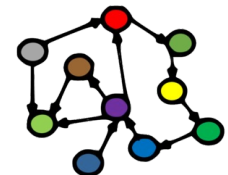


Welcome to INFO216:
Knowledge Graphs
Spring 2024

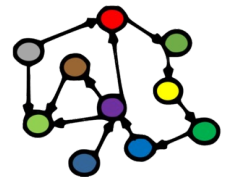
Andreas L Opdahl
<Andreas.Opdahl@uib.no>

Session 12: KGs and Large Language Models

- Themes:
 - What are Large Language Models (LLMs)
 - Combining KGs and Large Language Models (LLMs)
 - retrieval augmented knowledge fusion
 - end-to-end KG construction
 - LLM-augmented KG to text generation
- Readings:
 - No mandatory readings beyond these slides
 - Supplementary resource in the wiki <http://wiki.uib.no/info216>:
 - Pan et al. (2024): Unifying large language models and knowledge graphs: A roadmap.
 - Vaswani et al. (2017): Attention is all you need.



Readings



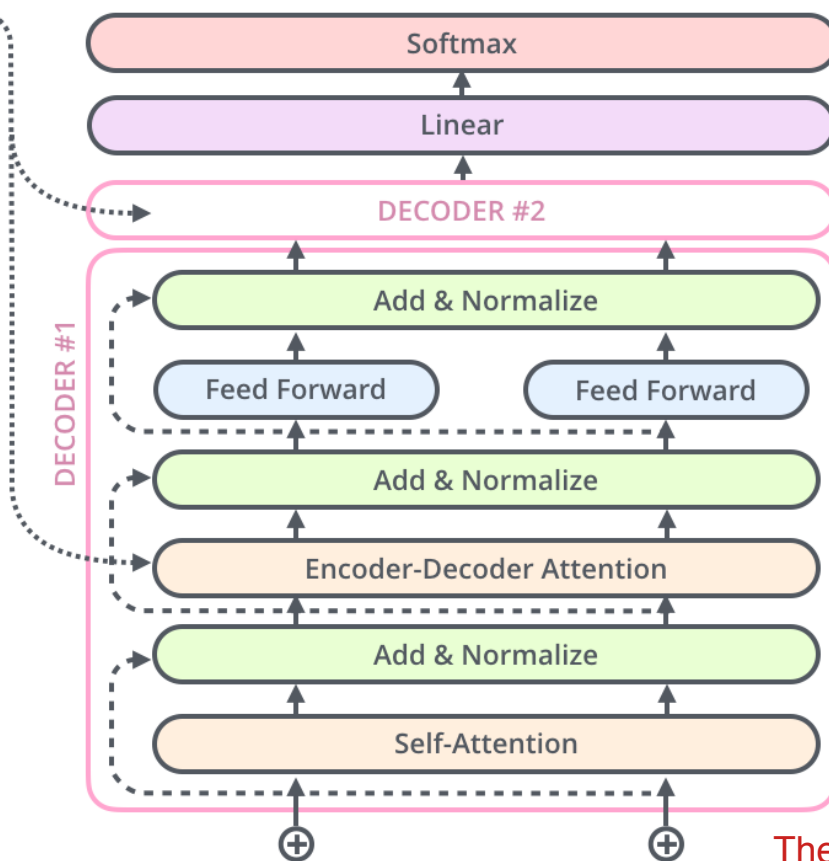
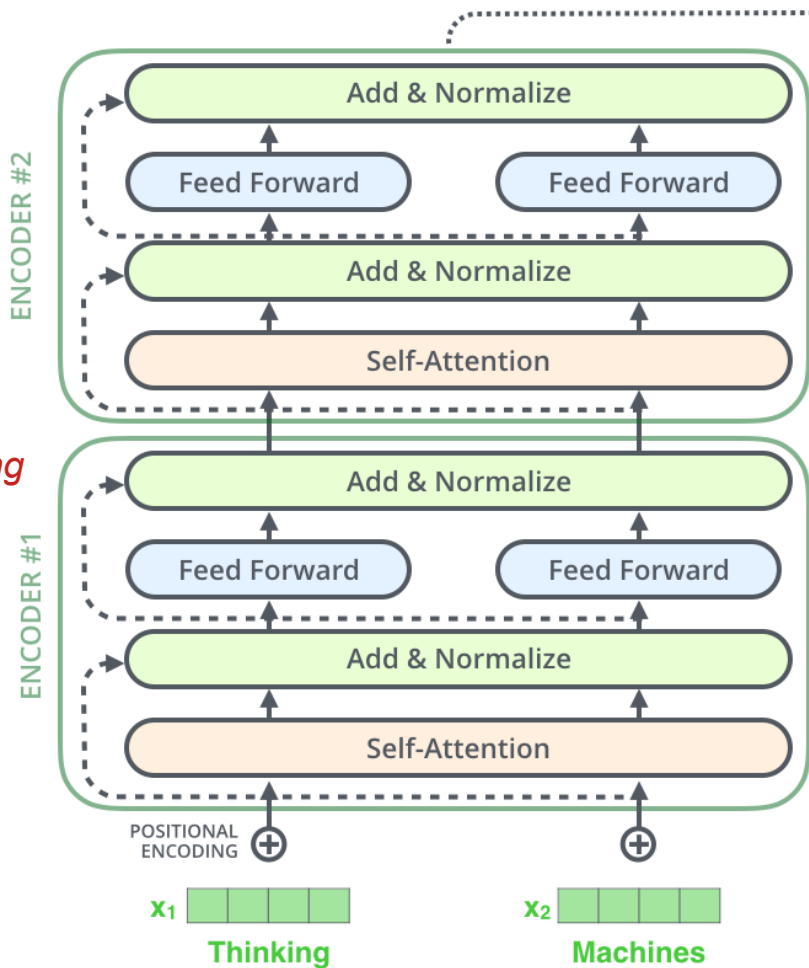
Large Language Models (LLMs)

Transformers

- A deep neural network architecture
 - many *encoder layers* transform the input text into an abstract representation (encoding as embeddings / vectors)
 - many *decoder layers* generate an output text from the abstract representation (encoding)
 - *attention modules* in the encoder and decoder layers are trained to make use of the context of each word
- Basic training
 - the transformer learns to regenerate its input text as output text via the abstract representation
 - *word masking*: predict the missing (masked) words (tokens) in the input text
 - *autoregression*: predict the next word in the input text
- Also different types of additional training: fine tuning on specific tasks and instruction tuning

