AI in the age of distrust

Lahmann, H.C.

ETHICS AND ARMED FORCES

CONTROVERSIES IN MILITARY ETHICS AND SECURITY POLICY

ISSUE 01/2024

AI and Autonomy in Weapons: War and Conflict out of Control?

SPECIAL

Military, Man, Machine

10 TH ANNIVERSARY

ETHICS AND ARMED FORCES

zebis

AI IN THE AGE OF DISTRUST

Author: ***Henning Lahmann***

Abstract

"Cognitive warfare", the malign use of information to manipulate target audiences in open and democratic societies, is considered one of the most urgent policy challenges today. The digital transformation has enabled states such as Russia or China to directly influence Western electorate through social media and other digital channels of communication. The development of ground-breaking artificial intelligence tools to generate and disseminate text and synthetic media at great speed is expected to further exacerbate the problem in the near future. At the same time, researchers have started working on concepts for AI-supported "early warning systems" for cognitive warfare, systems that utilise cutting-edge machine learning algorithms to detect, monitor, and even counter disinformation campaigns by adversarial actors. However, as long as extensive scholarship in the cognitive and social sciences produces only scarce evidence about the causal mechanics of misleading information and the degree of risk such conduct actually poses, such interventions threaten to infringe on communicative rights such as freedom of expression and freedom of information without in fact making Western societies more resilient against attempts of adversarial influencing. Without a solid evidentiary basis, the ongoing securitisation and externalisation of the problem of potentially harmful speech online may end up sacrificing the very rights and values such technological countermeasures ostensibly seek to protect.

"Russia just helped swing a European election", a breathless headline on *Foreign Policy* recently alleged.[1] In the article, Berlin-based journalist Paul Hockenos describes how the campaign for the presidential election in Slovakia was marred by a barrage of "anti-Ukrainian and pro-Russian disinformation" that ultimately "could have been the lever that turned the result so dramatically", ending with the win for the pro-Kremlin candidate Peter Pellegrini. In the larger story of Moscow's malign influence over politics in Western societies, the Slovakia episode has been only the latest instance in an ostensible string of campaigns that have been meddling with democratic decision-making processes in Europe and North America since at least 2016, when the shocking results of the UK Brexit referendum and Trump's surprise victory in the US presidential election catalysed a wholly new era of hybrid threats that seek to poison the minds of Western electorates. While the focus so far has been on Russia's efforts to sow informational distrust and to heighten polarization in the open societies of the West, more recently, Beijing has emerged as another potent actor keen on mimicking the apparent success of coordinated information operations against the same targets. According to researchers and government officials, the New York Times reported in early April, China has started to adopt some of the same tactics that Russia used in the build-up to the 2016 election that brought Trump to power in order to influence the upcoming vote in November 2024. Similar warnings of a growing threat from China can be heard in European capitals.

Another part of this worrisome narrative is the oft-repeated clarion call that despite Moscow's and Beijing's demonstrated intent to further exploit Western societies' informational discord, governments in Washington, Brussels, Paris, or Berlin remain woefully oblivious to this ongoing "information war", reluctant to crack down on such behaviour due to misplaced concerns about freedom of speech and information. In this telling, time is running out if we want to preserve our democracies, as things will only further deteriorate once the full poten-

tial of artificial intelligence (AI) to assist in the generation and dissemination of harmful and misleading information, thanks to the rapid development and availability of ever-more powerful large language models (LLMs), will begin to get exploited by those malicious agents from abroad. In light of this gloomy prospect, what indeed is to be done to safeguard democracies from the scourge of cognitive warfare? This brief essay will address this question not by attempting to come up with another set of policy recommendations – as such already abound – but by taking a step back, interrogating the foundational conceptual assumptions that underlie the issue, and questioning whether we have in fact embarked on the right path toward solving the riddle of today's information disorder.

Manipulation through information

The idea to influence and even manipulate public opinion of an adversary's civil society is of course as old as warfare itself. Every military campaign or even just any political dispute among states has been accompanied by attempts to steer attitudes and sentiments abroad in order to weaken the enemy or just to persuade reluctant potential allies. Propaganda, originally a rather neutral concept without principled negative connotations, has always been part of statecraft and thus largely accepted as par for the course. However, with the emergence of electronic means of instant and large-scale communication via online interfaces and especially with the rise of today's dominant social media platforms such as Facebook, X (formerly Twitter), or TikTok, a development that goes hand in hand with traditional mass media's decline in relevance for the formation of public opinion in democratic societies,[2] something is perceived to have shifted.

