NVIDIA® Riva that includes customized features.</p>	<p>2 hours</p> <h3>Integrating Sensors With NVIDIA DRIVE</h3> <p>Find out how to integrate automotive sensors into your applications using NVIDIA DRIVE®.</p>	

<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

What do you want to learn today?

Filters

Level +

Format +

Topics -

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

Free / Paid +

Language +

Generative AI



Sort by: --

Showing 19 results

Generative AI x

Generative AI

All Courses

Self-paced

Generative AI Explained

Free
02:00

Self-paced

Generative AI with Diffusion Models

\$90
08:00

Instructor-Led

Generative AI with Diffusion Models

08:00

Self-paced

Augment your LLM Using

Self-paced

Introduction to Transformer-

Instructor-Led

Rapid Application

Building RAG Agents with LLMs

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

Self-paced Course

Building RAG Agents with LLMs

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



About Course Objectives Topics Covered Course Outline Stay Informed Contact Us

Continue Learning

About this Course

This course is free for a limited time.

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

Course Details

Duration: 08:00

Price: Free

Level: Technical - Intermediate

Subject: Generative AI/LLM

Language: English

Course Prerequisites:

Introductory deep learning knowledge, with comfort

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

Generative AI Explained

Self-paced Course

Generative AI Explained

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

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

About this Course

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

Learning Objectives

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

Course Details

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

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

Introduction to Transformer-Based Natural Language Processing

Self-paced Course

Introduction to Transformer-Based Natural Language Processing

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

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

About this Course

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

Course Details

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

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

Generative AI with Diffusion Models

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

Self-paced Course

Generative AI with Diffusion Models

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

About Course

Objectives

Topics Covered

Course Outline

Stay Informed

Contact Us

Continue Learning

About this Course

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

Learning Objectives

Course Details

Duration: 08:00

Price: \$90

Subject: Generative AI/LLM

Language: English

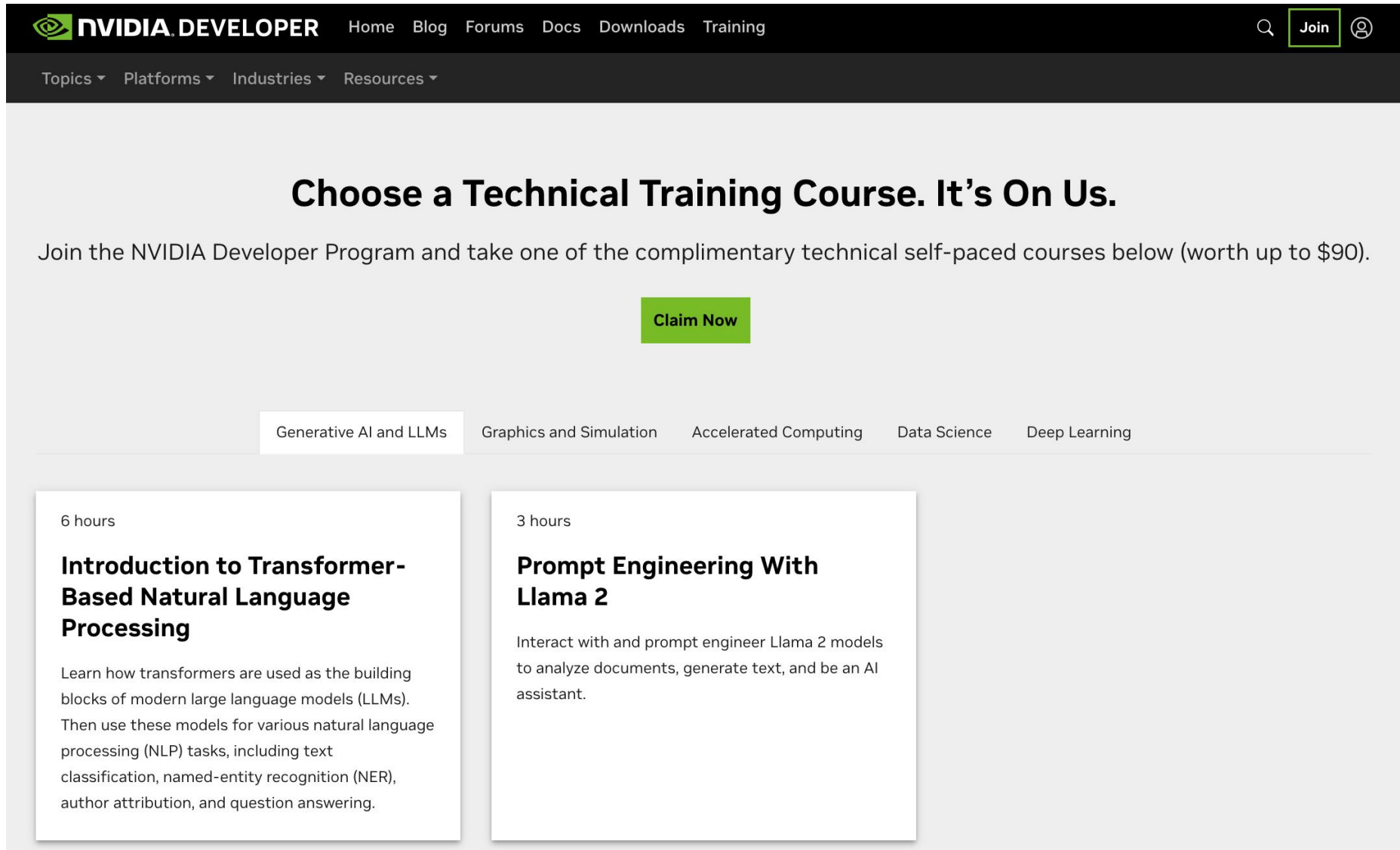
Course Prerequisites:

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

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)

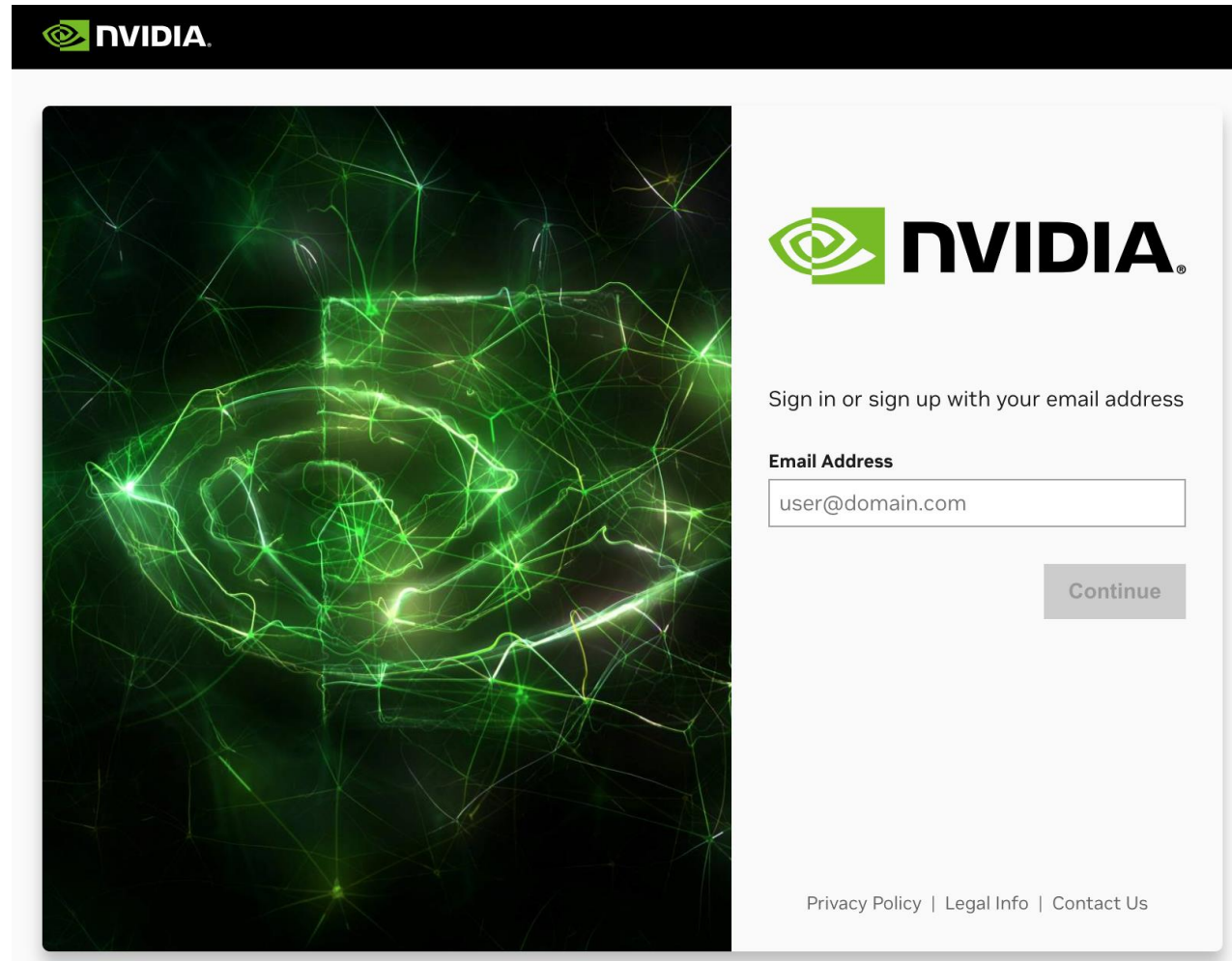


The screenshot shows the NVIDIA Developer Program website. At the top, there is a navigation bar with the NVIDIA logo and the text "NVIDIA DEVELOPER". To the right of the logo are links for "Home", "Blog", "Forums", "Docs", "Downloads", and "Training". Further right is a search icon, a "Join" button,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


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


Join the NVIDIA Developer Program

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



The screenshot shows the NVIDIA Developer Program sign-up page. It features a black header with the NVIDIA logo. The main content area is split into two columns. The left column contains a large, abstract image of a glowing green neural network or data visualization. The right column contains the NVIDIA logo, a sign-in/sign-up prompt, an email address input field with the placeholder 'user@domain.com', a 'Continue' button, and a footer with links for 'Privacy Policy', 'Legal Info', and 'Contact Us'.

 NVIDIA.

 NVIDIA.

Sign in or sign up with your email address

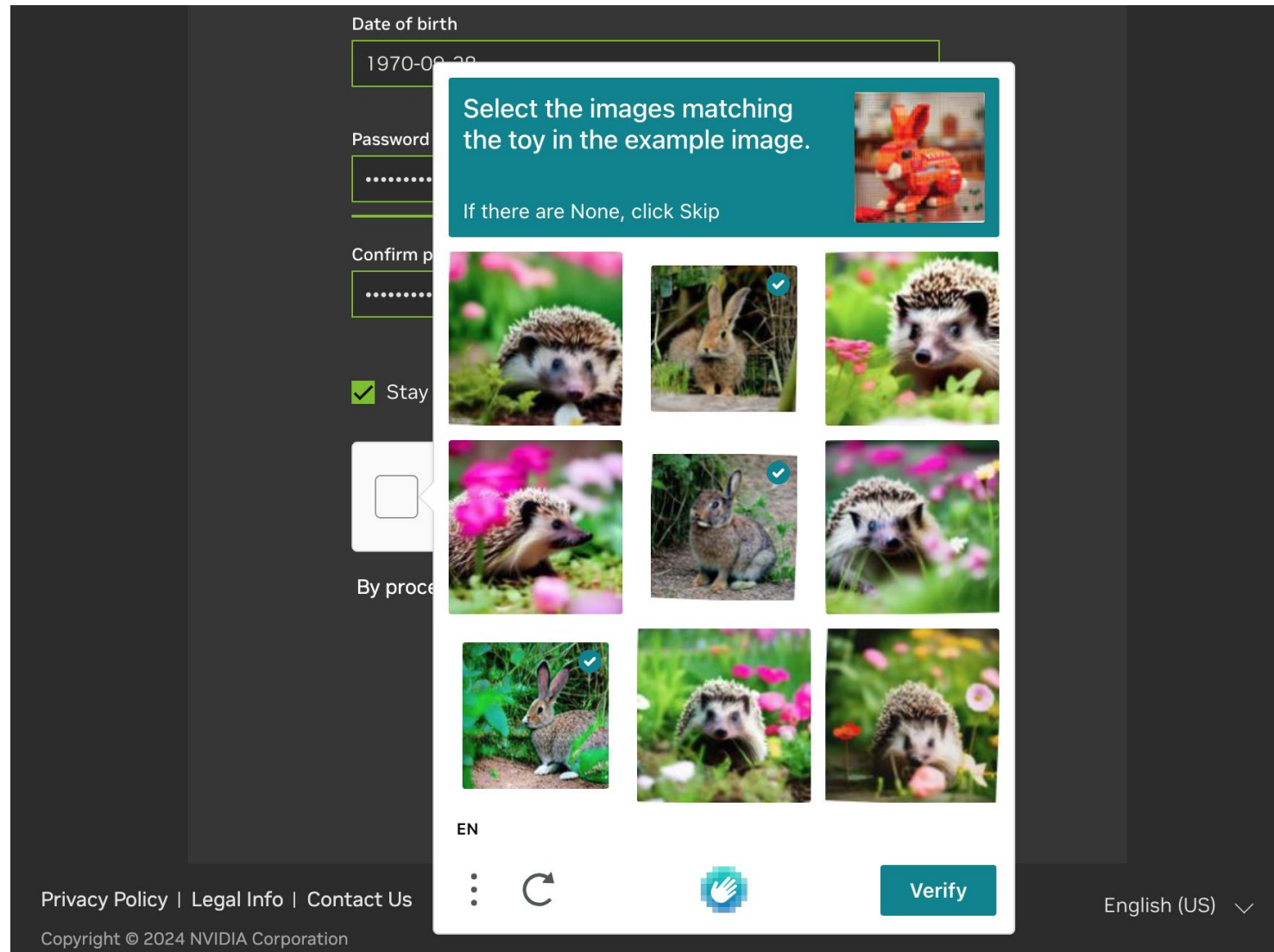
Email Address

[Continue](#)

[Privacy Policy](#) | [Legal Info](#) | [Contact Us](#)

Join the NVIDIA Developer Program

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



The screenshot displays a registration form with fields for "Date of birth" (1970-00-20), "Password", and "Confirm password". A "Stay" checkbox is checked. A CAPTCHA challenge is overlaid, asking the user to "Select the images matching the toy in the example image." The example image shows a red toy rabbit. The challenge grid contains nine images: three rabbits and six hedgehogs. The rabbits in the top-middle, middle-middle, and bottom-left positions are marked with blue checkmarks. At the bottom of the CAPTCHA window, there are icons for a menu, a refresh button, a hand icon, and a "Verify" button. The background shows a "By proceeding" button and a "Privacy Policy | Legal Info | Contact Us" link. The footer includes "Copyright © 2024 NVIDIA Corporation" and "English (US) ▾".

Join the NVIDIA Developer Program

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

 **NVIDIA. DEVELOPER** DEEP LEARNING INSTITUTE PROGRAM BENEFITS

SELECT YOUR FREE COURSE.


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


English ▾

- Integrating Sensors with NVIDIA DRIVE®
- Getting Started with Deep Learning
- Deploying a Model for Inference at Production Scale
- Get Started with Highly Accurate Custom ASR for Speech AI
- Introduction to Graph Neural Networks
- Introduction to Transformer-Based Natural Language Processing
- Prompt Engineering with LLaMA-2 (Access Expires Dec. 5th 2025)
- Generative AI with Diffusion Models**
- Building Real-Time Video AI Applications
- Introduction to Robotic Simulations in Isaac Sim

Generative AI with Diffusion Models

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

 Certificate Available

 Duration: 08:00

[Continue >](#)

Join the NVIDIA Developer Program

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

 NVIDIA DEVELOPER DEEP LEARNING INSTITUTE PROGRAM BENEFITS

SELECT YOUR FREE COURSE.


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


English ▾

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

Getting Started with Deep Learning

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

 Certificate Available

 Duration: 08:00

[Continue >](#)

Join the NVIDIA Developer Program

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

 NVIDIA DEVELOPER DEEP LEARNING INSTITUTE PROGRAM BENEFITS

SELECT YOUR FREE COURSE.

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

English

Get Started with Highly Accurate Custom ASR for Speech AI

Introduction to Graph Neural Networks

Introduction to Transformer-Based Natural Language Processing

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

Generative AI with Diffusion Models

Building Real-Time Video AI Applications

Introduction to Robotic Simulations in Isaac Sim

Introduction to Physics-informed Machine Learning with Modulus

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

Synthetic Data Generation for Training Computer Vision Models

Generative AI with Diffusion Models

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


 Certificate Available

 Duration: 08:00

Continue >

NVIDIA Deep Learning Institute (DLI)

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


Search 

Monthly Activity

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

Skills

Certificates


No Certificates
You don't have any certificates yet.

Courses in Progress

Self-paced

Generative AI with Diffusion Models

0% Completed
08:00

Generative AI with Diffusion Models

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

Generative AI with Diffusion Models > Start Here > 0: Server Access

Generative AI with Diffusion Models

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

0: Server Access

[Bookmark this page](#)

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

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

1: From U-Nets to Diffusion

[Bookmark this page](#)

Theory

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

Deep Learning Institute (DLI)

Monthly Activity

Skill Points	0
Time Spent	
Courses in Progress	16
Courses Completed	12
Watched Videos	
Assessments	

Skills

Certificates

- Introduction to Transformer-Based Natural Language Processing
- Building RAG Agents with LLMs**
- Building RAG Agents with LLMs
- Accelerating End-to-End Data Science Workflows
- Generative AI with Diffusion Models
- Building Agentic AI Applications with LLMs

Completed Courses

View more

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



Certificate of Completion

This certificate is awarded to

Min-Yuh Day

for successfully completing

Building RAG Agents with LLMs

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

Greg Estes

Vice President, NVIDIA

Issue Date: : December 8, 2024

Certification ID: ed-qOCIMQatzU8SNUNxgw |

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

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

NVIDIA Deep Learning Institute (DLI)

All Self-Paced Courses

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

[Share Generative AI/LLM Courses](#)

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

Building RAG Agents with LLMs

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

Self-paced Course

Building RAG Agents with LLMs

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



About Course Objectives Topics Covered Course Outline Stay Informed Contact Us

Continue Learning

About this Course

This course is free for a limited time.

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

Course Details

Duration: 08:00

Price: Free

Level: Technical - Intermediate

Subject: Generative AI/LLM

Language: English

Course Prerequisites:

Introductory deep learning knowledge, with comfort

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

Building RAG Agents with LLMs

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

Building RAG Agents with LLMs

Course Progress Bookmarks Updates

Building RAG Agents with LLMs Introduction Introduction

Building RAG Agents with LLMs

Introduction

Environment and LLMs

LangChain

Documents and Embeddings

Retrieval-Augmented Generation

Next Steps

Feedback

Previous

Next



Building RAG Agents with LLMs

Introduction



Building RAG Agents with LLMs

Deep Learning Institute

Find Training

Self Paced Courses

Instructor-Led Workshops

Educator Programs

Enter

Training Home

My Learning

NVIDIA Account

Logout

Building RAG Agents with LLMs

Course

Progress

Bookmarks

Updates

Building RAG Agents with LLMs

Introduction

Introduction

Building RAG Agents with LLMs

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

Building RAG Agents with LLMs

Course Progress Bookmarks Updates

Building RAG Agents with LLMs Environment and LLMs Environment [0, 1, 2]

Building RAG Agents with LLMs

Introduction

Environment and LLMs

Environment [0, 1, 2]

Part 1: Course Environment

Part 2: LLM Services

LangChain

Environment [3, 4]

Part 3: LangChain

Part 4: Running States

Documents and Embeddings

Environment [5, 6]

Previous

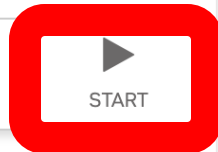
Next

Bookmark this page

Welcome to **Building RAG Agents with LLMs**. In this first section, we will get introduced to the overall course environment, LLM services, and recommended workflows!

This tab contains the course environment for this section, which will contain the notebooks for the next two videos! Please click through the videos in the remaining tabs to watch the material and work through the exercises!

Please click the "Start" button to start up your own private server for hands-on coding practice. It will take a few minutes to start up, so go ahead and click it now and then proceed to the next video! After a few minutes when the server has loaded, click "Launch" to access the code labs.



Building RAG Agents with LLMs

Building RAG Agents with LLMs

Course Progress Bookmarks Updates

Building RAG Agents with LLMs Environment and LLMs Environment [0, 1, 2]

Building RAG Agents with LLMs

Introduction

Introduction

Course Slides

Environment and LLMs

Environment [0, 1, 2]

Part 1: Course Environment

Part 2: LLM Services

LangChain

Environment [3, 4]

Part 3: LangChain

Previous

Next

welcome to **BUILDING RAG AGENTS WITH LLMs**. In this first section, we will get introduced to the overall course environment, LLM services, and recommended workflows!

This tab contains the course environment for this section, which will contain the notebooks for the next two videos! Please click through the videos in the remaining tabs to watch the material and work through the exercises!

Please click the "Start" button to start up your own private server for hands-on coding practice. It will take a few minutes to start up, so go ahead and click it now and then proceed to the next video! After a few minutes when the server has loaded, click "Launch" to access the code labs.



This Lab 0 : 01 : 06 / 2 : 00 : 00

Course 13 : 45 : 51 / 32 : 00 : 00

LAUNCH

STOP TASK

Building RAG Agents with LLMs

Browser address: Not Secure 54.211.119.108/lab/lab

Menu: File Edit View Run Kernel Tabs Settings Help

File browser search: Filter files by name

Name	Last Modified
chatbot	yesterday
composer	yesterday
docker_router	yesterday
frontend	yesterday
imgs	yesterday
llm_client	yesterday
slides	yesterday
00_jupyterlab.ipynb	yesterday
01_microservices.ipynb	yesterday
02_llms.ipynb	yesterday
99_table_of_contents.ipynb	yesterday

Launcher options:

- Notebook: Python 3 (ipykernel)
- Console: Python 3 (ipykernel)
- Other: Terminal, Text File, Markdown File, Python File, Show Contextual Help

Bottom status: Simple 0 \$ 0 Launcher 1

Building RAG Agents with LLMs

The screenshot shows a JupyterLab environment. On the left is a file browser with a search bar and a list of files and folders. The main area is a notebook editor with a 'Launcher' tab selected. The notebook content features the NVIDIA logo and the text 'DEEP LEARNING INSTITUTE'. Below this is the title 'Notebook 0: JupyterLab' and a paragraph of text. At the bottom, a green arrow points to the 'Menu bar' at the very bottom of the interface.

