

ON THE RADAR:



Mapping the Tools, Tactics and Narratives of Tomorrow's Disinformation Environment



DISINFO
RADAR



DEMOCRACY
REPORTING
INTERNATIONAL

ON THE RADAR:




Mapping the Tools, Tactics and Narratives of Tomorrow's Disinformation Environment

About Democracy Reporting International

DRI is an independent organisation dedicated to promoting democracy worldwide. We believe that people are active participants in public life, not subjects of their governments. Our work centres on analysis, reporting and capacity-building. For this, we are guided by the democratic and human rights obligations enshrined in international law. Headquartered in Berlin, DRI has offices in Lebanon, Libya, Myanmar, Pakistan, Sri Lanka, Tunisia, and Ukraine.

About Disinfo Radar

As part of the Disinfo Radar project, DRI will examine three core pillars of disinformation:

-  Emerging technological tools used to produce disinformation
-  New tactics for propagating manipulated content
-  Untold stories harnessing these tools and tactics to frame false narratives

For more information on the project click [here](#).

Acknowledgements

This report was written by Jan Nicola Beyer, Digital Democracy Research Coordinator, and Lena-Maria Böswald, Digital Democracy Programme Officer. ForSet designed the layout of this publication.

Date: June 2022

This report is part of the Disinfo Radar project funded by the German Federal Foreign Office. Its contents do not necessarily represent the position of the German Federal Foreign Office.



Federal Foreign Office

Table of Contents

1. Executive Summary	4
Tools	4
Tactics	4
Stories	5
Future Developments	5
2. Glossary	7
3. Introduction: The Progress of Disinformation	9
Accelerating Threat Factors	10
4. Tools	13
4.1 Deepfakes	13
4.1.1 Fake Images	13
4.1.2 Synthetic Audio	14
4.1.3 Synthetic Videos	15
4.2 Text Prediction	16
5. Tactics	17
Stories Retold	17
Cross-Platform Sharing	18
Abusing Algorithmic Recommender Systems	19
Coordinated Sharing Behaviour	19
Astroturfing	20
Content Laundering	20
Embedding Social Media into News Coverage	21
Inauthentic News Factories	21
6. Stories	22
The Russian War Against Ukraine: Anti-Western Narratives and the Fog of War	22
Traces of Deepfakes?	25
The COVID-19 Pandemic: Science as a Weapon of Disinformation	26
Elections and Gendered Disinformation: Constraining the Public Sphere	29
7. Predicting Future Threats	33
Approaching the Tipping Point	33
Extending the Global Reach	35
Potential Future Risk Scenarios	36
8. Conclusion	38



TOOLS TACTICS FUTURE DEVELOPMENTS STORIES

Executive Summary

Artificial intelligence (AI) and machine learning (ML) technologies have the potential to deepen the threat already posed by disinformation campaigns. While much research has been done covering individual characteristics of disinformation, little of this has brought together the different technical and tactical elements involved in the fabrication of false narratives.

This report, therefore, focuses on dissecting the major threats to our online information ecosystem into three components: technological tools, popular tactics and the resulting disinformation narratives. In doing so, it identifies potential new disinformation threats and offers forecasts on further developments in the disinformation sphere.



Tools

Current advances in technology hold the potential to reshape the speed, quality and quantity of the spread of disinformation. As it becomes easier and cheaper for amateurs to manipulate videos and images, highly realistic “deepfakes” generated using AI are likely to become common in the future.

There have also been major advances in generating synthetic audio content and synthetic text using natural language generation. While text prediction tools continue to struggle to produce long, logically consistent texts, models have shown a high level of sophistication when applied to short texts, rendering them a potential tool for future disinformation efforts. This will become especially relevant when used in combination with synthetic audio or video. In general, the current practice of combining various synthetic sources might reshape the information playing field and the way informational evidence is created, as can be seen in the rise of text-to-image conversion technology.



Tactics

Despite the advances in disinformation tools, their use can follow long-established forms of disinformation (i.e., misdating, mislocating and misrepresenting content). Newer tactics build on three pillars:

- Preparing content dissemination through inauthentic news factories or proxy sources;
- Fabricating false narratives by embedding false social media posts into (in)authentic news content; and
- Spreading disinformation by cross-platform sharing, coordinated sharing behaviour and the abuse of algorithmic recommender systems

Stories

Looking at three cases where there has been a prevalence of disinformation narratives (disinformation campaigns related to Russia's invasion of Ukraine, the COVID "infodemic" and false narratives during elections), this report examines the evolution of actors, strategies and tactics in the disinformation sphere. The case studies reveal that, in recent disinformation campaigns, innovations in tactics have played a more central role than innovations in tools. The use of synthetic content, in particular, has played a rather peripheral role. In both pro-Kremlin propaganda related to Ukraine and COVID disinformation, actors have drawn on patterns of spreading disinformation that emphasize strategy rather than cutting-edge technology. The only domain in which deep and cheap fakes have become prevalent is gendered disinformation.

Future Developments

Extrapolating from this research, this report identifies the following trends that might reshape the nature of how disinformation is spread in the future:

- **Falling entry barriers.** While the cost of training highly complex ML models requires computational resources that only states or major business actors can provide, the entry barriers to using these models are slowly falling. The proliferation of open-source resources and the decreasing need for large scale datasets used to turn ML-based applications into weapons of disinformation pose a major threat.
- **Advances in prevention also increase disinformation sophistication.** The open-source nature of many detection mechanisms results in a situation in which disinformation actors are steadily learning and improving their toolkit.
- **Quality increases across the board.** A striking development is the ever-evolving research into and investment in AI and ML models. This might lead to the thorough, sophisticated use of deep fake technology and easily accessible tools that do not require significant human resources or hardware.
- **Merging of multiple tools.** A plausible scenario in the foreseeable future would be, for example, the appearance of text generated with natural language processing (NLP) models, complemented with synthetic images or videos. This would allow

disinformation actors to produce more coherent fake evidence that will be increasingly hard to debunk.

- **Authentic coordinated spreading.** With greater advances in automatic detection, disinformation actors might increasingly opt for human sources to circulate content. The recent trend to spread disinformation via influencers is an example of this.

Beyond these imminent future risks, this report identifies long-term scenarios that could reshape the nature of disinformation. These include the use of emotional recognition technology, quantum prime factorisation algorithms, and AI as a seemingly universal, standardised concept.



BOTS ARTIFICIAL
DISINFORMATION INTELLIGENCE
DEEP LEARNING
ASTROTURFING
FACT-CHECKING

Glossary

Artificial Intelligence

The development and ability of computer systems to perform computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages.

Astroturfing

Organised activity on the Internet that is intended to create a false impression of a widespread, spontaneously arising, grassroots movement in support of or in opposition to something (such as a political policy) but that is in reality initiated and controlled by a concealed group or organization.

Bots

Social media accounts that are operated entirely by computer programs and are designed to generate posts and/or engage with content on a particular platform. Researchers and technologists take different approaches to identify bots, using algorithms or simpler rules based on the number of posts per day.

Cheapfakes

Low-level manipulation of audio-visual material (created with accessible software) to alter content.

Coordinated Inauthentic Behaviour

Groups of pages or people work together to mislead others about who they are or what they are doing in the online environment.

Deepfakes

Highly sophisticated manipulation of audio-visual media using AI-driven technology.

Deep Learning

Deep learning is a class of machine learning algorithms that uses multiple layers to progressively extract higher-level features from the raw input.

Disinformation

False information that is deliberately created or disseminated with the express purpose to cause harm.

