



Exploring the Potential of Linguistic Linked Data in the LLM Era

Patricia Martín Chozas, PhD.

Postdoctoral Researcher at Ontology Engineering Group
Assistant Professor at Universidad Politécnica de Madrid

✉ pmchozas@fi.upm.es

🐦 [@pmchozas](https://twitter.com/pmchozas)

📅 28th February 2024

📍 ENDORSE follow up, online

O1

Overview of LLOD & NLP

O2

Pros & Cons of LLM

O3

Uses of LLOD by LLM

01

“Traditional” Data

Heterogeneous inner structure

Heterogeneous/private formats

Isolated resources

Different access points

01

Linked (Open) Data

Standardised inner
structure

Standardised
publication formats

Interlinked

Single access points

01

Linguistic Linked (Open) Data

Ontolex

For
Dictionaries

SKOS

For
Thesauri

**TBX in RDF
Termlex**

For
Terminologies

OLiA

For Ling.
Annotation

NIF

For
Corpora

01

Generic

WordNet
BabelNet
ConceptNet
Wikidata
DBpedia
DBnary

Specific

UNESCO
STW
TheSoz
ILO
EuroVoc

- **Translations**
- **Synonyms**
- **Definitions**
- **Hierarchical relations**
- ...

01

English

Traduzca al...



bn:00030630n | [Sustantivo](#) [Concepto](#) | Categorías: Labour law, Pages using sidebar with the chil...

EN **employment contract** • employment agreement



Contract between [employer](#) and [employee](#) WordNet 3.0

+ See more



DEFINICIONES

RELACIONES

FUENTES

English



Más idiomas...

EN Contract between [employer](#) and [employee](#) WordNet 3.0 & Open English WordNet

An employment contract or contract of employment is a kind of [contract](#) used in [labour law](#) to attribute [rights](#) and responsibilities between parties to a bargain. [Wikipedia](#)

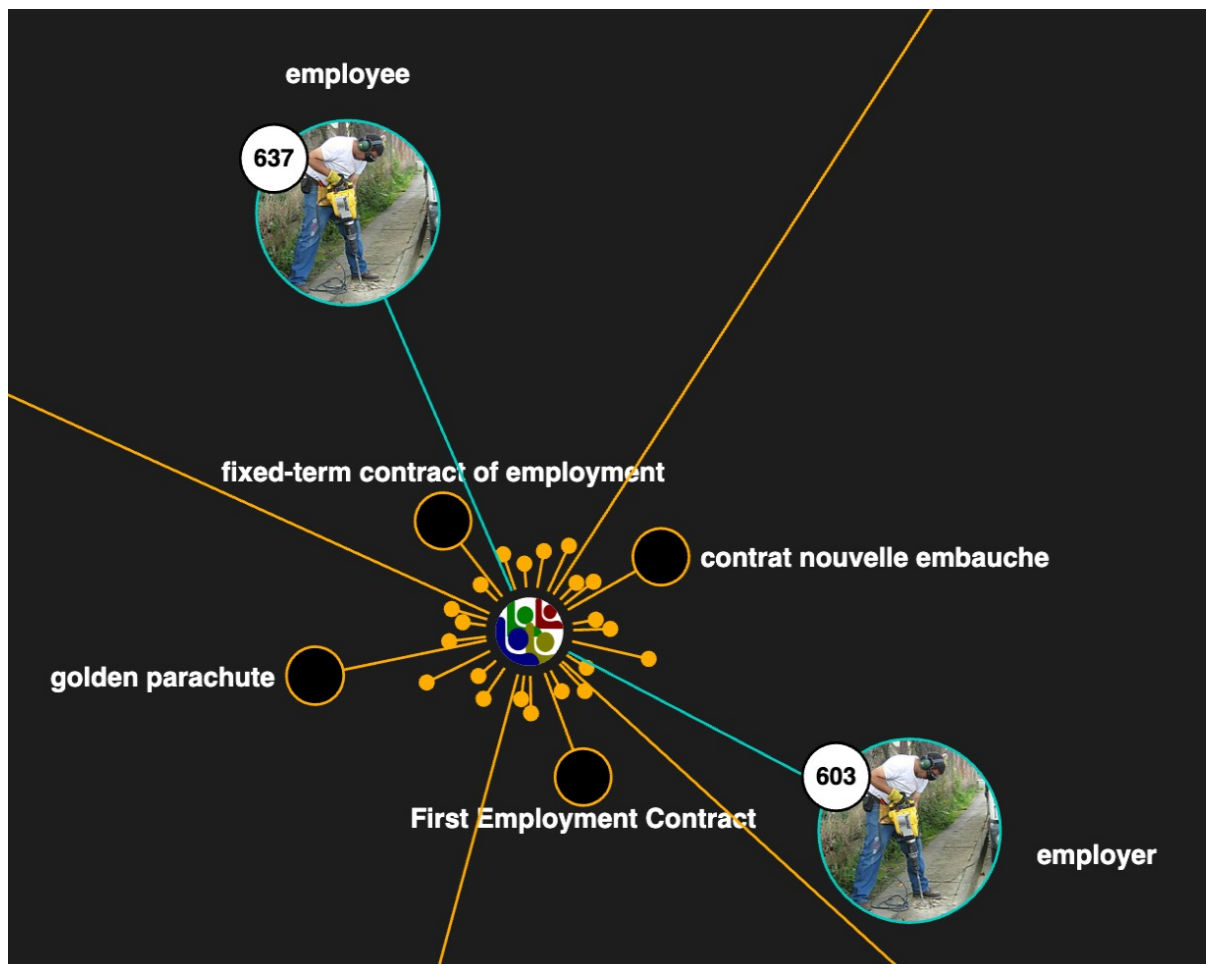
Agreement between employer and employee on terms of work and compensation [Wikidata](#)

