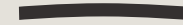




PLATFORM  
PEACEFUL  
CONFLICT  
TRANSFORMATION



Berghof Foundation

# PEACEBUILDING AND DISINFORMATION:

Taking stock and  
planning ahead

This publication is part of the portfolio of the Working Group Peace and Development (FriEnt) and supported by the Federal Ministry for Economic Cooperation and Development (BMZ).



**About this paper:** This paper was written by Grace Connors and Emma Baumhofer and was commissioned by the Berghof Foundation and the Plattform Zivile Konfliktbearbeitung / Platform Peaceful Conflict Transformation.

The authors would like to acknowledge the contributions of Berghof Foundation and colleagues at the Platform Peaceful Conflict Transformation and FriEnt: Cora Bieß, Ute Finckh-Krämer, Jennifer Menninger, Beatrix Austin, and Nina Strumpf, as well as the interviewees (listed in Annex A).

**To cite:** Connors, Grace & Emma Baumhofer 2025. *Peacebuilding and disinformation: Taking stock and planning ahead*. Berghof Foundation and Platform Peaceful Conflict Transformation.

**Disclaimer:** The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of Berghof Foundation or Platform Peaceful Conflict Transformation.

**Design and layout:** Amberdesign, Berlin

Published by

**Berghof Foundation**

Berghof Foundation Operations gGmbH  
Lindenstraße 34  
10969 Berlin  
Germany

[www.berghof-foundation.org](http://www.berghof-foundation.org)



Platform Peaceful Conflict  
Transformation  
Großbeerenstr. 13a  
10963 Berlin  
Germany

[www.pzkb.de](http://www.pzkb.de)

© 2025. All rights reserved.

# **PEACEBUILDING AND DISINFORMATION:**

Taking stock and  
planning ahead

Grace Connors and Emma Baumhofer



# Table of contents

Contents	5
Note from the authors	6
Executive summary	7
Recommendations	8
<b>I. Framing the topic of disinformation and peacebuilding</b>	<b>10</b>
Defining disinformation	11
The role of peacebuilders	12
The role of platforms	16
<b>II. The landscape of peacebuilding interventions against disinformation</b>	<b>18</b>
Case study: Vote Women and Future of Work	19
Case study: Reducing Online Conflict Community	21
Case study: #vrschwrng and Digital.Truth	22
Case study: Social Media Peace Agreements, Centre for Humanitarian Dialogue	23
Key learnings from the case studies	25
<b>III. What does cognitive science say about disinformation?</b>	<b>26</b>
Why do humans share false information?	26
Why are humans susceptible to disinformation?	27
How does this inform disinformation response?	29
<b>IV. Conclusion and recommendations</b>	<b>34</b>
Annexes	37
A – List of interviewees	37
B – The authors	37
References	38

# Note from the authors

## Potential bias and limitations

As peacebuilders, we would be remiss not to mention the inherent biases associated with this work. First, from the perspective of the authors, who hail from institutions in the United States and Switzerland, respectively. The examples and strategies shared below reflect conscious work on our part to present a holistic and realistic view of the current status of peacebuilding approaches to disinformation, although we recognise that this perspective is limited by the exposure we have to the topic. We welcome input and reflection from global peacebuilders working to combat disinformation, as your insights will only help improve our work in this field, which is the ultimate goal of this project.

Second, it is critical to recognise that the technologies we examine are themselves inherently biased in the ways they are created. Digital technology, while a great potential force for good in the world, is rooted in inequality and forms of violence. The authors suggest books such as Mirca Madianou's *Technocolonialism*, Frances Haugen's *The Power of One*, Joy Buolamwini's *Unmasking AI*, Kate Crawford's *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*, and Cathy O'Neil's *Weapons of Math Destruction* as great initial sources for unpacking the ways technology has fueled global harms and potential ways to move forward.

It is also because of this history that we note that the solutions we share below are not expected to apply to every country or conflict context. Peacebuilding practice emphasises the importance of localisation in conflict transformation, and given that disinformation proliferates in different ways in different spaces, we encourage that the lessons shared below be adapted and applied based on local peacebuilding practices.

Regardless of these limitations, this work was deemed necessary given that the technologies that perpetuate disinformation and disrupt social cohesion are woven into our daily lives in ways that no longer separate online and offline spaces. Rather, conflict is reinforced by and cycles between the two so critically that, as Julia Schiwal of USIP notes, “peacebuilding analysis and practice that fails to appreciate this shift will be painfully limited in its capacity to have enduring relevance and offer insight.”<sup>1</sup> We aim to appreciate that shift and help share insights for the future work of peacebuilders.

---

1 Schiwal, Julia 2023. A New Approach for Digital Media, Peace and Conflict. United States Institute for Peace. [www.usip.org/publications/2023/02/new-approach-digital-media-peace-and-conflict](https://www.usip.org/publications/2023/02/new-approach-digital-media-peace-and-conflict).

# Executive summary

**Disinformation, understood here as false or misleading information shared with the intent to harm**, is a pervasive challenge that undermines trust, social cohesion, and peace. It is an issue worked on in many sectors and yet, despite immense effort and research, the results can feel underwhelming. With limited resources to address this massive challenge, there can be a tendency to return to known strategies or repeat what others have done, without interrogating their efficacy. At this juncture, this study asks: **What strategies are peacebuilders currently employing to mitigate the spread of disinformation? What lessons have peacebuilders learned from previous efforts? How can research from other sectors, such as cognitive science, improve and inform those efforts?**

This study contributes to scholarship on disinformation by highlighting the unique ways *peacebuilders can contribute* to the interdisciplinary response that is needed to address this global issue. In particular, it highlights the role and value of the **conflict analysis and transformation, mediation and dialogue, reconciliation, storytelling, and moral imagination** skills of peacebuilders.

It also explores the strategies that peacebuilders have been using thus far to tackle disinformation, such as media literacy training, storytelling campaigns to counter false narratives, reconciliation programs, trust building in institutions, and participation in trusted flagger programs, to name a few. Four case studies provide deeper insights into how peacebuilding organisations tackle disinformation worldwide:

- ≡ **Pollicy's Pan-African Vote: Women and Future of Work** programs provide capacity building for women to counter gendered disinformation;
- ≡ **Mercy Corps' Reducing Online Conflict Community (ROCC)** emphasises an interdisciplinary systems approach in Nigeria;
- ≡ **Berghof Foundation's #vrschwrgn toolkit and Digital.Truth** empower youth and their caretakers to address conspiracy theories and participate in intergenerational dialogues;
- ≡ **and the Centre for Humanitarian Dialogue's social media codes of conduct** contribute to reducing the generation of disinformation while building trust among conflict parties.

Additionally, the study critically investigates the role of technology in producing and spreading disinformation, highlighting inflection points where peacebuilders can respond. Much of this work recognises that peacebuilders must go beyond “firefighting” techniques that address the symptoms of untrustworthy information environments and **bring in more aspects of cognitive science to peacebuilding programming that addresses the root causes of disinformation and conflict.**

Cognitive science research shows that humans share false information because they are seeking to belong or to build a reputation, have developed a habit of sharing, have high trust in their information source, have a strong emotional reaction to the information, or hold a desire to disrupt the social order. Similarly, people can be susceptible to disinformation narratives when they fail to think critically about the information they consume, are influenced by their perception of social consensus and their personal or political beliefs,

repeatedly see false information, or are emotionally manipulated. **Thus, peacebuilding responses need to address the specific motivations or influences of three broad types of actors: disinformation generators, disinformation spreaders, and disinformation recipients.**

The study encourages peacebuilders, policymakers, and platform operators to redesign systems that amplify false narratives, create new platforms built on peacebuilding principles, and adopt integrated approaches to combat disinformation and build resilience. With this in mind, the following recommendations come out of this analysis:

## Recommendations

### 1. Encourage a systems-based strategy for addressing disinformation.



Cross-sector partnerships and intervention strategies are critical in reducing disinformation's harms. Funders can support interdisciplinary research and programming, policymakers can hold platforms accountable for their role in disinformation amplification, peacebuilders can create diverse communities of practice, and technologists can integrate proven strategies into platforms.

### 2. Leverage cognitive science to prioritise preventative programming at the source of disinformation.



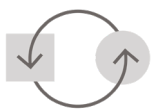
Cognitive science demonstrates many of the motivations or susceptibilities humans have to disinformation. These lessons can inform peacebuilding work to move interventions beyond “firefighting” initiatives that address the repercussions of disinformation to ones which tackle disinformation's root causes. Critical here is localised programming to address the specific economic or political conflict drivers that make communities more vulnerable to disinformation.

### 3. Aim for gender-transformative approaches to disinformation response.



Peacebuilders should protect and increase the resilience of women and marginalised groups by increasing access to proven programs designed for these groups, while addressing the role of masculinities and misogyny in disinformation production.

### 4. Integrate strategies into broader mediation and policy efforts.



Practical efforts to reduce disinformation can serve many purposes and have a stronger impact when integrated into larger strategies. Peacebuilders and their partners should aim to bring disinformation mitigation tactics into additional spaces of their work, such as national education curriculums or mediation dialogues.



### 5. Decolonise disinformation work by investing in local research and practice.



Many peacebuilders around the world are engaged in practical, effective, and localised disinformation work. The toolkits and lessons learned from these programs should be amplified to help other regions benefit from established building blocks, while further research is completed on local information ecosystems.

### 6. Advocate for and incentivise tech platform product and policy changes to reduce disinformation.



Many disinformation interventions have been validated by existing public research, such as accuracy prompts and content labelling, which should now be integrated into all tech platforms to mitigate disinformation's harms. Policymakers can support this work by incentivizing implementation and innovation of tactics that promote prosocial communication.

### 7. Imagine and design new prosocial tech platforms and build bridges with the technology community.



While encouraging change within existing tech platforms, peacebuilders can work with technologists on designing new ones that control for disinformation's harms and promote prosocial values. Technologists and peacebuilders can also work together to understand the opportunities and harms of existing and new technologies.

# I. Framing the topic of disinformation and peacebuilding

A cross-platform,<sup>2</sup> coordinated Russian campaign<sup>3</sup> against humanitarian workers in Syria to cast doubt over the evidence of war crimes they are collecting.<sup>4</sup> The “Disinformation Dozen,” or the twelve individuals who alone spread 65% of disinformation about Covid-19 in the United States.<sup>5</sup> Disinformation-for-hire firms promoting falsehoods amidst contested elections in Venezuela, Mexico, and Bolivia.<sup>6</sup> A sitting U.S. president removed from Twitter (now X) for inciting violence in the U.S. Capitol based on false information.<sup>7</sup> Disinformation campaigns naming trusted flaggers<sup>8</sup> “censorship machines” to engender public scepticism of the EU’s Digital Services Act.<sup>9</sup> These are all examples of disinformation on social media that has proliferated into direct, structural, and cultural violence in recent years.

Since the creation of social media, it has been leveraged by malign actors to perpetuate the spread of disinformation. This is not the sole cause

or responsibility of social media platforms, as disinformation campaigns have long played a role in human history. Yet social media’s affordances of access and scale, and limitations of content moderation practices, have allowed it to massively increase the global influence of disinformation.<sup>10</sup>

This was seen as recently as this month, when Meta came under intense scrutiny for its decision to end its fact-checking program on its platforms in the US in favor of adopting a “community notes” model.<sup>11</sup> With this new practice, platform users – rather than the platform itself – are charged with identifying false information and adding context to misleading posts, shared in crowdsourced “notes” on individual posts. The decision immediately sparked fears about the potential impact this decision will have on disinformation worldwide, with the expectation that Meta will likely also suspend its fact-checking programs in Latin America, Europe,<sup>12</sup> and Asia.<sup>13</sup> Mark Zuckerberg, Meta’s CEO, in part

2 Cross-platform here refers to how the Russian campaign in Syria operated across multiple social media platforms to fully pollute the information environment.

3 **Wilson, Tom & Kate Starbird** 2020. Cross-platform disinformation campaigns: Lessons learned and next steps. Harvard Misinformation Review. [misinfreview.hks.harvard.edu/article/cross-platform-disinformation-campaigns/](https://misinfreview.hks.harvard.edu/article/cross-platform-disinformation-campaigns/).

4 **The Syria Campaign** 2017. Killing the Truth. [thesyriacampaign.org/wp-content/uploads/2017/12/KillingtheTruth.pdf](https://thesyriacampaign.org/wp-content/uploads/2017/12/KillingtheTruth.pdf).

5 **Center for Countering Digital Hate** 2022. “The Disinformation Dozen.” [drive.google.com/file/d/101J0AzKiGG29Y02UaxzDFy63QgxUQ-fz/view](https://drive.google.com/file/d/101J0AzKiGG29Y02UaxzDFy63QgxUQ-fz/view).

6 **Cryst, Elena, Esteban Ponce de León, Daniel Suárez Pérez & Shelby Perkins** 2022. Bolivarian Factions: Facebook Takes Down Inauthentic Assets. Stanford Internet Observatory. [purl.stanford.edu/qb823mb8849](https://purl.stanford.edu/qb823mb8849).

7 **Allyn, Bobby & Tamara Keith** 2021. Twitter Permanently Suspends Trump, Citing 'Risk Of Further Incitement Of Violence'. NPR. [npr.org/2021/01/08/954760928/twitter-bans-president-trump-citing-risk-of-further-incitement-of-violence](https://npr.org/2021/01/08/954760928/twitter-bans-president-trump-citing-risk-of-further-incitement-of-violence).

8 Trusted flaggers are national actors identified by the DSA with expertise in detecting illegal and harmful content. They are trusted to submit a report on this content at least once annually, with these reports shown priority by online platforms.

9 **Klotsonis, David** 2024. Trusted Flaggers in the DSA: Challenges and Opportunities. Center for Democracy and Technology. [cdt.org/insights/trusted-flaggers-in-the-dsa-challenges-and-opportunities/](https://cdt.org/insights/trusted-flaggers-in-the-dsa-challenges-and-opportunities/).

10 **Benaissa Pedriza, Samia** 2021. Sources, Channels and Strategies of Disinformation in the 2020 US Election: Social Networks, Traditional Media and Political Candidates. Journalism and Media. [doi.org/10.3390/journalmedia2040036](https://doi.org/10.3390/journalmedia2040036).

11 **Kaplan, Joel** 2025. More Speech and Fewer Mistakes. Meta. [about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/](https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/).

12 **Ramis, Jorge** 2025. Latin American Fact-Checkers Brace for Meta's Next Moves. Wired. [wired.com/story/hispanic-fact-checkers-react-meta-disinformation/](https://www.wired.com/story/hispanic-fact-checkers-react-meta-disinformation/).

13 **Beltran, Sam** 2025. Meta’s US fact-checking shutdown sparks fears of disinformation crisis in Asia. South China Morning Post. [scmp.com/week-asia/politics/article/3294518/metas-us-fact-checking-shutdown-sparks-fears-disinformation-crisis-asia](https://www.scmp.com/week-asia/politics/article/3294518/metas-us-fact-checking-shutdown-sparks-fears-disinformation-crisis-asia).

attributed the decision to the challenge Meta has faced in its over-regulation of political content since 2016 given bias in fact-checkers themselves. This massive disinvestment in fact-checking by Meta, coupled with its attribution to political bias, is thus expected to have far-reaching impacts on the global information environment by undermining user trust and amplifying opportunities for disinformation to proliferate.

Meta's decision will likely amplify additional related challenges often faced on social media, including mal- and misinformation, cyber campaigns, and hate speech, among others. Each of these challenges deal with information accuracy and/or use. For example, malinformation is "information based on fact but used out of context to mislead, harm, or manipulate."<sup>14</sup> This is innately related to disinformation, which also aims to mislead or harm, but through the use of false information. Both of these types of information can be used in hate speech, which is language used to discriminate against an individual or a group based on factors related to their identity.<sup>15</sup> All of these challenges can then be leveraged in coordinated cyber campaigns which seek to disrupt functioning society by widely spreading specific narratives or political agendas.

Therefore, analysis of one of these areas is never undertaken in isolation, and most, if not all, peacebuilding responses to disinformation simultaneously address others from this list. With that said, this study focuses on disinformation because the authors view it as a space of intersection for these related themes. In particular, this study highlights the specific and unique ways that peacebuilders contribute to combatting disinformation and looks to ways that this work can evolve in the future. It largely does so through an examination of four case studies on current

peacebuilding approaches to disinformation, highlighting lessons learned from these experiences. It further includes an exploration of how cognitive science can help peacebuilders understand human susceptibility to disinformation so this community can design better programs and policies to address root causes. This all culminates in a series of practical recommendations on how the field can develop this work.

## Defining disinformation

Disinformation refers to false or misleading information that is created and disseminated with "malign intent and is shared in order to cause harm."<sup>16</sup> This includes the direct harm caused by intentionally spread falsehoods about individuals, groups, or events, and indirect harms, such as casting doubt on information ecosystems, eroding trust in institutions, and undermining democracy.<sup>17</sup> The objectives of disinformation are multiple, ranging from manipulation and radicalisation, discrediting or diminishing opponents, exacerbating existing conflicts by pitting groups against each other, creating confusion, reducing empathy, or contributing to post-truth environments.

The malign intent is what distinguishes disinformation from misinformation. While misinformation also pertains to false or misleading information, it does not require malintent from its creators or disseminators. As peacebuilding scholars Kristina Hook and Ernesto Verdeja note, though, "the boundaries are porous...many purveyors of misinformation believe what they are sharing and are not intentionally spreading false information."<sup>18</sup> One malicious actor may intentionally share disinformation, which then

14 **CISA.** Mis-, Dis-, and Malinformation. [cisa.gov/sites/default/files/publications/mdm-incident-response-guide\\_508.pdf](https://www.cisa.gov/sites/default/files/publications/mdm-incident-response-guide_508.pdf).

15 **United Nations.** What is hate speech? [un.org/en/hate-speech/understanding-hate-speech/what-is-hate-speech](https://www.un.org/en/hate-speech/understanding-hate-speech/what-is-hate-speech).

