Application for funding for a research project:

Beyond Fact-Checking: Detecting Frames and Disinformation in News and Social Media Content with Computational Methods

Applicant: Dr. Peter Maurer, Karlstad University (GMK)

1. Project description and goals

Today's (online-) news and social media are filled with frames and narratives aimed at political influence in support of a political line and/or goal. This project aims to develop a model for automated detecting of frames in news and social media content at a large scale, using computational methods. Computational methods represent a significant innovation in the field of communication science. They rely on machine learning to automate the analysis of texts, thereby performing tasks that, until recently, required human review. By automating frame detection, the project responds to the increased efforts to influence public opinion through the dissemination of 'framed information' on the internet and the related fact that AI has opened the way for mass production of disinformation and biased information that looks like real news.

In a nutshell, frames are interpretations woven into news and other types of public communication suggesting what an issue is all about. If framing is successful, people adopt frames and thus acquire a certain understanding of an issue. That way, frames influence the formation of opinions about an issue and can motivate political behavior in terms of whom to support, what to do, etc. Framing activities can polarize public opinion, and this becomes relevant especially for issues where domestic and international actors have high stakes in shaping citizens' views about a problem or conflict, and some may also use it to polarize societies. So, detecting the frames that are used together with certain issues and tracing them across platforms and channels is crucial.

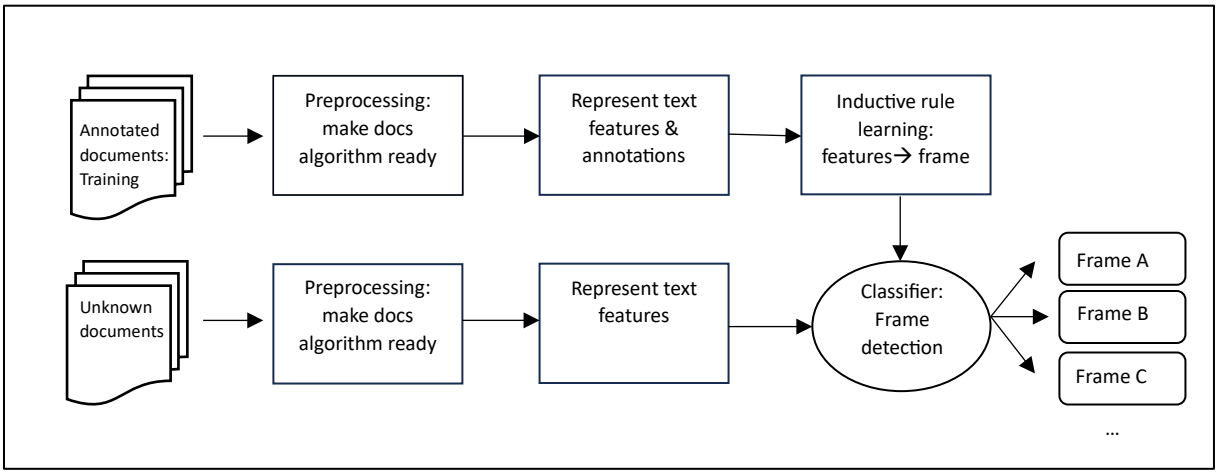
The initial focus, in this project, is the detection of frames used in narratives related to the Russia-Ukraine conflict, as well as two non-violent conflicts, over green policies and immigration. In all of these conflicts, political actors and a range of other stakeholders employ framing techniques to influence public opinion. For example, one way of framing the conflict between Russia and Ukraine would be to present Russia as a victim of Western encirclement whereas a counter-frame would interpret it as a consequence of Russian imperialism under Putin. With respect to green policies, to take another example, a frame might focus on the role of climate change in extreme weather events and present green policies as a solution while another might relate them to economic risks. Immigration-related discourse may allude to the discriminatory treatment of immigrants in the receiving society or on the capacity of the latter to integrate them successfully.

Hence, the first purpose of this project is to automatize the detection of frames in large amounts of news coverage and social media content about the aforementioned public issues. This can be achieved through computational methods like supervised machine learning (SML) which allow an automated and scalable mapping of frames. Based on SML, we will train a frame classifier able to detect frames specific to a given issue in different types of digital text documents (including transcribed verbal content from videos) by learning how to associating content features with frames (Figure 1). Furthermore, the metadata collected alongside the documents enables us to compare the dissemination of frames across diverse media outlets and platforms.

The second aim is the comparison of frames in the information ecology between Sweden and Germany. Therefore, the multilingual project team will develop a frame classifier for news and social media content in Swedish and German. This allows us to compare how frame setting and frame spreading for the same issues functions in the information ecologies of countries with different size, political-cultural background, and international influence. The project will collect the frames in parallel periods, also

going back in time, to monitor which frames emerge or disappear, when and where they originate, and how they spread in each country.

Figure 1: Steps in Supervised Machine Learning for Frame Detection



Note: Frame detection relies on training an algorithmic classifier capable of recognizing frames in different sorts of text, including transcribed video content. Once developed and calibrated, the detection tool can process a substantial volume of textual data from both, traditional media and social media platforms.

