上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

rLLM: Relational Table Learning with LLMs

Weichen Li, Xiaotong Huang, Jianwu Zheng, **Zheng Wang**, Li Pan, Jianhua Li

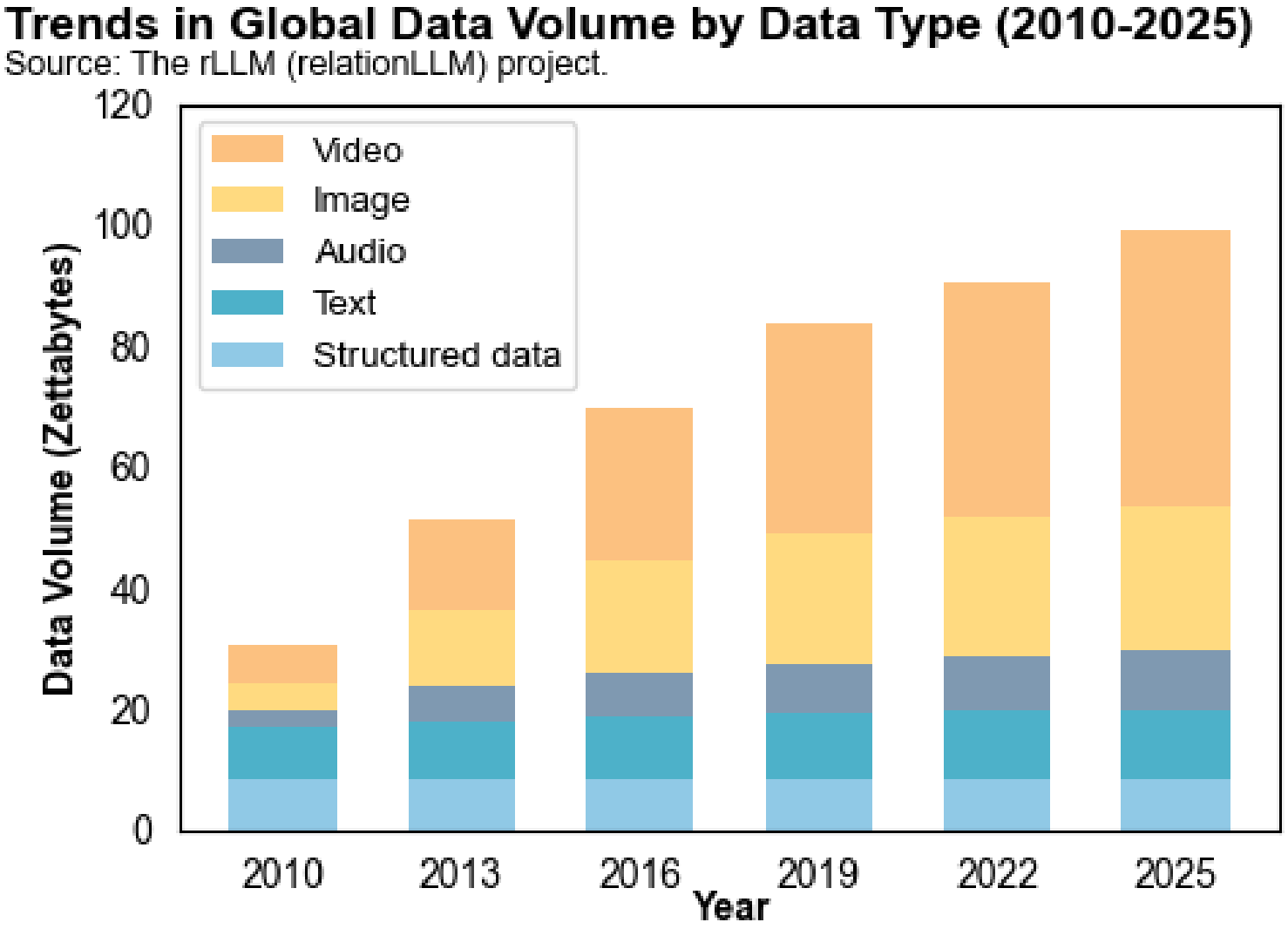
wzheng@sjtu.edu.cn

Shanghai Jiao Tong University

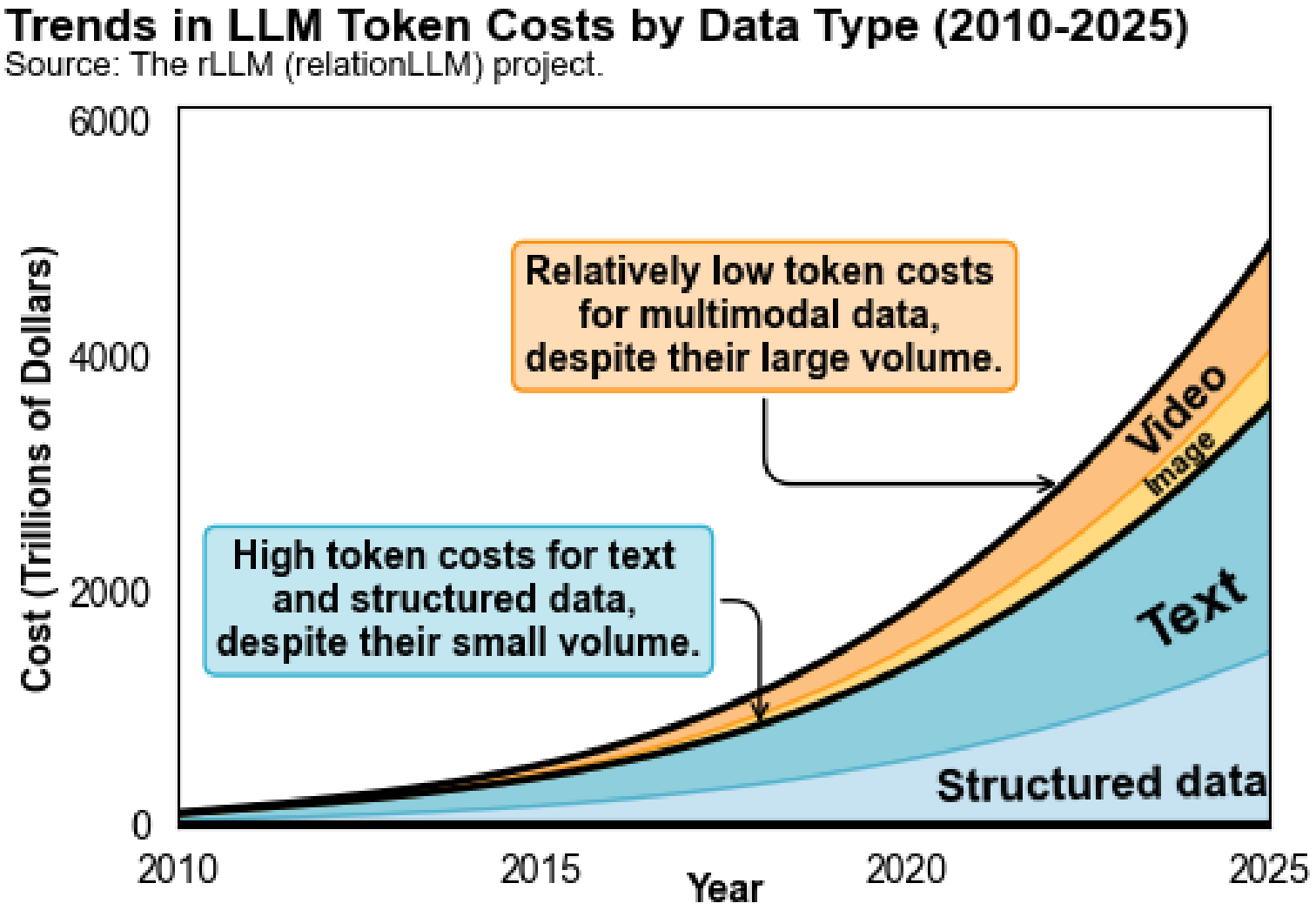