False or misleading information, whether disseminated intentionally by malicious actors or simply the by-product of people's biases, ignorance, or craving for sensational stories, now spreads through our digitally networked societies at a speed hitherto inconceivable. At least since the British citizenry opted for the end of EU membership and the US electorate voted Donald Trump into the White House in 2016, the newly disrupted landscape of information has come to be framed as one of the most urgent challenges of our time. Indeed, the World Economic Forum Global Risks Perception Survey for 2023-24 lists the issue of "false information" at the very top of global concerns.[3]

False or misleading information now spreads through our digitally networked societies at a speed hitherto inconceivable

Of course, the problem goes far beyond the manipulation of electoral processes; during the height of the Covid-19 pandemic, the rapid spread of health-related disinformation allegedly led citizens to ignore public health guidelines such as lockdown orders or mask mandates, resort to useless or even harmful remedies like ingesting hydroxychloroquine or bleach, or stay away from vaccination campaigns. More recently, a marked decrease in public support for Ukraine's defensive efforts against Russia's aggression in European societies has been attributed to a concerted campaign controlled and steered by Moscow.[4]

Aside from more overarching notions like "disinformation" or "computational propaganda", the new realities of our information disorder as part of a larger, emerging conflict between Western, democratic societies on the one hand and a growing number of authoritarian states on the other have been captured with terms like "hybrid" or "cognitive warfare". The former describes the ostensibly novel strategy – chiefly employed by Russia but increasingly also by China – to combine traditional, kinetic means of military conduct with "unconventional instruments of power and tools of subversion … to exploit the vulnerabilities of an antagonist and achieve synergistic effects".[5] More specifically, the latter has been defined as a non-kinetic form of warfare that relies on novel information and communication technologies and exhibits key features such as the "targeting of entire populations (…), its focus on changing a population's behaviour by way

of changing its way of thinking rather than merely by the provision of discrete bits of false information in respect to specific issues (…), its reliance on increasingly sophisticated psychological techniques of manipulation (…), and its aim of destabilising institutions, especially governments, albeit often indirectly by way of initially destabilising epistemic institutions, such as news media organisations and universities".[6]

Outside of the strictly militarised realm of discourse, the External Action Service of the European Union has tried to appreciate the distinctly adversarial, strategic, and external dimensions of today's information disorder with the concept of "Foreign Information Manipulation & Interference [FIMI]", which it defines as "a pattern of behaviour that threatens or has the potential to negatively impact values, procedures and political processes. Such activity is manipulative in character, conducted in an intentional and coordinated manner. Actors of such activity can be state or non-state actors, including their proxies inside and outside of their own territory".[7] What these concepts have in common is an understanding of information as effectively weaponised, exhibiting capacities to undermine an adversary in ways practically impossible prior to the digital transformation; the means of generation and dissemination of such manipulative information are not epiphenomenal but an essential feature and indeed a necessary condition for this age of information disorder in the open and democratic societies in the West.

The emphasis on the technological side of the equation somewhat downplays the actual substance of cognitive manipulation

The dawn of AI-enabled cognitive warfare

In light of this emphasis on the technological side of the equation that somewhat downplays the actual substance of cognitive manipulation – be it misleading information about a political candidate, false narratives about the threat posed by a novel virus, or the motives and capabilities of the Ukrainian armed forces – it becomes apparent why the development and today wide availability of LLMs have triggered a wave of renewed concern regarding the future of societal cohesion in democratic states. Based on the principle of deep learning as one of the cutting-edge variations of machine learning, LLMs are algorithms trained on massive amounts of textual data from the internet and other sources in order to recognise and interpret natural human language by learning the statistical relationships between words or "tokens" and then to gain the ability to generate human-sounding, meaningful text based on an input by consecutively predicting the most probable sequence of words or "tokens". Mainly thanks to vastly increased computing power and ever-larger datasets derived from corpora extracted from online sources, the latest generation of such LLMs, for example ChatGPT 4 by OpenAI, Gemini by Google, or Claude by Anthropic, demonstrate an impressive array of capabilities to respond to prompts with well-phrased and reasonably complex output.

