



POLITECNICO
MILANO 1863

**Pattern Recognition and
Human Language Technology**
Research Center



**UNIVERSITAT
POLITÀCNICA
DE VALÈNCIA**



Conspiracy vs Critical Thinking Using an Ensemble of Transformers with Data Augmentation Techniques

ANGELO MAXIMILIAN TULBURE¹

MARIONA COLL ARDANUY²

¹ Politecnico di Milano, Italy

² Universitat Politècnica de València, Spain

11/09/2024

OUTLINE

- 01 Task Description**
- 02 Dataset Description**
- 03 Experimental setup**
- 04 Task 1**
- 05 Task 2**
- 06 Evaluation Results**
- 07 Official Ranking**

Task Description

The project focuses on two primary tasks:

1. Binary Classification:

Determine whether a text is Conspiratorial or Critical

2. Span-Level Detection:

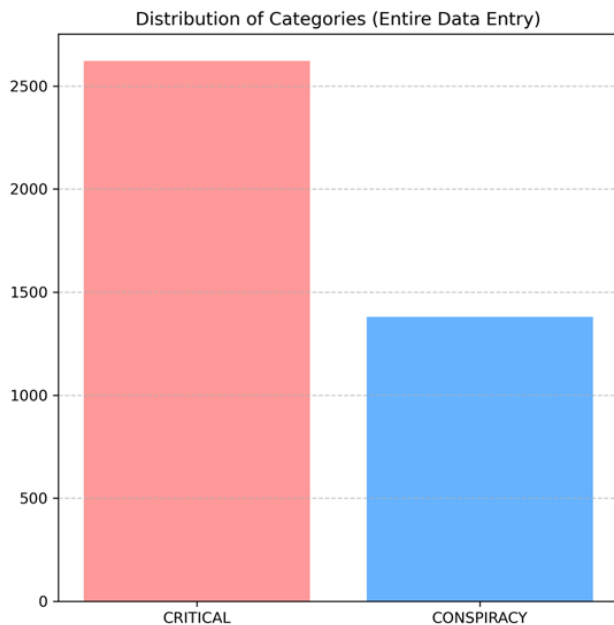
Detect elements of oppositional narratives in the texts
(AGENT, FACILITATOR, VICTIM, CAMPAIGNER, OBJECTIVE, NEGATIVE_EFFECT)

Datasets provided:

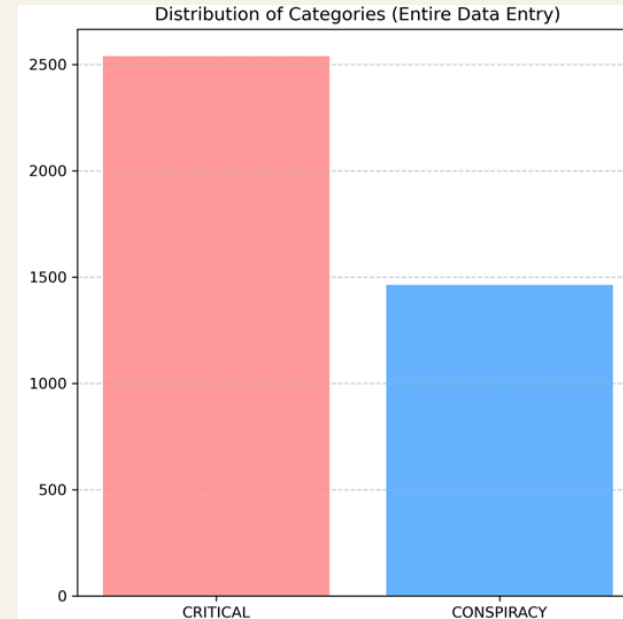
- 5K Telegram comments in *English* (4K for training and 1K for testing)
- 5K Telegram comments in *Spanish* (4K for training and 1K for testing)

Task 1 - Dataset Description

English Dataset



Spanish Dataset



Label	English	(%)	Spanish	(%)
CRITICAL	2621	65.53%	2538	63.45%
CONSPIRACY	1379	34.48%	1462	36.55%

