

Step-by-step Claim Verification Using LLMs and Knowledge Graphs

Ivana Hacajová

12.2.2024, Master Thesis Kick Off Presentation

Chair of Software Engineering for Business Information Systems (sebis)
Department of Computer Science
School of Computation, Information and Technology (CIT)
Technical University of Munich (TUM)
www.matthes.in.tum.de

Motivation

Research questions

Methods

- Question guided step-by-step verification
- Predicate logic
- Knowledge graphs
- Various evidence sources

Datasets

Initial findings

Timeline

Need to verify claims

Internet is flooded by often misleading or untrue claims about health, climate change, politics, public affairs, ...

Explainability

Simply stating if a claim is supported or refuted is not enough, some kind of reasoning should be provided, ideally with sources

Domain specificity

The recent claim verifiers using LLMs were not tested on domain specific datasets

RQ1

How can the use of LLMs help claim verification?

RQ2

Does leveraging knowledge from knowledge graphs and structured reasoning improve performance?

RQ3

How do different domains compare in this task?

Question based step-by-step claim verification

QACheck [1]

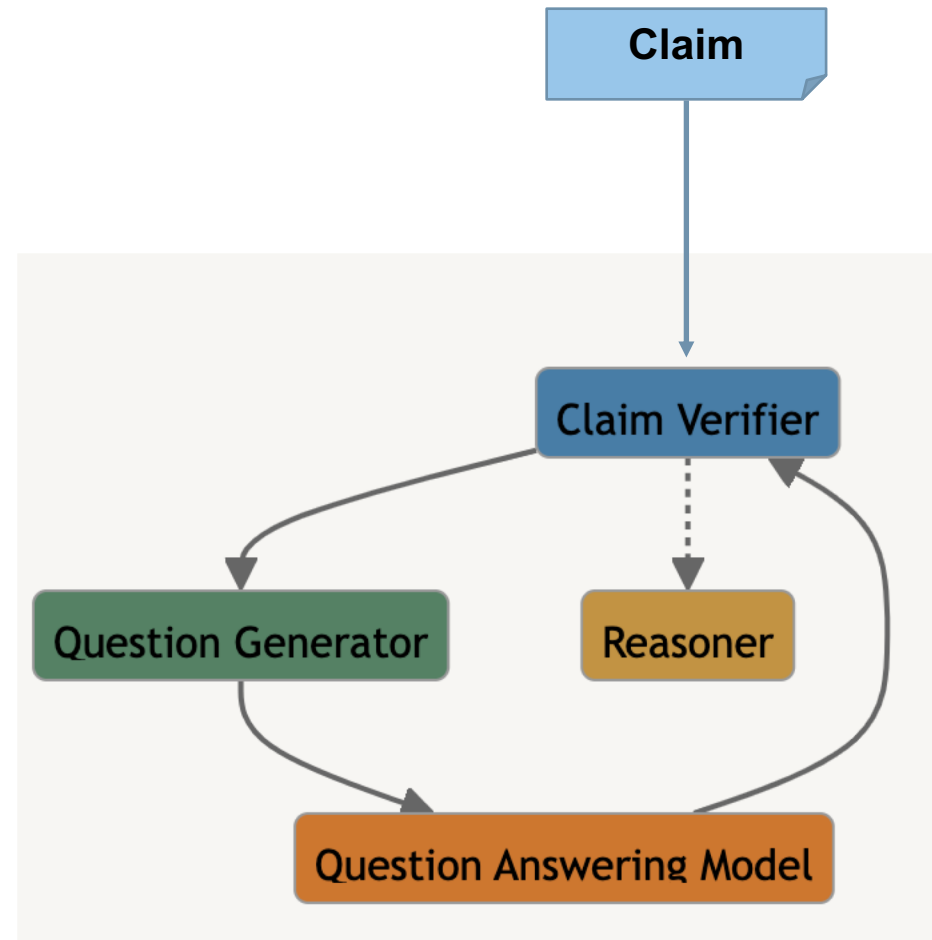
- LLM backbone of the fact checker
- Generates follow up questions based on already gathered evidence (question + answer pairs)

Claim Verifier – Is there enough evidence to say if the claim is refuted or supported?

