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

fit@hcmus

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

fit@hcmus

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

NLPA - Course Introduction

COURSE EVALUATION

fit@hcmus

Course Evaluation

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

fit@hcmus

NLPA - Course Introduction

COURSE OUTLINE

fit@hcmus

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

fit@hcmus

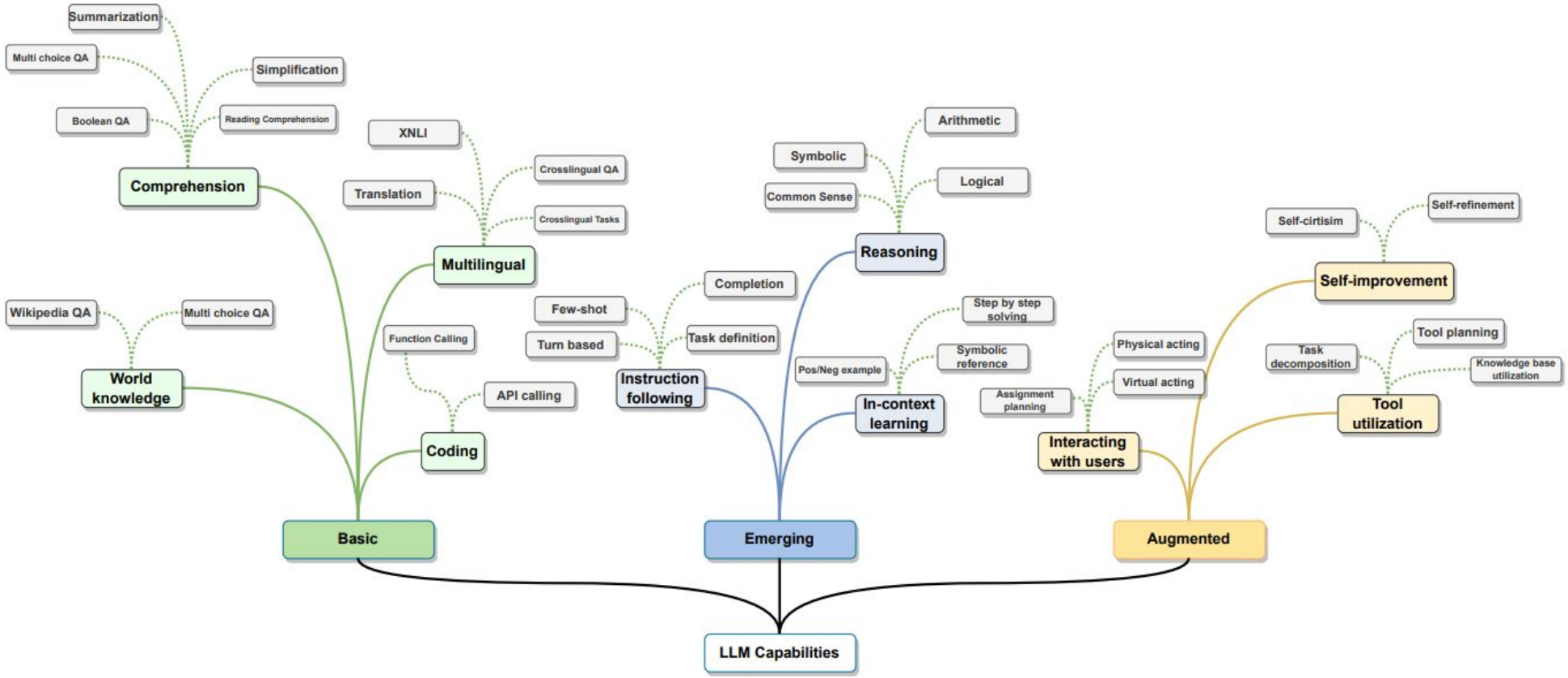
Lecture 2. Language Models

- 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

fit@hcmus

Lecture 2. Language Models

- Capabilities of LLMs

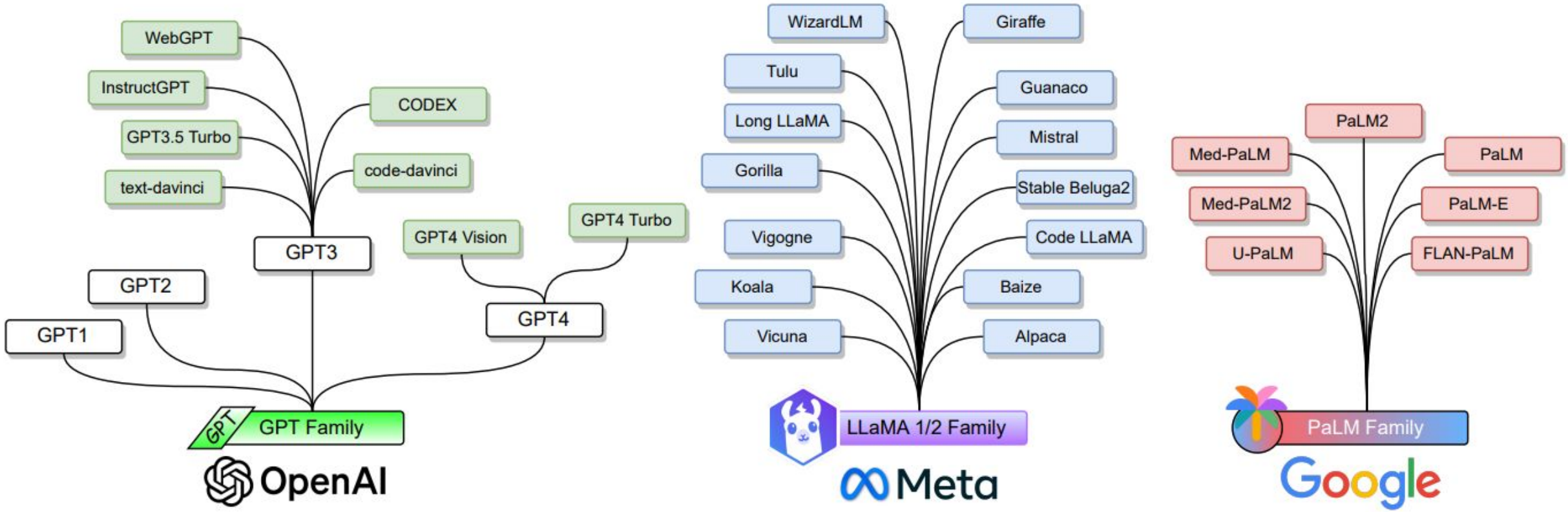


Nguồn: https://arxiv.org/pdf/2402.06196

fit@hcmus

Lecture 2. Language Models

