TIQ:
A Benchmark for Temporal Question Answering with Implicit Time Constraints

Zhen Jia [1], ***Philipp Christmann*** [2], *Gerhard Weikum* [2]

[1] **Southwest Jiaotong University,** Chengdu, China
[2] **Max Planck Institute for Informatics**, Saarbrücken, Germany

西南交通大學
Southwest Jiaotong University

TempWeb @ WWW 2024, Singapore

MAX PLANCK INSTITUTE
FOR INFORMATICS

Outline

★ Motivation

★ Construction of TIQ Benchmark

★ Characteristics of TIQ Benchmark

★ Experiments

★ Conclusion

Outline

★ Motivation

★ Construction of 🕓TIQ Benchmark

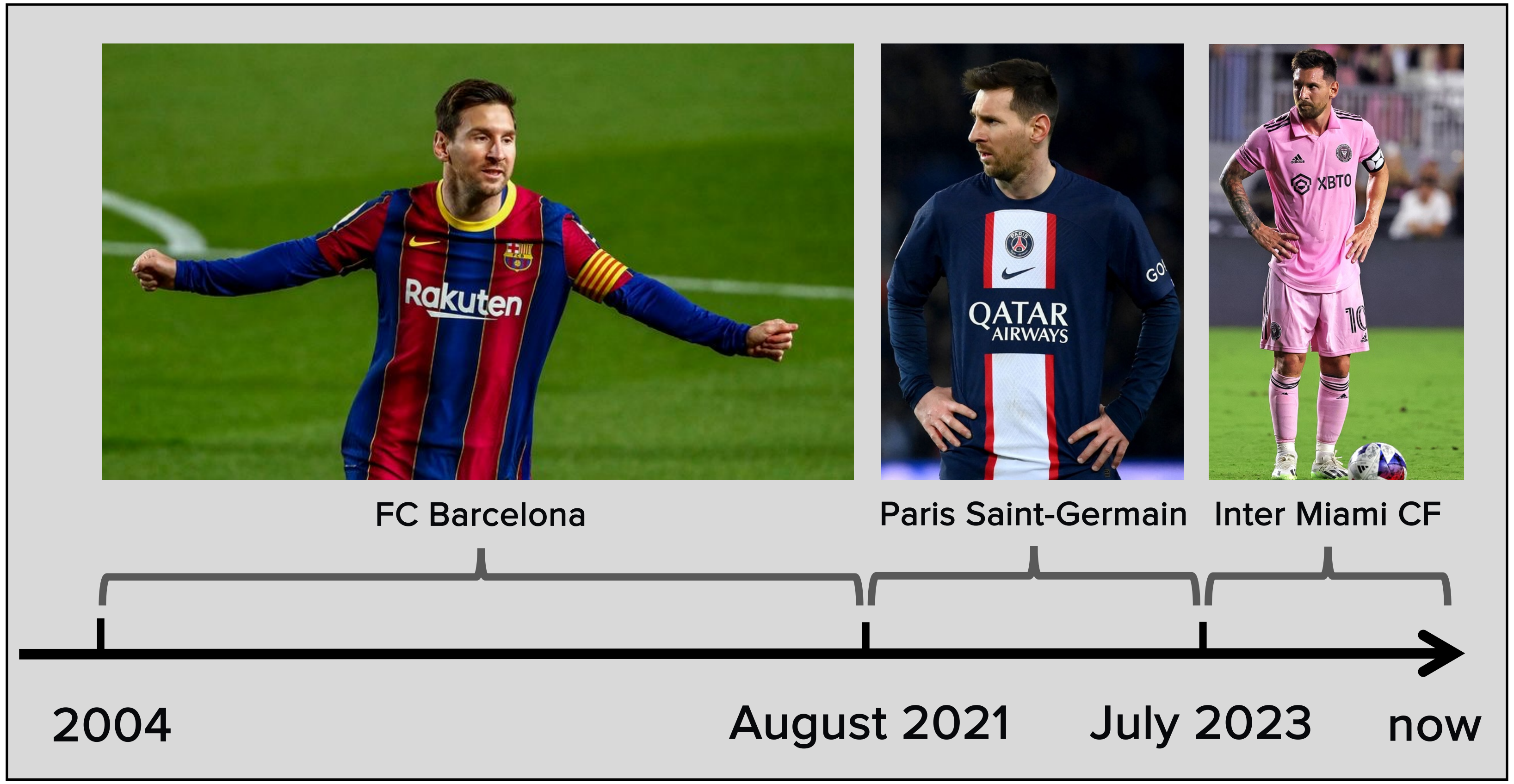
★ Characteristics of 🕓TIQ Benchmark

★ Experiments

★ Conclusion

Temporal questions

Which football club did Messi join in 2023?



⇒ Many user questions are time-sensitive

⇒ Temporal condition can be explicitly given (“*in 2023*”)

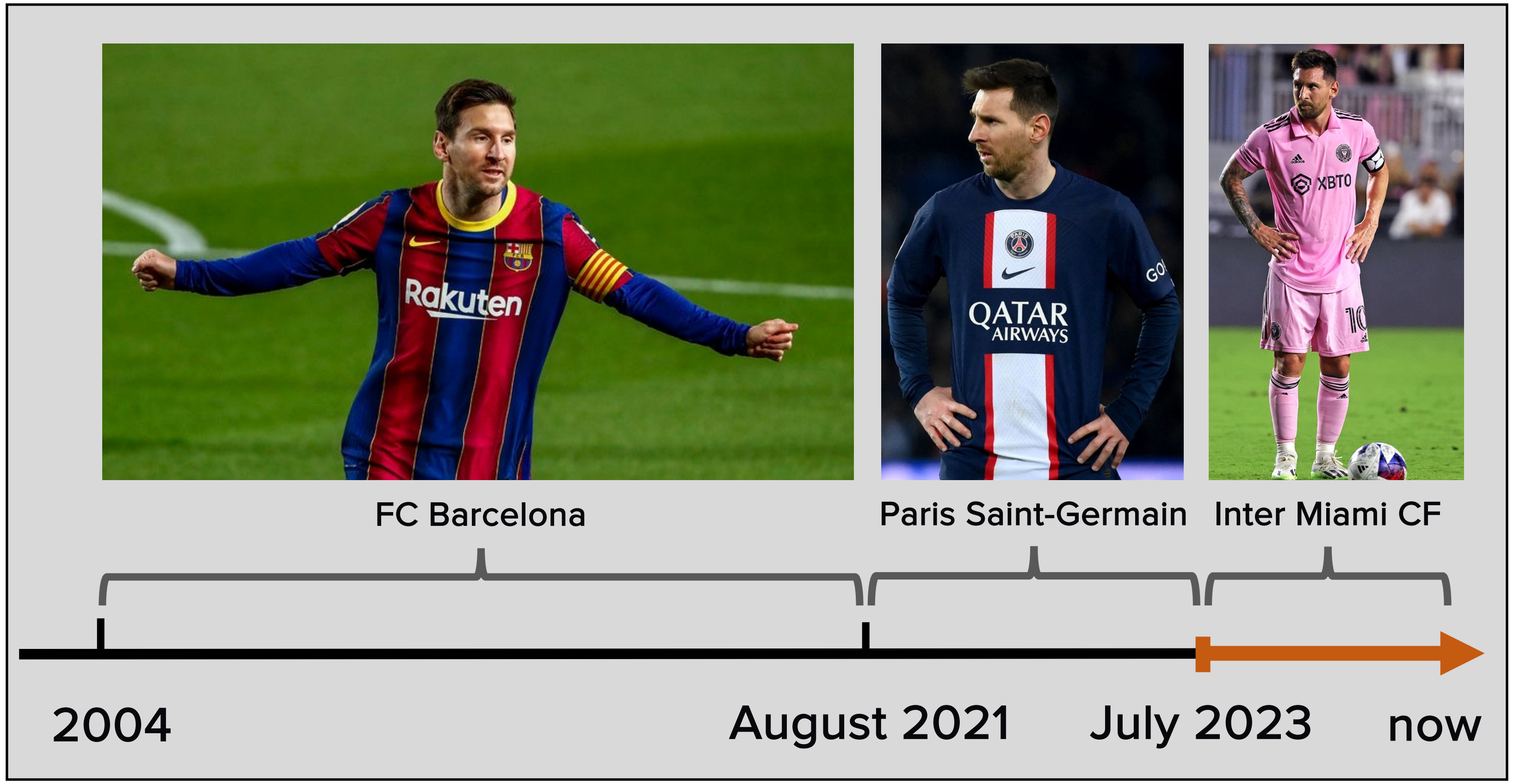
Implicit questions

Which football club did Messi join ***after Paris Saint-Germain****?*

⇒ Natural to provide temporal condition implicitly

Implicit questions

Which football club did Messi join after Paris Saint-Germain?