Question Generator – Based on gathered evidence generates a followup question to verify the claim

Question Answering Model – Searches for information to answer the question and stores the evidence

Reasoner – Given the claim and gathered evidence, decides if the claim is true



[1] Liangming Pan, Xinyuan Lu, Min-Yen Kan, Preslav Nakov.: ["QACHECK: A Demonstration System for Question-Guided Multi-Hop Fact-Checking"](#)

ChatGPT

Retrieves Wikipedia article and the relevant part from its internal „memory“

Google Search

Search the question using Google and use the answer box or first result's snippet as evidence

HuggingChat

Opensource model (mixtral) with access to the internet to search for the answer and providing the sources



Provide knowledge-grounded reasoning

Example

- Backbone: GPT-4
- Evidence source: HuggingChat

Marcus Rashford plays in the city he was born in.

True

Marcus Rashford was born in Wythenshawe, Manchester, England. Marcus Rashford plays for the Premier League club Manchester United and the England national team. Therefore, the final answer is: True.

Where was Marcus Rashford born?

Wythenshawe, Manchester, England.

Marcus Rashford was born in Wythenshawe, Manchester, England. Retrieved from: <https://www.sportskeeda.com/player/marcus-rashford> https://simple.wikipedia.org/wiki/Marcus_Rashford <https://www.thefamouspeople.com/profiles/marcus-rashford-38571.php> <https://www.blackheroesfoundation.org/people/marcus-rashford/>

Where does Marcus Rashford play?

Marcus Rashford plays for Premier League club Manchester United and the England national team.

Marcus Rashford plays as a forward for Premier League club Manchester United and the England national team. Retrieved from: https://en.wikipedia.org/wiki/Marcus_Rashford <https://www.planetsport.com/soccer/marcus-rashford>

Example

- Backbone: GPT-4
- Evidence source: HuggingChat

Nicole Kidman and Tom Cruise starred in a film directed by the director of The Shining

True

The director of The Shining is Stanley Kubrick. Nicole Kidman and Tom Cruise starred in a film directed by Stanley Kubrick. Therefore, the final answer is: True.

Who is the director of The Shining?

Stanley Kubrick

The director of The Shining is Stanley Kubrick. Retrieved from: <https://movieweb.com/stanley-kubrick-actually-had-a-second-choice-for-jack-torrence/> [https://theshining.fandom.com/wiki/The_Shining_\(film\)](https://theshining.fandom.com/wiki/The_Shining_(film))

Did Nicole Kidman and Tom Cruise star in a film directed by Stanley Kubrick?

Yes

Yes, Nicole Kidman and Tom Cruise starred in the Stanley Kubrick film "Eyes Wide Shut," which was released in 1999. Retrieved from: <https://www.sheknows.com/entertainment/articles/2829974/nicole-kidman-tom-cruise-rules-eyes-wide-shut/> https://en.wikipedia.org/wiki/Eyes_Wide_Shut

Predicate logic

FOLK [2]

- LLM generates predicates from the claim
- Has been shown such symbolic representation boosts reasoning capabilities
- Evidence consists of questions, answers and predicates

