The Summary Report from the Workshop I: Solving Ethical & Technical Challenges of Fake News

Fake News & other AI Challenges for the News Media in the 21st Century

SAILLABS
TECHNOLOGY

**Welcome to the world
of Global Media**

www.sail-labs.com

Contents:

1)Introduction

As the term fake news can be perceived as an interdisciplinary challenge, the first workshop aimed towards investigating both technical and social aspects of this phenomenon. The panel called "Ethical and Technical Challenges of Fake News" was designed to include specialists with different backgrounds, who offered their perspectives from various professional angle.

The workshop tried to answer questions such as:

- Can addressing and possible clarification of the terminology used to refer to fake news (for instance misinformation, disinformation, junk news etc.) help combating the phenomenon?

- What are the methodologies and practices needed to differentiate misinformation from the genuine news?

- How can journalists communicate the truth successfully in a digital world of clickbaits and falsehood?

- With so much information shared throughout the Internet, can technology alone fight the fake news?

- Sources, authors and search engines: what are the components of credibility we need to take into consideration when rating the veracity of online content and can they be measured?

2) Panelists

- **Rosemary Wolfe (moderator),** Consultant specializing in international media and monitoring services

- **Gerhard Backfried**, Head of Research at SAIL LABS Technology

- **Allan Hanbury**, Professor for Data Intelligence at Technical University of Vienna

- **Bettina Paur**, Researcher and Assistant at the University of Vienna

- **Nadejda Komendantova**, Theme Coordinator of Governance in Transition at the International Institute for Applied Systems Analysis

3)Presentations

Gerhard Backfried

The first speaker in the workshop was Gerhard Backfried, who discussed technical and social elements in the relation to the term computational propaganda. According to him, fake news is an interdisciplinary phenomenon which involves fields such as communication science, psychology, computer science and linguistics. Gerhard introduced terms associated with fake news, such as propaganda, misinformation, disinformation, malinformation, manipulation and deep fakes. He mentioned that on the one hand, fake news is cheap to produce, but on the other is costly to distinguish and debunk. Moreover, the phenomenon confirms the prior beliefs and preferences of recipients, which is known as confirmation bias. Fake news can have many forms ranging from textual fake news (articles, tweets) to fake images and fake videos. Gerhard believes that the solutions to the fake news problem should be tackled by interdisciplinary teams and the approach should combine diverse technical measures and human assessment.

Regarding the fake news itself, its messages are usually shorter and can frequently have emotional content. For instance, messages can urge people to do something about some situation and become a part of a movement. They try to create feelings which are mostly negative and stay with the recipient well beyond the message itself.

What is important is that language is at its core – if you dominate the terminology, you can frame issues. If this happens repeatedly, it is possible to create connotations, so that when people hear a keyword next time, they will automatically associate other words and concepts with it.

Both technical and social aspects should be addressed in order to prevent fake news from spreading. According to Gerhard, the suggestions concerning technical elements of fake news detection could include crowd-sourced fact-checking and automation. What can be automated is the source identification, origin, content etc., while the social aspects could cover factual reporting or educational programs promoting media literacy.

For more details regarding computational propaganda you can check Gerhard's presentation:

http://www.lt-innovate.org/sites/default/files/B004I%20001%20Gerhard%20Backfried.pdf

Or read an interview with Gerhard:

https://www.sail-labs.com/2018/11/26/gerhard_backfried/

Allan Hanbury

The second speaker was Allan Hanbury, who spoke about credibility of sources, content and algorithms, with the aim to determine credibility automatically. Although, credibility is a complex phenomenon, four key aspects can be identified: expertise, trustworthiness, quality and reliability.

According to Allan, expertise and trustworthiness can be attributed to the source of the message, while quality and reliability refer to the content itself. Subsequently, the various problems associated with the four aspects were presented.

The first aspect is expertise, which focuses on the knowledge of the source. The potential problem associated with this aspect was illustrated through an example suggesting that the automatic evaluation of expertise through algorithms could, for instance, consider people who are ideologically biased as experts based on their previous work.

The second feature of credibility is trustworthiness, which can be regarded as *goodness* or *morality* of the source. Moreover, reputation represents another concept that needs to be defined, but how is reputation different from trustworthiness?

