upgrade
democracy

September 2024

Two years of Upgrade Democracy

7 chapters on our findings

BertelsmannStiftung

Contents

1. Two years of Upgrade Democracy: What we have done and what we have learned

By Cathleen Berger and Kai Unzicker

Click here for the article:

Our project started with a question: Would the super election year of 2024 also be a ‘super year’ for disinformation? In 2024, around half of the world’s population was called to the polls, including people in India, Indonesia, the European Union, South Africa and the U.S. Not only are we witnessing many crucial elections, but public discourse and sentiment were and remain tense due to wars, economic crises, and the climate crisis — a window of opportunity for manipulation attempts. Our own representative survey confirms: The population is concerned. 84 % of Germans see disinformation as a major problem for our society and 81 % fear its influence on democracy. It was therefore obvious that we needed to meet these concerns with specific and targeted solutions while also strengthening societal resilience.

We addressed a wide range of topics relating to both, the spread and countermeasures against disinformation with the help pf various impulses, publications, and practical projects. It’s time to condense and bundle the results, experiences, and findings from two years of Upgrade Democracy.

This is why, we are launching a six-part series of articles today, in which we present our key observations. Primarily geared towards political decision-makers, we point to the critical levers where action is needed. The time of merely observing is over. Healthy, resilient digital discourse must be actively shaped. This requires more political attention, more recognition for civil society networks, and more reliable funding.

What the following articles will address

Our team went on an investigation: In a large-scale, international research endeavour and with valuable support from regional research partners, we mapped trends, actors, and patterns in the (dis)information ecosystem and analysed similarities and differences between the regions. In their article, Cathleen and Charlotte summarise their key observations and **call on political decision-makers to pay closer attention and integrate international perspectives into their own actions.**

To the article →

We take hope from the more than 230 approaches, ideas, and solutions that we were able to identify in all parts of the world and that are all working towards resilient democracies on a daily basis. As inspiring as the **variety of initiatives and the methods they use,** they need to be **developed in a more sustainable and targeted manner,** as Joachim explains in his article.

To the article →

Methodologically, countermeasures should not be limited to fact-checking, but **media and democracy literacy must rather be considered together.** Approaches such as "trust-checking" can provide a useful impetus here, as Julia describes in her article.

To the article →

Notably in digital spaces, it is important that proposals are backed by reliable data, so that platform responsibility does not remain a mere demand but is effectively implemented. To achieve this, research on digital discourses must be expanded. In her article, Cathleen therefore calls for the **establishment of a central point of contact for independent, continuous research on digital platforms.**

To the article →

However, the issue of platform responsibility cannot just be about the dominant platforms. We must also ask ourselves how **decentralised alternatives can be strengthened and better positioned for healthier, democratic discourses – including from political actors.** Charlotte provides ideas for this in her article.

To the article →

Yet, it is ultimately also important how we want to shape the digital public socially and politically. **What are our visions for a healthy and productive public sphere that supports democracy and contributes to mutual understanding?** Based on current trends and with a view to different future scenarios, Kai sheds light on what it takes to shape a healthy digital public sphere.

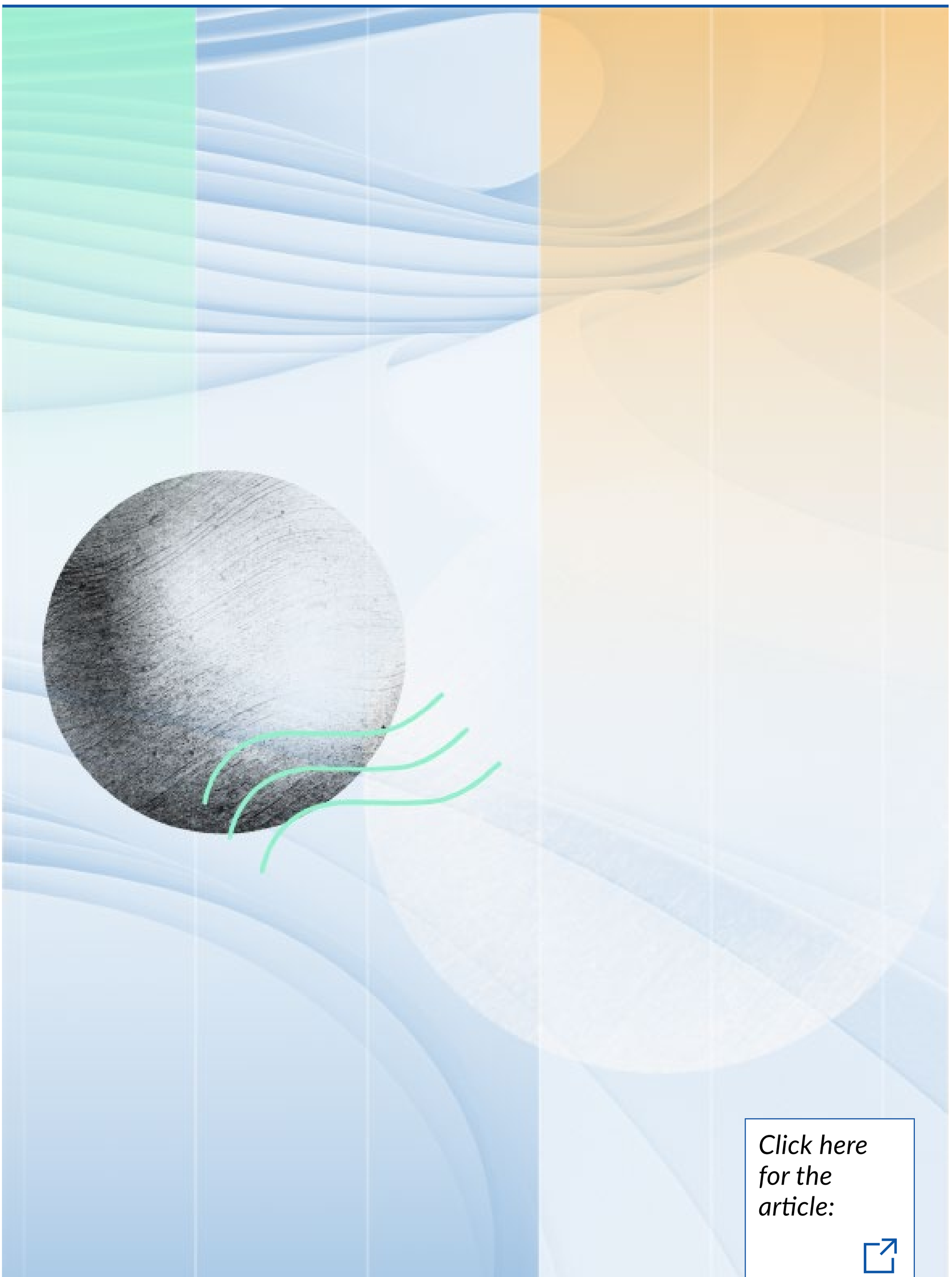
To the article →

Political action is required

We have learned a lot. And: There is still a lot to do. Above all, political action is required. We must not just observe the problem; we must find solutions. There is ***no one*** solution and ***no one*** player that can solve everything. Rather, a network of protagonists must work together against malign influence and for a democratic public. These networks must be established, maintained, and financially supported. To counter the disinformation industry, we need a resilient ecosystem that exemplifies and promotes democratic “controversy”. For governments and political decision-makers, this means that the debate about potential restrictions and the protection of freedom of expression must be conducted openly and honestly. International experiences are an important corrective here, as is a strong, well-connected civil society.

2. If you want to successfully counter disinformation, you need a global network

By Cathleen Berger and Charlotte Freihse



Click here for the article:

Disinformation does not stop at national borders—counter-measures must therefore be just as international and interlinked. Based on comprehensive, international research, we identify key challenges in dealing with disinformation and call on political decision-makers to provide greater support for resilient civil society networks.

Our international research and the analyses by our regional research partners have vividly shown that the actors spreading disinformation are becoming increasingly professionalised, technically savvy and networked. A disinformation industry has emerged that is fuelled not only by known actors but also by offers from PR and marketing agencies. It recruits vulnerable, economically weak parts of the population as "keyboard warriors" for their dishonest purposes.

If agitators and aggressors become more professional, the protagonists who oppose these campaigns must be even better organised, coordinated, and resilient.

There is no shortage of ideas and potential anywhere in the world, only the structures and existing resources need to be adapted to the digital reality and the speed associated with it. The response to the disinformation industry must be an ecosystem of protagonists and strategies that acts in a cross-sectoral, coordinated, and global manner as no single actor or measure alone can successfully work towards healthy digital discourse.

What is: Topics and the spread of disinformation

Analyses from all parts of the world show that disinformation prepares the ground for its influence over a long period of time by emotionalising socially controversial topics or distracting people with side issues. A direct attack on the integrity of elections, the trustworthiness of democratic institutions, or the credibility of individual candidates is often just the last drop in a slowly rising ocean. In this respect, elections can act as a catalyst and be the target of disinformation, but countermeasures are not only needed prior to elections but on an ongoing basis. A wide range of countermeasures, approaches, and protagonists must be combined to thwart disinformation campaigns: Prebunking, monitoring, demonetisation, debunking, regulation, and more. The toolbox is and must be versatile (see also Joachim's article → p. 14).

