From Tools to Agents:
Key Technologies for Language-Driven Action **Agents**

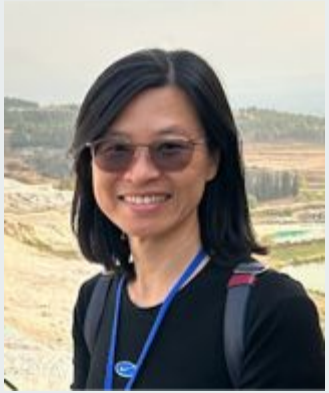
Prof. Chia-Hui Chang (張嘉惠)

National Central University, Taiwan

Convenor for the Intelligence Computing Discipline @ NSTC

Co-Chair for AI Center of Excellence @ NSTC

Jan. 20, 2025



About Me

- Dept. of CSIE, NCU
- Office of Institutional Research, NCU
- **Convenor for the Intelligence Computing discipline @ NSTC**
- **Co-Chair for AI Center of Excellence @ NSTC**

國立中央大學
National Central University

NSTC 國家科學及技術委員會
National Science and Technology Council

https://www.aclclp.org.tw/

中華民國人工智慧學會
Taiwanese Association for Artificial Intelligence

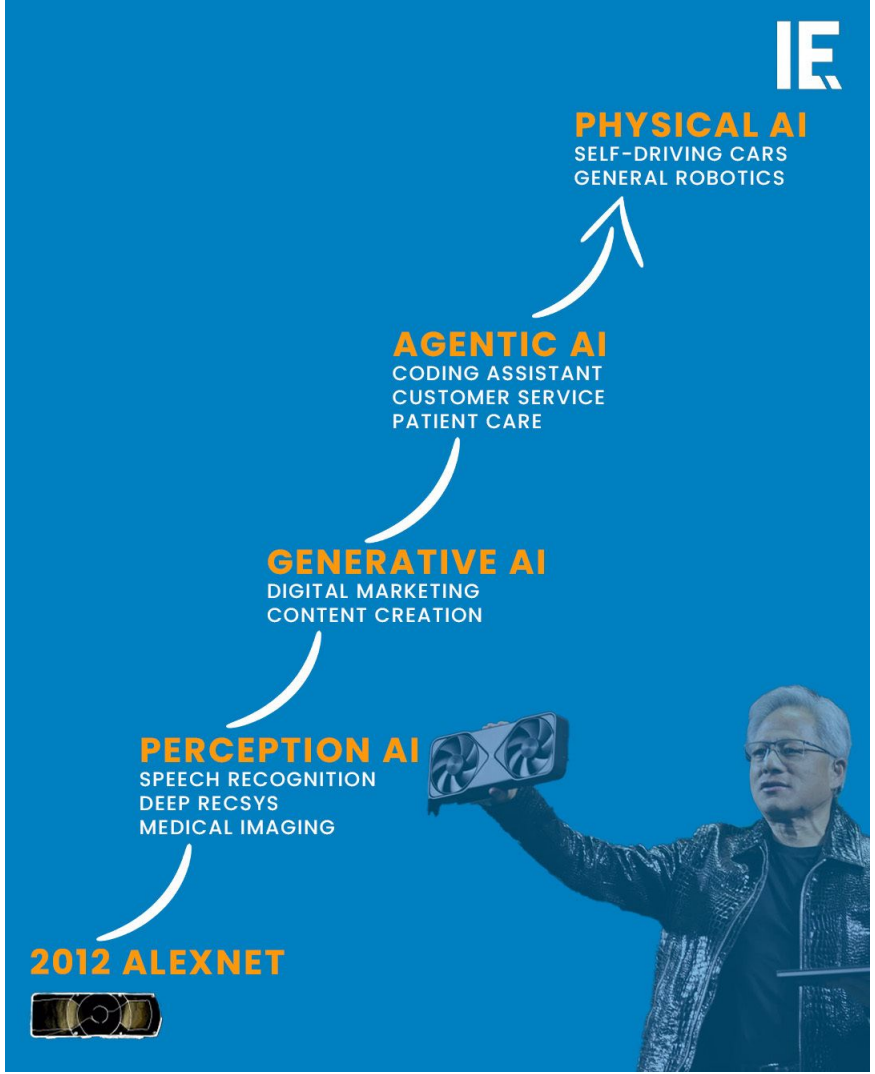
https://www.taai.org.tw/

Outline

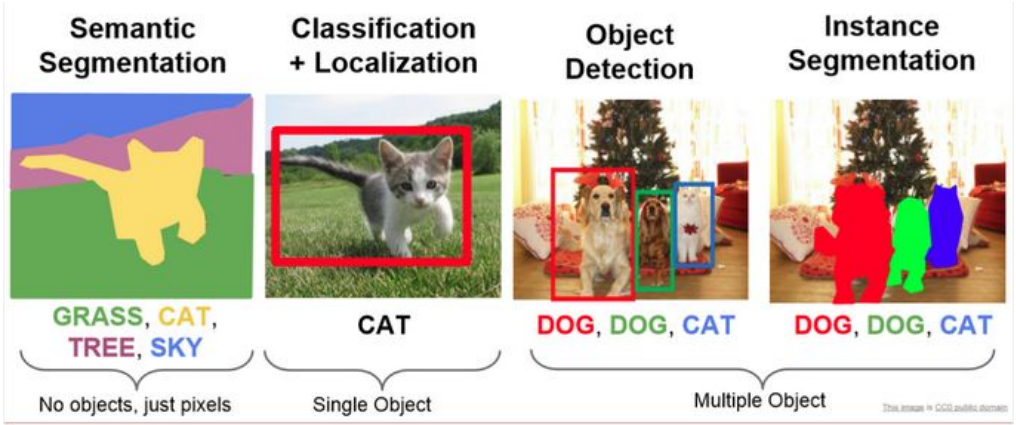
- Background
  - 4 Phases of AI
- **What are AI Agents?**
  - Example: GUI Agents, Robots
  - Agents for Software Development
- Course Administration, Virtual TA by EduACT
- Your turn

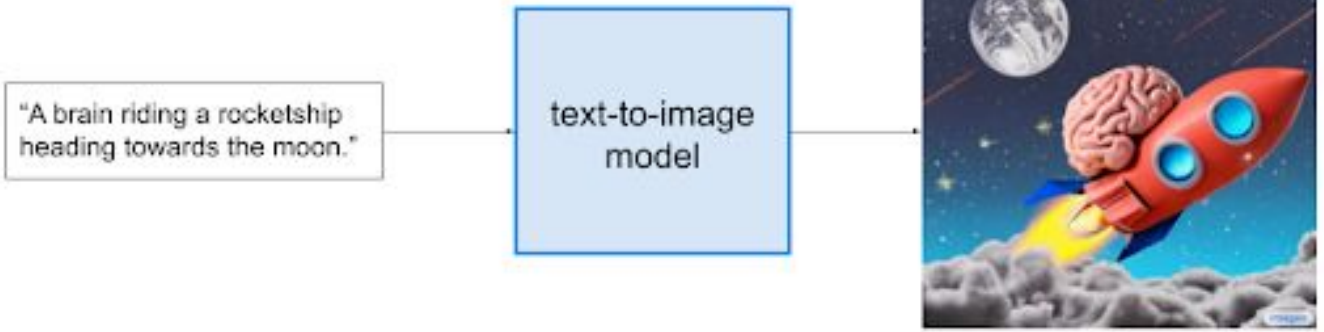
The Evolution of AI: From AlexNet to Reality

NVIDIA CEO Jensen Huang Keynote at CES 2025



Vision: Perception vs. Generative





Natural Language Understanding

Elon Musk PERSON apparently wasn't aware that his company SpaceX had a Facebook ORG page. The SpaceX and Tesla PRODUCT CEO has responded to a comment on Twitter GPE calling for him to take down the SpaceX, Tesla and Elon Musk ORG official pages in support of the #deletefacebook movement by first ORDINAL acknowledging he didn't know one existed, and then following up with promises that he would indeed take them down.

He's done just that, as the SpaceX NORP Facebook page is now gone, after having been live earlier today DATE (as you can see from the screenshot included taken at around 12:10 PM ET) TIME .

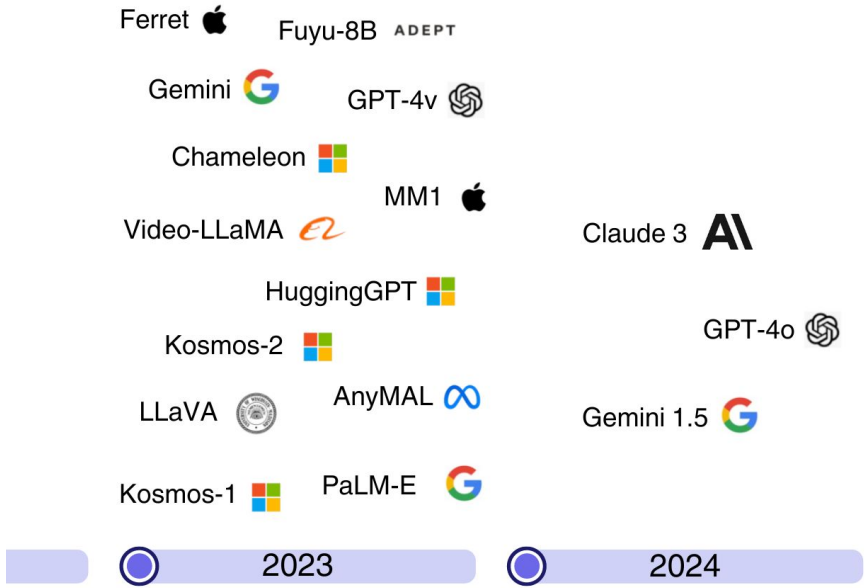
- Named Entity Recognition
- Relation Extraction
- Event Extraction
- Sentiment Analysis

- Part of Speech Tagging
- Chunking
- Dependency parsing

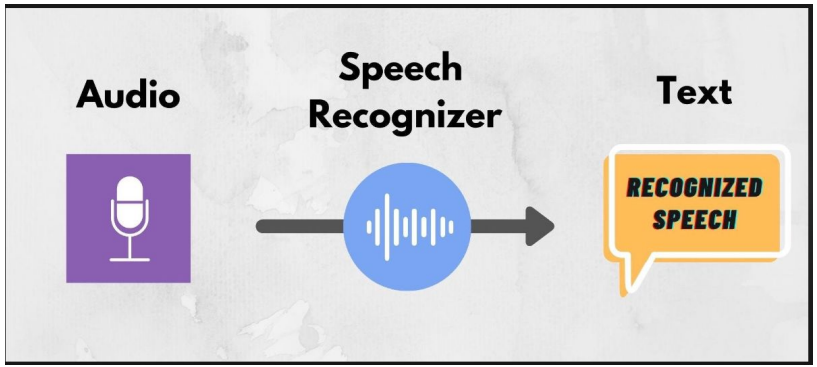
Natural Language Generation

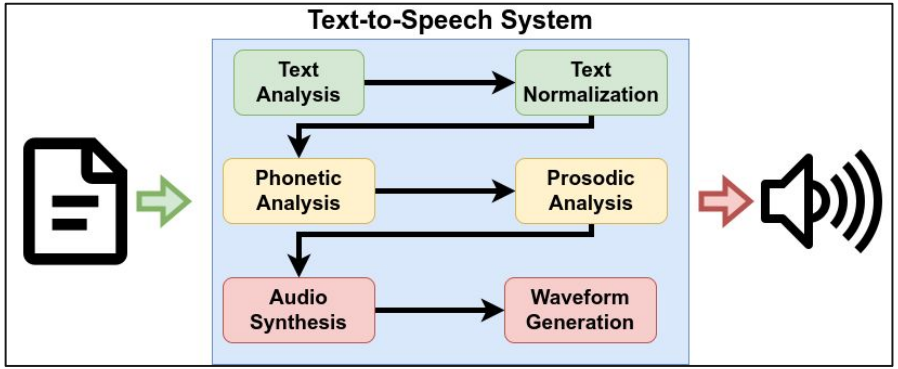
Conversation extract:

User	*I want to review a film*
LUCID	*Sure, I'd be happy to help with that. Can you please tell me the name of the film you'd like to review?*
User	*Bobby, stop playing with your food and eat your dinner!*
LUCID	*It seems like you're busy at the moment. Let's continue this conversation after dinner.*
User	*Sorry about that, my son was misbehaving. The film I want to review is The Dark Knight*
LUCID	*Great, you would like to review "The Dark Knight". Is that correct?*
User	*Yes, that's correct. My review is: I found the storyline of this movie a bit hard to follow.*