A category of contract used in labor law to attribute right and responsibilities between parties to a bargain. [Wiktionary](#)

Category of contract. [Wiktionary \(translation\)](#)

01

Inner Links



External Links

English

>

Más idiomas...



WordNet 3.0 & Open English WordNet

EN

employment contract, employment agreement



Wikipedia

EN

employment contract



Wikidata

EN

employment contract



Wiktionary

EN

employment contract

01

WordNet for
bullying **detection**
(Jahan et al. 2022)

DBnary for **NER**
(Vuth & Serasset.
2023)

UMLS for medical
tasks **prediction**
(Winter et al. 2022)

Dbpedia and
Eurovoc for **Relex**
(Shishaev et al.
2023)

Wikidata for **QA**
(Han et al. 2022)

DBpedia for **QA**
(Elahi et al. 2021)

- Rules
- Statistical Models
- Vectorial Models
- CNN
- LSTM (RNN)
- Transformers
- Attention mechanisms

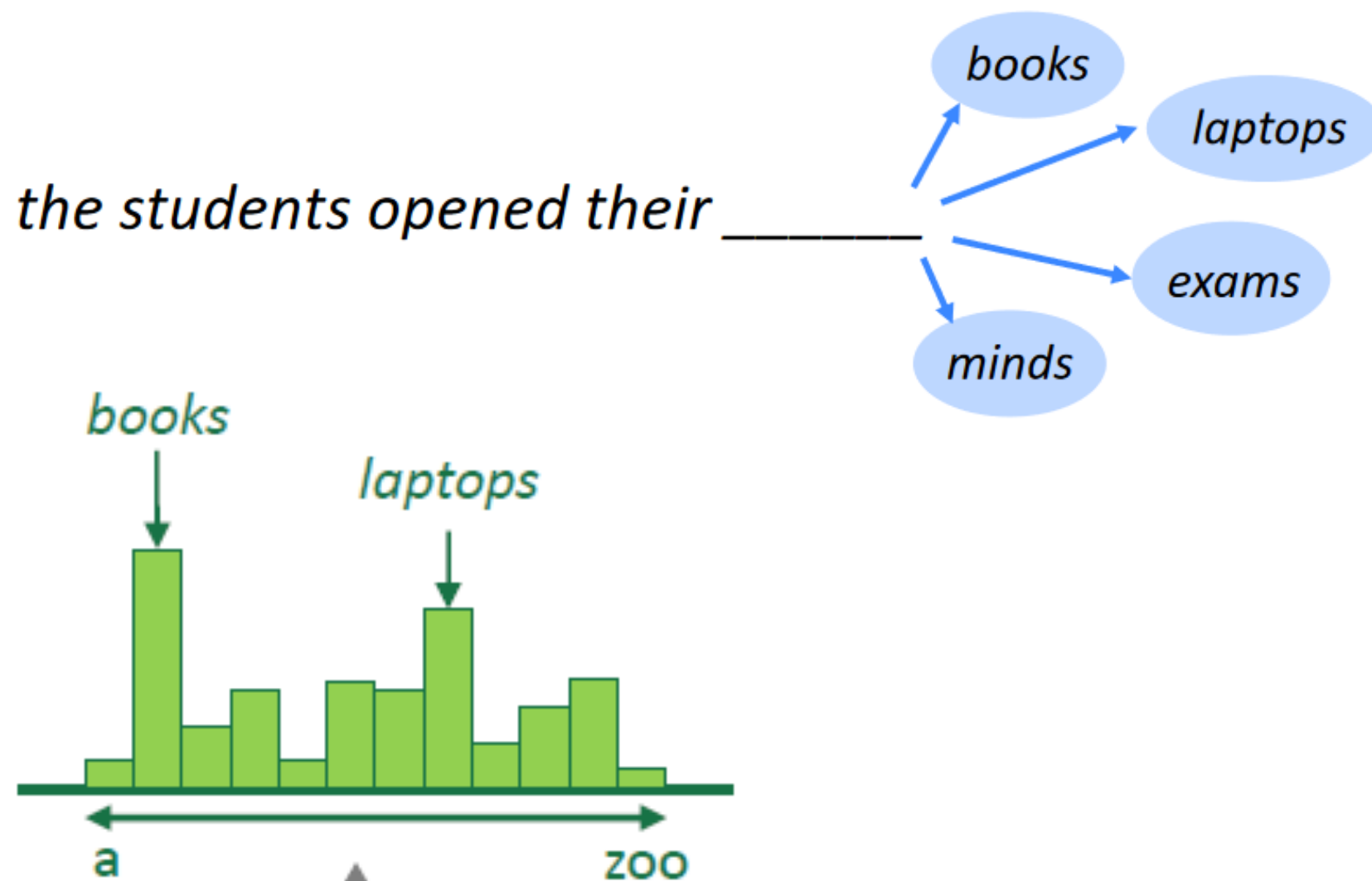
02

- Rules
- Statistical Models
- Vectorial Models
- CNN
- LSTM (RNN)
- Transformers
- Attention mechanisms