On the one hand, the digital spaces in which disinformation spreads are globally connected and are largely based on large, private-sector platforms such as YouTube, TikTok, Instagram or WhatsApp. On the other hand, usage patterns and preferences vary

considerably between individual countries and regions: LINE is available almost exclusively in Asia, TikTok is growing particularly rapidly in Europe, WhatsApp dominates in Africa, and the picture is mixed in Latin America. While digital publics overlap, disinformation campaigns use differing channels. Research documents that there are gaping holes in the responses of platforms, which interpret regulatory provisions as narrowly as possible while also applying their own rules vaguely, e. g. in the form of "copy & paste" procedures for varying contexts — especially in countries that are not considered lucrative markets from a platform perspective.

What we observe: Data, capacities, technological developments

Our understanding of the spread and influence of disinformation is based on the continuous monitoring of patterns, actors, and attempts to influence discourse on digital platforms. Access to data for research purposes could hardly be more crucial for developing evidence-based proposals and countermeasures. However, there are glaring gaps in the reliability, comparability, and analysis of data and platforms, especially with regard to non-European research that is not covered by the Digital Services Act (for more details, see Cathleen's post → p. 33).

The strength and resilience of civil society organisations is vital for the success of countermeasures and the promotion of healthy digital public discourse. The range of tasks for civil society protagonists is growing

worldwide — their expertise is in demand when it comes to regulation and platform oversight, they act as fact-checkers, offer trainings for media and digital literacy, monitor digital discourse, educate, bring people together, and fill gaps wherever they come to light. At the same time, their scope for action is shrinking due to dwindling resources, political repression, strategic lawsuits, targeted attacks and more, which are putting an enormous strain on already hard-pressed civil society protagonists worldwide.

Technological changes, such as artificial intelligence (AI), have become an integral part of the digital public. AI has also become a regular companion in election campaigns — not only for the manipulative purposes of disinformation but also as a tool in the campaigns of political candidates. Existing supervisory structures that ensure the transparency and fairness of political advertising need to be upgraded in many places to provide adequate responses to these new technological developments.

What needs to be done now: Politicians must listen to international perspectives and support networking

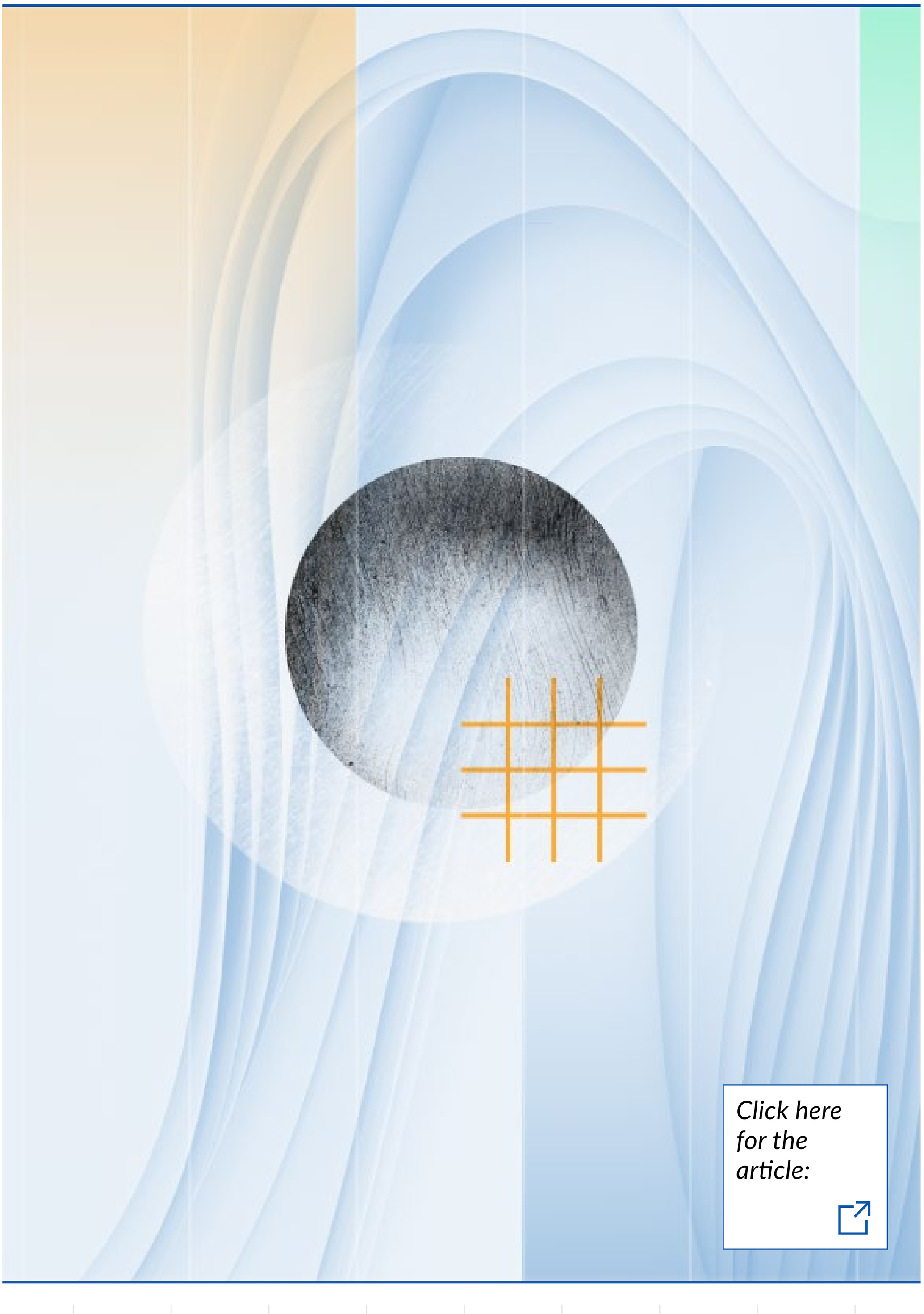
For the field of protagonists worldwide to professionalise as successfully as the disinformation industry, political decision-makers must emphasise the fundamental value of networking and cooperation formats. And fund them.

Maintaining and activating networks is time-consuming and labour-intensive, which must be reflected in the funding provided and in public recognition received. For example, philanthropy and democratic governments must provide long-term support and build on existing successes instead of constantly “chasing” innovations and the latest technological trends in their requests for proposals. Not only do we need diverse, international perspectives to make smart policy decisions for a healthy digital public sphere. The resilience of our democracies also depends on the resilience of civil society engagement. All over the world and in interaction with each other.

Click here for the seven-part report series with in-depth analyses.

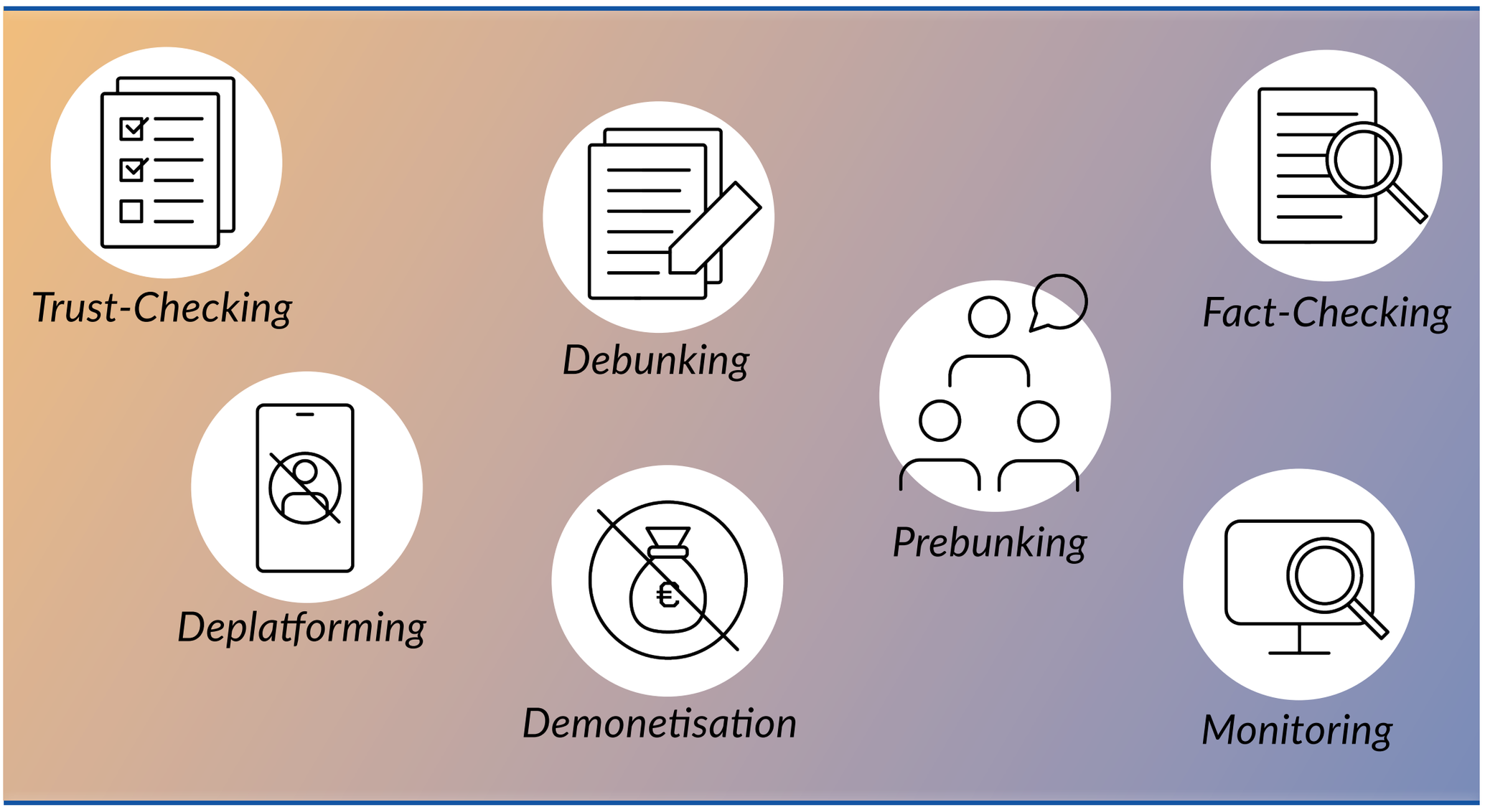
3. Prebunking or fact-checking? What matters is a comprehensive approach

By Joachim Rother



Click here for the article:

As diverse as disinformation strategies are, so are the methods to counter them. However, a look across the world reveals: In actual practice, one can hardly talk about a variety of methods. This needs to change.



