CLEF 2024 GRENOBLE

25 years of Cross Language / Conference and Labs Evaluation Forum

2000 Lisbon

2001 Darmstadt

2002 Rome

2003 Trondheim

2004 Bath

2005 Vienna

2006 Alicante

2007 Budapest

2008 Aarhus

2009 Corfu

CheckThat! 2024

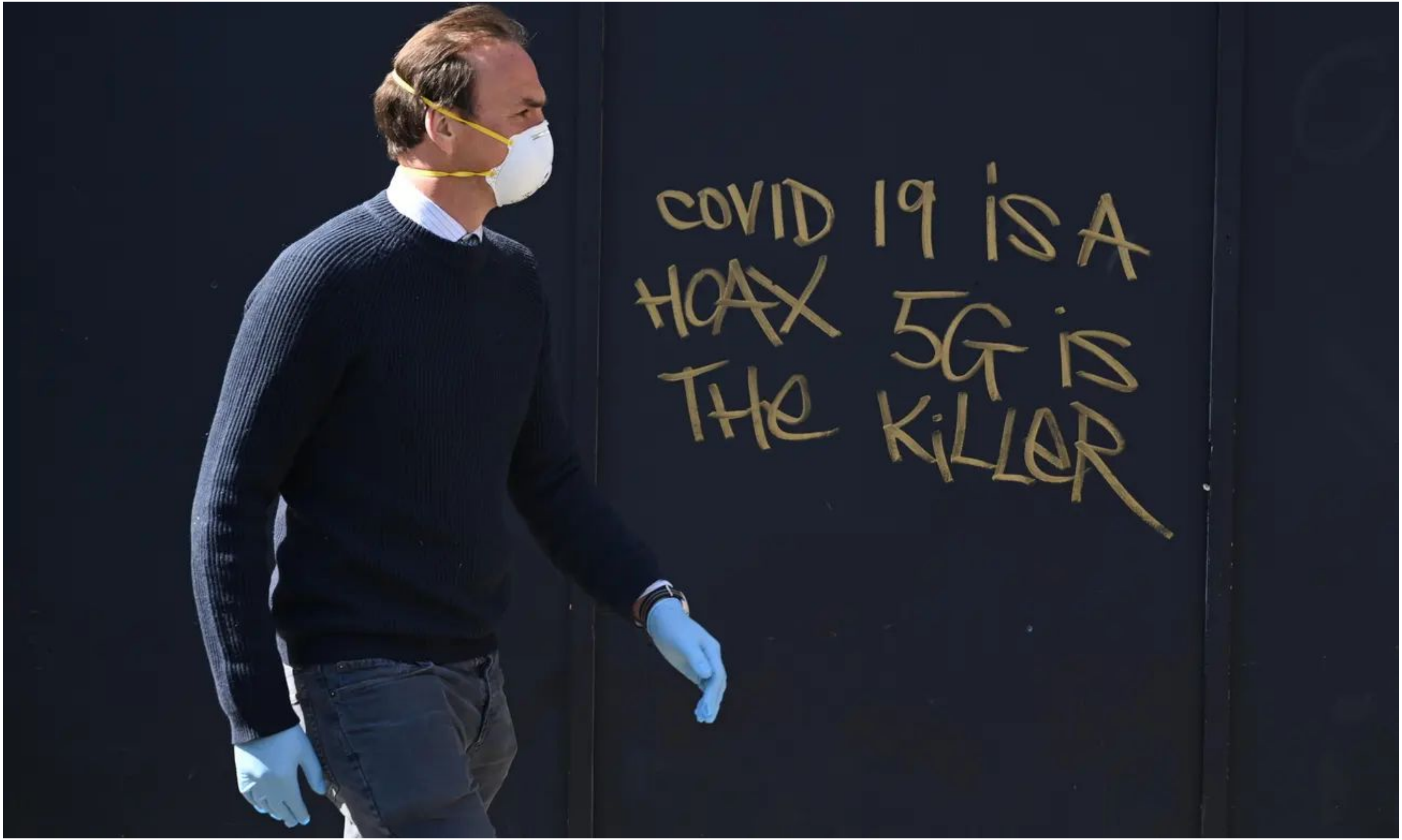
7th edition

Check-Worthiness, Subjectivity, Persuasion, Roles, Authorities, and Adversarial Robustness

http://checkthat.gitlab.io

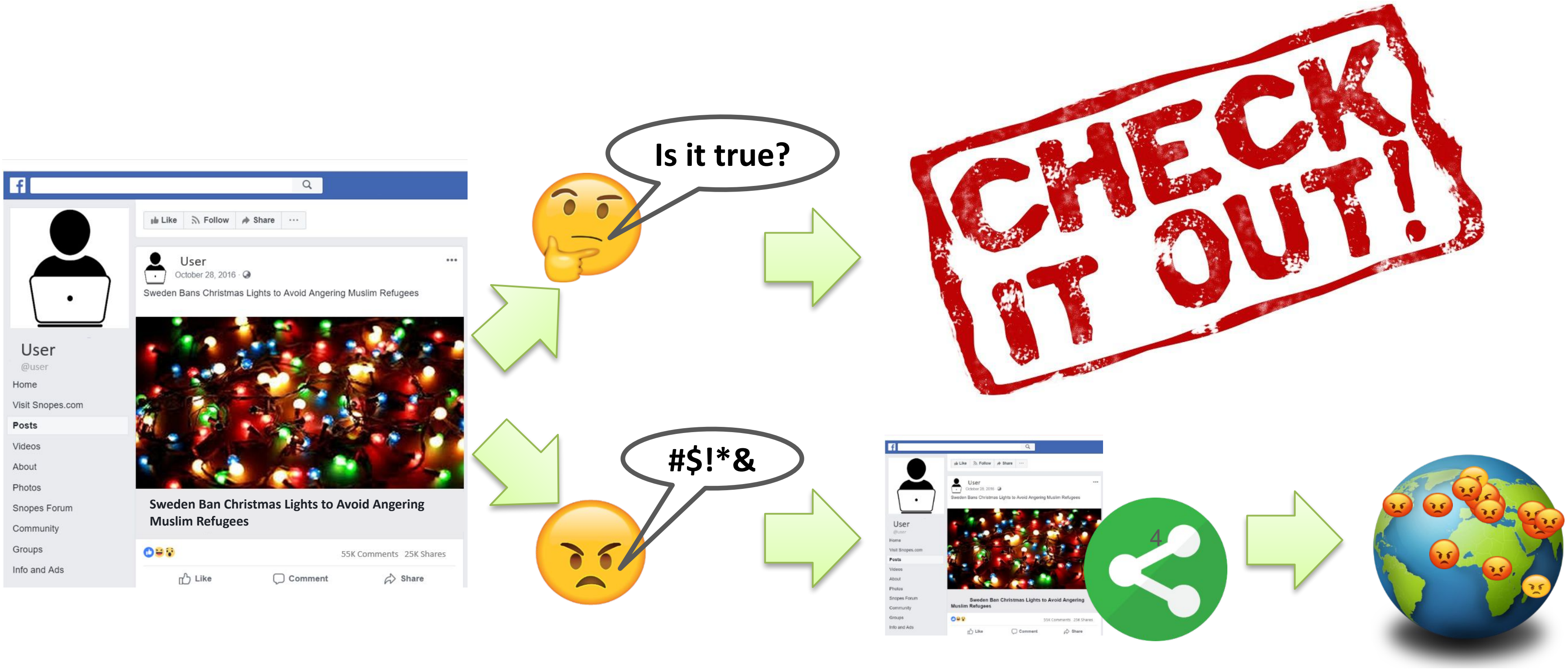
CLEF 2024 Extended Lab Overview

10th September, 2024 Grenoble, France



Photograph: Neil Hall/EPA https://www.theguardian.com/world/2020/apr/26/5g-coronavirus-and-contagious-superstition

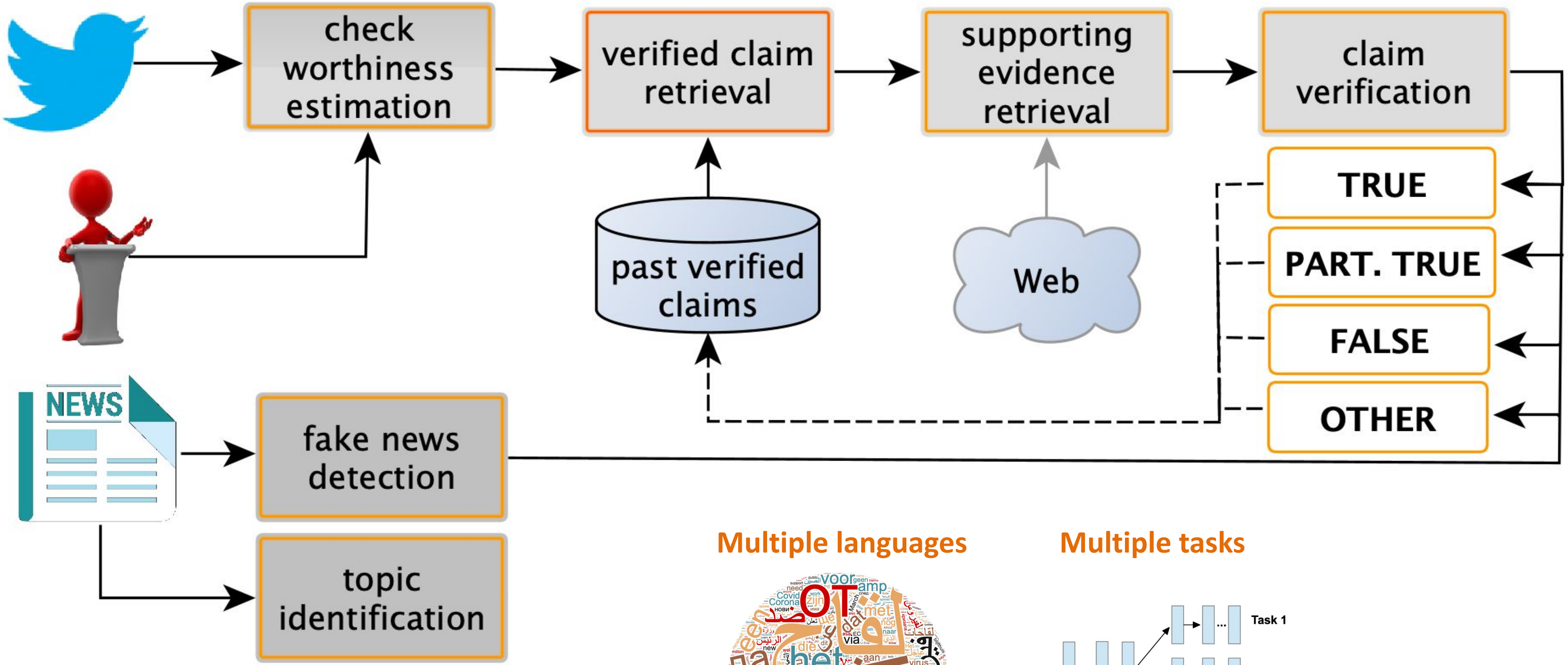
How?