End-To-End Encryption

End-to-end encryption is a secure communication process that prevents third parties from accessing data transferred from one endpoint to another.

Fact-Checking

The process of determining the truthfulness and accuracy of official published information such as politicians' statements and news reports.

Information Manipulation

The strategies employed by a source or producer of information to deceive the receiver or consumer into interpreting that information in an intentionally false way. The user thinks they are genuinely receiving the information but in fact they are deceived by its manipulation.

Generative Adversarial Networks

A generative adversarial network, or GAN, is a deep neural network framework that is able to learn from a set of training data and generate new data with the same characteristics as the training data.

Machine Learning

The use and development of computer systems that are able to learn and adapt without following explicit instructions, by using algorithms and statistical models to analyse and draw inferences from patterns in data.

Misinformation

Refers to incorrect or misleading information presented as fact, either intentionally or unintentionally.

Recommender System

Recommender systems are a subclass of machine learning that generally deal with ranking or rating products, content or users.

Social Media Amplification

When content is shared, either through organic or paid engagement, within social channels, thereby increasing word-of-mouth exposure. Amplification works by getting content promoted (amplified) through proxies. Each individual sharer extends the messaging to their personal network, who can then promote it to their network and so on.

Sock Puppet

An online account that uses a false identity designed specifically to deceive. Sock puppets are used on social platforms to inflate another account's follower numbers and to spread or amplify content to a mass audience. The term is considered by some to be synonymous with the term "bot".

Synthetic Media

Synthetic media (can refer to audio, text, image or video) refers to data and media artificially produced, manipulated and modified by automated means, especially through the use of artificial intelligence algorithms.



Introduction: The Progress of Disinformation

The term “disinformation” has been on everyone’s lips since long before the recent rise in pro-Kremlin propaganda in the context of Russia’s full-fledged invasion of Ukraine. Disinformation and its related concepts “misinformation and malinformation” are two other sides of the same coin, so to speak, and can create the same impact: undermining the notion of truth.¹

Artificial intelligence (AI), machine learning (ML) and deep learning (DL) techniques have the potential to deepen the threat posed by disinformation campaigns.² With the advancement of regulation and detection mechanisms, disinformation actors are driven to greater lengths to shield their identities and refine the production of false facts or narratives. They accomplish this primarily through the use of advanced tools and tactics to fabricate disinformation and through compelling and bias-confirming narratives. It is vital to understand what the most common tools and tactics employed by disinformation actors to fabricate disinformation are, and which overarching narratives are currently threatening the online information environment.

When this report talks about *tools*, it refers to the **technical foundation of disinformation content** that can appear in (audio-)visual or textual form. This is about the message. *Tactics*, on the other hand, are defined as **techniques or strategies to propagate messages** as subtly as possible. This is about messaging/dissemination. An **amalgamation of tools and tactics** and how they are used to distort facts and manipulate opinions can be studied in the context of disinformation *narratives*. Illuminating case studies embedded within such narratives help to better comprehend the extent of the threat that tools, tactics and stories pose to democratic discourse and processes.

-
- 1 Carme Colomina, Héctor Margalef, and Richard Youngs, [“The Impact of Disinformation on Democratic Processes and Human Rights in the World”](#), European Parliament, April 2021.
 - 2 Noémi Bontridder and Yves Poulet, [“The role of artificial intelligence in disinformation”](#), Cambridge University Press, 25 November 2021.

This report will do the following:

- (a) dissect the major threats to our online information ecosystem into components;
- (b) discuss potential new disinformation threats; and
- (c) forecast developments in the disinformation sphere. Before doing so, it will take a closer look at which political and societal factors foster the spread of disinformation.

Accelerating Threat Factors

Different factors increase the risk that information manipulation poses to society:



State actors have access to advanced AI



Encrypted messenger applications facilitate undetected dissemination



Dynamic information environments conceal disinformation



Passive ecosystems are not properly regulated



Confirmation bias is a root disinformation cause

State disinformation actors have access to advanced AI. Far-reaching foreign influence campaigns have become a “new way of warfare for the 21st century”.³ This new method of warfare has not only been successfully implemented by both Russia and China, but these state actors have excelled in what the Chinese Communist Party calls “external propaganda”. The war in Ukraine has not only led to China and Russia seemingly joining forces in their respective false information warfare⁴ against “the

³ Fiona Hill and Clifford G. Gaddy, *Mr. Putin: Operative in the Kremlin* (Geopolitics in the 21st Century) (Washington, DC: Brookings Institution Press, 2013), Ch. 1.

⁴ David Bandurski, [“China and Russia Are Joining Forces to Spread Disinformation”](#), Brookings, 11 March 2022.

West”, but it has also indirectly allowed them to write their own “technology-driven playbooks for authoritarian rule”⁵ grounded in AI. Compounding the political will to disinform, one can say that AI technology advances more rapidly in less democratic countries, mostly due to the availability of data collected on citizens for the training of AI models. Disinformation actors in these countries have proven their technical capabilities not only by using technology such as deepfakes and bot farms, but also through strategic thinking. For example, Russia has “outsourced” the bulk of its disinformation fabrication to content generation “farms”, which are seemingly state-independent agencies. It also uses verified social media accounts of its international embassies to seed disinformation and avoid content removals from platforms.⁶

Messaging platforms are effective tools for disinformation dissemination.

Popular encrypted messaging platforms (EMAs), or messenger applications based on end-to-end encryption, such as WhatsApp, Telegram and Signal, not only allow for online conversations largely free from surveillance, but also serve as catalysts for more covert disinformation campaigns. This form of messaging among friends and, often, tightly connected networks renders dissemination in a secured environment almost untraceable and invisible.⁷

A dynamic, constantly changing information environment facilitates disinformation concealment.

The ever-evolving social media landscape and trends towards audio and/or visual materials provide disinformation actors with many playing fields, and research, analysis and detection are always struggling to catch up. TikTok, a social network application for sharing short videos with 1 billion active users, has quickly gained dominance as a source of news among young people. Recent research by the Institute for Strategic Dialogue has found that Russian state-controlled media outlets such as RT and Sputnik News have used TikTok’s features to spread disinformation. Their content on TikTok is not blocked or labelled as emanating from a state agency.⁸ Spotify, on the other hand, the world’s leading music streaming service, has been the centre of a controversy, because a host of a top-rated podcast aired false claims and conspiracies about the COVID pandemic.⁹ The promotion of a conspiracy theory claiming that the US is funding the development of dangerous biological weapons in Ukraine by prominent podcasters is the latest example of how podcasting

-
- 5 Alina Polyakova and Chris Meserole, [“Exporting Digital Authoritarianism: The Russian and Chinese Models”](#), Brookings, August 2019.
 - 6 Ciarán O’Connor, [“After Bucha, Here’s What to Expect from the Next Phase of Russian Disinformation in Ukraine”](#), Institute for Strategic Dialogue, 6 May 2022.
 - 7 Jacob Gursky and Samuel Woolley, [“Countering Disinformation and Protecting Democratic Communication on Encrypted Messaging Applications”](#), Brookings, 11 June 2021.
 - 8 Ciaran O’Connor, [“#Propaganda: Russia State-Controlled Media Flood TikTok With Ukraine Disinformation”](#), Institute for Strategic Dialogue, 2 March 2022.
 - 9 Umberto Bacchi, [“Joe Rogan, Spotify: how tech firms tackle Covid misinformation”](#), Thomson Reuters Foundation, 2 February 2022.

has become a largely overlooked catalyst for mis- and disinformation.¹⁰ The nature of podcasting makes oversight challenging and offers the potential for false narratives to spread.

