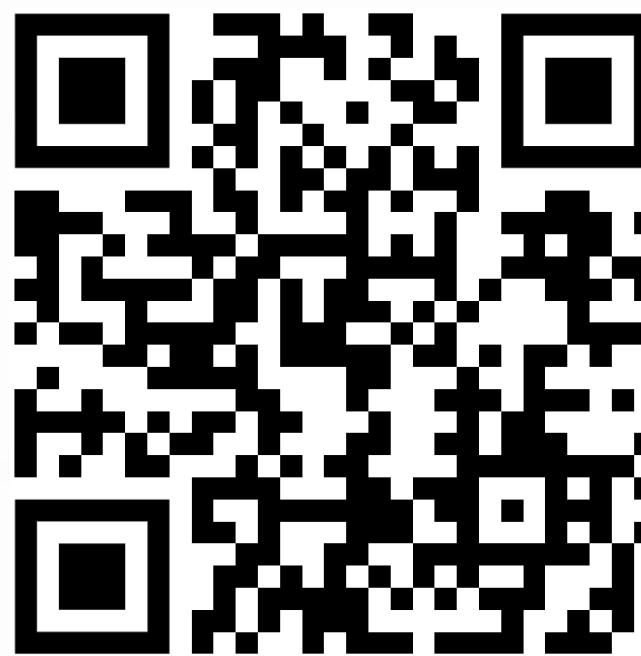


# Computational Fact-Checking of Online Discourse: Scoring Scientific Accuracy in Climate Change Related News Articles

Tim Wittenborg, Constantin S. Tremel, Oliver Karras, Sören Auer



Slides



L3S

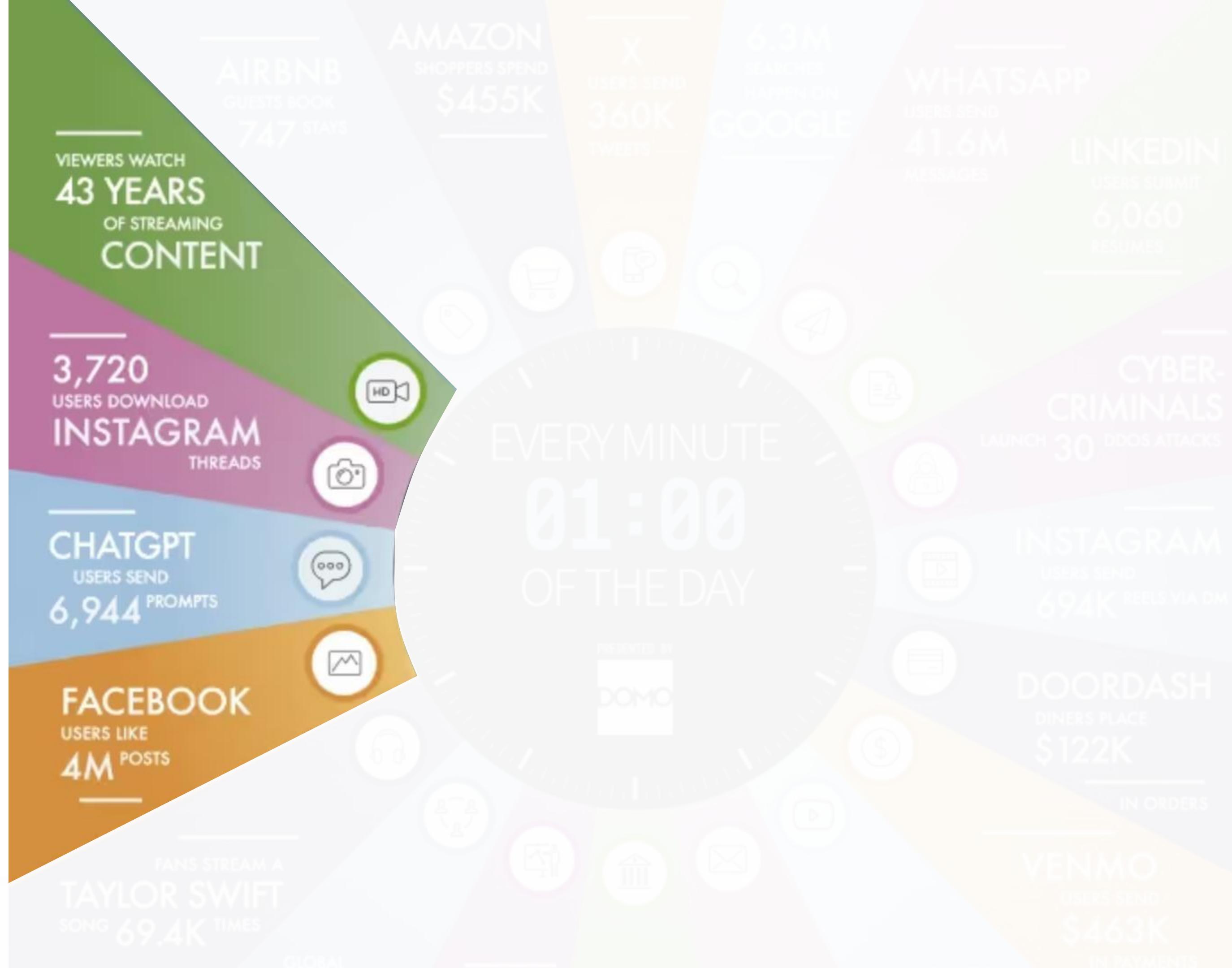


Leibniz  
Universität  
Hannover

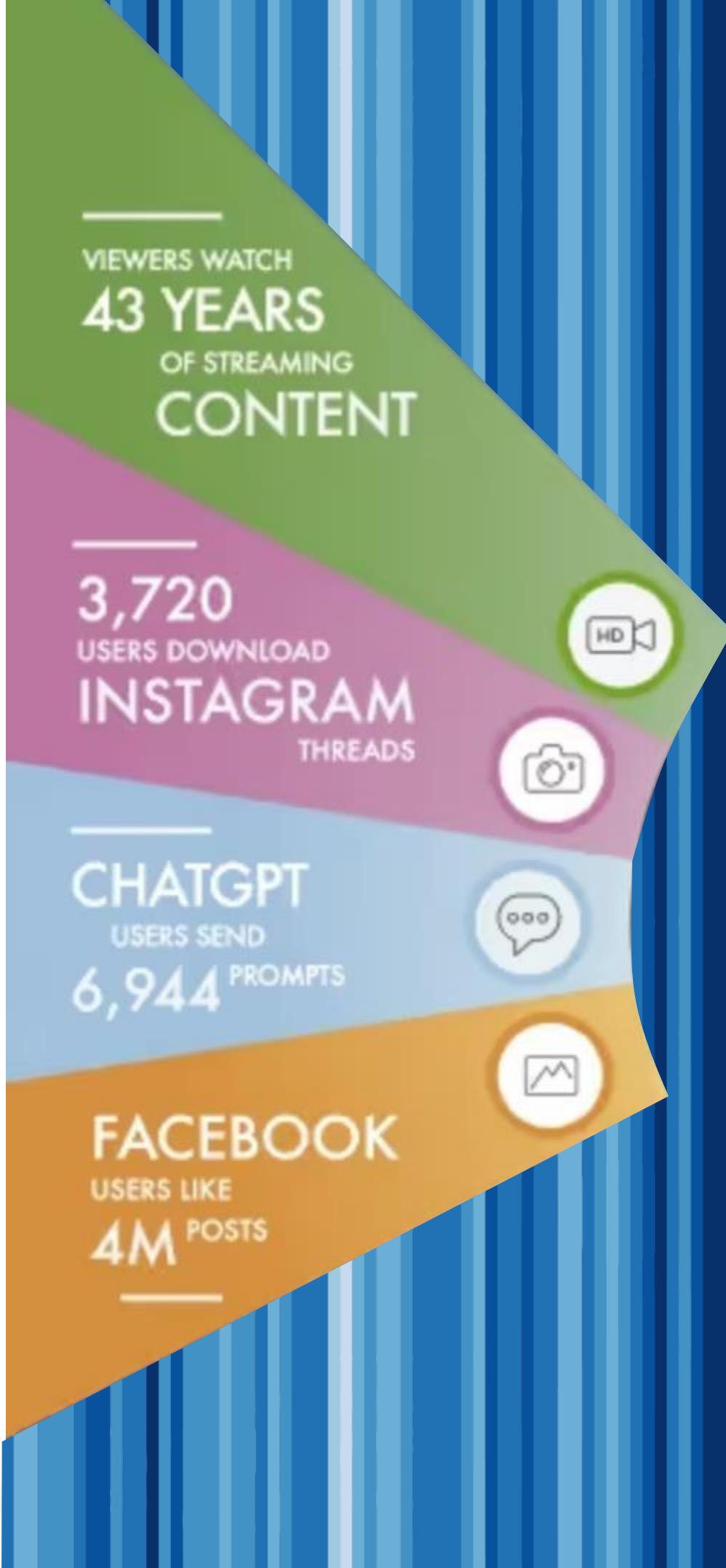




<https://www.domo.com/learn/infographic/data-a-never-sleeps-11>



<https://www.domo.com/learn/infographic/data-a-never-sleeps-11>



Ed Hawkins, climate scientist at University of Reading. CarlinMack created SVG version, [20181204 Warming stripes \(global, WMO, 1850-2018\) - Climate Lab Book \(Ed Hawkins\)](#), cropped, [CC BY-SA 4.0](#)

<https://www.domo.com/learn/infographic/data-never-sleeps-11>

# State of the Art: Manual Fact-Checking in distributed networks

- e.g. International Fact-Checking Network (IFCN):