Name	Last Modified
chatbot	yesterday
composer	yesterday
docker_router	yesterday
frontend	yesterday
imgs	yesterday
llm_client	yesterday
slides	yesterday
00_jupyterlab.ipynb	yesterday
01_microservices.ipynb	yesterday
02_llms.ipynb	yesterday
99_table_of_contents.ipynb	yesterday

Notebook 0: JupyterLab

We use [JupyterLab](#) to manage our environment for this hands-on lab. The [JupyterLab Interface](#) is a dashboard that provides access to interactive iPython notebooks, as well as the folder structure of our environment and a terminal window into the Ubuntu operating system. The view includes a **menu bar** at the top, a **file browser** in the **left sidebar**, and a **main work area** initially open to the "Launcher" page.

>

Menu bar

Simple 0 \$ 3 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 00_jupyterlab.ipynb 1

Building RAG Agents with LLMs

- Environment and LLMs
 - Environment [0, 1, 2]
 - Part 1: Course Environment
 - Part 2: LLM Services
- LangChain
 - Environment [3, 4]
 - Part 3: LangChain
 - Part 4: Running States
- Documents and Embeddings
- Retrieval-Augmented Generation
- Next Steps
- Feedback

DEEP LEARNING INSTITUTE

This Lab 0 : 10 : 32 / 2 : 00 : 00 STOP TASK

Course 13 : 55 : 17 / 32 : 00 : 00 LAUNCH

Stopping Task

Click confirm to stop this task.

Building RAG Agents with LLMs

Building RAG Agents with LLMs

Introduction

Environment and LLMs

LangChain

Documents and Embeddings

Environment [5, 6]

Part 5: Documents

Part 6: Embeddings

Retrieval-Augmented Generation

Environment [7, 8, Assessment]

Part 7: Vector Stores

Part 8: Evaluation

Next Steps

Previous

Next

[Bookmark this page](#)

In this section, we will combine all of our prior efforts to integrate and evaluate retrieval-augmented generation pipelines! Along the way, you will also get the opportunity to work through the assessment, which will involve Gradio, LangServe, FAISS, RAG, and Evaluation! **Good Luck!**

Please click the "Start" button to start up your own private server for hands-on coding practice. It will take a few minutes to start up, so go ahead and click it now and then proceed to the next video! After a few minutes when the server has loaded, click "Launch" to access the code labs.



This Lab 0 : 15 : 39 / 4 : 00 : 00



Course 14 : 12 : 18 / 32 : 00 : 00

LAUNCH

STOP TASK

ASSESS TASK

Building RAG Agents with LLMs

← → ↻ Not Secure 34.227.20.149/lab/lab/tree/08_evaluation.ipynb ☆ 📄 🔍 ⋮

File Edit View Run Kernel Tabs Settings Help

Filter files by name 🔍

Name	Last Modified
chatbot	yesterday
composer	yesterday
docker_router	yesterday
frontend	yesterday
imgs	yesterday
llm_client	yesterday
slides	yesterday
solutions	yesterday
00_jupyterlab.ipynb	yesterday
01_microservices.ipynb	yesterday
02_llms.ipynb	yesterday
03_langchain_intro.ipynb	yesterday
04_running_state.ipynb	yesterday
05_documents.ipynb	yesterday
06_embeddings.ipynb	yesterday
07_vectorstores.ipynb	yesterday
08_evaluation.ipynb	yesterday
09_langserve.ipynb	yesterday
64_guardrails.ipynb	yesterday
99_table_of_contents.ipynb	yesterday

DEEP LEARNING INSTITUTE

Notebook 8 [Assessment]: RAG Evaluation

Welcome to the last notebook of the course! In the previous notebook, you integrated a vector store solution into a RAG pipeline! In this notebook, you will take that same pipeline and evaluate it using numerical RAG evaluation techniques incorporating LLM-as-a-Judge metrics!

Learning Objectives:

- Learn how to integrate the techniques from prior notebooks to numerically approximate the goodness of your RAG pipeline.

Simple 0 \$ 12 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 08_evaluation.ipynb 1 🔔

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



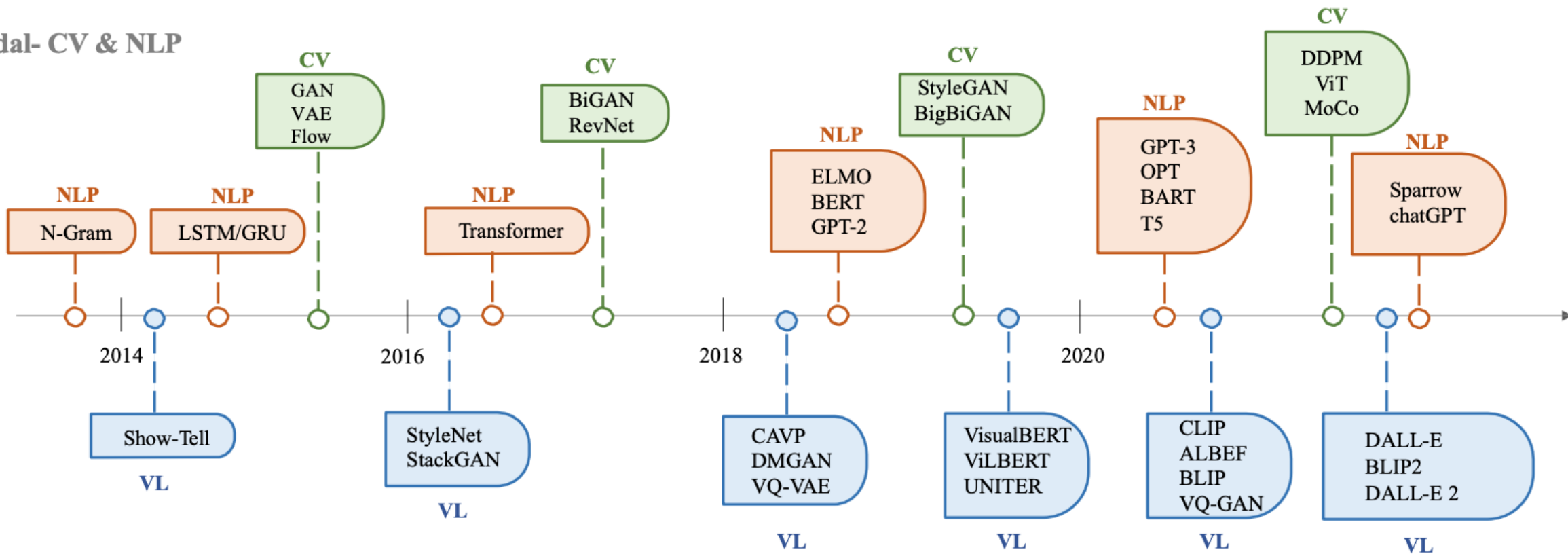
Figure 2



 **OpenAI DALL·E 2**

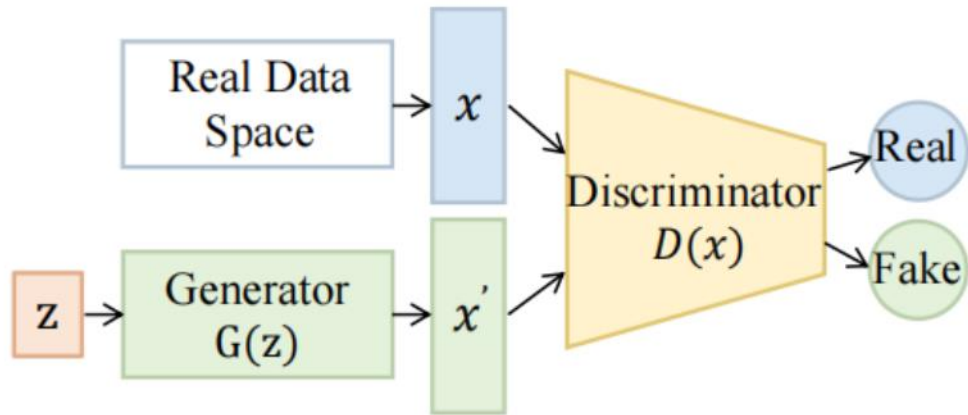
The history of Generative AI in CV, NLP and VL

Unimodal- CV & NLP

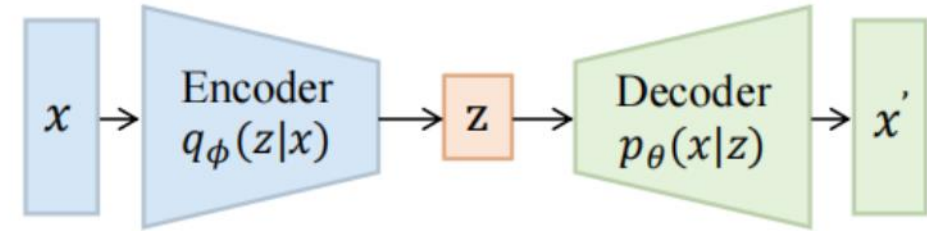


Multimodal – Vision Language

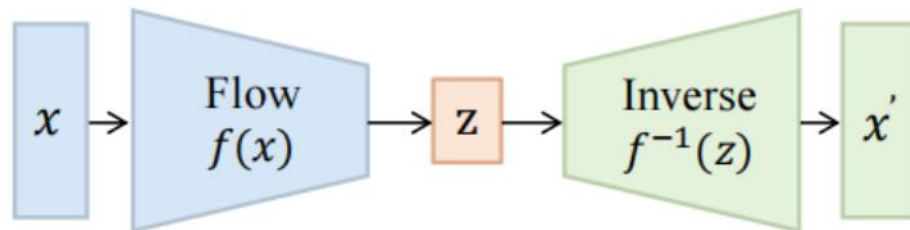
Categories of Vision Generative Models



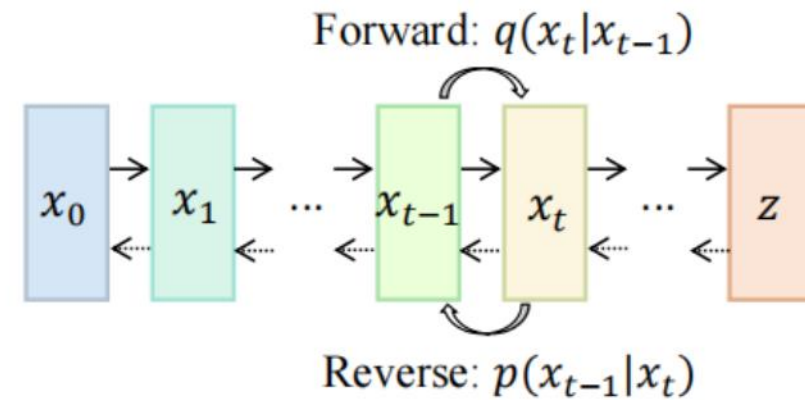
(1) Generative adversarial networks



(2) Variational autoencoders

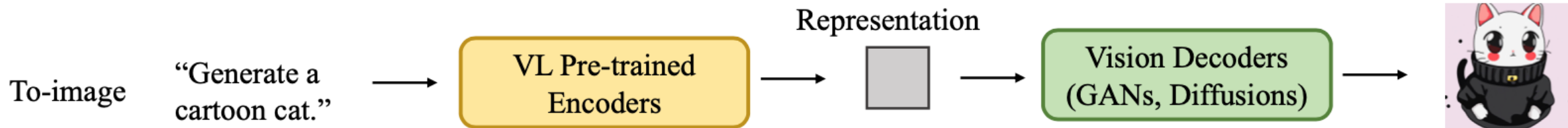
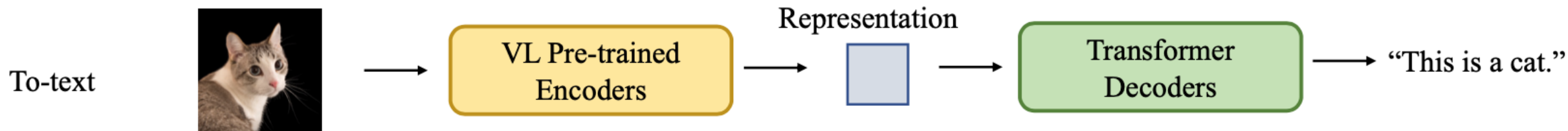
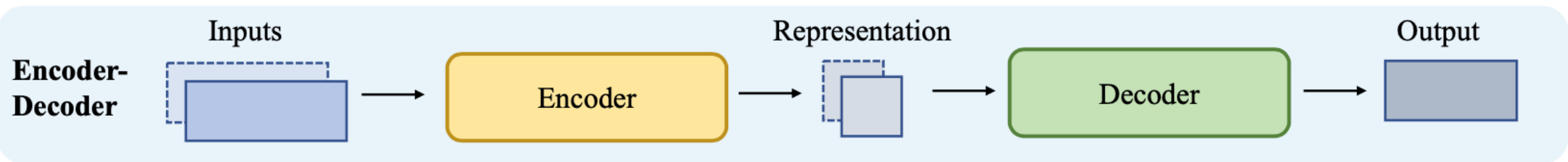


(3) Normalizing flows

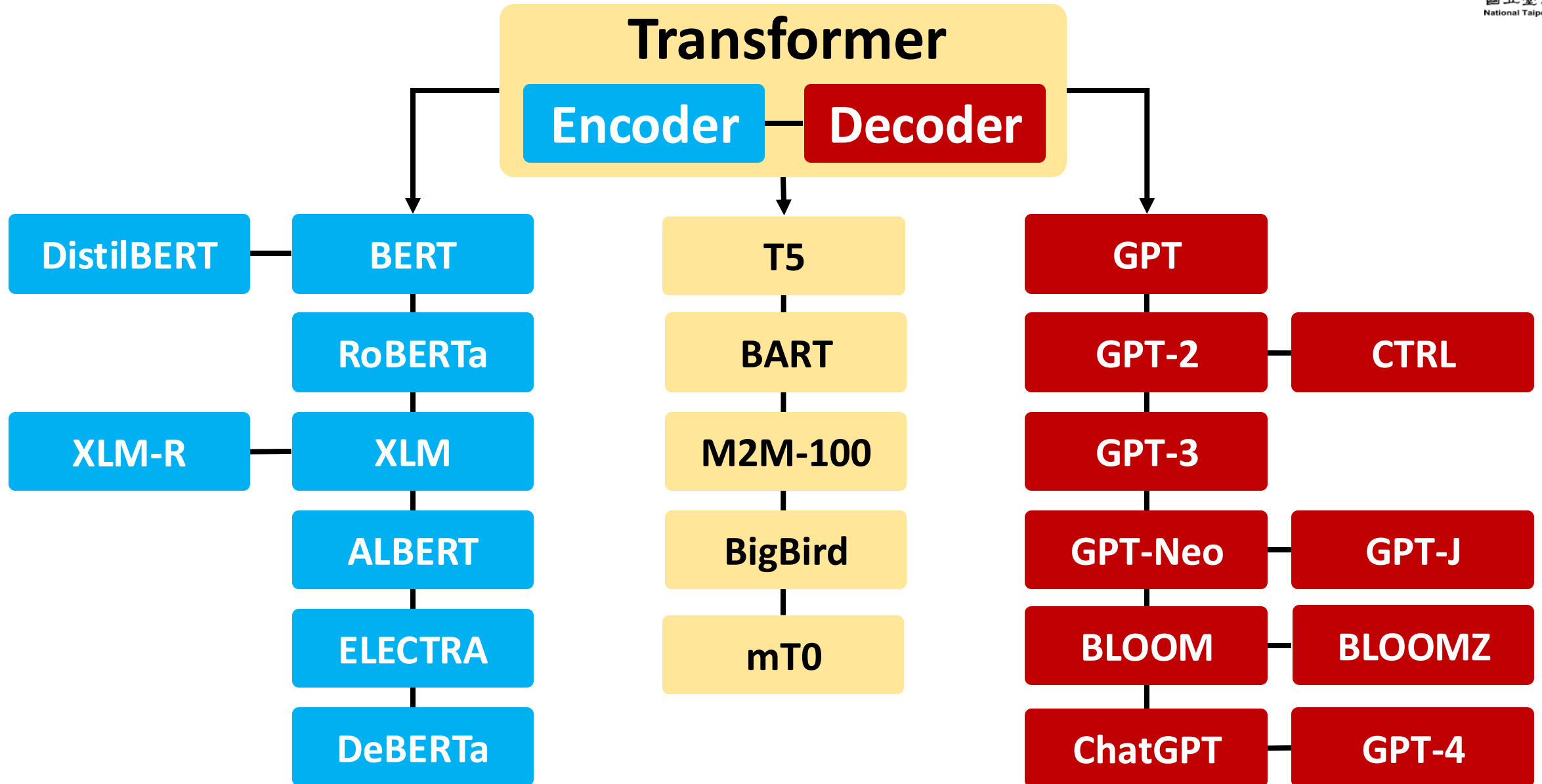


(4) Diffusion models

The General Structure of Generative Vision Language

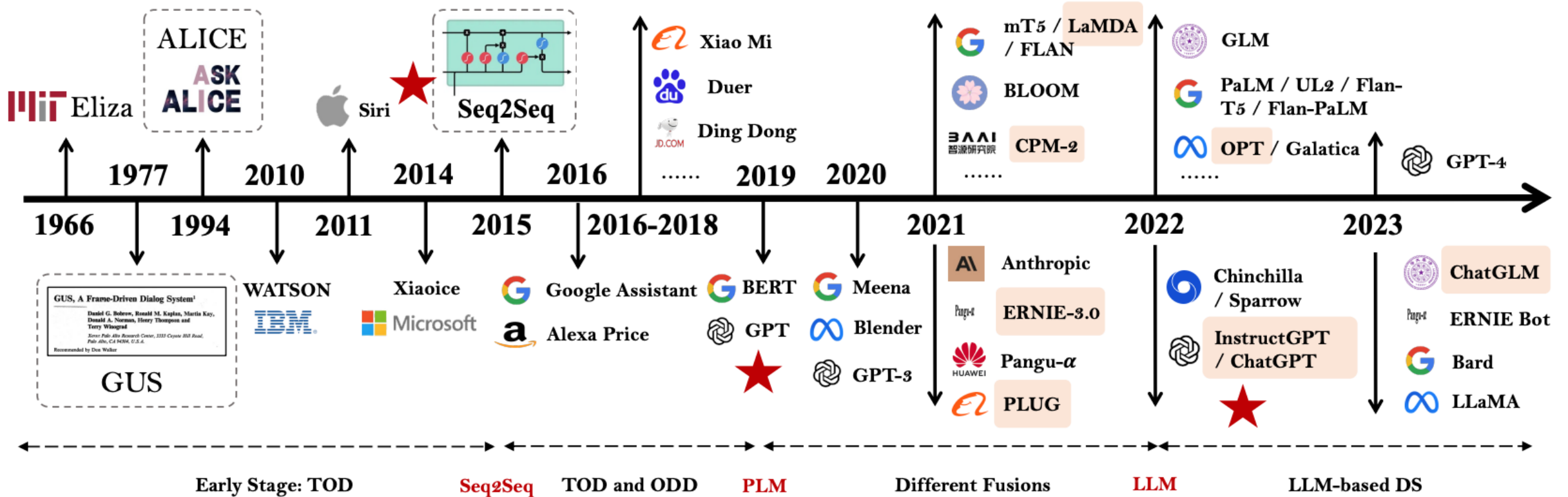


Transformer Models



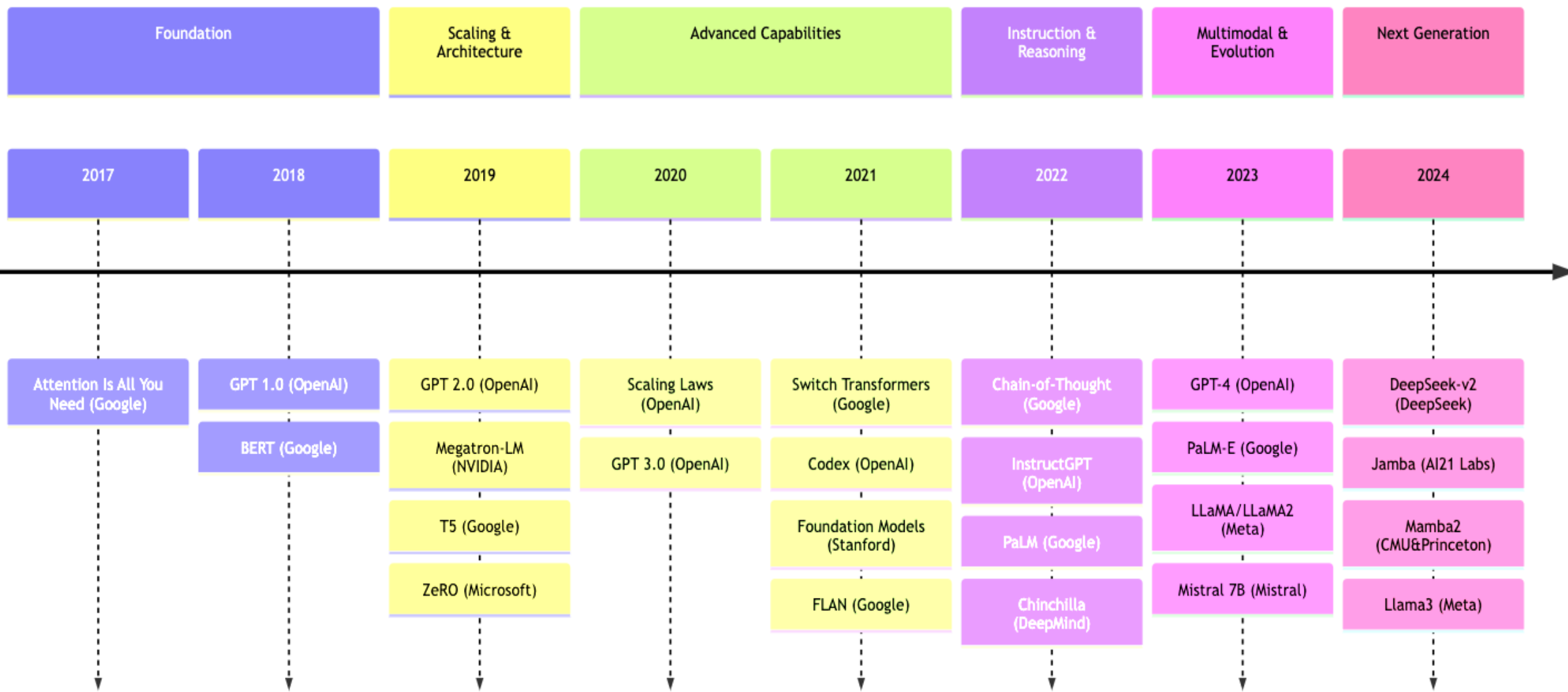
The Development of LM-based Dialogue Systems