Transformers

The encoder is great at *understanding* text

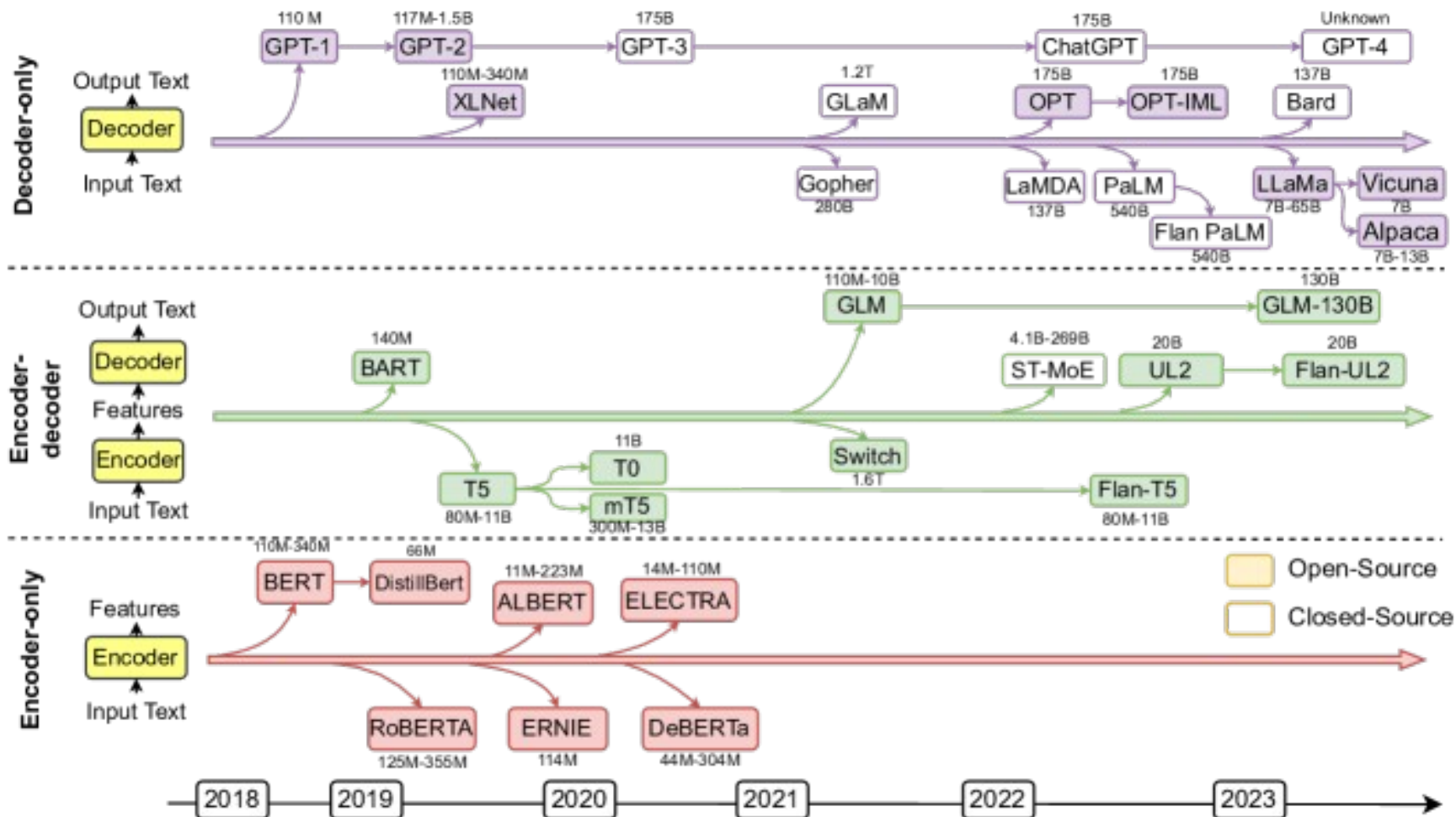


The decoder is great at *generating* text

Transformer variants

- **Autoencoder models**
 - encoder + decoder is trained to reconstruct masked input via the abstract representation (encoding)
 - then the decoder is thrown away
 - fine-tuning for tasks like dimensionality reduction, text classification, named entity recognition
- **Encoder + decoder**
 - can be trained like an autoencoder and the decoder is kept
 - versatile and flexible in generating texts and can be instruction aligned
- **Autoregressive models**
 - decoder (only) is trained to predict the next token from previous tokens
 - well-suited for tasks like language generation
 - can be instruction aligned

