

PAN Task on **Oppositional thinking analysis:** **Conspiracy theories vs critical thinking narratives**

Damir Korenčić, BERTa Chulvi, Xavier Bonet, Mariona Taulé,
Francisco Rangel, Paolo Rosso



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



UNIVERSITAT DE
BARCELONA

symanto
psychology ai

Grenoble, 11th September 2024

Information warfare:

Foreign information manipulation interference

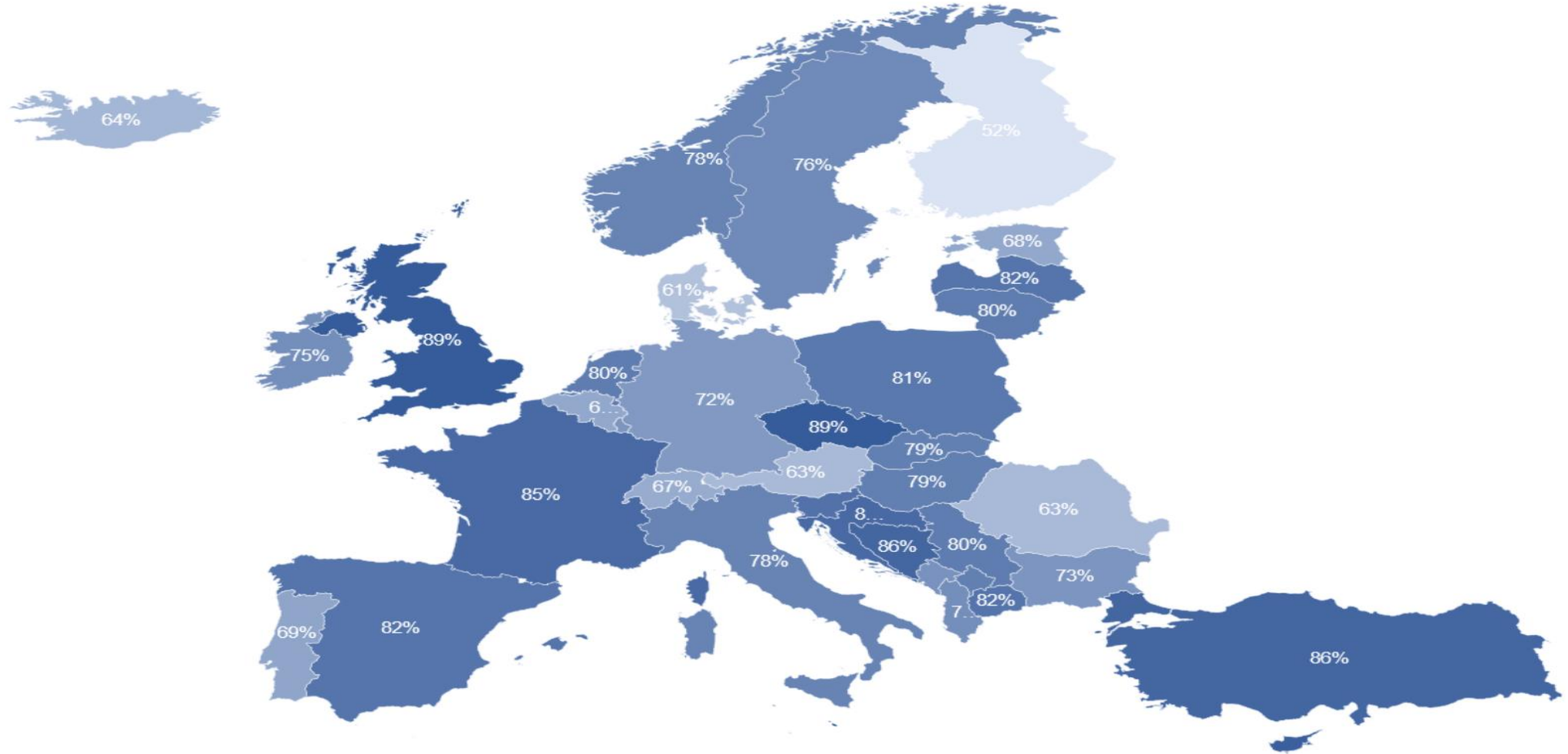
- **European External Action Service (EEAS)**
- Early detection of **Tactics, Techniques, and Procedures (TTP)**
- Indicators: **harmful, not illegal, manipulative, intentional, coordinated**
- **Action plan for the European democracy** by the EC (December 2020): aim is to **fight against disinformation**
- **DISARM Framework**: aim is the standardization of the model of the analysis of the **behaviours in disinformation attacks (detection, analysis, answers)**
- Threats of **manipulation of information and interference from foreign agents**
- Monitoring 616 channels: 40% linked to Russian and Chinese media (Russian's attempt to **create polarization in Western democracies**)