Task 1 - Text example

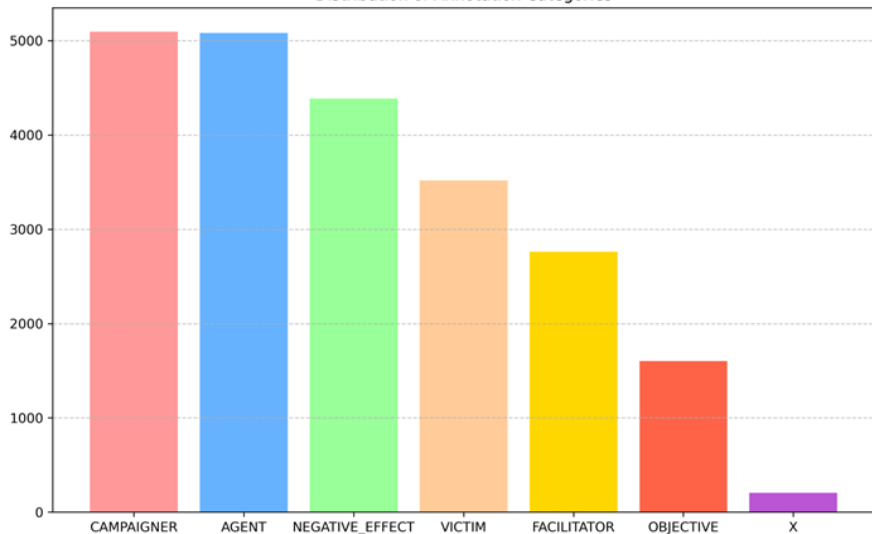
Text	Category
The new lockdown is primarily about giving the police leeway to be more draconian in enforcing it. Expect many more examples of wanton authoritarianism and brutality to follow. https://summit.news/2021/01/06/lockdown-cops-to-stop-people-in-the-street-issue-fines-target-anti-lockdown-anti-vaccine-protesters/	CRITICAL
New Study finds Covid Vaccination can cause Children to suffer Vaccine-Associated Enhanced Disease. Fully vaccinated children are suffering brain swelling, blood clots, sepsis, toxic shock syndrome, and autoimmune disease following Covid-19 injection. https://expose-news.com/2022/06/08/study-covid-vaccinated-children-suffer-vaed-pfizer-fda/ Join us at: t.me/AgentsOfTruth t.me/AgentsOfTruthChat	CRITICAL
Devastating news for every trusting human who has now injected a substance whose inventor stated no human or animal blood can survive longer than 8 years if Graphene Oxide is in it. It is inside the Pfizer and other Covid-19 vaccines & now in Billions of human beings. SHARE! JOIN: https://t.me/QNewsOfficialTV	CRITICAL

Text	Category
ITS IMPORTANT WE DON'T FORGET WHAT HAPPENED - THE CULT ARE RETRENCHING AND WILL BE BACK WITH SPREADING THEIR HIV AVIAN FLU - MARBURG - LASSA FEVER BLEEDING COVER FOR THE 5G ATTACKING THE VACCINATED	CONSPIRACY
WHY THEY INSTALLED THE URBAN RADAR WITH TARGET ACQUISITION CAPABILITY - COVID 19 VACCINE ANTENNA READY FOR THE MASS MURDER - INFO @ SAVEUSNOW.ORG.UK BE THE RESISTANCE	CONSPIRACY
PfizerGate Scandal: The Worldwide Cover-up of Data to disguise the fact Covid-19 Vaccines cause VAIDS. Health authorities around the world are manipulating figures in an attempt to hide from the general public that the Covid-19 injections are causing the fully vaccinated to develop Vaccine Acquired Immune Deficiency Syndrome; and we can prove it. . .	CONSPIRACY

Task 2 - Dataset Description

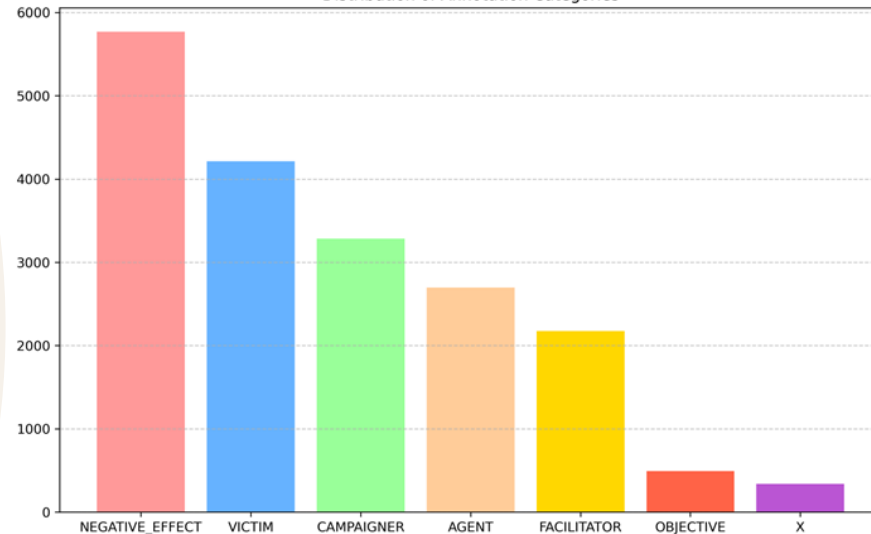
English Dataset

Distribution of Annotation Categories



Spanish Dataset

Distribution of Annotation Categories



Label	English	(%)	Spanish	(%)
CAMPAIGNER	5096	22.70%	3285	17.63%
AGENT	5082	22.63%	2698	14.47%
NEGATIVE_EFFECT	4387	19.54%	5770	30.96%
VICTIM	3517	15.67%	4213	22.61%
FACILITATOR	2763	12.31%	2174	11.67%
OBJECTIVE	1602	7.14%	493	2.65%

Task 2 - Text example

Span Text	Category
the globalists	AGENT
THE CULT	AGENT
Pfizer	AGENT
the police	FACILITATOR
inventor	FACILITATOR
Health authorities around the world	FACILITATOR
Fully vaccinated children	VICTIM
every trusting human who has now injected a substance whose inventor stated no human or animal blood can survive longer than 8 years if Graphene Oxide is in it	VICTIM
Billions of human beings	VICTIM

