

Natural Language Processing Applications

Course Introduction



fit@hcmus

KHOA CÔNG NGHỆ THÔNG TIN
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

- ❑ Course Introduction
- ❑ Course Evaluation
- ❑ Course Outline
- ❑ References



NLPA - Course Introduction

COURSE INTRODUCTION



Course Information

- ❑ Course name: Natural Language Processing Applications
- ❑ Credits: 4
 - ❑ Lecture: 45 hrs
 - ❑ Lab: 30 hrs
 - ❑ Self-study: 90 hrs
- ❑ Knowledge block: Elective - Computer Science
- ❑ Prerequisite: Introduction to Natural Language Processing

NLPA - Course Introduction

COURSE OBJECTIVES



Course Objectives

- ❑ Understand the basic applications of Natural Language Processing (NLP)
- ❑ Develop skills in describing, analyzing and modeling a real-world NLP application
- ❑ Know the methods/measures used to evaluate NLP applications
- ❑ Build some basic NLP applications





fit@hcmus

NLPA - Course Introduction

COURSE EVALUATION



Course Evaluation

- ❑ Lecture:
 - ❑ Assignments:
 - Quizzes: 5%
 - Homework: 5%
 - ❑ Projects:
 - Project1 (Seminar): 30%
 - Project2 (Application): 40%
- ❑ Lab:
 - ❑ Weekly lab work: 20%



NLPA - Course Introduction

COURSE OUTLINE



Lecture 1. Introduction

- ❑ Introduction to NLP Applications
- ❑ Development status of NLP Applications
- ❑ Examples of real NLP applications



Lecture 2. Language Models

- ❑ Introducing Small/Large Language Models
- ❑ Capabilities of LLMs
- ❑ Major LLM Examples
- ❑ How LLMs Work
- ❑ Applications in Real Life
- ❑ Limitations
- ❑ Challenges



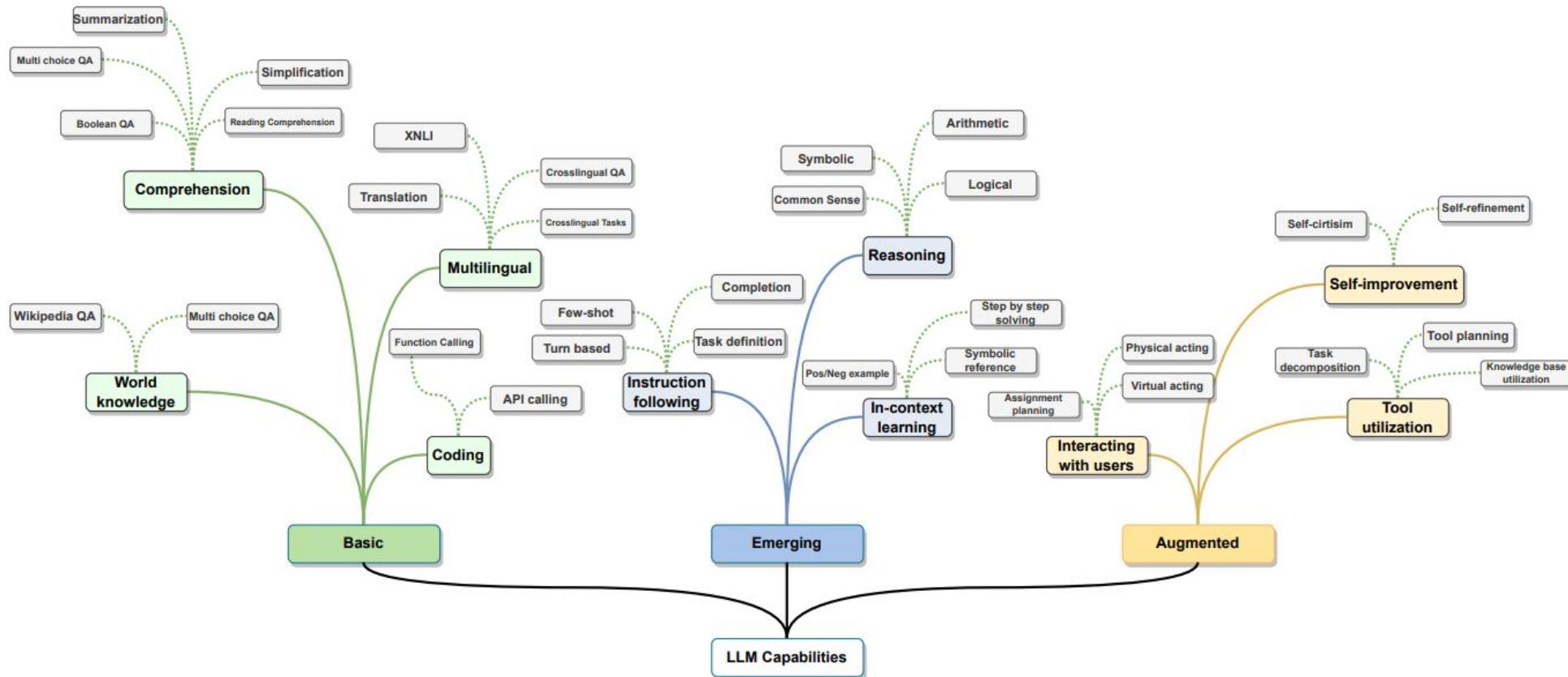
Lecture 2. Language Models

- ❑ Introducing Small/Large Language Models
 - Definition of Large Language Models (LLMs)
 - Brief history and evolution (e.g., GPT, BERT, T5)
 - Importance of LLMs in AI and everyday applications