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

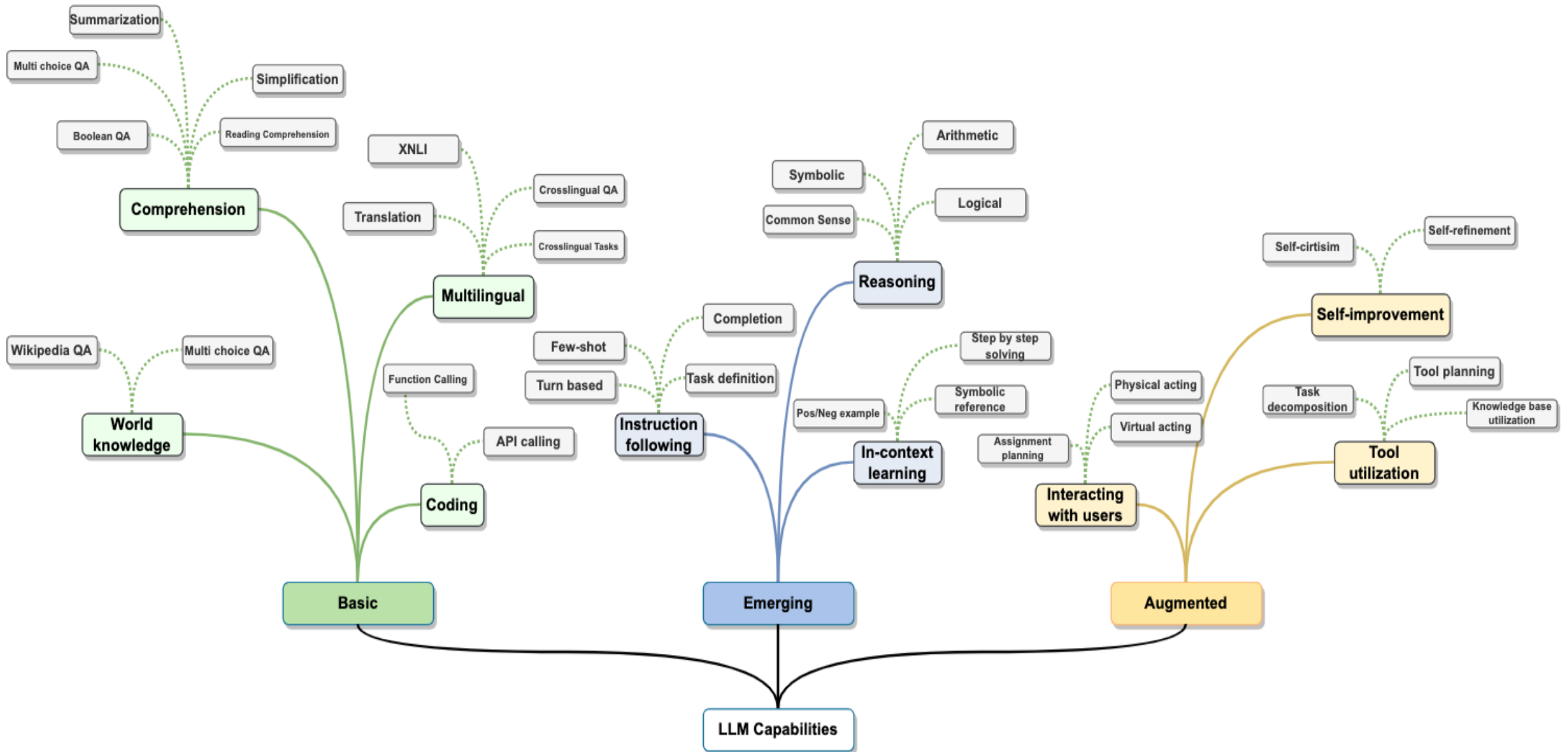


Task-oriented DS (TOD), Open-domain DS (ODD)

Major GenAI LLMs Research Milestones (2017-2024)



LLM Capabilities

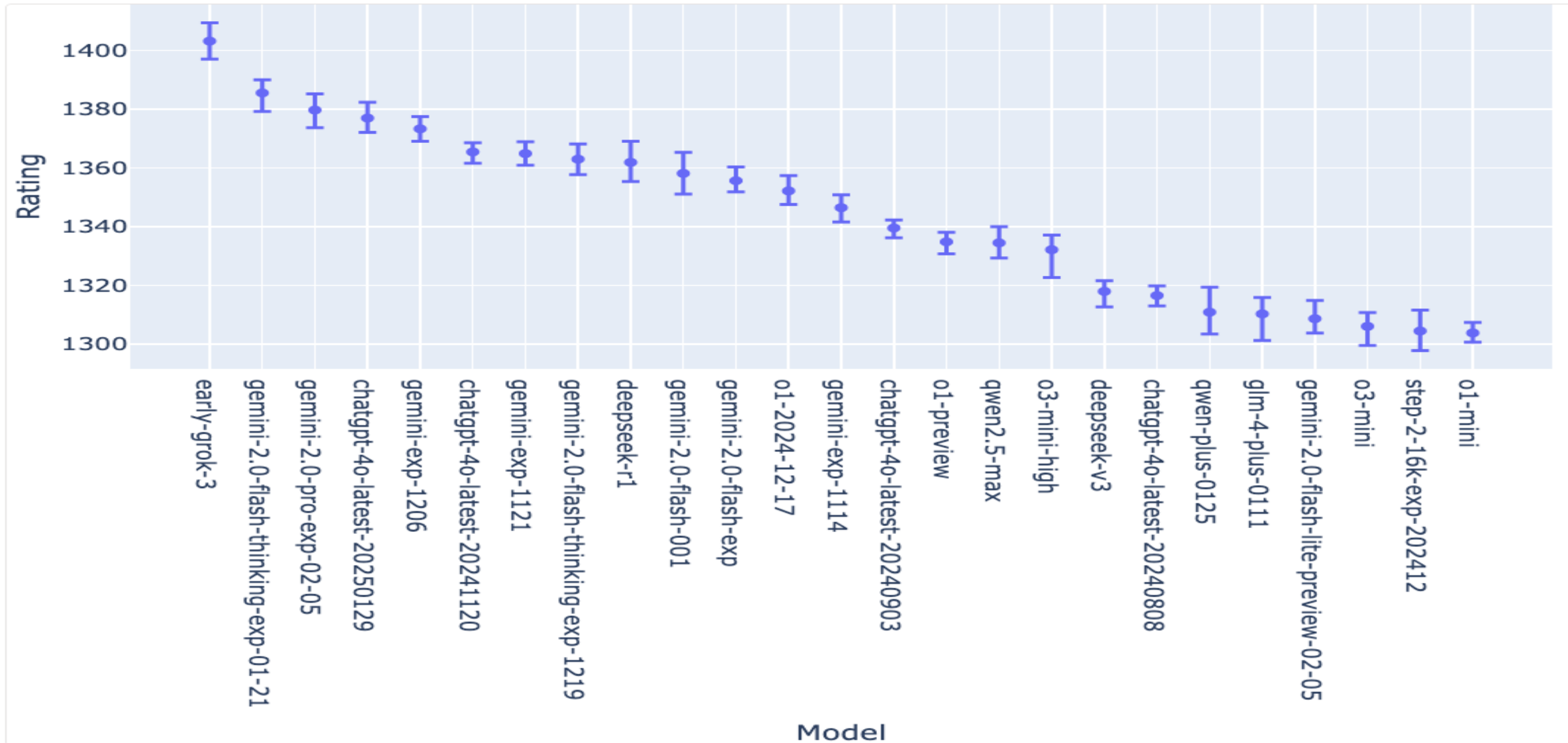


lmarena.ai Chatbot Arena Leaderboard

Rank* (UB)	Rank (StyleCtrl)	Model	Arena Score	95% CI	Votes	Organization	License
1	1	chocolate (Early Grok-3)	1403	+6/-6	9992	xAI	Proprietary
2	3	Gemini-2.0-Flash-Thinking-Exp-01-21	1385	+4/-6	15083	Google	Proprietary
2	3	Gemini-2.0-Pro-Exp-02-05	1380	+5/-6	13000	Google	Proprietary
2	1	ChatGPT-4o-latest (2025-01-29)	1377	+5/-5	13470	OpenAI	Proprietary
5	3	DeepSeek-R1	1362	+7/-7	6581	DeepSeek	MIT
5	8	Gemini-2.0-Flash-001	1358	+7/-7	10862	Google	Proprietary
5	3	o1-2024-12-17	1352	+5/-5	17248	OpenAI	Proprietary
8	7	o1-preview	1335	+3/-4	33169	OpenAI	Proprietary
8	8	Qwen2.5-Max	1334	+5/-5	9282	Alibaba	Proprietary
8	7	o3-mini-high	1332	+5/-9	5954	OpenAI	Proprietary
11	11	DeepSeek-V3	1318	+4/-5	19461	DeepSeek	DeepSeek
11	13	Qwen-Plus-0125	1311	+9/-7	5112	Alibaba	Proprietary
11	14	GLM-4-Plus-0111	1310	+6/-9	5134	Zhipu	Proprietary
11	13	Gemini-2.0-Flash-Lite-Preview-02-05	1309	+6/-5	10262	Google	Proprietary
12	12	o3-mini	1306	+5/-6	12179	OpenAI	Proprietary

Imarena.ai Chatbot Arena Leaderboard

Confidence Intervals on Model Strength (via Bootstrapping)



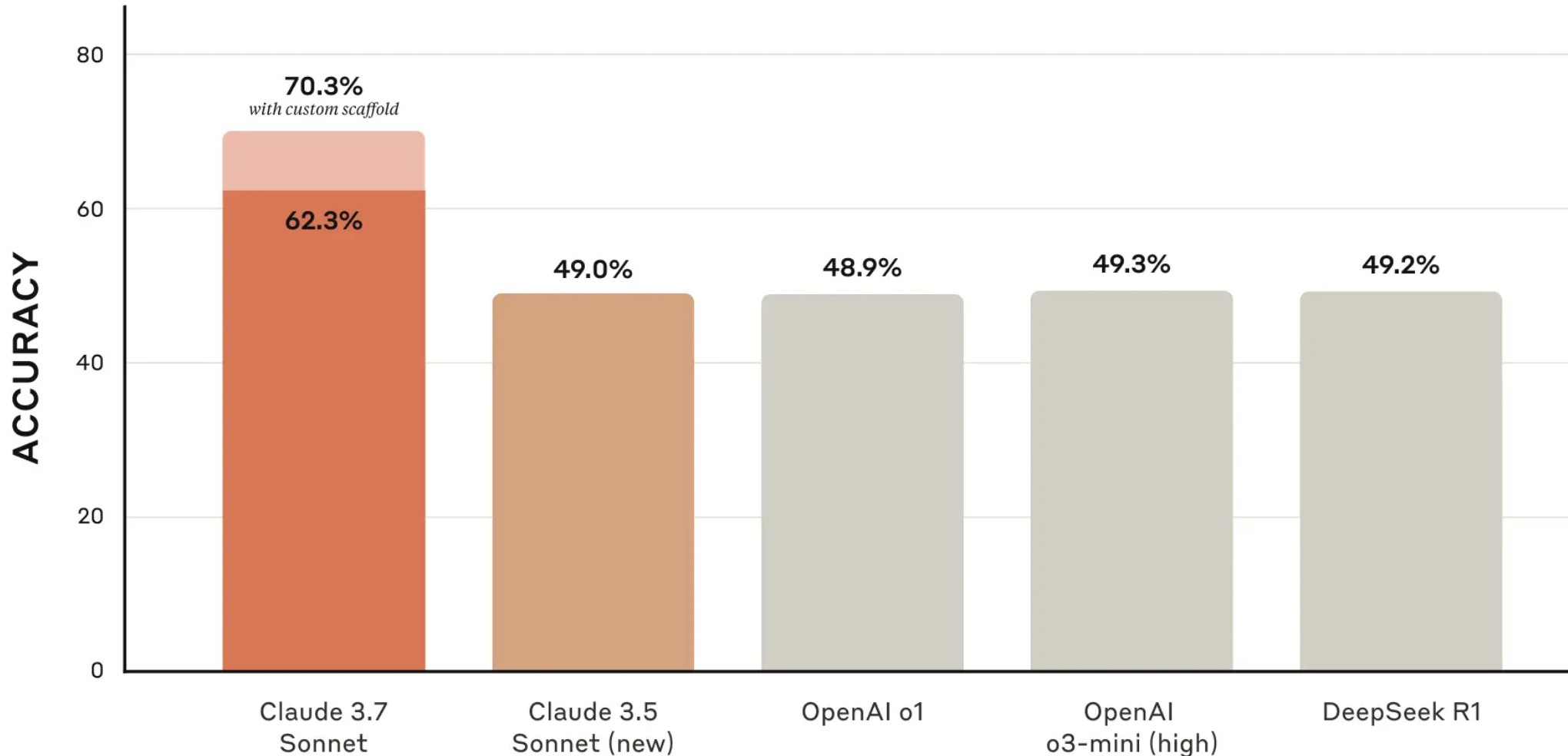
Claude 3.7 Sonnet, Claude 3.5 Sonnet, OpenAI, DeepSeek, and Grok

	Claude 3.7 Sonnet <i>64K extended thinking</i>	Claude 3.7 Sonnet <i>No extended thinking</i>	Claude 3.5 Sonnet <i>(new)</i>	OpenAI o1 ¹	OpenAI o3-mini ¹ <i>High</i>	DeepSeek R1 <i>32K extended thinking</i>	Grok 3 Beta <i>Extended thinking</i>
Graduate-level reasoning <i>GPQA Diamond³</i>	78.2% / 84.8%	68.0%	65.0%	75.7% / 78.0%	79.7%	71.5%	80.2% / 84.6%
Agentic coding <i>SWE-bench Verified²</i>	—	62.3% / 70.3%	49.0%	48.9%	49.3%	49.2%	—
Agentic tool use <i>TAU-bench</i>	—	Retail 81.2%	Retail 71.5%	Retail 73.5%	—	—	—
	—	Airline 58.4%	Airline 48.8%	Airline 54.2%	—	—	—
Multilingual Q&A <i>MMMLU</i>	86.1%	83.2%	82.1%	87.7%	79.5%	—	—
Visual reasoning <i>MMMU (validation)</i>	75%	71.8%	70.4%	78.2%	—	—	76.0% / 78.0%
Instruction-following <i>IFEval</i>	93.2%	90.8%	90.2%	—	—	83.3%	—
Math problem-solving <i>MATH 500</i>	96.2%	82.2%	78.0%	96.4%	97.9%	97.3%	—
High school math competition <i>AIME 2024³</i>	61.3% / 80.0%	23.3%	16.0%	79.2% / 83.3%	87.3%	79.8%	83.9% / 93.3%

Claude 3.7 Sonnet and Claude Code

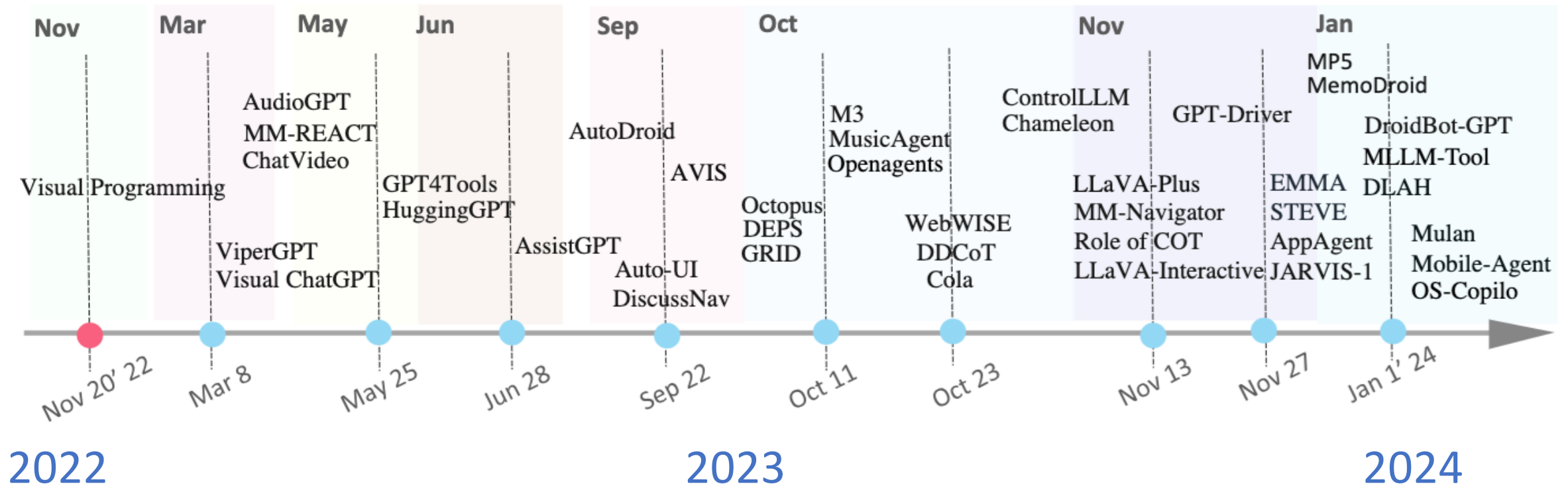
Software engineering

SWE-bench verified



LLM-powered Multimodal Agents

Large Multimodal Agents (LMAs)

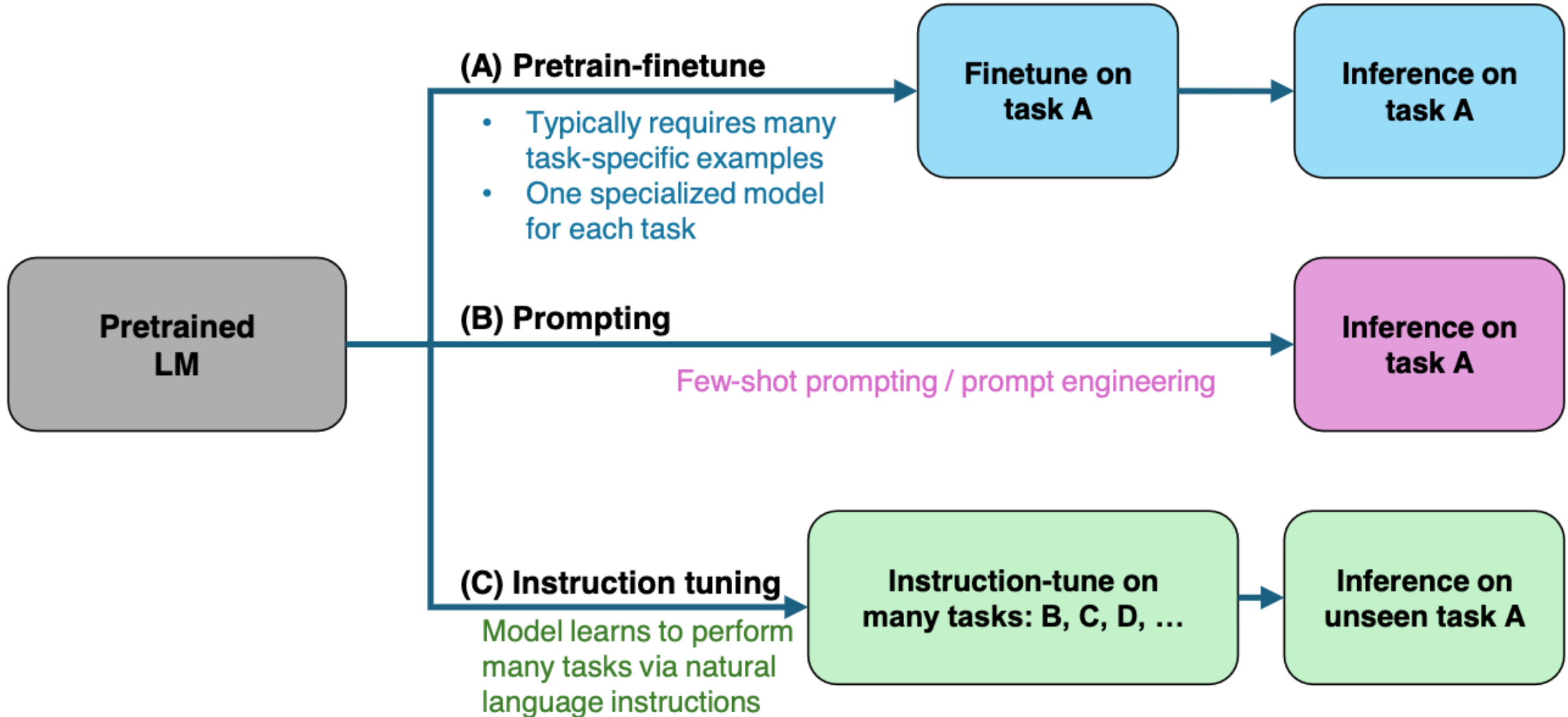


Four Paradigms in NLP (LM)

Paradigm	Engineering	Task Relation
a. Fully Supervised Learning (Non-Neural Network)	Feature (e.g. word identity, part-of-speech, sentence length)	
b. Fully Supervised Learning (Neural Network)	Architecture (e.g. convolutional, recurrent, self-attentional)	
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)	

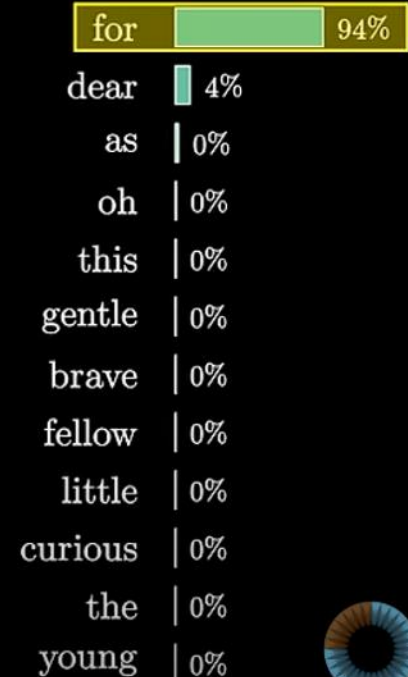
Large Language Models (LLM)

Three typical learning paradigms



Transformers (how LLMs work)

Behold, a wild pi creature, foraging in its native habitat of mathematical formulas and computer code! With its infinite digits and irrational tendencies, this strange creature is beloved by mathematicians and tech enthusiasts alike. Approach with caution, for attempting to calculate its exact value may lead to madness! But do not be afraid, **for**