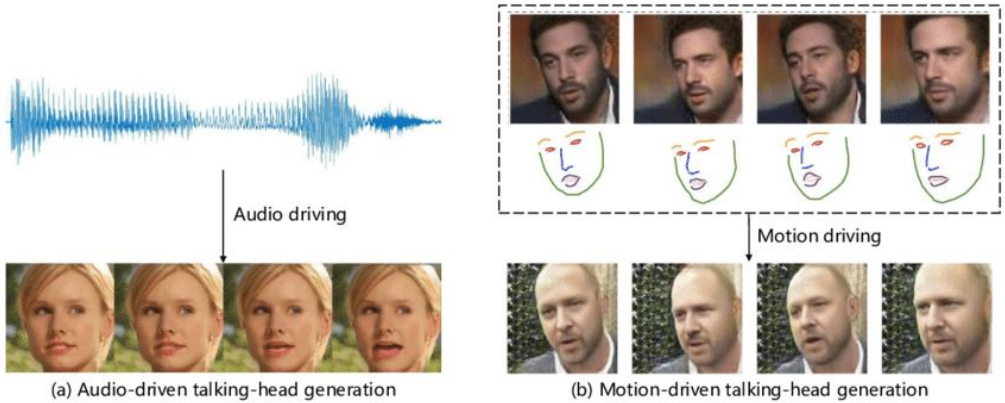


Speech: Recognition vs. Synthesis



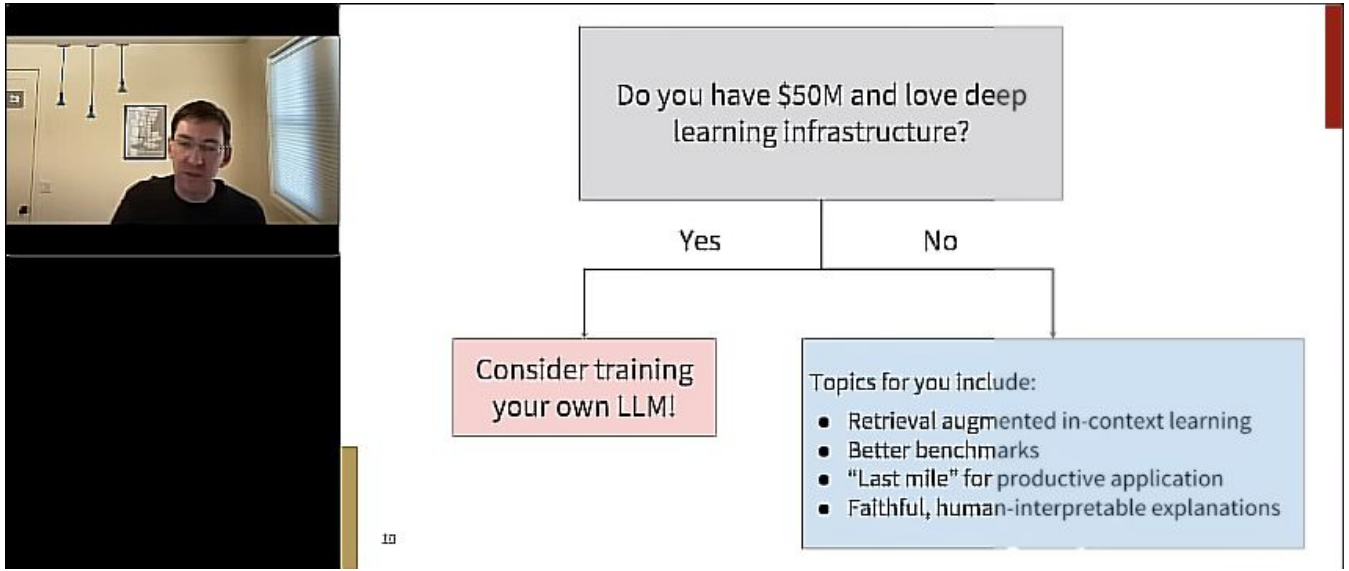


With image generation



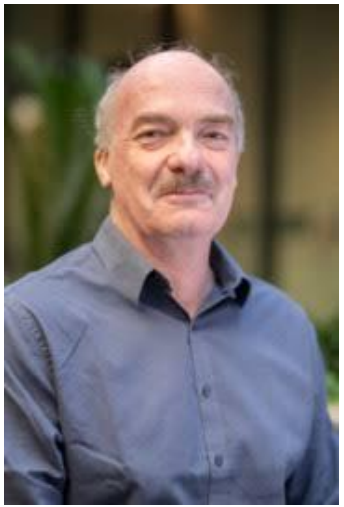
What we can contribute to NLU after GPT3?

Christopher Potts@Stanford



What are the Last Mile productive applications?

We're in a GenAI world … let's do responsible work

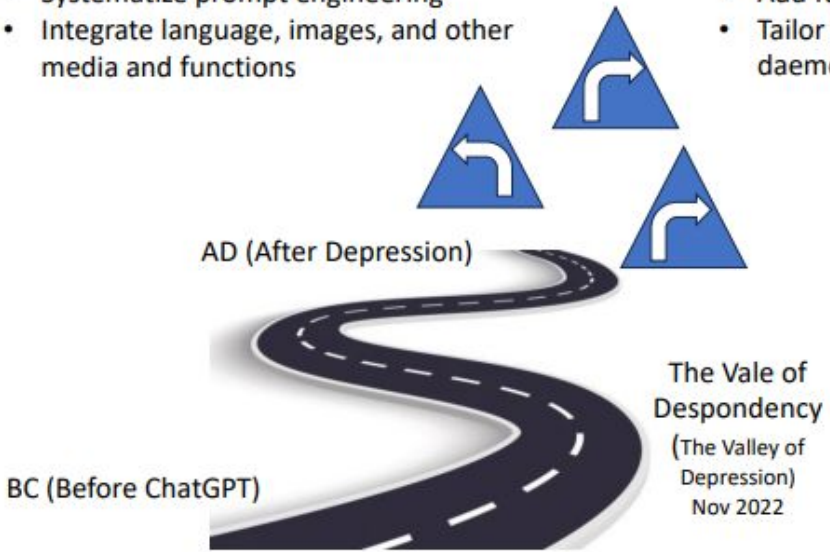


NLP engineering: Make LLMs usable

- Build smaller and cheaper LLMs
- Systematize prompt engineering
- Integrate language, images, and other media and functions

NLP applications: Make LLMs useful

- Tune LLMs to domains and companies for enterprise processing
- Add functionality and agency in the world
- Tailor LLMs to people to be their personal daemons/amanuenses in everyday life

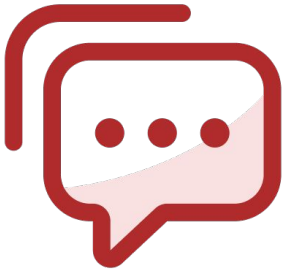


NLP research: Make LLMs understandable

(or at least, be solid engineering)

- Fix the problems with LLMs
- Get explanations how LLMs do what they do
- Formalize them well enough for autonomy, assurance, and ethics

Prof. Eduard Hovy University of Melbourne and CMU



Audience Q&A

AI Agents

The Rise and Potential of Large Language Model Based Agents: A Survey

https://arxiv.org/pdf/2309.07864

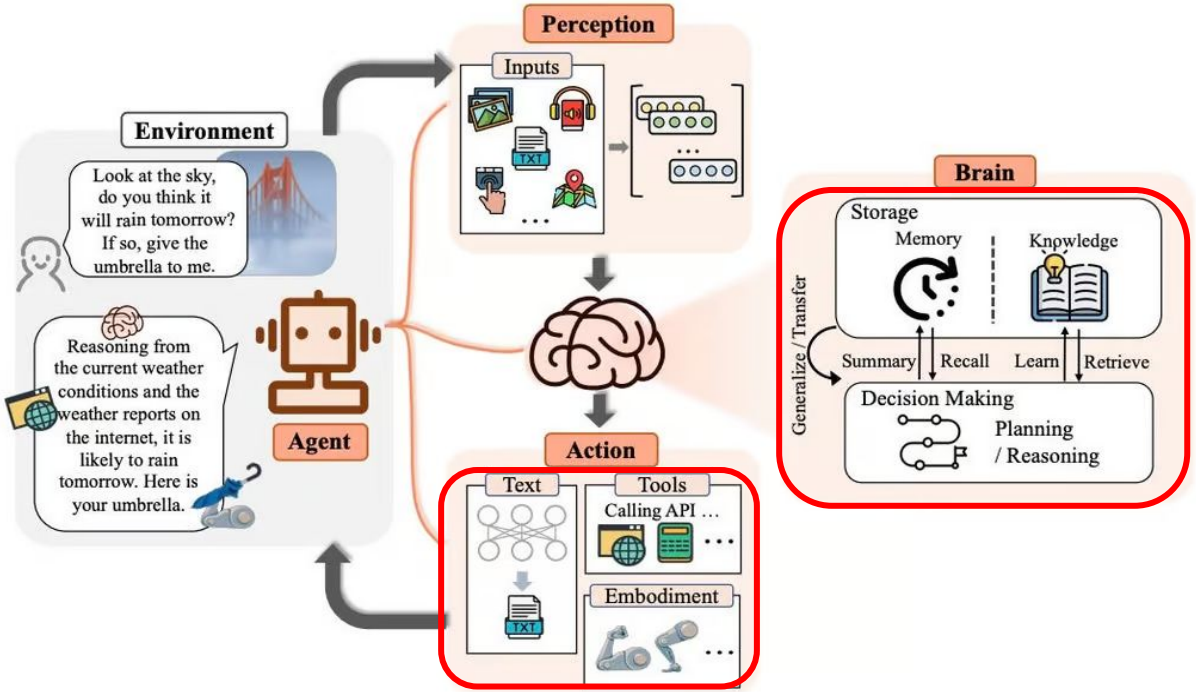
Zhiheng Xi*†, Wenxiang Chen*, Xin Guo*, Wei He*, Yiwen Ding*, Boyang Hong*,
Ming Zhang*, Junzhe Wang*, Senjie Jin*, Enyu

Rui Zheng, Xiaoran Fan, Xiao Wang, Limao Xiong, Yuhao Z
Changhao Jiang, Yicheng Zou, Xiangyang Liu, Zha