Span Text	Category
summit	CAMPAIGNER
expose - news	CAMPAIGNER
anti - lockdown - anti - vaccine - protesters	CAMPAIGNER
population control	OBJECTIVE
to destroy our way of life	OBJECTIVE
THE MASS MURDER	OBJECTIVE
The new lockdown is primarily about giving the police leeway to be more draconian in enforcing it	NEGATIVE_EFFECT
Fully vaccinated children are suffering brain swelling, blood clots, sepsis, toxic shock syndrome and autoimmune disease following Covid-19 injection	NEGATIVE_EFFECT
THE CULT ARE RETRENCHING AND WILL BE BACK WITH SPREADING THEIR HIV AVIAN FLU - MARBURG - LASSA FEVER BLEEDING COVER FOR THE 5G ATTACKING THE VACCINATED	NEGATIVE_EFFECT

Experimental Setup

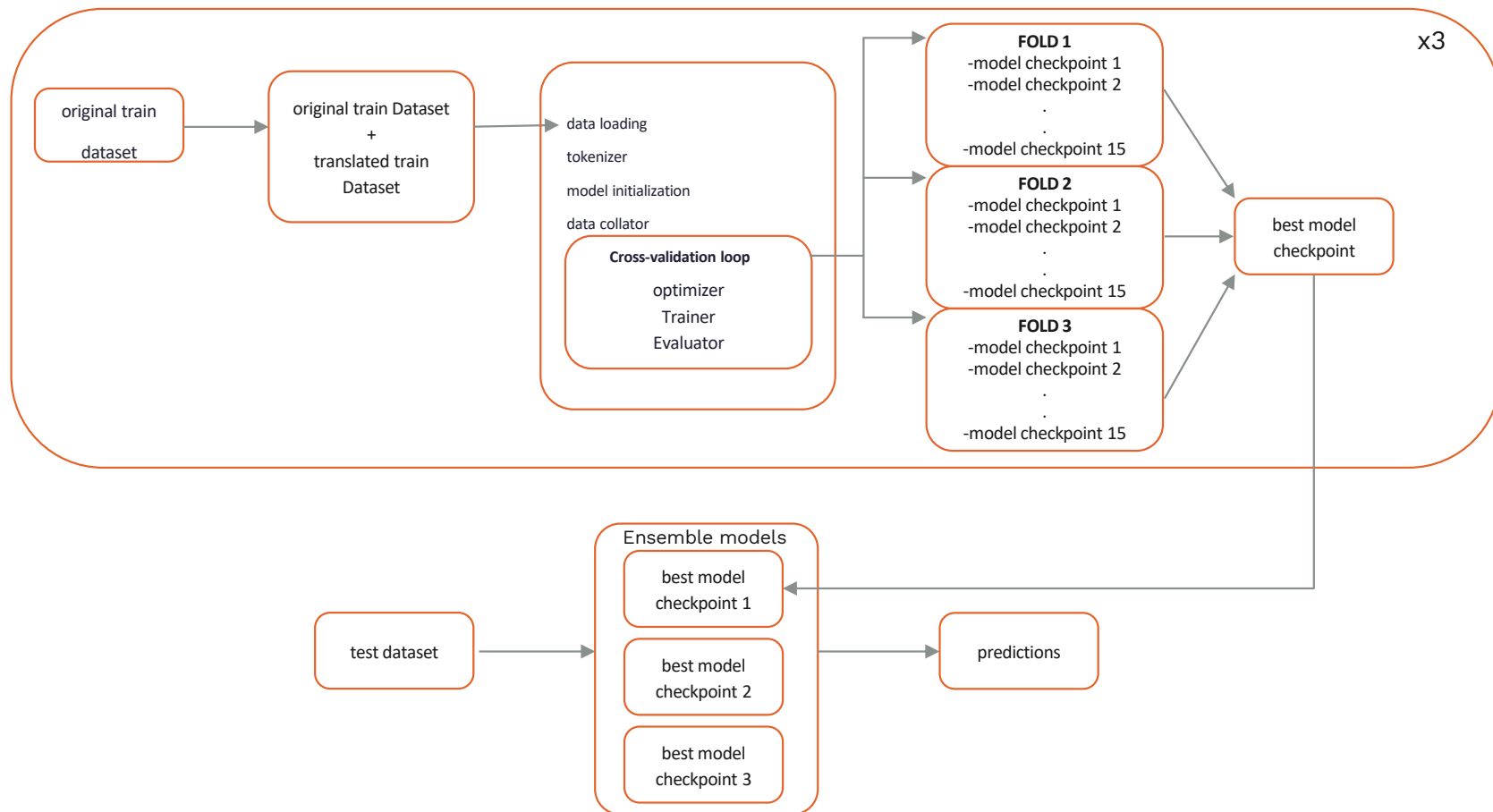
Experiments were conducted on an NVIDIA GeForce GTX 1080

learning_rate	linear_learning_rate_scheduler
evaluation_strategy	epoch
save_strategy	epoch
num_train_epochs	15
weight_decay	0.01
load_best_model_at_end	True
metric_for_best_model	f1
gradient_accumulation_steps	4
per_device_train_batch_size	batch_size ^(*)
per_device_eval_batch_size	batch_size

