Surveilling and Combating Misinformation using Knowledge Graphs

Prof. Raphael Troncy

INSTITUT CARNOT
Télécom & Société numérique

EURECOM
Sophia Antipolis

Misinformation has serious impact on societal topics

- **Nature**: Climate change denial, Australian bushfires, etc.

- **Health**: COVID "infodemic", Vaccines, etc.

- **Politics**: US presidential elections, Brexit, Ukraine war, etc.

Donald J. Trump ✔
@realDonaldTrump

Think of it, a modestly successful comedian, Volodymyr Zelenskyy, talked the United States of America into spending $350 Billion Dollars, to go into a War that couldn't be won, that never had to start, but a War that he, without the U.S. and "TRUMP," will never be able to settle. The United States has spent $200 Billion Dollars more than Europe, and Europe's money is guaranteed, while the United States will get nothing back. Why didn't Sleepy Joe Biden demand Equalization, in that this War is far more important to Europe than it is to us — We have a big, beautiful Ocean as separation. On top of this, Zelenskyy admits that half of the money we sent him is "MISSING." He refuses to have Elections, is very low in Ukrainian Polls, and the only thing he was good at was playing Biden "like a fiddle." A Dictator without Elections, Zelenskyy better move fast or he is not going to have a Country left. In the meantime, we are successfully negotiating an end to the War with Russia, something all admit only "TRUMP," and the Trump Administration, can do. Biden never tried, Europe has failed to bring Peace, and Zelenskyy probably wants to keep the "gravy train" going. I love Ukraine, but Zelenskyy has done a terrible job, his Country is shattered, and MILLIONS have unnecessarily died – And so it continues.....

17.8k ReTruths   **65.4k** Likes   Feb 19, 2025, 4:47 PM

- **Persuasion techniques**
  - Push **agendas** and **narratives**
  - Leveraging **emotions**
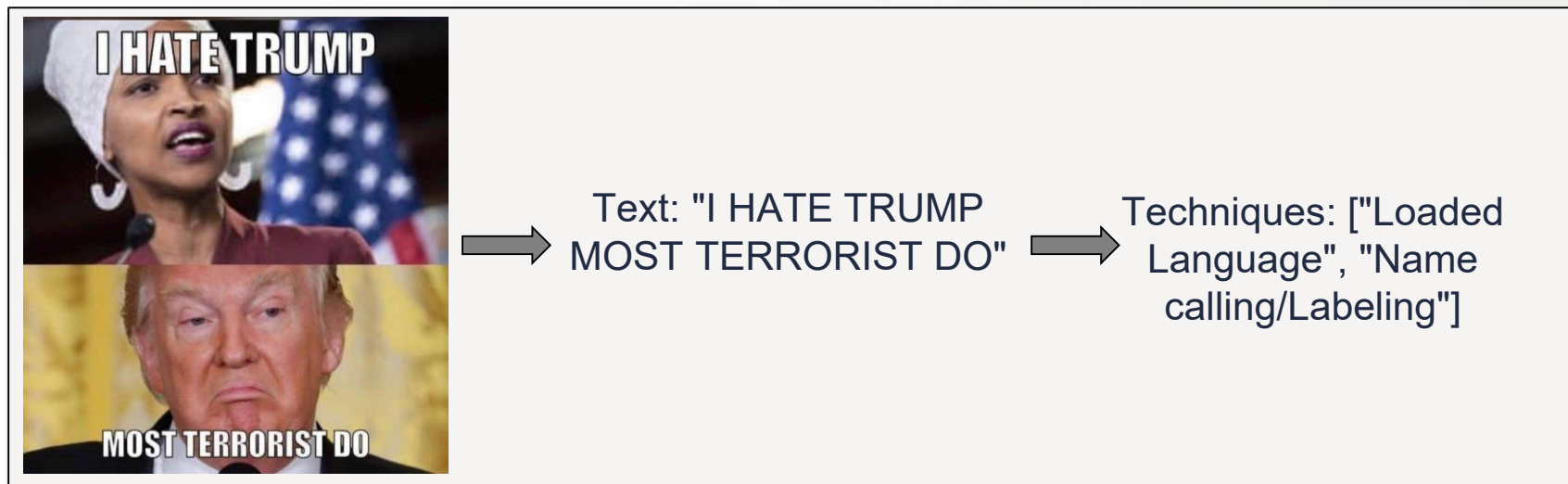 - Using **fallacies**

- **Conspiracy theories**
  - Malevolent **schemes**
  - Disproved or unproven **accusations**
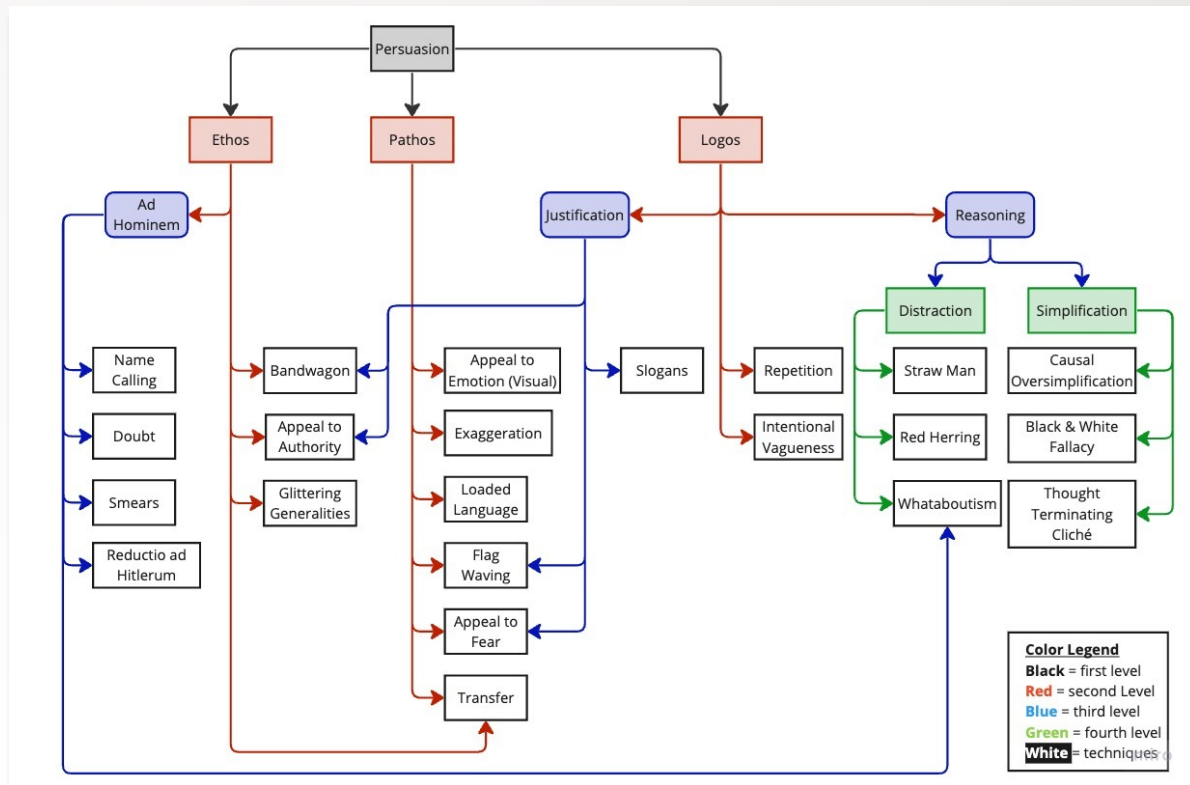
Persuasion Techniques