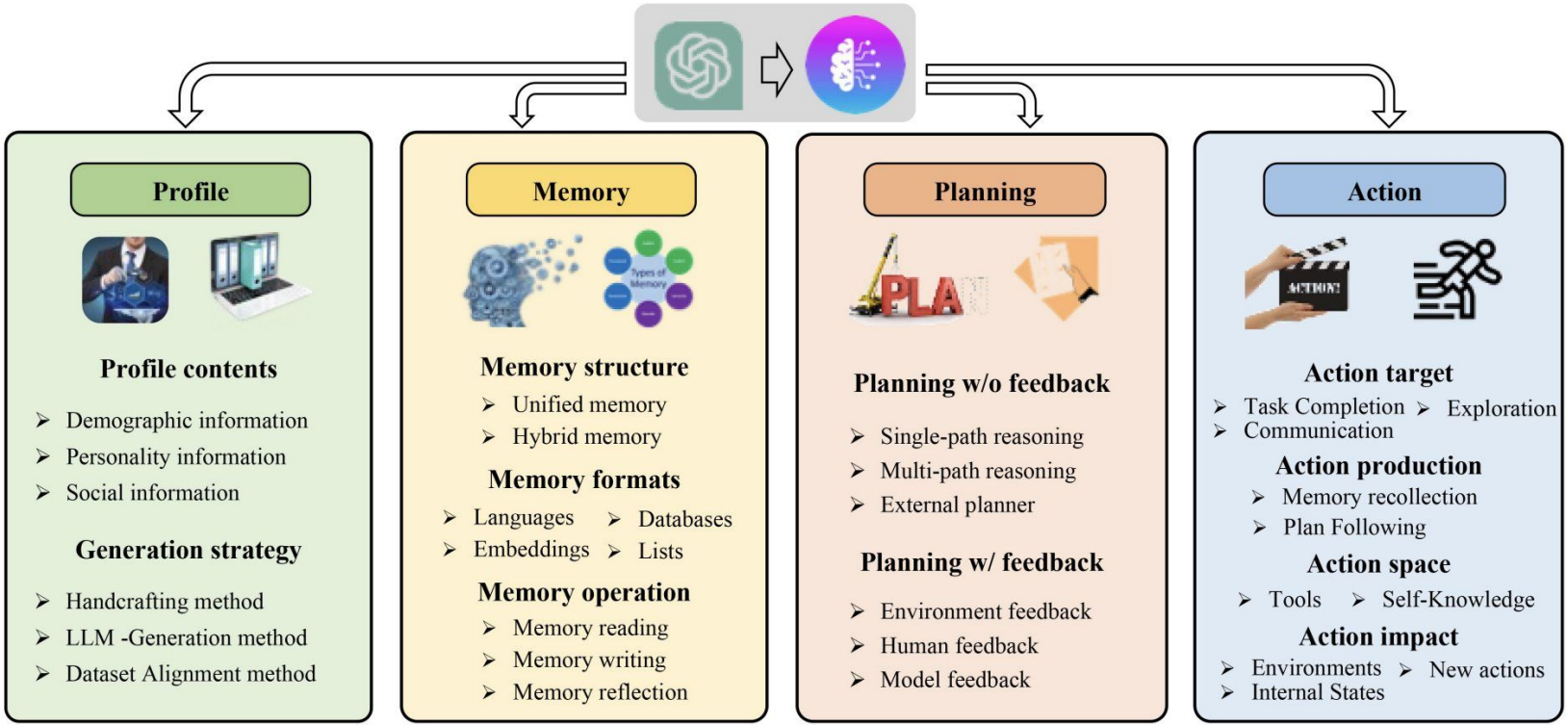
Shihan Dou, Rongxiang Weng, Wensen Che

Qi Zhang†, Wenjuan Qin, Yongyan Zheng, Xipeng Qiu, Xuanjin

Fudan NLP Group



Structure of an Agent

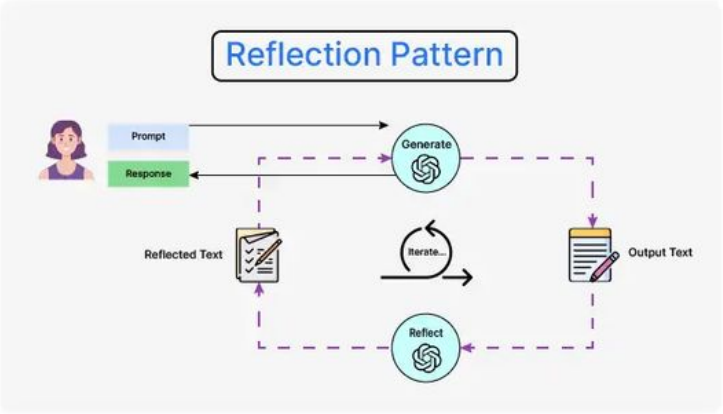


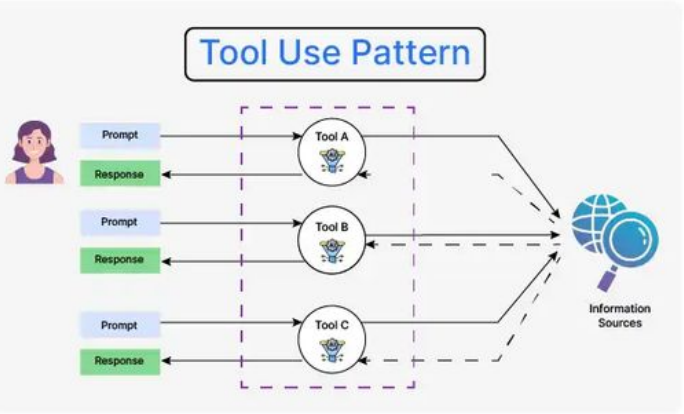
Agentic Reasoning Design Patterns

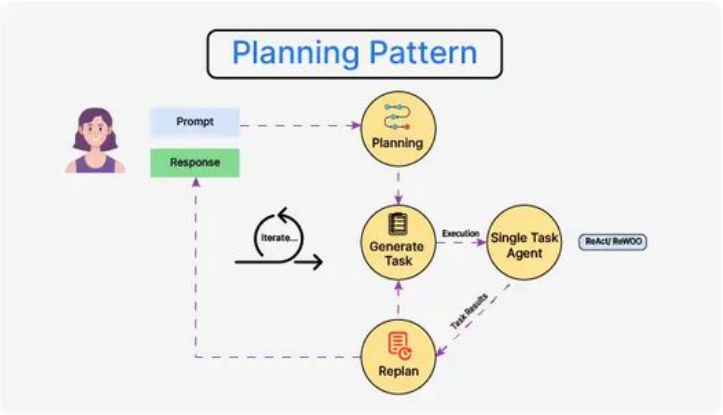
By Andrew Ng @ Sequoia AI Ascent 2024

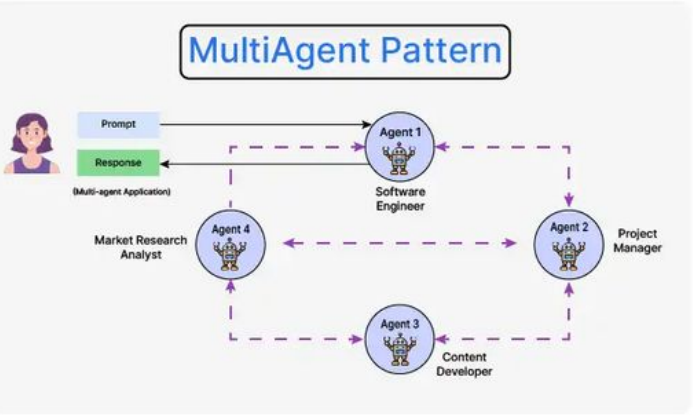
- Planning
- Tool Use
- Reflection
- Multi-agent collaboration

Agentic Design Patterns





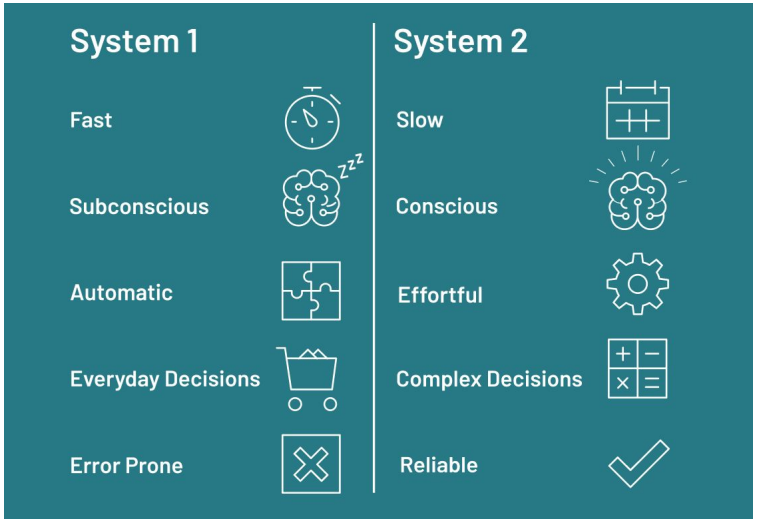


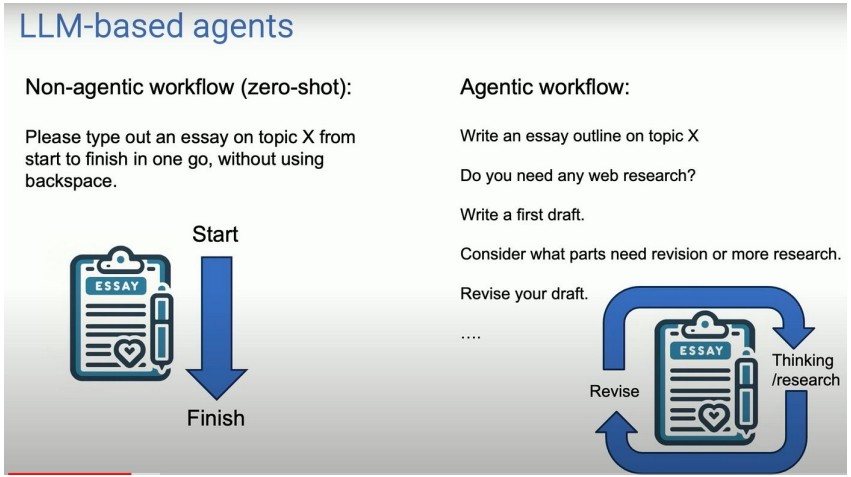


Reflection Agents

Reflection is a prompting strategy used to improve the quality and success rate of agents and similar AI systems..

[Reflection Agents]



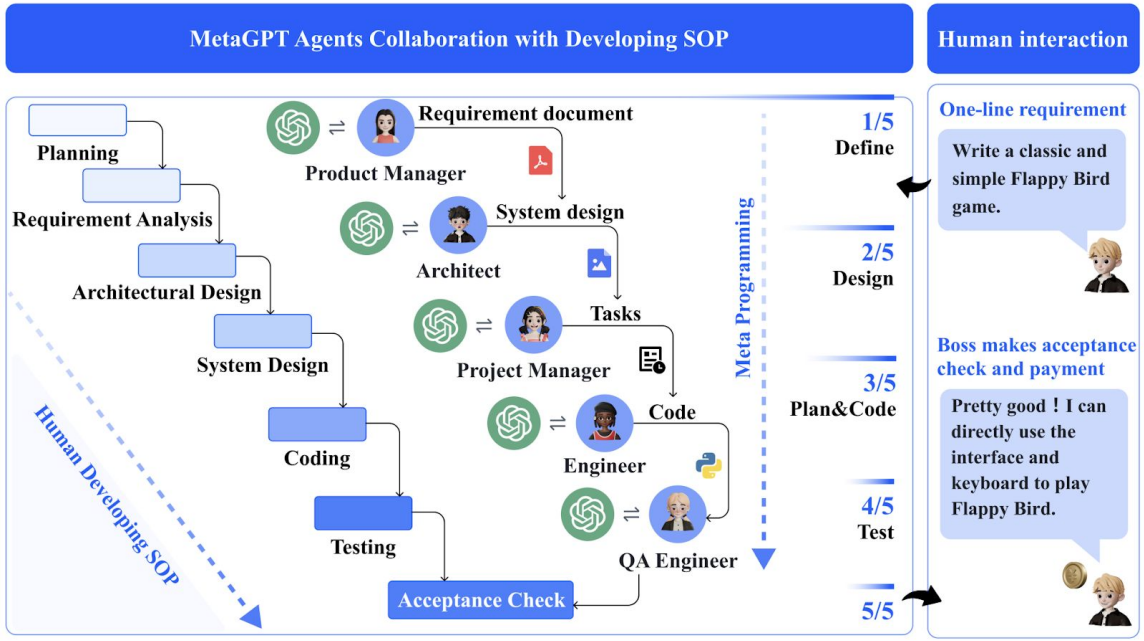


Multi-Agent Collaboration

MetaGPT:
AI Software Company

where LLMs act as

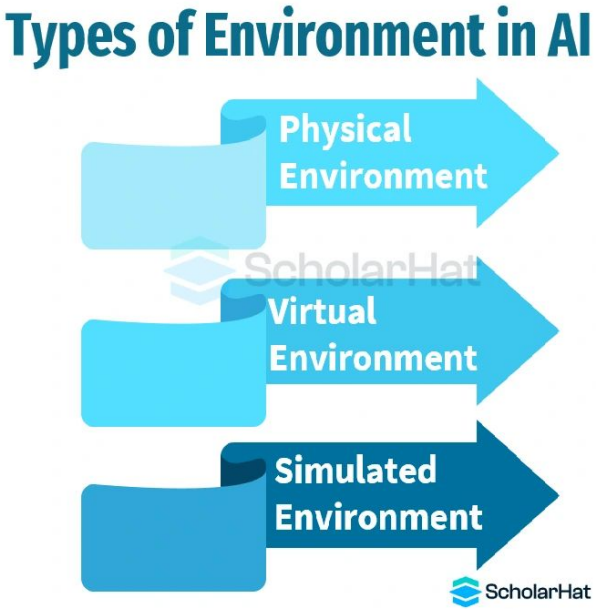
- Product Manager
- Architect
- Project Manager
- Engineer
- QA Engineer



Environment: Three Worlds

The Environment: Three Worlds

Physical world



Omniverse world

Digital world

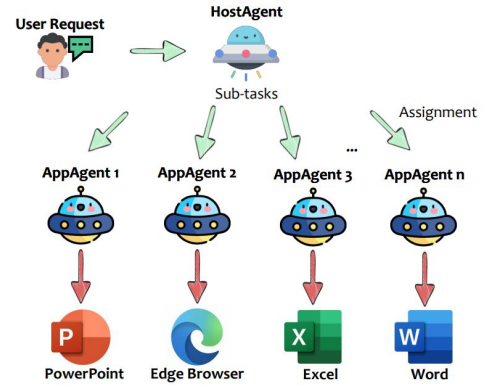


Fig. 21: The multi-agent architecture employed in UFO [17]. Figure adapted from the original paper.

Digital World

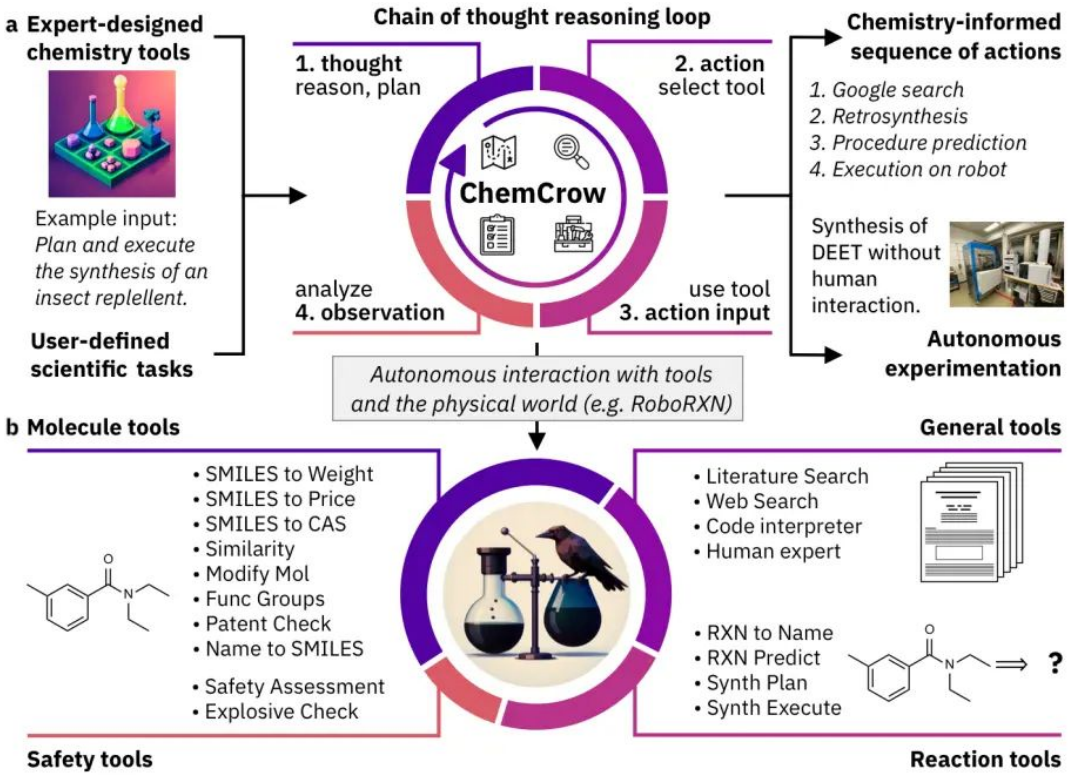
Augmenting large language models with chemistry tools

https://arxiv.org/abs/2304.05376

Andres M. Bran[12*] Sam Cox[3*] Oliver Schilter[24]
Carlo Baldassari[4] Andrew D. White[3] Philippe Schwaller[12]
[1] Laboratory of Artificial Chemical Intelligence (LIAC), ISIC, EPFL
[2]National Centre of Competence in Research (NCCR) Catalysis, EPFL
[3] Department of Chemical Engineering, University of Rochester
[4] Accelerated Discovery, IBM Research – Europe
*Contributed equally.
andrew.white@rochester.edu
philippe.schwaller@epfl.ch

Require APIs to

- Web Search
- LitSearch
- Python REPL
- Name2SMILES
- SMILES2Price
- Name2CAS
- Similarity
- PatentCheck
- FuncGroups
- SMILES2Weight



Gorilla: Large Language Model Connected with Massive APIs

NeurIPS 2024 poster, UC Berkeley

Systems and Algorithms for Integrating LLMs with Applications, Tools, and Services

Rather have the user at the center, Gorilla enables users to interact with a wide range of services through LLMs. Gorilla is an open-source, state-of-the-art LLM that invokes API calls to interact with services!

