

# VerifAI: Verified Generative AI (Vision)

**Nan Tang**

*HKUST (GZ)*

*China*

**Chenyu Yang**

*HKUST (GZ)*

*China*

**Ju Fan**

*Renmin Univ.*

*China*

**Lei Cao**

*Univ. of Arizona/MIT*

*USA*

**Yuyu Luo**

*HKUST (GZ)*

*China*

**Alon Halevy**

*Amazon*

*USA*

# Outline

- **What**
- **How**
- **Results**
- **Opportunities**

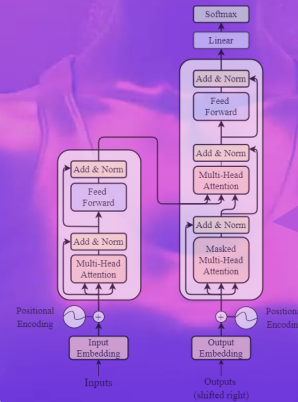


image  
table  
text  
graph

2017

2018: GPT-1 and GPT-2  
2020: GPT-3 (Wow)  
2022: ChatGPT (Amazing)  
2023: GPT-4 (Unbelievable)

.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....



**Attention Is All You Need**

---

Ashish Vaswani\*  
Google Brain  
avaswani@google.com

Noam Shazeer\*  
Google Brain  
noam@google.com

Niki Parmar\*  
Google Research  
niki@google.com

Jakob Uszkoreit\*  
Google Research  
usz@google.com

Llion Jones\*  
Google Research  
llion@google.com

Aidan N. Gomez\*<sup>†</sup>  
University of Toronto  
aidan@cs.toronto.edu

Lukasz Kaiser\*  
Google Brain  
lukaszkaizer@google.com

Illia Polosukhin\*<sup>‡</sup>  
illia.polosukhin@gmail.com

**Abstract**

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

---

\*Equal contribution. Listing order is random. Jakob proposed replacing RNNs with self-attention and started the effort to evaluate this idea. Ashish, with Illia, designed and implemented the first Transformer models and has been crucially involved in every aspect of this work. Noam proposed scaled dot-product attention, multi-head attention and the parameter-free position representation and became the other person involved in nearly every detail. Niki designed, implemented, tuned and evaluated countless model variants in our original codebase and tensor2tensor. Llion also experimented with novel model variants, was responsible for our initial codebase, and efficient inference and visualizations. Lukasz and Aidan spent countless long days designing various parts of and implementing tensor2tensor, replacing our earlier codebase, greatly improving results and massively accelerating our research.

<sup>†</sup>Work performed while at Google Brain.  
<sup>‡</sup>Work performed while at Google Research.

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

100,000+ citations

# Large Language Models

(a) ChatGPT for

*tuple generation/completion/augmentation*



You

Please help me fill in the missing values denoted by 'NULL' in this table and provide me with the completed one:

[title]: United States House of Representative Elections **2004 in Ohio**

| district | incumbent | candidates |

| Ohio's 1st congressional district | NULL | NULL |

| Ohio's 2nd congressional district | NULL | NULL |

| Ohio's 3rd congressional district | NULL | NULL |



district	incumbent	candidates
Ohio's 1st congressional district		
Ohio's 2nd congressional district		
Ohio's 3rd congressional district		

Please fill these values

# Do you trust the results?

## Hallucination

- Inaccurate decision-making
- Spread misinformation
- Privacy violation
- Legal liabilities
- Damage to trust
- ... ..

## Downstream tasks?

## Errors can be propagated ...



- **LLMs: Generative AI**
  - Learn a distribution from finite data
  - Generalize to infinite data points
- **Goals of Generative AI**
  - Generalizable and natural
  - Not factual
- **LLMs vs. data compression models?**
  - Is not
  - Can not

**Question**  
Can we  
remove  
hallucination?

**NO**





# Outline

- **What: Verify the outputs of LLMs (True/False)**
- **How**
- **Results**
- **Opportunities**



## Generative AI



### (a) ChatGPT for *tuple generation/completion/augmentation*

晨宇 You

Please help me fill in the missing values denoted by 'NULL' in this table and provide me with the completed one:

[title]: United States House of Representative Elections 2004 in Ohio

| district | incumbent | candidates |

| Ohio's 1st congressional district | NULL | NULL |

| Ohio's 2nd congressional district | NULL | NULL |

| Ohio's 3rd congressional district | NULL | NULL |



district	incumbent	candidates
Ohio's 1st congressional district	Steve Chabot (R)	Steve Chabot (R), John Cranley (D), James Berns (Libertarian)
Ohio's 2nd congressional district	Rob Portman (R)	Rob Portman (R), Charles W. Sanders (D), Edward S. Focke (Green)
Ohio's 3rd congressional district	Tony P. Hall (D)	Tony P. Hall (D), Richard E. Stobbs (R), John Fockler (Libertarian)



### (b) ChatGPT for *text generation*

MC

Does Meagan Good play a role in Stomp the Yard?



No, Meagan Good does not appear in the 2007 film "Stomp the Yard." "Stomp the Yard" is a dance drama film directed by Sylvain White, and Meagan Good is not listed as part of the main cast or credited with a role in the film. The main cast of "Stomp the Yard" includes actors such as Columbus Short, Chris Brown, Ne-Yo, and Brian White, among others. Meagan Good is an American actress known for her roles in films such as "Think Like a Man," "Eve's Bayou," and "Waist Deep," but she did not appear in "Stomp the Yard."





