

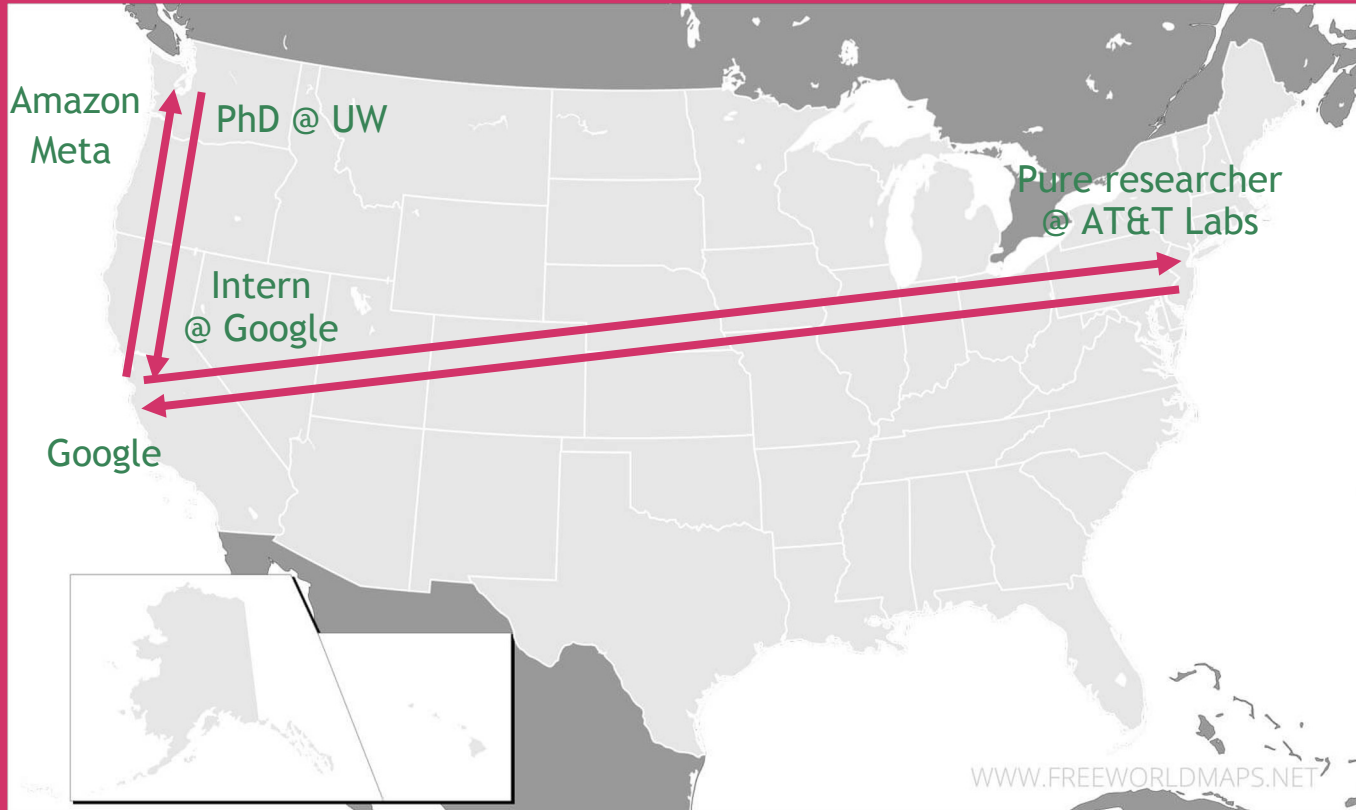
# How Knowledgeable Are LLMs?