Example

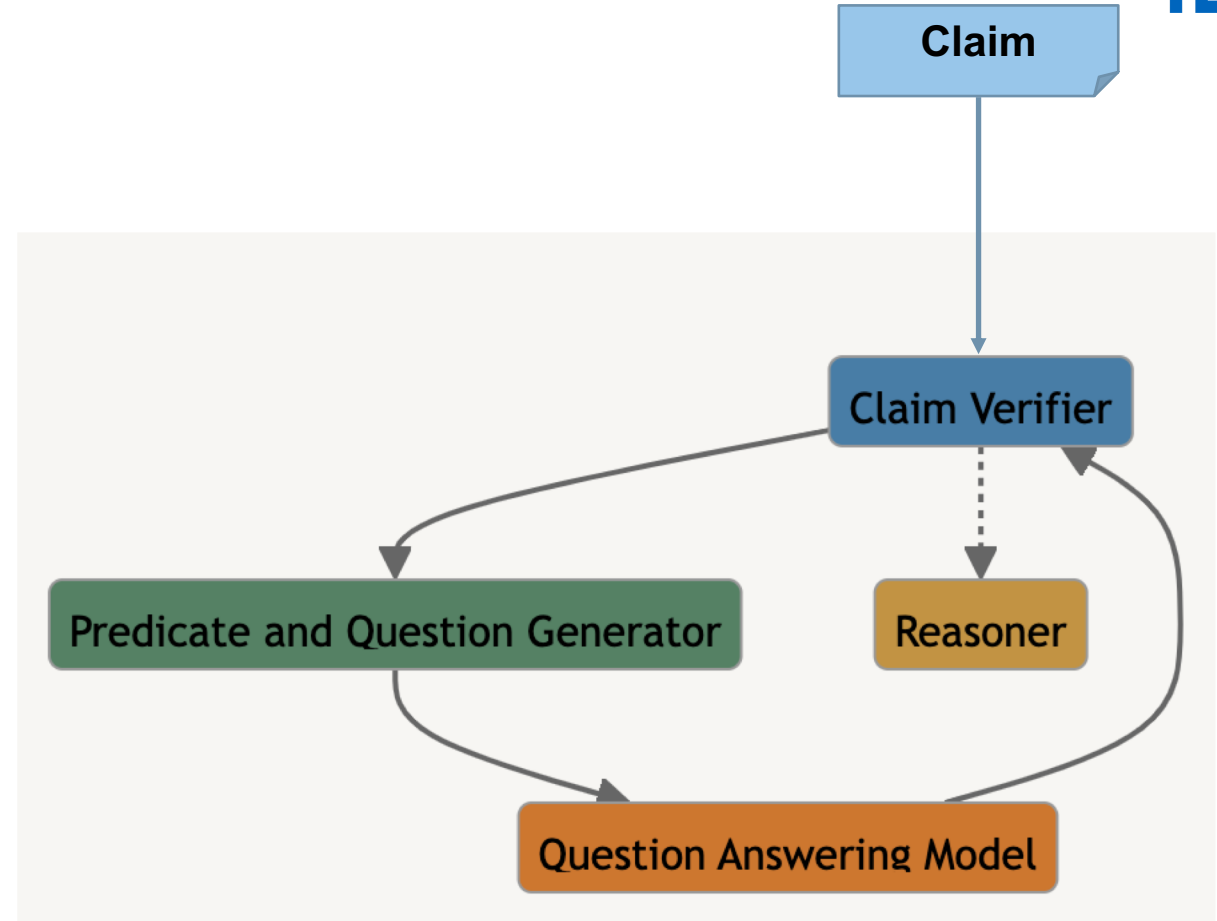
Claim:

TUM and LMU are both located in Munich.

Predicates:

Location(TUM, Munich)

Location(LMU, Munich)



[2] Haoran Wang, Kai Shu: [Explainable Claim Verification via Knowledge-Grounded Reasoning with Large Language Models](#)

Knowledge graphs

KG-GPT [3]

- LLM generates questions and extracts 2 main entities from them
- Retrieves a subgraph based on the entities
- Finds answer in the subgraph using LLM

Example

Claim:

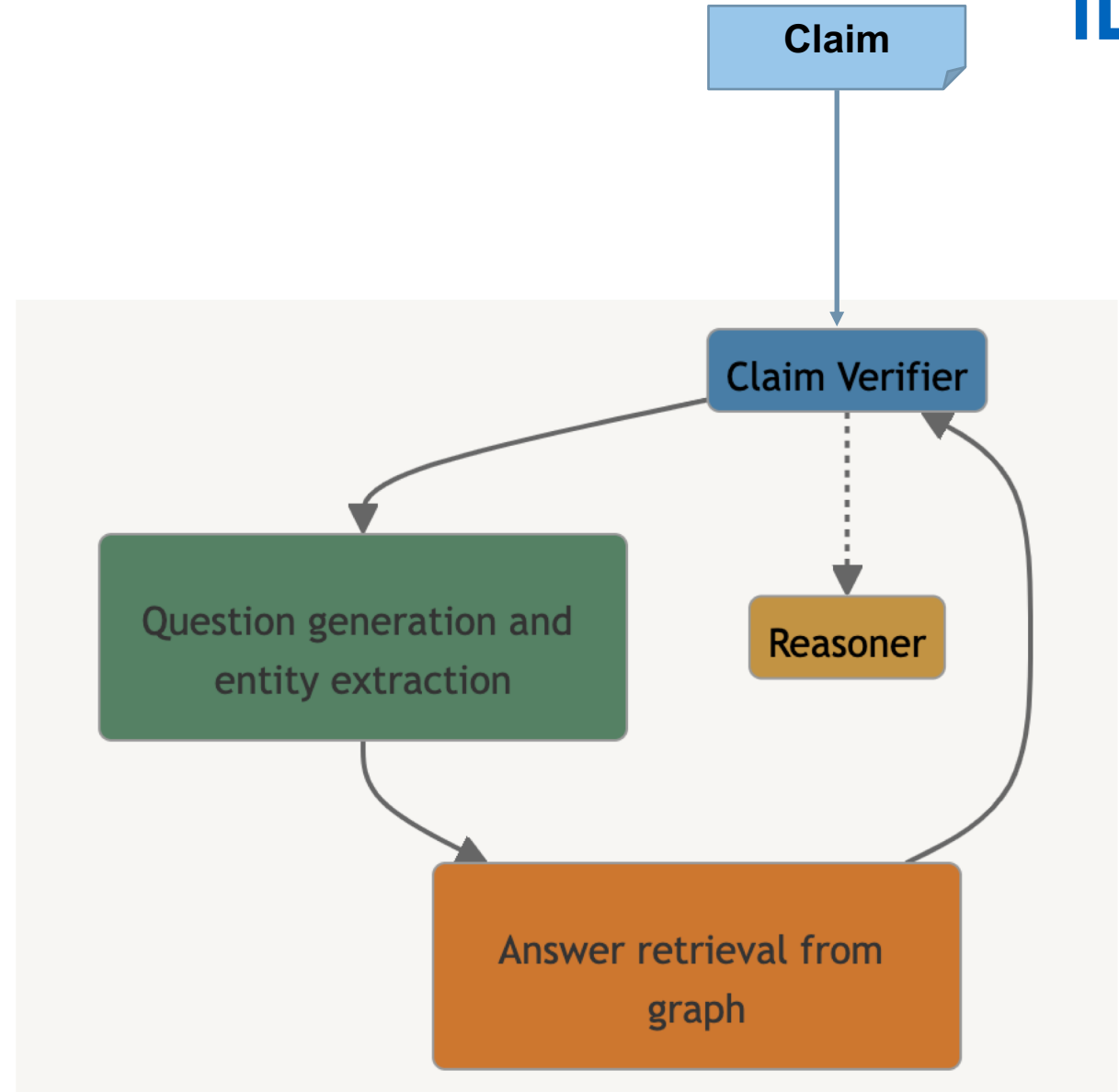
Bruno Fernandes plays for Manchester United.

Extracted entities and relations:

*Bruno Fernandes, Manchester United
[club, spouse, birthPlace]*

Semantically most similar relation:

club



[3] Jiho Kim, Yeonsu Kwon, Yohan Jo, Edward Choi :[KG-GPT: A General Framework for Reasoning on Knowledge Graphs Using Large Language Models](#)

HOVER - Multihop

The writer of for "Can You Feel the Love Tonight" from "The Lion King" and Leslie Marmon Silko are not both lyricists.