Outline

- **Background**
- rLLM (relationLLM)
- Use case

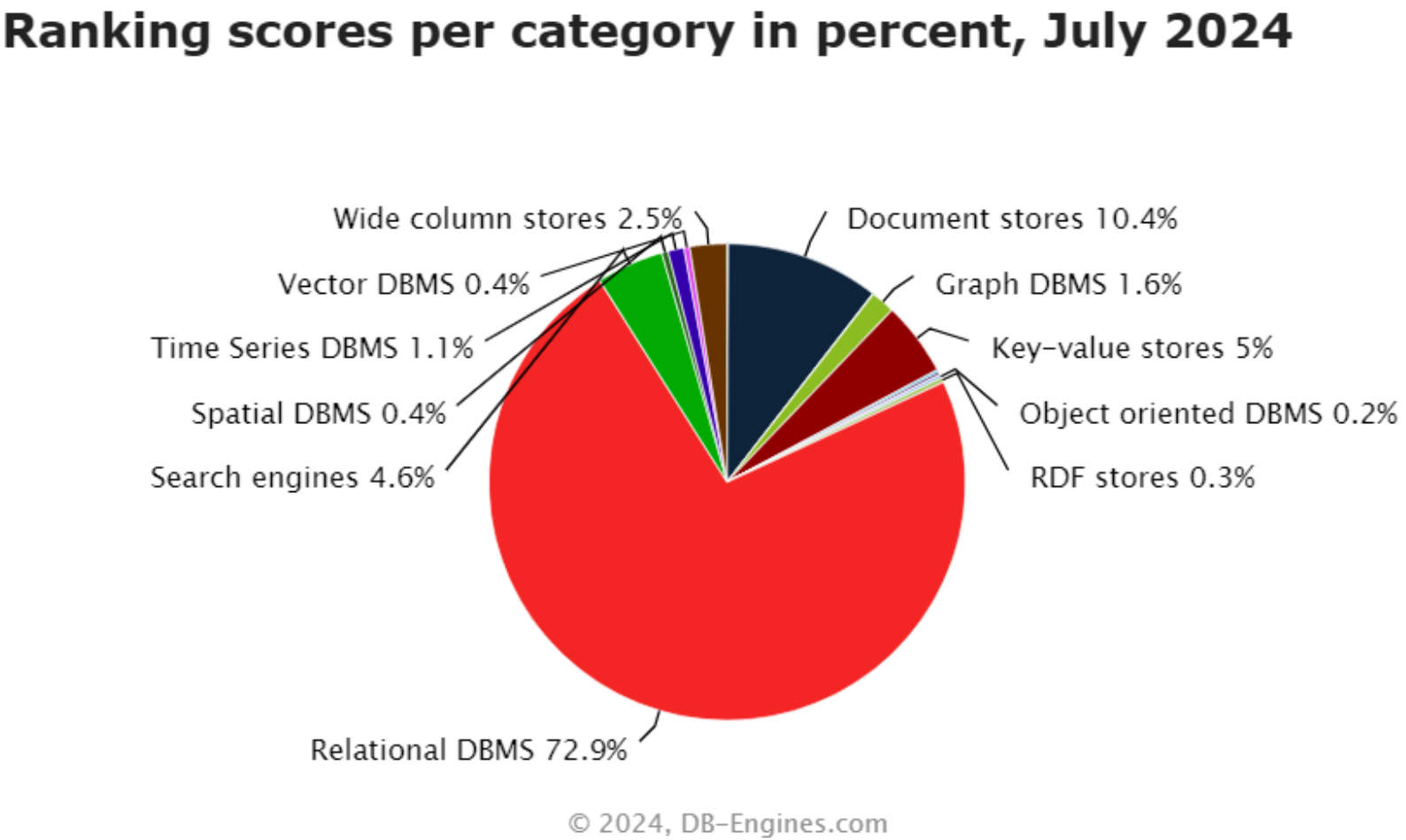
Big Data Trends



LLMs are Costly