Xin Luna Dong

8/2023

This talk does not represent the company's point of view

# About Me: My Journey



# About Me: KGs in My Journey

Amazon  
Product KGs

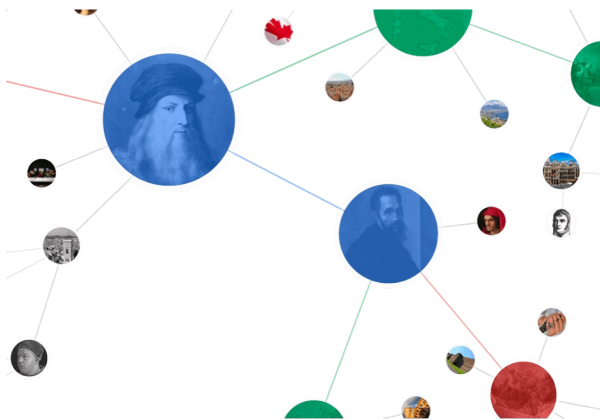
Personal KGs

Google KGs



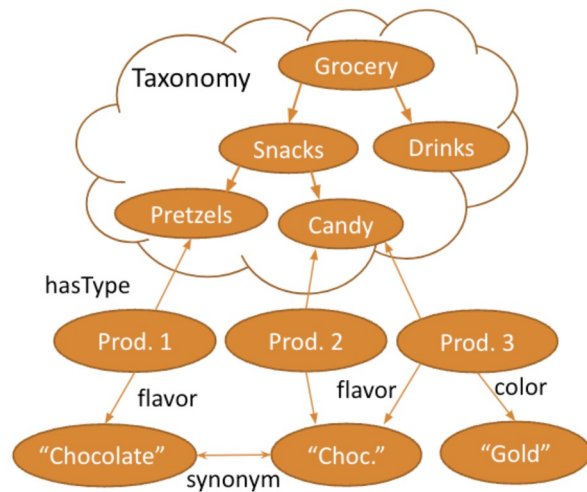
# Generations of Knowledge Graphs

## 1. Entity-Based KGs



Google Generic KGs

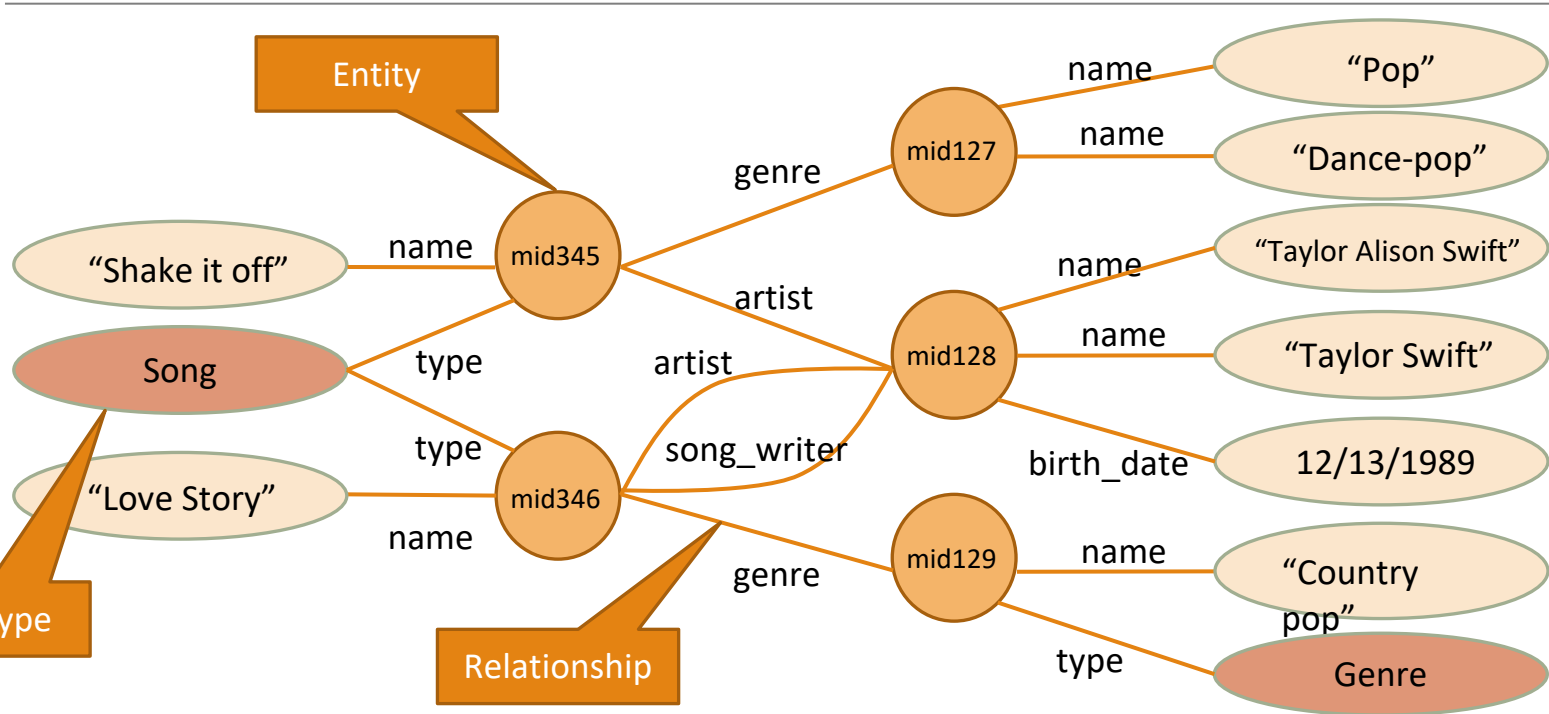
## 2. Text-Rich KGs



Amazon Product KGs

# Generation #1: Entity-Based Knowledge Graphs

# Entity-Based KG Example



# Google KG Shows in 25% Google Search in 2015

The image shows a Google search interface for 'taylor swift songs'. The search bar contains the text 'taylor swift songs' and is decorated with a pair of orange sunglasses. Below the search bar are filters for 'Videos', 'Lyrics', 'Latest', 'Popular', 'Number one', '2000s', 'Chronological', 'Images', and '1989'. The search results show 'About 236,000,000 results (0.65 seconds)'. The main result is for 'Taylor Swift', an American singer-songwriter, with tabs for 'Overview', 'Songs', 'Albums', 'Events', 'Videos', and 'Listen'. The 'Songs' section lists several tracks with their respective album covers and release years. A 'Listen' sidebar offers streaming options on YouTube, Spotify, iHeart, and Apple Music. An 'About' section provides biographical information, including her birth date (December 13, 1989), height (5' 11"), and parents (Andrea Swift, Scott Kingsley Swift). A 'Songfacts' link is also visible at the bottom left.

Google

taylor swift songs

Videos Lyrics Latest Popular Number one 2000s Chronological Images 1989 All filters Tools

About 236,000,000 results (0.65 seconds)

**Taylor Swift**  
American singer-songwriter

Overview Songs Albums Events Videos Listen

### Songs

<b>Bad Blood</b> 1989 · 2014		<b>I Knew You Were Trouble</b> Red · 2012	
<b>Cruel Summer</b> Lover · 2019		<b>Blank Space</b> 1989 · 2014	
<b>Don't Blame Me</b> reputation · 2017		<b>Shake it Off</b> 1989 · 2014	
<b>Look What You Made Me Do</b> reputation · 2017		<b>Anti-Hero</b> Midnights · 2022	
<b>You Belong With Me</b> Fearless · 2008		<b>Enchanted</b> Speak Now · 2010	

Show more

**Songfacts**  
<https://www.songfacts.com/songs/taylor-swift>

### Listen

YouTube Spotify iHeart Apple Music

### About

[taylorswift.com](https://taylorswift.com)

Taylor Alison Swift is an American singer-songwriter. Recognized for her songwriting, musical versatility, artistic reinventions, and influence on the music industry, she is a prominent cultural figure of the 21st century. [Wikipedia](#)

**Born:** December 13, 1989 (age 33 years), West Reading, PA  
**Height:** 5' 11"  
**Siblings:** Austin Swift  
**Parents:** Andrea Swift, Scott Kingsley Swift

Feedback

# Entity-Based KGs

## Characteristics

- Ontology (types, relationships) manually defined w. clear semantics
- Entities are named-entities, w. no overlap

## Key idea

Create a graph of entities and relationships to represent the world





# Challenges: Heterogeneous Data Everywhere

IMDB



**Anahí**  
Actress | Music Department | Soundtrack

Anahí was born in Mexico. She's had roles in Tu y Yo, in which she played a 17 year old girl while she was 13, and Vivo Por Elena, in which she played Talita, a naive and innocent teenager. Anahí lives with her mother and sister name Marychelo. She hopes to become a fashion designer one day, and is currently pursuing a career in singing.  
[See full bio >](#)

**Born:** May 14, 1982 in Mexico City, Distrito Federal, Mexico

More at IMDbPro >  
Contact Info: [View manager](#)

WikiData

**Anahí Puentes** (Q169461)

Mexican singer-songwriter and actress

Mia

[In more languages](#) [Configure](#)

