

DeFaktS: A German Dataset for Fine-Grained Disinformation Detection through Social Media Framing

Authors: Shaina Ashraf, Isabel Bezzaoui, Ionut Andone, Alexander Markowetz, Jonas Fegert, Lucie Flek

- What is disinformation
- Why disinformation detection is important?
- Intro to DefaktS Project
- Data collection
- Annotation scheme
- Experiments
- Results
- What comes next

What is disinformation

Misinformation

When false info
is shared
unintentionally
but no harm is
meant

When false info
is deliberately
shared to
cause harm

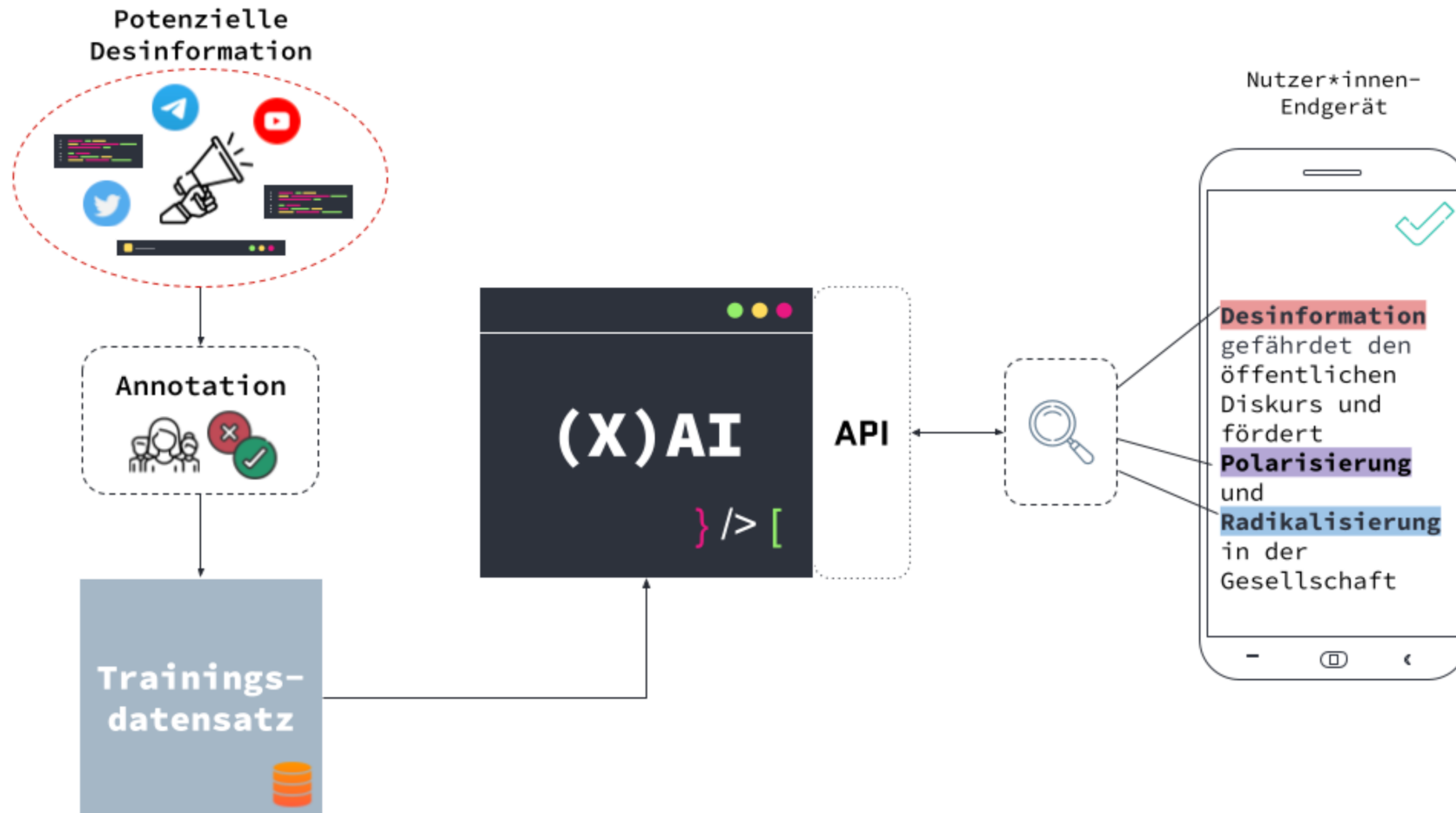
Info that is
partially or totally
true, but spread
with malicious
intent

