Report from Dagstuhl Seminar 22172

Technologies to Support Critical Thinking in an Age of Misinformation

Tilman Dingler[*1], **Benjamin Tag**[*2], **and Andrew Vargo**[*3]

1 **The University of Melbourne, AU.** tilman.dingler@unimelb.edu.au
2 **The University of Melbourne, AU.** benjamin.tag@unimelb.edu.au
3 **Osaka Metropolitan University, JP.** awv@omu.ac.jp

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 22172 "Technologies to Support Critical Thinking in an Age of Misinformation". This seminar brought together experts from computer science, behavioural psychology, journalists, and policy makers to examine and define the challenges of misinformation and fake news in the internet and social networks. This included discussions of what constitutes misinformation, technological advances for both spreading and mitigating misinformation, and discussions around policies that can be created and implemented to address propagators, both active and passive, of misinformation. The goal of this report is to summarize and present the various challenges and options for the development and implementation of technologies to support critical thinking.

Seminar April 24–27, 2022 – http://www.dagstuhl.de/22172
2012 ACM Subject Classification Human-centered computing → Human computer interaction (HCI); Human-centered computing → Social networks
Keywords and phrases Cognitive Security, Misinformation, Bias Computing
Digital Object Identifier 10.4230/DagRep.12.4.72

1 Executive Summary

Andreas Dengel
Laurence Devillers
Tilman Dingler
Koichi Kise
Benjamin Tag

The Dagstuhl Seminar on "Technologies to Support Critical Thinking in an Age of Misinformation" ran over a course of three days in April 2022. Each day focused on one specific aspect of the problem of Misinformation and the role technologies play in its worsening and mitigation.

Day 1 put the overall seminar goals and an introduction to the topic into its focus. All participants introduced themselves and gave a concrete example of an important challenge they have identified. The collected challenges were organized and later used as core challenges for group work activities, here Regulations/Policies, Human Factors and Platforms, and Critical Thinking. Over the course of the three days three groups worked on defining challenge statements (Day 1), ideas to solve the issue (Day 2), and concrete Research Questions and Project/Collaboration proposals (Day 3).

[*] Editor / Organizer

Technologies to Support Critical Thinking in an Age of Misinformation, *Dagstuhl Reports*, Vol. 12, Issue 4, pp. 72–95
Editors: Andreas Dengel, Laurence Devillers, Tilman Dingler, Koichi Kise, and Benjamin Tag
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

The theoretical underpinnings of all group discussions and activities were provided by a series of presentations that were topically organized. Day 1 was centered around how the problem of misinformation has evolved and why misinformation is so successful these days. A historical overview was given by keynote speaker Prof. Emma Spiro, which concluded with the key insights that Networks and platforms shape information flow and that attention dynamics matter. The second keynote talk of the day was given by Prof. Andreas Dengel that put light on the crucial role that images and their power to convey information that is tainted with emotional information, and how technology (e.g, CNNs) can be used to detect those, classify them, and can potentially correct them.

On day 2, the participants zeroed in on the role technology plays. Session 1, started with a keynote by Prof. Niels van Berkel on the role of Artificial Intelligence, and Human-AI interaction. Looking at Technology, Society, and Policy on a larger scale, van Berkel identified the core issue that there exists a lack of literacy on the tech side as well as on the regulatory side, a potential consequence of the lack of qualified tech personnel on regulatory bodies. Keynote 2, by Prof. Laurence Devillers, looked at how technology is used to misinform, deceive, and change public opinion, while proposing solutions, such as Nudging and Boosting techniques, how Human-Ai interaction should be better understood, and how research and industry must work together to mitigate the problem of lacking literacy. In session 2 of the day, Prof. Albrecht Schmid led an open, provocative discussion that served as a brainstorming session for the upcoming group work, mainly focussing on the role of platforms and technology. The third keynote was given by Prof. Stephen Lewandowsky who gave a detailed account of the role of human cognition and the larger impact of misinformation on democratic societies. He identified pressure points and proposes countermeasures that are effective but need to be scaled up through improved and coordinated cross-country regulation. Day 2 ended with a Misinformation Escape Room group activity (demo), led by Dr. Chris Coward, which aims at teaching players the power of misinformation and the complexity of the problem.

Day 3 featured the keynote by Roger Taylor which strongly focussed on the way misinformation is regulated globally, and how regulatory frameworks (Digital Service Act) and effective regulation can help to mitigate the misinformation problem. As an advisor to the UK government, and an expert in responsible AI programs and data ethics, Roger Taylor put a light on pain points in the bureaucracy and the misaligned aims of technology development and research, and politics.

2 Table of Contents

3 Overview of Talks

3.1 Misinformation Escape Room: A gamified approach to building resilience to misinformation

Chris Coward (University of Washington – Seattle, US)

URL www.lokisloop.org

"While facts make an impression, they just don't matter for our decision-making, a conclusion that has a great deal of support in the psychological sciences" [3]. This statement poses a fundamental challenge for the field of media and information literacy (MIL), a largely rationalist approach to learning that presumes the underlying problem to be solved is a skills deficit. The emergence of disinformation has not dislodged this conviction for many MIL scholars and practitioners. As one meta review summarizes, there is a prevailing belief that "the bulk of disinformation on the Internet could be combated with basic evaluation skills" [2]. At the same time we have witnessed a growing chorus questioning this conviction, with the most significant shortcoming concerning the psychological dimensions of disinformation, including the role of personal beliefs, social identity, emotion, confirmation bias, motivated reasoning, and epistemic beliefs [1].

In response to these observations and interviews with librarians with front-line experience helping patrons navigate misinformation, a research team led by Chris Coward at the University of Washington designed a misinformation escape room as an immersive, social, and active learning environment. In this Dagstuhl session we will play the escape room, followed by a discussion of the project's goals and research findings.

References

1 Lewandowsky, S. The "Post-Truth' World, Misinformation, and Information Literacy: A Perspective From Cognitive Science. In S. Goldstein (Ed.), Informed Societies (1st ed., pp. 69 – 88). 2019. Facet. https://doi.org/10.29085/9781783303922.006
2 Sullivan, M. C. Why librarians can't fight fake news. Journal of Librarianship and Information Science, 096100061876425. 2018. https://doi.org/10.1177/0961000618764258
3 Wardle, C. and Derakhshan, H. Information Disorder: Toward an interdisciplinary framework for research and policy making. Council of Europe report. 2019. Retrieved from: https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c

3.2 What the world thinks: Trending Topics and Multimedia Opinion Mining

Andreas Dengel (DFKI – Kaiserslautern, DE)

The Internet is full of opinion hidden under tons of irrelevant and unstructured data. From micro-blogging platforms like Twitter to video repositories such as YouTube the users express sentiments about products, brands, institutions and governments. Moreover, users tag other users' opinions: they submit comments in the same micro-blog, link them with other media content, submit brief comments to the videos or just click the "like" option. In the past few years, there has been a huge increment of interest in the analysis of this type of content by opinion consumers, such as companies and media organizations. Among others, companies aim at mining this collective opinion in order to know what people think and how they feel about products and services. The main drawback of current solutions is that they only consider the textual content, ignoring other sources and modalities of opinion and its cross-media relationships.

