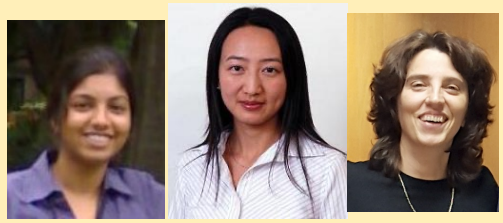
Data and AI for trustworthy digital democracy

Ioana Manolescu

Overview of our research

Big Data Management and Analytics



Models, methods and tools for digital democracy

Natural language processing

Challenges in data journalism

Le Monde

radiofrance

Digital data sources are **heterogeneous**

- ❑ Acquisition driven by *interest* and *opportunity*, not quality/model/schema/platform etc.
- ❑ No time / means to design an integrated schema
  Also: data sets are dynamic; application need may vary
- ❑ Non-technical users, no system literacy
- ❑ Datasets often features common actors, concepts of interest to the journalist

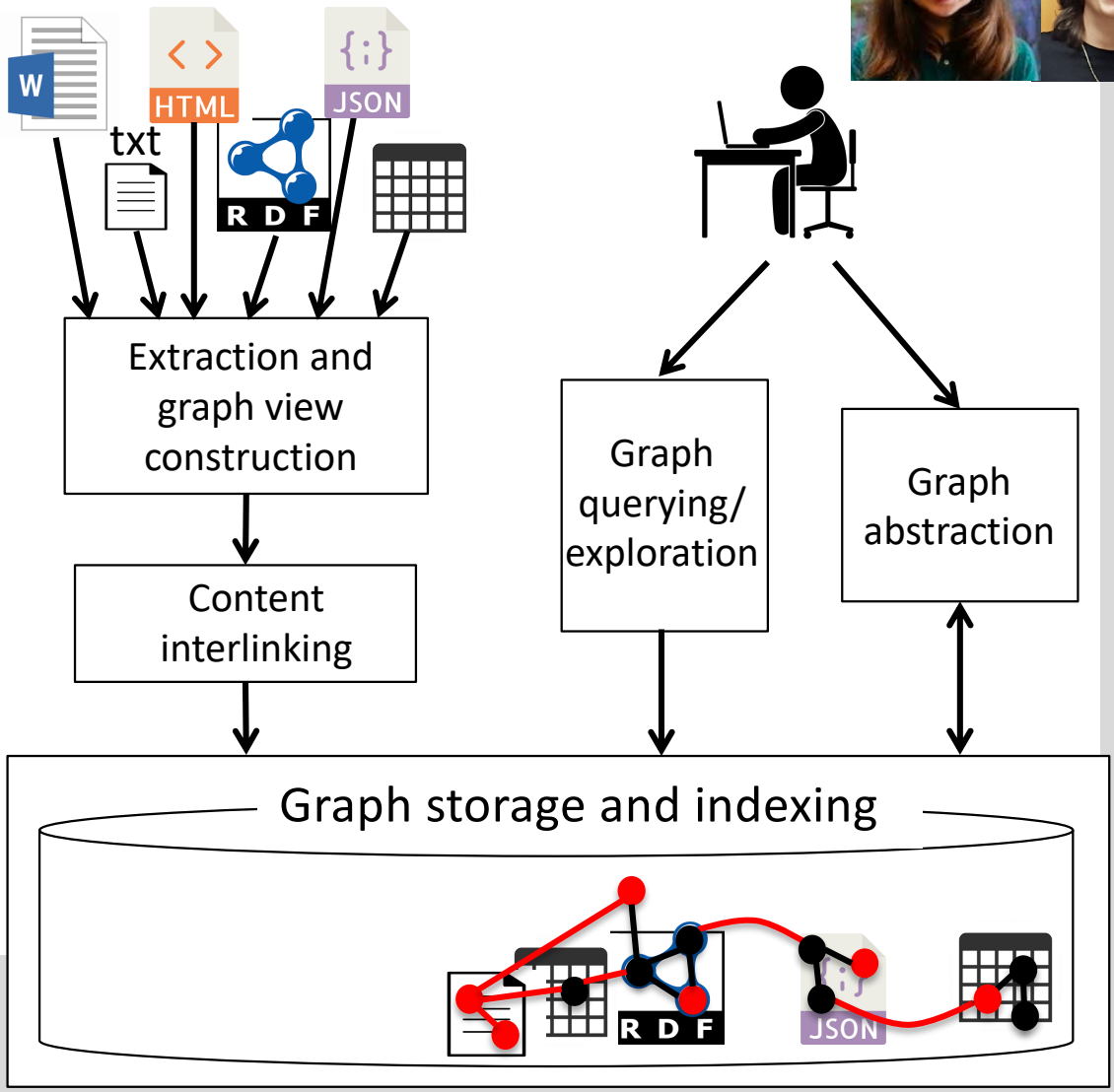
Inria

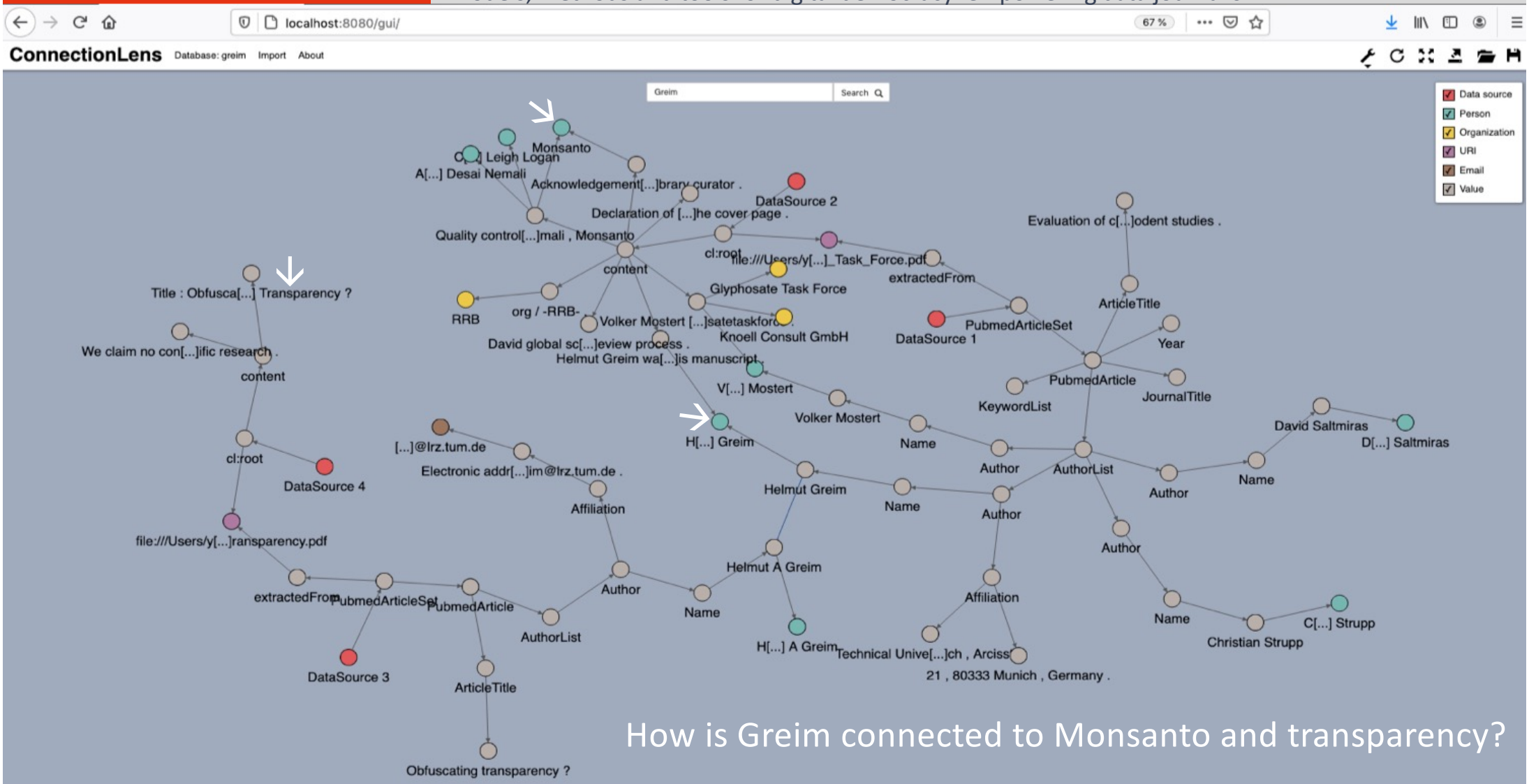
Digital arenas

- Integrate any data in a **graph**
- Based on common **entities**, and **relationships extracted**
- **Querying** for connections between heterogeneous data sources (keywords+structure)