Evolution of LLMs

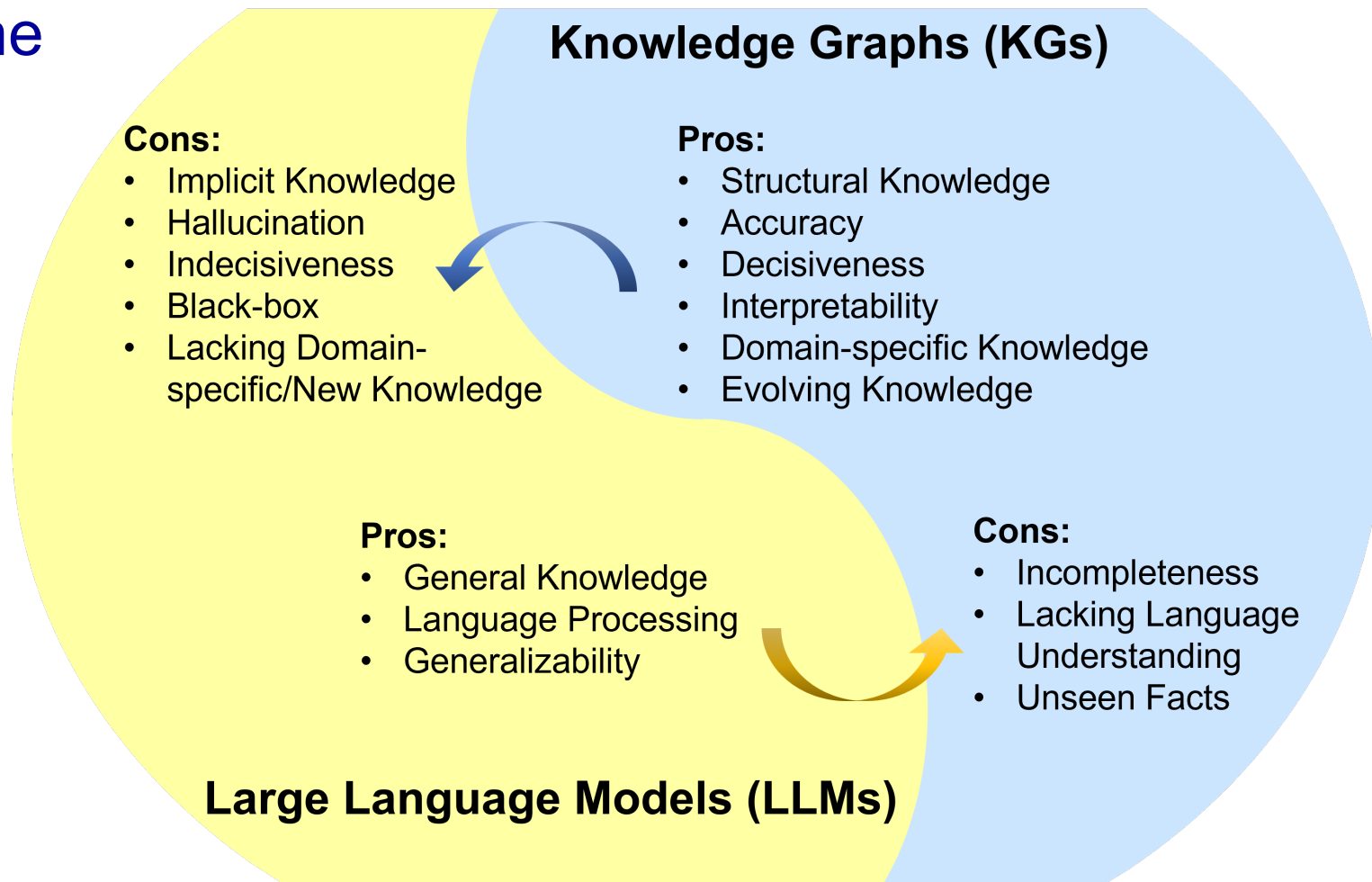


Transformer model sizes

Model	Layers	Dimensions	Context Window	Parameters	Encoder	Decoder
BERT-Base	12	768	512	110M	x	
BERT-Large	24	1024	512	340M	x	
T5-Base	12	768	512	220M	x	x
T5-3B	24	1024	512	3B	x	x
<u>GPT-3</u>	96	12288	2048	175B		x
ChatGPT (<u>GPT-3.5</u>)	96	12288	2048	~175B		x
Flan-T5-Base	12	768	512	220M	x	x
Flan-T5-11B	32	2048	2048	11B	x	x
<u>GPT-4</u>	96	25088	32768	~1T		x
<u>LLaMA-7B</u>	32	1024	2048	7B		x
<u>LLaMA-65B</u>	64	8192	4096	65B		x
<u>Vicuna-7B</u>	32	4096	2048	7B	x	x
<u>Vicuna-65B</u>	80	8192	4096	65B	x	x
Gemini-7B	32	2048	2048	7B		x
Gemini 1.0 Pro			32768			x
Gemini 1.5 Pro (Preview)			"multiple millions"	"~ <u>GPT-4</u> "		x

KGs + LLMs

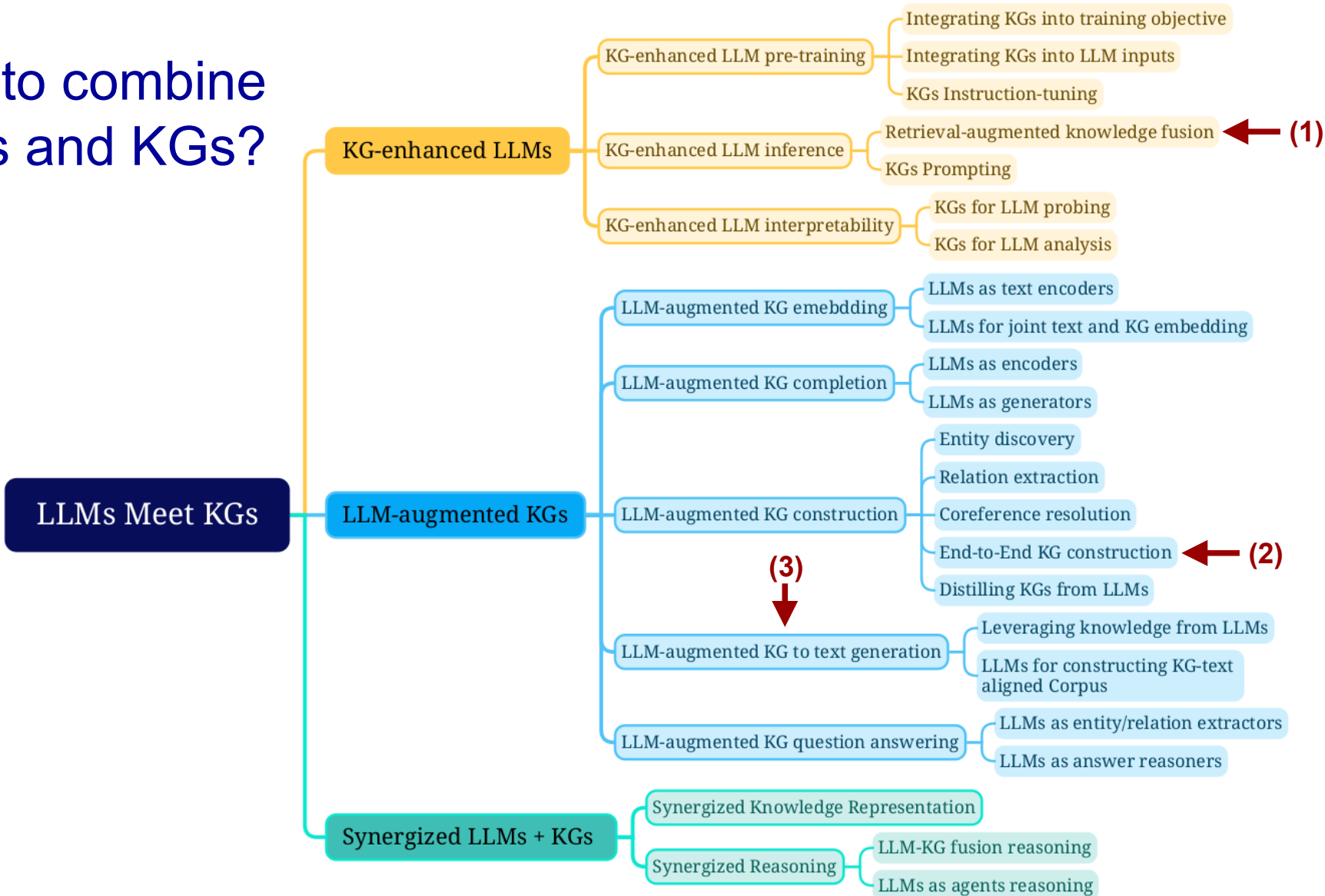
Why combine LLMs and KGs?