Algorithmic recommender systems, adtech: “passive ecosystems” that are not properly regulated. Recommendation algorithms, which rank and rate content based on data (such as user preferences) and are the foundation of the business model of social media platforms, can also be catalysts for disinformation. In particular, they may rank disinformation high in user feeds. AI-based algorithms determine the order in which a user will see online content. The regulation of algorithmic ranking is in its infancy.¹¹ Also, advertising technology (ad-tech) companies amplify, promote and capitalise off of disinformation in placing advertisements on disinformation platforms.¹² The EU’s Digital Services Act (DSA) will compel very large social media platforms to identify risks of algorithmic ranking and to increase transparency.¹³

The “emotional need to be right”¹⁴ allows disinformation to spread. Confirmation bias is a double-edged sword: It helps us to quickly classify new pieces of information, but it may also lead us to draw the wrong conclusions and hinder us from accepting information that does not conform to our biases. Regardless of how obscure the information we are consuming, as long as it supports pre-existing ideas and does not challenge our beliefs, we are prone to accept it.

¹⁰ Jessica Brand, Valerie Wirtschafter, and Adya Danaditya, [“Popular Podcasters Spread Russian Disinformation about Ukraine Biolaps”](#), Brookings TechStream, 23 March 2022.

¹¹ Michael Meyer-Resende, [“How Do Social Media Algorithms Rank Content?”](#), Democracy Reporting International, April 2022.

¹² Clare Melford and Craig Fagan, [“Cutting the Funding of Disinformation: The Ad-Tech Solution”](#), The Global Disinformation Index, May 2019.

¹³ See Michael Meyer-Resende, [“How Do Social Media Algorithms Rank Content?”](#), Democracy Reporting International, April 2022.

¹⁴ Marina Weisband, speech at the G7 Conference “Facts vs Disinformation”, 06 April 2022.

As the technical foundation and channel of disinformation content, tools to propagate false narratives in (audio-) visual or textual form have been emerging in recent years. What such tools to disseminate disinformation have we already seen? The following section provides an overview of the most common emerging tools and dives into their technical components.

4.1 Deepfakes

4.1.1 Fake Images

Images have already played an important role in disinformation campaigns for some time, as they speak more directly to emotions and perceptions than text. Furthermore, they can be easily presented in a decontextualised context, effectively contributing to lies or half-truths (see the tactics section).

The ability to create new visual content from scratch is now making significant progress. In particular, Generative Adversarial Networks (GANs) for creating realistic photo images continue to gain in popularity. It is now possible to easily create and edit visual content without leaving perceptible visible traces.¹⁵ GAN-based models are based on a system of two neural networks – one for creating the false images (the generator) and one for detecting them (the discriminator). These networks reflect an iterative, self-learning structure that continuously improves the degree of realism of computer-generated visual content, posing a challenge for visual forensics.¹⁶ The way AI-generated photo content can be used to disseminate disinformation campaigns was demonstrated vividly when a bot swarm produced hundreds of fake profile pictures for a variety of fake Facebook accounts in early 2019 using a GAN based algorithm (StyleGAN2).¹⁷

¹⁵ Ning Yu et al., [“Attributing Fake Images to GANs: Learning and Analyzing GAN Fingerprints”](#), Computer Science, 20 November 2018.

¹⁶ Katerina Sedova et al., [“AI and the Future of Disinformation Campaigns”](#), Center for Security and Emerging Technology, December 2021.

¹⁷ Alfonsas Juršėnas et al., [“The Double-Edged Sword of AI: Enabler of Disinformation”](#), NATO Strategic Communications Centre of Excellence, Riga, December 2021.

The advances in artificial image creation do not stop at GANs, as text-to-image conversion technology is a new trend. Also based on neural networks, such tools can be used to create images from text descriptions. Among these, DALL-E, a neural network based on 12 billion parameters developed by OpenAI Lab, has generated promising results. Despite a lack of credibility to date, visual content based on this technology could become important to disinformation efforts, in that it would enable AI-produced imagery in direct response to a given false news narrative. The advancement of research in this area could eventually allow for the manufacturing of fake evidence by fabricating multiple images from a given scene (i.e., from different angles or at different moments in time).¹⁸

4.1.2 Synthetic Audio

The disinformation toolkit has not only been extended to visual material; there have also been major advances with respect to synthetic audio content, particularly when it comes to speech synthesis. By using computer technology, a speech synthesis system generates synthetic, artificially generated human speech. This area of computer science is also known as text-to-speech (TTS), and is defined as the process of automatically converting written text into audio signals.¹⁹ The proliferation of such audio deepfakes could pose a considerable risk, as they can be embedded with biometrics and, thus, used to manipulate identity-verification systems based on speech.²⁰ Although TTS technology has become part of everyday consumer electronics, such as Google Home, Apple Siri, and Amazon Alexa, mass adoption has yet to occur.²¹ While there are sophisticated models available, such as Tacotron63, Wavenet64 and DeepVoice3, a realistic imitation of a target voice still requires a large amount of noise-cleaned audio samples, which hinders user-friendliness.

Aside from text-to-speech, there have been tremendous advances in voice conversion/swapping. These are speech-to-speech deep fakes based on audio data, rather than text. In other words, voice conversion involves converting the audio waveform of a source speaker into that of a target speaker without altering the linguistic content.²² As opposed to text-to-speech deep fakes, the advantage of speech-to-speech deep fakes lies in the fact that they can replicate typical speech patterns, as well as the intonation of a given person, making them particularly vulnerable to abuse. This remains challenging

¹⁸ Ibid.

¹⁹ Karolina Kuligowska et al., [“Speech Synthesis Systems: Disadvantages and Limitations”](#), *International Journal of Engineering and Technology*, Vol. 7, No. 2, May 2018.

²⁰ Alfonsas Juršėnas et al., [“The Double-Edged Sword of AI: Enabler of Disinformation”](#), NATO Strategic Communications Centre of Excellence, Riga, December 2021.

²¹ European Parliamentary Research Service, [“Tackling deepfakes in European policy”](#), July 2021.

²² Alfonsas Juršėnas et al., [“The Double-Edged Sword of AI: Enabler of Disinformation”](#), NATO Strategic Communications Centre of Excellence, Riga, December 2021.

in its application, however, due to the need to train on a large amount of matched source and target audio material. Yet, the associated barriers have been falling due to a variety of AI applications that are easy to access, making so-called audio cloning a realistic prospect.²³

4.1.3 Synthetic Videos

Disinformation utilising audio-visual tools, such as synthetic videos, poses a particular challenge, since research indicates that videos are much more likely to be believed and shared than either text or audio versions of the same news item.²⁴ At the same time, the interplay of both audio and video content makes synthetic videos one of the most challenging forms of deep fakes. Although sophisticated algorithms have been developed to create videos from scratch, the majority of synthetic video content that circulates on the Internet relies on rather simple neural image transfer.²⁵ They are often employed in video content for exchanging facial features to mimic an individual's facial characteristics. Open-source software solutions, such as FakeApp, DFaker, faceswap-GAN, faceswap or DeepFaceLab, have greatly facilitated the production of such videos.²⁶ Due to the low level of credibility of this output, this type of deep fake has primarily been applied to pornographic content.²⁷

The difficulty in creating a convincing disinformation campaign continues to be that a large amount of video data is required to create a convincing impersonation. This explains why such campaigns often target public figures for whom such content exists. Recent developments indicate that the barriers to creating convincing content are slowly eroding. For example, a group of computer scientists in the US developed an all-in-one tool that allows for the creation of new video material from existing footage by modifying spoken text, facial expressions and gesture intensity.²⁸ Not only was the new material realistic and credible, but the researchers used only 30 seconds of video training material to achieve such convincing results.²⁹

²³ European Parliamentary Research Service, [“Tackling deepfakes in European policy”](#), *op. cit.*, note 21.