- Major LLM Examples

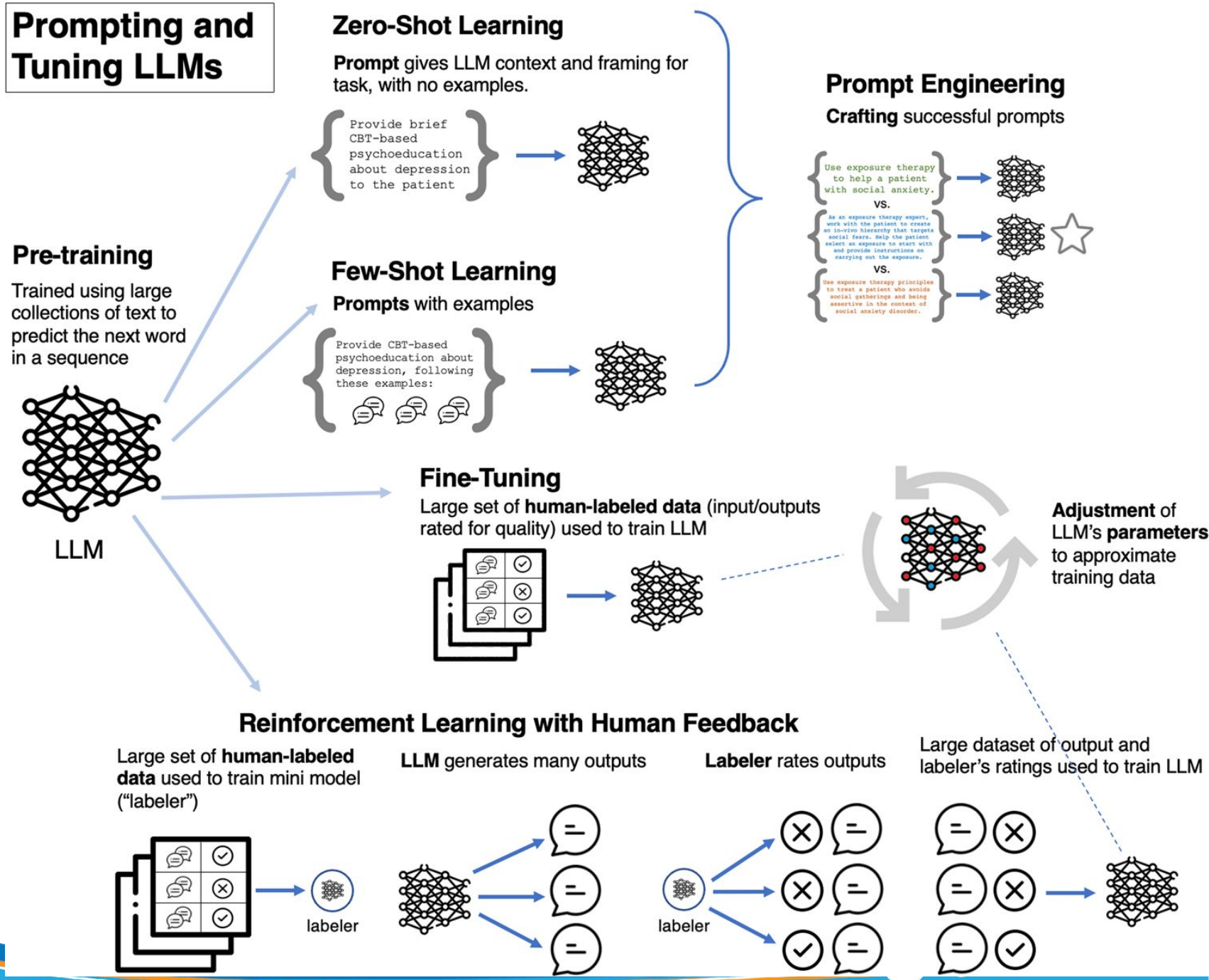


Nguồn: https://arxiv.org/pdf/2402.06196

fit@hcmus

Prompting and Tuning LLMs

- How LLMs Work

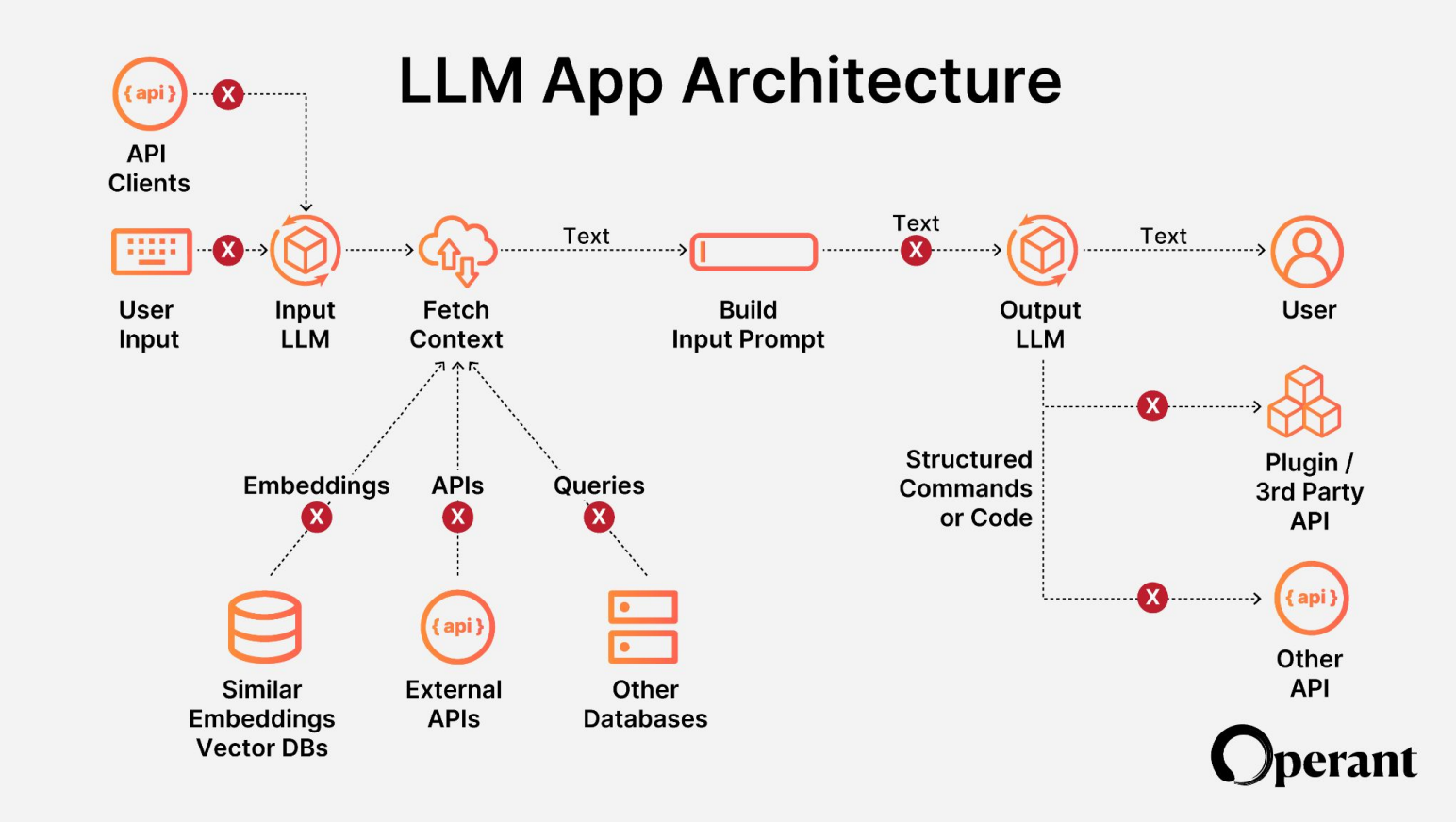


Nguồn: https://arxiv.org/pdf/2402.06196

fit@hcmus

Lecture 2. Language Models

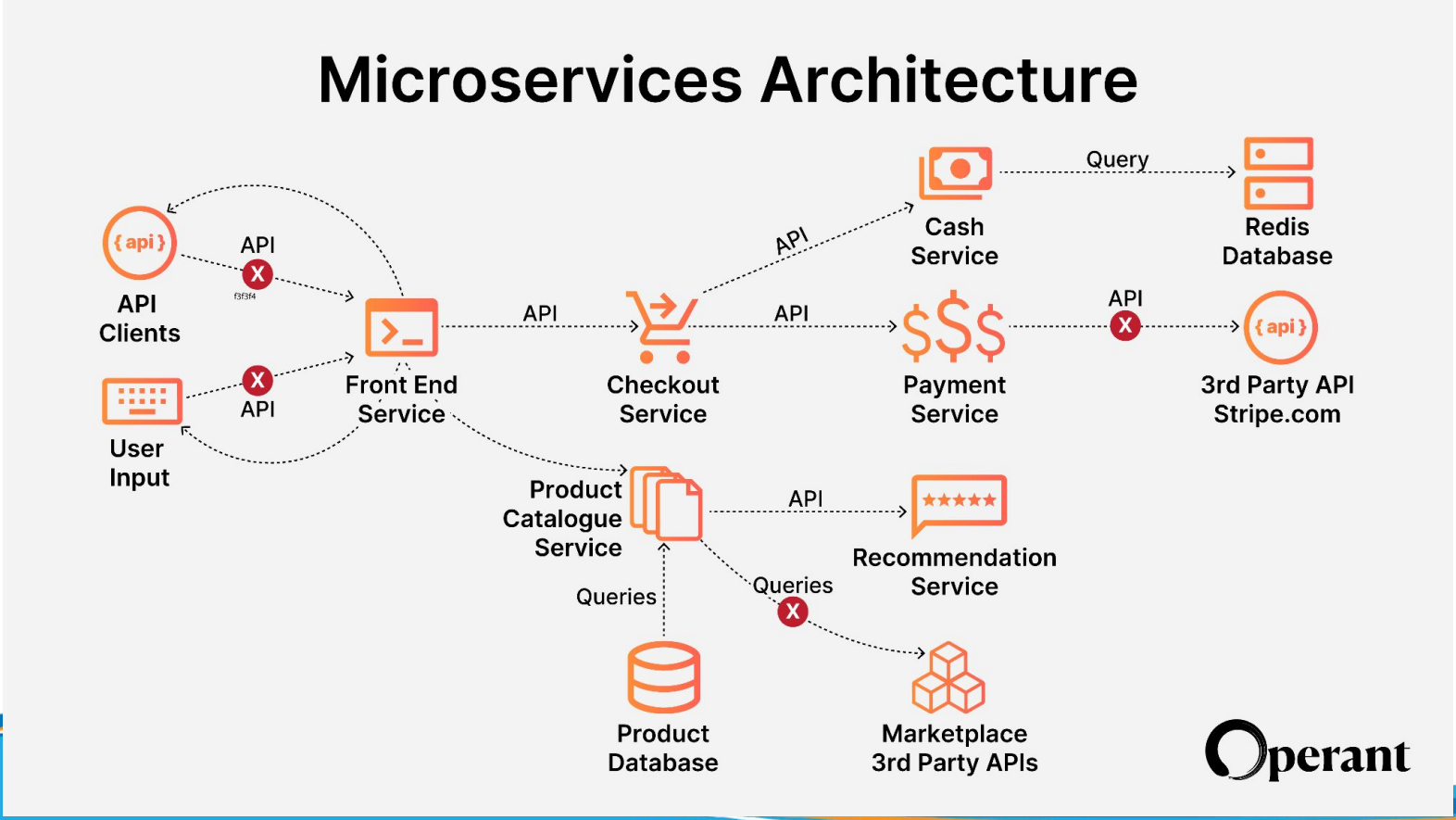
- Applications in Real Life