Malinformation

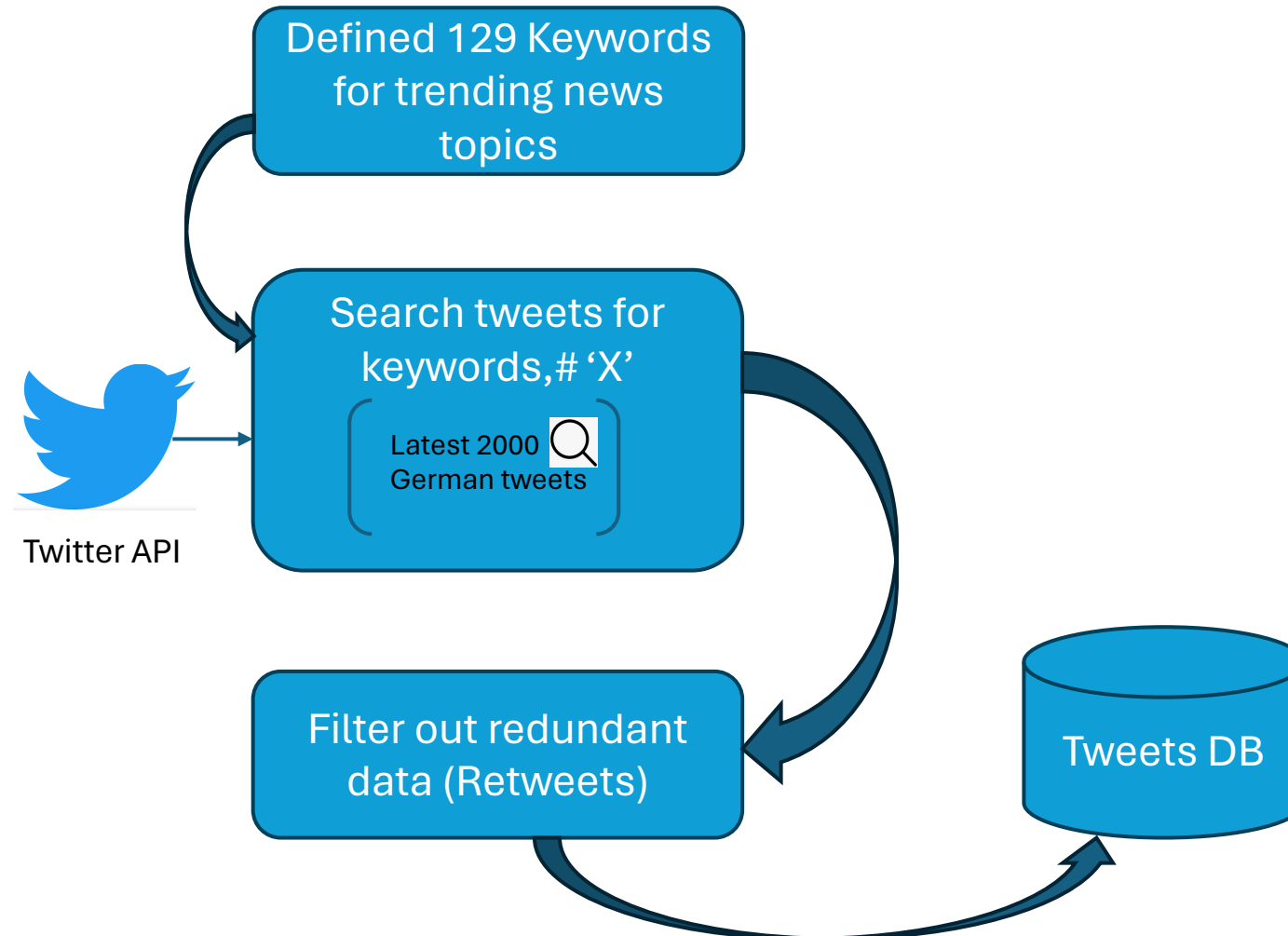
Why disinfo detection is important?

- Disinformation campaigns are designed to:
 - influence public and political opinion
 - manipulate emotions, create confusion, and sow mistrust
 - create chaos, disunity, discord & violence in public
 - disseminate government or partisan propaganda, attack the opposition
 - influence elections
 - deflect or divert discussion or criticism from important issues
 - suppress the participation of certain individuals through personal attacks or harassment
 - undermine trust in media or institutions
 - as tool to war against certain countries