Like
Follow
Share
User
October 28, 2016
Sweden Bans Christmas Lights to Avoid Angering Muslim Refugees
User
@user
Home
Visit Snopes.com
Posts
Videos
About
Photos
Snopes Forum
Community
Groups
Info and Ads
Sweden Ban Christmas Lights to Avoid Angering Muslim Refugees
55K Comments 25K Shares
Like
Comment
Share
Is it true?
CHECK IT OUT!
#$!*&

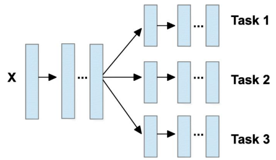
The CheckThat! Lab @ CLEF

The CheckThat! Lab (2018-2022) in a Nutshell



Multiple languages

Multiple tasks



Participation

2018 Tasks	Teams	Runs	Papers
1. Check-worthiness	7	21	5
2. Fact-checking	5	14	4
Total	**9**	**35**	**8**

2019 Tasks	Teams	Runs	Papers
1. Check-worthiness	12	21	8
2. Evidence & Factuality	4	36	4
Total	**14**	**57**	**12**

2020 Tasks	Teams	Runs	Papers
1. Check-worthiness	15	54	
2. Verified claim retrieval	9	20	
3. Evidence retrieval	1	2	
4. Claim verification	1	2	
Total	**23**	**86**	**16**

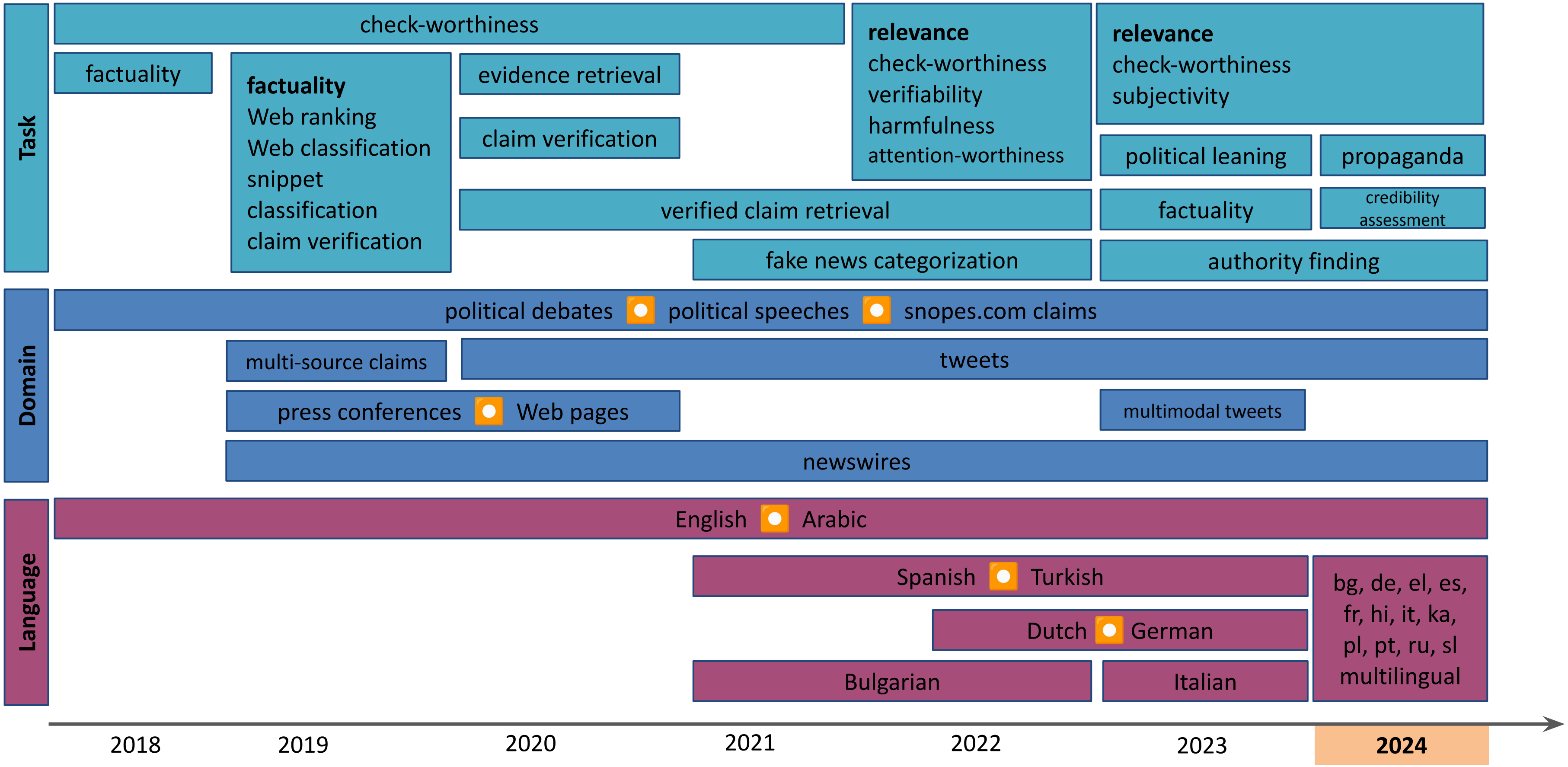
2021 Tasks	Teams	Runs	Papers
1. Check-worthiness	15	74	10
2. Verified claim retrieval	5	16	4
3. Fake news detection	27	139	13
Total	47	229	**27**

2022 Tasks	Teams	Runs	Papers
1. Check-worthiness	18	210	13
2. Verified claim retrieval	7	37	3
3. Fake news detection	26	126	15
Total	**51**	**373**	**31**

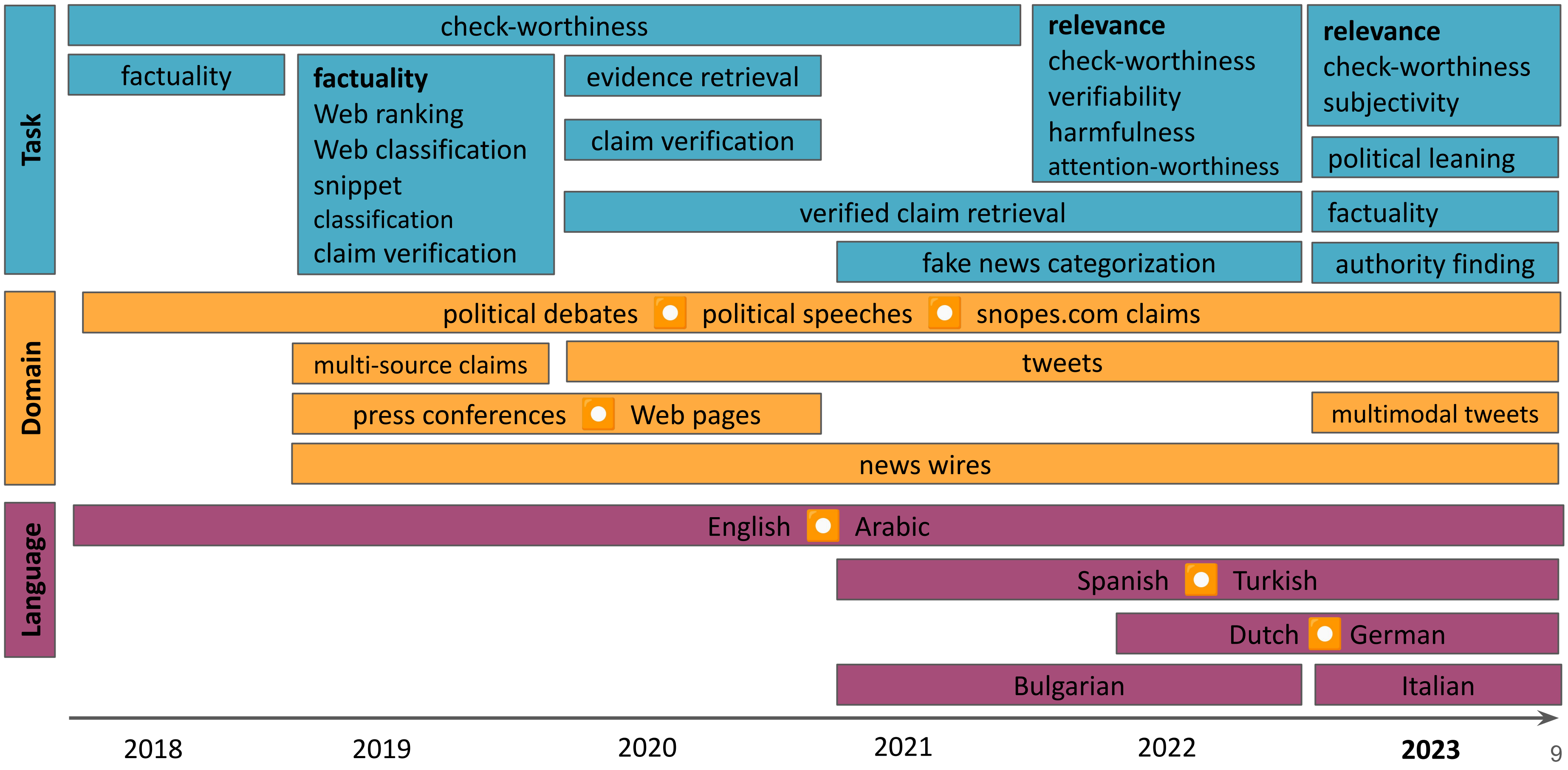
2023 Tasks	Teams	Runs	Papers
1. Check-worthiness	19	155	12
2. Subjectivity	12	88	10
3. Bias	6	41	4
4. Factuality	6	28	4
5. Authority	2	4	1
Total	**45**	**316**	**31**

2024 Tasks	Teams	Runs	Papers
1. Check-worthiness	28	236	19
2. Subjectivity	15	113	11
3. Persuasion Techniques	2	-	2
4. Hero, villain, and victim	-	-	-
5. Authority	5	16	3
6. Adversarial Robustness	6	6	6
Total	**46**	**294**	**36**