This talk addresses the challenge of opinion mining of multimedia content from the Web. This comprises a multi-modal analysis of social media streams and their underlying network dynamics considering different media channels such as Twitter, YouTube, Flickr, Google, and Wikipedia. It specifically proposes solutions for:

1. The detection of trending topics from a large set of dynamic data streams. These trending topics were able to be clustered, tracked and aggregated over social media channel and over time. In particular the combination of statistical methods with linked open data was of help to achieve this goal (see [3, 4, 6]).
2. The content-based multimedia analysis on single modalities and combination of multi-modalities. Here, the early shift from traditional approaches towards deep learning proofed to be in particular successful. Key element of analysis were Adjective Noun Pairs (ANP) providing a mid-level representation for visual content and foundation for sentiment analysis. The idea of ANPs was further extended towards Verb-Noun-Pairs to capture temporal concepts present in audio and video streams (see [1, 2, 5, 7])

References

1 M. Al-Naser, S. M. Chanijani, S. S. Bukhari, D. Borth, and A. Dengel. What makes a Beautiful Landscape beautiful: Adjective Noun Pairs Attention by Eye-Tracking and Gaze Analysis. In ACM Workshop on Affect and Sentiment in Multimedia (ASM), 2015.

2 B. Bischke, D. Borth, and A. Dengel. Large-Scale Social Multimedia Analysis. In Vrochidis Stefanos and Huet Benoit and Chang Ed and Kompatsiaris Ioannis, editor, Big Data Analytics for Large-Scale Multimedia Search. Wiley & Sons, Ltd., 2018. (to appear).

3 S. Elkasrawi, H. Elwy, S. Bauman, C. Reuschling, and A. Dengel. Prediction of Social Trends Using Nearest Neighbours Time Series Matching and Semantic Similarity. In Advances in Data Mining, 16th Industrial Conference, ICDM 2016, Poster Proceedings, 2016. (to appear).

4 S. Fuchs, D. Borth, and A. Ulges. Trending topic aggregation by news-based context modeling. In Joint German/Austrian Conference on Artificial Intelligence (Künstliche Intelligenz), pages 162 – 168. Springer, 2016.

5 J. Folz, C. Schulze, D. Borth, and A. Dengel. Aesthetic Photo Enhancement using Machine Learning and Case-Based Reasoning. In ACM Workshop on Affect and Sentiment in Multimedia (ASM), 2015.

6 A. Koochali, S. Kalkowski, A. Dengel, D. Borth, and C. Schulze. Which languages do people speak on flickr?: A language and geo-location study of the yfcc100m dataset. In Proceedings of the 2016 ACM Workshop on Multimedia COMMONS, pages 35 – 42. ACM, 2016.

7 S. Kalkowski, C. Schulze, A. Dengel, and D. Borth. Real-time Analysis and Visualization of the YFCC100m Dataset. In ACM Multimeda MMCOMMONS Workshop, 2015.

3.3 AI-enhanced nudging mechanism using affective computing: ethical issues

Laurence Devillers (CNRS – Orsay, FR & Sorbonne University – Paris, FR)

Ethics, Goals, and Societal impact have always been central subjects since the early days of the field of research on artificial intelligence such as affective computing. But currently, the new uses of social robots, affective conversational agents (chatbots), and, more generally, the so-called "affectively intelligent" digital environments in fields as diverse as health, education, insurance, transport, or economics reflect a phase of significant change in human-machine relations, amplify the necessity to keep great attention in ethical dimensions of these systems. What ethical issues arise from the development of affective computing with chatbot/robot interaction? Does it raise the crucial issue of trust? How will humans co-learn, co-create and co-adapt with the Machine? Notably, how will vulnerable people be protected against potential threats of the machine? During an interaction, we adapt our linguistic behaviors but also our prosodic and gestural behaviors and our conversational strategies. This multi-level adaptation can have several functions: reinforcing engagement in interaction, emphasizing our relationship with others, and showing empathy. Anthropomorphism introduces many challenges, among them ethical, uncanny valley, practical implementation, and user mind-reading problems. The anthropomorphic goal of "just like a human-to-human conversation". The designers of conversational agents seek for many to imitate, simulate the dialogical behavior of humans, and users spontaneously anthropomorphize the conversational agents" capacities and lend them human understanding. Thus, the Dilemma of the researchers is, on the one hand, to achieve the highest performance with conversational virtual agents and robots (close to or even exceeding human capabilities) but on the other hand, to demystify these systems by showing that they are "only machines".

Conversational agents and social robots using autonomous learning systems and affective computing will change the game around ethics. We need to build long-term experimentation to survey Human-Machine Co-evolution and to build "ethics by design" chatbots and robots. In the chair HUMAAINE (head: L. Devillers, LISN-CNRS, France), we aim to study the Human-Machine Affective interactions and relationships, in order to audit and measure the potential influence of intelligent and affective systems on humans, and finally to go towards a conception of "ethical systems", by design or not and to propose evaluation measures. For this purpose, the planned scientific work focuses on the detection of social emotions in a human voice, and on the study of audio and spoken language "nudges" [1, 2], intended to induce changes in the behavior of the human interlocutor.

Nudging is an ethically highly problematic topic. A digital nudge is an almost imperceptible incentive in the design of a digital system to drive behavior that is supposed to improve personal or collective well-being. Digital nudges use personal data and biometric sensors to profile and encourage people to take unintended actions while using familiar online technologies such as email, pop-ups, SMS, web interfaces, smart watches, mobile apps, IoT, home appliances, smart cars, chatbots, robots, etc. However, when a digital nudge is enhanced by Artificial Intelligence systems (so-called AI-enhanced nudge) using machine learning and affective computing technologies based on cognitive biases and behavioral science, its potential is immense. While its usage can be beneficial for an individual or the society, the AI-enhanced nudge persuasive power and intrusive capacity can also cause subliminal manipulations and profound and long-lasting changes in the behavior of users, especially children, and vulnerable people. If AI-enhanced Nudges are (intentionally or not) misused, they may become dangerous and raise serious ethical issues that can undermine the level of distrust in AI-enhanced nudging systems. AI-enhanced Nudging is already a reality influencing the actions and behaviors of thousands of people in the fields of education, health, gaming, gambling, hospitality, smart cities, security, justice, etc. However, this "soft" manipulation of behavior and emotions raises ethical questions to which no standard today provide direct answers. As AI-enhanced Nudging systems are flourishing in the market, it is widely believed that their design and the use of them, in the short or long term, ought to establish human and social responsibility through auditable behaviors under a typical set of conditions. This could help to build trustworthiness within a sustainable market. There is an urgent need to create a shared terminology and consensual processes and methodologies to mitigate and ethically adjust the enormous ability of people's manipulation provided by AI technologies such as affective computing to digital nudge [3].

References

1 H. Ali Mehenni, S. Kobylyanskaya, I. Vasilescu, L. Devillers, Nudges with a conversational agent or social robot: a first experiment with children at a primary school, IWSDS 2020

2 N. Kalashnikova, S. Pajak, F. Le Guel, I. Vasilescu, G. Serrano, and L. Devillers, *Corpus Design for Studying Linguistic Nudges in Human-Computer Spoken Interactions*, Language Resources and Evaluation Conference 2022 (LREC 2022), Marseille, France 2022, June 2022.