Attention in Transformers

Query
1,572,864

$$\begin{bmatrix} -3.7 & +3.9 & -2.4 & -6.3 & -9.4 & -8.6 & +3.6 & -0.9 & \dots & +0.7 \\ +7.9 & +9.7 & -5.6 & +3.2 & -4.7 & -9.5 & +5.1 & -3.6 & \dots & -2.3 \\ +1.7 & +6.6 & +2.6 & +7.4 & -4.5 & +5.9 & -6.2 & +9.0 & \dots & +3.7 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -5.6 & +8.9 & +4.6 & -4.9 & -5.7 & +0.4 & -9.4 & -5.8 & \dots & -1.5 \end{bmatrix}$$

Key
1,572,864

$$\begin{bmatrix} -2.5 & -0.7 & -4.4 & +1.7 & +7.2 & -7.6 & +0.3 & -7.3 & \dots & +4.3 \\ -2.1 & +1.3 & -6.3 & -7.0 & -0.2 & -2.9 & +8.7 & +5.3 & \dots & +4.9 \\ +8.0 & -8.2 & +1.0 & +1.7 & +9.1 & -4.1 & -5.1 & -7.9 & \dots & -9.6 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ +8.5 & +3.4 & +5.6 & -4.3 & +1.7 & -8.6 & -0.3 & +9.5 & \dots & +7.5 \end{bmatrix}$$

Value
12,288

$$12,288 \times 12,288 = 150,994,944$$
$$\begin{bmatrix} -3.2 & +9.1 & -5.3 & +8.9 & +8.7 & +5.9 & +2.6 & +7.4 & \dots & -4.1 \\ +6.9 & +2.3 & -9.6 & -3.0 & -7.0 & +9.5 & -0.4 & -0.1 & \dots & +2.8 \\ -2.6 & -7.2 & +6.4 & -6.1 & +0.2 & -5.5 & -8.0 & +7.2 & \dots & +9.4 \\ +9.1 & +8.0 & +5.4 & -3.3 & -8.3 & -1.8 & -5.3 & -7.3 & \dots & -8.8 \\ +4.5 & -9.7 & +5.4 & -7.0 & -8.3 & -8.1 & +3.4 & -5.0 & \dots & -1.6 \\ +1.1 & +7.1 & +4.5 & -4.5 & -7.3 & -8.8 & -3.9 & -4.7 & \dots & -0.9 \\ +3.6 & +3.9 & -4.3 & -2.4 & -6.3 & +5.7 & -8.8 & +3.9 & \dots & +5.5 \\ +5.5 & -4.8 & -2.5 & +1.7 & -4.5 & -2.6 & -6.0 & -0.8 & \dots & -9.0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ +5.9 & -8.4 & +0.4 & -3.8 & +1.5 & +9.1 & +2.9 & -9.2 & \dots & -1.4 \end{bmatrix} \begin{bmatrix} +0.2 \\ +0.7 \\ +3.6 \\ -4.4 \\ -7.3 \\ -2.1 \\ +9.0 \\ -6.2 \\ \vdots \\ +0.9 \end{bmatrix} = \begin{bmatrix} -198.6 \\ +73.1 \\ -28.2 \\ +119.4 \\ -4.4 \\ +215.7 \\ +91.8 \\ -29.1 \\ -5.6 \\ \vdots \\ -5.1 \end{bmatrix}$$

16:48 / 26:09 • Counting parameters >

How might LLMs store facts

GPT-3		Total weights:
Embedding	$12,288 \times 50,257$ $d_embed * n_vocab$	$= 617,558,016$
Key	$128 \times 12,288 \times 96 \times 96$ $d_query * d_embed * n_heads * n_layers$	$= 14,495,514,624$
Query	$128 \times 12,288 \times 96 \times 96$ $d_query * d_embed * n_heads * n_layers$	$= 14,495,514,624$
Value	$128 \times 12,288 \times 96 \times 96$ $d_value * d_embed * n_heads * n_layers$	$= 14,495,514,624$
Output	$12,288 \times 128 \times 96 \times 96$ $d_embed * d_value * n_heads * n_layers$	$= 14,495,514,624$
Up-projection	$49,152 \times 12,288 \times 96$ $n_neurons * d_embed * n_layers$	$= 57,982,058,496$
Down-projection	$12,288 \times 49,152 \times 96$ $d_embed * n_neurons * n_layers$	$= 57,982,058,496$
Unembedding	$50,257 \times 12,288$ $n_vocab * d_embed$	$= 617,558,016$

Large Language Models explained briefly

What follows is a conversation between a user and a helpful, very knowledgeable AI assistant.

User: Give me some ideas for what to do when visiting Santiago.

AI Assistant: Sure,

Large Language Model

Token	Probability
,	53%
!	38%
thing	7%
.	0%
!	0%
-	0%
!	0%
	0%
I	0%
thing	0%
-	0%
,I	0%
...	0%

1:49 / 8:47 · What are large language models? >

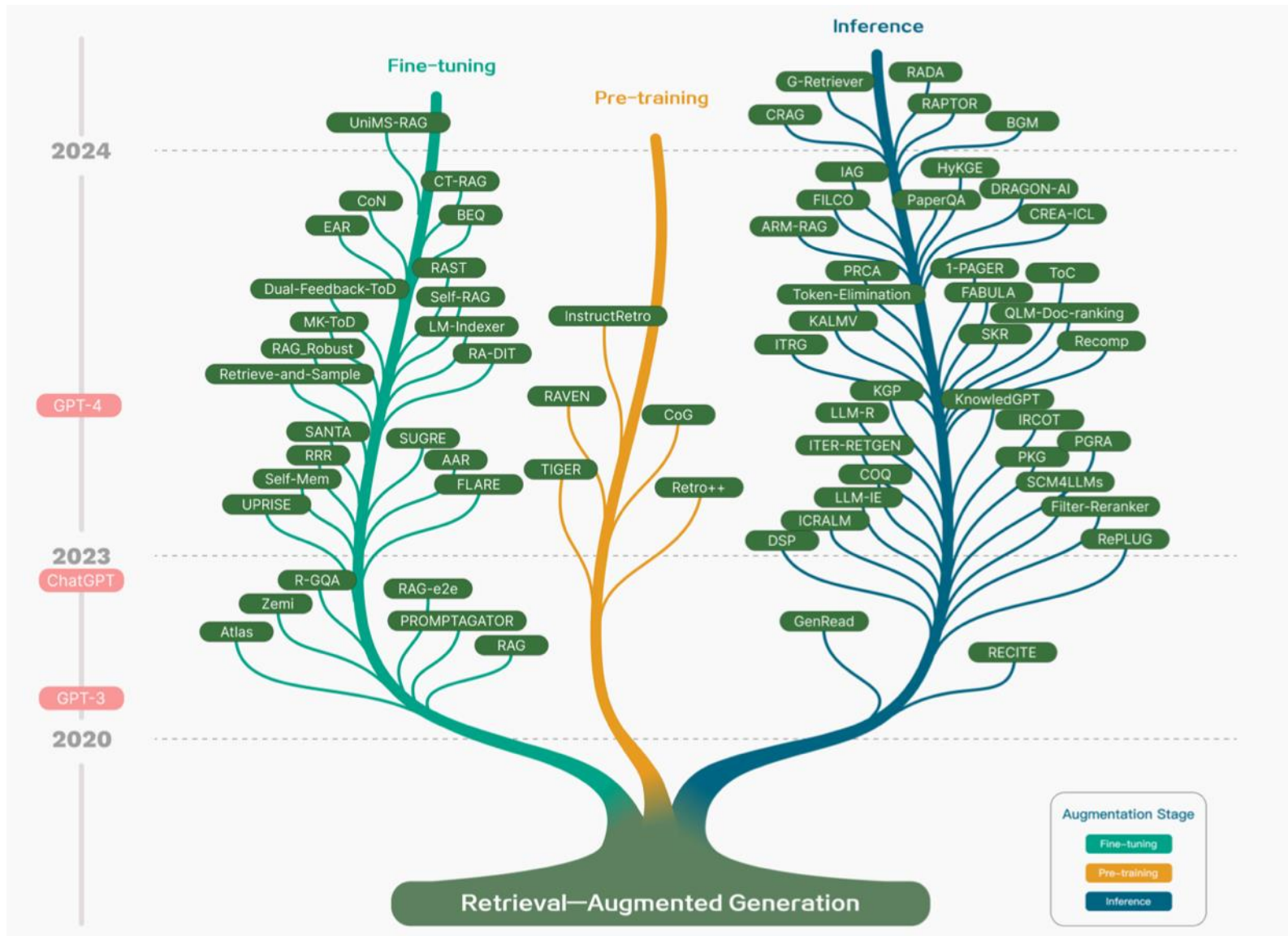
Source: 3Blue1Brown (2024), Large Language Models explained briefly, <https://www.youtube.com/watch?v=LPZh9BOjkQs>

RAG LLM

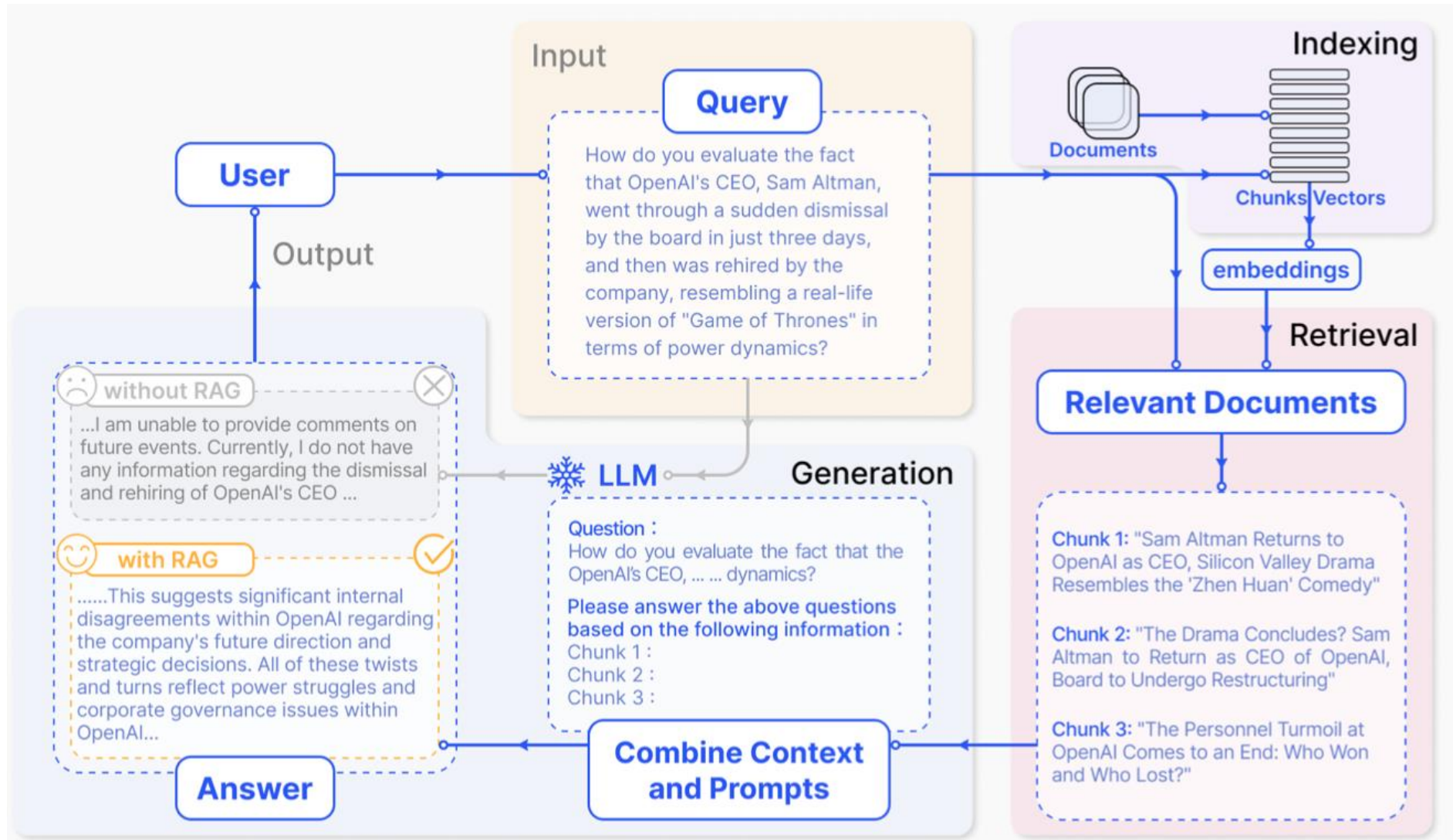
Dialogue Systems

Technology Tree of RAG Research

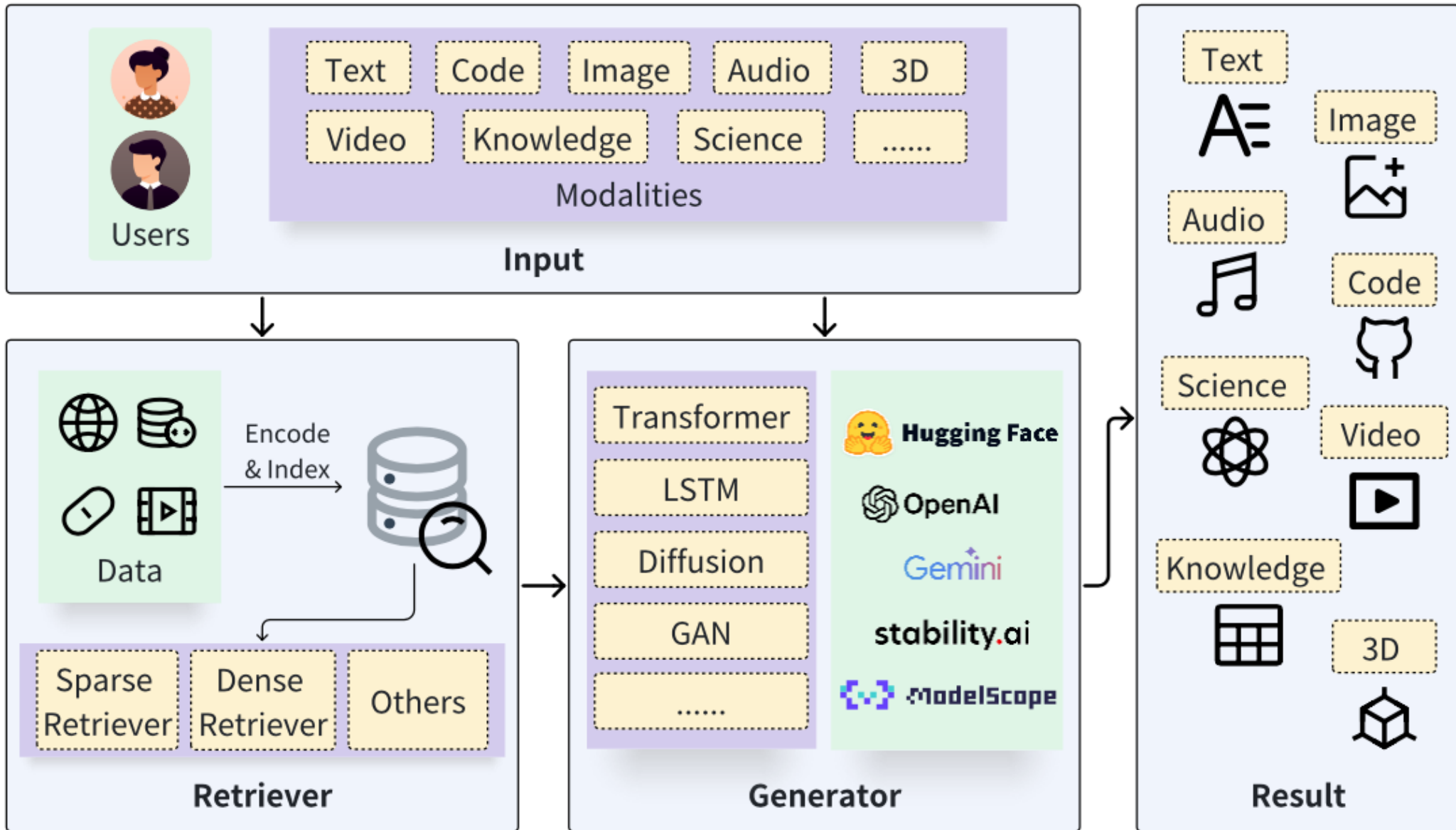
Retrieval-Augmented Generation (RAG) for Large Language Models (LLMs)



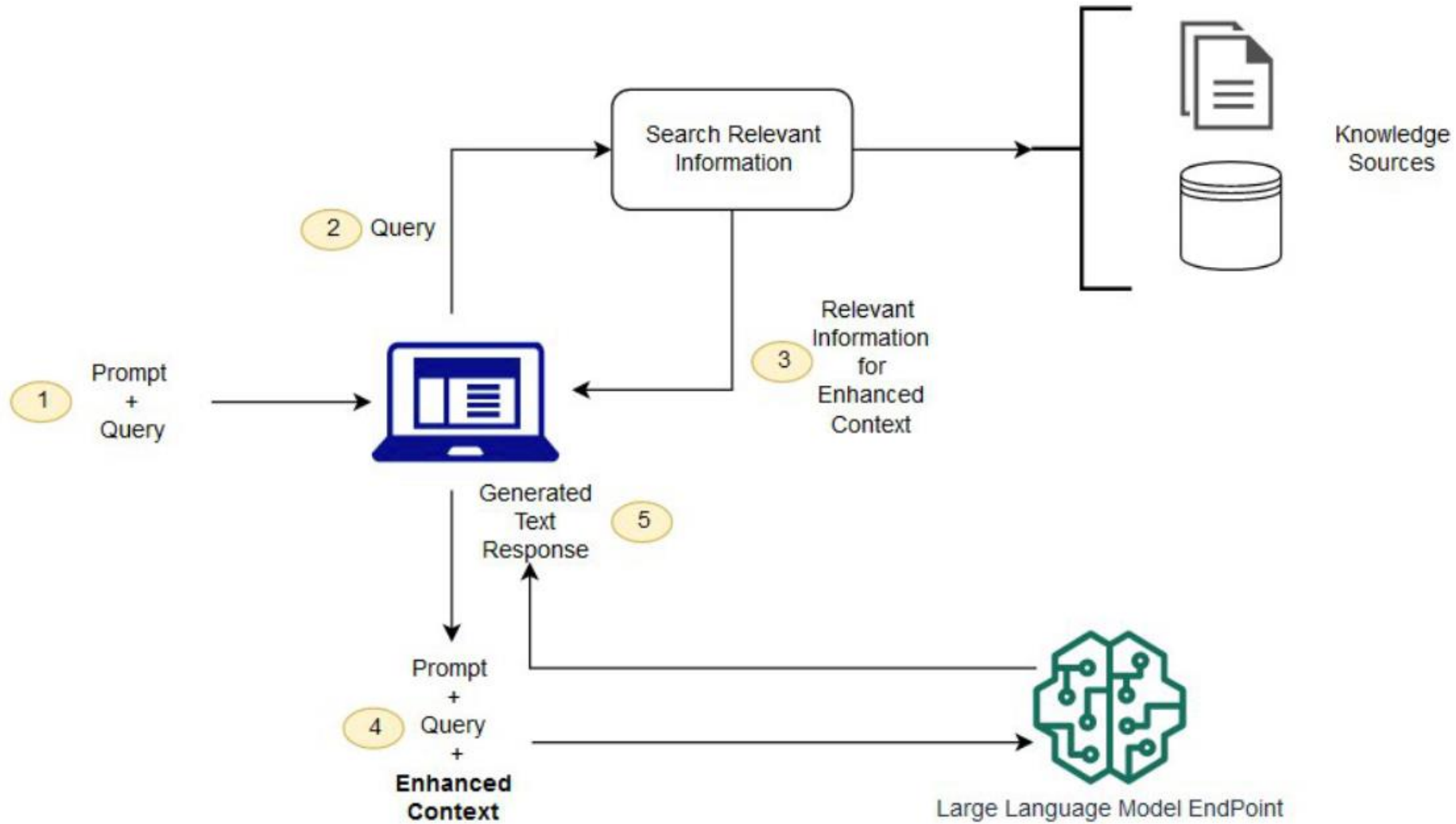
Retrieval-Augmented Generation (RAG) for Large Language Models (LLMs)



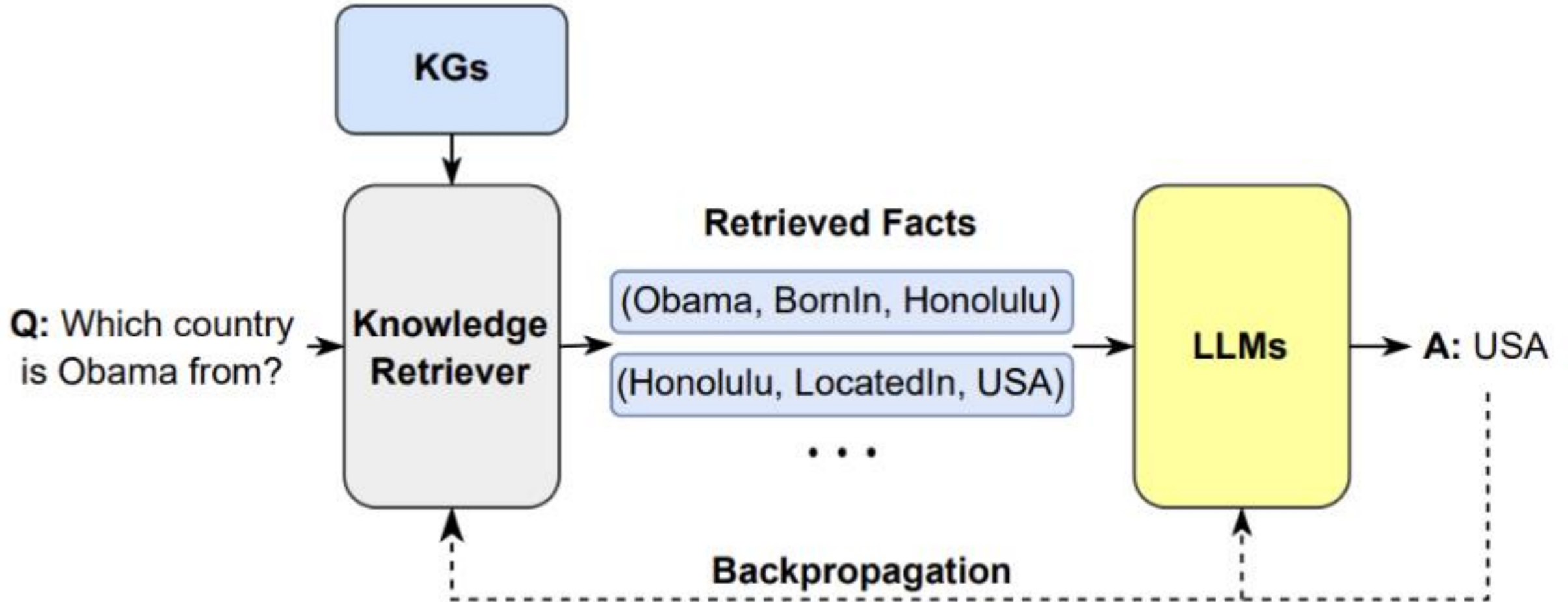
Retrieval-Augmented Generation (RAG) Architecture



