Resonant
Voices
Initiative

Getting Through:
Re-thinking Counternarratives

Resonant
Voices
Initiative

2020

This report has been produced as part of the Resonant Voices Initiative in the EU, funded by the European Union's Internal Security Fund – Police

GETTING THROUGH

Rethinking Counternarratives

TABLE OF CONTENTS

Executive Summary

Counternarratives or counternarrative campaigns have become a staple of strategies to counter – through communication - threats to the national security and democracy, from violent extremism and terrorism, disinformation and foreign influence, to hate crimes and bias-motivated incidents.

While present-day counternarratives evolved as a response to some of the externalities or unintended consequences of widespread internet use and were meant to compensate imbalances created by technological advances and the exploitation of new platform affordances, they quickly became plagued with the same problems affecting the entire information ecosystem. These problems include ethical issues surrounding audience profiling, privacy concerns in relation to tracking and micro-targeting of at-risk individuals, overreliance on algorithms and automated tools to achieve scale. The above are just a fraction of the issues that the field has yet to grapple with in a meaningful way.

The main objective of our report is to challenge the dominant "policy narrative" of counternarratives, focusing on the context of the European Union. We identify missing dimensions and opportunities for enhancing strategic communication efforts by civil society in during the current confidence crisis in the digital ecosystem and in democratic institutions. These include directly engaging with broader social changes to address the issues of concern, more intentional community participation and dialogue, collaboration across disciplines and expertise silos to develop new models and formats of communication campaigns, and a more critical and proactive engagement with the information infrastructure, rules, and platform affordances.

The report seeks to start a wider stakeholder discussion on addressing illegal and harmful online content associated with anti-democratic efforts in the EU, informing the EU-level policies, as well as national policies of the EU countries, with a likely spillover effect to EU candidate or future candidate countries.

We propose future directions for the required policy change that would boost our ability to defend and build resilience against threats that democratic societies are facing from terrorists, authoritarians, populists at home or abroad who are weaponising the internet and exploiting audience vulnerabilities.

These proposed changes focus on reframing of the overall mission and are presented around four challenges that strategic communication can help address.

We formulate four challenges for strategic communication:

- Supporting long term substantive policy reforms and social change processes, explaining costs and benefits of proposed solutions and advantages of messy problem solving in democratic societies, contrasting them to quick fixes and shortcuts promised by populists and propagandists, to audiences that are resistant to truth, facts, evidence and are sceptical of scientific methods.

- Addressing the appeal and resonance of problematic, dangerous narratives, in addition to engaging with their producers, their messages, and their ideologies, through long term multi-level campaigns with audience participation to support community building forces.

- Supporting new models of collaboration, linking existing strategic communication initiatives and efforts to build and sustain broader alliances that expand boundaries and definitions of communication campaigns.

- Explaining benefits, challenges of and limits to technological solutions for detecting, removing content, and suppressing its circulation, as well as banning users and networks at scale, and about risks related to privacy, surveillance, the use of AI and automatic filtering to both public and decision-makers.

Introduction

Democracy is under attack. Whether domestically or via foreign actors, European countries are witnessing a wave of populist politicians and ethno-nationalists, terrorists and violent extremists, using similar tools to influence and manipulate citizens: propaganda, mis/disinformation, and hate speech. While their agenda may not always be overt, deeply unsettling results have followed: from tangible events such as electoral victories of undemocratic parties, terrorist attacks and inter-communal violence; to more pernicious effects such as radicalisation, polarization, extremism and an overall erosion of social cohesion and trust. From an audience's perspective, it can be difficult to decipher who says what with what intentions. At times cacophonic, at others highly privatised, overall the digital communication spaces which audiences navigate are ruled by the opaque and market-driven incentives of a few monopolistic actors.

At the same time, initiatives to solve these issues have proliferated, and there is a wide range of specialised expertise dispersed in several corners of the society. From journalists and factcheckers, to experts in debunking Kremlin propaganda; and from human rights defenders documenting hate against Muslims, to technology activists, the list of highly specialised answers against these 21st century plagues is long. Often, these actors run into similar systemic, structural challenges, pertaining to a similar, overarching informational environment that is increasingly complex, and in which malicious actors thrive.

Counternarratives in this report are understood as both a method, and a policy to address a wide range of security and democracy threats. In an attempt to address the issue of fragmentation, this report aims to provide a holistic perspective on the main challenges that prevent counternarratives potential to be fully utilised. Combining theoretical explorations with practitioners' experience and reviewing of legal and policy frameworks, the report also aims to address some the most prominent critiques around counternarratives, such as measuring impact, theoretical grounding, and the overall lack of a strategy. At the same time, far from advocating a one-size fits all solution, or promoting technological quick fixes, this paper suggests delving into hard questions, that require long-term thinking, and that revolve essentially around the meaning of community-building and social change.

The first part provides an overview of the complex context in which counternarratives are evolving, with theoretical, informational, and governmental parameters that provide both challenges and affordances that require to be acknowledged. The second part then turns guidance and best practices, broken down into the different steps that need to be taken in order to build what is often lacking, a strategy. At last, future directions for the field to reach measurable social change are suggested.

What are counternarratives?

This report aims to systematise and connect the experiences and learnings of various radicalisation/violent extremism/terrorism, disinformation, and hate counternarratives projects. This multidisciplinary approach to counternarratives requires an introduction addressing etymological considerations and presenting the use of this term in a broader historical, methodological, and strategic context.

Discourses in context

Researchers have summed up the capacity for narratives to absorb, or engage, the audience in a state where the storyline distracts the audience from its own reality or perception, in a way that triggers both a cognitive and emotional reaction. Like a novel, a narrative has the power to "transport" the reader into a different state.[1]

A narrative, then, "is a cohesive, causally linked sequence of events that takes place in a dynamic world subject to conflict, transformation, and resolution through non-habitual, purposeful actions performed by characters."[2]

But narratives should also be understood more broadly as essentially relational and linked to a particular context, culture, and set of power relations, as a tool that also helps to form a particular identity and/or community, that helps individuals navigate uncertainty in a way that can be unifying or divisive.

Countering harmful speech

On the most basic of levels, narratives or speech that require countering are generally thought of as those that are outright illegal, and those that are not illegal but harmful or dangerous. There is a plethora of both academic and legal definitions of hate speech[3] and dangerous speech[4], terrorist propaganda and disinformation[5]. Expertise and practices around countering it evolved with a focus on the particular type of speech and the particular definition of related threat – countering hate, countering mis/disinformation, countering violent extremist/terrorist propaganda.

A more colloquial definition of counternarratives, chosen for this report, refers to communication campaigns providing content with information and stimuli that counter the effects of nefarious influences and the manipulation efforts of adversarial actors by correction or compensation. It also takes into consideration a useful definition of counternarratives, which has been developed by Grossman, who argues that they are storylines that "resist, reframe, divert, subvert, or disable other stories and other voices that vie for or already command discursive power"[6]. In this sense, counternarratives aim to challenge discourses that have gone dominant within a particular context, and that present a risk for violence or extremist radicalisation based on falsehoods or hyper-partisanship.

---

[1] See: Moyer-Gusé, E. (2008). Toward a theory of entertainment persuasion: Explaining the persuasive effects of entertainment-education messages. *Communication Theory*, 18(3), 407-425. doi:10.1111/j.1468-2885.2008.00328.x;
and Moyer-Gusé, E., & Dale, K. (2017). Narrative persuasion theories. In R. Patrick (Ed.), *The International Encyclopedia of Media Effects* (Vol. 3, pp. 1345-1354). Chichester, John Wiley & Sons, Inc. doi: 10.1002/9781118783764.wbieme0082; Green, M. C. (2004). Transportation into narrative worlds: The role of prior knowledge and perceived realism. *Discourse Processes*, 38(2), 247-266. doi: 10.1207/s15326950dp3802_5

[2] Braddock, K. and Dillard, J. (2016) Meta-analytic evidence for the persuasive effect of narratives on beliefs, attitudes, intentions, and behaviors, *Communication Monographs* 83(4):446-467.

[3] Sellars, A. (2016). Defining hate speech. *Berkman Klein Center Research Publication*, (2016-20), 16-48.

[4] Benesch, S., Buerger, C., Glavinic, T., & Manion, S. (2018). Dangerous Speech: A Practical Guide. *Dangerous Speech Project*.

[5] Derakhshan, H., & Wardle, C. (2017). Information disorder: definitions. *AA. VV., Understanding and addressing the disinformation ecosystem*, 5-12.

[6] Grossman, M. (2015). Disenchantments: counter-terror narratives and conviviality. In Mansouri, F. (Ed.), *Cultural, Religious and Political Contestations: The Multicultural Challenge* (pp. 71- 89). Cham: Springer. Available at: https://doi.org/10.1007/978-3-319-16003-0

Historically, counternarratives belong to a long line of communication-based responses to social conflict, that predate the internet. At the same time, narratives considered harmful and dangerous – from misinformation to propaganda - have also been effectively deployed long before the advent of digital communications.

But the rise of social media platforms and technology have supercharged the possibilities for spreading all sorts of discourses: on both a macro and micro level, massively and globally, or privately and in a hyper-targeted way.

Counternarratives and the national security

In the last decade, terrorist organisations have quickly adopted technology-enabled communication strategies to distribute their propaganda materials and recruit new members. This has in turn, shifted the focus of national security institutions to counterterrorism interventions with the following objectives: on the one hand, halting the spread of terrorist materials, and limiting internet access to terrorists; and on the other hand, favouring communication that counters terrorist ideology.

As a result, counternarratives have effectively become part of national security strategies, against terrorism, violent extremism, and radicalisation, with an overwhelming focus on online communication. Consequently, early application of counternarratives in the context of online jihadist propaganda and recruitment have sparked justified criticism.[7]

In addition to terrorism and terrorist propaganda, governments have also developed expertise and practices to address other threats to security and democracy and associated speech, particularly hate speech, and disinformation.

Nowadays, counternarratives evoked in the context of national security usually appear alongside alternative or positive narratives as a composite term in policy documents. They are also sometimes associated with government strategic communications, or counter-messaging.[8]

According to this categorization, alternative narratives are made of positive stories about social values, tolerance, openness, freedom and democracy, whereas counter-narratives would engage with or respond to the extremist ideology. Counternarratives, the authors argued, should be outsourced by the government to credible messengers, and campaigns should be implemented by civil society organization. This position is the one reflected in the current practice of counternarrative campaigns aimed against terrorism.[9]

[7] See for example: Rosand, E., Winterbotham, E. (2020) Do counter-narratives actually reduce violent extremism? *Brookings Institution* Available at : https://www.brookings.edu/blog/order-from-chaos/2019/03/20/do-counter-narratives-actually-reduce-violent-extremism/

[8] The categorisation of narratives as either counter, alternative or positive first appeared in a report "Review of Programs to Counter Narratives of Violent Extremism" by the Institute for Strategic Dialogue in 2013, in which he authors proposed a 'counter-messaging spectrum' comprising government strategic communications, alternative narratives and counter-narratives. Cited in United Nations' Counter-Terrorism Committee Executive Directorate (UN CTED) Analytical Brief Countering terrorist narratives online and offline

[9] Briggs, R. and Feve, S. (2013) Review of Programs to Counter Narratives of Violent Extremism *Institute for Strategic Dialogue* Available at: https://www.dmeforpeace.org/peacexchange/wp-content/uploads/2018/10/Review-of-Programs-to-Counter-Narratives-of-Violent-Extremism.pdf

Definitions used in the report

This report uses the term counternarratives to refer to all government-sponsored communication implemented by third parties in pursuing public policy goals related to the protection of security and democracy. As such, all counternarrative practices as part of a government policy must be based on respect and protection of freedom of expression and firmly rooted in the wider human rights framework.

While the report focuses on government and platform policy of supporting and promoting narratives to counter extremist radicalisation and violence, it attempts to draw comparisons and links with similar policies to address other types of speech considered harmful, especially hate speech and disinformation.

Furthermore, it takes a multidisciplinary approach that draws on research findings from diverse fields such as persuasion, social psychology, political science, journalism studies, marketing, advertising, strategic communication, and media and communications more generally. It also discusses practical insights from fields such as development, peace and democracy building, social cohesion, and journalism.

Obstacles and Opportunities for Counternarratives

The consensus around counternarratives is that they lack a set of clearly articulated theories[10]. Therefore, the report in this section examines the theoretical foundations and the empirical evidence informing counternarrative campaigns. It reviews key concepts, highlights the complexity of how beliefs are shaped, influenced by information in the media ecosystem, and the knowledge gaps around it. In then turns focus to key theories and evidence informing the terrorism counternarratives policy and practice, as well the regulatory, technological and policy context within which they are implemented.