²⁴ Matt Swayne, [“Video Fake News Believed More, Shared More than Text and Audio Versions”](#), Pennsylvania State University, 08 September 2021.

²⁵ Yuezun Li et al., [“Toward the Creation and Obstruction of DeepFakes”](#), in Christian Rathgeb et al. (eds.), *Handbook of Digital Face Manipulation and Detection* (Cham: Springer International Publishing, 2022).

²⁶ Ibid.

²⁷ Ian Sample, [“What Are Deepfakes – and How Can You Spot Them?”](#), *The Guardian*, 13 January 2020.

²⁸ Alfonsas Juršėnas et al., [“The Double-Edged Sword of AI: Enabler of Disinformation”](#), NATO Strategic Communications Centre of Excellence, Riga, December 2021.

²⁹ Ibid.

4.2 Text Prediction

Major advances have also been made in the area of natural language processing (NLP) and natural language generation (NLG). Furthermore, the latter has become increasingly powerful as a result of advances in language generation systems, including the Generative Pre-Trained Transformer 3 (GPT-3), developed by OpenAI, and its Chinese-language equivalent, PanGu-Alpha, developed by Huawei.³⁰

Typically, generated language models are constructed using statistical techniques, including neural networks (NNs), which are used to estimate the likelihood of a sequence of words occurring together. In order to achieve high accuracy, the GPT-3 was trained with 175 billion parameters and read terabytes of text from the entire Internet.³¹ GPT-3's strength lies in the fact that it requires only a few text samples to generate convincing, human-like text output, placing it in the category of few-shot learning (FSL) models.³² It could be argued that the GPT-3 has the potential for severe harm as, contrary to its predecessor GPT-2, it does not require precise parameter tuning to produce biased text such as hate speech or propaganda.³³ GPT-3-based applications can, however, also produce biased content without intent, since the black-box-like nature of neural networks makes preventing such outcomes difficult ex-ante, and difficult to fix ex-post.

Despite impressive developments, even the GPT-3 model continues to struggle to produce long, logically consistent texts, since it is unable to evaluate textual meaning, which has led to some contradictory statements in longer outputs. Nevertheless, when applied to short texts, the model is highly sophisticated, making the text output difficult to distinguish from text written by humans, and making it a powerful tool for disinformation campaigns on platforms that employ short messages (for example, Twitter).³⁴ Although, once trained, models like GPT-3 become cheap (in terms of data) and easy to employ, the initial training requires computational resources that limit future development to either governmental actors or private companies. Despite these limitations, published research on GPT-3 immediately sparked follow-up projects (in particular, in Russia and China).³⁵

³⁰ Katerina Sedova et al., [“AI and the Future of Disinformation Campaigns”](#), Center for Security and Emerging Technology, December 2021.

³¹ Pina Merkert, [“Freie API für GPT-3: Türöffner für Fake-News und Vorurteile?”](#), *c't Magazin*, 7. December 2021.

³² Alfonsas Juršėnas et al., [“The Double-Edged Sword of AI: Enabler of Disinformation”](#), NATO Strategic Communications Centre of Excellence, Riga, December 2021.

³³ Alfonsas Juršėnas et al., [“The Double-Edged Sword of AI: Enabler of Disinformation”](#), NATO Strategic Communications Centre of Excellence, Riga, December 2021.

³⁴ Joe Uchill, [“GPT-3 Disinformation Campaigns Could Look Increasingly Realistic”](#), *Scmagazine*, 4 August 2021.

³⁵ Alfonsas Juršėnas et al., [“The Double-Edged Sword of AI: Enabler of Disinformation”](#), NATO Strategic Communications Centre of Excellence, Riga, December 2021.

Despite advancements in disinformation tools, their use can still follow long-established patterns. In particular, in violent conflicts (for example, in Syria and Ukraine), the objective has often been to create physical evidence to legitimise the claims of combatants.

Stories Retold

In the current phase of the Russian war against Ukraine, for example, synthetic content (e.g., AI-generated pictures) has been outpaced by misdated, misrepresented or mislocated authentic sources (e.g., by showing authentic satellite pictures with altered timestamps).³⁶ The 4D model of disinformation campaigns devised by Ben Nimmo, a non-resident senior fellow at the Atlantic Council's Digital Forensic Research Lab, complements this technique by identifying four strategies that are often combined: dismiss (insult your opponent), distort (twist the facts), distract (deflect the blame) and dismay (intimidate your opponent).³⁷



Figure 1. Ben Nimmo's classification of pro-Kremlin disinformation campaigns (Source: [Twitter](#)).

³⁶ Eliot Higgins, "[New July 17th Satellite Imagery Confirms Russia Produced Fake MH17 Evidence](#)", Bellingcat, 12 June 2015.

³⁷ DFRLab, "[PROPAGANDA: Dismiss, Distort, Distract, & Dismay](#)", 24 June 2017.

The last resort is often to modify or create content from scratch with computer-based methods. In this way, the 2020 presidential election in the US showed that dis- and misinformation can be proliferated through relatively mundane tactics. A number of false claims about the election went viral on social media, based on “friend of a friend” claims – the principle of disseminating a message without questioning its origin because the source is known. The response of social media platforms to this type of dis- and misinformation has been inconsistent at best.³⁸ Based on human agency, trolling is another old but frequently used tactic in information wars. The only difference today is that troll farms operate more openly, actively inciting followers to disseminate propaganda.³⁹



Figure 2. A VICE investigation found that Cyber Front Z, a Russian troll farm accused of spreading pro-Kremlin misinformation on social media, publicly encourages its followers to post propaganda on Telegram (Source: VICE News).

This is not, however, to say that, in light of the speed of technical innovations, tactical approaches to disinformation are waning. The following identifies some of the recent trends in the dissemination of disinformation.

Cross-Platform Sharing

The efforts by certain social media platforms to combat disinformation have prompted those spreading it to start shopping for other venues. To avoid platform moderators,

³⁸ Karolina Koltai, [“‘Friend of a Friend’” Stories as a Vehicle for Misinformation](#), Election Integrity Partnership, 26 October 2020.

³⁹ [“Cyber Front Z, the Pro-Kremlin Troll Army Spreading Propaganda Online”](#), Institute for Strategic Dialogue, 1 May 2022.

misleading and false content is shifted to less regulated platforms, such as Gab, Telegram or Parler, while maintaining its visibility on more traditional social media sites through cross-platform sharing. Due to cross-referencing and link sharing, content hosted on less regulated platforms is widely available on more established social media platforms.⁴⁰ This was the case in the context of the 2021 German elections, where YouTube videos with questionable content were the most shared across platforms.⁴¹ This multi-platform dissemination of manipulated information, therefore, requires increased attention from platforms, and possibly regulation.