Relational databases domain the world data

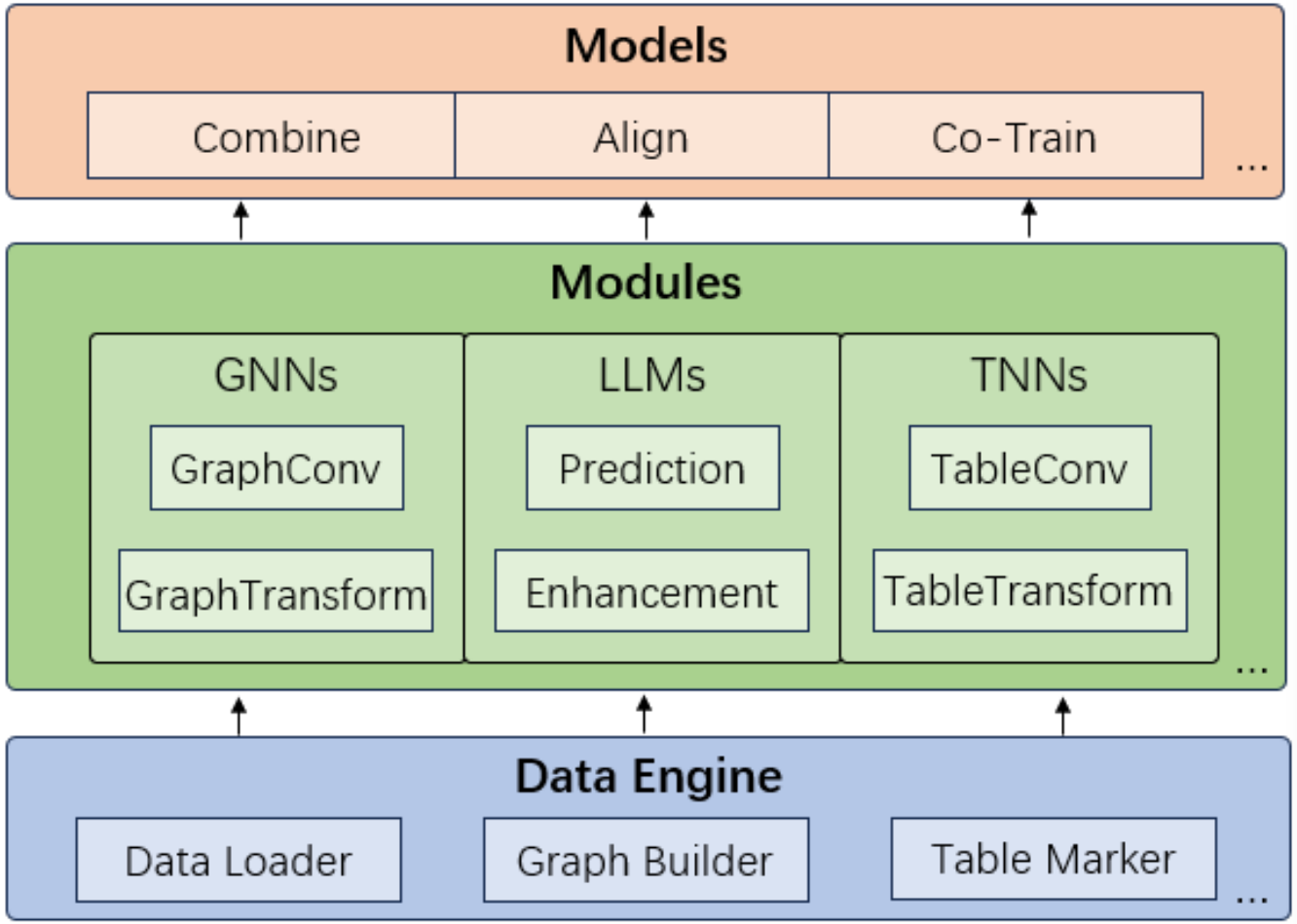


In the area of LLM, we need to pay great attention to Relational Table Learning (RTL).