16 **Brooks, Adrienne, Will Ferroggiario & Lisa Inks** 2021. Social Media, Conflict, and Peacebuilding: Issues and Challenges. Mercy Corps. [mercycorps.org/sites/default/files/2022-01/Social-Media-Discussion-Paper-9-Dec-1.pdf](https://www.mercycorps.org/sites/default/files/2022-01/Social-Media-Discussion-Paper-9-Dec-1.pdf)

17 **Starbird, Kate** 2019. Disinformation's spread: bots, trolls and all of us. *Nature*. [nature.com/articles/d41586-019-02235-x](https://www.nature.com/articles/d41586-019-02235-x).

18 **Hook, Kristina & Ernesto Verdeja** 2022. Social Media Misinformation and the Prevention of Political Instability and Mass Atrocities. Stimson Center. [stimson.org/2022/social-media-misinformation-and-the-prevention-of-political-instability-and-mass-atrocities/](https://www.stimson.org/2022/social-media-misinformation-and-the-prevention-of-political-instability-and-mass-atrocities/).

gets reshared thousands of times by unsuspecting users, or another actor may weaponise a piece of widely accepted misinformation and turn it into disinformation; the two forms of manipulation are responsive to each other. This is particularly true in conflict settings. A UN panel in South Sudan reported that “social media has been used by partisans on all sides...to exaggerate incidents, spread falsehoods and veiled threats or post outright messages of incitement,” a clear example of politically-motivated disinformation becoming misinformation when reshared widely by local social media users.<sup>19</sup>

This case in Sudan demonstrates one type of actor who engages in disinformation creation and spread, namely governments, their proxies, or conflict actors serving to disrupt or maintain the current political agenda. A report from the Oxford Internet Institute found that 81 governments and political parties used social media to spread political disinformation in 2020 alone.<sup>20</sup> Other examples include foreign actors interfering in another country to disrupt functioning society, with much international attention given to Russia in this respect;<sup>21</sup> actors seeking to make financial gain through sharing disinformation,<sup>22</sup> typically through participation in troll farms;<sup>23</sup> and actors amplifying a specific point of view or ideological agenda, or some combination thereof.

## The role of peacebuilders

A policy brief from CeMAS funded by the Alfred Landecker Foundation offers a series of perspectives from actors engaged in disinformation response, namely the information, security, technological, social science, and democracy spaces.<sup>24</sup> The authors point out that an integrated approach drawing on perspectives from each of these spaces is required to systematically respond to disinformation. For example, social science helps explain the nature, spread, and impact of disinformation so that it can be contained more effectively. Policy recommendations utilise this explanation to promote better responses to disinformation, such as citizen media literacy. A democracy-centered perspective maintains focus on institutional trust and the concept of truth. In place of short-term impact, interventions in this area aim for resilient societies in the long term.

Adding to this list, we offer a peacebuilding lens. **From a peacebuilding perspective, disinformation is a threat to social cohesion; undermines trust in individuals, communities, and institutions; and furthers global violence and structural harms.** The first lesson peacebuilders learn is that peace is not just the absence of violence,<sup>25</sup> rather that building a thriving prosocial world requires promoting human flourishing. Disinformation detracts from this vision of a peaceful world. Responses to disinformation from this perspective therefore take

19 Knopf, Payton, Andrews Atta-Asamoah, Andrei Kolmakov, Ann Oosterlinck & Klem Ryan 2016. Letter dated 15 November 2016 from the Panel of Experts on South Sudan established pursuant to Security Council resolution 2206 (2015) addressed to the President of the Security Council. United Nations Security Council. [documents.un.org/doc/undoc/gen/n16/350/68/pdf/n1635068.pdf](https://documents.un.org/doc/undoc/gen/n16/350/68/pdf/n1635068.pdf).

20 Bradshaw, Samantha, Hannah Bailey & Philip Howard 2021. Industrialized Disinformation: 2020 Global Inventory of Organised Social Media Manipulation. Working Paper 2021.1. Oxford, UK: Project on Computational Propaganda. [demotech.oii.ox.ac.uk/wp-content/uploads/sites/12/2021/02/CyberTroop-Report20-Draft9.pdf](https://demotech.oii.ox.ac.uk/wp-content/uploads/sites/12/2021/02/CyberTroop-Report20-Draft9.pdf).

21 Meaker, Morgan 2024. Russia Is Targeting Germany With Fake Information as Europe Votes. Wired. [wired.com/story/european-union-elections-russia-germany-disinformation-campaigns/](https://www.wired.com/story/european-union-elections-russia-germany-disinformation-campaigns/).

22 Scholtens, Michael, Pedro Pizano, Max Karpawich & Guthrie Kuckes 2024. The Disinformation Economy. The Carter Center. [cartercenter.org/resources/pdfs/news/peace\\_publications/democracy/the-disinformation-economy-mccain-may-2024.pdf](https://www.cartercenter.org/resources/pdfs/news/peace_publications/democracy/the-disinformation-economy-mccain-may-2024.pdf).

23 Silverman, Craig & Lawrence Alexander 2016. How Teens In The Balkans Are Duping Trump Supporters With Fake News.” Buzzfeed News. [buzzfeednews.com/article/craigsilverman/how-macedonia-became-a-global-hub-for-pro-trump-misinfo#.rhNWoPjEbG](https://www.buzzfeednews.com/article/craigsilverman/how-macedonia-became-a-global-hub-for-pro-trump-misinfo#.rhNWoPjEbG).

24 Lamberty, Pia & Lea Frühwirth 2023. Information manipulation as a complex challenge. CeMAS. [cemas.io/en/publications/integrative-model-disinformation/2023-06-14\\_Policy\\_Brief\\_Integrative\\_model\\_disinformation.pdf](https://www.cemas.io/en/publications/integrative-model-disinformation/2023-06-14_Policy_Brief_Integrative_model_disinformation.pdf).

25 Galtung, Johan 1967. Theories of Peace: A Synthetic Approach to Peace Thinking. Oslo: International Peace Research Institute.

a whole-of-system approach, aiming to transform settings in which disinformation exists, mediate between parties involved in disinformation, and prevent disinformation from occurring in the first place.

Multiple peacebuilding strengths and skills critically relate to this perspective:



### 1. Conflict analysis and transformation –

Defined by Berghof Foundation as “a complex process of constructively changing relationships, attitudes, behaviours, interests and discourses in violence-prone conflict settings [that] addresses and changes underlying structures, cultures and institutions that encourage and condition violent political and social conflict over the long term.”<sup>26</sup> Conflict transformation does not aim to eliminate conflict, but to improve the state of conflict from violence to justice.<sup>27</sup>



**2. Mediation and dialogue** – In peacebuilding this is typically seen in the context of peace agreement negotiation, where peacebuilders sit within often seemingly intractable conflicts to produce proactive peace.<sup>28</sup> Success in mediation draws on additional peacebuilding strengths, such as community building, the creation of inclusive spaces for dialogue, and multilateralism.



**3. Reconciliation** – Reconciliation moves past resolving conflict disputes to fostering long-term healing and social cohesion amongst impacted parties. As defined by the United States Institute for Peace, “reconciliation is the long-term process by which the parties to a violent dispute build trust, learn to live cooperatively, and create a stable peace. It can happen at the individual level, the community level, and the national level.”<sup>29</sup>



**4. Storytelling** – Stories have a profound influence over the human imagination and allow people to perceive and understand cultures, communities, and conflicts in a more accessible way. Peacebuilders recognise this as a potential tool for change, as noted by Kirithi Jayakumar, “storytelling is ...[a] means by which a community might examine values embedded in its traditional stories with an eye to abandoning strife” that allows for bridge building through experienced mutuality.<sup>30</sup>



**5. Moral imagination** – John Paul Lederach writes that peacebuilders have skills in “moral imagination,” which he understands as “the capacity to imagine something rooted in the challenges of the real world yet capable of giving

26 **Berghof Foundation** 2020. Berghof Glossary on Conflict Transformation: 20 Notions for Theory and Practice. [yumpu.com/en/document/view/63621335/berghof-glossary-on-conflict-transformation](https://yumpu.com/en/document/view/63621335/berghof-glossary-on-conflict-transformation).

27 **Clements, Kevin** 2004. Towards Conflict Transformation and a Just Peace. In: Austin, A., Fischer, M., Ropers, N. (eds) Transforming Ethnopolitical Conflict. VS Verlag für Sozialwissenschaften, Wiesbaden. [doi.org/10.1007/978-3-663-05642-3\\_21](https://doi.org/10.1007/978-3-663-05642-3_21).

28 **Kriesberg, Louis, Terrell Northrup & Stuart Thorson** 1989. Intractable Conflicts and Their Transformation. Syracuse University Press.

29 **United States Institute of Peace**. Reconciliation: Truth, Justice, Peace, Mercy. [usip.org/public-education-new/reconciliation-truth-justice-peace-mercy](https://usip.org/public-education-new/reconciliation-truth-justice-peace-mercy).

30 **Jayakumar, Kirithi** 2015. Storytelling for peace. Peace Insight. [peaceinsight.org/en/articles/storytelling-peace/?location=&theme=culture-media-advocacy](https://peaceinsight.org/en/articles/storytelling-peace/?location=&theme=culture-media-advocacy).

birth to that which does not yet exist.”<sup>31</sup> This is a unique type of creativity, which accepts existing conflict but is unafraid to radically reimagine what a prosocial society could look like within that context.

Each of these skills plays a unique role as peacebuilders respond to the lifecycle of disinformation, from its creation, to its spread, and ultimately to its lasting impact on society. They also complement the skills utilised by other actors engaged in disinformation response.

To demonstrate this integrated approach, the authors adapt the hypothetical example shared in the brief of the nonprofit Center for Monitoring, Analysis, and Strategy (CeMAS), in which Russia spreads disinformation about Ukrainian refugees setting fire to homes in Germany in order to disrupt functioning society. From the peacebuilding perspective, this disinformation is likely to detract from prosocial flourishing in German and Ukrainian communities, as well as aid in the normalisation of discrimination against Ukrainian refugees worldwide. Therefore, this scenario requires a peacebuilding response. The orange arrows on the graph represent interventions that actors from the CeMAS list might undertake in response to this type of disinformation. The green arrows are additions by the authors of this study to illustrate the peacebuilding perspective and possible interventions, though it is important to note that this is not an exhaustive list.

---

31 Lederach, John Paul 2005. *The Moral Imagination: The Art and Soul of Building Peace*. New York: Oxford University Press.

## Application of the integrative approach for dealing with a narrative

Perspective abbreviations: **So**: Social science / **Se**: Security / **T**: Technology / **I**: Information / **D**: Democracy / **P**: Peacebuilding

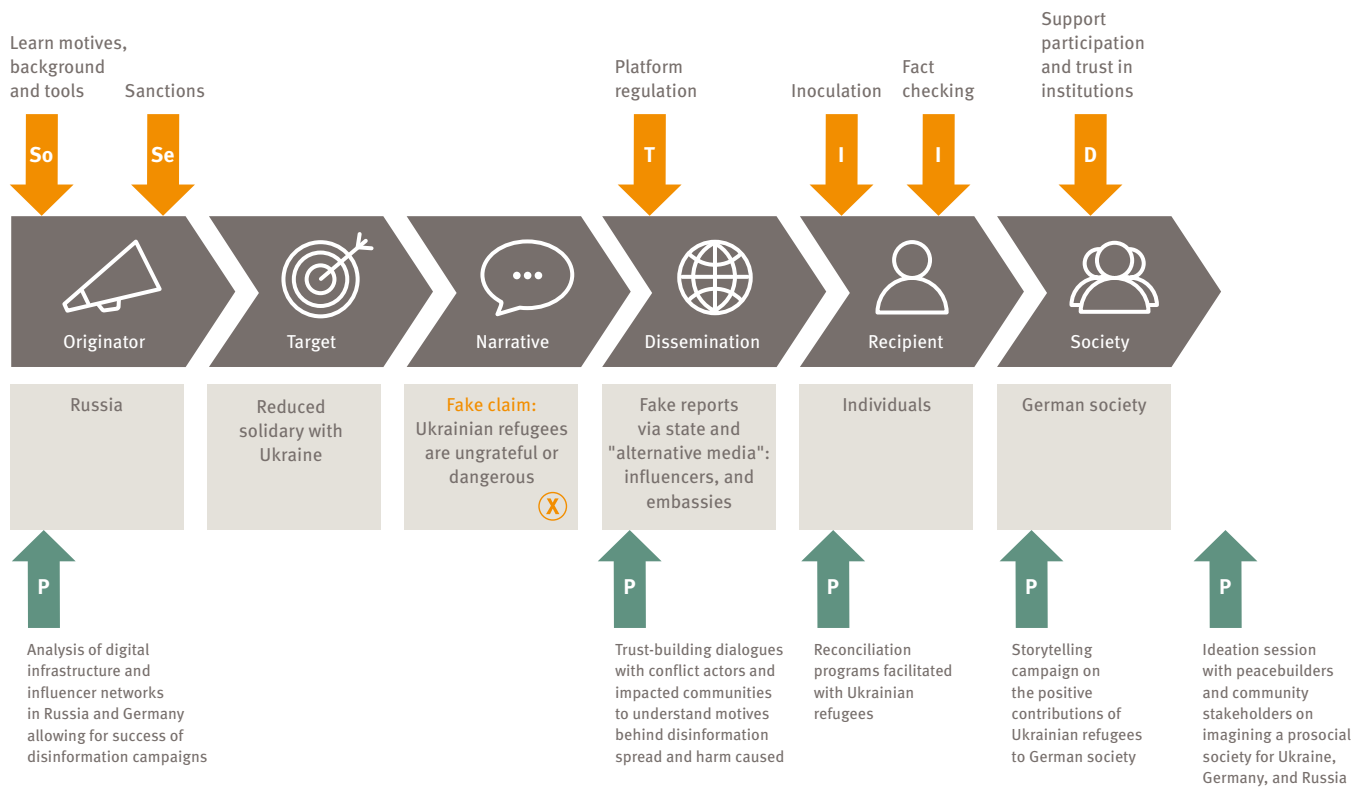


Figure 1: Complementary Peacebuilding Approaches to Addressing Disinformation Lifecycle, adapted from CeMAS<sup>32</sup>

From the CeMAS approach, the social science perspective aids in understanding motives, backgrounds, and tools used by the conflict actor, Russia; the security perspective advocates for sanctions against Russia; the technology perspective encourages platform changes and regulation; the information perspective works on inoculation and fact-checking programs; and the democracy perspective works on trust-building programming. The peacebuilding approach aligns well with these goals. For example, to assist in the democratic strengthening of trust in institutions, peacebuilders can lead storytelling campaigns to influence public perception of Ukrainian refugees and rebuild public support of institutions and

immigration policy. Given strengths in mediation, they can also host trust-building dialogues with conflict actors to understand their motivations behind disinformation creation and with impacted stakeholders to understand disinformation spread and the harms caused, which together can inform recommendations made by the technology perspective to platforms for regulation and redesign. In unpacking disinformation spread from the perspective of Ukrainian refugees and local populations, these dialogues should ideally bring to light any manufactured mistrust between these communities and allow the community to begin rebuilding social cohesion.

<sup>32</sup> Lamberty, Pia & Lea Frühwirth 2023. Information manipulation as a complex challenge. CeMAS. [cemas.io/en/publications/integrative-model-disinformation/2023-06-14\\_Policy\\_Brief\\_Integrative\\_model\\_disinformation.pdf](https://cemas.io/en/publications/integrative-model-disinformation/2023-06-14_Policy_Brief_Integrative_model_disinformation.pdf).

This ability to build trust and convene stakeholders is critical in peacebuilding responses to disinformation, as it allows peacebuilders to gather community intelligence and synthesise points of commonality to reimagine society. This is the moral imagination and conflict transformation in action, which is a process that undoubtedly involves stakeholders from tech platforms, security and information spheres, democratic institutions, and social science.

## The role of platforms

While understanding the unique skills peacebuilders hold in disinformation response, it is also critical to understand the technology landscape within which that response operates. Social media platforms amplify disinformation through their design and their ineffective content regulation policies. On design, platforms such as Facebook and Instagram are criticised for their algorithms which reward exploitative content<sup>33</sup> and equate content popularity with legitimacy,<sup>34</sup> their content display methods which often hide or remove context associated with content,<sup>35</sup> their mechanisms for seamlessly resharing content regardless of accuracy,<sup>36</sup> and much more. On policy, numerous studies have proven that the emphasis on content moderation is ineffective, exacerbated by the fact that these departments are consistently

understaffed and force employees to work in harsh conditions,<sup>37</sup> bad at detecting borderline or reclaimed<sup>38</sup> speech,<sup>39</sup> governed by biased content regulation,<sup>40</sup> and distant from product design teams.<sup>41</sup>

These platform critiques provide opportunities for peacebuilding interventions against disinformation. Peacebuilders Helena Puig Larrauri of BuildUp and Maude Morrison of the Centre for Humanitarian Dialogue (HD) understand that technology creates the enabling conditions for conflict drivers that disrupt social cohesion, and outline a series of digital conflict drivers and potential levels of peacebuilding response. The pyramid below represents their framework.

The top layer of the pyramid represents the most visible signals of conflict dynamics on social platforms. The subsequent layers move deeper into the root causes, motivations, and issues related to the conflicts themselves. As shown, disinformation is in the second layer, given its less-visible nature to the average social media user, but strong ability to contribute to and reinforce conflict. At the lower levels are more abstract topics such as human communication and human neurology, because, as noted by Puig Larrauri and Morrison, “digital conflict drivers touch on some of the deepest roots of the human condition –

33 **Bundtzen, Sara** 2022. “Suggested for You”: Understanding How Algorithmic Ranking Practices Affect Online Discourses and Assessing Proposed Alternatives. Institute for Strategic Dialogue. [isdglobal.org/wp-content/uploads/2022/12/Understanding-How-Algorithmic-Ranking-Practices-Affect-Online-Discourses-and-Assessing-Proposed-Alternatives.pdf](https://isdglobal.org/wp-content/uploads/2022/12/Understanding-How-Algorithmic-Ranking-Practices-Affect-Online-Discourses-and-Assessing-Proposed-Alternatives.pdf).