Abusing Algorithmic Recommender Systems

The loss of visibility for disinformation when content is moved to less regulated platforms can be mitigated by ad-tech and the manipulation of algorithmic recommendation systems. The latter, the exploitation of recommender systems, is a key challenge. These recommender systems are used by companies to select and rank content for users, with the overall aim to keep users on their platforms for as long as possible. For example, disinformation actors can exploit the fact that ML-based recommendation systems rely on the identification of clusters of users with similar interests and attitudes. Through fine-tuning disinformation for specific clusters, this content can be amplified and can cause strong emotional reactions in user networks without triggering condemnation or complaints regarding violations of terms of service.⁴² Content fine-tuning is only one way in which recommender systems have been abused. Another has been the attempt to create artificial bubbles of users with extremist views by using bots and cyborgs numerous enough to encourage high rankings of harmful content by recommendation systems.⁴³

Coordinated Sharing Behaviour

It takes a village to manipulate algorithms, but rapid-fire super spreaders⁴⁴ and “disinformers”⁴⁵ have found a way. The former react to each other’s messages within a split second and promote each other’s content repeatedly, while the latter are

⁴⁰ Alexandre Alaphilippe, [“Disinformation Is Evolving to Move under the Radar”](#), Brookings, 4 February 2021.

⁴¹ Jesse Lerke and Finn Klebe, [“Election Monitor Germany 2021 – Research Brief #1”](#), Democracy Reporting International, 20 July 2021.

⁴² Katerina Sedova et al., [“AI and the Future of Disinformation Campaigns”](#), Center for Security and Emerging Technology, December 2021.

⁴³ Ibid.

⁴⁴ Marcel Schliebs et al., [“China’s Public Diplomacy Operations: Understanding Engagement and Inauthentic Amplification of PRC Diplomats on Facebook and Twitter”](#), Oxford, UK: Programme on Democracy & Technology, 01 May 2021.

⁴⁵ Maria Giovanna Sessa, [“Disinformation Self-Proclaimed Experts: Spreading COVID-19 Disinformation under the Guise of Expertise”](#), EU DisinfoLab, 18 January 2022.

individuals with large followings in conspiracy and disinformation communities who proclaim themselves to be experts. Facebook has regulated inauthentic coordinated behaviour⁴⁶ for artificially boosting content. But what can be done when systematic sharing is conducted by authentic or duplicate accounts that are inherently harmful and spread disinformation? The “Stop the Steal” campaign to overturn the 2020 US presidential election was only the tip of the iceberg of authentic coordinated behaviour, that is, real people with authentic profiles sharing the same content, thus undermining trust in US democracy.⁴⁷

Astroturfing

What goes hand in hand with coordinated sharing behaviour is a practice known as astroturfing: an online multi-account operation that falsely gives the impression of wider grassroots support.⁴⁸ Research by the Institute for Strategic Dialogue indicates a network of pro-Putin groups on Facebook who occasionally disseminate pro-Kremlin disinformation, with more than 650,000 members combined and overlapping “fans” listed as page admins, many with multiple accounts under the same name. This technique paints a picture of an authentic grassroots movement while at the same time launching a coordinated, seemingly inauthentic Putin support campaign.

Content Laundering

Sharing across platforms and the exploitation of recommender systems are just two of the methods by which disinformation has been spread without being subjected to content moderation. This also has been achieved by using content laundering, which involves channelling content through or obtaining endorsements by proxies. In particular, content said to be disseminated by Russia’s foreign military intelligence agency (GRU) and its Internet research agency (IRA) has been laundered. In June 2020, for example, the EU DisinfoLab exposed a partnership between Observateur Continental, a French-language website engaged in the spread of false claims concerning COVID-19 in France, and companies close to the GRU.⁴⁹ Similarly, PeaceData, a fake news site run by the IRA, hired independent journalists during the 2020 US elections, contributing to the dissemination of false information.⁵⁰

⁴⁶ [“Inauthentic Behavior”](#), Meta Transparency Center, accessed 25 April 2022.

⁴⁷ Shannon Bond and Bobby Allyn, [“How the ‘Stop the Steal’ Movement Outwitted Facebook Ahead of the Jan. 6 Insurrection”](#), NPR, 22 October 2021.

⁴⁸ Moustafa Ayad, [“The Vladimirror Network: Pro-Putin Power-Users on Facebook”](#), Institute for Strategic Dialogue, 13 April 2022.

⁴⁹ Maria Giovanna Sessa, [“Disinformation Self-Proclaimed Experts: Spreading COVID-19 Disinformation under the Guise of Expertise”](#), EU DisinfoLab, 18 January 2022.

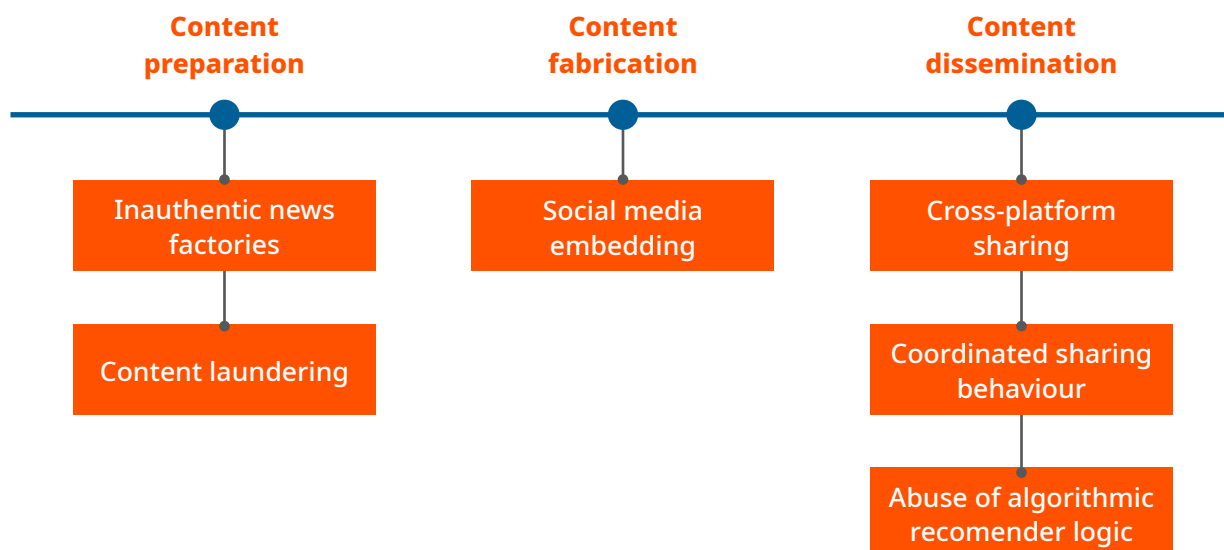
⁵⁰ Elizabeth Dwoskin and Craig Timberg, [“Facebook Takes down Russian Operation That Recruited U.S. Jour-](#)

Embedding Social Media into News Coverage

All media embed social media content into news, for example, to show commentary or witness statements, or evidence related to the events covered. This method is also abused by agents of disinformation, for example, by the IRA to interfere in countries such as the Central African Republic, Syria, Libya, Sudan, Mozambique, and Zimbabwe.⁵¹ Embedding has been done through blending in false social media accounts and the reproduction of content in pro-government media. Hence, through a combination of spreading disinformation through fake and real social media accounts, and then embedding that content into coverage by pro-Kremlin media like Sputnik, Russian state propaganda has used new forms of dissemination.⁵²

Inauthentic News Factories

The integration of social media content has also played a role in the creation of inauthentic news factories. For example, the newly created Russian RIA FAN media conglomerate employs opaque figures and an arsenal of Kremlin-friendly Telegram channels claiming to report objective news.⁵³



[nalists, amid Rising Concerns about Election Misinformation](#)", *Washington Post*, 1 September 2020.

