POLITECNICO
MILANO 1863

Pattern Recognition and
Human Language Technology
Research Center



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

# Conspiracy Theories vs Critical Thinking Using an Ensemble of Transformers and Large Language Models

Angelo Maximilian Tulbure

Advisor: Prof. Mark James Carman
Co-advisors: Prof. Paolo Rosso, Mariona Coll Aldanuy (UPV)

11/12/2024

# PAN 2024 Oppositional Thinking Analysis Tasks Description



The challenge 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:

- 5.000 Telegram comments in *English* (4.000 for training and 1.000 for testing)
- 5.000 Telegram comments in *Spanish* (4.000 for training and 1.000 for testing)

# Research Questions

**RQ1. How effectively can transformer models and Large Language Models distinguish between conspiracy theories and critical thinking narratives, and which factors impact their performance?**

*Subquestions:*

- What are the main challenges in differentiating linguistic and rhetorical features between conspiratorial and critical thinking narratives?
- How do the performances of different models vary across languages?

**RQ2. How feasible is it to accurately identify text spans that correspond to key elements within oppositional narratives?**

*Subquestions:*

- What are the primary difficulties in detecting partially overlapping narrative elements in multilingual datasets?
- How does the context in which narrative elements appear influence their identification accuracy across different models?

**RQ3. To what extent do data augmentation techniques enhance model performance across the two tasks?**

*Subquestions:*

- Which data augmentation techniques contribute most significantly to improving model performance in binary classification and span-level detection?
- How does data augmentation impact the model's ability to generalize across new or unseen narrative structures?

# Critical Thinking vs Conspiracy Theories

**CRITICAL THINKING:** critical messages that question major decisions in the public health domain, but do not promote a conspiracist mentality.

**CONSPIRACY THEORIES :** Texts that view the pandemic or public health decisions as a result of a malevolent conspiracy by secret, influential groups.

# Elements of Oppositional Narratives

**Agents** (A) : Responsible for the action or negative effect described in the comment.

**Victims** (V) : Those who suffer the consequences of the actions of the agents/facilitators.

**Facilitators** (F) : Collaborate with the agents and contribute to the execution of their goals.

**Campaigners** (C) : Those who oppose the mainstream narrative.

**Objectives** (O) : The objectives the agents are trying to achieve.

**Negative Effects** (E) : The negative consequences suffered by the victims.

# Annotation Examples

## Critical Thinking

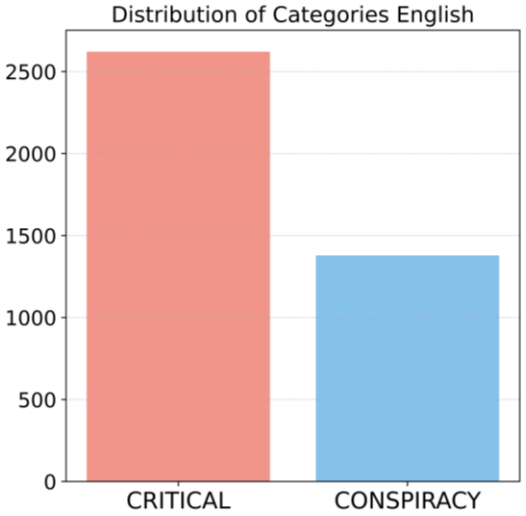
https :// twitter . com / [redacted] Hospitals Should Hire , Not Fire , Nurses with Natural Immunity by Dr Martin Kulldorff C By pushing vaccine mandates O , White House chief medical advisor Dr. Anthony Fauci A is questioning the existence of natural immunity after Covid disease . In doing so , he is 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 ?

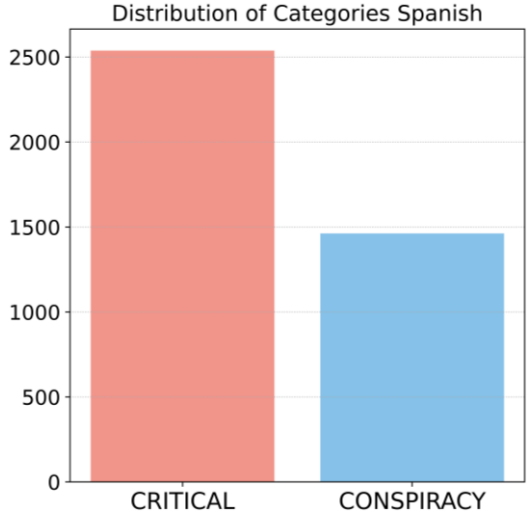
## Conspiracy Theory

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 .

What distinguishes conspiracy from critical narratives? A computational analysis of oppositional discourse - Korenčić - 2024 - Expert Systems - Wiley Online Library

# Task 1 – Dataset Description

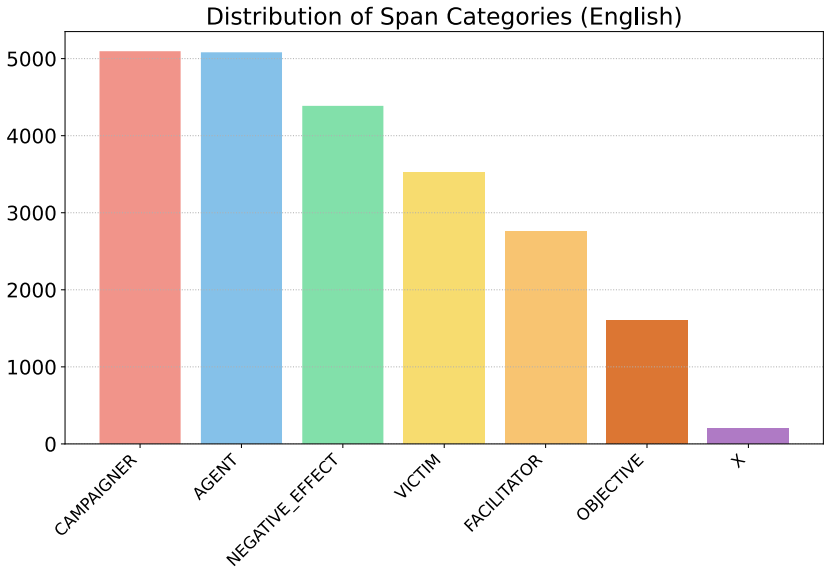


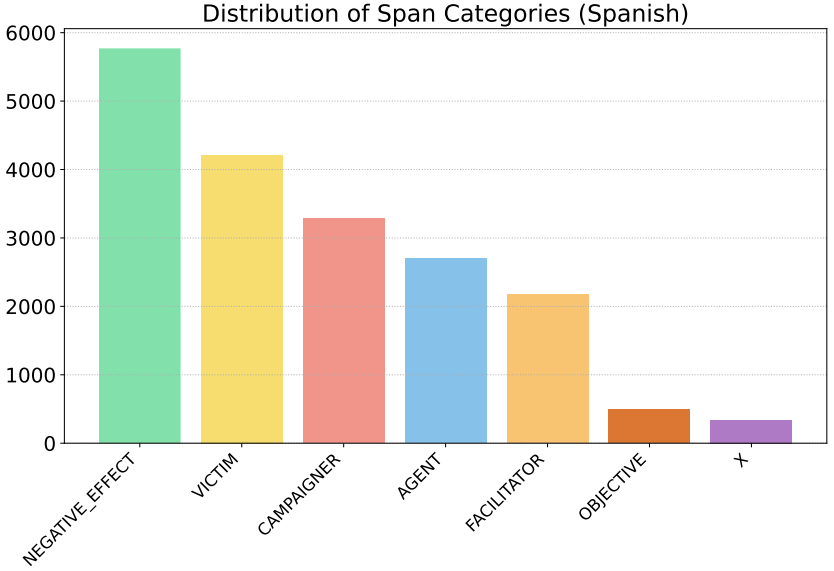




| Label | English | (%) | Spanish | (%) |
|---|---|---|---|---|
| CRITICAL | 2621 | 65.53% | 2538 | 63.45% |
| CONSPIRACY | 1379 | 34.48% | 1462 | 36.55% |

| | Dataset | Mean | Std Deviation | Min | Max |
|---|---|---|---|---|---|
| CRITICAL | English | 476 | 479.3 | 78 | 4695 |
| | Spanish | 641.2 | 577.8 | 99 | 4761 |
| CONSPIRACY | English | 742.9 | 740.2 | 88 | 4346 |
| | Spanish | 1112 | 945.3 | 123 | 4313 |

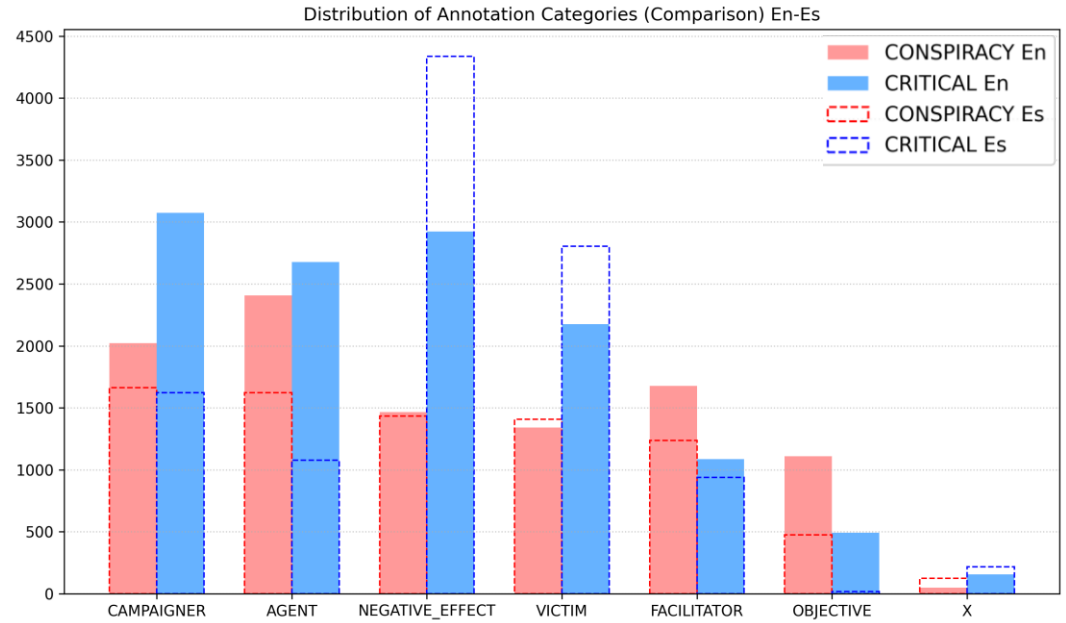
# Task 2 – Dataset Description




| 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% |

# Distribution of Annotation Categories




## English Dataset

| Span Category | CONSPIRACY | (%) | CRITICAL | (%) |
|---|---|---|---|---|
| NEGATIVE_EFFECT | 1465 | 33.39% | 2922 | 66.61% |
| AGENT | 2406 | 47.34% | 2676 | 52.66% |
| CAMPAIGNER | 2023 | 39.68% | 3073 | 60.32% |
| VICTIM | 1341 | 38.13% | 2176 | 61.87% |
| FACILITATOR | 1677 | 60.70% | 1086 | 39.30% |
| OBJECTIVE | 1110 | 69.29% | 492 | 30.71% |
| X | 50 | 24.27% | 156 | 75.73% |