(*) DynamicBatchSizeCallback

Task 1 - Approach

- Fine-tuning models with various hyperparameters
- StratifiedKFold cross validation
- Data augmentation via translation-based techniques:
 - **English**: Translated Spanish dataset to English
(Helsinki-NLP/opus-mt-es-en)
 - **Spanish**: Translated English dataset to Spanish
(Helsinki-NLP/opus-mt-en-es)
- *Run 1* : Predictions using the best model checkpoint
- *Run 2* : Predictions using Soft Voting Ensembling method



Task 1 English - Data Augmentation

"No tengo idea, pero si ya se ha vacunado el 92%de los Españoles, mi intuición me dice que no hace falta la obligación."

Helsinki-NLP/opus-mt-es-en

"I have no idea, but if 92% of the Spaniards have already been vaccinated, my intuition tells me that the obligation is not necessary."

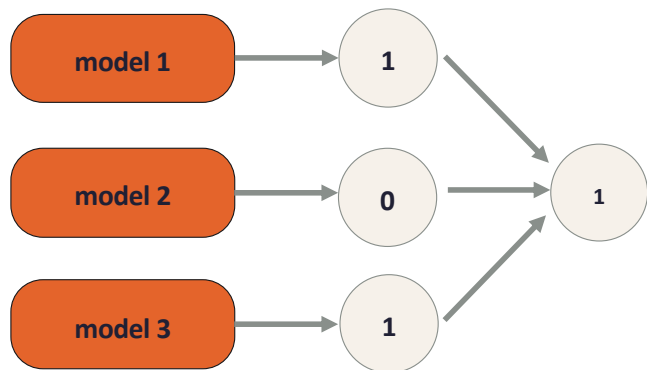
Task 1 Spanish - Data Augmentation

"2021 : They wanted to know your vaccination status and see your papers to be allowed to go to restaurants. 2023 : They don't want you to know the vaccination status of someone who died suddenly."

Helsinki-NLP/opus-mt-en-es

"2021 : Querían saber su estado de vacunación y ver sus papeles para que se les permita ir a los restaurantes. 2023 : No quieren que usted sepa el estado de vacunación de alguien que murió de repente."