51 Lukas Mejia and Clint Watts, ["The Illusion of a Russian Media Empire: How Anonymous Bloggers and Obfuscated Identities Power the Troll Factory's Successor"](#), Alliance For Securing Democracy, 19 July 2021.

52 ["In Bed with Embeds: How a Network Tied to IRA Operations Created Fake 'Man on the Street'"](#), Stanford Internet Observatory, 2 December 2021.

53 Lukas Mejia and Clint Watts, ["The Illusion of a Russian Media Empire: How Anonymous Bloggers and Obfuscated Identities Power the Troll Factory's Successor"](#), Alliance For Securing Democracy, 19 July 2021.



UKRAINE FACTS

DEEPEAKE ELECTIONS AND GENDERED DISINFORMATION FOCUS COVID-19

Stories

In the following section, we present some of the specific narratives that have been created recently through the use of mis- and disinformation. The aim of this section is to identify to what extent tools and tactics have or have not been applied in current disinformation narratives, and to infer from this potential future developments. The case studies here reveal that, in recent disinformation campaigns, innovations in tactics have played a more central role than innovations in tools, with the use of generic content playing only a peripheral role. Nevertheless, the next section shows that this may be a temporary situation, as the cost of accessing ML technologies is decreasing, while the quality of generic content continues to rise.

The Russian War Against Ukraine: Anti-Western Narratives and the Fog of War

In addition to the onset of physical hostilities, the Russian invasion of Ukraine on February 24, 2022, sparked an information war. Specifically, Russian propaganda efforts have sought to distort information about the purpose, progress and human cost of its military operations in Ukraine. War rhetoric emanating from Russia has to be viewed within the broader context of anti-Western rhetoric that the country has continued to engage in over the past decade, particularly during the Euromaidan protests in Ukraine in 2013 and 2014, during which it interpreted the protests as being manipulated by Western actors.⁵⁴ Nevertheless, the recent war efforts have brought a very specific set of narratives.⁵⁵

Depicting the war as a response. Russian propaganda efforts have tried to depict its military operation as a necessary response, rather than an attack. From the beginning of the military campaign, pro-Russian messaging has been centred on the the legitimacy of the Russian intervention. Its stated purpose has been to “liberate” and “de-Nazify”

⁵⁴ Madeline Roache and Sophia Tewa, [“Russia-Ukraine Disinformation Tracking Center”](#), *NewsGuard*, 2 April 2022.

⁵⁵ Maria Giovanna Sessa, [“Ukraine Conflict Disinformation: Worldwide Narratives and Trends”](#), EU Disinfo-Lab, 14 March 2022.

Ukraine, as well as to stop alleged Ukrainian atrocities and genocide against “ethnic Russians” in Donbas.

Stoking fear of Ukrainian refugees. Russian propaganda efforts have not only been targeted against the Ukrainian leadership, but also against ordinary citizens. There have been several instances in which Russia has used disinformation to portray refugees as violent and a threat to their host countries (e.g., Poland).

Refuting Ukrainian identity. As Russian propaganda efforts have primarily focused on delegitimizing the Ukrainian leadership, depicting them as alcoholics, drug addicts, and fascists, they have recently turned their attention to Ukrainian cultural identity. Attempts have been made by pro-Russian outlets and trolls to dispel the idea of a Ukrainian identity distinct from that of Russia.

Shielding against accusations. With mounting evidence of widespread war crimes, much of the propaganda has been oriented towards portraying the evidence as dubious or fabricated either by Ukrainians or the West. Reports of mass atrocities committed in the city of Butcha were immediately decried by Russia as having been fabricated by Ukrainian soldiers.

Instrumental to the Russian war propaganda has been a long-standing evolution in the tools and tactics it has employed; in fact, when it comes to the Russian propaganda machinery, there has been a long build-up of capabilities. At least since the annexation of Crimea in 2014, Russia has pursued the development of an elaborate structure for cyber influence operations (CIOs). Certain specific aspects are present in the Russian approach:

Reliance on volume. Russian propaganda is focused on volume, rather than consistency.⁵⁶ As such, the focus has been on creating a massive amount of textual, visual, audio and video materials, disseminated both through traditional means (such as state-owned media) and other channels (such as social media). Rather than crafting a coherent narrative or technically challenging content, the tactic seems to rely on the recognition effect (of simply employing endlessly repeated claims).⁵⁷ Perhaps this is why outright synthetic content (e.g., deep fakes) has played a relatively peripheral role, and a synthetically produced video that appeared, depicting the Ukrainian Prime Minister, Volodymyr Zelenskyy, was swiftly uncovered as a fake.

Blending of human and non-human agency. While Russian propaganda has relied on automated means to disseminate its messages (i.e., through bots), those behind the propaganda appear to have realised that human-to-human communication can be the

⁵⁶ Christopher Paul and Miriam Matthews, “[The Russian ‘Firehose of Falsehood’ Propaganda Model](#)”, RAND Corporation, 11 July 2016.

⁵⁷ Ibid.

most effective method of spreading disinformation. In particular, during the onset of the war efforts, Russian-language influencers played a crucial role in furthering Russian propaganda.⁵⁸

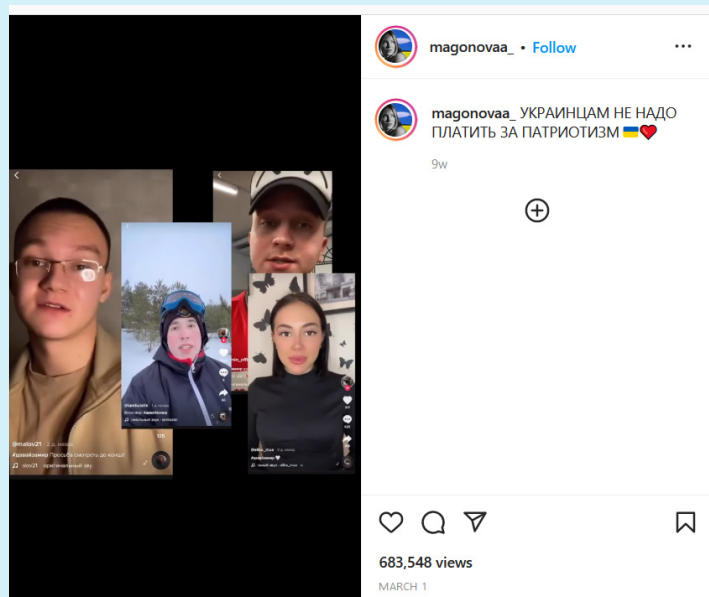


Figure 3. A video compilation on Instagram featuring a series of prominent Russian influencers reading the same script, which attempts to excuse the war in Ukraine by promoting the falsehood that Ukraine had perpetrated a genocide against Russian speakers in the Donbas region over the preceding eight years.⁵⁹

Since Russia has increased restrictions on the freedom of foreign social media websites, Russian-speaking VK and Telegram have become the primary platforms for the dissemination of such messages.⁶⁰

Domestic focus. Even though many of the messages promoted by Russian CIOs throughout the war have not found resonance with Western audiences, this may not necessarily indicate a failure of Russian propaganda efforts. An estimated 90% of propaganda produced in Moscow is intended for domestic consumption.⁶¹ According to a large number of reports, domestic efforts have been successful, with the Russian population being largely unaware of the purpose, progress and implications of Russian war efforts in Ukraine.

Regionally targeted campaigns. The sheer amount of disinformation coming out of Russia allows for regionally targeted campaigns, with context-specific messaging, despite a predominantly domestic focus. Russian messaging has targeted the West by promoting disunity among Western alliances (NATO, the EU, etc.) while messaging targeting the Global South has taken advantage of anti-colonial sentiments,⁶² trying to portray the assault against a smaller neighbour as an attack against “the West”.