Language	Label	Description
English	Anahí Puentes	Mexican singer-songwriter and actress
Chinese	阿纳希·普恩特	No description defined
Spanish	Anahí Puentes	Cantante, compositora y actriz mexicana

**date of birth** 7 November 1983 [edit](#)

1 reference imported from **Italian Wikipedia**

[+ add reference](#)  
[+ add value](#)

Are “Born” and “date of birth” the same? → **Schema Alignment**

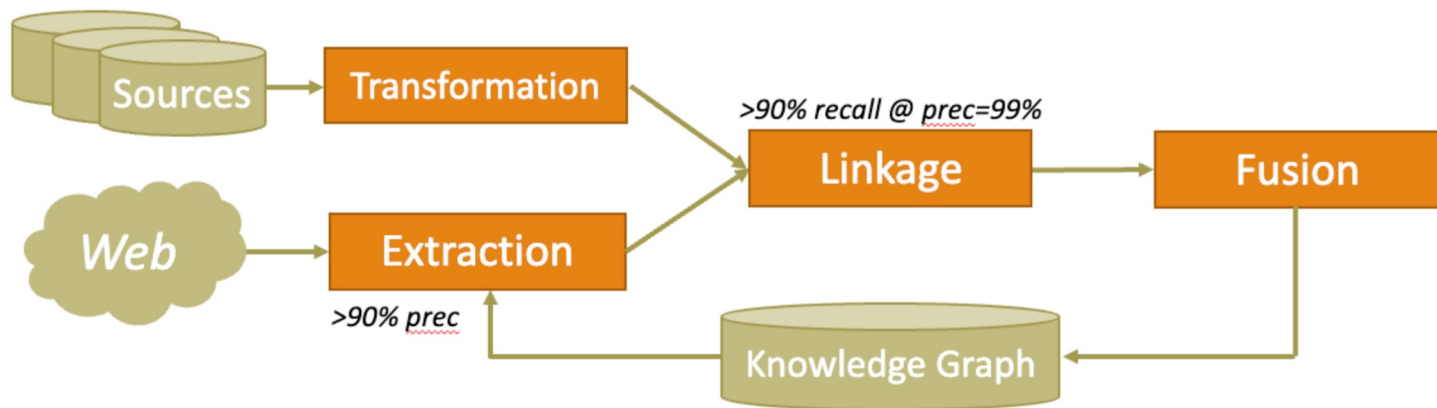
Are they the same person? → **Entity Linkage**

Why “May 14, 1982” vs “7 November 1983”? → **Data fusion**

**Heterogeneity, Heterogeneity, Heterogeneity**

# Constructing Entity-Based KGs

## Framework:

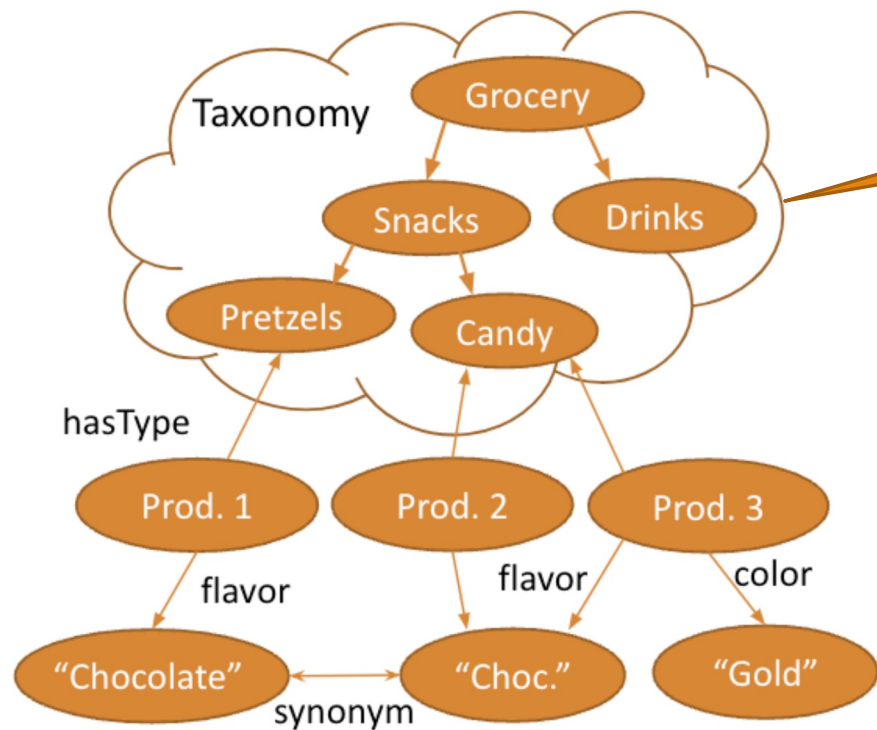


**Examples:** Google KG (500B facts), Bing Satori, Alexa KG, Yago, DBPedia.



# Generation #2: Text-Rich Knowledge Graphs

# Text-Rich KG Example



Taxonomy

**Other domains:**  
Bioinformatics, Health,  
Geography, Local business,  
Events, etc.

# Example: Providing Structured Product Information



Roll over image to zoom in

<b>Brand</b>	Cetaphil
<b>Ingredients</b>	Water, Cetyl Alcohol, Propylene Glycol, Iodopropynyl Butylcarbamate, 2-Bromo-2-Nitropropane-1, 3-Diol, Sodium Lauryl Sulfate, Stearyl Alcohol, Methylparaben, Propylparaben, Sodium Citrate, Butylparaben, Allantoin, Zinc Gluconate.
<b>Scent</b>	Fragrance free
<b>Additional Information</b>	Non-Comedogenic, Fragrance-free, Natural
<b>Skin Type</b>	Sensitive


## About this item

- Gentle for everyday use; Cetaphil gentle skin cleansing cloths will leave your skin feeling clean, refreshed and balanced after every use
- Removes makeup & dirt: Thoroughly remove makeup and dirt, leaving skin clean
- Mild & non irritating: Soap free formulation won't strip skin of its natural protective oils and emollients


# Example: Providing Choices


Visit the Bigmuscles Nutrition Store ★★★★★ 6,270  
Bigmuscles Nutrition Crude Whey 1kg, Whey Protein Concentrate 80%, 24g Protein, 5.5g BCAA, 4 g Glutamine