175 members



**Full Fact**  
from **United Kingdom**  
Verified on 19.1.2024



**Deutsche Welle**  
from **Germany**  
Verified on 30.4.2025



**Science Feedback**  
from **France**  
Verified on 10.5.2024



**Reuters**  
from **United States**  
Verified on 23.4.2025



**FactCheck.org**  
from **United States**  
Verified on 13.3.2024



**The Washington Post Fact Checker**  
from **United States**  
Verified on 23.8.2024



**CORRECTIV**  
from **Germany**  
Verified on 19.1.2024

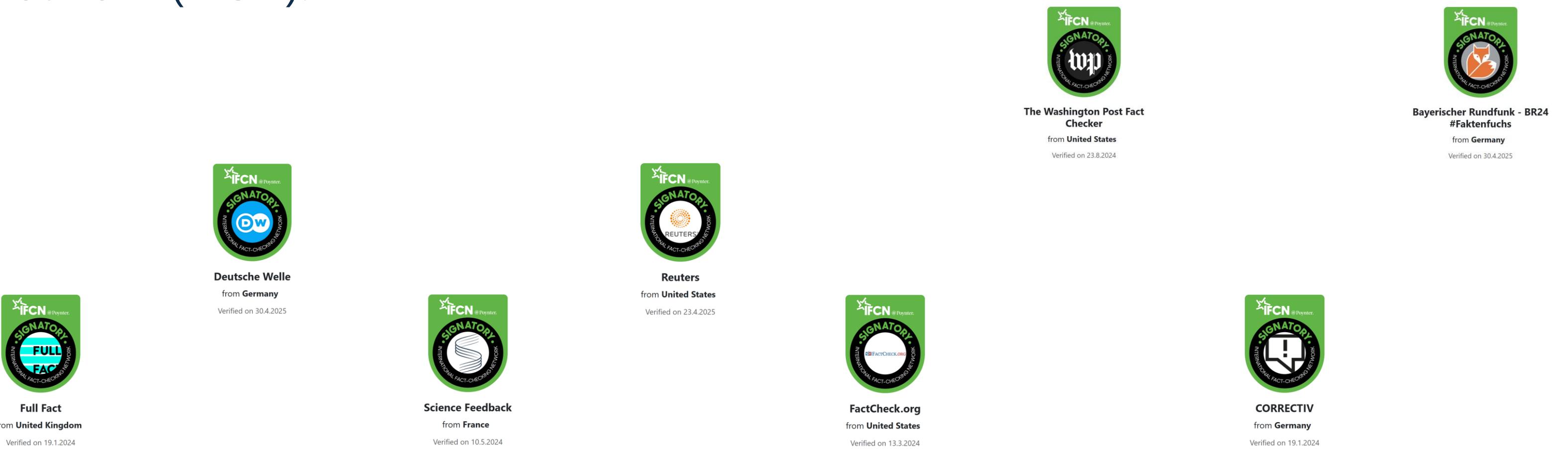


**Bayerischer Rundfunk - BR24**  
**#Faktenfuchs**  
from **Germany**  
Verified on 30.4.2025

# State of the Art: Manual Fact-Checking in distributed networks

- e.g. International Fact-Checking Network (IFCN):

175 isolated members



Ground Truth

# Ideal World: Knowledge Infrastructure curates Ground Truth KG

ipcc

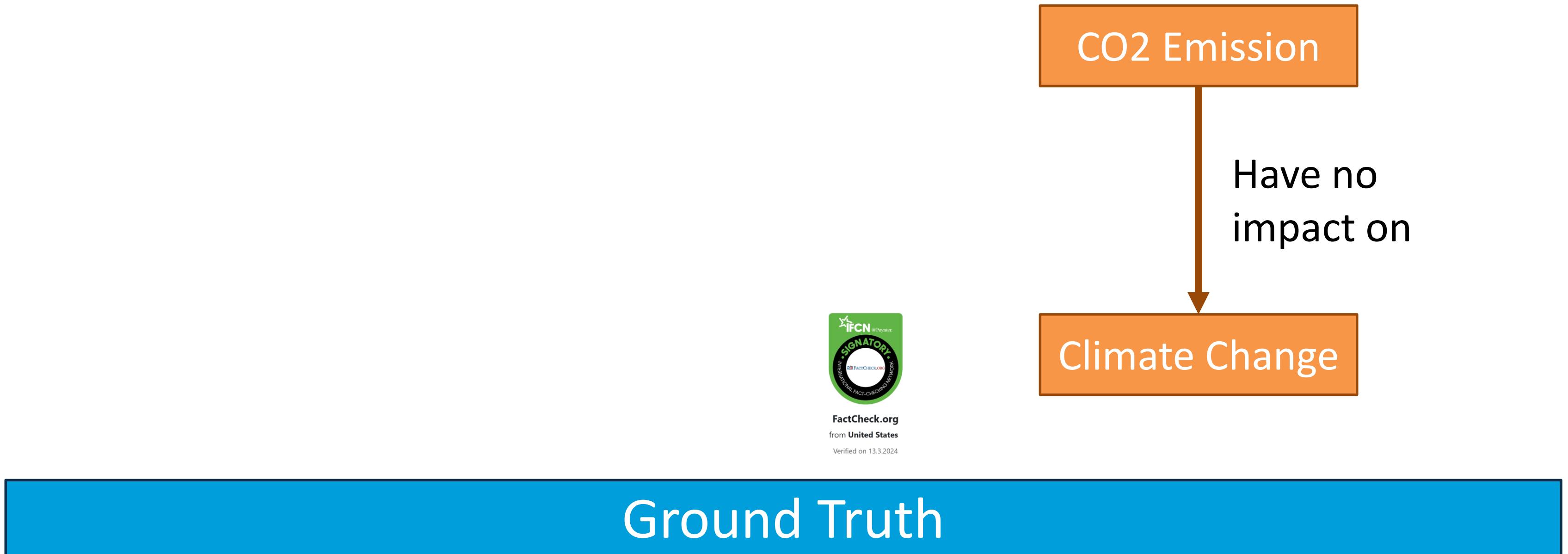
INTERGOVERNMENTAL PANEL ON  
climate change



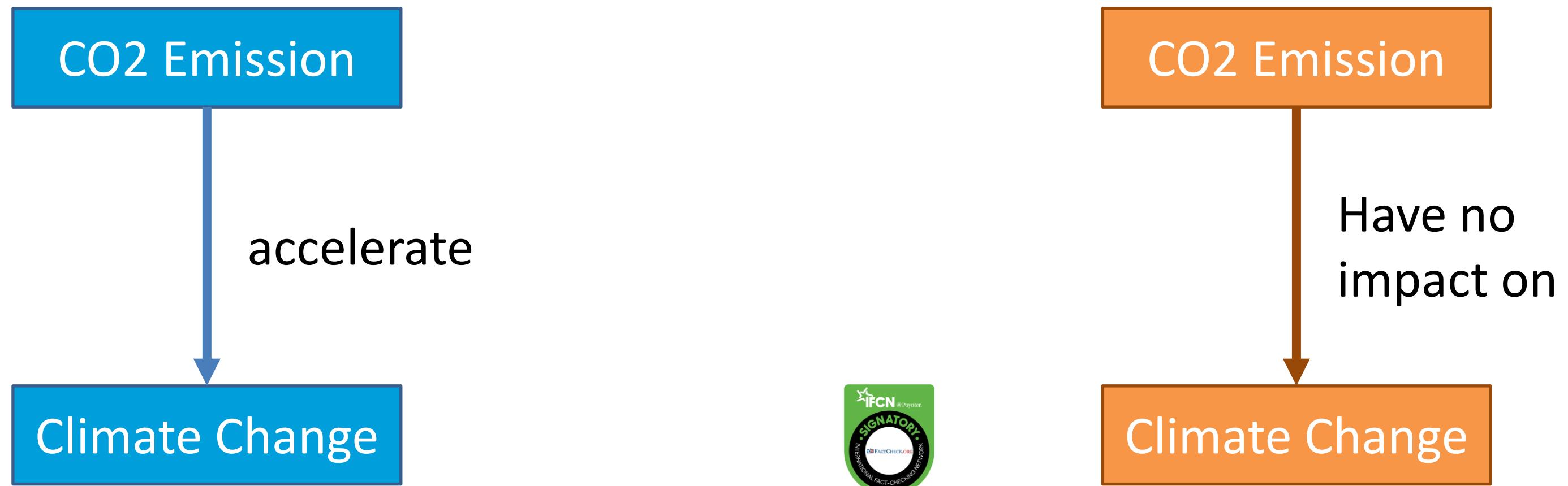
FactCheck.org  
from United States  
Verified on 13.3.2024