DefaktS Project Overview



Data Collection

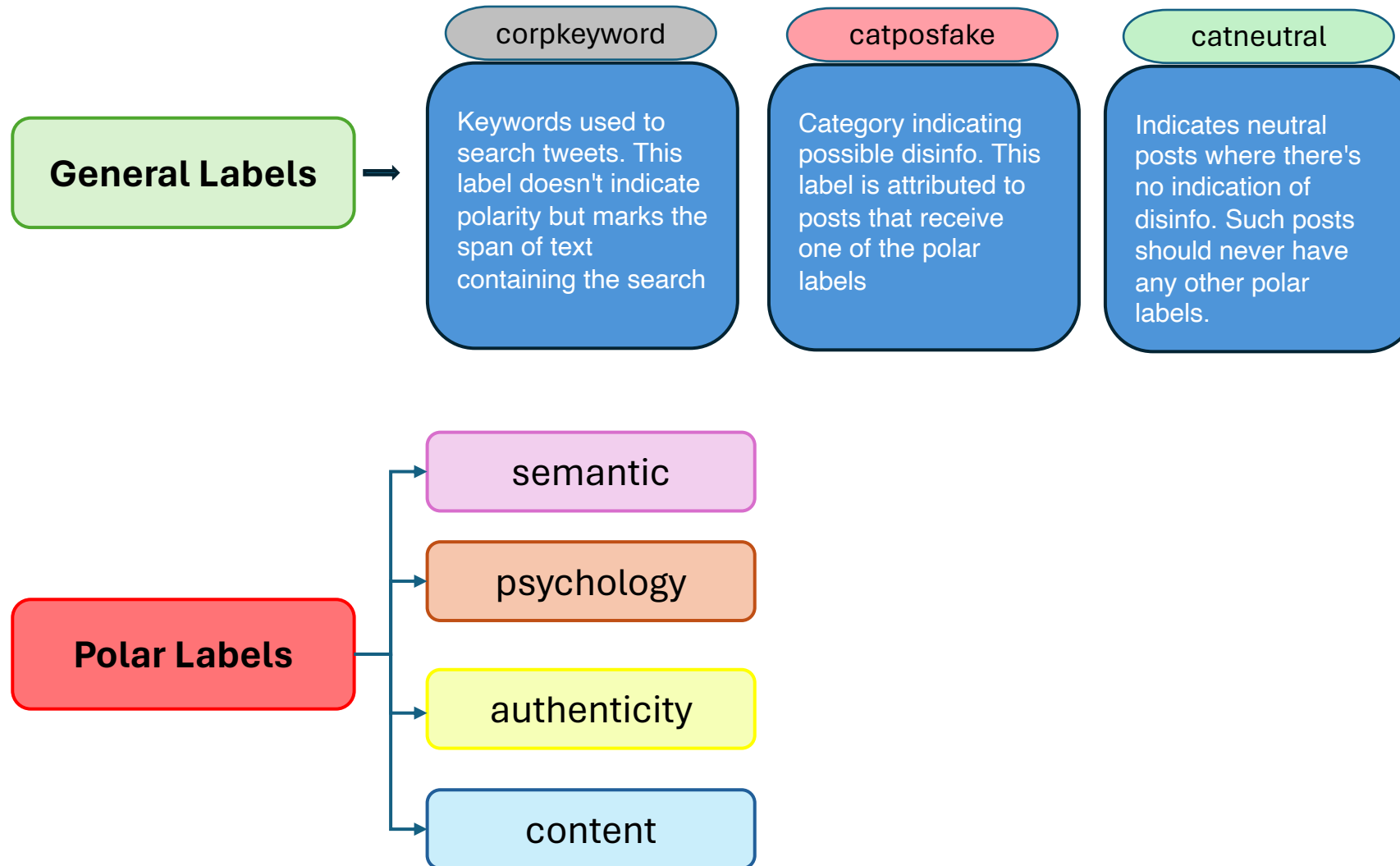


attributes	values
Total posts	105,855
Labelled posts	20,008
Unique Users	44,486
Avg Length(char)	187
Avg Length(words)	24
Average Retweets	4
Number of tweets with URLs	65,889