34 **Deb, Anamitra, Stacy Donohue & Tom Glaisyer** 2017. Is Social Media a Threat to Democracy?. The Omidyar Group. [omidyargroup.com/wp-content/uploads/sites/7/2017/10/Social-Media-and-Democracy-October-5-2017.pdf](https://omidyargroup.com/wp-content/uploads/sites/7/2017/10/Social-Media-and-Democracy-October-5-2017.pdf).

35 **Krafft, Peaks & Joan Donovan** 2020. Disinformation by Design: The Use of Evidence Collages and Platform Filtering in a Media Manipulation Campaign. Political Communication. [doi.org/10.1080/10584609.2019.1686094](https://doi.org/10.1080/10584609.2019.1686094).

36 **Allen, Jeff** 2022. Misinformation Amplification Analysis and Tracking Dashboard. Integrity Institute. [integrityinstitute.org/blog/misinformation-amplification-tracking-dashboard](https://integrityinstitute.org/blog/misinformation-amplification-tracking-dashboard).

37 **Stackpole, Thomas** 2022. Content Moderation Is Terrible by Design. Harvard Business Review. [hbr.org/2022/11/content-moderation-is-terrible-by-design](https://hbr.org/2022/11/content-moderation-is-terrible-by-design).

38 Borderline speech refers to language that does not break platform community standards guidelines, but is considered inappropriate in conventional standards. Reclaimed speech refers to words and phrases previously used as slurs or insults against a community that have been taken back by that community for use in empowering or neutral ways.

39 **Heldt, Amélie** 2020. Borderline speech: caught in a free speech limbo?. Internet Policy Review. [policyreview.info/articles/news/borderline-speech-caught-free-speech-limbo/1510](https://policyreview.info/articles/news/borderline-speech-caught-free-speech-limbo/1510).

40 **Endres, Dorothea, Luisa Hedler & Kebene Wodajo** 2023. Bias in Social Media Content Management: What Do Human Rights Have to Do with It? AJIL Unbound. [doi:10.1017/aju.2023.23](https://doi.org/10.1017/aju.2023.23).

41 **Iyer, Ravi** 2022. Content Moderation is a Dead End. Designing Tomorrow. [psychoftech.substack.com/p/content-moderation-is-a-dead-end](https://psychoftech.substack.com/p/content-moderation-is-a-dead-end).



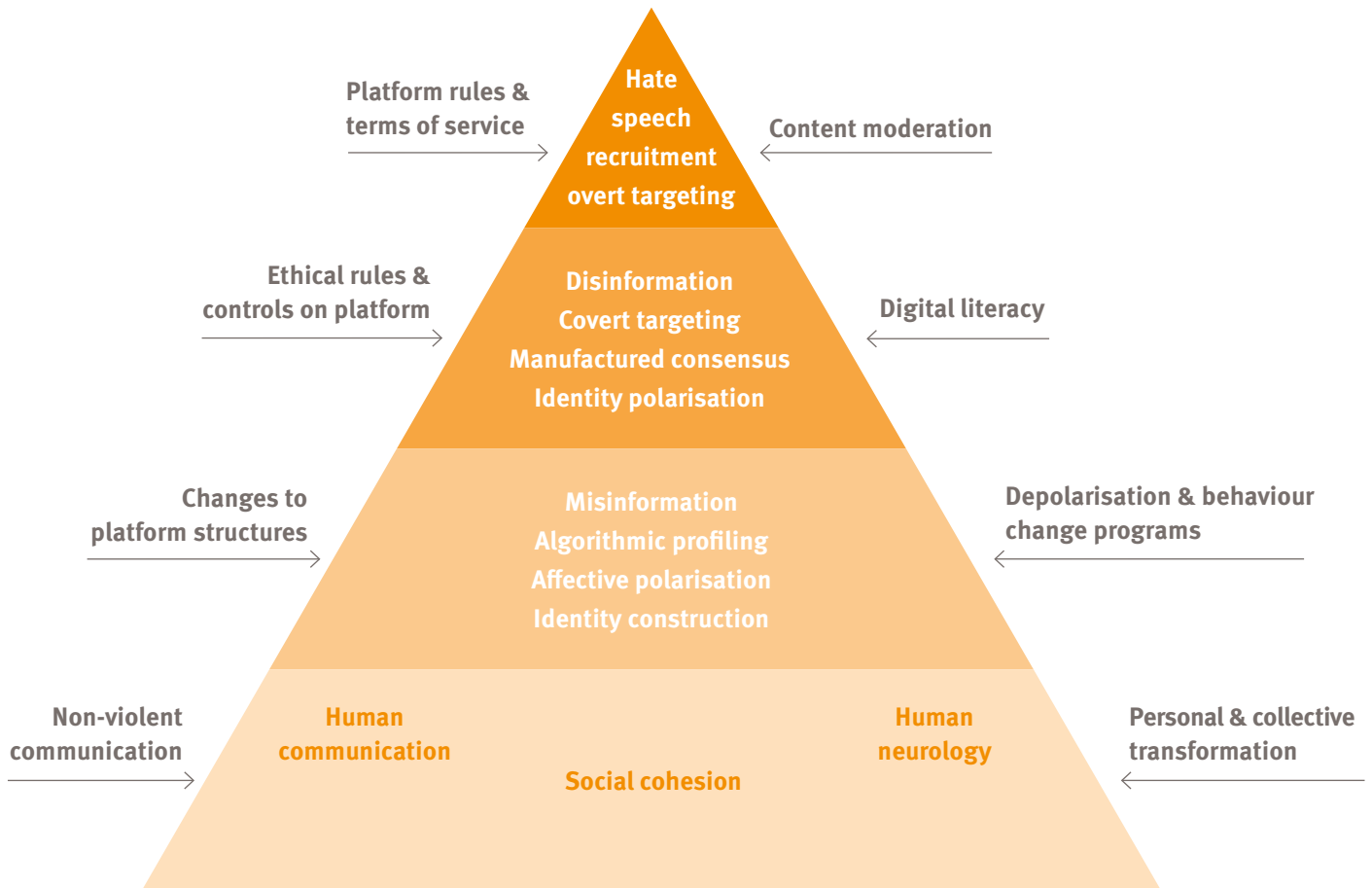


Figure 2: Pyramid of Digital Conflict Drivers and Respective Peacebuilding Responses, adapted from Build UP <sup>42</sup>

our mode of communication, our neurology and, ultimately, how we live together.”<sup>43</sup> As will be discussed later in this paper, this framework aids in demonstrating how understanding the human response to disinformation can help inform practical approaches to combat it.

Peacebuilders have established several invaluable methodologies relevant to issues toward the top of this pyramid, but focusing only at the top prevents the ability to manage the roots of the conflict itself, or those lower layers of the pyramid. As peacebuilders gain more experience combatting disinformation, it’s time to take stock of what has been learned and encourage the peacebuilding community to design programs that address issues further down the pyramid.

42 Puig Larrauri, Helena & Maude Morrison 2022. Understanding Digital Conflict Drivers. In: Hoda Mahmoudi, Michael Allen & Kate Seaman (eds). *Fundamental Challenges to Global Peace and Security*. Palgrave Macmillan. doi.org/10.1007/978-3-030-79072-1\_9.

43 Ibid.

## II. The landscape of peacebuilding interventions against disinformation

Peacebuilders have been working on responding to disinformation's harms for many years, understanding the challenges it proposes to social cohesion, functioning democracies, and conflict patterns. Current approaches include a focus on building communities resilient to polarisation and online harms, such as BuildUp's Digital Maps program in the MENA region,<sup>44</sup> the United Nation's Kinshasa Digital Army,<sup>45</sup> and Search for Common Ground's Digital Community Stewards course.<sup>46</sup> They also include "games for peace" such as Gali Fakta, developed by researchers in the United States and Indonesia.<sup>47</sup> During COVID-19, many peacebuilders aided in the facilitation of online storytelling and fact-checking movements to combat disinformation, as seen in countries such as Burundi<sup>48</sup> and Finland.<sup>49</sup> Peacebuilders have also worked with technical developers to create tools to report disinformation, such as the Sentinel Project's Hagiga Wahid in South Sudan.<sup>50</sup> Prior case studies<sup>51</sup> have examined some of these approaches,

as well as others from the broader disinformation response landscape, and this section aims to add to that discourse.

The approaches below have all identified entry points to combat the challenge of disinformation to varying degrees of success and impact. Despite these efforts, disinformation, and human susceptibility to it, persists. Further, the potential risks of new technologies, particularly Artificial Intelligence (AI), are now gaining the attention of peacebuilders, which will inform and influence future disinformation responses. The following case studies come from interviews with four peacebuilding organisations about their approaches to disinformation.

44 **BuildUp & British Council** 2022. Digital Maps Reports. [howtobuildup.org/wp-content/uploads/2023/02/DMaps\\_Report\\_2022.pdf](https://howtobuildup.org/wp-content/uploads/2023/02/DMaps_Report_2022.pdf).

45 **United Nations** 2023. Building a digital army: UN peacekeepers fight deadly disinformation. [news.un.org/en/story/2023/08/1139682](https://news.un.org/en/story/2023/08/1139682).

46 **Priscilla, Dharini** 2022. About the Digital Community Stewards Training. [cnxus.org/resource/about-the-digital-community-stewards-training/](https://cnxus.org/resource/about-the-digital-community-stewards-training/).

47 **Facciani, Matthew, Denisa Apriliawati & Tim Weninger** 2024. Playing Gali Fakta inoculates Indonesian participants against false information. [misinfocoreview.hks.harvard.edu/article/playing-gali-fakta-inoculates-indonesian-participants-against-false-information/](https://misinfocoreview.hks.harvard.edu/article/playing-gali-fakta-inoculates-indonesian-participants-against-false-information/).

48 **Anderson, Miriam & Madeline Eskandari** 2024. Fake News and Gendered Public Labor: Burundian Peace Activists Combat COVID-19 Disinformation. [doi.org/10.1093/isr/viae041.8](https://doi.org/10.1093/isr/viae041.8).

49 **Heikkilä, Melissa** 2020. Finland taps social media influencers during coronavirus crisis. [politico.eu/article/finland-taps-influencers-as-critical-actors-amid-coronavirus-pandemic/](https://politico.eu/article/finland-taps-influencers-as-critical-actors-amid-coronavirus-pandemic/).

50 **The Sentinel Project** 2020. Managing Misinformation to Build Peace in South Sudan. [reliefweb.int/report/south-sudan/managing-misinformation-build-peace-south-sudan](https://reliefweb.int/report/south-sudan/managing-misinformation-build-peace-south-sudan).

51 **Bateman, Jon & Dean Jackson** 2024. Countering Disinformation Effectively: An Evidence-Based Policy Guide. Carnegie Endowment for International Peace. [carnegie-production-assets.s3.amazonaws.com/static/files/Carnegie\\_Countering\\_Disinformation\\_Effectively.pdf](https://carnegie-production-assets.s3.amazonaws.com/static/files/Carnegie_Countering_Disinformation_Effectively.pdf).

## Vote: Women and Future of Work

**Where & when:** Pan-African, 2021–present

**Organisation:** Pollicy

**Type of Intervention:** Resiliency Programming

**Interviewed:** Bonnita Nyamwire, Director of Research

Pollicy, a feminist collective of technologists, data scientists, creatives, and academics, conducted a research project ahead of the 2021 general elections in Uganda. It found that women politicians were more likely to experience online violence, disinformation, and hate speech than their male counterparts, and were less likely to use social media platforms, such as Twitter and Facebook, to engage with their constituents.<sup>52</sup> Digging into that online violence, they found that mis- and disinformation as a form of violence tended to utilise harmful gender stereotypes in their narratives, prompting a critical response. This sparked the creation of two programs: Vote: Women and Future of Work.

**Vote: Women** is a program that focuses on women in political leadership, equipping them with digital literacy skills while raising awareness about technology-facilitated gender-based violence, including gendered disinformation.<sup>53</sup> The focus on equipping women with digital literacy skills, as opposed to addressing the perpetrators themselves, came from a recognition that current cybersecurity laws preventing online harassment in countries such as Uganda have not adequately ended the current abuse of women online.<sup>54</sup> Thus, women need to be prepared to protect themselves on the platforms as they currently exist while policymakers can demand legal interventions for better platform regulation in the future. **Vote: Women's** first cohort included 20 women in Uganda

and 20 in Tanzania, who for 6 months engaged in a curriculum on digital safety, digital content creation, stakeholder engagement, and online campaigns. The second round of the program expanded to include 30 women in the cohorts and added Senegal as a participating country given their upcoming elections. One important aspect of this approach is the access these participating women have inside the government. For example, Doreen Nyanjura, the Deputy Lord Mayor of Kampala, has participated in the course and can bring the learnings into her office and her interactions with other female politicians.

The **Future of Work** program takes a similar curriculum-based approach, but focuses on women in the media, given that they are also often exposed to online harms.<sup>55</sup> The content covers digital storytelling, social media management, online advocacy and safety, and content creation. It operates in seven countries, namely Mozambique, the Democratic Republic of the Congo, Uganda, Kenya, Tanzania, Cameroon, and Senegal. The program has also developed an Afro-feminist scorecard<sup>56</sup> for each participating country to assess their progress towards an equitable internet for all, including actions civil society organisations, government, and private sector organisations may take to assist this work.

There are three main challenges faced within both programs. First, limitations in funding mean that the programs primarily reach women in urban areas even though Pollicy's research shows there is a wide gender divide between urban and rural women in terms of digital access and digital literacy. This means that women in rural areas are more susceptible to online harms such as disinformation and technology-facilitated gender-based violence. Second, there are limits to focusing on women and

52 **Kakande, Arthur, Garnett Achieng, Neema Iyer, Bonnita Nyamwire, Sandra Nabulega & Irene Mwendwa** 2021. Amplified Abuse: Report on Online Violence Against Women in the 2021 Uganda General Election. Pollicy. [vawp.pollicy.org/](http://vawp.pollicy.org/).

53 **Pollicy** 2021a. **Vote: Women**. [votewomen.pollicy.org/](http://votewomen.pollicy.org/).

54 **Offiong, Adie Vanessa** 2023. 'People find us easy targets': Women politicians face a torrent of online abuse but say they won't stop their work. CNN. [cnn.com/2023/05/25/africa/uganda-women-politicians-online-abuse-as-equals-intl-cmd/index.html](https://www.cnn.com/2023/05/25/africa/uganda-women-politicians-online-abuse-as-equals-intl-cmd/index.html).

55 **Pollicy** 2021b. **Future of Work**. [pollicy.org/futureofwork/the-idea/](http://pollicy.org/futureofwork/the-idea/).

56 **Nyamwire, Bonnita & Dércio Tsandzana** 2024. Afrofeminist Internet Scorecard. Pollicy. [pollicy.org/futureofwork/wp-content/uploads/2024/08/Afrofem-Internet-Scorecard.pdf](http://pollicy.org/futureofwork/wp-content/uploads/2024/08/Afrofem-Internet-Scorecard.pdf).

not broadening the scope to include girls. Access to these trainings, inclusion in spaces such as Pollicy’s annual DataFest conference, and an increase of in-classroom digital education must increase for younger generations. Third, the majority of the training content has been developed in English, so there is a need to consider languages such as French and Swahili, which are commonly used in countries where the program may be implemented. Looking beyond the African continent, more languages could be adopted to allow for broader

expansion of the program. Bonnita Nyamwire notes that “in most of the global South countries, ... you find that the challenges for women are the same: access to digital technology, a digital gap in use of digital technology, barriers from using smartphones. They are all almost all the same, so a multi-pronged approach to address these barriers, as well as language used in digital literacy training, can be adopted.”<sup>57</sup>

## Games as a resiliency tool against disinformation

To respond directly to the challenge of fake news and disinformation, Pollicy also developed a tool called the Choose Your Own Fake News Game.<sup>58</sup> It is a simple “choose your own adventure” game meant to educate online users on the potential harms of mis- and disinformation. Players choose between three characters and make choices based on localised scenarios. For example, “Jo is a 35-year-old shopkeeper from Kitale who provides phone charging services and loves talking about politics.” Players help Jo navigate election rumours that stoke violence in his neighbourhood. Unfortunately, developing games like this is expensive and hard to scale. Pollicy found that not many people interacted with this game because of the urban-rural digital divide in East Africa.

Similarly, psychologists Jon Roozenbeek and Sander van der Linden developed the Breaking Harmony Square game.<sup>59</sup> They describe the approach as psychological “inoculation,” like a vaccine: by being exposed to techniques used to spread disinformation, players build up resilience to these practices. A research experiment proved the game to be effective at increasing players’ confidence in assessing misinformation and decreasing their chance of resharing it.<sup>60</sup> Despite this positive assessment, researchers do not agree on the overall effectiveness of this approach.<sup>61</sup>

57 Nyamwire, Bonnita, (Director of Research at Pollicy), Personal interview by author. Virtual, November, 28, 2024.

58 Pollicy 2020. Choose Your Own Fake News Game. [chooseyourownfakenews.com/](http://chooseyourownfakenews.com/).

59 Roozenbeek, Jon & Sander van der Linden 2020a. Breaking Harmony Square. [harmonysquare.game/](http://harmonysquare.game/).

60 Roozenbeek, Jon & Sander van der Linden 2020b. Breaking Harmony Square: A game that “inoculates” against political misinformation. Harvard Misinformation Review. [misinforeview.hks.harvard.edu/article/breaking-harmony-square-a-game-that-inoculates-against-political-misinformation/](https://misinforeview.hks.harvard.edu/article/breaking-harmony-square-a-game-that-inoculates-against-political-misinformation/).

61 Modirrousta-Galian, Ariana & Philip Higham 2023. Gamified inoculation interventions do not improve discrimination between true and fake news: Reanalyzing existing research with receiver operating characteristic analysis. *Journal of Experimental Psychology*. [doi.org/10.1037/xge0001395](https://doi.org/10.1037/xge0001395).