## Spanish Dataset

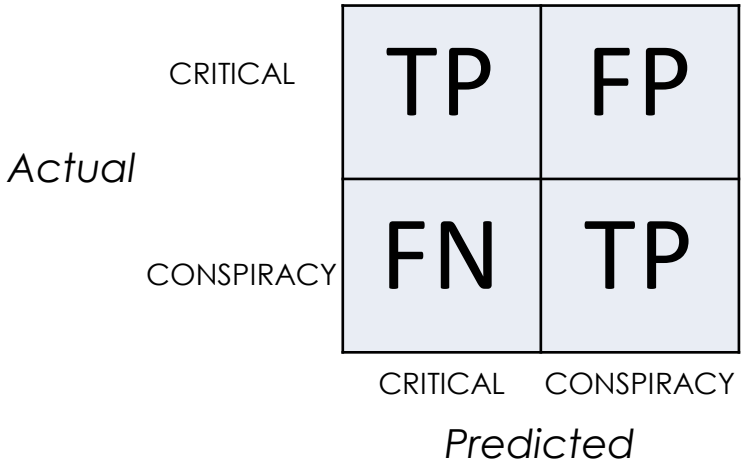
| Span Category | CONSPIRACY | (%) | CRITICAL | (%) |
|---|---|---|---|---|
| NEGATIVE_EFFECT | 1433 | 24.83% | 4337 | 75.17% |
| AGENT | 1621 | 60.07% | 1077 | 39.93% |
| CAMPAIGNER | 1662 | 50.59% | 1623 | 49.41% |
| VICTIM | 1408 | 33.42% | 2805 | 66.58% |
| FACILITATOR | 1236 | 56.87% | 938 | 43.13% |
| OBJECTIVE | 475 | 96.34% | 18 | 3.66% |
| X | 123 | 36.18% | 217 | 63.82% |

# Evaluation Metrics

## Binary Classification

### *Matthews Correlation Coefficient*

$$\text{MCC} = \frac{(TP \cdot TN) - (FP \cdot FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$




## Span-level Detection

### *Span-F1 Metric*

$$P(S,T) = \frac{1}{|S|} \cdot \sum_{d \in D} \sum_{s \in S_d, t \in T_d} \frac{|s \cap t|}{|t|}$$

$$R(S,T) = \frac{1}{|T|} \cdot \sum_{d \in D} \sum_{s \in S_d, t \in T_d} \frac{|s \cap t|}{|t|}$$

$$F1(S,T) = 2 \cdot \frac{P(S,T) \cdot R(S,T)}{P(S,T) + R(S,T)}$$

# Experimental Framework

Experiments were conducted on an NVIDIA GeForce GTX 1080

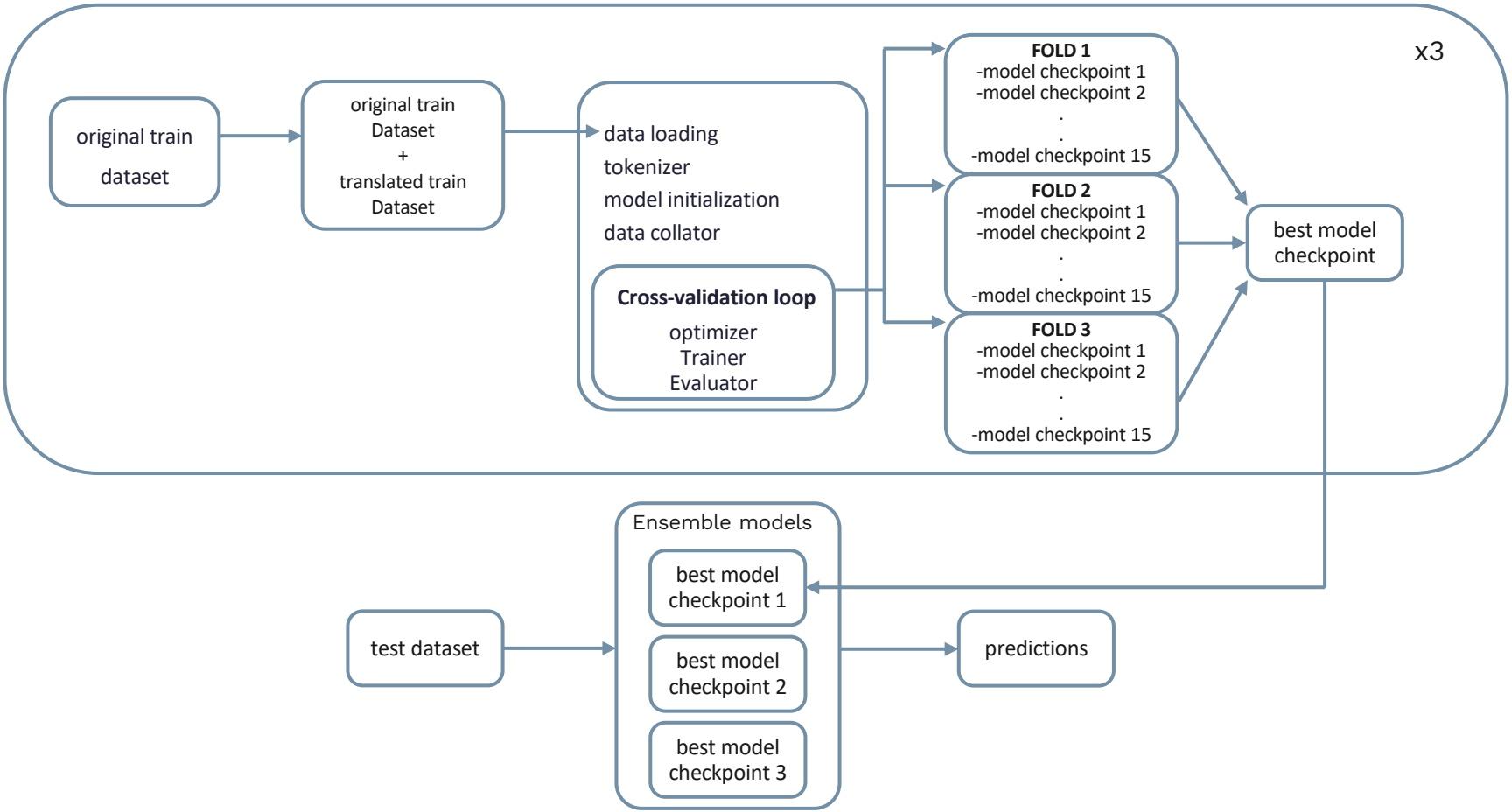
| | |
|---|---|
| learning rate | *linear_learning_rate_scheduler* |
| evaluation_strategy | *epoch* |
| save_strategy | *epoch* |
| num_train_epochs | *15* |
| weight_decay | *0.01* |
| load_best_model_at_end | *True* |
| metric_for_best_model | *f1* |
| gradient_accumulation_steps | *4* |
| per_device_train_batch_size | *batch_size* (*) |
| per_device_eval_batch_size | *batch_size* (*) |

(*) DynamicBatchSizeCallback

# Task 1 - Approach

- Fine-tuning models with various hyperparameters
- Stratified K-Fold 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 - Approach



# Task 1 – Data Augmentation

## *Spanish → English*

"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."

## *English → Spanish*

"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 – Experiments on the Train Dataset