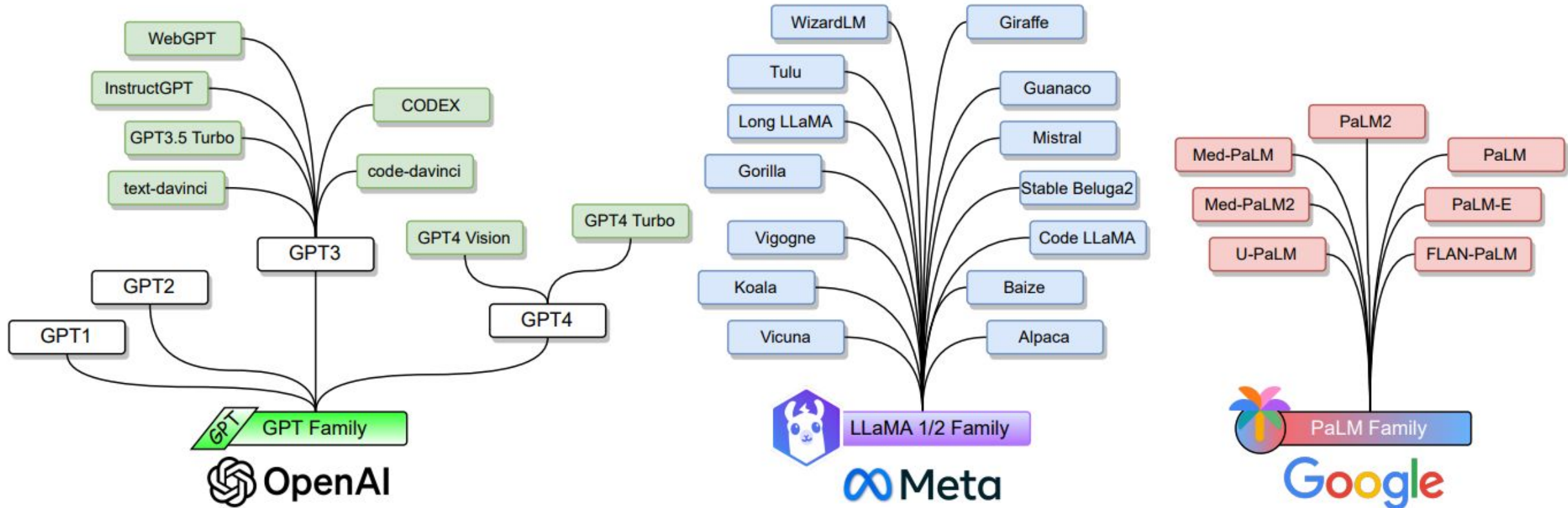
Lecture 2. Language Models

Capabilities of LLMs



Lecture 2. Language Models

Major LLM Examples

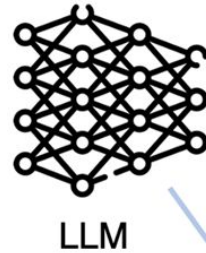


How LLMs Work

Prompting and Tuning LLMs

Pre-training

Trained using large collections of text to predict the next word in a sequence



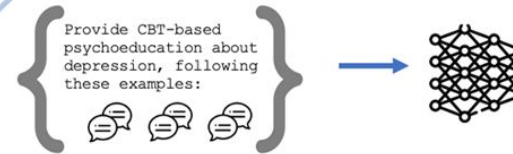
Zero-Shot Learning

Prompt gives LLM context and framing for task, with no examples.



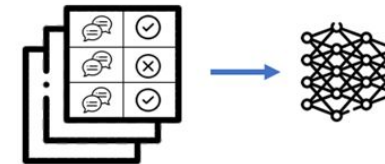
Few-Shot Learning

Prompts with examples



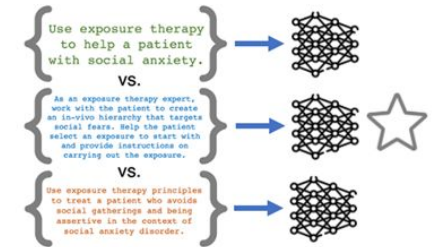
Fine-Tuning

Large set of **human-labeled data** (input/outputs rated for quality) used to train LLM



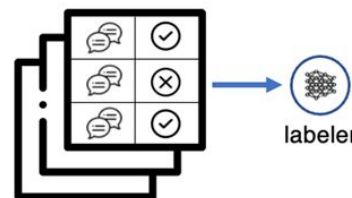
Prompt Engineering

Crafting successful prompts

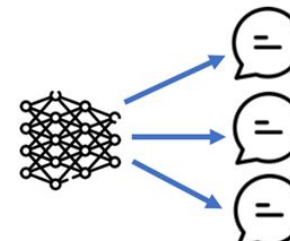


Reinforcement Learning with Human Feedback

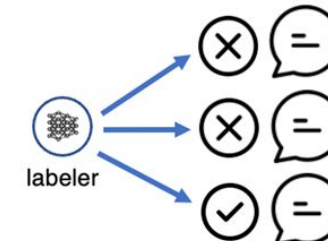
Large set of **human-labeled data** used to train mini model ("labeler")