There also seems to be a general consensus around the fact that both platforms and governments fall short at providing effective measures to limit the spread of hate speech, disinformation, and extremist propaganda. Counternarratives, as part of the solution, rely on an ambiguous technological ecosystem, which is constantly evolving and is subject to evolving regulation. The report aims to highlight the main challenges of implementing counternarrative campaigns in the context of fast-changing technology, policies and regulation, and suggest how they can be better addressed in the future.

What do we know about how beliefs are shaped?

A short review of academic debates, the relevant frameworks and available empirical evidence that practitioners could benefit from greater engagement with, is presented in this section. Some of the theoretical concepts and empirical findings may already be familiar and implicitly used, some would deserve to be more thoroughly appreciated. A clear theoretical grounding, acknowledging what is known and what is not known could ultimately lead to better results, and return on investment[11].

The review focuses on the complexity of individual media effects from media exposure in the digitally mediated, networked communication and cross-media practices, and effects of narrative-based messaging, in light of various cognitive and information processing biases.

While this review is by no means exhaustive and may offer a simplified perspective of complex academic debates, it could offer some actionable insights useful for campaign strategies.

Magic bullet theory vs. media effects

The most common assumption behind counternarratives is that passive exposure to information is sufficient to influence an audience's attitudes, beliefs and behaviours, which can be linked to magic

[10] Ahmed, M., Bindner, L., Bright, J., Busher, J., Coyer, K., Crosset, V., Davies, H., Gallacher, J., Ganesh, B., Gluck, R., Lee, B., Pohjonen, M., and Reeve, Z., (2019) Extreme Digital Speech: Contexts, Responses and Solutions *Vox Pol* p15. Available at: https://www.voxpol.eu/download/vox-pol_publication/DCUJ770-VOX-Extreme-Digital-Speech.pdf

[11] van Eerten, J. J., Doosje, B., Konijn, E., de Graaf, B. A., & de Goede, M. (2017). Developing a social media response to radicalization: The role of counter-narratives in prevention of radicalization and de-radicalization. *Amsterdam: University of Amsterdam.* Available at: https://dspace.library.uu.nl/handle/1874/360002

bullet theory, or hypodermic needle model[12]. While largely disproven, it has regained popularity with the rise of big data and the possibility of delivering micro targeted messages on social media and it underpins much of current online advertising and campaigning practice. This assumption presents three problems: it shuns the complexity of how messages may change behaviours, it relies on measurement that is based on commercial and short-term metrics that have limited use in capturing social change; and it assumes that audiences are passive and homogenous.

Behind this simplistic model lays a complex theory: media effects. Media effects are "deliberative and non-deliberative short- and long-term within-person changes in cognitions, emotions, attitudes, beliefs, physiology, and behaviour that result from media use."[13] While this concept has been discussed over decades of communications research, it can be narrowed down to its particular micro perspective: a specific message could change an individual's perspective in a short amount of time[14].

There is a consensus around the difficulty of measuring such effects, particularly when it comes to "curative" effects of exposure to counternarratives, due to the wide range of variables that come into play - long term or short term - that relate to discourses, but also to broader systemic, cultural and societal factors and particular local contexts and individual traits, that all contribute to shaping beliefs and attitudes.

Narrative persuasion

Counternarrative campaigns are seeking to influence audiences, changing their "hearts and minds" by immersing them in narratives, i.e. stories, using a variety of genres and formats. As such, they adapt their messaging strategy from the field of entertainment-education[15]. The narrative transportation approach "distinguishes entertainment-education message processing from that of overtly persuasive messages."[16] Narrative transportation is explained as a "mental state that produces enduring persuasive effects without careful evaluation of arguments." Audience engaged in a story experiences "vicarious cognitive and emotional response to the narrative."[17]

There is ample empirical evidence to support the idea that transportation into a narrative can increase acceptance of the messages contained in a narrative and that narrative transportation can cause affective and cognitive responses, as well as changes in beliefs, attitudes and intentions. For example, a recent study comparing various journalism formats showed that narrative stories that framed the

[12] Ahmed, M., Bindner, L., Bright, J., Busher, J., Coyer, K., Crosset, V., Davies, H., Gallacher, J., Ganesh, B., Gluck, R., Lee, B., Pohjonen, M., and Reeve, Z., (2019) Extreme Digital Speech: Contexts, Responses and Solutions

[13] For an overview if various media effect theories: Valkenburg, P M. and Jochen P. (2013) "The Differential Susceptibility to Media Effects Model." *Journal of Communication* 63: pp 221-243

[14] Napoli, P. M. (2014). Measuring media impact. *The Norman Lear Center.* Available at: https://learcenter.org/pdf/measuringmedia.pdf

[15] Moyer-Gusé, E. (2008) Toward a Theory of Entertainment Persuasion: Explaining the Persuasive Effects of Entertainment-Education Messages. *Communication Theory* 18: pp 407-425.

[16] This notion of narrative involvement has been given several different labels across the literature, including absorption, transportation, engagement, immersion, and engrossment: see Bandura, A. (2004). Social cognitive theory for personal and social change by enabling media. In A. Singhal, M. J. Cody, E. M. Rogers, & M. Sabido (Eds.), *Entertainment-education and social change: History, research, and practice* pp.75–96. Mahwah, NJ:Lawrence Erlbaum; Gerrig, R. J. (1993). Experiencing narrative worlds. New Haven, CT: Yale University; Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, 79, pp 701–721.; Slater, M. D., & Rouner, D. (2002). Entertainment-education and elaboration likelihood: Understanding the processing of narrative persuasion. *Communication Theory*, 12, pp173–191. All of the above as found in Moyer-Gusé, E. (2008). *Toward a theory of entertainment persuasion*

[17] van Laer, T., Ruyter, K., Visconti, L M. and Wetzels, M. (2014) The Extended Transportation-Imagery Model: A Meta-Analysis of the Antecedents and Consequences of Consumers' Narrative Transportation. *Journal of Consumer Research*, 40(5), 797-817

issues through the experiences of individuals produced more favourable evaluations of stigmatised groups[18].

On the other hand, it is understood that overtly persuasive messaging can cause reactance – a perceived threat to one's freedom to choose their own attitudes and behaviours - and be rejected or even cause a boomerang effect as a result, even if the message recommendation is in the receiver's best interest[19].

Cognitive and information processing biases

This existing evidence of the effects of narrative persuasion largely supports the theory of change of counternarrative campaigns. However, pre-existing beliefs and attitudes, as well as peers, significantly influence media choice and information processing. The concepts presented below, drawing from psychology studies, explore the psychological and social mechanisms that are at play when beliefs are formed, in the context of misinformation and propaganda appeal. These biases are currently not sufficiently reflected in the design of counternarrative campaigns.

- **Confirmation Bias**

Confirmation bias[20] pushes people to look for information that validates their prior beliefs. It is a form of motivated reasoning[21], which pushes individuals to arrive at a particular conclusion by balancing accuracy motivations to find the correct answer with directional motivations – a preference to arrive at a particular conclusion that is consistent with one's belief or attitude.

- **Selective and Cross-Cutting Exposure**

Studies on selective exposure[22], a concept based on the theory of cognitive dissonance, have found evidence that people tend to prefer information that is consistent with their attitudes, and to filter out inconsistent information. Additionally, there is some evidence suggesting that selective exposure is stronger for those holding more extreme views[23].

At the same time, it seems that cross-cutting exposure, which is the exposure to opposing viewpoints, that are counter to those people hold, is also quite common[24]. Overall, there is mixed evidence that both selective and cross-cutting exposure contribute to polarisation.

- **Endorsement**

In addition to our individual biases, peers also influence information processing. Endorsement is a mechanism of social validation whereby the trust placed in peers affects whether information will be

[18] Oliver, M. B., Dillard, J. P., Bae, K., & Tamul, D. J. (2012). The Effect of Narrative News Format on Empathy for Stigmatized Groups. *Journalism & Mass Communication Quarterly*, 89(2), 205–224. https://doi.org/10.1177/1077699012439020

[19] Moyer-Gusé, E. (2008). *Toward a theory of entertainment persuasion*

[20] Boyd, D. (2018) You Think You Want Media Literacy… Do You? *SXSW Edu Keynote*. Available at: https://points.datasociety.net/you-think-you-want-media-literacy-do-you-7cad6af18ec2

[21] Kunda, Z. (1990) "The case for motivated reasoning. *Psychological bulletin* 108(3):480

[22] Knobloch-Westerwick, S and Meng, J. (2009). Looking the Other Way: Selective Exposure to Attitude-Consistent and Counterattitudinal Political Information. *Communication Research* - COMMUN RES. 36. 426-448. Available at: https://doi.org/10.1177/0093650209333030.

[23] Stroud, N. J. (2010). Polarization and partisan selective exposure. Journal of Communication, 60, 556–576.

[24] Matthes, J., Knoll, J., Valenzuela, S., Hopmann, D N., and Sikorski, C von. (2019). A Meta-Analysis of the Effects of Cross-Cutting Exposure on Political Participation. *Political Communication*. 1-20. Available at: https://doi.org/10.1080/10584609.2019.1619638.

perceived as credible or not[25]. Concretely, the perceived credibility and authority of some sources endorsing information may be more influential on someone's behaviour, attitudes and beliefs than the simple accuracy of the message.

- **Affect and Emotion**

In the context of terrorist propaganda, dis/misinformation and hate speech, there is a historically understudied role played by affect and emotion. It is increasingly being acknowledged as a major factor in giving rise to populism and polarisation across the ideological spectrum in countries across the world[26]. This is partially because affectively and emotionally charged narratives have been found to be more effective in the context of societies where large groups have experienced the failure of rationality in politics and bureaucracy to improve their economic or their moral well-being. Affect and emotion-based discourses can bypass rational discourse and create a 'direct connect' with audiences[27].

Media ecosystem

Understanding the reception side of communication and the processing of messages also requires paying attention to media repertoires[28] and increasingly personalised yet diffused information consumption. This consumption is taking place in complex and fragmented media landscapes.

Most counternarrative creators develop their stories and messaging in response to terrorist propaganda, with a goal to prevent or counteract individual exposure to them. As this communication is mostly distributed and consumed using social media platforms, the concern becomes less about the specific output of one group, or a certain category of content, and more about its spread and interaction in the media ecosystem.

Information and narratives do not only circulate in digital spaces but also through mainstream media outlets, and public figures. Called transmediality[29], this mechanism highlights how misinformation can quickly gain traction through different mediums, somehow even reaching public figures who transmit them back to their audiences, and how it can, in the long term, contribute to shaping erroneous beliefs.

[25] Mena, P., Barbe, D., & Chan-Olmsted, S. (2020). Misinformation on Instagram: The Impact of Trusted Endorsements on Message Credibility. *Social Media + Society*. https://doi.org/10.1177/2056305120935102

[26] Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, *114*(28), 7313-7318.

[27] Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146-1151. Vosoughi, S, Roy, D,. and Aral, S. (2018) The spread of true and false news online *Science* Vol. 359, Issue 6380, pp. 1146-1151 Available at:https://science.sciencemag.org/content/359/6380/1146

[28] Hasebrink, U. & Popp, J. (2006) Media repertoires as a result of selective media use. A conceptual approach to the analysis of patterns of exposure. *Communications*, 31, pp 369-387. Available at: https://www.researchgate.net/publication/240753317_Media_repertoires_as_a_result_of_selective_media_use_A_conceptual_approach_to_the_analysis_of_patterns_of_exposure

[29] Banaji, S., Bhat, R., Agarwal, A., Passanha, N., and Sadhana-Pravin, M. (2019). WhatsApp vigilantes: an exploration of citizen reception and circulation of WhatsApp misinformation linked to mob violence in India. *LSE Blogs*. Available at: https://blogs.lse.ac.uk/medialse/2019/11/11/whatsapp-vigilantes-an-exploration-of-citizen-reception-and-circulation-of-whatsapp-misinformation-linked-to-mob-violence-in-india/

For example, recent studies have shown that European media has systematically framed the arrival of migrants in Europe as a "crisis"[30], and failed to "humanise" them[31], a narrative that has in turn been widely amplified by malicious online actors. In the US, the right-wing media ecosystem was able to shape discussions in mainstream media, thus providing more attention to the immigration agenda of Donald Trump[32].