3 L. Devillers, E. Panai, Ad-hoc group 6: AI-Enhanced nudges (AFNOR/CEN-CENELEC/JTC21)

3.4 From Cognition-Aware to Bias-Aware Systems

Tilman Dingler (The University of Melbourne, AU)

Joint work of Tilman Dingler, David A. Eccles, Martin Pielot, Benjamin Tag

With advancements in sensing and processing power and more sophisticated machine-learning methods, computing systems can increasingly detect and monitor human activities. Computers that consider the context in which they are used can support their users according to their current location, activities, and intent, a field coined context-aware computing [9]. In our work, we have extended this notion to also include the user's cognitive context to build systems that help users increase their ability to effectively process information according to their current mental state. By utilising phone sensor data, for example, we have trained machine-learning algorithms to detect when people are attentive to their phones [3] and seek stimulation [8]. Insights into when a person is bored or focused can provide us with a better understanding of when people are more productive and when downtimes occur: during highly focused states, devices in the user's environment can be advised to prevent interruptions in order to help people focus better. Beyond in-situ assessments of cognitive states, we have developed tools and methods to elicit users' circadian rhythms of alertness [1, 2, 10], which describe systematic cognitive performance fluctuations throughout the day. Awareness of these rhythms opens up a whole range of opportunities to suggest content, schedule a day full of work, or generally recommend activities whose cognitive requirements match the user's current state [5]. The resulting tools and algorithms give insights into the user's internal body clock, which helps people to better schedule, for example, learning sessions or generally activities that require high focus. In recent years, however, it has become apparent that effectively dealing with information is not necessarily a matter of consuming more in less time but of the quality of the information processing itself. Society is transitioning into an era where computing pervades all aspects of people's lives, with humans and their cognitive processes at the centre of it. The rise of fake news and the interplay between bad actors, fast dissemination through social media, and people's receptivity to emotionally charged content present an ever-growing challenge to individuals, society, and our democratic institutions [7]. When looking at what can be done about its reception, we have identified several preventative interventions to bolster people against fake news. These include media literacy training, psychological inoculation, and transaction cost economics [6]. Further, receptivity to fake news often comes from the prevalence of cognitive biases. They play an important role in how information is perceived and processed, a fact that can be both utilised and exploited by computing systems. A prominent example of a cognitive bias is the confirmation bias, i.e., the tendency to seek out information that confirms our existing perspectives and notions. We have recently established a strand of research to use sensors to detect the occurrence of cognitive biases. Computing systems capable of detecting cognitive biases, which we call bias-aware systems, can help people increase their awareness of and mitigate their effects as well as inform recommender systems to introduce a more balanced news diet. One of the main challenges is the collection of ground truth, i.e., ensuring we can successfully induce and measure the occurrence of cognitive biases for observational and experimental research. Therefore, we developed a tool to collect ground truth on people's implicit preferences that can be adapted to any thematic issue, such as opinions on climate change, feminism, or political ideologies [4]. The Dagstuhl Seminar on "Technologies to Support Critical Thinking in an Age of Misinformation" was born out of the realisation that the problem of fake news can only be addressed in a truly interdisciplinary fashion as it involves the technology through which fake news spread, the human who creates, receives and shares it, and the regulatory bodies who are looking for ways of reeling in its spread. Throughout the 3-day seminar, our team sat down with behavioural psychologists, government advisers, and technologists to discuss the human element in this triangle of technology, human, and government. The goal of our bias detection research is to allow people to increase their awareness of their innate biases and allow systems to help mitigate them. Media literacy training, on the other hand, can help current and future generations of technology users critically process online information. Future computing systems need to be designed responsibly to consider people's cognitive biases and help them bolster against their cognitive vulnerabilities. Technology is thus the ouroboros of fake news, i.e., its enabler and mitigator.

References

1. Dingler, Tilman, Albrecht Schmidt, and Tonja Machulla. *Building cognition-aware systems: A mobile toolkit for extracting time-of-day fluctuations of cognitive performance.* Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1, no. 3 (2017): 1-15.
2. Dingler, Tilman, Ken Singer, Niels Henze, and Tonja-Katrin Machulla. *Extracting Daytime-Dependent Alertness Patterns from Mobile Game Data.* In 22nd International Conference on Human-Computer Interaction with Mobile Devices and Services, pp. 1-6. 2020.
3. Dingler, Tilman, and Martin Pielot. *I'll be there for you: Quantifying Attentiveness towards Mobile Messaging.* In Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services, pp. 1-5. 2015.
4. Dingler, Tilman, Benjamin Tag, David A. Eccles, Niels van Berkel, and Vassilis Kostakos. *Method for Appropriating the Brief Implicit Association Test to Elicit Biases in Users.* In

CHI Conference on Human Factors in Computing Systems, pp. 1-16. 2022.

5 Dingler, Tilman, Dominik Weber, Martin Pielot, Jennifer Cooper, Chung-Cheng Chang, and Niels Henze. *Language learning on-the-go: opportune moments and design of mobile microlearning sessions.* In Proceedings of the 19th international conference on human-computer interaction with mobile devices and services, pp. 1-12. 2017.

6 Eccles, David A., Sherah Kurnia, Tilman Dingler, and Nicholas Geard. *Three Preventative Interventions to Address the Fake News Phenomenon on Social Media.* In Proceedings of ACIS, 2021.

7 Lazer, David MJ, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger et al. *The science of fake news.* Science 359, no. 6380 (2018): 1094-1096.

8 Pielot, Martin, Tilman Dingler, Jose San Pedro, and Nuria Oliver. *When attention is not scarce-detecting boredom from mobile phone usage.* In Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing, pp. 825-836. 2015.

9 Schilit, Bill, Norman Adams, and Roy Want. *Context-aware computing applications.* In 1994 first workshop on mobile computing systems and applications, pp. 85-90. IEEE, 1994.

10 Tag, Benjamin, Andrew W. Vargo, Aman Gupta, George Chernyshov, Kai Kunze, and Tilman Dingler. *Continuous alertness assessments: Using EOG glasses to unobtrusively monitor fatigue levels In-The-Wild.* In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1-12. 2019.

3.5 Technology and Democracy: Cognitive Remedies

Stephan Lewandowsky (University of Bristol, GB)

Democracy is in retreat or under pressure worldwide. Even in countries with strong democracies, polarization is increasing, and the public sphere is awash in misinformation and conspiracy theories. Many commentators have blamed social media and the lack of platform governance for these unfortunate trends, whereas others have celebrated the Internet as a tool for liberation, with each opinion being buttressed by supporting evidence. One way to resolve this paradox is by identifying some of the pressure points that arise between the architecture of human cognition and the online information landscape, and their fallout for the well-being of democracy. Two such pressure points arise from the algorithmic curation of content and the prevalence of misinformation and disinformation on social medial. Virtually everything users see on the internet is curated by intelligent algorithms (e.g., the newsfeed on Facebook or Twitter). These algorithms are designed by platforms without public accountability or auditing, with the primary intent of keeping users engaged longer by satisfying their presumed preferences. While preference satisfaction by itself is not a threat to democracy, it can become problematic if extremist or conspiratorial content keep a person engaged longer, because the platforms are then incentivized to prevent more and more potential harmful content. Algorithms may thus at least indirectly imperil our democracy when people are being radicalized or are presented with misinformation not because they want to, but because platforms are making money by facilitating it. Algorithmic content curation can additionally be problematic if people's personal data are used to identify sensitive attributes, such as their personality or sexual orientation, which political operatives can then exploit by presenting messages to people that exploit their personal vulnerabilities. This process is known as microtargeting and it comes with a number of attributes that may imperil democracy. In the absence of any regulation, one possible countermeasure involves "boosting" people's ability to detect on their own when they might be targeted by manipulative messages. A existence proof of boosting showed that once people were given information about their own personality along the introversion-extraversion spectrum, they were better able to identify advertisements that were aimed at them based on their personality. Similar "boosting" approaches can also equip people to become resilient to misinformation and disinformation online. This approach is known as inoculation and it entails warning people ahead of time that they might be misled, and providing them with information about the misleading rhetorical techniques they are likely to encounter. Inoculation has been shown to be effective in numerous different domains, from anti-vaccination messages to radicalization attempts and conspiracy theories. In all cases, people's ability to detect when they are manipulated was significantly enhanced by inoculation. Notwithstanding the success of such cognitive countermeasures, they are insufficient to counter the immense asymmetry in power between platforms and users that currently exists and that gives rise to the pressure points between cognition and technology. It requires deep structural change and smart regulation to create a new Internet with democratic credentials.