### (a) ChatGPT for tuple generation/completion/augmentation



Generative AI

晨宇 You

Please help me fill in the missing values denoted by 'NULL' in this table and provide me with the completed one:

[title]: United States House of Representative Elections **2004 in Ohio**  
district	incumbent	candidates
Ohio's 1st congressional district	NULL	NULL
Ohio's 2nd congressional district	NULL	NULL
Ohio's 3rd congressional district	NULL	NULL

district	incumbent	candidates
Ohio's 1st congressional district	Steve Chabot (R)	Steve Chabot (R), John Cranley (D), James Berns (Libertarian)
Ohio's 2nd congressional district	Rob Portman (R)	Rob Portman (R), Charles W. Sanders (D), Edward S. Focke (Green)
Ohio's 3rd congressional district	Tony P. Hall (D)	Tony P. Hall (D), Richard E. Stobbs (R), John Fockler (Libertarian)



#### Steve Chabot

Article Talk

From Wikipedia, the free encyclopedia

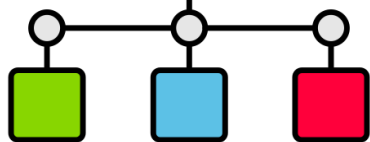
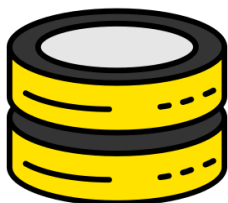
**Steven Joseph Chabot** (/ˈʃæbət/ *SHAB-ət*; born January 22, 1953) is an American politician and lawyer who represented **Ohio's 1st congressional district** in the **United States House of Representatives** from 1995 to 2009 and again from 2011 to 2023. A member of the **Republican Party**, he lost his 2022 reelection bid to Democrat **Greg Landsman**. Until his election loss, he was the dean of **Ohio's GOP delegation to the House of Representatives**, after the retirement of former Speaker **John Boehner**.<sup>[1]</sup>

#### General election results [ edit ]

##### Ohio's 3rd Congressional District election, 2004<sup>[1]</sup>

Party	Candidate	Votes	%
<span>Republican</span>	<b>Mike Turner (Incumbent)</b>	197,290	62.29
<span>Democratic</span>	Jane Mitakides	119,448	37.71
<b>Total votes</b>		<b>316,738</b>	<b>100</b>
<span>Republican hold</span>			

Data Lake



### (b) ChatGPT for text generation

MC

Does Meagan Good play a role in Stomp the Yard?



No, Meagan Good does not appear in the 2007 film "Stomp the Yard." "Stomp the Yard" is a dance drama film directed by Sylvain White, and **Meagan Good is not listed as part of the main cast or credited with a role in the film.** The main cast of "Stomp the Yard" includes actors such as Columbus Short, Chris Brown, Ne-Yo, and Brian White, among others. Meagan Good is an American actress known for her roles in films such as "Think Like a Man," "Eve's Bayou," and "Waist Deep," but she did not appear in "Stomp the Yard."

#### Stomp the Yard

Article Talk

From Wikipedia, the free encyclopedia

The film stars **Columbus Short**, **Meagan Good**, **Darrin Henson**, **Brian White**, **Laz Alonso**, and **Valarie Pettiford**, with **Harry Lennix**, and, in their film debuts, R&B singers **Ne-Yo** & **Chris Brown**. *Stomp the Yard* was filmed in **Atlanta, Georgia**, on the campuses of **Morris Brown College**, **Georgia Institute of Technology**, **Morehouse College**, and **Clark Atlanta University**, and in the **MAK Historic District** of **Decatur, Georgia**. Elsewhere Short, Alonso and Brown had also starred together in the film *This Christmas*.

#### Meagan Good's Filmography

##### Film [ edit ]

Year	Title	Role	Notes
2005	<i>Brick</i>	Kara	
	<i>Venom</i>	Cece	
	<i>Roll Bounce</i>	Naomi Phillips	
2006	<i>Miles from Home</i>	Natasha Freeman	
	<i>Waist Deep</i>	Coco	
2007	<b><i>Stomp the Yard</i></b>	April Palmer	

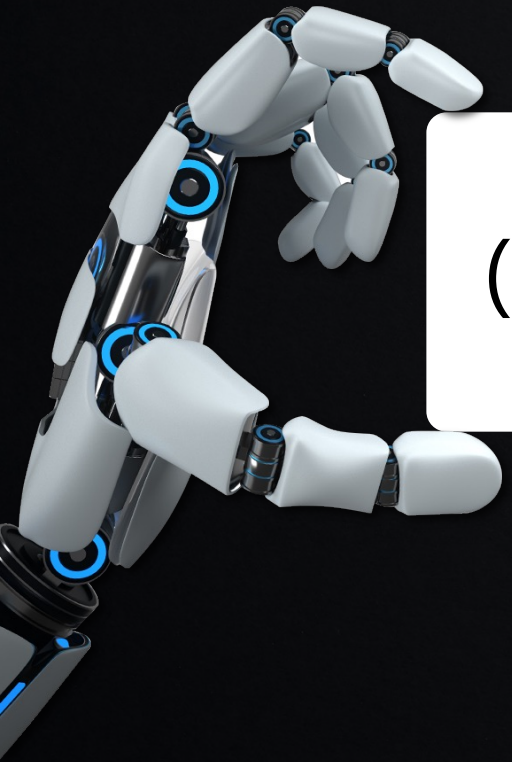




# Outline

- **What: Verify the outputs of LLMs (True/False)**
- **How**
- **Results**
- **Opportunities**





(Q, A)

Retrieve **relevant** dataset(s)  
to verify **factual (not  
subjective)** answers



- textual files
- table
- (sub)graphs
- images
- HTML files
- a combination



# Outline

- **What: Verify the outputs of LLMs (True/False)**
- **How: Retrieval**
- **Results**
- **Opportunities**



# Retrieval

- Retrieval-augmented generation (RAG)
- VerifAI: Retrieval-based matching
- Question-Answer Similarity
- Answer-Answer Similarity (matching)