Some uses of LLMs and KGs.

Name	Category	LLMs	KGs	URL
ChatGPT/GPT-4	Chat Bot	✓		https://shorturl.at/cmsE0
ERNIE 3.0	Chat Bot	✓	✓	https://shorturl.at/sCLV9
Bard	Chat Bot	✓	✓	https://shorturl.at/pDLY6
Firefly	Photo Editing	✓		https://shorturl.at/fkzJV
AutoGPT	AI Assistant	✓		https://shorturl.at/bkoSY
Copilot	Coding Assistant	✓		https://shorturl.at/IKLUV
New Bing	Web Search	✓		https://shorturl.at/bimps
Shop.ai	Recommendation	✓		https://shorturl.at/aICY/
Wikidata	Knowledge Base		✓	https://shorturl.at/lyMY5
KO	Knowledge Base		✓	https://shorturl.at/sx238
OpenBG	Recommendation		✓	https://shorturl.at/pDMV9
Doctor.ai	Health Care Assistant	✓	✓	https://shorturl.at/dhIKU

How to combine LLMs and KGs?

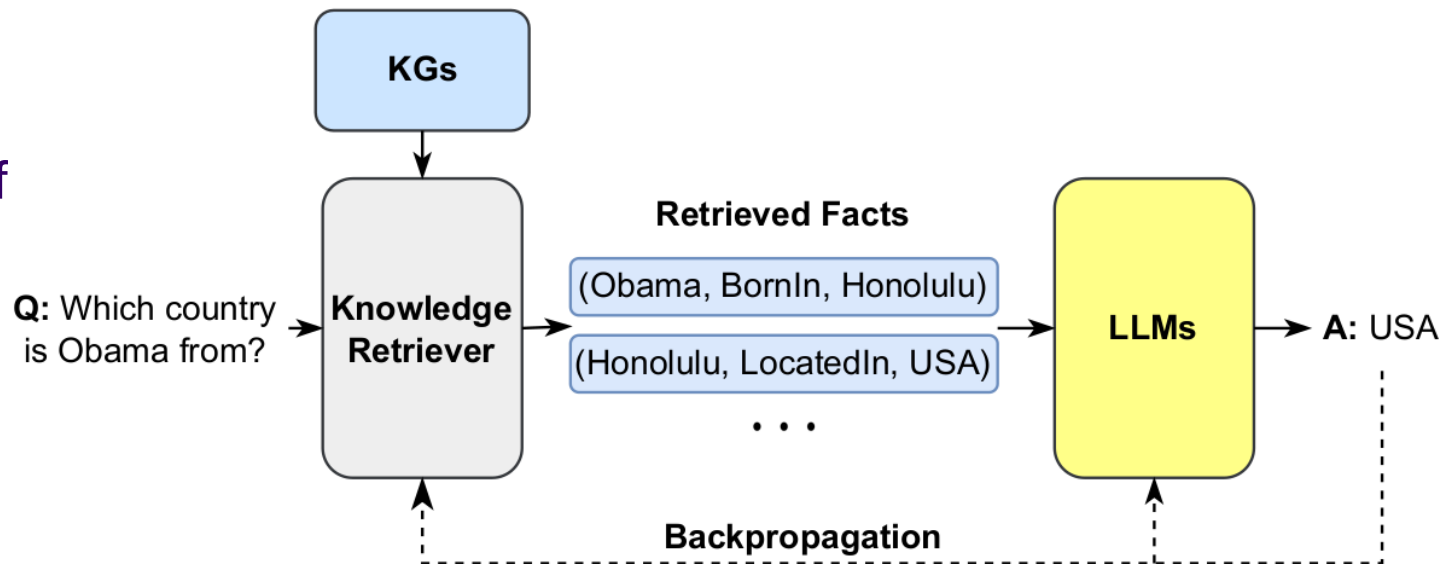


Retrieval-augmented knowledge fusion

- Problem
 - LLMs lack specific knowledge
 - LLMs hallucinate, may not generalise well
 - LLMs are not continuously updated
- Solution
 - keep text and knowledge and text spaces separate
 - retrieve relevant KG facts as supplementary input to LLM
- Graph Retrieval Augmented Generation (GRAG)
 - aka “Retrieval-Augmented Knowledge Fusion”
- Retrieval Augmented Generation (RAG) can also be done with plain text infusion

Graph RAG

- 1) Receive prompt P
- 2) Identify entities in prompt P
- 3) Query KG about entities
- 4) Receive related KG facts
- 5) Enhance prompt P with KG facts
- 6) Send enhanced P to LLM
- 7) Receive response
- 8) Perhaps: align response text with KG facts
- 9) Return response to user





You

Who won the Eurovision Song Contest 2024 and with which song?

Graph RAG example



ChatGPT Classic

As of my last update in December 2023, I don't have the results of the Eurovision Song Contest 2024. You might want to check the latest updates on a reliable news source or the official Eurovision website for the most recent information.





You

Who won the Eurovision Song Contest 2024 and with which song?



ChatGPT Classic

As of my last update in December 2023, I don't have the results of the Eurovision Song Contest 2024.

You might want to check the latest updates on a reliable news source or the official Eurovision website for the most recent information.



