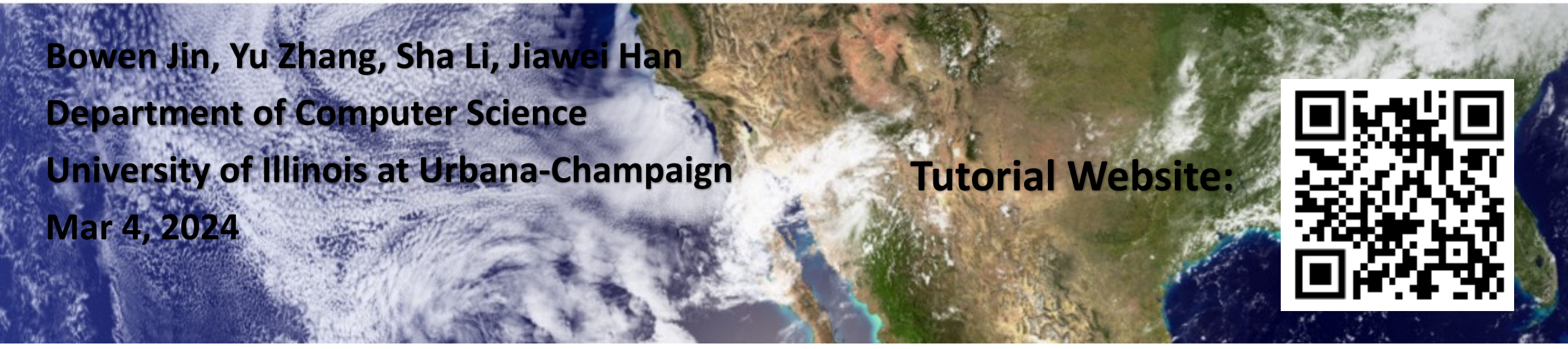




# Summary & Future Directions

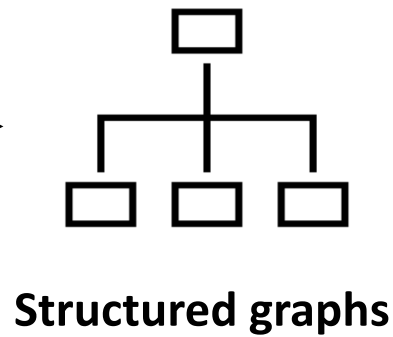


**Bowen Jin, Yu Zhang, Sha Li, Jiawei Han**  
**Department of Computer Science**  
**University of Illinois at Urbana-Champaign**  
**Mar 4, 2024**

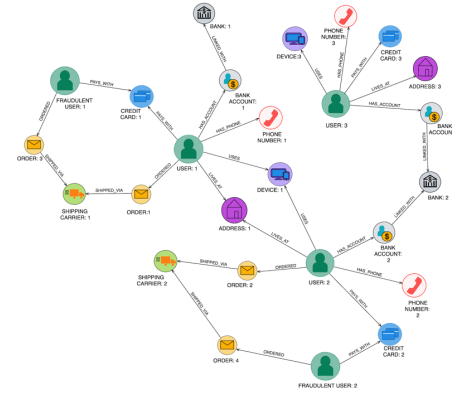
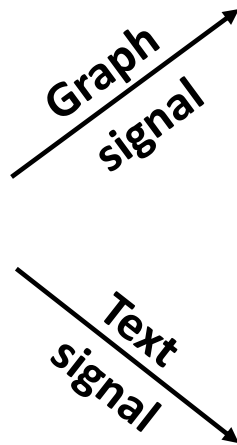
**Tutorial Website:**



# Our Roadmap of This Tutorial



**Part I: Enhancing Text with Graph Structure**



**Part II: Graph Mining with Large Language Model**



**Part III: Text Mining with Structure Information**



**Part IV: Summary & Looking Forward**

# Summary: bridging Text data and Graph data

---

- ❑ Leverage the Power of (Large) Language Models to Transform Unstructured Text into Graphs for Knowledge Discovery
- ❑ Mining Graphs with LLMs
  - ❑ LLM-based Graph Reasoning
  - ❑ Joint Graph/Text Representation Learning and Generation
- ❑ Text Mining with Structure-information
  - ❑ Classification/QA/Pretraining with Structured Information
- ❑ Still a lot of work to do on the topic of “LLMs on graphs”