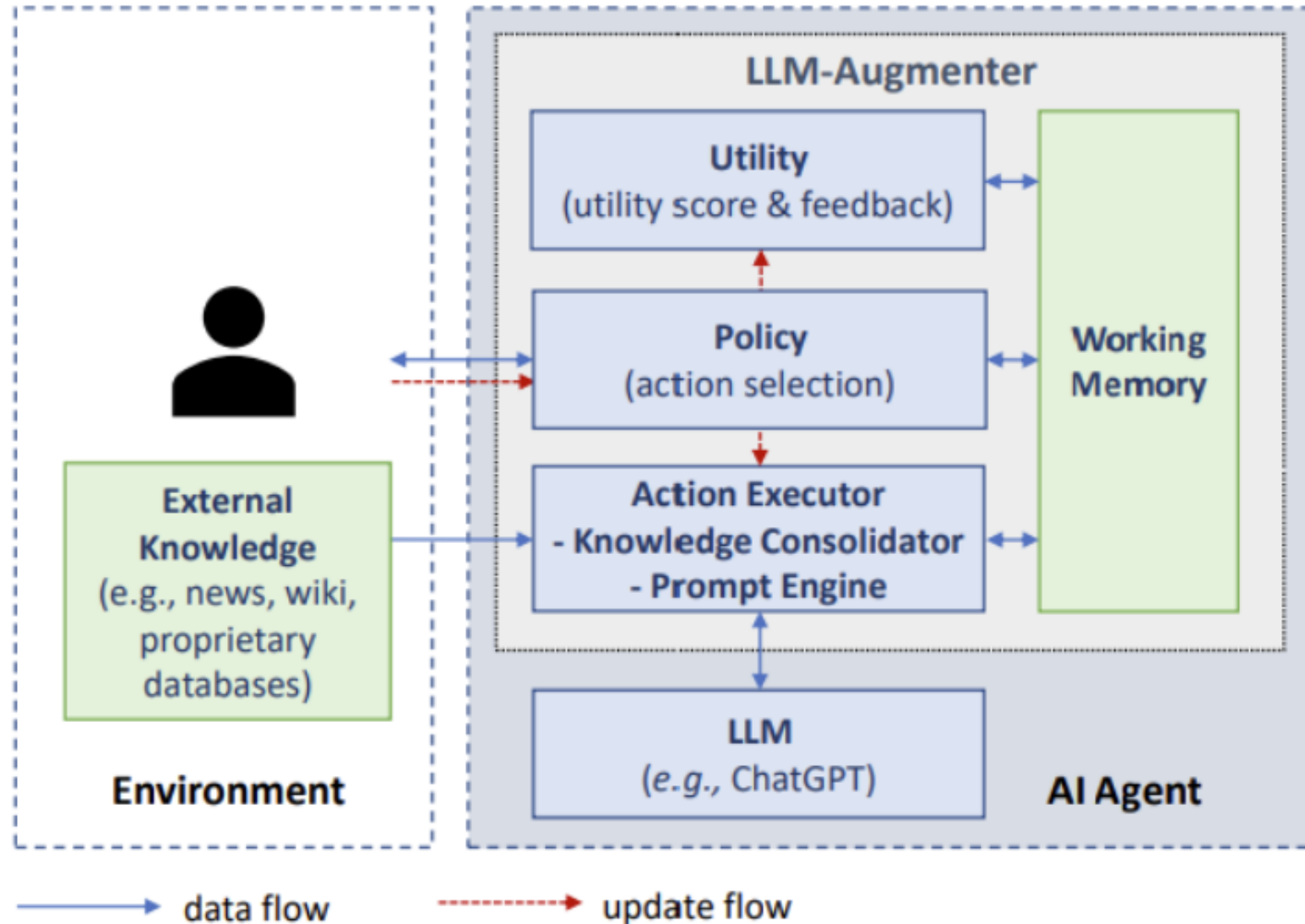
Synthesizing RAG with LLMs for Question Answering Application



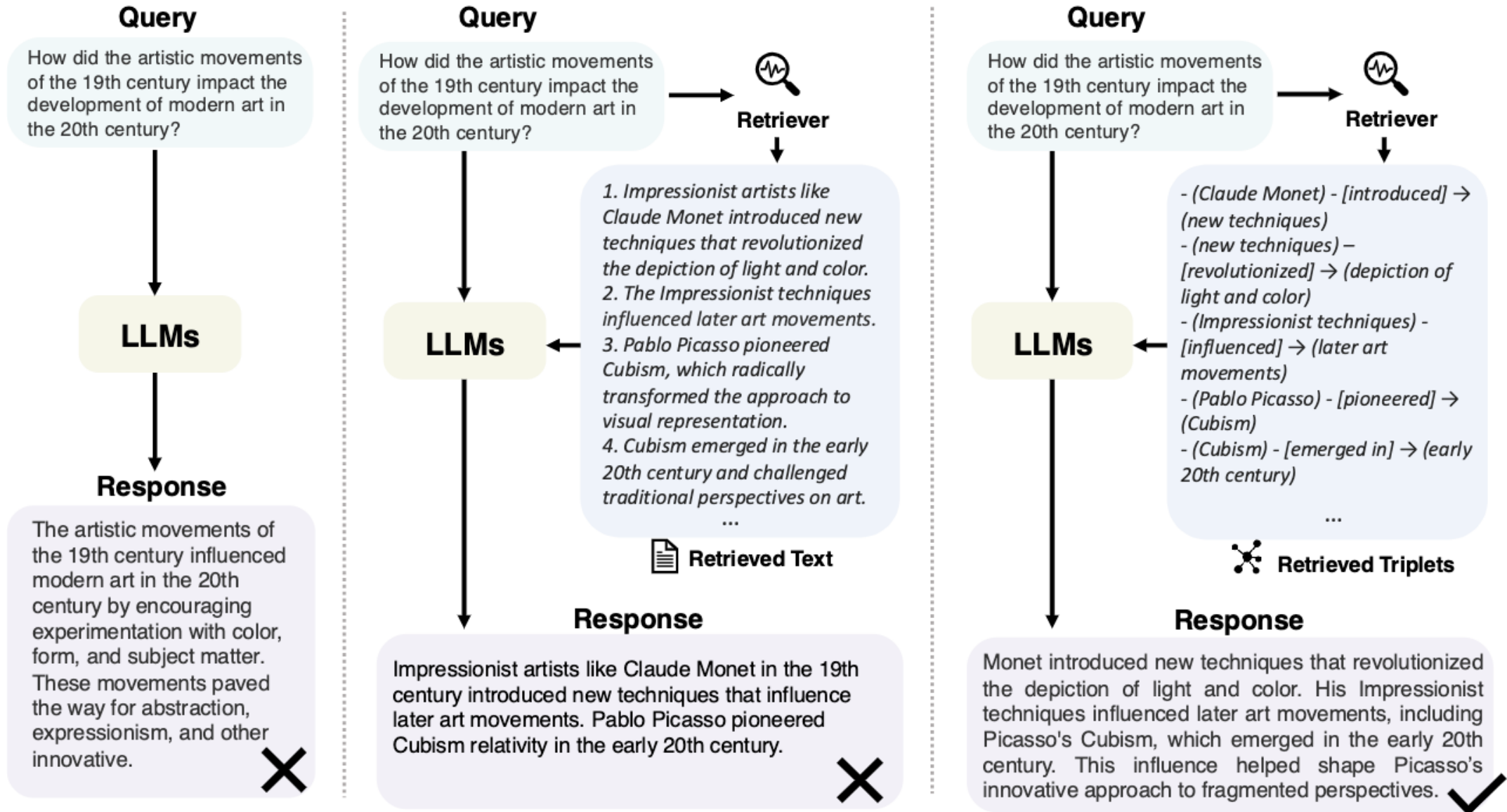
Synthesizing the KG as a Retriever with LLMs



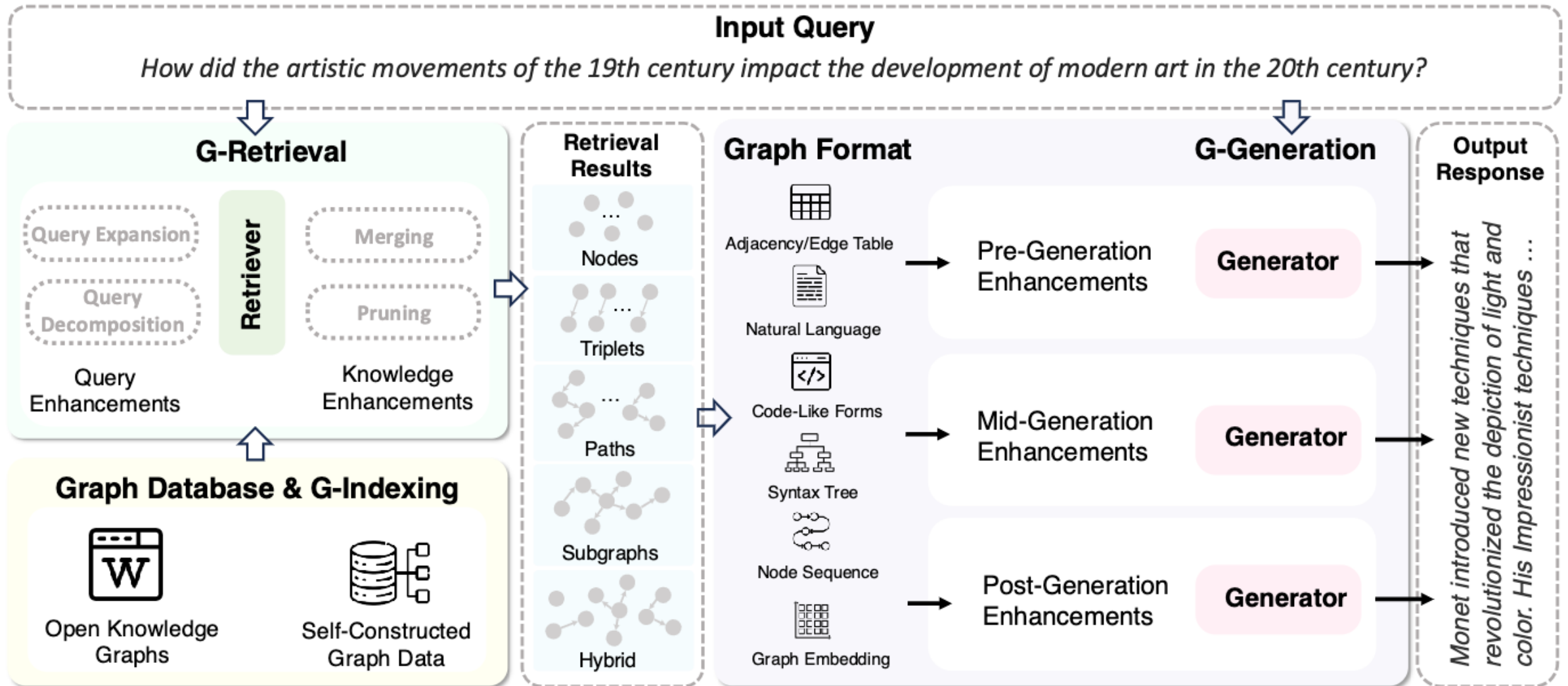
A LLM-based Agent for Conversational Information Seeking



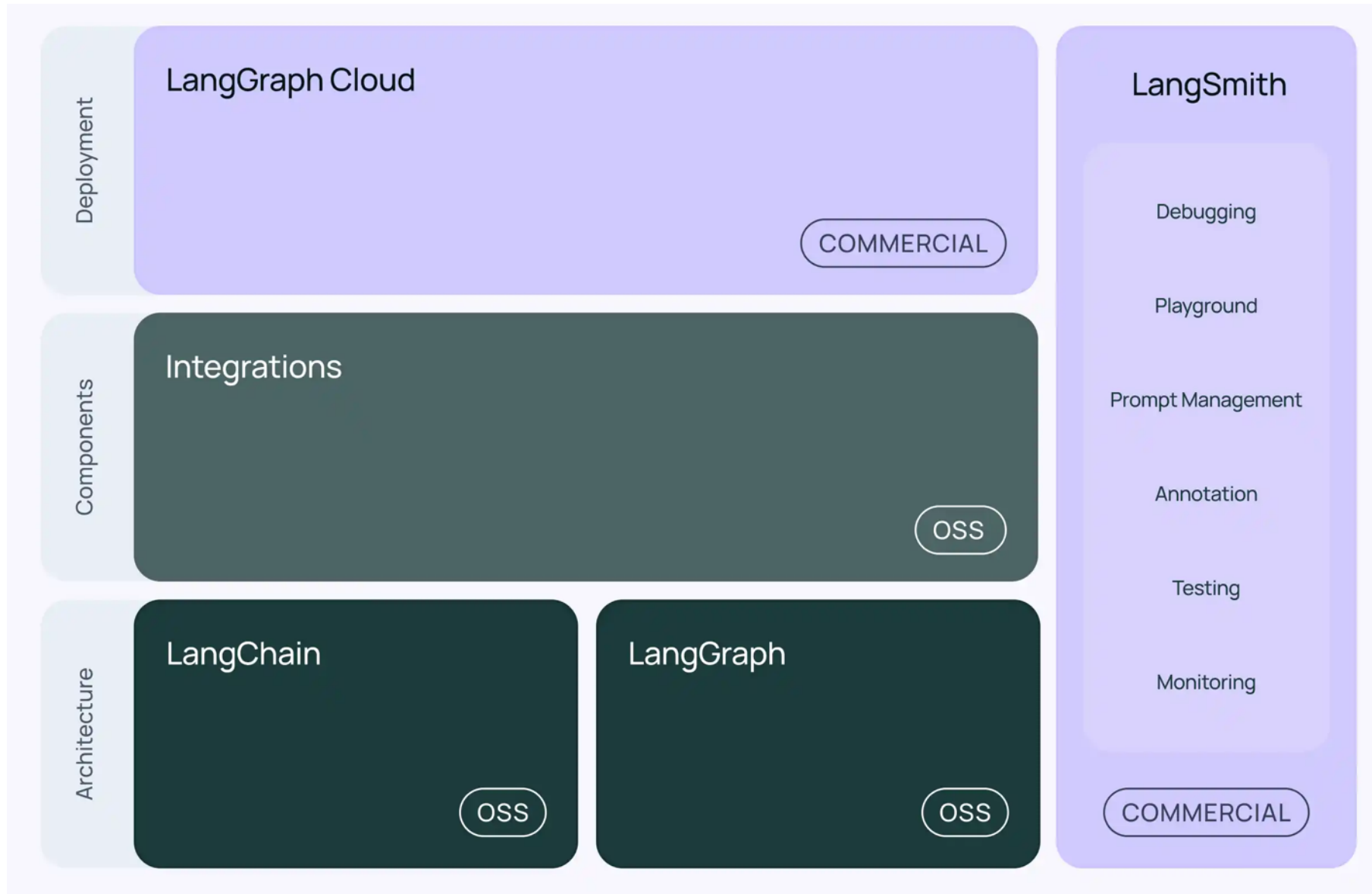
Direct LLM, RAG, and GraphRAG



GraphRAG Framework for Question Answering

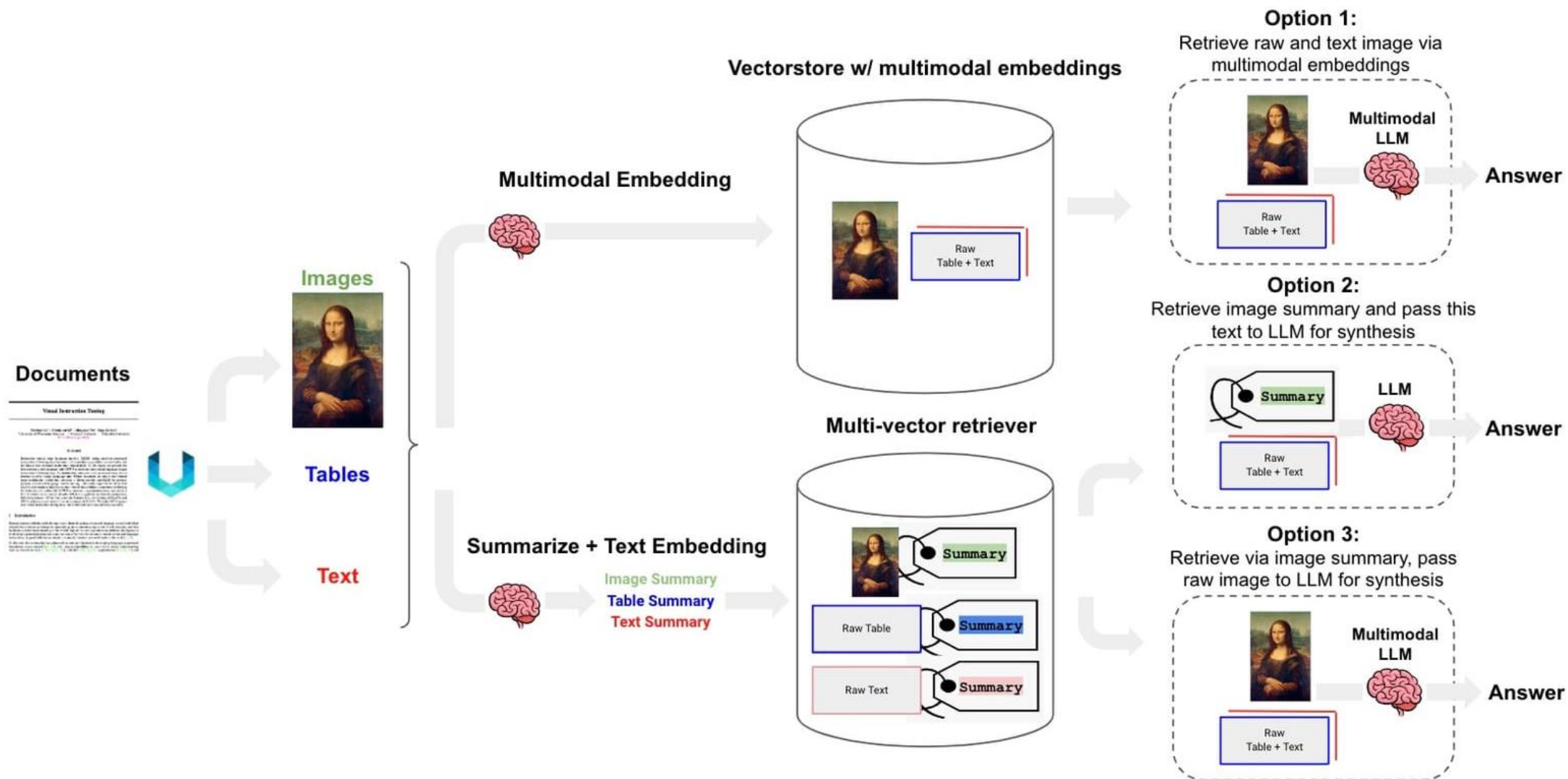


LangChain Architecture



Multimodal LLM RAG

Multi-Vector Retriever for RAG



Evaluating RAG with Ragas Metrics

ragas score

generation

faithfulness

how factually accurate is
the generated answer

answer relevancy

how relevant is the generated
answer to the question

retrieval

context precision

the signal to noise ratio of retrieved
context

context recall

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

ESG

ESG:

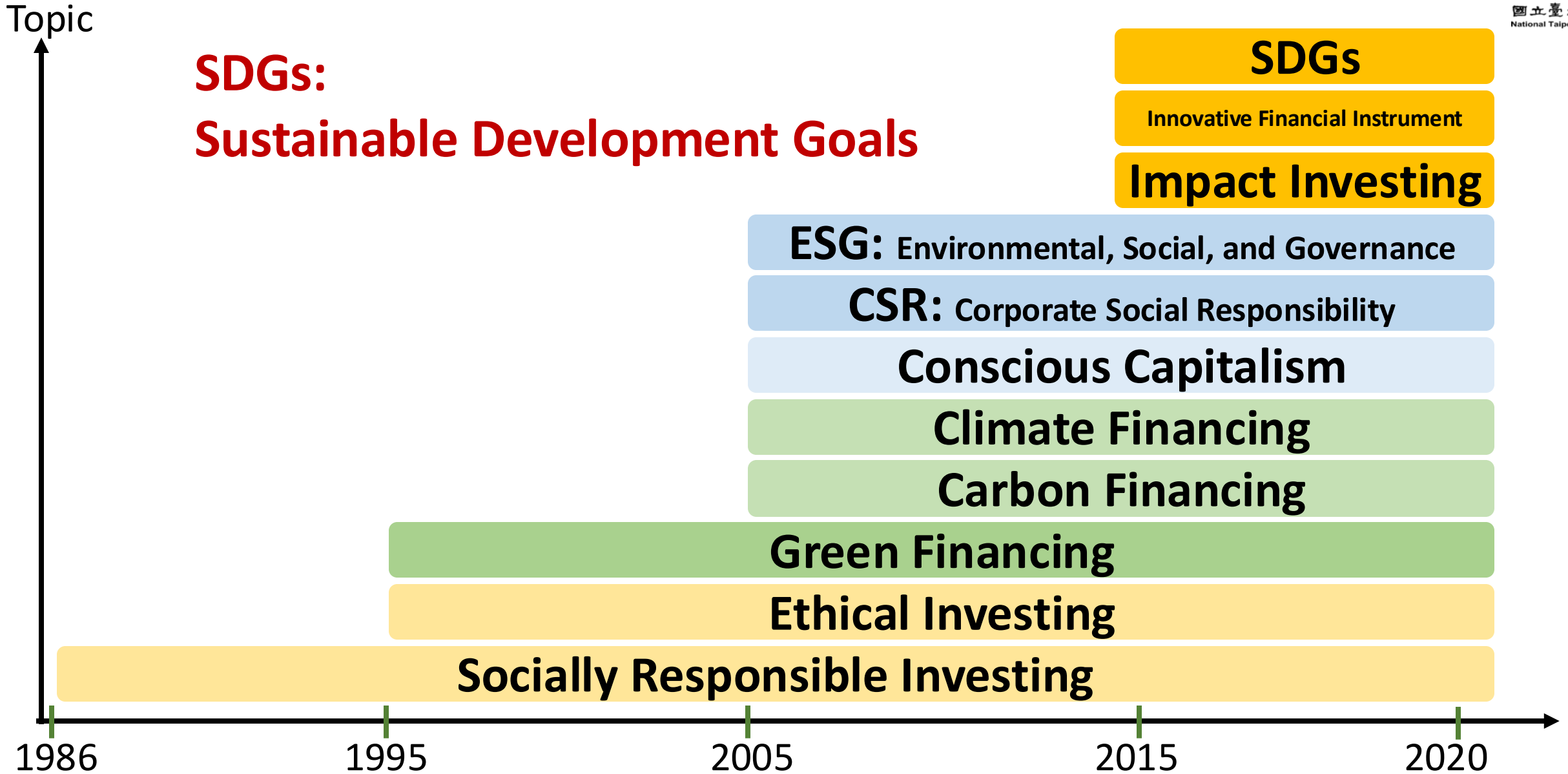
Environmental

Social

Governance

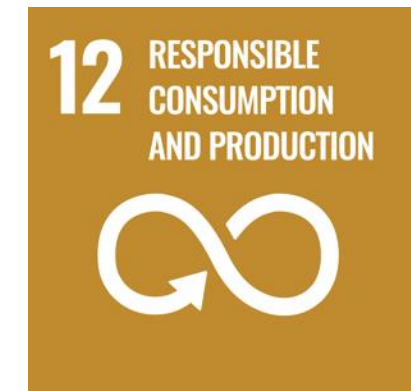
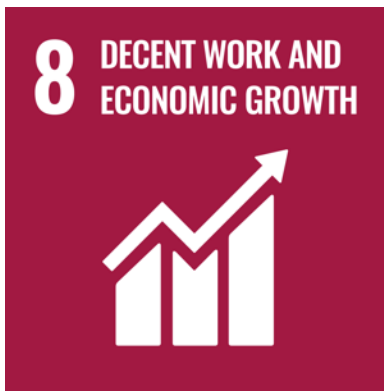
CSR: Corporate Social Responsibility

Evolution of Sustainable Finance Research



Source: Kumar, S., Sharma, D., Rao, S., Lim, W. M., & Mangla, S. K. (2022). Past, present, and future of sustainable finance: Insights from big data analytics through machine learning of scholarly research. *Annals of Operations Research*, 1-44.