Ground Truth

# Ideal World: Knowledge Infrastructure curates Ground Truth KG



# Ideal World: Knowledge Infrastructure curates Ground Truth KG



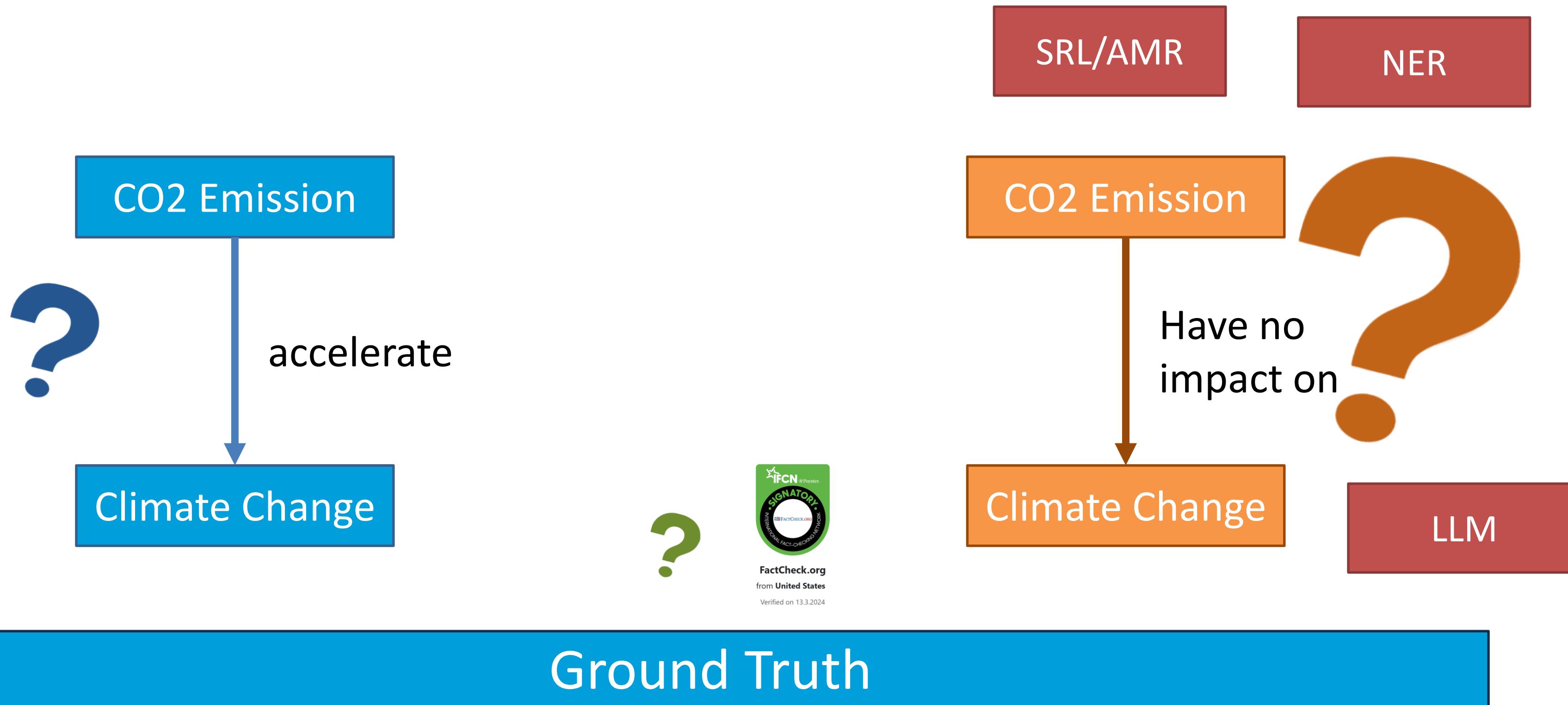
Ground Truth

# Ideal World: Knowledge Infrastructure curates Ground Truth KG

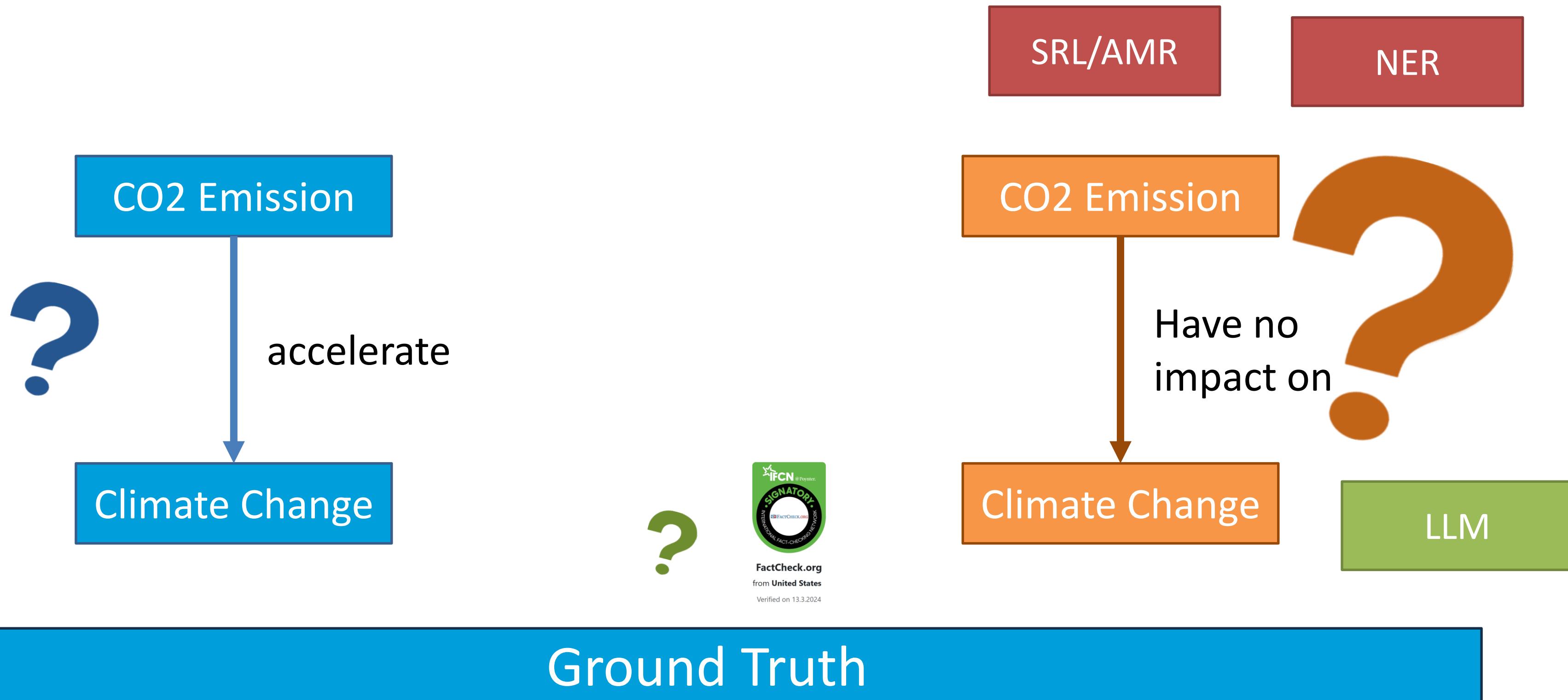


Ground Truth

# Ideal World: Knowledge Infrastructure curates Ground Truth KG



# Ideal World: Knowledge Infrastructure curates Ground Truth KG



# Related Work: Knowledge Graphs + NLP + LLMs

Approach	Data Source	Fact-Checking Method	Domain Focus	Open Source	Limitations
Ciampaglia et al. (2015)	DBpedia	Semantic proximity in knowledge graphs (path length, degree of nodes)	General	Partially	Limited contextual understanding; scalability concerns
Thorne et al. (2018) FEVER	Wikipedia	Textual entailment using NLP models; human annotations	General	Yes	Requires structured textual evidence; limited to Wikipedia domain
Leippold et al. (2025)	Scientific climate texts	LLM-based claim verification (blackbox)	Climate Change	No	persistent knowledge preservation and process transparency issues
Dess`i et al. (2022) SCICERO	Computer science literature	Deep learning + NER for research knowledge graph generation	Scientific Knowledge	Yes	Domain-specific; not focused on media claim validation
This Work (Tremel, 2024)	Climate reports (e.g., IPCC), online media	LLM-based triple extraction + symbolic verification in KG	Climate Change	Yes	Limited ground truth; scalability and context handling remain challenges

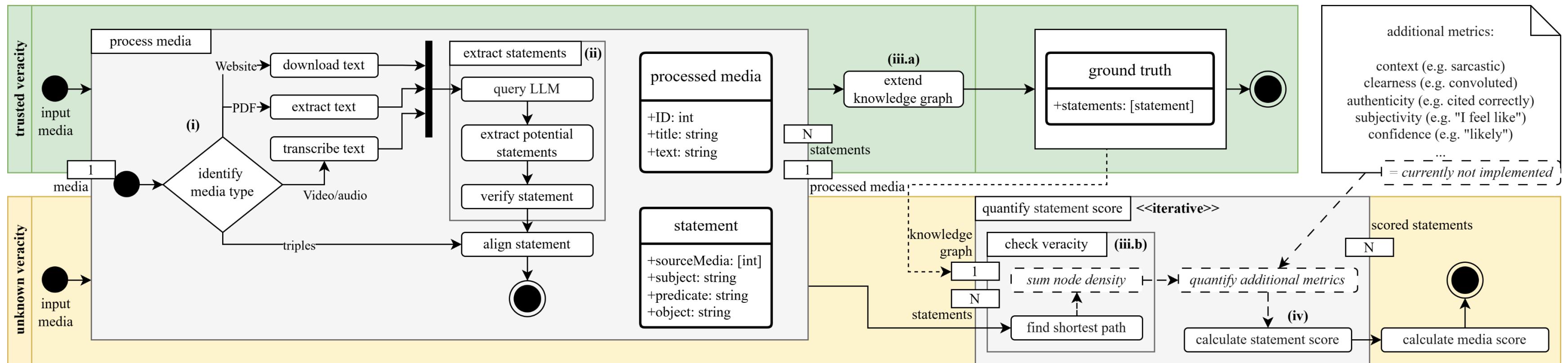


How can **natural language processing**  
and **knowledge graphs** help quantify the  
scientific accuracy of secondary literature  
in the context of climate change?