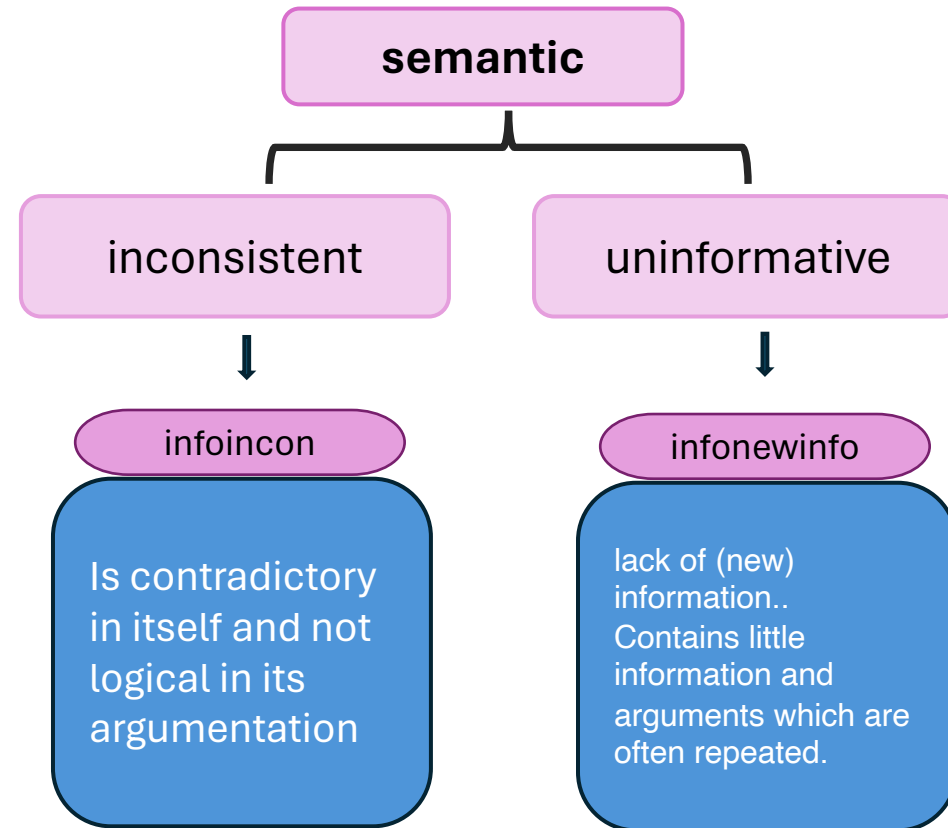


Distribution of Topics in Dataset

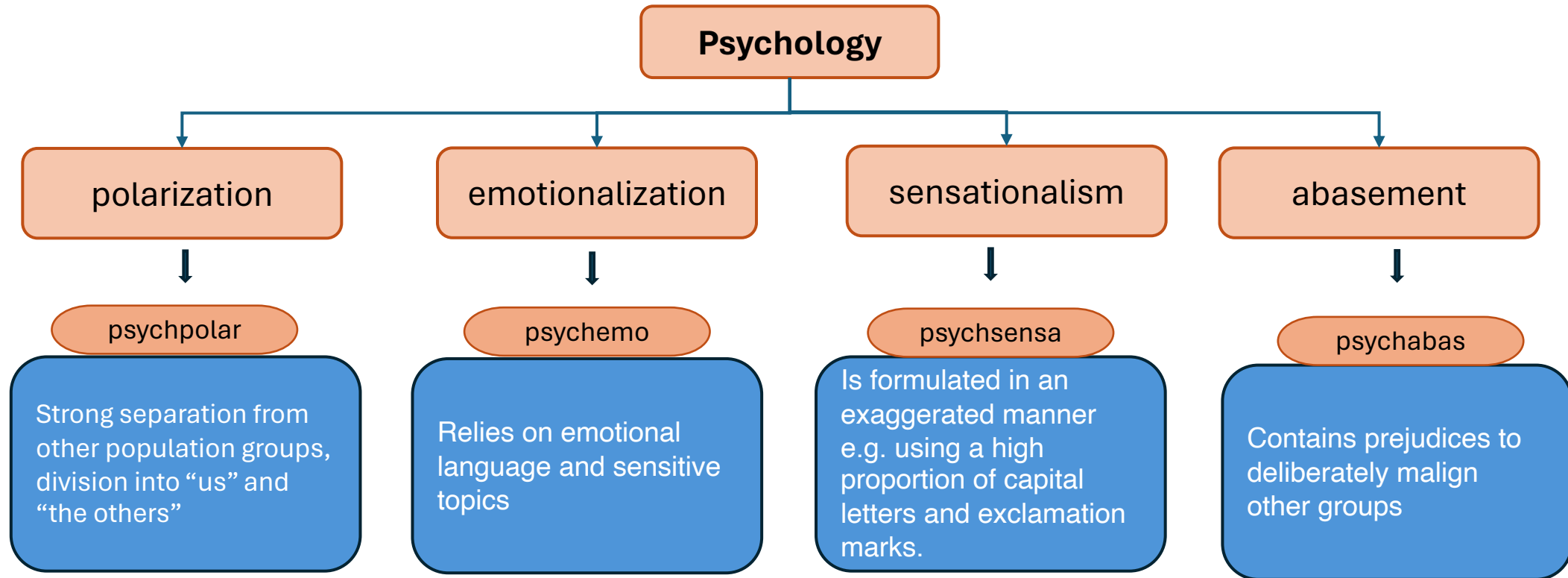
Annotation Scheme



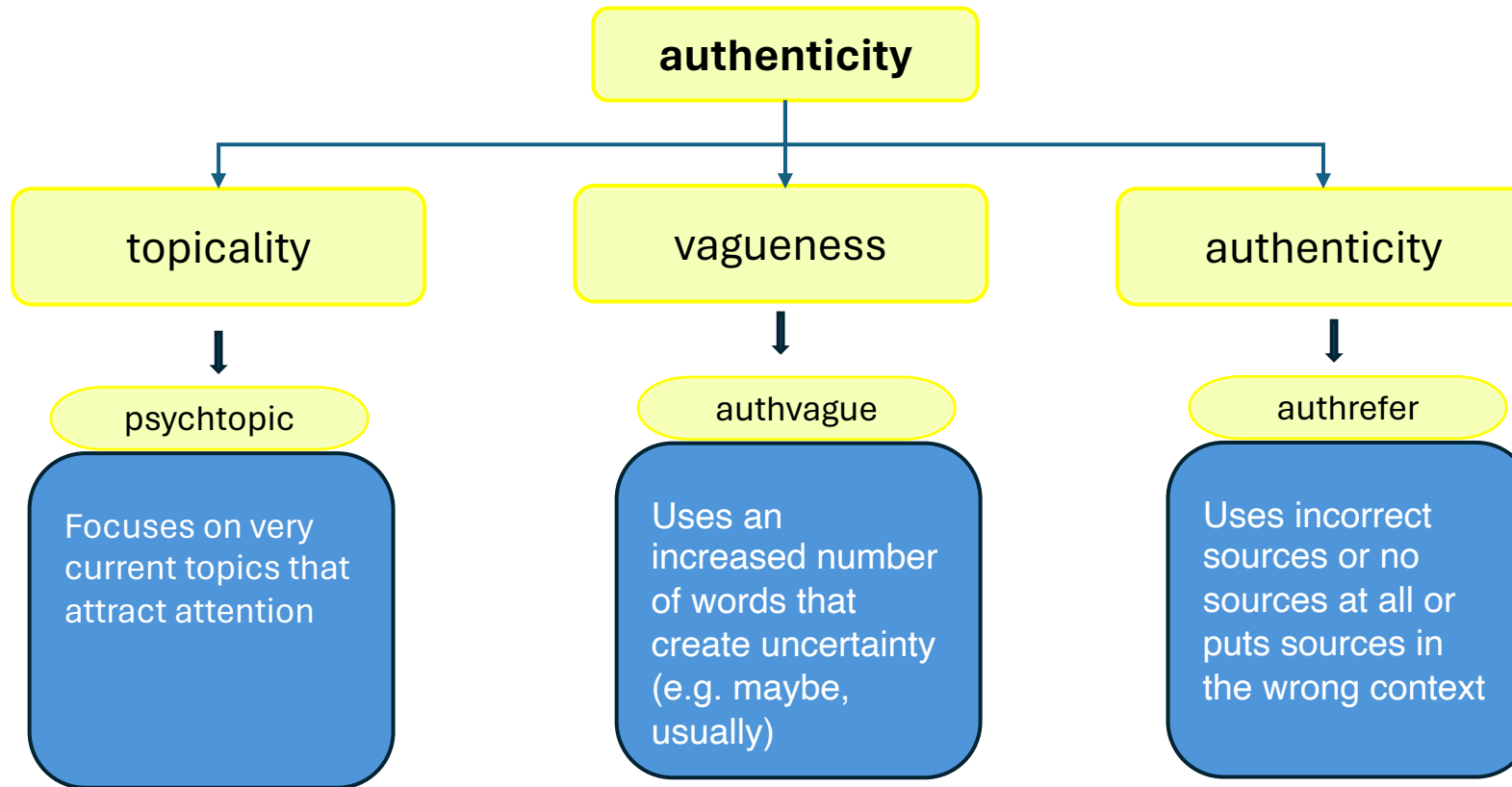
Semantic features



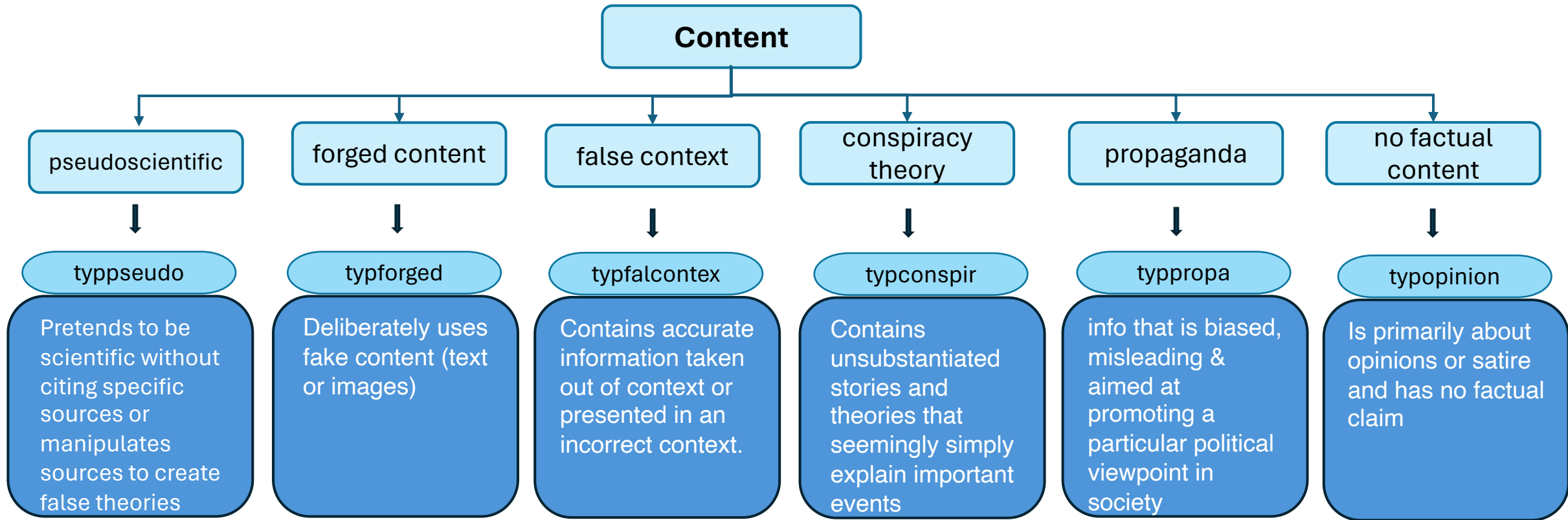
Psychology features



Authenticity features



Content type features



Annotated Examples

#BerlinWahlen2023 psychtopic seitdem Grün regiert bin ich gegen Briefwahl typopinion. Wer auf Wahlplakaten lügt, wer mit Machttöten statt reden möchte, Lützerath, Spaltung, Hetze, den Verfassungsschutz corpkeyword missbraucht psychpolar, egal was meine deutschen Wähler denken, traue ich alles zu, ist unwählbar. 12.2. 51 typopinion