⇒ Natural to provide temporal condition implicitly

⇒ QA system has to understand temporal condition to answer question

⇒ Challenging for existing QA systems

Existing benchmarks for temporal QA

- ☆ Most benchmarks have **only few implicit questions**
  - ☆ TempQuestions (209), TimeQuestions (1,476), TempQA-WD (154)
- ☆ Some have more implicit questions, but **limited** in question **intents**
  - ☆ CronQuestions, TempReason: Both **template-based** and questions derived from **only 5/10 KB-relations**

Existing benchmarks for temporal QA

- ☆ Most benchmarks have **only few implicit questions**
  - ☆ TempQuestions (209), TimeQuestions (1,476), TempQA-WD (154)
- ☆ Some have more implicit questions, but **limited** in question **intents**
  - ☆ CronQuestions, TempReason: Both **template-based** and questions derived from **only 5/10 KB-relations**

CronQuestions

1. Member of sports team
2. Position held
3. Award received
4. Spouse
5. Employer

TempReason

1. Member of sports team
2. Position held
3. Employer
4. Political party
5. Head coach
6. Educated at
7. Chairperson
8. Head of government
9. Head of state
10. Owned by

Existing benchmarks for temporal QA

- ☆ Most benchmarks have **only few implicit questions**
  - ☆ TempQuestions (209), TimeQuestions (1,476), TempQA-WD (154)
- ☆ Some have more implicit questions, but **limited** in question **intents**
  - ☆ CronQuestions, TempReason: Both **template-based** and questions derived from **only 5/10 KB-relations**
- ☆ Questions are derived from **single source only**
  - ☆ TempQuestions, TimeQuestions, TempQA-WD, TempTabQA, CronQuestion, TempReason

TIQ: Temporal Implicit Questions

- New benchmark for temporal QA with 10,000 implicit questions
- Questions derived from multiple knowledge sources
  - Wikipedia text
  - Wikipedia infoboxes
  - Wikidata Knowledge Base

TIQ: Temporal Implicit Questions

★ Derived using automatic construction pipeline

★ Configurable among various dimensions

- ✓ Temporal scope of questions
- ✓ Domain diversity of topic entities
- ✓ Ratio of prominent vs. long-tail entities
- ✓ Question complexity
- ✓ Total number of questions
- ✓ …

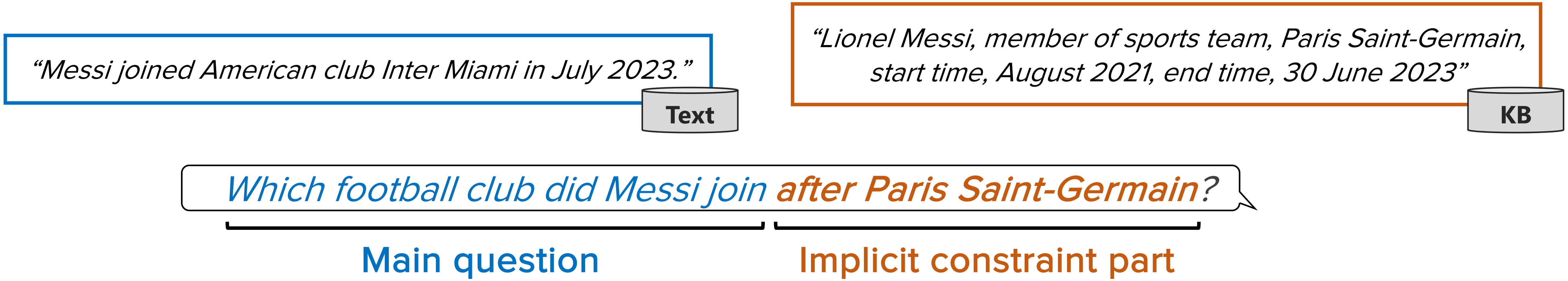
Outline

- Motivation
- Construction of TIQ Benchmark
- Characteristics of TIQ Benchmark
- Experiments
- Conclusion

Construction of TIQ benchmark

Key idea #1: Implicit questions consist of **main question** and implicit constraint

⇒ Initialize parts individually, based on different information snippets



"Messi joined American club Inter Miami in July 2023."

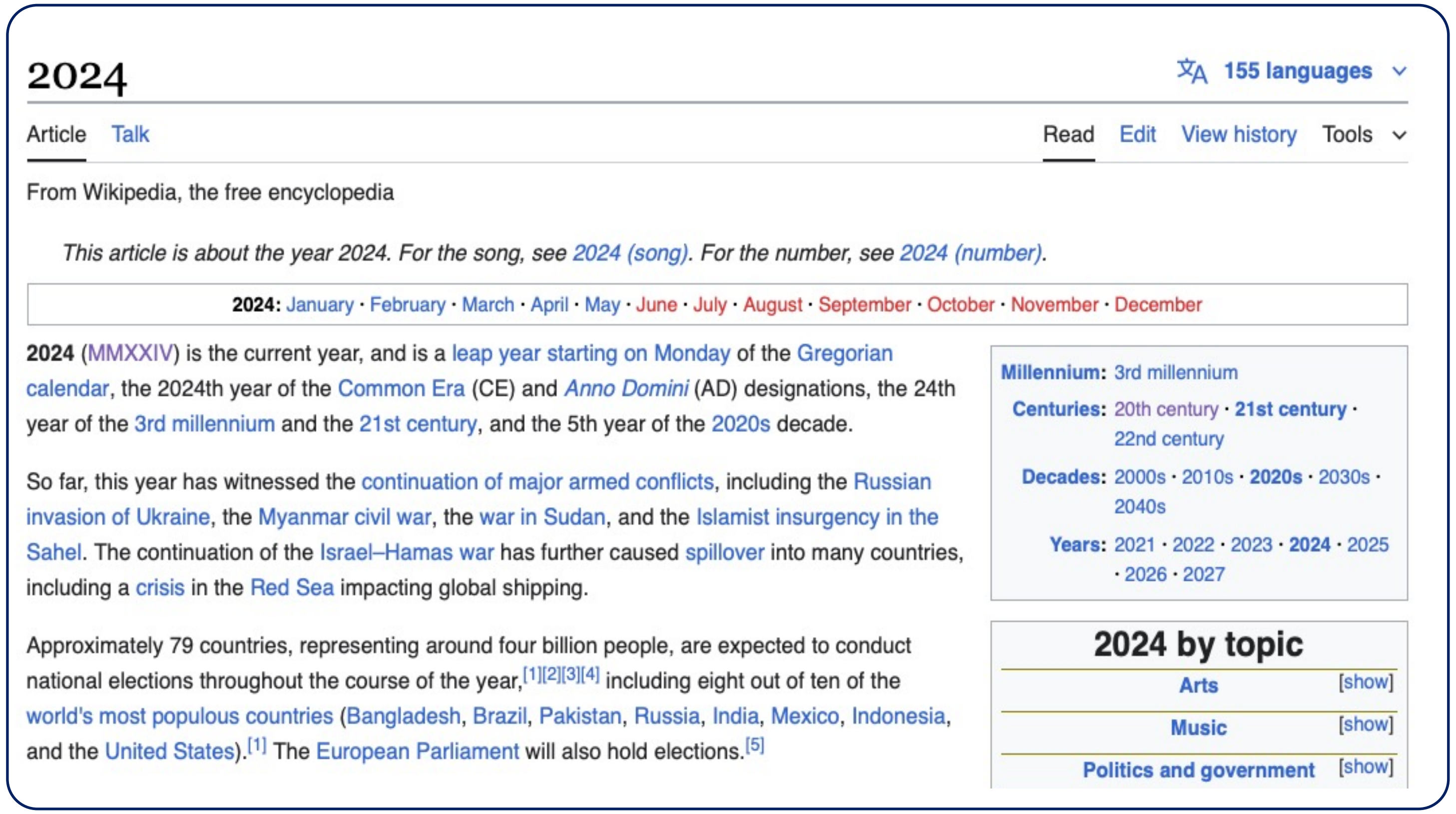
"Lionel Messi, member of sports team, Paris Saint-Germain, start time, August 2021, end time, 30 June 2023"

Which football club did Messi join after Paris Saint-Germain?

Construction of TIQ benchmark

Key idea #2: Start from Wikipedia year pages

⇒ Provide salient information



2024

155 languages

Article Talk Read Edit View history Tools

From Wikipedia, the free encyclopedia

This article is about the year 2024. For the song, see 2024 (song). For the number, see 2024 (number).

2024: January · February · March · April · May · June · July · August · September · October · November · December

2024 (MMXXIV) is the current year, and is a leap year starting on Monday of the Gregorian calendar, the 2024th year of the Common Era (CE) and *Anno Domini* (AD) designations, the 24th year of the 3rd millennium and the 21st century, and the 5th year of the 2020s decade.

So far, this year has witnessed the continuation of major armed conflicts, including the Russian invasion of Ukraine, the Myanmar civil war, the war in Sudan, and the Islamist insurgency in the Sahel. The continuation of the Israel–Hamas war has further caused spillover into many countries, including a crisis in the Red Sea impacting global shipping.

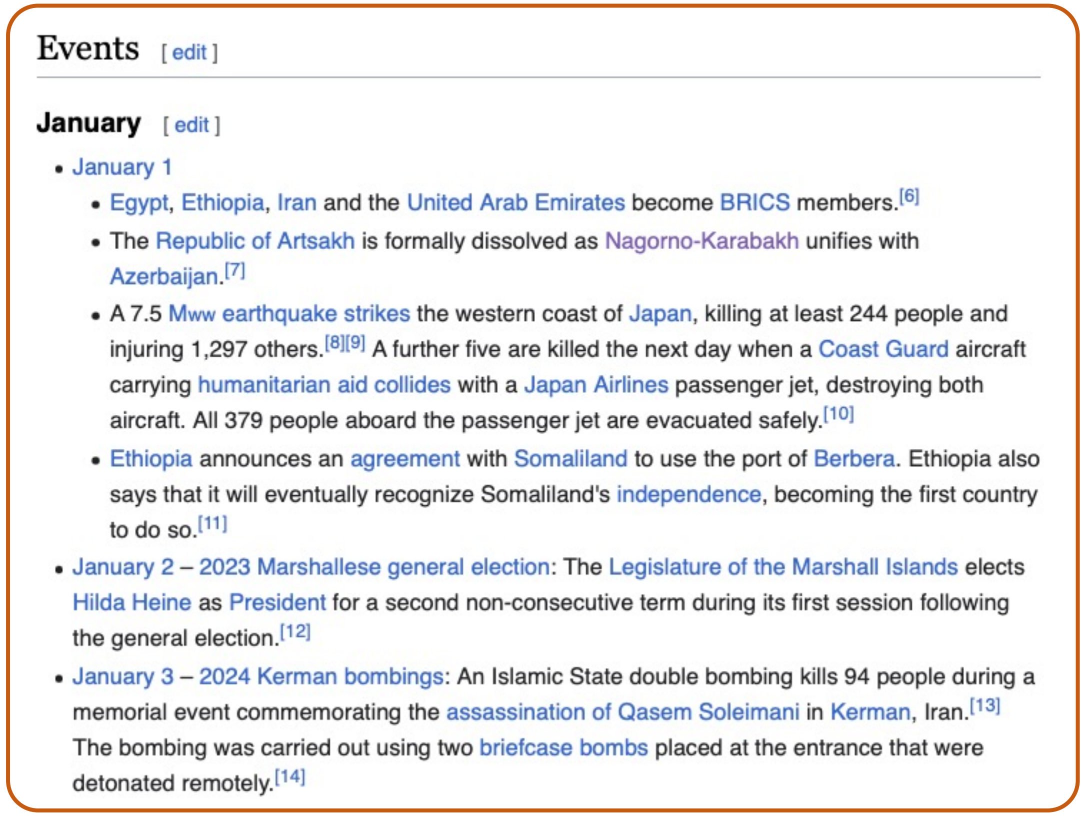
Approximately 79 countries, representing around four billion people, are expected to conduct national elections throughout the course of the year,[1][2][3][4] including eight out of ten of the world's most populous countries (Bangladesh, Brazil, Pakistan, Russia, India, Mexico, Indonesia, and the United States).[1] The European Parliament will also hold elections.[5]