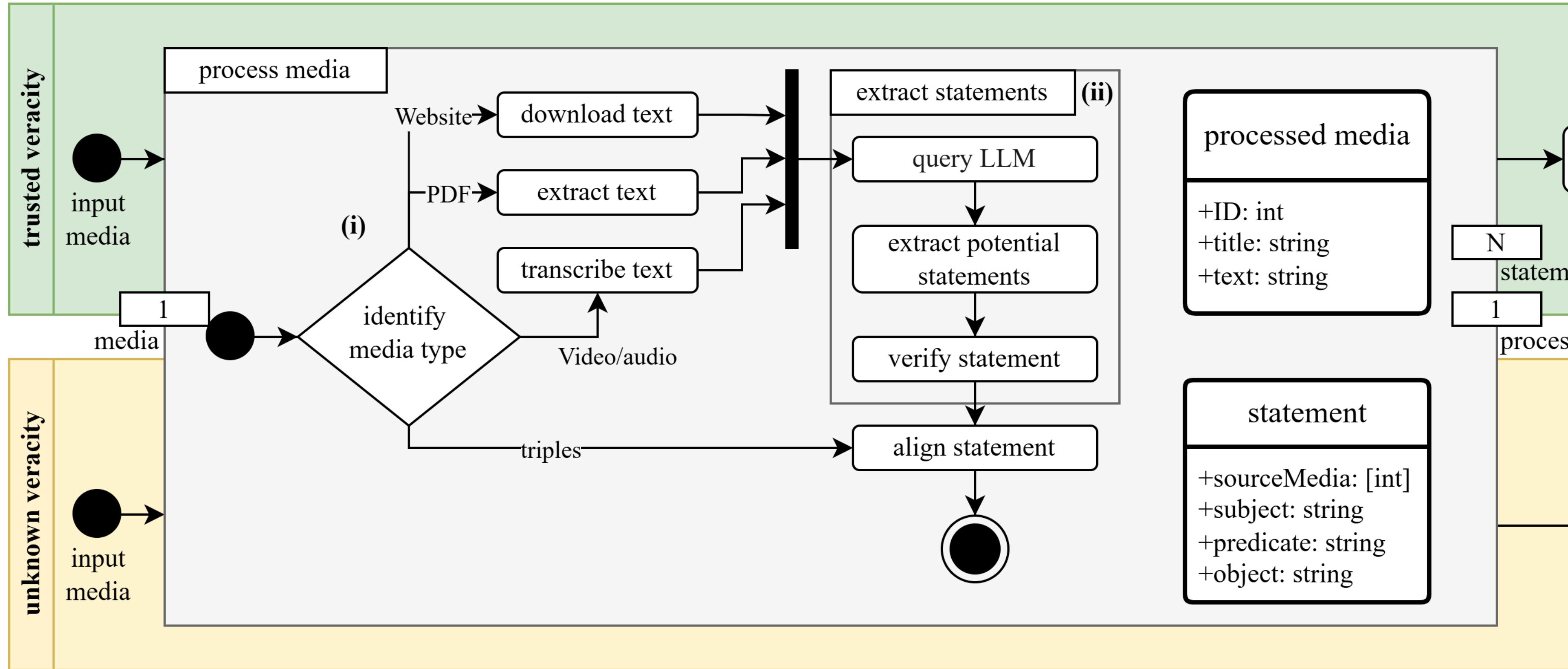
# Our contributions:

1. An **overview** of existing and required **ground truth knowledge graphs** for the climate change domain, alongside current **fact-checking support**, identifying the potential and limits of **computational fact-checking**.
2. A **streamlined workflow** from ground-truth **semantification** to equally processing popular media and **scoring scientific accuracy**, available as a modular open-source implementation.
3. Two **evaluations** of this tool from both **experts** and **users**, substantiating the necessity and usability, as well as an indication of required future work towards sustainably scaling fact-checking amid the information flood.

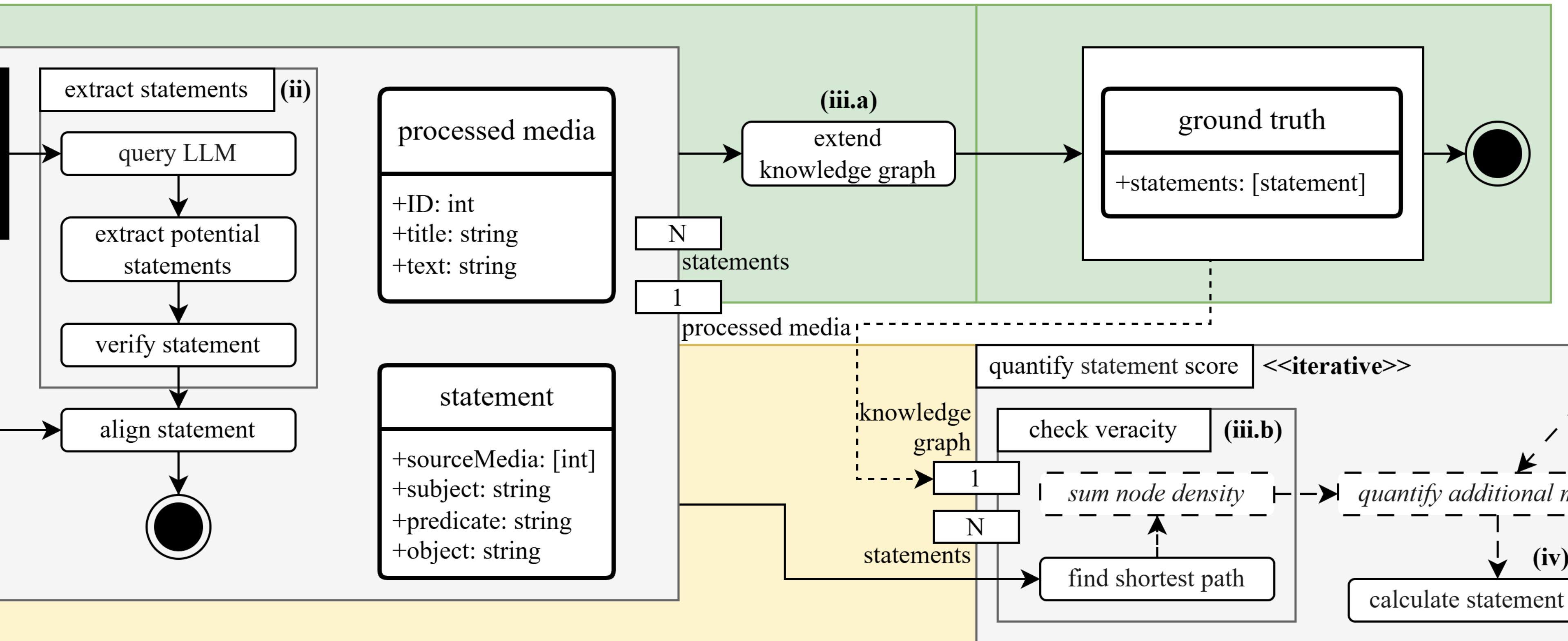
# Our workflow



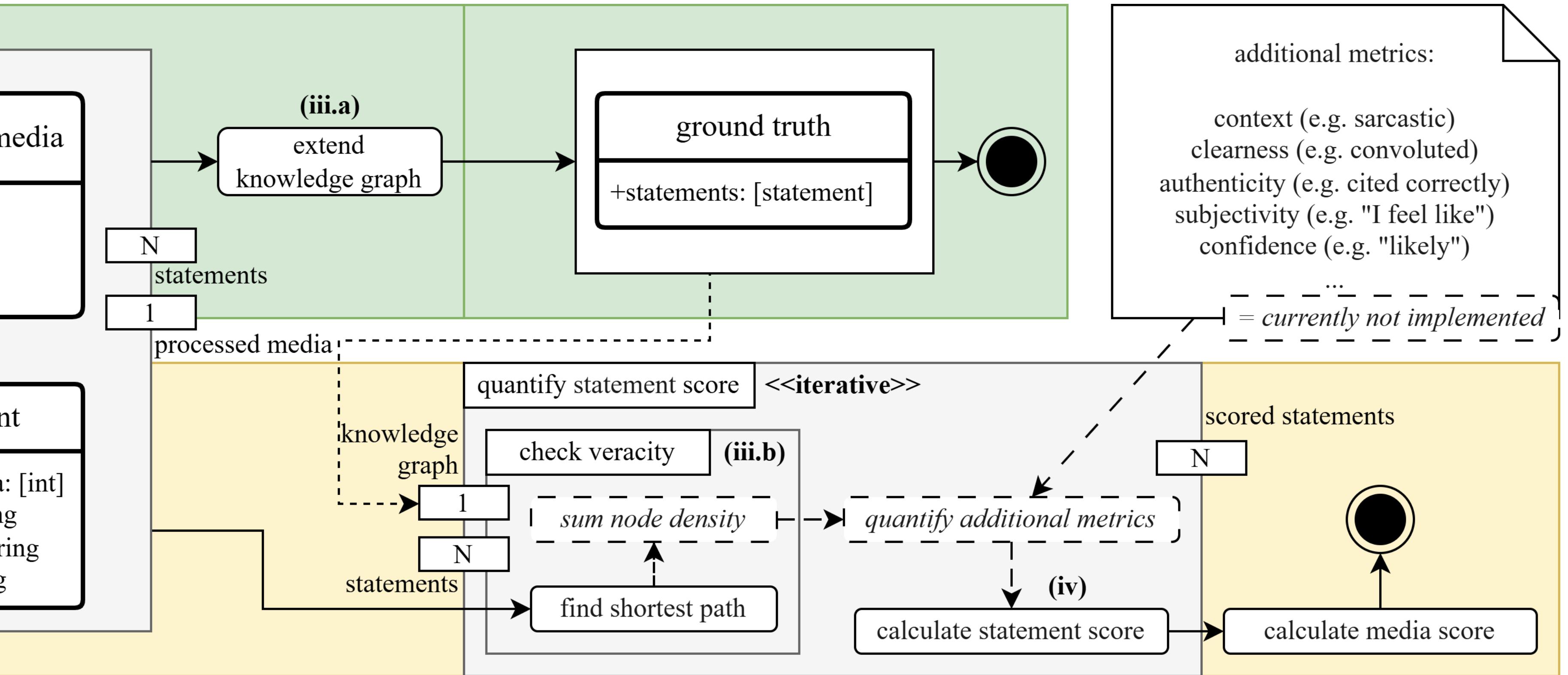
# Our workflow



# Our workflow



# Our workflow



# Example

## article text body

The effects of human-caused global warming are happening now, are irreversible for people alive today, and will worsen as long as humans add greenhouse gases to the atmosphere.[En español](#)[Takeaways](#)We already see effects scientists predicted, such as the loss of sea ice, melting glaciers and ice sheets ...