<https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:52020DC0790>

https://www.eeas.europa.eu/eeas/tackling-disinformation-foreign-information-manipulation-interference_en

https://www.eeas.europa.eu/eeas/1st-eeas-report-foreign-information-manipulation-and-interference-threats_en

The perception of disinformation as a problem in Europe



Badillo-Matos A. et al. (2023). **Analysis of the Impact of Disinformation on Political, Economic, Social and Security Issues, Governance Models and Good Practices: The cases of Spain and Portugal.** IBERIFIER, Pamplona.



IBERIFIER

Iberian Digital Media
Observatory

European Digital Media Observatory

Project IBERIFIER Plus - 101158511
co-funded by the EC under the call
DIGITAL-2023-DEPLOY-04

National and multinational hubs

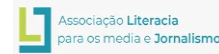


COORDINATOR



Universidad
de Navarra | FACULTAD DE
COMUNICACIÓN

PARTNERS




EDMO Task Force on 2024 EU Elections

EDMO Task Force On 2024 European Elections



EDMO Task Force on 2024 EU Elections

 **Several false narratives about climate change are on the scene.** In recent days, a remarkable assortment of false narratives about climate change has resurfaced in various countries. False stories range from usual denialism to conspiracy theories (e.g. chemtrails), from painting measures to address the climate crisis as unfair to attacking climate activists, politicians and supporters as hypocrites or Nazis. These false stories were detected in: PL, DE, HR, EL, ES.

Conspiracy theories



UNIVERSITY OF
BIRMINGHAM

Search

Home > News

Finding order in a chaotic world: Understanding the Trump assassination conspiracy theories

Following the attempted assassination of Donald Trump, Professor Lisa Bortolotti explains how and why conspiracy theories thrive in the wake of shocking events.