<https://t.co/7kuWdYjclM>

#BerlinElections2023 Since the Green Party has ruled, I have been against postal voting. Anyone who lies on election posters, who wants to use power to kill instead of talk, Lützerath, division, agitation, who abuses the protection of the constitution, no matter what my German voters think, I believe everything is unelectable. 12.2. 51 typopinion
<https://t.co/7kuWdYjclM>

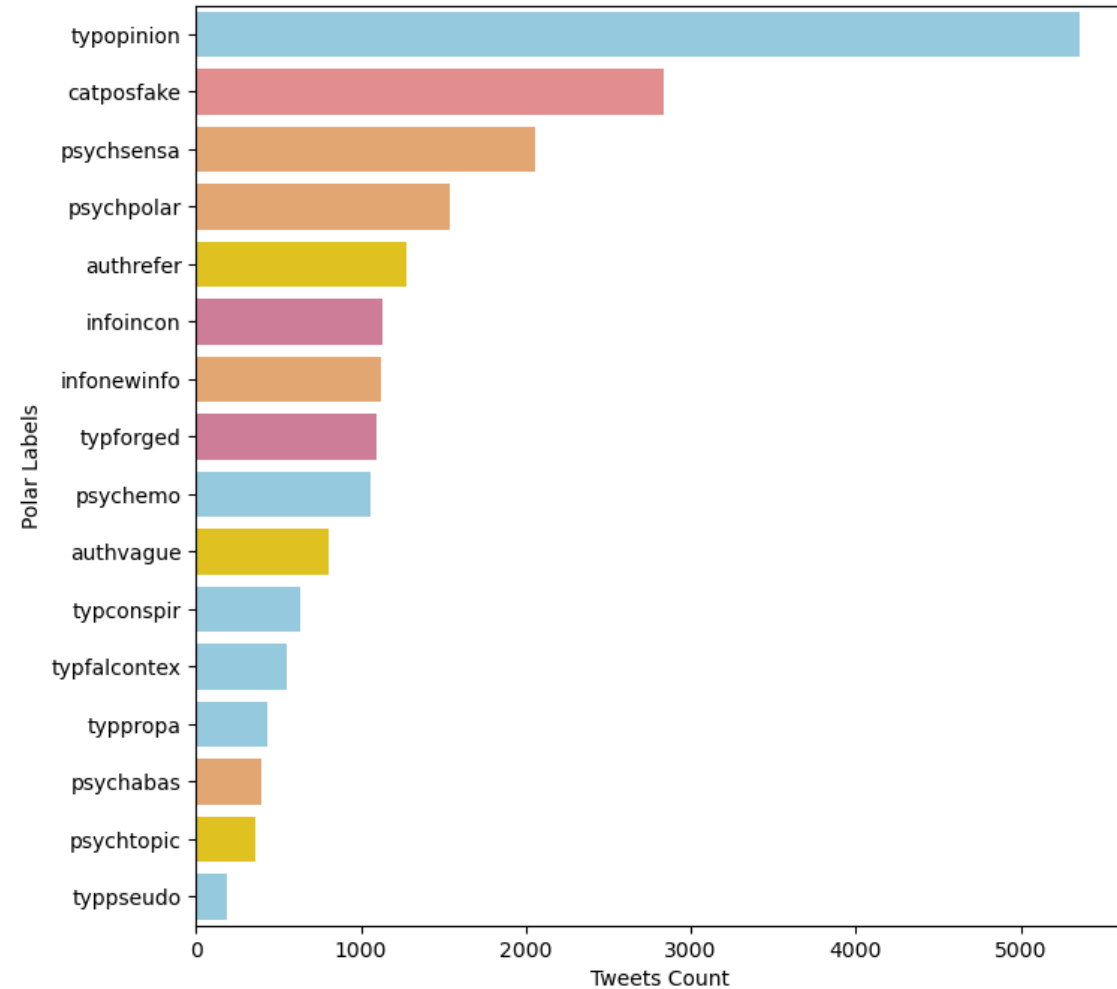
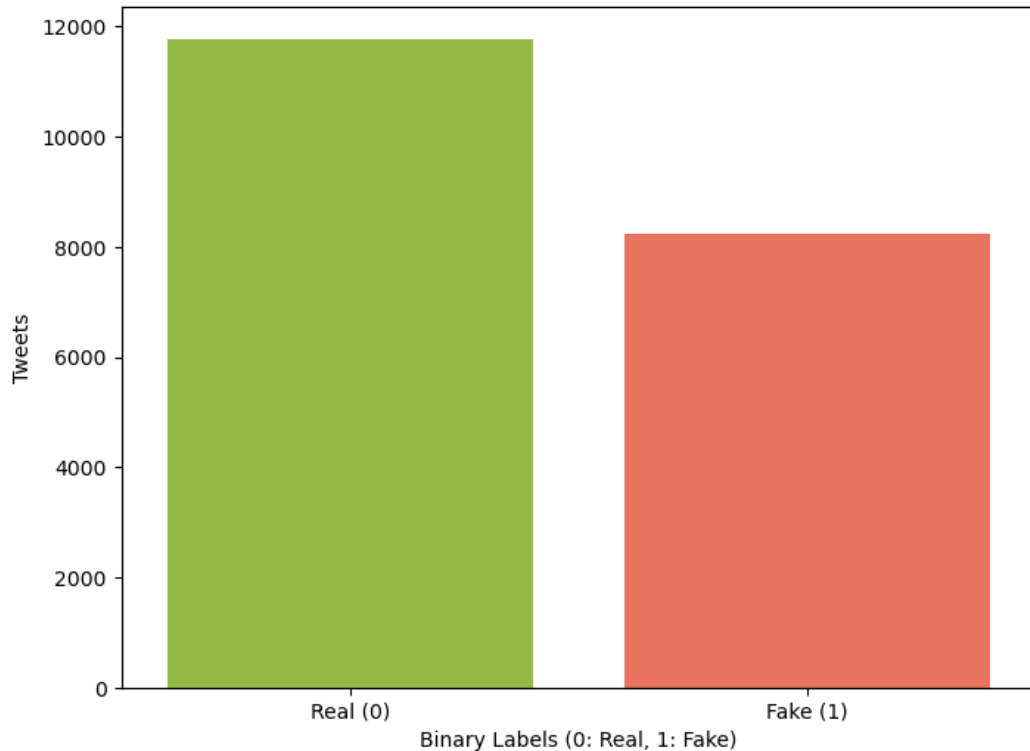
Steffen Kontré sollte sich entscheiden: Parlament oder #Putin.
Ansonsten raus aus dem #Bundestag! #AfDVerbotSofort catneutral