Sustainable Development Goals (SDGs)



Sustainable Development Goals (SDGs)

Partnership

Peace

Prosperity

People

Planet



ESG to 17 SDGs

ENVIRONMENT



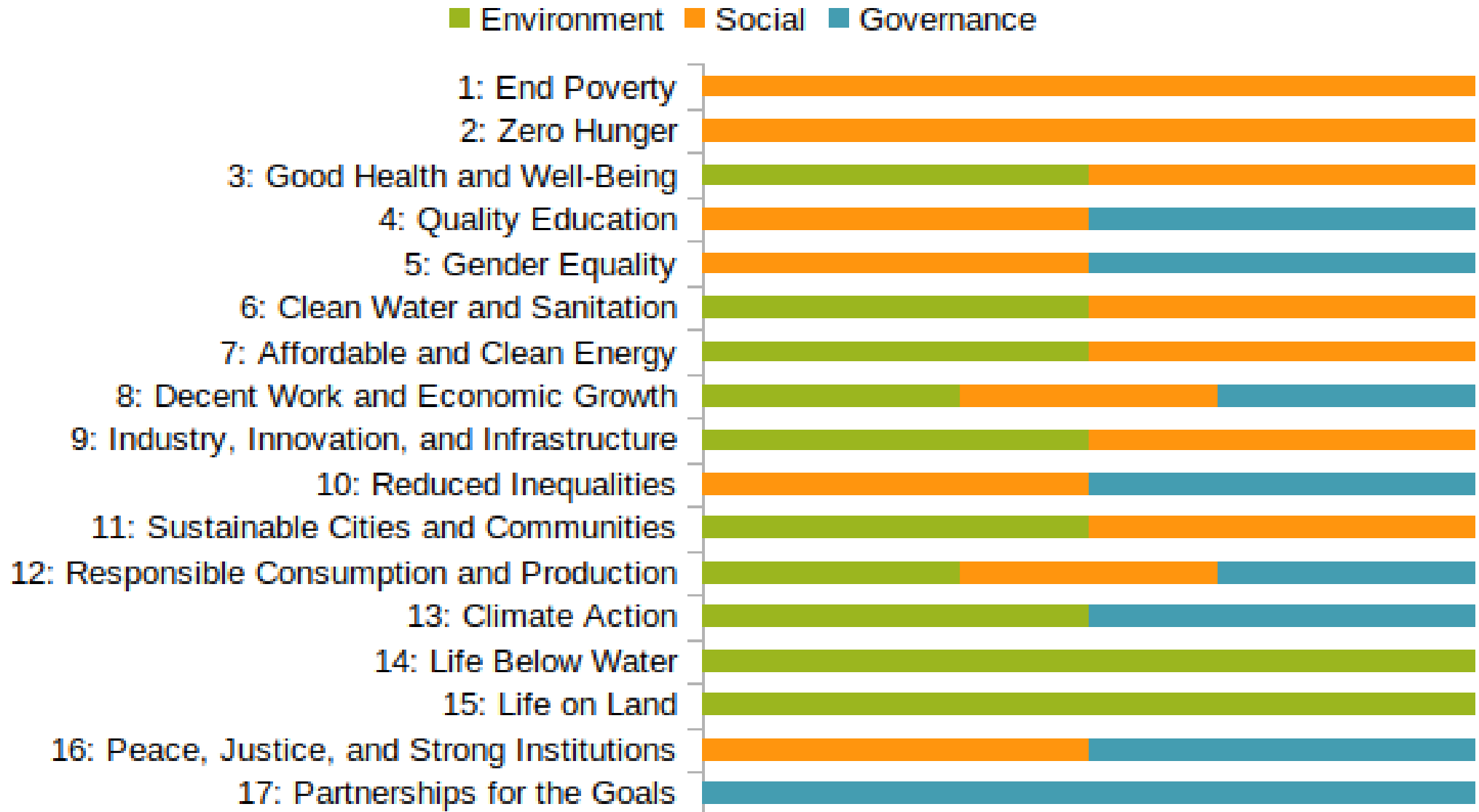
SOCIAL



GOVERNANCE



ESG to 17 SDGs



Sustainable Development

Net-Zero Transformation

- **Ambition**

- Aligned to achieving global net zero by no later than 2050 & to limit warming to 1.5° C

- **Governance**

- Accountability driven from the top

- **Strategy**

- Embedded and aligned net zero into company strategy

- **Enterprise**

- Key operating model changes in support of transformation

- **Supply chains**

- Transformed net zero supply chains

- **Innovation**

- Developed innovation and technologies to deliver net zero

- **Finance**

- Financing the net zero transformation

- **Transparency**

- Communicating action

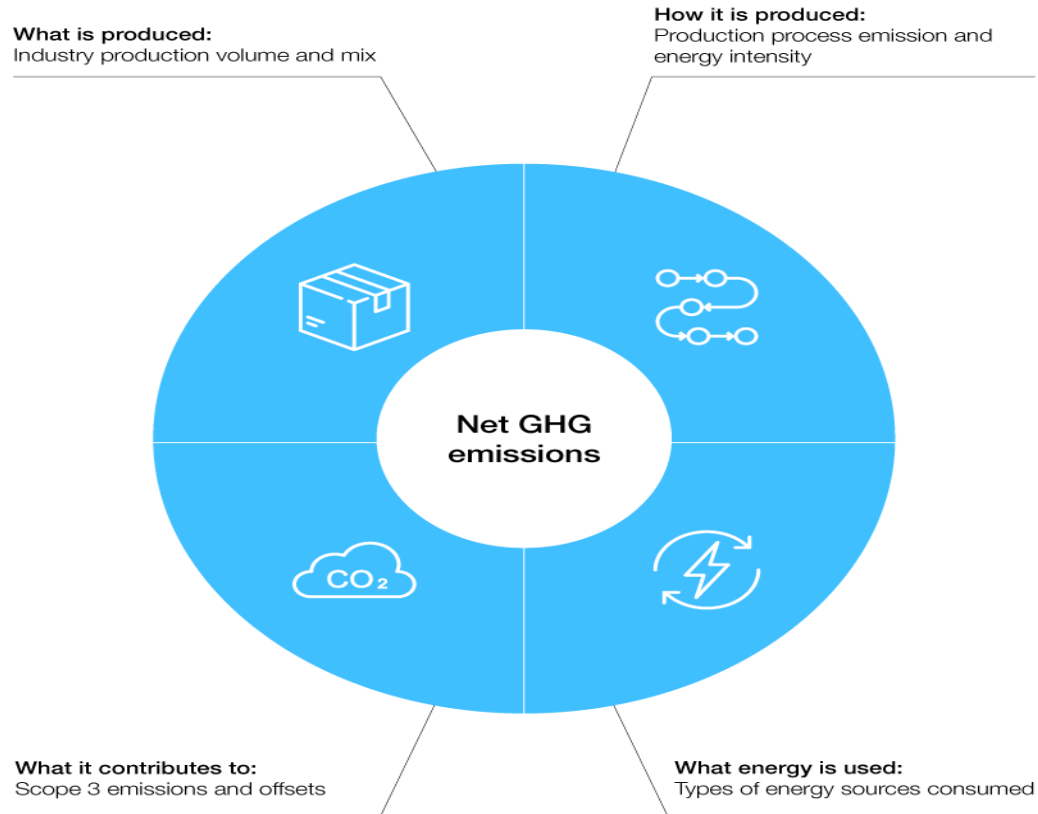
- **Engagement**

- Enhancing the pace and scale of net zero action

Net-Zero Transformation Enablers

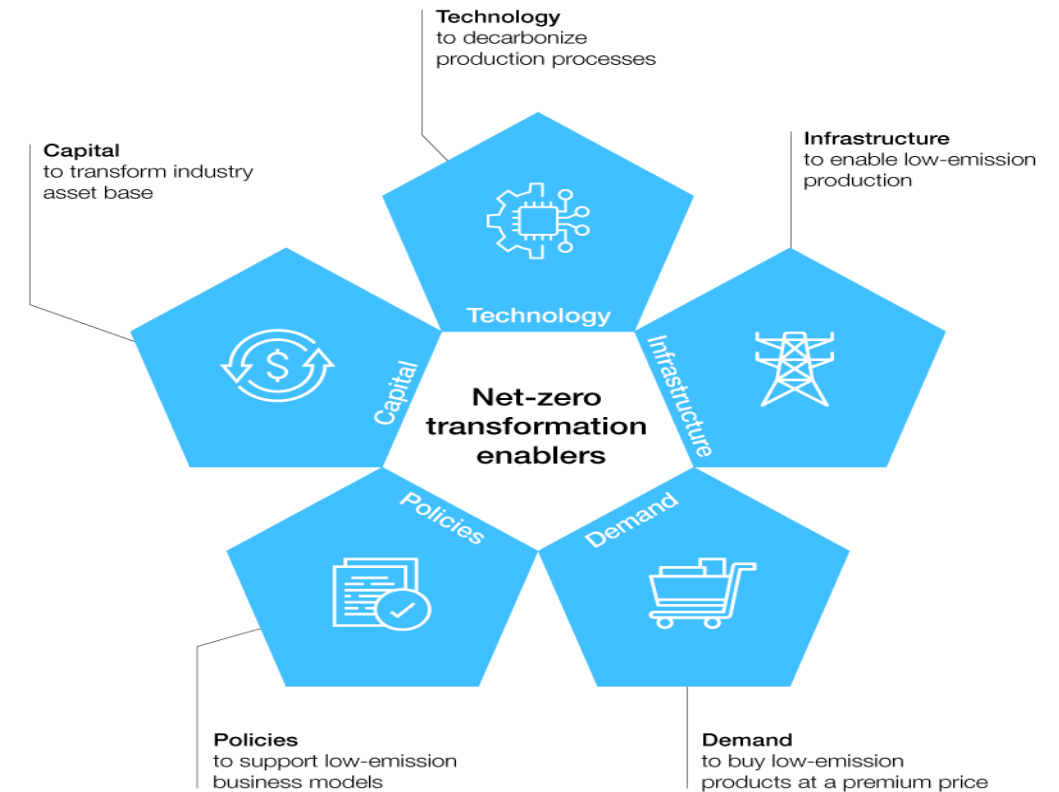
Net-zero industry performance

The four drivers of industry net greenhouse gas (GHG) emissions:

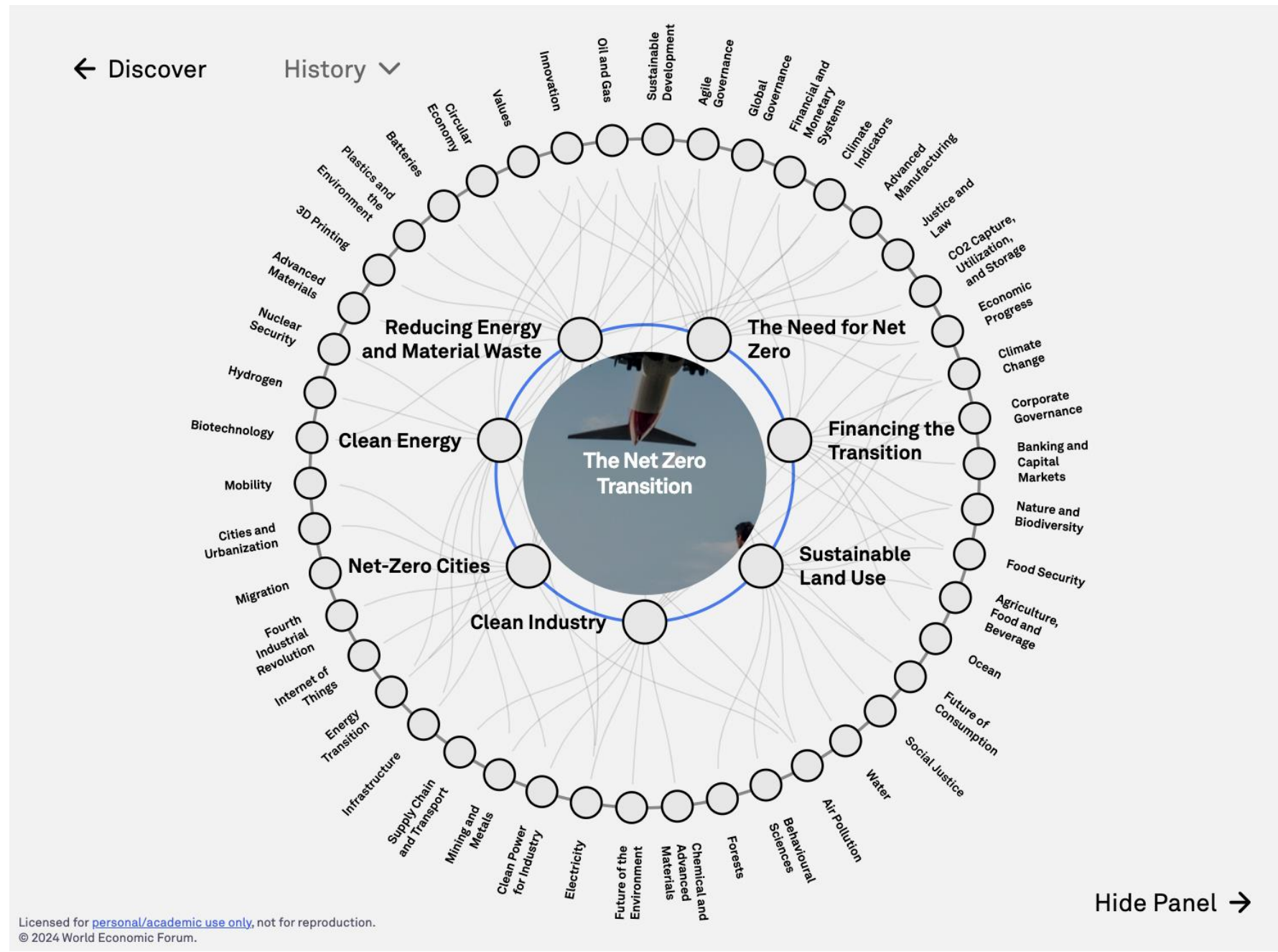


Net-zero industry readiness

The five enabling dimensions of industry net-zero transformation:

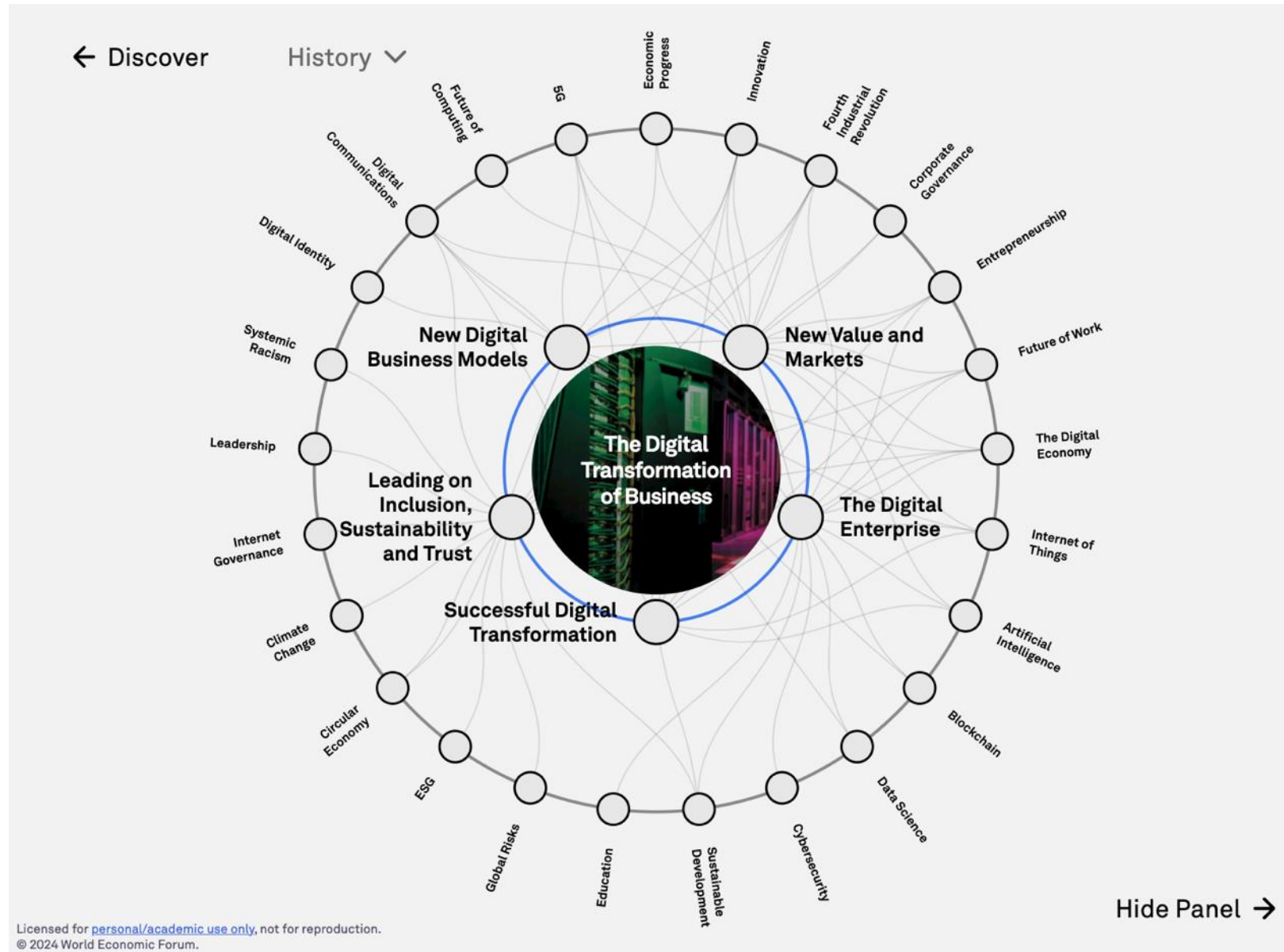


The Net Zero Transition



Licensed for [personal/academic use only](#), not for reproduction.
© 2024 World Economic Forum.

The Digital Transformation of Business



Digital Transformation

Dimensions	Categories
A. BUSINESS MODELS	<ul style="list-style-type: none"> Business Process Innovation Business Strategy
B. DIGITAL BUSINESS	<ul style="list-style-type: none"> Digital Culture, Literacy and Skills Digital Economy Innovation and Socio-technical Shared Values
C. TECHNOLOGIES	<ul style="list-style-type: none"> Technology and Innovation Management Artificial Intelligence Big Data Internet of Things Industry 4.0
D. SUSTAINABILITY	<ul style="list-style-type: none"> Sustainable Business Sustainable Competitive Advantage Sustainable Development Sustainable Innovation
E. HUMAN RESOURCES	<ul style="list-style-type: none"> Employee Experience Career Dynamics
F. SMART CITIES	<ul style="list-style-type: none"> Sustainable Smart Manufacturing Digital Manufacturing

ESG Challenges and Opportunities

- **Challenges**

- **Fragmented and unstructured ESG data.**
- **Lack of standardization and transparency.**
- **Timeliness of data availability.**

- **Opportunities**

- **Rising demand for actionable ESG insights.**
- **Innovation in sustainable solutions and policies.**
- **Generative AI as a tool for transformation.**

Sustainability and ESG Data Analytics



Generative AI for ESG Data Analytics

- **Data Integration and Enrichment:**
 - Synthesizing structured and unstructured ESG data.
- **Automated Reporting and Insight Generation:**
 - Tailored ESG reports and insights for stakeholders.
- **Scenario Modeling and Forecasting:**
 - Simulating potential risks and opportunities.
- **Addressing Bias and Ensuring Accountability:**
 - Transparent, fair, and ethical AI deployment.

Generative AI and LLMs for Sustainability and ESG Data Analytics



Sustainability Innovation with Generative AI

- **Sustainable Product Design:**
 - Eco-friendly designs minimizing waste and energy.
- **Policy Formulation and Implementation:**
 - AI-driven simulations for effective policies.
- **Stakeholder Engagement and Awareness:**
 - Communicating ESG strategies with compelling AI-driven visuals.