## English

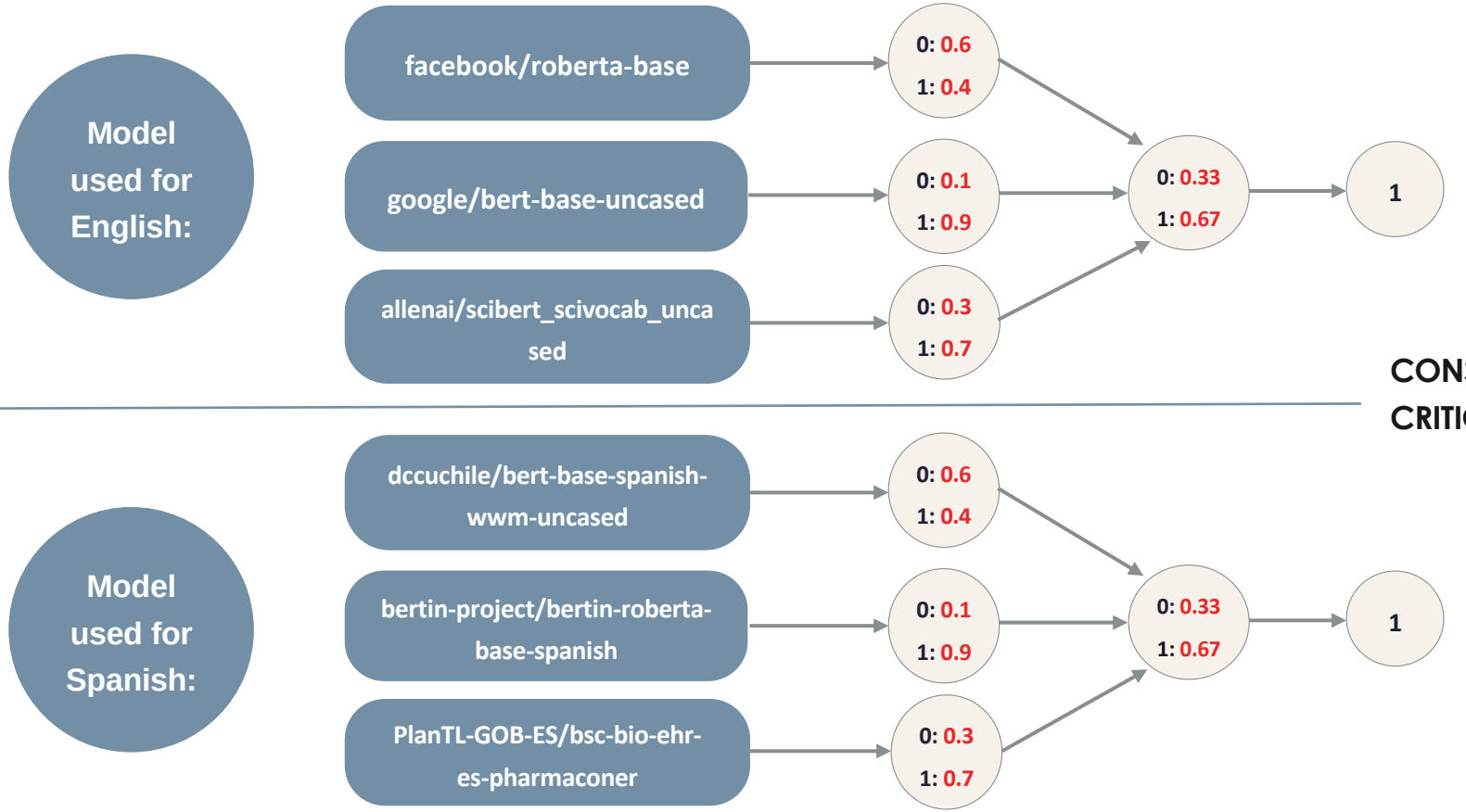
| Model | Epochs | DA* | Acc. | F1 | F1-macro | Prec. | Recall |
|---|---|---|---|---|---|---|---|
| scibert_scivocab_uncased | 15 | ✓ | 0.8637 | 0.8973 | 0.8475 | 0.8734 | **0.9225** |
| bert-base-uncased | 15 | ✓ | 0.8688 | 0.9005 | 0.8539 | 0.8813 | 0.9205 |
| roberta-base | 15 | ✓ | 0.8793 | 0.9072 | 0.8672 | 0.8994 | 0.9151 |
| bert-large-uncased | 5 | ✓ | 0.8631 | 0.8638 | 0.8401 | 0.8645 | 0.8631 |
| distilbert-base-uncased | 5 | ✓ | 0.8629 | 0.8637 | 0.8518 | 0.8640 | 0.8633 |
| distilbert-base-uncased | 3 | ✓ | 0.7901 | 0.7891 | 0.7730 | 0.7887 | 0.7895 |
| ensemble-3-models_HV | - | ✓ | 0.9031 | 0.9101 | 0.8801 | 0.9181 | 0.9022 |
| ensemble-3-models_SV | - | ✓ | **0.9041** | **0.9134** | **0.8916** | **0.9223** | 0.9048 |
| CT-bert | 4 | ✓ | 0.7701 | 0.7412 | 0.7102 | 0.7507 | 0.7317 |
| clf1-basic | - | | 0.846 | 0.881 | 0.831 | 0.875 | 0.887 |
| clf1-spacy | - | | 0.849 | 0.846 | 0.831 | 0.845 | 0.847 |
| clf1-spacy_pos | - | | 0.844 | 0.849 | 0.824 | 0.842 | 0.856 |
| clf2-basic | - | | 0.848 | 0.844 | 0.821 | 0.856 | 0.844 |
| clf2-spacy | - | | 0.865 | 0.848 | 0.844 | 0.863 | 0.854 |
| clf2-spacy_pos | - | | 0.866 | 0.868 | 0.843 | 0.872 | 0.860 |
| clf1-basic | - | ✓ | 0.849 | 0.851 | 0.834 | 0.853 | 0.848 |
| clf1-spacy | - | ✓ | 0.860 | 0.862 | 0.843 | 0.867 | 0.854 |
| clf1-spacy_pos | - | ✓ | 0.858 | 0.861 | 0.841 | 0.868 | 0.852 |
| clf2-basic | - | ✓ | 0.884 | 0.886 | 0.870 | 0.885 | 0.889 |
| clf2-spacy | - | ✓ | 0.888 | 0.890 | 0.872 | 0.888 | 0.893 |
| clf2-spacy_pos | - | ✓ | 0.886 | 0.887 | 0.870 | 0.888 | 0.892 |
| ensemble-3-clf2_HV | - | ✓ | 0.720 | 0.725 | 0.612 | 0.743 | 0.724 |
| ensemble-3-clf2_SV | - | ✓ | 0.731 | 0.718 | 0.644 | 0.738 | 0.724 |
| ensemble-2_models-1_clf2_HV | - | ✓ | 0.889 | 0.871 | 0.865 | 0.879 | 0.869 |
| ensemble-2_models-1_clf2_SV | - | ✓ | 0.885 | 0.867 | 0.861 | 0.870 | 0.865 |
| ensemble-1_model-2_clf2_HV | - | ✓ | 0.887 | 0.865 | 0.859 | 0.871 | 0.856 |
| ensemble-1_models-2_clf2_SV | - | ✓ | 0.838 | 0.881 | 0.874 | 0.887 | 0.874 |

## Spanish

| Model | Epochs | DA* | Acc. | F1 | F1-macro | Prec. | Recall |
|---|---|---|---|---|---|---|---|
| bertin-roberta-base-spanish | 15 | ✓ | 0.8613 | 0.8992 | 0.8384 | 0.8462 | **0.9593** |
| bert-base-spanish-wwm | 15 | ✓ | 0.8687 | 0.9005 | 0.8539 | 0.8813 | 0.9205 |
| bsc-bio-ehr-es-pharmaconer | 15 | ✓ | 0.8669 | 0.8973 | 0.8542 | 0.8934 | 0.9012 |
| roberta-base-bne | 10 | ✓ | 0.8675 | 0.8674 | 0.8550 | 0.8672 | 0.8675 |
| roberta-base | 5 | ✓ | 0.8555 | 0.8529 | 0.8369 | 0.8546 | 0.8555 |
| ensemble-3_models_SV | - | ✓ | **0.8971** | **0.9248** | **0.8841** | **0.9020** | 0.9488 |
| ensemble-3_models_HV | - | ✓ | 0.8802 | 0.9172 | 0.8798 | 0.8910 | 0.9451 |
| alberto-base-spanish | - | ✓ | 0.8612 | 0.8610 | 0.8428 | 0.8609 | 0.8613 |
| classifier1(basic) | - | - | 0.807 | 0.7945 | 0.784 | 0.792 | 0.797 |
| classifier1-(spacy) | - | - | 0.814 | 0.7990 | 0.788 | 0.800 | 0.798 |
| classifier1-(spacy_pos) | - | - | 0.816 | 0.7994 | 0.789 | 0.804 | 0.795 |
| classifier2-(basic) | - | - | 0.809 | 0.7886 | 0.774 | 0.805 | 0.773 |
| classifier2-(spacy) | - | - | 0.818 | 0.7985 | 0.772 | 0.818 | 0.780 |
| classifier2-(spacy_pos) | - | - | 0.829 | 0.8118 | 0.783 | 0.836 | 0.789 |
| classifier1-(basic) | - | ✓ | 0.871 | 0.8604 | 0.810 | 0.857 | 0.864 |
| classifier1-(spacy) | - | ✓ | 0.870 | 0.8575 | 0.812 | 0.858 | 0.857 |
| classifier1-(spacy_pos) | - | ✓ | 0.875 | 0.8625 | 0.815 | 0.864 | 0.861 |
| classifier2-(basic) | - | ✓ | 0.877 | 0.886 | 0.861 | 0.891 | 0.881 |
| classifier2-(spacy) | - | ✓ | 0.896 | 0.885 | 0.864 | 0.896 | 0.875 |
| classifier2-(spacy_pos) | - | ✓ | 0.894 | 0.883 | 0.862 | 0.893 | 0.874 |
| ensemble-3-clf2_HV | - | ✓ | 0.772 | 0.748 | 0.724 | 0.788 | 0.712 |
| ensemble-3-clf2_SV | - | ✓ | 0.798 | 0.778 | 0.729 | 0.815 | 0.745 |
| ensemble-2_models_1_clf_HV | - | ✓ | 0.818 | 0.799 | 0.734 | 0.831 | 0.771 |
| ensemble-2_models_1_clf_SV | - | ✓ | 0.832 | 0.818 | 0.783 | 0.852 | 0.786 |
| ensemble-1_model_2_clf_HV | - | ✓ | 0.820 | 0.803 | 0.749 | 0.836 | 0.773 |
| ensemble-1_models_2_clf_SV | - | ✓ | 0.838 | 0.823 | 0.764 | 0.848 | 0.799 |

DA* = Data Augmented

# Task 1 – Submitted Models



# Task 2 - Approach

We treated it as a token classification problem:

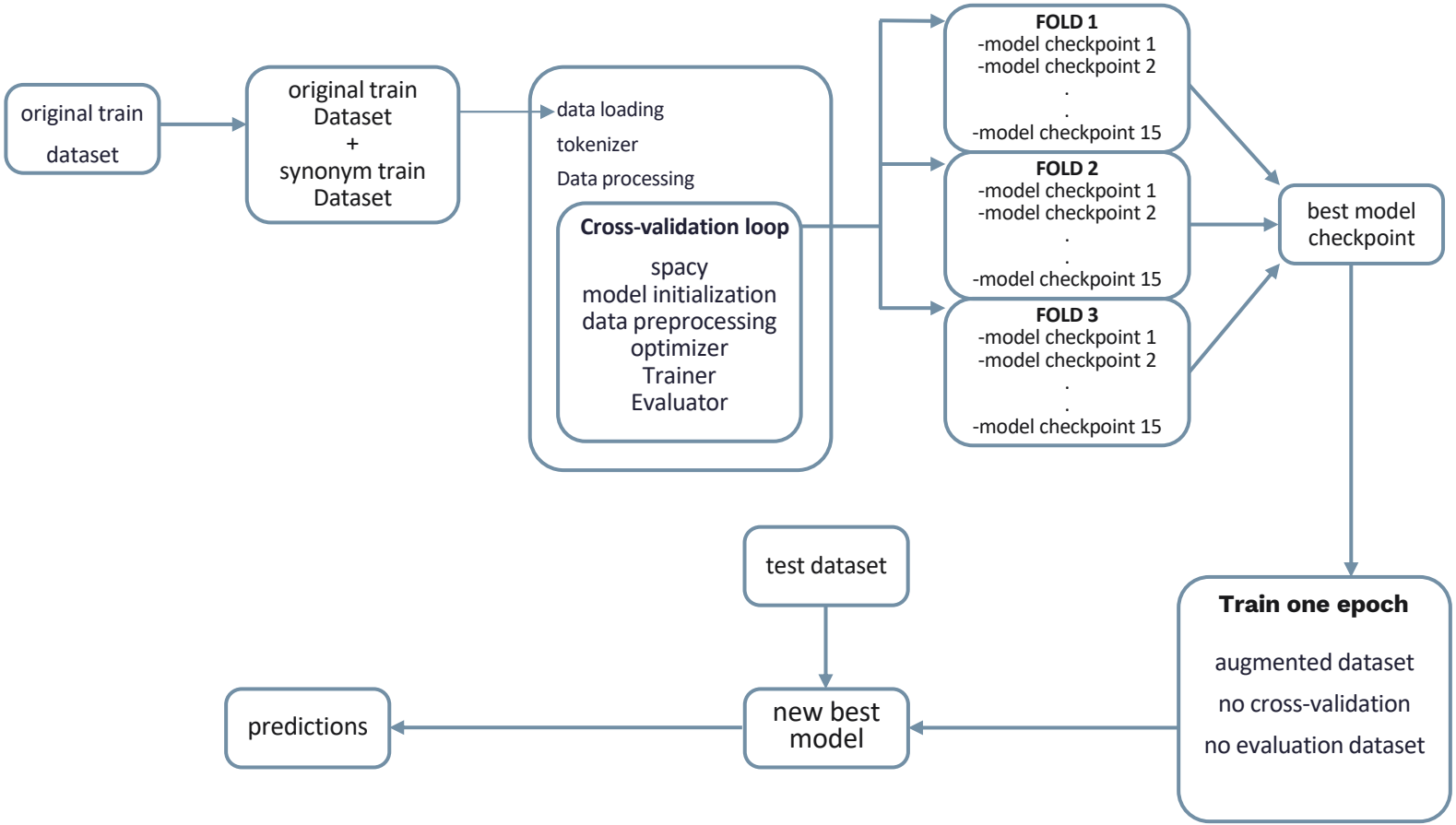
- *Fine-tuning transformer models with a token classification head*
- *Sentence-level processing to avoid text truncation*
- *Data augmentation via synonym replacement*:

-**English**: static word embeddings (word2vec) and SpaCy (GoogleNews-vectors-negative300)

-**Spanish**: FastText and SpaCy (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 – Data Augmentation

## English

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 ***drive*** to ***immunize*** these ***kids*** is nothing ***less*** than a ***nightmarish*** *experiment* with ***unforeseen*** ***repercussion*** . " — Rep. Louie Gohmert ( R - Texas ) # TheDefender."

## Spanish

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 ' ***peligro*** ***alto*** ' de ***endometriosis*** para las ***niñas***.

# Task 2 – Data Augmentation

"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": 0,
  "end_char": 145,
  "start_spacy_token": 0,
  "end_spacy_token": 21
},
{
  "span_text": "to ***vaccinate*** these ***children***",
  "category": "OBJECTIVE",
  "start_char": 35,
  "end_char": 62,
  "start_spacy_token": 7,
  "end_spacy_token": 11
},
{
  "span_text": "these ***children***",
  "category": "VICTIM",
  "start_char": 48,
  "end_char": 62,
  "start_spacy_token": 9,
  "end_spacy_token": 11
},

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

# Task 2 – Data Augmentation

"text": “I’m ***profoundly*** ***worried*** that the ***drive*** to ***immunize*** these ***kids*** is nothing ***less*** than a ***nightmarish*** experiment with ***unforeseen*** ***repercussions***“, - Rep. Louie Gohmert ( R – Texas ) # TheDefender.”

“annotations": [

{
  "span_text": "I’m ***profoundly*** ***worried*** that the ***drive*** to ***immunize*** these ***kids*** is nothing ***less*** than a ***nightmarish*** experiment with ***unforeseen*** ***repercussions*** ",
  "category": "NEGATIVE_EFFECT",
  "start_char": 0,
  "end_char": 144,
  "start_spacy_token": 0,
  "end_spacy_token": 21
},
{
  "span_text": "to ***immunize*** these ***kids***",
  "category": "OBJECTIVE",
  "start_char": 41,
  "end_char": 60,
  "start_spacy_token": 7,
  "end_spacy_token": 11
},
{
  "span_text": "these ***kids***",
  "category": "VICTIM",
  "start_char": 50,
  "end_char": 60,
  "start_spacy_token": 9,
  "end_spacy_token": 11
},

# Task 2 - Experiments on Train Dataset

**Model used for English:**

| Model | Epochs | Accuracy | F1 | Precision | Recall |
|---|---|---|---|---|---|
| facebook/roberta-base | 15 | **0.8897** | **0.5890** | **0.5684** | **0.6112** |
| facebook/roberta-base | 10 | 0.8210 | 0.4212 | 0.3943 | 0.4520 |
| distilbert-base-uncased | 5 | 0.6932 | 0.2881 | 0.2893 | 0.2869 |
| distilbert-base-uncased | 10 | 0.8457 | 0.4551 | 0.4279 | 0.4860 |
| distilbert-base-uncased | 15 | 0.8563 | 0.5645 | 0.5388 | 0.5928 |
| bert-base-uncased | 15 | 0.8723 | 0.5278 | 0.4976 | 0.5619 |
| dmis-lab/biobert-v1.1 | 15 | 0.8439 | 0.4608 | 0.4353 | 0.4894 |
| dslim/bert-base-NER | 12 | 0.7578 | 0.3412 | 0.2928 | 0.4088 |
| xlnet-base-cased | 15 | 0.8856 | 0.5716 | 0.5522 | 0.5924 |
| covid-twitter-bert | 4 | 0.7401 | 0.3301 | 0.3279 | 0.3323 |

Submitted model

**Model used for Spanish:**

| Model | Epochs | Accuracy | F1 | Precision | Recall |
|---|---|---|---|---|---|
| PlanTL-GOB-ES/roberta-base-bne | 15 | **0.9326** | **0.5849** | **0.5833** | **0.5866** |
| PlanTL-GOB-ES/roberta-base-bne | 8 | 0.9011 | 0.4565 | 0.4562 | 0.4568 |
| bert-base-spanish-wwm-uncased | 10 | 0.8213 | 0.4201 | 0.4093 | 0.4315 |
| bertin-roberta-base-spanish | 12 | 0.8598 | 0.5844 | 0.5856 | 0.5832 |
| alberto-base-spanish | 8 | 0.7841 | 0.3841 | 0.3833 | 0.3848 |
| distilbert-base-spanish-uncased | 15 | 0.8047 | 0.5101 | 0.5019 | 0.5185 |
| electra-small-spanish | 5 | 0.7288 | 0.3421 | 0.3332 | 0.3513 |
| xlm-roberta-base | 5 | 0.7981 | 0.4012 | 0.3834 | 0.4206 |

Submitted model

# Evaluation Results

**Binary Classification Task**

| | Task 1 | | | |
|---|---|---|---|---|
| | MCC | F1-macro | F1-conspiracy | F1-critical |
| *English_baseline* | *0.7964* | *0.8975* | *0.8632* | *0.9318* |
| English_submitted_run1 | 0.7574 | 0.8769 | 0.8338 | 0.9200 |
| English_submitted_run2 | 0.7872 | 0.8917 | 0.8536 | 0.9297 |
| *Spanish_baseline* | *0.6681* | *0.8339* | *0.7872* | *0.8806* |
| Spanish_submitted_run1 | 0.6147 | 0.7950 | 0.7179 | 0.8720 |
| Spanish_submitted_run2 | 0.6722 | 0.8293 | 0.7699 | 0.8887 |

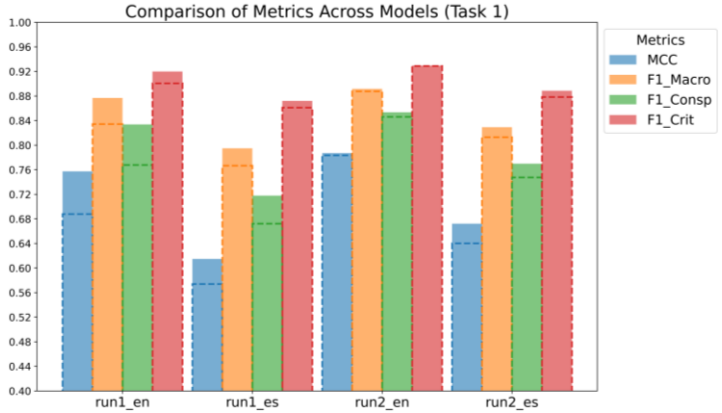
**Span-level Detection Task**

| | Task 2 | | | |
|---|---|---|---|---|
| | span-P | span-R | span-F1 | micro-span-F1 |
| *English_baseline* | *0.5323* | *0.4684* | *0.6334* | *0.4998* |
| English_submitted_run1 | 0.5832 | 0.6856 | 0.6293 | 0.6074 |
| English_submitted_run2 | 0.5859 | 0.6790 | 0.6279 | 0.6120 |
| *Spanish_baseline* | *0.4533* | *0.5621* | *0.4934* | *0.4952* |
| Spanish_submitted_run1 | 0.5997 | 0.6193 | 0.6089 | 0.6051 |
| Spanish_submitted_run2 | 0.6159 | 0.6129 | 0.6129 | 0.6108 |

# Experiments with and without data Augmentation

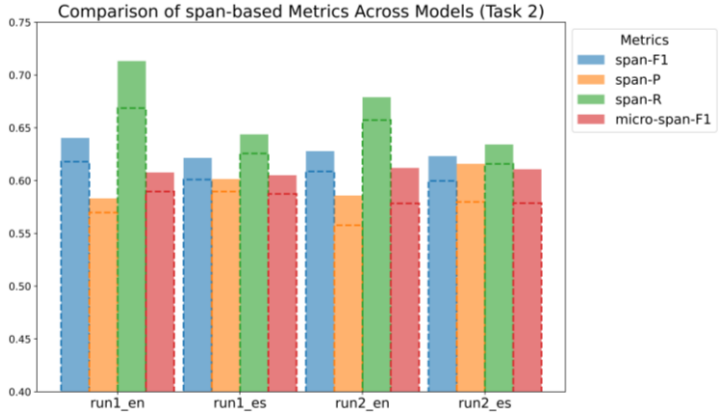
## Task 1

| Model | Data Aug | MCC | F1-Macro | F1-Consp | F1-Crit |
|---|---|---|---|---|---|
| **ENGLISH** | | | | | |
| run1 | ✓ | 0.7574 | 0.8769 | 0.8338 | 0.9200 |
| run1 | | 0.6880 | 0.8344 | 0.7679 | 0.9008 |
| run2 | ✓ | 0.7872 | 0.8917 | 0.8536 | 0.9297 |
| run2 | | 0.7836 | 0.8877 | 0.8463 | 0.9291 |
| **SPANISH** | | | | | |
| run1 | ✓ | 0.6148 | 0.7950 | 0.7179 | 0.8721 |
| run1 | | 0.5740 | 0.7667 | 0.6723 | 0.8612 |
| run2 | ✓ | 0.6722 | 0.8293 | 0.7699 | 0.8887 |
| run2 | | 0.6402 | 0.8131 | 0.7477 | 0.8785 |