The last two aspects of credibility are associated with the content – quality and reliability. Quality can be estimated through the number of spelling errors, typos, style etc. Although, the potential problem associated with assessing quality in an isolated manner is that even a text of low quality can be written by trusted experts and vice versa. The final feature of credibility is reliability, which refers to something being perceived as dependable and consistent in quality.

Allan implied that these four aspects of credibility also apply to the credibility of search engines. In this case expertise refers to the track record of the provider of the system – for instance, how long has the system been in operation? The trustworthiness, on the other hand, is reflected by the fact that the search engine does not filter out results, or does not show biased results. The quality can be measured in terms of effectiveness and efficiency, while reliability is represented by the consistency of results that the system produces over a range of different datasets.

For more details regarding credibility of sources, content and algorithms you can check Allan's presentation:
http://www.lt-innovate.org/sites/default/files/B004I%20002%20Allan%20Hanbury.pdf

Bettina Paur

Bettina Paur, the third speaker in the workshop, held a speech concerning fake news and the challenge of ethics in journalism. Bettina started with the historical roots of fake news itself, noting that in the context of communication studies this phenomenon can be found in essays from 17th century. Bettina proceeded with examples, such as the broadcast of a radio adaptation of Herbert George Wells' drama "The War of the Worlds", which adopted a radio news format and led to panic by listeners, who believed that the story of Martian invasion was part of the news.

On the other hand, as Bettina suggested, concealing information can be also perceived as fake news. For instance, the fact that the former US President Roosevelt spent most of his term bound to a wheelchair was not considered as public information. This would probably not be possible nowadays, as even if journalists would cooperate, not every smartphone-user would. Combined with the power of social media, journalists would need to cover the story.

Fake news became a frequently used phrase after the 2016 Presidential elections in the USA. As Bettina notes, in the final three months of the presidential campaigns, 20 top-performing false stories regarding elections generated more than 8 million shares, reactions or comments on Facebook.

Bettina also stated that there is no agreement regarding the definition of fake news. The term can be understood as a news satire or parody, fabrication, manipulation, advertising or propaganda. If we want to counter fake news, we need to develop precise definitions, so that potential algorithms would not eliminate for instance satire or other forms of humor.

The digitalization of news has challenged the traditional perception of news, as nowadays non-journalists can reach a mass audience through social media, which has both advantages and disadvantages. As Bettina suggests, the most problematic content uses people's emotions (especially anger, fear etc.), which drives sharing among people who want to belong to their online communities. To understand fake news itself, we need to understand the ritualistic function of communication and employ a multidisciplinary approach, where technological companies, governments, media organizations and citizens collaborate.

For more details regarding fake news and the ethics in journalism you can check Bettina's presentation:

http://www.lt-innovate.org/sites/default/files/B004I%20004%20Bettina%20Paur.pdf

Or read an interview with Bettina: https://www.sail-labs.com/2018/12/04/interview-with-bettina-paur/

Nadejda Komendantova

The last speaker was Nadejda Komendantova, who spoke about co-creating misinformation resilient societies.

Nadejda suggested that the misinformation itself does not intend to cause harm, as opposed to disinformation, which is false and deliberately harming. Nowadays, the Internet and particularly social media have made it easier for misinformation to be spread and made the phenomenon more complex.

Co-Inform project is based on the assumption that co-creation experiences are the basis of a value. The aim of the project is to co-create solutions towards detecting and combating misinforming posts on social media, support and persuade the misinformation-resilient behavior, understand and predict which misinforming content is likely to be spread, provide policy-makers with misinformation analyses, help with connecting public using social media, etc.

According to Nadejda, the goal of Co-Inform project is to study misinformation in a multidisciplinary environment, while the co-creation itself requires both understanding the technical side and human factors like decision-making, institutional structure and the perceptions of different stakeholders. The project aims to examine the sources of disinformation, but also focuses on the perception of recipients and tries to understand how different social groups perceive disinformation. As Nadejda suggests, the way forward is both the development of new technologies and the research on cognitive biases of different recipients.

The Co-Inform project has two tools – one is the browser plugin, which raises awareness about misinforming content, and the other is the dashboard that could be used by journalists and policy-makers, because its aim is to detect misinformation, its origins and the location.