Mapping the ESG Standards Landscape

- **The most prevalent ESG reporting frameworks**
 - **GRI (Global Report Initiative)**
 - **CDP (Carbon Disclosure Project)**
 - **SASB (Sustainability Accounting Standards Board)**
 - **ISSB (International Sustainability Standards Board)**
 - **TCFD (Task Force on Climate-related Financial Disclosures)**
- **How companies choose**
 - **Materiality, industry-specific standards, investor alignment**

GRI (Global Report Initiative)



Standards ▾

How to use the GRI Standards ▾

Reporting support ▾

Public policy & partnerships ▾

About GRI ▾

News ▾

Goals and targets database

Sign In

Search 🔍

Donate Now



The global leader for impact reporting

Welcome to GRI. For over 25 years, we have developed and delivered the global best practice for how organizations communicate and demonstrate accountability for their impacts on the environment, economy and people.

We provide the world's most widely used sustainability reporting standards, which cover topics that range from biodiversity to tax, waste to emissions, diversity and equality to health and safety. As such, GRI reporting is the enabler for transparency and dialogue between companies and their stakeholders.

[Access the GRI Standards →](#)

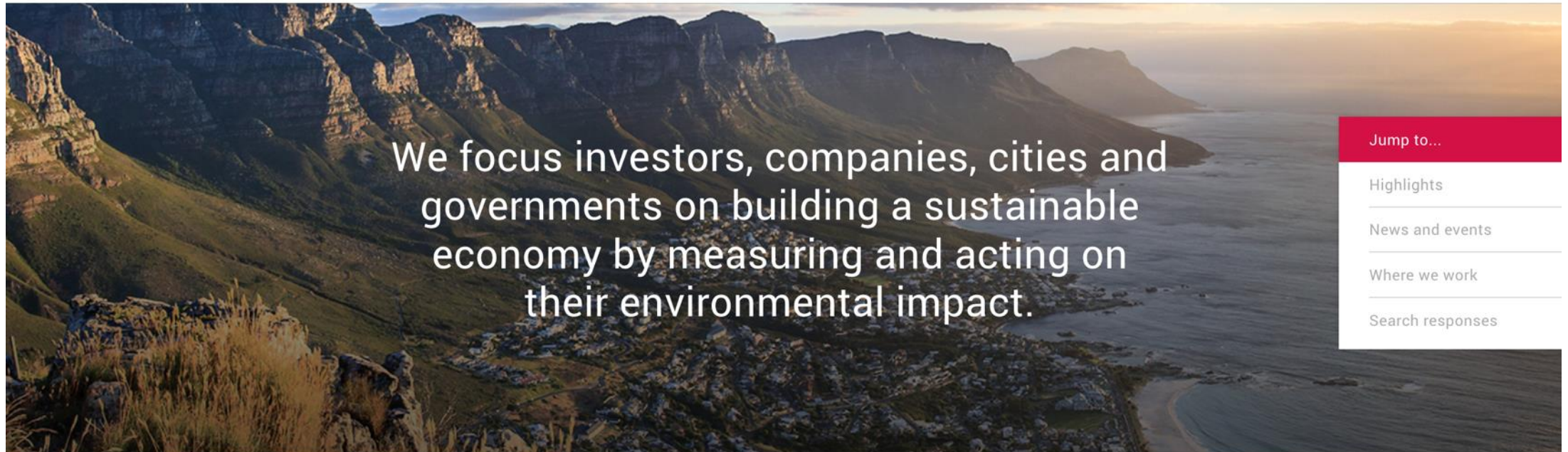
Feedback

CDP (Carbon Disclosure Project)



[Guidance & questionnaires](#) [Contact](#) [Regional websites](#) [Language](#)

[About us](#) [Our work](#) [Why disclose?](#) [Become a member](#) [Data and insights](#)



CDP is a not-for-profit charity that runs the global disclosure system for [investors](#), [companies](#), [cities](#), [states and regions](#) to manage their environmental impacts. Over the past 20 years we have created a system that has resulted in unparalleled engagement on environmental issues worldwide. Find out more about [how we work](#).

<https://www.cdp.net/>

SASB (Sustainability Accounting Standards Board)

IFRS Foundation

Other Resources: [The ISSB](#) [Integrated Reporting Framework](#)



[✉ Subscribe](#) [↓ Download Standards](#)

[About](#) [SASB Standards](#) [Using the SASB Standards](#) [Pathway to ISSB](#) [Education](#) [Membership](#)

An aerial photograph of a landscape. On the left, there's a river winding through green fields. On the right, there's a dense forest of trees with some yellow foliage. A road or path crosses the river.

SASB Standards: Your pathway to ISSB

[Learn more](#)

<https://sasb.org/>

ISSB (International Sustainability Standards Board)



ABOUT US | IFRS ACCOUNTING | IFRS SUSTAINABILITY

Home > International Sustainability Standards Board

International Sustainability Standards Board

ABOUT

MEMBERS

MEETINGS

RESOURCES

NEWS

About the International Sustainability Standards Board

The Trustees of the IFRS Foundation announced the formation of the International Sustainability Standards Board (ISSB) on 3 November 2021 at COP26 in Glasgow, following strong market demand for its establishment. The ISSB is developing—in the public interest—standards that will result in a high-quality, comprehensive global baseline of sustainability disclosures focused on the needs of investors and the financial markets.

Sustainability factors are becoming a mainstream part of investment decision-making. There are increasing calls for companies to provide high-quality, globally comparable information on sustainability-related risks and opportunities, as indicated by feedback from many consultations with market

Related information

[Sustainability FAQs](#)

[General Sustainability-related Disclosures project](#)

[Climate-related Disclosures project](#)

[Consolidated organisations](#)

<https://www.ifrs.org/groups/international-sustainability-standards-board/>

TCFD

(Task Force on Climate-related Financial Disclosures)



<https://www.ifrs.org/sustainability/tcfd/>



ABOUT US | IFRS ACCOUNTING | IFRS SUSTAINABILITY

Home > ISSB and TCFD

ISSB and TCFD

The Financial Stability Board has announced that the work of the TCFD has been completed, with the ISSB's Standards marking the '**culmination of the work of the TCFD**'.

Companies applying IFRS S1 *General Requirements for Disclosure of Sustainability-related Financial Information* and IFRS S2 *Climate-related Disclosures* will meet the TCFD recommendations as the recommendations are fully incorporated into the ISSB's Standards.

Companies can continue to use the **TCFD recommendations** should they choose to do so, and some companies may still be required to use the TCFD recommendations. Using the recommendations is a good entry point for companies as they move to use the ISSB's Standards.

The IFRS Foundation has **published a comparison** of the requirements in IFRS S2 and the TCFD recommendations.

Related Information

[IFRS Foundation welcomes culmination of TCFD work and transfer of TCFD monitoring responsibilities to ISSB from 2024](#)

[Comparison: IFRS S2 Climate-related Disclosures with the TCFD Recommendations](#)

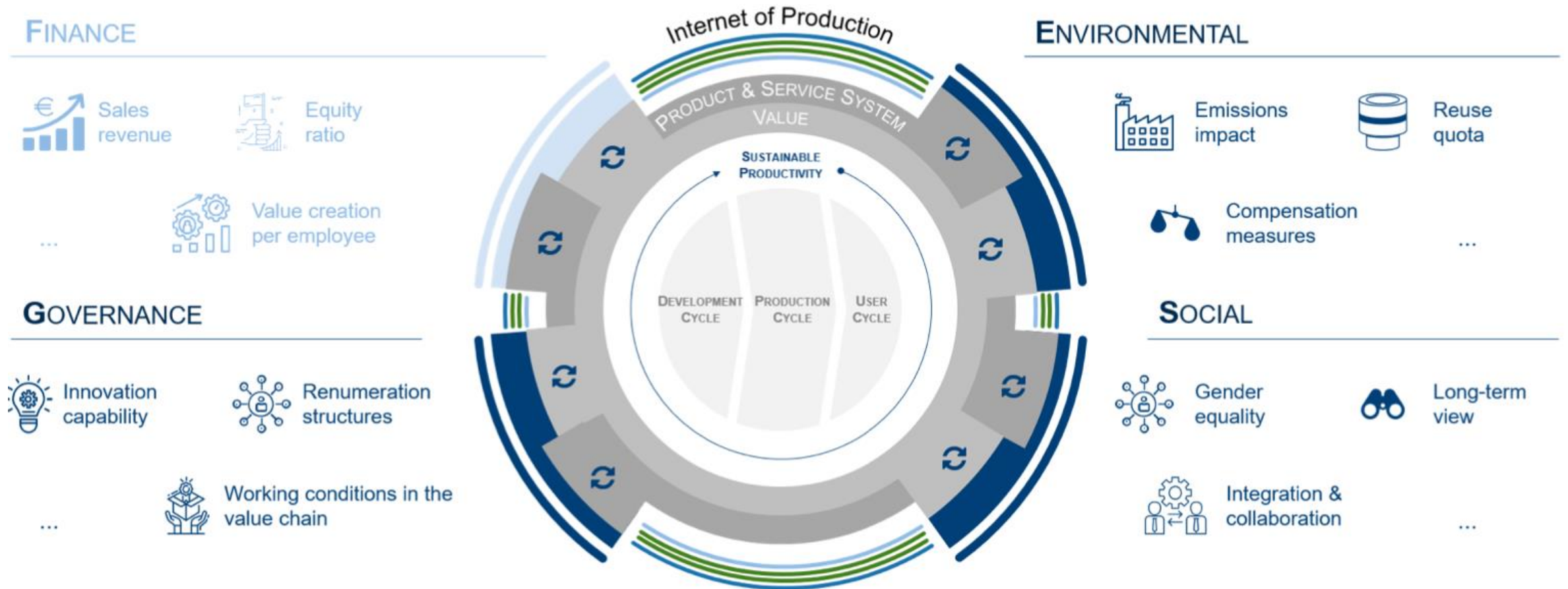
[Resource: Making the transition from TCFD to ISSB](#)

[IFRS Sustainability Standards Navigator](#)

<https://www.fsb-tcfd.org/>

Sustainable Resilient Manufacturing

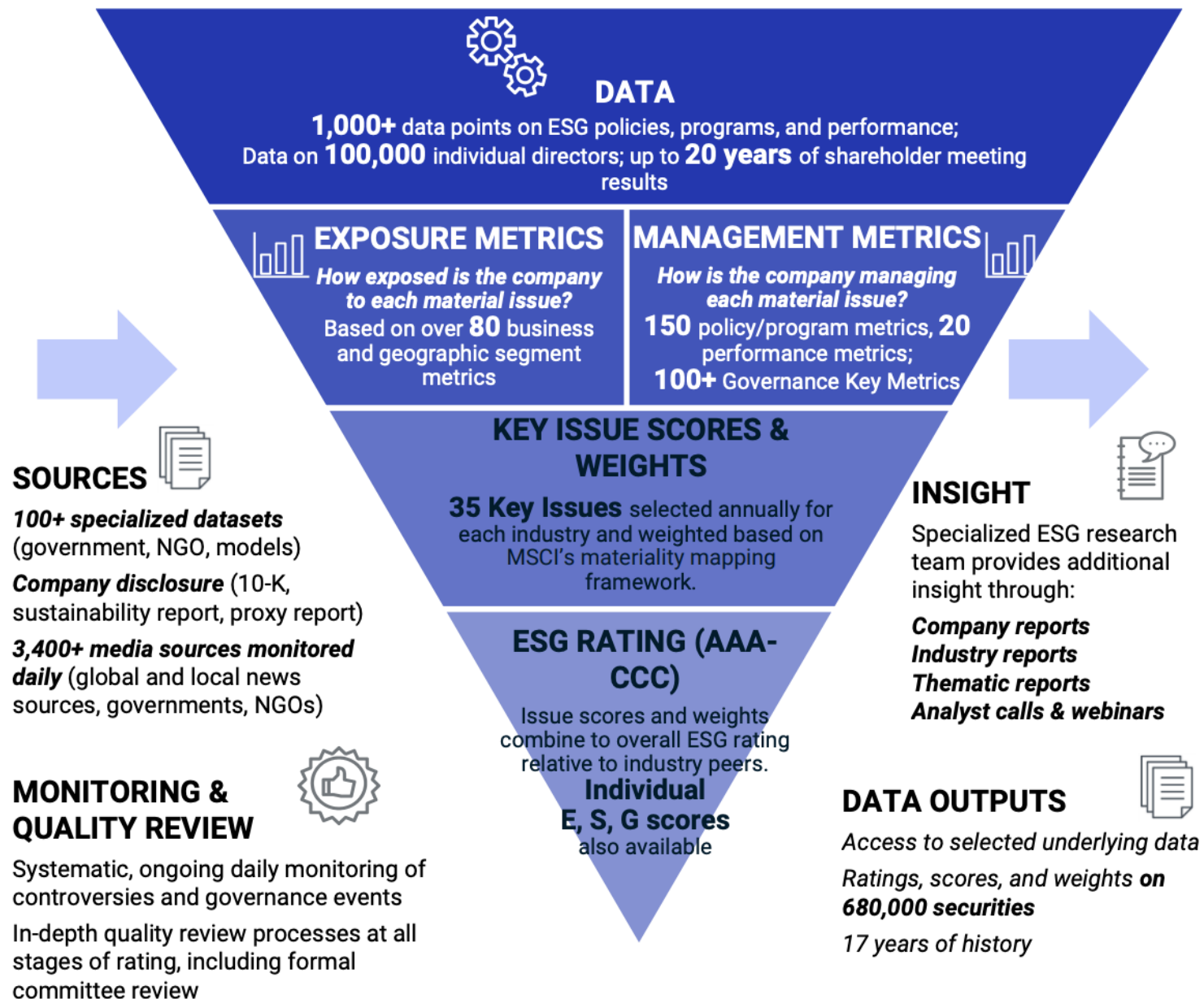
ESG



ESG Indexes

- **MSCI ESG Index**
- **Dow Jones Sustainability Indices (DJSI)**
- **FTSE ESG Index**

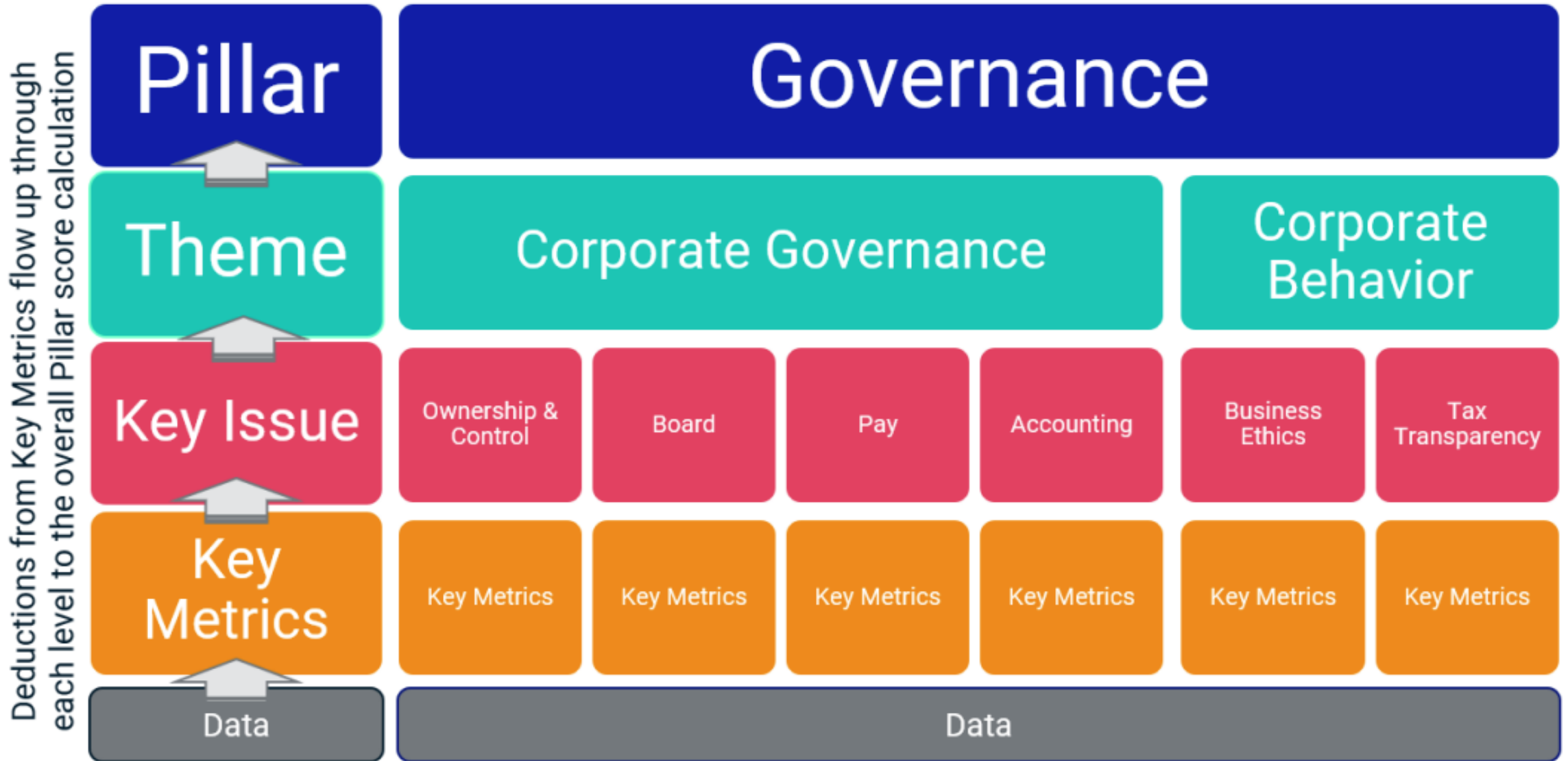
MSCI ESG Rating Framework



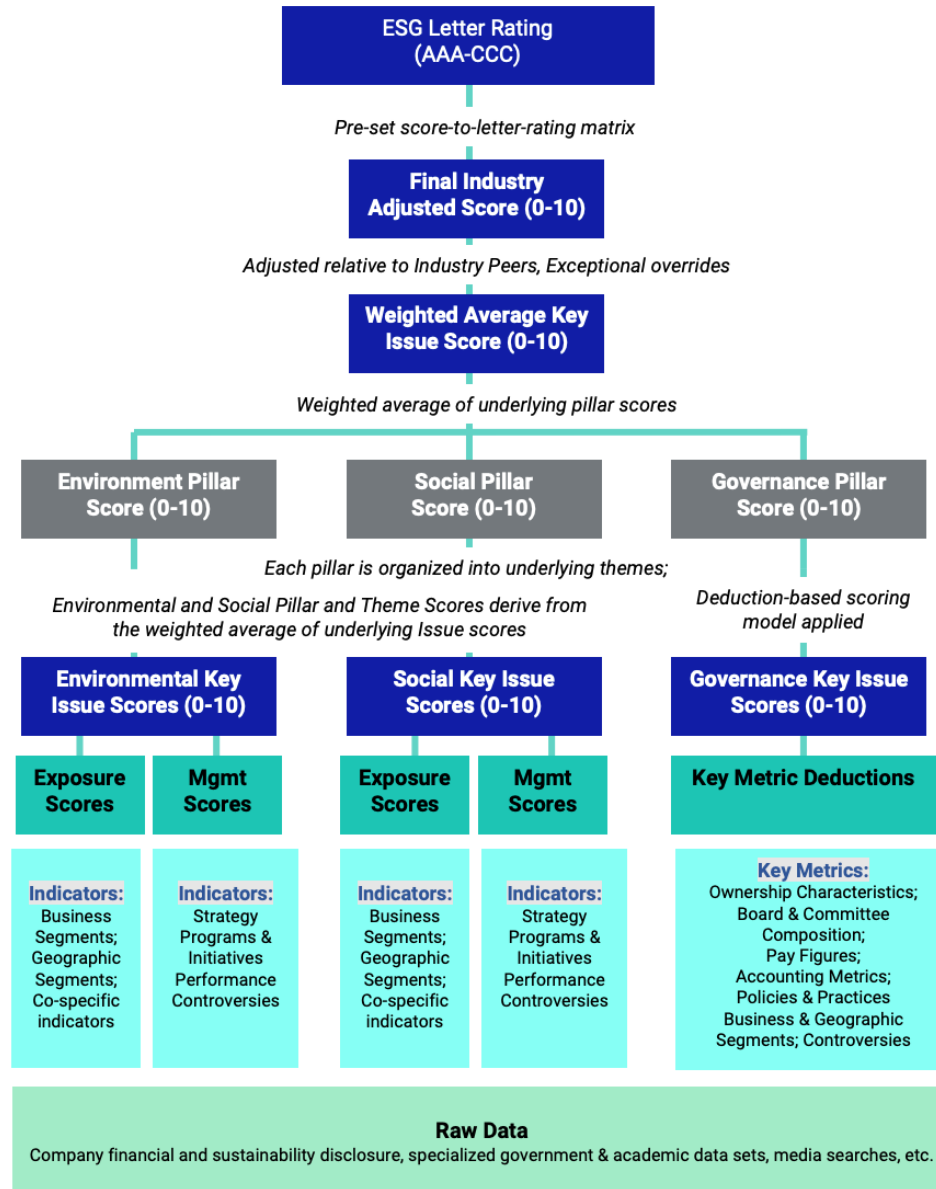
MSCI ESG Key Issue Hierarchy

3 Pillars	10 Themes	35 ESG Key Issues	
Environment	Climate Change	Carbon Emissions Product Carbon Footprint	Financing Environmental Impact Climate Change Vulnerability
	Natural Capital	Water Stress Biodiversity & Land Use	Raw Material Sourcing
	Pollution & Waste	Toxic Emissions & Waste Packaging Material & Waste	Electronic Waste
	Environmental Opportunities	Opportunities in Clean Tech Opportunities in Green Building	Opportunities in Renewable Energy
Social	Human Capital	Labor Management Health & Safety	Human Capital Development Supply Chain Labor Standards
	Product Liability	Product Safety & Quality Chemical Safety Consumer Financial Protection	Privacy & Data Security Responsible Investment Health & Demographic Risk
	Stakeholder Opposition	Controversial Sourcing Community Relations	
	Social Opportunities	Access to Communications Access to Finance	Access to Health Care Opportunities in Nutrition & Health
Governance	Corporate Governance	Ownership & Control Board	Pay Accounting
	Corporate Behavior	Business Ethics Tax Transparency	