Eurovision Song Contest 2024 (Q114565318)

68th edition of the Eurovision Song Contest

ESC 2024 | Eurovision 2024 | ESC2024 | 4אאגטעכז | עכווטו



▼ In more languages

Configure

Language	Label	Description	Also known as
English	Eurovision Song Contest 2024	68th edition of the Eurovision Song Contest	ESC 2024 Eurovision 2024 ESC2024 4אאגטעכז עכווטו
Norwegian Bokmål	Eurovision Song Contest 2024	68. utgave av Eurovision Song Contest	ESC 2024 Eurovision 2024
Norwegian	No label defined	No description defined	
Norwegian Nynorsk	Eurovision Song Contest 2024	68. utgåve av Eurovision Song Contest	ESC 2024 Eurovision 2024

Graph RAG example

Use a Named Entity Recogniser (NER), like DBpedia Spotlight:

- <https://www.wikidata.org/entity/Q114565318>
- retrieve <https://www.wikidata.org/entity/Q114565318.ttl>
 - size: >110Kb, >4000 lines

AN

You

Who won the Eurovision Song Contest 2024 and with which song?

Please use the following trusted facts as background:

###

sLabel; pwLabel; oLabel

Eurovision Song Contest 2024; alternative label; Eurovision 2024

Eurovision Song Contest 2024; alternative label; ESC2024

Eurovision Song Contest 2024; country; Sweden

Eurovision Song Contest 2024; instance of; Eurovision Song Contest edition

Eurovision Song Contest 2024; genre; music television

Eurovision Song Contest 2024; follows; Eurovision Song Contest 2023

...

###

+ >4000 more lines

Graph RAG example

Use a Named Entity Recogniser (NER), like DBpedia Spotlight:

- <https://www.wikidata.org/entity/Q114565318>
- retrieve <https://www.wikidata.org/entity/Q114565318.ttl>
 - size: >110Kb, >4000 lines



ChatGPT Classic

The message you submitted was too long, please reload the conversation and submit something shorter.



```
PREFIX wd: <http://www.wikidata.org/entity/>
```

```
SELECT ?sLabel ?pwLabel ?oLabel WHERE {
```

```
  BIND(wd:Q114565318 AS ?s)  
  ?s ?p ?o .
```

```
  FILTER (  
    (STRSTARTS(STR(?p), STR(wdt:)) &&  
     STRSTARTS(STR(?o), STR(wd:)) || DATATYPE(?o) = xsd:string)  
    ||  
    (?p IN (rdfs:label, skos:altLabel, schema:description) &&  
     DATATYPE(?o) = rdf:langString && LANG(?o) = "en")  
  )
```

```
  BIND (IRI(REPLACE(STR(?p), STR(wdt:), STR(wd:))) AS ?pw)  
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
```

```
} LIMIT 100
```

Do a Wikidata SPARQL
query to only retrieve
“mostly truthful” triples
about wd:Q114565318 .

*Remember
Lab 6...*

AN

You

Who won the Eurovision Song Contest 2024 and with which song?

Please use the following trusted facts as background:

###

sLabel; pwLabel; oLabel

Eurovision Song Contest 2024; alternative label; Eurovision 2024

Eurovision Song Contest 2024; alternative label; ESC2024

Eurovision Song Contest 2024; country; Sweden

Eurovision Song Contest 2024; instance of; Eurovision Song Contest edition

Eurovision Song Contest 2024; genre; music television

Eurovision Song Contest 2024; follows; Eurovision Song Contest 2023

...

###

+ 70 more lines

**ChatGPT Classic**

The winner of the Eurovision Song Contest 2024 was **Nemo**. Unfortunately, the specific song they won with isn't listed in the information you provided. For more details about their winning song and performance, you might need to check additional sources like the official Eurovision website or a reliable media outlet covering the event.



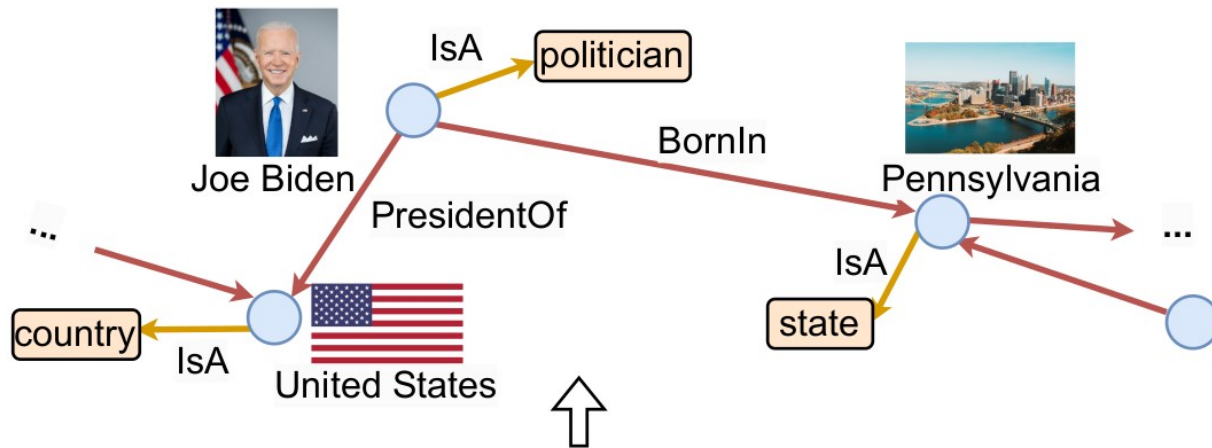
Use a Named Entity Recogniser (NER), like DBpedia Spotlight:

- <https://www.wikidata.org/entity/Q114565318>
- retrieve <https://www.wikidata.org/entity/Q114565318.ttl>
 - size: >110Kb, >4000 lines
- do a Wikidata SPARQL query to only retrieve “mostly truthful” triples about wd:Q114565318
 - size: 4.6Kb, 76 lines