SemEval-2024: Detection of persuasion techniques in memes

Text: "I HATE TRUMP MOST TERRORIST DO" → Techniques: ["Loaded Language", "Name calling/Labeling"]

- Train: 5000 memes
- Test: 1000 memes

20 different classes using a **hierarchical** structure

- **Models:** BERT, BERT-HarMe, RoBERTa, AlBERT, DeBERTa, DistilBERT
- **Losses:** CE, BCE, Focal Loss, Hierarchical Loss (HL)
- **Data:** 2024, 2021, GPT-augmented
- **Output Classes:** 20, 28

$$\mathcal{L}_{BCE}^{a} = y^a \cdot log\sigma(\max(\{x^c\}_{c \in child(a)})) + (1 - y^a) \cdot log(1 - \sigma(\max(\{x^c\}_{c \in child(a)})))$$

$$\mathcal{L}_{HL} = \mathcal{L}_{BCE} + \alpha \cdot \sum_{a \in \mathcal{A}} \mathcal{L}_{BCE}^{a}$$
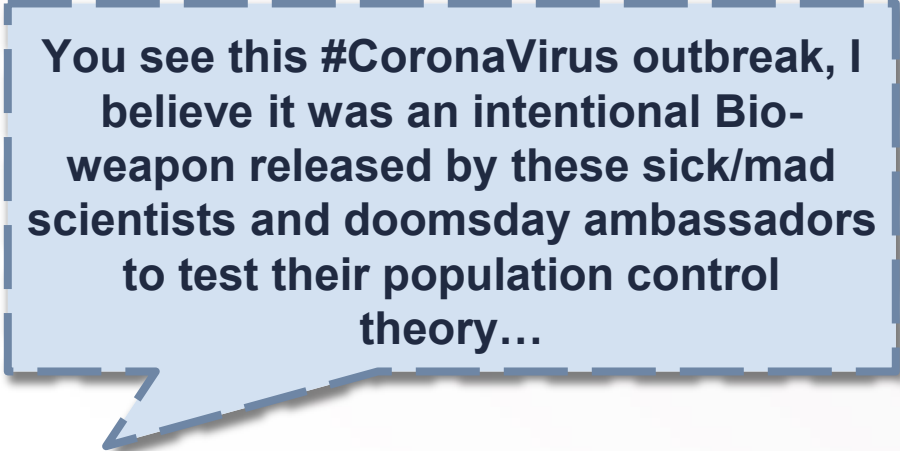
Our approach:

Ensembling of BERT-based models

"EURECOM at SemEval-2024 Task 4: Hierarchical Loss and Model Ensembling in Detecting Persuasion Techniques", Peskine Y., Troncy R., Papotti P., Proceedings of the 18th International Workshop on Semantic Evaluation

	Team	F1H
1	**914isthebest**	**0.752**
2	BCAmirs	0.699
3	OtterlyObsessedWithSemantics	0.697
4	TUMnlp	0.674
5	GreyBox	0.670
6	NLPNCHU	0.663
7	Puer	0.660
8	EURECOM	0.655
9	SuteAlbastre	0.652
10 / 33	UMUTeam	0.648

Code shared on github: https://github.com/D2KLab/semeval-2024-task-4

Conspiracy Theories

MediaEval: Detection of 9 COVID-related conspiracy theories in tweets

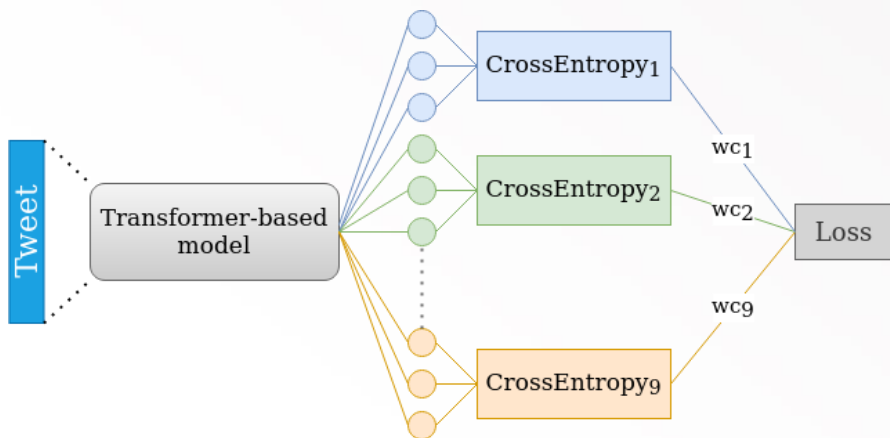Suppressed Cures, Antivax, Fake virus, etc.

> **You see this #CoronaVirus outbreak, I believe it was an intentional Bio-weapon released by these sick/mad scientists and doomsday ambassadors to test their population control theory…**

Our winning approach:
Ensembling of Covid-Twitter-BERT models

"Detecting COVID-19-Related Conspiracy Theories in Tweets", Peskine Y., Alfarano G., Harrando I., Papotti P., Troncy R., Multimedia Benchmark Workshop (MediaEval 2021)

	Team	MCC
1	**D2KLab**	**0.733**
2	SELAB_HCMUS	0.648
3	Deltamap	0.632
4	SELAB-HMUS-Junior	0.599
5	Upsilon Labs	0.454
6	MG-UCB	0.450
7	FakeINA	0.446
8	OTS-UEC	0.413
9/17	Delta_IITKGP	0.347

Code shared on github: https://github.com/D2KLab/mediaeval-fakenews
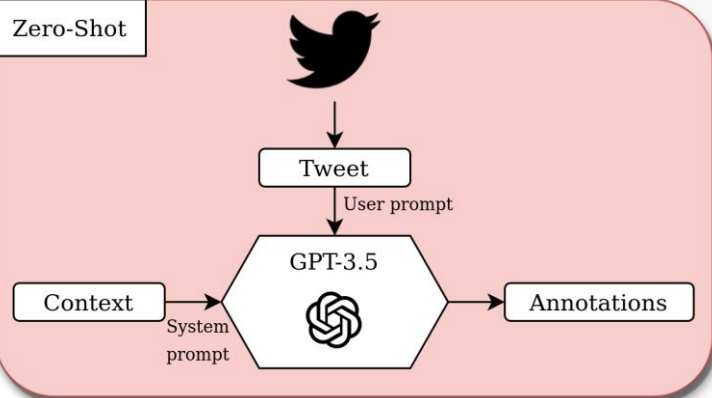
9

Definitions Matters

Can LLMs detect conspiracy theories?