1.1. The increased power of framing in the digital information ecology

Although framing has always existed in modern political communication, it has become more pervasive and ubiquitous in the digital information ecology. Furthermore, the capacity of frames to shape the opinions of recipients by influencing their information processing renders them a potent instrument of public communication for actors pursuing political interests. Therefore, the frame classifier will be able to detect frames in media coverage as well as in the content of social media platforms. Strategic narratives and disinformation spread comparably unfiltered especially in this arena. Moreover, frame sponsors, like political actors, activists, and influencers, sometimes appear in the guise of neutral information brokers on social media platforms. Hence, the risk of being exposed to disinformation while seeking valid information about political conflicts is high when using these platforms. In response to its crucial position in the digital information ecology, the detection tool will be specifically calibrated to identify frames in social media communication. The material that can be analyzed by the tool includes both shorter and longer postings, which are prevalent on Twitter and Telegram, transcriptions of video content from YouTube and TikTok, as well as content from AI search-engines and chatbots. In the domain of social media, the focus will be on the public social media accounts of media outlets, political parties, and a diverse array of political actors, activists, and influencers with high social media activity and a substantial following.

2. Expected benefits for media professionals, society, and researchers

Apart from making a contribution to computational communication research, the project has positive effects in at least six areas:

First, large-scale frame mapping can serve as a monitor and an early warning system to spot biased or one-sided information and (AI-generated) disinformation through framing across the vast information ecosystem. It can assess the relevance of specific frames in relation to the aforementioned and potentially other issues on different platforms and trace them back to the actors who propagate them. Such an early warning system is especially relevant to spot disinformation campaigns that operate through misleading

frames used to communicate about an event or topic. It is important to invest in automated, scalable frame detecting tools, because fact checking as a method to evaluate the accuracy of statements is too labor-intensive for large amounts information and not always able to handle the type of manipulation associated with frames.

Second, the data obtained with frame mapping may counteract eroding public trust in the traditional media by facilitating fact-based discussion. Given that parts of the public have developed the perception that the traditional media cover certain issues in a one-sided manner, frame mapping could provide a solid and constant check on these claims. This may help the media, either to adjust the coverage or to refute unwarranted criticism.

Third, related to the previous point, mapping frames in media coverage can help media professionals to become conscious about successful attempts at instrumentalization by frame sponsors. Evidently, journalists are under high risk in this regard because the sponsors and promoters of frames are among their sources. So, constant frame monitoring by an incorruptible tool might facilitate the upholding of standards of objectivity and balance and may protect the media from inadvertently adopting one-sided narratives.

Fourth, automated frame detection ties in with the goal to enhance digital resilience against propaganda and disinformation. Detecting frames means to assess how misleading a text about a topic is and this creates awareness of biases in the information system. Media outlets can use this information to balance the coverage of an issue and include more viewpoints, while platform companies can use it to flag misleading content or filter out blatant disinformation.

Fifth, AI search-engines and AI chatbots are increasingly available and used by citizens and journalists to search information about political issues. If the trend continues, frames in AI generated info will have a growing influence on the knowledge and opinions the public develops concerning the issues and conflicts shaping our time. It is thus important to monitor the frames contained in the provided information with adequate frame detecting tools to assess the societal impact of AI-generated information.

Sixth, frame mapping and tracing increases the fairness in political discourse through enhancing transparency concerning the discursive power of various actors in the formation of public opinion. This will improve the public's media literacy, political knowledge and ability to recognize propaganda campaigns.

Last not least, this project will contribute to communication research. The results will be of interest ot other researchers engaged in the automation of the extraction of the political meaning in various types of text. One the one hand, the project will provide researchers studying political framing a validated tool in the domains studied (Russia-Ukraine, green policies, migration, etc.). On the other hand, the project will help to enhance the development of similar tools for other domains. Especially, it will contribute to the development of measurement tools adapted for social media communication, a type of communication that has been less in the focus of framing research, and it will advance multilingual frame detection, which is an important resource for comparative studies and non-English communication. Moreover, the project stimulates innovative research designs that combine computational text and network analysis. It will also bring research on frames and disinformation strategies closer together and further the knowledge about discursive power in the digital era, two hotly debated topics of contemporary communication research. The project team will make fully available, through a website, all the steps and tools so the research community can test them and develop them further.

3. Embedding in existing research frameworks

The project will benefit from the expertise assembled in the NODE research group where it ties in with the research program *Vänner och fiender: tillit och misstänksamhet i det nya medielandskapet.* The applicant is a regular participant of NODE meetings, which facilitates exchange and collaborations. The project implementation will greatly benefit from exchanges with Michael Karlsson and Henrik Örnebring, directors of NODE. The NODE environment is also an opportunity for the PhD student.

Apart from NODE, the applicant is a member of the GEOMEDIA center, where colleagues conduct an ongoing project about digital media infrastructures. The project leader will create synergies and seek exchanges with the GEOMEDIA project group as well. Beyond, the applicant is part the EU funded COST Action CA21129 "OPINION – Measuring Opinionated Communication", a network of international researchers dedicated to the development of innovative methods for the automated measurement of opinion in digital text. Moreover, OPINION hands out grants for early career researchers like PhD students for conference participation and short term visits to other universities and research groups working on similar projects.

Lastly, the applicant has established work relationships with Christian Nuernbergk, Professor of Media and Communication at Trier University (Germany) and his team. Exchanges and research visits are envisaged in relation with of the German part of the analysis.

4. Team

1. Peter Maurer (Senior Lecturer, Principal investigator) 30% research time, 10% administration and supervision time for 4 years

2. N.N. (Doctoral student), 100% research for 4 years.