Due to the additional fact that current LLMs are able to generate such content at striking speed, some experts expect them to soon be deployed as increasingly effective tools for manipulation in a way that further exacerbates open societies' predicament in this age of informational discord. Calling it nothing less than an "existential threat", Bradley Honigberg noted in 2022 that "AI-generated synthetic media and convincing AI-enhanced chatbots now offer threat actors a growing array of persuasive, tailored, and difficult-to-detect messaging capabilities".[8] Apart from textual disinformation produced by LLMs, another set of generative AI tools can create "deepfakes", synthetic visual or audio-visual media that can be used to make persons such as celebrities or politicians seem to say things or act in ways that do not correspond to reality in order to manipulate audiences into believing false information about and lose trust in the persons so portrayed. Allegedly, the candidate for

prime minister of Slovakia's liberal Progressive party, Michal Šimečka, already fell victim to such tactics ahead of the country's parliamentary elections in September 2023, an incident that just like the meddling in the presidential election mentioned above was allegedly initiated by Moscow.[9]

Experiments with the latest generation of LLMs have demonstrated that these are capable of generating persuasive content that is context-aware, matches the tone and conversational style of particular target audiences and mimics the hallmarks of legacy media, making their output seemingly authentic and increasingly difficult to distinguish from credible journalism. In a recent review study, three researchers demonstrated how the LLM ChatGPT can be used to create and disseminate a highly detailed and believable piece of disinformation at every stage of the disinformation lifecycle: from prompt generation and the making up of false information based on the prompt through refinement, packaging, the creation of seemingly legitimate social media accounts for the purpose of dissemination, the development of a dissemination strategy, all the way to the dissemination itself, fake engagement, amplification, and even subsequent adaptation of the narrative, if necessary.[10]

If this theoretical exploration turns out to be not only correct but also feasible in practice, it will mean that LLMs are in fact highly vulnerable to be manipulated into acting as the perfect instruments for further ramping up disinformation campaigns. In that sense, it seems indeed prudent to assume that we are currently at yet another inflection point. Combined with the capabilities of machine learning models to analyse vast amounts of social media traffic to identify emerging trends that can be exploited to further spread desired messaging and to target specific sub-groups that seem more amenable to certain manipulative narratives, the emerging possibilities for carrying out a strategy of cognitive warfare appear to paint a rather gloomy picture for the future of democracy and our open societies.

Combating disinformation campaigns with AI

So, what is to be done? Apart from recent legislative efforts on the level of the European Union to conceive of the scourge of disinformation as a challenge for the regulation of online platforms, with accompanying initiatives to further strengthen the resilience of political will-formation and democratic decision-making processes within the Union,[11] in 2015 the European External Action Service established the "East Stratcom Task Force", whose principal project is the website and database "EUvsDisinfo", which is entirely devoted to exposing and counter-narrating Russian attempts to influence public opinion in Europe by means of information operations. Similarly, in 2014 NATO set up its Strategic Communications Centre of Excellence in Riga, Latvia; less focused on real-time debunking of false Russian narratives, it is nonetheless primarily focused on studying and developing strategies to counter Moscow's brand of cognitive warfare.

In a recent review study, three researchers demonstrated how ChatGPT can be used to create and disseminate a highly detailed and believable piece of disinformation at every stage of the disinformation lifecycle

With the proliferation of AI technologies and faced with the prospect of an ever-widening landscape of informational threats, however, the focus has more recently turned to the idea of utilising machine learning principles to detect and counter cognitive warfare campaigns.

Based on the premise that a "proper defence requires at the very least an awareness that a cognitive warfare campaign is underway" as well as "the ability to observe and orient before decision-makers can decide to act", a team of students in the NATO Review in 2021 described a quite influential proposal of a "cognitive warfare monitoring and alert system".[12] Such a tool is envisioned to work with machine learning models that would be able not only to detect any such adversarial activities across social networks and online media outlets by way of

identifying suspicious patterns of conduct but furthermore to autonomously track and monitor their progress. In doing so, such a system could automatically generate continuously updated reports for further processing that form the basis of measures to counter the malicious conduct. The idea is that such real-time monitoring is necessary to shape appropriate responses that are capable of preventing any negative effects such campaigns might otherwise cause.

The use of AI to counter cognitive warfare does not have to end with detection and monitoring, however. One step further, there have been considerations to use the power of machine learning to assist human fact-checkers in analysing and flagging individual pieces of content more quickly.[13] It has even been suggested that such systems could be directly involved in counteraction, for example by deleting content that is being disseminated as part of a disinformation campaign. Indeed, in theory at least, LLMs could possibly be utilised to generate information or whole narratives that seek to retort a campaign on the substantive level, directly refuting the misleading information with a factually correct counterpart that spreads through the same channels and targeted at the same, previously identified audiences.