Steffen Kontré should decide: Parliament or #Putin. Otherwise, get out of the #Bundestag! #AfDVerbotImmediately

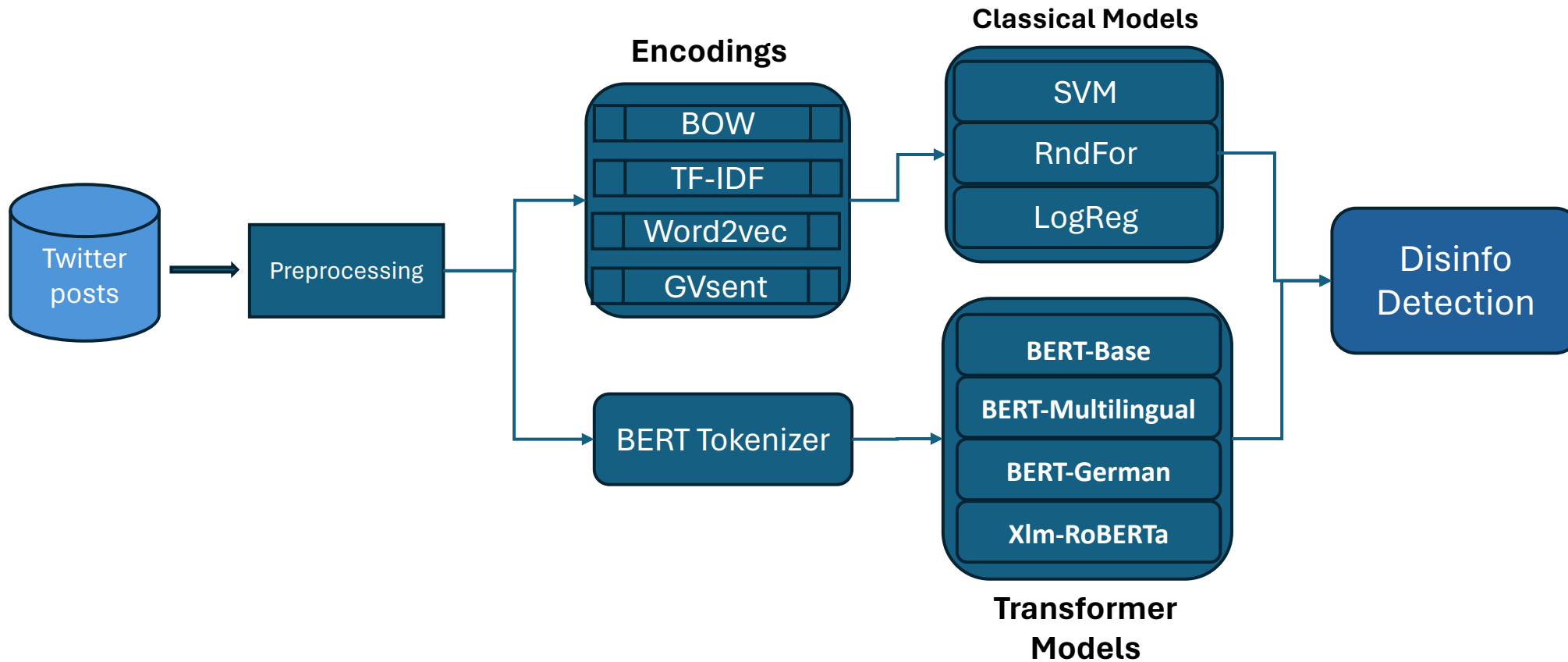
Wie wahr; Die EU fördert statt Kinder und Familie Abtreibung corpkeyword und LGBTQ. infonewinfo Linksradikale psychpolar Klima Hysterie psychemo Immigration. Vor allem werden verfolgte Christen sehr benachteiligt. typopinion

How true; The EU promotes abortion and LGBTQ instead of children and families. Left radical climate hysteria immigration. Above all, persecuted Christians are very disadvantaged.