Millennium:	3rd millennium
Centuries:	20th century · **21st century** · 22nd century
Decades:	2000s · 2010s · **2020s** · 2030s · 2040s
Years:	2021 · 2022 · 2023 · **2024** · 2025 · 2026 · 2027

2024 by topic

Arts	[show]
Music	[show]
Politics and government	[show]

Wikipedia year pages



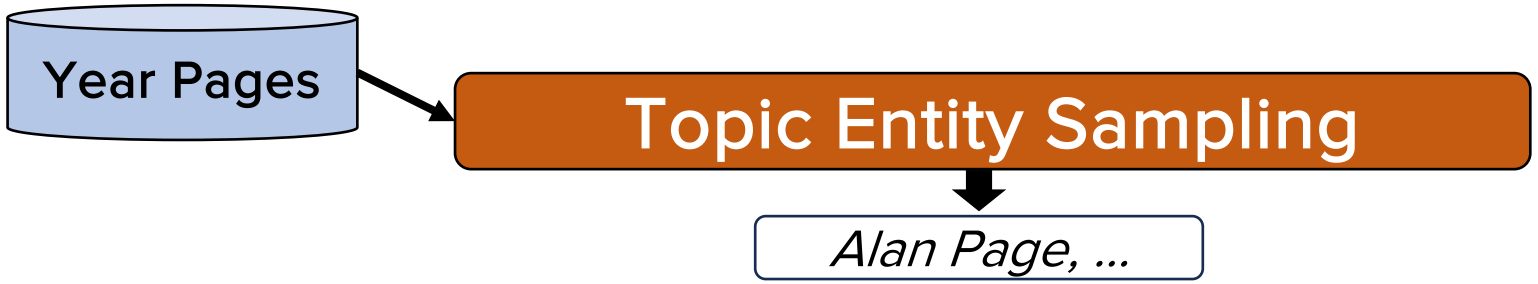
Events [edit]

January [edit]

- January 1
  - Egypt, Ethiopia, Iran and the United Arab Emirates become BRICS members.[6]
  - The Republic of Artsakh is formally dissolved as Nagorno-Karabakh unifies with Azerbaijan.[7]
  - A 7.5 Mww earthquake strikes the western coast of Japan, killing at least 244 people and injuring 1,297 others.[8][9] A further five are killed the next day when a Coast Guard aircraft carrying humanitarian aid collides with a Japan Airlines passenger jet, destroying both aircraft. All 379 people aboard the passenger jet are evacuated safely.[10]
  - Ethiopia announces an agreement with Somaliland to use the port of Berbera. Ethiopia also says that it will eventually recognize Somaliland's independence, becoming the first country to do so.[11]
- January 2 – 2023 Marshallese general election: The Legislature of the Marshall Islands elects Hilda Heine as President for a second non-consecutive term during its first session following the general election.[12]
- January 3 – 2024 Kerman bombings: An Islamic State double bombing kills 94 people during a memorial event commemorating the assassination of Qasem Soleimani in Kerman, Iran.[13] The bombing was carried out using two briefcase bombs placed at the entrance that were detonated remotely.[14]

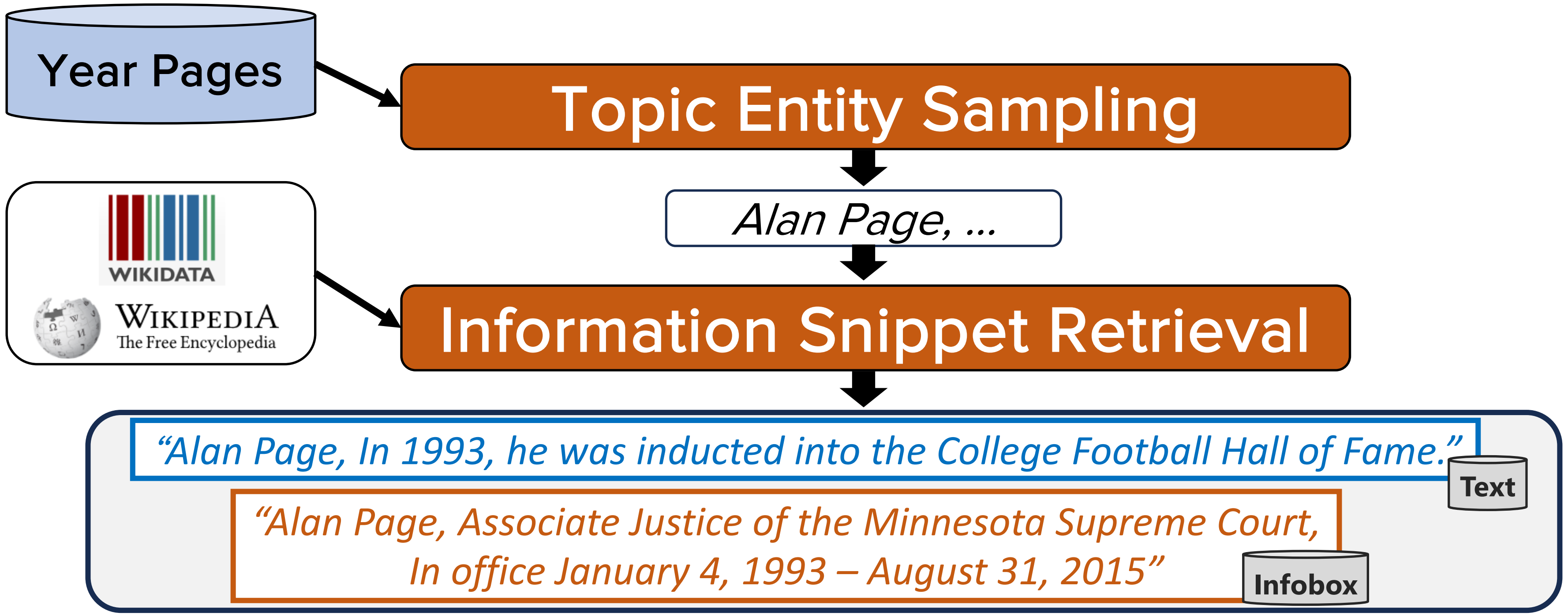
List of notable events (with dates)

Construction of TIQ benchmark



Sample topic entities from Wikipedia year pages

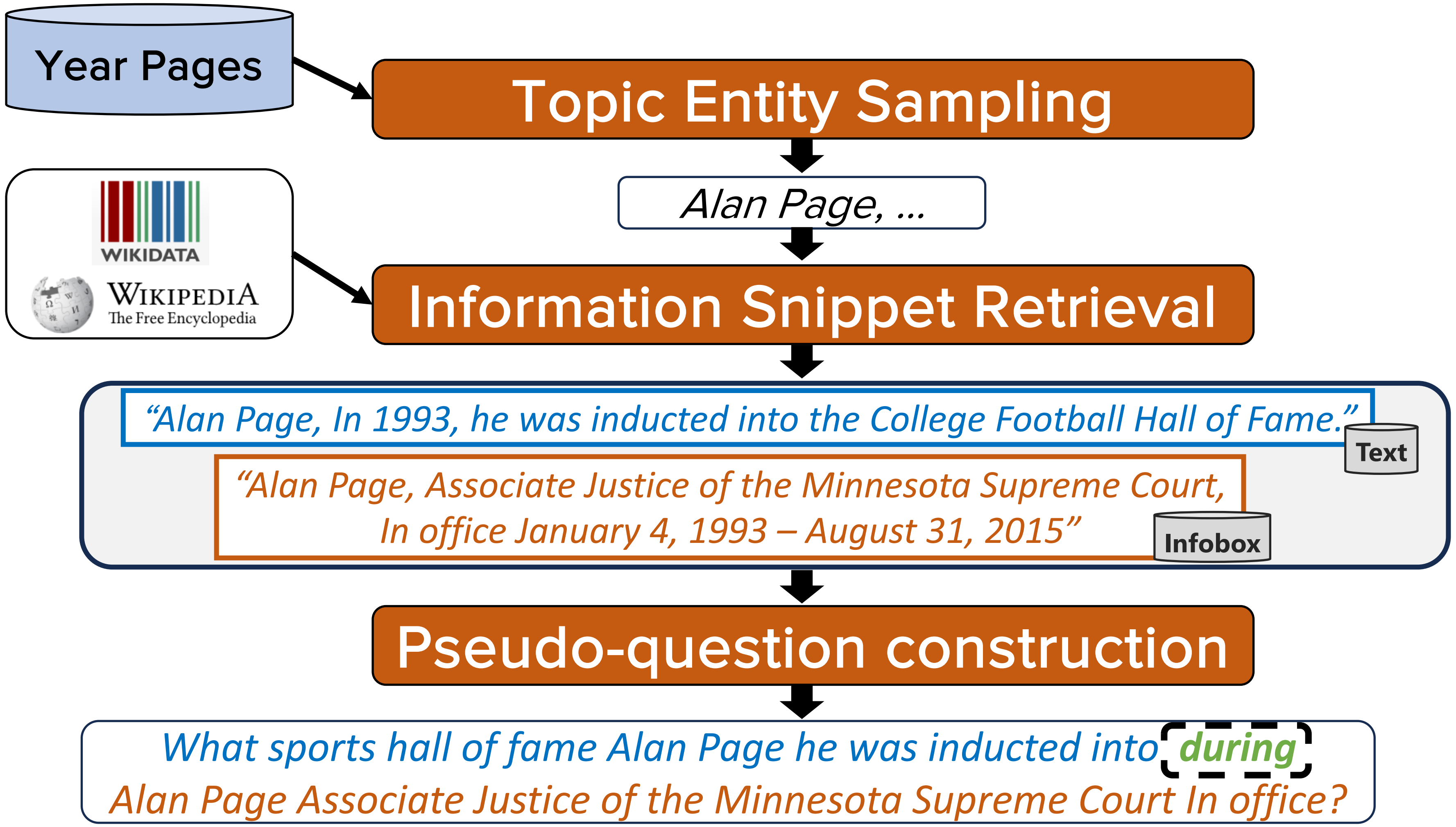
Construction of TIQ benchmark


Year Pages
Topic Entity Sampling
Alan Page, ...
WIKIDATA
Wikipedia The Free Encyclopedia
Information Snippet Retrieval
"Alan Page, In 1993, he was inducted into the College Football Hall of Fame."
Text
"Alan Page, Associate Justice of the Minnesota Supreme Court, In office January 4, 1993 – August 31, 2015"
Infobox

Sample topic entities from Wikipedia year pages

Retrieve information snippets from Wikipedia (text and infoboxes), and Wikidata (KB)

Construction of TIQ benchmark


Year Pages
Topic Entity Sampling
Alan Page, …
WIKIDATA
WIKIPEDIA The Free Encyclopedia
Information Snippet Retrieval
Text
Infobox
Pseudo-question construction
during

"Alan Page, In 1993, he was inducted into the College Football Hall of Fame."