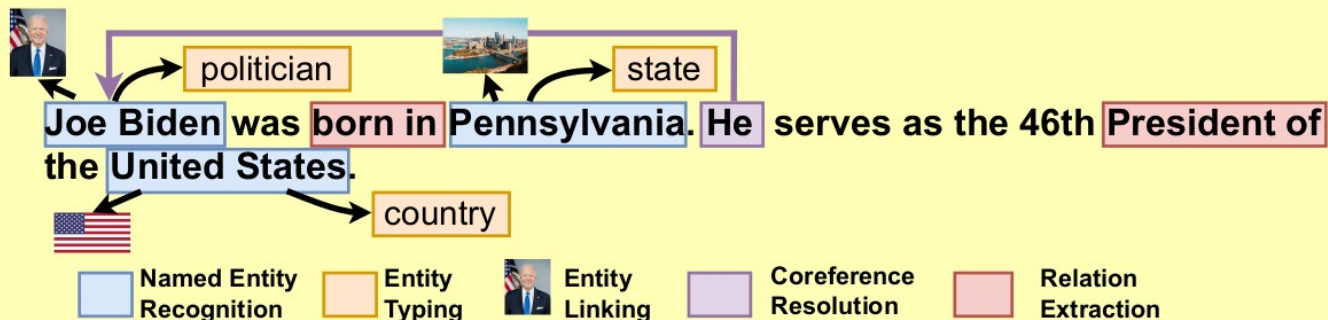
LLM-augmented KG Construction

- Piecewise KG construction
 - 1) entity discovery (incl NER)
 - 2) coreference resolution
 - 3) relation extraction
- End-to-End KG construction
 - can also be piecewise

Knowledge Graph



LLM-based Knowledge Graph Construction



Text: Joe Biden was born in Pennsylvania. He serves as the 46th President of the United States.

End-to-End KG construction example



ChatGPT Classic

The winner of the Eurovision Song Contest 2024 was **Nemo**. Unfortunately, the specific song they won with isn't listed in the information you provided. For more details about their winning song and performance, you might need to check additional sources like the official Eurovision website or a reliable media outlet covering the event.



Who won the Eurovision Song Contest 2024 and



Switzerland wins Eurovision, Ireland comes sixth

Switzerland's Nemo was crowned the Eurovision Song Contest 2024 winner following an eventful day in Malmö, Sweden, while Ireland's Bambie Thug finished sixth place.

The Swiss non-binary artist triumphed at this year's contest with their track The Code, marking the third time Switzerland triumphed at the contest. The country won the very first edition of the Eurovision Song Contest in 1956 and in 1988 with Celine Dion's Ne Partez Pas Sans Moi.

Tearfully accepting the trophy, Nemo said: "I want to say thank you so much – I hope this contest can live up to its promise and continue to stand for peace and dignity for every person in this world."



...and with this text, we could do plain (non-graph) RAG

The screenshot shows a Google search interface. The search bar contains the text "Who won the Eurovision Song Contest 2024 and". Below the search bar, the "All" tab is selected. The search results show a snippet from RTE.ie, dated "2 hours ago", with the text: "Switzerland's Nemo was crowned the Eurovision Song Contest 2024 winner following an eventful day in Malmö, Sweden, while Ireland's Bambie Thug finished sixth place. The Swiss non-binary artist triumphed at this year's contest with their track The Code, marking the third time Switzerland triumphed at the contest." The word "winner" in the snippet is highlighted in blue. At the bottom of the search results, there are links for "About featured snippets" and "Feedback".

Step 1: Identify entities and concepts

AN

You

What are the main concepts and entities mentioned in the following text?

###

Switzerland wins Eurovision, Ireland comes

Switzerland's Nemo was crowned the Eurovision Song Contest 2024 in a highly eventful day in Malmö, Sweden, while Ireland's

The Swiss non-binary artist triumphed at marking the third time Switzerland triumphed at the edition of the Eurovision Song Contest in Malmö, Sweden, with the song "Sans Moi."

"Tearfully accepting the trophy, Nemo said: "The Eurovision Song Contest can live up to its promise and connect people from all over the world. I am proud to represent my country and my person in this world."

###



ChatGPT

The main concepts and entities mentioned in the text are:

1. **Eurovision Song Contest 2024** - A music competition event where countries from Europe (and sometimes beyond) compete in singing.
2. **Switzerland** - The country that won the contest.
3. **Nemo** - The non-binary artist from Switzerland who won the contest.
4. **The Code** - The song performed by Nemo, which won the contest.
5. **Ireland** - The country whose representative finished sixth.
6. **Bambie Thug** - The artist representing Ireland, who finished sixth.
7. **Malmö, Sweden** - The location where the Eurovision Song Contest 2024 was held.
8. **Celine Dion** - Previous winner of the Eurovision Song Contest for Switzerland in 1988.
9. **Ne Partez Pas Sans Moi** - The song performed by Celine Dion that won the contest in 1988.
10. **Peace and dignity** - Values mentioned by Nemo in their acceptance speech, emphasizing the hope for the contest to promote these ideals globally.

Step 1: Identify entities and concepts in JSON syntax

AN

You

What are the main concepts and entities mentioned in the following text?

###

Switzerland wins Eurovision, Ireland comes sixth

Switzerland's Nemo was crowned the Eurovision Song Contest 2024 winner following an eventful day in Malmö, Sweden, while Ireland's Bambie Thug finished sixth place.

The Swiss non-binary artist triumphed at this year's contest with their track The Code, marking the third time Switzerland triumphed at the contest. The country won the very first edition of the Eurovision Song Contest in 1956 and in 1988 with Celine Dion's Ne Partez Pas Sans Moi.

Tearfully accepting the trophy, Nemo said: "I want to say thank you so much – I hope this contest can live up to its promise and continue to stand for peace and dignity for every person in this world."

###

Please provide the response in valid JSON format with the following schema: `[{'name': value, 'description': value}, {'name': value, 'description': value}, ...]`

Step 1: Identify entities and concepts in JSON syntax

AN

You

What are the main concepts and entities mentioned in the following text?



ChatGPT

json

Copy code

```
[
  {"name": "Eurovision Song Contest 2024", "description": "A music competition event"},
  {"name": "Switzerland", "description": "The country that won the contest."},
  {"name": "Nemo", "description": "The non-binary artist from Switzerland who won the contest"},
  {"name": "The Code", "description": "The song performed by Nemo, which won the contest"},
  {"name": "Ireland", "description": "The country whose representative finished sixth"},
  {"name": "Bambie Thug", "description": "The artist representing Ireland, who finished sixth"},
  {"name": "Malmö, Sweden", "description": "The location where the Eurovision Song Contest 2024 was held"},
  {"name": "Celine Dion", "description": "Previous winner of the Eurovision Song Contest"},
  {"name": "Ne Partez Pas Sans Moi", "description": "The song performed by Celine Dion"},
  {"name": "Peace and dignity", "description": "Values mentioned by Nemo in their song"}
]
```

Step 2: Identify relations between the entities and concepts

AN

You

What are the main relations found between these entity and concept names:

###

Eurovision Song Contest 2024

Switzerland

Nemo

The Code

Ireland

Bambie Thug

Malmö, Sweden

Celine Dion

Ne Partez Pas Sans Moi

Peace and dignity

###

in the following text:

###

Switzerland wins Eurovision, Ireland comes sixth

Switzerland's Nemo was crowned the Eurovision Song Contest 2024 winner following an eventful day in Malmö, Sweden, while Ireland's Bambie Thug finished sixth place.