Task 1 - Ensembling Models

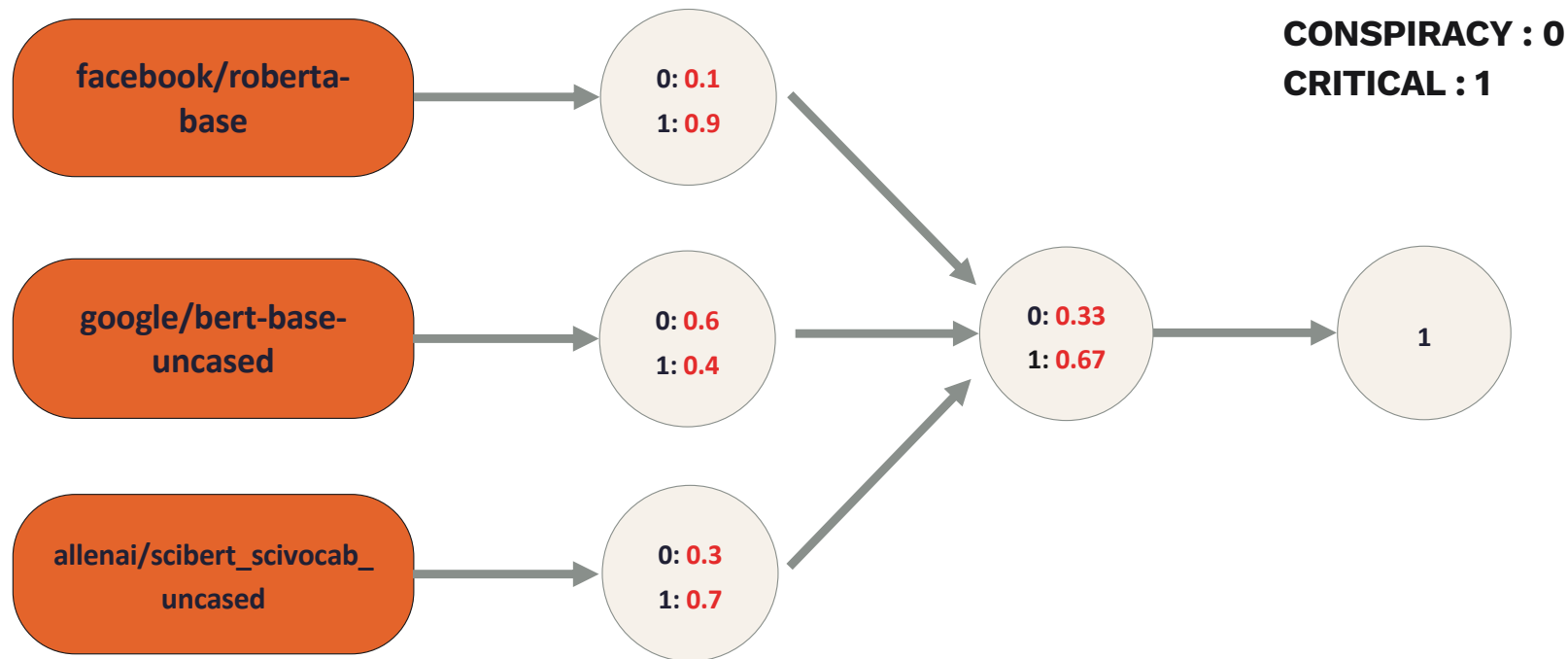


Hard Voting

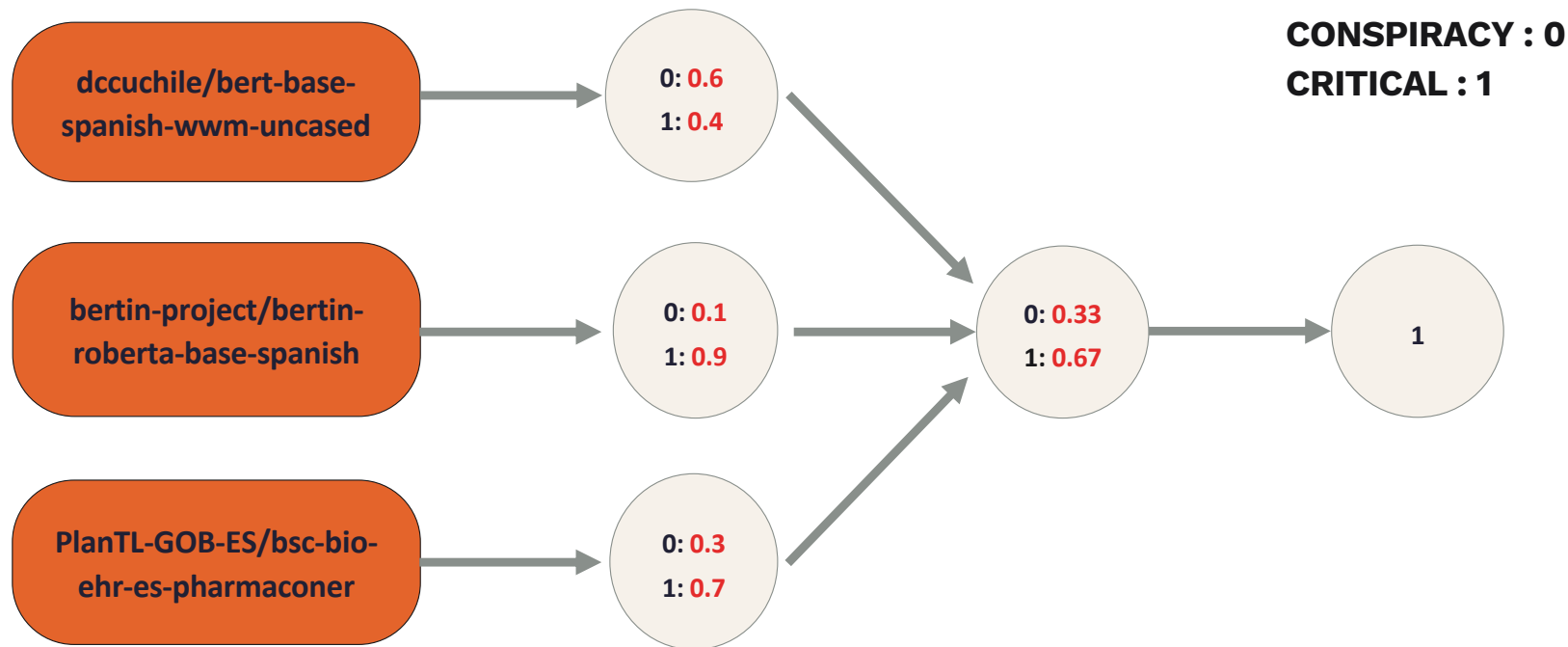


Soft Voting

Task 1 English - Models used



Task 1 Spanish - Models used

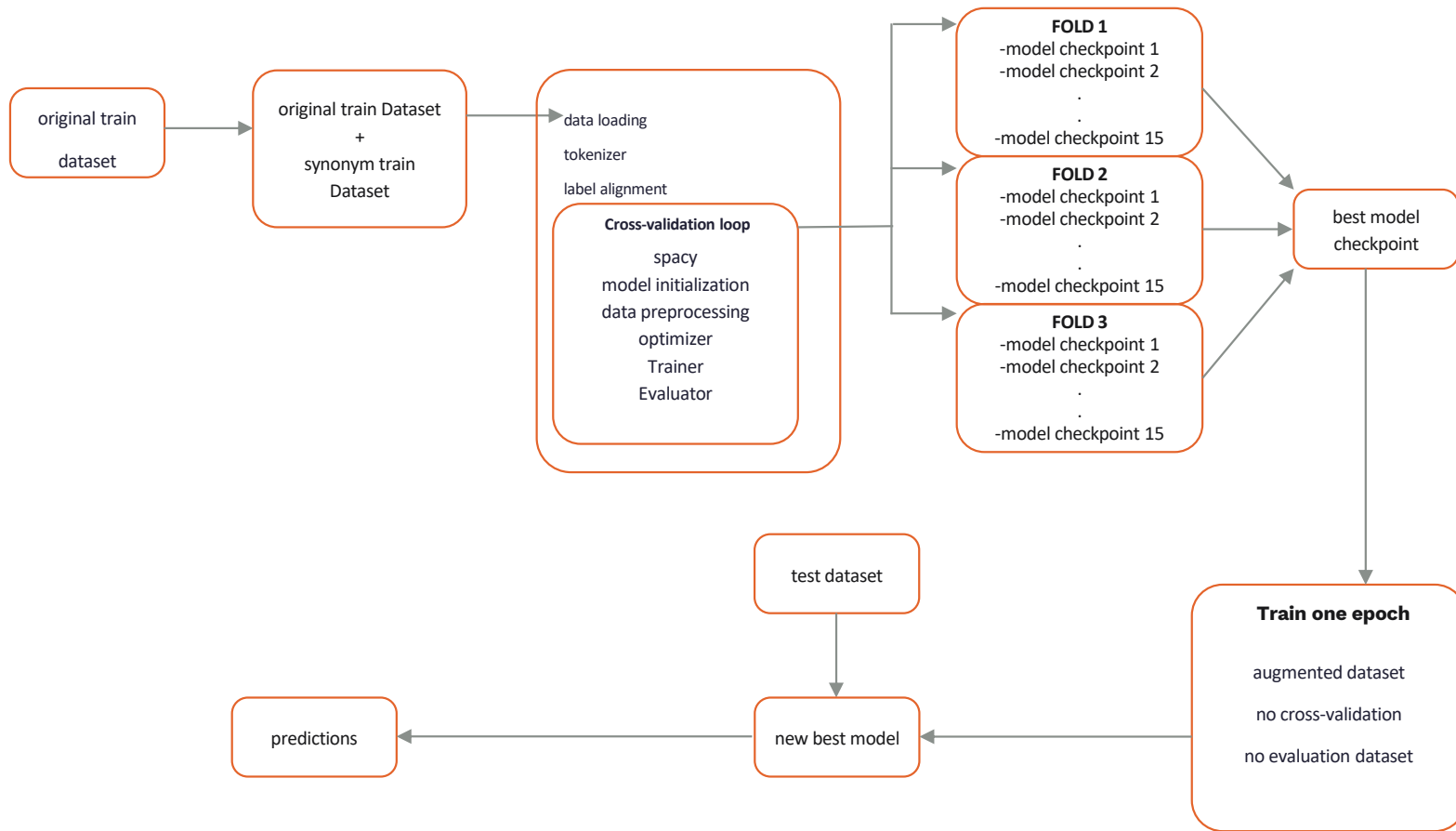


Task 2 - Approach

We treated it as a token classification problem:

- *Fine-tuning a transformer model with a token classification head*
- *Sentence-level processing to avoid text truncation*
- *Data augmentation via synonym replacement:*
 - English:** static word embeddings (word2vec) and SpaCy
(GoogleNews-vectors-negative300)
 - Spanish:** Translated English dataset to Spanish
(FastText embeddings from the Spanish Unannotated Corpora)
- *Run 1* : Using the best model checkpoint
- *Run 2* : Additional training with the augmented dataset

Task 2 - Approach



Task 2 English - Data Augmentation

I'm *deeply concerned* that the *push* to *vaccinate* these *children* is nothing *more* than a *dystopian experiment* with unknown *consequences* .” — Rep. Louie Gohmert (R - Texas) # TheDefender."

word2vec and SpaCy (GoogleNews-vectors-negative300)

I'm *profoundly worried* that the *pushing* to *vaccinated* these *kids* is nothing *less* than a *dystopic experiments* with unknown *repercussion* .” — Rep. Louie Gohmert (R - Texas) # TheDefender."

"text": "I 'm **deeply concerned** that the **push** to **vaccinate** these **children** is nothing **more** than a **dystopian experiment** with unknown **consequences** ." — Rep. Louie Gohmert (R - Texas) # TheDefender."

"annotations": [

```
{
  "span_text": "I 'm deeply concerned that the push to vaccinate these children is
    nothing more than a dystopian experiment with unknown consequences",
  "category": "NEGATIVE_EFFECT",
  "start_char": 2,
  "end_char": 135,
  "start_spacy_token": 1,
  "end_spacy_token": 22
},
{
  "span_text": "to vaccinate these children",
  "category": "OBJECTIVE",
  "start_char": 38,
  "end_char": 65,
  "start_spacy_token": 8,
  "end_spacy_token": 12
},
{
  "span_text": "these children",
  "category": "VICTIM",
  "start_char": 51,
  "end_char": 65,
  "start_spacy_token": 10,
  "end_spacy_token": 12
},
}
```

———— word2vec and SpaCy (GoogleNews-vectors-negative300) ———→

"text": "I 'm **profoundly worried** that the **pushing** to **vaccinated** these **kids** is nothing **less** than a **dystopic experiments** with unknown **repercussion** ." — Rep. Louie Gohmert (R - Texas) # TheDefender."