Gorilla: Large Language Model Connected with Massive APIs

Gorilla Open Functions v2

1. Ask about Github Star history
2. Gorilla output valide REST API
3. Navigate to website

- Colab

Code Generation

Multiple Functions	Parallel Functions	Function Relevance Detection	Rest API
User: Prompt: `What is 2 + 3?` Function: `[add(int a, int b), mult(int a, int b)]` Agent: **`add(a=2, b=3)`**	User: Prompt: `What is (2 + 3) and (4 + 5)?` Function: `[add(int a, int b)]` Agent: **`[add(a=2, b=3), add(a=4, b=5)]`**	User: Prompt: `What is 2*3?` Function: `[add(int a, int b)]` Agent: **`Error. The user asks for adding but we only have multiplication.`**	User: Prompt: `get weather in Berkeley?` Function: `[requests.get(str url, str loc)]` Agent: **`requests.get(url=, loc = 'Berkeley')`**

What if no API is available?

e.g. existing tools that are designed for humans WWW

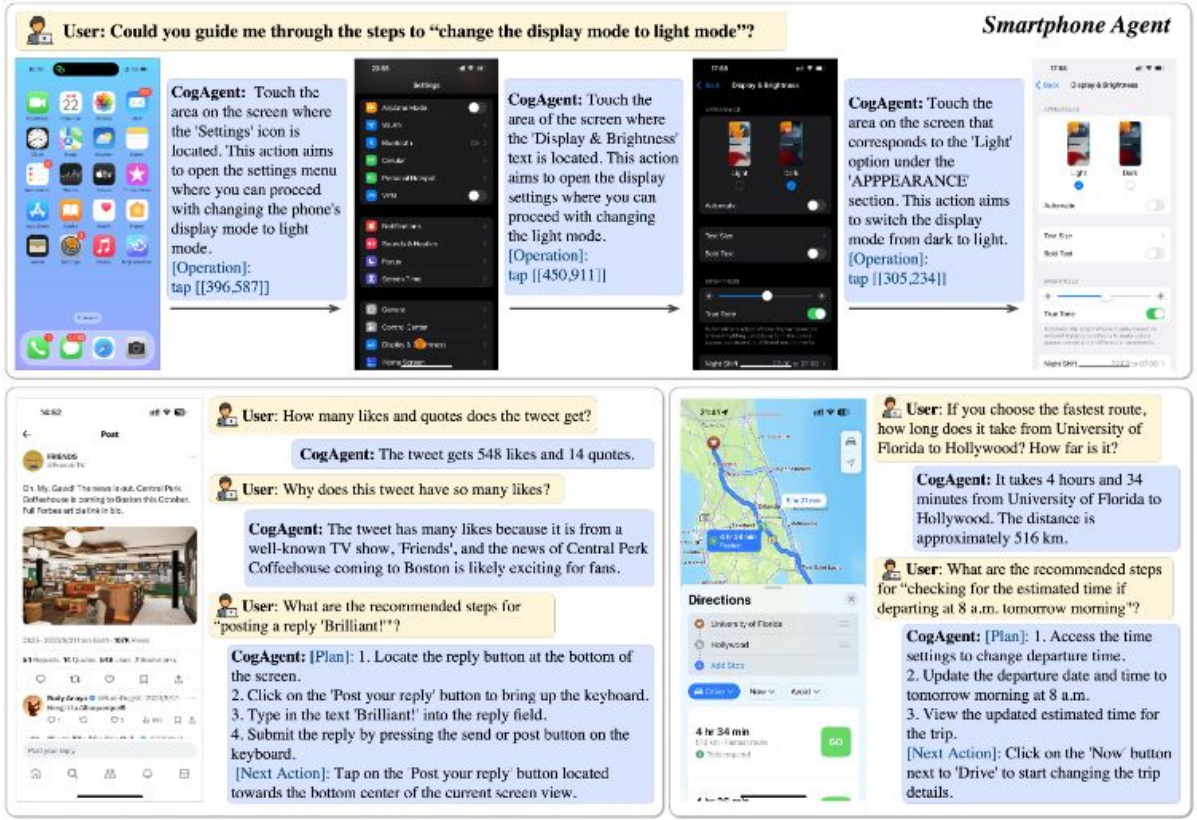
CogAgent: A Visual Language Model for GUI Agents [CVPR2024]

CogAgent

- Visual Language Model based on CogVLM.
- CogAgent-18B 11B visual parameters and 7B language parameters

GUI agents for smartphones are important especially for busy persons or people with poor eyesight or elderly people

http://36.103.203.44:7861/



Large Language Model-Brained GUI Agents: A Survey

Chaoyun Zhang, Shilin He, Jiaxu Qian, Bowen Li, Liqun Li, Si Qin, Yu Kang, Minghua Ma, Guyue Liu, Qingwei Lin, Saravan Rajmohan, Dongmei Zhang, Qi Zhang

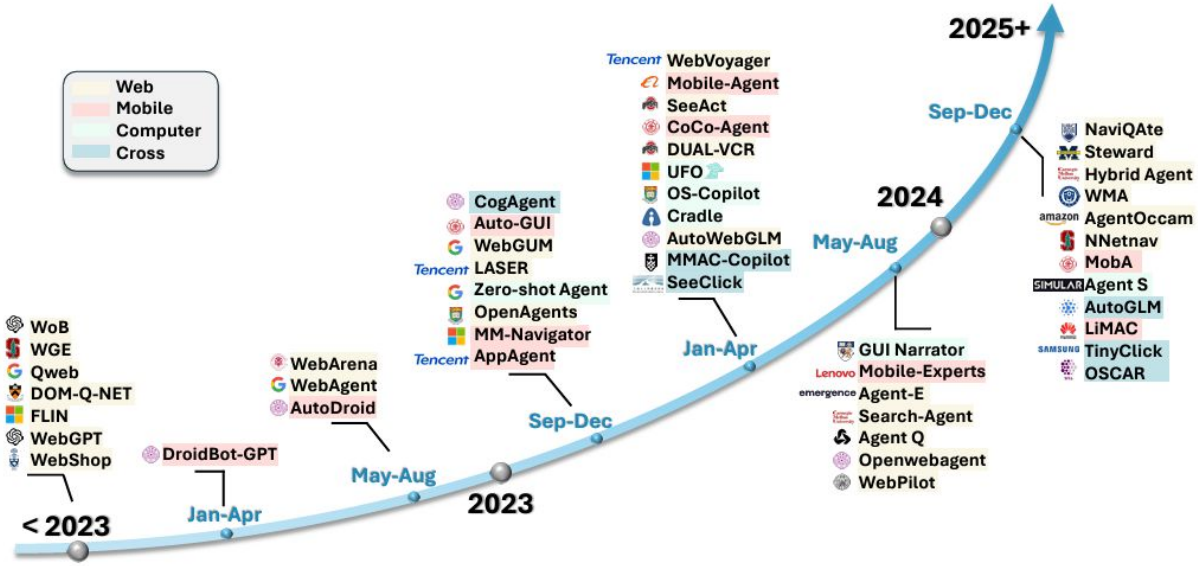
https://arxiv.org/html/2411.18279

From

Graphical UI

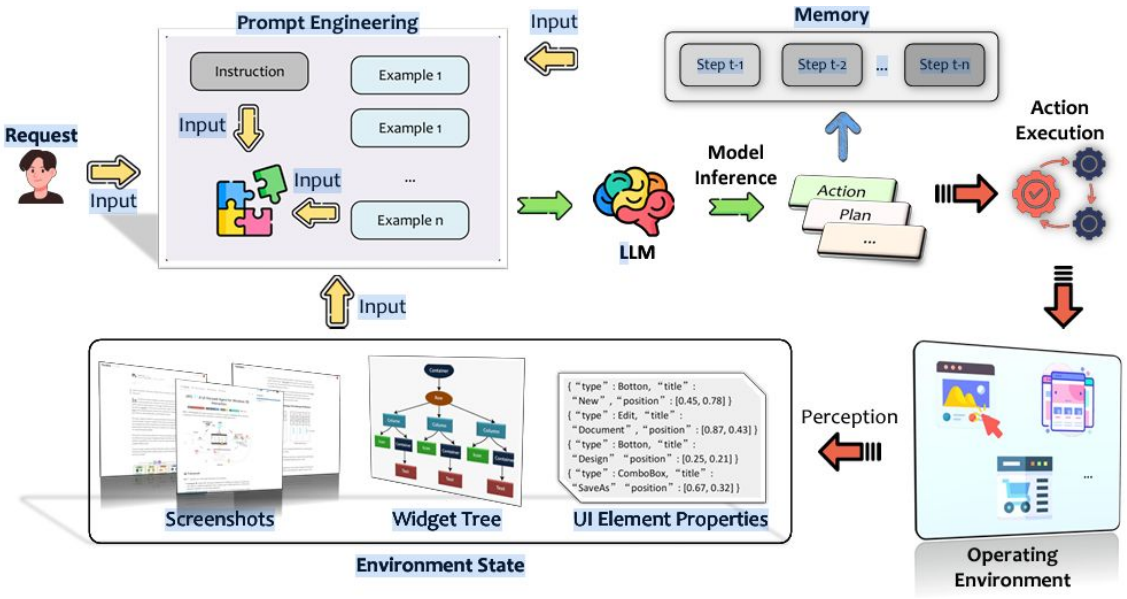
to

Conversational UI



Web Agents

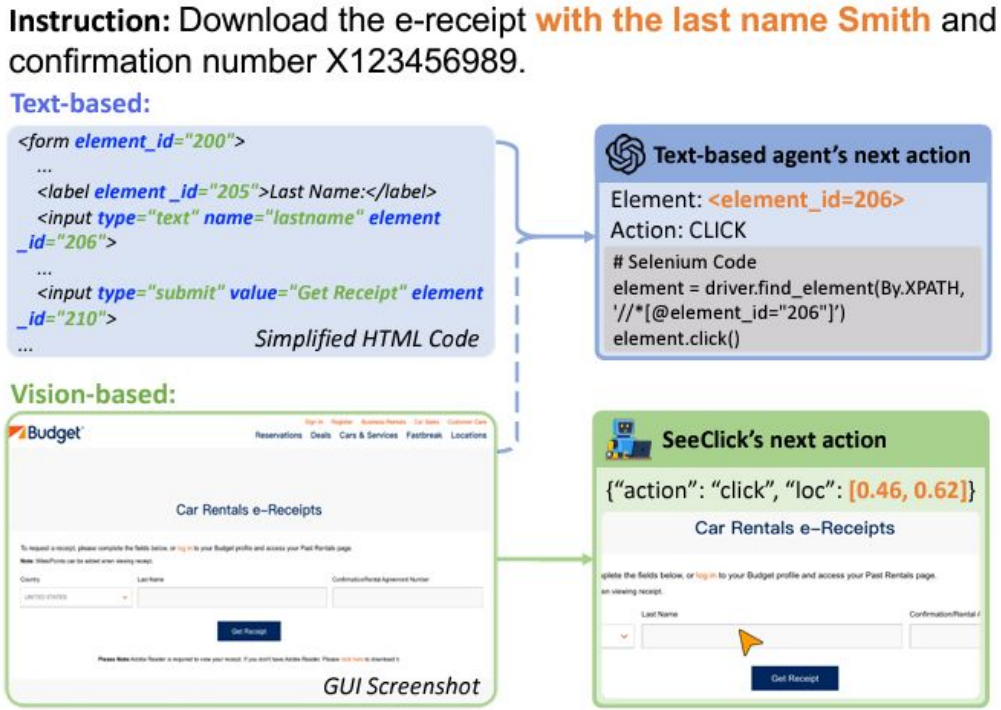
From
Graphical UI
to
Conversational UI



SeeClick: Harnessing GUI Grounding for Advanced Visual GUI Agents

Kanzhi Cheng[♢♡ *] **Qiushi Sun**[♡] **Yougang Chu**[♢] **Fangzhi Xu**[♡]
Yantao Li[♢] **Jianbing Zhang**[♢ †] **Zhiyong Wu**[♡ †]
[♢]National Key Laboratory for Novel Software Technology, Nanjing University
[♡]Shanghai AI Laboratory

- SeeClick only relies on screenshots for task automation.