MSCI Governance Model Structure



MSCI Hierarchy of ESG Scores

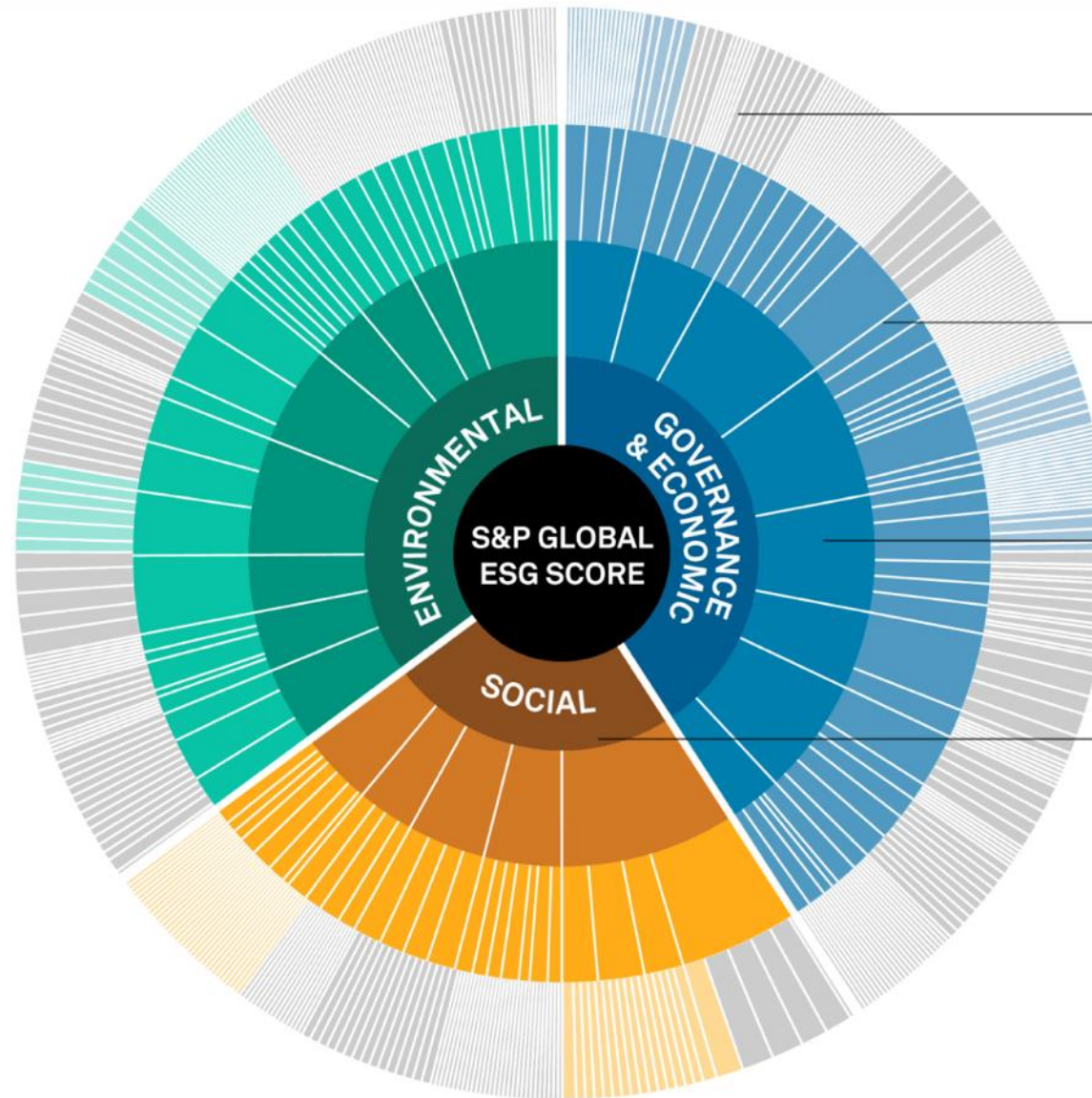


DJSI S&P Global ESG Score

8,000
Companies

90%
Global market capitalization

340,000+
Current Research Universe and Active Securities



Approx.
1,000
Datapoints

Assessed values, text, checkboxes, documents

Sources: Web-based questionnaire and company documents

130+
Questions

Weighted data point scores

Up to 50% industry-specific

Ave.
30+
Criteria scores

Weighted question scores

61 industry specific approaches, with tailored questions, criteria and related weightings

3
Dimension scores

Weighted criteria scores

Adjusted for corporate ESG controversies where applicable

1
S&P Global ESG Score

Sum of weighted dimension scores

FTSE Russell ESG Ratings



Sustainalytics

ESG Risk Ratings

Analyst-based
approach

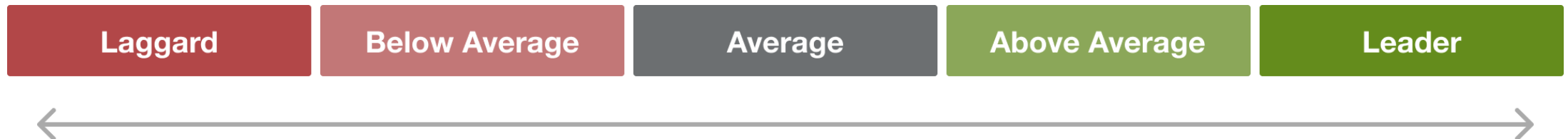
Sustainalytics' ESG Risk Ratings measure a company's exposure to industry-specific material ESG risks and how well a company is managing those risks.

Negligible	Low	Medium	High	Severe
0 - 10	10 - 20	20 - 30	30 - 40	40+

Truvalue ESG Ranks

Machine-based
approach

- **Truvalue Labs** applies **AI** to analyze over **100,000 sources** and uncover **ESG risks** and opportunities hidden in **unstructured text**.
- The ESG Ranks data service produces an overall company rank based on industry percentile leveraging the **26 ESG categories** defined by the **Sustainability Accounting Standards Board (SASB)**.
- The data feed covers **20,000+** companies with more than **13 years** of history.



Analyst-driven vs. AI-driven ESG

Analyst-driven ESG research

Derives ratings in a structured data model

Sustainalytics



Analyst role at the end of the process allows subjectivity to color results

AI-driven ESG research

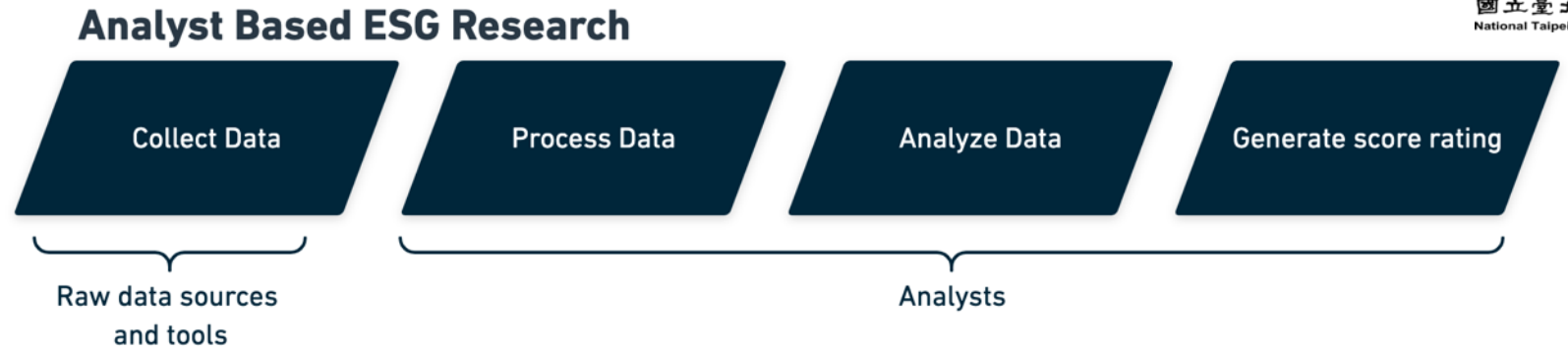
Derives signals from unstructured data

Truvalue Labs

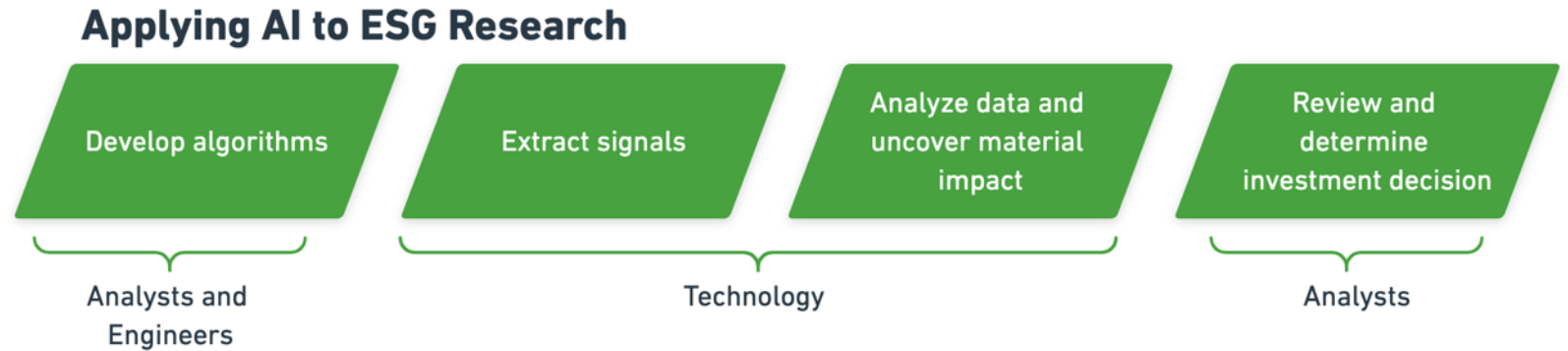


Analyst expertise at the beginning of the process produces consistent results

Analyst based ESG Research

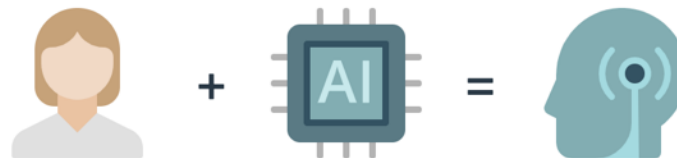


AI based ESG Research



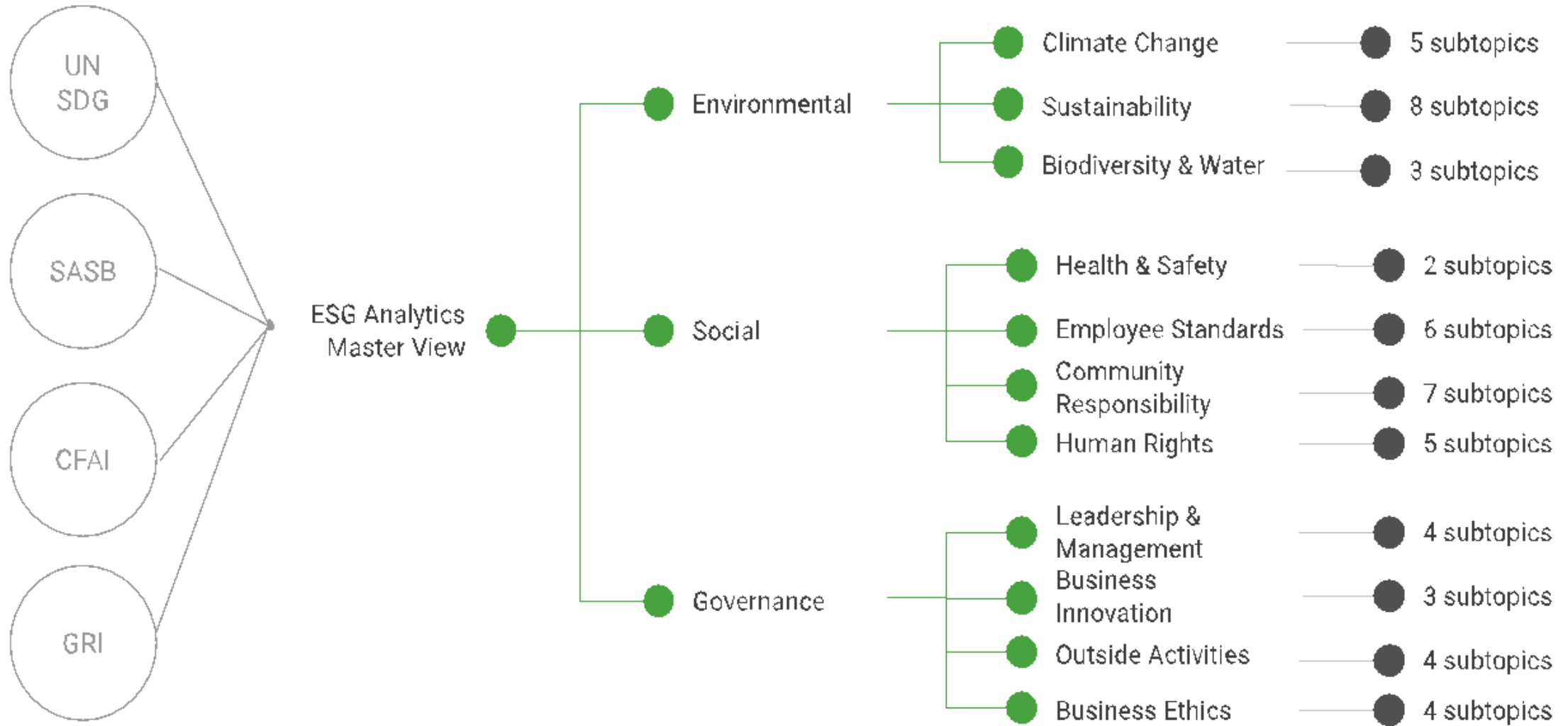
It would take an analyst over 5 years to do what our AI can in 1 week

Combining analysts with AI creates gives you the full picture



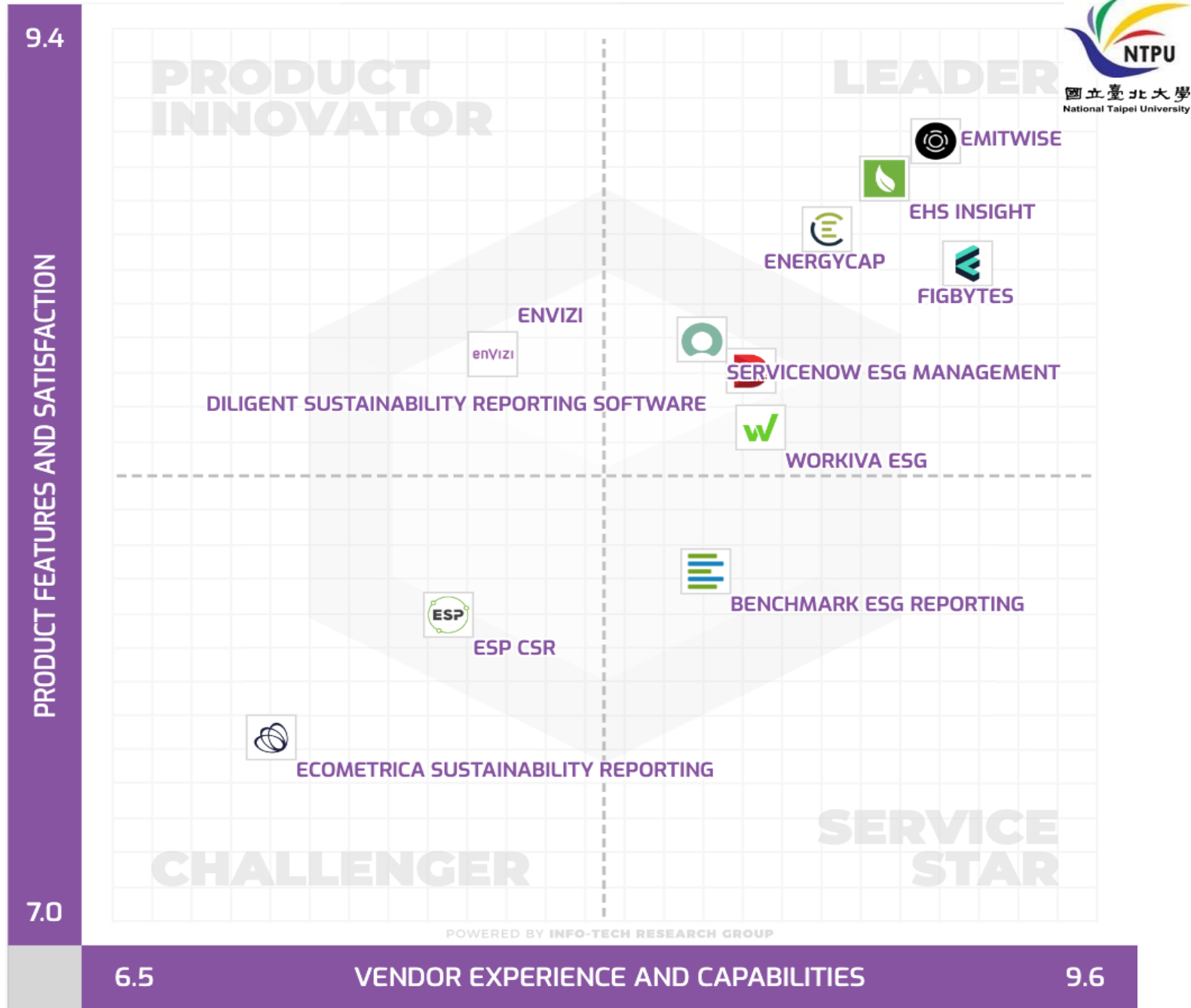
ESG ANALYTICS
Invest where it matters.

ESG Analytics: NLP Taxonomy



Top ESG Reporting Software

Environmental, Social and Governance (ESG) Reporting software or Sustainability software helps organizations manage their operational data, evaluate their impact on the environment and provide reporting to perform audits.



Future Directions

- Integrating blockchain, IoT, and digital twins.
- Democratizing AI tools for all stakeholders.
- Promoting collaboration among experts and communities.

Conclusion

- **Generative AI is transforming ESG analytics and sustainability innovation.**
- **Collaboration among researchers, policymakers, and innovators is key.**
- **Generative AI to build a sustainable future.**

Summary

- 1. Generative AI**
- 2. ESG**
- 3. Sustainable Development**

References

- Stuart Russell and Peter Norvig (2020), *Artificial Intelligence: A Modern Approach*, 4th Edition, Pearson.
- Numa Dhamani and Maggie Engler (2024), *Introduction to Generative AI*, Manning
- 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
- NVIDIA DLI (2024), *Building RAG Agents with LLMs*, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-15+V1
- NVIDIA DLI (2024), *Generative AI with Diffusion Models*, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1
- Denis Rothman (2024), *Transformers for Natural Language Processing and Computer Vision: Explore Generative AI and Large Language Models with Hugging Face, ChatGPT, GPT-4V, and DALL-E 3*, 3rd Edition, Packt Publishing
- Denis Rothman (2024), *RAG-Driven Generative AI: Build custom retrieval augmented generation pipelines with LlamaIndex, Deep Lake, and Pinecone*, Packt Publishing
- Jay Alammar and Maarten Grootendorst (2024), *Hands-On Large Language Models: Language Understanding and Generation*, O'Reilly Media
- Simon Thompson (2023), *Green and Sustainable Finance: Principles and Practice in Banking, Investment and Insurance*, 2nd Edition, Kogan Page.
- Chrissa Pagitsas (2023), *Chief Sustainability Officers At Work: How CSOs Build Successful Sustainability and ESG Strategies*, Apress.
- 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.
- 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.
- Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min et al. (2023) "A Survey of Large Language Models." arXiv preprint arXiv:2303.18223.
- Touvron, Hugo, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov et al. (2023) "Llama 2: Open Foundation and Fine-Tuned Chat Models." arXiv preprint arXiv:2307.09288 (2023).
- Rafailov, R., Sharma, A., Mitchell, E., Ermon, S., Manning, C. D., & Finn, C. (2023). Direct preference optimization: Your language model is secretly a reward model. arXiv preprint arXiv:2305.18290.
- Tunstall, Lewis, Edward Beeching, Nathan Lambert, Nazneen Rajani, Kashif Rasul, Younes Belkada, Shengyi Huang et al. "Zephyr: Direct Distillation of LM Alignment." arXiv preprint arXiv:2310.16944 (2023).
- Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., ... & Lowe, R. (2022). Training language models to follow instructions with human feedback. arXiv preprint arXiv:2203.02155.
- Gozalo-Brizuela, Roberto, and Eduardo C. Garrido-Merchan (2023). "ChatGPT is not all you need. A State of the Art Review of large Generative AI models." arXiv preprint arXiv:2301.04655 (2023).
- Wenliang Dai, Junnan Li, Dongxu Li, Anthony Meng Huat Tiong, Junqi Zhao, Weisheng Wang, Boyang Li, Pascale Fung, and Steven Hoi. (2023) "InstructBLIP: Towards General-purpose Vision-Language Models with Instruction Tuning." arXiv preprint arXiv:2305.06500 (2023).
- Shahab Saquib Sohail, Faiza Farhat, Yassine Himeur, Mohammad Nadeem, Dag Øivind Madsen, Yashbir Singh, Shadi Atalla, and Wathiq Mansoor (2023). "The Future of GPT: A Taxonomy of Existing ChatGPT Research, Current Challenges, and Possible Future Directions." *Current Challenges, and Possible Future Directions (April 8, 2023)* (2023).
- Longbing Cao (2022). "Decentralized ai: Edge intelligence and smart blockchain, metaverse, web3, and desc." *IEEE Intelligent Systems* 37, no. 3: 6-19.
- Qinglin Yang, Yetong Zhao, Huawei Huang, Zehui Xiong, Jiawen Kang, and Zibin Zheng (2022). "Fusing blockchain and AI with metaverse: A survey." *IEEE Open Journal of the Computer Society* 3 : 122-136.
- Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., ... & Lowe, R. (2022). Training language models to follow instructions with human feedback. arXiv preprint arXiv:2203.02155.

國立臺北大學 商學院碩士在職專班
企業倫理與永續發展

(Business Ethics and Sustainable Development)



Q & A

Generative AI for ESG and Sustainable Development (生成式 AI 於 ESG 與永續發展)

Time: 18:50-21:30, Tuesday, March, 11, 2025

企業倫理與永續發展 (Business Ethics and Sustainable Development)

任課教師：陳宥杉，戴敏育

戴敏育 教授 (Prof. Min-Yuh Day)

國立臺北大學 資訊管理研究所 教授

金融科技暨綠色金融研究中心 主任

永續辦公室 永續發展組 組長