"annotations": [

{

"span_text": "I 'm **profoundly worried** that the **pushing** to **vaccinated** these **kids** is nothing **less** than a **dystopic experiments** with unknown **repercussion**",

"category": "NEGATIVE_EFFECT",

"start_char": 2,

"end_char": 138,

"start_spacy_token": 1,

"end_spacy_token": 22

},

{

"span_text": "to **vaccinated** these **kids**",

"category": "OBJECTIVE",

"start_char": 41,

"end_char": 68,

"start_spacy_token": 8,

"end_spacy_token": 12

},

{

"span_text": "these **kids**",

"category": "VICTIM",

"start_char": 52,

"end_char": 68,

"start_spacy_token": 10,

"end_spacy_token": 12

},

Task 2 Spanish - Data Augmentation

El *ex vicepresidente* de Pfizer *vuelve* a *advertir* sobre las vacunas COVID y el '*riesgo severo*' de *infertilidad* para las *mujeres*

FastText embeddings from the Spanish Unannotated Corpora

El *convicto presidente* de Pfizer *regresa* a *alertar* sobre las vacunas COVID y el '*riesgos severa*' de *endometriosis* para las *niñas*

"text": "El **ex vicepresidente** de Pfizer **vuelve a advertir** sobre las vacunas COVID y el '**riesgo severo**' de **infertilidad** para las **mujeres**"

"annotations": [

```
{
  "span_text": "El ex vicepresidente de Pfizer",
  "category": "CAMPAIGNER",
  "start_char": 0,
  "end_char": 30,
  "start_spacy_token": 0,
  "end_spacy_token": 5
},
```

```
{
  "span_text": "Pfizer",
  "category": "AGENT",
  "start_char": 24,
  "end_char": 30,
  "start_spacy_token": 4,
  "end_spacy_token": 5
},
```

```
{
  "span_text": "el 'riesgo severo' de infertilidad para las mujeres",
  "category": "NEGATIVE_EFFECT",
  "start_char": 75,
  "end_char": 128,
  "start_spacy_token": 13,
  "end_spacy_token": 23
}
]
```

———— FastText embeddings from the Spanish Unannotated Corpora ————>

"text": "El **convicto presidente** de Pfizer **regresa a alertar** sobre las vacunas COVID y el '**riesgos severa**' de **endometriosis** para las **niñas**"

"annotations": [

```
{
  "span_text": "El convicto presidente de Pfizer",
  "category": "CAMPAIGNER",
  "start_char": 0,
  "end_char": 32,
  "start_spacy_token": 0,
  "end_spacy_token": 5
},
```

```
{
  "span_text": "Pfizer",
  "category": "AGENT",
  "start_char": 26,
  "end_char": 32,
  "start_spacy_token": 4,
  "end_spacy_token": 5
},
```

```
{
  "span_text": "el 'riesgos severa' de endometriosis para las niñas",
  "category": "NEGATIVE_EFFECT",
  "start_char": 76,
  "end_char": 129,
  "start_spacy_token": 13,
  "end_spacy_token": 23
}
```

```
]
```

Task 2 - Models used

**Model
used for
English:**

facebook/roberta-base

**Model
used for
Spanish:**

PlanTL-GOB-ES/roberta-base-bne

Task 1 - Evaluation results

	MCC	Subtask 1		
		F1-macro	F1-conspiracy	F1-critical
English_baseline	<i>0.7964</i>	<i>0.8975</i>	<i>0.8632</i>	<i>0.9318</i>
English_run1	<i>0.7574</i>	<i>0.8769</i>	<i>0.8338</i>	<i>0.9200</i>
English_run2	<i>0.7872</i>	<i>0.8917</i>	<i>0.8536</i>	<i>0.9297</i>
Spanish_baseline	<i>0.6681</i>	<i>0.8339</i>	<i>0.7872</i>	<i>0.8806</i>
Spanish_run1	<i>0.6147</i>	<i>0.7950</i>	<i>0.7179</i>	<i>0.8720</i>
Spanish_run2	<i>0.6722</i>	<i>0.8293</i>	<i>0.7699</i>	<i>0.8887</i>

Task 2 - Evaluation results

	span-P	span-R	Subtask 2	micro-span-F1
			span-F1	
English_baseline	<i>0.5323</i>	<i>0.4684</i>	<i>0.6334</i>	<i>0.4998</i>
English_run1	<i>0.5832</i>	<i>0.6856</i>	<i>0.6293</i>	<i>0.6074</i>
English_run2	<i>0.5859</i>	<i>0.6790</i>	<i>0.6279</i>	<i>0.6120</i>
Spanish_baseline	<i>0.4533</i>	<i>0.5621</i>	<i>0.4934</i>	<i>0.4952</i>
Spanish_run1	<i>0.5997</i>	<i>0.6193</i>	<i>0.6089</i>	<i>0.6051</i>
Spanish_run2	<i>0.6159</i>	<i>0.6129</i>	<i>0.6129</i>	<i>0.6108</i>