WebVoyager: Building an End-to-End Web Agent with Large Multimodal Models

Hongliang He[1,3]*, Wenlin Yao[2], Kaixin Ma[2], Wenhao Yu[2], Yong Dai[2], Hongming Zhang[2], Zhenzhong Lan[3], Dong Yu[2]

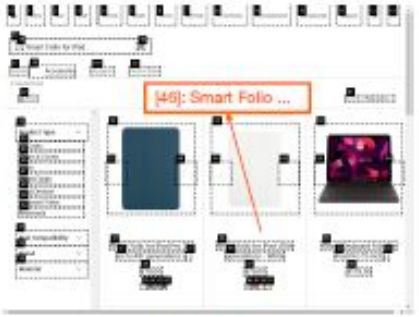
[1]Zhejiang University, [2]Ten

hehongliang@westlake.edu.

- Vision + Android OS API
- Vision + Windows OS API
- UAI + Selenium

Step 1: Click [10]

Step 2: Type [17]: Smart Folio for iPad



Step 3: Click [46]

Step 4: Click [22]

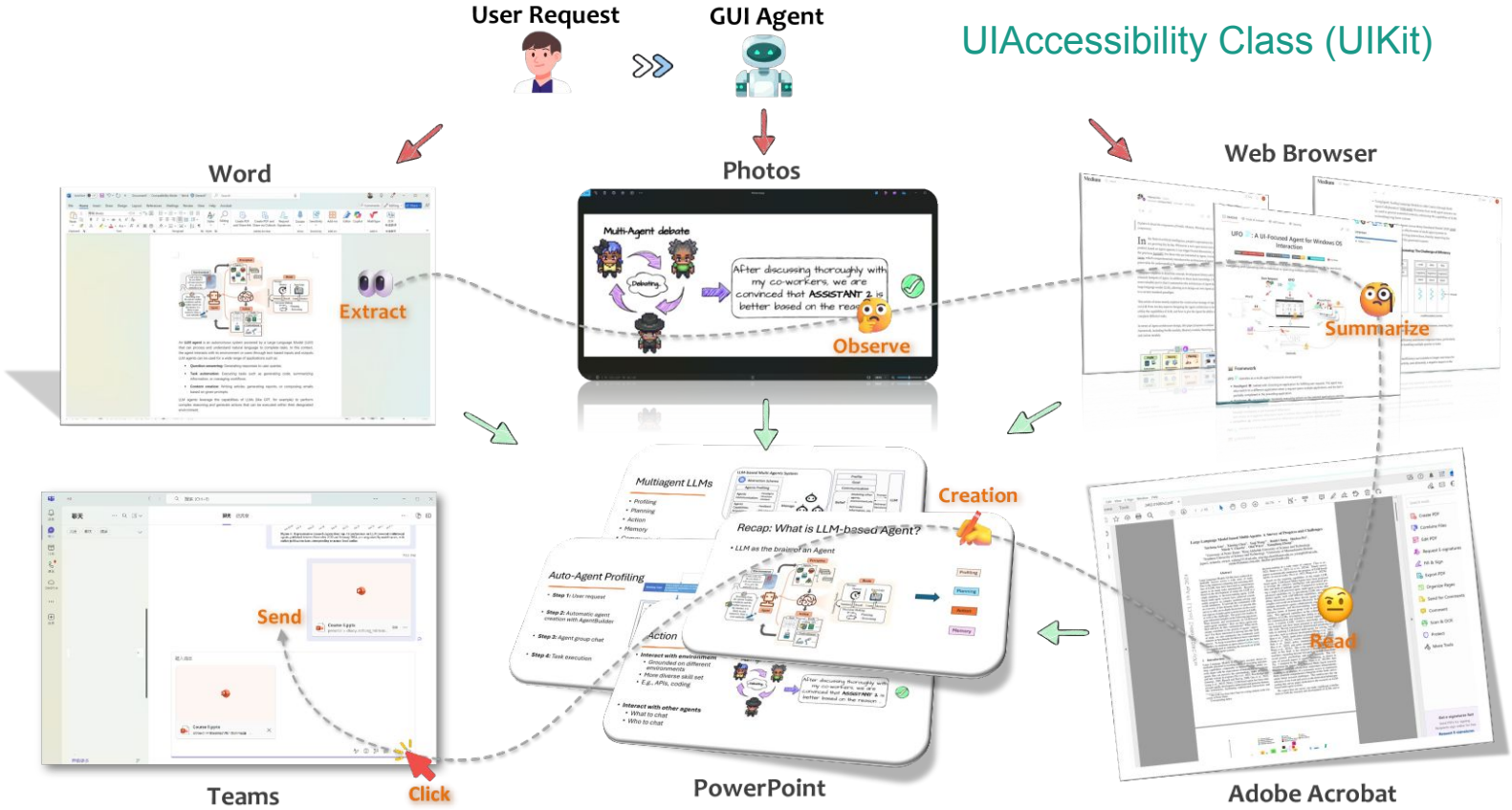
Step 5: Type [0]: 90038

Step 6: ANSWER

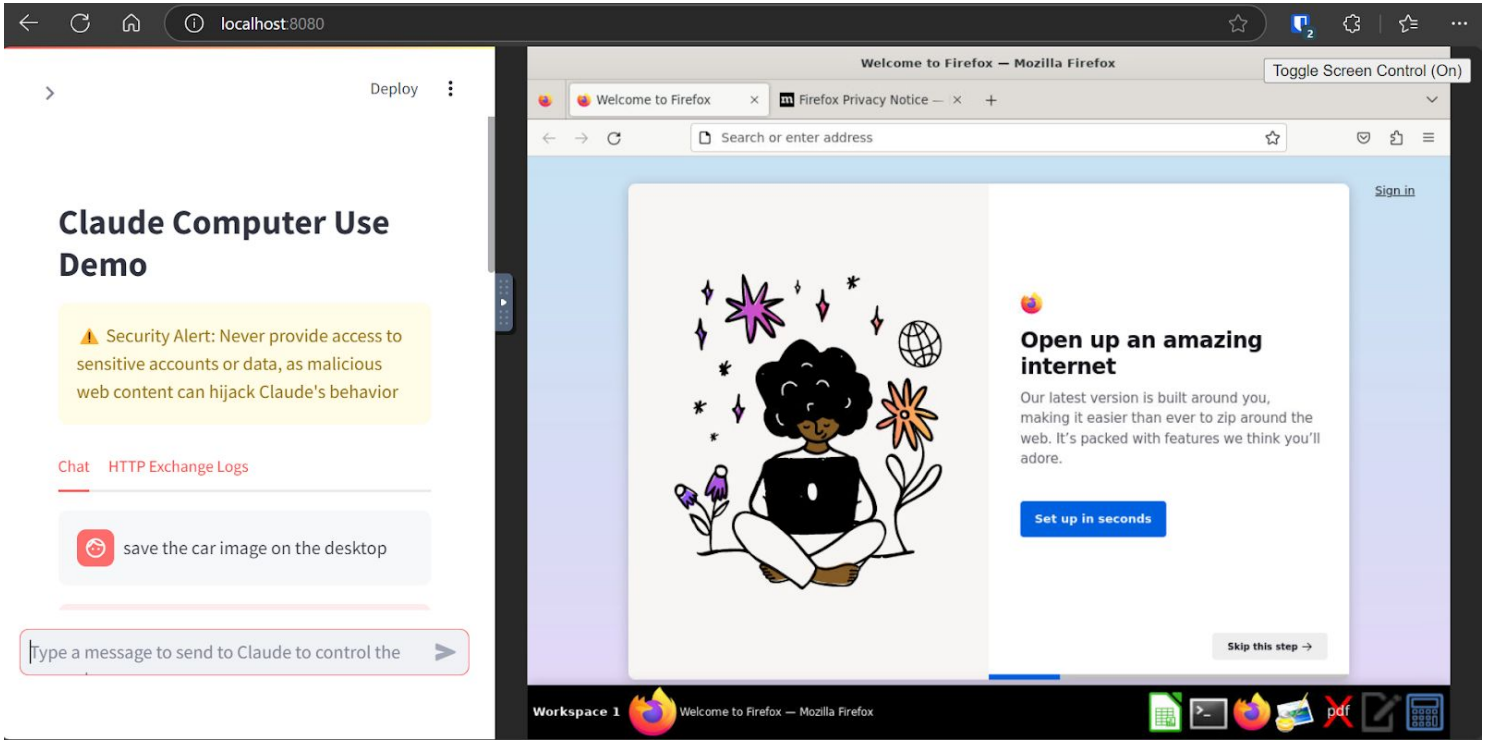
GUI Agents - Cross Applications

UIAccessibility | Apple Developer Documentation

UIAccessibility Class (UIKit)



Computer Use (within a Docker) by Anthropic

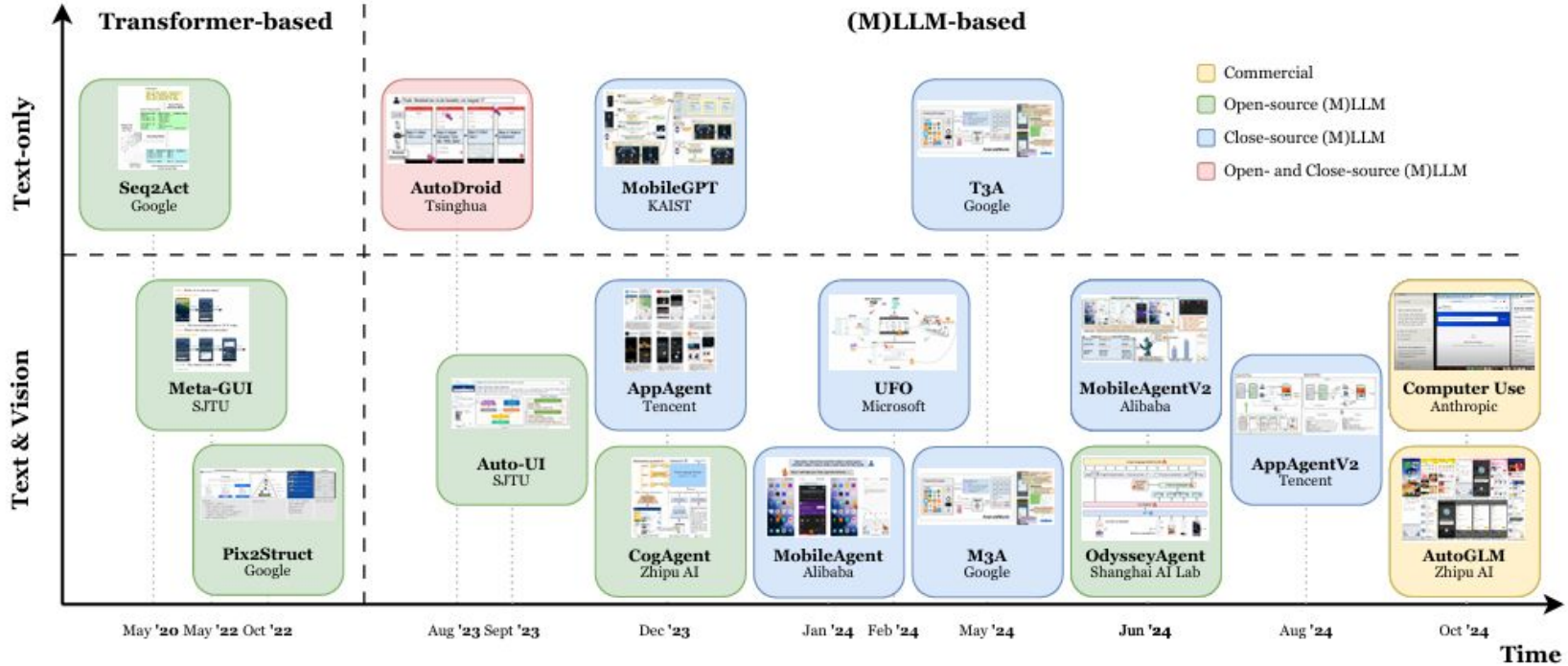


GUI Agents with Foundation Models: A Comprehensive Survey

Shuai Wang[1], Weiwen Liu[*1], Jingxuan Chen[1], Weinan Gan[1], Xingshan Zeng[1], Shuai Yu[1], Xinlong Hao[1], Kun Shao[1], Yasheng Wang[1], and Ruiming Tang[1]