40% off




● ○ ○ ○ ○ ○ ▷

Style name: **Crude Whey [Unflavored]** 


Flavor: **Unflavored** 

Size: **1Kg[2.2lbs]** 


₹ 779 (₹ 82.50 / 100 g) **Lightning Deal**  
MRP: ₹ 1,299.00 **Save ₹ 520**  



Visit the Bigmuscles Nutrition Store ★★★★★ 6,270  
Bigmuscles Nutrition Crude Whey 1kg, Whey Protein Concentrate 80%, 24g Protein, 5.5g BCAA, 4 g Glutamine


40% off




● ○ ○ ○ ○ ○ ▷


Style name: **Crude Whey [Unflavored]** 

Flavor: **Unflavored** 

<b>Unflavored</b>	Caffe Latte	Isola
₹ 779.00  (₹ 77.90 / 100 g) Deal is 34% Claimed	See available options	See a optio

Size: 

# Example: Explaining Recommendations



**KitchenAid KSM150PSE Artisan Tilt-Head Stand Mixer with Pouring Shield, 5-Quart, Empire Red**  
by KitchenAid  
★★★★☆ 6,331 customer reviews | 972 answered questions


List Price: \$429.99  
Price: **\$249.99 & FREE Shipping**  
You Save: **\$180.00 (42%)**




*i* Item is eligible for **6 Month Special Financing** with your Amazon.com Store Card. [Learn more](#)

**Note:** Not eligible for Amazon Prime. Available with free Prime shipping from [other sellers on Amazon](#).

**Only 14 left in stock.**  
**Estimated Delivery Date:** July 28 - Aug. 2 when you choose Standard at checkout.  
Ships from and sold by [GE Premier](#) in [easy-to-open packaging](#).

Color: **Empire Red**




**Share**   

Qty:

**\$249.99 + Fr**  
**Only 14 left in sto**  
**Prem**


Include 2-Year i  
**\$14.99**

 **Add**


Turn on 1-Click order!

**Ship to:**  
KEVIN DAVENPOF




**Other Sellers**

**\$264.99**  
  
Sold by: Amazon.com

**\$264.99**  
+ Free Shipping  
Sold by: Marcus AV

**\$289.00**  
  
Sold by: goldentech

**Similar But Different**

More Capacity	More Attachments	Different Brand
		
<input type="button" value="See more choices"/>	<input type="button" value="See Color Options"/>	<input type="button" value="See Color Options"/>

[KitchenAid® NSF Certified®](#) [KitchenAid KP26M1XER 6](#) [Hamilton Beach 63232 Electric](#)

# Text-Rich KGs

## Characteristics

- Ontology (types, relationships) very complex with overlaps and ambiguities; E.g., millions of product types
- Entities may not be named-entities, such as products  
E.g., “Onus 2 Colors Highlighter Stick, Shimmer Cream Powder Waterproof Light Face Cosmetics, creamy Self Sharpening Crayon STick Highlighter” vs. “Xin Luna Dong”
- Attribute values are oftentimes texts, with overlaps and ambiguities  
E.g., “Coffee” vs “Cappuccino” as icecream flavors

## Key Idea

Finding structure and modeling ambiguity from text sources



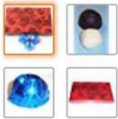


# Do We Need Different Techniques?

Scott's Cakes Dark Chocolate Toffee Cream Filling Candies with Dark Blue Foils in a 1 Pound Red Roses Box

by Scott's Cakes

Be the first to review this item



Price: **\$19.95** + \$14.95 shipping

You can get 5% back on all Amazon.com purchases with the Amazon Prime Store Card. No annual fee.

**Note:** Not eligible for Amazon Prime.

**In Stock.** Ships from and sold by Scott's Cakes.

**38 Flavors:** 1 lb. Red Rose Box



**19 Sizes:** Dark Blue Foils



**Get it Thu, Aug 22 - Tue, Aug 27**

**Get it Tue, Aug 20 - Fri, Aug 23** if you choose paid shipping at checkout.

📍 Deliver to Yaqing - Seattle 98109

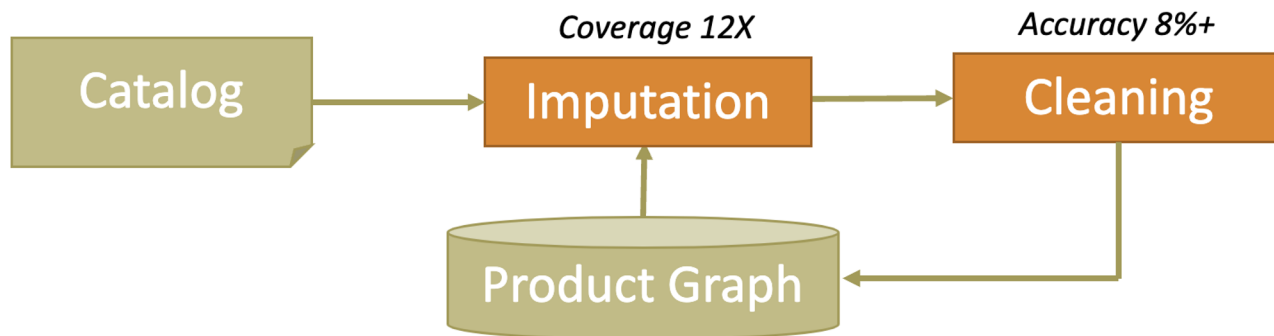
Qty: 1 ▼

Turn on 1-click ordering

Different challenges: **Unstructured** and **Noisy** product data

# Constructing Text-Rich KGs

## Framework:



**Examples:** Amazon Product Graph (1B facts)

**Generation #3: ???**

*Will LLM Replace KGs?*

# An Example



Tell me about famous people named Michael Jordan. Be brief



Sure, here are brief descriptions of three famous individuals named Michael Jordan:

1. **Michael Jordan (Basketball Player):** An American former professional basketball player, considered one of the greatest of all time. He led the Chicago Bulls to six NBA championships and won **five** MVP awards.
2. **Michael B. Jordan (Actor):** An American actor known for his roles in TV series like "The Wire," "Friday Night Lights," and films such as "Fruitvale Station," "Creed," and "Black Panther."
3. **Michael I. Jordan (Researcher):** An American scientist and professor specializing in machine learning, statistics, and artificial intelligence, currently teaching at the University of California, Berkeley.

# Another Example



Tell me famous people named Xin Luna Dong. Be brief



As of  
scie  
the c  
play  
How  
Luna



Tell me famous people named Luna Dong. Be brief



Tell me famous people named Xin Dong. Be brief



As of my last training data up to September 2021, there are no notable or famous individuals named Xin Dong that stand out. Please note that this is a common name in Chinese-speaking regions, so there may be many individuals with this name who aren't widely recognized internationally. There might also be notable people with this name who have gained prominence after my last training cut-off in 2021.