17 July 2024 • 5 min read

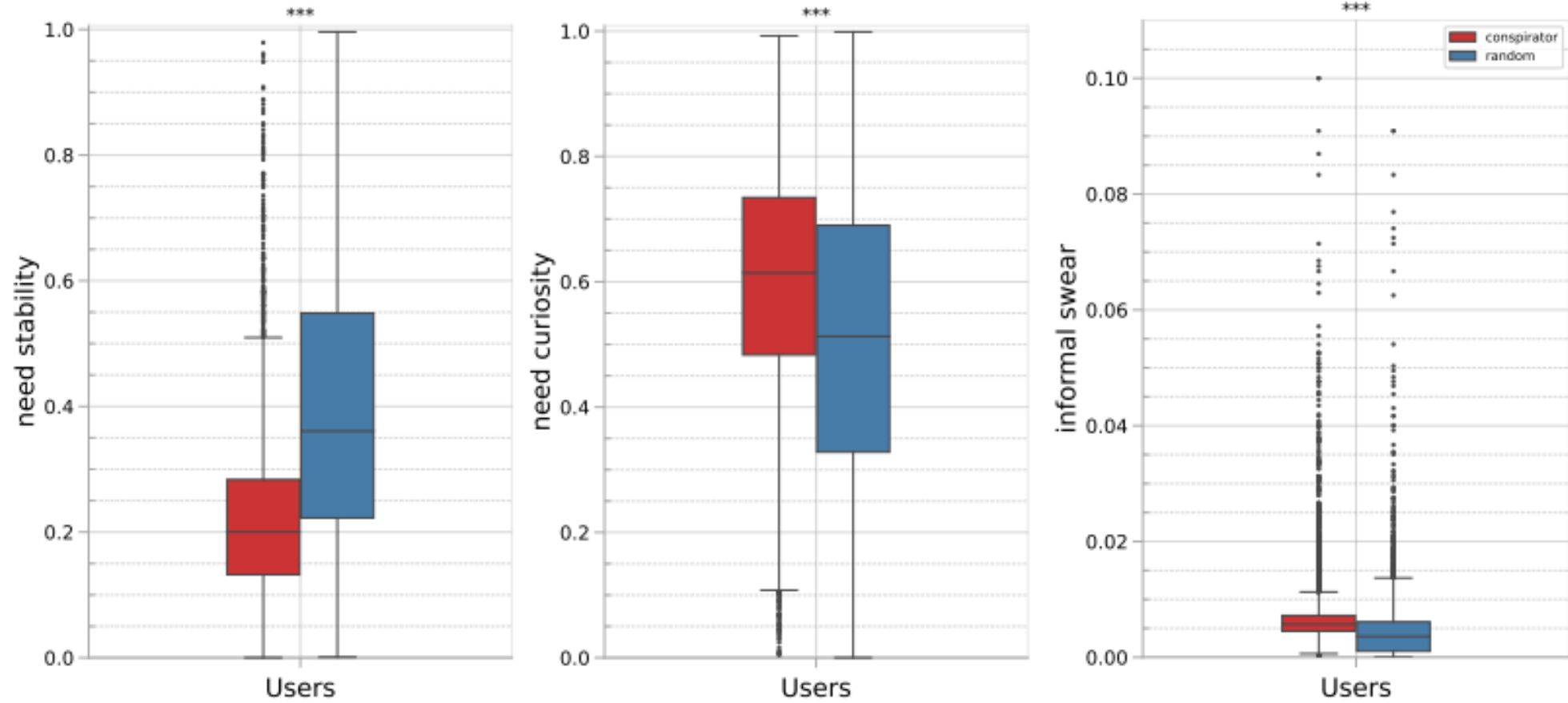
Conspiracy theories

- Causal explanations of significant events that present them as a **result of cover plots orchestrated by secret powerful and malicious groups**
- Greater use of **violent words** and the **emotional manifestation of anger**
- **COVID-19**
- **Qanon** (satanism and pedophilia)
- **Green tyranny** (cut rights and freedoms)
- **Great replacement** (Christian by Muslims)
- **Great reset** (plan of control of the population)
- **Secret Jewish plan**
- In **2M Reddit comments** only **5% users** were **conspiracy theorists** but they wrote **64% of comments: the most active user wrote 896k words = the double of The Lord of the Rings** (UNESCO, 2020)