"Alan Page, Associate Justice of the Minnesota Supreme Court, In office January 4, 1993 – August 31, 2015"

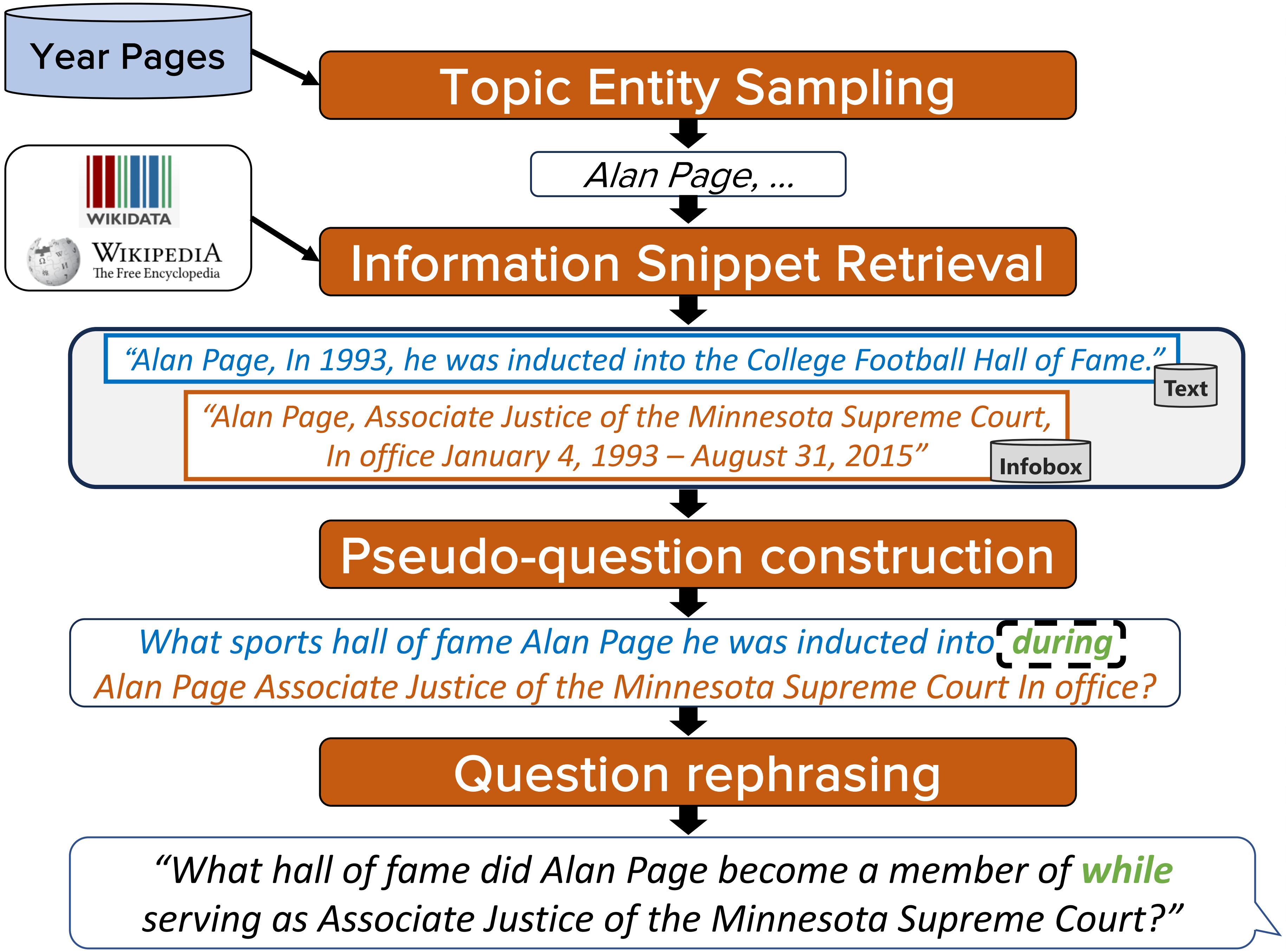
*What sports hall of fame Alan Page he was inducted into **during** Alan Page Associate Justice of the Minnesota Supreme Court In office?*

Sample topic entities from Wikipedia year pages

Retrieve information snippets from Wikipedia (text and infoboxes), and Wikidata (KB)

Connect snippets with temporal relation

Construction of TIQ benchmark



Topic Entity Sampling

Alan Page, …

Information Snippet Retrieval

“Alan Page, In 1993, he was inducted into the College Football Hall of Fame.”

“Alan Page, Associate Justice of the Minnesota Supreme Court, In office January 4, 1993 – August 31, 2015”

Pseudo-question construction

*What sports hall of fame Alan Page he was inducted into **during** Alan Page Associate Justice of the Minnesota Supreme Court In office?*

Question rephrasing

*“What hall of fame did Alan Page become a member of **while** serving as Associate Justice of the Minnesota Supreme Court?”*

Sample topic entities from Wikipedia year pages

Retrieve information snippets from Wikipedia (text and infoboxes), and Wikidata (KB)

Connect snippets with temporal relation

Rephrase the pseudo-question into a natural question

Outline

★ Motivation

★ Construction of TIQ Benchmark

★ Characteristics of TIQ Benchmark

★ Experiments

★ Conclusion

TIQ Example: #1

Topic entity

Alicia Keys

Evidence

“Alicia Keys, Keys followed up her debut with The Diary of Alicia Keys, which was released in December 2003.”

Text

“Norah Jones, award received, Grammy Award for Best New Artist, follows, Alicia Keys, point in time, 2003”

KB

Pseudo-question

*What album Alicia Keys Keys followed up her debut with which was released **during** Norah Jones award received Grammy Award for Best New Artist follows Alicia Keys*

Question

“What album did Alicia Keys release when Norah Jones won the Grammy Award for Best New Artist?”

TIQ Example: #2

Topic entity

Chris Brown

Evidence

“Chris Brown, His fifth album, Fortune, released in 2012, also topped the Billboard 200.”

Text

“Chris Brown, Chris Brown, Brown performing in Sydney, 2012”

Infobox

Pseudo-question

What album Chris Brown His fifth album released also topped the Billboard 200 during Chris Brown Chris Brown Brown performing in Sydney

Question

“Which album released by Chris Brown topped the Billboard 200 when he was performing in Sydney?”

TIQ Example: #3

Topic entity

Clarence Andrew Cannon

Evidence

"Clarence Andrew Cannon, occupation, teacher, start time, 1904, end time, 1908"

KB

"Clarence Cannon, He earned an LL.B. and joined the bar in 1908."

Infobox

Pseudo-question

What position Clarence Andrew Cannon occupation before Clarence Cannon He earned an LL.B . and joined the bar

Question

"What was Clarence Andrew Cannon's occupation before becoming a lawyer?"

TIQ Example: #4

Topic entity

Hulk Hogan

Evidence

"Hulk Hogan, He starred in his own television series, Thunder in Paradise, in 1994." Text

"Hulk Hogan, He was lured back to the ring when he signed with rival promotion World Championship Wrestling (WCW) in 1994." Text

Pseudo-question

What television series Hulk Hogan He starred in his own television series during Hulk Hogan He was lured back to the ring when he signed with rival promotion World Championship Wrestling (WCW)

Question

"What television series was Hulk Hogan starring in when he signed with World Championship Wrestling?"