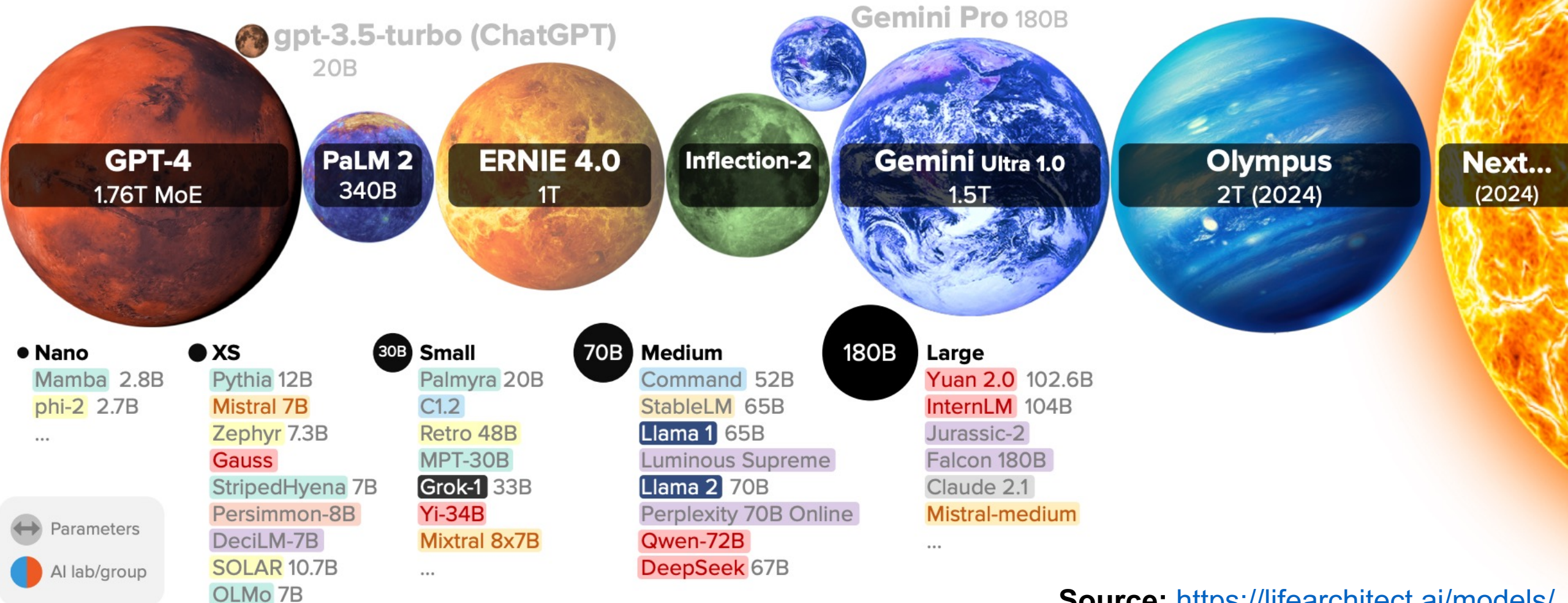
LLM

- Based on Transformers and Attention Mechanisms
- Probability distribution over words or word sequences

02



LARGE LANGUAGE MODEL HIGHLIGHTS (FEB/2024)



Source: <https://lifearchitect.ai/models/>

02

**LLM benefits
are...**

Multiple languages

Customization

Scalability

**Improvement over the
time***

**Efficiency in multiple
tasks***

02

**LLM limitations
are...**

**Hallucinations and
Biased results**

**Performance of small
models**

**Performance on small
corpora**

**Improvement over the
time***

**Efficiency in multiple
tasks***

03

Hallucinations

**Performance of small
models**

**Performance on small
corpora**

**Improvement over the
time***

**Efficiency in multiple
tasks***

**Structured
Knowledge**

Task

Entity Alignment: the task of finding entities in two knowledge bases that refer to the same real-world object.