## Reducing Online Conflict Community

**Where & when:** Nigeria, 2024

**Organisation:** Mercy Corps

**Type of Intervention:** Stakeholder Convening

**Interviewed:** Adrienne Brooks, Senior Advisor for Technology, Peace, & Governance

In February 2024, Mercy Corps convened the first-ever Reducing Online Conflict Community (ROCC) meeting in Abuja, Nigeria.<sup>62</sup> The convening followed over six months of careful stakeholder mapping of the digital information ecosystem in Nigeria, which ultimately brought together individuals representing over 30 organisations and perspectives. This included representatives from local civil society organisations, international NGOs, social media platforms, fact-checkers, journalists, and influencers. Their collective goal: understanding the online harms in the country, including disinformation, and their shared work in addressing these harms. By bringing together stakeholders who do not traditionally work together, the ROCC allowed each to see their role as part of a greater system, identifying gaps and pathways forward in collaboration.

This goal returned three established outcomes, namely an overall report of Nigeria's information ecosystem,<sup>63</sup> a report on best practices in addressing social media harms,<sup>64</sup> and an advocacy priorities document streamlining the language used by different stakeholders in their various response silos.<sup>65</sup> Each was crafted in the months following the February meeting by the sub-committees established at that session. While the best practices document came together quickly given each representative's extensive history of working in this space, the advocacy document took many months

of conversation to align the various narrative tools each organisation utilises. The intentional time spent on its creation meant that new voices, such as the local influencers, were represented and amplified on a country-wide scale in a new way. This time also built strong, trusting relationships. Members of the group continue to meet regularly, beyond the completion of the project, to explore new collaborations.

**This level of cross-sector networking was the critical success of the convening and key learning for future projects.** The variety of perspectives in the room - a unique approach for Mercy Corps and their peers, who typically work with local peacebuilding organisations - allowed people to understand how their work fits into a broader ecosystem of actors and encouraged them to connect with actors beyond their current focus areas. This approach could be better supported by funders, who typically focus on response efforts such as media literacy training, which, while important, ultimately reinforces existing silos between organisations and misses opportunities for creative collaboration. This systems approach has been seen in other conflict response spaces, such as a renewed focus on resiliency that has brought together representatives from the economic, agricultural, and governance sectors, and would benefit work addressing online harms and their impacts as well.

62 **Mercy Corps** 2024a. The Reducing Online Conflict Community. [mercy Corps.org/research-resources/reducing-online-conflict-community](https://mercy Corps.org/research-resources/reducing-online-conflict-community).

63 **Mercy Corps** 2024b. Nigeria's Digital Peacebuilding Ecosystem. [mercy Corps.org/research-resources/nigerias-digital-peacebuilding-ecosystem](https://mercy Corps.org/research-resources/nigerias-digital-peacebuilding-ecosystem).

64 **Mercy Corps** 2024c. Showing Up for Digital Peace. [mercy Corps.org/research-resources/showing-up-for-digital-peace](https://mercy Corps.org/research-resources/showing-up-for-digital-peace).

65 **Mercy Corps** 2024d. Advocacy Priorities - The Reducing Online Conflict Community. [mercy Corps.org/advocacy-priorities-reducing-online-conflict](https://mercy Corps.org/advocacy-priorities-reducing-online-conflict).

## #vrschwrng and Digital.Truth

**Where & when:** Germany, 2020 – 2024

**Organisation:** Berghof Foundation

**Type of Intervention:** Educational Programming

**Interviewed:** Nicole Rieber, Head of Peace Education Unit

#vrschwrng is an interactive toolkit in development since 2020 by Berghof Foundation which aims to equip youth aged 16 to 20 with skills to identify, address, and emotionally respond to conspiracy theories.<sup>66</sup> It was born out of a previous project from Berghof Foundation on hate speech and disinformation, which identified an uptick in conspiracy theories in Germany with the advent of the COVID-19 pandemic and a subsequent opportunity for response.

Conspiracy theories can be vehicles for disinformation and play a similar role in the deterioration of trust in institutions and news online.<sup>67</sup> Using a common peacebuilding analogy, Nicole Rieber compared conspiracy theories to an onion: the layers represent a series of false facts or rumours related to an event or individual, with the outer layers being smaller falsehoods that build into the inner layers of an abstruse theory. Disinformation can play a role in building these layers by lowering individuals' abilities to identify false information or lowering the barriers to belief in wild theories. Ultimately, **“conspiracy theories are the psychological mechanism addressing the specific needs of people, and of course, disinformation can also do that but doesn't have to,”** as Rieber commented. Many techniques used to respond to conspiracy theories also work against disinformation, namely, critical thinking and resiliency against falsehoods.

These skills are taught in the toolkit's curriculum, which is shared with youth in peer-to-peer workshops based on five primary topics: conspiracy theories, conspiracy theories and their dangers, conspiracy theories and social media, dealing with conflicts and conspiracy theories, and democracy and peaceful coexistence. The content for each session is developed in a participatory process whereby the students tailor the course to relevant information for their context. The workshop has been held about 130 times across Germany, reaching around 3000 youth, and initial external monitoring reports from the University of Mainz are returning positive results.

A critical finding from the first few years of student exchanges was that students did not have spaces at home or school to talk about these topics. Thus, #vrschwrng inspired the Digital.Truth project, which facilitates intergenerational workshops on conspiracy theories and disinformation between youth and their parents or teachers.<sup>68</sup> One iteration of the dialogue revealed that **“young people... [were] afraid of what the parents and the teachers do online because they have no media literacy, and the parents and the teachers were scared of what their children do.”**<sup>69</sup>

The primary challenge to this work is funding, as the current project collaboration for both projects is due to expire at the end of the year. Opportunities for flexible funding are critical to ensuring the lasting impact of these interventions. Additionally, the training sessions are only run for one day, whereas skills such as critical thinking and resiliency need to be practised consistently to take hold in online users. Policymakers and educators should consider implementing these skills in daily curriculums.

According to Rieber, while her media literacy work is critical, peacebuilders need to move beyond these responses and ask grander questions, like “How do

66 Berghof Foundation 2020. #vrschwrng: An Interactive Toolkit Against Conspiracy Theories. [vrschwrng.de/](https://vrschwrng.de/)

67 Xiao, Xizhu, Porismita Borah & Yan Su 2021. The dangers of blind trust: Examining the interplay among social media news use, misinformation identification, and news trust on conspiracy beliefs. *Public Understanding of Science*. [journals.sagepub.com/doi/pdf/10.1177/0963662521998025](https://journals.sagepub.com/doi/pdf/10.1177/0963662521998025).

68 Berghof Foundation 2023. Digital.Truth – A Workshop Offering for Adults. [vrschwrng.de/digitale-wahrheiten/](https://vrschwrng.de/digitale-wahrheiten/).

69 Rieber, Nicole, (Head of Peace Education Unit at Berghof Foundation), Personal interview by author. Virtual, November, 27, 2024.

we want to live together?” if the community aims to ultimately address the challenges of disinformation and conspiracy theories.

## Social Media Peace Agreements, Centre for Humanitarian Dialogue

**Where & when:** Various, 2021 – present  
**Organisation:** Centre for Humanitarian Dialogue (HD)  
**Type of Intervention:** Mediation Dialogues  
**Interviewed:** Karen Banaa, Project Associate

Since the first phase of the genocide against the Rohingya began in Myanmar in 2017, it has been widely accepted that Facebook played a critical role in fomenting hatred against the Rohingya people that encouraged the outsized violence. The UN Special Rapporteur on human rights in Myanmar went as far as to call the platform a “beast” for the harm it had enacted on the country.<sup>70</sup> Centre for Humanitarian Dialogue’s Adam Cooper and Maude Morrison were in Myanmar during the violence and saw this as a critical inflexion point: mediators must begin addressing the impacts of social media amidst their efforts on the ground.

The core idea was to replicate HD’s traditional mediation work with conflict actors to address and mediate online behaviour as well. They conceptualise a “supply chain of disinformation” which has three phases: production, distribution, and consumption. HD’s work sits primarily in the first two phases.

In the production phase, they go straight to the producers of online disinformation – conflict actors themselves – to discuss social media issues or risks and try to establish a set of principles for acceptable online behaviour, as “there are no global norms for how people should behave online in conflict situations; this exists for offline behaviour, but not online.”<sup>71</sup> HD has facilitated these “social media codes of conduct” in Nigeria,<sup>72</sup> Kosovo,<sup>73</sup> Bosnia and Herzegovina,<sup>74</sup> Thailand,<sup>75</sup> and a few other countries where HD is present.

In the distribution phase, HD works to establish relationships with social media platforms - the “distributors” of content - to encourage them to direct more resources towards high-risk contexts, which are often under-resourced within these companies, where harmful social media behaviour exacerbates offline conflict. This work includes using scenario exercises based on HD’s experiences in various conflict contexts to see how platforms and peacemaking organisations could collaborate to prevent and mitigate digital threats to peace. Karen Banaa noted that these exercises have helped platforms and mediators build a shared understanding of the concrete risks social media poses to mediation efforts and mediators themselves, and helped clarify the roles of different teams within social media companies and the tools at their disposal.

This relationship-building with actors across the conflict spectrum is critical, as it allows for more holistic responses to peacemaking from all parties. Ahead of the 2023 general elections in Thailand, HD held a months-long dialogue process between the 30 main political parties in the country and the elections commission, in which they formulated a

70 **BBC** 2018. UN: Facebook has turned into a beast in Myanmar. [bbc.com/news/technology-43385677](https://www.bbc.com/news/technology-43385677).

71 Rieber, Nicole, (Head of Peace Education Unit at Berghof Foundation), Personal interview by author. Virtual, November, 27, 2024.

72 **Centre for Humanitarian Dialogue** 2021a. HD brokers landmark social media peace agreement in central Nigeria. [hdcentre.org/news/hd-brokers-landmark-social-media-peace-agreement-in-central-nigeria/](https://www.hdcentre.org/news/hd-brokers-landmark-social-media-peace-agreement-in-central-nigeria/).

73 **Centre for Humanitarian Dialogue** 2021b. HD secures social media conduct commitments for Kosovo elections. [hdcentre.org/news/hd-secures-social-media-conduct-commitments-for-kosovo-elections/](https://www.hdcentre.org/news/hd-secures-social-media-conduct-commitments-for-kosovo-elections/).

74 **Centre for Humanitarian Dialogue** 2022. HD Citizens’ Charter in Bosnia and Herzegovina sets standards for social media conduct in run-up to elections. [hdcentre.org/news/hd-citizens-charter-in-bosnia-and-herzegovina-sets-standards-for-social-media-conduct-in-run-up-to-elections/](https://www.hdcentre.org/news/hd-citizens-charter-in-bosnia-and-herzegovina-sets-standards-for-social-media-conduct-in-run-up-to-elections/).

75 **EngageMedia** 2023. Thai political parties sign election code of conduct, pledge to uphold human and digital rights. [engagemedia.org/2023/thailand-election-code-conduct/](https://www.engagemedia.org/2023/thailand-election-code-conduct/).

social media annexe to the official elections code of conduct. The annexe contained concrete clauses related to the online behaviour of the candidates, with their adherence monitored through a robust body that included researchers, civic tech groups, election commission members, and digital investigations experts. The approach of this body was based on HD's learnings from Nigeria, Bosnia, and Kosovo;<sup>76</sup> for example, they learned of the need to monitor both the politicians' accounts and their proxy accounts run by local influencers or other third party actors. Each day, the monitoring body reviewed posts from the identified actors, checking for any violations of the clauses. Problematic content that also violated platform standards was reported to social media platforms for review, and the overall findings of the monitoring body were shared with the election commission and key stakeholders.

**Working with the Elections Commission of Thailand demonstrated the importance of including these agreements within a broader governmental or mediation process. This is the key direction in which HD is evolving with its approach to these processes.** They're exploring how dialogue on the topic of social media behaviour can be an entry point for wider mediation efforts as it may be seen as more of a "low-hanging fruit." Engaging in dialogue on these topics can help conflict parties build trust that can lead to further dialogue on more sensitive conflict issues. Now, HD is working to integrate social media into their broader peacemaking efforts, both to limit social media's role in undermining peace efforts and to ensure that conflict actors take more accountability for their actions online.

From Banaa's perspective, **more attention should be given to the production phase of the disinformation supply chain to help prevent such harms from occurring and spreading in the first place**, not just focusing on countering disinformation that is already online. Funders need to manage their expectations of what success will look like given that the field of digital mediation is still relatively new, and provide peacemakers the space to innovate and adapt their work.

As shown, these interventions all took a unique approach to combating disinformation by leveraging peacebuilding skills. Each has already had immediate impacts on their target community, such as the resiliency of the African female political leaders Pollicy has reached and the behaviour regulation of Thai political officials through HD's facilitation of the social media annexe. Each is likely to have continued impacts in the future as well, through the projects that will grow out of the ROCC convening and the critical thinking skills instilled in the students engaged in #vrschwrng. **A critical lesson that many of the programs either directly noted or demonstrated, though, was the need for integrated approaches that get at the larger question of why disinformation persists, rather than primarily combating its impacts.**

Cognitive science can play a role in achieving this goal. This is already seen in peacebuilding approaches such as the game design of Harmony Square and gestured toward in the lower layers of the pyramid framework created by BuildUp and HD. To move peacebuilding responses to disinformation further down the pyramid cited above, the following section explores how cognitive science can help peacebuilders understand human susceptibility to disinformation and how to tailor responses accordingly.

---

76 BuildUp, Humanitarian Dialogue & UN Department of Political and Peacebuilding Affairs 2024. Monitoring of Social Media Provisions in Peace Agreements. [hdcentre.org/wp-content/uploads/2024/05/Final\\_final\\_tabletopoutcomes.pdf](https://hdcentre.org/wp-content/uploads/2024/05/Final_final_tabletopoutcomes.pdf).



## Key learnings from the case studies

1. **There is a need for ecosystem-wide network building, both within peacebuilding organisations and across stakeholder spaces.** To avoid “reinventing the wheel,” all actors engaged in disinformation response should build partnerships across the ecosystem of harms. Examples can be seen in the EU Disinformation Lab,<sup>77</sup> the Council on Technology and Social Cohesion,<sup>78</sup> the CyberPeace Institute,<sup>79</sup> and The Psychology of Technology Research Network.<sup>80</sup>
2. **There is a demonstrated need for adaptive funding for disinformation interventions.** Peacebuilders need consistent, long-term funding for programs that allow them to respond to pervasive harms using the context gained on the ground, rather than short-term grants with burdensome reporting requirements.
3. **Elections are a great hook for engaging attention towards disinformation campaigns and gaining support for responses.** This trend holds for many cases considered in this paper and relates strongly to public concerns about information security and trustworthiness ahead of periods of potential risk, such as elections.
4. **The “buzz” of artificial intelligence shifts focus away from existing programs responding to disinformation.** Multiple interview participants noted the advent of AI has drawn attention away from the current challenges of disinformation when the extent of its impacts is still yet unknown.<sup>81</sup> While an important technology to consider, and one which will undoubtedly impact the disinformation landscape, funders and organisations should continue to prioritise established programs focusing on disinformation. Not only will this allow for the long-term benefits of this work to arise, but also allows for these peacebuilders to address the challenges of AI themselves, as they are already critically aware of its impacts at the local level.

77 **EU Disinfo Lab.** [disinfo.eu/](https://disinfo.eu/) [accessed 14 January 2025].

78 **Council on Technology and Social Cohesion.** What if technology fostered trust and collaboration instead of driving polarization and violence? [techandsocialcohesion.org/](https://techandsocialcohesion.org/).

79 **CyberPeace Institute.** [cyberpeaceinstitute.org/](https://cyberpeaceinstitute.org/) [accessed 14 January 2025].

80 **Psychology of Technology Research Institute.** Mission. [psychoftech.org/about](https://psychoftech.org/about).

81 **Simon, Felix, Sacha Altay & Hugo Mercier** 2023. Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown. *Harvard Misinformation Review*. [misinfreview.hks.harvard.edu/article/misinformation-reloaded-fears-about-the-impact-of-generative-ai-on-misinformation-are-overblown/](https://misinfreview.hks.harvard.edu/article/misinformation-reloaded-fears-about-the-impact-of-generative-ai-on-misinformation-are-overblown/).

### III. What does cognitive science say about disinformation?

As digital technologies have developed, so has research from cognitive science on how these technologies interact with our societies. This knowledge is critical when designing peace interventions for disinformation, as, in addition to addressing the ways those technologies encourage and amplify disinformation, peacebuilders must also address the human-level drivers and impact of disinformation. This was recognised by the United Nations in their new 2.0 strategy for achieving the Sustainable Development Goals, which integrates behavioural science into their programming.<sup>82</sup>

#### Why do humans share false information?<sup>83</sup>

Studies show that relatively few people will intentionally share false information they find online.<sup>84</sup> Rather, disinformation is typically spurred by a handful of users, either individually or in coordinated attacks to pollute information environments. The disinformation spreads when it is received by audiences that are willing to reshare it. This is why it is important to examine why people share false information regardless of intention, as this has implications for disinformation response.

#### 1. Social belonging and reputation building –

One study notes that the sense of group belonging that is provided by social media platforms inadvertently encourages the sharing of false information, as users have a “**psychological need to keep groups informed and to remain connected,**” and will share all information with their groups on social media regardless of its accuracy.<sup>85</sup> Fear of missing out, or “FOMO”, is also positively associated with the spread of false information. This is related to in-group dynamics, or the desire to share gossip online to stay socially relevant to your group(s). Laziness and information fatigue also play a role in disinformation spread, as users want to upkeep their social status online through sharing information, but do not want to put in the effort to fact-check that information.<sup>86</sup>

#### 2. Habitual sharing –

Humans tend to build routines, and this includes our habits of sharing information online. Researchers find that **users online build a habit of sharing information with little consideration for the accuracy of that information**, and these habitual sharers have a tendency to share news irrespective of alignment with their political beliefs.<sup>87</sup> The same study finds that users can also build positive habits and learn

82 **United Nations** 2023. Evolving for Impact: Skills and Culture for Tomorrow. [un.org/two-zero/en](https://un.org/two-zero/en).