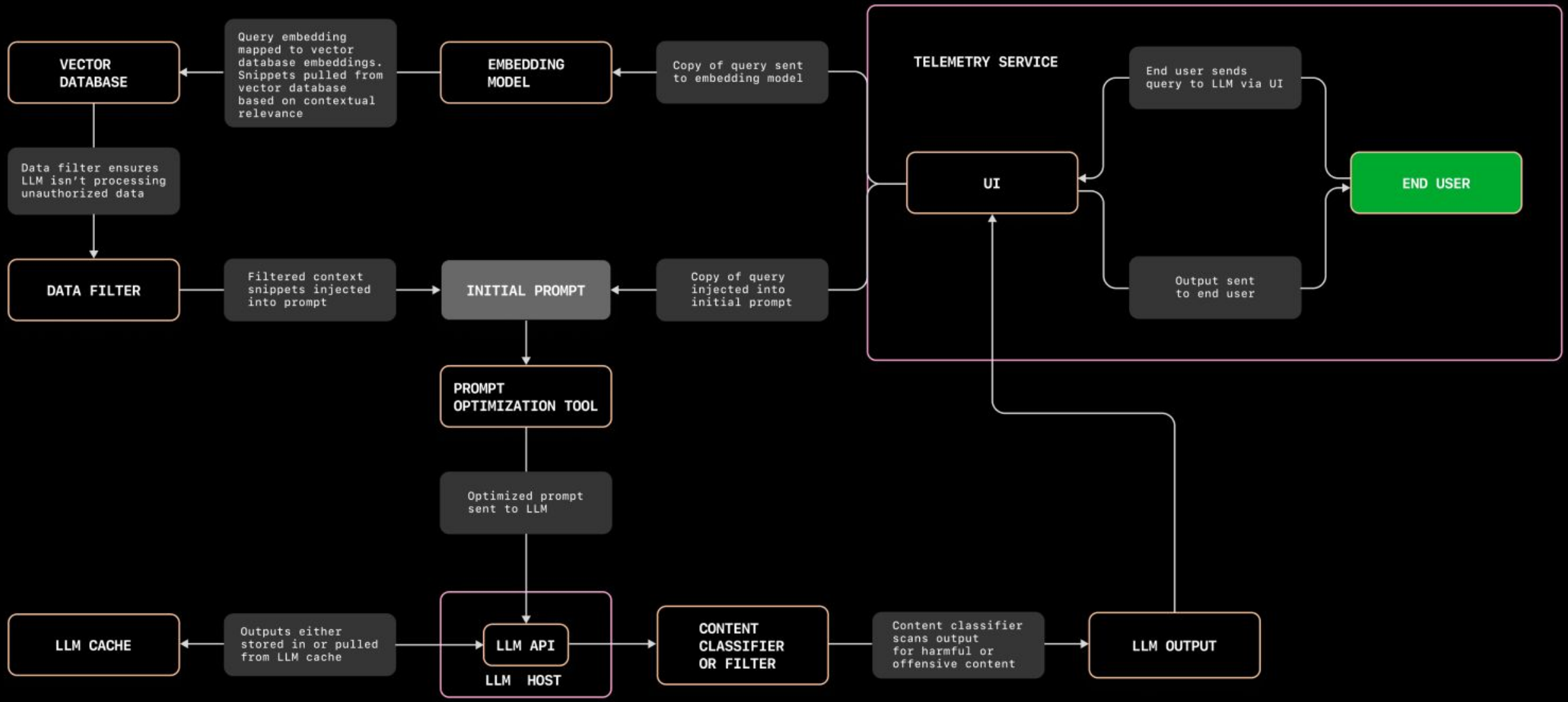
Nguồn: https://www.operant.ai/art-kubed/defense-against-the-cyber-dark-arts-managing-llms-as-if-they-were-ai-house-elves

Lecture 2. Language Models

- Applications in Real Life



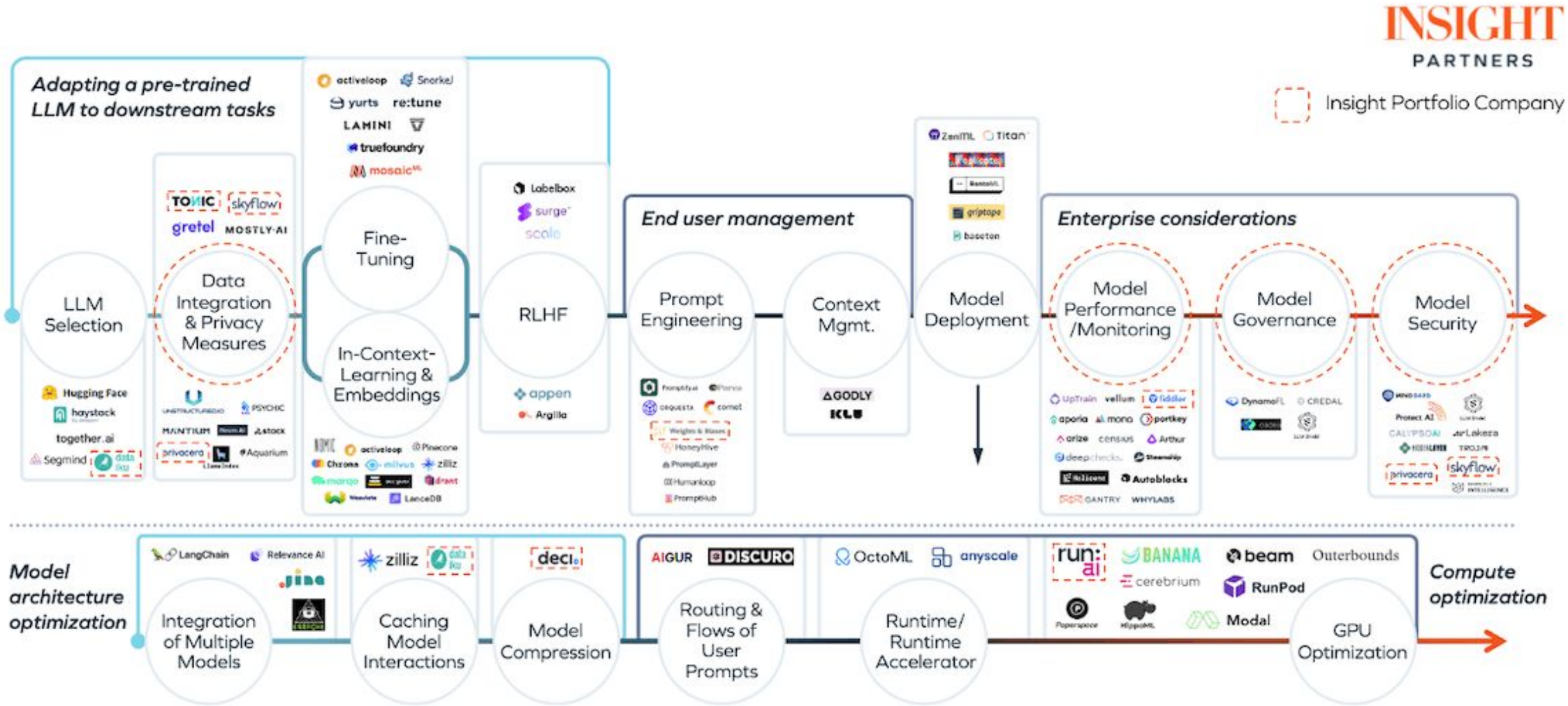
Nguồn: https://www.operant.ai/art-kubed/defense-against-the-cyber-dark-arts-managing-llms-as-if-they-were-ai-house-elves