# Future Directions

## □ Better Benchmark Datasets

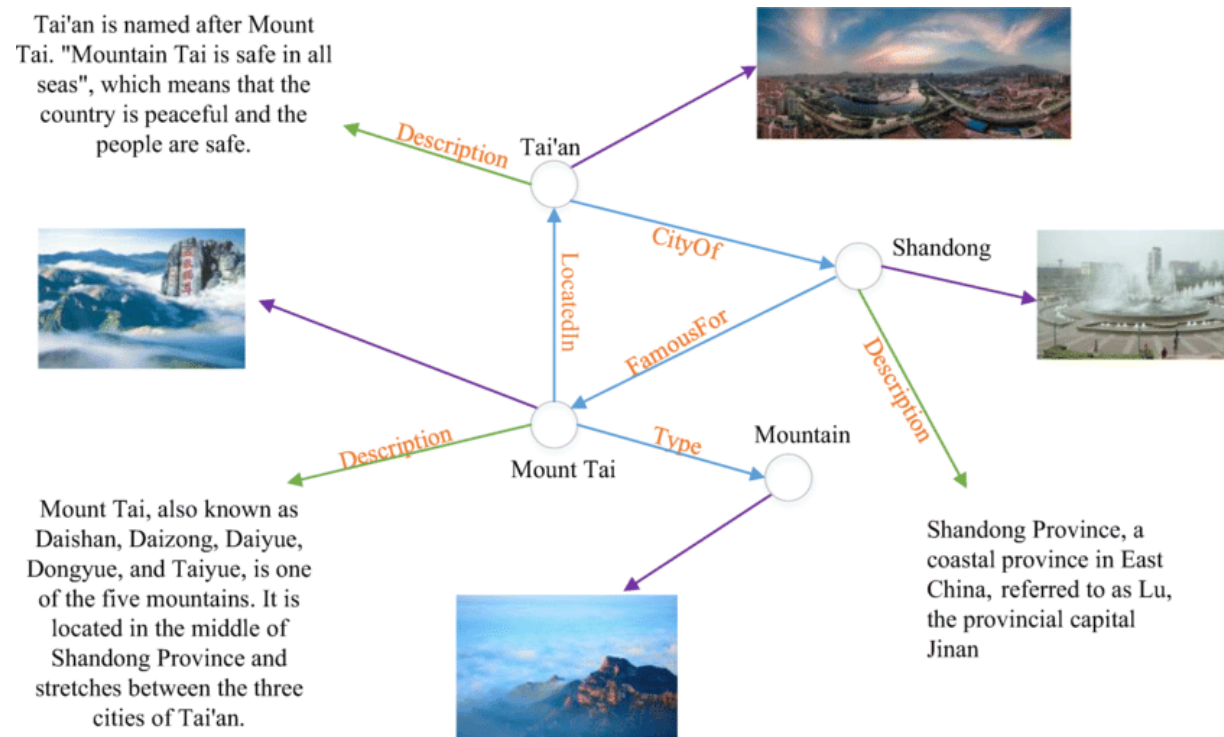
- More diverse datasets from multiple domains are needed (e.g., legal and healthcare).

Data collection in Section 5 for text-attributed graphs. Task: “NC”, “UAP”, “LP”, “Rec”, “EC”, “RG” denote node classification, user activity prediction, link prediction, recommendation, edge classification, and regression task.

	Text.	Data	Year	Task	# Nodes	# Edges	Domain	Source & Notes
Node		ogb-arxiv	2020.5	NC	169,343	1,166,243	Academic	OGB [188]
		ogb-products	2020.5	NC	2,449,029	61,859,140	E-commerce	OGB [188]
		ogb-papers110M	2020.5	NC	111,059,956	1,615,685,872	Academic	OGB [188]
		ogb-citation2	2020.5	LP	2,927,963	30,561,187	Academic	OGB [188]
		Cora	2000	NC	2,708	5,429	Academic	[10]
		Citeseer	1998	NC	3,312	4,732	Academic	[11]
		DBLP	2023.1	NC, LP	5,259,858	36,630,661	Academic	www.aminer.org/citation
		MAG	2020	NC, LP, Rec RG	~ 10M	~ 50M	Academic	multiple domains [12] [13]
		Goodreads-books	2018	NC, LP	~ 2M	~ 20M	Books	multiple domains [14]
		Amazon-items	2018	NC, LP, Rec	~ 15.5M	~ 100M	E-commerce	multiple domains [15]
		SciDocs	2020	NC, UAP, LP, Rec	-	-	Academic	[51]
		PubMed	2020	NC	19,717	44,338	Academic	[16]
		Wikidata5M	2021	LP	~ 4M	~ 20M	Wikipedia	[17]
		Twitter	2023	NC, LP	176,279	2,373,956	Social	[53]
Edge		Goodreads-reviews	2018	EC, LP	~ 3M	~ 100M	Books	multiple domains [14]
		Amazon-reviews	2018	EC, LP	~ 15.5M	~ 200M	E-commerce	multiple domains [15]
		Stackoverflow	2023	EC, LP	129,322	281,657	Social	[74]