Disinformation from Veles: A small town gains fame

2016. Shortly before the U.S. presidential elections. Donald Trump is a few months away from becoming U.S. President. Far away from that, in the small town of Veles in North Macedonia, some teenagers experiment with websites, filling them with random headlines copied from major media outlets, and realise: The articles are generating clicks. And quite a lot of them. The model catches on. The websites become more numerous and professional. Some now appear to resemble legitimate news outlets. Five to ten articles per website are published every day, and although most of the pro-Trump articles make little sense or contain no truth, many of them spread like wildfire. From the middle

of nowhere in Europe, public opinion in the U.S. is getting influenced, and some of these teenagers in the economically-struggling North Macedonia suddenly earn money: Between August and November 2016, over 16,000 USD through Google AdSense payouts. It is only when The Guardian and Buzzfeed publish investigations revealing that at least 100 websites registered in the small town in North Macedonia are churning out disinformation about the U.S. elections, that Google demonetises the websites. The advertising revenue dries up, and the operators lose interest.

Anyone who thinks of Russia, China, or Iran when it comes to disinformation campaigns will be surprised by the monetary motives of the Veles example, as the motives for creating and spreading disinformation vary greatly. Whether targeted political influence or purely economic interest, countermeasures must consider the mechanisms and context of specific disinformation efforts to be effective.

The dilemma of choice: Which method is the right one?

The toolbox of countermeasures to mitigate disinformation is versatile. What is striking, however, is that most methods only address disinformation when it is already out and difficult to rein in, including fact-checking or debunking.

Prebunking, on the other hand, attempts to prepare people for disinformation or specific misleading narratives before they even encounter them. The goal is to build

resilience through sensitisation, thereby undermining the impact of disinformation. How such prevention against disinformation can work technically is demonstrated by Google subsidiary Jigsaw with its video campaigns: Video snippets that address specific disinformation and warn against it are played as so-called pre-rolls before the actual content. The problem: Prebunking is labour-intensive, must be tailored to specific topics of disinformation, and its effectiveness is limited. According to one study, the proportion of people who could recognise manipulative content after watching a prebunked video increased by an average of 5 percentage points.

In contrast to prebunking, debunking focuses on correcting disinformation once it has already been published. Unlike fact-checking, the strength of debunking lies in placing content and sources within a larger context and identifying patterns through which disinformation is spread in major sectors such as climate or gender. Debunking is practised very successfully by numerous projects worldwide, such as AltNews (India), Mafindo (Indonesia), or Africa Check (South Africa). Typically, these corrections are published and disseminated in comprehensive counterstatements after extensive research. However, this also highlights the challenges of this method: Debunking is labour-intensive and time-consuming, and by the time the counterstatement is published, the original false information is usually several days old. This is problematic because studies show that false information on social media

generate 90 % of their engagement on the first day — far too quickly for debunking to keep up.

Fact-checks, however, can have a much quicker impact, often taking only a few hours and requiring much less time. In this method, statements or reports are verified for their truthfulness and evaluated through confirmation, correction, or rejection. Fact-checks promote accountability among public figures and encourage verifying the truthfulness of information before it is published or shared. Fact-checks are conducted according to journalistic standards, such as those defined by the International Fact-Checking Network of the Poynter Institute or the European Fact-Checking Standards Network.

Since Donald Trump's first candidacy in 2016, fact-checking has become the most widely used method globally in the fight against disinformation. In 2023, the Fact-Checking Census by the Duke Reporters' Lab counted over 400 institutions that are active in fact-checking in about 69 languages across more than 100 countries. Our own international research, based on desktop research, expert interviews, and workshops on five continents, also highlights the dominance of fact-checking as a method. Of more than 230 registered initiatives, more than half are involved in fact-checking to some extent.

Despite these good examples from around the world, the effectiveness of fact-checking is subject of great debate. Apart from the mental strain on fact-checkers,

the sheer volume of disinformation is too great, the fact-checks themselves are too slow, and their measurable impact is too limited. And another problem rattles the fact-checking method: Disinformation actors are hijacking the tool and simply publishing their own "fact-checks." By taking advantage of the trust-building effect of fact-checking, political polarisation can apparently be spread much more easily, as shown by the case of CheckYourFact.com, a right-wing conservative fact-checking outlet of former Fox News host Tucker Carlson.

Follow the money!

So, are the resources exhausted? Are we powerless against the flood of disinformation? There is hope, as a look at the international landscape of protagonists reveals gaps worth examining more closely. The aforementioned global mapping of anti-disinformation initiatives, with over 200 entries, lists only four organisations (Check My Ads Institute, Global Disinformation Index, Konspirátori, Sleeping Giants Brazil) that use demonetisation as a primary tool in the fight against disinformation.

This is noteworthy because demonetisation fundamentally differs from all the basic concepts mentioned so far, as it targets the incentive that often leads to the spread of disinformation in the first place: Economic interest. When social media accounts or websites are identified as sources of disinformation, platforms or hosts can cut off their funding by, for example, drying up advertising revenue through Google AdSense.

Why this can be a sensible measure is shown by a look at the numbers: The NGO Global Disinformation Index analysed 20,000 domains spreading disinformation in a study and found that ad tech companies had placed ads worth 235 million USD on these sites.

In a system where success is measured by clicks and page views, disinformation content can be monetised relatively easily, as a recent collaborative article by the Centre for Media Pluralism and Media Freedom (CMPF) and the European Digital Media Observatory (EDMO) highlighted. This type of content is quick and cheap to produce and is often prioritised by platforms because it generates high reach under the guise of free speech through emotionalisation, controversy, clickbait, or decontextualisation. While this is not illegal, platforms can decide to take action and penalise the respective accounts—such as by withholding their advertising revenue. However, this is not without problems, especially when advertising revenue is blocked without transparent reasons. As early as 2018, an article in the SZ concluded that "demonetisation [...] would be the bogeyman among professional YouTubers," as the lack of transparency in such measures could suddenly deprive entire livelihoods. Despite vehement demands from the EU Commission's Vice-President Věra Jourová to enforce demonetisation measures on platforms, the major platforms are still reluctant to consistently implement demonetisation measures due to ongoing criticism, not least from their content creators.

Demonetisation takes many forms: Public pressure is needed

Some are not satisfied with this. Organisations such as the Global Disinformation Index (UK), Sleeping Giants (Brazil), or Konspiratori (Slovakia) evaluate websites or accounts with high reach for their trustworthiness and the reliability of the information provided. If accounts or websites are suspected of spreading disinformation, this is made public, and advertisers or the platforms themselves are urged to stop placing ads there or to block the revenue accordingly. After all, most brands want to avoid being discredited by dubious advertising partners.

The fact that this method of demonetisation through public pressure can be a sharp sword is impressively demonstrated by the example of the teenagers from Veles mentioned at the beginning: When the money dries up, it is often no longer worth maintaining the channel. Unlike many other methods, demonetisation thus goes beyond merely treating the symptoms and, if successful, can tackle a key driver of disinformation at its root and bring it to a halt on the respective channel. However, scientific research on the method of demonetisation is still in its early stages, and beyond anecdotal evidence, such as the successful work of Sleeping Giants USA against Breitbart News, no reliable statements can currently be made about the medium- or even long-term effectiveness of the method.

There is no Swiss army knife against disinformation

As shown, the toolbox against disinformation does not

offer a single tool that, even in a customised application, can counter the entire spectrum of disinformation. The good news is: It doesn't have to. Because complementarity is key.

