

Project title: Detecting misinformation techniques on social media

Project leader: Dr. G. Spanakis, Assistant Professor, Department of Advanced Computing Sciences, Maastricht University

Promotor: Dr. G. Weiss, Full Professor, Department of Advanced Computing Sciences, Maastricht University

Proposal (250 words):

Introduction: Social media platforms have become an integral part of everyday life and activities of most people, providing new forms of communication and interaction. Online users now share information with each other easily via their computers, phones etc. However, this massive user-generated content has also led to the growth of misinformation: From search engines (where ranking of information plays a role into any decision making task) to news outlets and social media the main question is: Are users able to distinguish between unauthorized/unreliable information and correct/incorrect information? If not, how can we build techniques that assist in fighting misinformation?

Hypothesis and Objectives: The known fact is that social media facilitate spreading misinformation through different ways (specific communities that spread misinformation, fake accounts, social bots, filter bubbles, light moderation etc.). The main objective of this PhD thesis is to explore the characteristics of misinformation (in e.g. news, advertising etc.) on different social media platforms and build models that are able to detect such phenomena while providing a framework (technological, ethical and legal) for correcting these.

Setting and Methods: The PhD project starts with the collection of datasets from open social media (e.g. Twitter, Facebook, Youtube, Instagram, etc.) that suffer from misinformation. The next step would be the description and classification of misinformation phenomena (e.g. fake news spread on social media or misleading advertising promotion). State of the art NLP models and techniques will be used to represent the datasets (e.g. BERT, T5) and solve the task of detection (e.g. graph neural networks).

Impact: Social media platforms are affected by the proliferation of low-quality content, which hinders the capacity of technological solutions to since they become unintended means for the amplification and massive distribution of online harm. Addressing misinformation mechanisms becomes essential not only to mitigate the negative effects of the diffusion of unwanted content, but also to increase the user-perceived quality of social media and technology in general. Techniques of this project can help users in performing informed decision making in the context of online misinformation or other forms of online harm. Generally speaking, this PhD project can be of significant societal and economic impact and value.

Requirements candidate: Highly motivated student with good English communication skills and proactive and resolute attitude. Very good programming skills (python, torch/tensorflow/keras) and background in machine learning and NLP.

Keywords: Machine learning, Natural Language Processing, Social Media.

Top 5 selected publications:

1. V. Kolev, G. Weiss, G. Spanakis: FOREAL: RoBERTa Model for Fake News Detection based on Emotions. In Proceedings of the 14th International Conference on Agents and Artificial Intelligence – Volume 2, pages 429-440, 2022
2. T.C. Bertaglia, C. Goanta, G. Spanakis: Business Model Prevalence in Influencer Marketing on Instagram, IEEE Workshop on Technology and Consumer Protection (ConPro '20)
3. M. Meyers, G. Weiss, G. Spanakis: Fake News Detection on Twitter Using Propagation Structures. 2nd Multidisciplinary International Symposium on Disinformation in Open Online Media (MISDOOM 2020). Leiden, Netherland
4. M. Trusca, G. Spanakis: Hybrid Tiled Convolutional Neural Networks (HTCNN) for Text Sentiment Classification. In Proceedings of the 12th International Conference on Agents and Artificial Intelligence – Volume 2: ICAART, ISBN 978-989-758-395-7, pages 506-513, Malta, 22-24 February 2020 (**best paper award**)
5. F. Krebs, B. Lubascher, T. Moers, P. Schaap, G. Spanakis: Social Emotion Mining Techniques for Facebook Posts Reaction Prediction. Proceedings of the 10th International Conference on Agents and Artificial Intelligence (ICAART 2018), pp. 211-220, Funchal, Portugal, 16-18 January 2018 (**31 citations**)