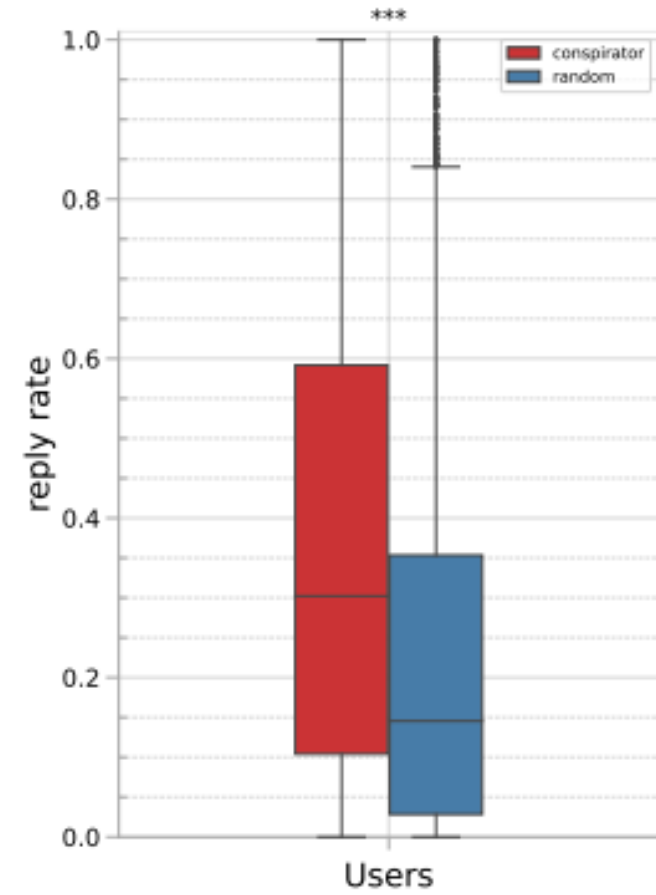
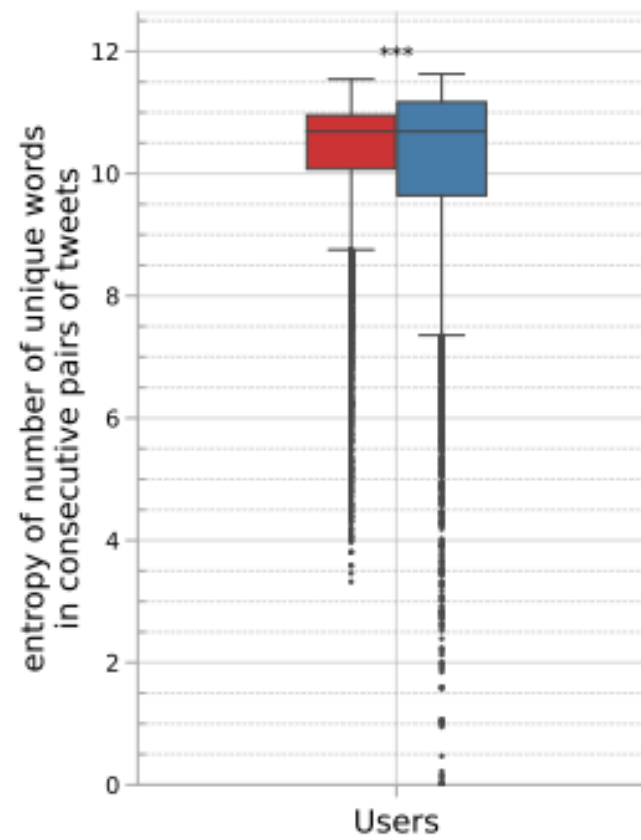
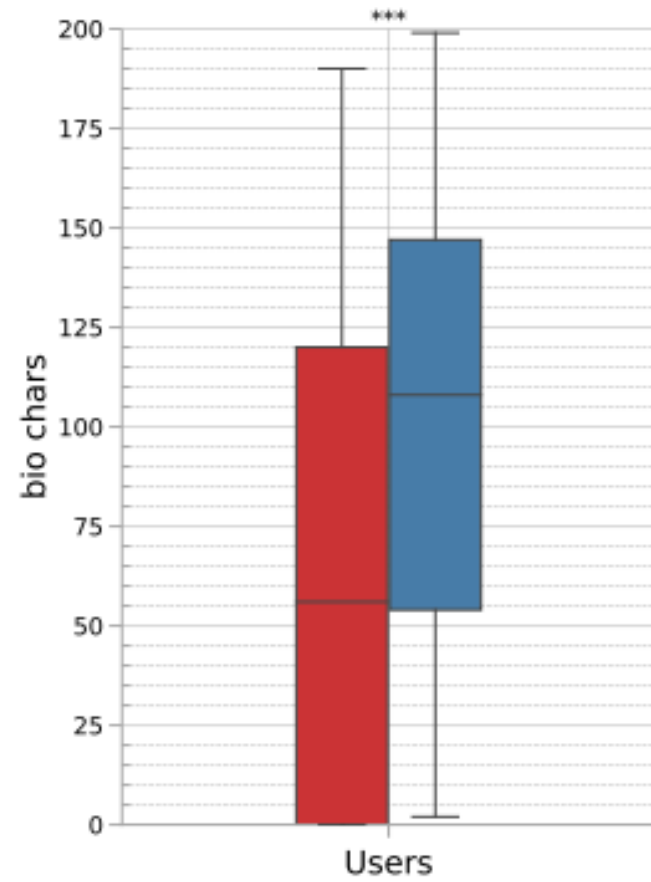
Psycholinguistics features