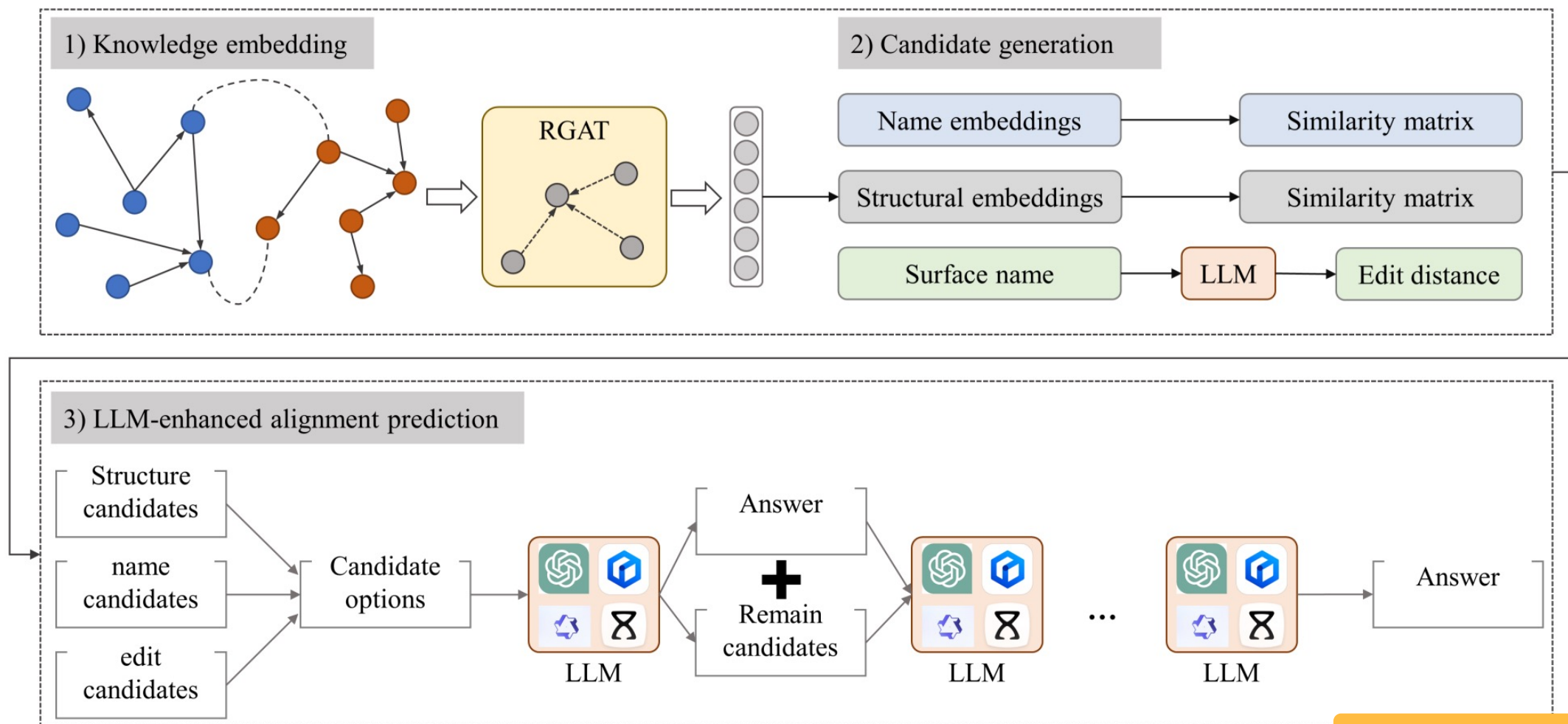
Approach

KG: Domain specific
GAN: Graph Attention Networks
LLM: ERNIE

Yang, L., Chen, H., Wang, X., Yang, J., Wang, F. Y., & Liu, H. (2024). Two Heads Are Better Than One: Integrating Knowledge from Knowledge Graphs and Large Language Models for Entity Alignment. arXiv preprint arXiv:2401.16960.

03

Task: Entity Alignment. Approach: KG + GAN + LLM (ERNIE)



Yang et al., 2023.

Task

NL2SPARQL: the task of translating Natural Language into SPARQL (query language)