TIQ: Key statistics

- 10,000 implicit questions

  6,000 train
  2,000 dev
  2,000 test

- Derived from 10,000 topic entities

- Temporal relations

  “before”: 14%
  “during”: 66%
  “after”: 20%

TIQ: Balance of entity popularity

★ Balanced fraction of long-tail entities vs. prominent entities

long-tail entity: < 20 KB-facts

prominent entity: > 500 KB-facts

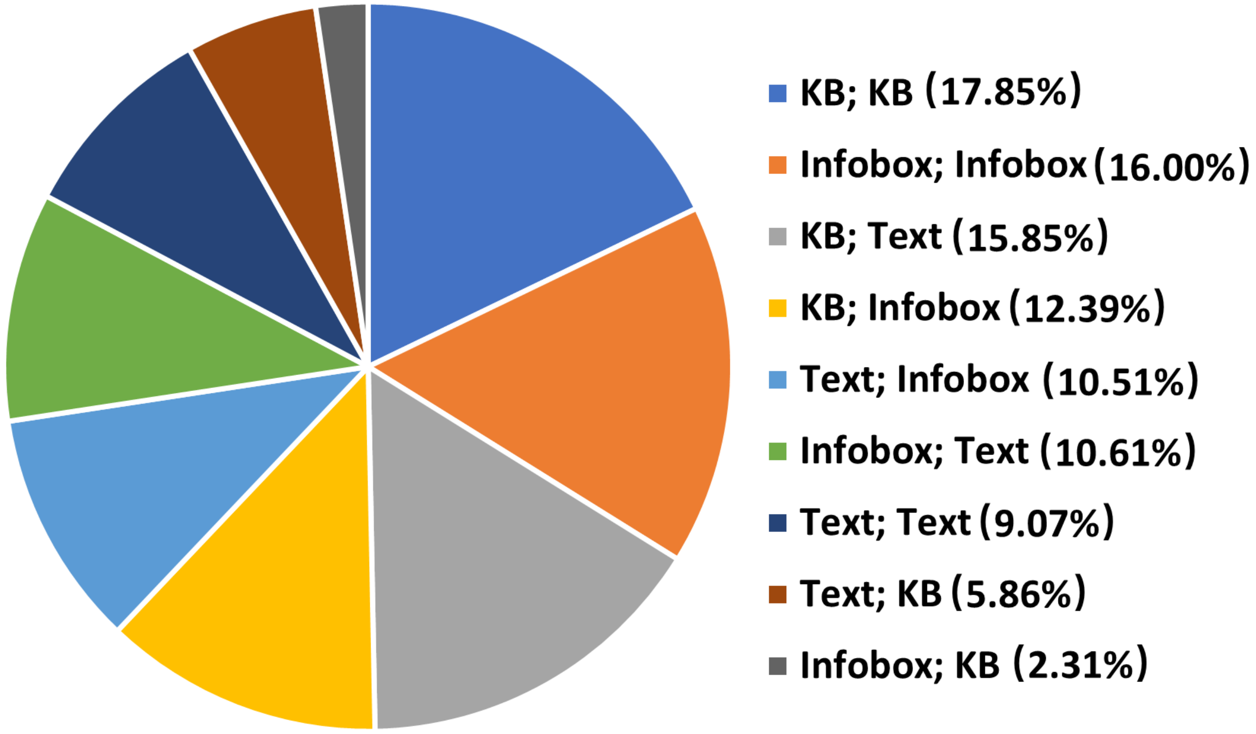
★ 2,542 long-tail topic entities

★ 2,613 prominent topic entities

⇒ Actively controlled when sampling topic entities

⇒ Challenging test bed for LLMs

TIQ: Diverse source combinations



⇒ Source combinations balanced as well

⇒ Not actively controlled

TIQ: Further statistics (more in the paper)

- Temporal values

  12,094 years

  538 months

  7,992 dates

- Year pages: [1801, 2025]

- 2.45 entities per question

- 17.96 words per question

Outline

- ★ Motivation
- ★ Construction of 🕓TIQ Benchmark
- ★ Characteristics of 🕓TIQ Benchmark
- ★ Experiments
- ★ Conclusion

Experimental setup

★ Experiments with diverse range of QA systems

- ★ Generative LLMs: InstructGPT, GPT-4
- ★ Heterogeneous QA: UNIQORN, UniK-QA, EXPLAIGNN
- ★ Temporal QA: EXAQT, FAITH (WWW2024)

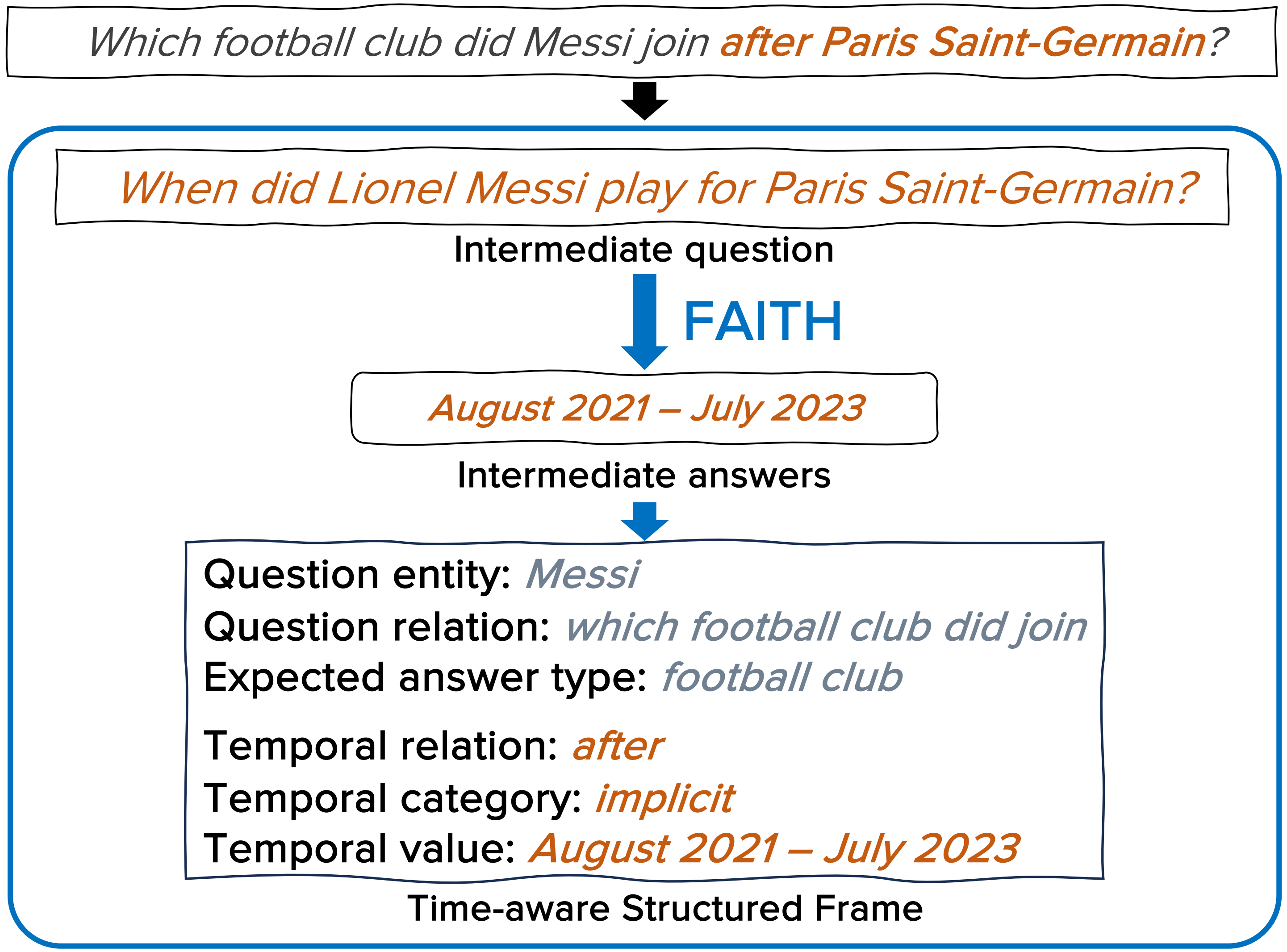
FAITH (WWW2024)

Presented on Thursday, 16 May in Central Ballroom
(Poster Session from 16:30-18:00 on Security, Semantics, Social, Systems)

Which football club did Messi join after Paris Saint-Germain?

FAITH (WWW2024)

Presented on Thursday, 16 May in Central Ballroom
(Poster Session from 16:30-18:00 on Security, Semantics, Social, Systems)



Which football club did Messi join after Paris Saint-Germain?

When did Lionel Messi play for Paris Saint-Germain?

Intermediate question

FAITH

August 2021 – July 2023

Intermediate answers

Question entity: *Messi*
Question relation: *which football club did join*
Expected answer type: *football club*

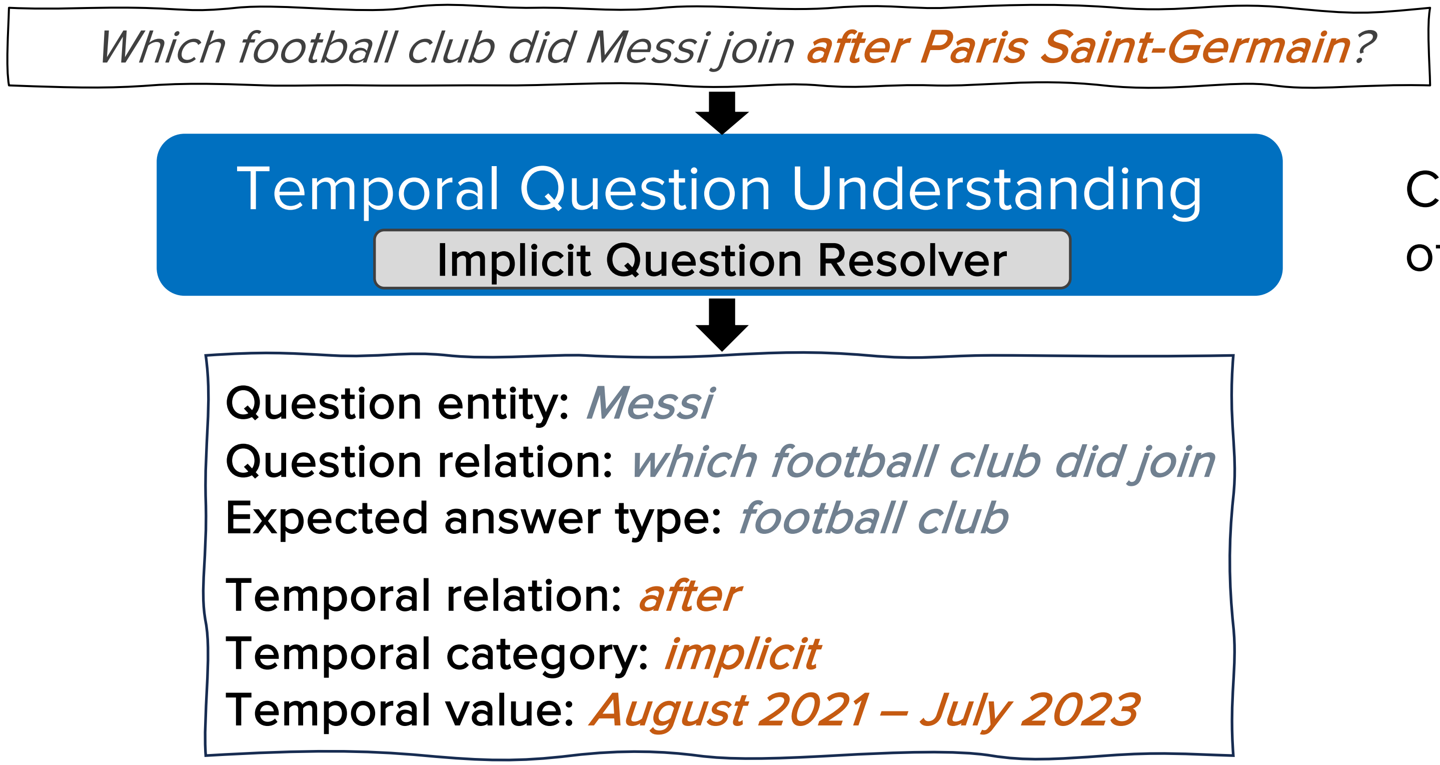
Temporal relation: *after*
Temporal category: *implicit*
Temporal value: *August 2021 – July 2023*

Time-aware Structured Frame

Temporal Question Understanding

FAITH (WWW2024)

Presented on Thursday, 16 May in Central Ballroom
(Poster Session from 16:30-18:00 on Security, Semantics, Social, Systems)



Which football club did Messi join after Paris Saint-Germain?