# Example

## article text body

The effects of human-caused global warming are happening now, are irreversible for people alive today, and will worsen as long as humans add greenhouse gases to the atmosphere. En español Takeaways We already see effects scientists predicted, such as the loss of sea ice, melting glaciers and ice sheets ...



## extraction prompt

## candidate triples

1. [Climate Change, encompasses, Global warming]
2. [Global warming, refers to, long-term warming]
3. [Humans, caused, major climate changes]
4. [Humans, emitting, greenhouse gases]
5. [Greenhouse gases, contribute to, Global warming]
6. ...

# Example

## candidate triples

1. [Climate Change, encompasses, Global warming]
2. [Global warming, refers to, long-term warming]
3. [Humans, caused, major climate changes]
4. [Humans, emitting, greenhouse gases]
5. [Greenhouse gases, contribute to, Global warming]
6. ...

## base form prompt

## intermediate triples

1. [Climate Change, ~~encompasses~~, Global warming]  
[Climate Change, encompass, Global warming]
2. [Global warming, refers to, long-term warming]
3. [Humans, caused, major climate changes]
4. [Humans, emitting, greenhouse gases]
5. ...

# Example

## intermediate triples

1. [Climate Change, ~~eneompasses~~, Global warming]  
[Climate Change, encompass, Global warming]
2. [Global warming, refers to, long-term warming]
3. [Humans, caused, major climate changes]
4. [Humans, emitting, greenhouse gases]
5. ...

## synonym prompt

## intermediate triples

1. [~~Climate Change~~, encompass, ~~Global warming~~]  
[climateChange, encompass, globalWarming]
2. [globalWarming, referTo, longTermWarming]
3. [humanActivity, cause, climateChange]
4. [humanActivity, emit, greenhouseGas]
5. ...

# Example

## intermediate triples

1. [~~Climate Change~~, encompass, ~~Global warming~~]  
[climateChange, encompass, globalWarming]
2. [globalWarming, referTo, longTermWarming]
3. [humanActivity, cause, climateChange]
4. [humanActivity, emit, greenhouseGas]
5. ...

## *ontology mapping*

## final triples

1. [**ex:climateChange**, **ex:encompass**, **ex:globalWarming**],
2. ['**ex:globalWarming**', '**ex:referTo**', '**ex:longTermWarming**']
3. ['**ex:humanActivity**', '**ex:cause**', '**ex:climateChange**']
4. ['**ex:humanActivity**', '**ex:emit**', '**ex:greenhouseGas**']
5. ['**ex:globalWarming**', '**ex:cause**', '**ex:longTermWarming**']
6. ...

# UI Mock-Up

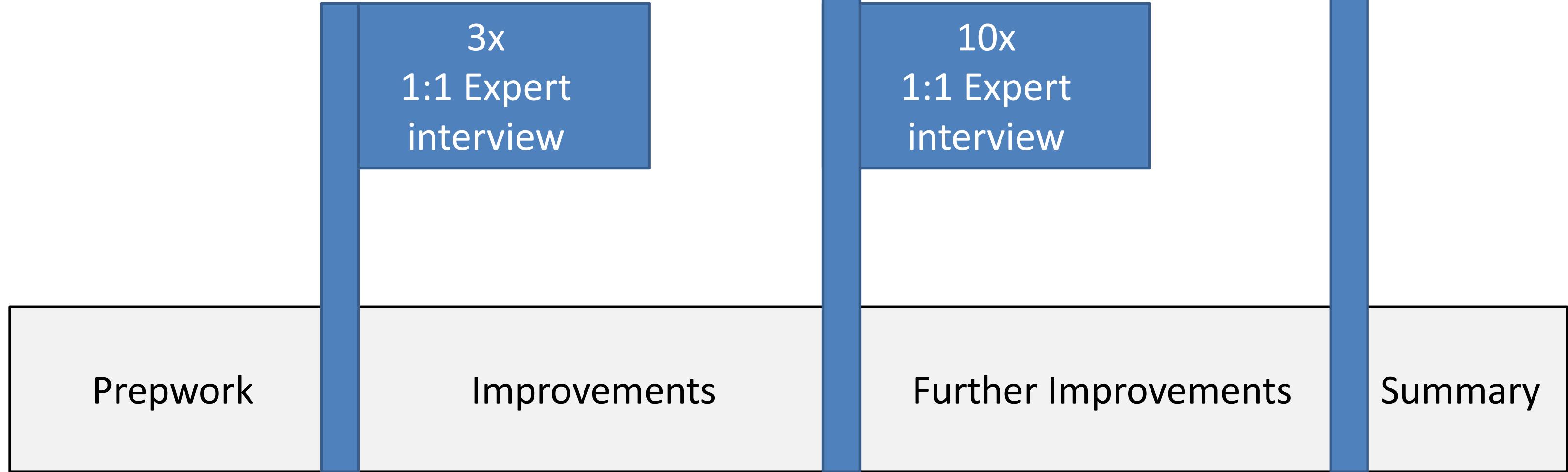
**"The scientific evidence is unequivocal:  
climate change is a threat to human  
wellbeing and the health of the planet. Any  
further delay in concrete  
miss the brief, rapidly closing  
window of opportunity to secure a liveable future."**

The scientific evidence is unequivocal: climate change is a threat to human wellbeing and the health of the planet.  
*"Climate change is a threat to human well-being and planetary health (very high confidence). There is a rapidly closing window of opportunity to secure a liveable and [read more](#) [go to source](#)*



# Evaluation

# Iterative Evaluation



# Expert presentation (N=27)

regarding agreement

- with our consolidated findings
- their prioritization of future challenges

Concluding statement	# Agree
The optimal method for extracting triples is currently unclear.	9
LLMs have limitations and should not be used without proper checks to identify non-reproducible or hallucinated triples.	16
Scientific accuracy checks must take into account the context of statements, which cannot be represented by a single triple.	13
This tool is more likely to be used as an integrated rather than a stand-alone tool.	4
As long as "a perfect algorithm to check against the truth" is not achievable: Having an indication of what is more or less likely to be accurate is already helpful.	8
Challenge whose solution brings the greatest benefit	# Votes
Handling the semantic alignment of natural language.	6
Handling LLM hallucinations and lacking reproducibility.	9
Fully automated triple extraction.	0
Keeping the context of statements, especially in empirical research.	11
Making statements about statements to show their confidence or time validity.	1

# Expert presentation (N=27)

regarding agreement

- with our consolidated findings
- their prioritization of future challenges

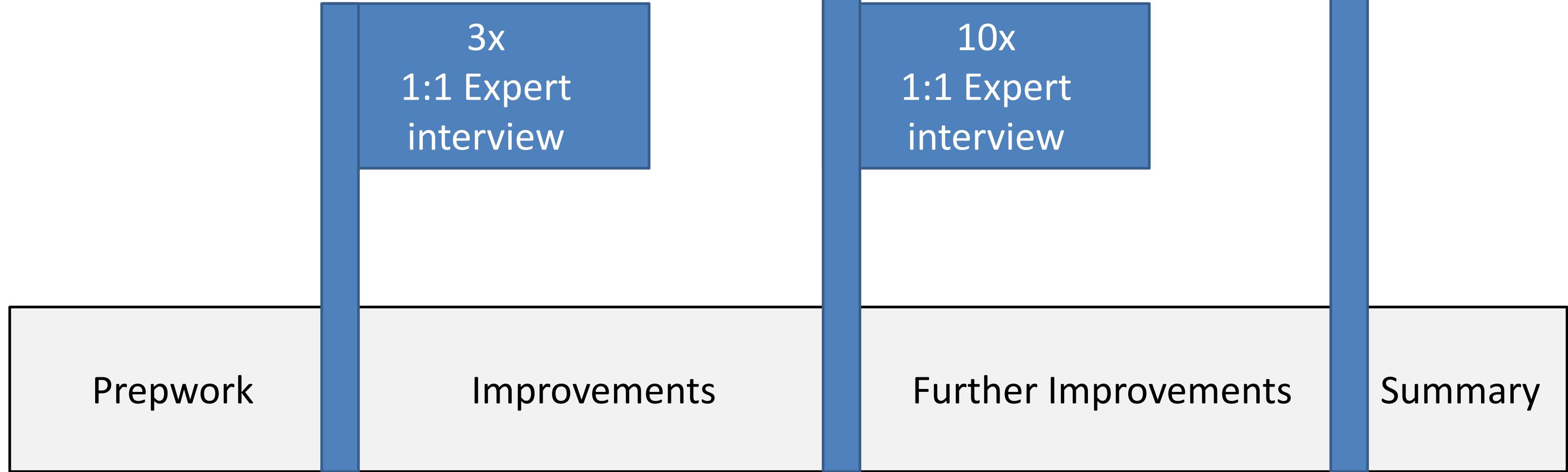
Concluding statement	# Agree
The optimal method for extracting triples is currently unclear.	9
LLMs have limitations and should not be used without proper checks to identify non-reproducible or hallucinated triples.	16
Scientific accuracy checks must take into account the context of statements, which cannot be represented by a single triple.	13
This tool is more likely to be used as an integrated rather than a stand-alone tool.	4
As long as "a perfect algorithm to check against the truth" is not achievable: Having an indication of what is more or less likely to be accurate is already helpful.	8
Challenge whose solution brings the greatest benefit	# Votes
Handling the semantic alignment of natural language.	6
Handling LLM hallucinations and lacking reproducibility.	9
Fully automated triple extraction.	0
Keeping the context of statements, especially in empirical research.	11
Making statements about statements to show their confidence or time validity.	1