Approach

KG: Wikidata and domain specific KG

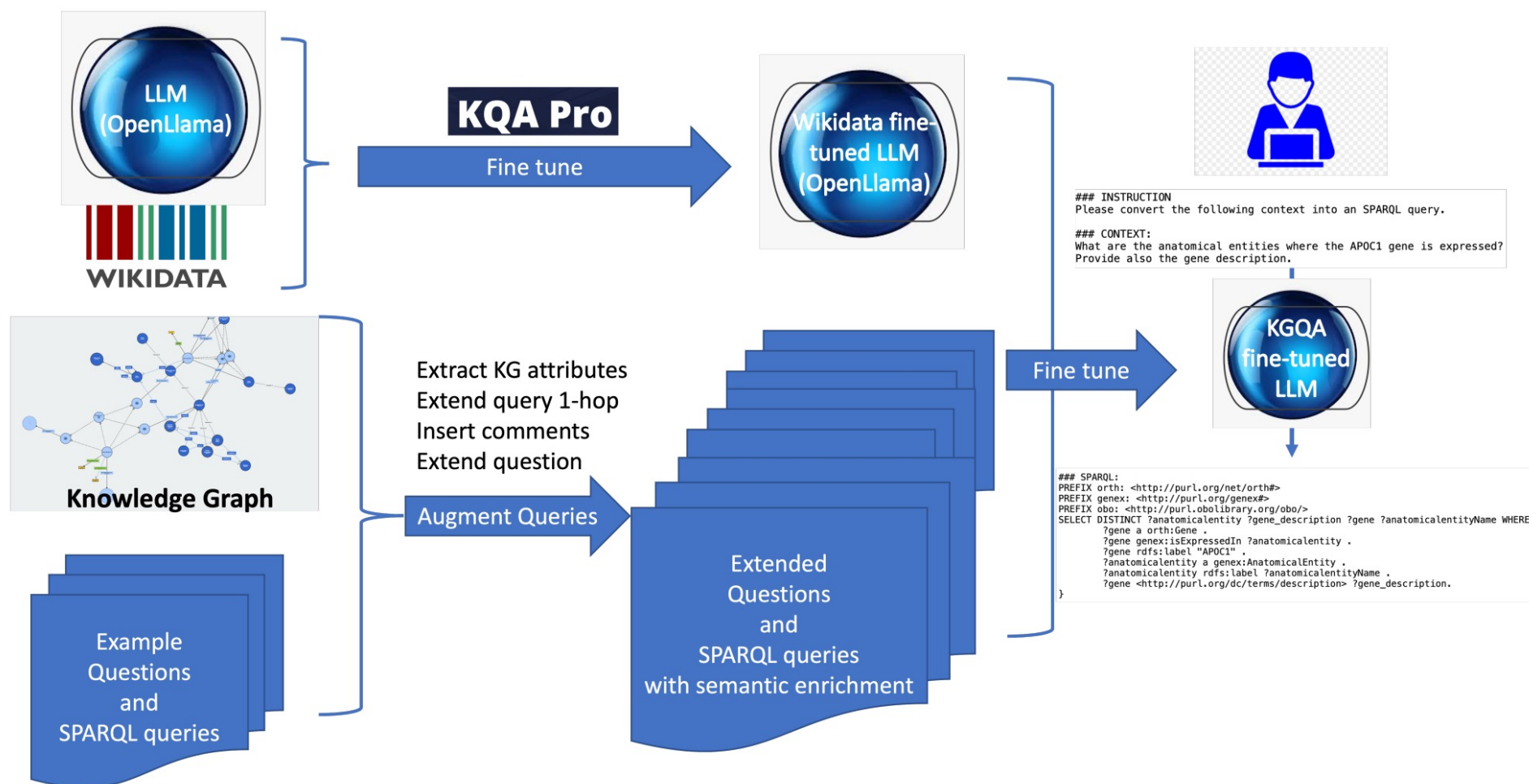
Fine tuning and Data Augmentation

LLM: OpenLLAMA

Rangel, J. C., de Farias, T. M., Sima, A. C., & Kobayashi, N. (2024). SPARQL Generation: an analysis on fine-tuning OpenLLaMA for Question Answering over a Life Science Knowledge Graph. *arXiv preprint arXiv:2402.04627*.

03

Task: NL2SPARQL. Approach: Wikidata + KG (DA) + LLM (OpenLLAMA)



Rangel et al., 2024.

Task

Disease Prediction: the task of assigning probabilities of certain diseases by analysing clinical records