Temporal Question Understanding

Implicit Question Resolver

Question entity: *Messi*
Question relation: *which football club did join*
Expected answer type: *football club*

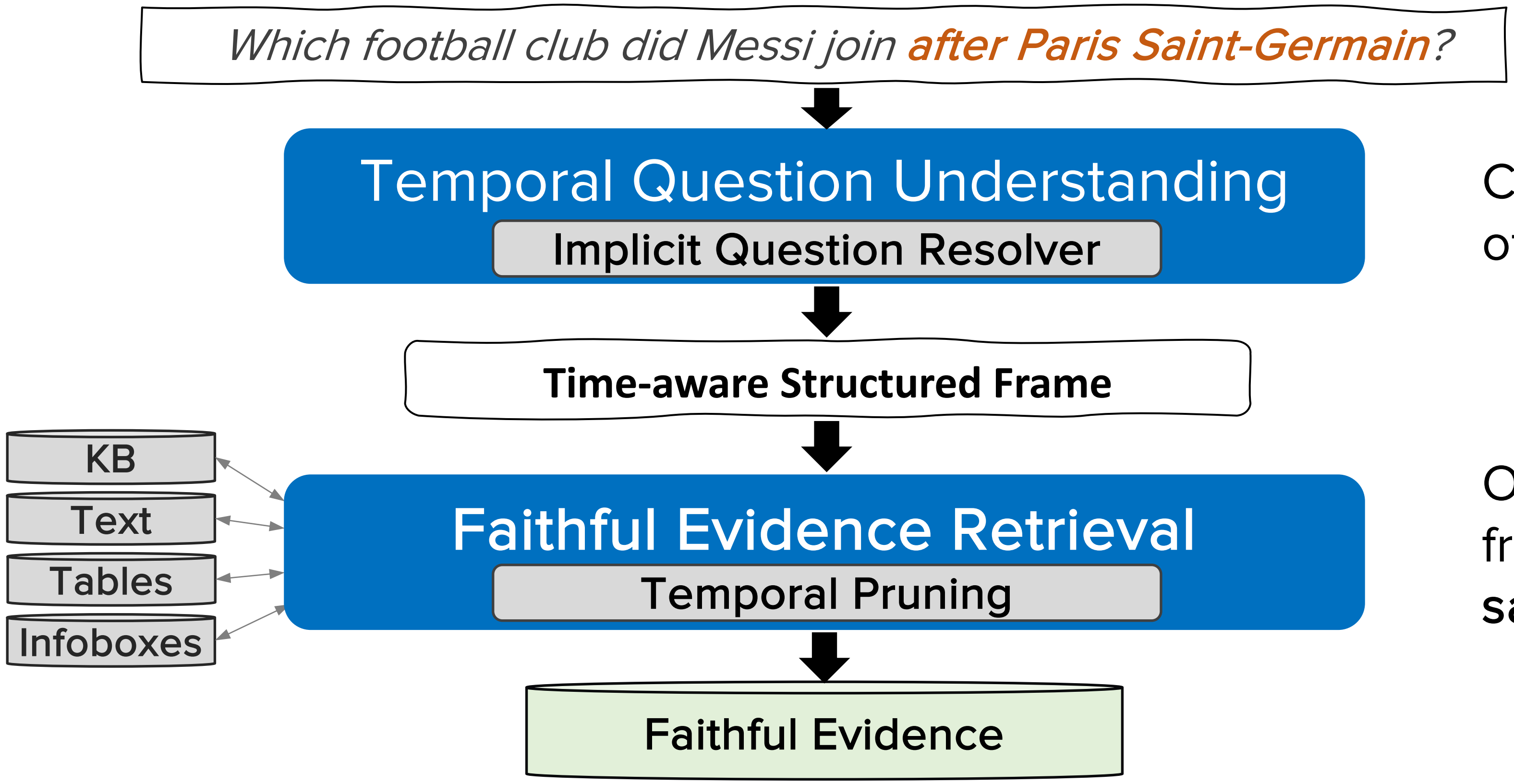
Temporal relation: *after*
Temporal category: *implicit*
Temporal value: *August 2021 – July 2023*

Time-aware Structured Frame

Capture **temporal information need** of user intent in a structured manner

FAITH (WWW2024)

Presented on Thursday, 16 May in Central Ballroom
(Poster Session from 16:30-18:00 on Security, Semantics, Social, Systems)

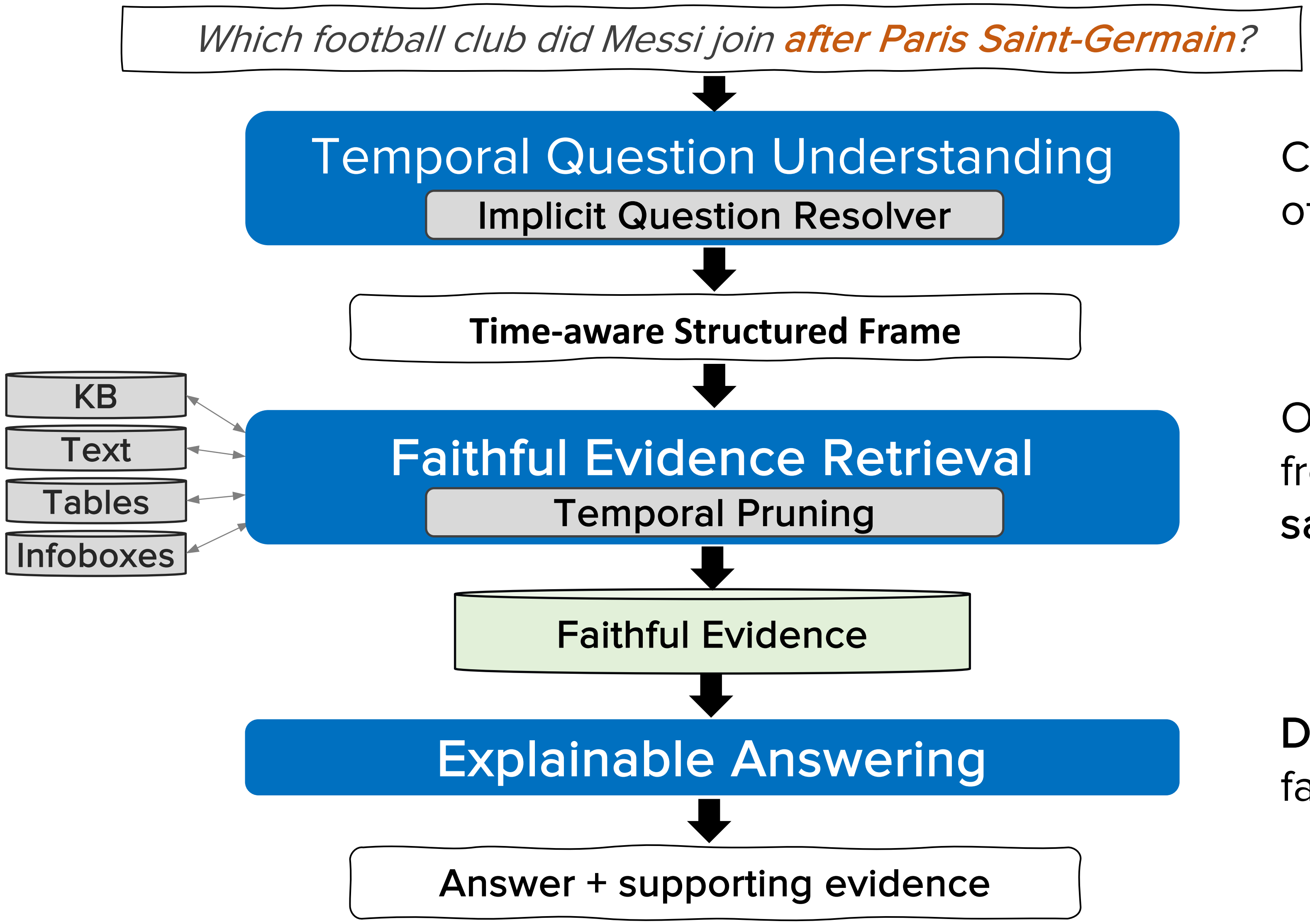

Which football club did Messi join after Paris Saint-Germain?
Temporal Question Understanding
Implicit Question Resolver
Time-aware Structured Frame
KB
Text
Tables
Infoboxes
Faithful Evidence Retrieval
Temporal Pruning
Faithful Evidence

Capture **temporal information need**
of user intent in a structured manner

Obtain **set of evidences**
from **heterogeneous sources**
satisfying the **temporal conditions**

FAITH (WWW2024)

Presented on Thursday, 16 May in Central Ballroom
(Poster Session from 16:30-18:00 on Security, Semantics, Social, Systems)



Which football club did Messi join after Paris Saint-Germain?
Temporal Question Understanding
Implicit Question Resolver
Time-aware Structured Frame
KB
Text
Tables
Infoboxes
Faithful Evidence Retrieval
Temporal Pruning
Faithful Evidence
Explainable Answering
Answer + supporting evidence

Capture **temporal information need** of user intent in a structured manner

Obtain **set of evidences** from **heterogeneous sources** **satisfying** the **temporal conditions**