The CLEF CheckThat! Lab:Tasks, Lang & Data



Our Main Focus in 2018-2023



The Verification Pipeline and 2024 Tasks

T1 Check-worthiness estimation

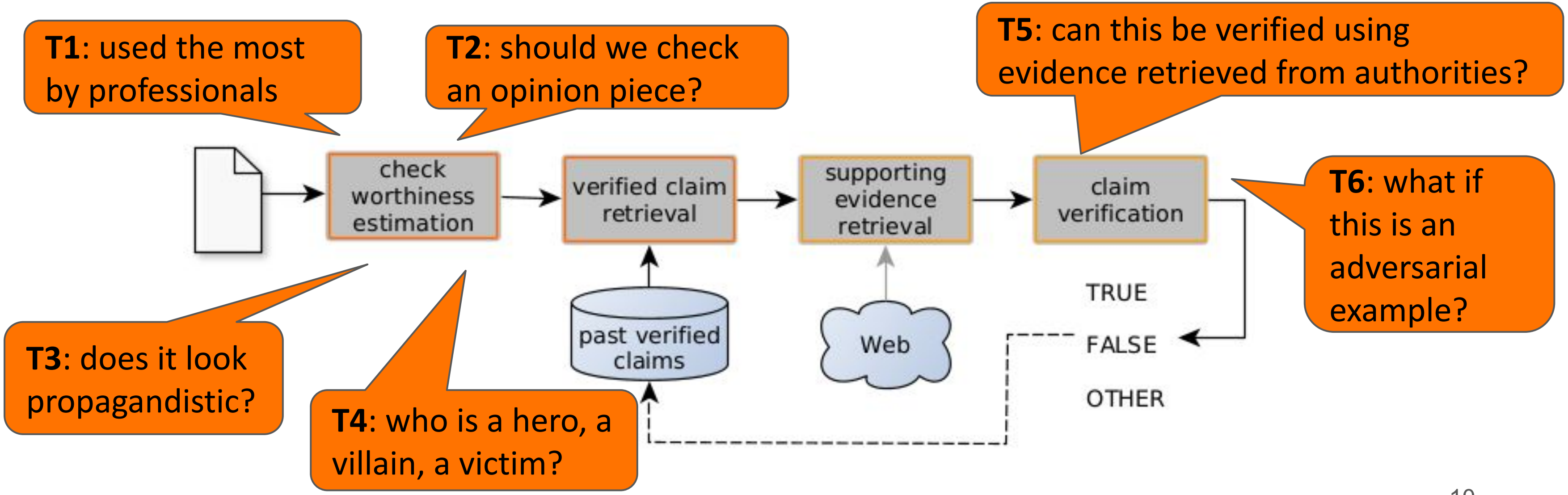
T2 Subjectivity in news

T3 Persuasion in news

T4 Hero, villain, and victim in memes

T5 Rumor Verification using evidence from authorities

T6 Robustness of Credibility Assessment with Adversarial Examples



Task 1: Check-Worthiness Estimation of Multigenre Content

Motivation

By Robert Farley

Posted on August 17, 2023 | Corrected on December 13, 2023

In recent speeches touting so-called Bidenomics, President Joe Biden has repeatedly cited the statistic that "unemployment has been below 4% for the longest stretch in over 50 years."

Olympic Boxer Imane Khelif Featured on Vogue Algeria Front Covers

بدأت مشاهد المذبحة الكبرى بالوصول
ليلة البارحة قطع الاحتلال الإسرائيلي الإنترنت بالكامل وبدأ بقصف هستيري حتى أن المتحدث باسم جيشه صرح بأنه "يهاجم غزة بقوة عظيمة"
وسائل الإعلام التقليدية غير قادرة على نقل الصورة بالوضوح الذي تنقله هواتف الناشطين
الآن بدأت مشاهد مذبحة البارحة تصل
الأرقام تتحدث عن حوالي 500 شهيد في ليلة واحدة وكما يظهر في الفيديوهات معظم المصابين من الأطفال
#GazaGenocide
#GazaBleedingWorldSleeping

Translated from Arabic by Google

Scenes of great massacre began to arrive
Last night, the Israeli occupation cut off the entire Internet and began hysterical bombing. Its army spokesman even declared that it was "attacking Gaza with great force."
Traditional media are unable to transmit the image with the clarity that activists' phones transmit
Now the scenes of yesterday's massacre are starting to arrive
The numbers speak of about 500 martyrs in one night, and as shown in the videos, most of the injured were children
#GazaGenocide
#GazaBleedingWorldSleeping

Task Description

Which asks to detect whether a given text snippet from multigenre content, in a form of **a tweet or a sentence from a political debate or speech**, is **worth fact-checking**.

Transcription

ar nl en es

Data

Training, development and dev-test subsets for the 2024 edition by re-using all the data released in 2023

Test Sets:

- **Arabic:** Tweets using keywords relevant to the war on Gaza, that started in October 2023.
- **Dutch:** $1k$ messages between January 2021 and December 2022 on climate change and its associated debate
- **English:** Transcribed sentences that did not appear in Arslan et al. (2020)
- **Spanish:** No test set

ar nl en es

Data

Data Splits	Arabic		Dutch		English		Spanish	
	Yes	No	Yes	No	Yes	No	Yes	No
Train	2,243	5,090	405	590	5,413	17,087	3,128	16,862
Dev	411	682	102	150	238	794	704	4,296
Dev-test	377	123	316	350	108	210	509	4,491
Test	218	392	397	603	88	253	-	-
Total	3,249	6,287	1,220	1,693	5,847	18,344	4,341	25,649

Results

Arabic			Dutch			English		
	Team	F1		Team	F1		Team	F1
1	IAI Group	0.569	1	TurQUaz	0.732	1	FactFinders	0.802
2	OpenFact	0.557	2	DSHacker	0.730	2	OpenFact	0.796
3	DSHacker	0.538	3	IAI Group	0.718	3	Fraunhofer SIT	0.780
4	TurQUaz	0.533	4	Mirela	0.650	4	mjmanas54	0.778
5	SemanticCuetSync	0.532	5	Zamoranesis	0.601	5	ZHAW_Students	0.771
6	mjmanas54	0.531	6	FC_RUG	0.594	6	SemanticCuetSync	0.763
7	Fired_from_NLP	0.530	7	OpenFact	0.590	7	SINAI	0.761
8	Madussree	0.530	8	HYBRINFOX	0.589	8	DSHacker	0.760
9	pandas	0.520	9	mjmanas54	0.577	9	IAI Group	0.753
10	HYBRINFOX	0.519	10	DataBees	0.563	10	Fired_from_NLP	0.745
11	Mirela	0.478	11	JUNLP	0.550	11	TurQUaz	0.718
12	DataBees	0.460	12	Fired_from_NLP	0.543	12	HYBRINFOX	0.711
13	Baseline	0.418	13	Madussree	0.482	13	SSN-NLP	0.706
14	JUNLP	0.212	14	Baseline	0.438	14	Checker Hacker	0.696
			15	pandas	0.308	15	NapierNLP	0.675
			16	SemanticCuetSync	0.218	16	Mirela	0.658
						18	DataBees	0.619
						19	Trio_Titans	0.600
						20	Madussree	0.583
						21	pandas	0.579
						22	JUNLP	0.541
						23	Sinai and UG	0.517
						24	grig95	0.497
						25	CLaC	0.494
						26	Aqua_Wave	0.339
						27	Baseline	0.307

Approaches

- Transformers were most popular.
- Monolingual and multilingual transformers
- Several teams used LLMs: LLaMA, Mistral, Mixtral, and GPT

Team	Language			Models																			Misc			
	Arabic	Dutch	English	LLama2	LLama 3	Mixtral	Mistral	GEITje	GPT-3.5	GPT-4	Gemini	BERT	RoBERTa	BERTweet	XLM-r	ALBERT	DistilBERT	DeBERTa	Electra	AraBERT	BERTje	GPT-3	Data aug	Preprocessing	Data Pruning	Info. Extraction
Aqua_Wave [10]			26									✓	✓										✓			
Checker Hacker [14]			14																				✓			
CLaC [29]			25								✓												✓			
DataBees [81]	12	10	18									✓	✓		✓	✓	✓			✓	✓			✓		
DSHacker [28]	3	2	8						✓	✓			✓		✓											
FactFinders [49]			1	✓	✓	✓																			✓	
FC_RUG [92]		6						✓																		
Fired_from_NLP [15]	7	12	10									✓	✓							✓						
Fraunhofer SIT [91]			3																							✓
HYBRINFOX [23]	10	8	12									✓	✓													✓
IAI Group [1]	1	3	9						✓	✓			✓		✓											
JUNLP [76]	14	11	22									✓														✓
Mirela [20]	11	4	16												✓		✓									
OpenFact [77]	2	7	2															✓								
pandas [85]	9	15	21									✓												✓		
SemanticCUETSync [60]	5	16	6									✓	✓				✓									
SINAI [89]			7						✓				✓									✓	✓			
SSN-NLP [27]			13									✓	✓		✓			✓						✓		✓
Team_Artists [53]	6	9	4									✓	✓	✓					✓	✓						
Trio_Titans [67]			19										✓			✓	✓							✓		
TurQUaz [12]	4	1	11		✓		✓		✓	✓	✓				✓											

Summary/Main takeaways/Highlights

The task attracted significant participation, with **75 registered teams**

- 13, 15 and 26 teams participated for Arabic, Dutch and English, respectively

Performances:

- Performances of the systems are relatively higher for English, followed by Dutch and Arabic

- Performances suggest that there is a room for improvement for English and low resource languages.

- Interests have been increasing over the years…

CT! Lab	Content Type	Modality	Language	Papers
CT-2018 [40]	Debate	Text	Ar, En	5
CT-2019 [41]	Debate, Web pages	Text	Ar, En	8
CT-2020 [42]	Tweet	Text	Ar, En	10
CT-2021 [43, 44]	Tweet, debate	Text	Ar, Bg, En, Es, Tr	10
CT-2022 [35, 45]	Tweet	Text	Ar, Bg, En, Nl, Es, Tr	13
CT-2023 [46, 47]	Tweet	Text, Image	Ar, En	12
CT-2024	Tweet, debate	Text	Ar, En, Nl	19

Task 2: Subjectivity in News Articles

Motivation

Subjective sentences often include elements that make them more difficult to analyze by machine learning models.

Objective sentences => Fact-checking pipeline

Subjective sentences => Additional processing

- Opinion piece: discard information
- Contains fact:
  - extract the objective version
  - flag it as a feature?

The event, which organisers had envisaged as a celebration of a new, progressive era, turned into a chaotic nightmare.

There is yet everywhere a deficit in the public revenue because the shrinkage in everything taxable was so sudden and violent.

Task Description

Given a sentence, extracted either from a news article, determine whether it is influenced by the subjective view of its author (class SUBJ) or presents an objective view of the covered topic (class OBJ).

Offered in **five** languages:

Arabic, Bulgarian, English, German, and Italian

Also offered in a **multilingual setting**.