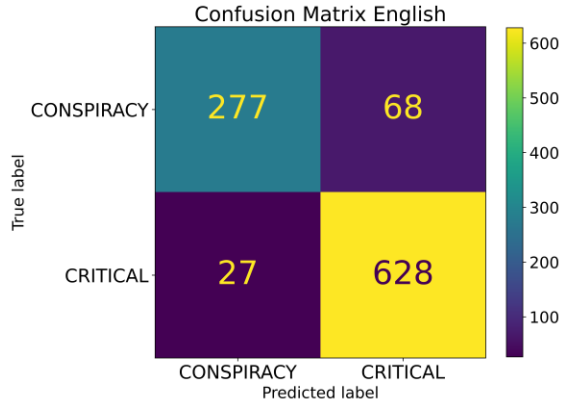


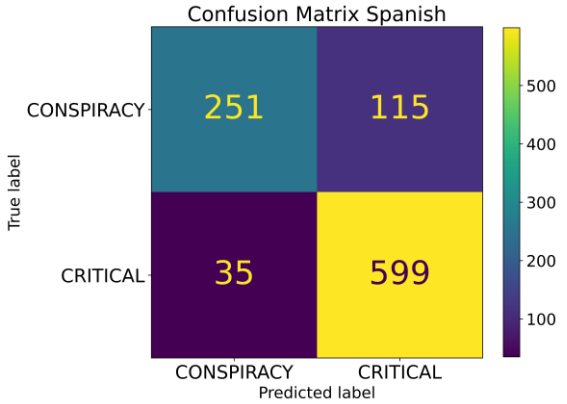

## Task 2

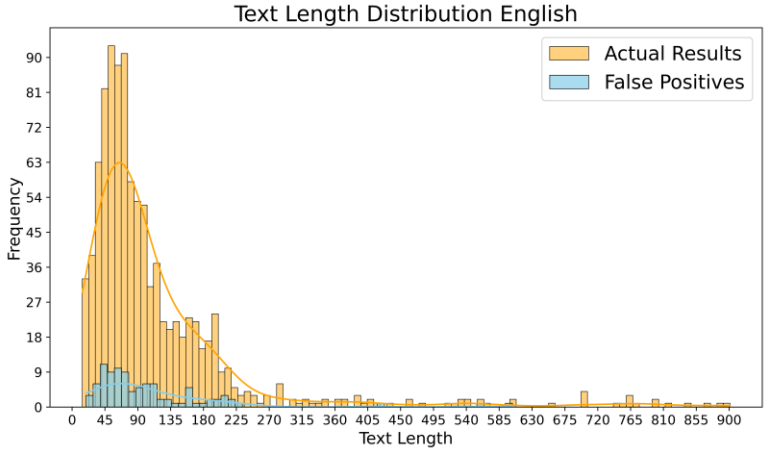
| Model | Data Aug | span-F1 | span-P | span-R | micro-span-F1 |
|---|---|---|---|---|---|
| **ENGLISH** | | | | | |
| run1 | ✓ | 0.6404 | 0.5832 | 0.7133 | 0.6077 |
| run1 | | 0.6180 | 0.5698 | 0.6687 | 0.5897 |
| run2 | ✓ | 0.6279 | 0.5859 | 0.6790 | 0.6120 |
| run2 | | 0.6087 | 0.5578 | 0.6574 | 0.5784 |
| **SPANISH** | | | | | |
| run1 | ✓ | 0.6215 | 0.6015 | 0.6438 | 0.6051 |
| run1 | | 0.6010 | 0.5897 | 0.6257 | 0.5874 |
| run2 | ✓ | 0.6232 | 0.6159 | 0.6342 | 0.6108 |
| run2 | | 0.5998 | 0.5799 | 0.6158 | 0.5787 |

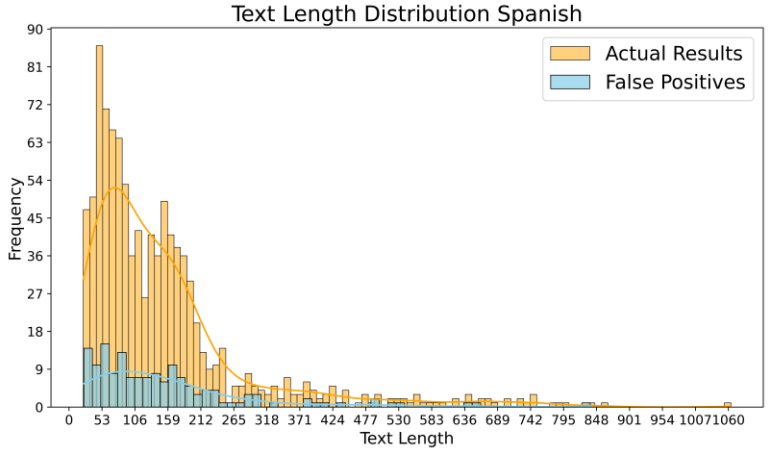


# Error Analysis Task 1

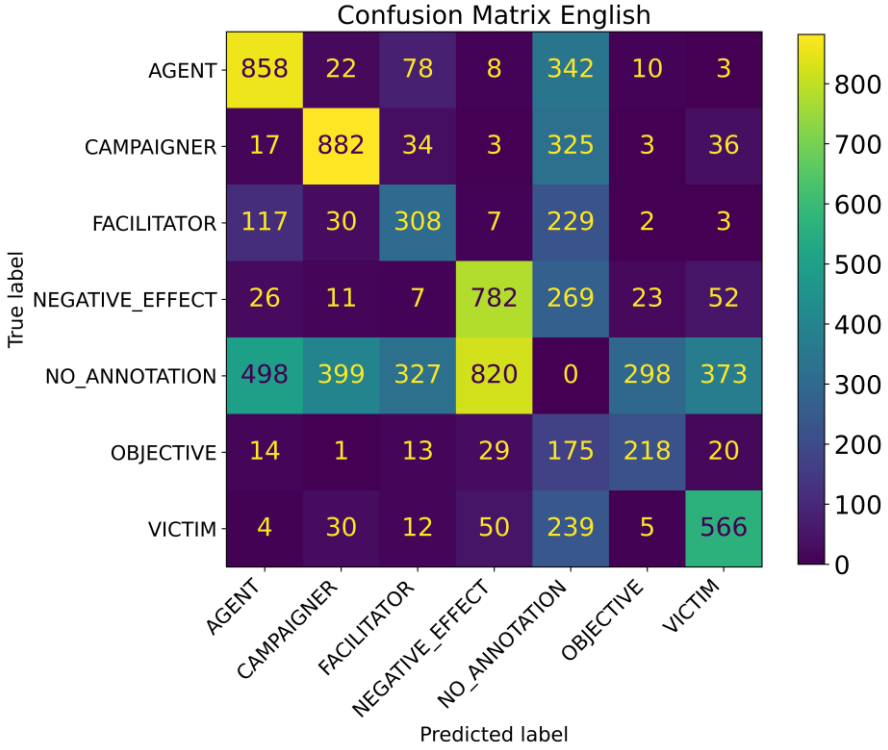


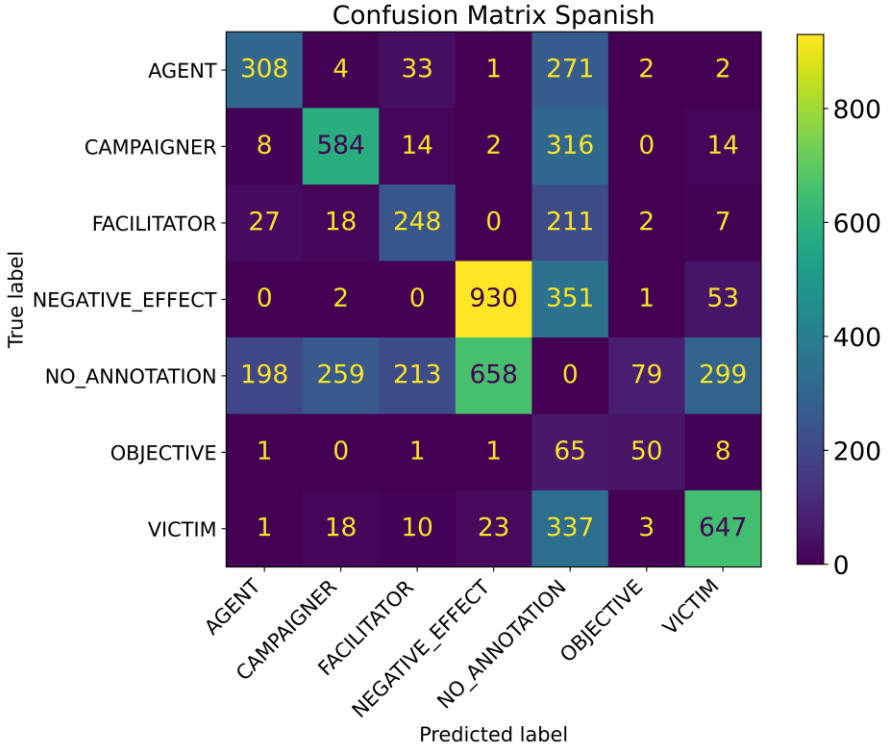



# Error Analysis Task 2



# Official Ranking Task 1

**TASK 1 - ENGLISH**

| POSITION | TEAM | MCC | F1-MACRO | F1-CONSPIRACY | F1-CRITICAL |
|---|---|---|---|---|---|
| | *baseline-BERT* | *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 | nlpln | 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%

**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 |
| | *baseline-BERT* | *0.6681* | *0.8339* | *0.7872* | *0.8806* |

~0.4%

*Total Participating Teams: 82*

# Official Ranking Task 2

**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 |
| | *baseline-BETO* | *0.5323* | *0.4684* | *0.6334* | *0.4998* |

~11%

~12%

**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 |
| | *baseline-BETO* | *0.4934* | *0.4533* | *0.5621* | *0.4952* |

# Further Experiments with Multilingual Transformers Models and Large Language Models

**Methods used:**

***- Multilingual Transformer Models :***

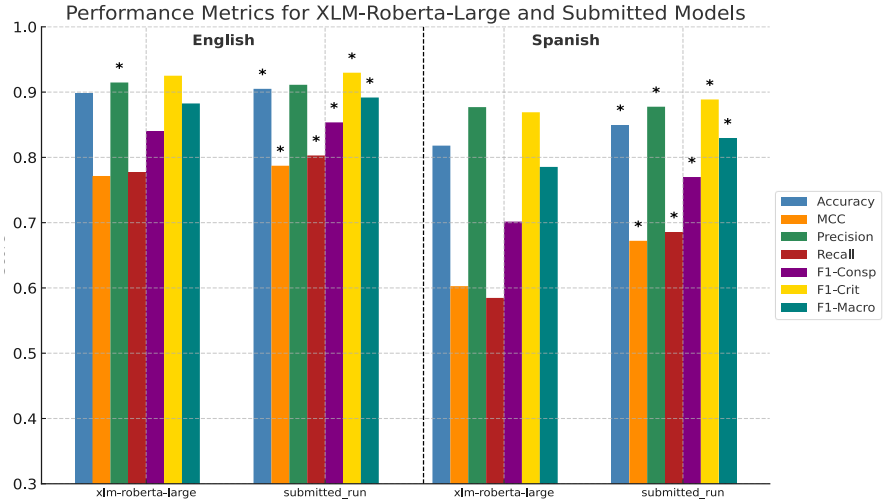
- Leverage the 2 datasets simultaneously