[1]Huawei Noah's Ark Lab
{wangshuai231, liuweiwen8}@huawei.com

https://arxiv.org/abs/2411.04890

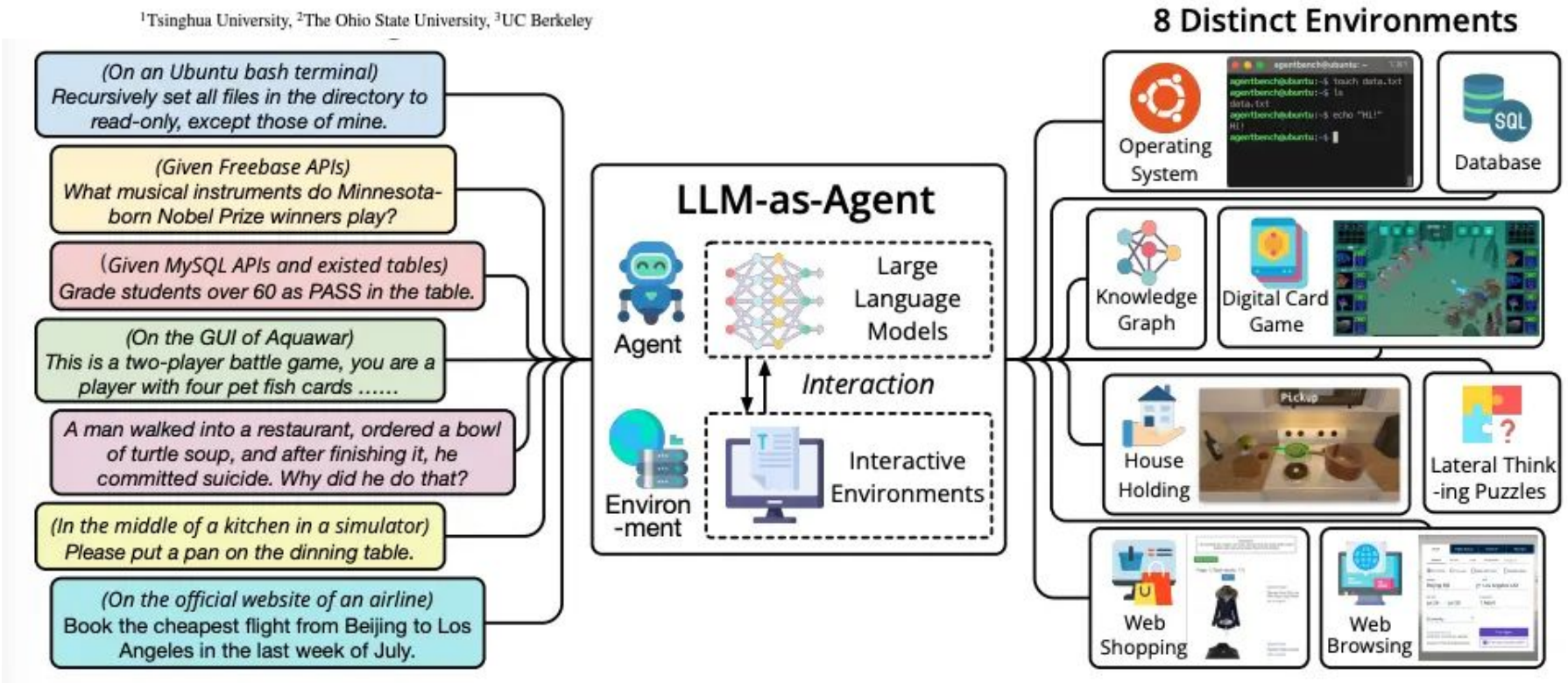


AgentBench: Evaluating LLMs as Agents

Xiao Liu[1,*], Hao Yu[1,*], Hanchen Zhang[1], Yifan Xu[1], Xuanyu Lei[1], Hanyu Lai[1], Yu Gu[2], Hangliang Ding[1], Kaiwen Men[1], Kejuan Yang[1], Shudan Zhang[1], Xiang Deng[2], Aohan Zeng[1], Zhengxiao Du[1], Chenhui Zhang[1], Sheng Shen[3], Tianjun Zhang[3], Yu Su[2], Huan Sun[2], Minlie Huang[1], Yuxiao Dong[1], Jie Tang[1]

[1]Tsinghua University, [2]The Ohio State University, [3]UC Berkeley

https://arxiv.org/abs/2308.03688

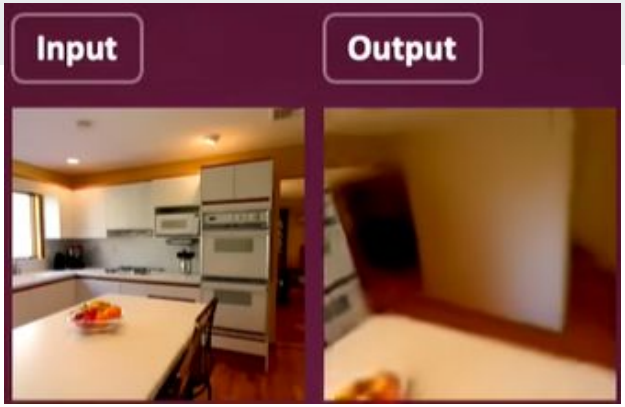


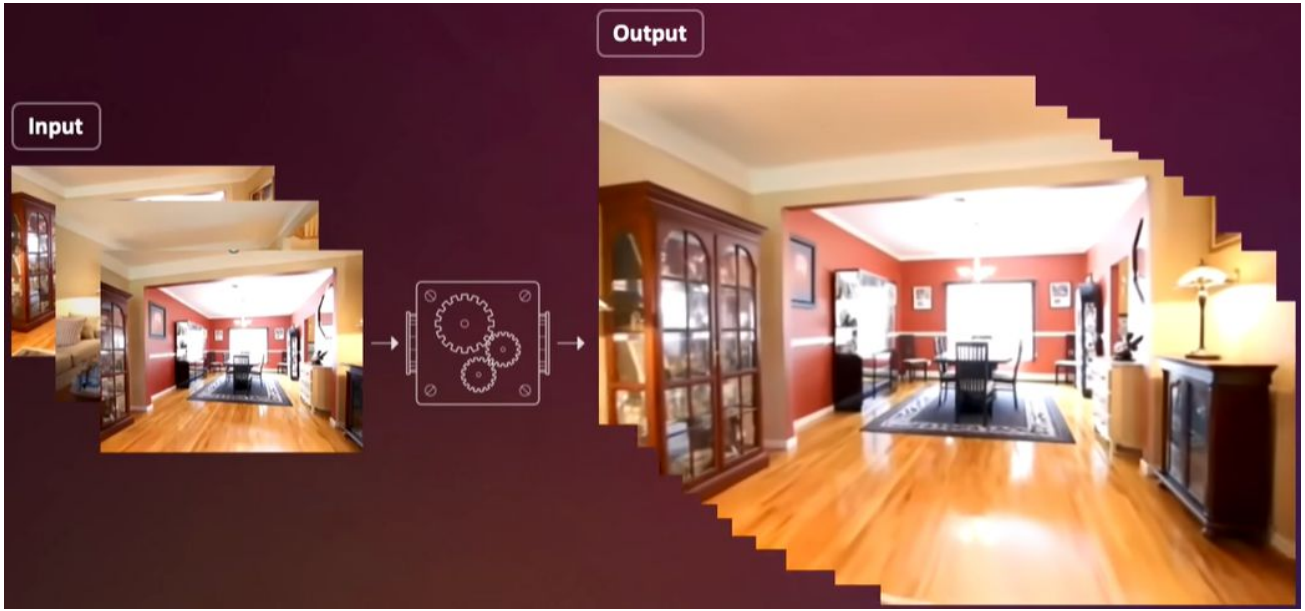
Physical World

Spatial Intelligence

- **From 2D to 3D**
- Seeing is for doing

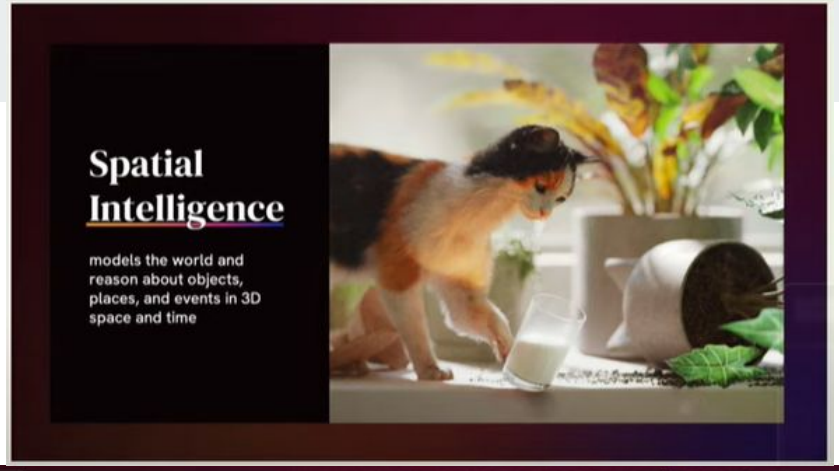
Fei Fei Li, Professor at TED, Apr. 2024



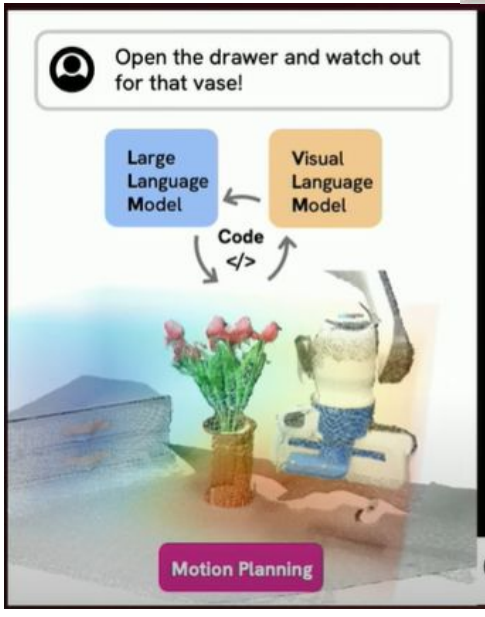


Spatial Intelligence

- From 2D to 3D
- **Seeing is for doing**



Fei Fei Li, Professor | Data + AI Summit 2024



Robotic Learning

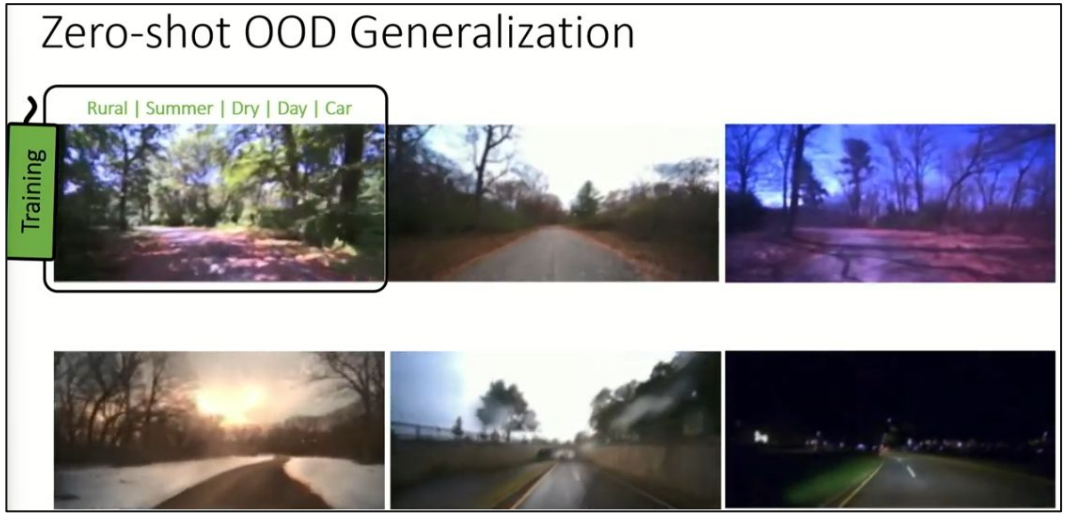
Imitation Learning
Reinforcement Learning

CVPR24 Tutorial | Chelsea Finn: Humanoids and Robot Generalists

Text to Drive
(World Model for Automobile)

Language Augmented Driving Liquid Network

[Text to Drive | Daniela Rus | MIT 2024]

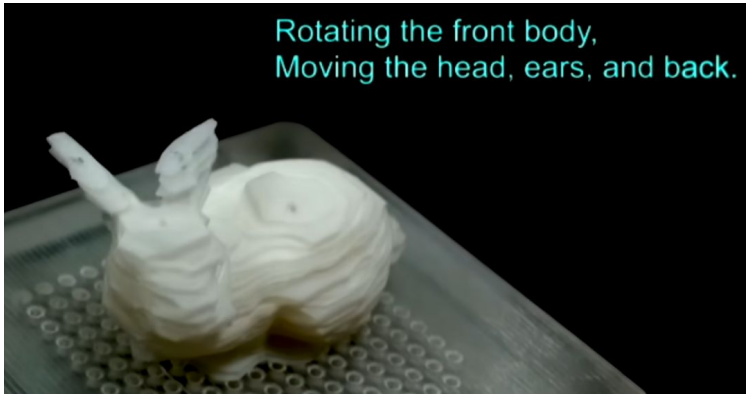


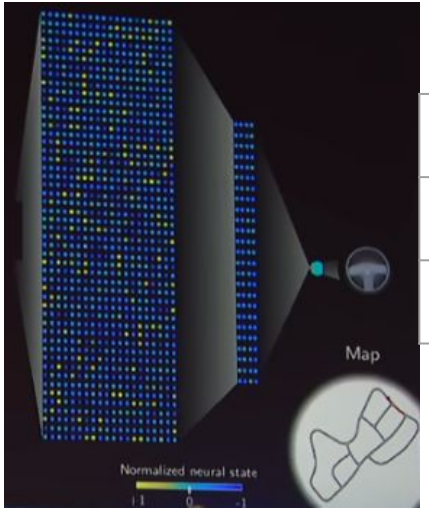
Physical Intelligence

- Think: Liquid Networks
- Design: Text/Image to Robot
- Learn from Human

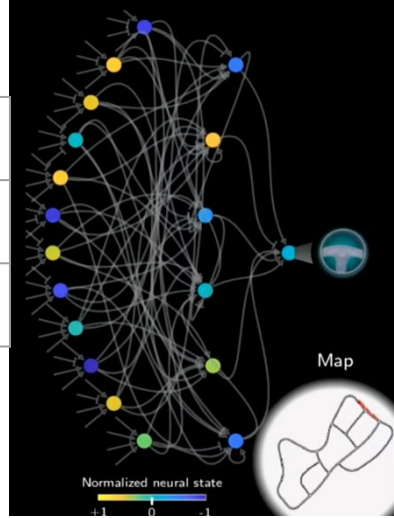
Text to Drive | Daniela Rus | MIT 2024

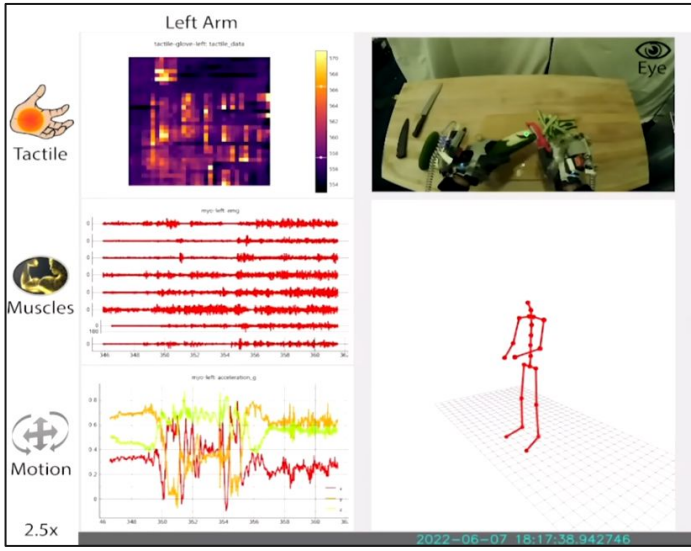
Generate design (shape, material, actuators, sensors, control program)