- Experiment with **GPT-3.5**

- **Zero-shot** settings
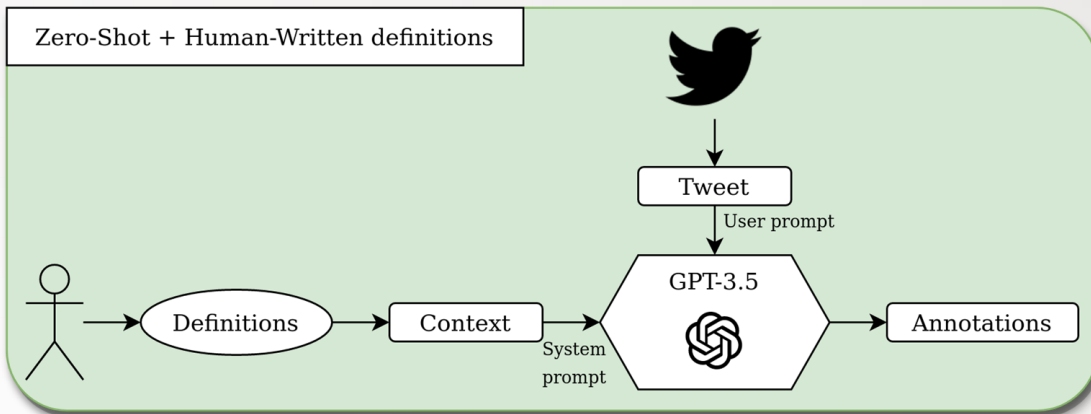
- Explore impact of different **definitions** of classes

"Definitions matter: Guiding GPT for multi-label classification", Peskine Y., Korenčić D., Grubisic I., Papotti P., Troncy R., Rosso P., Conference on Empirical Methods in Natural Language Processing (EMNLP 2023)
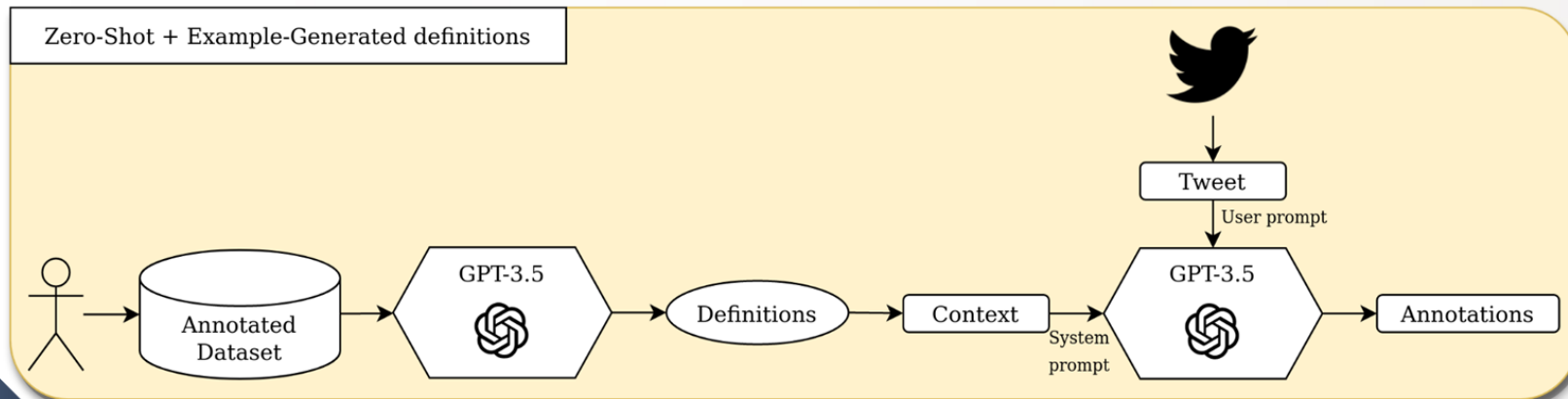
Definitions Matters

Classification Results

- **Definitions** improve classification performance

- **Human-Written** (HW) perform best on average

- Best **Example Generated** (EG) definitions outperform their HW counterparts

Suppressed Cures (SUP), Behaviour Control (BHC), Anti-Vaccination (AVX), Fake Virus (FAK), Intentional Pandemic (INT), Harmful Radiation (HAR), Depopulation (DEP), New World Order (NWO), Satanism (SAT)