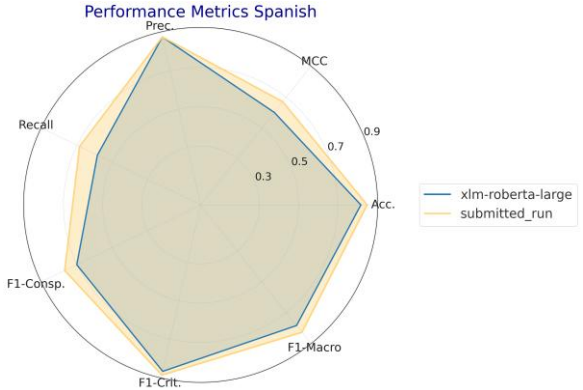
***- Large Languale Models :***

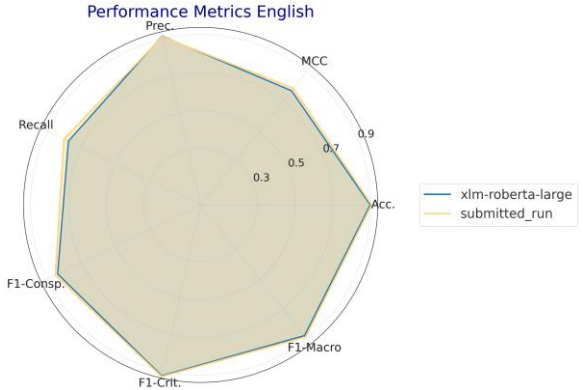
- Zero shot learning (text-generation mode)
- Few shot learning (text-generation mode)
- Zero shot learning (text-classification mode)
- Few shot learning (text-classification mode)
- Fine-Tuning LLMs with QLoRA
- Fine-Tuning LLMs with Zero-shot prompting
- Fine-Tuning LLMs with Few-shot prompting

# Multilingual Transformer Models Task 1

| Model | Acc. | MCC | Prec. | Recall | F1-Consp | F1-Crit | F1-Macro |
|---|---|---|---|---|---|---|---|
| **ENGLISH** | | | | | | | |
| xlm-roberta-large | **0.8980** | **0.7715** | 0.9147 | **0.7768** | **0.8401** | **0.9251** | **0.8826** |
| bert-base-multilingual | 0.8760 | 0.7237 | **0.9300** | 0.6928 | 0.7940 | 0.9113 | 0.8527 |
| xlm-roberta-base | 0.8750 | 0.7191 | 0.9015 | 0.7159 | 0.7981 | 0.9095 | 0.8538 |
| ensemble-3_multilingual | 0.8900 | 0.7555 | 0.9401 | 0.7275 | 0.8203 | 0.9207 | 0.8705 |
| *submitted_run* | *0.9050* | *0.7872* | *0.9112* | *0.8029* | *0.8536* | *0.9297* | *0.8917* |
| **SPANISH** | | | | | | | |
| xlm-roberta-large | **0.8180** | **0.6027** | 0.8770 | **0.5847** | **0.7016** | **0.8691** | **0.7854** |
| bert-base-multilingual | 0.7730 | 0.5011 | 0.8564 | 0.4563 | 0.5954 | 0.8423 | 0.7188 |
| xlm-roberta-base | 0.7970 | 0.5632 | **0.9137** | 0.4918 | 0.6394 | 0.8587 | 0.7491 |
| ensemble-3_multilingual | 0.8010 | 0.5687 | 0.8957 | 0.5164 | 0.6551 | 0.8602 | 0.7576 |
| *submitted_run* | *0.8500* | *0.6722* | *0.8776* | *0.6858* | *0.7699* | *0.8887* | *0.8293* |



# Zero and Few shot Learning Task 1 – *Text Generation Mode* -

**prompt = “<s>[INST] «SYS»**

**You are an expert critical thinker specialized in analyzing public health narratives, particularly regarding the COVID-19 pandemic.**

**You are tasked with classifying a text into one of these two categories:**

**1. CRITICAL THINKING: Texts that question major public health decisions but do not suggest that secret, powerful, or malevolent groups are behind these decisions.**

**2. CONSPIRACY THEORIES: Texts that imply public health decisions, especially those regarding the COVID-19 pandemic, are part of a plot orchestrated by secret, powerful, and malevolent groups.**

**Here below there are some examples:**

**– – – Start of the Examples – – –**

**<EXAMPLES>**

**– – – End of the Examples – – –**

**Now, say if the following text is ’CRITICAL’ or ’CONSPIRACY’:**

**«/SYS»**

**{user_text}**

**[/INST]</s>”**

## Zero-shot Learning

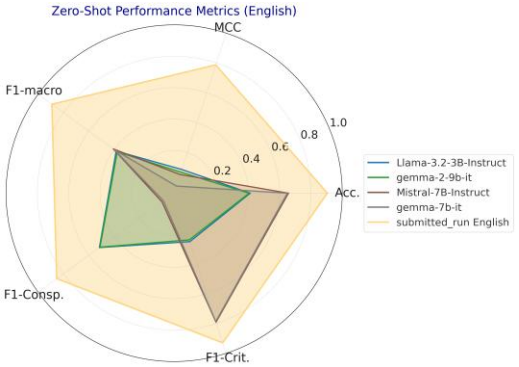
| Model | Acc. | MCC | F1-macro | F1-Consp. | F1-Crit. |
|---|---|---|---|---|---|
| | | | ENGLISH | | |
| Llama-3.2-3B-Instruct | 0.5010 | 0.1124 | 0.5010 | 0.5035 | 0.4985 |
| gemma-2-9b-it | 0.6400 | **0.4602** | **0.6394** | **0.6545** | 0.6245 |
| Mistral-7B-Instruct | 0.3670 | 0.0921 | 0.2953 | 0.5201 | 0.0705 |
| gemma-7b-it | **0.6500** | -0.0246 | 0.3993 | 0.0113 | **0.7874** |
| *submitted_run* | *0.9050* | *0.7872* | *0.8917* | *0.8536* | *0.9297* |
| | | | SPANISH | | |
| Llama-3.2-3B-Instruct | **0.6470** | 0.1260 | 0.4671 | 0.1575 | **0.7767** |
| gemma-2-9b-it | 0.5600 | **0.3265** | **0.5517** | **0.6127** | 0.4907 |
| Mistral-7B-Instruct | 0.4800 | 0.0761 | 0.4777 | 0.5122 | 0.4433 |
| gemma-7b-it | 0.3730 | 0.0099 | 0.2883 | 0.5338 | 0.0427 |
| *submitted_run* | *0.8500* | *0.6722* | *0.8293* | *0.7699* | *0.8887* |

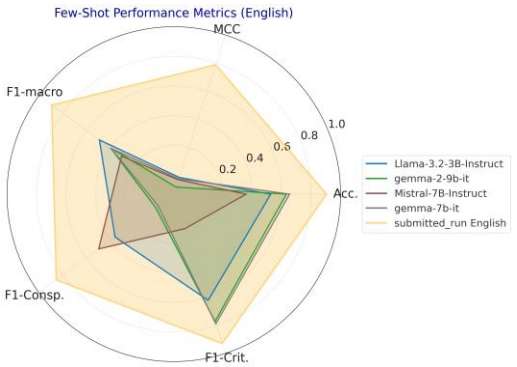
## Few-shot Learning

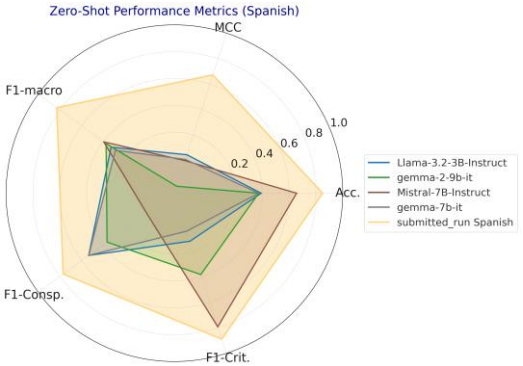
| Model | Acc. | MCC | F1-macro | F1-Cons. | F1-Crit. |
|---|---|---|---|---|---|
| | | | ENGLISH | | |
| Llama-3.2-3B-Instruct | **0.6970** | 0.2559 | 0.5635 | 0.3221 | **0.8049** |
| gemma-2-9b-it | 0.6870 | **0.5049** | **0.6869** | **0.6822** | 0.6916 |
| Mistral-7B-Instruct | 0.5090 | 0.2164 | 0.5039 | 0.5540 | 0.4538 |
| gemma-7b-it | 0.5410 | 0.1208 | 0.5361 | 0.4883 | 0.5839 |
| *submitted_run* | *0.9050* | *0.7872* | *0.8917* | *0.8536* | *0.9297* |
| | | | SPANISH | | |
| Llama-3.2-3B-Instruct | **0.6480** | 0.1315 | 0.4750 | 0.1737 | **0.7764** |
| gemma-2-9b-it | 0.5910 | **0.3322** | **0.5859** | **0.6181** | 0.5597 |
| Mistral-7B-Instruct | 0.6410 | 0.0914 | 0.4337 | 0.0911 | 0.7763 |
| gemma-7b-it | 0.6440 | 0.1158 | 0.4744 | 0.1759 | 0.7730 |
| *submitted_run* | *0.8500* | *0.6722* | *0.8293* | *0.7699* | *0.8887* |