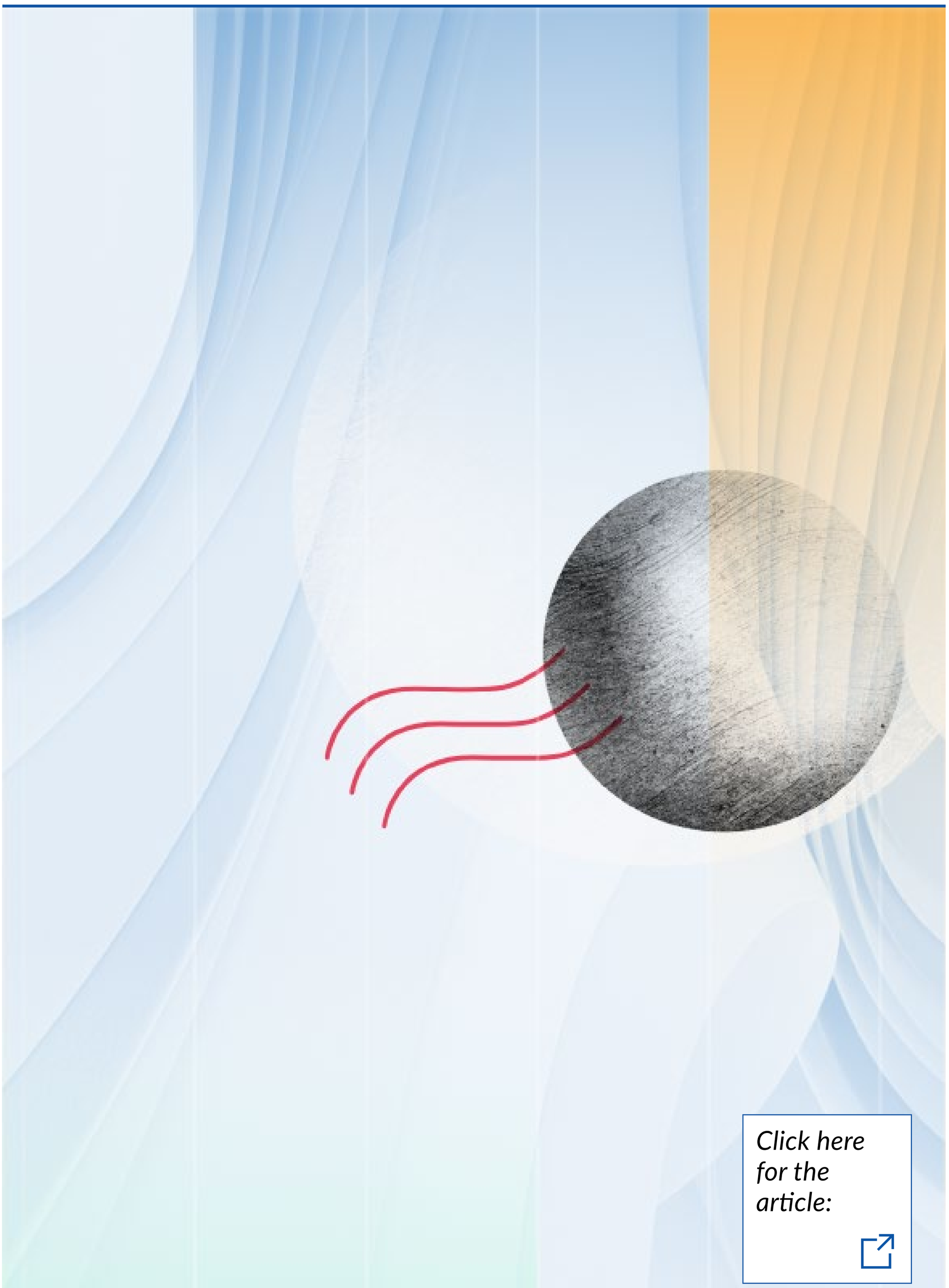
An information ecosystem that wants to successfully defend itself against disinformation in the long term must rely on a plurality of methods and a meaningful interplay of different mechanisms. More is not necessarily better, as the case of fact-checking shows. Rather, methods should be coordinated so that disinformation in its different phases — both before it is created (media literacy, prebunking) and after it is spread (debunking, fact-checking, or, to some extent, demonetisation) — has as difficult of a time as possible to unleash its destructive effects.

Our international research shows that the options here are not yet exhausted. Demonetisation emerges as a strategy in the international comparison that, despite its potentially significant impact, has so far received relatively little attention and is therefore underutilised. It seems that this strategy, due to its approach, is capable of severely undermining the profitability of disinformation. Demonetisation addresses a gap in the current practice of most anti-disinformation strategies and should be much more widely applied in the future than it has been.

Click here for an overview of the global initiatives:

4. News literacy: Strong against disinformation

By Julia Tegeler



Click here for the article:

To curb disinformation, the news literacy of media users is crucial. Strengthening it among the population is challenging. Low-threshold and everyday educational programmes – such as those we tested in the "faktenstark" pilot project – and closer collaboration among stakeholders present promising approaches.

News literacy: Why it's important and where it falls short

How well people can assess the quality and reliability of digital information influences what content they trust and on what basis they form an opinion. Anyone who wants to curb the spread of disinformation should therefore also focus on citizens' digital news and information literacy. It comprises a bundle of skills that enable participation in democratic digital public spheres: Those who are news literate understand what role digital public spheres play for democracy and how they function. They can recognise their own information needs, research digital information, and assess its

relevance, reliability, and intention. Furthermore, they can engage with the content of the information, process it, comment on it, share it. In Germany and other European countries, the level of news literacy in the population is rather low and varies greatly depending on age, level of education, and media usage behaviour. Political attitudes and values also influence how news are consumed and evaluated. A study by the Stiftung Neue Verantwortung states that, for far too long, citizens have been left alone to find their own way in increasingly complex media environments. Although there are many initiatives and promising approaches worldwide that promote digital news literacy, their impact is too often limited. There are several reasons for this:

- **Hard-to-reach target groups:** Certain target groups — such as adults in general, elderly people, people with little education — are difficult to reach. For this reason, many programmes focus on children and young people, while adults are still often neglected.

- **Limited resources:** Educational work suffers from scarce resources, precarious working conditions, and limited funding. This makes it difficult to sustainably establish proven concepts and limits their widespread application.

- **Higher demands due to digital public spheres:** Digital public spheres place higher demands on the media literacy of individuals because user-

generated content is displacing professionally prepared and verified information. Information is often no longer shared and categorised by journalists. Instead, each media user must decide for themselves which information is trustworthy and which they want to believe, consume, or share.

- **Inadequate concepts:** Many educational programmes still fall short conceptually. Teaching skills for checking and evaluating information is important. However, more is needed for citizens to be able to counteract the democratically threatening influences of disinformation: In addition to strategies for self-help, technological, social, and cognitive skills as well as democratic competence are needed.

Four fields of action to strengthen news literacy

To tackle these challenges, initiatives to strengthen digital news and information literacy should focus more strongly on four fields of action.

1. Create a variety of offers that are relevant to everyday life to reach as many people as possible

To reach as many people as possible, a wide range of educational offers are needed: From low-threshold information events and educational workshops for a broad target group to digital street work or in-depth training for specific groups. The offers should convey content that is relevant to everyday life and meet people where they are. Particularly hard-to-reach target groups such as the poorly educated, the elderly, or

people with low media confidence can be reached in places they regularly visit, such as community centres, their workplaces, sports and other outfits, and leisure facilities. Cooperation with social institutions, employers, local initiatives, and public libraries is crucial for reaching people in their daily lives. In our pilot project, called faktenstark, we held workshops on dealing with disinformation at various locations. We went to vocational schools, volunteer training days, sports and other outfits, public libraries, trade unions, charities, and even breweries. This allowed us to reach different target groups: From trainees and volunteers to employees of civil society organisations, local politicians, and pensioners. We used a modular concept to adapt the workshops in terms of time and content. Experience has shown that such low-threshold educational offers can reach and bolster the confidence of many people who have fundamental trust in democracy but wonder what they can do to counter disinformation on the internet or in their own environment.

To reach adults in general, corresponding educational offers should also be integrated into further vocational training. The Business Council for Democracy's programme is a successful example of this. For work with children and young people and the target group of educational professionals, there are many good materials, and further training offers from initiatives such as klicksafe, klickwinkel, weitklick as well as from our two faktenstark cooperation partners codetekt and Amadeu Antonio Foundation

2. Pool resources to improve sustainability and reach

Collaboration can be an answer to limited resources: Numerous national and international initiatives and projects pursue similar goals and have developed effective approaches. And yet they often work in parallel – partly because there is not enough time in everyday working life to exchange ideas and initiate collaborations. However, it is precisely through closer cooperation that strengths can be pooled, and synergies created. Successful approaches could be combined, scaled up, and integrated more sustainably into existing structures. Through a coordinated approach, educational institutions, civil society organisations and initiatives, media companies, and political decision-makers can each contribute their strengths. The aim should be to join forces and take a coordinated approach to disseminate effective initiatives to strengthen news literacy and anchor them sustainably in existing structures.

At faktenstark, for example, three different cooperation partners have joined forces. This enabled us to use technical know-how to develop digital tools such as our Trust-O-Mat and our chat bot Klaro, as well as media education expertise, and comprehensive expertise in political education work. We were also able to draw on various networks for the communication of the project and the implementation of workshops.

3. Teach self-help strategies to strengthen skills in dealing with disinformation

Media users should be taught strategies with which

they can assess the trustworthiness of information themselves in everyday life and recognise manipulative content. Approaches such as trust-checking and pre-bunking are promising here. In our faktenstark-work-shops, we therefore worked with both approaches to strengthen participants' ability to deal with disinforma-tion. Trust-checking enables media users to assess the credibility of information quickly and as objectively as possible using journalistic quality criteria. Aspects such as sources, content, medium, quotes, and visual repre-sentations are checked. Pre-bunking aims to educate people about how disinformation works. The idea is to "vaccinate" media users against disinformation and thus build up immunity even before they encounter specific cases.

To this end, pre-bunking explains typical manipulative techniques and misleading narratives so that people can then recognise when information is based on such narratives or contains certain rhetorical patterns. How effective pre-bunking is is still under discussion. Ulti-mately, it is also advisable to use a variety of methods in educational work to deal with disinformation and to teach different strategies. This is not only the task of educational institutions or civil society initiatives. The media can also make an important contribution by ex-plaining journalistic quality criteria and showing how they research, check, and process information. Collab-orative approaches in which players from the media, education and research work together, such as the "UseTheNews" initiative to promote news literacy in the digital age, are particularly promising.