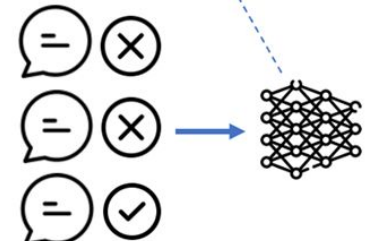
LLM generates many outputs



Labeler rates outputs

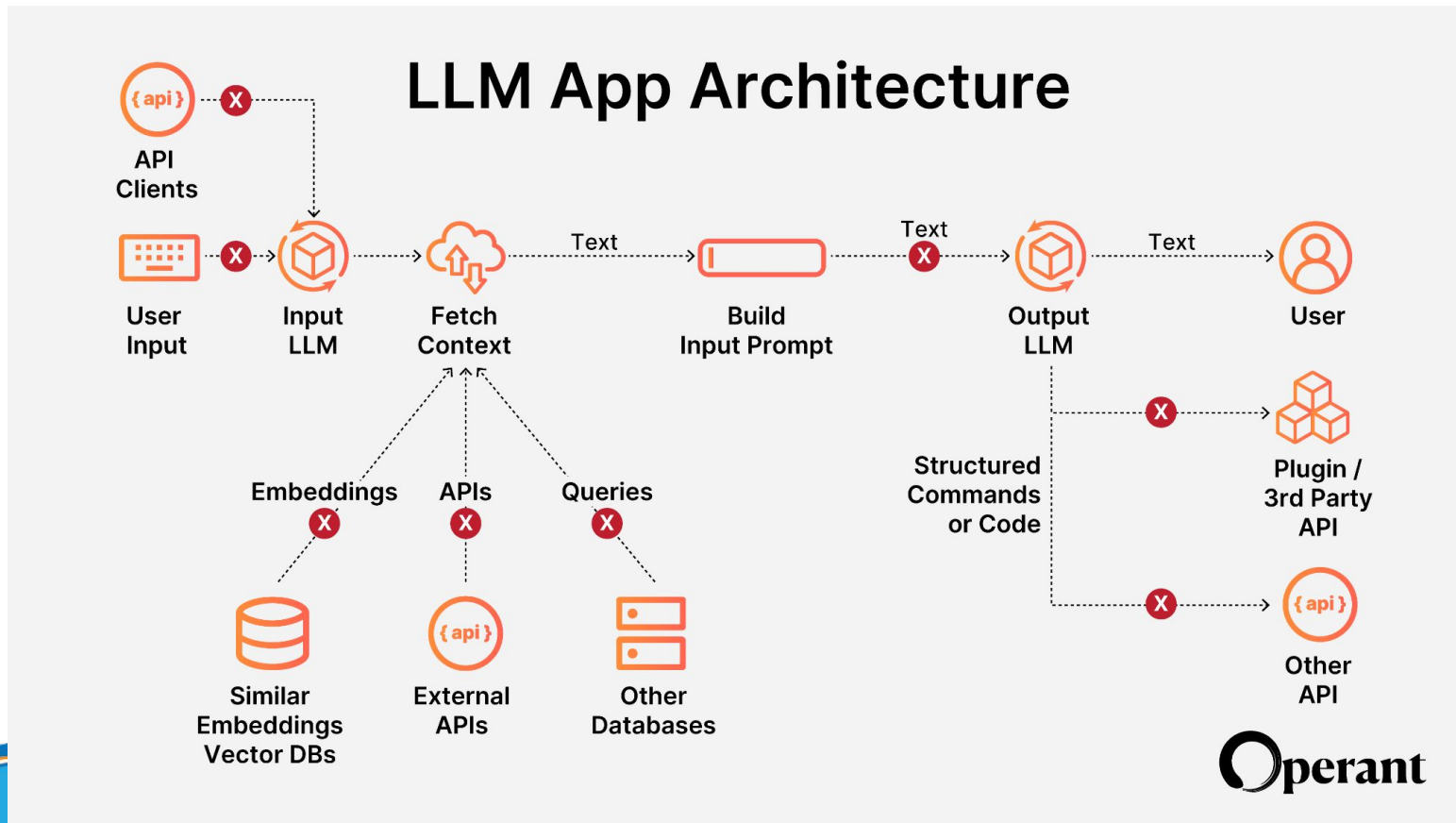


Large dataset of output and labeler's ratings used to train LLM



Lecture 2. Language Models