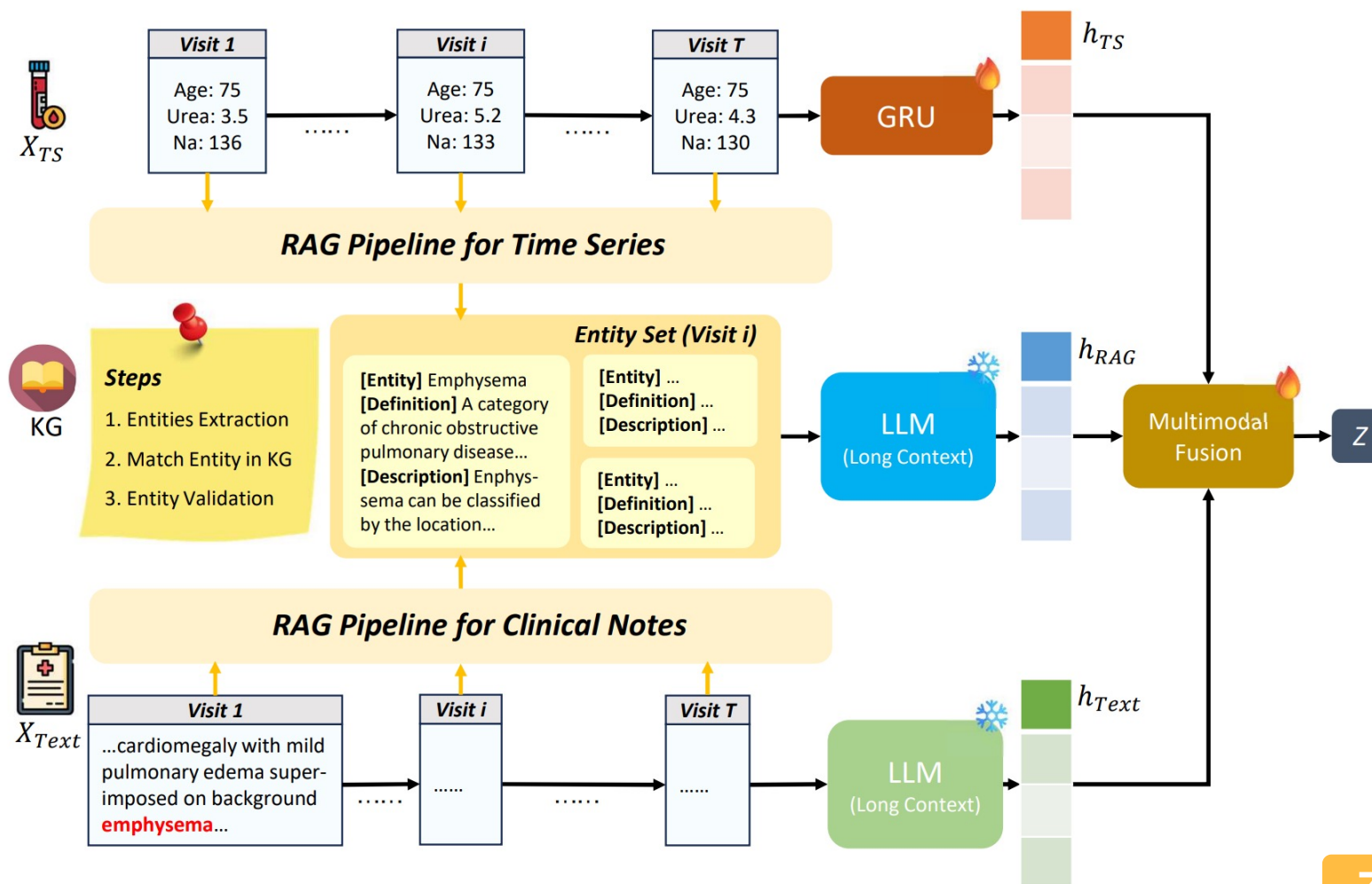
Approach

KG: Domain specific
RAG: Retrieval Augmented Generation
LLM: ?

Zhu, Y., Ren, C., Xie, S., Liu, S., Ji, H., Wang, Z., ... & Pan, C. (2024). REALM: RAG-Driven Enhancement of Multimodal Electronic Health Records Analysis via Large Language Models. arXiv preprint arXiv:2402.07016

03

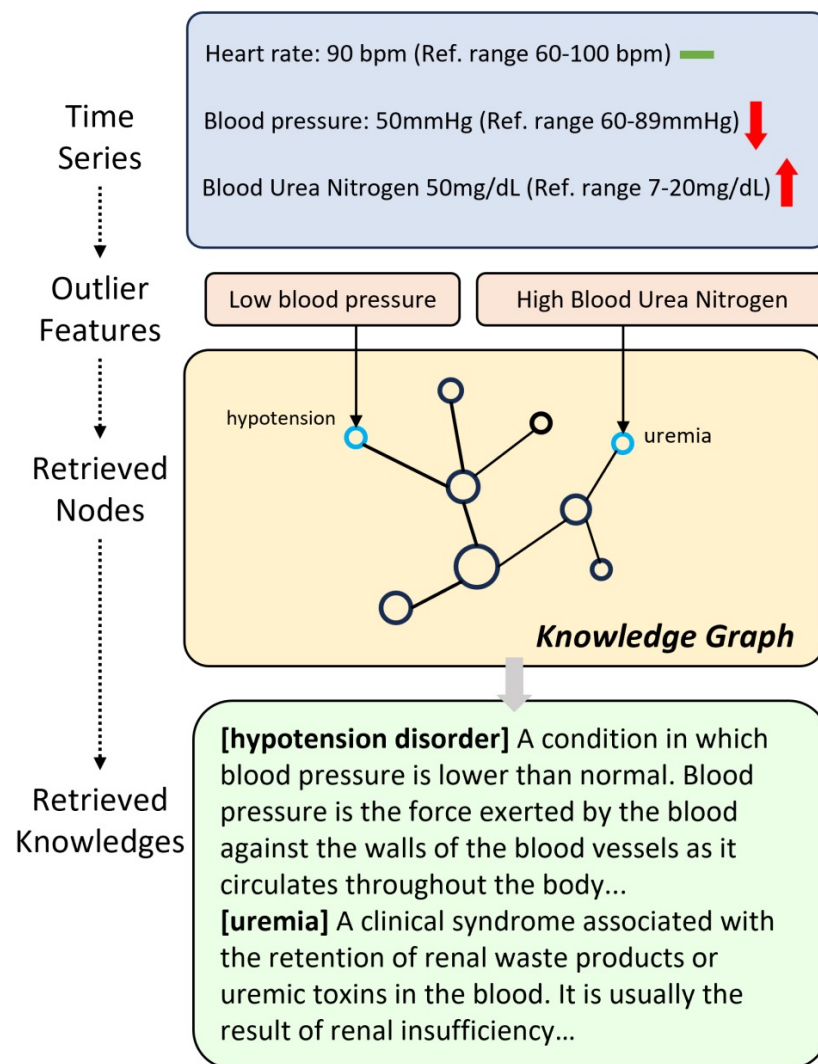
Task: Healthcare prediction. Approach: KG + RAG + LLM



Zhu et al., 2024.

O3

Task: Healthcare prediction. Approach: KG + RAG + LLM (OpenLLAMA)



Zhu et al., 2024.

Task

Node Classification: the task of predicting the label or category of a node in a graph. Useful for recommendation systems (i.e).

Approach

KG: Linguistic KG
KD: Knowledge Distillation
LLM: LLAMA, Mistral and others

Hu, S., Zou, G., Yang, S., Zhang, B., & Chen, Y. (2024). Large Language Model Meets Graph Neural Network in Knowledge Distillation. arXiv preprint arXiv:2402.05894.