# Expert interview (N=10)

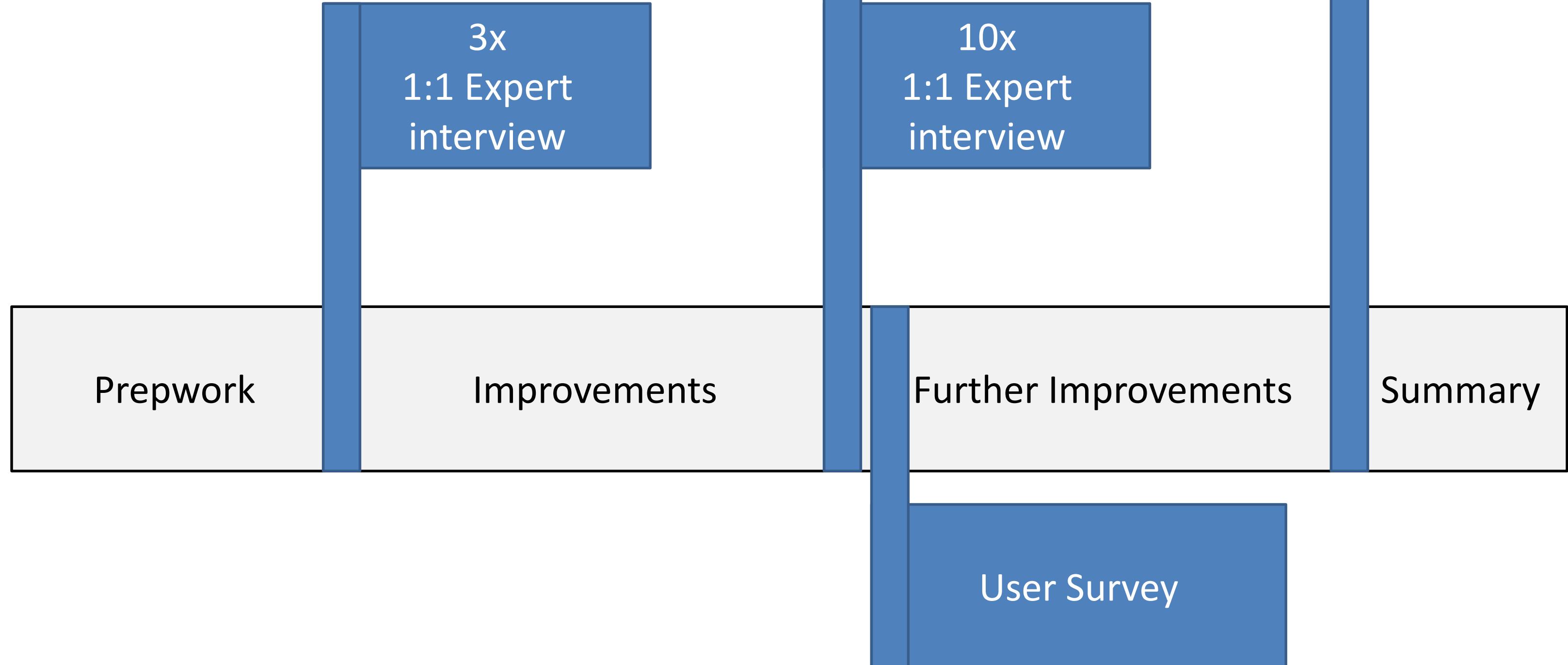
for detailed feedback

ID	Statement
1	Using the tool as a personal assistant can be helpful. It could communicate with a browser, a PDF reader, and other sources of information through a trusted graph.
2	The use of LLMs and assembling information in knowledge graphs is often considered an up-to-date approach.
3	When focusing on supporting claims, incoming data may be more relevant. Incoming nodes could be given more weight than outgoing nodes, similar to how SEO rankings work on websites.
4	Graph walks can be utilized to match claims by expanding the connections of the trusted knowledge graph and counting the number of hops required to reach a similar triple to the one being checked.
5	A common approach uses embeddings to compare the distances between different concepts.
6	The bottleneck lies in the triple extraction phase. Semantic parsing presents a challenge due to the syntax, particularly when using RDF.
7	There is a consensus that LLMs can be useful in specific domains. However, their effectiveness is debatable and depends on the requirements. Using them for independent tasks fast-tracks development. Creating interfaces early on opens the possibility for more efficient and or reliable solutions for the subtask.
8	A combination of NER and LLMs has potential. A mediator is necessary to ensure consistency in moderating statements or types of statements, particularly in cases involving trusted sources and popular media that require verification.
9	LLMs have limitations and should not be used without proper checks to identify non-reproducible or hallucinated triples.

# Iterative Evaluation

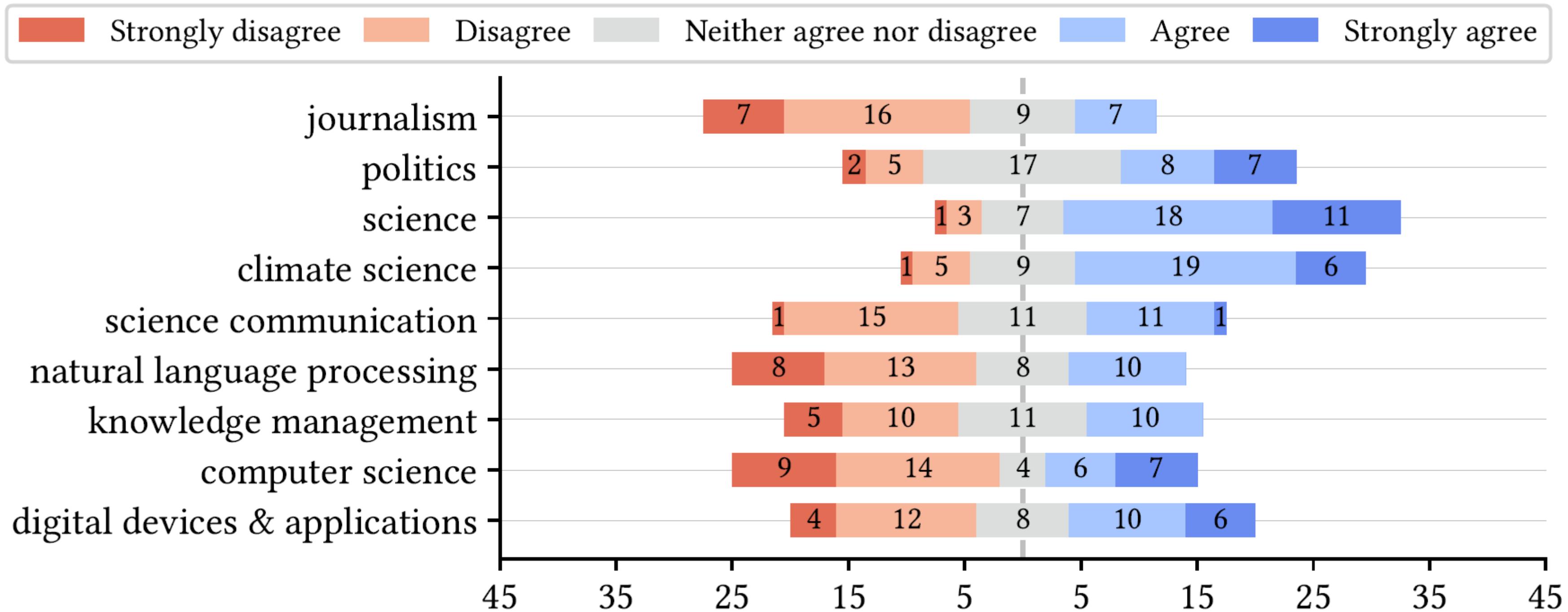


# Iterative Evaluation



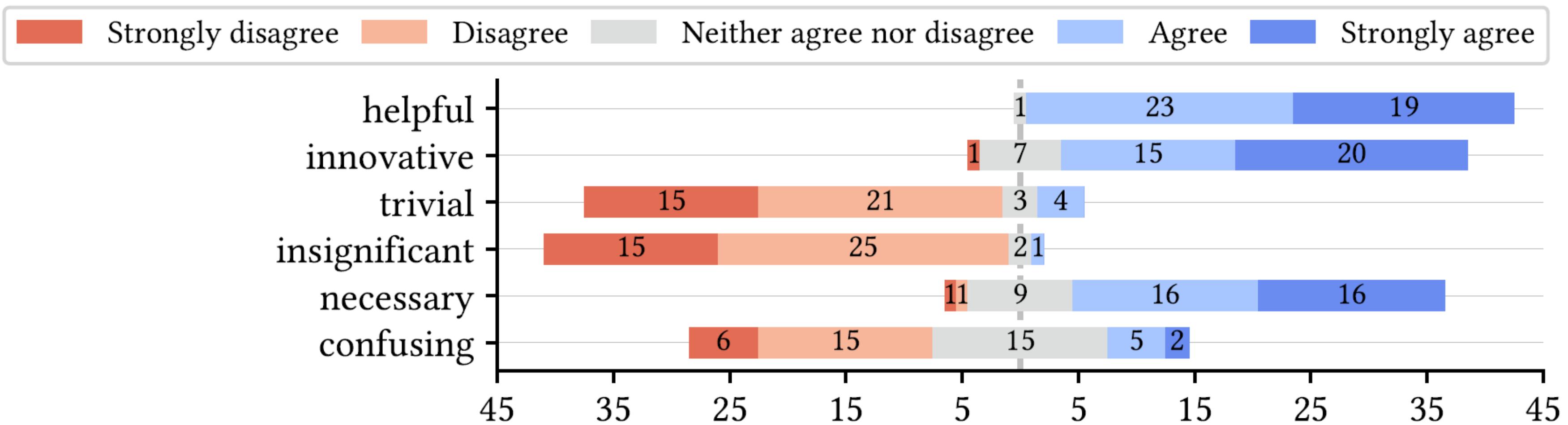
# User survey (N=43)

I am experienced in the following area ...