83 False information is understood here as inaccurate or misleading information. It is used in place of disinformation in this section because the majority of cognitive science research examines the spread and belief of information through its audience regardless of the original publisher or creator's intent. Therefore, disinformation and misinformation can be equally applied in this work, which is why it was abstracted to false information.

84 **Guess, Andrew, Jonathan Nagler & Joshua Tucker** 2019. Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Scientific Advancements*. [pubmed.ncbi.nlm.nih.gov/30662946/](https://pubmed.ncbi.nlm.nih.gov/30662946/).

85 **Talwar, Shalini, Amandeep Dhir, Dilraj Singh, Gurnam Singh Virk & Jari Salo** 2020. Sharing of fake news on social media: Application of the honeycomb framework and the third-person effect hypothesis. *Journal of Retailing and Consumer Services*. [sciencedirect.com/science/article/pii/S0969698920306433](https://sciencedirect.com/science/article/pii/S0969698920306433).

86 **Talwar, Shalini, Amandeep Dhir, Puneet Kaur, Nida Zafar & Melfi Alrasheedy** 2019. Why do people share fake news? Associations between the dark side of social media use and fake news sharing behavior. *Journal of Retailing and Consumer Services*. [sciencedirect.com/science/article/pii/S0969698919301407?via%3Dihub](https://sciencedirect.com/science/article/pii/S0969698919301407?via%3Dihub).

87 **Ceylan, Gizem, Ian Anderson & Wendy Wood** 2022. Sharing of misinformation is habitual, not just lazy or biased. *PNAS*. [pnas.org/doi/epdf/10.1073/pnas.2216614120](https://pnas.org/doi/epdf/10.1073/pnas.2216614120).

to identify true information.

**3. Trust** – A study on WhatsApp finds that users with high trust in the information found on the app were likely to share that information without checking it for accuracy first.<sup>88</sup> This has led to the social platform taking steps to mitigate widespread information sharing by placing forwarding limits on messages, which has allegedly reduced “highly forwarded” messages by 70 per cent.<sup>89</sup> Another study also finds that high levels of **trust in institutions lead users to further spread unverified information shared from these actors in an act of “blind” trust.**<sup>90</sup>

**4. Outrage** – A recent study found that misinformation evokes more outrage in readers than trustworthy news sources, and, simultaneously, that **outrage provokes the sharing of misinformation at least as much as of trustworthy news and enhances reasons unrelated to information accuracy for sharing misinformation.**<sup>91</sup> These reasons include signalling public loyalty to a moral position or political group, which has a lower social cost barrier of sharing potential misinformation (i.e. “outrageous if true”) given their emotional resonance with viewers. This outrage is often provoked by misinformation centring concerns about the economy, injustice, identity, security, and other emotionally charged topics, understanding well that individuals who resonate with these concerns are seeking a reason for their perceived grievances and policy shortcomings.

**5. “Need for chaos”** – In a nationwide survey of the US, researchers asked participants whether they knowingly shared false information on social media about politics. Of the 14% of respondents who say yes, the authors find a correlation between them and their support for extremist groups, antagonistic characters, and “need for chaos,”<sup>92</sup> defined as a desire to disrupt the current social order based on one’s social context.<sup>93</sup> These users “indiscriminately share hostile political rumors as a way to unleash chaos and mobilise individuals against the established order that fails to accord them the respect that they feel they personally deserve.” Therefore, **there is a relationship between intentional false information sharing and the desire to dispute or change one’s social status.**

### Why are humans susceptible to disinformation?

Myriad factors influence a person’s susceptibility to disinformation, ranging from predisposition to interpersonal community dynamics. Four qualities stand out:

**1. Failure to use critical thinking** – The Cognitive Reflection Test studies whether participants think analytically when presented with a reasoning challenge versus following their gut instincts. One study finds that individuals who score higher for critical thinking on the test are less likely to believe fake news and are better able to discern fake news

88 Talwar, Shalini, Amandeep Dhir, Puneet Kaur, Nida Zafar & Melfi Alrasheedy 2019. Why do people share fake news? Associations between the dark side of social media use and fake news sharing behavior. *Journal of Retailing and Consumer Services*. [sciencedirect.com/science/article/pii/S0969698919301407?via%3Dihub](https://www.sciencedirect.com/science/article/pii/S0969698919301407?via%3Dihub).

89 Porter, Jon 2020. WhatsApp says its forwarding limits have cut the spread of viral messages by 70 percent. *The Verge*. [theverge.com/2020/4/27/21238082/whatsapp-forward-message-limits-viral-misinformation-decline](https://www.theverge.com/2020/4/27/21238082/whatsapp-forward-message-limits-viral-misinformation-decline).

90 van Zoonen, Ward, Vilma Luoma-aho & Matias Lievonon 2024. Trust but verify? Examining the role of trust in institutions in the spread of unverified information on social media. *Computers in Human Behavior*. [sciencedirect.com/science/article/pii/S0747563223003436?via%3Dihub](https://www.sciencedirect.com/science/article/pii/S0747563223003436?via%3Dihub).

91 McLoughlin, Killian, William Brady, Aden Goolsbee, Ben Kaiser, Kate Klonick, M.J. Crockett 2024. Misinformation exploits outrage to spread online. *Science*. [science.org/doi/10.1126/science.adl2829](https://www.science.org/doi/10.1126/science.adl2829).

92 Littrell, Shane, Casey Klofstad, Amanda Diekman, John Funchion, Manohar Murthi, Kamal Premaratne, Michelle Seelig, Daniel Verdear, Stefan Wuchty & Joseph Uscinski 2023. Who knowingly shares false political information online?. *Harvard Misinformation Review*. [misinforeview.hks.harvard.edu/article/who-knowingly-shares-false-political-information-online/](https://misinforeview.hks.harvard.edu/article/who-knowingly-shares-false-political-information-online/).

93 Petersen, Michael Bang, Mathias Osmundsen & Kevin Arceneaux 2023. The “Need for Chaos” and Motivations to Share Hostile Political Rumors. *American Political Science Review*. [cambridge.org/core/journals/american-political-science-review/article/need-for-chaos-and-motivations-to-share-hostile-political-rumors/7E50529B41998816383F5790B6E0545A](https://www.cambridge.org/core/journals/american-political-science-review/article/need-for-chaos-and-motivations-to-share-hostile-political-rumors/7E50529B41998816383F5790B6E0545A).

regardless of political ideology. Thus, **“people fall for fake news because they fail to think; not because they think in a motivated or identity-protective way.”**<sup>94</sup> Further research suggests that **this impact may be limited to Western contexts**, given the finding that there was no relation between a high test score and rejection of epistemically suspect beliefs in Japanese individuals.<sup>95</sup> This both supports the commitment of peacebuilders to localisation when implementing practice and reflects the urgent need for research on disinformation susceptibility from global contexts.

Receptivity to pseudo-profound news is also linked to being “reflexively open-minded,”<sup>96</sup> understood as unquestioningly accepting of a wide variety of claims and information. The researchers argue for the need for education to encourage individuals to switch from reflexive to **reflective open-mindedness which includes the critical questioning of one's beliefs.**

**2. Perception of social consensus** – One study highlights how individual susceptibility is impacted by social cues of group consensus and source cues of perception of trustworthiness. **Visible social cues such as likes, retweets, and comments have an impact when a user perceives those judgments**

**as indicative of wide social consensus.**<sup>97</sup> Implicit social cues endorsing a headline’s reliability, such as comments beneath a post, also have an impact on susceptibility in users who actively pay attention to those cues, but cues that discredit false information do not have a strong impact.

**3. Prior exposure** – Cognitive science posits that **repetition relates to perceived accuracy.** Researchers find that the same principle holds for false information: “A single prior exposure to fake-news headlines was sufficient to measurably increase subsequent perceptions of their accuracy.”<sup>98</sup> This effect did not significantly diminish when study participants learned that the headlines were disputed by fact-checkers, nor when the headlines did not align with their political beliefs. The effect continues with multiple exposures, with a similar study finding that the first two repetitions have the highest effect on perception.<sup>99</sup>

**4. Partisanship and group identity** – There is varying evidence about the impact of political parties on the belief of false information. In some studies, it is proven to be a determining factor for disinformation susceptibility<sup>100</sup> and sharing,<sup>101</sup> a factor particularly heightened both in polarised contexts<sup>102</sup> and for conservative-leaning users,<sup>103</sup>

94 Pennycook, Gordon & David Rand 2019a. Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*. doi.org/10.1016/j.cognition.2018.06.011.

95 Majima, Yoshimasa, Alexander Walker, Martin Turpin & Jonathan Fugelsang 2020. Culture and Epistemically Suspect Beliefs.” *PsyArXiv*. doi.org/10.31234/osf.io/qmtn6.

96 Pennycook, Gordon & David Rand 2019b. Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of Personality*. doi.org/10.1111/jopy.12476.

97 Traberg, Cecilie, Trisha Harjani, Jon Roozenbeek & Sander van der Linden 2024. The persuasive effects of social cues and source effects on misinformation susceptibility. *Scientific Reports*. doi.org/10.1038/s41598-024-54030-y.

98 Pennycook, Gordon, Tyrone Cannon & David Rand 2018. Prior Exposure Increases Perceived Accuracy of Fake News. *American Psychological Association*. doi.org/10.1037/xge0000465.

99 Fazio, Lisa, Raunak Pillai & Deep Patel 2022. The Effects of Repetition on Belief in Naturalistic Settings. *American Psychological Association*. doi.org/10.1037/xge0001211.

100 Pennycook, Gordon & David Rand 2019a. Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*. doi.org/10.1016/j.cognition.2018.06.011.

101 Osmundsen, Mathias, Alexander Bor, Peter Bjerregaard Vahlstrup, Anja Bechmann & Michael Bang Petersen 2021. Partisan Polarization Is the Primary Psychological Motivation behind Political Fake News Sharing on Twitter. *American Political Science Review*. doi.org/10.1017/S0003055421000290.

102 Pereira, Andrea, Elizabeth Harris & Jay van Bavel 2021. Identity concerns drive belief: The impact of partisan identity on the belief and dissemination of true and false news. *Group Processes & Intergroup Relations*. doi.org/10.1177/13684302211030004.

103 Nikolov, Dimitar 2021. Right and left, partisanship predicts (asymmetric) vulnerability to misinformation. *Harvard Misinformation Review*. [misinformationreview.hks.harvard.edu/article/right-and-left-partisanship-predicts-asymmetric-vulnerability-to-misinformation/](https://misinformationreview.hks.harvard.edu/article/right-and-left-partisanship-predicts-asymmetric-vulnerability-to-misinformation/).

yet in others, it is tempered by other factors.<sup>104</sup> Some studies find that users are likely to accept false information when it aligns with their political beliefs<sup>105</sup> or that a user's perception of accuracy is biased toward partisan leaning.<sup>106</sup> While the extent to which political beliefs impact susceptibility to false information is varied, the amount of research on this topic, coupled with the above findings on social identity, suggests that this is a factor which should be considered in disinformation response.

**5. Reliance on emotion** – Research has shown that **individuals who rely more on emotion over reason to discern information validity are more susceptible to fake news.**<sup>107</sup> Fake news, understood here as “demonstrably false information that is published and circulated as truth in service of a political or economic agenda,” can be understood as a form of disinformation given the intent of the circulating actor.<sup>108</sup> Prior studies have shown that fake news typically contains negative emotional language<sup>109</sup> and is capable of influencing its viewer's behaviour.<sup>110</sup> Research has also shown that social media is designed to manipulate users' emotions to retain attention<sup>111</sup> and that online users are more drawn to negative content online.<sup>112</sup> Thus, the potential harm of fake news on emotionally-driven online users is high, and must be considered in response.

## How does this inform disinformation response?

There are many categories and motivations of actors involved in disinformation, from those who generate false information to those who disseminate it, knowingly or otherwise. They may be motivated by political ambition, a sense of injustice, financial incentives, or a wish to perpetuate “chaos”. Since these actors have differing goals, it is necessary to tailor peacebuilding efforts towards the motivation of the actor.

**1. Disinformation generators** – As defined previously in this paper, disinformation generators are individuals, countries, or organisations who create and disseminate disinformation for reasons such as influencing domestic politics, disrupting foreign affairs, making significant financial gain, and advancing a certain ideological or moral agenda. Peacebuilding efforts must get to the roots of the conflict in which these actors are engaged to understand the structural factors which are encouraging them to foment conflict. This is conflict transformation at its heart, and it requires much effort in research, stakeholder mapping, narrative work, and mediation and dialogue.

Through rigorous stakeholder mapping and dialogue, peacebuilders can identify disinformation generators at a local level, then work to facilitate dialogues with these conflict actors to outline their

104 **Weeks, Brian** 2015. Emotions, Partisanship, and Misperceptions: How Anger and Anxiety Moderate the Effect of Partisan Bias on Susceptibility to Political Misinformation. *Journal of Communication*. doi.org/10.1111/jcom.12164.

105 **Traberg, Cecilie, Trisha Harjani, Jon Roozenbeek & Sander van der Linden** 2024. The persuasive effects of social cues and source effects on misinformation susceptibility. *Scientific Reports*. doi.org/10.1038/s41598-024-54030-y.

106 **Stein, Jonas, Marc Keuschnigg & Arnout van de Rijt** 2024. Partisan belief in new misinformation is resistant to accuracy incentives. *PNAS Nexus*. doi.org/10.1093/pnasnexus/pgae506.

107 **Martel, Cameron, Gordon Pennycook & David Rand** 2020. Reliance on emotion promotes belief in fake news. *Cognitive Research: Principles and Implications*. doi.org/10.1186/s41235-020-00252-3.

108 **Zimdars, Melissa & Kembrew McLeod** 2020. Fake News: Understanding Media and Misinformation in the Digital Age. The MIT Press. doi.org/10.7551/mitpress/11807.001.0001.

109 **Taboada, Maite** 2021. Authentic Language in Fake News. *Insights from the Social Sciences*. items.ssr.org/beyond-disinformation/authentic-language-in-fake-news/.

110 **Bastick, Zack** 2021. Would you notice if fake news changed your behavior? An experiment on the unconscious effects of disinformation. *Computers in Human Behavior*. doi.org/10.1016/j.chb.2020.106633.

111 **Center for Humane Technology** 2020. How Social Media Hacks Our Brains. [humanetech.com/insights/how-social-media-hacks-our-brains](https://humanetech.com/insights/how-social-media-hacks-our-brains).

112 **Kohout, Susann, Sanne Kruikemeier & Bert Bakker** 2023. May I have your Attention, please? An eye tracking study on emotional social media comments. *Computers in Human Behavior*. doi.org/10.1016/j.chb.2022.107495.

motivations for disinformation. Potential responses to these motivations can be identified through the facilitation of deliberative democratic dialogues, wherein participants can work together to identify alternative solutions to their needs. Then, peacebuilders can encourage tech companies to adopt these policies or interventions, which requires much trust and relationship building with these platforms.

Where dialogues are not possible, peacebuilders can work on identifying and reporting disinformation actors, strengthening local political institutions, and amplifying alternative sources of information to civilians.

**2. Disinformation spreaders** – For disinformation spreaders, the impact on peacebuilding response is dependent on the intention behind the actor.

For those who knowingly spread disinformation, the concept of outrage and a “need for chaos” must be considered critically here. It is clear that these actors are responding to some level of distrust in the current political or social system, some sense of bias or misgiving that they are subject to arbitrarily. This is why they are susceptible to disinformation generated by foreign or domestic actors who aim to disrupt existing social norms, as these actors prey on and amplify their frustrations and

emotions. Any solution that aims to get to the root of disinformation among this group must therefore address these concerns: **Why do people feel disillusioned by the current social structure? What needs of theirs are not being fulfilled that encourage an openness to disinformation?** This again is conflict transformation work, and responses for both peacebuilders and platforms should leverage the tactics noted above.

For those that unknowingly or unintentionally spread disinformation, peacebuilding response must address their mindless habit of spreading information, blind trust in major institutions, and/or adherence to group culture and need for a positive reputation. Initiatives that address these reasons have been proven to work to varying degrees: media literacy campaigns,<sup>113</sup> prebunking efforts,<sup>114</sup> debunking efforts,<sup>115</sup> accuracy prompts,<sup>116</sup> and content labelling,<sup>117</sup> among others.<sup>118</sup> As much research has emerged in recent years on the efficacy of these intervention tactics, **tech platforms should begin integrating these findings into their designs and policymakers should begin requiring their use and application.** Tools such as the Prosocial Design Network’s Interventions Library have increased the ease of access to this research.<sup>119</sup> On misinformation, they have rated the potential success of various popular interventions, finding that accuracy prompts and community

113 **Breakstone, Joel, Mark Smith, Priscilla Connors, Teresa Ortega, Darby Kerr & Sam Wineburg** 2021. Lateral reading: College students learn to critically evaluate internet sources in an online course. *Harvard Misinformation Review*. [misinforeview.hks.harvard.edu/article/lateral-reading-college-students-learn-to-critically-evaluate-internet-sources-in-an-online-course/](https://misinforeview.hks.harvard.edu/article/lateral-reading-college-students-learn-to-critically-evaluate-internet-sources-in-an-online-course/).

114 **Roozenbeek, Jon, Sander van der Linden & Thomas Nygren** 2020. Prebunking interventions based on “inoculation” theory can reduce susceptibility to misinformation across cultures. *Harvard Misinformation Review*. [misinforeview.hks.harvard.edu/article/global-vaccination-badnews/](https://misinforeview.hks.harvard.edu/article/global-vaccination-badnews/).