- **Emotions:** Plutchik's 8 emotional categories (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust)
- **Sentiment** (polarity: positive, negative)
- **Personality traits** via **IBM Personality Insights: Big 5 Traits (OCEAN), 5 Values** (conservation, hedonism, openness to change, self-enhancement, self-transcendence), and **12 Needs** (challenge, closeness, curiosity, excitement, harmony, ideal, liberty, love, practicality, self-expression, stability, structure)
- **Linguistic patterns** via **LIWC tool** (pronouns, personal concerns, time focus, cognitive processes, informal language, affective processes)

Psycholinguistics features



Other features



Tackling COVID-19 conspiracy on Twitter

- Shared task at MediaEval 2022
- **Twitter** data: scraping, keyword-filtering, cleaning, annotation
- User graph: nodes are users, edges are user-user interactions
- **Text-based detection** of conspiracy theories (task 1)
- **Graph-based conspiracy spreader detection** (task 2)
- **Conspiracy categories:** suppressed cures, behaviour and mind control, antivax, fake virus, intentional pandemic, harmful radiation or influence, population reduction, new world order, and satanism
- Text-conspiracy relation: support, mention, no-mention

<https://github.com/konstapo/2022-Fake-News-MediaEval-Task>

Langguth J., Schroeder D.T., Filkuková P., Brenner S., Phillips J., Pogorelov K. (2023).

COCO: An Annotated Twitter Dataset of COVID-19 Conspiracy Theories. Journal of Computational Social Science.

Taxonomies on conspiracy theories

- **Focus:**

- **outsiders vs insiders** (exogroup vs endogroup) as **friend/enemy schema**
- **Social Identity Theory** that gives to the individual a social identity and a **sense of belonging**

- **Drawbacks:**

- it mixes **actions** and **actors**, i.e. groups of people (**social categories**): an event (e.g. AIDS) may provoke the *action* of a *social group*
- **actors** with **consequences** and **objectives** (labelled the three of them with just one label: **insiders**)
- mixing **actors and actions** cannot capture an intergroupal conflict, just friend/enemy schema

Conspiracy narrative vs critical thinking

- Importance of not confuse **critical and conspiracy narratives: high risk of pushing people towards conspiracy communities**
- **Social psychologist** on board and **linguists** for the annotation
- **"Us vs them"** narrative
- **Insiders** include **campaigners** and **victims**
- **Outsiders** include **agents** and **facilitators**
- Categories at **span level**
- **Domain-agnostic**: it could be applied to other conspiracy theories (e.g. climate change)

Conspiracy narrative vs critical thinking

- **Agents (A)**
- **Objectives (O)**
- **Consequences (E)**
- **Victims (V)**
- **Campaigners (C):** those who oppose the mainstream narrative
- **Facilitators (F):** collaborators with conspiracy propagators (conspiracy narrative) vs implementing measures dictated by the authorities (critical thinking)

Oppositional narrative



Conspiracy Theory

Private owned WHO A with investors like Bill Gates A can declare a new pandemic out of thin air anytime they want and the world governments ruled by their puppets F as well as their media F starts with the constant fear mongering E , getting people V to get their pharma companies A injections and drugs that are magically ready in light speed, clear induction that they have been ready for the orchestrated fake pandemics, long before they start with the constant fear mongering E by the media F and governments F . To those awake already C , we know their games and agenda O , but sadly most people V fall for it, again and again and pay a hefty price, often with their health, lives, the loss of their loved ones E . These are very evil beings A , intent on destroying us O regular people V .

Critical Thinking

<https://twitter.com/...> Hospitals Should Hire , Not Fire , Nurses with Natural Immunity by Dr Martin Kulldorff C By pushing vaccine mandates O , White House chief medical advisor Dr. Anthony Fauci A is questioning the existence of natural immunity after Covid disease . In doing so , he is following the lead of CDC director Rochelle Walensky , who questioned natural immunity A in a 2020 Memorandum published by The Lancet . By instituting vaccine mandates , university hospitals F are now also questioning the existence of natural immunity after Covid disease . This is astonishing . I work at Brigham and Women 's Hospital in Boston , which has announced that all nurses , doctors and other health care providers V will be fired if they do not get a Covid vaccine E . Last week I spoke with one of our nurses . She worked hard caring for Covid patients , even as some of her colleagues left in fear at the beginning of the pandemic . Unsurprisingly , she got infected , but then recovered . Now she has stronger and longer - lasting immunity than the vaccinated work - from - home hospital administrators who are firing her for not being vaccinated F . If university hospitals can not get the medical evidence right on the basic science of immunity , how can we trust them with any other aspects of our health ?