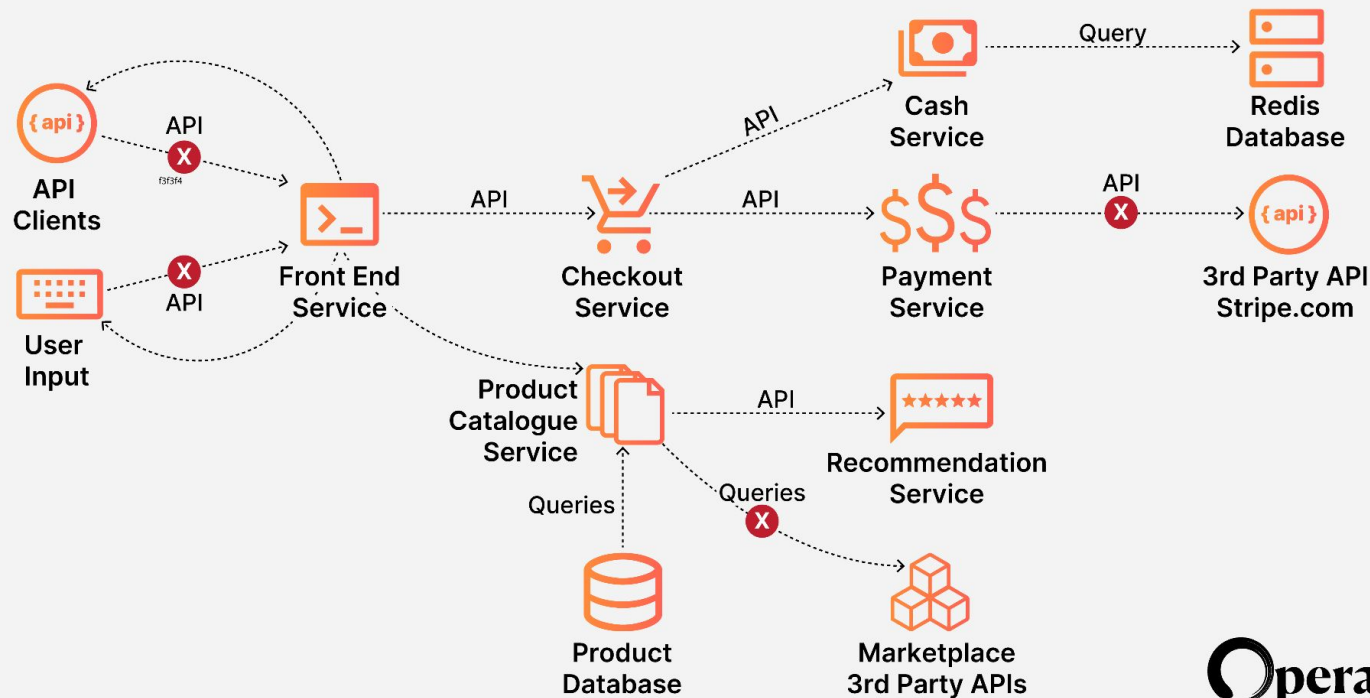
❑ Applications in Real Life

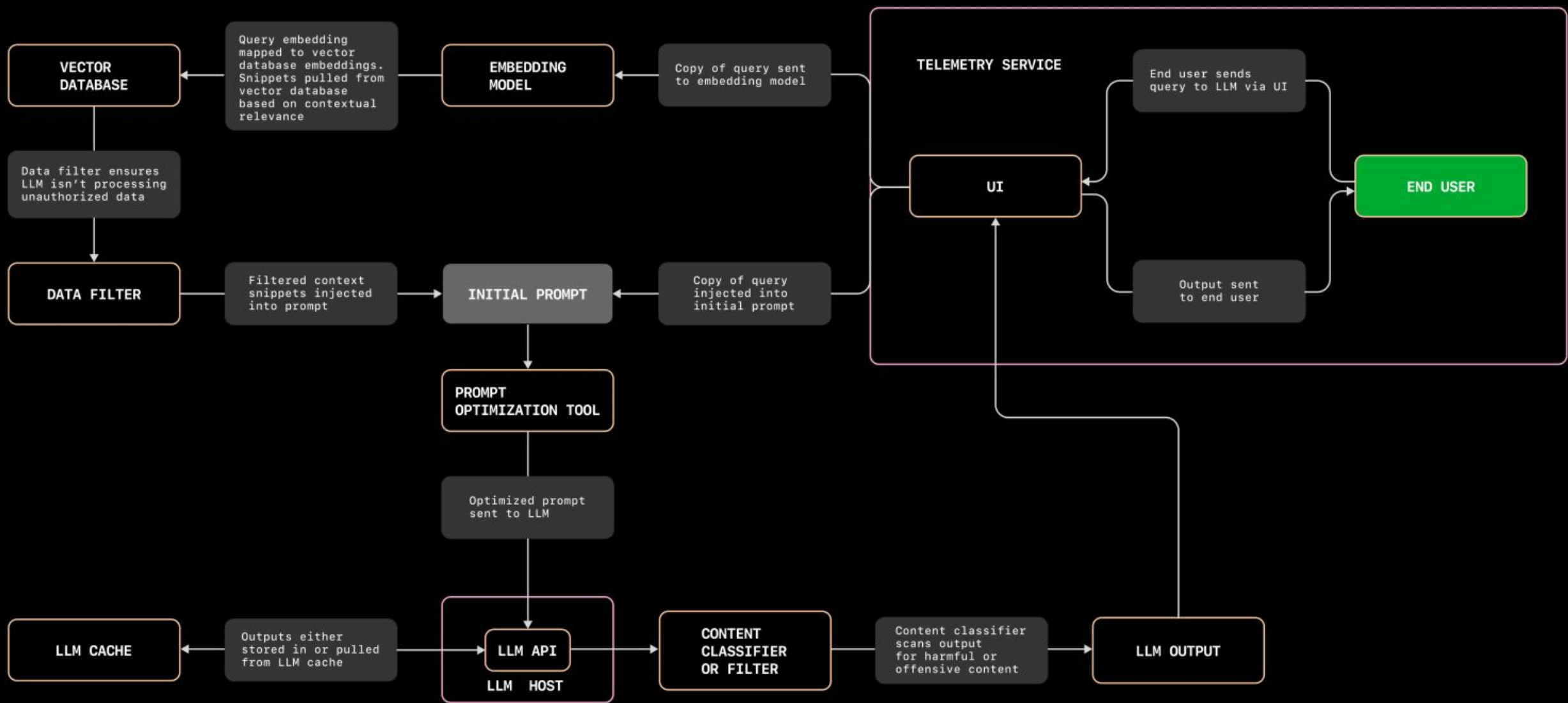


Lecture 2. Language Models

❑ Applications in Real Life

Microservices Architecture



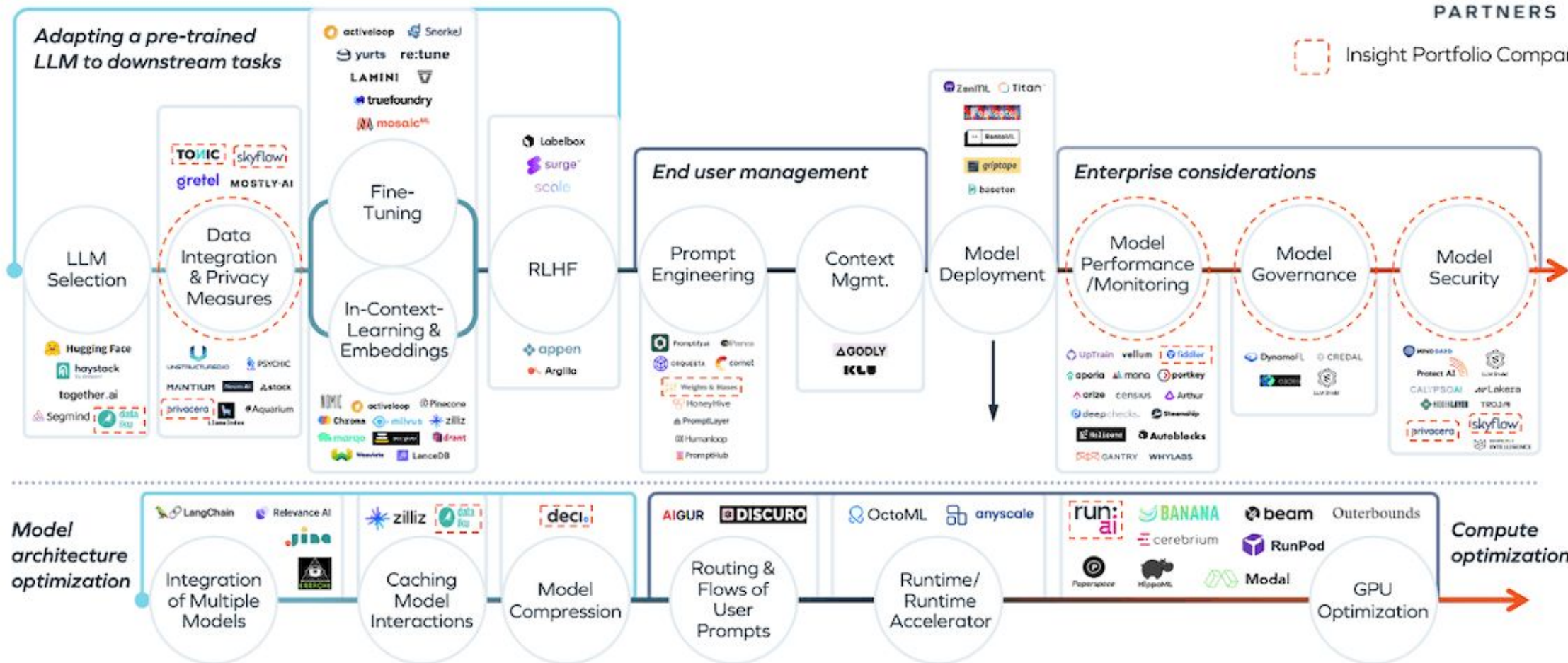


THIS DIAGRAM REPRESENTS THE ARCHITECTURE OF TODAY'S LLM APPLICATION. THE DIFFERENT COMPONENTS CAN BE ROUGHLY GROUPED INTO THREE CATEGORIES: USER INPUT, INPUT ENRICHMENT TOOLS AND PROMPT CONSTRUCTION, AND EFFICIENT AND RESPONSIBLE AI TOOLING.