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.

Nguồn: https://github.blog/ai-and-ml/llms/the-architecture-of-todays-llm-applications/

LLMOps



Nguồn: https://www.leanix.net/en/blog/llmops-enterprise-architecture

fit@hcmus

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

fit@hcmus

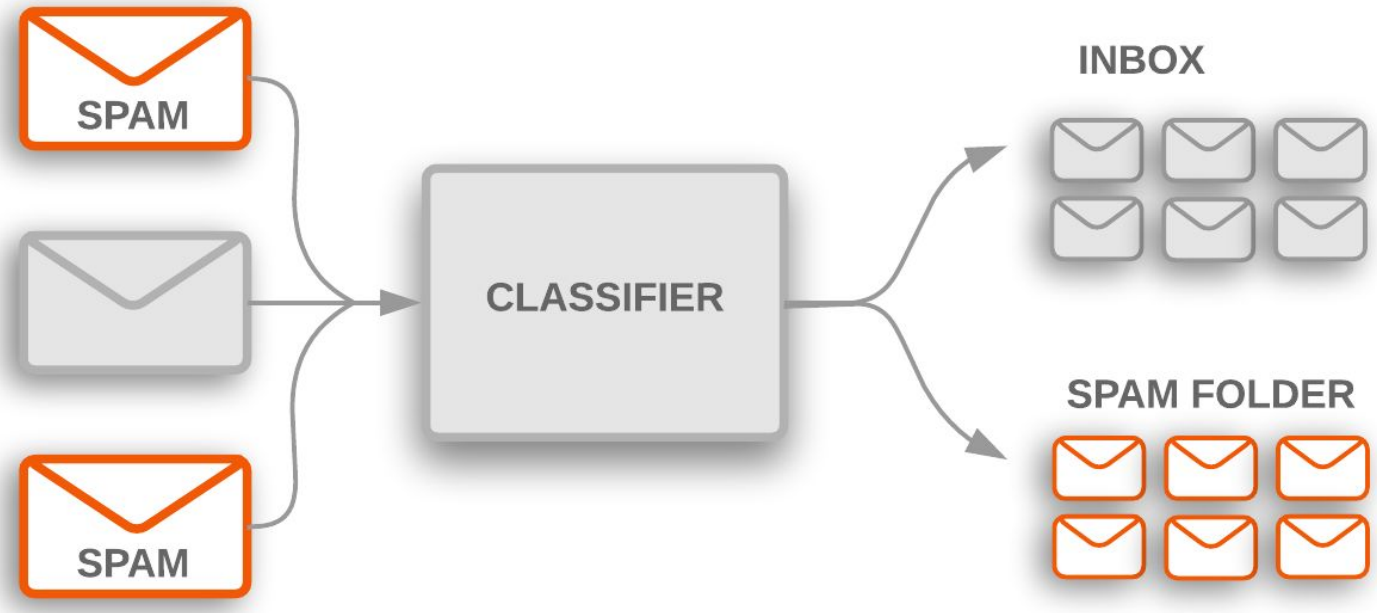
Lecture 3. Text Classification

- Introduction
- Common Applications
- Future Trends

fit@hcmus

Lecture 3. Text Classification

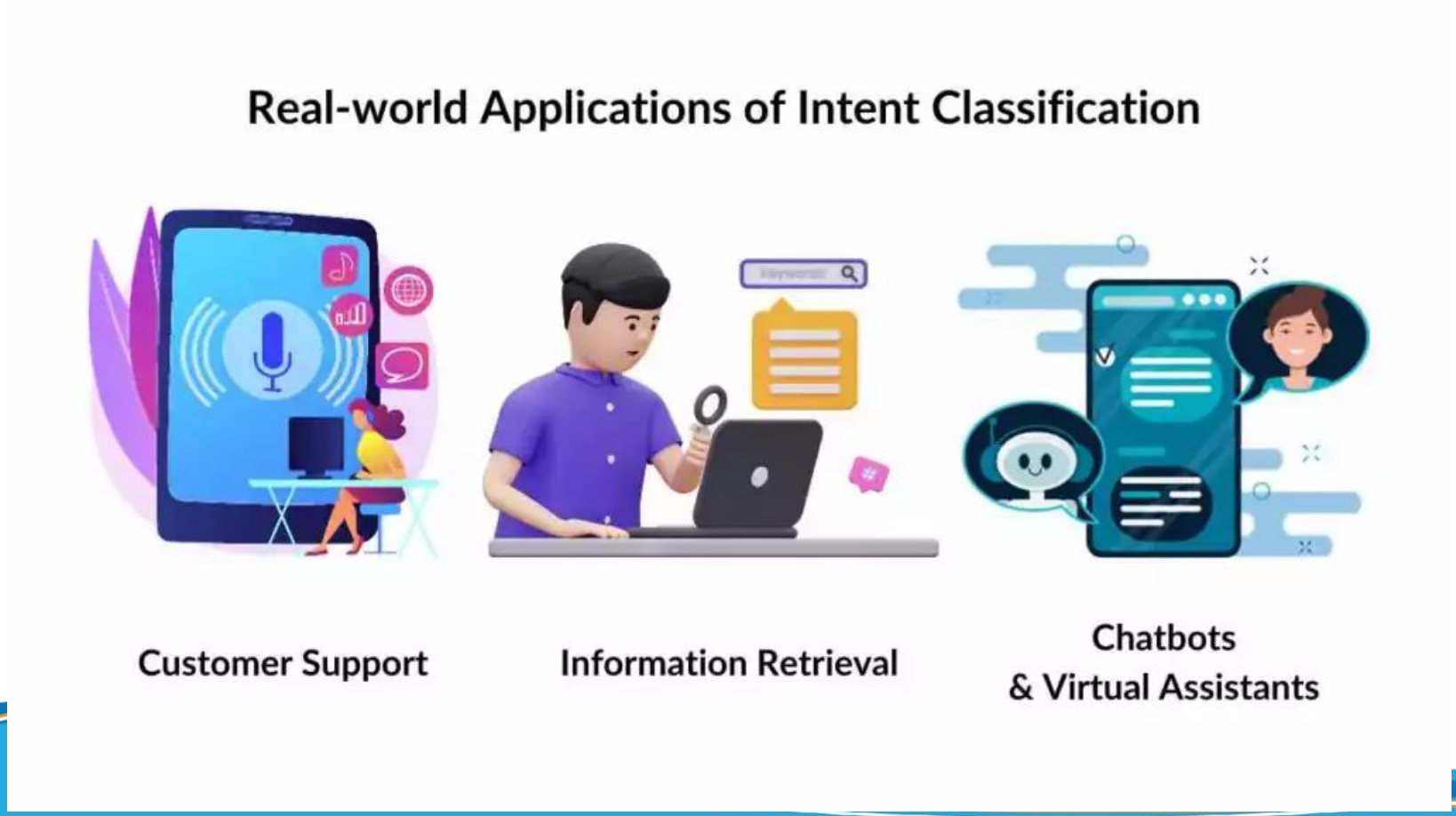