Importantly, the current gaps and shortcomings of mainstream media may also play a significant role in the spread of harmful messages. In the last decade, established news outlets have faced a deep crisis with severe democratic consequences, including a steady decline in the overall media's trustworthiness. Scholars have highlighted several factors that have led to a democracy deficit, providing a fertile ground for misinformation to spread. First, the financial crisis may have accelerated a trend toward decreasing resources for centrist, public service media that could offer a strong voice against populism[33].The rise of commercialism in mainstream news may have also contributed to weakening of professional journalism[34]. Second, scholars have noted the decline of robust local newspapers across Europe[35], and have linked this trend to the rise of polarisation: as newsrooms are concentrated in urban areas, they fail to acknowledge the realities and voices of citizens living in rural areas[36]. At the same time, news agenda setting may be highly influenced by an increasingly concentrated media ownership, which is detrimental for media plurality[37]. Thus, the concerns of many citizens go unaddressed, creating a vacuum that risks being exploited by malicious actors. As a result, in a time of high uncertainty such as during the ongoing COVID-19 crisis, trust in news across Europe is at its lowest[38] and falsehoods evermore contagious.

Countering Terrorism Through Narratives

Theories and evidence presented in the short review in the previous section have broader applicability in public safety, health, and political campaigns, or in advertising commercial products. This section critically examines their application in the context of counterterrorism.

To understand and evaluate the persuasion of both terrorist messaging and narratives and their counternarratives, policymakers and practitioners rely on another key theoretical concept: radicalisation (the key mechanism); a catch-all term capturing beliefs, attitudes and behaviours consistent with increased engagement with violent extremism. It has been understood as an increased acceptance of violence and/or an increased alignment with terrorist group's ideology, and policies

[30] Chouliaraki, L., Georgiou, M., Zaborowski, R., and Oomen, W. A. (2017). The European 'migration crisis' and the media: a cross-European press content analysis. *The London School of Economics and Political Science*, London, UK.

[31] Chouliaraki, L., & Stolic, T. (2017). Rethinking media responsibility in the refugee 'crisis': a visual typology of European news. *Media, Culture & Society*, *39*(8), 1162–1177. https://doi.org/10.1177/0163443717726163

[32] Benkler, Y., et al. 2017. "Study: Breitbart-Led Right-Wing Media System Altered Broader Media Agenda," *Columbia Journalism Review*. Available at: https://www.cjr.org/analysis/breitbart-media-trump-harvard-study.php

[33] Freedman, D. (2018). Populism and media policy failure. *European Journal of Communication*, *33*(6), 604-618.

[34] Pickard, V. (2018). When commercialism Trumps democracy: Media pathologies and the rise of the misinformation society. *Pablo J. Boczkowski/Zizi Papacharissi (Hg.): Trump and the Media. Cambridge, Mass./London*, 195-202.

[35] Nielsen, R. K. (Ed.). (2015). *Local journalism: The decline of newspapers and the rise of digital media*. Bloomsbury Publishing.

[36] Ramsay, G., & Moore, M. (2016). Monopolising local news: Is there an emerging local democratic deficit in the UK due to the decline of local newspapers. *Centre for the Study of Media, Communication and Power*.

[37] Media Reform Coalition. (2015). Who Owns the UK Media. Retrieved from https://www.mediareform.org.uk/wp-content/uploads/2019/03/FINALonline2.pdf

[38] Ferraresi, M. (2020, April 24).As Europe Confronts Coronavirus, the Media Faces a Trust Test. *Nieman Reports.* Retrieved from https://niemanreports.org/articles/a-trust-test-for-the-media-in-europe/

advocated by them. To achieve their goal, terrorists use persuasive narratives to stimulate the process of radicalisation, proceeding from the adoption of radical beliefs to involvement in terrorism.

Radicalisation is used as the key explanatory mechanism of how individuals become involved in terrorism In the majority of government CT/CVE strategies, despite the evidence that most radicals never transition to violence and those who do are not always motivated by their beliefs.[39] This approach has also been subject to mounting criticism due to its flawed application, on the grounds of conflating violence with radical ideology, predominant association with one religious ideology in particular, and ignoring or downplaying the broader socio-economic, political and cultural context.[40,41]

Counternarratives, as one part of the counterterrorism/countering violent extremism policies are subject to the same criticism, aggravated by their visibility. In the absence of an acknowledgement of and a wider strategy to address underlying grievances, while continuing to approach violent extremism as a threat associated with particular minority groups, counternarratives risk causing more harm than good.

"The question to ask is what the added value is of these programmes, considering factors such as collapsing educational institutions, corruption, discriminatory governance and lack of a national vision, lack of policies to ensure the basic collective and individual freedoms, control and territorial occupation systems."[42]

Research does seem to indicate that exposure to extreme right-wing content online is linked to motivations that concern individual and societal factors, answering emotional and material needs.[43] This suggests the need for counternarratives to be engaged with society-wide concerns.

"Once a conflict problem is rebranded as a 'violent extremism' problem, it can be hard to see beyond the assumption that the problem lies exclusively with 'extremists'. The grievances that may drive violent movements become little more than nefarious narratives used to exploit vulnerable people, who do not understand the true facts and their own interests."[44]

A corollary problem is the lack of conceptual clarity surrounding the notion of 'vulnerability' to radicalisation. "People vulnerable to radicalisation' may include 'vulnerable people', but 'vulnerable people' are not necessarily 'vulnerable to radicalisation.'"[45] This can lead to problematic targeting and

[39] Schuurman, B., and Taylor, M. (2018) "Reconsidering Radicalization: Fanaticism and the Link Between Ideas and Violence." *Perspectives on Terrorism*, vol. 12, no. 1, pp. 3–22.

[40] Hemmingsen, A-S., and Castro, K I. (2017) The Trouble with Counter-Narratives, *Danish Institute for International Studies*, Available at: https://css.ethz.ch/content/dam/ethz/special-interest/gess/cis/center-for-securities-studies/resources/docs/DIIS_RP_2017_1.pdf

[41] Abu-Nimer, M. (2018) Alternative Approaches to Transforming Violent Extremism: The Case of Islamic Peace and Interreligious Peacebuilding in: Austin, B., and Giessmann H J., (eds). *Transformative Approaches to Violent Extremism.* Berghof Handbook Dialogue Series No. 13. Berlin: Berghof Foundation.

[42] Abu-Nimer, Alternative Approaches to Transforming Violent Extremism p.6

[43] Odağ, Ö., Leiser, A., and Boehnke, K. (2019) Reviewing the role of the internet in radicalisation processes, *Journal for Deradicalisation*, 21, pp261-300

[44] Attree, L (2017) Shouldn't YOU be countering violent extremism? *Saferworld* Available at: https://saferworld-indepth.squarespace.com/

[45] Corner, E; Bouhana, N and Gill, P (2018) The multifinality of vulnerability indicators in lone-actor terrorism. *Psychology, Crime and Law , 25 (2) pp. 111-132*

biased radicalisation risk assessment among practitioners[46], including those working in the counternarrative space.

"Delivering messaging or counter-narratives against specific individuals or groups is very risky and the opposite of a conflict prevention or peacebuilding approach."[47]

Further, the current approaches still overwhelmingly focus on deconstructing terrorist narratives, a strategy that is reactive by default[48]. By seeking to deconstruct them, counternarratives may end up repeating and amplifying extremist messaging[49].

Finally, the broader problem created by the securitised understanding of radicalisation, is that an alternative, positive meaning describing "an intensification of political engagement and the drawing of political frontiers, not necessarily aimed at undermining the state through terrorism"[50] became tainted, despite its centrality to activism and social change. Inclusionary social change through non-violent means to protect and/or expand rights of individuals and communities might be the only alternative to terrorism.

In the following section, technological aspects of communication and related policy and regulation dilemmas are briefly discussed, with a view of their impact on the counternarrative practice.

Platforms and Technology

Both a megaphone for harmful speech, and an opportunity for counternarratives to be promoted more effectively, the role of platforms and technology in counternarrative campaigns is important, yet ambiguous, and its use and associated risks must be carefully examined by counternarrative practitioners.

On the one hand, platforms have democratised publishing and enabled access to unprecedented volumes and variety of information. They created a communication infrastructure, with design features and rules, that incentivise certain behaviours to achieve their advertising oriented, commercial objectives.

On the other hand, the collaterals of this development are often at odds with the objectives of public security or democracy protection policies and programmes. They provide affordances for content creation by fringe political actors[51], allowing them to connect easily with disenfranchised audiences, and ample tools to spread their messages.

---

[46] de Weert, A., and Eijkman, A M., (2019) Early detection of extremism? The local security professional on assessment of potential threats posed by youth, *Crime, Law and Social Change*, 73:491-507 Available at: https://www.readcube.com/articles/10.1007/s10611-019-09877-y

[47] Ernstorfer, A (2019) Conflict Sensitivity in Approaches to Preventing Violent Extremism: Good intentions are not enough *UNDP Reflection Paper*. Available at: http://www.pvetoolkit.org/media/1216/conflict-sensitivity-in-approaches-to-pve.pdf

[48] Braddock, K and Horgan, J. (2015). Towards a Guide for Constructing and Disseminating Counternarratives to Reduce Support for Terrorism. *Studies in Conflict & Terrorism*. 39. pp381-404 Available at: https://doi.org/10.1080/1057610X.2015.1116277

[49] Schmitt, J., Rieger, D., Rutkowski, O., and Ernst, J. (2018). Counter-messages as Prevention or Promotion of Extremism?! The Potential Role of YouTube Algorithms. Vol 68, Issue 4, Pages 780–808, https://doi.org/10.1093/joc/jqy029

[50] Karakatsanis, Leonidas, and Marc Herzog. "Radicalisation as Form: Beyond the Security Paradigm." *Journal of Contemporary European Studies* 24.2 (2016): 199–206.

[51] Munger, K., & Phillips, J. (2019). A supply and demand framework for YouTube politics. *Preprint*.

These dynamics enable divisive and inflammatory speech to travel quickly on social media platforms. In that sense, manipulators and propagandists may always be one step ahead of counternarrative campaigns, as these design features might play into their advantage.

Counternarrative practitioners have sought to use the same tools to spread their messages. What practitioners should keep in mind is first that social media metrics may not adequately reflect social change, as they are based on commercial and advertising success benchmarks; and second that extreme speech is very likely to achieve far more impressive metrics in terms of reach and engagement. One study has shown that in some cases, fake news Facebook interactions figures were even higher than those of established newsbrands[52].

Content recommendation

Platforms decide what users see[53], based on personalised recommendations, defined as "algorithmic selection by service providers of 'content' served to individuals or groups according to some determination made by the service provider of relevance, interest, importance, popularity, and so on to those individuals or groups."[54]

Personalisation of content is based on the large quantity of data that platforms have at their disposals about their users. In addition to privacy concerns, the underlying problem is that the public does not know what and how decisions are being made that lead recommendation algorithms to favour some content over another[55]. Studies seem to indicate that recommendation algorithms feed both on technological input and on the audience's choices, although in a nuanced way.

There is also mounting evidence that algorithmic recommendation systems can cause harm, for example through amplifying harmful content, or exploiting existing user vulnerabilities. Research has shown that exposure to extremist content online plays a role in enhancing already existing right-wing populist sentiment[56]. Respondents with already existing anti-immigrant attitudes tend to look for content that confirms their views on that matter, thus feeding further these attitudes as well as their anxiety[57]. Dynamics of group endorsement in closed groups, whether online or offline, contribute to increased polarisation.[58] Another study has shown that diminishing exposure to Facebook had had the effects of augmenting their well-being and diminished political polarisation[59].

[52] Fletcher, R., Cornia, A., Graves, L., & Nielsen, R. K. (2018). Measuring the reach of "fake news" and online disinformation in Europe. Reuters Institute factsheet.

[53] See Appendix 2 for overview of these recommendation systems on major platforms

[54] Cobbe, J., and Singh, J., (2019). Regulating Recommending: Motivations, Considerations, and Principles, *European Journal of Law and Technology*, 10(3),

[55] Leerssen, P. (2020). The Soap Box as a Black Box: Regulating transparency in social media recommender systems. Available at: https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=3544009

[56] Heiss, R., & Matthes, J. (2020). Stuck in a nativist spiral: Content, selection, and effects of right-wing populists' communication on Facebook. *Political Communication*, *37*(3), 303-328.