4. Combine news and democracy literacy to strengthen real resilience

Teaching media users how to apply strategies for checking and assessing information is an important first step. However, this alone is not enough to mitigate the influence of disinformation that threatens democracy. Educational programmes should not only teach skills, but also encourage people to engage with the topic of "disinformation" and create an understanding of the disinformation ecosystem. It is important to understand the structures, goals, strategies, and modes of action of disinformation and to shed light on the reasons why many people believe it. This is also a central aspect of educational work in our pilot project faktenstark: In the workshops, we not only teach specific strategies for dealing with disinformation – such as trust-checking – but also take a broader view of disinformation – especially regarding democracy.

Another aspect that is important in educational work against disinformation is the promotion of basic democratic attitudes and skills. This includes a willingness to inform oneself about politics, an appreciation of freedom of opinion, freedom of the press, and journalism, basic trust in democracy, and tolerance towards other opinions. Those who regularly inform themselves about political developments can recognise disinformation more easily. Those who value freedom of expression recognise the importance of diverse information and are more likely to engage with different perspectives. This helps to avoid jumping to conclusions and to counteract confirmation bias.

In a nutshell, promoting news literacy alone ultimately falls short. It is about strengthening democratic competence. This requires continuous and integrated educational work and is a task for all of us in society. In addition to educational institutions, politics, the media, and businesses are also called upon here. Politicians should not only conduct objective, fact-based, and fair debates, but also create good framework conditions for political education. In terms of corporate democratic responsibility, companies can facilitate participation and promote programmes that support critical thinking and social engagement. Media can provide information about journalistic formats ("What is a commentary?", "What characterises a factual article?") or explain journalistic working methods (How is news collected, checked, and presented?). This transparency can strengthen trust in the media and make it clear how important quality journalism is for an informed society and that it offers more reliable information than other sources.

Conclusion

Information and news literacy in the digital society are crucial for mitigating the spread of disinformation and strengthening democratic discourse. The current challenges require more effort than before: Relatable and diverse educational programmes, more collaboration, the teaching of self-help strategies, and integrated approaches that understand news literacy as a component of political education are promising here.

More information about our pilot project can be found here:

5. Platform accountability through independent research: What is, what is missing, what is next

By Cathleen Berger

Click here for the article:

The key to successfully mitigating disinformation could lie in moving from anecdotally analysing attacks on digital discourse to do so in a continuous, data-driven manner. Sounds simple, but faces numerous hurdles in. Going forward, we need a hub for knowledge and data management to fill serious gaps.

Role and limits of monitoring social media platforms

In January 2024, the Federal Foreign Office investigated a large-scale disinformation campaign by Russia. The strategic communication team identified over 50,000 bot accounts that were automatically spreading false information about the war in Ukraine and mutually reinforcing each other. Since 2022, the EU DisinfoLab, Correctiv and others have been uncovering new cases of the so-called 'doppelganger' campaign, in which content from established, far-reaching newspapers such as Der Spiegel, Zeit Online, Le Monde and others is copied and then replaced with individual,

manipulated content and disinformation. Even if the URLs of the pages vary, the deceptions are often not recognisable at first glance, so that the supposedly 'serious' but false articles are sometimes distributed widely on platforms and in messengers.

Such attacks are detected by monitoring digital discourse, which involves checking posts, interactions and trends for anomalies. Without this monitoring, numerous attacks on our discourse would remain undetected – a major threat to the quality of and trust in our information ecosystem. In recent years, there has been a visible increase in the number of civil society organisations that use or want to use monitoring to propose more concrete changes to platforms and support the mitigation of threats. The demands for better, more reliable and easier access to platform data for independent research are as diverse as the respective contexts: Be it CeMAS, the Institute for Strategic Dialogue, the Mozilla Foundation, Democracy Reporting International, Aos Fatos, Soch Fact Pakistan, Media Monitoring Africa or the Coalition for Independent Tech Research. There are now dozens, if not hundreds, of organisations worldwide that integrate monitoring into their work and help to better understand and defuse disinformation campaigns.

Successes such as those mentioned above are heartening – and yet they are anecdotes. Anti-disinformation successes identify individual attacks, can reveal bot armies, and networks or sensitise people to patterns in disinformation narratives. However, they

remain selective and are structurally limited in their reach and speed. There are several reasons for this.

Platform Governance: Private actors with public responsibility

One reason for this lies in the tension between the public sphere and private sector platforms. In the digital public sphere, more and more of our societal discourse is taking place on social media platforms, which, as private actors, have a disproportionate influence on our lived realities. Of course, this also entails a high level of responsibility — which has been repeatedly and emphatically emphasised and demanded by numerous experts in recent years. Where platforms initially based their rules for content moderation on internal business guidelines, today many such decisions are either legally pre-structured, supervised, or critically assessed from the outside.

The pressure to act responsibly has increased noticeably. At the same time, the imbalance of power persists: Despite new requirements and increasing regulation in many countries around the world, our insight into the functional logic and available data of platforms remains limited. This is another reason why the importance and necessity of independent research and data analysis is a highly relevant topic that has been the subject of intense debate, including political debate, at least since the negotiations surrounding the Digital Services Act (DSA).

According to the DSA, disinformation is considered a 'systemic risk' to democracy. The DSA obliges the dominant platforms, such as TikTok, YouTube, Facebook, and LinkedIn, to take decisive action to counter the risks. Platforms are required to prevent the spread of disinformation and to design and curate their services in such a way that risks can be minimised, violations can be tracked, and countermeasures can be evaluated. Civil society organisations and academics play a central role in the implementation of these obligations. They act as a corrective, an early warning system, and a source of inspiration for how to foster healthy digital discourse. To fulfil these functions, civil society organisations need access to platform data. This is regulated in Article 40 of the DSA – the clearest framework to date for data access for research purposes on platforms.

Data access for research purposes: Data protection, applicability, coordination

Despite the supposedly clear legal framework, inconsistencies and uncertainties emerge in practice. Monitoring is based on data that the platforms already collect about their users, information flows, interactions, or the effect of their design choices. Such data can be used, for example, to visualise networks, track the virality of individual posts, and trace patterns of 'comment bots' and pseudo-accounts.

However, a lot of platform data is also sensitive as it contains personal information, individual preferences, or direct messages between individuals. Data protection

experts express justified concerns if too much data is collected, stored, or analysed. Strict guidelines for independent research are therefore necessary to prevent intrusions into the privacy of individuals and disproportionate surveillance of our digital discourse. Awareness of data protection and the sensitivity to the impact of monitoring varies, especially among civil society organisations, and often depends a great deal on the context, national legal traditions, and the urgency or pressure that civil society faces in different countries around the world. Due to a lack of resources and/or limited capacity, data protection and ethics advice is not always institutionalised in civil society organisations and individual analysts must often assess and decide sensitive issues on their own.

There are also legal gaps when it comes to mandatory access to platform data. For example, journalists, non-affiliated researchers, and research consortia with non-European partners cannot refer directly to the DSA. Researchers outside the EU therefore often resort to commercial marketing tools or web scraping to conduct their analyses. This is legally vague and analytically limited, both with regard to the comparability of results and the possibilities of filtering data and preparing it in a methodologically sound manner.

In addition, our research and exchanges with experts show that every organisation and every research network develop its questions, the research design, and, in case of doubt, the respective code for the collection, evaluation, and analysis of platform data from

scratch. This means that existing knowledge is rarely built upon, and each organisation sets up their respective monitoring effort independently. This presents a chicken-and-egg problem: On the one hand, the capacities and competencies of civil society organisations are too limited to conduct long-term, legally compliant, and interlinked research based on platform data, so that they concentrate on smaller, anecdotal monitoring projects. On the other hand, their resources and leverage are limited not least by the fact that they do not build on existing knowledge and complement each other in their work because they lack the time and capacity to network internationally and continuously and keep up to date.

Platforms interpret the obligation to allow access data differently

A further hurdle arises from the lack of comparability between the platforms and their way of implementing the various legal provisions. Each platform provider implements existing legislation in their own interests and context.

For example, Meta's platforms, Facebook and Instagram, were accessible to researchers for years via a tool called 'CrowdTangle'. Organisations from all over the world could apply to Meta for access to the tool. For around two years, applications were no longer possible, and Meta recently announced that the tool would be shut down. A new access point set up to comply with the implementation of the DSA limits access to data from the last three months and has been

criticised by first-time users for its limited options. Although current developments and major campaigns can be monitored, historical or regional comparisons are not possible. The clear criticism from civil society regarding the shutdown of CrowdTangle underlines the high relevance of independent, reliable monitoring for their work, but also shows that many of them are overwhelmed by learning and setting up new access points with the necessary speed and urgency due to a lack of resources and skills.

TikTok and YouTube, which for a long time were only accessible to researchers through workarounds, e. g. web scraping or data donation methods, are also gradually setting up access points for research purposes on the basis of the DSA. However, both are limited in their respective ways. For example, TikTok only allows access for researchers from the U.S. and the EU, a maximum of 10 researchers are allowed to join together in a network, and initial experience suggests that access is not reliable and is faulty in places. Although YouTube makes its data research programme available worldwide, the sheer volume of data and the daily growth are so large that long-term, comparative studies usually do not have sufficient server capacity for evaluation, which is why anecdotal research is particularly present here. The costs for server and storage capacities for long-term research are often prohibitive for civil society organisations anyway, not only, but especially for YouTube.