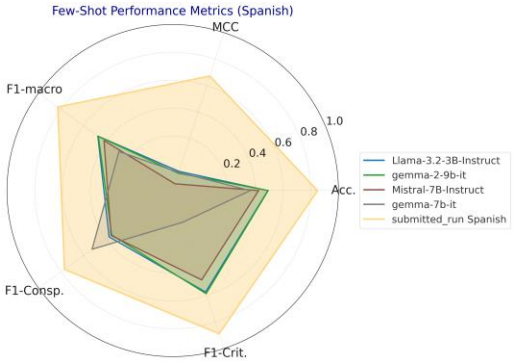
# Zero and Few shot Learning Task 1 – *Text Classification Mode* -

| Model | Acc. | MCC | F1-macro | F1-Consp. | F1-Crit. |
|---|---|---|---|---|---|
| **ZERO SHOT** | | | | | |
| **English** | | | | | |
| Llama-3.2-3B-Instruct | 0.4130 | **0.0866** | 0.3846 | **0.5169** | 0.2522 |
| gemma-2-9b-it | 0.4060 | 0.0683 | 0.3764 | 0.5123 | 0.2404 |
| Mistral-7B-Instruct | **0.6570** | 0.0526 | **0.4074** | 0.0228 | **0.7920** |
| gemma-7b-it | 0.6500 | -0.0246 | 0.3993 | 0.0113 | 0.7874 |
| **Spanish** | | | | | |
| Llama-3.2-3B-Instruct | 0.3920 | **0.0464** | 0.3283 | **0.5352** | 0.1214 |
| gemma-2-9b-it | 0.3740 | -0.2007 | 0.3739 | 0.3651 | 0.3826 |
| Mistral-7B-Instruct | **0.6563** | 0.0000 | **0.3962** | 0.0000 | **0.7924** |
| gemma-7b-it | 0.3730 | 0.0099 | 0.2883 | 0.5338 | 0.0427 |
| **FEW SHOT** | | | | | |
| **English** | | | | | |
| Llama-3.2-3B-Instruct | 0.5260 | **-0.0127** | **0.4922** | 0.3612 | 0.6232 |
| gemma-2-9b-it | 0.6320 | -0.0813 | 0.3973 | 0.0213 | 0.7734 |
| Mistral-7B-Instruct | 0.3590 | -0.0296 | 0.3057 | **0.4980** | 0.1134 |
| gemma-7b-it | **0.6540** | -0.0230 | 0.3954 | 0.0001 | **0.7908** |
| **Spanish** | | | | | |
| Llama-3.2-3B-Instruct | 0.4920 | **-0.0481** | **0.4730** | 0.3728 | 0.5731 |
| gemma-2-9b-it | **0.4940** | -0.0624 | 0.4678 | 0.3496 | **0.5859** |
| Mistral-7B-Instruct | 0.4250 | -0.1420 | 0.4181 | 0.3547 | 0.4815 |
| gemma-7b-it | 0.3650 | -0.0557 | 0.2847 | **0.5243** | 0.0451 |
| *submitted_run English* | *0.9050* | *0.7872* | *0.8917* | *0.8536* | *0.9297* |
| *submitted_run Spanish* | *0.8500* | *0.6722* | *0.8293* | *0.7699* | *0.8887* |



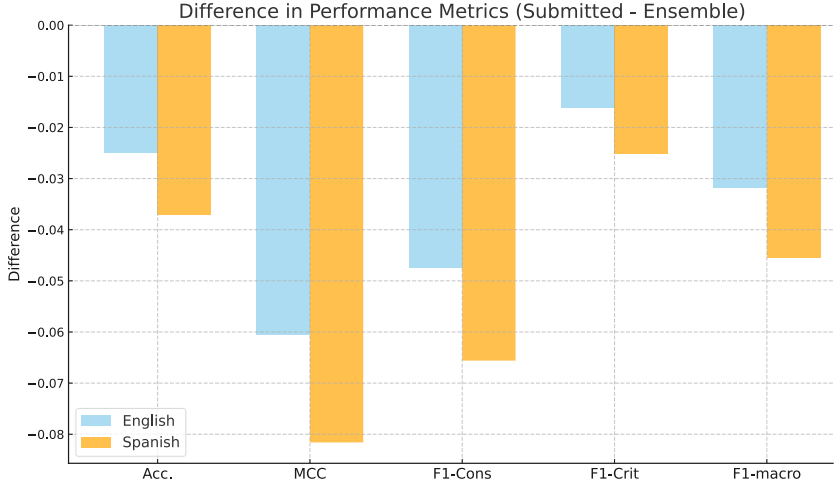



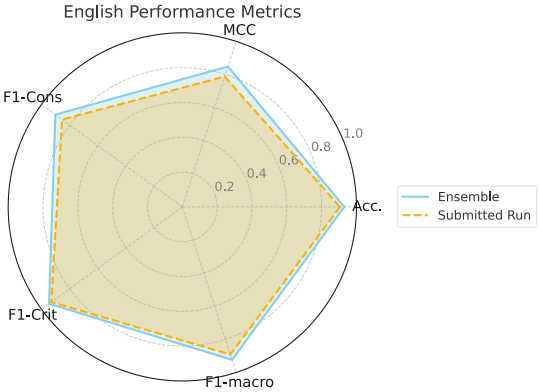
# Fine-Tuning LLMs Task 1

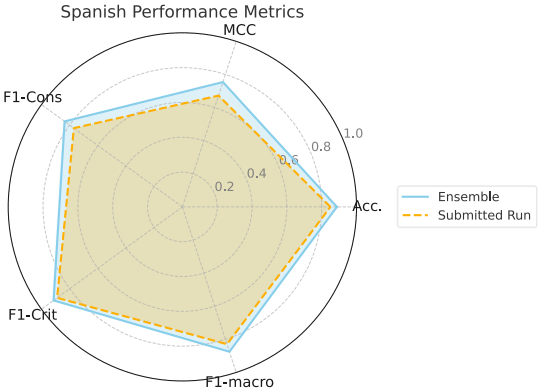
| Model | Acc. | MCC | F1-Cons | F1-Crit | F1-macro |
|---|---|---|---|---|---|
| **ENGLISH** | | | | | |
| LLAMA 3.2-3B_NoAug | 0.9140 | 0.8077 | 0.8677 | 0.9363 | 0.9020 |
| LLAMA 3.2-3B_DataAug | 0.9110 | 0.8009 | 0.8624 | 0.9342 | 0.8983 |
| LLAMA 3.2-3B-Instruct | 0.8930 | 0.7632 | 0.8243 | 0.9231 | 0.8737 |
| gemma-2-9b-it | 0.9170 | 0.8144 | 0.8744 | 0.9380 | 0.9062 |
| LLAMA 3.1-8B | 0.9050 | 0.8037 | 0.8728 | 0.9242 | 0.8985 |
| ensemble_3_LLMs | 0.9210 | 0.8236 | 0.8819 | 0.9406 | 0.9113 |
| ensemble_4_LLMs | **0.9300** | 0.8477 | **0.9011** | **0.9458** | **0.9235** |
| *submitted_run* | *0.9050* | *0.7872* | *0.8536* | *0.9297* | *0.8917* |
| **SPANISH** | | | | | |
| LLAMA 3.2-3B_NoAug | 0.8020 | 0.5770 | 0.7346 | 0.8421 | 0.7883 |
| LLAMA 3.2-3B_DataAug | 0.8260 | 0.6234 | 0.7119 | 0.8754 | 0.7936 |
| LLAMA 3.2-3B-Instruct | 0.8290 | 0.6268 | 0.7255 | 0.8758 | 0.8007 |
| gemma-2-9b-it | 0.8510 | 0.6737 | 0.7759 | 0.8884 | 0.8322 |
| LLAMA 3.1-8B | 0.8620 | 0.7003 | 0.8073 | 0.8925 | 0.8499 |
| ensemble_3_LLMs | 0.8540 | 0.6810 | 0.7774 | 0.8914 | 0.8344 |
| ensemble_4_LLMs | **0.8870** | 0.7538 | 0.8355 | **0.9139** | **0.8747** |
| *submitted_run* | *0.8500* | *0.6722* | *0.7699* | *0.8887* | *0.8293* |

BEST PERFORMING TEAMS

| Team | MCC | F1-Cons | F1-Crit | F1-macro |
|---|---|---|---|---|
| **ENGLISH** | | | | |
| *IUCL* | *0.8388* | *0.8947* | *0.9441* | *0.9194* |
| **SPANISH** | | | | |
| *SINAI* | *0.7429* | *0.8319* | *0.9091* | *0.8705* |



# Fine-Tuning LLMs with Zero and Few shot Learning Task 1

***First Method*** [preproc] : embed prompt during preprocessing *([prompt] + [text])*
***Second Method*** [train_ctx] : dynamically integrate prompts during training

**prompt = “**

**<s>[INST] You are tasked with classifying whether a given text is a CRITICAL statement questioning public health decisions, or a CONSPIRACY narrative alleging a secret plot related to the pandemic.**

**Here below there are some examples:**

**– – – Start of the Examples – – –**

**- - Start Examples of CRITICAL texts - -**

**<EXAMPLES_CRITICAL>**

**- - End Examples of CRITICAL texts - -**

**- - Start Examples of CONSPIRACY texts - -**

**<EXAMPLES_CONSPIRACY>**

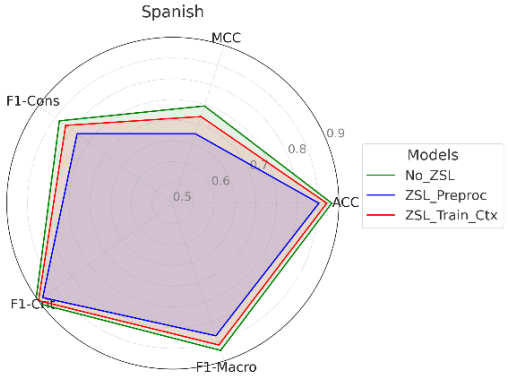
**- - End Examples of CONSPIRACY texts - -**

**– – – End of the Examples – – –**

**</s>[/INST] ”**

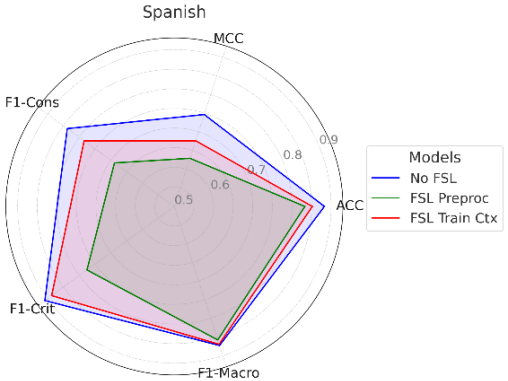
## Fine-tuning with ZSL

| Model | Acc. | MCC | F1-Cons | F1-Crit | F1-Macro |
|---|---|---|---|---|---|
| **ENGLISH** | | | | | |
| Llama-3.2-3B_No_ZSL | **0.9110** | **0.8009** | **0.8624** | **0.9342** | **0.8983** |
| Llama-3.2-3B_ZSL_preproc | 0.9050 | 0.7883 | 0.8494 | 0.9306 | 0.8900 |
| Llama-3.2-3B_ZSL_train_ctx | 0.9100 | 0.7987 | 0.8649 | 0.9325 | 0.8987 |
| **SPANISH** | | | | | |
| Llama-3.2-3B_No_ZSL | **0.8830** | **0.7465** | **0.8377** | **0.9085** | **0.8731** |
| Llama-3.2-3B_ZSL_preproc | 0.8520 | 0.6760 | 0.7849 | 0.8872 | 0.8360 |
| Llama-3.2-3B_ZSL_train_ctx | 0.8710 | 0.7196 | 0.8190 | 0.8998 | 0.8594 |