115 **Bruns, Hendrik, François Dessart, Michal Krawczyk, Stephan Lewandowsky, Myrto Pantazi, Gordon Pennycook, Philipp Schmid & Laura Smillie** 2024. Investigating the role of source and source trust in prebunks and debunks of misinformation in online experiments across four EU countries. *Scientific Reports*. [doi.org/10.1038/s41598-024-71599-6](https://doi.org/10.1038/s41598-024-71599-6).

116 **Pennycook, Gordon & David Rand** 2022. Accuracy prompts are a replicable and generalizable approach for reducing the spread of misinformation. *Nature Communications*. [doi.org/10.1038/s41467-022-30073-5](https://doi.org/10.1038/s41467-022-30073-5).

117 **Nassetta, Jack & Kimberly Gross** 2020. State media warning labels can counteract the effects of foreign misinformation. *Harvard Misinformation Review*. [misinforeview.hks.harvard.edu/article/state-media-warning-labels-can-counteract-the-effects-of-foreign-misinformation/](https://misinforeview.hks.harvard.edu/article/state-media-warning-labels-can-counteract-the-effects-of-foreign-misinformation/).

118 Media literacy campaigns are social media campaigns that encourage social media users to think about their online habits, such as checking news sources. Prebunking efforts are messages that warn online users of common misinformation narratives with the goal to prevent their susceptibility to those narratives. Debunking efforts warn of these narratives after they have already been disseminated. Accuracy prompts are cues associated with social media posts that encourage users to think about the accuracy of that information. Content labelling are labels applied to social media posts, including fact-checking labels, which show a post has been reviewed, and source credibility labels, which indicate the trustworthiness of the source of a post.

119 **Prosocial Design Network**. Browse interventions. [prosocialdesign.org](https://prosocialdesign.org) [accessed 12 December 2024].

notes are “convincing”; misinformation literacy is “likely”; and fact-checking labels, pre-bunking, and source credibility labels are “tentative.”<sup>120</sup> Similar results were found by the Carnegie Endowment, who report that tactics such as media literacy education, fact-checking, labelling social media content, and counter-messaging strategies all are modestly or significantly researched and proven to have significant to modest effects on the information environment.<sup>121</sup>

As companies work to integrate these interventions, **peacebuilders should improve their programming efforts through integrating lessons from cognitive science that get to the roots of users' interactions online.** This includes actions such as adding reflective open mindedness training to media literacy education programs and facilitating storytelling campaigns that encourage social media users to critically think about the institutions they trust for their information.

**3. Disinformation recipients** – Disinformation recipients are those individuals susceptible to believing disinformation for reasons such as a failure to use critical thinking, partisan or group bias, or prior exposure to false narratives. Much peacebuilding work concentrates in this space: media literacy, counter-messaging and critical thinking campaigns, inoculation games, bridge-building dialogues, and community programming to increase resiliency, among others. Looking at cognitive science, one suggestion is that this work could be improved with increased localised research examining misinformation susceptibility across countries and cultures, as intervention effectiveness can vary. Within partisan contexts,

**peacebuilders can further lend unique support to identifying language that resonates with various parties and facilitate bridge-building dialogues and campaigns to encourage a more holistic information environment.**

Dealing with the challenge of perceived accuracy through repetition is more difficult given that users take in information through many channels and false information can cycle through online and offline spaces. For example, research into encouraging users to pause before resharing shows that it is generally effective in mitigating the spread of false information, but this effect is limited for false news which users have previously seen.<sup>122</sup> Efforts such as the one WhatsApp has integrated to limit the rapid forwarding of information could help with the impact of this effect on one platform, but ultimately, **tech companies will need to seriously consider becoming interoperable in order for them to be able to track users' exposure to information across the information ecosystem.** Interoperability is when companies can transfer and share data across platforms, which would make it possible to track a user's exposure to information across Facebook, Instagram, LinkedIn, and so forth.<sup>123</sup> Policymakers can play a role in developing legislation that requires this sharing of information, such as how the EU included this for messaging apps, but not social media apps, in their 2022 Digital Markets Act.<sup>124</sup> Peacebuilders can assist in advocacy efforts for the inclusion of interoperability requirements – understood here as the work of the moral imagination – and in research on its potential harms, while also continuing to increase civilian access to media literacy training.

120 **Prosocial Design Network** 2025. Primer: What We Know About Effective Misinformation Interventions. [prosocialdesign.org/blog/primer-what-we-know-about-effective-misinformation-interventions](https://prosocialdesign.org/blog/primer-what-we-know-about-effective-misinformation-interventions) [accessed 11 January 2025].

121 **Bateman, Jon & Dean Jackson** 2024. Countering Disinformation Effectively: An Evidence-Based Policy Guide. Carnegie Endowment for International Peace. [carnegie-production-assets.s3.amazonaws.com/static/files/Carnegie\\_Countering\\_Disinformation\\_Effectively.pdf](https://carnegie-production-assets.s3.amazonaws.com/static/files/Carnegie_Countering_Disinformation_Effectively.pdf).

122 **Fazio, Lisa** 2020. Pausing to consider why a headline is true or false can help reduce the sharing of false news. Harvard Misinformation Review. [misinforeview.hks.harvard.edu/article/pausing-reduce-false-news/](https://misinforeview.hks.harvard.edu/article/pausing-reduce-false-news/).

123 **Arnao, Zander** 2022. Why Social Media Needs Mandatory Interoperability. Tech Policy Press. [techpolicy.press/why-social-media-needs-mandatory-interoperability/](https://techpolicy.press/why-social-media-needs-mandatory-interoperability/).

124 **Morton, Fiona M. Scott** 2024. It's time for the European Union to rethink personal social networking. Bruegel. <https://www.bruegel.org/policy-brief/its-time-european-union-rethink-personal-social-networking>.

Ultimately, proactively getting ahead of the disinformation cycle requires addressing the motivations of generators, needs and habits of spreaders, and information identification skillsets of recipients, while simultaneously designing and advocating for better online environments that cannot be easily weaponized by conflict actors. The graph below outlines this work.

As shown in the overlapping spaces on this map, there is much work left to be done in terms of responding to the challenge of disinformation and the many actors with which to engage in this process. The final section aims to leverage the findings from the case studies and cognitive science review to outline these critical next steps.



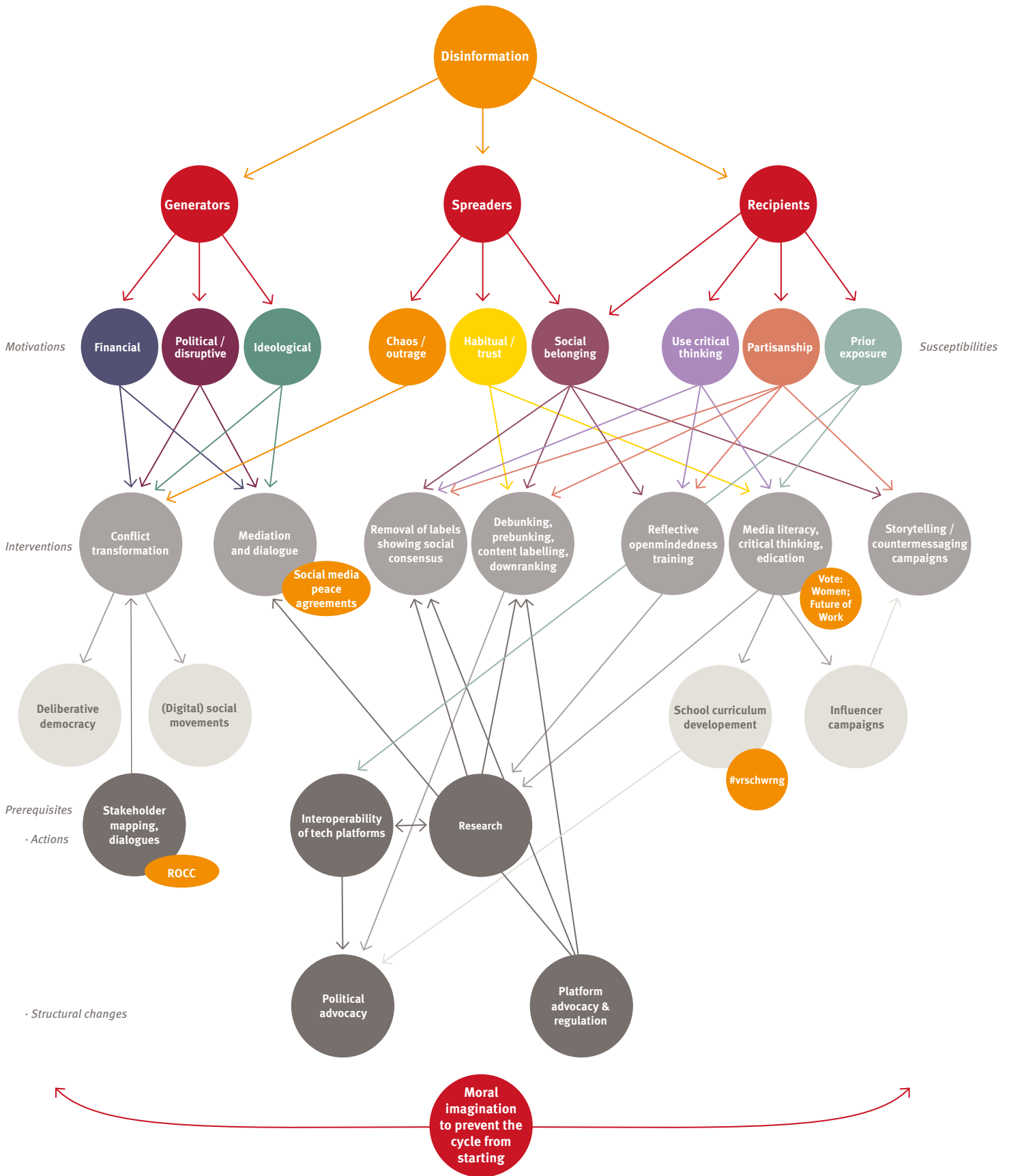


Figure 3: Mapping Peacebuilding Disinformation Interventions by Actor Type

## IV. Conclusion and recommendations

This study looks back on peacebuilding approaches to disinformation to take stock of what has been tried and document lessons learned to make future work more impactful. It has been shown that peacebuilding brings unique skills—such as conflict transformation, mediation and dialogue, reconciliation, storytelling, and moral imagination—to the fight against disinformation. The impact of these skills is amplified when intersecting with other sectors, like technology, sociology, democracy, and security. From this work, an **overarching key learning emerges: the need to take an integrated and interdisciplinary approach to shut down the supply chain of disinformation.**

Thus, the following recommendations centre on how to build these integrated networks across peacebuilding, technology, policy, and social science sectors. Essential to this approach is embracing creative and transformative responses to disinformation that ask deep questions about societal values and coexistence. Peacebuilding responses that engage with these questions contribute to reshaping the broader social conditions that will hamper the success of disinformation.



### 1. Encourage a systems-based strategy for addressing disinformation.

Persistent strains on funding and shrinking academic spaces, as seen in the dismantlement of the Stanford Internet Observatory,<sup>125</sup> emphasise that cross-sector partnerships will be key to the evolution of disinformation response.

- Pool resources and strategies by collaborating through “anti-disinformation” ecosystems like those mentioned in the case studies.
- Hold digital platforms accountable for their role in amplifying false information through regional policy instruments, such as the European Digital Services Act.
- Push donors to provide long-term, flexible funding to ensure the sustainability and continued effectiveness of interdisciplinary approaches.



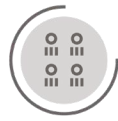
### 2. Leverage cognitive science to prioritise preventative programming at the source of disinformation.

In the supply chain of disinformation, more attention needs to be paid to the initial production and dissemination phases, examining both why disinformation is created and why people are susceptible to it. This informs both the approach of peacebuilders and of technology platforms.

- Move beyond “firefighting” initiatives that address the *repercussions* of disinformation to tackle the root causes and motivations for it.
- Localise programs to address the specific economic or political conflict drivers that make communities vulnerable to disinformation or that make individuals more open to intentionally spreading it, for either financial or social reasons.

125 Newton, Casey & Zoë Schiffer 2024. The Stanford Internet Observatory is being dismantled. Platformer. [platformer.news/stanford-internet-observatory-shutdown-stamos-diresta-sio/](https://platformer.news/stanford-internet-observatory-shutdown-stamos-diresta-sio/).

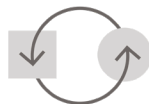
- Include a gender analysis lens in preventative programming and research.



### 3. Aim for gender-transformative approaches to disinformation response.

Because the nature of disinformation is to amplify harm and take advantage of vulnerability, it innately contributes to gender and minority discrimination. Disinformation mitigation strategies should both protect and increase the resilience of women and marginalised groups.

- Increase access to intervention programs for women and girls, such as the one facilitated by Pollicy in the case study.
- Further research and address the role of masculinities and misogyny in disinformation production.
- Push for platform design changes to make digital spaces more inclusive and equitable so they are less susceptible to spreading gendered disinformation.



### 4. Integrate strategies into broader mediation and policy efforts.

In many cases, a practical effort to reduce disinformation serves a dual purpose when it is part of a larger strategy. As noted in the Centre for Humanitarian Dialogue case study, social media behaviour can be “low hanging fruit” in a dialogue process, while greatly contributing to the reduction of disinformation spread by conflict parties. Likewise, media literacy campaigns achieve better impact at scale.

- Utilise codes of conduct as a trust-building exercise in dialogues with conflict parties while

mitigating disinformation generation.

- Integrate media literacy and inoculation games into national education programs, such as the #vrschwrrng toolkit and Vote: Women programs, to ensure these skills are practised consistently.



### 5. Decolonise disinformation work by investing in local research and practice.

Disinformation looks different across countries and communities, both in its content and drivers. Despite this, most current disinformation interventions and the research behind them, including in cognitive science, largely originate from Global North countries,<sup>126</sup> leading to their limited success outside of these contexts. At the same time, better dissemination of peacebuilding interventions and lessons against disinformation is needed so that organisations can build on each other’s work, adapt to their contexts, and accelerate their impact. It can also be difficult for smaller localised programs, although proven to be effective, to gain broader attention and secure long-term funding.

- Amplify localised peacebuilding work against disinformation so that other regions can benefit from established building blocks and avoid reinventing the wheel.
- Support practical efforts to adapt proven peacebuilding strategies to the localised needs of similar contexts, as shared in the Pollicy case study.



### 6. Advocate for and incentivise tech platform product and policy changes to reduce disinformation.

There is extensive public research demonstrating the effectiveness of popular disinformation

126 Lai, Samantha 2024. Understanding the Information Environment: Insights from the Majority World. Carnegie Endowment. [carnegieendowment.org/research/2024/12/global-north-majority-world-misinformation?lang=en&center=india](https://carnegieendowment.org/research/2024/12/global-north-majority-world-misinformation?lang=en&center=india).

interventions, such as accuracy prompts, community notes, content labelling, and counter-messaging strategies. Tech companies need to use this research to improve their designs to prevent the harmful impacts of disinformation on their platforms. Article 35 of the EU's Digital Services Act has begun this process by requiring online platforms to test and adapt their algorithmic systems, including their recommender structures.<sup>127</sup> Future legislation can build upon and strengthen these requirements, aiming to prevent the ability for platform leadership to massively disrupt the information environment, as seen with Meta's choice to end its fact-checking program.

- Push for policies that pressure technology platforms to integrate proven strategies against disinformation in their designs, share learnings from internal research, and give access to their data.
- Continue researching how specific product changes can protect trusted information spaces and diminish the spread of disinformation.
- Create policies that incentivise innovations in prosocial communication, such as examining the potential impact of platform interoperability in reducing the spread of disinformation.
- Reward technology platforms that prove to contribute to prosocial communication.



## 7. Imagine and design new prosocial tech platforms and build bridges with the technology community.

The design affordances, values, moderation practices, and economic structures of major

tech platforms are conducive for disinformation. In addition to advocating for changes to the existing platforms, peacebuilders can contribute to designing new communication technologies and business models that disrupt disinformation supply chains. Bluesky, a new competitor to X that prioritises user autonomy and transparency,<sup>128</sup> is one example, among others like New\_Public,<sup>129</sup> or Sparkable.<sup>130</sup>

Building alternatives to the major social media platforms could help to diversify information sources, present alternative economic and design models to the public who can demand widespread adoption, and reduce the need to collaborate with uncooperative tech platforms. That being said, practitioners and policymakers should be wary of techno-solutionism. No technology or platform on its own will prevent conflict.

- Support peacetechnology incubators and foster local tech for good ecosystems.
- Build bridges between technologists and peacebuilders to help translate between perspectives and expertise. For example, peacebuilders need help adapting disinformation programming to navigate new threats and appeal to donors, and technologists need to understand conflict contexts to code more conflict sensitive safeguards into their technologies.
- Work across disciplines to tackle thorny issues like unpacking the opportunities and harms of advances in AI.

127 **European Union** 2022. Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act). [eu-digital-services-act.com/Digital\\_Services\\_Act\\_Article\\_35.html](https://eur-lex.europa.eu/eli/reg/2022/2065/oj).

128 **Sircar, Anisha** 2024. Bluesky Vs. X: Can The Decentralized Platform Dethrone Elon Musk's Revamped Twitter? *Forbes*. [forbes.com/sites/anishasircar/2024/11/21/bluesky-vs-x-can-the-decentralized-platform-dethrone-elon-musk-x-twitter/](https://www.forbes.com/sites/anishasircar/2024/11/21/bluesky-vs-x-can-the-decentralized-platform-dethrone-elon-musk-x-twitter/).

129 **New\_Public**. Reimage social media. [newpublic.org/](https://newpublic.org/) [accessed 15 January 2025].

130 **Sparkable**. You won't regret time spent here. [sparkable.cc/](https://sparkable.cc/) [accessed 15 January 2025].

# Annexes

## A – List of interviewees

Banaa, Karen, (Project Associate at Centre for Humanitarian Dialogue), Personal interview by author, November, 29, 2024.

Brooks, Adrienne, (Senior Advisor for Technology, Peace, & Governance at Mercy Corps), Personal interview by author. Virtual, November, 15, 2024.

