

# Language Model, RoBERTa and T5 implementations for Detection of Propaganda Techniques in News Articles.

Antoine Basseto, Giacomo Camposampierro and Andrea Pinto

ETH Zurich - Swiss Federal Institute of Technology

{abasseto, gcamposampie, pintoa}@ethz.ch

## 1 Introduction

The proliferation of online misinformation has led to a significant amount of research into the automatic detection of fake news. However, most efforts have focused on verifying the factual correctness of a piece of information, and little attention has been paid to the propaganda techniques used by malicious actors to spread their message.

In this context, *Task 11 of SemEval-2020*<sup>1</sup> aims to fill this gap. This shared task provides a well-annotated dataset of new articles enabling us to develop propaganda detection models. The focus of the task is broken down into two well-defined sub-tasks, namely 1—*Span identification (SI)* for detecting the propaganda text fragments in the news articles and 2—*technique classification TC* for detecting the type of propaganda used in a given text span.

## 2 Objectives and Goals

The objective of this Natural Language Processing project would be to develop three model architectures for these two sub tasks and provide an error analysis for these models. Namely, we would like to implement a basic Language Model, RoBERTa<sup>2</sup> — a state-of-the-art pre-trained model based on the Transformer architecture and T5<sup>3</sup> — another state-of-the-art Transformer based architecture that uses a text-to-text approach. We then want to perform an in-depth analysis of the obtained results with the models to deduce which architecture performs better and eventually also understand why it has better performances. We also want to see if we would have any improvement tracks in the end.

<sup>1</sup>The official task webpage: <https://propaganda.qcri.org/semeval2020-task11/>

<sup>2</sup>RoBERTa paper and code: <https://huggingface.co/roberta-base>

<sup>3</sup>T5 paper and code: <https://paperswithcode.com/method/t5>

## 3 Implementation

Because the shared task is already a closed topic, many professional teams have made their way and produced a paper with satisfactory results. [Giovanni Da San Martino and Nakov \(2020\)](#) paper summarizes in a condensed way these results and the outline of the winning teams' implementations.

The Language Model is very scholar and is implemented for academic purposes. RoBERTa has been integrated in the Transformers Ensembles sets of various winning—so we consider it interesting to add it to our analysis. T5 on the other hand would be a brand new implementation recommended in the conclusions of the *aschern Anton Chernyavskiy and Nakov (2020)* team's project paper to further the topic research.

We will implement all the models from scratch but will probably take inspiration from their code work publicly available on their project GitHub repository<sup>4</sup>.

## 4 Milestones

Is this section really useful ?

## References

Dmitry Ilvovsky Anton Chernyavskiy and Preslav Nakov. 2020. [aschern at semeval-2020 task 11: It takes three to tango: Roberta, crf, and transfer learning](#).

Henning Wachsmuth Rostilav Petrov Giovanni Da San Martino, Alberto Barrón-Cedeño and Preslav Nakov. 2020. [Semeval-2020 task 11: Detection of propaganda techniques in news articles](#).

<sup>4</sup>*aschern* team GitHub repository: [https://github.com/aschern/semeval2020\\_task11](https://github.com/aschern/semeval2020_task11)