3.6 Regulation of Misinformation

Roger Taylor (Open Data Partners – London, GB)

Different countries are taking very different approaches to regulation of social media to combat misinformation. Singapore has passed a law against telling lies which allows the government to order the take down of material regarded as untrue. China is seeking to register anyone who comments publicly about key political issues online. The European Union is bringing in regulation that makes social media responsible for harms which include harms to democracy and civic discourse as well as harms to fundamental rights. The UK is proposing more limited regulation that focuses on immediate harm to individuals rather than harm to society (but which still might capture medical misinformation). The US is adopting a more laissez-faire attitude based on giving primacy of freedom of speech. However, within the US, individual platforms are implementing their own governance mechanisms in recognition of public pressure for change.

These regulatory strategies do not specifically call for action on critical thinking. However, platforms may respond to the European regulatory proposals by adopting measures such as misinformation vaccination. (Also, there are, in some territories, complementary strategies on media literacy alongside regulatory proposals – the EU strategy on disinformation).

Key limitations to the successful implementation of regulation are: A lack of social consensus around the meanings of the words used (e.g. "harm to public discourse" "psychological harm"). Lack of agreed mechanisms that are capable of determining whether such harm has occurred. Lack of mechanisms for determining responsibility. (Regulations require platforms to balance rights to free speech against risk of harm. The issue for regulators is whether they have found the right balance. This requires a determination of whether or not they could have done better in balancing these risks which, in turn, requires an understanding of what is possible in order to make a sound assessment of responsibility.)

Each of these issues is exacerbated by the rapidly changing and hugely heterogenous nature of the harms being addressed, as well as the complexity of the environment that is being regulated.

Regulators will likely have to adopt an approach based on identifying the most egregious harms and using rough and ready measures to assess the responsibility of the platform. The degree to which this will significantly impact disinformation and online harms is uncertain.

Regulators would be wise to adopt a strong stance at the outset with regard to the data access provisions in the EU regulations. They should set out a long term strategy to establish

- relatively objective/consensual approaches to categorising and monitoring misinformation;
- research methods to understand the impact of misinformation on individuals and on democracy;
- and mechanisms for understanding the relative impact of different types of remedy including media literacy and critical thinking.

3.7 Move Slow and Fix Things: Algorithms, Systems, and Design

Niels van Berkel (Aalborg University, DK)

The efforts toward technologies to support critical thinking in an age of misinformation require a collaborative effort across the fields of Technology, Society, and Policy. In this appetiser talk, I will outline some of the primary challenges faced in each area, pointing to promising research that indicates opportunities for moving forward.

TECHNOLOGY
Challenges faced within the technology field include biases, biased algorithms, and black box decision-making. Therefore, it is critical to recognise the real-world consequences of algorithmic systems. Examples include disparities in AI skin cancer diagnoses between different skin colours and discrimination built into the design of a fraud detection system of the Dutch tax authorities. Recent work by Huszár et al. analyses the amplification of tweets by elected legislators from major political parties in seven countries [1]. Their results show that the mainstream political right enjoys higher algorithmic amplification. While highlighting the possibility of assessing the impact of algorithmic-driven recommendation systems, it also raises new questions, including; Should distribution always be a perfect 50/50 split? Are politics as black and white as left / right? What can we do about this technological bias?

SOCIETY
Disinformation, filter bubbles, and an increased polarisation in politics and beyond are amongst the challenges currently faced in society. Current events, such as the Russian invasion of Ukraine, bring to the front the societal challenges related to disinformation. Disinformation also plays a significant role domestically, with the US being a famous example of the growing ideological divide between Democrats and Republicans. In democratic countries with a multi-party democratic system, such as The Netherlands, polarisation can take a different form – with traditional parties finding it increasingly challenging to distinguish themself from one another. Recent work by Broockman and Kalla studies the effect of paying Fox News viewers to regularly watch CNN (two politically opposed media channels) [2]. Compared to a control group of Fox News viewers, the study finds more nuanced political

beliefs and knowledge of current events. The authors highlight how the skewing of media has had a broader and negative impact on how US society functions. Highlighting the opportunity for viewers to obtain more nuanced viewpoints once presented with an alternative media source, the study raises new relevant questions, including the potential for long-term effects and the impact of removing the financial incentives offered in the study.

POLICY
With the growing impact of technology in an increasingly unstable world, policy is often looked at as the instrument to bring back some stability. Simultaneously, policy can be perceived as a slow-moving instrument which struggles in dealing with local versus global issues. In this context, we increasingly see the global impact of national and international governmental organisations. In particular, the US and the EU are at the forefront of developing AI policies and research plans. A recent review by Jobin et al. studies the global landscape of AI ethics guidelines [3]. Their results highlight eleven unique ethical principles that are discussed within these guidelines, including "transparency", "justice", and "privacy". Their study shows an apparent interest of both governmental organisations and industry in developing ethics guidelines while also presenting questions for further research. For example; How do we implement these guidelines in products and services? Can we match policy with outcomes? How do these, primarily developed in Europe and North America, impact the rest of the world?

References

1 Huszár, F., Ktena, S. I., O'Brien, C., Belli, L., Schlaikjer, A., & Hardt, M. (2022). *Algorithmic amplification of politics on Twitter*. Proceedings of the National Academy of Sciences, 119(1).

2 Broockman, D., & Kalla, J. (2022). *The manifold effects of partisan media on viewers' beliefs and attitudes: A field experiment with Fox News viewers*. OSF Preprints.

3 Jobin, A., Ienca, M., & Vayena, E. (2019). *The global landscape of AI ethics guidelines*. Nature Machine Intelligence, 1(9), 389-399.

4 Working groups

On the first day of the seminar, participants jointly collected and discussed a range of challenges that misinformation on digital platforms presents. Throughout the day, we arrived at six common themes:

1. **Critical Thinking**: what are critical thinking skills students and, more generally, online users need to have to navigate online platforms? How can critical reflection be prompted? How can we design platforms to help users break down filter bubbles?
2. **Regulation and Policies**: How does an empirical approach towards research and policy development with regard to misinformation look like? How can we systematically gather evidence of the impact on interventions on false beliefs? Which measures (algorithm policing, platform policies, education) are most impactful? How do we regulate online discourse without breaking the fundamentals of a pluralistic society?
3. **Discourse *vs.* Censorship**: How can policing be done without silencing non-wanted voices? What does *good* or *healthy* online discourse look like? Research mixed with activism can be problematic. Certain opinions are difficult to express at universities. Does the notion of safe spaces lead to places where contrary opinions are not expressed anymore? What is the cost of not talking to each other?

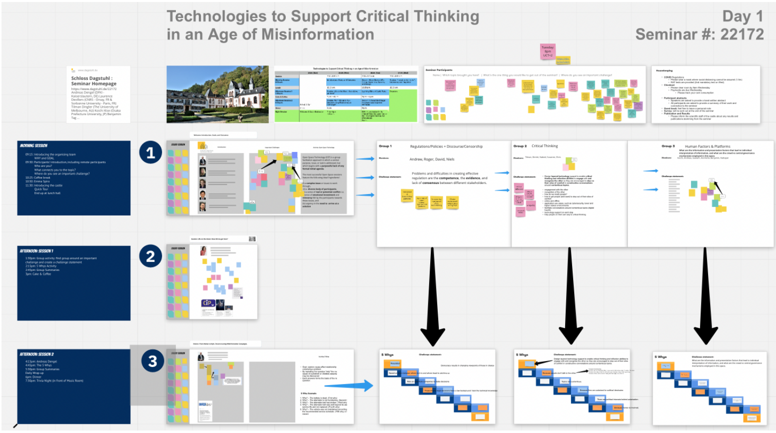

Figure 1 The Miro Digital Whiteboard Used to Facilitate Interaction and Group-Work for the In-Person and Remote Participants.