Traditional	Liquid
>100K	19
Frozen	Adapt

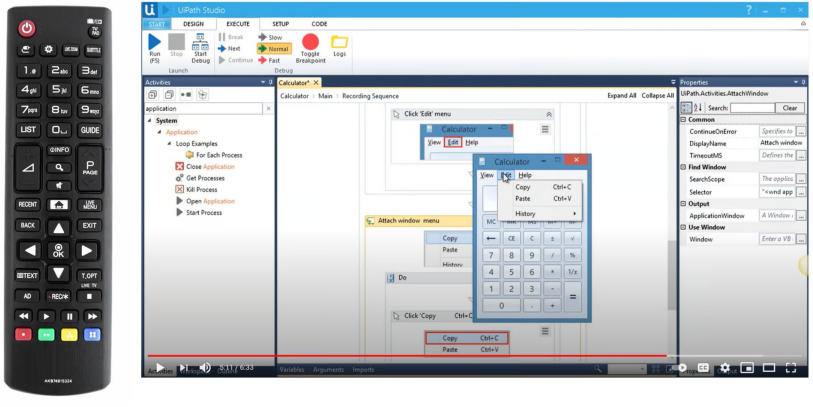




Opportunities

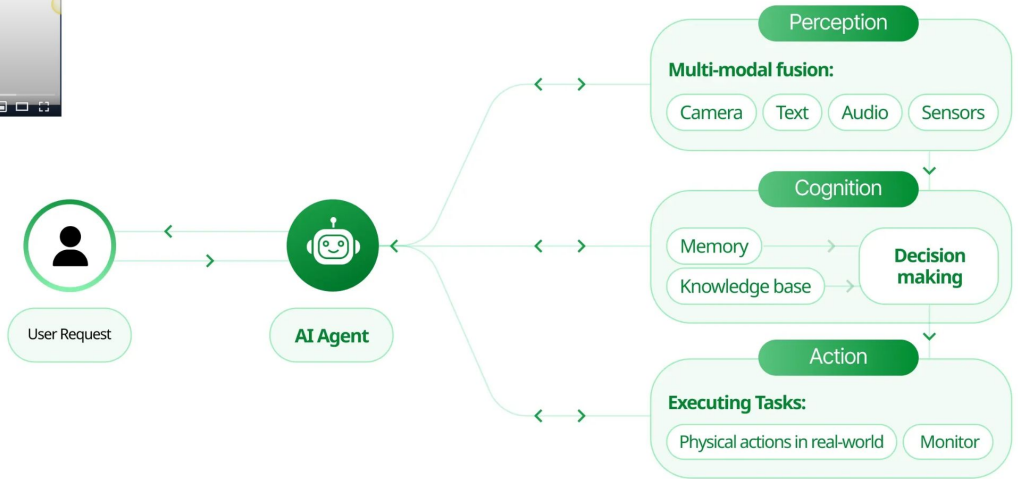
- Agentic AI
  - **Civilization of RPA**
  - **Better HCI for all information systems**
  - **CUI 2025**
- Sovereign and Sustainable AI
  - **Data and Model Sovereignty**
  - **Economic Independence**
  - **DeepSeek R1, Grok 3**
- Safety AI
  - **Evaluation of AI Systems**
  - **Ethical and Cultural Alignment**

Tools vs. Agents



With many hidden functions!

AI Agents– Brain, Perception, Tools, Agents

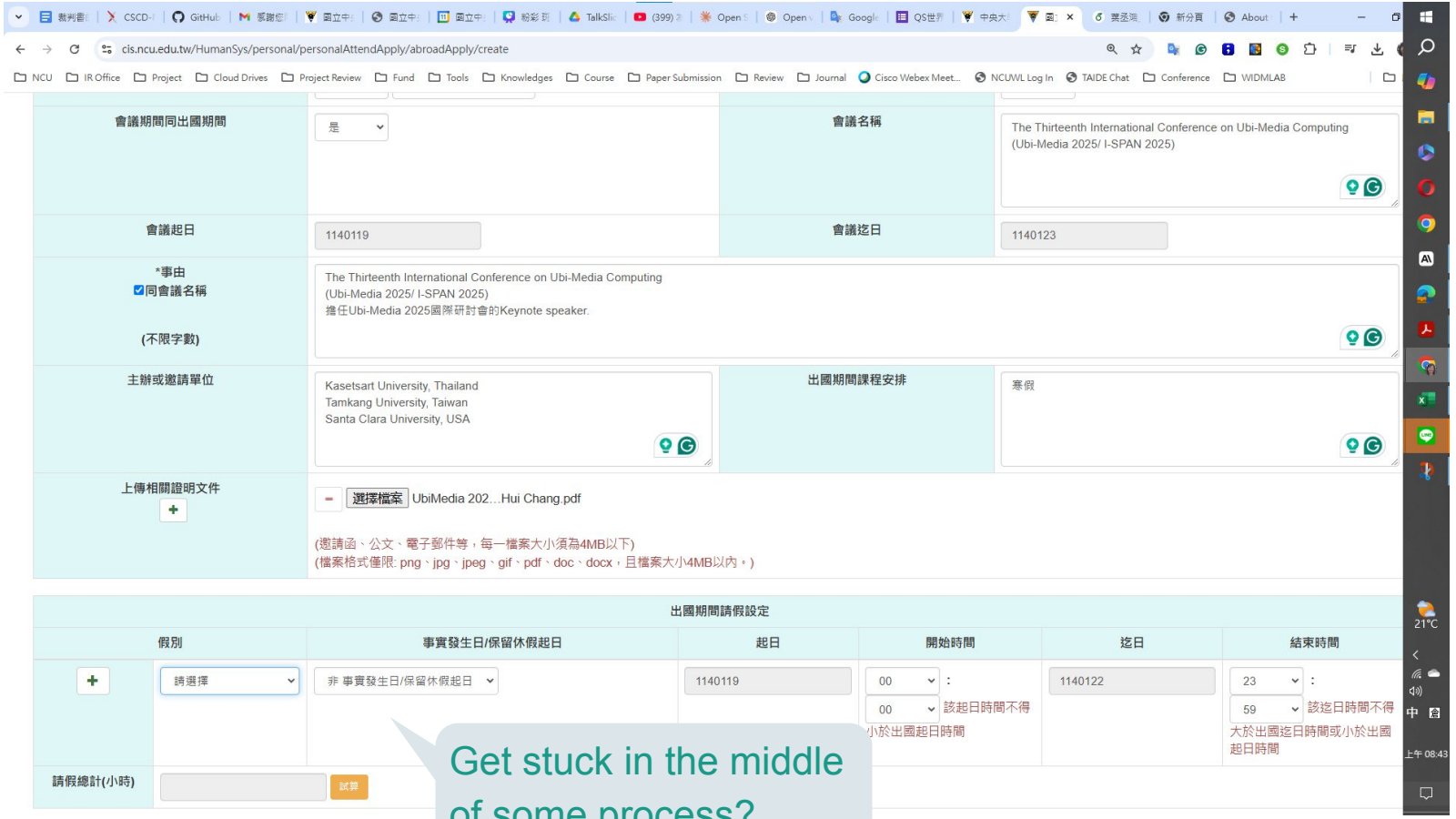


搜尋系統為什麼不好用？

We need better search systems

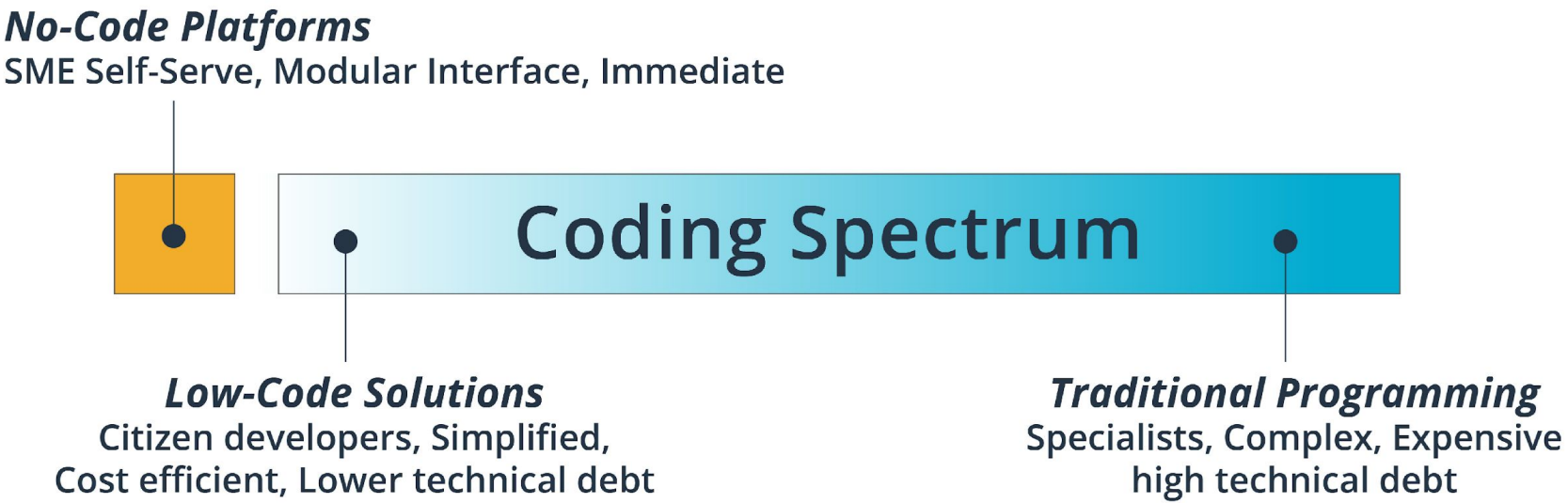
我們是不是常常在尋找網站中的某個功能?

We need better UI and UX


cis.ncu.edu.tw/HumanSys/personal/personalAttendApply/abroadApply/create
會議期間同出國期間
是
會議名稱
The Thirteenth International Conference on Ubi-Media Computing (Ubi-Media 2025/ I-SPAN 2025)
會議起日
1140119
會議迄日
1140123
*事由
同會議名稱
(不限字數)
The Thirteenth International Conference on Ubi-Media Computing (Ubi-Media 2025/ I-SPAN 2025)
擔任Ubi-Media 2025國際研討會的Keynote speaker.
主辦或邀請單位
Kasetsart University, Thailand
Tamkang University, Taiwan
Santa Clara University, USA
出國期間課程安排
寒假
上傳相關證明文件
選擇檔案
UbiMedia 202...Hui Chang.pdf
(邀請函、公文、電子郵件等，每一檔案大小須為4MB以下)
(檔案格式僅限: png、jpg、jpeg、gif、pdf、doc、docx，且檔案大小4MB以內。)
出國期間請假設定
假別
事實發生日/保留休假起日
起日
開始時間
迄日
結束時間
請選擇
非 事實發生日/保留休假起日
1140119
00
00
該起日時間不得小於出國起日時間
1140122
23
59
該迄日時間不得大於出國迄日時間或小於出國起日時間
請假總計(小時)
試算
Get stuck in the middle of some process?