Oppositional thinking analysis: Conspiracy theories vs critical thinking narratives

- **XAI-DisInformation dataset:** 10k messages of Telegram in  
- Oppositional non-mainstream views on the **COVID-19 pandemic**
- 1st task: **conspiracy theories vs critical thinking narratives (Matthew's correlation coefficient)**
- 2nd task: **text-span recognition of elements of oppositional narratives (macro-averaged span-F1)**
- **82 teams** participated



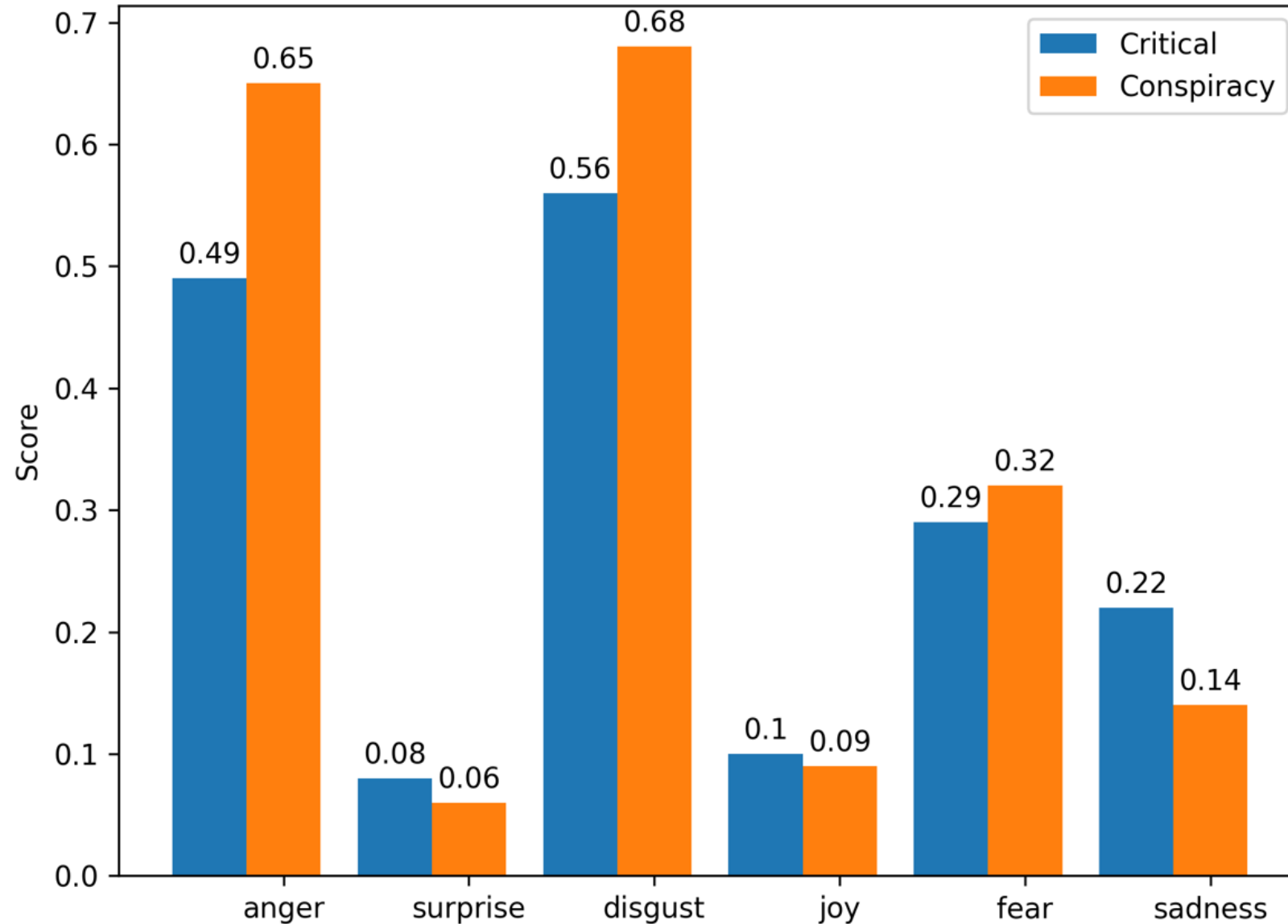
XAI-DisInfodemics:
eXplainable AI for
disinformation and conspiracy
detection during infodemics
(PLEC2021-007681)

Korenčić D., Chulvi B., Bonet X., Taulé M., Rosso P., Rangel F. (2024).