# Another Example



Tell me famous books for knowledge graphs




Sure, here are some popular books on the subject of knowledge graphs:

1. "Knowledge Graph and Semantic Computing: Semantic, Knowledge, and Linked Big Data" by various authors. This book contains selected papers from leading researchers on the topic.
2. "Knowledge Graphs: Methodology, Tools and Selected Use Cases" by Sherif Sakr and Marcin Wylot. This book provides an accessible, yet comprehensive introduction to the field of knowledge graphs.
3. "Knowledge Graphs: Fundamentals, Techniques, and Applications" by Raghava Mutharaju, Sarasi Lalithsena, Pavan Kapanipathi. A comprehensive look into various aspects of knowledge graphs, including the technical aspects and practical applications.

4. "Linked Data: Structured Data on the Web" by David Wood, Marsha Zaidman, Luke Ruth, Michael Hausenblas. This book offers a practical introduction to linked data and how it's used in semantic web technologies, which are closely related to knowledge graphs.
5. "Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL" by Dean Allemang, Jim Hendler. This book isn't specifically about knowledge graphs, but it provides foundational knowledge about semantic web technologies that are critical to understanding and working with knowledge graphs.
6. "Foundations of Data Science" by John Hopcroft and Ravindran Kannan. This one covers a broad range of data science topics, including knowledge graphs.
7. "Graph Databases" by Ian Robinson, Jim Webber, Emil Eifrem. This book is all about graph databases, which provide the underlying technology for many knowledge graphs.

# Key Questions to Ask Regarding Hallucination

- How ***reliable*** are LLMs when answering factual questions?
  - Do LLMs perform ***equally well*** across different types of factual knowledge?
  - Do well-known LLM tricks work?
    - Increasing ***model size***
    - ***Instruction tuning***
    - ***Few-shot*** in-context learning
- 

## Head-to-Tail Benchmark

<b>Domain</b>	<b>Sources</b>	<b># Templates</b>	<b># Instances</b>
Movies	IMDb	13	3,093
Books	Goodreads	4	3,000
Sports	SR/Olympic Sports	7	3,150
Academics	MAG, DBLP	13	2,946
<hr style="border-top: 1px dashed black;"/>			
Open	DBpedia	393	9,132
<b>Total</b>		<b>430</b>	<b>21,321</b>

Table 2: The overall statistics of Head-to-Tail.



## Head-to-Tail Benchmark

Domain	Sources	# Templates	# Instances
Movies	IMDb	13	3,093
Books	Goodreads	4	3,000
Sports	SR/Olympic Sports	7	3,150
Academics	MAG, DBLP	13	2,946
Open	DBpedia	393	9,132
<b>Total</b>		<b>430</b>	<b>21,321</b>

Table 2: The overall statistics of Head-to-Tail.

# Questions

---

Answer the following questions in as few words as possible. Say "unsure" if you don't know.

Question: What is the capital of China?

Answer: Beijing

Question: What is the captical of Wernythedia?

Answer: unsure

Question: {QUESTION}

Answer:

---

# Correctness Checking

## 3 categories

- Correct
- Incorrect
- Unsure

## 3 metrics

- Accuracy =  $\text{corr} / \text{all}$
- Hallucination rate =  $\text{Incorr} / \text{all}$
- Missing rate =  $\text{Unsure} / \text{all}$

---

You need to check whether the prediction of a question-answering system to a question is correct. You should make the judgment based on a list of ground truth answers provided to you. Your response should be "correct" if the prediction is correct or "incorrect" if the prediction is wrong.

Question: Who authored The Taming of the Shrew (published in 2002)?

Ground truth: ["William Shakespeare", "Roma Gill"]

Prediction: W Shakespeare

Correctness: correct

Question: Who authored The Taming of the Shrew (published in 2002)?

Ground truth: ["William Shakespeare", "Roma Gill"]

Prediction: Roma Shakespeare

Correctness: incorrect

Question: {QUESTION}

Ground truth: {GROUND\_TRUTH}

Prediction: {PREDICTION}

Correctness:

---

98% consistent w.  
human labels

# Q1. How Reliable Are LLMs on Factual Questions?

Domain	A <sub>LM</sub>	H <sub>LM</sub>	Miss.
Specific	19.2	11.8	69.0
Open	22.1	14.8	63.2
All	20.4	13.1	66.5

Table 3: Performance of ChatGPT on Head-to-Tail. Numbers are in percentage (%).

Similar for *specific* domains and *general* domains

Very low accuracy. High missing rate; hallucination rate reasonable

# Entity Distributions

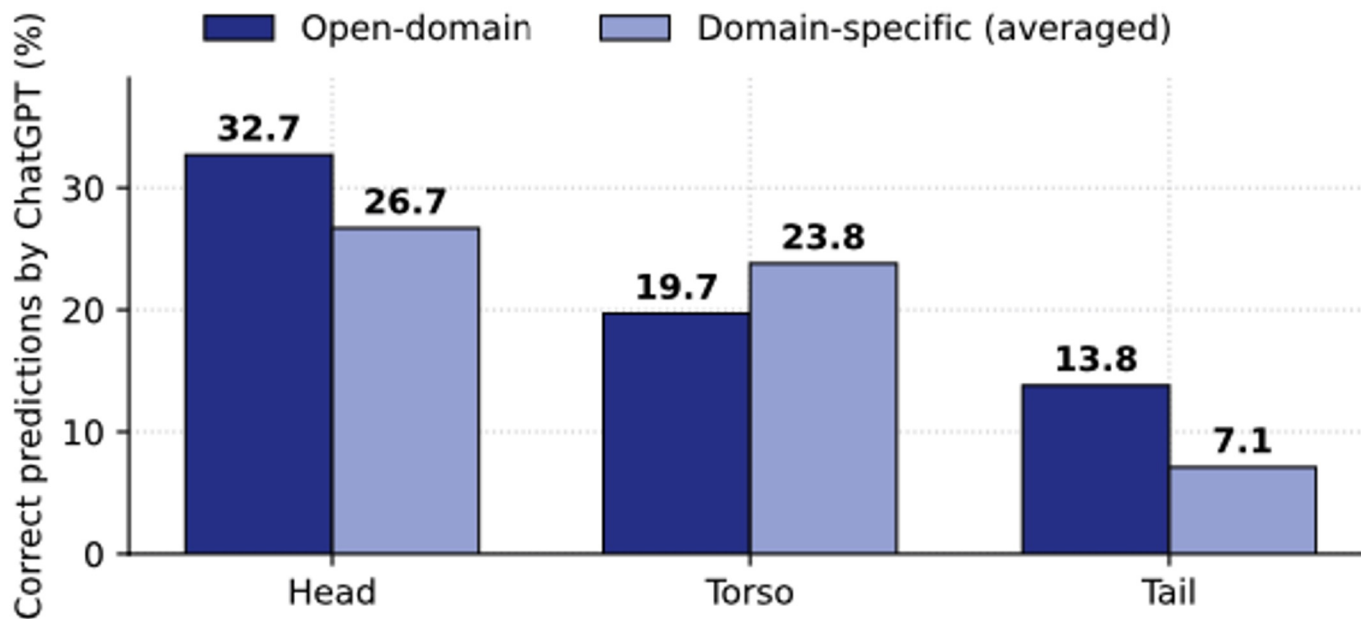
	IMDb		Goodreads	SR/Olympic Sports
	Title	Person	Book	Athlete
Head	767 ( 0.01)	34,903 ( 0.48)	847 ( 0.05)	3,150 ( 2.31)
Torso	4,113 ( 0.05)	87,645 ( 1.21)	13,118 ( 0.75)	7,304 ( 5.35)
Tail	7,536,482 (99.94)	7,111,496 (98.31)	1,723,765 (99.20)	126,134 (92.35)