Label Distributions



Experiments



Results

Binary Class

Features	SVM	RndFor	LogFor
TF-IDF	0.76	0.74	0.81
BOW	0.78	0.72	0.80
Word2Vec	0.47	0.52	0.46
GVSent	0.64	0.44	0.58

Fine-grained Class

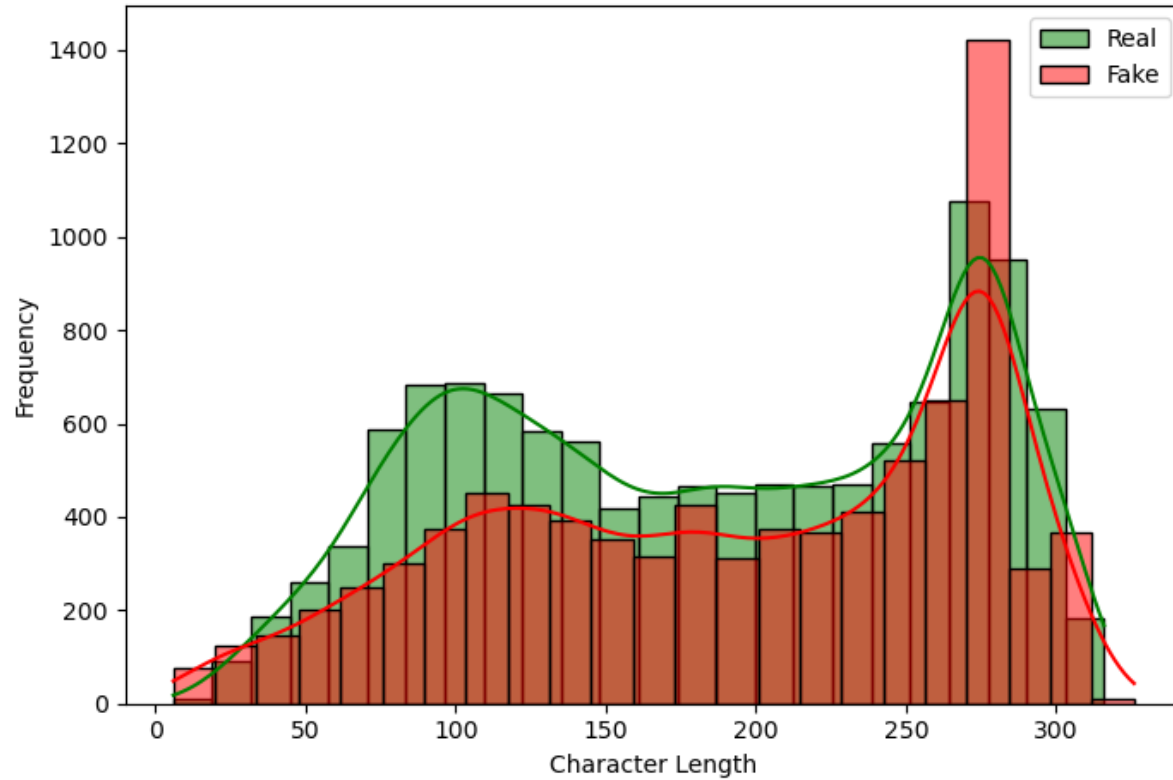
Features	SVM	RndFor	LogFor
TF-IDF	0.40	0.48	0.54
BOW	0.50	0.48	0.54
Word2Vec	0.23	0.27	0.29
GVSent	0.27	0.28	0.29

Transformer Models

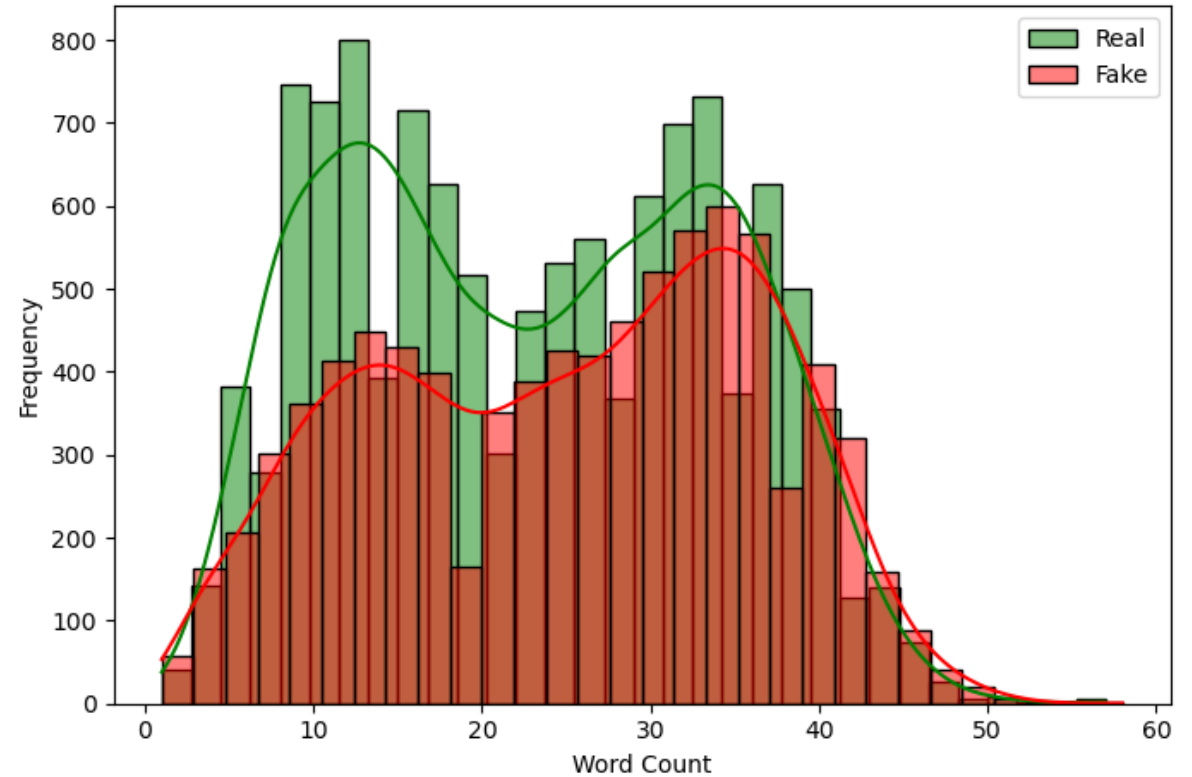
Model	Binary	Fine-grained
BERT-base	0.78	0.49
BERT-Multilingual	0.80	0.61
BERT-German	0.86	0.65
Roberta	0.82	0.58