From Low-Code to No-Code

https://www.nextw.com/what-is-no-code-ap

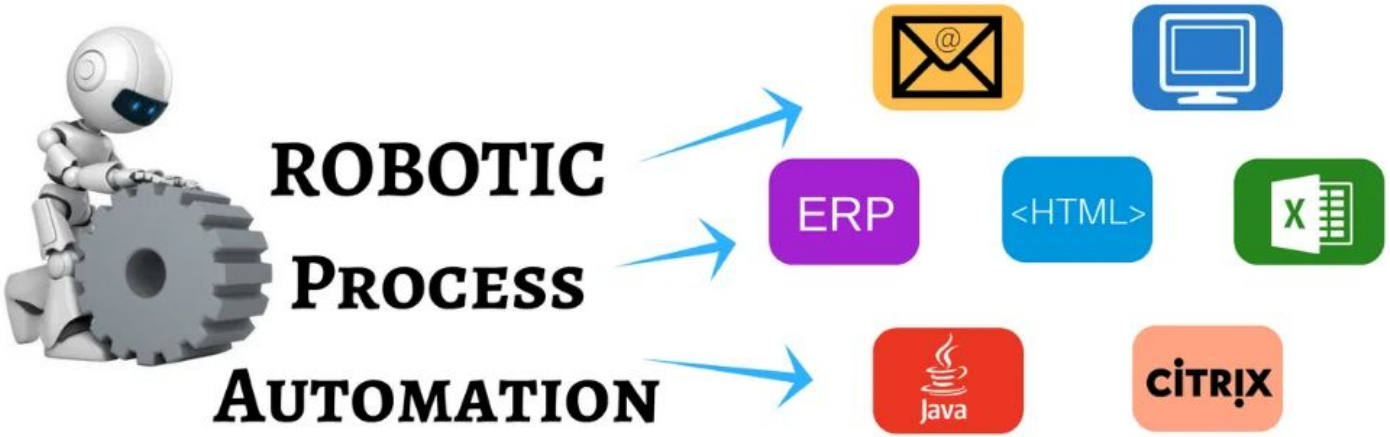


Your Turn - Discussion

Topics:

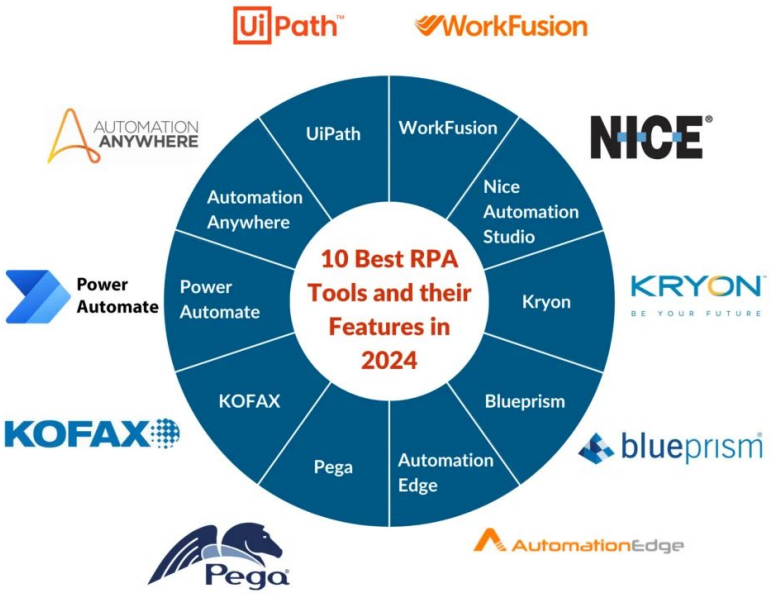
- Difference of AI Agents and Agentic AI?
- Agentic AI in X?

Applications in RPA



RPA Tools

RPA creation may takes days, weeks, or months



	Automation Anywhere	ElectroNeek	UiPath	WinAutomation by Softomotive
Star Rating	★★★★½	★★★★★	★★★★½	★★★★½
Visual Editor	8.7	9.5	8.9	8.6
Meets Requirements	8.7	9.4	8.9	8.8
Ease of Use	8.8	9.6	9.0	8.6
Ease of Setup	8.2	9.5	8.6	9.0
Bot Scheduling	8.9	8.9	8.9	8.8
Attended Automation	8.7	9.9	8.8	9.0
Unattended Automation	8.5	9.6	8.7	8.4
Bot Performance Analytics	8.7	9.0	8.7	7.7

Overcome the barrier of RPA creation

Devices

Old Interfaces
- GUIs
- Displays
- Apps
- Websites

AI Interface
The Generative AI models trained and tuned for human communications.

Traditional Agents
- Digital process automation
- Rule-based 'bots (RPAs)
- Reinforcement Learning

AI Agents
AI models (generative, predictive and others) and supporting software that accomplish work on our behalf.

Applications & Data

Source: Forrester Report, "Change The Interface; Change The World"

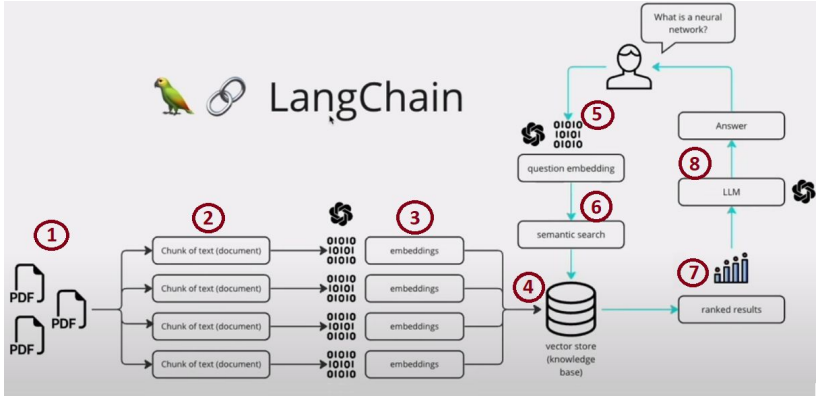
Thank you!

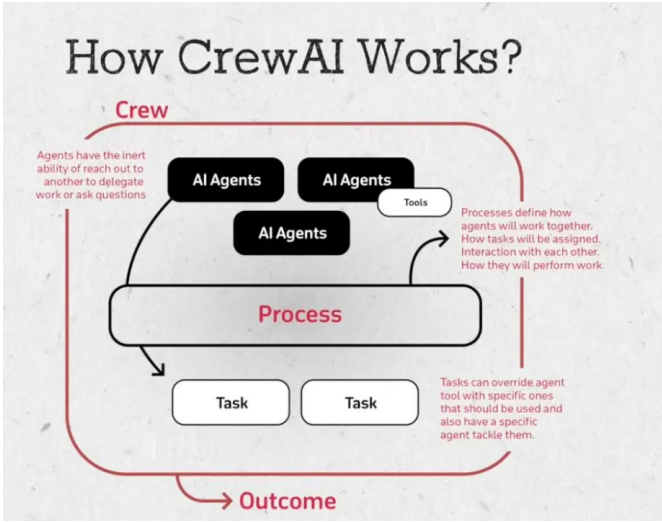
Thank you for your time and attention.

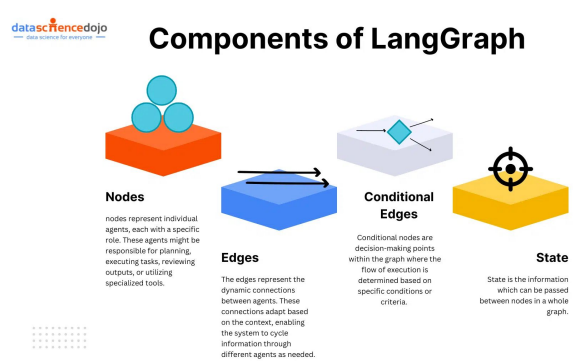
If you have any questions or would like to discuss this topic further, please feel free to email me.

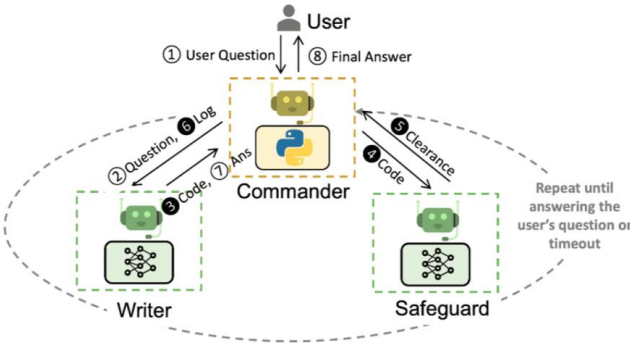
chiahui@g.ncu.edu.tw

Agent Development Frameworks

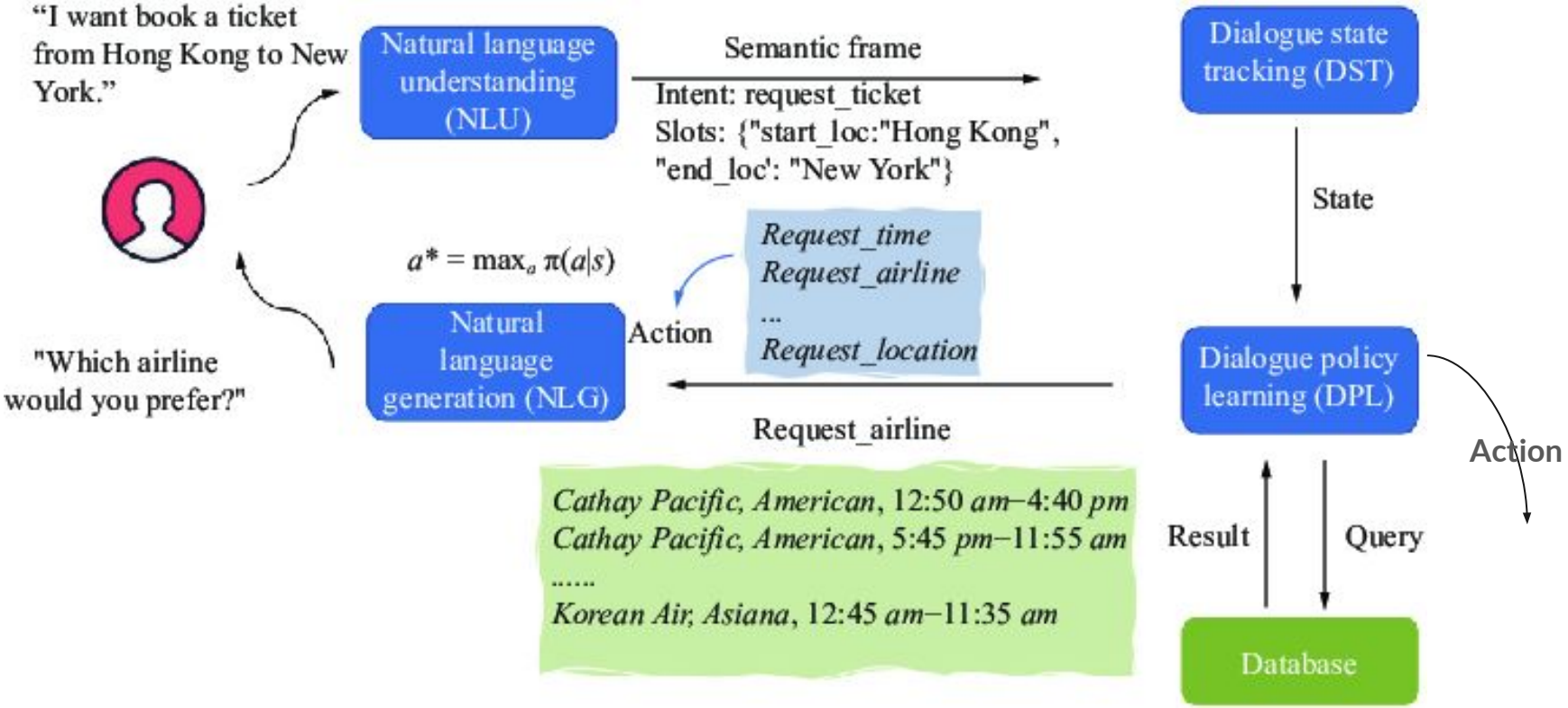




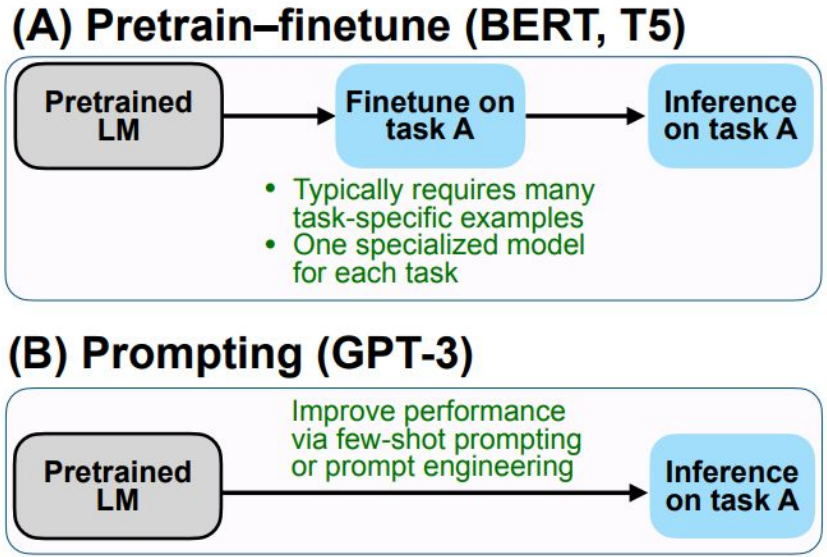


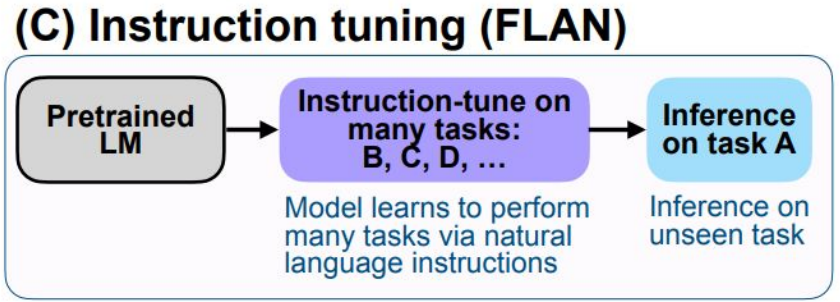


TOD: Task-Oriented Dialog Systems



Pre-training vs. Fine-Tuning vs. In-context Learning





https://openreview.net/pdf?id=gEZrGCozdqR

Challenges of EduACT

1. **The agent creator might have overlooked something.**

   The platform need to handle situations not considered by the creator.

2. **Users might not always follow agents' guidances.**

- Relevance score measures whether users respond well to agents
- Distance to the goal