03

Task: Classification. **Approach:** LKG for KD (GNN+LLM) (LLAMA+Mistral)

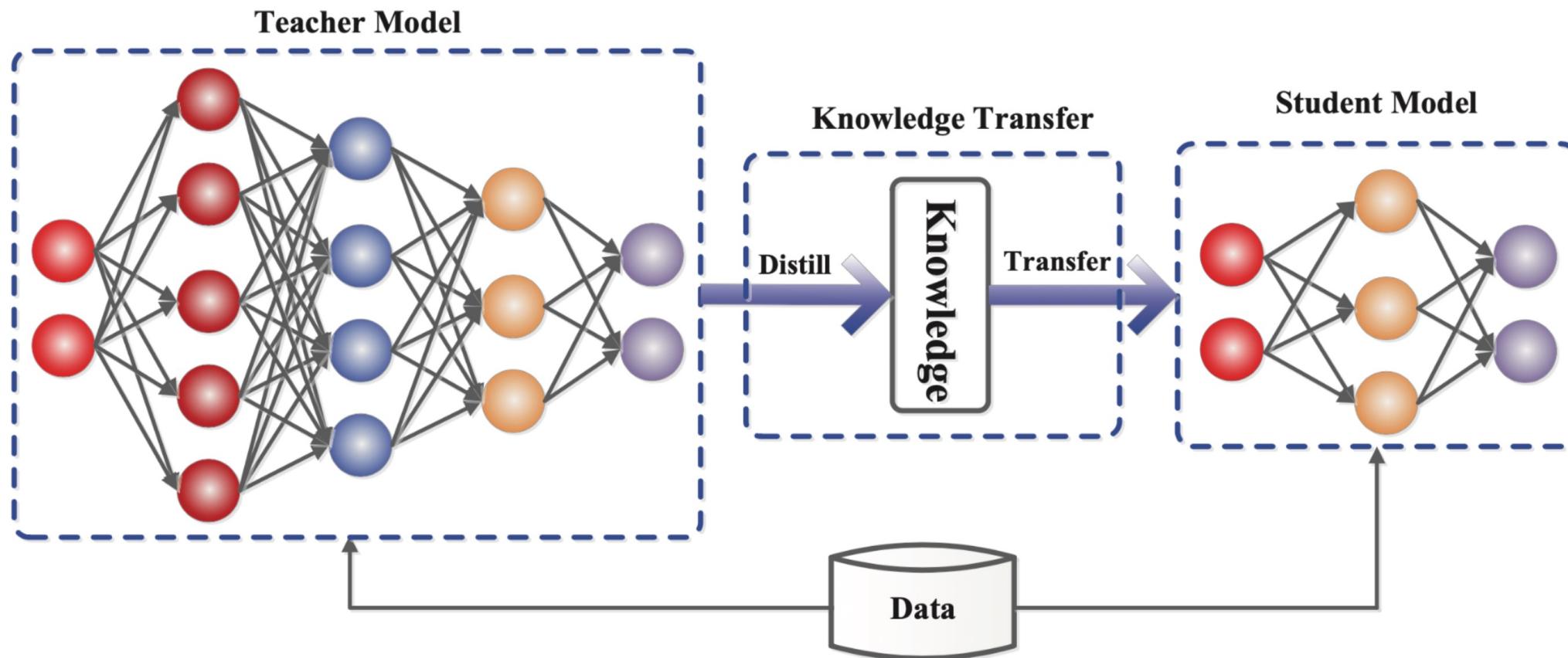


Image from: Jianping Gou et al., 2021

Hu et al., 2024.

Task

Question Answering: the task in which a computer needs to process questions and answers in natural language.