LLMOps

INSIGHT
PARTNERS

 Insight Portfolio Company



Lecture 2. Language Models

❑ Future Trends

- Combining LLMs with Knowledge Graphs
- LLMs in multimodal applications (text, images, videos)
- Open-source and lightweight models for broader accessibility
- Advances in reinforcement learning with human feedback (RLHF)
- Improved interpretability and control



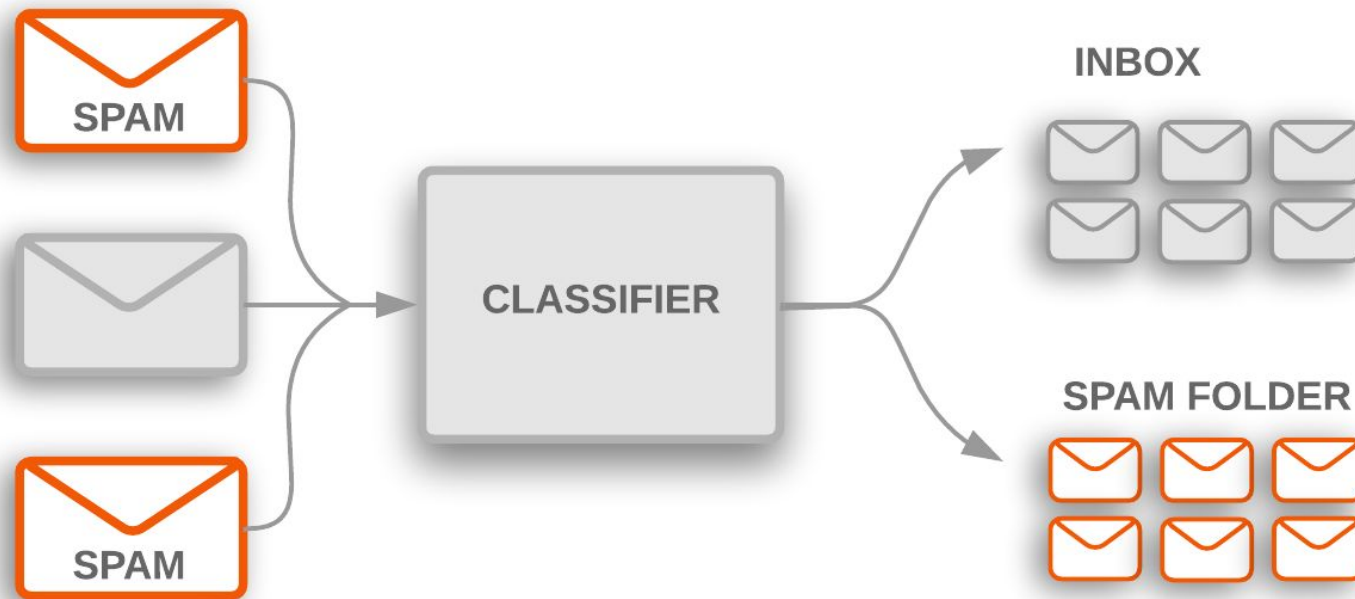
Lecture 3. Text Classification

- ❑ Introduction
- ❑ Common Applications
- ❑ Future Trends



Lecture 3. Text Classification

❑ Introduction



Lecture 3. Text Classification

❑ Common Applications

Real-world Applications of Intent Classification



Customer Support



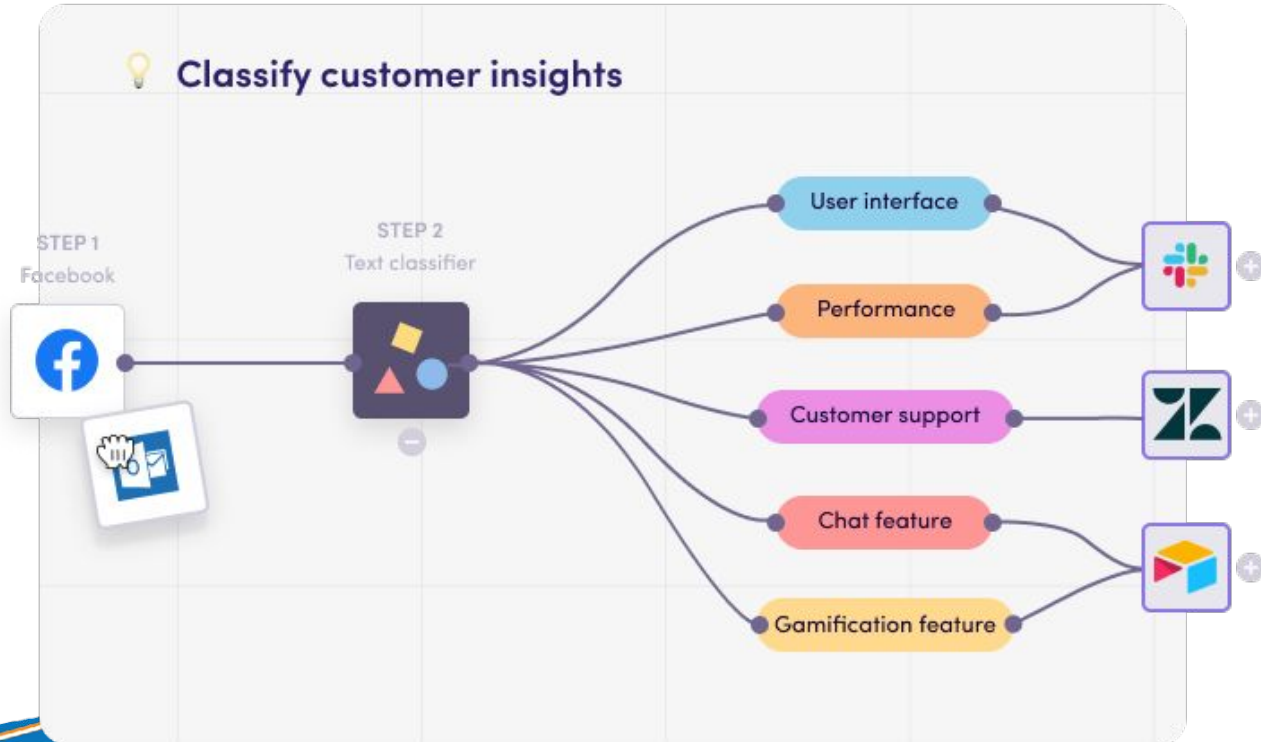
Information Retrieval



Chatbots
& Virtual Assistants