[57] Heiss, R., & Matthes, J. Stuck in a nativist spiral: Content, selection, and effects of right-wing populists' communication on Facebook

[58] Lee, E.J. (2007). Deindividuation effects on group polarization in computer-mediated communication: The role of group identification, public-self-awareness, and perceived argument quality. *Journal of Communication, 57*(2), 385-403.

[59] Allcott, H., Braghieri, L., Eichmeyer, S., & Gentzkow, M. (2020). The welfare effects of social media. *American Economic Review*, *110*(3), 629-76.

It remains difficult to contain the spread of harmful content as the system is designed to favour it. Whether via closed Facebook groups, so-called YouTube rabbit holes, or echo-chambers, platforms have been criticised for inadvertently facilitating exposure to harmful content and aiding polarisation and radicalisation[60]. There are concerns that they may be architecturally built in a way that favours extreme speech; and the system's vulnerabilities seem to have effectively been hijacked by malicious actors, who may be highly digitally literate[61], and proficient at exploiting platforms policy greyzones.

"Prioritising for engagement is likely to favour content that produces an emotional response and therefore may be controversial, shocking, or extreme, as people tend to be drawn to this content."[62]

"Viral outrage for many algorithm-driven services is a key driver of value, with products and applications that are designed to maximise attention and addiction."[63]

While this type of content generates engagement, individual motives why people engage with it remain hidden. It is therefore important not to conflate interest with endorsement, or with other reasons people might have for clicking on fear or anxiety-inducing titles for example.

Content moderation

All social media and other internet services users accept company's terms of service, which guide their conduct and the actions company can take on their content and accounts. Primarily designed around suppressing bad content and detecting abuse at scale, content moderation relies on both automation and human evaluation, when reviewing large volumes of third-party, i.e. user-generated content, which platforms, as intermediaries, carry, and which presents potential liability for them.

To encourage growth in the early days, platforms were given "immunity" from liability for what their users post, provided they take an action upon notification that there is a problem with user content[64]. There are a variety of mechanisms and processes that are constantly evolving to support this moderation of user-generated online content, and content moderation is a growing field of practice of a particular importance for counternarrative practitioners who are using platforms to connect with their audiences.

The content moderation policies and processes do not only enforce the rules and decisions that govern content removals, they influence the scale and context in which a message is seen, through content ranking and downranking, surfacing or hiding content in search results or news feeds.

[60] Hao, K. (2019) DeepMind is asking how AI helped turn the internet into an echo chamber, *MIT Technology Review* Available at: https://www.technologyreview.com/2019/03/07/65984/deepmind-is-asking-how-google-helped-turn-the-internet-into-an-echo-chamber/

[61] Banaji, S., *et al* WhatsApp vigilantes.

[62] Cobbe, J *et al* Regulating Recommending: Motivations, Considerations, and Principles

[63] Opinion3/2018, EDPS Opinion on online manipulation and personal data, *European Data Protection Supervisor* p.3 Available at: https://edps.europa.eu/sites/edp/files/publication/18-03-19_online_manipulation_en.pdf

[64] Kuczerawy, A., (2018). From 'Notice and Take Down' to 'Notice and Stay Down': Risks and Safeguards for Freedom of Expression. Frosio, G (ed), *The Oxford Handbook of Intermediary Liability Online*, Forthcoming, Available at: https://ssrn.com/abstract=3305153

The policies and processes to promote "desirable" speech and counternarratives, are not systematically considered and platform policies in this space are underdeveloped, despite repeated commitments, discussed further below.

Behavioural microtargeting

A report by the UK government has flagged that current advertising practices around targeting do not match the OECD human-centred principles on AI, namely because they exploit people's vulnerabilities and prevent autonomy, and its use of data collection and profiling may increase discrimination[65]. "The potential for discrimination in targeted advertising arises from the ability of an intentionally malicious—or unintentionally ignorant—advertiser could leverage such data to preferentially target (i.e., include or exclude from targeting) users belonging to certain sensitive social groups (e.g., minority race, religion, or sexual orientation)."[66]

Since the Cambridge Analytica scandal, there have been multitude of reports, by the media, researchers[67], and governments[68] on how microtargeted advertising has been weaponised, and the complicity of platforms in this process. Furthermore, it has been shown that the lack of transparency around microtargeting practices is known to raise suspicions of surveillance which may lead the audience to change their behaviour.[69] On the other hand, there are also recognised benefits of targeting, and cases when microtargeting "for good" is justifiable or desirable. Targeting people based on "risky" search terms and other online behaviour or interest proxies, is one such example.

Counternarrative campaigns frequently rely on social media platforms to deliver targeted content to their audiences. It is not uncommon for counternarrative campaigns to advertise on platforms, channelling the government funding to platforms to rectify or prevent harms platforms are seen as co-creating. Both a potential amplifier of counternarratives, and a megaphone for dangerous speech, social media platforms, driven by advertising-based business incentives, profit from both sides[70].

Government Policy and Regulation

In response to issues presented in the previous section, Government's approaches to addressing harmful online speech and address the role of technology are fast evolving. The drive to limit the spread of terrorist content, misinformation, and hate speech has motivated a lot of government

[65] Taylor, R. (2020). Online Targeting - Final Report and Recommendations. *Center For Data Ethics and Innovation*, UK Government. Available at: https://www.gov.uk/government/publications/cdei-review-of-online-targeting/online-targeting-final-report-and-recommendations

[66] Speicher, T., Ali, M., Venkatadri, G., Ribeiro, F., Arvanitakis, G., Benevenuto, F., Gummadi, K P., Loiseau, P. and Mislove, A., (2018). "Potential for Discrimination in Online Targeted Advertising." *Proceedings of Machine Learning Research: - Conference on Fairness, Accountability, and Transparency*, 81:1–15. New York, United States, 2018.

[67] Angwin, J., Varner, M., and Tobin, A., (2017) Machine Bias: Facebook Enabled Advertisers to Reach 'Jew Haters' *Pro Publica.*

[68] Democracy disrupted? Personal information and political influence. *Information Commissioner's Office* (2018)

[69] *Richards, N. (2015). Intellectual privacy: Rethinking civil liberties in the digital age*. Oxford University Press, USA., Dobber, T., Trilling, D., Helberger, N., & de Vreese, C. (2019). Spiraling downward: The reciprocal relation between attitude toward political behavioral targeting and privacy concerns. new media & society, 21(6), 1212-1231.Cited in Dobber, T., Ó Fathaigh, R., & Zuiderveen Borgesius, F. (2019). The regulation of online political micro-targeting in Europe. *Internet Policy Review*, *8*(4).

[70] See for example: Purnell, J., Horowitz, J. (2020) Facebook's Hate-Speech Rules Collide With Indian Politics, The Wall Street Journal

regulatory activity in recent years. Efforts both in the in EU[71] and globally included the adoption of voluntary codes of practice and regulation to address the threats presented by online speech proliferated. Their aim is to define online speech that is illegal or harmful[72] and guide and enforce its removal, blocking and filtering by the intermediaries, i.e. platforms, with speed and at scale[73].

Building accountability and trust in the digital ecosystem

- **CONTENT MODERATION TRANSPARENCY**

Given the potential for unintentional bias in the application of these new regulations and the potential for their intentional abuse by authoritarian governments, the stakes for democracy are high. Requiring robust transparency with regards to removals and other actions taken on content from both platforms and governments and insisting that illegal and harmful content regulation and moderation strictly adheres to the human rights standards have been at the centre of debate. [74, 75]
Both the debate itself and the enforcement of these rules have direct consequences for counternarrative practitioners.

- **DATA PRIVACY AND ONLINE TARGETING TRANSPARENCY**

Another important area with implications for counternarrative practice where regulators increasingly require transparency, exists around user-facing disclaimers of online targeting[76]. A self-regulatory EU Code of Practice on Disinformation contains provisions on transparency about political and issue-based advertising with an intention to enable users' understanding of the reasons for being targeted, but the recent monitoring report[77] indicates very slow progress in compliance. The GDPR is falling short of holding platforms into account on that matter, as data protection authorities lack the tools and resources to enforce it[78]. Researchers and activists have also repeatedly asked for improvements in the ad transparency.

- **ALGORITHMIC TRANSPARENCY AND THE IMPACT OF ML/AI**

Despite ubiquitous data harvesting taking place, counternarrative practitioners do not generally have access to data on the prevalence, scale, volume, size of networks, patterns of network distribution, which would allow to assess the impact of different types of harmful content on the society. In this context, it is also near impossible to estimate the impact of counternarratives.

[71] such as the Law on Countering Online Hatred, more commonly known as the Avia law in France, or The Network Enforcement Act, known as NetzDG in Germany

[72] Keller, D. (2018). Inception Impact Assessment: Measures to Further Improve the Effectiveness of the Fight Against Illegal Content, *Stanford Center for Internet and Society, Stanford Law School*. Available at: https://cyberlaw.stanford.edu/files/publication/files/Commission-Filing-Stanford-CIS-26-3_0.pdf

[73] Douek, E., (2020) The Rise of Content Cartels *Knight First Amendment Institute: Columbia University*, Available at: https://knightcolumbia.org/content/the-rise-of-content-cartels

[74] Protecting Free Expression in the Era of Online Content Moderation, (2019) *Access Now*, Available at: https://www.accessnow.org/cms/assets/uploads/2019/05/AccessNow-Preliminary-Recommendations-On-Content-Moderation-and-Facebooks-Planned-Oversight-Board.pdf

[75] Santa Clara Principles on Transparency and Accountability in Content Moderation, Available at: https://santaclaraprinciples.org/

[76] Leerssen, P. *The Soap Box as a Black Box*.

[77] ERGA Report on disinformation: Assessment of the implementation of the Code of Practice (2020), The European Regulators Group for Audiovisual Media Services, Available at: https://erga-online.eu/wp-content/uploads/2020/05/ERGA-2019-report-published-2020-LQ.pdf

[78] Dobber, T. & Ó Fathaigh, R. & Zuiderveen Borgesius, F. J. (2019). The regulation of online political micro-targeting in Europe. *Internet Policy Review*, *8*(4). Available at: https://policyreview.info/articles/analysis/regulation-online-political-micro-targeting-europe

Many concerns about privacy in relation to online targeting will only be exacerbated with the advance of AI, as AI developers require access to large quantities of user data as training data. Mozilla warns that the use of AI in many consumer products and services "creates significant collective risks related to bias, misinformation, and corporate surveillance."[79]

Journalists, researchers and civil society projects such as Algorithmwatch[80] and Algotransparency[81] have ambitiously attempted to uncover the workings of recommendation systems.[82] In this way, algorithms can be made more transparent by exposing both their human and the technological aspects[83], explaining to the audience how they operate in a non-neutral way.

Supporting media literacy

Media literacy is often promoted as a broad-based, [non-technological] solution against polarising discourse and the challenge of disinformation online, as well as to extremist radicalisation. But seeing as both consumers and creators of problematic online narratives can be very digitally literate, this approach cannot only focus on the media, or technology[84]. Given the aforementioned importance of pre-existing, potentially polarised or extreme beliefs, it has been advised that critical digital literacy, should focus on both improving the understanding of technology and the media, but also on building skills to engage in both institutional and non-institutional politics, such as for example, involvement in alternative media or activism[85], and developing democratic norms in a way that builds individual resilience to violent extremism[86].

At the same time, given the aforementioned risks associated with irresponsible technology use, and the sensitivity of the data collection that could jeopardize the target audience, critical digital literacy skills are also needed by practitioners around privacy and the handling of personal data[87].

Supporting the development of counternarratives

Counternarratives and communication initiatives more broadly appear alongside the previously listed government approaches to addressing threats to national security, public safety, and liberal democracy, exacerbated by the use and abuse of communication technology.

[79] Ricks, B., and Surman, M. (2020) Creating Trustworthy AI. *Mozilla* Foundation. Available at: https://drive.google.com/file/d/1LD8pBC-cu7bkvU-9v-DZEyCmpWED7W7Z/view

[80] Duportail, J., Kayser-Bril, N., Schacht, K. and Richard, É (2020) Undress or fail: Instagram's algorithm strong-arms users into showing skin *Algorithm Watch*. Available at: https://algorithmwatch.org/en/story/instagram-algorithm-nudity/

[81] Available at: https://algotransparency.org/

[82] Leerssen, P. The Soap Box as a Black Box.