Approach

Different KI approaches

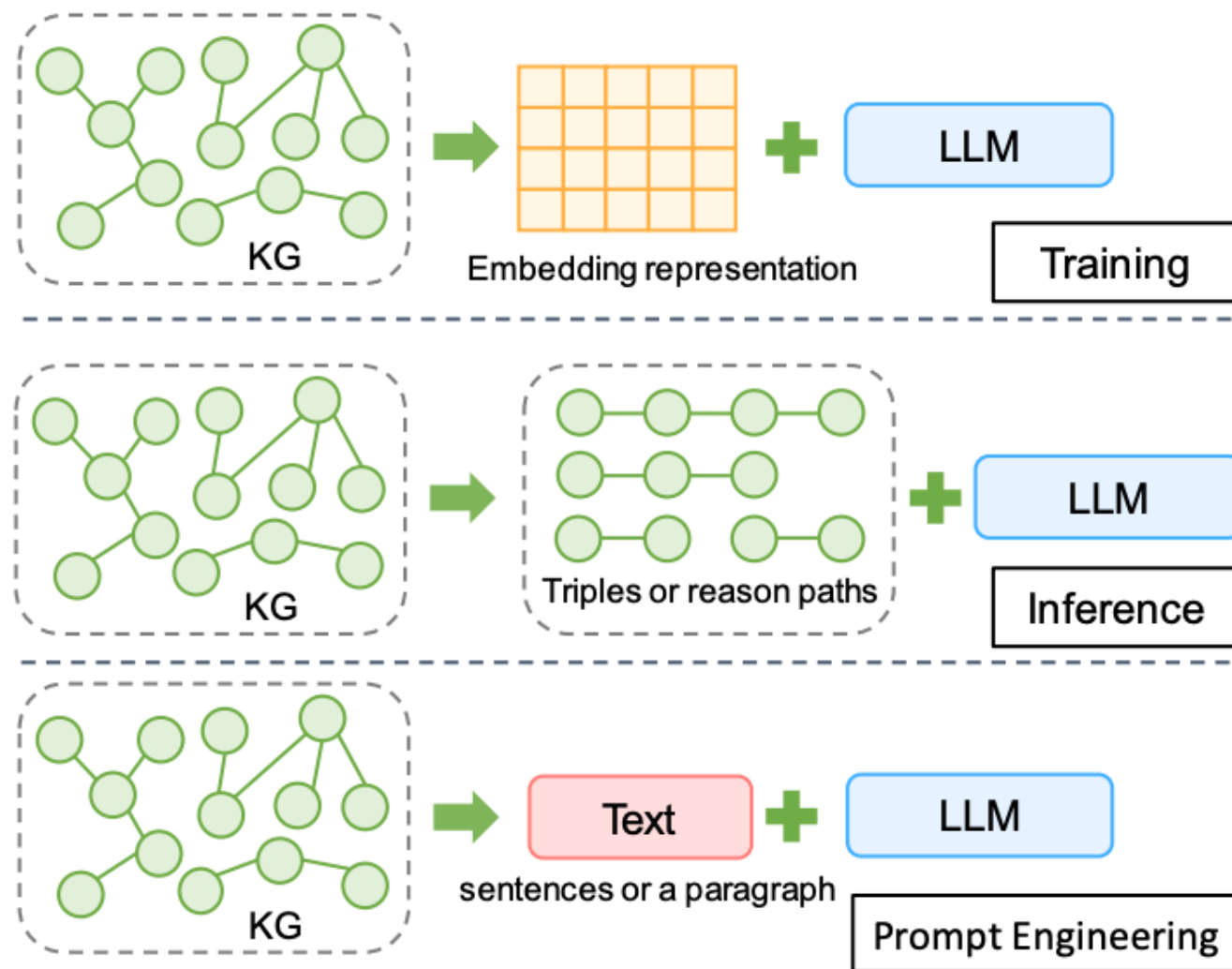
KG: Wikidata

LLM: ChatGPT and Vicuna

Dai, X., Hua, Y., Wu, T., Sheng, Y., & Qi, G. (2024). Counter-intuitive: Large Language Models Can Better Understand Knowledge Graphs Than We Thought. arXiv preprint arXiv:2402.11541.

O3

Task: QA. Different approaches: Wikidata + LLMs (CGPT+Vicuna)



Dai et al., 2024.

What conclusions could we extract from this analysis?

1. Factual Knowledge > Linguistic Knowledge

2. Focus on semantic relations and context

3. Towards Richer Lang. Resources

4. LK to enrich KGs

5. LK not in LLM but around

6. Lack of work on small corpora

- Structured knowledge still have a role to play
- We may not need to follow every trend, but to prove the usefulness of what we do despite the trend
- Not all problems need to be solved by LLM

“You don’t need an ocean liner to cross the Manzanares.”

Dr. Pablo Calleja, Senior Researcher at OEG



- Han, K., Ferreira, T. C., & Gardent, C. (2022, June). Generating questions from wikidata triples. In 13th Edition of its Language Resources and Evaluation Conference.
- Winter, B., Figueroa, A., Löser, A., Gers, F. A., & Siu, A. (2021). KIMERA: Injecting Domain Knowledge into Vacant Transformer Heads.
- Jahan, M. S., Beddiar, D. R., Oussalah, M., & Mohamed, M. (2022, June). Data Expansion Using WordNet-based Semantic Expansion and Word Disambiguation for Cyberbullying Detection. In Proceedings of the Thirteenth Language Resources and Evaluation Conference (pp. 1761-1770).
- Hakimov, S., Oto, S. A., & Dogdu, E. (2012, May). Named entity recognition and disambiguation using linked data and graph-based centrality scoring. In Proceedings of the 4th international workshop on semantic web information management (pp. 1-7).
- Vuth, N., & Serasset, G. (2023, September). DBnary2Vec: Preliminary Study on Lexical Embeddings for Downstream NLP Tasks. In Proceedings of the 4th Conference on Language, Data and Knowledge (pp. 417-427).
- Elahi, M. F., Ell, B., Grimm, F., & Cimiano, P. (2021). Question Answering on RDF Data based on Grammars Automatically Generated from Lemon Models. In SEMANTICS Posters&Demos.



Exploring the Potential of Linguistic Linked Data in the LLM Era

Patricia Martín Chozas, PhD.

Postdoctoral Researcher at Ontology Engineering Group
Assistant Professor at Universidad Politécnica de Madrid

✉ pmchozas@fi.upm.es

🐦 [@pmchozas](https://twitter.com/pmchozas)

📅 28th February 2024

📍 ENDORSE follow up, online