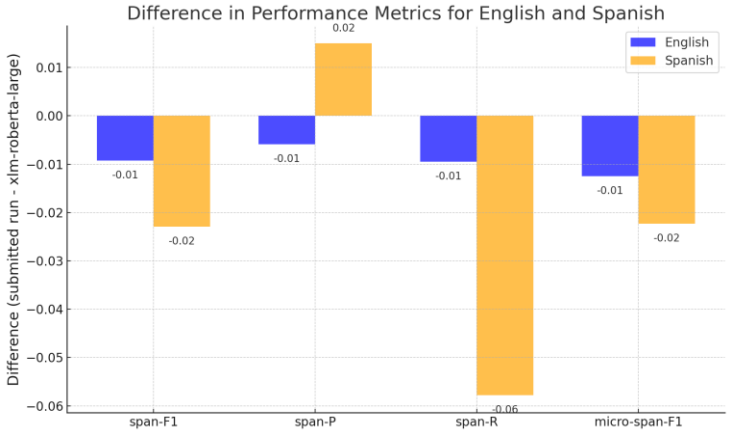

## Fine-tuning with FSL

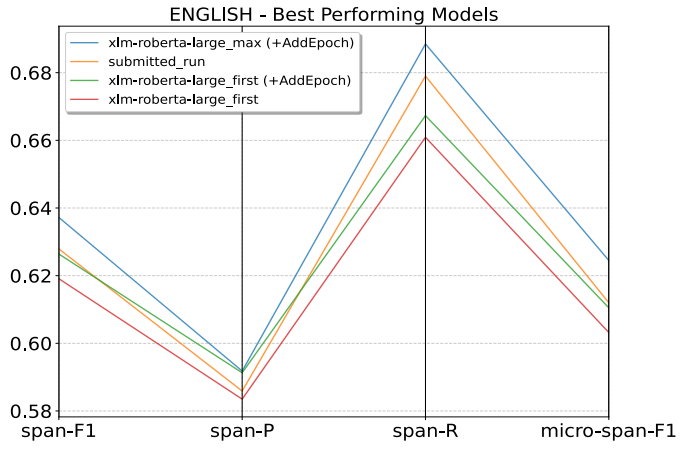
| Model | Acc. | MCC | F1-Cons | F1-Crit | F1-Macro |
|---|---|---|---|---|---|
| **ENGLISH** | | | | | |
| Llama-3.2-3B_No_FSL | **0.9110** | **0.8009** | **0.8624** | **0.9342** | **0.8983** |
| Llama-3.2-3B_FSL_preproc | 0.8840 | 0.7401 | 0.8135 | 0.9158 | 0.8647 |
| Llama-3.2-3B_FSL_train_ctx | 0.8910 | 0.7552 | 0.8305 | 0.9197 | 0.8751 |
| **SPANISH** | | | | | |
| Llama-3.2-3B_No_FSL | **0.8830** | **0.7465** | **0.8377** | **0.9085** | **0.8731** |
| Llama-3.2-3B_FSL_preproc | 0.8330 | 0.6291 | 0.6888 | 0.7756 | 0.8581 |
| Llama-3.2-3B_FSL_train_ctx | 0.8520 | 0.6761 | 0.7846 | 0.8870 | 0.8698 |



# Multilingual Transformer Models Task 2

| Model | AddEpoch | span-F1 | span-P | span-R | micro-span-F1 |
|---|---|---|---|---|---|
| | | **ENGLISH** | | | |
| xlm-roberta-base_first | | 0.6002 | 0.5400 | 0.6827 | 0.5843 |
| xlm-roberta-base_max | | 0.6021 | 0.5421 | 0.6841 | 0.5862 |
| xlm-roberta-base_first | ✓ | 0.6005 | 0.5488 | 0.6682 | 0.5846 |
| xlm-roberta-base_max | ✓ | 0.6013 | 0.5494 | 0.6693 | 0.5854 |
| bert-base-multilingual_first | | 0.5691 | 0.5337 | 0.6118 | 0.5532 |
| bert-base-multilingual_max | | 0.5699 | 0.5345 | 0.6125 | 0.5540 |
| bert-base-multilingual_first | ✓ | 0.5859 | 0.5325 | 0.6570 | 0.5700 |
| bert-base-multilingual_max | ✓ | 0.5857 | 0.5327 | 0.6562 | 0.5698 |
| xlm-roberta-large_first | | 0.6191 | 0.5835 | 0.6609 | 0.6032 |
| xlm-roberta-large_max | | 0.6191 | 0.5840 | 0.6602 | 0.6032 |
| xlm-roberta-large_first | ✓ | 0.6264 | 0.5913 | 0.6673 | 0.6105 |
| xlm-roberta-large_max | ✓ | **0.6372** | **0.5918** | **0.6885** | **0.6245** |
| *submitted_run* | ✓ | *0.6279* | *0.5859* | *0.6790* | *0.6120* |
| | | **SPANISH** | | | |
| xlm-roberta-base_first | | 0.6041 | 0.5483 | 0.6786 | 0.5882 |
| xlm-roberta-base_max | | 0.6050 | 0.5490 | 0.6797 | 0.5891 |
| xlm-roberta-base_first | ✓ | 0.6020 | 0.5563 | 0.6596 | 0.5861 |
| xlm-roberta-base_max | ✓ | 0.6039 | 0.5587 | 0.6608 | 0.5880 |
| bert-base-multilingual_first | | 0.5651 | 0.5384 | 0.5952 | 0.5492 |
| bert-base-multilingual_max | | 0.5665 | 0.5416 | 0.5940 | 0.5506 |
| bert-base-multilingual_first | ✓ | 0.5793 | 0.5434 | 0.6226 | 0.5634 |
| bert-base-multilingual_max | ✓ | 0.5801 | 0.5439 | 0.6236 | 0.5642 |
| xlm-roberta-large_first | | 0.6330 | 0.6093 | 0.6581 | 0.6171 |
| xlm-roberta-large_max | | 0.6341 | **0.6113** | 0.6580 | 0.6182 |
| xlm-roberta-large_first | ✓ | 0.6355 | 0.6047 | 0.6702 | 0.6196 |
| xlm-roberta-large_max | ✓ | **0.6358** | 0.6049 | **0.6707** | **0.6331** |
| *submitted_run* | ✓ | *0.6129* | *0.6199* | *0.6129* | *0.6108* |



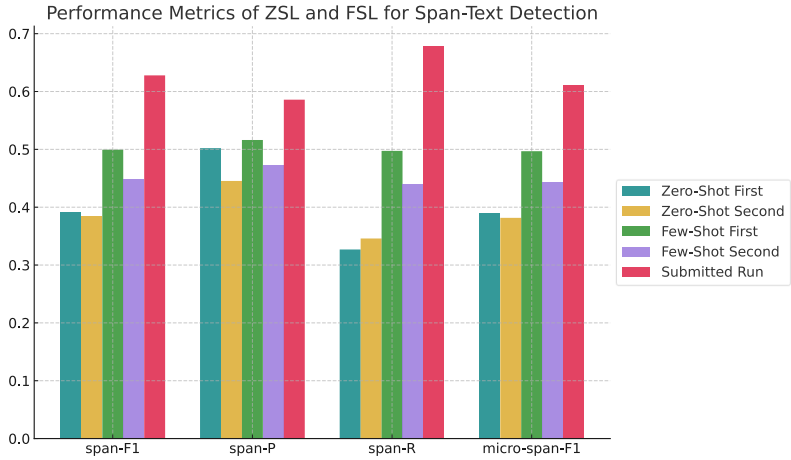
# Zero and Few shot Learning – Task 2

## First prompt:

prompt = "<s>[INST]
You are an expert in detecting elements of the texts. Since conspiracy narratives are a special type of causal explanation, your task consists in the recognition of text spans corresponding to the key elements of a text.
**Step 1:** Identify all of the negative effects mentioned in the text and relate them to the oppositional narrative. A negative effect is a harmful consequence or negative impact related to conspiracy theories or critical aspects. Put these negative effects in the same form that they appear in the text in different lines with the keyword "**NEGATIVE_EFFECT**".
…
[/INST]</s> "

## Second prompt:

prompt = "<s>[INST]
You are an expert in detecting elements of the texts. Since conspiracy narratives are a special type of causal explanation, your task consists in the recognition of text spans corresponding to the key elements of a text.
- Identify all the negative effects mentioned in the text and relate them to the oppositional narrative. They are the negative consequences suffered by the victims as a result of the actions and decisions of those in power and/or their collaborators. Put these negative effects in the same form that they appear in the text in different lines with the keyword "**NEGATIVE_EFFECT**".
…
Provide an output like this:
"annotations": [{
"span_text":
"category":
"start_char":
"end_char": }]
[/INST]</s> "




| Model | Prompt | span-F1 | span-P | span-R | micro-span-F1 |
|---|---|---|---|---|---|
| **ZERO-SHOT** | | | | | |
| gemma-2-9b-it | first | **0.3923** | **0.5021** | 0.3269 | **0.3897** |
| gemma-2-9b-it | second | 0.3843 | 0.4454 | **0.3459** | 0.3816 |
| **FEW-SHOT** | | | | | |
| gemma-2-9b-it | first | **0.4995** | **0.5162** | 0.4974 | **0.4968** |
| gemma-2-9b-it | second | 0.4485 | 0.4723 | **0.4395** | 0.4426 |
| *submitted_run* | | *0.6279* | *0.5859* | *0.6790* | *0.6120* |

# Conclusions

## TASK 1 – BINARY CLASSIFICATION

### Best Performances

*Ensemble of fine-tuned LLMS*

Fine-tuning LLMs without ZSL or FSL

Fine-tuning LLMs with FSL prompt-context

Fine-tuning LLMs with FSL prompt-inside-text

Fine-tuning LLMs with ZSL prompt-context

Fine-tuning LLMs with ZSL prompt-inside-text

Fine-tuning LLMs with FSL prompt-context

Fine-tuning LLMs with FSL prompt-inside-text

FSL LLMs

ZSL LLMs

## TASK 2 – SPAN-LEVEL CLASSIFICATION

### Best Performances

*Multilingual models*

*Monolingual models*

*Few shot learning LLMs*

*Zero shot learning LLMs*

# Thanks for your attention!

## Contacts:

angelomaximilian.tulbure@mail.polimi.it

angelotulbure29@outlook.it