[83] Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *new media & society*, *20*(3), 973-989. Available at: https://doi.org/10.1177/1461444816676645

[84] Banaji, S *et al* Whatsapp Vigilantes

[85] Polizzi, G. (2018, August 29). Critical digital literacy: Ten key readings for our distrustful media age. *Parenting for a Digital Future, the London School of Economics.* Retrieved from https://blogs.lse.ac.uk/parenting4digitalfuture/2018/08/29/critical-digital-literacy/

[86] Swedish Media Council. (2013). Pro-Violence and Anti-Democratic Messages on the Internet.

[87] Latonero, M., Hiatt, K., Napolitano, A., Clericetti, G., Penagos, M. (2019). *Digital Identity in the Migration & Refugee Context: Italy Case Study* [Report]. Data and Society. https://datasociety.net/library/digital-identity-in-the-migration-refugee-context/

In the context of counterterrorism strategies, counternarratives have been used as far back as 2005[88,89] but their prominence rose alongside the rise of social media platforms. The emphasis on the development of counternarratives, as part of government counterterrorism strategies has accelerated post-2014, with ISIS' ascent[90]. In 2017, the UN's Comprehensive International Framework to Counter Terrorist Narratives was adopted, structuring recommendations for state action around legal and law enforcement measures (regulation and prosecution); public-private partnerships; and the development of counter-narratives[91].

Public private initiatives and multi-stakeholder fora, such as the, EU Internet Forum, Global Internet Forum to Counter Terrorism (GIFCT), Tech Against Terrorism, EU RAN, and all major international organisation with security roles and mandates, starting from the UN agencies such as UNESCO[92] and UNDP, OSCE[93], Council of Europe have heavily promoted the idea that counternarratives are an effective way to prevent violent extremism and terrorism and invested in campaign development and capacity building.

From 2016 the same institutions, bodies, and agencies have also been actively involved in countering foreign influence and mis/disinformation, most of them also having extant regulations, policies and programmes in place related to hate speech.

When it comes to online hate speech, the EU's Code of Conduct highlights the role of civil society "in the field of preventing the rise of hatred online, by developing counter-narratives promoting non-discrimination, tolerance and respect, including through awareness-raising activities."[94] Meanwhile, IT Companies are "to intensify their work with CSOs to deliver best practice training on countering hateful rhetoric and prejudice and increase the scale of their proactive outreach to CSOs to help them deliver effective counter speech campaigns."[95] The European Commission, in cooperation with Member States, is to contribute to this endeavour by taking steps to map CSOs' specific needs and demands in this respect.

The more recently formulated strategy for fighting disinformation in the EU is focused on factchecking and quality journalism. Only in one footnote, did the report of the Independent High-level Group on

[88] Casebeer W., and Russell J. A.., (2005) "Storytelling and Terrorism: Towards a Comprehensive 'Counter-Narrative Strategy,'" *Strategic Insights* 4(3), pp. 1–16. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.116.7615&rep=rep1&type=pdf

[89] Corman, S (2008) "Understanding the Role of Narrative," pp36–43 in Corman S R., Tretheway, A., and Goodall Jr, H. L. (Eds), *Weapons of Mass Persuasion: Strategic Communication to Combat Violent Extremism* (New York: Peter Lang Publishing);Goodall Jr, H.L. (2010) Counter-Narrative: How Progressive Academics Can Challenge Extremists and Promote Social Justice. London: *Taylor and Francis*; Halverson, J., Goodall H. L.,., and Corman, S., (2011) Master Narratives of Islamic Extremism. London *Palgrave Macmillan*

[90] The Hedayah-ICCT organised expert meeting minutes Developing Effective Counter-Narrative Frameworks for Countering Violent Extremism, provides a unique snapshot of the moment in time and the thinking behind CVE CN See: Developing Effective Counter-Narrative Frameworks for Countering Violent Extremism: Meeting Note (2014) *Hedayah and International Centre for Counter-Terrorism*. Available at: https://www.icct.nl/download/file/Developing%20Effective%20CN%20Frameworks_Hedayah_ICCT_Report_FINAL.pdf

[91] Available at: https://www.un.org/sc/ctc/news/document/s2017375-comprehensive-international-framework-counter-terrorist-narratives/

[92] Preventing Violent Extremism Worldwide, UNESCO. Available at: https://en.unesco.org/sites/default/files/unesco_in_action-pve_worldwide-en.pdf

[93] Youth Engagement to Counter Violent Extremism and Radicalization that Lead to Terrorism Report on Findings and Recommendations (2013) *OSCE Secretariat*. Available at: https://www.osce.org/files/f/documents/c/b/103352.pdf; Holmer. G *et al* (2018) The Role of Civil Society in Preventing and Countering Violent Extremism and Radicalization that Lead to Terrorism: A Guidebook for South-Eastern Europe. *OSCE Secretariat.* Available at: https://www.osce.org/files/f/documents/2/2/400241_1.pdf

[94] EU Code of Conduct On Countering Illegal Hate Speech Online (2016), *EU Commission*. Available at: https://ec.europa.eu/info/sites/info/files/code_of_conduct_on_countering_illegal_hate_speech_online_en.pdf

[95] EU Code of Conduct On Countering Illegal Hate Speech Online (2016), *EU Commission*.

Fake News and Online Disinformation outlining a multi-dimensional approach the EU should take, refer to research suggesting that detailed counter-messages and alternative narratives are often more effective than corrections in countering disinformation[96].

Despite appearances and despite overall endorsement and commitment to this practice under the above-mentioned instruments, the promotion of counter-narratives as a government and platform sponsored activity is not an uncontested policy area. It raises numerous ethical and methodological concerns, some of which are discussed in this report. In reference to counternarratives, the UN Rapporteur's 2018 report stated that "pressure for such approaches runs the risk of transforming platforms into carriers of propaganda well beyond established areas of legitimate concern."[97]

Way forward: responsible use of technology

The current state of research shows a lack of decisive evidence on the actual impact of media exposure to harmful content on audiences' behaviour, and on processes such as radicalisation and polarisation.[98,99,100,101,102] Furthermore, there is a possibility that effects and reach of such content may have been overstated, compared to the reach of mainstream and trusted news[103]. Overhyping the effects associated with disinformation and extremism has had some negative consequences, leading some governments to implement restrictive policies that serve a partisan agenda[104].

Automated solutions and technological tools that claim to do "good" have been hailed as simple and neutral solutions to prevent the dissemination of violent extremist speech. This is mostly prevalent in the grey literature that constitutes most of the research on CVE. Yet this grey literature may be funded by actors who have vested commercial interests in these solutions[105]. Overall, there is growing evidence that this technology is far from neutral and that its claim that it can predict (and prevent) social behaviours might be overstated[106].

[96] Report of the independent High level Group on fake news and online disinformation: A multi-dimensional approach to disinformation (2018) *European Commission*. Available at: http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=50271

[97] Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (2018) *Human Rights Council, UNGA*. Available at: https://undocs.org/A/HRC/38/35, p.8

[98] Tucker, J., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D. & Nyhan, B., (2018) Social Media, Political Polarization, And Political Disinformation: A Review Of Scientific Literature. *Hewlett Foundation* Available at: https://hewlett.org/library/social-media-political-polarization-political-disinformation-review-scientific-literature/

[99]Ferguson, K. (2016) Countering violent extremism through media and communication strategies: A review of the evidence *Partnership for Conflict, Crime & Security Research* Available at: http://www.dmeforpeace.org/peacexchange/wp-content/uploads/2018/10/Countering-Violent-Extremism-Through-Media-and-Communication-Strategies.pdf

[100] Odağ Ö., Leiser, A., Boehnke, K., (2019). Reviewing the role of the internet in Radicalisation processes, *Journal for Deradicalisation*, 21. pp261-300 Available at: https://journals.sfu.ca/jd/index.php/jd/article/view/289

[101] Milt, K *et al* (2017) Countering Terrorist Narratives, *Policy Department for Citizens' Rights and Constitutional Affairs* Available at: https://www.europarl.europa.eu/RegData/etudes/STUD/2017/596829/IPOL_STU(2017)596829_EN.pdf

[102] Fletcher, R., and Jenkins, J. (2019) Polarisation and the news media in Europe, *European Parliamentary Research Service* Available at : https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2019-03/Polarisation_and_the_news_media_in_Europe.pdf

[103] Fletcher, R., Cornia, A., Graves, L., & Nielsen, R. K. (2018). Measuring the reach of "fake news" and online disinformation in Europe. *Reuters institute factsheet*. Available at: https://reutersinstitute.politics.ox.ac.uk/our-research/measuring-reach-fake-news-and-online-disinformation-europe

[104] Lim, G. (2020) Securitize/Counter-Securitize: The Life and Death of Malaysia's Anti-Fake News Act, *Data and Society*, Available at: https://datasociety.net/wp-content/uploads/2020/03/Counter-securitize.pdf

[105] Van Eert et al *Developing a social media response to radicalization*

[106] Nelson, L. K. (2019). To measure meaning in big data, don't give me a map, give me transparency and reproducibility. *Sociological Methodology*, 49(1), 139-143. Available at: https://journals.sagepub.com/doi/full/10.1177/0081175019863783

Additionally, the collateral risks attached to the use of technology may further exacerbate conditions conducive to the rise of extremism and further diminish trust. There have been cases of technology, developed initially to prevent terrorist recruitment, that has been used for broader societal surveillance purposes which, if used by the wrong hands, can present serious risks to democracy, by targeting journalists[107] or immigrants[108]. Given the fact that transparency is an important part of the solution against the spread of harmful content and given the potential harmful effects of opaque technological tools, both governments, platforms, and counternarratives practitioners should promote trust by adhering to high transparency standards.

When designing effective strategies for counternarratives campaigns, recent findings and advances in the theory, the quickly evolving technologies powering the information ecosystem with its affordances and developing rules and regulation to reign in the most visible challenges to security and democracy, need to be considered. The next part of this report will focus on how to effectively harness the aforementioned insights in order to enhance the impact of narratives.

[107] Srivastava, M (2019) WhatsApp voice calls used to inject Israeli spyware on phones *The Financial Times* Available at: https://www.ft.com/content/4da1117e-756c-11e9-be7d-6d846537acab

[108] Waldman, P., Chapman, L., Peterson, J. (2018) Palantir targets commercial market with tools for War on Terror *Privacy International*. Available at: https://privacyinternational.org/examples/2752/palantir-targets-commercial-market-tools-war-terror

The Art and Craft of Counternarrative Campaigns

The objective in this part of the report is to provide recommendations based on the evidence and discussion presented in the previous sections. The report also highlights good practices and includes lessons from counter hate speech, misinformation, extremism projects and initiatives in various parts of the world. Suggestions are aimed at practitioners working to create, implement and measure the impact of a counternarratives campaigns and at the donors and policymakers who guide and support them.

The recommendations are primarily applicable to government funded strategic communication campaigns aimed at achieving public policy goals – such as the protection of national security, protection of democracy – implemented by civil society organisation in the European Union.

They are drafted to be complementary to and build on the available guidelines and existing practices,[109] which lay out the standard elements and processes of campaign development, from setting goals, understanding audiences, developing messages, selecting messengers and medium, and measuring impact.

Providing Roadmaps: Invisible Parts of the Campaign [Objectives, Audience]

This report focuses on the government and platform sponsored and supported counternarratives implemented as part of national security strategies. As mentioned in the previous part of the report, a "pivot" to counternarratives involved the inclusion of an assortment of community groups and organisations in the implementation of CT/CVE Strategies. In this context the communication role is effectively outsourced to the civil society organisations. Business/brand driven or activist driven initiatives and movements that offer counternarratives as part of their agenda, or as their marketing or community building strategies are not the focus of this report.

As such, the objectives at the highest level are defined by the government and those who fund counternarrative initiatives have the responsibility of engaging with the implementing organisations on the strategic level. This engagement should include discussing the expectations and limitations of scale, depth, and duration of individual campaigns and interventions, specifically outlining which effects and impacts can be realistically achieved. The donors should also step up support in the measurement of these effects, through data collection and reporting at an appropriate level, so findings can be compared or aggregated.

The speech being promoted through counternarrative campaigns is government sponsored, and there are implications of government funding despite the use of usual caveats, disclaimers, and in particular in cases when funding source information is concealed. There are different concerns and risks for

[109] Such as recommended by the EU Radicalisation Awareness Network, Hedayah, and others

recipient organisations depending on whether the funds come from the domestic or foreign government, or from the EU.

Campaign objectives

Even though campaign's objectives are expected to be specific, hyperlocal, and contextualised, they need to be justified in the context of high-level goals and government priorities. The standard approach is that each project must have a definition of what constitutes violent extremism in the local context[110] and its own theory of change. This still gives a very broad room to define objectives anywhere from tackling ideologies to addressing root causes, from preventing specific outcomes to decreasing vulnerability or risk, to intervening in processes that are not very well understood and highly individualised, like radicalisation, which makes aggregation or comparison of effects difficult.