## A Deep Look into Neural Ranking Models for Information Retrieval

Jiafeng Guo<sup>a,b</sup>, Yixing Fan<sup>a,b</sup>, Liang Pang<sup>a,b</sup>, Liu Yang<sup>c</sup>, Qingyao Ai<sup>c</sup>, Hamed Zamani<sup>c</sup>, Chen Wu<sup>a,b</sup>, W. Bruce Croft<sup>c</sup>, Xueqi Cheng<sup>a,b</sup>

<sup>a</sup>University of Chinese Academy of Sciences, Beijing, China

<sup>b</sup>CAS Key Lab of Network Data Science and Technology, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

<sup>c</sup>Center for Intelligent Information Retrieval, University of Massachusetts Amherst, Amherst, MA, USA

### Abstract

Ranking models lie at the heart of research on information retrieval (IR). During the past decades, different techniques have been proposed for constructing ranking models, from traditional heuristic methods, probabilistic methods, to modern machine learning methods. Recently, with the advance of deep learning technology, we have witnessed a growing body of work in applying shallow or deep neural networks to the ranking problem in IR, referred to as neural ranking models in this paper. The power of neural ranking models lies in the ability to learn from the raw text inputs for the ranking problem to avoid many limitations of hand-crafted features. Neural networks have sufficient capacity to model complicated tasks, which is needed to handle the complexity of relevance estimation in ranking. Since there have been a large variety of neural ranking models proposed, we believe it is the right time to summarize the current status, learn from existing methodologies, and gain some insights for future development. In contrast to existing reviews, in this survey, we will take a deep look into the neural ranking models from different dimensions to analyze their underlying assumptions, major design principles, and learning strategies. We compare these models through benchmark tasks to obtain a comprehensive empirical understanding of the existing techniques. We will also discuss what is missing in the current literature and what are the promising and desired future directions.



... like "Mama's Song," Carrie Underwood won the fourth season of the American singing competition television series "American Idol" in 2005 ...

You are wrong!!!

**Question:**

Which contest did the singer of the song Mama's Song win?

**Ground Truth: American Idol**

- Passage 1: The Eurovision Song Contest 2001 was ...
- Passage 2: The Eurovision Song Contest 1996 was ...
- Passage 3: The Junior Eurovision Song Contest 2010 was ...

VerifAI



1. LLM

The singer of the song "Mama's Song" did not win any specific contest related to the song. The song was released as a single by country music artist Carrie Underwood and was well-received by fans and critics alike.



2. RAG

The singer of the song "Mama's Song" won the Junior Eurovision Song Contest in 2010. This is evident from the passage: "The contest was won by Vladimir Arzumanyan from Armenia"