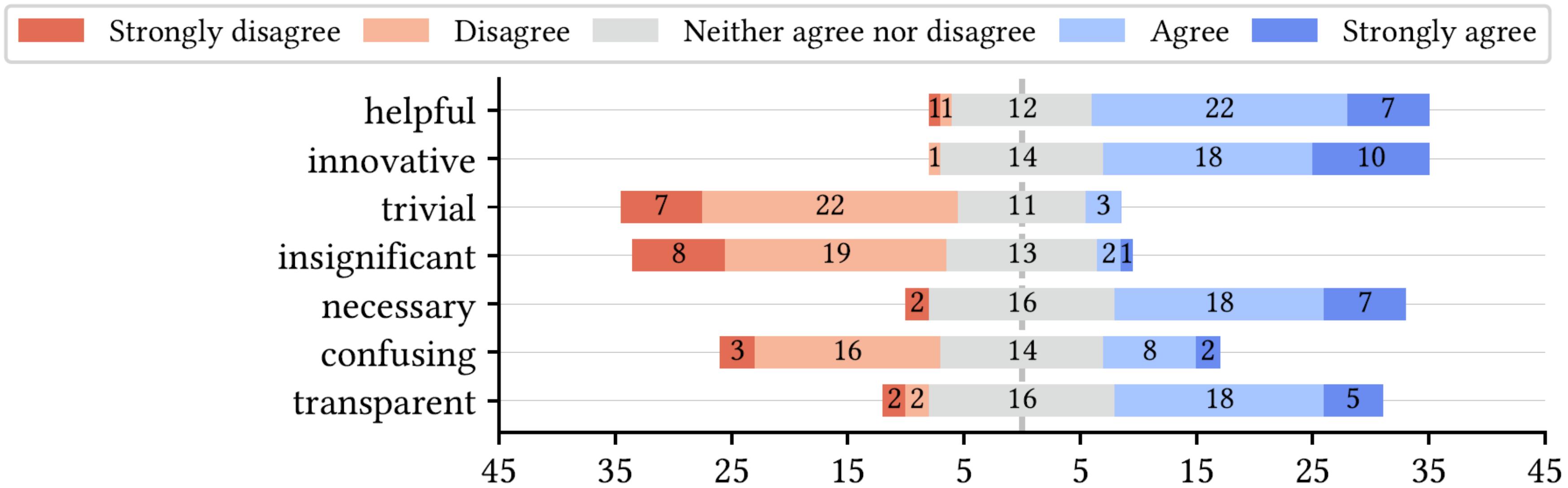
# User survey (N=43)

I think a scientific accuracy score is ...



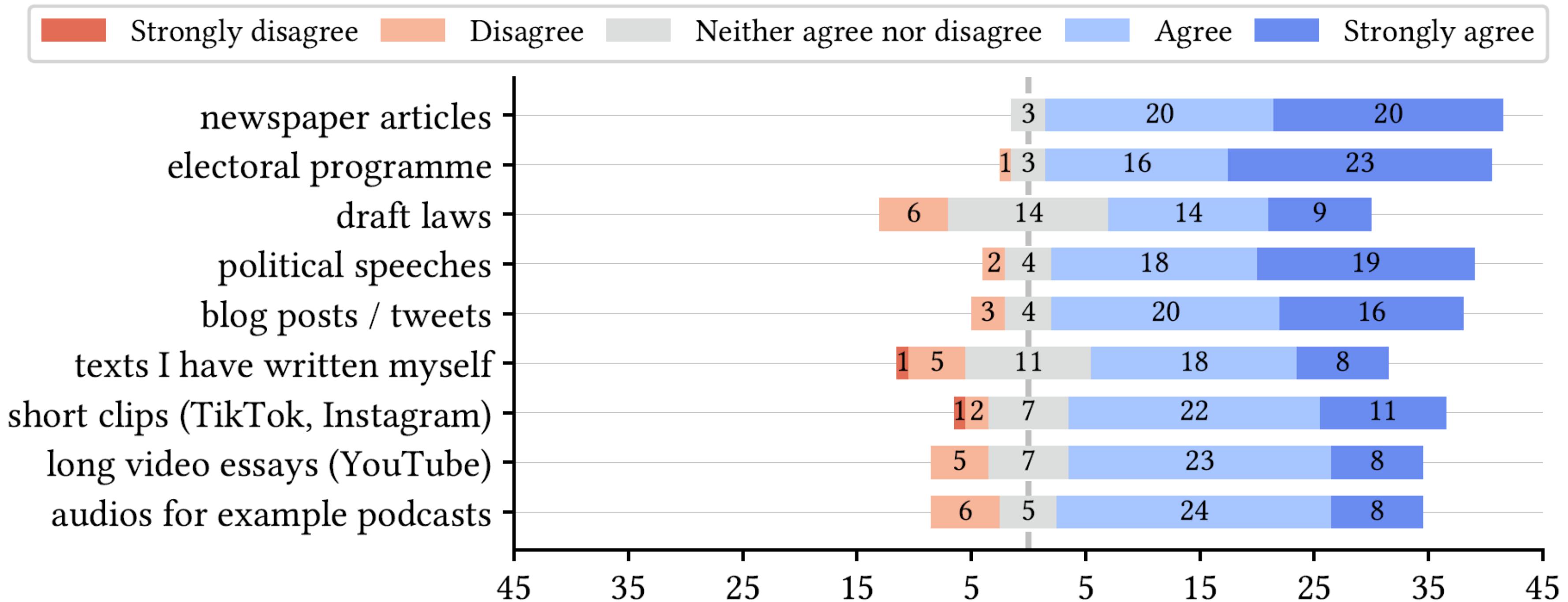
# User survey (N=43)

I think the tool in the current state is ...



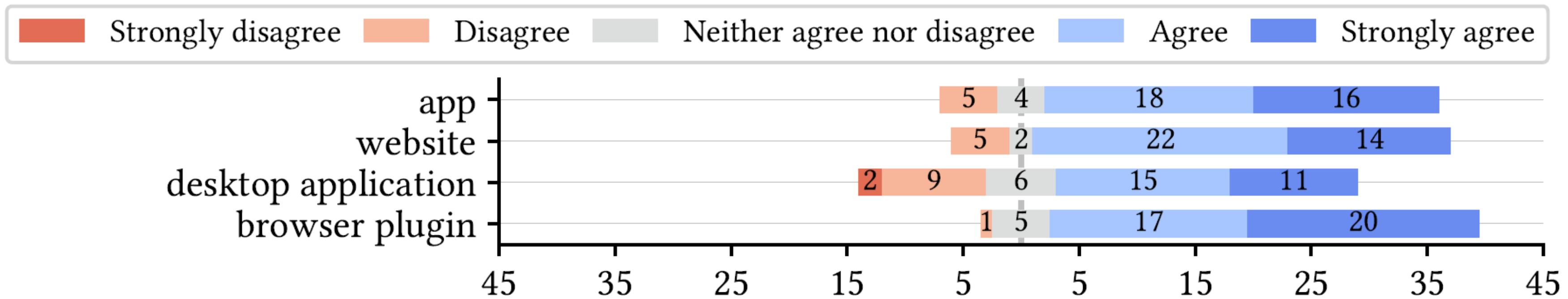
# User survey (N=43)

I would like to use this tool to check ...



# User survey (N=43)

I would like to use this tool as a ...