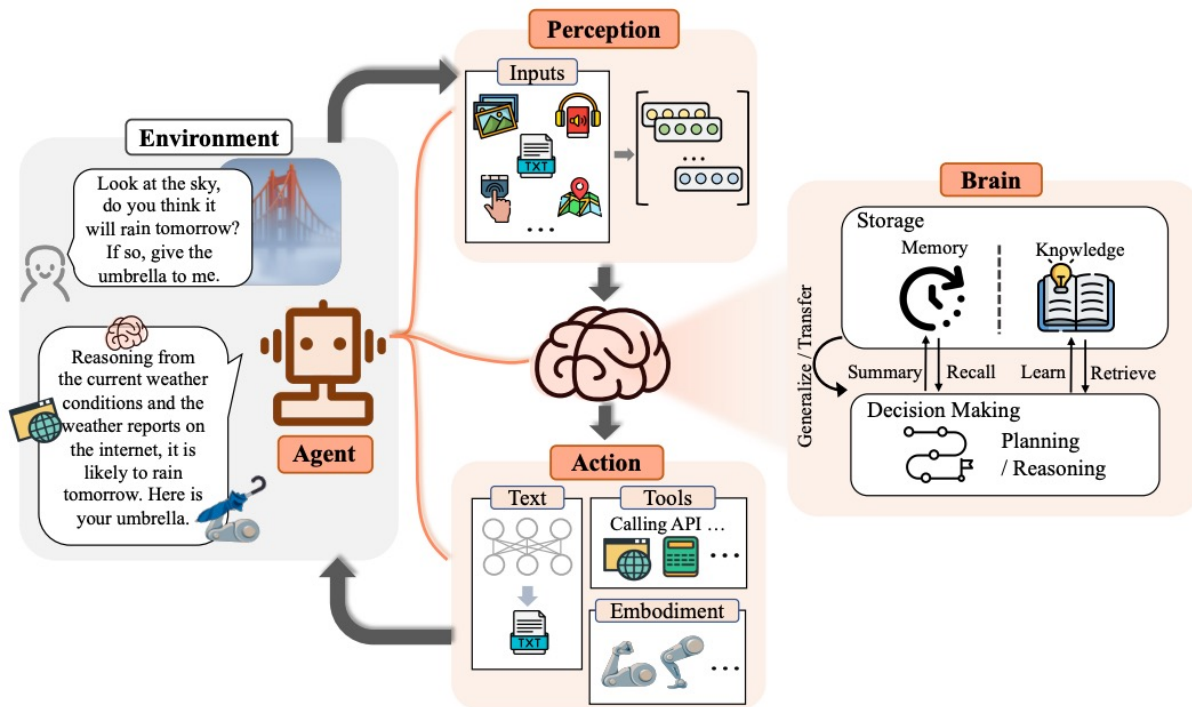
# Future Directions

- Multi-Modal Learning on Graphs with Vision-Language Models
  - Multi-Modal data (e.g., image, text, audio, etc) are linked as a graph.