Derive the answer from the faithful evidences

Experimental setup

★ Experiments with diverse range of QA systems

  ★ **Generative LLMs:** InstructGPT, GPT-4

  ★ **Heterogeneous QA:** UNIQORN, UniK-QA, EXPLAIGNN

  ★ **Temporal QA:** EXAQT, FAITH (WWW2024)

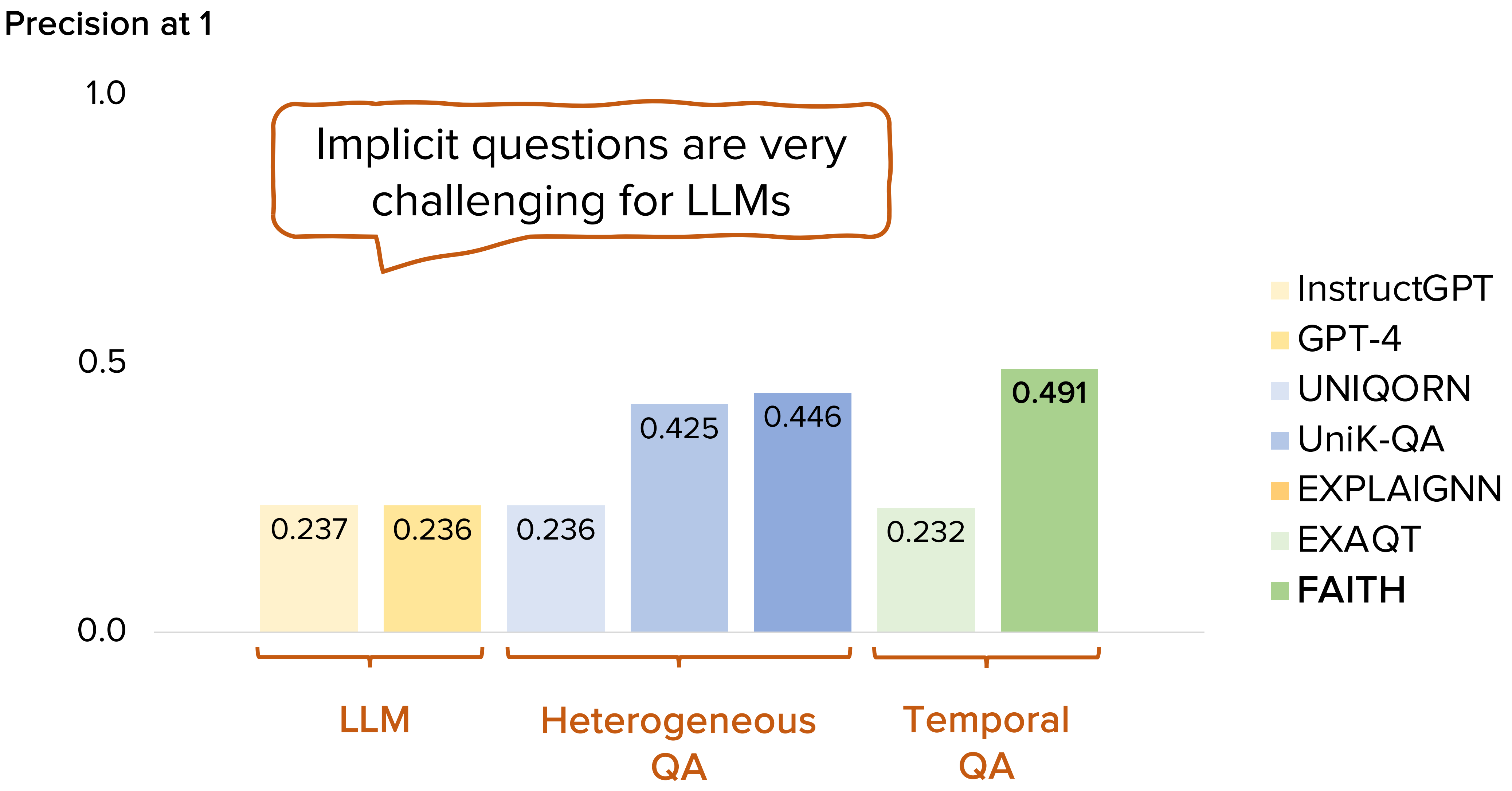
★ Metrics

  ★ Precision at 1 (**P@1**)