	MAG			DBLP	DBpedia
	Article	Conference	Journal	Scholar	-
1,827,710 ( 0.70)	257 ( 1.63)	225 ( 0.46)	79,521 ( 2.44)	103,564 ( 1.30)	
9,386,034 ( 3.60)	965 ( 6.12)	1,266 ( 2.58)	500,778 (15.36)	1,255,113 (15.77)	
249,311,539 (95.70)	14,550 (92.25)	47,546 (96.96)	2,680,704 (82.20)	6,600,206 (82.93)	

When counted by popularity, majority of entities are long-tail

## Q2. Performance Similar for Head/Torso/Tail?



## Q2. Performance Similar for Head/Torso/Tail?

Domain	Head		Torso		Tail	
	$A_{LM}$	$H_{LM}$	$A_{LM}$	$H_{LM}$	$A_{LM}$	$H_{LM}$
Movies	51.3	11.5	46.4	16.6	6.4	11.8
Books	20.1	26.3	22.5	29.2	8.0	19.2
Sports	28.1	9.0	22.9	7.0	12.7	5.6
Academics	5.9	3.0	2.3	1.7	0.8	0.9
Open	32.7	20.8	19.7	13.3	13.8	10.2
All	29.2	16.0	22.0	13.5	9.9	9.7

Tail domains have lower accuracy

1. Accuracy consistently dropping from head to torso to tail
2. Hallucination rate might be higher for torso

## Q2. Performance Similar for Head/Torso/Tail?

Model	Head & Torso		Tail	
	A <sub>LM</sub>	H <sub>LM</sub>	A <sub>LM</sub>	H <sub>LM</sub>
ChatGPT	18.6	14.2	22.3	14.8

Table 5: Comparison of LLMs' knowledgeability about head, torso, and tail attributes in A<sub>LM</sub> (%) and H<sub>LM</sub> (%) using open-domain instances from Head-to-Tail.

Interestingly, not significant differences for head / torso / tail **attributes**



## Q3. Do Normal LLM Tricks Help?

Increasing model size does not appear to help significantly

Instruction-tuned counterparts have higher missing rate

Model	A <sub>LM</sub>	H <sub>LM</sub>	Miss.
LLaMA (7B)	14.9	78.3	6.7
LLaMA (13B)	16.7	82.1	1.1
LLaMA (33B)	20.5	78.0	1.5
LLaMA (65B)	20.0	79.7	0.3
Vicuna (7B)	11.9	78.6	9.5
Vicuna (13B)	11.7	63.3	25.0
Flan-T5 (3B)	3.6	16.8	79.7
Flan-T5 (11B)	6.8	18.4	74.8
Falcon (7B)	11.4	59.2	29.4
Falcon (40B)	12.6	43.4	44.0
Falcon-Instruct (7B)	8.1	57.9	34.0
Falcon-Instruct (40B)	10.4	36.0	53.7

## Q4. Would Few-Shot In-Context Learning Help?

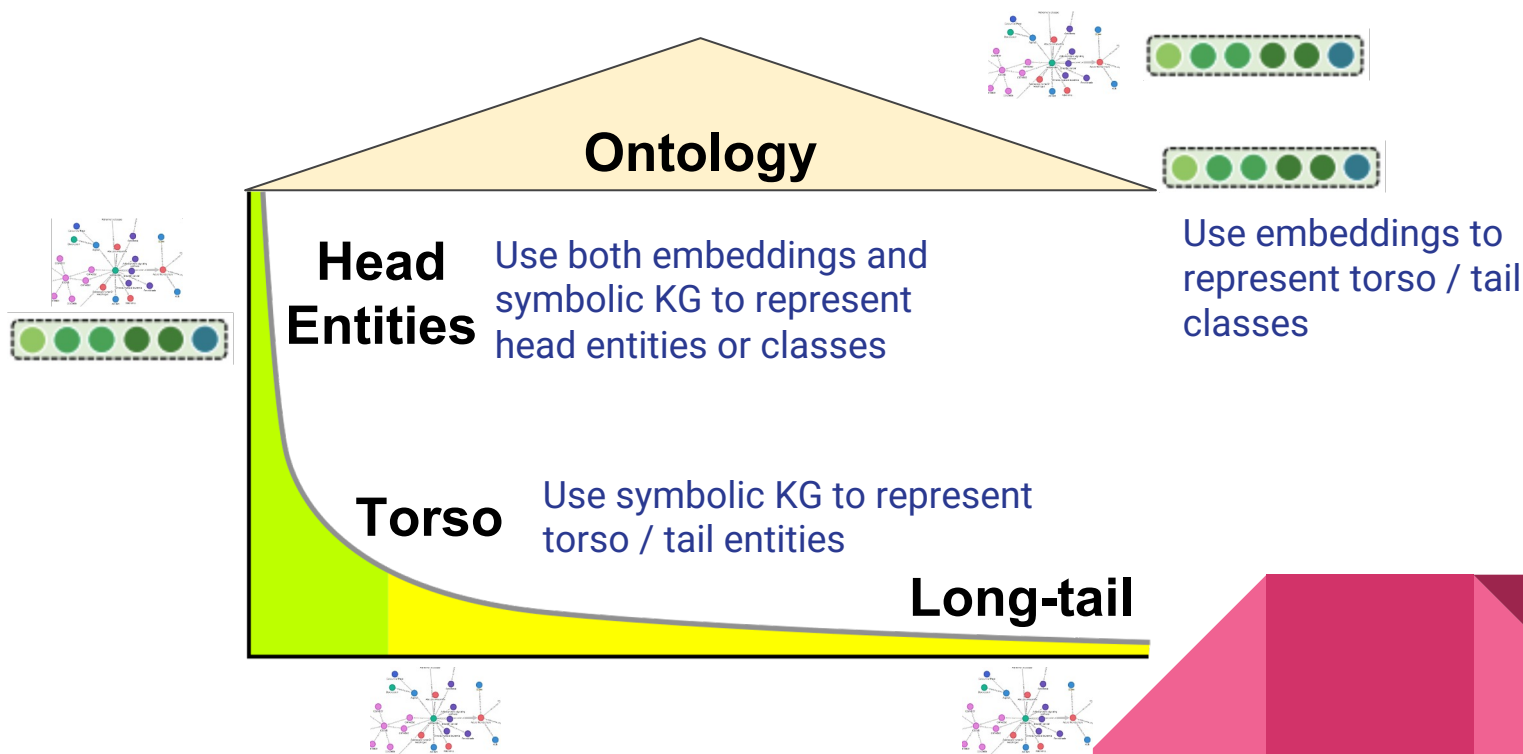
Zero-shot and Few-shot have similar performances