4. **Human Factors**: how can we model human cognition/emotion to make reliable predictions of behaviour? What is a meaningful cognitive architecture to enable students to think critically? And to what extend can fake information be used to achieve/do more, *i.e.*, taking benevolent advantage of it?
5. **Platforms**: how can we control misinformation across different platforms? What happens when platforms intervene but fail? Can research lead to an early detection of where technology contributes and where technology goes wrong? What role do algorithms play, and how can we design *better* ones?
6. **Bad Information**: a common problem seems to be that people don't make decisions based on facts. How can we reconcile this and bring facts back into the decision-making process?

After some discussion, we identified some overlaps and merged the *Human Factors* with *Platforms* and *Regulation and Policies* with *Discourse vs. Censorship*. The *Bad Information* scheme seemed to be ubiquitous, hence we settled on three final themes, around which we formed the following working groups.

4.1 A Governance Framework for faster Technology Regulation

David Eccles (The University of Melbourne, AU)

Joint work of Eccles, David; Taylor, Roger; van Berkel, Niels; Vargo, Andrew.

4.1.1 Challenge Statement

Problems and difficulties in creating effective regulation are the **competence**, the **evidence**, and **lack of consensus** between different stakeholders.

4.1.2 Working Group

Our role as an information systemist and human computer interaction researchers is on the intersection between people, their processes, data, and technology. We are interested in empowering individuals to better understand how their data and the technology they use reinforces an information and knowledge asymmetry. I spent most of my time at the Dagstuhl

Seminar in the governance stream as the threat to democratic processes and institutions is evident from my research on fake news on social media phenomenon. Residing in a country which has compulsory voting for all citizens and permanent residents over the age of 18, it has an impact on elections as political parties have to move to the centre not the extremes of liberal and conservative issues to win elections and govern [5]. Democracy is under direct threat from social media platform technologies implicit and explicit role in the fake news phenomenon. Today's world wide web is no longer the place of free and open exchange of information and ideas as envisioned by its creators [2]. The world wide web and in particular social media platforms that have arisen because of the internet's capabilities in their current form and function represent a threat to democratic processes such as elections and democratic institutions fraying the separation of powers (legislatures, executive, and judiciary) [1]. Arguments can readily be made for legislative intervention using individual, societal, and government reasons [3]. A regulatory intervention for economic reasons is rarely justified except when there is a deformity in the function of a free market. Examples of market failure include its devolution to a private exchange excluding new entrants, situations where there is extreme information and or data asymmetry in the relationship between parties creating an unfair commercial and / or negotiating advantage, or an oligarchy, monopoly or duopoly exists being able to manipulate market price through collusion in supply and demand [6]. The overwhelming dominance of five key vendors Meta (formerly Facebook), Apple, Google, Microsoft and Amazon in the world wide web advertising and social media platforms demonstrates an economic failure of the market. Interventions of this kind are not new (e.g., Standard Oil, Bell Corporation), however, any legislative intervention in markets is not without unintended outcomes. Interventions may bring benefit, may bring negative, or even catastrophic unforeseen consequences [4]. In the existing market of social media platforms and the world wide web this can be demonstrated by the unintended consequences of the EU's GDPR legislation. We propose a governance model of individual's sensitive private user internet data that separates user data from its application and use. We believe this model would be a minimum imposition on technology vendors, restore the market imbalance allowing for greater competition and new entrants, and lesson the data and information asymmetry between social media users and vendors. In this model data at rest would not reside with any commercial vendor but with a statutory authority as does user statistics for bodies such as the U.S. Census Bureau, and is already in place for sensitive personal data as in the case of the Australian Digital Health Agency.

References

1. Allcott, H., & Gentzkow, M. Social Media and Fake News in the 2016 Election. 2017. The Journal of Economic Perspectives, 31(2), 211 – 235.
2. Hern, A. Tim Berners-Lee on 30 years of the world wide web: 'We can get the web we want', Interview. 2019. The Guardian Retrieved from https://www.theguardian.com/technology/2019/mar/12/tim-berners-lee-on-30-years-of-the-web-if-we-dream-a-little-we-can-get-the-web-we-want
3. House of Commons Digital. Disinformation and "fake news": Final Report. Westminster House of Commons (UK Government). 2019. Retrieved from https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/1791.pdf
4. Merton, R. K. The Unanticipated Consequences of Purposive Social Action. 1936. American Sociological Review, 1(6), 894 – 904.
5. Swire, T. B., Ecker, U. K. H., Lewandowsky, S., & Berinsky, A. J. They Might Be a Liar But They're My Liar: Source Evaluation and the Prevalence of Misinformation. 2020. Political Psychology, 41(1), 21 – 34.
6. Williamson, O. E. The economic institutions of capitalism: firms, markets relational contracting. 1987. Free Press.

4.2 Working Group on Critical Thinking

Tilman Dingler (University of Melbourne, AU)

4.2.1 Challenge Statement

How can we design layered technology support to enable critical thinking and reflection abilities to engage with and recognise the other so they are encouraged to step out of their silos of comfort in constructive conversations around contentious topics?

4.2.2 Working Group

This working group focused on what can be done to foster critical thinking abilities in online users and ways of bringing media literacy education into young people's curricula.

Other than relying on government regulation, platform policies, or technological interventions, the idea was to strengthen the critical thinking abilities of online users and prevent cognitive vulnerabilities. One of the premises of acknowledging or accepting other people's viewpoints is to engage with *the other* in the first place. The group discussed the necessity of stepping out of people's comfort zone and making an effort to get to know and understand other people's perspectives. Social media and also universities more generally tend to form silos of comfort where like-minded people discuss topics from one congruent angle. But the online world is made up of a plethora of perspectives, which reflects the plurality of our society. The risk of these silos or echo chambers [1] is that people with differing viewpoints do not engage with and, as a result, further alienate each other. Democracy is built on a pluralistic society where viewpoints need to be discussed in the open to reach an agreement or compromise. Being exposed to other viewpoints and the ability to critically engage, understand and find compromise are crucial for members of our society. To build and foster this ability, critical thinking skills and media literacy should be systematically integrated into all levels of our education system. We should encourage students to leave their comfort zones in an attempt to understand, learn, and re-evaluate their own standpoints.

The group collated a range of techniques and interventions to teach critical thinking abilities, including:

- Inoculation: the goal is to build psychological immunity against misinformation. This technique has famously been applied by Roozenbeek *et al.* [2] in their *Bad News* game, where players have to apply the tricks of the trade of misinformation to get their fictional message out as wide as possible. The idea is that by getting exposed to common mechanisms of creating and spreading misinformation, receivers of inoculation training acquire the ability to spot and resist manipulation.
- Debunking / Prebunking: This is the attempt to correct misinformation after (debunking) or prior to (prebunking) exposure. This can be as simple as uncovering argument-supporting facts or finding flaws in the argumentation itself. Cook and colleagues [3], however, discuss in detail why people often struggle with correcting misinformation and inaccurate beliefs and why debunking misinformation is not always straight forward. For one, there is the risk of "backfire effects," which arises when, rather than being refuted, people double down on their preconceived notions. And second, there is the role of worldviews in accentuating the persistence of misinformation. Lewandowsky *et al.* created a series of writings on how to apply debunking effectively [4].