Outline

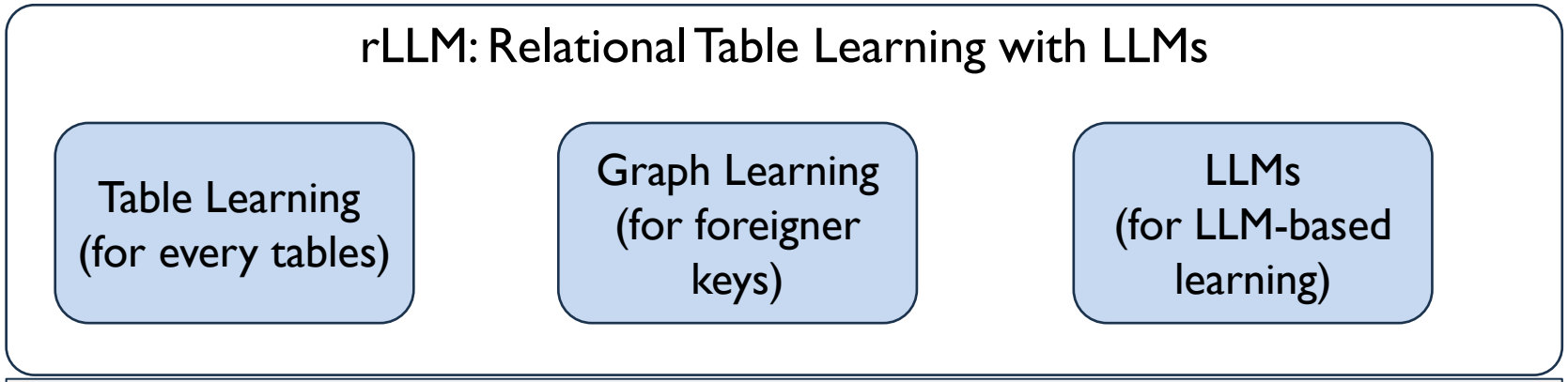
- Background
- **rLLM (relationLLM)**
- Use case

Overview of rLLM



The rLLM Overall Architecture

The Key of rLLM (relationLLM)



rLLM (relationLLM) is an easy-to-use Pytorch library for Relational Table Learning with LLMs, by performing two key functions:

- Breaks down state-of-the-art GNNs, LLMs, and TNNs as standardized modules.
- Facilitates novel RTL model building in a "combine, align, and co-train“ way.

Project page: https://github.com/rllm-project/rllm

Highlight Features:

- **LLM-friendly**: Modular interface designed for LLM-oriented applications, integrating smoothly with LangChain and Hugging Face transformers.
- **One-Fit-All Potential**: Processes various graphs (like social/communication/e-commerce graphs) by treating them as multiple tables linked by foreigner keys.
- **Novel Datasets**: Introduces three new relational table datasets useful for RTL model design. Includes the standard classification task, with examples.
- **Community Support**: Maintained by students and teachers from Shanghai Jiao Tong University and Tsinghua University. Supports the SJTU undergraduate course "Content Understanding (NIS4301)" and the graduate course "Social Network Analysis (NIS8023)".

Outline

- Background
- rLLM (relationLLM)
- **Use case**

Three new relational table datasets with standard classification tasks

- TML1M is derived from the classical MovieLens1M dataset.
- TLF2K is derived from the classical LastFM2K dataset.
- TACM12K is derived from the ACM heterogeneous graph dataset.