- Introduction



fit@hcmus

Lecture 3. Text Classification

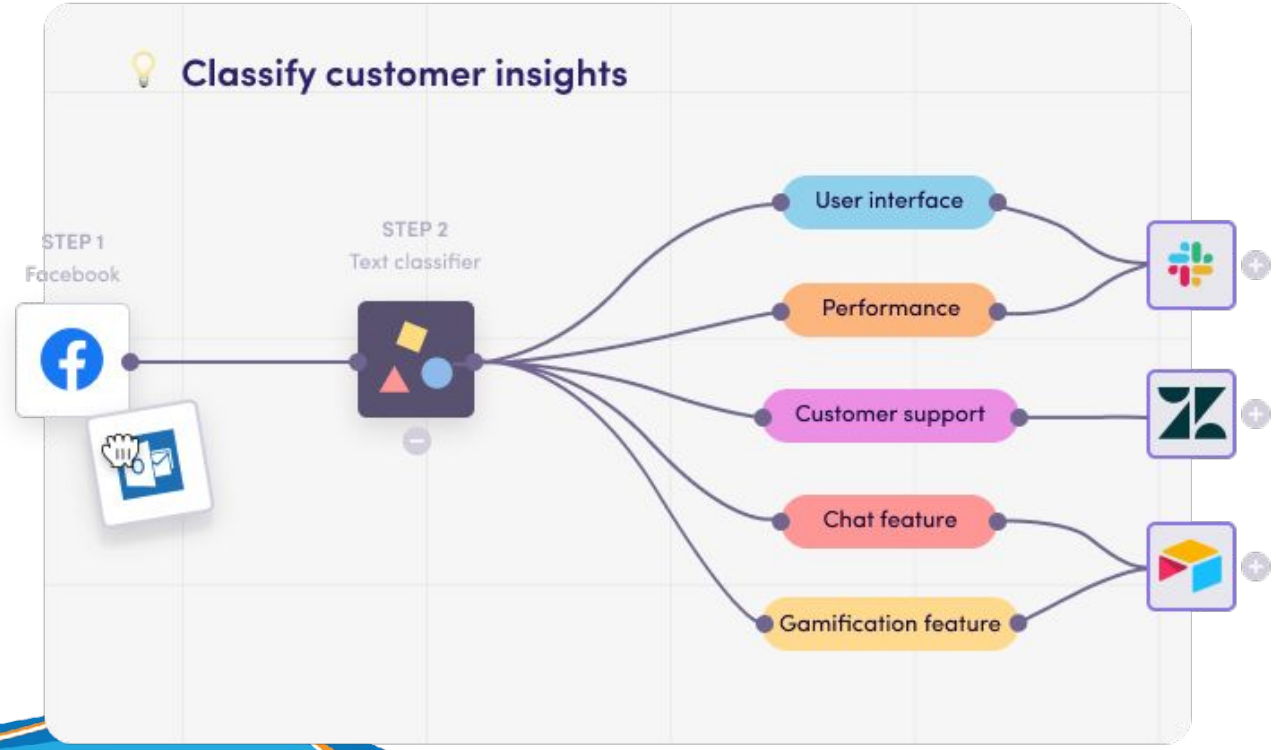
- Common Applications



fit@hcmus

Lecture 3. Text Classification

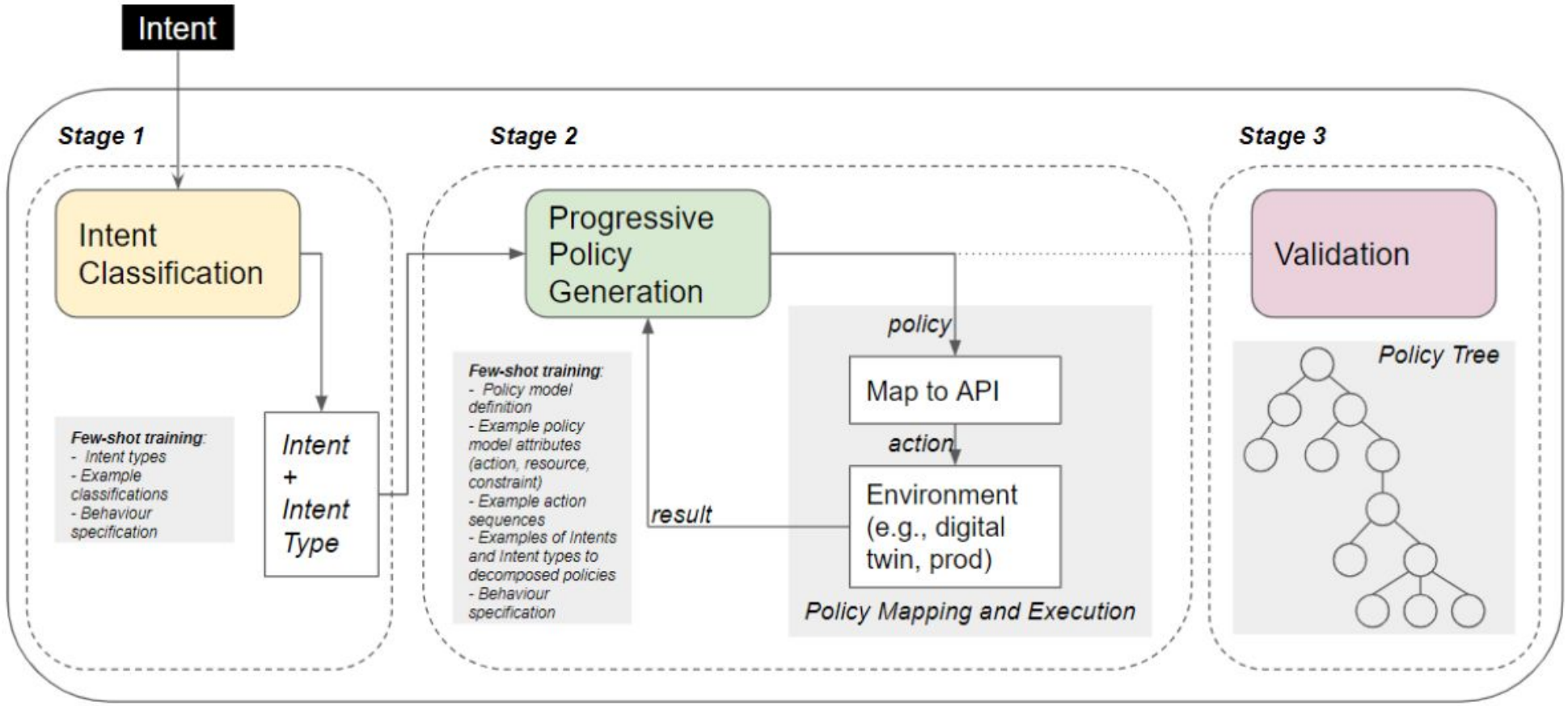
- Common Applications



fit@hcmus

Lecture 3. Text Classification

- Common Applications



Nguồn: https://arxiv.org/html/2402.10067v1

fit@hcmus

Lecture 4. Text Similarity

Lecture 5. Text Summarization

Lecture 6. Machine Translation

Lecture 7. Question-Answering