Examples

English	*While it's misguided to put all focus or hope onto one section of the working class, we can't ignore this immense latent power that logistics workers possess.*	SUBJ
	Workers would have a 24 percent wage increase by 2024, including an immediate 14 percent raise.	OBJ
Arabic	الدكتور سامي الخيمي واللواء بهجت سليمان سفيران للأسد في حرب لفظية طاحنة.	SUBJ
	وكما هو معلوم فوجود الأوزون يحمي الحياة على الأرض من الأشعة فوق البنفسجية المنبعثة من الشمس.	OBJ

Data

	Training			Development			Development-Test			Test		
Language	**Total**	**OBJ**	**SUBJ**	**Total**	**OBJ**	**SUBJ**	**Total**	**OBJ**	**SUBJ**	**Total**	**OBJ**	**SUBJ**
Arabic	1,185	905	280	297	227	70	445	363	82	748	425	323
Bulgarian	729	406	323	106	59	47	208	116	92	250	143	107
English	830	532	298	219	106	113	243	116	127	484	362	122
German	800	492	308	200	123	77	291	194	97	337	226	111
Italian	1,613	1,231	382	227	167	60	440	323	117	513	377	136

Results

Rank	Team	F1
Arabic		
1	IAI Group	0.495
2	Nullpointer †	0.491
3	(baseline)	0.485
4	SemanticCuetSync	0.480
5	Tonirodriguez	0.465
6	HYBRINFOX	0.455
7	JUNLP	0.362
German		
1	Nullpointer	0.791
2	IAI Group	0.730
3	(baseline)	0.699
4	HYBRINFOX	0.697
Multilingual		
–	Nullpointer *	0.712
1	HYBRINFOX	0.685
2	(baseline)	0.670
3	IAI Group	0.629

Rank	Team	F1
Bulgarian		
1	(baseline)	0.753
2	Nullpointer	0.717
3	HYBRINFOX	0.715
4	IAI Group	0.582
5	JUNLP	0.364
Italian		
1	JK_PCIC_UNAM	0.792
2	HYBRINFOX	0.784
3	Nullpointer	0.743
4	(baseline)	0.650
5	IAI Group	0.586

Rank	Team	F1
English		
1	HYBRINFOX	0.744
2	Tonirodriguez	0.737
3	SSN-NLP	0.712
4	Checker Hacker	0.708
5	JK_PCIC_UNAM	0.708
6	SINAI	0.703
7	FactFinders	0.695
8	Vigilantes	0.695
8	Eevvgg	0.695
9	Nullpointer	0.689
10	Indigo	0.639
11	(baseline)	0.635
12	SemanticCuetSync	0.627
13	JUNLP	0.560
14	CLaC	0.450
15	IAI Group	0.449

† Team involved in the preparation of the data.
* Submitted after the official deadline.

Approaches

Team	Language						Model														Misc			
	Multilingual	Arabic	Bulgarian	English	German	Italian	BERT	RoBERTa	DistilBERT	Gemini	mBERT	mDeBERTa	Sentence-BERT	SetFit	Mistral-7B-Instruct	XLM RoBERTa	DeBERTa	BART	Llama	Sentiment-Analysis-BERT	Data Augmentation	Translating data	Multi-lingual Training	Feature Selection
Checker Hacker [36]				4				✓													✓			
CLaC [37]				14						✓											✓			
Eevvgg [38]				8			✓																	✓
FactFinders				7											✓									
HYBRINFOX [39]	1	6	3	1	4	2		✓	✓		✓											✓		✓
IAI Group [40]	3	1	4	15	2	5		✓								✓								
Indigo [41]				10									✓	✓										
JK_PCIC_UNAM [42]				5		1	✓																	✓
JUNLP		7	5	13			✓				✓													
Nullpointer [35]	-	2	2	9	1	3														✓		✓		
SemanticCuetSync [43]		4		12								✓							✓					
SINAI				6				✓																
SSN-NLP [44]				3				✓																✓
Tonirodriguez [45]		5		2								✓				✓	✓	✓					✓	
Vigilantes				8			✓																	

- The run was submitted after the official deadline, therefore not part of the official ranking.

Summary

- Transformers were most popular, both monolingual and multilingual.
- Strategies for data augmentation relied on LLMs.
- Strategies for addressing multilinguality include translation of data, multilingual training

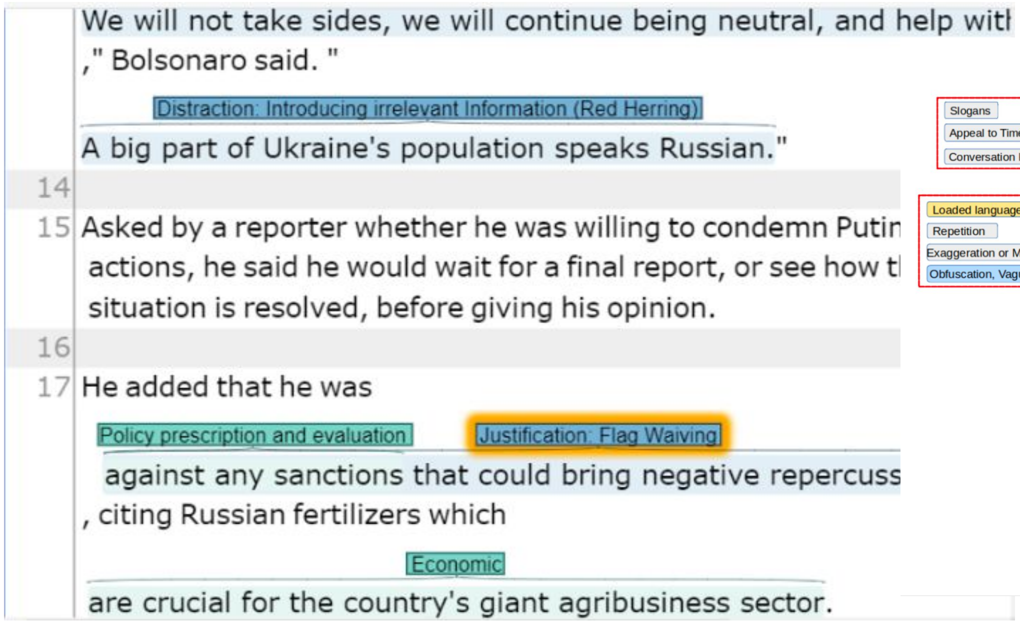
Task 3: Persuasion Techniques

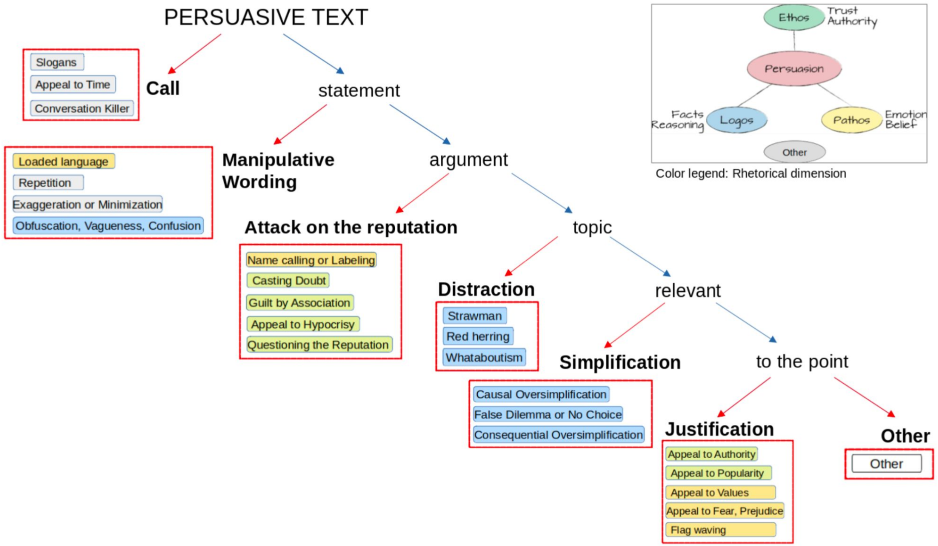
Motivation

Recognizing the various techniques used in news articles to influence readers' opinions on important topics.

Task Description

Given a news article and a list of 23 persuasion techniques, identify the spans where each technique occurs





Data

	Training set	Development set	Test set
English	X	X	X
French	X	X	
Italian	X	X	
German	X	X	
Russian	X	X	
Polish	X	X	
Spanish		X	
Greek		X	
Georgian		X	
Arabic			X
Portuguese			X
Slovenian			X
Bulgarian			X

	Training		Development	
language	#documents	#spans	#documents	#spans
English	536	9,002	54	1,775
French	211	6,831	50	1,681
German	177	5,737	50	1,904
Italian	303	7,961	61	2,351
Polish	194	3,824	47	1,491
Russian	191	4,138	72	944
Georgian	-	-	29	218
Greek	-	-	64	691
Spanish	-	-	30	546

	Test			
language	#documents	#paragraphs	#spans	α
Arabic	1,527	1,642	2,197	-
Bulgarian	100	916	1,732	0.197
English	98	2,174	2,599	0.168
Slovenian	100	1,478	4,591	0.470
Portuguese	104	1,501	1,727	0.587

Evaluation: Partial Matching

Fact: humanity will be extinct by 2025

Fact: humanity will be extinct by 2025

By traditional measures this is not a match

We propose a evaluation measure based on partial matching

$$I(p,g) = \begin{cases} 1, & \text{if } \frac{|p \cap g|}{|g|} \geqslant 0.5 \text{ and } |p| \leqslant 2 \cdot |g| \\ \frac{|p \cap g|}{|g|} \in (0,1), & \text{if } \frac{|p \cap g|}{|g|} \in (0,0.5) \text{ and } |p| \leqslant 2 \cdot |g| \\ \frac{|p \cap g|}{|p|} \in (0,1), & \text{if } \frac{|p \cap g|}{|g|} \in (0,1] \text{ and } |p| > 2 \cdot |g| \text{ and } |p| \leqslant 4 \cdot |g| \\ 0, & \text{otherwise} \end{cases}$$

Results

Rank	Team	F1 micro	F1 macro	Rank	Team	F1 micro	F1 macro
English				**Portuguese**			
1	UniBO	0.092	0.061		PersuasionMultiSpan*	0.132	0.120
	PersuasionMultiSpan*	0.078	0.086	1	UniBO	0.107	0.073
2	Baseline	0.009	0.001	2	Baseline	0.002	
Bulgarian				**Slovenian**			
	PersuasionMultiSpan*	0.132	0.128		PersuasionMultiSpan*	0.153	0.127
1	UniBO	0.114	0.081	1	UniBO	0.123	0.075
2	Baseline	0.009	0.002	2	Baseline	0.003	0.002
Arabic							
1	Mela	0.301	0.080				
2	UniBO	0.108	0.068				
	PersuasionMultiSpan*	0.028	0.059				
3	Baseline	0.021	0.006				