It is obvious that there is a categorical difference between a private actor enacting algorithmic policies in regard to its own products, and a state actor doing the same on a much larger scale and not limited to a singular channel

Most of these ideas are not exactly new. For quite a while already, the leading social media platforms have been deploying machine learning algorithms to detect and counter such conduct on their services. For example, Meta – the company that owns Facebook, Instagram, WhatsApp, and Threads – has systems in place to automatically uncover what its policies call "coordinated inauthentic behavior (CIB)", defined as "coordinated efforts to manipulate public debate for a strategic goal where fake accounts are central to the operation"; the identification of such accounts usually leads to their instant removal. Other platform providers such as TikTok or YouTube have similar mechanisms in place. Further, predictive machine learning tools are today widely used by social media companies for the purpose of algorithmic content moderation, which means that these systems parse all content posted to the respective platforms and automatically delete any items that are deemed to contradict the provider's terms and conditions, which frequently comprises disinformation at least if it is considered potentially harmful.[14]

What distinguishes such efforts by individual digital service companies from the described ideas to let a state or supra-national entity like the EU or NATO deploy AI-supported systems to counter cognitive warfare campaigns is that the latter would by definition enact their measures across the different platforms and websites. It is precisely the point of such an algorithmic tool to be able to monitor data traffic along digital networks writ large, most likely extending to the websites of media outlets, so that it can detect and counter multi-platform campaigns that seek to exploit a variety of communication channels in order to exert influence on Western audiences, as the most sophisticated and complex of today's adversarial campaigns already do. However, it is obvious that there is a categorical difference between a private actor enacting such algorithmic policies in regard to its own products on the one hand, and a state actor doing the same on a much larger scale and not limited to a singular channel, on the other. This categorical difference cuts across the legal, ethical, and political dimensions.

The mechanics of disinformation

One of the most critical issues in the context of formulating any type of response to the current information disorder is that virtually all of these policies work on the basis of a set of assumptions about the mechanics of misleading information that do not hold up to scientific scrutiny. The theories that implicitly or explicitly inform the design of countermeasures

seem to assume a very straightforward causal relationship between the dissemination of information that seeks to mislead its audience and some negative outcome. But the problem with this assumption is that such clear causal link is yet to be proven. Despite an already vast and quickly growing mountain of studies in the cognitive and social sciences about the mechanics of misleading information and other varieties of cognitive warfare, we still have very little concrete evidence that such adversarial efforts have any effects on the recipients' behaviour at all;[15] if anything, a good amount of research points to the opposite.[16]

What we can prove is that adversarial actors, mainly Russia and China, have certainly been trying to influence public opinion in Western countries by way of tactics associated with cognitive warfare, and continue to do so. But as researchers already noted with regard to the election of Trump in 2016, "evidence of sustained effort is not the same as evidence of impact or prevalence".[17] This is not because empirical research in this area is exceedingly complex, but rather because common assumptions about what type of information actually manages to influence audience behaviour may largely be wrong. Rather than being gullible recipients of misleading information, ready to act on the false beliefs that are formed after exposure, Hannah Arendt's observation still by and large holds up: "What convinces masses are not facts, and not even invented facts, but only the consistency of the system of which they are presumably part."[18] Indeed, cognitive science research confirms that a recipient can likely be influenced only in the case that a new piece of information – whatever its truth value – connects to convictions that are already part of their belief system.[19]

Crucially, this picture is not expected to change with the increasing employment of LLMs to create and disseminate false and misleading information. Even if the use of such machine learning algorithms leads to an increase in the quantity, quality, and personalisation of false and misleading information, it has recently been shown that "existing research suggests at best modest effects of generative AI on the misinformation landscape", so that growing concerns as to the dangers of these new technologies for the information ecosystems in democratic countries are "overblown".[20] In light of this, frequent journalistic accounts of how Moscow has just determined the outcome of yet another election in a Western country with a concerted influence campaign, as cited at the outset, are misguided at best and dangerously alarmist at worst.

The perils of overreacting

That we get the science of the mechanics of manipulating information right matters because any intervention on the state level incurs considerable costs for fundamental rights, in particular if it is done with the help of AI. For one, an algorithmic early warning system for cognitive warfare campaigns has obvious privacy and data protection implications. This is because a model as envisaged can only function if it is fed with vast amounts of input data both for training purposes and during actual deployment. As its main goal is the detection of unusual activities in the informational realm in digital networks, such data will by definition comprise large quantities of personal data of the users of social media platforms and other websites.