- Building Empathy: the idea to put oneself into the shoe of another and walk a thousand miles in it. Seeing the world from someone else's perspective is crucial to understanding where the other is coming from. Connecting and empathising are the key to meaningful conversations, especially around critical topics. Only when we understand the other we can have a generative dialogue, i.e., allow the other person to contribute to our knowledge and understanding of the world around us.
- Moderation: to facilitate contentious conversations, moderation might be necessary to bring people with different views to the table. Philosopher Jürgen Habermas formulated the *ideal speech situation*, a set of basic rules that are based on reason and evidence.

Efforts to integrate critical thinking education effectively into curricula of all levels need to be based in pedagogy and philosophy. For online learning and discourse, this requires a research agenda around measuring the effectiveness of different approaches. Which skills are more successful in fighting off misinformation? How can we implement digital interventions that build spaces to meet people with opposing opinions to allow them to learn about the existence of other opinions and help them value those? And how can we measure the effectiveness of these interventions? These questions built the basis for solution proposals that members of the working group would like to take forward, such as:

1. Collect existing interventions and gather empirical evidence for their effectiveness. The *Prosocial Design Network*[1] is an online space where ideas and empirical evidence supporting those is being gathered. The group would use the platform to collect ideas and inspiration for future studies.
2. Opportunistic education: social media platforms, such as Twitter or Facebook, already flag potential mis- or harmful information. The effect of such flags on users' critical thinking skills should be assessed. This requires, however, that platforms will collaborate with researchers on conducting such experiments and data collections. Such collaborations could be open the door to changes in regulation.
3. Tools and Methods to test interventions: we need standardised ways to test the effectiveness of new interventions to allow benchmarking and comparison.
4. Integration of critical thinking and media literacy programs into all levels of the schooling systems. While media literacy training has made its way into middle school education, the quickly changing nature of the online discourse requires a frequent revisiting of these contents and techniques. In schools, universities and vocational training.

References

1 Pariser, Eli. The filter bubble: What the Internet is hiding from you. penguin UK, 2011.
2 Roozenbeek, Jon, Sander van der Linden, and Thomas Nygren. "Prebunking interventions based on the psychological theory of "inoculation" can reduce susceptibility to misinformation across cultures." Harv Kennedy Sch Misinformation Rev 2020b 1 (2020).
3 Cook, John, Ullrich Ecker, and Stephan Lewandowsky. "Misinformation and how to correct it." Emerging trends in the social and behavioral sciences: An interdisciplinary, searchable, and linkable resource (2015): 1-17.
4 Lewandowsky, Stephan, John Cook, Ullrich Ecker, Dolores Albarracin, Michelle Amazeen, Panayiota Kendou, Doug Lombardi et al. The debunking handbook 2020. 2020.

[1] `https://www.prosocialdesign.org/`

4.3 Human Factors and Platforms

Benjamin Tag (The University of Melbourne, AU)

4.3.1 Challenge Statement

What are the **information and presentation factors** that lead to individual interpretation of information? What are the **crowd vs. central** governance mechanisms employed in this space?

4.3.2 Working Group

The members of work group "Human-Factors and Platforms" were researchers in Human-Computer Interaction, Misinformation, and directors of research for commercial entities. The expertise of the group members covers document analysis, social media analysis, artificial intelligence, as well as research in trust, safety, and algorithmic responsibility. The work group focused on the question: "What are the information and presentation factors that lead to individual interpretation of information, and what are the crowd vs central governance mechanisms employed in this space." During the initial discussion session, the members identified a series of problems underlying and deriving from this core challenge. The main challenge that all members identified as crucial is that academic research and industry partners have to collaborate better, i.e., more openly. Misinformation mostly spreads through platforms built, maintained, and promoted by a relatively small group of companies. However, while these platforms are connected to certain degrees, it is extremely difficult to control the cross-platform migration of misinformation. As it will be difficult to design regulations that satisfy the needs of different platforms as well as that of public and regulators, the group agreed that the human factor rather than the infrastructural aspect of the spread of misinformation should be put into the research focus, here especially a better understanding of human cognitive architecture, and the development of tools to support critical thinking. Because the most powerful way to stop misinformation is arming people against them. This, however, should be supported by technical solutions, such as providing meta-data, automated fact-checking, and providing warnings. Following this initial definition of the problem statement, the work group used the 5-Whys technique to probe the causes of why there is no solution yet to clearly identifying what is false and what is right. One problem is that we cannot measure truth, therefore, we cannot fully trust and rely on sources, links, and services. These often lack full transparency, and tend to hide a purpose or motivation, e.g., biasing information in favor of large sponsors. The final conclusion of this exercise is that people or companies often try to come ahead of others, e.g., to gain advantages in funding through advertising, making them more powerful. The philosophical conclusion of this exercise was that you can create power by creating your own reality, as many recent political campaigns have shown.

Based on these findings, the group started to dive into the solution exploration through a reverse thinking exercise. When it comes to Human Factors, the most promising solutions to understanding why misinformation is read, believed, and distributed by humans, are "silent" solutions. This means that researchers should use unobtrusive and non-invasive sensing solutions to make sense of human cognition, without altering the human's environment and context, which potentially leads to altered behavior, e.g., through the Hawthorn Effect. Many of the necessary sensors and technologies are already integrated in computers, smartphones, and wearable devices, such as smart watches. It is therefore not necessary to develop new sensing solutions. Rather, are researchers in academia and industry required to collaborate and better identify mutual needs and rights. Here, especially the access to information collected by platforms, e.g., Facebook, is deemed extremely helpful to researchers, allowing them to create insights on human behavior. The group also discussed the importance of these collaborations intensively. Academic and industry researchers in the group agreed that both sides have to better communicate timelines (industry and academia differ substantially), the creation of necessary output, and the creation of IP. Finally, the work group summarized these discussions in two research questions:

1. How can we teach people to have a healthy mix of trust and mistrust towards machines/machine output?
2. How can physiological data be used to make interactions with platforms more intuitive/simple – while also protecting privacy?

These research questions informed the group's last task, that of defining a set of short-, mid-, and long-term projects that help tackling the identified problems. The short term projects aim at sensing the platform impact on the user. A core issue is to develop methods that help to anonymize (physiological) user data without making them useless in order to protect privacy, which increases trust in the platforms, while enabling the full data analysis spectrum to create actionable insights. Mid term projects shall take advantage of the insights, and help researchers and developers to build intervention and nudging systems that help users follow a more balanced information diet, while allowing for data to be shared with, e.g., coaches or data analysis tools. To tackle the question of quantifying truth and identifying truth, the members discussed the idea to crowd-source information from events that make the news. Today, the majority of users carries smart devices that allow for filming, reporting, and so on. However, the big challenge here is to protect the privacy of these citizen journalists to protect them from becoming victims of targeted campaigns. Image processing and smart algorithms (NLP) will allow for an analysis of these large amounts of data. Based on the short term project, an effective way to better understand the impact platforms have on humans, is the development of an affective middle-ware. This can not only be used for better understanding of the impact, but also as a source for individualizing apps, news distribution and provide a more balanced information intake. Last but not least, the group members agreed, that in the long run, we have to aim at understanding, i.e., quantifying, how people create truth out of information, and how they decide what information should be prioritized over other.

5 Open problems

5.1 Intellectual humility: a virtue worth pursuing in public discourse?

Nabeel Gillani (MIT – Cambridge, US)

Intellectual humility is the recognition that what we believe might, in fact, be wrong. What would it mean to have more intellectual humility in our online public discourse? Could it improve how we communicate with, perceive, and ultimately treat one another? How might we design for greater intellectual humility (e.g. through changes / additions to online discourse platforms)? This lightning talk poses these questions and offers examples of how we might design tools and systems for fostering greater intellectual humility in order to foster group discussion about the potential merits and pitfalls of more intellectual humility in our online lives.