Overview of the Oppositional Thinking Analysis PAN Task at CLEF 2024.

Working Notes of CLEF 2024 – Conference and Labs of the Evaluation Forum.

Conspiracy theorists vs critical thinkers: emotions



cardiffnlp/twitter-roberta-base-emotion-latest

Higher use of **hurtful words** by conspiracy theorists (**HurtLex**)

Bassignana E., Basile V., Patti V. (2018). **Hurtlex: A Multilingual Lexicon of Words to Hurt**. In Proc. of the Fifth Italian Conf. on Computational Linguistics (CLiC-It 2018)

Results (above transformer baselines): task 1

English		Spanish	
TEAM	MCC	TEAM	MCC
IUCL [56]	0.8388	→ SINAI [41]	0.7429
AI_Fusion	0.8303	auxR	0.7205
→ SINAI [41]	0.8297	RD-IA-FUN [40]	0.7028
ezio [44]	0.8212	Elias&Sergio	0.6971
hinlole [53]	0.8198	AI_Fusion	0.6872
Zleon [48]	0.8195	zhengqiaozeng [52]	0.6871
virmel	0.8192	virmel	0.6854
inaki [47]	0.8149	trustno1	0.6848
yeste	0.8124	Zleon [48]	0.6826
auxR	0.8088	ojo-bes	0.6817
Elias&Sergio	0.8034	tulbure [54]	0.6722
theateam	0.8031	sail [50]	0.6719
trustno1	0.7983	nlpln [55]	0.6681
→ DSVS [46]	0.7970	<i>baseline-BETO</i>	<i>0.6681</i>
ojo-bes	0.7969		
sail [50]	0.7969		
RD-IA-FUN [40]	0.7965		
<i>baseline-BERT</i>	<i>0.7964</i>		

- **IUCL**: DeBERTa fine-tuned with a subset of 16k of the LOCO dataset
- **SINAI**: fine-tuned GPT-3.5 turbo (ES), fine-tuned LLaMA (EN)
- **DSVS**: Ensemble of Transformers
- **TU-Berlin**: LLM-generated text's context and argumentation

Miani A., Hills T., Bangerter A. (2021). **LOCO: The 88-million-word Language Of Conspiracy Corpus**, Behavior research methods, pp. 1–24.

Results (above transformer baselines): task 2

English TEAM	span-F1	Spanish TEAM	span-F1
→ tulbure [54]	0.6279	→ tulbure [54]	0.6129
Zleon [48]	0.6089	Zleon [48]	0.5875
hinlole [53]	0.5886	AI_Fusion	0.5777
oppositional_opposition	0.5866	virmel	0.5616
AI_Fusion	0.5805	CHEEXIST	0.5621
virmel	0.5742	miqarn	0.5603
miqarn	0.5739	DSVS [46]	0.5529
TargaMarhuenda	0.5701	TargaMarhuenda	0.5364
ezio [44]	0.5694	Elias&Sergio	0.5151
zhengqiaozeng [52]	0.5666	hinlole [53]	0.4994
Elias&Sergio	0.5627	<i>baseline-BETO</i>	<i>0.4934</i>
DSVS [46]	0.5598		
CHEEXIST	0.5524		
rfenthusiasts	0.5479		
ALC-UPV-JD-2	0.5377		
<i>baseline-BERT</i>	<i>0.5323</i>		

- **tulbure**: RoBERTa + data augmentation replacing words in the text-spans by synonyms or semantically-related words
- **Zleon**: several transformers + multi-task learning

Brandolini's law, aka the bullshit asymmetry principle (2013):

The amount of energy needed to refute something stupid is greater than that needed to generate it.

(but we have to do it, manually or with the help of AI LLMs)

proso@dsic.upv.es