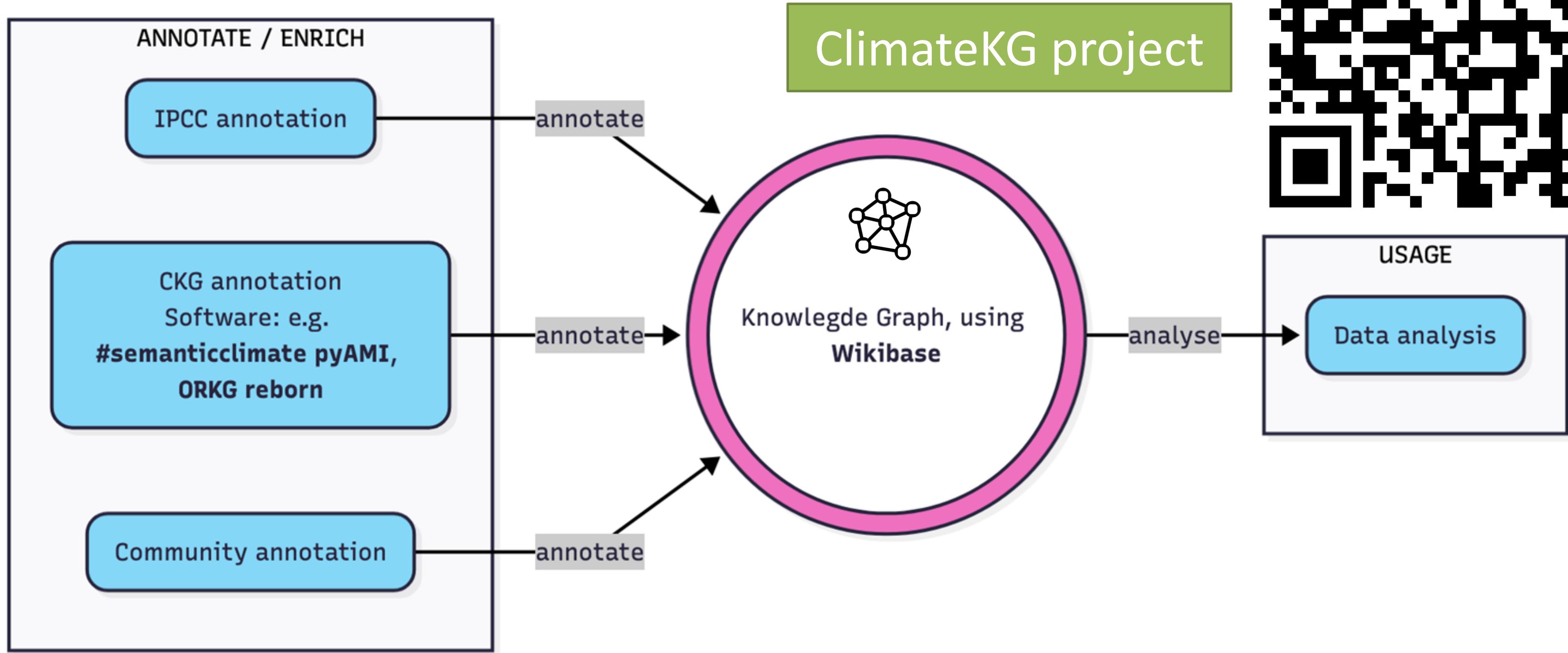


# Summary

# Technical Overview

1. Data Input: e.g. web pages, PDFs, audio/video
2. Text Extraction: BeautifulSoup, PDFMiner, OpenAI Whisper
3. Statement Extraction: explored SRL/AMR, NER-based extraction, and LLM-based extraction
  - Chosen for prototype: LLMs + Human-in-the-loop
  - LLMs tested: Mistral, Llama-2, ChatGPT; ChatGPT 3.5 used in prototype
4. Ground truth: No existing suitable, Climate KG project promising, but just started
  - Manually created MVP based on IPCC
  - KG: GraphDB, Turtle (RDF), SPARQL queries, manual validation for KG building
5. Accuracy score: Ideally find exact match via SPARQL
  - If no match: theoretical approaches include path length, graph walks, embeddings
  - Final score  $s_{acc}$  = weighted sum; prototype currently uses veracity-only ( $w_{ver} = 1$ )

# Limitations & Future work: Better KGs!



# Limitations & Future work: reliable efficient statement extraction

## Statement extraction with LLM

- easy
- unreliable
- resource expensive

➤ rebound effect!

Willenbacher, M., Hornauer, T., Wohlgemuth, V. (2022). Rebound Effects in Methods of Artificial Intelligence. In: Wohlgemuth, V., Naumann, S., Behrens, G., Arndt, HK. (eds) Advances and New Trends in Environmental Informatics. ENVIROINFO 2021. Progress in IS. Springer, Cham. [https://doi.org/10.1007/978-3-030-88063-7\\_5](https://doi.org/10.1007/978-3-030-88063-7_5)



# Summary

## Current State

**Knowledge graphs** are state-of-the-art interoperable knowledge representations. When no exact match is found, embeddings or graph walks should be utilized, particularly with increased weighting of incoming edges.

**LLMs** are state-of-the-art for scalable statement extraction, but unreliable for semantic parsing and require the identification of non-reproducible or hallucinated triples. NER mediators could complement LLMs to ensure consistency in moderating statements, instances, and classes.

**Our approach** is suitable as a personal assistant and should focus on the interoperability of tools.

# Summary

## Research Finding

*Natural language processing and knowledge graphs can help quantify the scientific accuracy of secondary literature by semantifying its content and matching it against ground truth representations. Yet, scaling is inhibited by a lack of FAIR climate change ground truth and unreliable semantification. Future work needs to collaboratively advance three interconnected inhibitors: FAIR ground truth curation, particularly in critical knowledge infrastructure; semantification accuracy, particularly triple extraction; energy-efficient reuse, particularly regarding AI-accelerated rebound effects.*

## Current State

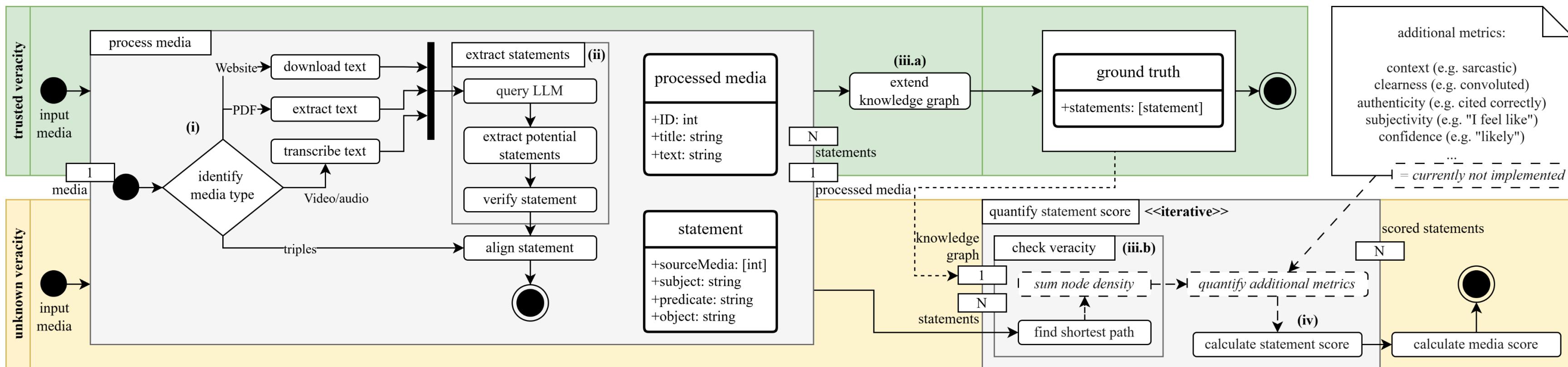
**Knowledge graphs** are state-of-the-art interoperable knowledge representations. When no exact match is found, embeddings or graph walks should be utilized, particularly with increased weighting of incoming edges.

**LLMs** are state-of-the-art for scalable statement extraction, but unreliable for semantic parsing and require the identification of non-reproducible or hallucinated triples. NER mediators could complement LLMs to ensure consistency in moderating statements, instances, and classes.

**Our approach** is suitable as a personal assistant and should focus on the interoperability of tools.

## Research Finding

Natural language processing and knowledge graphs can help quantify the scientific accuracy of secondary literature by semantifying its content and matching it against ground truth representations. Yet, scaling is inhibited by a lack of FAIR climate change ground truth and unreliable semantification. Future work needs to collaboratively advance three interconnected inhibitors: **FAIR ground truth curation**, particularly in critical knowledge infrastructure; **semantification accuracy**, particularly triple extraction; **energy-efficient reuse**, particularly regarding AI-accelerated rebound effects.



# Contact



Tim Wittenborg



tim.wittenborg@I3s.uni-hannover.de



www.I3s.de



slides, code, data



@I3s\_luh



@I3s-research-center



@forschungszentrumI3s



co

**You**

You will perform the open information extraction task. You will identify the named entities in the content and then extract the relations between them. Based on the provided testimony, you will return triples, which are formatted as [named entity A, relation, named entity B]. START of the testimony:

Human activities, principally through emissions of greenhouse gases, have unequivocally caused global warming, with global surface temperature reaching 1.1°C above 1850–1900 in 2011–2020. Global greenhouse gas emissions have continued to increase, with unequal historical and ongoing contributions arising from unsustainable energy use, land use and land-use change, lifestyles and patterns of consumption and production across regions, between and within countries, and among individuals (high confidence).

END of the testimony. The extracted triples formatted as .ttl are:

```
1 @prefix ex: <http://example.org/> .  
2  
3 ex:HumanActivities  
4   ex:cause ex:GlobalWarming .  
5  
6 ex:GreenhouseGasEmissions  
7   ex:cause ex:GlobalWarming .  
8  
9 ex:GlobalSurfaceTemperature  
10  ex:reach "1.1°C above 1850-1900 in 2011-2020"^^ex:Temperature .  
11  
12 ex:GlobalGreenhouseGasEmissions  
13   ex:increaseContinuously true .  
14  
15 ex:UnsustainableEnergyUse  
16   ex:contributeTo ex:GreenhouseGasEmissions .  
17  
18 ex:LandUseAndLandUseChange  
19   ex:contributeTo ex:GreenhouseGasEmissions .  
20  
21 ex:LifestylesAndPatterns  
22   ex:contributeTo ex:GreenhouseGasEmissions .  
23  
24 ex:Contributions  
25   ex:ariseFrom ex:UnsustainableEnergyUse, ex:LandUseAndLandUseChange, ex:LifestylesAndPatterns ;  
26   ex:varyAcross ex:Regions ;  
27   ex:varyBetweenAndWithin ex:Countries ;  
28   ex:varyAmong ex:Individuals .  
29  
30 ex:Confidence  
31   ex:hasLevel "high confidence"^^ex:ConfidenceLevel .  
32
```