Sometimes the overall, high-level objectives campaigns aspire to, are preventing or countering specific outcomes, like terrorist violence or hate crimes, sometimes they refer to intervening in processes, such as polarisation or radicalisation. At times they are only addressing them in the context of specific extremist group ideology, propaganda, and recruitment messaging.
When objectives are not clearly defined at the level of donor, there is the trickle-down effect of ambiguity to audience and messenger choice, message creation, dissemination strategy, and measurement.

Roles and expectations

When governments' security objectives are tied to specific ideologies, their efforts focus predominantly on jihadi terrorism, the disproportionate targeting of minorities by counternarratives may lead to stigmatisation, contribute to radicalisation through increasing sense of victimisation on one side and increasing the resonance of anti-Muslim narratives among the general population[111].
In both the UK and the U.S., the government strategies to counter violent extremism have been criticised for singling out Muslim communities, using vulnerability assessment criteria applicable to almost anyone, generating many "false positives."[112] They both essentialised and homogenised audiences[113] in a deceptive way that further jeopardised their safety[114] and opened the door for unjustified surveillance practices by private technology companies[115].

According to Anita Ernstorfer, the author of UNDP's reflection paper on conflict sensitivity in preventing violent extremism, "[M]any partners and communities do not relate to the language and

[110] Holdaway, L., and Simson, R., (2018) Improving the impact of preventing violent extremism programming: A toolkit for design, monitoring and evaluation, UNDP / International Alert, Available at: https://www.undp.org/content/dam/undp/library/Global%20Policy%20Centres/OGC/PVE_ImprovingImpactProgrammingToolkit_2018.pdf

[111] McDonnell, T. & Bail, C. & Tavory, I., (2017). A Theory of Resonance. Sociological Theory. 35. 1-14. 10.1177/0735275117692837

[112] Patel, F., and Koushik, M., (2017) Countering Violent Extremism, *Brennan Center for Justice*. Available at: https://www.brennancenter.org/sites/default/files/2019-08/Report_Brennan%20Center%20CVE%20Report_0.pdf

[113] For more on why homogenising audiences is problematic, refer to: Mohanty, C. (1988). Under Western eyes: Feminist scholarship and colonial discourses. *Feminist review*, *30*(1), pp61-88. Available at: https://journals.sagepub.com/doi/abs/10.1057/fr.1988.42

[114] Shafi, A and Qureshi A. (2020) Stranger than Fiction: How 'Pre-Crime' Approaches to "Countering Violent Extremism" institutionalise Islamophobia; A European Comparative Study, *Transnational Institute*. Available at: https://www.tni.org/files/publication-downloads/web_strangerthanfiction.pdf

[115] Why Countering Violent Extremism Programs Are Bad Policy (2019). *Brennan Centre for Justice*. Available at: https://www.brennancenter.org/our-work/research-reports/why-countering-violent-extremism-programs-are-bad-policy

framing around 'extremism' and perceive labelling certain people and groups as 'radical' as insulting, exclusionary or missing the point of the issues at hand altogether."[116]

In the design phase of a campaign, roles and expectations in the relationship between the implementer and the sponsor of campaign, and their collective responsibility towards the target audience needs to be carefully thought through and spelled out.
An audience's trust is the most valuable asset and both the precursor to, and the objective of, any communication effort. Social movements and community-based organisations are rich in access and credibility with their audiences and have an authentic voice on issues of concern to the audience. They face the challenge of balancing their mission and objectives, the need to protect their reputation, against the donor's expectations, and the funding conditions.

The proponents of government-sponsored counternarratives delivered by civil society recognised early on that training and supporting established local community-based organisations with "built-in" audiences to convey the message, is not only more effective, it can also facilitate long term impact and sustainability beyond the duration of one campaign.

Availability of funding for these types of campaigns meant the increase of the number and variety of civil society stakeholders involved in the implementation of government strategies to counter violent extremism. The ideal model epitomizing this approach are small scale campaigns that capitalise on already established audiences. Implemented by community-level organisations, youth clubs, women's group messaging to their members, relying on influential, authentic, and credible members of the community, such as priests,[117] imams, mothers, teachers as messengers. They complement in-person interactions with online messaging. For these kinds of campaigns, audience definition is a straightforward matter.

But many organisations implementing government sponsored counternarrative campaigns focus on creation of content and its distribution to audiences that do not already belong to their constituency and which they do not have real-world ties to. For them, defining who their audience is constitutes an essential step. Regardless of objectives, methods, and resources, this process should adhere to the following guiding principles.

- Avoiding ethnic or religion-based profiling.
- Segmenting audience according to their positions, attitudes on issues and according to communication practices and habits.
- Relying on ethically conducted audience research[118].
- Respecting privacy and not jeopardising audience's safety.
- Assuming a primarily two-way communication focusing on dialogue instead of distribution.

[116] Ernstorfer A., (2019) *Conflict Sensitivity in Approaches to Preventing Violent Extremism*

[117] Kaal, H. (2016). Politics of place: political representation and the culture of electioneering in the Netherlands, c.1848–1980s. *European Review of History*, *23*(3), 486–507. Available at: https://doi.org/10.1080/13507486.2015.1086314

[118] Brown, R. H. (2016). *Defusing hate: a strategic communication guide to counteract dangerous speech*. United States Holocaust Memorial Museum, Simon-Skjodt Center for the Prevention of Genocide.

Problem Solving Through Stories: Visible Parts of the Campaign [Message, Messenger, Medium, and Audience Engagement]

The overall purpose of strategic communication, generally speaking, is to develop a message and target it in line with the campaign's objectives, to suggest a course of action to an audience that was its intended recipient. Having the audience follow that suggestion is part art and part science - both of which are discussed in this section.

It has already been mentioned that trust is a prerequisite to any communication effort. To build long-term trust, audiences should be involved in every step of the campaign design, as co-creators, co-producers, communicators, and distributors of campaign message. Trust building requires adopting the following values: authenticity, transparency, positivity, diversity, consistency and shared mission[119]. Counternarrative campaigns should formulate a clear set of objectives with transparent intentions, arrived at through a participatory process, discussing the values and course of action that are being promoted to a target audience.

When engaging with audiences, counternarrative creators should promote understanding of risks and consequences of false, manipulative divisive communication, knowledge and skills of how to resist, stories showcasing benefits of living in a society that adheres to values like freedom and tolerance, democracy and openness, while at the same time recognising audience's concerns, difficulties, and frustrations.

The points below discuss in more detail the application of these values and seek to address the gaps in current practices, from message development, to managing audience reactions.

Storytelling and listening [Message]

Message development should begin with listening to diverse points of views among the audience on the issues so as to construct authentic communication and adapt broader imaginaries to local contexts and concerns. Shaping messages in counternarratives should always be done hand-in-hand with members of the target audience. Prior to creating campaign messages and calls to action, it can be useful to consider the discourses underpinning the content and messages that are meant to be countered. Actually, listening to audience members on sensitive, politically charged, divisive issues can be complementary to social listening, which might omit important views and voices. This will ensure that the appropriate language, slogans, and pain points are considered, as part of a broader ecosystem of speech that is not just made of social media discourses and news narratives, but also of in-person interactions and lived experiences.

There is a need for counternarratives to engage with familiar and recognisable political and social discourses to be able to resonate with audiences and through stories that appeal to emotions and that are not overtly persuasive. While some members of the audience might be convinced by debunks and factchecks, this may not be enough.

[119] McKinley, E G., and Green-Barber, L., (2019) Engaged Journalism: Practices for Building Trust, Generating Revenue and Fostering Civic Engagement, *Impact Architects*. Available at: https://mediaimpact.issuelab.org/resource/engaged-journalism-practices-for-building-trust-generating-revenue-and-fostering-civic-engagement.html

A time-consuming method[120], often consisting of short, emotionless reports, factchecking is defined as "the practice of systematically publishing assessments of the validity of claims made by public officials and institutions with an explicit attempt to identify whether a claim is factual"[121]. While increasingly used to correct the effects of misinformation, this strategy's results have been ambiguous. It has been demonstrated that factchecks are effective with an audience mostly when the latter already trusts the messenger[122], and when it does not affect too much their pre-existing beliefs[123]. Otherwise, given the complexity of how beliefs are shaped, simply being exposed to corrected facts is not enough, and may sometimes even backfire[124]. At the same time, one study has shown that the presence of moral-emotional language in social media messages did increase their diffusion[125], particularly within groups with similar ideologies, and another has found that novelty, negative emotions and politically oriented topics participated in the diffusion of falsehoods[126].

For counternarrative campaigns, what these learnings on how information travels seem to indicate, is the importance of addressing issues in terms of their emotional resonance with audiences. The participatory creative design of counternarrative campaigns should serve the purpose of incorporating not just facts, but also affect and emotion, and how these are linked to group dynamics and ideological beliefs.

The framing theory, which focuses on how social movements construct messages and present claims to the target audience in the process of mobilisation, offers useful insights during the process of message development.[127] Experienced grievances and anxieties, as a result of economic, cultural, and social changes are ripe for exploitation by extremists and populists, who are channelling them into fear and resentment of designated culprits, such as elites or ethnic, racial and religious minorities.[128] The search for identity, meaning, or justice is typically also understood as motivating individuals and driving appeal of extremist messaging and propaganda in the radicalisation process.[129]

The mobilisation success of social movements is predicated on the resonance of frames through which they are diagnosing the problem, present solutions and motivate public to action, with the audience's

[120] Graves, D. (2018). Understanding the promise and limits of automated fact-checking.
[121] Walter, N., Cohen, J., Holbert, R. L., & Morag, Y. (2020). Fact-checking: A meta-analysis of what works and for whom. *Political Communication*, *37*(3), 350-375.
[122] Margolin, D. B., Hannak, A., & Weber, I. (2018). Political fact-checking on Twitter: When do corrections have an effect?. *Political Communication*, *35*(2), 196-219.
[123] Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological science in the public interest*, *13*(3), 106-131.
[124] Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, *32*(2), 303-330.
[125] Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, *114*(28), 7313-7318.
[126] Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146-1151.
[127] Snow, D., and Benford, R. (1988). Ideology, Frame Resonance and Participant Mobilization. *International Social Movement Research*. 1. pp197-217. Available at: https://ssc.wisc.edu/~oliver/SOC924/Articles/SnowBenfordIdeologyframeresonanceandparticipantmobilization.pdf
[128] Bonikowski, B. (2017). Ethno-nationalist populism and the mobilization of collective resentment. *The British Journal of Sociology*. 68 Suppl 1. Available at: https://onlinelibrary.wiley.com/doi/10.1111/1468-4446.12325
[129] Feddes, A. R., Nickolson, L., & Doosje, B. (2015). Triggerfactoren in het radicaliseringsproces [Trigger factors in the radicalisation process]. *Social Stability Expertise Unit, Dutch Ministry of Social Affairs & Employment* Available at: https://www.socialestabiliteit.nl/documenten/publicaties/2015/10/13/triggerfactoren-in-het-radicaliseringsproces
See also van Eerten *et al* (2017). Developing a social media response to radicalization: The role of counter-narratives in prevention of radicalization and de-radicalization.

lived experiences and their prevalent understanding of what is happening.[130] Similarly counternarratives should seek to provide a competing, alternative framing to problems and frustrations that citizens are facing. The framing effort to develop resonant messages can thus be imagined as collective search for ideas that offer solutions to problems experienced by audiences better than existing or competing narrative frames.[131]

Storytellers, protagonists, and influencers [Messengers]

Individual and group stories, fictional or testimonies, have been used in social communications for a long time, and can be powerful at rendering vivid accounts of what an alternative to violent extremism or hate might look like.

In the case of campaigns using accounts of survivors, victims, or families linked to violent extremism, individual testimonies have been demonstrated to have a positive impact, thus humanising and destigmatising that topic[132]. Showcasing migrant stories, who are often depicted by mainstream narratives as a group, or in impersonal ways, these individualised accounts have been shown to generate empathetic responses.[133] Extrapolating the narratives to groups can also prove effective, for example in order to avoid a framing that may show the story as an exceptional event, and to avoid individualising the responsibility of vulnerable people.[134]

When tapping into these personal histories, ethical considerations are paramount so as to avoid re-traumatization and exposing them to other risks. Power dynamics, prevailing attitudes, situational factors should all be considered to protect privacy and security of campaign messengers. Fictional accounts and composite characters based on thorough audience research are also an option in cases where risks of harm are considered too high.

