Exploring applications of multimodal models to fake news detection

Prerequisites: intermediate knowledge of PyTorch (or willingness to acquire it)
Contact: camilla.penzo@finances.gouv.fr

Le Pôle d'Expertise de la Régulation Numérique (PEReN) is placed under the joint authority of the French Ministers of Economy, Culture and Digital Technology and was created to answer needs and provide support to state services that have digital platform regulatory powers. Now more than ever, understanding data and its environment has become essential to understand how digital platforms work, and to set and adapt their regulations. At PEReN we conduct data science research projects with an exploratory and/or a scientific goal. Our projects range from machine to deep learning, focusing on image and video analysis to natural language processing.

In particular, several of our projects focus on social media platforms, given that these have largely replaced the conventional way in which information is spread. Today, very large social media platforms (e.g. Facebook, YouTube, etc) have an undoubtedly significant role in deeply shaping the way information is shared among people worldwide. Unfortunately, on top of providing billion users access to news and entertainment, social-networking platforms also build virtual grounds for the most diverse types of disinformation. Despite the blurry lines intrinsic in the complicated definition of *disinformation (aka fake news),* events of recent years have shown the central role of social media platforms in amplifying political disinformation, public health misconceptions, and incitement of violence.

While manual observation is both expensive and time-consuming, automatic fake news detection using deep learning is an efficient way to contrast their widespread dissemination. Early works on fake news detection merely focused on text-only or image-only content analysis. On the other hand, modern news and social media posts are almost always composed of both texts and images and thus unimodal methods neglect their correlation. In fact, for some fake news, a real image can be combined with rumours, or actual facts can be paired up with a fake image. Feature analysis that accounts simultaneously for both image and text is required for a more accurate fake news detection.

CLIP [1] is a multimodal model that combines knowledge of language concepts with semantic knowledge of images. It was trained on a variety of image-text pairs to predict the most relevant text snippet given an image, and vice versa. Very recently it was shown [2] that multimodal feature analysis offers complementary information that seem to be helpful also in fake news detection.

At PEReN we are firstly interested in extending this work to other available multimodal models, and check whether fake news detection accuracy can be improved. Secondly, we would like to train the resulting architecture on an in-house French fake news dataset.

The student will be first of all responsible for reproducing the results obtained in [2] by implementing the FakeNewsDetection-CLIP architecture described in the paper (see Fig. 2). Then they will be training the obtained architecture with the publicly available English fake news datasets, Gossipcop and Politifact [3]. Once the results of the paper [2] will be obtained, we would like the student to apply a post-training optimal transfer method to improve the latent space alignment of the CLIP model, and check whether this can improve the fake news detection task. Eventually, other multi-modal models will be considered.

Secondly, depending on the student's interest and the time available, the newly produced PyTorch implementation will be used to train a *French*FakeNewsDetection-CLIP method using a French fake news dataset in the process of being collected during a scraping project currently running at PEReN.

To be noted that these are tentative guidelines, that can be adjusted according to the evolution of the project. Also, we welcome suggestions/comments/requests from the student and the teaching staff.

[1] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, 2021. *Learning transferable visual models from natural language supervision*. In International Conference on Machine Learning. PMLR, 8748–8763

[2] Yangming Zhou, Qichao Ying, Zhenxing Qian, Sheng Li, Xinpeng Zhang, 2022. Multimodal Fake News Detection via CLIP-Guided Learning. https://arxiv.org/pdf/2205.14304.pdf

[3] Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. 2020. *Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media*. Big data 8, 3 (2020), 171–188.