Code shared on github: https://github.com/dkorenci/gpt-def-zeroshot
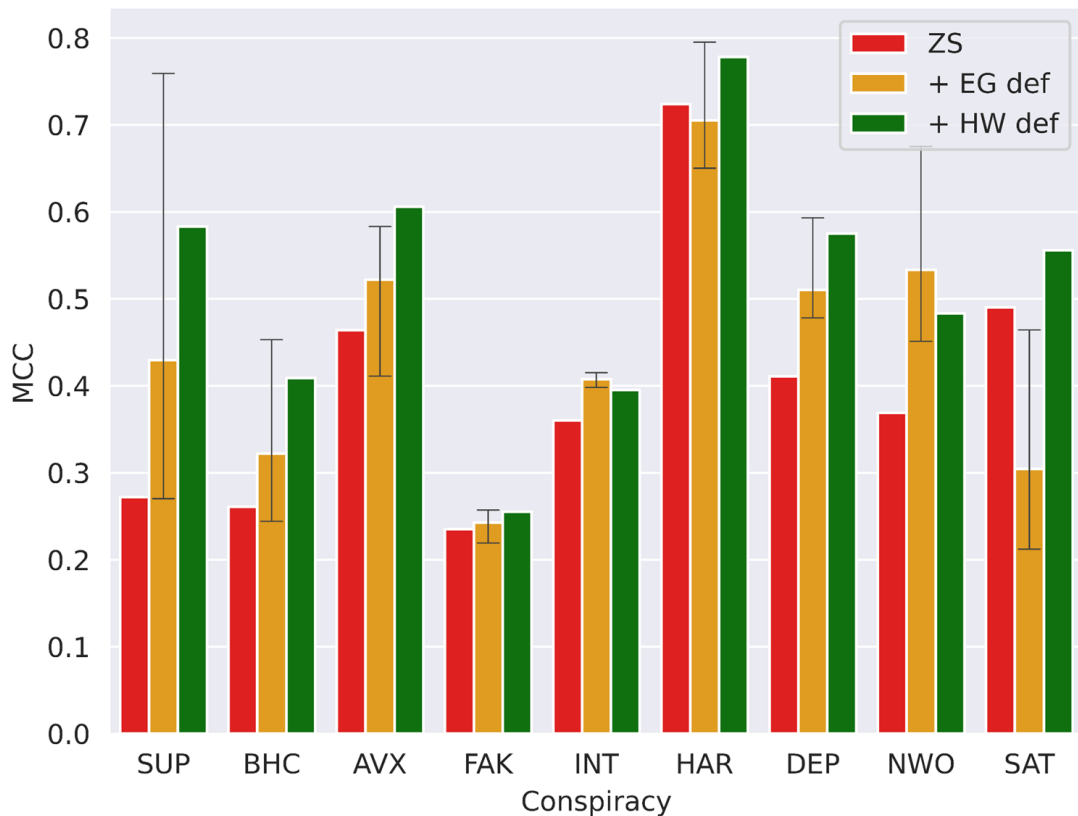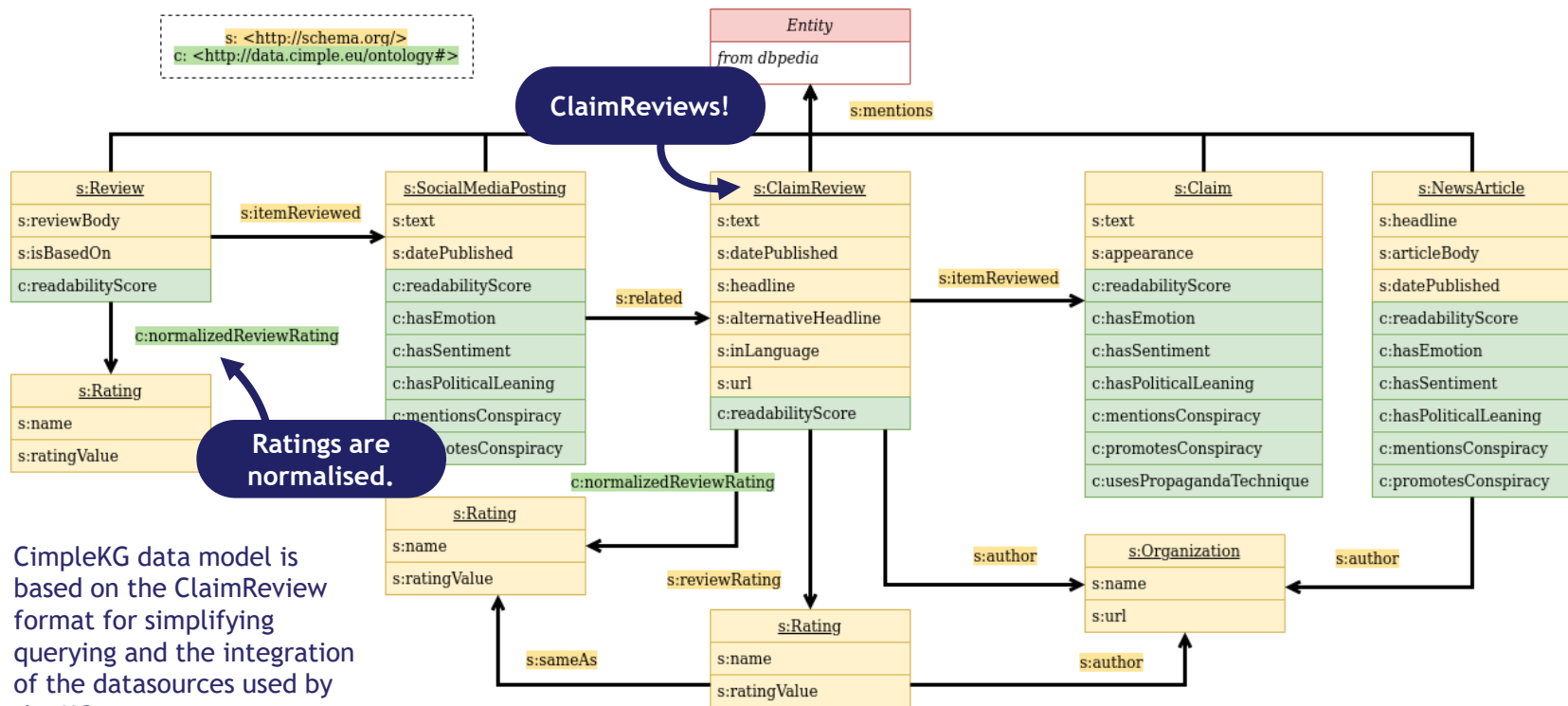
CimpleKG ClaimReviews data collection pipeline.
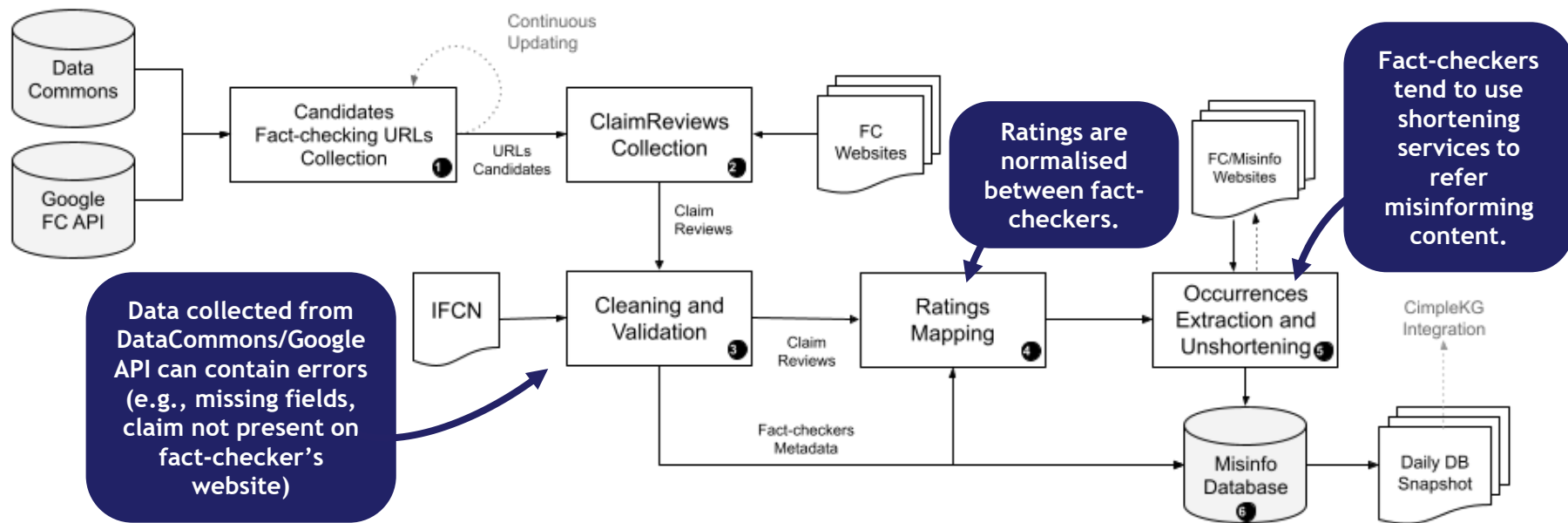
Data collection and processing pipeline for gathering Claim Reviews. The Claim Review ratings are normalised to a set of 5 fixed values: **not verifiable**, **not credible**, **mostly credible**, **credible** and **uncertain.**

The CimpleKG data model.

s: <http://schema.org/>
c: <http://data.cimple.eu/ontology#>

Entity
from dbpedia

ClaimReviews!