5.2 Open problems I found at the seminar

Koichi Kise (Osaka Prefecture University, JP)

The issue of misinformation and disinformation is not just technical but related to different factors such as human cognition, and social sciences. What are correct and fake are not always clearly defined, nor shared by all people. They are often relative, depending on the standpoints of people. Some people can be easily affected by the given (mis/dis)information, but some cannot change their way of thinking even if it is better. This seminar was a good starting point for me to think about the issue with the help of talented participants from a wide variety of fields. It is mandatory to discuss the issue with such people to avoid tunnel vision. At the beginning of the seminar, it was sometimes difficult to understand well what speakers from different fields say. But the seminar provided me with many ways to find out the solutions, for example, by having meals together, the short excursion, and the game called the escape room. Some open problems I found interesting during the seminar are:

1. mechanism of human cognition that produces, being affected by mis- and dis- information.
2. computational models that reveal justifications of human beliefs about information. It would be a good starting point to accept the fact that no recognition is possible without prejudice, as the "ugly duckling theorem" tells us.

5.3 My Background and Work on Critical Online Reasoning

Dimitri Molerov (Universität Mainz, DE)

I attended the Dagstuhl Seminar "Technologies to Support Critical Thinking in an Age of Misinformation" by recommendation from Prof. Andreas Dengel's office. His collaboration with my supervisor Prof. Olga Zlatkin-Troitschanskaia (economics education (educational assessment) JGU Mainz) on the cross-university initiative Positive Learning in the Age of Information (`https://www.plato.uni-mainz.de/`) had led to two interdisciplinary Springer volumes [1], including contributions by seminar participants. The "Age of Misinformation" phrase from prior presentations may have inspired the seminar's title, too. The focus was on scoping improvements to learning in higher education in the face of increasing self-directed online learning, as well as a need for more evidence-based reasoning in regard to the Internet (Asking not only what you know, but how the discipline found out about it). Two aspects have made it a priority topic in our research group. A) Media use surveys in the initiative showed students' use of online sources for learning surpassed their use of offline resources (e.g., scripts, textbooks) (this was even pre-pandemic)[2]; B) the Internet as an uncurated space for learning inputs (and mostly ignored space for educational research, apart from work on curated e-learning) with all its high- and low-quality information and its preselection and distribution, attention-grabbing, addiction-reinforcing, and polarization mechanisms partly opposing learning preconditions (see HumaneTech). The general idea has been that students as Internet users need a specific skill-set for successfully acquiring reliable knowledge – the umbrella for the collaboration has been on how to model, measure, technologically support and foster necessary skills for students' self-directed learning online.

In our research group, we have meanwhile specified a skill-set for assessment as Critical Online Reasoning [3]. The concept follows known phase models, e.g., Information Problem Solving on the Internet [4, 5], including a search/information acquisition facet, and a critical evaluation facet (modeled as identifying cues to credibility or deficiency in online information), but also specifies a critical reasoning facet (weighing evidence, drawing conclusions) and expanding on the "activation" and monitoring (when do we even apply critical reflection, given that it takes mental effort and we are all cognitive misers). In essence, the additional thinking and behavior one has to undertake to ascertain information quality when one suspects that perceived information and consulted sources may not be entirely dependable. Another framing would be skills for discriminating dependable information from misinformation online. Various assessments exist; a novel approach has been to adapt the Civic Online Reasoning Assessment [6], which features the actual Internet and (sometimes dubious) websites to be searched and evaluated, and thereby affords quick and ecologically valid creation of test item stimuli. My PhD research revolves around the adaptation, modeling, and test item design for such skills and connection to critical thinking skills. It was very heartening to meet colleagues who are systematizing the various approaches to fostering skills.

In this vein, the work of the computer-science and developer community has been complementary. For one, applications are built to implement educational theories and models into learning support tools, and critical thinking online is supported in many other ways, e.g., automatic detection of misinformation in specific media formats. As I mentioned, to detect misinformation, "someone has to do the thinking", the computer or the human user; and our job may be to (re)negotiate the share of each, as the information landscape keeps evolving, and help optimise human-computer interaction. The study on image reverse detection and automatic emotional labeling was an impressive presentation.

I came to Dagstuhl also to gauge interest in the following project, i.e., if someone would like to digitize and gamify labels for (mis)information. The inspiration is John Cook's work on FLICC – who collected arguments by climate deniers, distilled them into common persuasion techniques (e.g., argument patterns, tropes, fallacies) [7], optimised them didactically using graphic icons, and also used the labels in an educational multiple-choice quiz game (Cranky Uncle game) [8]. One next step could be to take such labels to actual social or news media and either offer a computational pre-labeling, or more interestingly, enable users to label their own and/or peer's content. Having assigned "epistemic labels", a user could review their often evasive initial judgments of a piece of content or even single statements made in a chat (e.g., reminding themselves of initial vague irritation) at a later point and come back to reflect on it more thoroughly.

Social uses are envisioned, as well, from learning games to live collaboration on information evaluation, such as in crowd-sourced fact-checking. Epistemic labels (defined in a publicly accessible scheme or library) can go beyond verifiable facts in also highlighting undesirable or baseless persuasion techniques, which are not strictly falsifiable, but still say misleading and would warrant a warning (e.g., a "citation needed"/reflection needed flag from Wikis or a friendly bias reminder). Epistemic labels can be implemented as emojis, but rather than emotional responses, they would represent cognitive judgment snippets. Here, a future design challenge can be to define epistemically grounded, generative labels and set fair rules for their interpretations, e.g., to gamify error culture or allow some space for people to learning democracies (Here, philosophers and logicians have mapped out a good part of the agenda and formalized reasonable discourse and truth conditions. The inquiry into truth conceptions is far from resolved, but a minimal consensus around wanting to be internally consistent and avoid basic fallacies can already be enough for developing tools and making progress in public discourse).

Disinformation campaigns are a severe risk on one end of a spectrum; on the other end, we find censorship, national/ally/block cultural media bubbles – which on a global scale can be as polarizing, as well: exchanging national for cross-national polarization). Equally some participants reflected back to me the apparently not so rare we-know-it-all-and-will-teach-you-the-right-way attitude or trap that we as designers of cognitive training, such as the inoculation approach, can fall into (and which I try to address more thoroughly in the PhD – the short response can be external bias checks, non-domination in education, and stressing user's capacity development as inherent self-interest). As designers, we may possibly ignore our own biographical, cultural, method biases – e.g., shouldn't it seem too one-sided to safeguard against Russian disinformation only, but remain blissfully unaware of own embedding in other national media diets and forget about past, recent (and maybe unknown) present propagandist efforts by governments, militaries, international business conglomerates, one's administration, and unwittingly participating compatriots. As one of four-five schools of critical thinking research – apart from logical (syllogisms, fallacies focus; recently computational argument), psychological (biases, emotions), educational (mix and content focus), and media scientific insights – the Frankfurt school of critical thinking has been strong in the humanities, and, e.g., with spin offs in critical pedagogy, offering criticisms of surrounding societal power structures that shape discourses. An integration into technology support and assessment seems to be still pending, e.g., in the form of a decision aid when to think critically about a range of granularities from the small everyday mental operations to large global systems (where algorithmic biases come up at pain points) to the metacognitive reflection of when not to overthink. Higher critical thinking requirements, involving criticism of self or own culture are difficult to assess within a government-dependent educational system and risky to design for responsibly. Do we need to acknowledge neo-Imperialism, as Noam Chomsky and Ray Dalio will have us, and a consequential imbalance of consumed cultural content, perhaps even embrace it as unavoidable anthropological evolution resulting of a human drive for power or excellence that aggregates, or do we reject domination attempts in the political, and particularly the digital sphere, as artifacts of last-century public administrative personnel that limits current human development? Reaching the big questions has been easy at Dagstuhl. Coming back out with organized and fun research designs was the grittier, but well-scaffolded part of the seminar.