* Post competition experiment from the organizers

Approaches

Team	Language					Models		Misc
	Ar	Bg	En	Pt	Sl	mBERT	DeBERTa	Data aug
Mela	✓					✓		
UniBO	✓	✓	✓	✓	✓		✓	✓

Summary/Main Takeaways/Highlights

- Mostly fine-tuned transformer-based model
- Multilingual transformers
- Strategies for data augmentation
- Strategies for two-stage classification process

Task 4: Detecting the hero, the villain, the victim in memes

Motivation

- Social media
  - Online information exchange
  - Room for manifestation
- Memes express:
  - Emotions[1]
  - Sarcasm[2]
  - Hate speech[3] and misinformation[4]
  - Offensiveness[5] and harmfulness[6]
- What about the semantic roles[7] within the memes

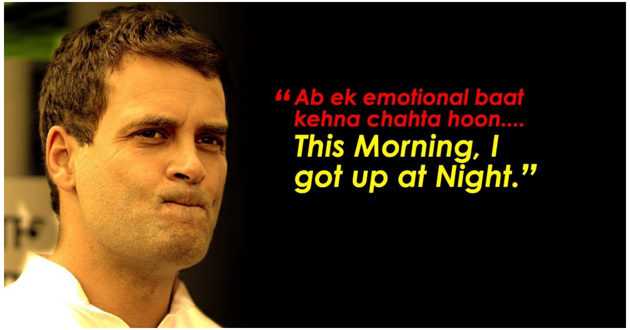
1. Sharma et al., SemEval-2020 Task 8: Memotion Analysis- the Visuo-Lingual Metaphor!, SemEval ’20
2. Kumar and Garg, Sarcasm detection in typographic memes, ICAESMT ’19,
3. Zhou et al., Multimodal learning for hateful memes detection, ICMEW ’21
4. Zidani and Moran, Memes and the spread of misinformation: Establishing the importance of media literacy in the era of information Disorder, Teaching Media Quarterly
5. Suryawanshi et al., Multimodal meme dataset (MultiOFF) for identifying offensive content in image and text, Workshop on Trolling, Aggression and Cyberbullying
6. Pramanick et al., Detecting harmful memes and their targets, ACL-IJCNLP ’21,
7. Sharma et al., Findings of the CONSTRAINT 2022 Shared Task on Detecting the Hero, the Villain, and the Victim in Memes, CONSTRAINT 2022

Task Description

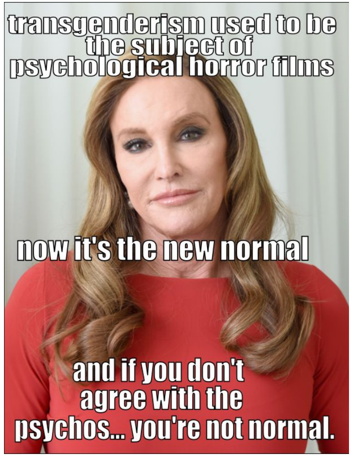
Objective: Predict roles ("hero", "villain", "victim", or "other") for each entity in a meme.

Input: Meme (image + extracted text) and list of entities.

Examples



(a) Rahul Gandhi (Villain) - EnHi



(b) Transgenders (Villain) - En

(c) ваксината (Villain) - Bg

❑ *Hero*: Entity presented in a positive light. Glorified for their deeds conveyed via the meme	❑ *Villain*: Portrayed negatively, associated with adverse traits like wickedness, cruelty, hypocrisy, etc.
❑ *Victim*: Portrayed as suffering the negative impact of an unfair act/wrongdoing.	❑ *Other*: The entity is not a hero, a villain, or a victim.

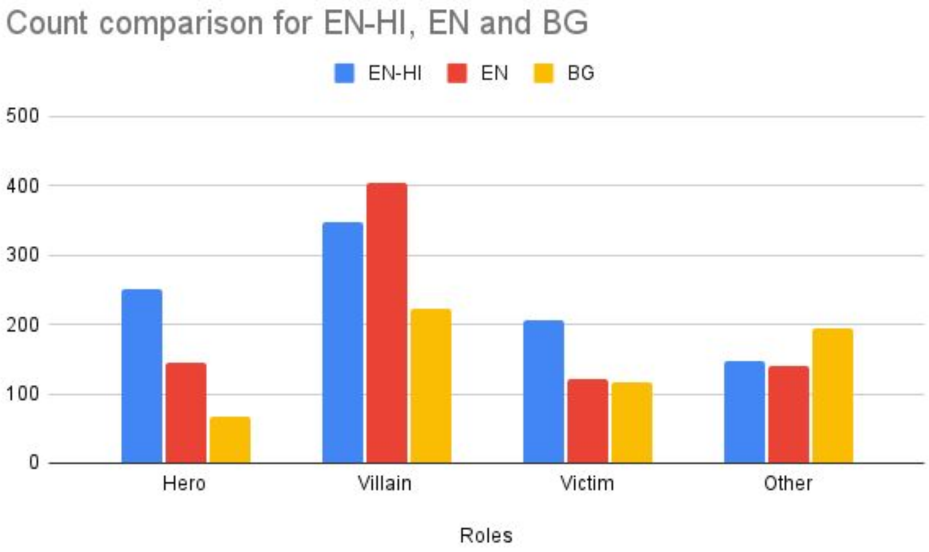
Annotation Guidelines and Training Data

S. No.	Annotation guidelines
1	Meme author's perspective needs to be considered as the frame of reference, while assigning roles.
2	Towards complete assimilation, both visual and textual cues should be factored-in.
3	Relevant background context should be acquired before assigning roles.
4	Ambiguous memes can be categorised as *other*.
5	A 3-point Likert scale based mental frame of reference, implying *negative*, *neutral* and *positive* sentiments involved, should steer connotation adjudication.
6	All reasonably *intelligible* (without ambiguity) entities referred must be considered as valid.
7	Entities with multiple interpretations should be categorised as *other*.
8	The role of the original speaker of a quote, expressed within a meme, must not be presumed.

Released for Training (En)			Official Test (hidden)		
Train	Dev	Devtest	En	Hi+En	Bg
3,711	468	501	500	500	227

Data: Testing

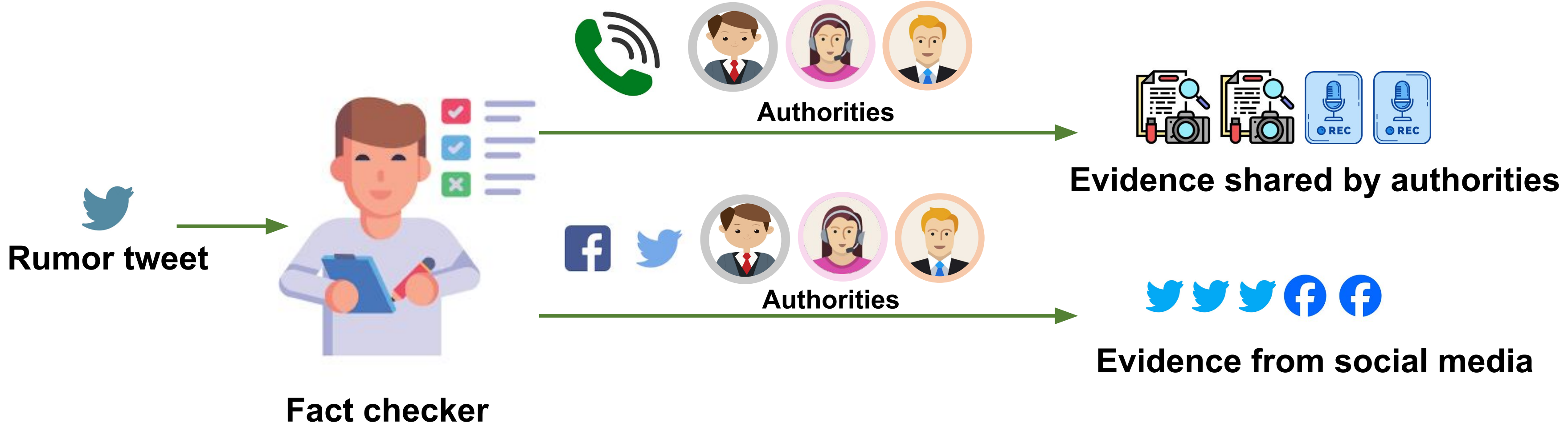
Roles	EN-HI	EN	BG
Hero	252	144	66
Villain	348	404	222
Victim	207	122	116
Other	148	141	195
Total	**955**	**811**	**599**



Task 5: Authority Evidence for Rumor Verification

Motivation

A trusted source of evidence for fact checking



Motivation

Rumor tweet

حمزة اليافعي
@hamzahalyafei

وباء كورونا وصل الى الامارات 75 إصابة في ابوظبي و 63 إصابة في دبي
تحذير للامتناع عن السفر الى الامارات حفاظًا على السلامه و عدم نقل الوباء .
اللهم أحفظ المسلمين في كل مكان...

Translated from Arabic by Google

The Corona epidemic reached the Emirates, with 75 cases in Abu Dhabi and 63 cases in Dubai
A warning to refrain from traveling to the Emirates in order to maintain safety and not transmit the epidemic.
May God protect Muslims everywhere...

5:36 PM · Jan 29, 2020

Authorities Twitter accounts

صاحب السمو الشيخ محمد بن راشد آل مكتوم

Follow

وزارة الصحة ووقاية المجتمع - MOHAP UAE
@mohapuae

وزارة الصحة ووقاية المجتمع هي جهة اتحادية ومنظمة للتشريعات الصحية في دولة الامارات العربية المتحدة

United Arab Emirates · mohap.gov.ae · Joined April 2010

World Health Organization
Regional Office for the Eastern Mediterranean

HEALTH FOR ALL

Following

WHO Regional Office for the Eastern Mediterranean
@WHOEMRO

Official Twitter account of the World Health Organization Eastern Mediterranean Regional Office.

Medical & Health · Cairo, Egypt · emro.who.int · Joined March 2011

Evidence tweets from authorities

وزارة الصحة ووقاية المجتمع - MOHAP UAE
@mohapuae

الوضع الصحي العام لا يدعو للقلق

تشخيص حالة إصابة بفيروس كورونا الجديد لأشخاص من عائلة واحدة قادمين من مدينة ووهان في جمهورية الصين الشعبية. وأن الحالة الصحية للمصابين مستقرة وتحت الملاحظة الطبية.

#وزارة_الصحة_ووقاية_المجتمع_الإمارات

Translated from Arabic by Google

The general health situation is nothing to worry about

A case of infection with the new Coronavirus was diagnosed for people from the same family coming from the city of Wuhan in the People's Republic of China. The health condition of the injured is stable and under medical observation.