Dataset	Tables [#row/#col]	Relation Tables	Label	Classes	#Train/#Val/#Test
TML1M	users [6,040/5] movies [3,883/11] ratings [1,000,209/4]	ratings: user-movie	Age range of user	7	[140/500/1000]
TLF2K	artists [9,047/10] user_artists [80,009/3] user_friends [12,717/3]	user_artists: user-artist user_friends: user-user	Genre of artist	11	[220/500/1000]
TACM12K	papers [12,499/5] authors [17,431/3] citations [30,789/2] writings [37,055/2]	citations: paper-paper writings: paper-author	Conference of paper	14	[280/500/1000]

An illustration RTL method - BRIDGE

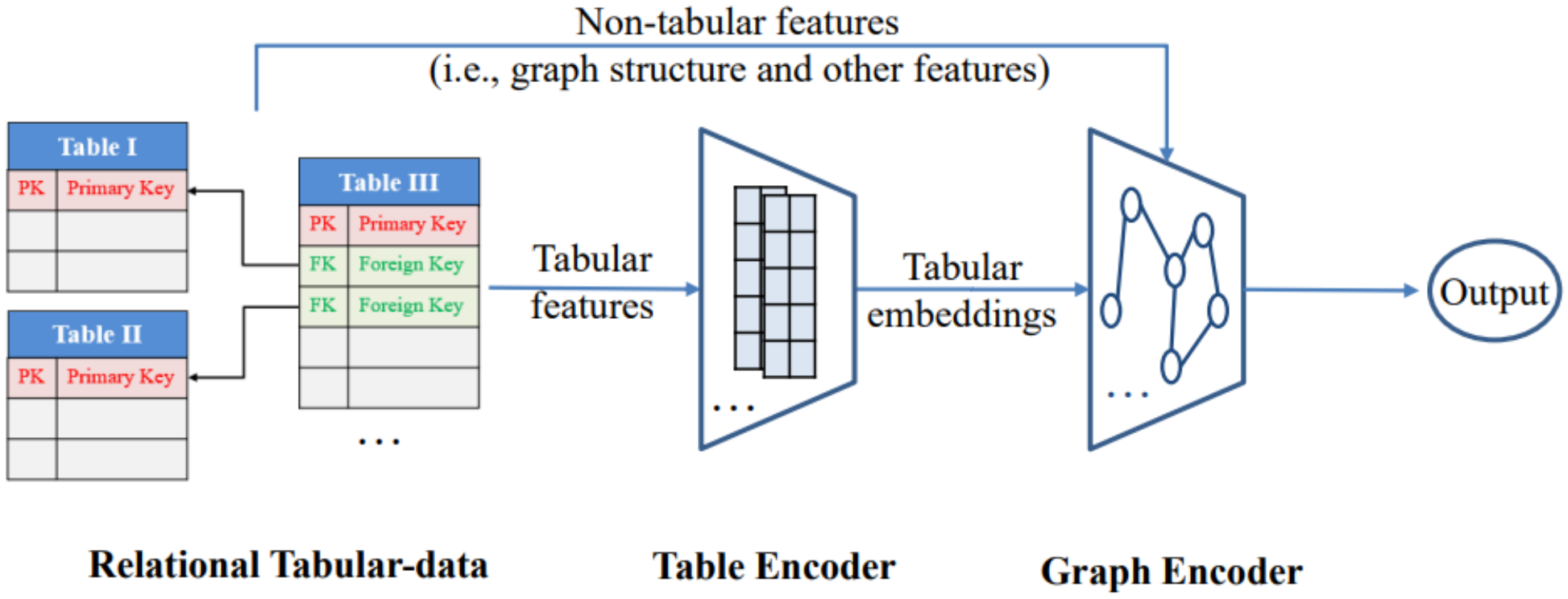


Figure 4: The architecture of BRIDGE

BRIDGE utilizes TNNs to process table data and leverages the "foreign keys" in relational tables to construct relationships between table samples, which are then analyzed using GNNs.

Pseudo-code of BRIDGE

```python
from rllm.nn.conv.graph_conv import GCNConv
from rllm.nn.conv.table_conv import TabTransformerConv

graph_encoder = GraphEncoder(GCNConv, ...)
table_encoder = TableEncoder(TabTransformerConv, ...)

class GraphTableBridge:
    def __init__(self, table_encoder, graph_encoder):
        self.table_encoder = table_encoder
        self.graph_encoder = graph_encoder

    def forward(self, table_data, non_table_data, adj):
        table_embeds = self.table_encoder(table_data)
        node_feats = COMBINE(table_embeds, non_table_data)
        return self.graph_encoder(node_feats, adj)
```

In practical implementations, the code lines are around 40.
Without rLLM, more than 400+ lines are needed!