The Swiss non-binary artist triumphed at this year's contest with their track The Code, marking the third time Switzerland triumphed at the contest. The country won the very first edition of the Eurovision Song Contest in 1956 and in 1988 with Celine Dion's Ne Partez Pas Sans Moi.

Tearfully accepting the trophy, Nemo said: "I want to say thank you so much – I hope this contest can live up to its promise and continue to stand for peace and dignity for every person in this world."

###

Please provide the response in valid JSON format with the following schema:

```
[{'from_name': value, 'to_name': value, 'relation': value, 'description': value}, {'from_name': value, 'to_name': value, 'relation': value, 'description': value}, ...]
```

Step 2: Identify relations between the entities and concepts

AN

You

What are the main relations found between these entity and concept names:

###

Eurovision Song Contest 2024

Switzerland

Nemo

The Code

Ireland

Bambie Thug

Malmö, Sweden

Celine Dion

Ne Partez Pas Sans Moi

Peace and dignity

###

in the following text:



ChatGPT

json

Copy code

```
[  
  {"from_name": "Switzerland", "to_name": "Eurovision Song Contest 2024", "relation": "host"},  
  {"from_name": "Nemo", "to_name": "Switzerland", "relation": "representative", "description": "Nemo is the representative of Switzerland in the Eurovision Song Contest 2024."},  
  {"from_name": "Nemo", "to_name": "The Code", "relation": "performer", "description": "Nemo is the performer of the song 'The Code'."},  
  {"from_name": "The Code", "to_name": "Eurovision Song Contest 2024", "relation": "song"},  
  {"from_name": "Ireland", "to_name": "Eurovision Song Contest 2024", "relation": "host"},  
  {"from_name": "Bambie Thug", "to_name": "Ireland", "relation": "representative", "description": "Bambie Thug is the representative of Ireland in the Eurovision Song Contest 2024."},  
  {"from_name": "Malmö, Sweden", "to_name": "Eurovision Song Contest 2024", "relation": "venue"},  
  {"from_name": "Celine Dion", "to_name": "Switzerland", "relation": "former winner", "description": "Celine Dion is a former winner of the Eurovision Song Contest, representing Switzerland in 1988."},  
  {"from_name": "Ne Partez Pas Sans Moi", "to_name": "Eurovision Song Contest 1988", "relation": "song"},  
  {"from_name": "Nemo", "to_name": "peace and dignity", "relation": "advocate", "description": "Nemo is an advocate for peace and dignity."}]
```

AN

You

Please list the triples in this JSON list:

```
###
```

```
...
```

```
###
```

Please provide the response in valid Python syntax with the following schema:

```
[(from_name, relation, to_name), (from_name, relation, to_name), ...]
```

**ChatGPT**

```
[
```

```
  ("Switzerland", "winner", "Eurovision Song Contest 2024"),
```

```
  ("Nemo", "representative", "Switzerland"),
```

```
  ("Nemo", "performer", "The Code"),
```

```
  ("The Code", "entry song", "Eurovision Song Contest 2024"),
```

```
  ("Ireland", "participant", "Eurovision Song Contest 2024"),
```

```
  ("Bambie Thug", "representative", "Ireland"),
```

```
  ("Malmö, Sweden", "host location", "Eurovision Song Contest 2024"),
```

```
  ("Celine Dion", "former winner", "Switzerland"),
```

```
  ("Ne Partez Pas Sans Moi", "winning song", "Eurovision Song Contest 1988"),
```

```
  ("Nemo", "advocate", "peace and dignity")
```

```
]
```

Graph RAG revisited



You

Graph RAG revisited

Who won the Eurovision Song Contest 2024 and with which song?

Please use the following trusted facts as background:

###

- ("Switzerland", "winner", "Eurovision Song Contest 2024"),
- ("Nemo", "representative", "Switzerland"),
- ("Nemo", "performer", "The Code"),
- ("The Code", "entry song", "Eurovision Song Contest 2024"),
- ("Ireland", "participant", "Eurovision Song Contest 2024"),
- ("Bambie Thug", "representative", "Ireland"),
- ("Malmö, Sweden", "host location", "Eurovision Song Contest 2024"),
- ("Celine Dion", "former winner", "Switzerland"),
- ("Ne Partez Pas Sans Moi", "winning song", "Eurovision Song Contest 1988"),
- ("Nemo", "advocate", "peace and dignity")

###



ChatGPT Classic

Switzerland won the Eurovision Song Contest 2024 with the song "The Code," performed by Nemo.



LLM-augmented KG-to-text Generation

AN

You

Please generate a text that contains the information in these triples, and no other information:

###

("Switzerland", "winner", "Eurovision Song Contest 2024"),
("Nemo", "representative", "Switzerland"),
("Nemo", "performer", "The Code"),
("The Code", "entry song", "Eurovision Song Contest 2024"),
("Ireland", "participant", "Eurovision Song Contest 2024"),
("Bambie Thug", "representative", "Ireland"),
("Malmö, Sweden", "host location", "Eurovision Song Contest 2024"),
("Celine Dion", "former winner", "Switzerland"),
("Ne Partez Pas Sans Moi", "winning song", "Eurovision Song Contest 1988"),
("Nemo", "advocate", "peace and dignity")

###



ChatGPT Classic

Switzerland emerged as the winner of the Eurovision Song Contest 2024, which was hosted in Malmö, Sweden. The winning entry for Switzerland was the song "The Code," performed by their representative, Nemo. Meanwhile, Ireland also participated in the contest, represented by Bambie Thug. Previously, Switzerland had won the contest with Celine Dion performing "Ne Partez Pas Sans Moi" in 1988. Nemo, apart from his performance, is recognized for advocating peace and dignity.