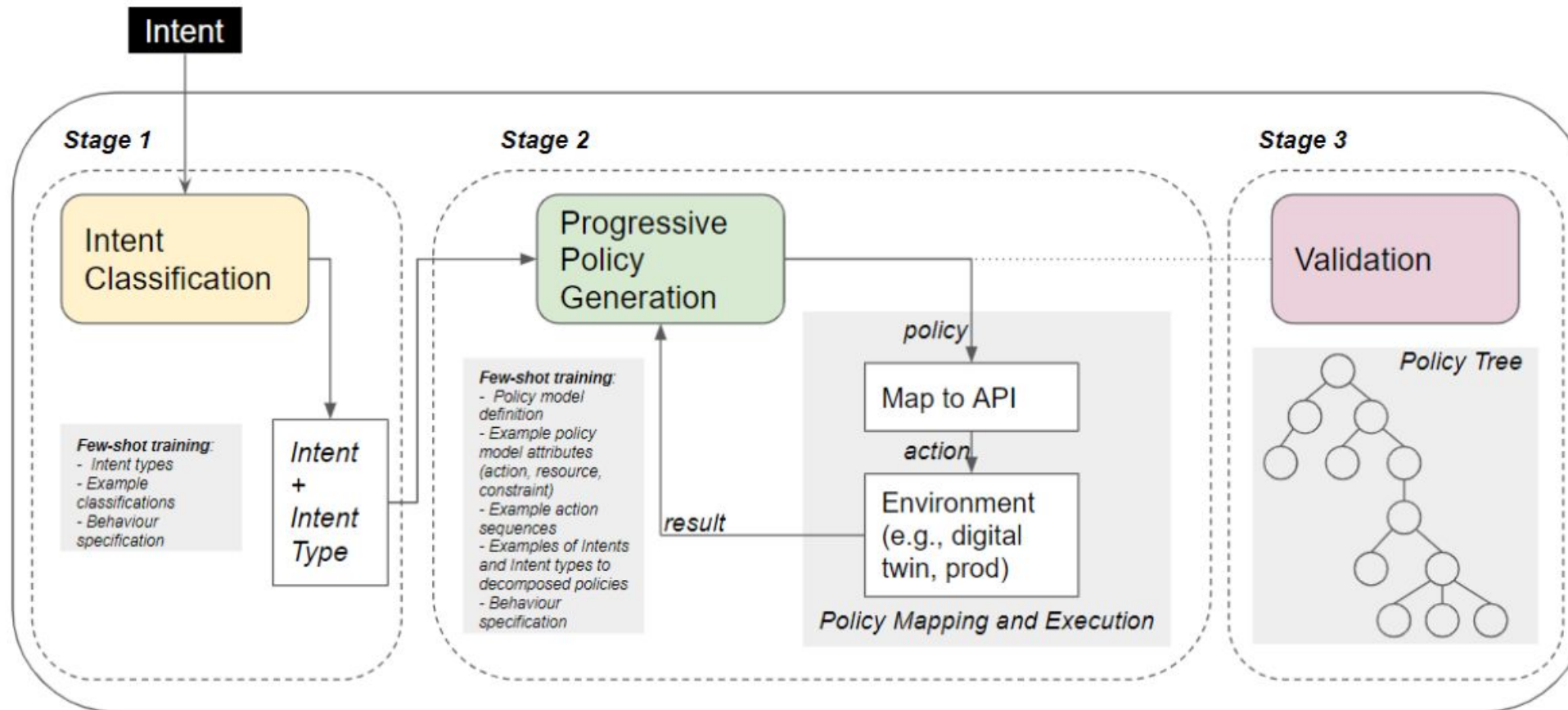
Lecture 3. Text Classification

Common Applications



Lecture 3. Text Classification

Common Applications



Lecture 4. Text Similarity

Lecture 5. Text Summarization

Lecture 6. Machine Translation

Lecture 7. Question-Answering



Lecture 8. Robustness in NLP

Article: Super Bowl 50

Paragraph: *“Peyton Manning became the first quarterback ever to lead two different teams to multiple Super Bowls. He is also the oldest quarterback ever to play in a Super Bowl at age 39. The past record was held by John Elway, who led the Broncos to victory in Super Bowl XXXIII at age 38 and is currently Denver’s Executive Vice President of Football Operations and General Manager. Quarterback Jeff Dean had jersey number 37 in Champ Bowl XXXIV.”*