Persuasive effects of narratives are stronger when the audience identifies with messengers or protagonists[135]. Evidence mentioned in the first part of the report suggested that close and trusted contacts tend to wield more influence over people's beliefs and behaviours.

On the other hand, the use of social media influencers or celebrities as messengers is a popular tactic, but risks and benefits of engaging these super spreaders should be carefully evaluated. The "right" influencer should be selected based on their credibility and potential to persuade the target audience in line with the overall goals of the campaign. For broad awareness raising campaigns, or for mobilisation of specific audiences, using social media influencers and celebrities who are known for their position on issues or for their commitment to values can increase reach and engagement significantly, but it can also increase polarisation, pushback, and make managing a potential fallout

[130] Snow, D and Benford, R. (1988). *Ideology, Frame Resonance and Participant Mobilization*. pp197-217.

[131] McDonnell, T., Bail, C. and Tavory, I. (2017). A Theory of Resonance. *Sociological Theory*. 35. pp1-14. Available at: https://journals.sagepub.com/doi/10.1177/0735275117692837

[132] Rrustemi, A. (2020). The Lifestories Method Handbook. Available at: https://hcss.nl/news/lifestories-method-handbook

[133] Chouliaraki, L., & Stolic, T. (2017). Rethinking media responsibility in the refugee 'crisis': A visual typology of European news. *Media, Culture & Society*, *39*(8), pp1162-1177. Available at: https://journals.sagepub.com/doi/abs/10.1177/0163443717726163

[134] Chouliaraki, L., & Stolic, T. Rethinking media responsibility in the refugee 'crisis': A visual typology of European news.

[135] Meng, C., Bell, R A., and Taylor, L D (2016): Narrator Point of View and Persuasion in Health Narratives: The Role of Protagonist–Reader Similarity, Identification, and Self-Referencing, Journal of Health Communication, *Journal of Health Communication* [International Perspectives] 21: 8 pp908-918. Available at: https://www.tandfonline.com/doi/abs/10.1080/10810730.2016.1177147

impossible. For communicating counter-attitudinal messages, or changing perceptions however, influencers who are not perceived as overtly partisan or biased are more suitable and they can be identified with the help of social network analysis.

Media and technology choices

Effective and responsible use of digital communication requires the understanding of technology and advanced digital literacy, to allow responsible balancing of campaign objectives with security and ethical considerations.

The choice of social media platforms as a medium for a counternarrative campaign is far from straightforward. The message dissemination conundrum has been complicated by the increased use of different social media platforms, on which audiences may behave differently, with different habits and discourses. At the same time, the aforementioned informational and algorithmic architecture provided by platforms may skew audience's choices in a significant yet opaque way. Individually, people also use social media with different objectives in mind, and some people may be entirely absent from it.

Thorough audience research in the campaign design phase, the mapping of media repertoires, social media use and behaviours should be conducted. Techniques of social network analysis can be applied to examine the ties between the target audience members, their connection to other groups, to understand who the influencers and connectors are, but also those are silent or excluded from it[136].

Additionally, given the aforementioned role of transmediality, and the way narratives travel via several media, beyond social media and the internet, the use of traditional media as part of the campaign strategy may bring significant benefits. Print media, broadcast TV, and radio still gather a significant number of viewers and listeners, local news also has the power to reinforce communities and civic sense[137,138]. In some countries and contexts, posters, handbills, or flyers are extremely important vectors in the larger media ecosystem, and often go on to influence how people engage with digital content.[139]

Campaigns should be discerning about the tools they chose to use and conduct a careful examination of the consequences of using invasive technological solutions, and consider whether the benefits outweigh their harms, in a way that is protective of the target audience and respectful of their consent and vulnerabilities[140]. This includes transparency about how campaign messages are promoted or targeted online, and in terms of funding.

[136] Lutkenhaus, R O., Jansz, J., Bouman, M , (2019) Mapping the Dutch vaccination debate on Twitter: Identifying communities, narratives, and interactions, Vaccine: X, Vol 1, 100019, Available at: https://doi.org/10.1016/j.jvacx.2019.100019.

[137] McKinley, E G., and Green-Barber, L. (2019) *Engaged Journalism*.

[138] Kornbluh, K., & Goodman, E. P. (2020). Safeguarding Digital Democracy.

[139] O'Neill, C. S., & Boykoff, M. (2012). The role of new media in engaging the public with climate change. In *Engaging the public with climate change* (pp. 259-277). Routledge.

[140] Privacy International. (2018). *The Humanitarian Metadata Problem: «Doing no Harm» in the Digital Era.* https://privacyinternational.org/sites/default/files/2018-12/The%20Humanitarian%20Metadata%20Problem%20-%20Doing%20No%20Harm%20in%20the%20Digital%20Era.pdf

Message spread and audience reactions

Campaign creators are often predominantly concerned with content production. Equally important to consider, and often overlooked in the design phase (and treated as an afterthought), is how audience reactions and content sharing behaviour will factor into message delivery and processing. There is a concern that audience comments might distort the message. A related, but separate concern is about the context audience views the message, which is hard to control on platforms and might be crucial for how the information is processed.

Some effective counter-speech efforts actually do not focus on production of new content, but on intervening in the discussions about other content to lower its toxicity and incivility. This has been through coordination of a large number of volunteers[141] and also using automatic response generation methods[142].

For campaigns that produce original content, the publishing of content is not the final step. Once a counternarrative content carrying the campaign's message is developed and published, most creators think about audience reactions, including sharing, only in the context of evaluation. But responding to audience reactions requires a more strategic treatment in the campaign planning phase and should always be a part of message development.

Particularly important in this context is careful mapping and identification of spoilers and those who will likely oppose a message. The reactions and types of conversations campaign content sparks should ideally be tested through focus groups before the launch, after a careful consideration of participant selection criteria. Some content and messages might be too controversial and potentially toxic and might be better presented to audiences in a more controlled environment, where it is possible to respond to concerns and correct misinterpretations. Uncontrolled viral spread does not always bring net positive effects.

There is a need to think carefully about moderating audience reactions as most platforms where content is published enable and encourage engagement through comments. Counternarratives may evoke hateful or even extremist comments themselves[143], particularly among the audiences that hold opposite views and beliefs. While disagreement and criticisms are often an indicator that the message is reaching the target audience, in order to not undermine the efforts of campaign creators, sensitive and strategic moderation efforts of audience comments by the campaign are needed. Campaigns should be prepared to respond to those comments that are not considered harmful or toxic and promptly delete or mute those that are.

Audiences might engage with content through the campaign channels, but the impact of an audience further sharing campaign content in the communities opposing the values, messages, and goals

[141] See for example: https://www.ichbinhier.eu/

[142] Qian, J., Bethke, A., Liu, Y., Belding-Royer, E.M., & Wang, W.Y. (2019). A Benchmark Dataset for Learning to Intervene in Online Hate Speech. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing,* Association for Computational Linguistics. pp4755–4764 Available at: https://www.aclweb.org/anthology/D19-1482

[143] Ernst, J., Schmitt, J. B., Rieger, D., Beier, A. K., Vorderer, P., Bente, G., & Roth, H.-J. (2017). Hate beneath the counter speech? A qualitative content analysis of user comments on YouTube related to counter speech videos. Journal for Deradicalization, 10, 1–49. Available at: https://journals.sfu.ca/jd/index.php/jd/article/view/91

promoted through the campaign should also be considered. The level of detail in the data available on how and where content is shared in general varies by platform and depends on user privacy settings, making it difficult to accurately gauge spread of the narrative, particularly across platform and across languages, but it is worth the effort to closely track this not only for the purposes of evaluation, but to intervene in line with the strategy.

Beyond individual campaigns, audience reactions can be factored in by policymakers to view counternarratives as a continual and on-going process. Participatory research mechanisms can be developed whereby audiences over time have a stake in the process of developing effective and sustainable counternarratives.

The story of impact [Measurement, Evaluation]

Creators wishing to measure attitudinal or behavioural change as the indicator of impact, should be aware of selective exposure and other cognitive biases and difficulties of attributing individual level media effects, and the privacy concerns this raises.

The choice of impact measurement needs to match the campaign strategy, the overall objective and the intervention logic. While it may be difficult to measure the high-level impact for small campaigns, isolating their effect from other influences and factors, the persuasiveness of campaign messages with target audiences can almost always be assessed, using opinion polls or focus groups.

A useful approach to help with campaign metrics is using the so-called markers - "unique identifiable elements of messages such as new words, phrases or novel behaviours that ideally model new realities to break oppressive power structures in society."[144] They are visible indicators of the behavioural uptake, i.e. of attainment of campaign goals, while also serving as a tag enabling tracking of online conversations by campaign evaluators and attribution of campaign effects.

Approaches that rest on a mix of quantitative and qualitative indicators and both online and offline data collection with the audience's consent, are preferred.

Social media performance figures should not be the sole or the most important impact metric. It has been mentioned already that measures of reach, engagement, clicks and views provided by social media platforms are weak indicators of social change and the definitions of these metrics have been criticised for their fickleness and opaqueness. Sometimes inflated by platforms[145], they are also poor indicators of behavioural change as the motivations behind a click or a view are still unclear. Most importantly, they ignore political and social relations, with which counternarratives aim to engage. Therefore, more weight should be given to metrics that are relevant and specific to the campaign's defined objective, than to the standardised platform metrics. It is also important to contextualise them, by comparing with the figures gained by harmful content, or with the actual size of the target audience.

[144] Lutkenhaus R.O., Jansz, J., and Bouman M., (2019). Toward spreadable entertainment-education: leveraging social influence in online networks. *Health Promotion International* Oxford: OUP. Available at: https://academic.oup.com/heapro/advance-article/doi/10.1093/heapro/daz104/5588513

[145] Allcott, H. *et al* (2020) *The Welfare effects of social media*

A triangulation of methods and data sources will always increase the reliability of findings[146]. For example, a combination of survey data, collected at different points, a sentiment analysis to analyse online content such as social media posts and comments and focus groups or, in-depth interviews with representatives of target audience. Participatory ethnographic action research, described in detail and used effectively by UNESCO, can provide useful methods to do so[147].

On the other side, campaigns also face the challenge of the data deluge, collecting too much data, or not quite knowing which data is important or how best to present it to demonstrate impact. To address these issues, a purposive data collection plan, specifying type of data that will be collected and how they will be analysed should be developed at the outset and reviewed regularly. Additionally, such a plan should address questions around data storage, access, personal data protection, and consent[148].

Measurement of impact is not only important to evaluate the actual success of a campaign, but also to provide benchmarks for funders. Harvesting too much noise might debilitate institutional ability to amplify vital signals. There needs to be a conversation around the validity of social media metrics which, although impressive, may not actually demonstrate impact. A similar conversation is happening in journalism, particularly among outlets funded by non-profits. While journalism and counternarrative campaigns, or advocacy more generally, have different purposes, they can learn from each other[149], and they sometimes have similar funders that require similar measures of return on investment and proof of impact[150].

Selected Campaign Tactics

Complementing the recommendations above, this last section provides selected evidence-based informed communication tactics for the counternarrative practitioners.

Using debunking carefully to avoid amplification

A dilemma between publishing or remaining silent about misinformation or hate speech should be resolved strategically[151]. For example, it might not be necessary to diffuse a factcheck to a broad audience when the said disinformation has not reached many people. Doing so may inadvertently provide further oxygen to it. In that case, it is useful to establish thresholds that determine at which point a narrative requires to be countered. Similarly, the audience at risk of being exposed to jihadi propaganda may be very limited and adopting too broad of a targeting of counternarrative may help promote it.

[146] Rrustemi, A. (2020). Measuring the Impact of the Lifestory Approach on Preventing and Countering Violent Extremism.
[147] Tacchi, J., Slater, D., & Hearn, G. (2003). Ethnographic action research. Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000139419)
[148] Latonero, M., Hiatt, K., Napolitano, A., Clericetti, G., Penagos, M. (2019). *Digital Identity in the Migration & Refugee Context: Italy Case Study* [Report]. Data and Society. https://datasociety.net/library/digital-identity-in-the-migration-refugee-context/
[149] Pitt, F., & Green-Barber, L. (2017). The Case for Media Impact: A Case Study of ICIJ's Radical Collaboration Strategy.
[150] Tofel, R. J. (2013). Issues Around Impact, ProPublica Annual Report.
[151] Benkelman, S. (2019) The Sound of Silence: Strategic Amplification *American Press Institute* Available at: https://www.americanpressinstitute.org/publications/reports/strategy-studies/the-sound-of-silence-strategic-amplification/

Scholars have drafted a handbook that shows best practices when drafting debunks[152].
If used by campaign, a careful three-step strategy can be followed:

- focusing on core facts rather than the myth to avoid the misinformation becoming more familiar;
- preceding by explicit warnings any mention of a myth to notify the reader that the upcoming information is false; and
- including an alternative explanation in the debunk that accounts for important qualities in the original misinformation."