Official Ranking

TASK 1 - ENGLISH

POSITION	TEAM	MCC	F1-MACRO	F1-CONSPIRACY	F1-CRITICAL
	<i>baseline-BERT</i>	0.7964	0.8975	0.8632	0.9318
18	aish_team	0.7917	0.8944	0.8580	0.9309
19	rfenthusiasts	0.7902	0.8948	0.8605	0.9291
20	Dap_upv	0.7898	0.8944	0.8593	0.9294
21	oppositional_opposition	0.7894	0.8935	0.8571	0.9300
22	RD-IA-FUN	0.7894	0.8947	0.8617	0.9276
23	miqarn	0.7881	0.8938	0.8593	0.9283
24	CHEEXIST	0.7875	0.8932	0.8576	0.9287
25	tulbure	0.7872	0.8917	0.8536	0.9297
26	XplaiNLP	0.7871	0.8922	0.8550	0.9294
27	TheGymNerds	0.7854	0.8923	0.8567	0.9278
28	nlpIn	0.7844	0.8922	0.8580	0.9263
29	RalloRico	0.7771	0.8879	0.8559	0.9198
30	LasGarcias	0.7758	0.8855	0.8447	0.9263

↓ ~1%

Official Ranking

TASK 1 - SPANISH

POSITION	TEAM	MCC	F1-MACRO	F1-CONSPIRACY	F1-CRITICAL
1	SINAI	0.7429	0.8705	0.8319	0.9091
2	auxR	0.7205	0.8572	0.8112	0.9032
3	RD-IA-FUN	0.7028	0.8497	0.8035	0.8960
4	Elias&Sergio	0.6971	0.8485	0.8087	0.8884
5	AI_Fusion	0.6872	0.8419	0.7931	0.8908
6	zhengqiaozeng	0.6871	0.8417	0.7925	0.8909
7	virmel	0.6854	0.8426	0.8022	0.8831
8	trustno1	0.6848	0.8400	0.7895	0.8906
9	Zleon	0.6826	0.8410	0.7955	0.8865
10	ojo-bes	0.6817	0.8395	0.8026	0.8764
11	tulbure	0.6722	0.8293	0.7699	0.8887
12	sail	0.6719	0.8299	0.7713	0.8884
13	nlpln	0.6681	0.8339	0.7872	0.8806
	<i>baseline-BERT</i>	<i>0.6681</i>	<i>0.8339</i>	<i>0.7872</i>	<i>0.8806</i>

↑ ~0.4%

Official Ranking

TASK 2 - ENGLISH

POSITION	TEAM	span-F1	span-P	span-R	micro-span-F1
1	tulbure	0.6279	0.5859	0.6790	0.6120
2	Zleon	0.6089	0.5537	0.6881	0.5856
3	hinlole	0.5886	0.5243	0.6834	0.5571
4	oppositional_opposition	0.5866	0.5347	0.6586	0.5344
5	AI_Fusion	0.5805	0.5585	0.6082	0.5437
6	virmel	0.5742	0.5235	0.6477	0.5540
7	miqarn	0.5739	0.5184	0.6462	0.5325
8	TargaMarhuenda	0.5701	0.5161	0.6477	0.5437
9	ezio	0.5694	0.5229	0.6340	0.5389
10	zhengqiaozeng	0.5666	0.5122	0.6485	0.5421
11	Elias&Sergio	0.5627	0.5149	0.6364	0.5248
12	DSVS	0.5598	0.5332	0.6012	0.5287
13	CHEEXIST	0.5524	0.4767	0.6845	0.5299
14	rfenthusiasts	0.5479	0.5381	0.5666	0.5408
15	ALC_UPV_JD_2	0.5377	0.4643	0.6562	0.4956
	<i>baseline-BETO</i>	<i>0.5323</i>	<i>0.4684</i>	<i>0.6334</i>	<i>0.4998</i>

↑ ~11%

Official Ranking

TASK 2 - SPANISH

POSITION	TEAM	span-F1	span-P	span-R	micro-span-F1
1	tulbure	0.6129	0.6159	0.6129	0.6108
2	Zleon	0.5875	0.5439	0.6474	0.5939
3	AI_Fusion	0.5777	0.5437	0.6189	0.5843
4	CHEEXIST	0.5621	0.5379	0.5995	0.5456
5	virmel	0.5616	0.4963	0.6584	0.5620
6	miqarn	0.5603	0.5117	0.6273	0.5618
7	DSVS	0.5529	0.5384	0.5785	0.5323
8	TargaMarhuenda	0.5364	0.5128	0.5710	0.5385
9	Elias&Sergio	0.5151	0.4864	0.5533	0.5231
10	hinlole	0.4994	0.4530	0.5740	0.4890
	<i>baseline-BETO</i>	<i>0.4934</i>	<i>0.4533</i>	<i>0.5621</i>	<i>0.4952</i>

↑ ~12%

**Thanks for your
attention!**

Any question?

Contacts:

angelomaximilian.tulbure@mail.polimi.it

angelotulbure29@outlook.it



Angelo Tulbure

MSc student in Computer Engineering at
Politecnico di Milano