For more details regarding co-creating misinformation resilient societies you can check Nadejda's presentation:

http://www.lt-innovate.org/sites/default/files/B004I%20003%20Nadejda%20Komendantova.pdf

4) Discussion

The discussion was moderated by Rosemary Wolfe. The first discussed topic regarded the ways to encourage people to question the media, which led to conversation about education. Allan Hanbury mentioned that teachers encourage students to investigate the sources of information, while Gerhard Backfried noted that students have computer science classes, but social media and their questioning is not part of the curriculum. Moreover, people have tendency to share information, which sometimes becomes automatic and that could be a problem. Bettina Paur added that people also share information to gain the feeling of identification with certain groups.

How to help people understand that they are being manipulated? Bettina Paur suggested that it is important that people understand the difference between information and entertainment and do not allow these two to merge together. Gerhard Backfried suggested the importance to question also the opposite side of the story when conducting

research or writing an article and noted that people should get wider spectrum of information. The problematic aspect is the fact that some people do not want to engage or be confronted with another perspectives.

Afterwards, the questions from audience were raised. The first topic with the interaction of audience concerned uncovering the truth and getting this truth to people who do not want to know it. Moreover, would it be possible to fact-check the whole world? Allan Hanbury pointed out to the similarity between virus detection and fake news detection and noted that the UK currently discusses the issue. Allan suggested the idea of the Internet browsers and alerts, so that when people would read an article, they should be able to see the alerts.

Although, the question is how to create the browser plugin itself, as there are many complications regarding the mechanics, the

different law in different countries, etc. Moreover, in order to achieve diversity, it is important that many different groups would need to collaborate. The antivirus detection deals with signatures providing insights, for instance when you are downloading something that you should not be. The same would apply in case of fake news. We will not be able to achieve some level of universal truth, although we will be able to know additional insights and signals whether something is trustworthy or not. This could depend on the source, on the text, on the fact who shared it, etc. Nadejda Komendantova added that it is also important whether the article was shared in traditional or social media and whether it contains national or local information. Nadejda agreed that comparing different sources of information is vital.

The idea of similarity between antivirus detection and fake news detection was developed further. Allan Hanbury suggested that treating fake news as a virus is a way how technology can help. Allan claimed that the reason why fake news is so aggressive is because it is profitable, easier to produce and it is optimized for social platforms that share engaging content. Therefore, the fake news detection should be applied as quick as possible – on the social platform itself.

The next topic regarded the threat of fabricated videos, which was brought up by Gerhard Backfried. Gerhard noted that this sub-field is very fast-paced. Rosemary Wolfe suggested the importance of emotional impact of such fabricated pictures or videos.

The last topic concerned the statement that some people who believe in fake news believe it even though they might know that the information is not true. Gerhard Backfried agreed that part of the problem is that some people are closed to arguments.

Afterwards, the conversation moved to ways of countering fake news on social media platforms themselves, as mentioned earlier, and reasons why this could be problematic. The problem lies in the fact that if social media providers would investigate the content and rank it somehow, people who would disagree with the labeling would probably move to another social media platforms. Moreover, another problematic aspect is that the technology can easily work for the profit of both sides – people countering fake news and people responsible for its creation.

5)Conclusions

- The idea of the need of multidisciplinary approach to fake news was mentioned throughout the workshop panel. Fake news can only be successfully combated when experts from different fields and with different backgrounds will cooperate. This cooperation is also needed between the governments and commercial sectors and also with the help of the public.

- The automation that AI brings could be used as a support for readers or recipients of news, meaning it could provide additional information on the specific article.

- The idea of the browser plugin was mentioned by Allan Hanbury, Nadejda Komendantova and before the workshop by speaker Iryna Gurevych.

- What can be problematic when combating fake news is the difference in terminology. There is a need for comprehensive and clear definitions, which can then be used by technology.

- Both technical and social aspects of fake news need to be covered in order to successfully fight them.

- The obstacles of the fake news prevention concerning human behavior include cognitive biases, information bubbles, etc.

- Social media provide fruitful environment for fake news spreading, as emotional and engaging content is successful on these kinds of platforms. Therefore, the fake news spreading should be stopped as soon as possible – on the social platforms themselves. Although, the question is how to alert the users about this harmful content without discouraging them to further use the platforms.

SAIL LABS Technology GmbH

Mariannengasse 14

A-1090 Vienna, Austria

Tel.: **+43-1-58095-0**
Fax: **+43-1-58095-580**

Email: **info4speech@sail-labs.com**

www.sail-labs.com