# Future Directions

## LLM Agent on Graphs

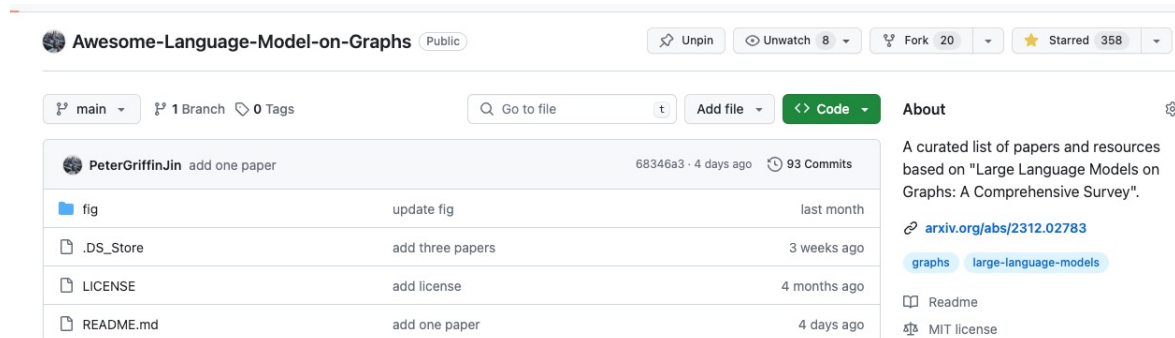
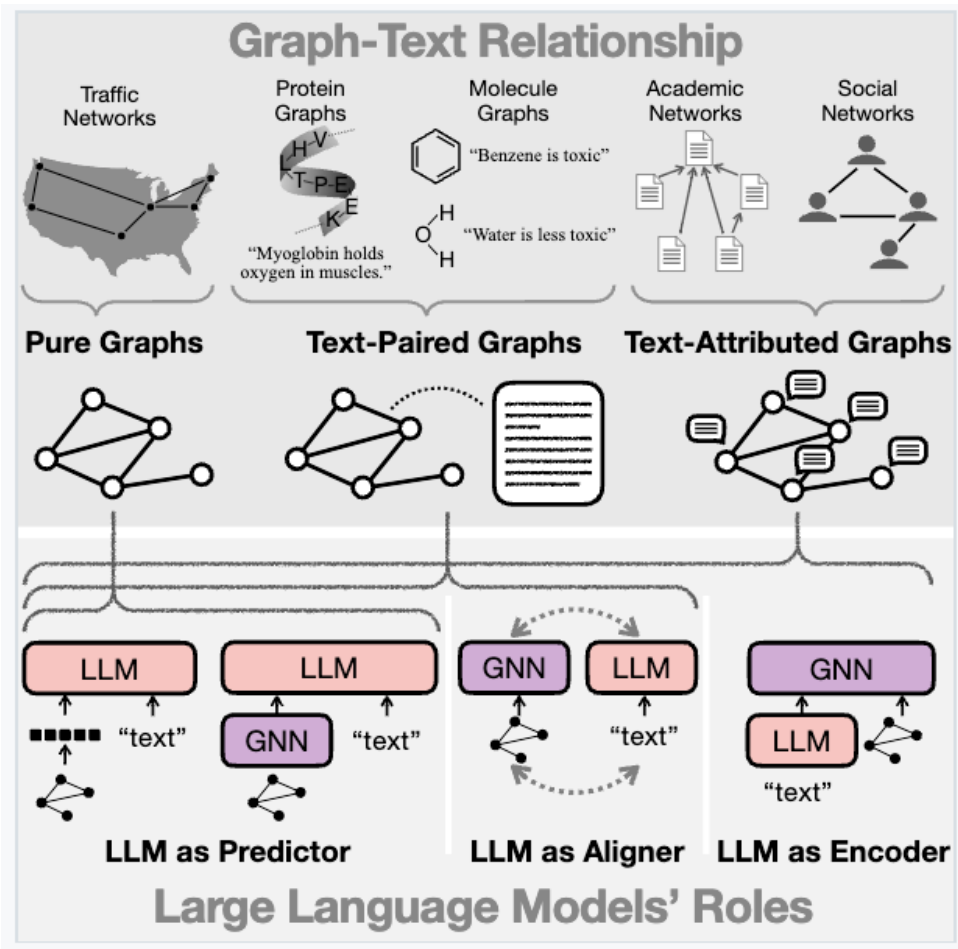


Graphs as the Environment;  
LLMs as the agents

# More Resources and Future Directions

□ A survey paper of LLM & graphs

□ A resource repo of LLM & graphs



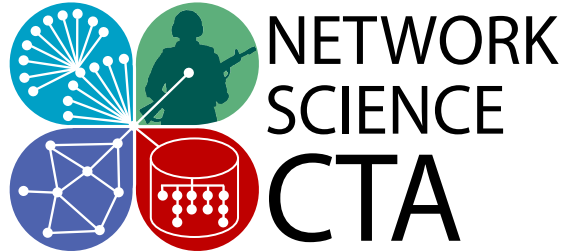
paper



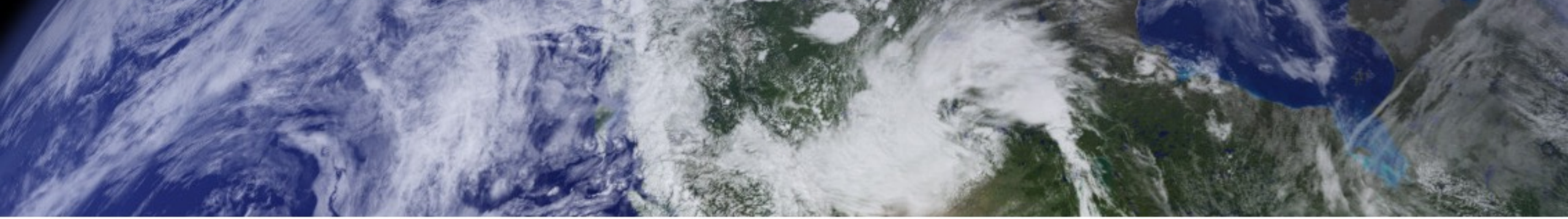
repo

# Acknowledgement

- Thanks for the research support from: ARL/NSCTA, NIH, NSF, DHS, ARO, DTRA







**Thank you !**  
**Q&A**