3. VerifAI

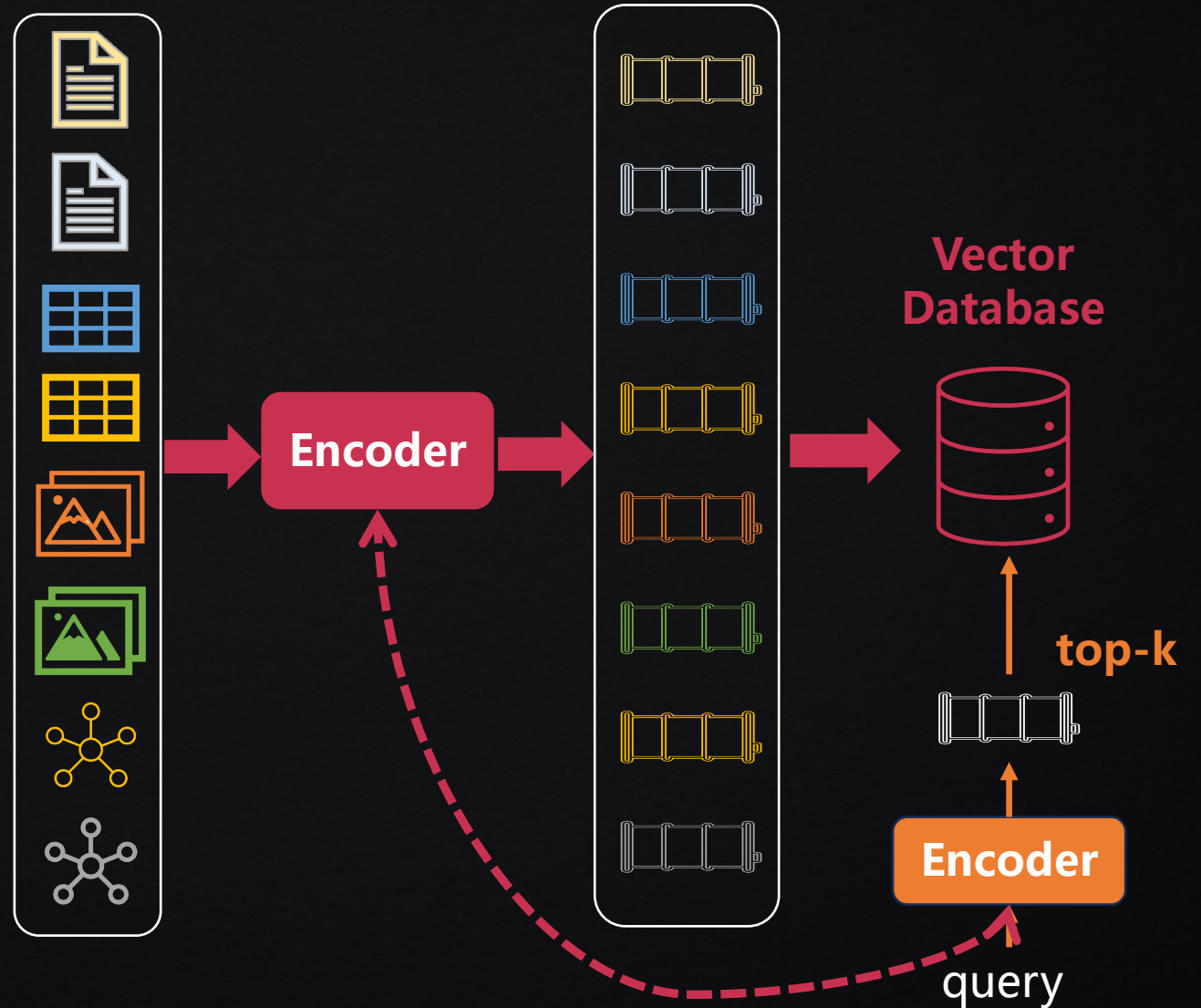
American Idol





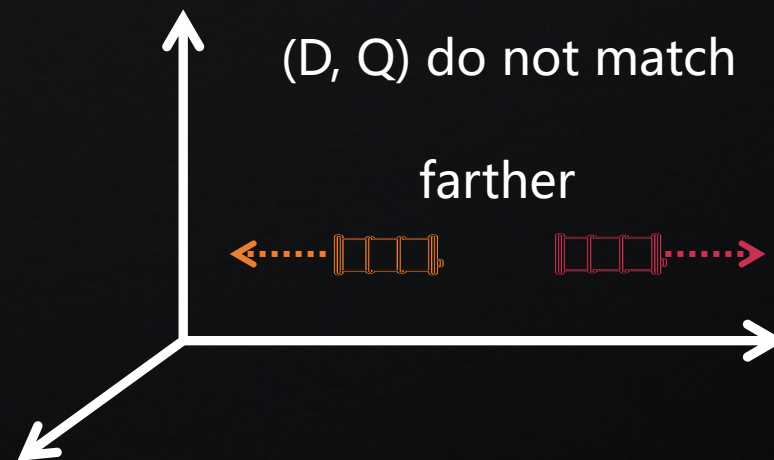
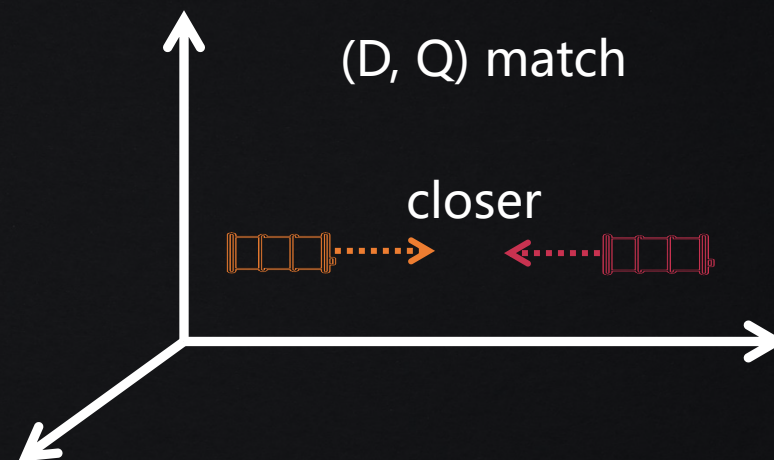
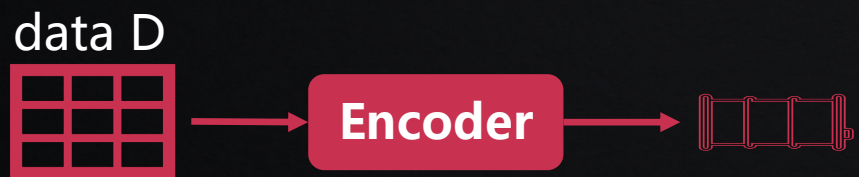
# Retrieval

- Given a query, find
- textual files
- table (tuples)
- (sub)graphs
- images
- HTML files
- or a combination thereof