[InfSystems22, ICDE23, VLDB23, ADBIS23, CoopIS23, EACL23, EMNLP23]
Invited keynotes [DASFAA22, DEBS22, ICFCA21, DOLAP21, ADBIS20, DATA20]

https://connectionstudio.inria.fr (code),
https://sourcessay.inria.fr





How is Greim connected to Monsanto and transparency?

How do we analyze statements?

A statement is made

How do we analyze statements?

A statement is made

What should we think of this?

How do we analyze statements?

A statement is made

What should we think of this?

Information?
Misinformation?
Dysinformation?

How do we analyze statements?

A statement is made

What should we think of this?

Information?
Misinformation?
Disinformation?

Evaluate
source authority
(past statements…)

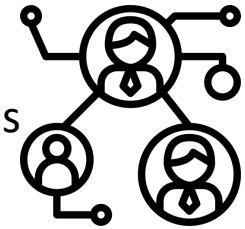
How do we analyze statements?

A statement is made

What should we think of this?

Information?
Misinformation?
Dysinformation?

Evaluate source connections



Evaluate source authority (past statements…)

How do we analyze statements?

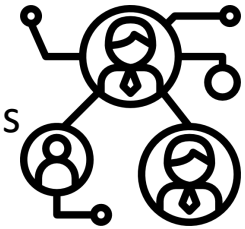
A statement is made

Consult reference data

What should we think of this?

Information?
Misinformation?
Dysinformation?

Evaluate source connections

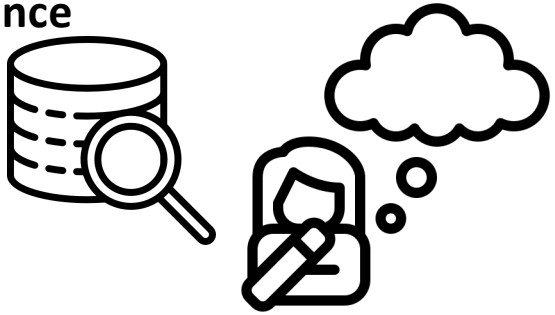


Evaluate source authority (past statements…)

How do we analyze statements?

A statement is made

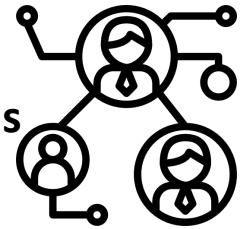
Consult **reference data**



What should we think of this?

Information?
Misinformation?
Disinformation?

Evaluate **source connections**



Evaluate **source authority (past statements…)**

How do we analyze statements?

A statement is made

Information?
Misinformation?
Disinformation?

source connections

source past statements

fact-checks

reference data

Fact-check retrieval pipeline

A statement is made

Consult **reference data**

What should we think of this?

Information?
Misinformation?
Disinformation?