Nyamwire, Bonnita, (Director of Research at Pollicy), Personal interview by author. Virtual, November, 28, 2024.

Rieber, Nicole, (Head of Peace Education Unit at Berghof Foundation), Personal interview by author. Virtual, November, 27, 2024.

## B – The authors

**Grace Connors** is a research assistant at the Do Good Institute at the University of Maryland, where she researches the intersection of philanthropy, nonprofits, and social impact. She also researches the proliferation of hate speech on social platforms in the PeaceTech and Polarization Lab at the University of Notre Dame. Her prior roles have included work in strategic communications as a Scoville Peace Fellow, monitoring and evaluation, and software engineering. Grace is passionate about the digital peacebuilding field, inspired by her BA in Computer Science and Peace Studies from the University of Notre Dame.

**Emma Baumhofer** is an independent consultant working in the fields of digital peacebuilding and interaction design. Before this, she led swisspeace's portfolio on technology, conflict, and peacebuilding, where her work centred on building bridges between technologists and peacebuilders to foster inclusive digital spaces and mitigate conflict. As a consultant Emma has worked with a wide range of international clients, offering expertise in product and community management, strategic communication, and innovation. She holds an MA and MSc in International & Global History from the London School of Economics and Columbia University and a BA in Anthropology and English from UC Berkeley.

# References

- Allen, Jeff** 2022. Misinformation Amplification Analysis and Tracking Dashboard. Integrity Institute. [integrityinstitute.org/blog/misinformation-amplification-tracking-dashboard](https://integrityinstitute.org/blog/misinformation-amplification-tracking-dashboard).
- Allyn, Bobby & Tamara Keith** 2021. Twitter Permanently Suspends Trump, Citing 'Risk Of Further Incitement Of Violence'. NPR. [npr.org/2021/01/08/954760928/twitter-bans-president-trump-citing-risk-of-further-incitement-of-violence](https://www.npr.org/2021/01/08/954760928/twitter-bans-president-trump-citing-risk-of-further-incitement-of-violence).
- Anderson, Miriam & Madeline Eskandari** 2024. Fake News and Gendered Public Labor: Burundian Peace Activists Combat COVID-19 Disinformation. [doi.org/10.1093/isr/viae041.8](https://doi.org/10.1093/isr/viae041.8).
- Arnao, Zander** 2022. Why Social Media Needs Mandatory Interoperability. Tech Policy Press. [techpolicy.press/why-social-media-needs-mandatory-interoperability/](https://techpolicy.press/why-social-media-needs-mandatory-interoperability/).
- Bastick, Zack** 2021. Would you notice if fake news changed your behavior? An experiment on the unconscious effects of disinformation. *Computers in Human Behavior*. [doi.org/10.1016/j.chb.2020.106633](https://doi.org/10.1016/j.chb.2020.106633).
- Bateman, Jon & Dean Jackson** 2024. Countering Disinformation Effectively: An Evidence-Based Policy Guide. Carnegie Endowment for International Peace. [carnegie-production-assets.s3.amazonaws.com/static/files/Carnegie\\_Countering\\_Disinformation\\_Effectively.pdf](https://carnegie-production-assets.s3.amazonaws.com/static/files/Carnegie_Countering_Disinformation_Effectively.pdf).
- BBC** 2018. UN: Facebook has turned into a beast in Myanmar. [bbc.com/news/technology-43385677](https://www.bbc.com/news/technology-43385677).
- Beltran, Sam** 2025. Meta's US fact-checking shutdown sparks fears of disinformation crisis in Asia. *South China Morning Post*. [scmp.com/week-asia/politics/article/3294518/metass-us-fact-checking-shutdown-sparks-fears-disinformation-crisis-asia](https://www.scmp.com/week-asia/politics/article/3294518/metass-us-fact-checking-shutdown-sparks-fears-disinformation-crisis-asia).
- Benaissa Pedriza, Samia** 2021. Sources, Channels and Strategies of Disinformation in the 2020 US Election: Social Networks, Traditional Media and Political Candidates. *Journalism and Media*. [doi.org/10.3390/journalmedia2040036](https://doi.org/10.3390/journalmedia2040036).
- Berghof Foundation** 2020. Berghof Glossary on Conflict Transformation: 20 Notions for Theory and Practice. [yumpu.com/en/document/view/63621335/berghof-glossary-on-conflict-transformation](https://yumpu.com/en/document/view/63621335/berghof-glossary-on-conflict-transformation).
- Berghof Foundation** 2023. Digital.Truth - A Workshop Offering for Adults. [vrschwrng.de/digitale-wahrheiten/](https://vrschwrng.de/digitale-wahrheiten/).
- Berghof Foundation** 2020. #vrschwrng: An Interactive Toolkit Against Conspiracy Theories. [vrschwrng.de/](https://vrschwrng.de/).
- Bradshaw, Samantha, Hannah Bailey & Philip Howard** 2021. Industrialized Disinformation: 2020 Global Inventory of Organised Social Media Manipulation. Working Paper 2021.1. Oxford, UK: Project on Computational Propaganda. [demtech.oii.ox.ac.uk/wp-content/uploads/sites/12/2021/02/CyberTroop-Report20-Draft9.pdf](https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/12/2021/02/CyberTroop-Report20-Draft9.pdf).
- Breakstone, Joel, Mark Smith, Priscilla Connors, Teresa Ortega, Darby Kerr & Sam Wineburg** 2021. Lateral reading: College students learn to critically evaluate internet sources in an online course. *Harvard Misinformation Review*. [misinfreview.hks.harvard.edu/article/lateral-reading-college-students-learn-to-critically-evaluate-internet-sources-in-an-online-course/](https://misinfreview.hks.harvard.edu/article/lateral-reading-college-students-learn-to-critically-evaluate-internet-sources-in-an-online-course/).
- Brooks, Adrienne, Will Ferroggiario & Lisa Inks** 2021. Social Media, Conflict, and Peacebuilding: Issues and Challenges. Mercy Corps. [mercy corps.org/sites/default/files/2022-01/Social-Media-Discussion-Paper-9-Dec-1.pdf](https://mercy corps.org/sites/default/files/2022-01/Social-Media-Discussion-Paper-9-Dec-1.pdf).
- Bruns, Hendrik, François Dessart, Michal Krawczyk, Stephan Lewandowsky, Myrto Pantazi, Gordon Pennycook, Philipp Schmid & Laura Smillie** 2024. Investigating the role of source and source trust in prebunks and debunks of misinformation in online experiments across four EU countries. *Scientific Reports*. [doi.org/10.1038/s41598-024-71599-6](https://doi.org/10.1038/s41598-024-71599-6).
- BuildUp & British Council** 2022. Digital Maps Reports. [howtobuildup.org/wp-content/uploads/2023/02/DMaps\\_Report\\_2022.pdf](https://howtobuildup.org/wp-content/uploads/2023/02/DMaps_Report_2022.pdf).

- BuildUp, Humanitarian Dialogue & UN Department of Political and Peacebuilding Affairs** 2024. Monitoring of Social Media Provisions in Peace Agreements. [hdcentre.org/wp-content/uploads/2024/05/Final\\_final\\_tabletopoutcomes.pdf](https://hdcentre.org/wp-content/uploads/2024/05/Final_final_tabletopoutcomes.pdf).
- Bundtzen, Sara** 2022. “Suggested for You”: Understanding How Algorithmic Ranking Practices Affect Online Discourses and Assessing Proposed Alternatives. Institute for Strategic Dialogue. [isdglobal.org/wp-content/uploads/2022/12/Understanding-How-Algorithmic-Ranking-Practices-Affect-Online-Discourses-and-Assessing-Proposed-Alternatives.pdf](https://isdglobal.org/wp-content/uploads/2022/12/Understanding-How-Algorithmic-Ranking-Practices-Affect-Online-Discourses-and-Assessing-Proposed-Alternatives.pdf).
- Center for Countering Digital Hate** 2022. “The Disinformation Dozen.” [drive.google.com/file/d/1o1j0AzKiGG29Y02UaxzDFy63QgxUQ-fz/view](https://drive.google.com/file/d/1o1j0AzKiGG29Y02UaxzDFy63QgxUQ-fz/view).
- Center for Humane Technology** 2020. How Social Media Hacks Our Brains. [humanetech.com/insights/how-social-media-hacks-our-brains](https://humanetech.com/insights/how-social-media-hacks-our-brains).
- Centre for Humanitarian Dialogue** 2021a. HD brokers landmark social media peace agreement in central Nigeria. [hdcentre.org/news/hd-brokers-landmark-social-media-peace-agreement-in-central-nigeria/](https://hdcentre.org/news/hd-brokers-landmark-social-media-peace-agreement-in-central-nigeria/).
- Centre for Humanitarian Dialogue** 2021b. HD secures social media conduct commitments for Kosovo elections. [hdcentre.org/news/hd-secures-social-media-conduct-commitments-for-kosovo-elections/](https://hdcentre.org/news/hd-secures-social-media-conduct-commitments-for-kosovo-elections/).
- Centre for Humanitarian Dialogue** 2022. HD Citizens’ Charter in Bosnia and Herzegovina sets standards for social media conduct in run-up to elections. [hdcentre.org/news/hd-citizens-charter-in-bosnia-and-herzegovina-sets-standards-for-social-media-conduct-in-run-up-to-elections/](https://hdcentre.org/news/hd-citizens-charter-in-bosnia-and-herzegovina-sets-standards-for-social-media-conduct-in-run-up-to-elections/).
- Ceylan, Gizem, Ian Anderson & Wendy Wood** 2022. Sharing of misinformation is habitual, not just lazy or biased. PNAS. [pnas.org/doi/epdf/10.1073/pnas.2216614120](https://pnas.org/doi/epdf/10.1073/pnas.2216614120).
- CISA. Mis-, Dis-, and Malinformation.** [cisa.gov/sites/default/files/publications/mdm-incident-response-guide\\_508.pdf](https://cisa.gov/sites/default/files/publications/mdm-incident-response-guide_508.pdf).
- Clements, Kevin** 2004. Towards Conflict Transformation and a Just Peace. In: Austin, A., Fischer, M., Ropers, N. (eds) Transforming Ethnopolitical Conflict. VS Verlag für Sozialwissenschaften, Wiesbaden. [doi.org/10.1007/978-3-663-05642-3\\_21](https://doi.org/10.1007/978-3-663-05642-3_21).
- Council on Technology and Social Cohesion.** What if technology fostered trust and collaboration instead of driving polarization and violence? [techandsocialcohesion.org/](https://techandsocialcohesion.org/).
- Cryst, Elena, Esteban Ponce de León, Daniel Suárez Pérez & Shelby Perkins** 2022. Bolivarian Factions: Facebook Takes Down Inauthentic Assets. Stanford Internet Observatory. [purl.stanford.edu/qb823mb8849.ceinstitute.org/](https://purl.stanford.edu/qb823mb8849.ceinstitute.org/) [accessed 14 January 2025].
- CyberPeace Institute.** [cyberpeaceinstitute.org/](https://cyberpeaceinstitute.org/) [accessed 14 January 2025].
- Deb, Anamitra, Stacy Donohue & Tom Glaisyer** 2017. Is Social Media a Threat to Democracy?. The Omidyar Group. [omidyargroup.com/wp-content/uploads/sites/7/2017/10/Social-Media-and-Democracy-October-5-2017.pdf](https://omidyargroup.com/wp-content/uploads/sites/7/2017/10/Social-Media-and-Democracy-October-5-2017.pdf).
- Endres, Dorothea, Luisa Hedler & Kebene Wodajo** 2023. Bias in Social Media Content Management: What Do Human Rights Have to Do with It? AJIL Unbound. [doi:10.1017/aju.2023.23](https://doi.org/10.1017/aju.2023.23).
- EngageMedia** 2023. Thai political parties sign election code of conduct, pledge to uphold human and digital rights. [engagemedia.org/2023/thailand-election-code-conduct/](https://engagemedia.org/2023/thailand-election-code-conduct/).
- EU Disinfo Lab.** [disinfo.eu/](https://disinfo.eu/) [accessed 14 January 2025].
- European Union** 2022. Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act). [eu-digital-services-act.com/Digital\\_Services\\_Act\\_Article\\_35.html](https://eu-digital-services-act.com/Digital_Services_Act_Article_35.html).
- Facciani, Matthew, Denisa Apriliawati & Tim Weninger** 2024. Playing Gali Fakta inoculates Indonesian participants against false information. [misinforeview.hks.harvard.edu/article/playing-gali-fakta-inoculates-indonesian-participants-against-false-information/](https://misinforeview.hks.harvard.edu/article/playing-gali-fakta-inoculates-indonesian-participants-against-false-information/).
- Fazio, Lisa** 2020. Pausing to consider why a headline is true or false can help reduce the sharing of false news. Harvard Misinformation Review. [misinforeview.hks.harvard.edu/article/pausing-reduce-false-news/](https://misinforeview.hks.harvard.edu/article/pausing-reduce-false-news/).
- Fazio, Lisa, Raunak Pillai & Deep Patel** 2022. The Effects of Repetition on Belief in Naturalistic Settings. American Psychological Association. [doi.org/10.1037/xge0001211](https://doi.org/10.1037/xge0001211).

- Galtung, Johan** 1967. *Theories of Peace: A Synthetic Approach to Peace Thinking*. Oslo: International Peace Research Institute.
- Guess, Andrew, Jonathan Nagler & Joshua Tucker** 2019. Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Scientific Advancements*. [pubmed.ncbi.nlm.nih.gov/30662946/](https://pubmed.ncbi.nlm.nih.gov/30662946/).
- Heikkilä, Melissa** 2020. Finland taps social media influencers during coronavirus crisis. [politico.eu/article/finland-taps-influencers-as-critical-actors-amid-coronavirus-pandemic/](https://politico.eu/article/finland-taps-influencers-as-critical-actors-amid-coronavirus-pandemic/).
- Heldt, Amélie** 2020. Borderline speech: caught in a free speech limbo?. *Internet Policy Review*. [policyreview.info/articles/news/borderline-speech-caught-free-speech-limbo/1510](https://policyreview.info/articles/news/borderline-speech-caught-free-speech-limbo/1510).
- Hook, Kristina & Ernesto Verdeja** 2022. Social Media Misinformation and the Prevention of Political Instability and Mass Atrocities. *Stimson Center*. [stimson.org/2022/social-media-misinformation-and-the-prevention-of-political-instability-and-mass-atrocities/](https://stimson.org/2022/social-media-misinformation-and-the-prevention-of-political-instability-and-mass-atrocities/).
- Iyer, Ravi** 2022. Content Moderation is a Dead End. *Designing Tomorrow*. [psychoftech.substack.com/p/content-moderation-is-a-dead-end](https://psychoftech.substack.com/p/content-moderation-is-a-dead-end).
- Jayakumar, Kirthi** 2015. Storytelling for peace. *Peace Insight*. [peaceinsight.org/en/articles/storytelling-peace/?location=&theme=culture-media-advocacy](https://peaceinsight.org/en/articles/storytelling-peace/?location=&theme=culture-media-advocacy).
- Kakande, Arthur, Garnett Achieng, Neema Iyer, Bonnita Nyamwire, Sandra Nabulega & Irene Mwendwa** 2021. Amplified Abuse: Report on Online Violence Against Women in the 2021 Uganda General Election. *Pollicy*. [vawp.pollicy.org/](https://vawp.pollicy.org/).
- Klotsonis, David** 2024. Trusted Flaggers in the DSA: Challenges and Opportunities. *Center for Democracy and Technology*. [cdt.org/insights/trusted-flaggers-in-the-dsa-challenges-and-opportunities/](https://cdt.org/insights/trusted-flaggers-in-the-dsa-challenges-and-opportunities/).
- Knopf, Payton, Andrews Atta-Asamoah, Andrei Kolmakov, Ann Oosterlinck & Klem Ryan** 2016. Letter dated 15 November 2016 from the Panel of Experts on South Sudan established pursuant to Security Council resolution 2206 (2015) addressed to the President of the Security Council. *United Nations Security Council*. [documents.un.org/doc/undoc/gen/n16/350/68/pdf/n1635068.pdf](https://documents.un.org/doc/undoc/gen/n16/350/68/pdf/n1635068.pdf).
- Kohout, Susann, Sanne Kruikemeier & Bert Bakker** 2023. May I have your Attention, please? An eye tracking study on emotional social media comments. *Computers in Human Behavior*. [doi.org/10.1016/j.chb.2022.107495](https://doi.org/10.1016/j.chb.2022.107495).
- Krafft, Peaks & Joan Donovan** 2020. Disinformation by Design: The Use of Evidence Collages and Platform Filtering in a Media Manipulation Campaign. *Political Communication*. [doi.org/10.1080/10584609.2019.1686094](https://doi.org/10.1080/10584609.2019.1686094).
- Kriesberg, Louis, Terrell Northrup & Stuart Thorson** 1989. *Intractable Conflicts and Their Transformation*. Syracuse University Press.
- Lai, Samantha** 2024. Understanding the Information Environment: Insights from the Majority World. *Carnegie Endowment*. [carnegieendowment.org/research/2024/12/global-north-majority-world-misinformation?lang=en&center=india](https://carnegieendowment.org/research/2024/12/global-north-majority-world-misinformation?lang=en&center=india).
- Lamberty, Pia & Lea Frühwirth** 2023. Information manipulation as a complex challenge. *CeMAS*. [cemas.io/en/publications/integrative-model-disinformation/2023-06-14\\_Policy\\_Brief\\_Integrative\\_model\\_disinformation.pdf](https://cemas.io/en/publications/integrative-model-disinformation/2023-06-14_Policy_Brief_Integrative_model_disinformation.pdf).
- Lederach, John Paul** 2005. *The Moral Imagination: The Art and Soul of Building Peace*. New York: Oxford University Press.
- Littrell, Shane, Casey Klofstad, Amanda Diekman, John Funchion, Manohar Murthi, Kamal Premaratne, Michelle Seelig, Daniel Verdear, Stefan Wuchty & Joseph Uscinski** 2023. Who knowingly shares false political information online?. *Harvard Misinformation Review*. [misinforeview.hks.harvard.edu/article/who-knowingly-shares-false-political-information-online/](https://misinforeview.hks.harvard.edu/article/who-knowingly-shares-false-political-information-online/).
- Majima, Yoshimasa, Alexander Walker, Martin Turpin & Jonathan Fugelsang** 2020. Culture and Epistemically Suspect Beliefs." *PsyArXiv*. [doi.org/10.31234/osf.io/qmtn6](https://doi.org/10.31234/osf.io/qmtn6).
- Martel, Cameron, Gordon Pennycook & David Rand** 2020. Reliance on emotion promotes belief in fake news. *Cognitive Research: Principles and Implications*. [doi.org/10.1186/s41235-020-00252-3](https://doi.org/10.1186/s41235-020-00252-3).