Over the past years, we have already seen what it means to let states decide what type of online content should be considered "disinformation" and thus taken down

Perhaps even more significantly, allowing a state or other authoritative entity to make distinctions between "good" and "bad" content online is potentially very detrimental for individual communication rights such as freedom of expression and freedom of information. Over the past years, we have already seen what it means to let states decide what type of online content should be considered "disinformation" and thus taken down. This issue goes beyond the challenges of evidence and of finding a definition of the concept that

is sufficiently precise so as to not be in conflict with rule-of-law considerations. Especially after the beginning of the Covid-19 global health crisis, a number of authoritarian-minded governments exploited the emerging discourses surrounding health-related disinformation to crack down on free speech.[21] These concerns should be taken seriously even if the algorithmic system is limited to detecting and monitoring possible adversarial campaigns; after all, once the algorithmic output predicts the existence of such conduct, pressure will be high to act against it by deleting associated accounts or information, whether manually or by means of another algorithm that automatically tackles such content. Obviously, the latter variant implicates communication rights even more severely. Even if adversaries in Moscow and Beijing try their utmost to influence citizens in Western democracies, it must be emphasised that such attempts do not override the rights to freedom of expression and information. It is the hallmark of a functioning liberal democracy to let citizens make their own choices as to the sources of information and even concerning the narratives they want to believe in, and for this reason it is generally not up to governments to paternalistically intervene in the process of political will-formation. Projected onto the collective dimension, such AI-supported informational interventionism ultimately implicates a people's right to self-determination.

It is the hallmark of a functioning liberal democracy to let citizens make their own choices as to the sources of information and even concerning the narratives they want to believe in

This note of caution points to a larger consideration. The political discourse in the West surrounding "disinformation", "cognitive warfare", or "foreign information manipulation and interference" implicitly proceeds from the assumption that our current "age of distrust" is both principally a security concern and a threat coming from abroad. But it is such securitisation and externalisation of the problem that has led to the search for solutions in the language of the military in the first place. A number of scholars have shown that many of the most hazardous false political narratives emerge and spread domestically, with outside actors merely exploiting and amplifying them.[22] If that is true, however, any real solution must primarily address the processes of increasing societal polarisation at home instead of seeking to disrupt the malicious machinations of some external actor. And finally, an "early warning system" is certainly desirable when it comes to the detection of incoming missiles or other adversarial armed activities. Whether it is an appropriate approach for the dealing with potentially problematic speech is much less clear.

The digital transformation, and especially the rise of social media, has been accompanied by an increasing fragmentation of the information ecosystems in the open societies of Western democracies, a development that adversarial states such as Russia and China have been trying to take advantage of for their own geopolitical ends. The emergence and expected ubiquity of LLMs may further contribute to an exacerbation of the situation, although their actual impact could ultimately turn out to be less grave than commonly feared. Most importantly, we should be cautious when formulating policy responses to the perceived information disorder, in particular as a possible use of algorithmic systems as an antidote may prove to be the cure that is worse than the disease. Instead, policymakers are well advised to focus on long-term strategies that might seem less technologically appealing but that would have the advantage of avoiding negative impacts on the very rights and values the fight against the information disorder seeks to protect.

The Author

***Henning Lahmann** is Assistant Professor at the Center for Law and Digital Technologies at Leiden University Law School. His research focuses on the intersection of digital technologies and international law. In 2021/22, Henning was a Hauser Post-Doctoral Global Fellow at NYU School of Law, with support from a research grant by the German Academic Exchange Service. He holds a doctoral degree in international law from the University of Potsdam, Germany.*