Goal: given a statement, search for previous FC on this or a similar topic

- ❑ Recent FCs: to save effort (avoid checking again)
- ❑ Older FCs: false information is often repurposed, e.g., criminal son of a French minister, etc.
- ❑ Also talking points from malicious agents.

The FC retrieval problem

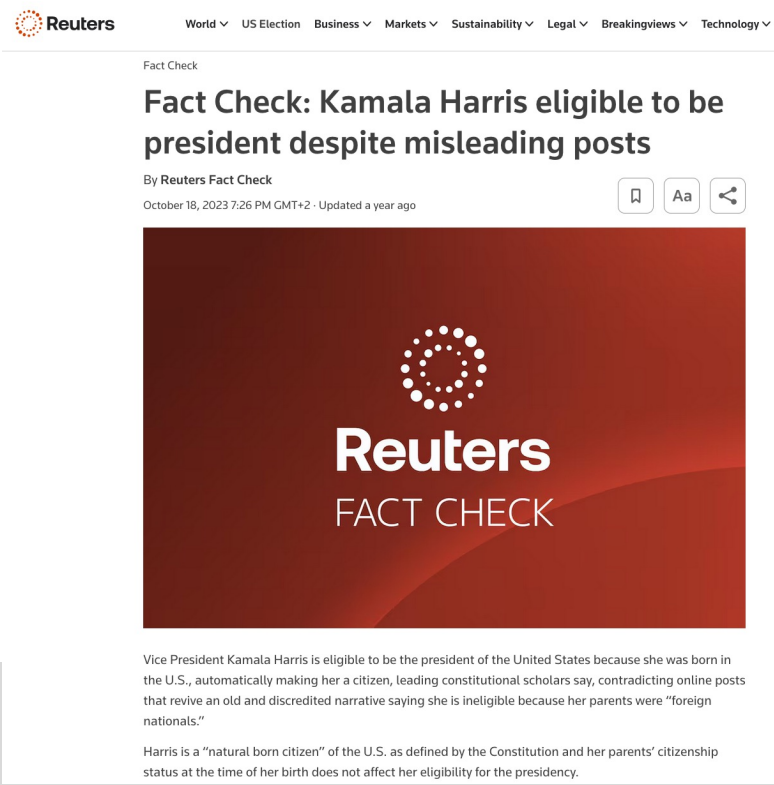
A **claim** is: a phrase, or a social media post, or an image (→ text)

A **fact-check** consists of: title, claimant, publisher, dateOfPublication, URLtoArticle, claimText, language and rating

ClaimText can be short-ish (a phrase)

URLtoArticle points to a long article

Title is typically related to claimText but not identical



The FC retrieval problem

A **claim** is: a phrase, or a social media post, or an image (→ text)

A **fact-check** consists of: title, claimant, publisher, dateOfPublication, URLtoArticle, claimText, language and rating

- ❑ Structure from de-facto standard (Google ClaimCheckReview)

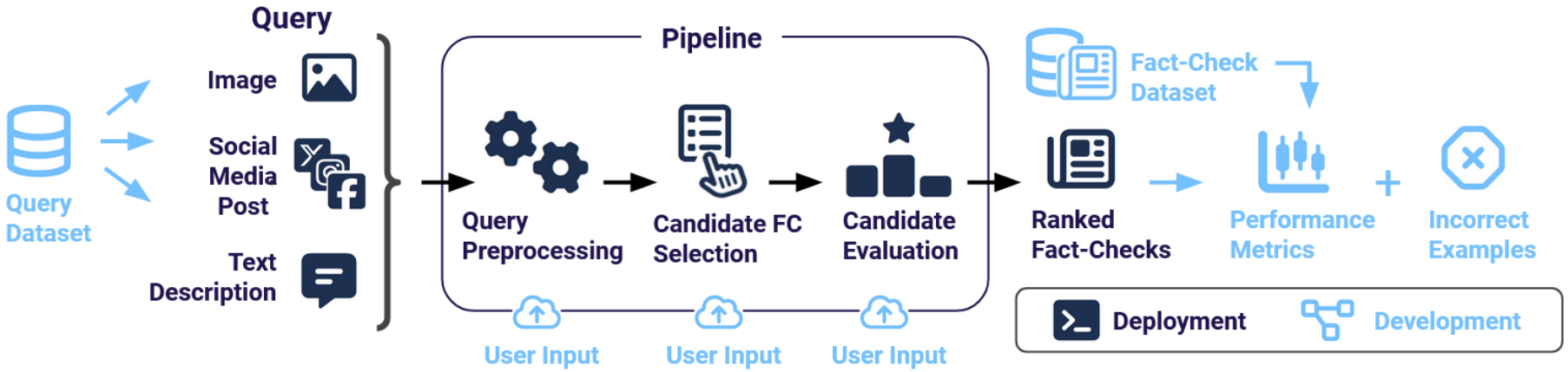
FC retrieval is a 2-step process:

1. Retrieval: out of a potentially large FC set, given a claim, find possible candidates. Probabilistic vs. neural methods
2. (Re-)ranking: determine the relevance of each candidate, to show the most relevant first. Nowadays , typically neural methods (transformers)

FactCheckBureau: a tool with many modes

We built our platform to support:

- ❑ Specialists (data scientists) building their pipelines: design, inspect, compare
- ❑ Non-expert users using a pipeline (journalists, collaboration with RadioFrance)



Corpus and some results

Corpus

$218K$ claim reviews in 14 languages, from 83 fact-checking agencies recognized by IFCN (International Fact-Checking Network).
Include links to FC articles and sometimes tweets (*)

❑ Narrowed to EN, FR

$9.1K$ tweets mentioned in FC articles + $8K$ recent tweets from prominent Members of the European Parliament ($4.7K$ EN, $1.2K$ FR)

→ $9.1K$ aligned pairs of (social media posts, FC articles)

Insights

Article retrieval (P2A) worked better than claim retrieval (P2C)

Also very precise → we can afford to retrieve very few articles
→ less effort to rank them

Augmenting tweets with generated image captions and/or OCR

On our EN corpus, OCR alone is best.
On others, OCR+caption is best…

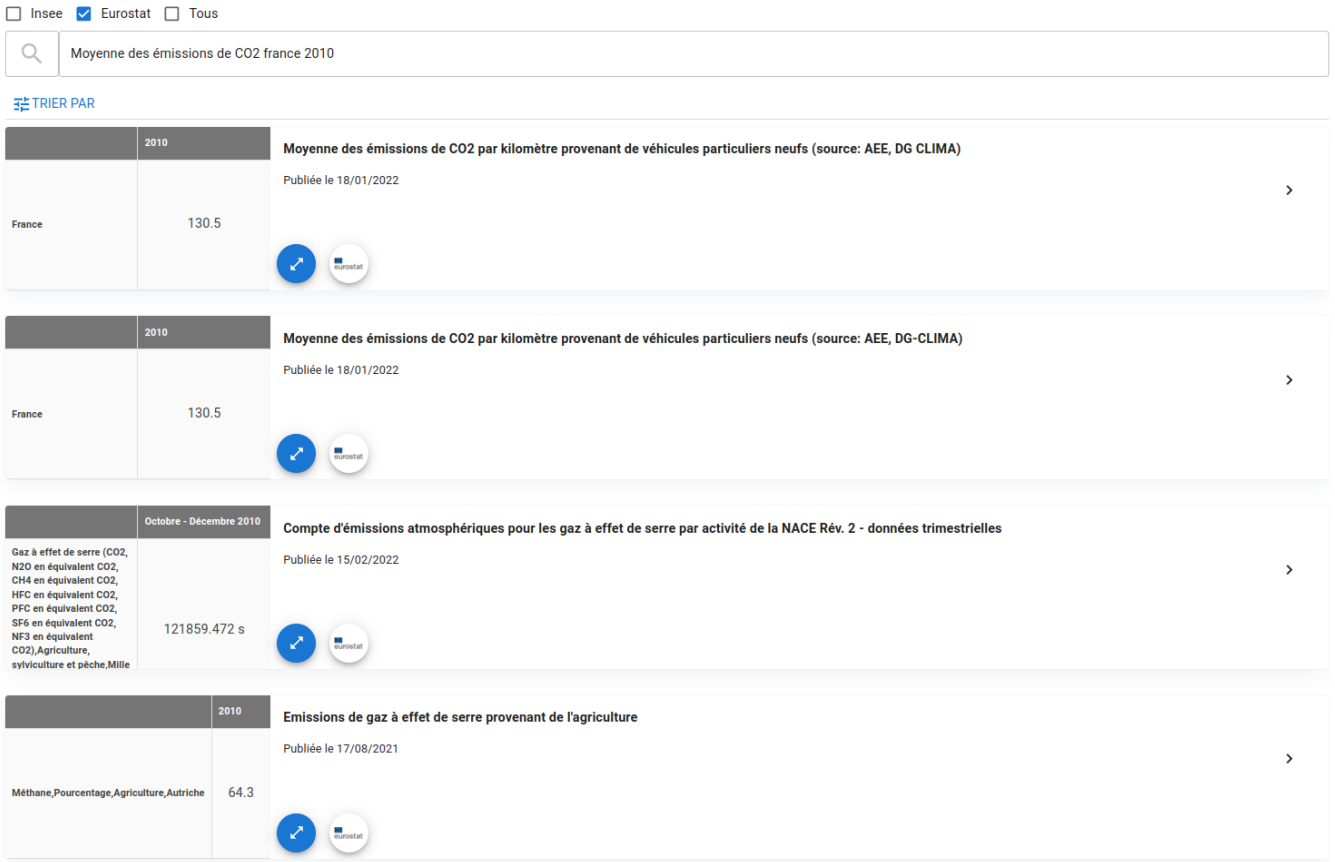
Fact-checking and analyzing debates

Heterogeneous statistic datasets (INSEE, Eurostat, etc.) contain valuable **insights**

Extracting and fact-checking statistic claims from social media [CIKM22, TTO22, CIKM24]

anr° Inria

Le Monde radiofrance



Propaganda detection

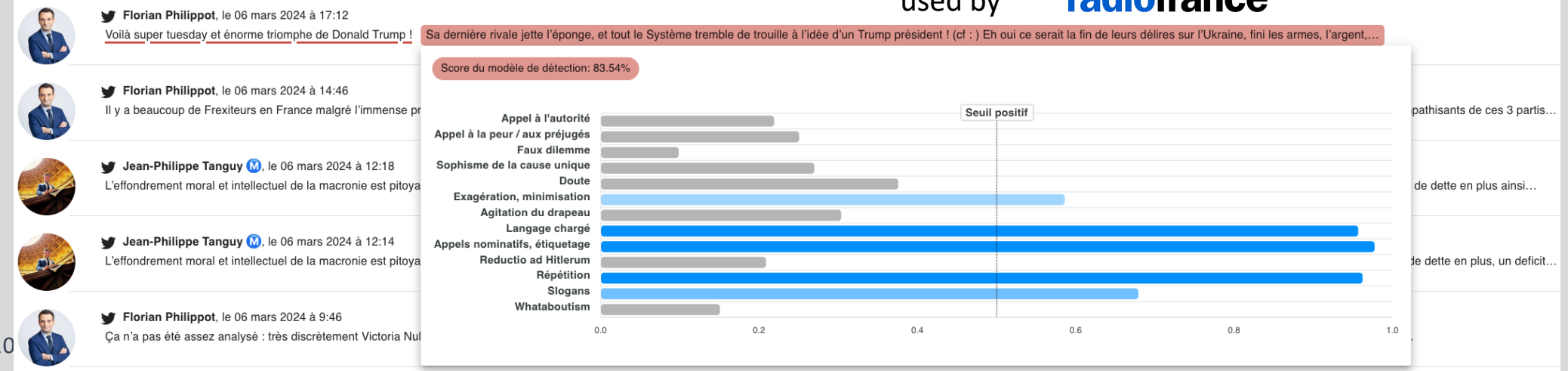
Public debates:

- Arguments (factual or not)
- Reasoning (valid or not)
- Various forms of propaganda

Propaganda benchmark from social networks (sub-Reddit) [EACL2021]

Propaganda detection

Deployed in StatCheck, used by radiofrance



Political advertising online

- ❑ + Communicate policy, engage with voters
- ❑ -- Risks: user targeting, spreading propaganda, disinformation

EU Digital Services Act: on platforms >= 45M users

- ❑ Political ads must be **declared**
- ❑ Researchers, NGOs **must have access** to the ads to analyze them

First **AI analysis of 400.000 political ads in 2022 in France** on Meta [WebConf23]

Expertise and advice for EU and FR institutions

erc

En raison de l'influence de TikTok, les juges roumains annulent la présidentielle

Le candidat prorusse arrivé en tête du premier tour du scrutin a dénoncé « un coup d'Eta

Par Mirel Bran (Bucarest, correspondant)
Publié le 07 décembre 2024 à 04h30, modifié le 07 décembre 2024 à 10h18 · Lecture 4 min. · Read in English

Offrir l'article

Article réservé aux abonnés

Inria

Analysis of political ads in France, 2022

	Gender		Age						
	Female	Male	13-17	18-24	25-34	35-44	45-54	55-64	65
Population* (baseline)	53.94%	46.06%	2.68%	14.24%	22.79%	18.33%	15.14%	13.12%	13.52%
International affairs	53.64%	46.36%	0.18%	6.26%	22.22%	19.83%	16.64%	15.89%	18.98%
Energy	35.61%	64.39%	0.01%	1.74%	25.76%	34.49%	27.35%	7.88%	2.77%
Immigration	65.65%	33.35%	0.18%	4.6%	16.68%	17.06%	18.88%	20.57%	22.02%
Law and crime	51.88%	48.12%	28.57%	24.52%	8.46%	10.97%	9.3%	8.33%	9.85%
Government operations	53.35%	46.65%	0.92%	30.26%	27.32	13.83%	9.7%	8.7%	9.23%
Cultural policy	51.53%	48.47%	3.17%	16.18%	23.32%	17.97%	15.63%	12.21%	11.52%
Social policy	65.34%	34.66%	0.93%	13.65%	19.74%	16.27%	14.51%	16.15%	18.74%
Education	59.19%	40.81%	11.4%	24.2%	16.6%	13.8%	9.67%	10.26%	14.79%
Environment	50.89%	49.11%	1.95%	10.78%	25.52%	21.91%	17.04%	12.17%	10.64%
Health	68.45%	31.55%	4.16%	8.62%	14.18%	18.12%	17.12%	18.42%	19.37%
Economy	44.24%	55.76%	0.01%	12.85%	22.25%	19.36%	15.67%	15.53%	14.34%
Human rights	59.31%	40.69%	8.74%	16.18%	20.09%	16.52%	13.54%	12 %	12.13%
Work and employment	61.27%	38.73%	0.84%	27.49%	19.03%	12.60%	14.79%	12.88%	12.37%
Urban and territorial policy	51.09%	48.91%	3.63%	6.32%	14.62%	15.90%	15.60%	17.65%	26.29%

Over-exposed to a topic

Under-exposed to a topic

Overview of our research

Big Data Management and Analytics

Exploring and querying RDF graphs

Exploration of large datasets

Taming heterogeneous data

Optimization for cloud data analytics

Data stream analytics

Statistic fact-checking

Digital Democracy

Safe online political advertising & marketing

Uncovering exposure and (mis)information on social media

Natural Language Processing

Information extraction and text generation

Argumentation mining

More information:
https://team.inria.fr/cedar/
https://team.inria.fr/cedar/publications/