Question: *“What is the name of the quarterback who was 38 in Super Bowl XXXIII?”*

Original Prediction: John Elway

Prediction under adversary: Jeff Dean

Lecture 8. Robustness in NLP

Article: **Nikola Tesla**

Paragraph: "In January 1880, two of Tesla's uncles put together enough money to help him leave Gospić for **Prague** where he was to study. Unfortunately, he arrived too late to enroll at Charles-Ferdinand University; he never studied Greek, a required subject; and he was illiterate in Czech, another required subject. Tesla did, however, attend lectures at the university, although, as an auditor, he did not receive grades for the courses."

Question: "What city did Tesla move to in 1880?"

Answer: **Prague**

Model Predicts: **Prague**

AddAny

Randomly initialize d words:

spring attention income **getting** reached

↓ Greedily change one word

spring attention income **other** reached

↓ Repeat many times

Adversary Adds: **tesla move move other george**

Model Predicts: **george**

AddSent

What city did **Tesla** move to in **1880**?

Prague

(Step 1)
Mutate
question

(Step 2)
Generate
fake answer

What city did **Tadakatsu** move to in **1881**?

Chicago

(Step 3)
Convert into
statement

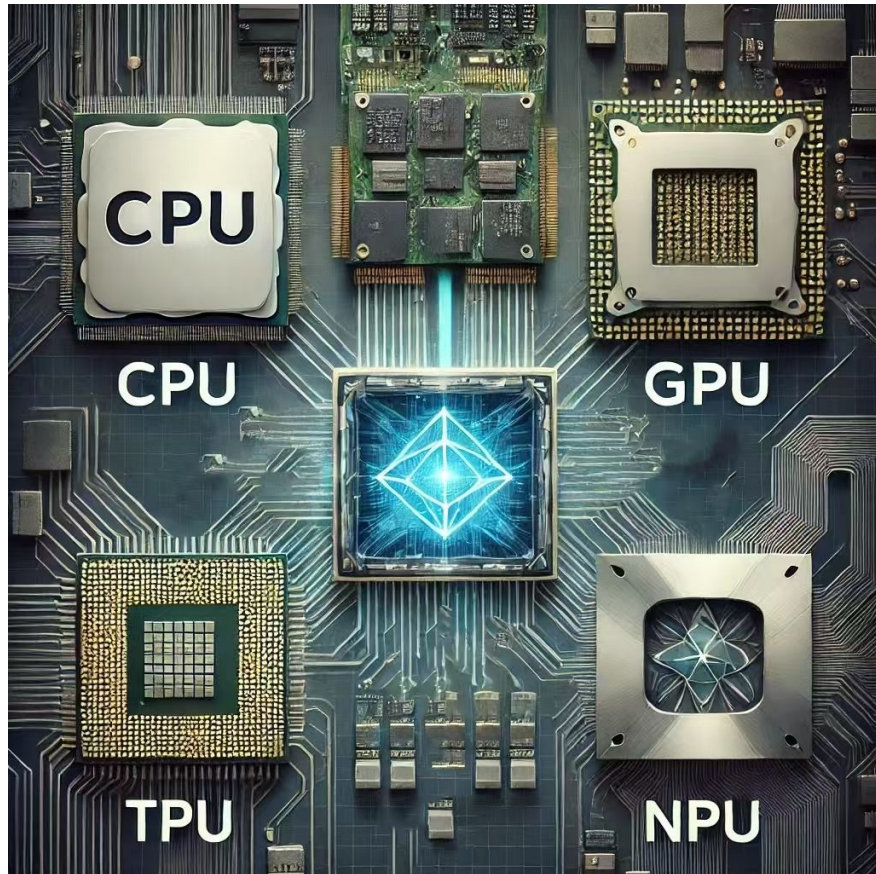
Tadakatsu moved the city of **Chicago** to in **1881**.
























(Step 4)
Fix errors with
crowdworkers,
verify resulting
sentences with
other crowdworkers

Adversary Adds: **Tadakatsu moved to the city of Chicago in 1881.**

Model Predicts: **Chicago**

Lecture 9. Hardware accelerated



						
CPU	GPU		TPU	NPU		
						
						
						

Lecture 9. Building NLP Applications

- ❑ Sample Applications
- ❑ Notes in Building NLP Applications



NLPA - Course Introduction

REFERENCES



Tài liệu tham khảo

- ❑ Vietnamese:
 - ❑ **Đinh Điền**, Xử Lý Ngôn Ngữ Tự Nhiên, 2006, NXB ĐHQG.
- ❑ English:
 - ❑ **Chris Manning and Hinrich Schütze**, Foundations of Statistical Natural Language Processing, MIT Press. Cambridge, MA: May 1999.
 - ❑ **Dan Jurafsky and James H. Martin**, Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Third Edition draft, 2020.
 - ❑ **Kevyn Collins-Thompson**, Computational Assessment of Text Readability, 2014.
 - ❑ **Jiapeng Wang and Yihong Dong**, Measurement of Text Similarity: A Survey, Information 2020.