Textual Distribution Real vs Fake posts

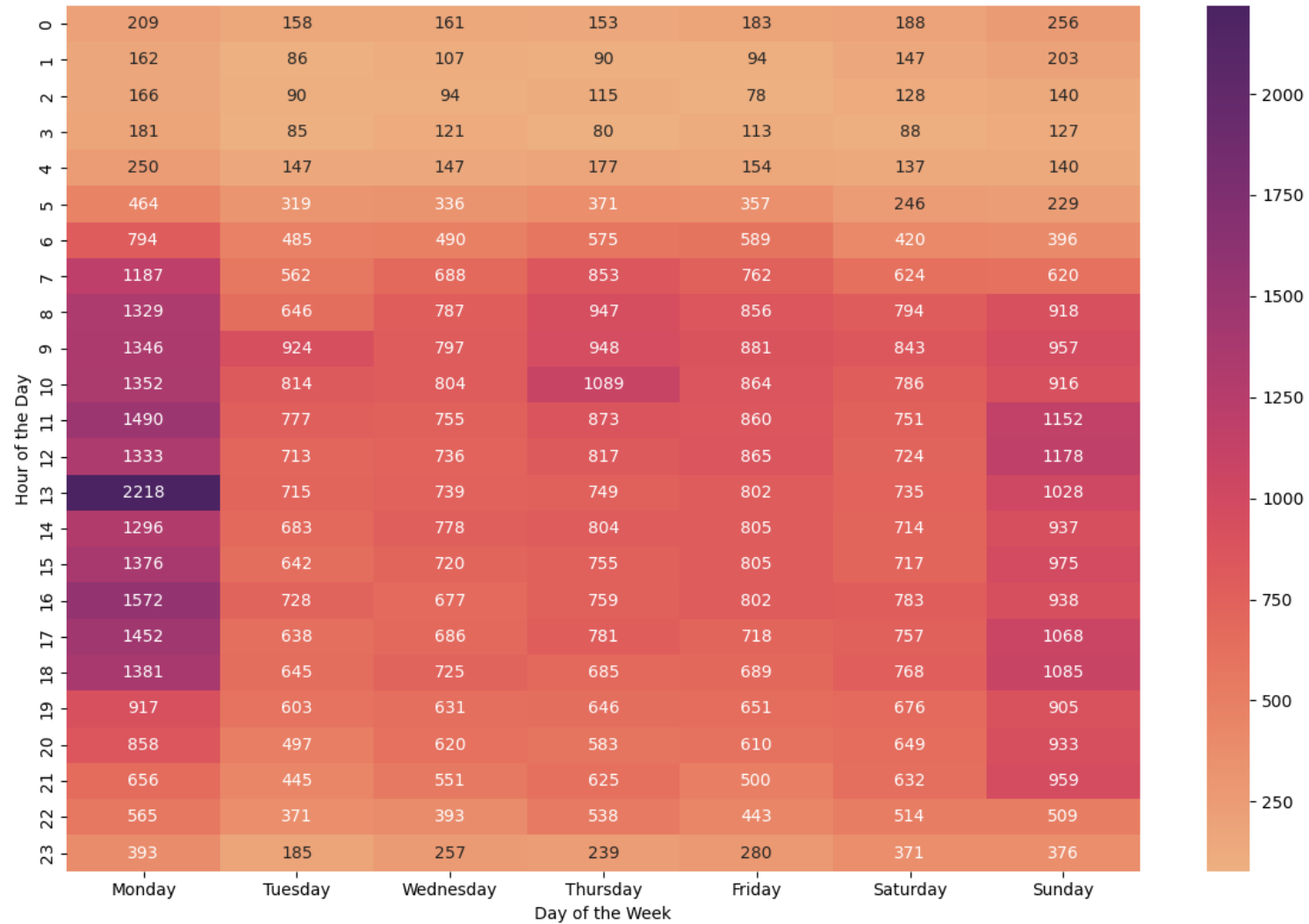
Distribution of Tweet Character Length



Distribution of Tweet Word Count



Tweet activity



What comes next..

- Leveraging advanced LLM's :
 - In-Context Learning and Prompt-Based approaches
 - Zero-shot , Few-shots
 - Instruction tuning
- Incorporating External Knowledge :
 - Knowledge graphs e.g., Wikidata, Wikipedia

****THANK YOU****
ANY QUESTIONS?