As each platform defines its own access, research across multiple platforms is extremely difficult and has not yet been standardised or harmonised. Analyses of the spread of narratives and manipulation tactics across platforms and between different networks are therefore virtually impossible. Entertainment on TikTok, shopping on Instagram, news on X or Threads – user behaviour is diverse and we can only understand the long-term effects of manipulation attempts if we research them.

Gaps that need to be filled

On the plus side, we know much more about the emergence of disinformation on platforms today than we did five years ago. And yet, existing gaps urgently need to be filled if we are to move beyond anecdotal knowledge to truly measurable, evidence-based successes in dealing with disinformation.

These challenges can be summarised as follows:

1. There is a lack of standardised research access to platform data. As platforms are privately organised, each one sets its own rules. This hinders research across multiple platforms. In addition, real-time analyses across different regions are currently limited or not possible independently.

2. Journalists, independent researchers, and especially non-European research are currently neglected, which limits the categorisation and assessment of systemic risks on social media platforms.

3. The capacities and competences of civil society are currently not yet sufficient to produce long-term analyses and thus make evidence-based design proposals for platforms. A lack of data protection and ethics supervision, too few training programmes, a lack of coordination and further development of existing research, limited access to comparative data, and the cost of sufficient server and storage capacity significantly limit the impact of independent monitoring.

The hurdles are great, but not insurmountable. In view of the high relevance of independent research and data analysis for our digital public, things can and must change quickly.

Outlook: A central hub for knowledge and data management

Many wheels need to mesh here: Legal improvements, pressure on platforms to assume their public responsibility through harmonised research access, and a strengthening and increase in civil society organisations in order to fulfil their role as a corrective and initiator. Political decision-makers, technology companies, and philanthropic organisations are called upon.

However, these solutions also require greater coordination between researchers and opportunities to share significantly more knowledge in order to shape digital discourse in a sustainable and trustworthy manner. Awareness of the need and the will alone will not be enough here. Instead, we need a central hub for

knowledge and data management for independent research on digital discourses. Such a (service) organisation would have to act as a hub on three levels:

1. as a source of knowledge that provides, among other things, templates for legal and ethical issues relating to research questions, data collection, evaluation, and storage;

2. as a data manager that prepares methods and approaches for monitoring and, as a scientific fiduciary, provides cleaned, pre-coded data on shared server capacities for research purposes; and

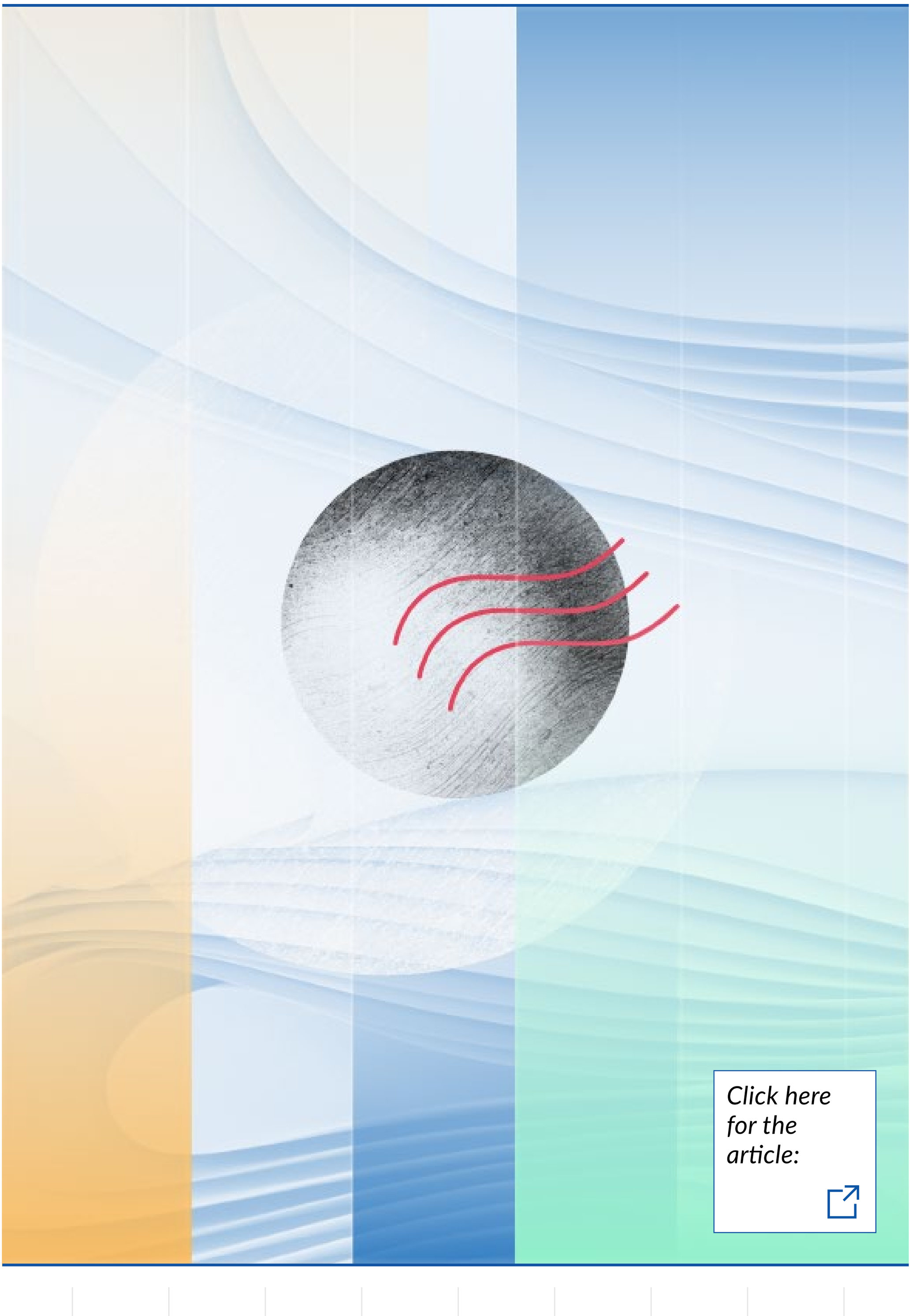
3. as a spokesperson that collects the experiences of monitoring organisations from all over the world and represents them in a coordinated form to platform providers and political decision-makers to call for future improvements.

With the Data Knowledge Hub for researching digital discourse, we have launched an initial pilot for a knowledge database. The further development and networking of this concept must and will occupy us in the future.

The Data Knowledge Hub can be found here:

6. Democracy by design: How decentralised alternatives can contribute to a better social media ecosystem

By Charlotte Freihse



Click here for the article:

In today's digital discourse landscape, privately-owned social media platforms dominate—and are increasingly subject to regulation. However, more is needed for sustainable change: the democratising potential of decentralised networks should be given more attention.

A healthy digital discourse is a key component of opinion formation, exchange, and information in democratic societies. Currently, however, this discourse takes place in privatised spaces, with the power over rules and conditions concentrated in the hands of a few, profit-oriented companies. Most of the regulation in this area focuses on the power of dominant platforms like Instagram, TikTok, and others. This is important because respective legislation—especially the EU's Digital Services Act (DSA)—increases transparency and introduces a range of new control mechanisms. However, there is no democratisation of platforms that more directly incorporates the interests of citizens. For that, we need other ideas.

One frequently discussed idea is the establishment of platform councils to promote more participation and inclusivity in shaping the digital space. This means involving individuals outside the company to better integrate and oversee fundamental rights and values on platforms (cf. Berger et al.; Kettemann et al.; Riedel; Pietron and Haas). Meta's Oversight Board is a first implementation example and certainly a step in the right direction, while many other platforms are still hesitant to introduce comparable governance mechanisms. However, the Meta Oversight Board also highlights the difficulty of integrating "democratic" elements into a privately-owned company: The board has limited power as it can only make non-binding recommendations. The same applies to its resources: It can only handle a few cases while millions of content moderation decisions are pending (notably during the super election year or with regard to non-English content), limiting its overall impact. In addition, there is a lack of transparency, especially regarding algorithms, which hampers long-term systemic changes. Although platform councils are a good idea, they face other challenges in practice: There is a risk that government regulatory agencies will be weakened, and responsibilities diversified. Furthermore, implementing democratic feedback mechanisms requires significant resources, including creating incentives for less privileged groups to promote inclusive participation (for more information, see our analysis conducted with other experts).

So, what now? Is that all, or how could a digital space with less monopolisation and power imbalances look? And what does that mean for users?

Users currently have little room for action and decision-making in the social media ecosystem: They have limited platform choices, inadequate protection of personal data, and a lack of transparency about the architecture and design of the platforms they use. Regulation can only partially solve this. What could a more democratic social media ecosystem look like for users? And what would be necessary for users to fully realise this potential?

Here are four ideas from decentralised platforms that could contribute to the democratisation of the social media ecosystem:

- **Data sovereignty and privacy:** Control over personal data is a central aspect of decentralised platforms. For example, many instances in the Fediverse do not allow third-party tracking, meaning users can regain control over their personal data depending on the instance they choose. This contrasts with centralised platforms that collect vast amounts of user data and often reuse it without user consent. Decentralised examples like Mastodon and Diaspora show that this strengthens user sovereignty and reduces the risk of mass surveillance and data misuse. However, users must understand the importance of this and know where and how to manage their data, as there are also

decentralised alternatives that allow data collection and analysis for advertising purposes, such as Meta's decentralised platform Threads (see Berger and Freihse). More power thus also requires more responsibility and competence. For decentralised platforms, this form of data sovereignty means they need new financing models. One possibility is a subscription model, which requires users to be willing to pay with money instead of their data.