How to align?  
“match” -> similar vectors

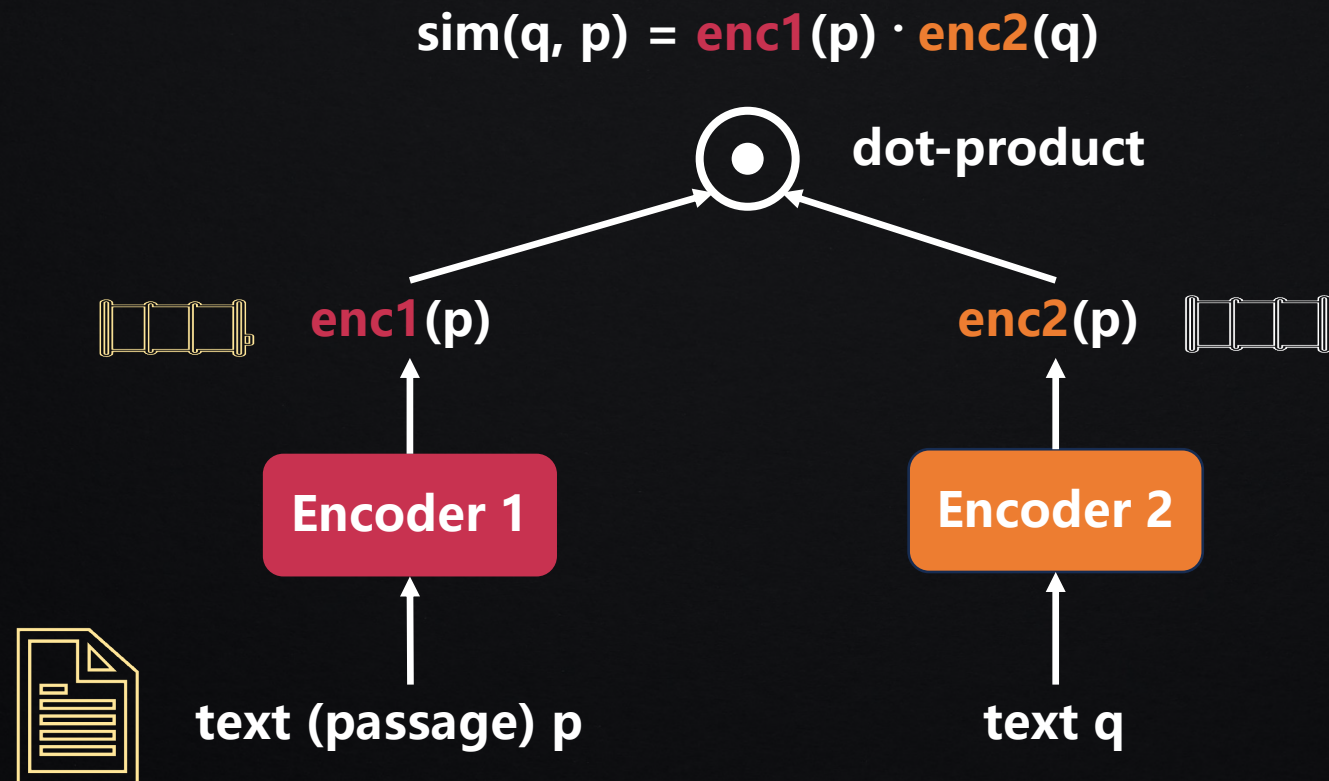






# Retrieval

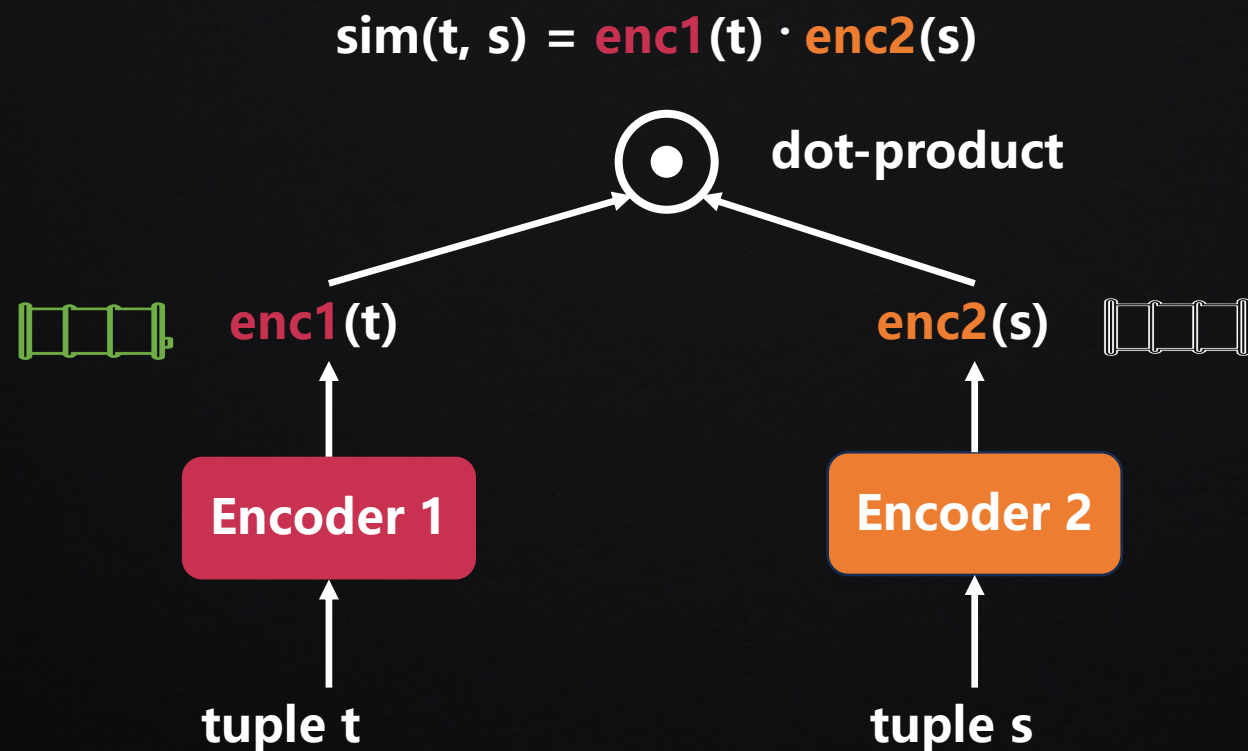
## Dense Passage Retriever (DPR) for Text-Text





# Retrieval

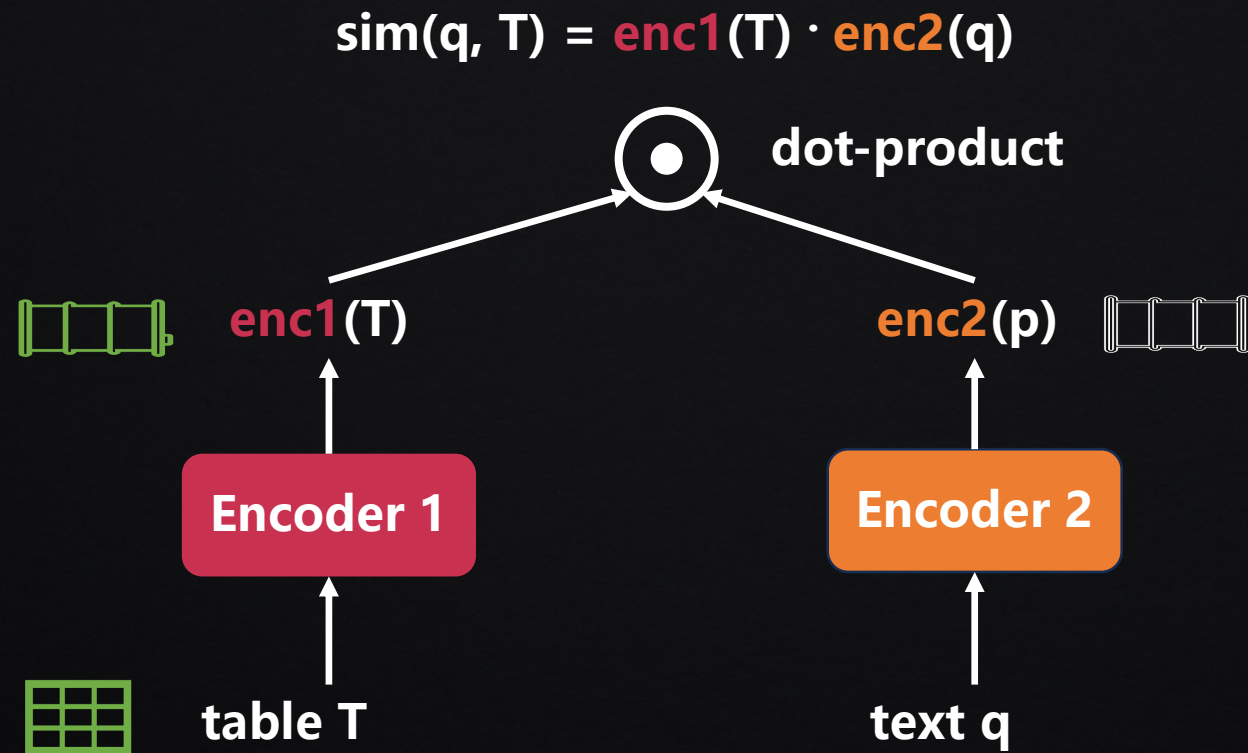
## Dense Passage Retriever (DPR) for Tuple-Tuple