- McLoughlin, Killian, William Brady, Aden Goolsbee, Ben Kaiser, Kate Klonick, M.J. Crockett** 2024. Misinformation exploits outrage to spread online. *Science*. [science.org/doi/10.1126/science.adl2829](https://doi.org/10.1126/science.adl2829).
- Meaker, Morgan** 2024. Russia Is Targeting Germany With Fake Information as Europe Votes. *Wired*. [wired.com/story/european-union-elections-russia-germany-disinformation-campaigns/](https://www.wired.com/story/european-union-elections-russia-germany-disinformation-campaigns/).
- Mercy Corps** 2024a. The Reducing Online Conflict Community. [mercycorps.org/research-resources/reducing-online-conflict-community](https://mercycorps.org/research-resources/reducing-online-conflict-community).
- Mercy Corps** 2024b. Nigeria's Digital Peacebuilding Ecosystem. [mercycorps.org/research-resources/nigerias-digital-peacebuilding-ecosystem](https://mercycorps.org/research-resources/nigerias-digital-peacebuilding-ecosystem).
- Mercy Corps** 2024c. Showing Up for Digital Peace. [mercycorps.org/research-resources/showing-up-for-digital-peace](https://mercycorps.org/research-resources/showing-up-for-digital-peace).
- Mercy Corps** 2024d. Advocacy Priorities - The Reducing Online Conflict Community. [mercycorps.org/advocacy-priorities-reducing-online-conflict](https://mercycorps.org/advocacy-priorities-reducing-online-conflict).
- Miller, Carl** 2022. Elon Musk Embraces Twitter's Radical Fact-Checking Experiment. *Wired*. [wired.com/story/elon-musk-embraces-twitters-radical-crowdsourcing-experiment/](https://www.wired.com/story/elon-musk-embraces-twitters-radical-crowdsourcing-experiment/).
- Modirrousta-Galian, Ariana & Philip Higham** 2023. Gamified inoculation interventions do not improve discrimination between true and fake news: Reanalyzing existing research with receiver operating characteristic analysis. *Journal of Experimental Psychology*. [doi.org/10.1037/xge0001395](https://doi.org/10.1037/xge0001395).
- Morton, Fiona M. Scott** 2024. It's time for the European Union to rethink personal social networking. *Bruegel*. [bruegel.org/policy-brief/its-time-european-union-rethink-personal-social-networking](https://bruegel.org/policy-brief/its-time-european-union-rethink-personal-social-networking).
- Nassetta, Jack & Kimberly Gross** 2020. State media warning labels can counteract the effects of foreign misinformation. *Harvard Misinformation Review*. [misinforeview.hks.harvard.edu/article/state-media-warning-labels-can-counteract-the-effects-of-foreign-misinformation/](https://misinforeview.hks.harvard.edu/article/state-media-warning-labels-can-counteract-the-effects-of-foreign-misinformation/).
- New\_Public**. Reimage social media. [newpublic.org/](https://newpublic.org/) [accessed 15 January 2025].
- Newton, Casey & Zoë Schiffer** 2024. The Stanford Internet Observatory is being dismantled. *Platformer*. [platformer.news/stanford-internet-observatory-shutdown-stamos-diresta-sio/](https://platformer.news/stanford-internet-observatory-shutdown-stamos-diresta-sio/).
- Nikolov, Dimitar** 2021. Right and left, partisanship predicts (asymmetric) vulnerability to misinformation. *Harvard Misinformation Review*. [misinforeview.hks.harvard.edu/article/right-and-left-partisanship-predicts-asymmetric-vulnerability-to-misinformation/](https://misinforeview.hks.harvard.edu/article/right-and-left-partisanship-predicts-asymmetric-vulnerability-to-misinformation/).
- Noorshams, Nina, Saurabh Verma & Aude Hofleitner** 2020. TIES: Temporal Interaction Embeddings For Enhancing Social Media Integrity At Facebook. *Meta Research*. [research.facebook.com/publications/ties-temporal-interaction-embeddings-for-enhancing-social-media-integrity-at-facebook/](https://research.facebook.com/publications/ties-temporal-interaction-embeddings-for-enhancing-social-media-integrity-at-facebook/).
- Nyamwire, Bonnita & Dércio Tsandzana** 2024. Afrofeminist Internet Scorecard. *Pollicy*. [pollicy.org/futureofwork/wp-content/uploads/2024/08/Afrofem-Internet-Scorecard.pdf](https://pollicy.org/futureofwork/wp-content/uploads/2024/08/Afrofem-Internet-Scorecard.pdf).
- Offiong, Adie Vanessa** 2023. 'People find us easy targets': Women politicians face a torrent of online abuse but say they won't stop their work. *CNN*. [cnn.com/2023/05/25/africa/uganda-women-politicians-online-abuse-as-equals-intl-cmd/index.html](https://www.cnn.com/2023/05/25/africa/uganda-women-politicians-online-abuse-as-equals-intl-cmd/index.html).
- Osmundsen, Mathias, Alexander Bor, Peter Bjerregaard Vahlstrup, Anja Bechmann & Michael Bang Petersen** 2021. Partisan Polarization Is the Primary Psychological Motivation behind Political Fake News Sharing on Twitter. *American Political Science Review*. [doi.org/10.1017/S0003055421000290](https://doi.org/10.1017/S0003055421000290).
- Parent, Deepa & Katie McQue** 2023. 'I log into a torture chamber each day': the strain of moderating social media. *The Guardian*. [theguardian.com/global-development/2023/sep/11/i-log-into-a-torture-chamber-each-day-strain-of-moderating-social-media-india](https://www.theguardian.com/global-development/2023/sep/11/i-log-into-a-torture-chamber-each-day-strain-of-moderating-social-media-india).
- Pennycook, Gordon & David Rand** 2019a. Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*. [doi.org/10.1016/j.cognition.2018.06.011](https://doi.org/10.1016/j.cognition.2018.06.011).
- Pennycook, Gordon & David Rand** 2019b. Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of Personality*. [doi.org/10.1111/jopy.12476](https://doi.org/10.1111/jopy.12476).
- Pennycook, Gordon & David Rand** 2022. Accuracy prompts are a replicable and generalizable approach for reducing the spread of misinformation. *Nature Communications*. [doi.org/10.1038/s41467-022-30073-5](https://doi.org/10.1038/s41467-022-30073-5).

- Pennycook, Gordon, Tyrone Cannon & David Rand** 2018. Prior Exposure Increases Perceived Accuracy of Fake News. American Psychological Association. doi.org/10.1037/xge0000465.
- Pereira, Andrea, Elizabeth Harris & Jay van Bavel** 2021. Identity concerns drive belief: The impact of partisan identity on the belief and dissemination of true and false news. *Group Processes & Intergroup Relations*. doi.org/10.1177/13684302211030004.
- Petersen, Michael Bang, Mathias Osmundsen & Kevin Arceneaux** 2023. The “Need for Chaos” and Motivations to Share Hostile Political Rumors. *American Political Science Review*. cambridge.org/core/journals/american-political-science-review/article/need-for-chaos-and-motivations-to-share-hostile-political-rumors/7E50529B41998816383F5790B6E0545A.
- Pollicy** 2020. Choose Your Own Fake News Game. chooseyourownfakenews.com/.
- Pollicy** 2021a. Vote: Women. votewomen.pollicy.org/.
- Pollicy** 2021b. Future of Work. pollicy.org/futureofwork/the-idea/.
- Porter, Jon** 2020. WhatsApp says its forwarding limits have cut the spread of viral messages by 70 percent. *The Verge*. theverge.com/2020/4/27/21238082/whatsapp-forward-message-limits-viral-misinformation-decline.
- Priscilla, Dharini** 2022. About the Digital Community Stewards Training. cnxus.org/resource/about-the-digital-community-stewards-training/.
- Prosocial Design Network**. Browse interventions. prosocialdesign.org [accessed 12 December 2024].
- Prosocial Design Network** 2025. Primer: What We Know About Effective Misinformation Interventions. prosocialdesign.org/blog/primer-what-we-know-about-effective-misinformation-interventions [accessed 11 January 2025].
- Psychology of Technology Research Institute**. Mission. psychoftech.org/about.
- Puig Larrauri, Helena & Maude Morrison** 2022. Understanding Digital Conflict Drivers. In: Hoda Mahmoudi, Michael Allen & Kate Seaman (eds). *Fundamental Challenges to Global Peace and Security*. Palgrave Macmillan. doi.org/10.1007/978-3-030-79072-1\_9
- Michael Allen & Kate Seaman (eds)**. *Fundamental Challenges to Global Peace and Security*. Palgrave Macmillan. doi.org/10.1007/978-3-030-79072-1\_9.
- Ramis, Jorge** 2025. Latin American Fact-Checkers Brace for Meta's Next Moves. *Wired*. wired.com/story/hispanic-fact-checkers-react-meta-disinformation/
- Roozenbeek, Jon & Sander van der Linden** 2020a. Breaking Harmony Square. harmonysquare.game/.
- Roozenbeek, Jon & Sander van der Linden** 2020b. Breaking Harmony Square: A game that “inoculates” against political misinformation. *Harvard Misinformation Review*. misinfreview.hks.harvard.edu/article/breaking-harmony-square-a-game-that-inoculates-against-political-misinformation/.
- Roozenbeek, Jon, Sander van der Linden & Thomas Nygren** 2020. Prebunking interventions based on “inoculation” theory can reduce susceptibility to misinformation across cultures. *Harvard Misinformation Review*. misinfreview.hks.harvard.edu/article/global-vaccination-badnews/.
- Schiwal, Julia** 2023. A New Approach for Digital Media, Peace and Conflict. United States Institute for Peace. usip.org/publications/2023/02/new-approach-digital-media-peace-and-conflict.
- Scholtens, Michael, Pedro Pizano, Max Karpawich & Guthrie Kuckes** 2024. *The Disinformation Economy*. The Carter Center. cartercenter.org/resources/pdfs/news/peace\_publications/democracy/the-disinformation-economy-mccain-may-2024.pdf.
- Silverman, Craig & Lawrence Alexander** 2016. How Teens In The Balkans Are Duping Trump Supporters With Fake News.” *Buzzfeed News*. buzzfeednews.com/article/craigsilverman/how-macedonia-became-a-global-hub-for-pro-trump-misinfo#.rhNWoPjEbG.
- Simon, Felix, Sacha Altay & Hugo Mercier** 2023. Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown. *Harvard Misinformation Review*. misinfreview.hks.harvard.edu/article/misinformation-reloaded-fears-about-the-impact-of-generative-ai-on-misinformation-are-overblown/.

- Sircar, Anisha** 2024. Bluesky Vs. X: Can The Decentralized Platform Dethrone Elon Musk's Revamped Twitter? *Forbes*. [forbes.com/sites/anishasircar/2024/11/21/bluesky-vs-x-can-the-decentralized-platform-dethrone-elon-musks-x-twitter/](https://forbes.com/sites/anishasircar/2024/11/21/bluesky-vs-x-can-the-decentralized-platform-dethrone-elon-musks-x-twitter/).
- Sparkable**. You won't regret time spent here. [sparkable.cc/](https://sparkable.cc/) [accessed 15 January 2025].
- Stackpole, Thomas** 2022. Content Moderation Is Terrible by Design. *Harvard Business Review*. [hbr.org/2022/11/content-moderation-is-terrible-by-design](https://hbr.org/2022/11/content-moderation-is-terrible-by-design).
- Starbird, Kate** 2019. Disinformation's spread: bots, trolls and all of us. *Nature*. [nature.com/articles/d41586-019-02235-x](https://nature.com/articles/d41586-019-02235-x).
- Stein, Jonas, Marc Keuschnigg & Arnout van de Rijt** 2024. Partisan belief in new misinformation is resistant to accuracy incentives. *PNAS Nexus*. [doi.org/10.1093/pnasnexus/pgae506](https://doi.org/10.1093/pnasnexus/pgae506).
- Taboada, Maite** 2021. Authentic Language in Fake News. *Insights from the Social Sciences*. [items.ssrc.org/beyond-disinformation/authentic-language-in-fake-news/](https://items.ssrc.org/beyond-disinformation/authentic-language-in-fake-news/).
- Talwar, Shalini, Amandeep Dhir, Dilraj Singh, Gurnam Singh Virk & Jari Salo** 2020. Sharing of fake news on social media: Application of the honeycomb framework and the third-person effect hypothesis. *Journal of Retailing and Consumer Services*. [sciencedirect.com/science/article/pii/S0969698920306433](https://sciencedirect.com/science/article/pii/S0969698920306433).
- Talwar, Shalini, Amandeep Dhir, Puneet Kaur, Nida Zafar & Melfi Alrasheedy** 2019. Why do people share fake news? Associations between the dark side of social media use and fake news sharing behavior. *Journal of Retailing and Consumer Services*. [sciencedirect.com/science/article/pii/S0969698919301407?via%3Dihub](https://sciencedirect.com/science/article/pii/S0969698919301407?via%3Dihub).
- The Sentinel Project** 2020. Managing Misinformation to Build Peace in South Sudan. [reliefweb.int/report/south-sudan/managing-misinformation-build-peace-south-sudan](https://reliefweb.int/report/south-sudan/managing-misinformation-build-peace-south-sudan).
- The Syria Campaign** 2017. Killing the Truth. [thesyriacampaign.org/wp-content/uploads/2017/12/KillingtheTruth.pdf](https://thesyriacampaign.org/wp-content/uploads/2017/12/KillingtheTruth.pdf).
- Traberg, Cecilie, Trisha Harjani, Jon Roozenbeek & Sander van der Linden** 2024. The persuasive effects of social cues and source effects on misinformation susceptibility. *Scientific Reports*. [doi.org/10.1038/s41598-024-54030-y](https://doi.org/10.1038/s41598-024-54030-y).
- United Nations** 2023. Building a digital army: UN peacekeepers fight deadly disinformation. [news.un.org/en/story/2023/08/1139682](https://news.un.org/en/story/2023/08/1139682).
- United Nations** 2023. Evolving for Impact: Skills and Culture for Tomorrow. [un.org/two-zero/en](https://un.org/two-zero/en).
- United Nations**. What is hate speech? [un.org/en/hate-speech/understanding-hate-speech/what-is-hate-speech](https://un.org/en/hate-speech/understanding-hate-speech/what-is-hate-speech).
- United States Institute of Peace**. Reconciliation: Truth, Justice, Peace, Mercy. [usip.org/public-education-new/reconciliation-truth-justice-peace-mercy](https://usip.org/public-education-new/reconciliation-truth-justice-peace-mercy).
- van Zoonen, Ward, Vilma Luoma-aho & Matias Lievonen** 2024. Trust but verify? Examining the role of trust in institutions in the spread of unverified information on social media. *Computers in Human Behavior*. [sciencedirect.com/science/article/pii/S0747563223003436?via%3Dihub](https://sciencedirect.com/science/article/pii/S0747563223003436?via%3Dihub).
- Weeks, Brian** 2015. Emotions, Partisanship, and Misperceptions: How Anger and Anxiety Moderate the Effect of Partisan Bias on Susceptibility to Political Misinformation. *Journal of Communication*. [doi.org/10.1111/jcom.12164](https://doi.org/10.1111/jcom.12164).
- Wilson, Tom & Kate Starbird** 2020. Cross-platform disinformation campaigns: Lessons learned and next steps. *Harvard Misinformation Review*. [misinforeview.hks.harvard.edu/article/cross-platform-disinformation-campaigns/](https://misinforeview.hks.harvard.edu/article/cross-platform-disinformation-campaigns/).
- Xiao, Xizhu, Porismita Borah & Yan Su** 2021. The dangers of blind trust: Examining the interplay among social media news use, misinformation identification, and news trust on conspiracy beliefs. *Public Understanding of Science*. [journals.sagepub.com/doi/pdf/10.1177/0963662521998025](https://journals.sagepub.com/doi/pdf/10.1177/0963662521998025).
- Zimdars, Melissa & Kembrew McLeod** 2020. Fake News: Understanding Media and Misinformation in the Digital Age. The MIT Press. [doi.org/10.7551/mitpress/11807.001.0001](https://doi.org/10.7551/mitpress/11807.001.0001).

**Platform Peaceful Conflict  
Transformation**

Großbeerenstr. 13a  
10963 Berlin  
Germany

[www.pzkb.de](http://www.pzkb.de)  
[kontakt@pzkb.de](mailto:kontakt@pzkb.de)



[@plattformzkb.bsky.social](https://bsky.app/profile/plattformzkb)



[/plattformzkb](https://www.linkedin.com/company/plattformzkb)

**Berghof Foundation  
Operations gGmbH**

Lindenstraße 34  
10969 Berlin  
Germany

[www.berghof-foundation.org](http://www.berghof-foundation.org)  
[info@berghof-foundation.org](mailto:info@berghof-foundation.org)



[@berghoffnd.bsky.social](https://bsky.app/profile/berghoffnd)



[@BerghofFnd](https://x.com/BerghofFnd)



[/BerghofFoundation](https://www.facebook.com/BerghofFoundation)



[/berghof-foundation](https://www.linkedin.com/company/berghof-foundation)



[@berghoffnd](https://www.instagram.com/berghoffnd)