- **Independence of decentralised instances:** Decentralised platforms like Mastodon allow users and/or groups to operate their own servers, minimising the potential for centralised control and censorship by a single organisation. This is not only interesting for individual users but also for media professionals and other institutions that want to contribute information to the discourse. However, to fully exploit this potential, significant resources are needed – both in terms of server costs and resources for content moderation, as well as the technical knowledge to set up and manage these servers. To prevent only well-equipped actors from entering the field, support and guidance are needed for marginalised actors with few resources. Financing models for decentralised platforms should take this into account.

- **Freedom of choice and diversification through interoperability:** Open protocols foster innovation and competition, as users are less dependent on changes and conditions on platforms. This can

encourage the development of new features and services that are better tailored to users' needs. PeerTube, for example, allows users to run their own video hosting instances, creating a diversified landscape of video content not restricted by the monopoly of a single platform like YouTube. Interoperability allows users to leave platforms more easily and gives smaller platforms and servers a better chance of being used. This leads to more diversity and plurality. However, for verified news content — especially from traditional media — this could mean greater difficulties in reaching users and being widely visible. Therefore, it is necessary to consider how to support quality information, especially from established media, on decentralised platforms.

- **Co-determination and self-governance:** In decentralised networks, users often have more say in the design and management of the platform. Participatory governance models allow users to influence rules, moderation policies, and technical developments. Platforms like Mastodon, which use participatory governance models, or concepts like an algorithm marketplace on the decentralised platform BlueSky, create a more democratic and user-centred environment (see Freihse and Sieker). This form of participation could lead to greater platform loyalty from users in an interoperable ecosystem. However, clear and accessible governance structures and the engagement and participation of users in decision-making processes are essential.

What needs to be done now

Our research over the past two years has repeatedly highlighted great potential: Those who demand resilience and healthy digital discourse in challenging democratic times must take action and assume responsibility. One lever to do this is to develop and implement approaches for redesigning our social media ecosystem. Regulation of existing platforms is no longer sufficient — it is time to promote decentralised alternatives to counter the tendencies of concentration in the platform economy. For policy-makers, this means investing money and actively promoting these developments by specifically supporting decentralised alternatives and creating conditions that enable the construction of a more democratic and diverse social media ecosystem.

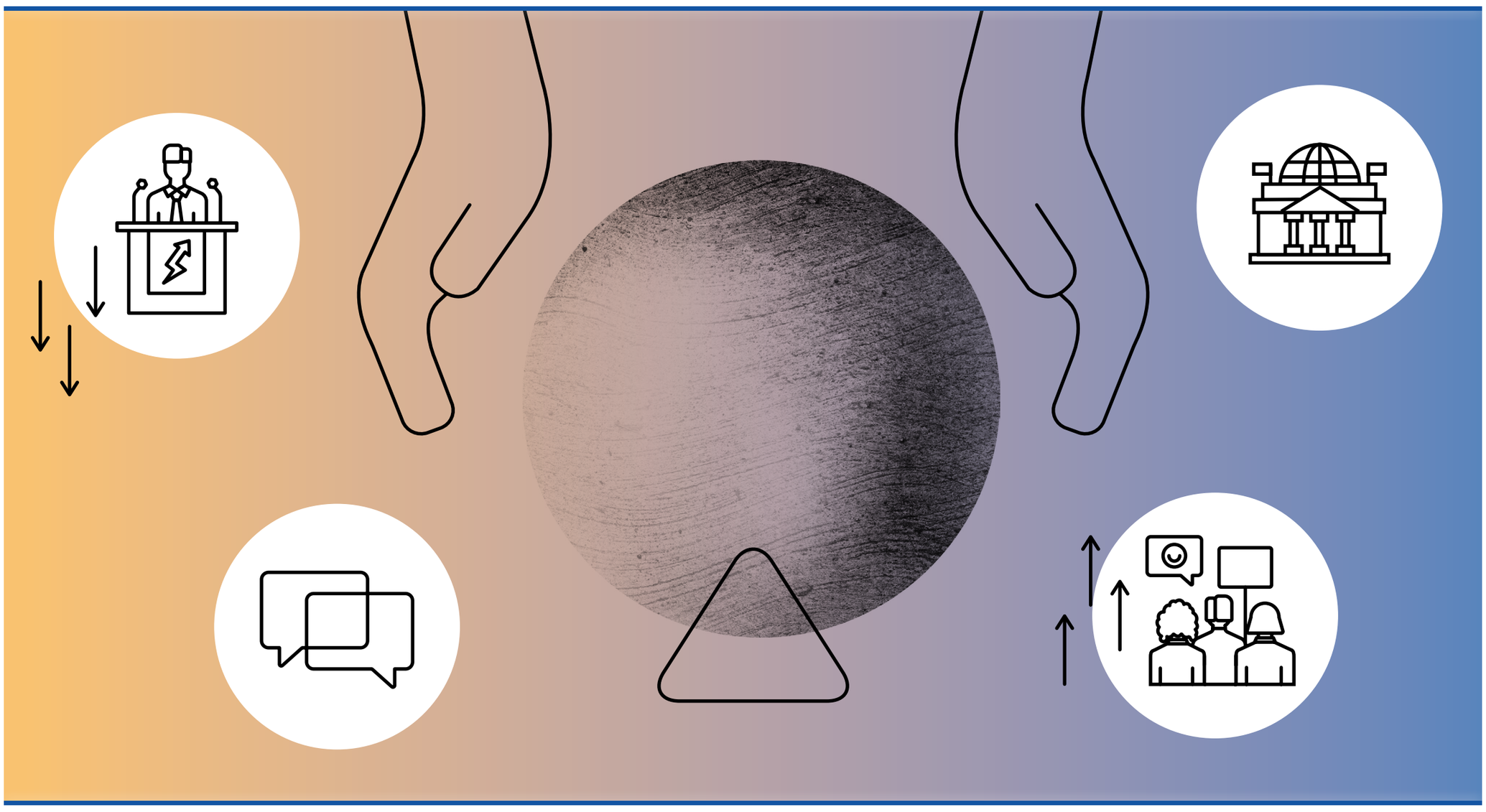
Click here for our impulse series:

7. The utopia(s) of digital discourse spaces

By Kai Unzicker

Click here for the article:

The digital public sphere was a promise. It was meant to bring about exchange, connection, and understanding. After more than two decades of social media and amid an AI hype, the initial optimism has given way to disappointment: Digital discourse spaces are now flooded with polarisation, hate, and misinformation. But what can we do to heal them? After two years of the Upgrade Democracy initiative, we present pathways towards healthier digital discourse spaces.



The initial hope was immense: Global connectivity and access to information promised a new era of reason and understanding. And indeed, digital communication has changed the world and repeatedly demonstrated its potential to fulfil these hopes. Would the Arab Spring, #MeToo, #MashaAmini, or FridaysForFuture have ever taken shape in this form without a global digital public sphere, with social media and mes˚aging services? How quickly could we access the world's knowledge without the hundreds of thousands of volunteers who

contribute to Wikipedia daily — free, multilingual, and in constant critical reflection? How many exams or DIY repairs would fail without support from countless YouTube videos? Undoubtedly, the digitised world has its bright side.

While the 20th century was marked by mass societies, where mass media (few-to-many) wielded great political influence, the 21st century has ushered in an era of individualised, decentralised online communication, where anyone can be both a sender and a receiver (many-to-many). This radical democratisation of the public sphere has made previously marginalised groups more visible and empowered. Yet, it has also led to a radicalisation of democratic discourse. When everyone has a voice and amplifiers, it rarely results in harmonious choir; more often, it is a cacophony of different tunes and tones. The immediacy of social media communication encourages emotions and distortions, which gain far more traction than in traditional, editorially curated media.

Digital communication and the digital public sphere, especially as embodied by profit-driven social media platforms operating on the logic of the attention economy, increasingly appear dysfunctional and even threatening to democracy. But must this be the case? What might a realistic vision of a functional digital public sphere look like — one that connects people, fosters constructive discourse, and promotes understanding rather than division? Clearly, this cannot be about realising the utopia of a rational, power-free, and civilised

digital discourse. The past two decades have shown how challenging that is. However, what would be a desirable improvement over the current state, achievable with the actual actors involved (companies, political institutions, users)? A thorough problem analysis is necessary to identify which steps can bring about this improvement.