  ★ Mean reciprocal rank

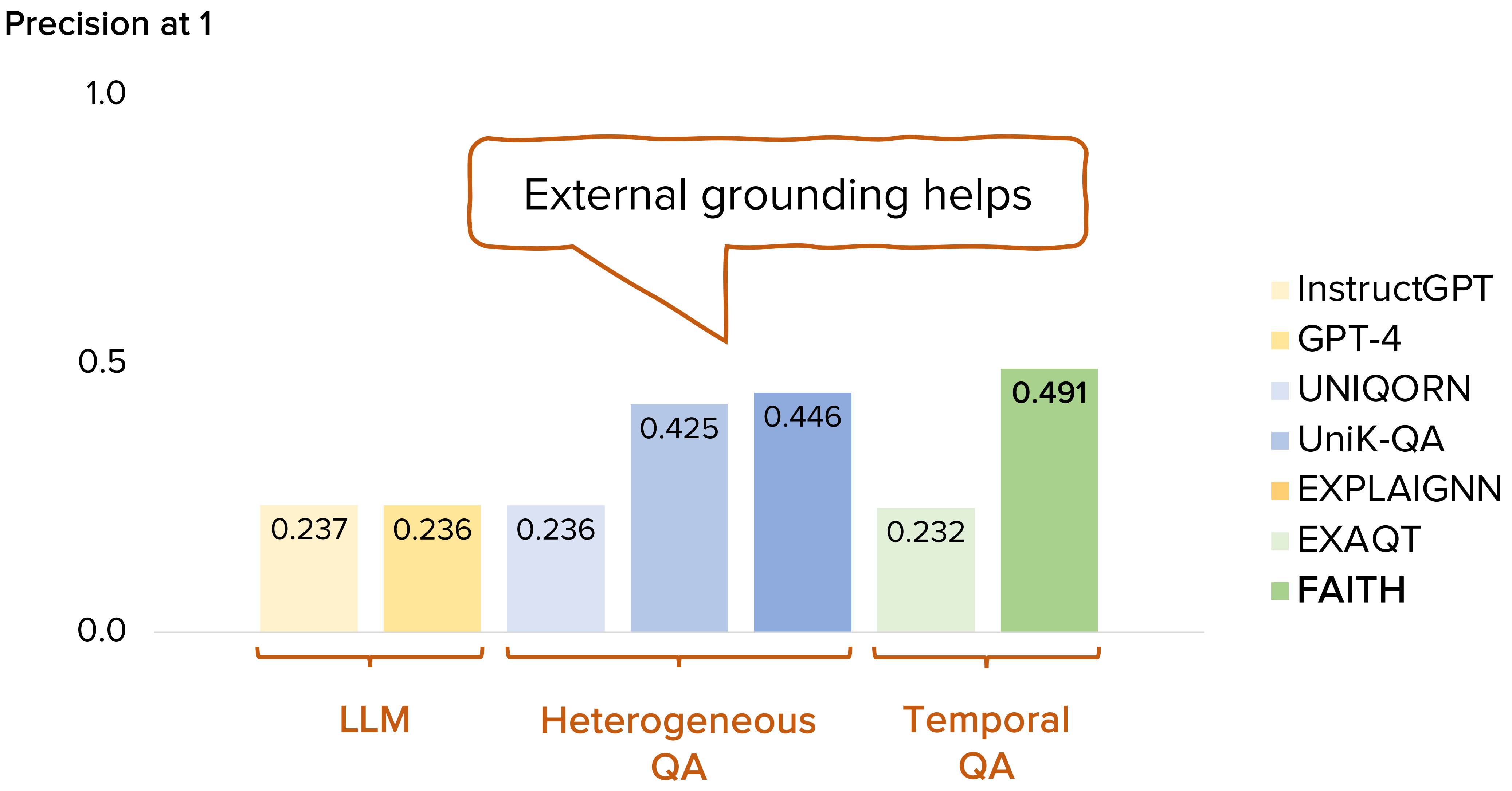
  ★ Hits at 5

Experimental results



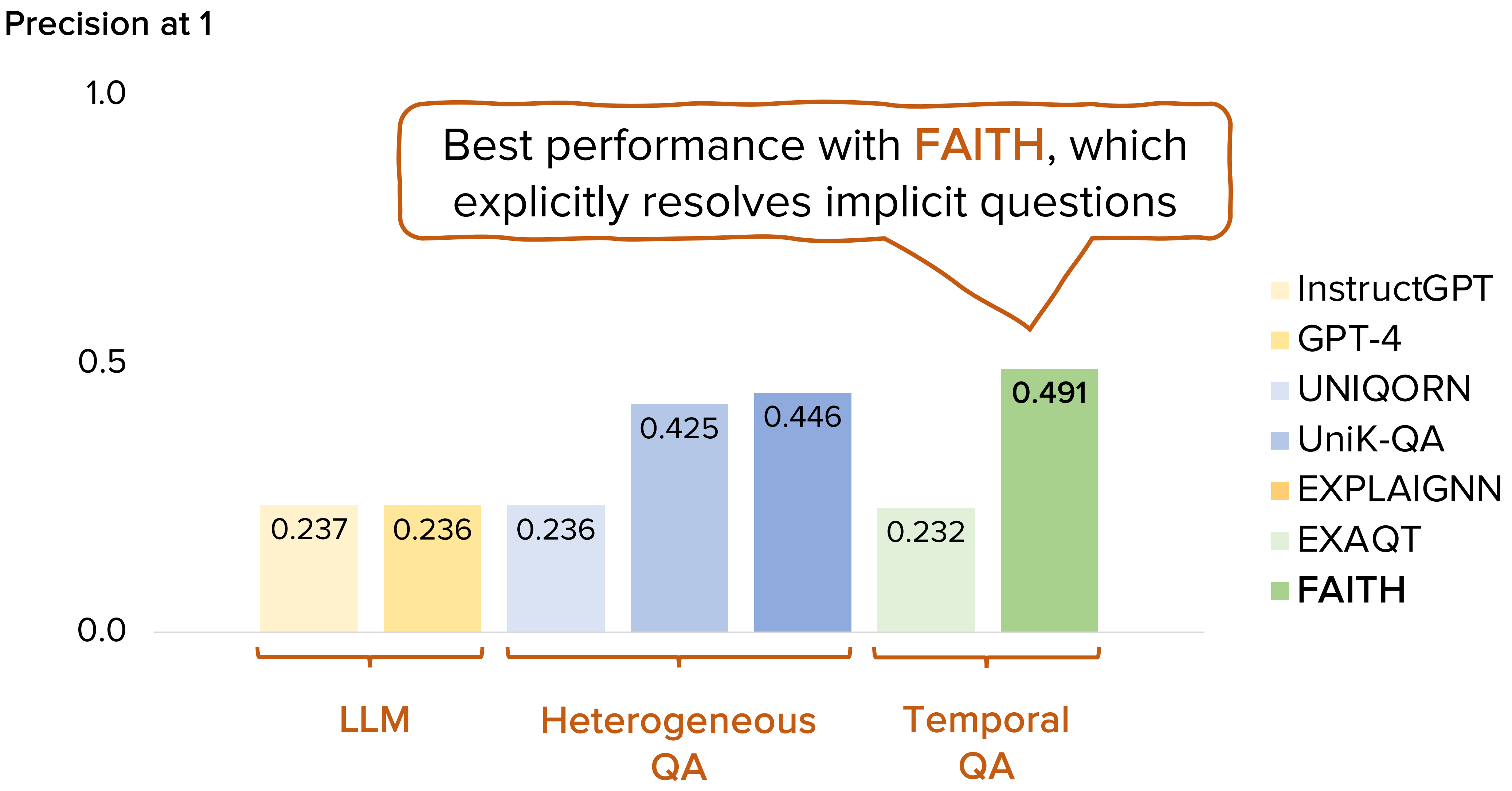
Implicit questions are very challenging for LLMs

Experimental results



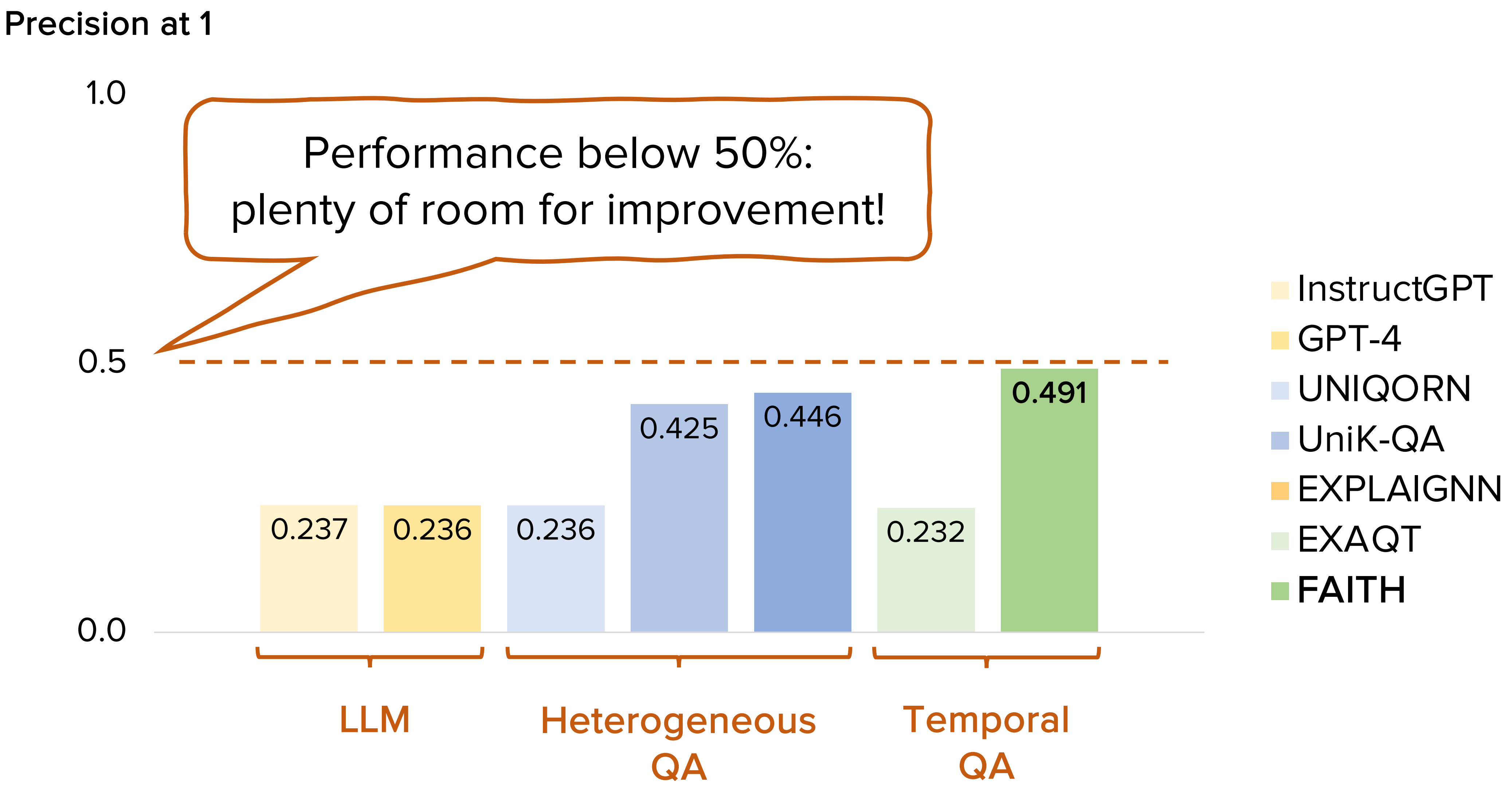
External grounding helps

Experimental results



Best performance with **FAITH**, which explicitly resolves implicit questions

Experimental results



Performance below 50%: plenty of room for improvement!

Common failure cases

★ 66.05%: no **LLM** able to answer

What club did Francisco Gento manage after Palencia?

Temporal constraint ignored

★ 34.80%: no **heterogeneous QA** method able to answer

Which university did Thomas Hunt Morgan attend after receiving his Bachelor of Science degree?

Temporal constraint misunderstood

★ 41.50%: no **temporal QA** method able to answer

What was the first line of the Saint Petersburg Metro to open when it began operations?

Evidence scoring issues

★ 16.65%: no method able to answer

During Marcone's Senior career, what football club did he play for while representing the Qatar Olympic team?

Other complexities
(long-tail entities, “during” and “while” in single question,...)

Outline

★ Motivation

★ Construction of TIQ Benchmark

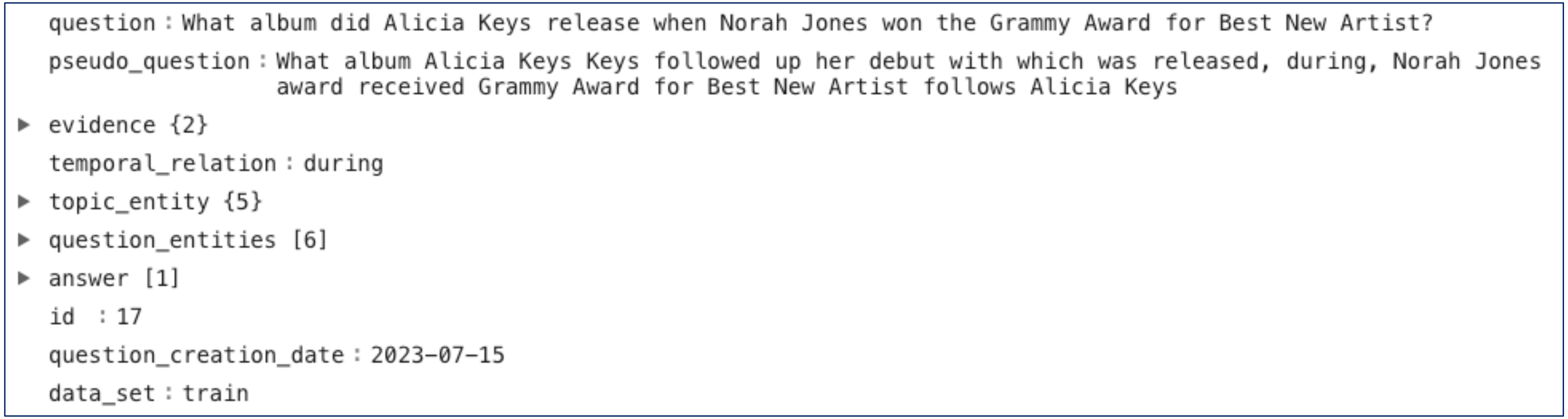
★ Characteristics of TIQ Benchmark

★ Experiments

★ Conclusion

Meta-data

- ★ Question
- ★ Answer (text, Wikidata KB, Wikipedia)
- ★ Information snippets (text, source, temporal values)
- ★ Question derivation (topic entity, temporal relation, pseudo-question)
- ★ Question entities
- ★ Question reference date



```
  question : What album did Alicia Keys release when Norah Jones won the Grammy Award for Best New Artist?
  pseudo_question : What album Alicia Keys Keys followed up her debut with which was released, during, Norah Jones
                    award received Grammy Award for Best New Artist follows Alicia Keys
▶ evidence {2}
  temporal_relation : during
▶ topic_entity {5}
▶ question_entities [6]
▶ answer [1]
  id  : 17
  question_creation_date : 2023-07-15
  data_set : train
```

Conclusion

- ★ Automatic method to construct implicit questions
  - ★ Highly configurable
  - ★ Accesses multiple sources
  - ★ Diverse information needs

- ★ 🕓TIQ benchmark: Temporal Implicit Questions
  - ★ New benchmark with 10,000 implicit temporal questions
  - ★ Derived from Wikipedia text, Wikipedia infoboxes, and Wikidata KB
  - ★ Challenging testbed for temporal QA systems
  - ★ Useful beyond QA?

★ Code and data: *qa.mpi-inf.mpg.de/tiq*

Thank you!