fit@hcmus

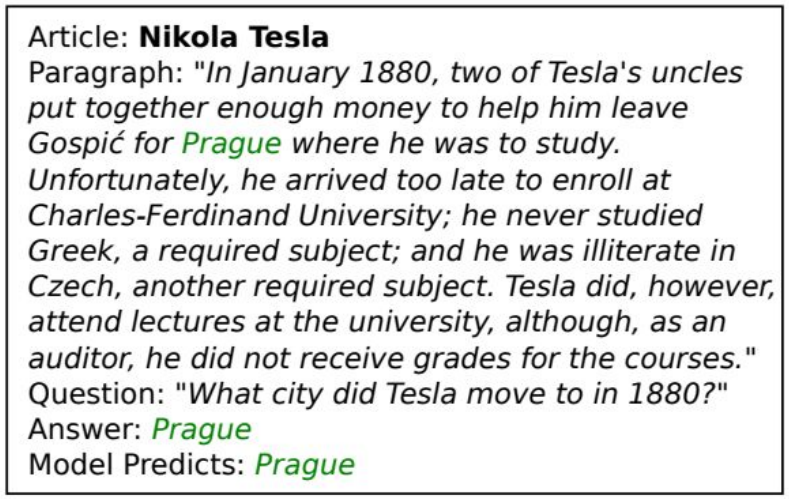
Lecture 8. Robustness in NLP

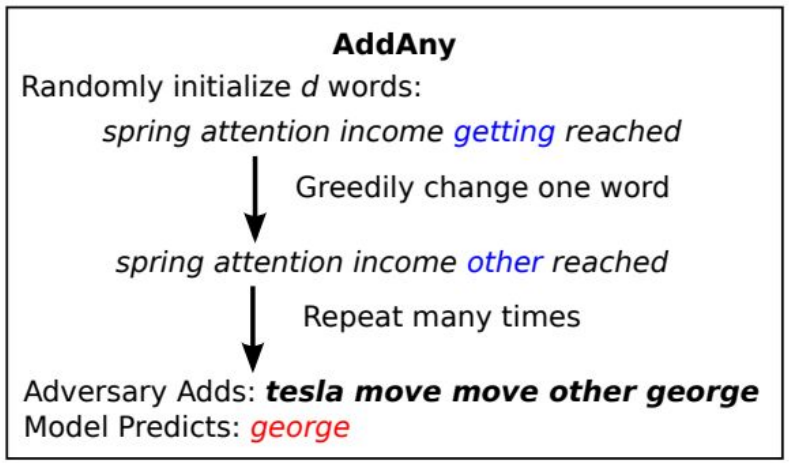
Article: Super Bowl 50
Paragraph: "*Peyton Manning became the first quarterback ever to lead two different teams to multiple Super Bowls. He is also the oldest quarterback ever to play in a Super Bowl at age 39. The past record was held by John Elway, who led the Broncos to victory in Super Bowl XXXIII 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