# Retrieval

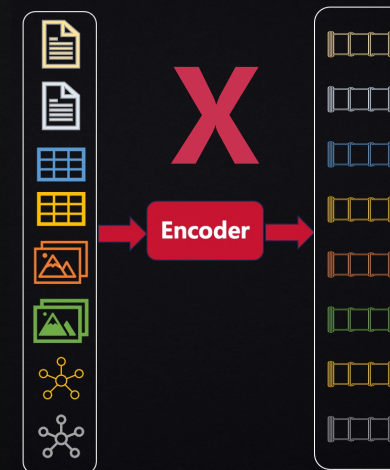
## Dense Passage Retriever (DPR) for Text-Table





# Retrieval over multi-modal data lakes

- **We know (a little bit about) how to**
  - align two encoders of the same/different data modalities
  - the main effort is to find positive/negative pairs
- **We do not know how to**
  - align multiple encoders
  - encode/index large tables
  - encode/index large graphs



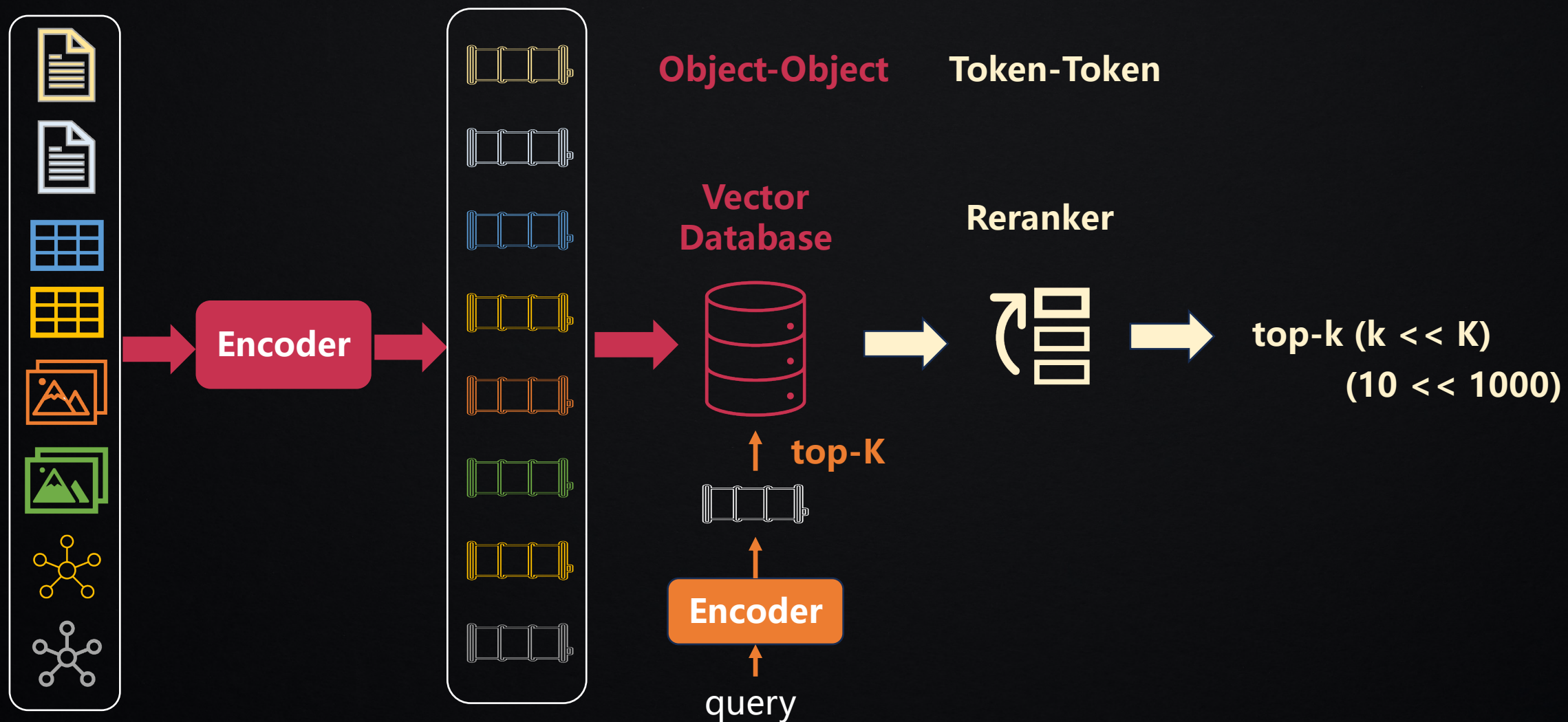


# Outline

- **What: Verify the outputs of LLMs (True/False)**
- **How: Retrieval Rerank**
- **Results**
- **Opportunities**