Averitec – Popular claims from the internet

In a letter to Steve Jobs, Sean Connery refused to appear in an apple commercial.

Climate-Fever - Climate

Global warming is driving polar bears towards extinction.

HealthFC - Health

Are there any potential risks or side effects associated with inhaling hive air for asthma, allergies, or other conditions?

HuggingChat performance boost

F1 score improved significantly on 2-hop claims from HOVER when using HuggingChat instead of GPT-3.5 for evidence retrieval from **55.67** to **63.03**

Ignoring evidence

In some instances, if evidence is provided, ChatGPT will use its internal knowledge anyway to verify the claim

GPT-4 vs GPT-3.5

GPT-4 is able to create much more simple first and follow up questions, which allows for retrieving higher quality evidence

+ simple web GUI for claim verification with an overview of the reasoning steps

Example of ignoring the provided evidence

GPT-3.5 remembers the president at its time of training (Miloš Zeman), who is different from the current one

Petr Pavel is the president of Czechia

False

The president of Czechia is Petr Pavel. Therefore, the final answer is: False.

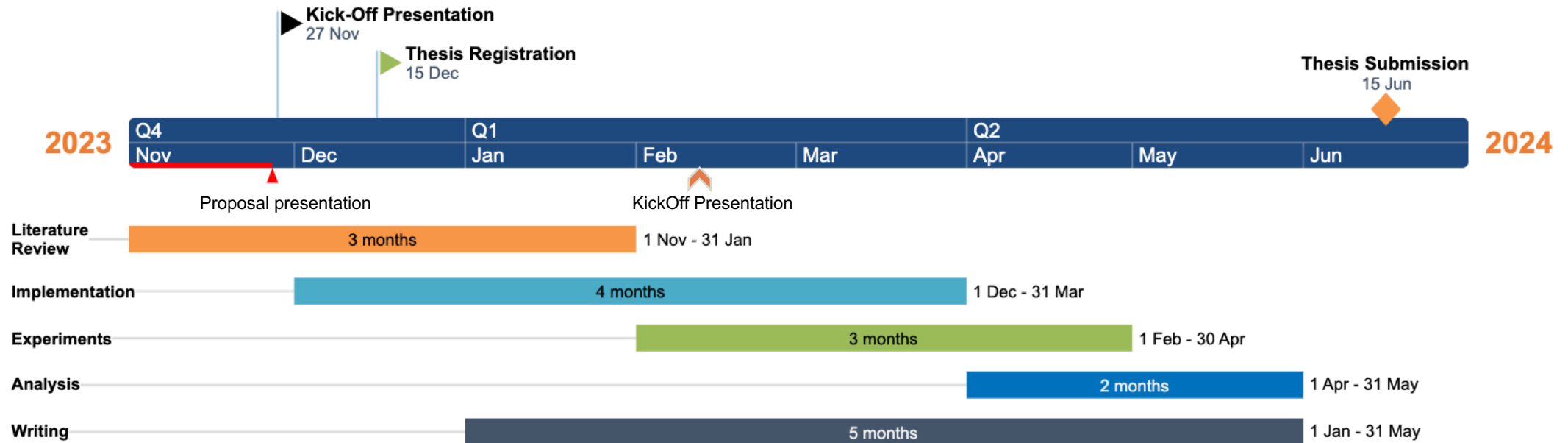
Who is the president of Czechia?

Petr Pavel

The first president of the Czech Republic was Václav Havel. The current president, Petr Pavel, was sworn into office on 9 March 2023 - 2 January 2024 Until 2012 ... Retrieved

from:https://en.wikipedia.org/wiki/List_of_presidents_of_the_Czech_Republic

Timeline





BSc.

Ivana Hacajová

Technical University of Munich (TUM)
TUM School of CIT
Department of Computer Science (CS)
Chair of Software Engineering for Business
Information Systems (sebis)

Boltzmannstraße 3
85748 Garching bei München

+49.89.289.

ivana.hacajova@tum.de
www.matthes.in.tum.de