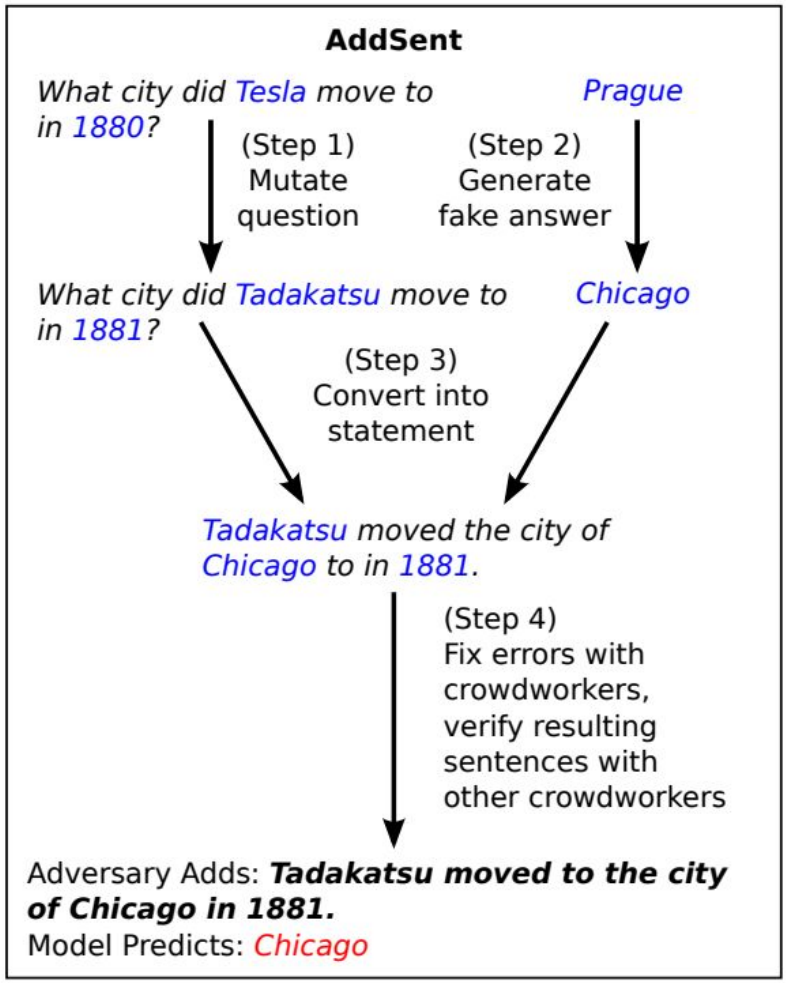
Nguồn: https://arxiv.org/pdf/1707.07328

fit@hcmus

Lecture 8. Robustness in NLP



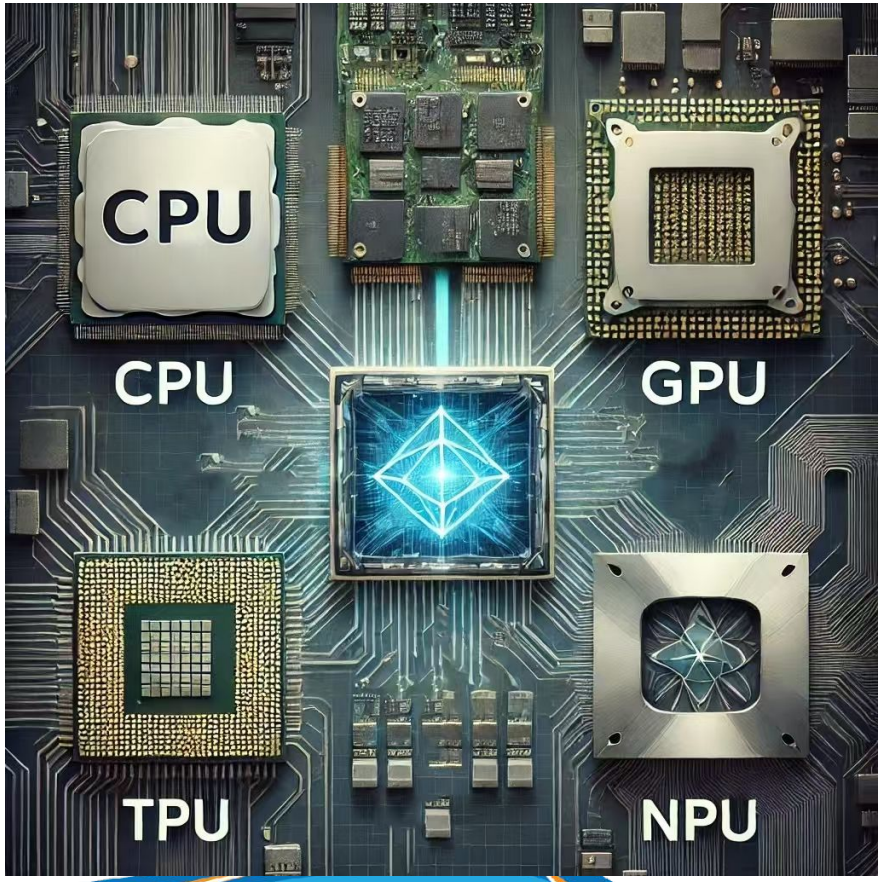




Nguồn: https://arxiv.org/pdf/1707.07328

fit@hcmus

Lecture 9. Hardware accelerated



CPU	GPU	TPU	NPU
			

Nguồn: https://www.seeedstudio.com/blog/2024/08/12/cpu-vs-gpu-vs-tpu-vs-npu/

fit@hcmus

Lecture 9. Building NLP Applications

- Sample Applications
- Notes in Building NLP Applications

NLPA - Course Introduction

REFERENCES

fit@hcmus

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.
  - J**iapeng Wang and Yihong Dong**, Measurement of Text Similarity: A Survey, Information 2020.