⁵⁸ Maria Giovanna Sessa, “Ukraine Conflict Disinformation: Worldwide Narratives and Trends”, EU Disinfo-Lab, 14 March 2022.

⁵⁹ For more information on the VICE News influencer investigation, see David Gilbert, “Russian TikTok Influencers Are Being Paid to Spread Kremlin Propaganda”, VICE News, 11 March 2022.

⁶⁰ Mark Scott, “As War in Ukraine Evolves, so Do Disinformation Tactics”, POLITICO, 10 March 2022.

⁶¹ Ibid.

⁶² Emman El-Badawy et al., “Security, Soft Power and Regime Support: Spheres of Russian Influence in Africa”, Tony Blair Institute for Global Change, 23 March 2022.

Consequently, this has not only relied on domestic channels, but has also benefited from Chinese media reproducing and amplifying Russian propaganda.⁶³

Relying on proxies. In spreading its message across the West, Russian efforts have relied heavily on proxies. As a result, fringe websites, sometimes from the extreme ends of the political spectrum, have played an important role in laundering Russian content online.⁶⁴

Traces of Deepfakes?

Deepfakes and cheapfakes differ in technical sophistication, barriers to entry and techniques.⁶⁵ When the technical sophistication increases, the wider public's ability to produce fakes decreases. The use of deepfakes in politics still remains relatively rare. This might be the reason why deepfakes have not yet played a pivotal role in Russia's war in Ukraine.

The first use of a crude manipulated video as part of a war propaganda effort emerged in March 2022 on Facebook and YouTube, showing Ukraine's President Volodymyr Zelenskyy with a voice different from his usual tone. The clip was not only posted to Telegram but amplified on the Russian social network VKontakte.⁶⁶ In this video, he asks

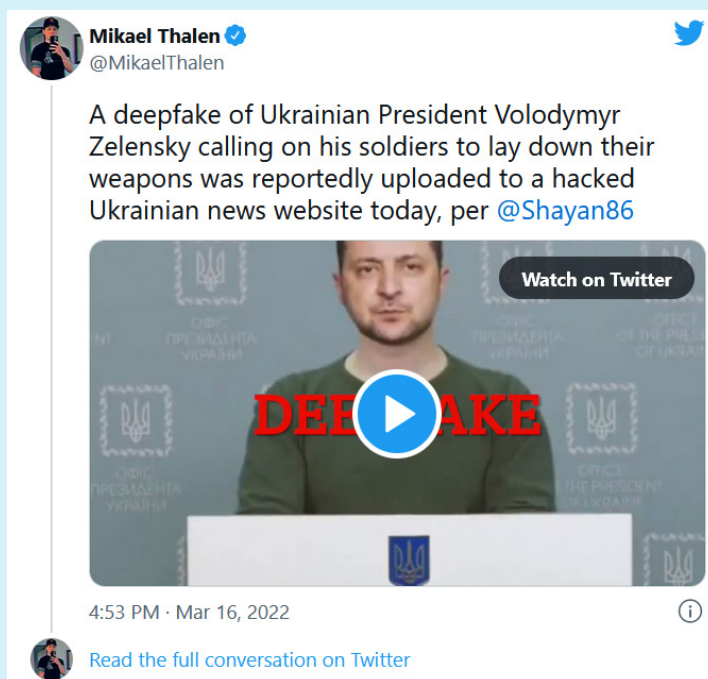


Figure 4. On 16 March, a fake video emerged on social media of a motionless version of Zelenskyy encouraging Ukrainian soldiers to surrender. The video was debunked very quickly by Zelenskyy himself (Source: Twitter).

⁶³ Helen Davidson, "[Close Ties Allow Russian Propaganda to Spread Swiftly through China, Report Claims](#)", *The Guardian*, 31 March 2022.

⁶⁴ Sara Fischer and Basu Zachary, "[Russian Disinformation Frenzy Seeds Groundwork for Ukraine Invasion](#)", *Axios*, 21 February 2022.

⁶⁵ Madeline Brady, "[Deepfakes: A New Disinformation Threat?](#)", Democracy Reporting International, 31 July 2020.

⁶⁶ Digital Forensic Lab, "[Russian War Report: Hacked News Program and Deepfake Video Spread False Zelenskyy Claims](#)", Atlantic Council, 16 March 2022.

Ukrainian troops to lay down their weapons. The cheaply produced video was easily spotted as manipulated content, as his head appears too large for and more pixelated than his body, and the message itself was too surprising to be true.

A few days after the manipulated video of Zelenskyy, another doctored video, also posted on 16 March, went viral on social media, supposedly showing Russian President Vladimir Putin announcing that the war in Ukraine was over.⁶⁷

Both videos were of poor quality and, therefore, were easily debunked and rebutted. This illustrates that it is still far more common to find less computationally advanced cheapfakes than sophisticated deepfakes for reasons of scale, cost and technical capacities.

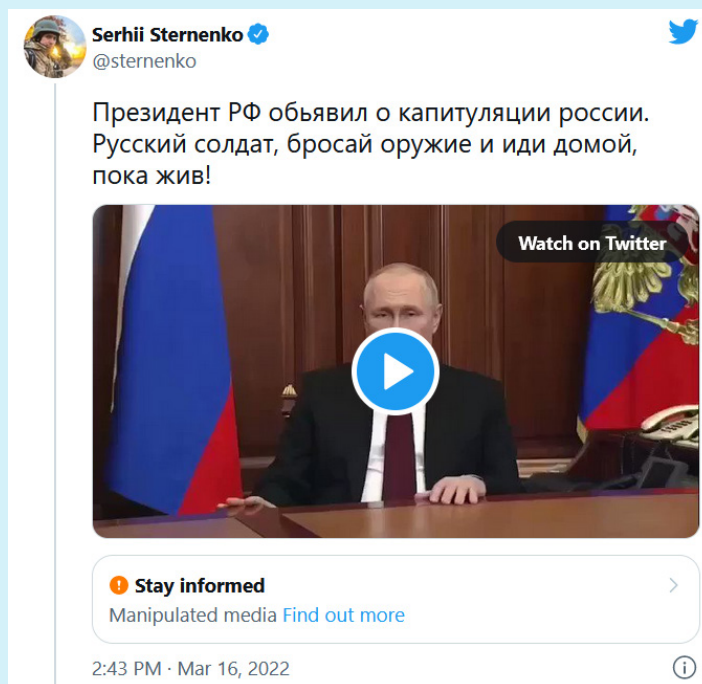


Figure 5. A fake video shows Russian President Vladimir Putin announcing the end of Russia's war with Ukraine. The video was created by manipulating Putin's mouth movements in a [real video](#) that was originally posted by the Kremlin on 21 February. (Source: Twitter)

The COVID-19 Pandemic: Science as a Weapon of Disinformation

Aside from the fact that the COVID-19 pandemic brought public life to a halt in many countries, the high degree of uncertainty, especially at the beginning of the pandemic, provided an ideal environment for disinformation to spread. Due to the sheer volume of accurate and inaccurate information available to the general public, the WHO coined the term “infodemic”. The connotation attached to the term derives from the idea that, with the overabundance of information, citizens are less able to filter for reliable content. As a result of this inability, false narratives and conspiracy theories quickly took hold.

⁶⁷ Dan Evon, “[Putin Deepfake Imagines Russian President Announcing Surrender](#)”, Snopes, 18 March 2022.