- Problem
 - make KGs useful in human interactions, e.g.:
 - story-telling
 - knowledge-based chat
- Solution
 - generate high-quality texts that accurately and consistently describe the KG information

KG-to-text Generation *with factual grounding*

AN

You

Please generate a text that narrates the information contained in these numbered triples and no other information:

###

1. ("Switzerland", "winner", "Eurovision Song Contest 2024"),
2. ("Nemo", "representative", "Switzerland"),
3. ("Nemo", "performer", "The Code"),
4. ("The Code", "entry song", "Eurovision Song Contest 2024"),
5. ("Ireland", "participant", "Eurovision Song Contest 2024"),
6. ("Bambie Thug", "representative", "Ireland"),
7. ("Malmö, Sweden", "host location", "Eurovision Song Contest 2024"),
8. ("Celine Dion", "former winner", "Switzerland"),
9. ("Ne Partez Pas Sans Moi", "winning song", "Eurovision Song Contest 1988"),
10. ("Nemo", "advocate", "peace and dignity")

###

At each point in the text where a triple is narrated, place the number of that triple in parentheses.

KG-to-text Generation *with factual grounding*

AN

You

Please generate a text that narrates the information contained in these numbered triples and no other information:

###

1. ("Switzerland", "winner", "Eurovision Song Contest 2024"),
2. ("Nemo", "representative", "Switzerland"),
3. ("Nemo", "performer", "The Code"),
4. ("The Code", "entry song", "Eurovision Song Contest 2024"),
5. ("Ireland", "participant", "Eurovision Song Contest 2024"),
6. ("Bambie Thug", "representative", "Ireland"),
7. ("Malmö, Sweden", "host location", "Eurovision Song Contest 2024"),
8. ("Celine Dion", "former winner", "Switzerland"),
9. ("Ne Partez Pas Sans Moi", "winning song", "Eurovision Song Contest 1988"),
10. ("Nemo", "advocate", "peace and dignity")

###

At each point in the text where a triple is narrated, place the number of that triple in parentheses.



ChatGPT Classic

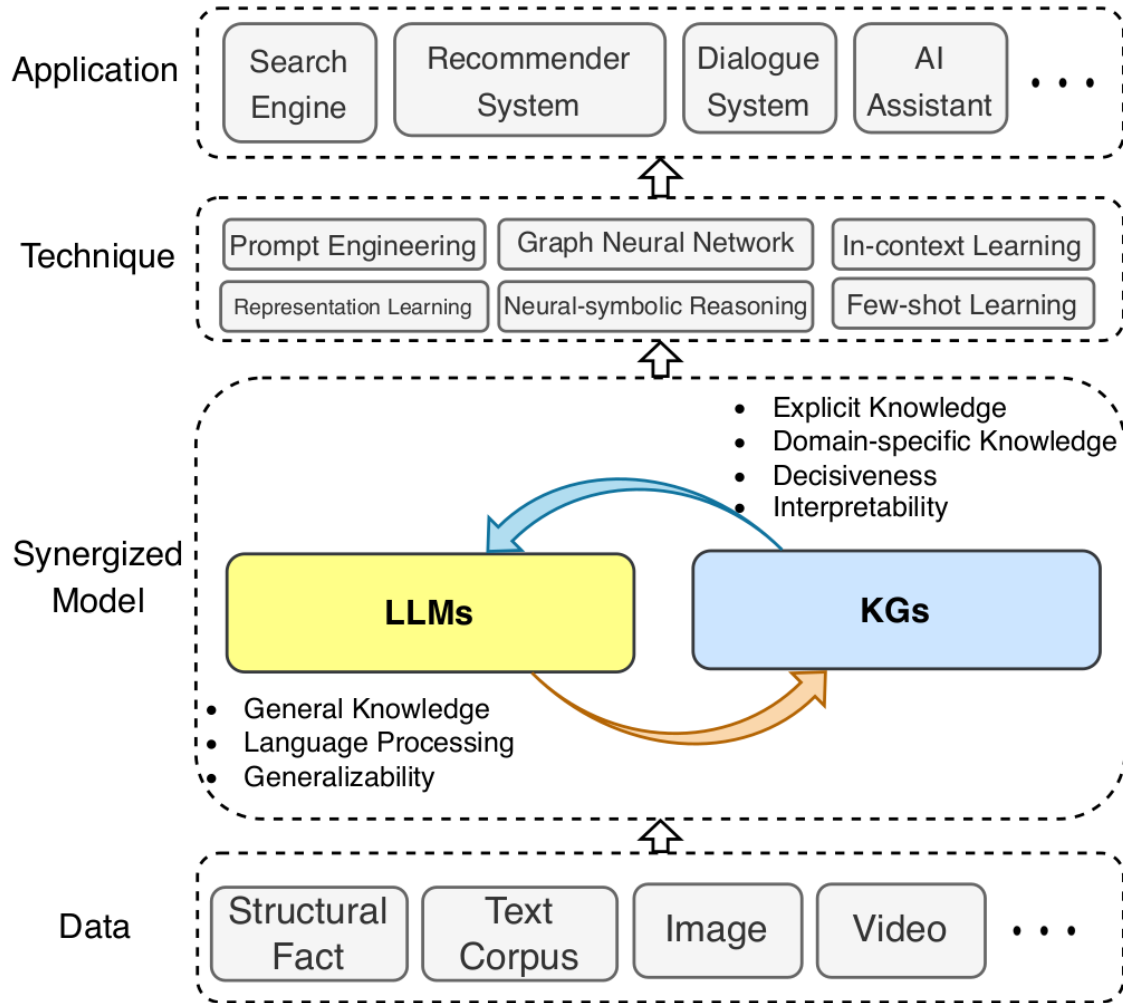
In 2024, the Eurovision Song Contest was hosted in Malmö, Sweden (7). That year, Switzerland triumphed as the winner of the contest (1), with the song "The Code" (4), which was performed by their representative, Nemo (2, 3). Not only is Nemo known for his musical talent, but he is also recognized for his advocacy for peace and dignity (10).

Ireland was also a participant in the contest (5), represented by Bambie Thug (6). Reflecting on Switzerland's past achievements, it is notable that they had previously won the contest when Celine Dion represented them in 1988 with the song "Ne Partez Pas Sans Moi" (8, 9).



KG-LLM synergy: Mutual enhancement

- Bidirectional reasoning driven by both data and knowledge



Hope to see you in
INFO320 this autumn

- and Best of Luck
on the Exam!