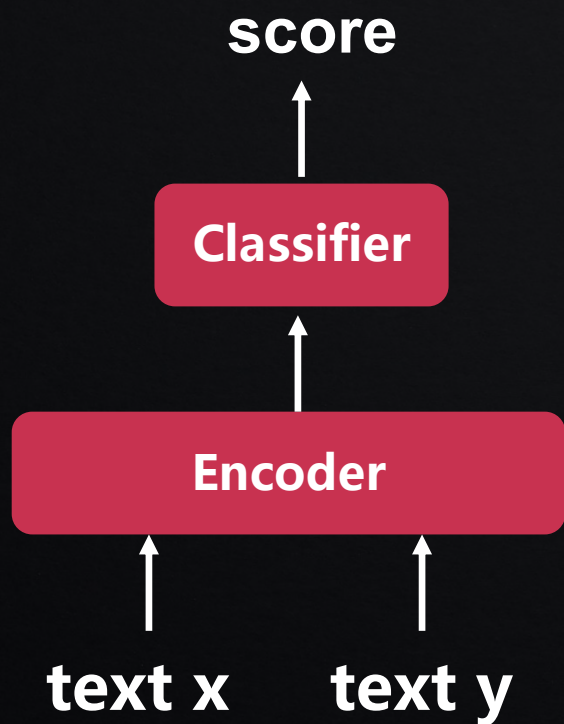
# Retrieval (Coarse-grained) -> Rerank (Fine-grained)