s:mentions

s:Review
s:reviewBody
s:isBasedOn
c:readabilityScore

s:itemReviewed

s:SocialMediaPosting
s:text
s:datePublished
c:readabilityScore
c:hasEmotion
c:hasSentiment
c:hasPoliticalLeaning
c:mentionsConspiracy
c:promotesConspiracy

s:related

s:ClaimReview
s:text
s:datePublished
s:headline
s:alternativeHeadline
s:inLanguage
s:url
c:readabilityScore

s:itemReviewed

s:Claim
s:text
s:appearance
c:readabilityScore
c:hasEmotion
c:hasSentiment
c:hasPoliticalLeaning
c:mentionsConspiracy
c:promotesConspiracy
c:usesPropagandaTechnique

s:NewsArticle
s:headline
s:articleBody
s:datePublished
c:readabilityScore
c:hasEmotion
c:hasSentiment
c:hasPoliticalLeaning
c:mentionsConspiracy
c:promotesConspiracy

c:normalizedReviewRating

s:Rating
s:name
s:ratingValue

Ratings are normalised.

c:normalizedReviewRating

s:Rating
s:name
s:ratingValue

s:reviewRating

s:Rating
s:name
s:ratingValue

s:sameAs

s:author

s:author

s:Organization
s:name
s:url

s:author

s:author

CimpleKG data model is based on the ClaimReview format for simplifying querying and the integration of the datasources used by the KG.

CimpleKG explorer.

- The CimpleKG data explorer enables searching and browsing for the KG.

- Claims and tweets can be discovered based on filters (language, ratings, entities, etc.).

➤ https://explorer.cimple.eu/

Misinfome Bot.

The misinfome bot is a Twitter/X social media bot that post misinformation corrections to misinformation spreaders.

- The bot uses CimpleKG to retrieve recent fact-checks.

- The bot uses a set of templates for correcting misinformation spreaders.

Burel, G., M. Tavakoli, and H. Alani. 2024. *Exploring the impact of automated correction of misinformation in social media. AI Magazine* 45: 227–245. https://doi.org/10.1002/aaai.12180

The CimpleKG data sources.

Dataset	Document Types	Nb. of documents
Claim-Reviews	Claims, Claim Reviews	199,964 claims 215,446 Claim-Reviews
AFP	News Article	193,933 news articles
Community Notes (Twitter)	Social Media Posts, Reviews	6,563 tweets 1,983 reviews 1,112 links to Claim-Reviews
CLEF CheckThat! 2022	Social Media Posts, Claim Reviews	1,196 tweets 1,198 links to Claim-Reviews
MediaEval 2022	Social Media Posts	2,702 tweets
Propaganda Corpus	Claims	1,908 claims

Updated daily

Statistics of the datasets integrated into CimpleKG

CimpleKG

A Continuously Updated Knowledge Graph on Misinformation, Factors and Fact-Checks

SPARQL Endpoint:
https://purl.org/net/cimplekg/sparql

KG Releases:
https://purl.org/net/cimplekg/knowledge-graph

K M i Knowledge Media Institute

EURECOM
Sophia Antipolis

ClimateSense

Climate Misinformation Surveillance and Analysis with Multidimensional GIS

The challenge of climate misinformation.

Climate mis/disinformation is the non-intentional/intentional dissemination of **false information related to climate change and climate action**:

- Undermine trust in science.
- Weaken public support for climate action.
- Hinder policy implementation and international cooperation on climate change.
- Undermine democracy.

How do we fight climate mis/disinformation effectively?

Global Initiative for Information Integrity on Climate Change
UN, UNESCO, Brazil

1) Check sources

Confirm the credibility of the source by verifying whether it references scientific institutions, peer-reviewed journals, and reputable news outlets.

2) Pay attention to language and tone

Sensationalist language, exaggerations and appeals to strong emotions are all red flags when checking for disinformation.

3) Fact-check

Make use of independent fact-checking websites available to help you verify claims.

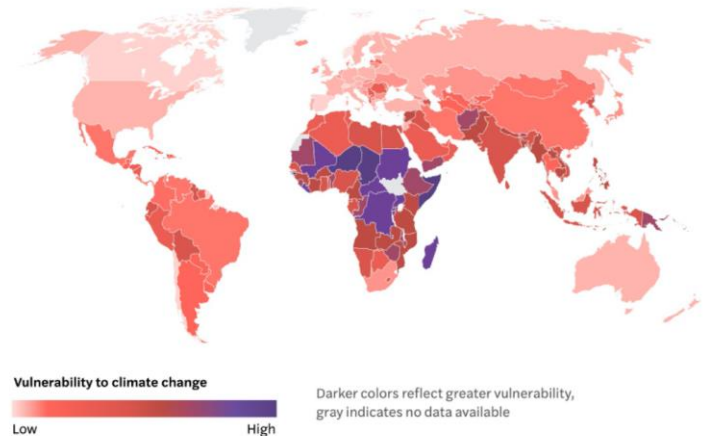https://climate.ec.europa.eu/eu-action/climate-disinformation_en

20

Misinformation vs. climate misinformation.

- **Scale and scope**: Climate change misinformation deals with a global, long-term issue, making it more complex than misinformation about localized or short-term events.
- **Scientific complexity**: The intricate nature of climate science provides fertile ground for misinterpretation and manipulation.
- **Political and economic interests**: Climate change is deeply intertwined with political and economic interests, leading to well-funded and organized misinformation campaigns.

The countries most vulnerable to climate change

The Notre Dame Global Adaptation Initiative measures overall vulnerability to climate change by considering exposure, sensitivity, and ability to adapt across six life-supporting sectors – food, water, health, ecosystem service, human habitat, and infrastructure

Vulnerability to climate change

Low High

Darker colors reflect greater vulnerability, gray indicates no data available

Source: Notre Dame Global Adaptation Initiative

Rising climate misinformation hinders effective policy.

21

Climate Misinformation, Fact-checking and Social Media

NEWS

Home | InDepth | Israel-Gaza war | War in Ukraine | Climate | UK | World | Business

Technology

Facebook and Instagram get rid of fact checkers

TikTok fails to enforce its own policies on climate misinformation during COP29

Published: 06 December 2024

SHARE

THE CONVERSAT

Academic rigour, journalistic flair

Arts + Culture | Business + Economy | Education | Environment | Health | Politics + Society | Science + Tech | World | Podcasts | Insights