How do we approach our task exactly? Researchers' work might also stretch beyond just picking a favorite approach between creating knowledge for the privileged few and to educate someone who does not know better. If we take a society-wide view, the work can be about strengthening mental capacities within one another (and building socially reinforcing systems), leaving meetings with the best available knowledge and skills within the largest possible N (including accepting a remainder group), and validating whether the individual reasoners' autonomy is preserved and strengthened and they feel their concerns have been truly addressed – which affords them some relaxation and emotional safety to approach more daunting questions of truth in information. As social psychology indicates, motives for misinformation consumption are often not cognitive, but a symptom of social and psychological conditions (e.g., power distance, lacking self-efficacy); however, pathologizing misinformation consumption would take away the individual's autonomy and the opportunity to tap into their resources, as well as blanket their possibly legitimate concerns. Telling non-scientists to "trust the science" ignores the many cases of misinformation in science, the somewhat fewer scandals, the somewhat large paradigm shifts under way in a given set of disciplines at any time, basic research failures, interest, and incentive structures, and the inability of outsiders to quantify the magnitudes. The confusion may seem daunting to resolve, but can be more easily referenced by distinguishing disciplinary/within-method knowledge gain (critical question: am I being methodologically rigorous, avoiding thought traps), from interdisciplinary/cross-method knowledge gain (critical thinking: does my method apply to the problem? How much do I gain from different approaches?).

Not only preventing misinformation intake, supporting reevaluation of contaminated mindware. It is illusory (and possibly limiting) to safeguard users from being exposed to or rejecting any and all encountered online misinformation; some of it will find a way to seep in. Perhaps, the discussion needs to shift to (tolerable) percentages and thresholds of within- and between person misinformation. It seems equally important to admit that users have already been confused from different sources and support them in learning to regularly reevaluate their acquired misconceptions and "contaminated mindware"[9]. Debiasing and debunking are two successful approaches discussed.

Another still undervalued bundle of approaches is highlighting and insisting on positive conversation and evidence standards and strengthening virtuous communication techniques and patterns. How can these be supported technologically needs further discussion? This one has the advantages of refocusing conversation from problems to existing communicative solutions, being less threatening to the ego, and addressing prevalent cultural skepticism in other's intellectual rigor with grounding.

Overall, the visit to Dagstuhl helped me better understand some of the concepts of prior work on interventions against misinformation (e.g., inoculation theory, pre/debunking, debiasing), get a glimpse of what is being done on the technology and legal side, and meet important proponents and seasoned experts (e.g., Stephan Lewandowsky), while fleshing out project ideas. The sense of community-building and not having to tackle huge challenges alone was as nourishing as the Dagstuhl menu, beautiful nature, and the deep calm of the information scientist. The Outing was a special treat – pondering on media consumption habits and intellectual humility – while having your (lack of) misinformation (trivia knowledge) handed back to you. Much appreciated!

At the seminar, our work group agreed that labeling of information quality at several layers was one important goal for future projects, though precise frameworks are still scarce.

My proximate contribution going forward has been to collect and attempt to classify types of dis- and misinformation, together with example cues, and the search process phase they occur in. I aim to provide this in wiki format. Feedback on how to make classes easily machine-referencable will be much appreciated.

References

1 Zlatkin-Troitschanskaia, O. (Ed.). Frontiers and Advances in Positive Learning in the Age of InformaTiOn (PLATO). Cham: Springer International Publishing. 2020

2 Maurer, M., Quiring, O., & Schemer, C. Media Effects on Positive and Negative Learning. in Zlatkin-Troitschanskaia, O., Wittum, G., & Dengel, A. (Eds.). Positive Learning in the Age of Information. Wiesbaden: Springer Fachmedien Wiesbaden. 2018.

3 Molerov D., Zlatkin-Troitschanskaia O., Nagel M., Brückner S., Schmidt S., Shavelson R.J. Assessing University Students' Critical Online Reasoning Ability: A Conceptual and Assessment Framework With Preliminary Evidence. Frontiers in Education. 2020. https://doi.org/10.3389/feduc.2020.577843

4 Brand-Gruwel, S., Wopereis, I., Vermetten, Y. Information problem solving by experts and novices: analysis of a complex cognitive skill. Computers in Human Behavior. 2005. https://doi.org/10.1016/j.chb.2004.10.005
5 Brand-Gruwel, S., Wopereis, I., Walrave, A. A descriptive model of information problem solving while using internet. Computers & Education. 2009. https://doi.org/10.1016/j.compedu.2009.06.004
6 Wineburg, S., McGrew, S. Evaluating information: The cornerstone of civic online reasoning. Stanford History Education Group. 2016. https://apo.org.au/node/70888.
7 Cook, J. Deconstructing climate science denial. Research handbook on communicating climate change, 62-78. 2020. https://doi.org/10.4337/9781789900408.00014
8 Cook, J. Cranky uncle vs. climate change: How to understand and respond to climate science deniers. Citadel Press. 2020.
9 Stanovich, K. E. The comprehensive assessment of rational thinking. Educational Psychologist, 51(1), 23-34. 2016.

5.4 Critical Thinking and Misinformation in Academic Research

Andrew Vargo (Osaka Prefecture University – Sakai, JP)

Most of the research regarding misinformation that is spread online typically focuses on news and fake news (generally of a political or societal nature). While this is certainly an important and interesting topic, my focus is in knowledge-sharing in technical domains. One of the most fruitful aspects of the Seminar was the wide-ranging discussions held about trust, authority, and misinformation in academic research. In an age in which numerous papers use opaque data mining techniques and questionable data science, it is difficult to assign veracity to each research article. Doing so requires both technical (the academic field may vary) and domain expertise (the area on which the data is extracted may require specialized information) from a reader to call into question what is likely true and what is possibly not. The group discussion was very interesting and participants explored their experiences with problematic research areas and claims. It seems that what we think we know is often a product of repetition. If venues publish and promote research that have questionable conclusions based on flawed methodology, this can perpetuate more research using the same flawed techniques. This eventually creates something that we think as objective truth in the field.

This has spurred an interest in investigating the network relationship between published papers we can identify as having questionable methodologies and conclusions and their impact on the wider research community. This hopefully will uncover how deeply misinformation spreads in academic research and allow us to develop critical thinking tools for academics.

Participants

- Chris Coward
University of Washington – Seattle, US
- Henriette Cramer
Spotify – San Francisco, US
- Andreas Dengel
DFKI – Kaiserslautern, DE
- Tilman Dingler
The University of Melbourne, AU
- David Eccles
The University of Melbourne, AU
- Nabeel Gillani
MIT – Cambridge, US
- Koichi Kise
Osaka Prefecture University, JP
- Dimitri Molerov
Universität Mainz, DE
- Albrecht Schmidt
LMU München, DE
- Gautam Kishore Shahi
Universität Duisburg-Essen, DE
- Benjamin Tag
The University of Melbourne, AU
- Roger Taylor
Open Data Partners – London, GB
- Niels van Berkel
Aalborg University, DK
- Andrew Vargo
Osaka Prefecture University – Sakai, JP
- Eva Wolfangel
Stuttgart, DE

Remote Participants

- Susanne Boll
Universität Oldenburg, DE
- Nattapat Boonprakong
The University of Melbourne, AU
- Laurence Devillers
CNRS – Orsay, FR & Sorbonne University – Paris, FR
- Stephan Lewandowsky
University of Bristol, GB
- Philipp Lorenz-Spreen
MPI for Human Development-Berlin, DE
- Emma Spiro
University of Washington – Seattle, US