10:19 AM · Jan 29, 2020

WHO Regional Office for the Eastern Mediterranean
@WHOEMRO

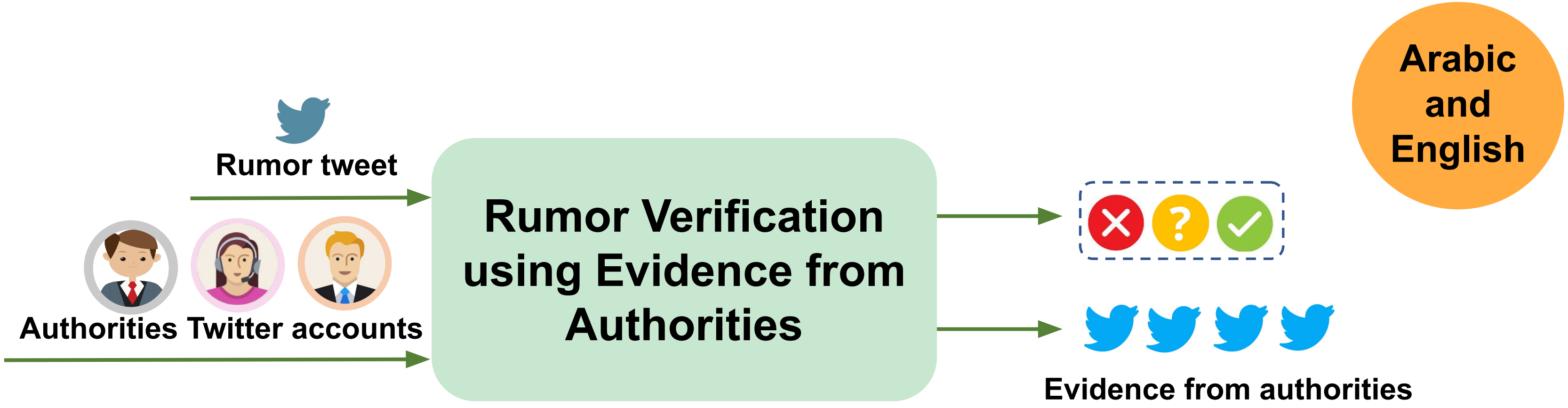
أكدت اليوم @WHO ظهور أولى حالات فيروس كورونا المستجد في إقليم شرق المتوسط، بالإمارات العربية المتحدة. عقب تأكيد @mohapuae في 29 يناير.
كان 4 أفراد من نفس العائلة من مدينة ووهان الصينية وصلوا إلى الإمارات في بداية يناير 2020، وتم إدخالهم المستشفى بعد تأكد إصابتهم ب #فيروس_كورونا.

Translated from Arabic by Google

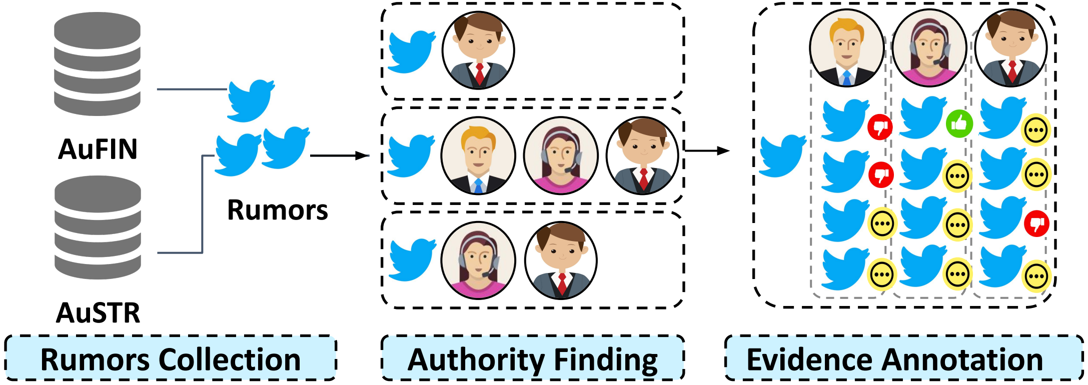
Today @WHO confirmed the emergence of the first cases of the new Corona virus in the Eastern Mediterranean Region, in the United Arab Emirates. Following confirmation of @mohapuae on January 29.
4 members of the same family from the Chinese city of Wuhan arrived in the Emirates at the beginning of January 2020, and were admitted to the hospital after their infection with #فيروس_كورونا was confirmed.

4:08 PM · Jan 29, 2020

Task Description



Data



	Arabic rumors	English rumors
Train	96	96
Dev	32	32
Test	32	32

Rumors	
SUPPORTS	30
REFUTES	64
NOT ENOUGH INFO	66
Authority tweets	
Authorities	692
Authority tweets	33705

The data is originally in Arabic but translated to English

Evidence Retrieval Results

English

Rank	Team (run ID)	MAP	Recall@5
1	bigIR^{+} (bigIR-MLA-En)	0.604	0.677
2	Axolotl (run_rr=llama_sp=llama_rewrite=3_boundary=0)	0.566	0.617
3	DEFAULT (DEFAULT-Colbert1)	0.559	0.634
4	IAI Group (IAI-English-COLBERT)	0.557	0.590
5	AuthEv-LKolb (AuthEv-LKolb-oai)	0.549	0.587
	Baseline	0.335	0.445

5 teams for English

Arabic

Rank	Team (run ID)	MAP	Recall@5
1	bigIR^{+} (bigIR-MLA-Ar)	0.618	0.673
2	IAI Group (IAI-Arabic-COLBERT)	0.564	0.581
	Baseline	0.345	0.423
3	SCUoL (SCUoL)	-	-

3 teams for English Arabic

Evaluation:

- Mean Average Precision (MAP) for evidence retrieval

Rumor Verification Results

English

Rank	Team (run ID)	m-F1	Strict m-F1
1	AuthEv-LKolb (AuthEv-LKolb-oai)	0.879	0.861
2	Axolotl (run_rr=llama_sp=llama_rewrite=3_boundary=0)	0.687	0.687
	Baseline	0.495	0.495
3	DEFAULT (DEFAULT-Colbert1)	0.482	0.454
4	bigIR$^+$ (bigIR-MLA-En)	0.458	0.428
5	IAI Group (IAI-English-COLBERT)	0.373	0.373

5 teams for English

Arabic

Rank	Team (run ID)	m-F1	Strict m-F1
1	IAI Group (IAI-Arabic-COLBERT)	0.600	0.581
2	bigIR$^+$ (bigIR-MLA-Ar)	0.368	0.300
3	SCUoL (SCUoL)	0.355	-
	Baseline	0.347	0.347

3 teams for Arabic

Evaluation:

- Macro-F1 and strict Macro-F1 for rumor verification

Approaches

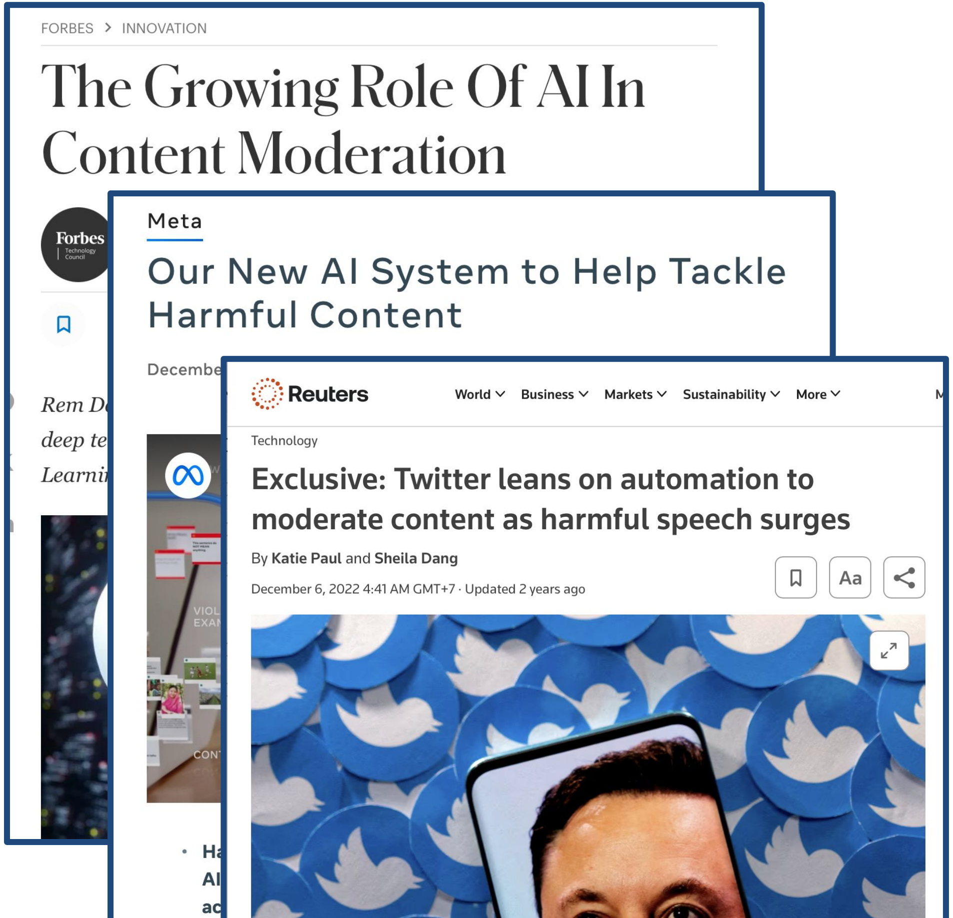
- 5 teams for English: **bigIR, IAI group, DEFAULT, Axolotl, AuthEv-LKolb**
- 3 teams for Arabic: **bigIR, IAI group, SCUoL**
- 2 teams participated in both languages.
- **Multiple approaches for evidence retrieval:**
  - Fine-tuned existing fact-checking models.
  - Adopted a zero-shot setup by leveraging existing pre-trained language models, LLMs, lexical retrieval such BM25, or combination of these models.
- **Different approaches for rumor verification:**
  - Fine-tuned existing fact-checking models.
  - Adopted a zero-shot setup using Large language models such as GPT-4 and Llama.

Summary/Main takeaways/Highlights

- For **evidence retrieval**, a fine tuned fact-checking model outperformed all models.
- For **rumor verification**, only the models adopting LLMs managed to outperform the baseline.
- The data is relatively small to train effective rumor verification models.

Task 6: Robustness of Credibility Assessment with Adversarial Examples (InCrediblAE)

Motivation



- ML is increasingly common in moderation of platforms with user-generated content
- Automatic credibility analysis can perform well, but is it vulnerable to motivated attackers?
- Let's check how easy it is to fool a text classifier by making small changes to text input!