In-domain few-shot (positive examples) increases quality but also hallucination rate

Domain	Few-shot		Zero-shot		In-domain Few-shot	
	A <sub>LM</sub>	H <sub>LM</sub>	A <sub>LM</sub>	H <sub>LM</sub>	A <sub>LM</sub>	H <sub>LM</sub>
Movies	51.3	11.5	51.1	12.5	<b>54.7</b>	<b>7.1</b>
Books	<b>20.1</b>	26.3	16.5	<b>17.8</b>	18.4	42.4
Sports	28.1	<b>9.0</b>	22.6	10.1	<b>42.9</b>	21.4
Academics	5.9	<b>3.0</b>	8.4	7.2	<b>20.4</b>	15.9
Open	32.7	<b>20.8</b>	32.6	24.7	<b>45.0</b>	27.8
All	29.2	<b>16.0</b>	28.2	17.4	<b>39.0</b>	24.3

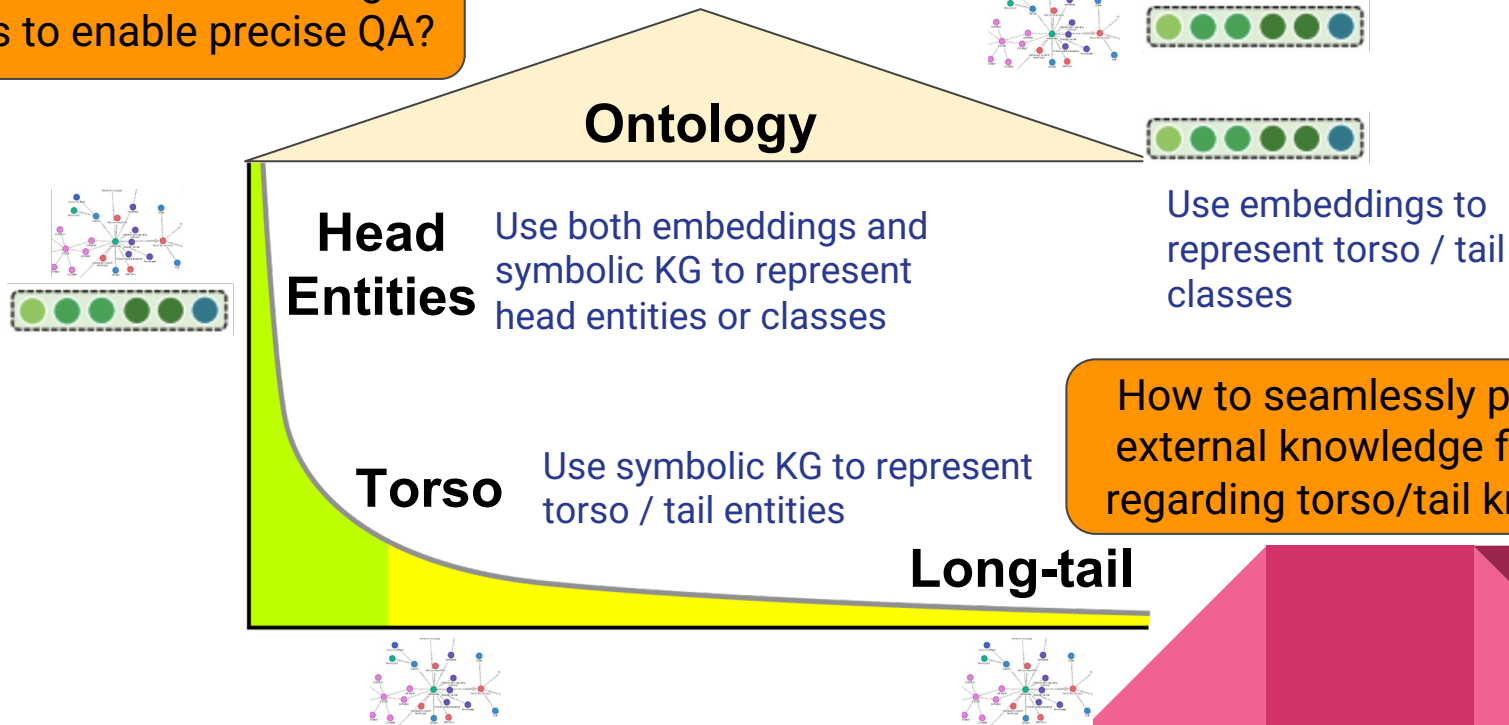
Table 7: Performance of ChatGPT with different prompts on Head-to-Tail. All numbers are in percentage (%).

# Next Generation of KG—Dual Neural KG



# Next Generation of KG—Dual Neural KG

How to infuse head knowledge into LLMs to enable precise QA?