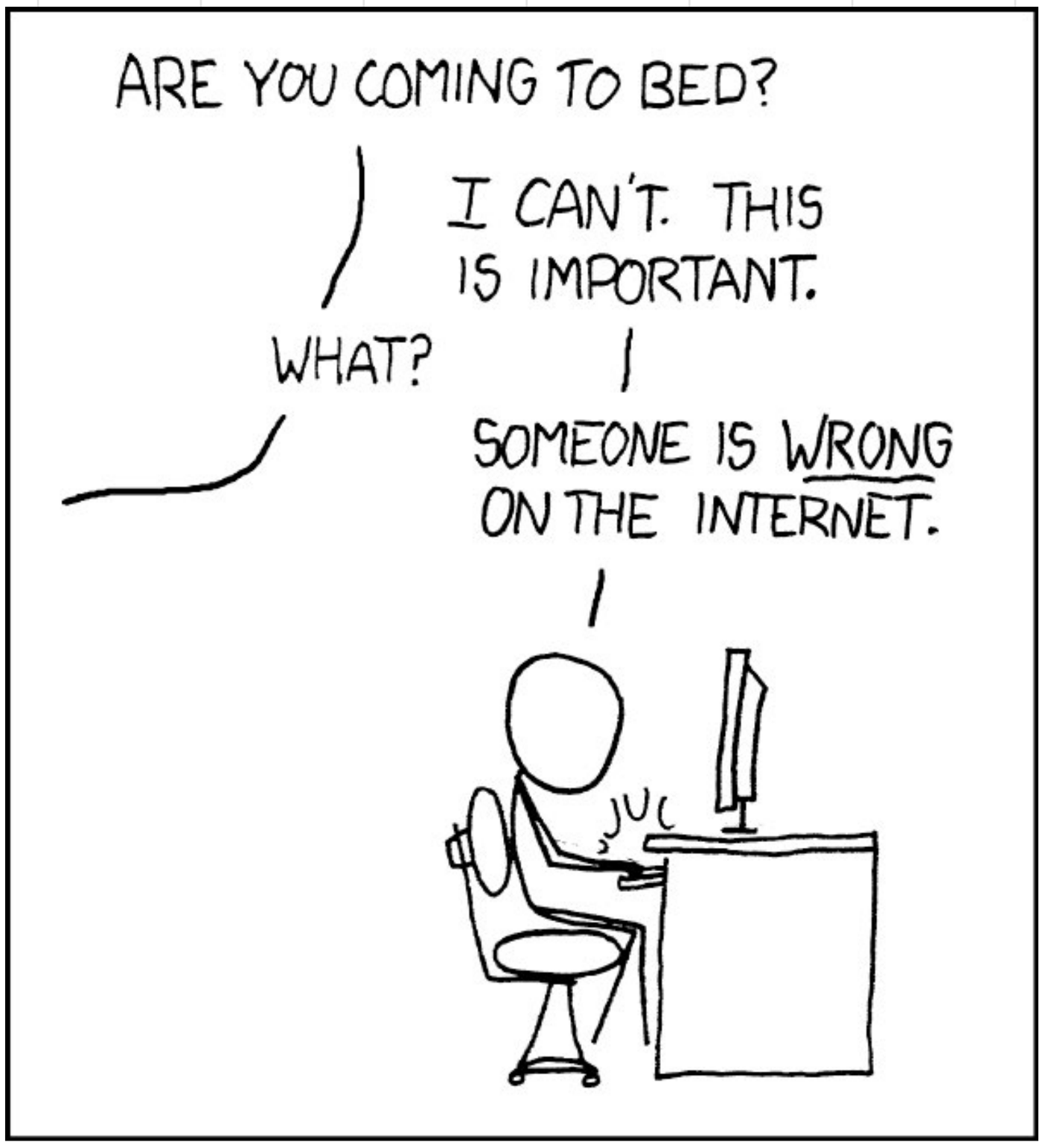
The players — An information ecosystem out of balance

It would be too simplistic to blame the woes of digital discourse on any single actor. Neither Big Tech's profit-driven operation of communication platforms nor political institutions' struggle to regulate the so-called "new territory" are solely at fault. Nor can authoritarian regimes or extremists, domestic or foreign, be held solely responsible for the state of digital discourse. It is also not just down to human psychology, which might inevitably lead to the worst forms of online interaction. In fact, it is the toxic interplay of all these actors, with their intentions, actions, reflexes, and mechanisms, that creates the current situation. The information ecosystem — comprising all the parts that can only exist together, and which form the public discourse space — is out of balance. For tech companies, the time users spend on their platforms equates to money. Therefore, they have optimised their algorithms to maximise this time. Newsfeeds are filled with one post after another, new content is suggested based on preferences, and the promise is that with every swipe or click, better entertainment, bigger surprises, fresher news, or more outrageous scandals await. It is like sugary treats or

fatty crisps: Once you start, it's hard to stop. But here, it is emotions at play, not sugar or fat. Anything that stirs emotions works. While this mechanism is already problematic, it would be less worrying if it were only about fun, entertainment, or sex. But — and here we encounter the unscrupulous and often malicious actors who have quickly grasped and exploited the mechanism — fear, anger, envy, and greed work even better as fuel on social media.

"Someone is wrong on the internet" or 90, 9, 1

Under these conditions, constructive debate, civilised exchange, or understanding is hard to achieve. However, it is an illusion to believe that digital discourse represents the entirety of public opinion. For a while, there was a tendency to read general sentiment from Twitter's (now X) trending topics or individual posts. Yet, the commonly cited 90, 9, 1 rule from online communities likely holds truer. While everyone can potentially participate in the digital public sphere, few do. The rule of thumb is that 90 per cent of users are passive. Around 9 per cent engage by commenting or using standardised forms of engagement like likes or shares. However, actual content creation is done by just a tiny fraction, roughly 1 per cent of users. These individuals often hold strong views. A famous XKCD meme illustrates this well: The 1 per cent of content creators and the 9 per cent who interact often do so either out of strong agreement or — perhaps more frequently — vehement disagreement. The meme shows a stick figure refusing to go to bed because, as it says, "Someone is wrong on the internet."



[Source: https://xkcd.com/386 Creative Commons, Attribution-NonCommercial 2.5 License]

This dynamic causes discussions on social media or in news site comment sections to be more polarising and conflict-driven, rather than constructive through the exchange of arguments. Articles, posts, videos, and comments that provoke strong emotional reactions get the most attention. As a result, online debates appear more polarised, conflict-ridden, and emotional than their real-world, face-to-face counterparts.

Less reporting, more opinion

In this environment, whether you are a private user, professional content creator, or media company, gaining reach means focusing on sensationalism, emotional appeal, and controversial opinions. The potential for comprehensive, transparent information and background quickly gives way to a reality dominated

by strong opinions and interpretations. The distinction between fact and opinion often gets lost. Calls for "freedom of speech" often merely mean claiming the right to assert anything without challenge, whether it is that bleach can cure Covid, that economic elites are replacing populations, or that climate change is a lie.

Is news consumption divided?

The trend towards emotionally charged and opinionated content has not spared editorial media. Traditionally, journalism's role — with its standards and ethics — has been to select what matters from the endless stream of reportable events. This has always included the interpretation of the world's state through commentary and opinion pieces. But as newsrooms and publishers come under increasing pressure, they are tempted to follow the logic of the attention economy. Clickbait, emotionalisation, and increasing polarisation are the result. While these phenomena aren't new to journalism (think "tabloids"), they continue to spread. Where successful, quality journalism — now reliant on solid financial backing — ends up behind paywalls. This threatens to further divide those who can afford curated news from those who rely on snippets and dubious websites. Some national or international quality outlets may still find sustainable business models and interested consumers in the future, but whether this will work for local and regional media remains uncertain.

The next step: AI-generated and personalised

While news media are still searching for ways to

monetise their online content, tech companies are one step ahead. Rather than offering reach to journalistic content, they play with ways to distribute news content directly. AI can provide users with personalised news—based on interests, location, or time of day—straight into their timeline or search results. Where the information comes from, whether it was researched by journalists, and whether your neighbour is receiving the same news or spin is becoming increasingly difficult to determine. What is more, the social aspect of social media is gradually disappearing. It is no longer your network shaping your worldview, but rather a smaller number of tech gatekeepers delivering the news, driven perhaps less by societal value and more by commercial potential. This presents another serious threat to the state of public discourse.

How to ease the digital discourse space

That said, it is not as if nothing can be done. The current state of the digital public sphere is the result of a complex interplay of various actors, and there is room for change. A key step is to hold platform operators—who dominate these digital spaces—more accountable. Platforms must be regulated, and the rules under which they employ algorithms must be made transparent and verifiable. The European Digital Services Act provides some initial tools for this. But this alone won't suffice. Further political measures are needed to curb the influence of disinformation, targeted polarisation, and malicious actors. This could be achieved through supporting independent fact-checkers and promoting media literacy in educational institutions.

Media organisations themselves must also take responsibility. This means adhering to their ethical standards and resisting the temptations of attention-driven economics. Journalistic quality must not be sacrificed for quick clicks. Ultimately, users themselves must take responsibility, critically questioning the content they consume and how they contribute to the quality of discourse.

While these measures aren't exhaustive, they represent important steps towards stabilising the digital discourse space. Achieving this will require collaboration between politics, corporate sector, and civil society. Only by working together can we turn the digital space into a place for exchange, understanding, and democratic participation. One thing is clear: Democracy needs a public sphere where critical discourse and political debate are possible. It will, however, collapse if that space is flooded with hate, conspiracy theories, and deliberate manipulation.

Further reading and exploration

- *Digital Turbulence: Challenges facing Democracies in Times of Digital Turmoil*

- *Social Media and Democracy. The State of the Field and Prospect for Reform*

- *Truth Decay and National Security. Intersections, Insights, and Questions for Future Research*

- *Generative Artificial Intelligence and Political Will-Formation*

- *Digital Discourses and the Democratic Public Sphere 2035*

Imprint:

Bertelsmann Stiftung
Carl-Bertelsmann-Straße 256
33311 Gütersloh

Responsible for the concept
Cathleen Berger
Co-Lead Upgrade Democracy
cathleen.berger@bertelsmann-stiftung.de
upgradedemocracy.de

Authors are listed in the individual chapters

Design:
nach morgen, Cologne