Task Description

- for each credibility assessment task *t* (e.g. propaganda detection)
  - for each victim classifier $f_{t,v}$ (e.g. fine-tuned BERT)
    - for each attack example x_i, e.g.
      *Puerto Rico's housing secretary, Fernando Gil, says the number of **homes** destroyed by the hurricane totals about 70,000 so far, (...)*
    - find an adversarial modification x_i^*, e.g.
      *Puerto Rico's housing secretary, Fernando Gil, says the number of **houses** destroyed by the hurricane totals about 70,000 so far, (...)*
    - such that $f_{t,v}(x_i) \neq f_{t,v}(x_i^*)$

=> Compute the victim confusion, example similarity and number of queries.

Data

- Five domains/tasks of credibility assessment:
  - News bias assessment (HN)
  - Propaganda detection (PR)
  - Fact checking (FC)
  - Rumour detection (RD)
  - COVID-19 misinformation detection (C19)

-> All based on previously published datasets.

- Three victim classifiers:
  - simple BiLSTM network,
  - fine-tuned BERT
  - surprise classifier (revealed in test phase): RoBERTa, adversarially fine-tuned.

Task	Training	Attack	Development	Positive
HN	60,235	400	3,600	50.00%
PR	12,675	416	3,320	29.42%
FC	172,763	405	19,010	51.27%
RD	8,694	415	2,070	32.68%
C19	1,130	595	0	42.55%

Approaches

- Six teams have submitted solutions:
  - **MMU_NLP** (Manchester Metropolitan University)
  - **TurQuaz** (TOBB University of Economics and Technology)
  - **TextTrojaners** (University of Zurich)
  - **Palöri** (University of Zurich)
  - **OpenFact** (Poznań University of Economics and Business)
  - **SINAI** (University of Jaén)

-> All have submitted papers

-> Go to the presentations to see their approaches – sessions on Wednesday: oral (**OpenFact**, **TextTrojaner** and **MMU_NLP**) and poster (**SINAI** and **TurQuaz**)

Results

- Automatic evaluation:
  - confusion score [$f_{t,v}(x_i) \neq f_{t,v}(x_i^*)$]
  - semantic similarity score [BLEURT]
  - character similarity score [Levenshtein]
  - BODEGA score
- Manual evaluation:
  - meaning preserved / changed / nonsensical
  - confidence [1-5]

#	Method	BODEGA avg.
1.	OpenFact	0.7458
2.	TextTrojaners	0.7074
3.	TurQUaz	0.4859
4.	Palöri	0.4776
5.	MMU_NLP	0.3848
6.	SINAI	0.3507
-	BERT-ATTACK	0.4261
-	DeepWordBug	0.2682

Team	% of Preserve the meaning
SINAI	99%
MMU_NLP	96%
TurQUaz	62%
Plagori	14%
OpenFact	11%
TextTrojaners	7%

Highlights

- Two word-focused approaches dominated the automatic evaluation: **OpenFact** and **Palöri**,
- In manual evaluation, the two character-focused approaches were judged as best at preserving meaning: **SINAI** and **MMU_NLP**
- **TextTrojaners** won some of the scenarios, but at the cost of very many queries (record: 15,458.12)
- **OpenFact** were overall winners, but did not submit the number of queries.
- Only the **TurQUaz** team attempted to prompt LLMs for adversarial examples, but the results were not encouraging.

All the data are available for more experiments in the BODEGA framework:
https://github.com/piotrmp/BODEGA/

CheckThat! Program

Programme (Grenoble time)

CT! oral session 1: Tuesday 10th September, 16:40 to 18:10

16:40 - Introduction to the CheckThat! Lab

17:25 - **Task 1**: Three talks on Check-Worthiness in Multigenre Content

CT! oral session 2: Wednesday 11th September, 14:00 to 15:30

14:00 - **Task 2**: Three talks on Subjectivity in News Articles

14:45 - **Task 5**: Three talks on Rumor verification using evidence from authorities

CLEF poster session 3: Wednesday 11th September, 15:30-16:30

CT! oral session 3: Wednesday 11th September, 16:30 to 18:00

16:30 - **Task 3**: One talk on Persuasion techniques

14:45 - **Task 6**: Three talks on Robustness of Credibility Assessment with Adversarial Examples

17:30 - **Invited talk**. Salim Hafid. Claims and Sources in Scientific Web Discourse (SciWeb)

Details on the CheckThat! website:

http://checkthat.gitlab.io/clef2024/#lab-programme

Organisation

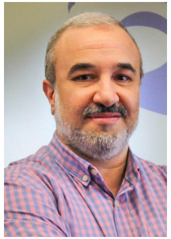

Alberto
Barrón-Cedeño


Firoj Alam


Julia Maria Struß


Preslav Nako


Tanmoy
Chakraborty


Tamer Elsayed


Piotr Przybyła


Tommaso Caselli


Giovanni
da San Martino


Fatima Haouari


Maram Hasanain


Chengkai Li


Jakub Piskorski

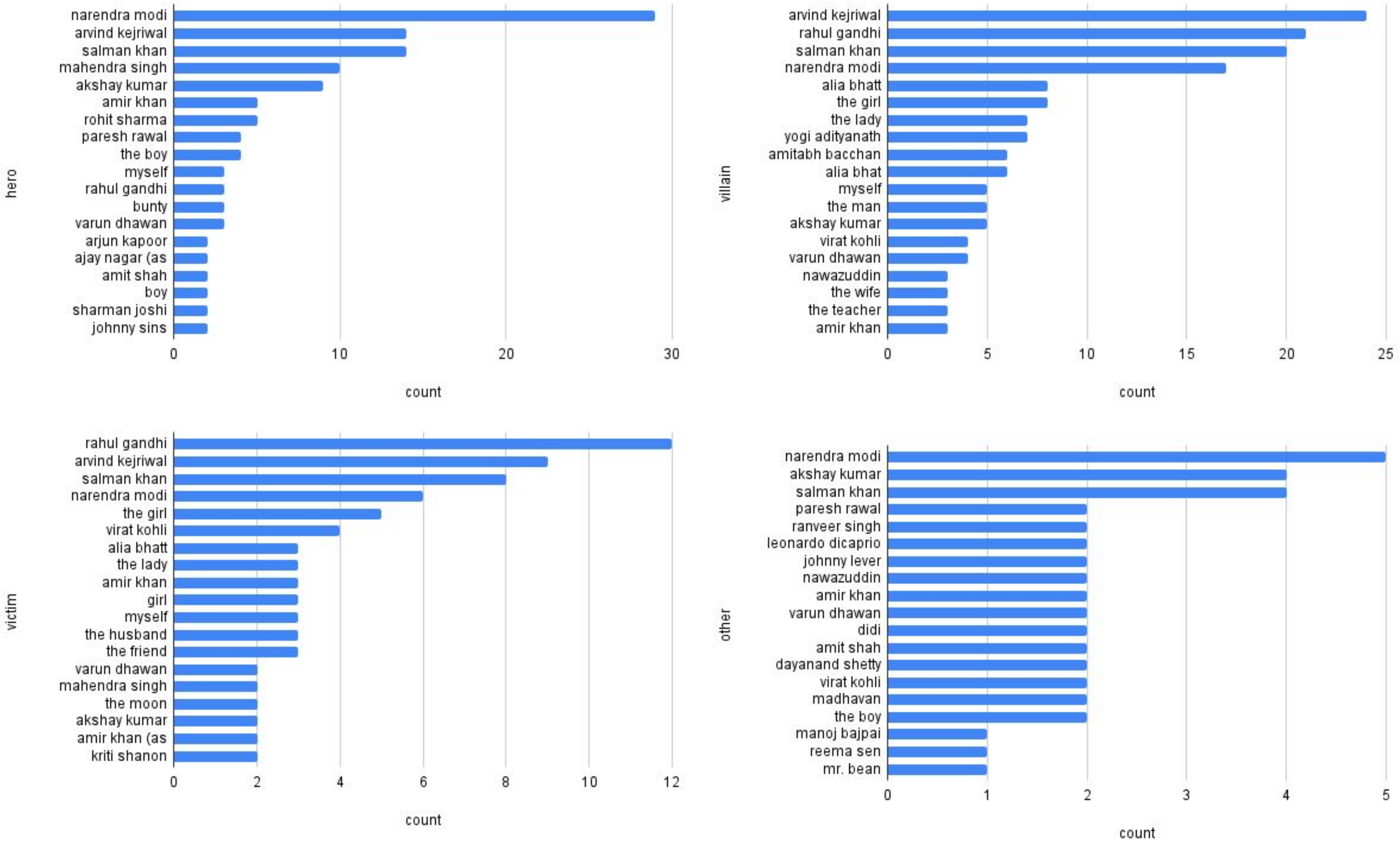

Federico Ruggeri


Xingyi Song


Reem Suwaileh

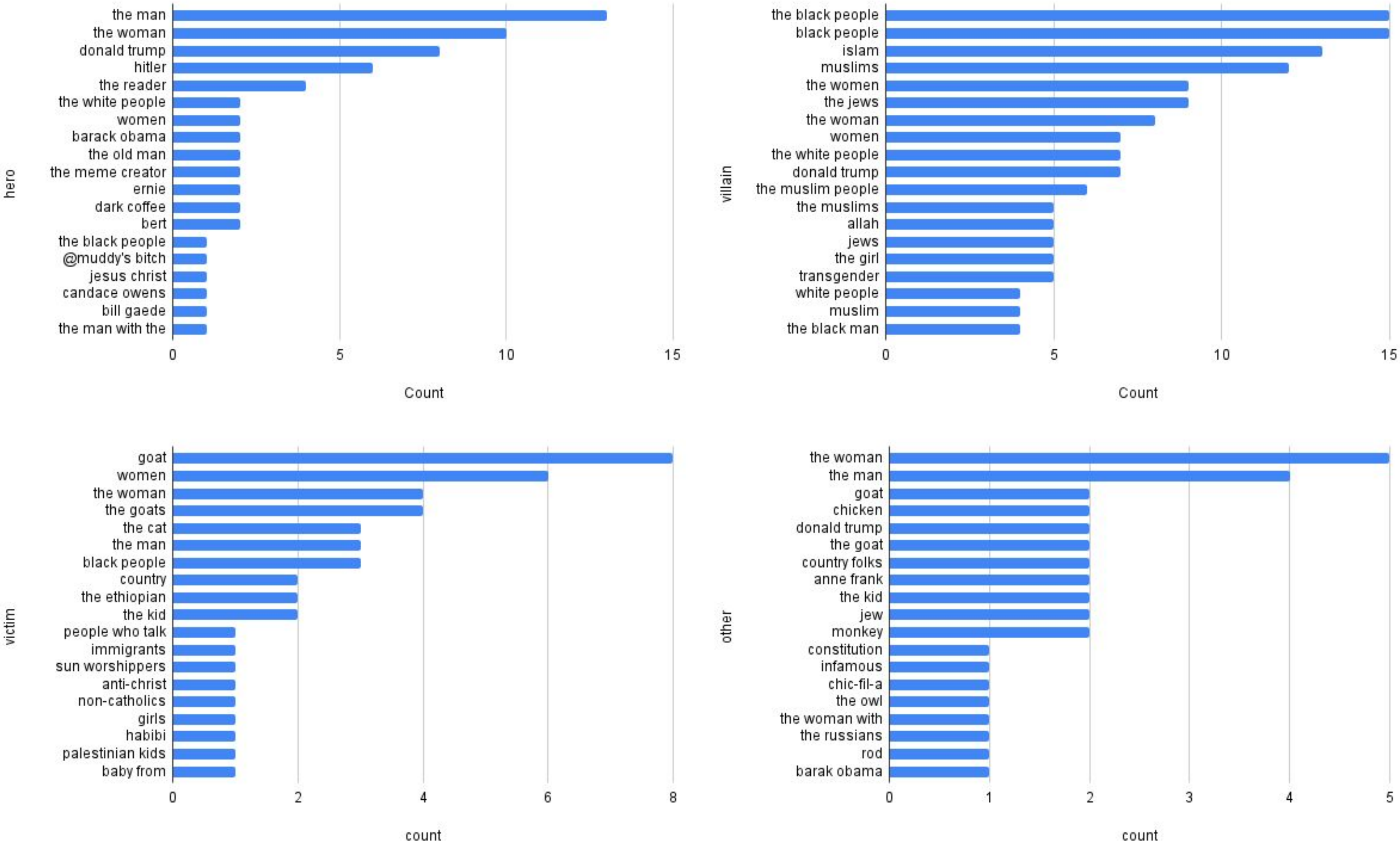
Testset (En-Hi)