Climate misinformation is rife on social media – and poised to get worse

Published: January 17, 2025 2.09pm GMT

Politically charged rumors and conspiracy theories about Helene flourish on X

OCTOBER 3, 2024 · 9:39 AM ET

HEARD ON ALL THINGS CONSIDERED

By Huo Jingnan

▶ 2-Minute Listen   + PLAYLIST   ≡ TRANSCRIPT   ⋯

Marine One, carrying President Biden, flies above a storm-impacted area near Asheville, N.C., on Wednesday. Biden was visiting North and South Carolina to survey damage after Hurricane Helene. Social media sites, especially X, were filled with politically charged rumors and conspiracy theories about the hurricane response.
Mandel Ngan/AFP/Getty Images

nman & Courtney Subramanian
ngton

22

Climate misinformation exposure

Media type	Share of respondents
Online (includes socials and messaging)	21%
Social media	16%
TV news	15%
Messaging apps	5%
Newspapers	5%
Radio news	5%
Elsewhere	12%
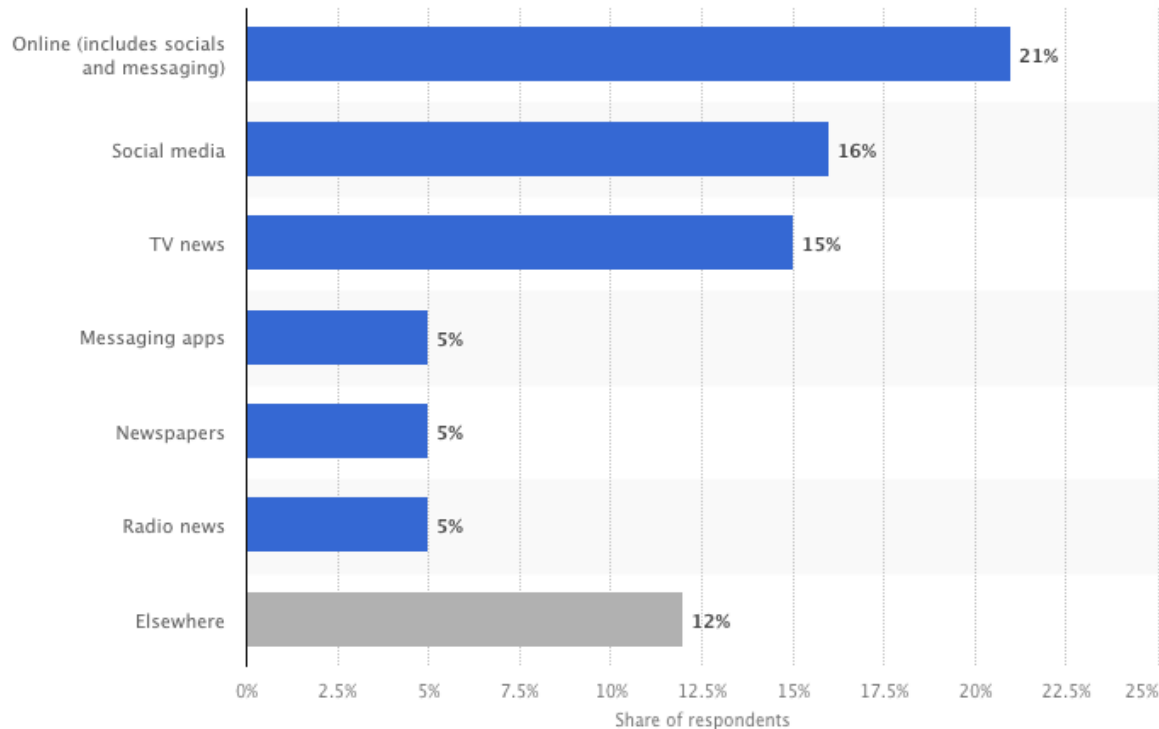
Share of people who think they came across false or misleading information about climate change in the U.S. in 2023, by media type

https://www.statista.com/statistics/1490059/opinion-climate-change-media-misinformation-occurrence-by-type-usa/

Frugal AI Challenge

Frugal AI Challenge
co-located with the 2025 AI Action Summit

https://frugalaichallenge.org/

Detecting Climate Change related Disinformation:
- **Textual** data
- Identify misleading or false information related to **climate change**
- **8 classes**: "climate change is not happening", "It is not human-caused", "science is unreliable", etc.
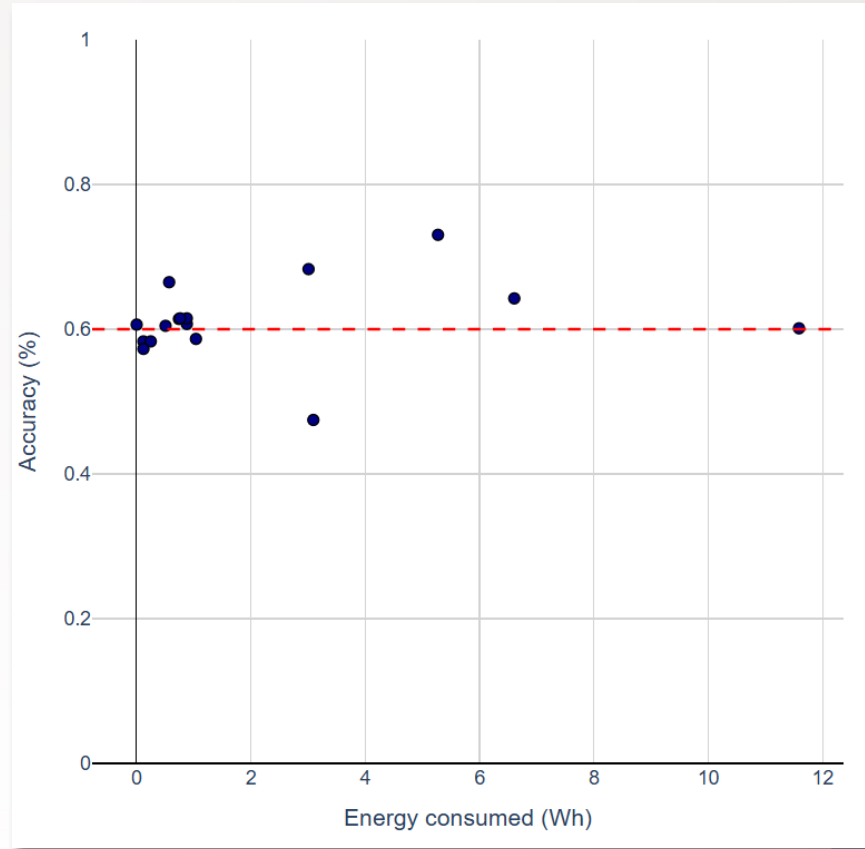
Evaluation:
- **Accuracy**
- **Energy consumption**

Proposed approaches:

- Fine-tuned BERT
- Embeddings + Classifier
- Fine-tuned LLMs

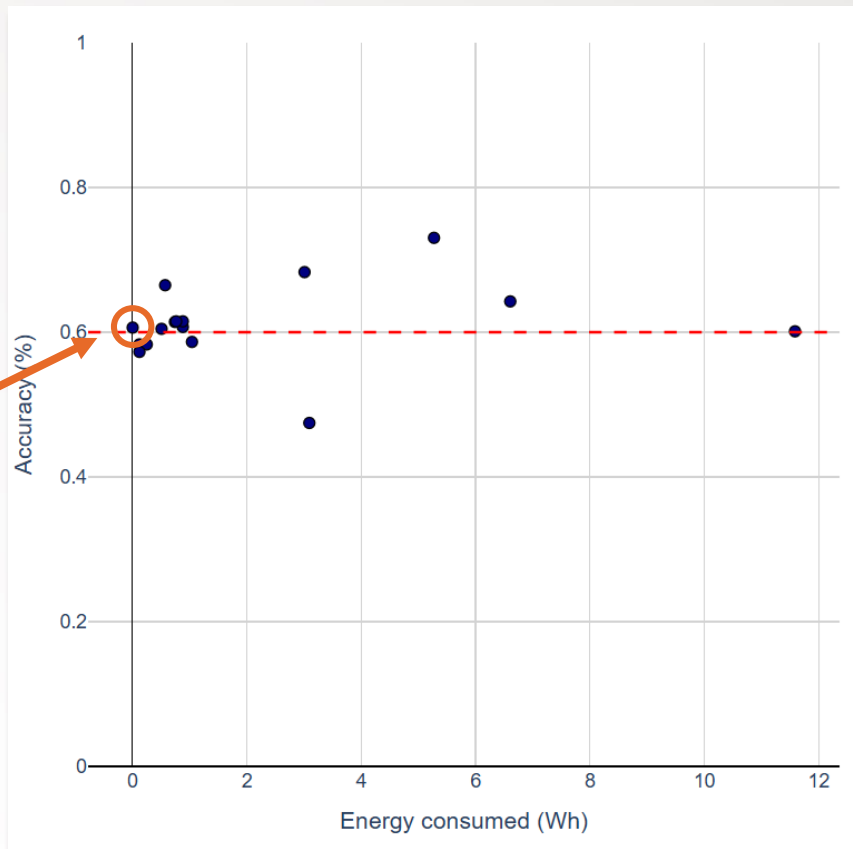**Proposed approaches:**
- Fine-tuned BERT
- Embeddings + Classifier
- Fine-tuned LLMs

Our **winning submission**:
- Sentence-BERT + MLP

Code shared on github: https://github.com/D2KLab/frugal-ai-challenge-eurecom

Project consortium.

Knowledge Media institute, The Open University, UK

www.open.ac.uk

EURECOM, France

www.eurecom.fr

Prague University of Economics and Business (VSE), Czech Republic

www.vse.cz

Kaunas University of Technologies (KTU), Lithuania

www.ktu.edu

Climate Misinformation Surveillance and Analysis with Multidimensional GIS

ClimateSense aims to tackle the issue of climate misinformation by leveraging the powers of Geographic Information Systems (GIS):

- **Develop a GIS** to integrate: Climate Data (temperature, precipitation, emissions).

- **Track climate misinformation** (fact-checks, social media).

- **Analyse correlations between misinformation and geographical/climate factors**.

- Develop predictive capabilities to counter misinformation proactively.
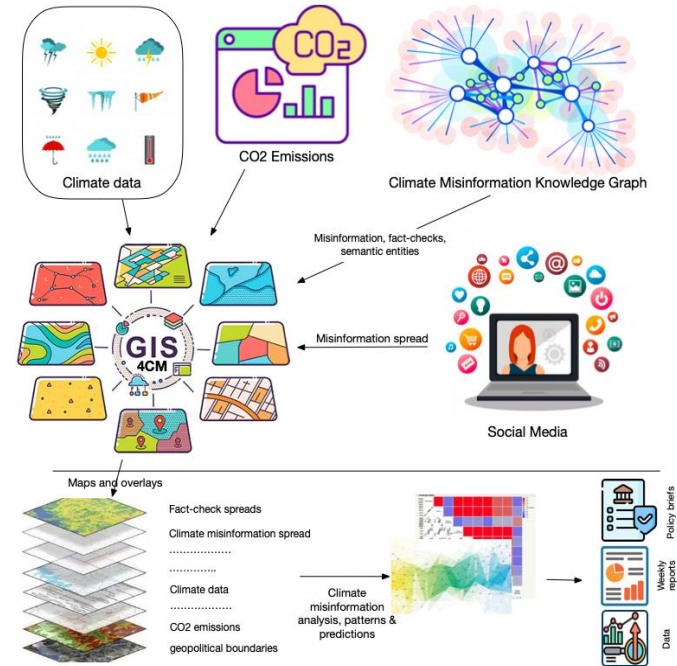
Misinformation spread reports

Following a similar approach to the Fact-checking Observatory (FCO), create reports around the **spread of climate misinformation**.

1. Reports about *misinformation spread*:
   a. **Spread of specific misinformation topic** based on SkepticalScience topics.
   b. Overall spread for a given time period (same as FCO).
   c. **Country report**: spread of misinformation for a country.
2. Reports about *media and fact-checkers coverage*:
   - **Same as above** but focus on media and fact-check analysis (does not require social media tracking).
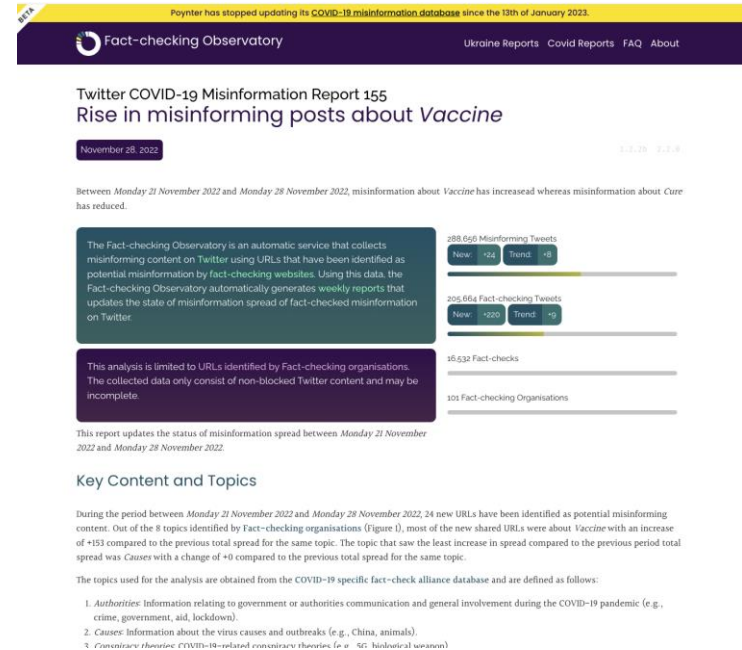
Facts and science search

Create a tool that identify scientific evidence for climate-related claim search:

1. Using CimpleKG, Skeptical Science KG, CORE and other data source and LLMs, create a tool that:
   a. Identify existing science related to the query.
   b. Make a judgment statement about the original claim.
   c. Identify scientific evidence.
2. Create a summary of the finding using LLMs and graphs.
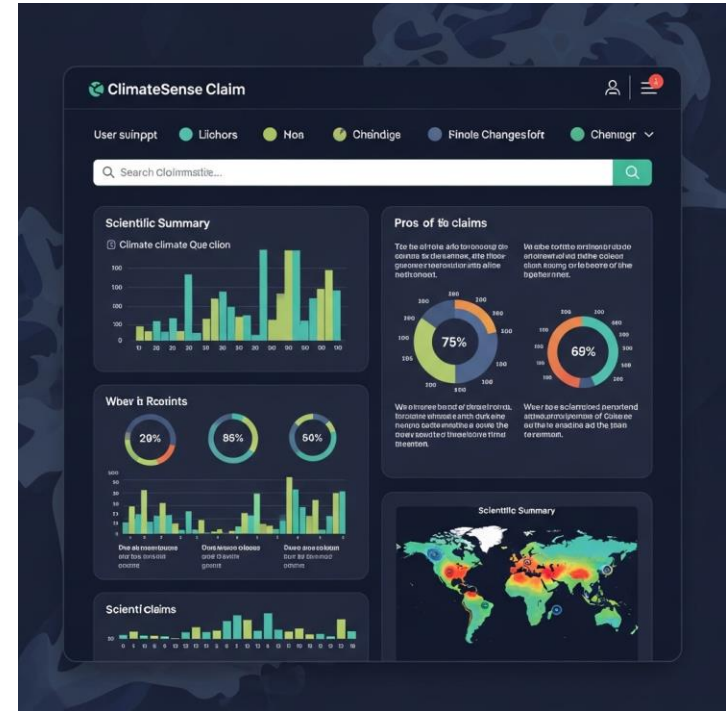
AI Generated mockup.

Lobby map and policy Impact

Create a tool that identify lobby interference and the impact of policies:

1. Use resources like **LobbyMap** and the **Climate Policy Database** to identify environmental policies and lobbies location.
2. Analyse **media discourse** and **official press releases** of lobbies and the historical implementation of policies over time.
3. Display influence and interference on maps by identifying lobbies interests.
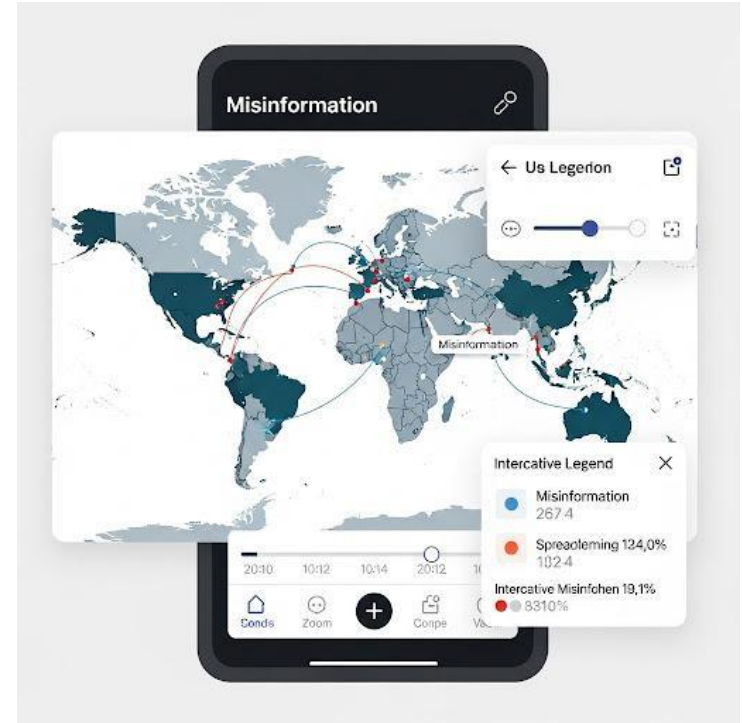

AI Generated mockup (Gemini doesn't know how to spell).

Temporal Misinformation and Communication

As written in the proposal, create a tool that visualise how misinformation spreads:

1. **Track misinformation claims in the media and on social media**.
2. Visualise misinformation paths including sources and targets.
3. Show the diffusion of misinformation.


AI Generated mockup.

Evaluation and impact

The project output will be evaluated by organising **workshops with key stakeholders** at relevant conference :

- Evaluation synchronised with the **COP conferences** (COP30, COP31 and COP32).
- Evaluation with **policymakers, fact-checkers, citizens, NGOs and business representatives**.
- Research outputs released as open source.

Project impact is expected in multiples areas such as :
- **Economic impact**: ClimateSense will contribute to reducing the cost of climate change by arming policymakers with powerful tools.
- **Societal impact**: ClimateSense will contribute to raising awareness of society to the propagation of misinformation.
- **Policy impact**: ClimateSense will create evidence-based policy brief in multiple languages.
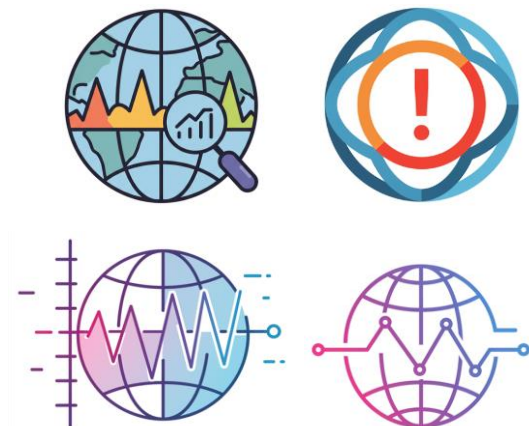
COP30
BELÉM · BRASIL · 2025

Google Search Central Blog

Home › Search Central › Google Search Central Blog

Simplifying the search results page 🔖▾

Was this helpful? 👍 👎

Send feedback

June 12, 2025

As part of our ongoing efforts to simplify the Google Search results page, we will be phasing out support for a few structured data features in Search. We regularly evaluate the usefulness of Search features, both for users and website owners.

We're phasing out these specific structured data types because our analysis shows that they're not commonly used in Search, and we found that these specific displays are no longer providing significant additional value for users. Removing them will help streamline the results page and focus on other experiences that are more useful and widely used.

This update won't affect how pages are ranked. This simplification means that for some results, the specific visual enhancements powered by these lesser-used markups will no longer appear, leading to a more streamlined presentation. The use of these structured data types outside of Google Search (and dependent features) is not affected.

The following structured data types will no longer be supported in Google Search results and will be phased out over the coming weeks and months:

- Book Actions

- Course Info   **https://developers.google.com/search/blog/2025/06/simplifying-search-results**

- Claim Review

- Estimated Salary

- Learning Video

- Special Announcement

Fact checks **Analysis** Policy & Impact About   Search

Donate

HOME / COMMENT

The web just got a little harder to trust

26 JUNE 2025

In a change buried in a little-known technical blog, Google has announced that it will abandon a key element of its tagging system for fact checks, effectively de-prioritising this content and making it harder for people to access reliable information.

For the last 10 years, fact checkers around the world have quietly been working on what is probably the largest structured data journalism project ever created. The project, called ClaimReview, is part of the plumbing of the internet. ClaimReview is a tagging system that lets search engines, apps and social media platforms find fact checks and show them in other places, like newsfeed or search results.

This is part of the wider Schema.org set of standards, such as Movie, Local business, or Recipe, which ensure that when you search for something online, you get specific, contextualised results.

https://fullfact.org/technology/the-web-just-got-a-little-harder-to-trust/

oogle   cinema listings in cardiff

All   Images   News   Short videos   Maps   Shopping   Videos   More