How to seamlessly plug-in external knowledge for QA regarding torso/tail knowl.?

# Meta's Assistant

Empowering connection to people and experiences in your life

Meta Quest 2



“Hey Facebook” (double press the button on your controller)

“Who’s online?”--meet up with friends

“Open Beat Saber”--jump straight in the game, and more.

Ray-Ban Stories



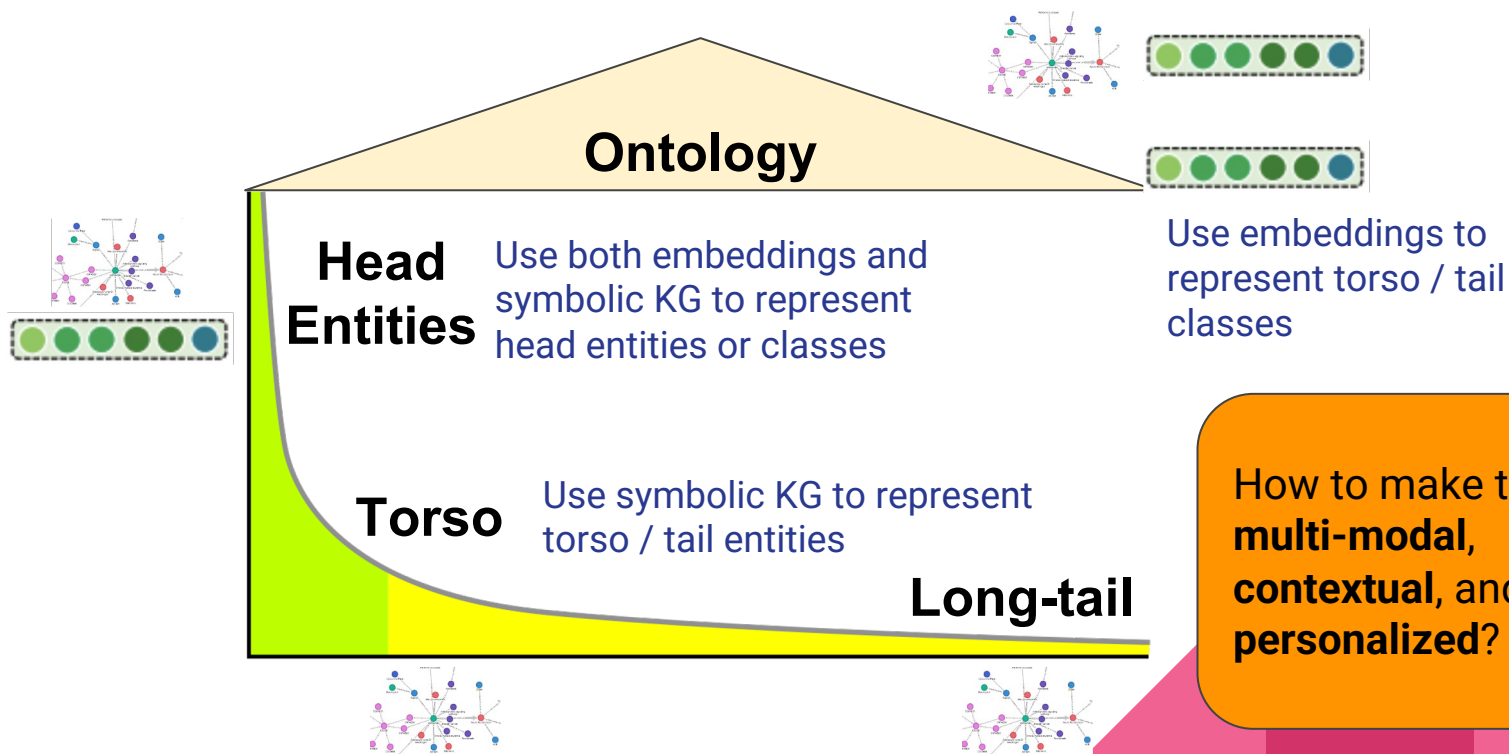
“Hey Facebook, take a picture” -- capture moments hands-free

“Hey Facebook”--call friends on Messenger, manage device settings, and more.

# Building an AR/VR Smart Assistant



# Next Generation of KG—Dual Neural KG



# Shameless Advertisement

## **Next-Generation Intelligent Assistants for AR/VR Devices**

Workshop "Multimodal Learning"  
8/7 (Monday) 1:10pm, Room: 102A

## **Tutorial: Towards Next-Generation Intelligent Assistants leveraging LLM techniques**

8/9 (Wednesday) 10am-1pm, Room: 202C

How to make this  
**multi-modal,**  
**contextual,** and  
**personalized?**



# Take-Aways. 3 Generations of KGs

## 1. Entity-Based KGs

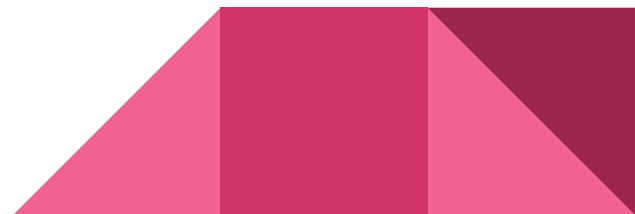
Resolving heterogeneity with **entity linkage** and **web knowledge extraction**

## 2. Text-Rich KGs

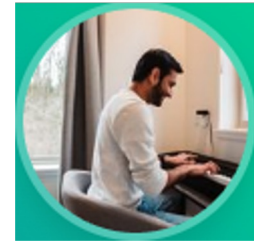
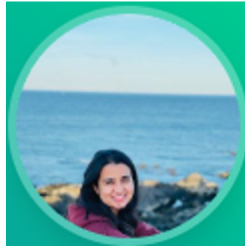
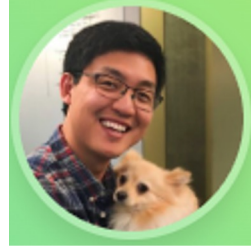
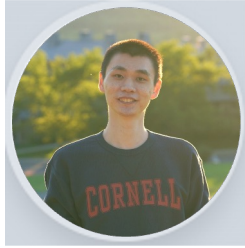
**Extractions** and **cleanings** from sparse and noisy source data, and handling semantics ambiguities

## 3. Dual Neural KGs

**Knowledge infusion** into LLMs, and seamless **knowledge augmentation** to facilitate LLMs



# Acknowledgement





# Thank You

Q&A?