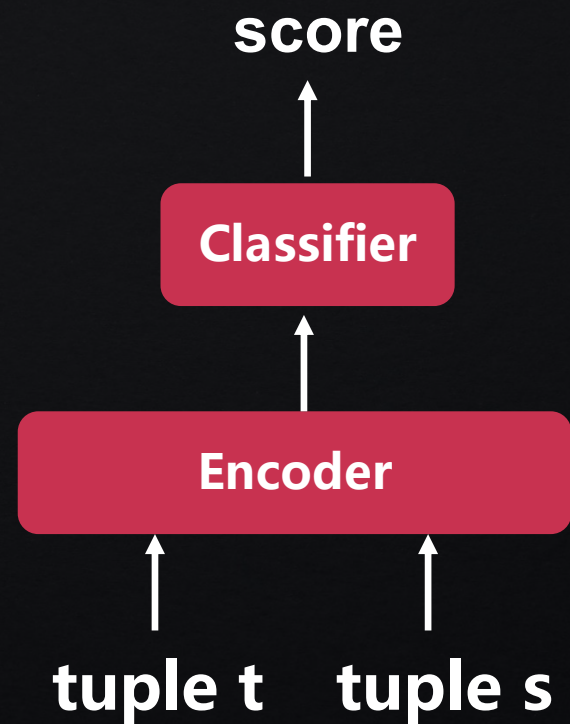


# Rerank

## Text-Text: Cross-Encoder (monoBERT)



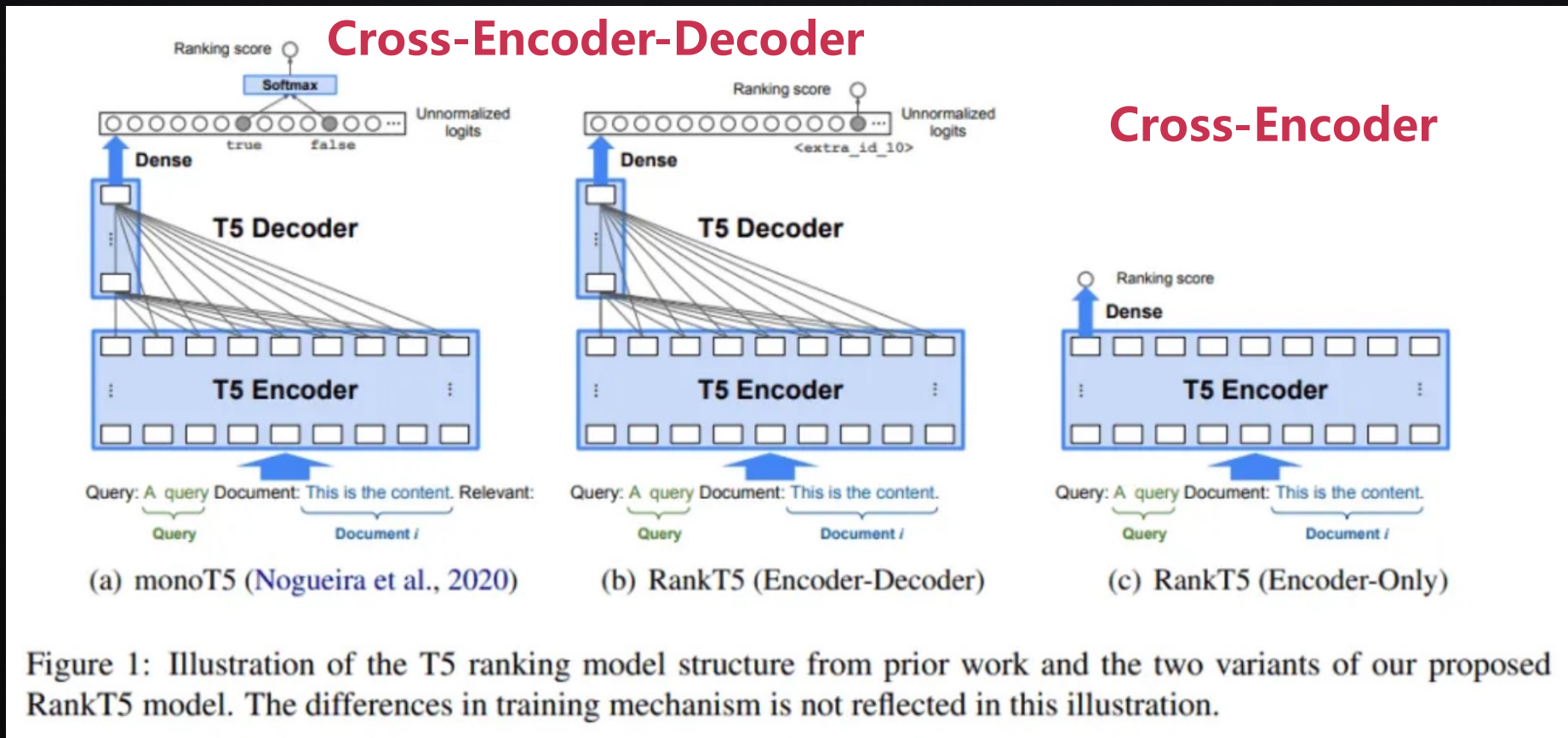
## Tuple-Tuple





# Rerank

## Text-Table: Cross-Encoder-Decoder (monoT5)



Note: Figure from the Google paper "RankT5: Fine-Tuning T5 for Text Ranking with Ranking Losses"



# Outline

- **What: Verify the outputs of LLMs (True/False)**
- **How: Retrieval Rerank Verifier**





# Outline

- **What: Verify the outputs of LLMs (True/False)**
- **How: retrieval, rerank, verifier**
- **Results**
- **Opportunities**

# Verify Tuple(s) via retrieving Tuple and Text



**Generated Tuple:** members of the 20th dail

constituency	name	party
Clare	Brendan Daly	Fianna Fáil (generated value)

## Retrieved Evidence and Verification

**E1:** Brendan Daly (2 February 1940 – 6 July 2023) was an Irish Fianna Fáil politician. He was a long-serving Teachta Dála(TD) for the Clare constituency, a government minister, and Senator.

**Verification result: True. Explanation:** The attribute 'party' value for Brendan Daly is Fianna Fáil, which matches with the claim.

## Retrieved Evidence and Verification

**E2:** members of the 20th dail

constituency	name	party
Tipperary South	Brendan Griffin	Fine Gael

**Verification result: Not Related. Explanation:** The evidence talks about the members of the 20th Dáil and the TDs elected by constituencies Tipperary South and Clare. It mentions Brendan Griffin as a member of the 20th Dáil from Tipperary South constituency, but it does not provide any information about Brendan Daly's party affiliation.



# Verify Tuple(s) via retrieving Tuple and Text

Generative AI task: **tuple completion**

**100** tuples with missing values

ChatGPT's accuracy in completion: **0.52**

Data lake: **13,796** textual files and 19,498 tables containing **269,622** tuples

Retrieval	Recall@3
Tuple	0.99
Text	0.58

**both** →

Verifier	Accuracy
ChatGPT	0.88

# Verify Text using via retrieving Tables



**Claim:** In 1954 u.s. open (golf), the cash prize for tommy bolt, fred haas, and ben hogan was 960 in total.

(**Ground Truth:** a false claim that should be **Refuted**)

## Retrieved Evidence and Verification

**Table E1:** 1954 u.s. open (golf)

place	player	country	score	to par	money
t1	ed furgol	united states	$71 + 70 + 71 + 72 = 284$	+ 4	1 6000
t2	gene littler	united states	$70 + 69 + 76 + 70 = 285$	+ 5	1 3600
t3	lloyd mangrum	united states	$72 + 71 + 72 + 71 = 286$	+ 6	1500
t3	dick mayer	united states	$72 + 71 + 70 + 73 = 286$	+ 6	1500
t5	bobby locke	south africa	$74 + 70 + 74 + 70 = 288$	+ 8	960
t6	tommy bolt	united states	$72 + 72 + 73 + 72 = 289$	+ 9	570
t6	fred haas	united states	$73 + 73 + 71 + 72 = 289$	+ 9	570
t6	ben hogan	united states	$71 + 70 + 76 + 72 = 289$	+ 9	570
t6	shelley mayfield	united states	$73 + 75 + 72 + 69 = 289$	+ 9	570
t6	billy joe patton (a)	united states	$69 + 76 + 71 + 73 = 289$	+ 9	1 0 1

**Verification result: Refuted. Explanation:** The cash prize for Tommy Bolt, Fred Haas, and Ben Hogan was \$570 each, totaling \$1710.

**Table E2:** 1959 u.s. open (golf)

player	country	year (s)	won	total	to par	finish
ben hogan	united states	1948, 1950, 1951, 1953	287	+ 7	t8	
cary middlecoff	united states	1949, 1956	294	+ 14	t19	
liack fleck	united states	1955	294	+ 14	t19	
liulius boros	united states	1952	297	+ 17	t28	
tommy bolt	united states	1958	301	+ 21	t38	


**Verification result: Not related.**



# Verify Text using via retrieving Tables

A controlled study: **1,300** textual claims from TabFact benchmark  
Retrieved from **19K Wikipedia Tables**

Retrieval	Recall@5
Table	0.88



Verifier	Reference	Accuracy
ChatGPT		0.54
ChatGPT	Tables	<b>0.91</b>

# Outline

- **What: Verify the outputs of LLMs (True/False)**
- **How: retrieval, rerank, verifier**
- **Results**
- **Opportunities**



# Opportunities (DB)

1. Table learning using LLMs
2. Table reasoning using LLMs
3. Table retrieval (small tables, large tables, database tables)
4. Cross-modal data discovery (aligning multiple encoders)
5. Cross-modal matching
6. Cross-modal reranking
7. Cross-modal reasoning

**Q & A**