Table 2: Classification accuracy.

Methods\Datasets	TML1M	TLF2K	TACM12K
Random	0.144±0.01	0.091±0.03	0.075±0.00
TabTransformer	0.347±0.02	0.137±0.08	0.142±0.01
TabNet	0.259±0.08	0.135±0.03	0.120±0.02
FT-Transformer	0.352±0.02	0.132±0.01	0.128±0.01
BRIDGE	**0.428±0.02**	**0.454±0.01**	**0.309±0.02**

How to try

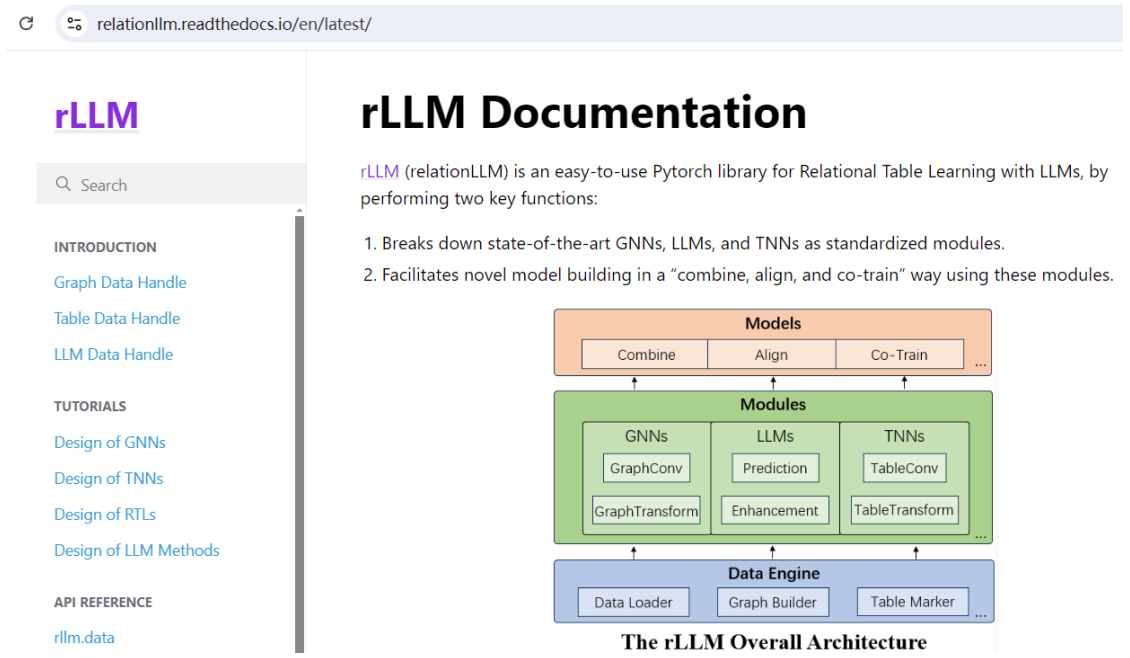
```
# cd ./examples

# set parameters if necessary


python bridge/bridge_tml1m.py

python bridge/bridge_tlf2k.py

python bridge/bridge_tacm12k.py
```



Project page: https://github.com/rllm-project/rllm

Core develop team

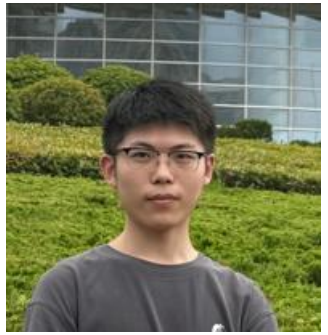
Zheng Wang
(Teacher)

Jianwu Zheng (PhD)

Feiyu Pan (PhD)

Zhe Kan (PhD)

Weichen Li (MS)



Ken Zhong (MS)

Zhichao Luo (MS)

Thanks for your time.
QA.