1 Hockenos, Paul (2024): Russia Just Helped Swing a European Election. Foreign Policy. https://foreignpolicy.com/2024/04/17/slovakia-president-pellegrini-russia-election-interference-disinformation/ (all internet references accessed April 30, 2024).
2 See e.g. Newman, Nic (2023): Young People Are Abandoning News Websites – New Research Reveals Scale of Challenge to Media. The Conversation. https://theconversation.com/young-people-are-abandoning-news-websites-new-research-reveals-scale-of-challenge-to-media-207659; Lipka, Michael and Shearer, Elisa (2023): Audiences Are Declining for Traditional News Media in the U.S. – With Some Exceptions. Pew Research Center. https://www.pewresearch.org/short-reads/2023/11/28/audiences-are-declining-for-traditional-news-media-in-the-us-with-some-exceptions/.
3 World Economic Forum (2024): The Global Risks Report 2024, p. 18. https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2024.pdf.
4 Digital Forensics Research Lab (2024): Undermining Ukraine: How Russia Widened Its Global Information War in 2023. https://www.atlanticcouncil.org/in-depth-research-reports/report/undermining-ukraine-how-russia-widened-its-global-information-war-in-2023/.
5 Bilal, Arsalan (2021): Hybrid Warfare – New Threats, Complexity, and 'Trust' as the Antidote. NATO Review. https://www.nato.int/docu/review/articles/2021/11/30/hybrid-warfare-new-threats-complexity-and-trust-as-the-antidote/index.html.
6 Miller, Seumas (2023): Cognitive Warfare: An Ethical Analysis. In: Ethics and Information Technology, 25, p. 1.
7 European External Action Service (2021): Tackling Disinformation, Foreign Information Manipulation & Interference. https://www.eeas.europa.eu/eeas/tackling-disinformation-foreign-information-manipulation-interference_en.
8 Honigberg, Bradley (2022): The Existential Threat of AI-Enhanced Disinformation Operations. Just Security. https://www.justsecurity.org/82246/the-existential-threat-of-ai-enhanced-disinformation-operations/.
9 Meaker, Morgan (2023): Slovakia's Election Deepfakes Show AI Is a Danger to Democracy. Wired. https://www.wired.com/story/slovakias-election-deepfakes-show-ai-is-a-danger-to-democracy/.
10 Barman, Dipto, Guo, Ziyi and Conlan, Owen (2024): The Dark Side of Language Models: Exploring the Potential of LLMs in Multimedia Disinformation Generation and Dissemination. In: Machine Learning with Applications 16, art. 100545. https://www.sciencedirect.com/science/article/pii/S2666827024000215.
11 European Commission (2024): Guidelines for Providers of VLOPs and VLOSEs on the Mitigation of Systemic Risks for Electoral Processes. https://digital-strategy.ec.europa.eu/en/library/guidelines-providers-vlops-and-vloses-mitigation-systemic-risks-electoral-processes.
12 Johns Hopkins University and Imperial College London (2021): Countering Cognitive Warfare: Awareness and Resilience. NATO Review. https://www.nato.int/docu/review/articles/2021/05/20/countering-cognitive-warfare-awareness-and-resilience/index.html.
13 Bateman, Jon and Jackson, Dean (2024): Countering Disinformation Effectively: An Evidence-Based Policy Guide. Carnegie Endowment for International Peace, p. 87.
14 Gorwa, Robert, Binns, Reuben and Katzenbach, Christian (2020): Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance. In: Big Data & Society 7 (1). https://journals.sagepub.com/doi/epub/10.1177/2053951719897945.
15 See only Bateman, Jon et al. (2021): Measuring the Effects of Influence Operations: Key Findings and Gaps from Empirical Research. Carnegie Endowment for International Peace; Maschmeyer, Lennart et al. (2023): Donetsk Don't Tell – 'Hybrid War' in Ukraine and the Limits of Social Media Influence Operations. In: Journal of Information Technology & Politics. https://www.tandfonline.com/doi/epdf/10.1080/19331681.2023.2211969?needAccess=true.
16 Mercier, Hugo and Altay, Sacha (2022): Do Cultural Misbeliefs Cause Costly Behavior? In: Sommer, Joseph et al. (eds.): The Cognitive Science of Belief: A Multidisciplinary Approach. Cambridge, pp. 193-208.
17 Benkler, Yochai, Faris, Robert and Roberts, Hal (2018): Network Propaganda. Oxford, p. 254.
18 Arendt, Hannah (1958): The Origins of Totalitarianism. Cleveland and New York, p. 352.
19 Jowett, Garth S. and O'Donnell, Victoria (2012): Propaganda and Persuasion. 5th ed. Los Angeles et al., p. 34.
20 Simon, Felix M., Altay, Sacha and Mercier, Hugo (2023): Misinformation Reloaded? Fears About the Impact of Generative AI on Misinformation Are Overblown. In: Harvard Kennedy School Misinformation Review 4, p. 3. https://misinforeview.hks.harvard.edu/wp-content/uploads/2023/10/simon_generative_AI_fears_20231018.pdf.
21 See Kaye, David (2020): Disease Pandemics and the Freedom of Opinion and Expression: Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression. United Nations. https://documents.un.org/doc/undoc/gen/g20/097/82/pdf/g2009782.pdf?token=vxfXRfLUEqLTf8Syoi&fe=true.
22 Benkler, Yochai, Faris, Robert and Roberts, Hal (2018), see endnote 17.