Acting before misinformation: inoculation and pre-bunking

Insights from the inoculation theory can be useful when thinking about messaging that counters misinformation or manipulative discourses. This theory uses the analogy of a vaccine where pre-emptive exposure, or knowledge that people can be misinformed on a certain topic, strengthen their resistance to further misinformation attacks[153]. This type of messages includes a warning of misinformation, and an explanation of how the misinformation may be constructed, prior to the exposure to the actual misinformation[154]. For practitioners, it may be useful to draft messages that expose how arguments can be flawed, in order to strengthen a target audience's resilience to them.

Communicating with respect and empathy

When people are upset, angry, fearful, outraged, under high stress, involved in conflict, or feel high concern, they often have difficulty processing information[155]. Controversial topics already loaded with sentiments - particularly in the context of Covid-19[156] - need to be addressed with empathy and kindness[157]. It is essential to acknowledge what is known and what is unknown with humility. As mentioned before, beliefs and information gathering is not just about facts, but also about identity and group membership. Avoiding an authoritative, know-it-all attitude regarding what is true or not in contexts of high uncertainty may avoid the pitfalls of rebuttals by an audience who may feel contradicted and further entrenched in false beliefs[158] that may still offer some reassurance in times of justified anxieties[159].

[152] Cook, J., & Lewandowsky, S. (2012). *The debunking handbook*. St. Lucia, Australia: University of Queensland.

[153] Zerback, T., Toepfl, F., and Knöpfle, M. (2020). The disconcerting potential of online disinformation: Persuasive effects of astroturfing comments and three strategies for inoculation against them. *New Media & Society*. Available at: https://journals.sagepub.com/doi/abs/10.1177/1461444820908530

[154] Lewandowsky, S., van der Linden, S., and Cook, J. (2019) Inoculating against fake news? *Economic and Social Research Council.* Available at: https://esrc.ukri.org/news-events-and-publications/news/news-items/inoculating-against-fake-news/, See also Cook, J., Lewandowsky, S., & Ecker, U. K. H. (2017).Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. PLOS ONE, 12(5), e0175799. Available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0175799

[155] Glik, D. (2007) Risk Communication for Public Health Emergencies, *Annual Review of Public Health* 28:1, 33-54. Available at: https://www.annualreviews.org/doi/abs/10.1146/annurev.publhealth.28.021406.144123

[156] Chen, Y. (2020) Empathy in the age of misinformation: An open letter to healthcare and science professionals *Medical News Today*. Available at: https://www.medicalnewstoday.com/articles/empathy-in-the-age-of-misinformation-an-open-letter-to-healthcare-and-science-professionals#2

[157] Basu, T (2020) How to talk to conspiracy theorists—and still be kind *MIT Technology Review.* Available at: https://www.technologyreview.com/2020/07/15/1004950/how-to-talk-to-conspiracy-theorists-and-still-be-kind/

[158] Ahmed, M., *et al* (2019) Extreme Digital Speech

[159] Makri, A (2020) Don't Just Debunk Covid-19 Myths. Learn From Them *LSE Blogs*. Available at :https://blogs.lse.ac.uk/impactofsocialsciences/2020/03/20/dont-just-debunk-covid-19-myths-learn-from-them/

Telling complex stories

Campaign creators operate in an adversarial and divided online environment, especially when engaging in conversations that relate to concerns about issues surrounded by a lot of uncertainty, like climate change, migration, or the global pandemic. The communicators might achieve better results if they follow a rather counterintuitive advice to complicate the narrative. According to research, reviving complexity might bring surprising benefits[160].

One way to surface complexity involving audiences is through the story-circles tactic, based on an iterative, cyclical communication strategy. Counternarrative interventions that function as a point of engagement and work on multiple levels, can incentivise audiences to engage in narrative exchange[161], through which the framing and the prevailing attitudes and behaviours can be progressively changed with participation of the audiences.

[160] Ripley, A (2019) Complicating the Narratives. *Solutions Journalism*. Available at: https://thewholestory.solutionsjournalism.org/complicating-the-narratives-b91ea06ddf63

[161] Lutkenhaus R.O., Jansz, J., and Bouman M., (2019). Toward spreadable entertainment-education: leveraging social influence in online networks. *Health Promotion International* Oxford

Future Directions

This report summarised the important points in, and also mapped the main points of, disconnect between the present day academic, content regulation and technology policy debates about counternarratives and their role within wider security policies to counter terrorism, violent extremism, hate, and foreign influence. In the second part, the report provided concrete recommendations on how to address these gaps in the strategic communication practice.

In the last part, we propose general future directions for the required policy change that would help to defend and build resilience against threats that democratic societies are facing from terrorists, authoritarians, populists at home or abroad who are weaponising the internet and exploiting audience vulnerabilities. These proposed changes focus on reframing of the overall mission and are presented around four challenges that strategic communication can help address.

There are four challenges for strategic communication in the situation of confidence crisis in the digital ecosystem and in democratic institutions. They concern collective problem solving in democracies, audience participation and supporting community building forces, collaboration in broader alliances and technology risks and potential.

Governments, platforms, and civil society organizations are invited to discuss these challenges and the proposed directions, iteratively translating these changed mission objectives into practical steps.

1. **Supporting democratic social change**

News reports and policy debates centre around the prevalence of propaganda, hate speech or misinformation on social media platforms, but the biggest gaps in our understanding relate to their effects and larger societal impacts. Intentional malicious use or unintended consequences of communication technology result in real-world harm, to individuals, communities, businesses and institutions.

In our strategies to prevent harm, we should be looking beyond describing and quantifying problems affecting platforms, the symptoms, i.e. the presence of certain type of problematic content and conduct and measure the success by indicators other than its absence. This can be achieved first of all by shifting the attention to the consequences, particularly to how online content and conduct interacts with histories, cultures and conflicts of different countries and communities, over time.

Next, the redefined mission of strategic communication should include the consideration of structural causes and barriers to collective action and social change to address the big challenges facing humanity today, like economic and climate change, migration or pandemics. Propagandists and manipulators successfully exploit the difficulties democratic societies are facing to grapple with these issues. Their online speech distorts the perception of the crises and associated risks, erodes trust and solidarity, making consensus and collaboration more difficult.

Strategic communication has an essential role to play in counteracting these destructive efforts and in facilitating collective problem solving in the situation of confidence crisis in the digital ecosystem and in democratic institutions. Effective communication about challenges of policy making and about benefits of participatory, collaborative solutions to these problems is the best line of defence against the narratives of anti-democratic players.

Key strategic communication challenge #1

Supporting long term substantive policy reforms and social change processes, explaining costs and benefits of proposed solutions and advantages of messy problem solving in democratic societies, contrasting them to quick fixes and shortcuts promised by populists and propagandists, to audiences that are resistant to truth, facts, and evidence and sceptical of the scientific method.

2. Building communities

Social media promised to radically expand participation in decision making and improve ability to hold government to account for unprecedented numbers of engaged citizens and communities. In the process, they ended up exposing everybody to hate speech, disinformation, terrorist propaganda that feed off of uncertainty, prejudice, divisions that pre-date and exist beyond the internet, leaving citizens feeling overwhelmed, irritated, distracted, and disenfranchised.

Narratives and messaging that persuade people to vote for ethno-nationalist parties or authoritarian leaders, disregard information from the health authorities, or sympathise with hate groups are, for the most part, a blend that cannot be easily classified and do not fall into neat categories of content or speech. These narratives and messages resonate because they are perceived as offering a solution to a problem, as providing an alternative to solutions presented by the mainstream media, politicians, and experts.

From the point of view of their target audience, the terms used to describe the problematic speech categories are often indistinguishable and offensive. This language serves as an easy marker, a cue to stop paying attention and disengaging and may end up alienating audiences over longer term. Beyond counter-arguing the messages of propagandists and manipulators and debunking the endless supply of content they produce; our focus should be on understanding their appeal for those we want to convince.

Disinformation, hate, and terrorism counternarratives should strive to effectively communicate and engage audiences on problems they face, explaining what the government and civil society are doing and what they themselves can do to support education, job creation, welfare, anti-corruption, rule of law, justice, and public safety. Strategic communication response to weaponization of communication and exploitation of grievances should be organised around collaborative, inclusive solutions to issues and frustrations citizens face.

Key strategic communication challenge #2

Addressing the appeal and resonance of problematic, dangerous narratives, in addition to engaging with their producers, their messages, and their ideologies, through long term multi-level campaigns with audience participation to support community building forces.

3. Collaborating across policy and functional domains

The real-world antecedents and consequences of hate speech, terrorist propaganda, and misinformation are strikingly similar, and so are the sources of their appeal to susceptible audiences. Bureaucracies and corporations adopted labels and compartmentalised different categories and subcategories of problematic speech, as well as policy efforts to address them.

Under the previous point, we discussed how the use of these labels can be negatively perceived by affected audiences. From the perspective of communication practitioners, they represent a wider problem of the 'siloisation' of knowledge and resources, which limits the scope for collaboration. While there is a need for deep expertise in and nuanced understanding of mis/disinformation, hate speech, terrorist propaganda, there is even a greater need for building common vocabularies and collaborative infrastructure, through which efforts of different types of organizations, individuals and communities to resist and push back can be better connected and amplified.

Further, governments and platforms made various commitments to supporting counternarratives and strategic communication under different instruments, yet this support presents only a minor part of present-day content regulation and moderation discussions. In the absence of a strategy, campaigns run the risk of being treated as band-aid solutions, good PR, feel-good vanity projects to demonstrate that "something" is being done. These communication campaigns, i.e. speech funded by the government and/or platforms should also be subject to public oversight, the same way content restrictions and removals are.

This presents a rationale for more comprehensive communication strategies and interventions, joining-up efforts to counter hate speech, terrorist propaganda and mis/disinformation on the internet. They already are linked at least formally, under the platform content moderation, trust and safety policies and functions, and more recently under some proposed regulatory frameworks. A more strategic deployment of communication-based efforts countering hate, terrorist propaganda, mis/disinformation by the government, platforms and civil society requires advancing the underdeveloped practices and addressing multiple disconnects that hamper their effectiveness today.

Key strategic communication challenge #3

Supporting new models of collaboration, linking existing strategic communication initiatives and efforts to build and sustain broader alliances that expand boundaries and definitions of communication campaigns.

4. Demanding public interest, rights-protecting technology

Technology and democratic institutions are seemingly locked in a downward spiral of diminishing trust, at least partially as a result of online public debate plagued by hate, falsehoods, and propaganda. This report argues we cannot hope to defend against these without restoring faith in both democracy and technology.

Previous future direction points discussed how strategic communication can help build trust in democratic reforms and social change that address real issues and citizen's concerns, by building

communities that collaborate on inclusive, sustainable policy solutions, while resisting and pushing back against attacks by anti-democratic forces.

This last direction for strategic communication policy to consider, addresses trust issues related to technology. Core digital rights concerns like censorship and freedom of expression became weaponised by ideologues and partisans to serve as flash points of polarized public debate. Going beyond speech and content related technology, wider concerns about data ownership, privacy, and surveillance, the use of AI and automation, but also wireless networking technology also become popular conspiracy theory material.

Campaigners cannot afford to be uncritical or agnostic users of communication infrastructures, as they depend on the underlying architecture, its products and features. They are delivering their messages in an increasingly complex and fast evolving cross-platform information ecosystem. The governing policies that influence behaviour of users around the world are often put in place in response to abuse related incidents and crises. Strategic communication around these crises, around content policies, explaining to the public the reasons and benefits of moderation decisions more effectively, might limit the ability of the punished abusers to use it as a badge of honour. Better communicating far reaching implications of technology for democracy and human rights is even more essential.

Key strategic communication challenge #4

Explaining benefits, challenges of, and limits to technological solutions for detecting and removing content, suppressing its circulation, as well as banning users and networks at scale, and about risks related to privacy, surveillance, the use of AI and automatic filtering communication to both public and decision-makers.