- **Polarized Portrayals**: Figures like Narendra Modi and Arvind Kejriwal appear as both heroes and villains, indicates polarizing public perception in memes.
- **Flexible Narratives**: Celebrities such as Salman Khan are depicted across hero, villain, and victim roles, expressing adaptability of public figures in meme storytelling.
- **Satire in Politics & Entertainment**: Memes heavily focus on political and entertainment figures; satire prominently used for public events and controversies.
- **Public Sentiment**: Victimization in memes also reflects public sentiment; figures like Rahul Gandhi often portrayed as victims in satire.
- **Code-Mixed Language**: Hindi-English code-mixing in memes makes them more relatable to bilingual audiences, expanding reach and engagement.


hero
narendra modi
arvind kejriwal
salman khan
mahendra singh
akshay kumar
amir khan
rohit sharma
paresh rawal
the boy
myself
rahul gandhi
bunty
varun dhawan
arjun kapoor
ajay nagar (as
amit shah
boy
sharman joshi
johnny sins
0
10
20
30
count
villain
arvind kejriwal
rahul gandhi
salman khan
narendra modi
alia bhatt
the girl
the lady
yogi adityanath
amitabh bacchan
alia bhat
myself
the man
akshay kumar
virat kohli
varun dhawan
nawazuddin
the wife
the teacher
amir khan
0
5
10
15
20
25
count
victim
rahul gandhi
arvind kejriwal
salman khan
narendra modi
the girl
virat kohli
alia bhatt
the lady
amir khan
girl
myself
the husband
the friend
varun dhawan
mahendra singh
the moon
akshay kumar
amir khan (as
kriti shanon
0
2
4
6
8
10
12
count
other
narendra modi
akshay kumar
salman khan
paresh rawal
ranveer singh
leonardo dicaprio
johnny lever
nawazuddin
amir khan
varun dhawan
didi
amit shah
dayanand shetty
virat kohli
madhavan
the boy
manoj bajpai
reema sen
mr. bean
0
1
2
3
4
5
count

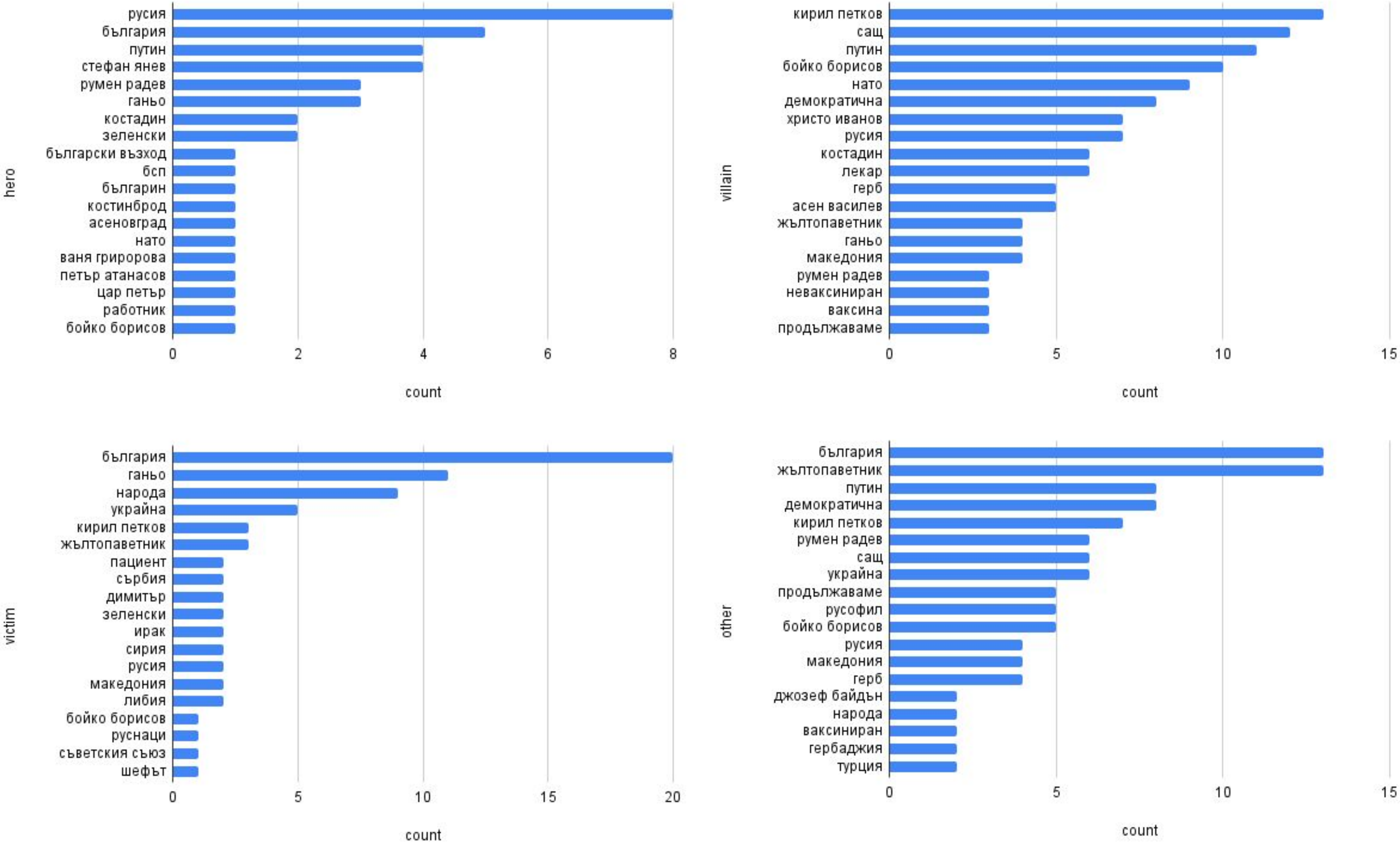
Test-set (En)

- **Cultural and Social Roles**: Familiar figures like political leaders or archetypal characters are elevated as heroes, while portraying marginalized groups as villains, reflecting societal biases.
- **Humor and Irony**: Victim range from *serious* (e.g., women) to *humorous* (e.g., goats), using satire or trivialization.
- **Gender Dynamics**: Both men and women are featured prominently, roles switches equally between hero and victim, highlighting traditional gender representations.
- **Political and Social Bias**: Memes amplify political and social biases, often portraying real-world groups as villains, serving as a mirror of *ideological viewpoints*.
- **Simplification of Issues**: Memes condense complex issues into simple hero-villain-victim narratives, which can perpetuate stereotypes or biases.


hero
the man
the woman
donald trump
hitler
the reader
the white people
women
barack obama
the old man
the meme creator
ernie
dark coffee
bert
the black people
@muddy's bitch
jesus christ
candace owens
bill gaede
the man with the
0
5
10
15
Count
villain
the black people
black people
islam
muslims
the women
the jews
the woman
women
the white people
donald trump
the muslim people
the muslims
allah
jews
the girl
transgender
white people
muslim
the black man
0
5
10
15
Count
victim
goat
women
the woman
the goats
the cat
the man
black people
country
the ethiopian
the kid
people who talk
immigrants
sun worshippers
anti-christ
non-catholics
girls
habibi
palestinian kids
baby from
0
2
4
6
8
count
other
the woman
the man
goat
chicken
donald trump
the goat
country folks
anne frank
the kid
jew
monkey
constitution
infamous
chic-fil-a
the owl
the woman with
the russians
rod
barak obama
0
1
2
3
4
5
count

Testset (Bg)

- **Humor and Irony**: The Bulgarian nation is often portrayed through a fictional literature character (Bay Ganyo) considered an exemplary image of an anti-hero: an uneducated, ignorant, egoistic and poor.
- **Gender Dynamics**: Men are prominently featured in both hero and victim roles, primarily because most big political party leaders are men..
- **Victimization:** The Bulgarian people are predominantly portrayed as victims of a corrupt government or specific political parties.
- **Simplification:** Memes simplify the political landscape into a choice between *A very bad villain* and a much *better Hero* alternative.


hero
русия
България
путин
стефан янев
румен радев
ганьо
костадин
зеленски
български възход
бсп
българин
костинброд
асеновград
нато
ваня гриророва
петър атанасов
цар петър
работник
бойко борисов
0
2
4
6
8
count
villain
кирил петков
сащ
путин
бойко борисов
нато
демократична
христо иванов
русия
костадин
лекар
герб
асен василев
жълтопаветник
ганьо
македония
румен радев
неваксиниран
ваксина
продължаваме
0
5
10
15
count
victim
българия
ганьо
народа
украйна
кирил петков
жълтопаветник
пациент
сърбия
димитър
зеленски
ирак
сирия
русия
македония
либия
бойко борисов
руснаци
съветския съюз
шефът
0
5
10
15
20
count
other
българия
жълтопаветник
путин
демократична
кирил петков
румен радев
сащ
украйна
продължаваме
русофил
бойко борисов
русия
македония
герб
джозеф байдън
народа
ваксиниран
гербаджия
турция
0
5
10
15
count