Conspiracies around the origins of the virus. Immediately after the outbreak of the pandemic, disinformation focused particularly on the origin of the virus. While the question of its origins remains open, and neither accidental nor intentional human responsibility can be ruled out, conspiracy theories made wide-ranging claims with no basis in facts. There were claims, for example, that the virus may have been created by elite actors to reduce population growth, and that it may have been accelerated by the proliferation of 5G wireless networks.⁶⁸

Dissemination of false medical advice. Almost from the beginning, there was a stream of information materials promoting dubious treatments and cures. At a time when the official health authorities were adjusting their strategies, often reversing their own positions, based on new scientific insights, disinformation actors exploited the general feeling of insecurity by promoting untested or outright harmful remedies and medicines. Misleading health information and hoaxes with false claims concerning treatment and prevention of the disease have had a detrimental effect on public health. The World Health Organization reports that in the first three months of 2020, nearly 6,000 people were hospitalised across the globe due to dis- and misinformation regarding the virus. It is estimated that at least 800 people have died as a result of misinformation related to COVID-19 during this period.⁶⁹

Undermining institutional trust. While dis- and misinformation promoting distrust of public officials was primarily targeted at health authorities at the beginning of the pandemic, the introduction of containment measures changed this focus. Increasingly, voices became prevalent that linked restrictions on personal and public freedoms to an attempt to undermine democracy and establish authoritarian structures. Ample evidence exists that such narratives were strongly amplified through foreign cyber influence operations emanating from China and Russia. Despite these attempts, research has shown that average institutional trust increased during the first months of the pandemic, reflective of a so-called “rallying around the flag” effect.⁷⁰ Overall institutional trust fell, however, as limits to individual liberties and public freedoms remained in place and were often inadequately or inconsistently communicated. Democracy Reporting International has published much analysis criticising COVID-19 restrictions around the world.⁷¹ This analysis has demonstrated that the problem with disinformation is that it transforms legitimate criticism and debate into part of a more fundamental narrative of systematic, targeted abuse, which insinuates that democratic governments plan to build authoritarian regimes.

⁶⁸ [“Tackling Coronavirus Disinformation”](#), European Commission, accessed 25 April 2022.

⁶⁹ [“Fighting Misinformation in the Time of COVID-19, One Click at a Time”](#), World Health Organisation, 27 April 2021.

⁷⁰ [“Trust in Public Institutions”](#), OECD, 09 July 2021.

⁷¹ See [“COVID19, Publications”](#), Democracy Reporting International.

A false counter-narrative. A number of authoritarian and semi-authoritarian states have exploited the spread of rising fears and the proliferation of disinformation for their own benefit. By exaggerating the risks associated with false narratives, it provided states such as Russia and China, as well as Hungary, with a pretext to implement emergency laws and the crackdown on members of the press, falsely portraying them as disinformation agents.⁷² In some of these cases, the state-driven narrative insisted that investigative journalism exposing institutional failures, as well as private bloggers expressing disagreements with the government, were false or fake news, presenting them as a threat to society, thus legitimising repressive measures.

Impersonation and inauthentic news. Although most of the language employed in disinformation has been emotive and appealing to fears, the primary tactic of disinformation actors has been to cast doubt on scientific insights, rather than to reject facts as a whole. Thus, for example, the use of hard-to-verify medical qualifications by many disinformation actors, as well as the spread of disinformation material that often imitated news outlets, increased the effort necessary for citizens to discern what information was reliable. The deliberate attempt to spark confusion, rather than outright denial of facts, resonates with trends from other disinformation topics, such as climate change, where science denial has gradually been replaced by solution disinformation targeting climate policy and renewable energy.⁷³

End-to-end encryption channels. Even though misinformation frequently surfaced on social media and video hosting websites (i.e., YouTube), the speed at which it was disseminated, even globally, has to be traced back, at least in part, to WhatsApp. Family and friend-based chat groups often served as collection pools for misleading or outright false information. An important tactic employed in such groups was the exchanging of friend-of-a-friend experiences, meant to exacerbate unjustified fears of containment measures, such as vaccines.

Foreign influence on national debates. During the crisis, several attempts have been made by outside states and even commercial actors to influence discussions in the EU and its Member States.⁷⁴ Despite the fact that both China and Russia disseminated false information about COVID-19, their tactics were quite different. While Beijing was primarily engaged in reputation management and used disinformation to counter criticisms of its crisis management, Russian propaganda was mainly designed to sow distrust in Western governments and health officials.⁷⁵

72 [“A Facade of Legality: COVID-19 and the Exploitation of Emergency Powers in Hungary”](#), International Commission of Jurists, Februar 2022; [“Covid-19: Six Chinese Defenders of Press Freedom Still in Detention”](#), Reporters Without Borders, 26 May 2020.

73 Carolyn Gramling, [“Climate Change Disinformation Is Evolving. So Are Efforts to Fight Back”](#), *Science News*, 18 May 2021.

74 [“Tackling Coronavirus Disinformation”](#), European Commission

75 [“Jabbed in the Back: Mapping Russian and Chinese Information Operations During COVID-19”](#), CEPA, 2 December 2021.

Memes as a tool of disinformation. The covid-related disinformation was characterised by the use of memes as a symbolic method of communicating inaccurate information. Social media sites have been flooded with memes suggesting, for example, that the number of infections increased following the advent of vaccines.



Figure 6. A meme with images of rapper Drake, here including fact-checking labels, has been used to promote false vaccine claims (Source: [BBC](#)).

Elections and Gendered Disinformation: Constraining the Public Sphere

Elections have been particularly marked by disinformation and misinformation. This was demonstrated during the 2020 United States presidential election, during which a variety of foreign and domestic efforts were made to spread false and misleading information.⁷⁶ There was widespread disinformation about the electoral process (e.g., communicating incorrect registration dates or requirements) in the run-up to the vote, and disinformation campaigns designed to make false accusations of voter fraud and institutional corruption with a view to delegitimize the election results.⁷⁷ The DRI report on the 2021 German election campaign online highlighted similar patterns in Germany, though less intensive.⁷⁸

It is important, however, to note that attempts to undermine institutional trust in electoral arrangements have been only one worrying trend when it comes to the use of disinformation during elections. Another is the prevalence of gendered disinformation

⁷⁶ “Foreign Threats to the 2020 U.S. Federal Elections”, Office of the Director of National Intelligence, 16 March 2021.

⁷⁷ “Evaluating Platform Election-Related Speech Policies”, Election Integrity Partnership, 28 October 2022.

⁷⁸ Democracy Reporting International, “What’s #BTW21 Got to Do with It? Taking stock of the German Election Campaign Online”, December 2021, p. 22-24.

defined by falsity, malign intent and coordination.⁷⁹ Examples from the 2019 European Parliament elections, the 2020 United States presidential election and the 2021 German federal elections show that women candidates are being targeted with harmful content at an alarming rate.⁸⁰ This form of cyber-misogyny is particularly disturbing, as it occurs at the intersection of disinformation and forms of online violence, such as abuse and harassment.⁸¹ Disinformation targeting women in politics has evolved into a global phenomenon with similar rhetorical patterns:

Depictions of incompetence. Women in politics have generally been targeted by gender-based narratives that amplify negative stereotypes or misconceptions. Typically, women are portrayed as being less competent and endowed with the wrong emotional disposition for public office (women are frequently depicted as being too emotional or polite). For example, tampered quotes and videos were both used to support allegations of knowledge gaps on climate change issues made against the Green Party's leader and Germany's current foreign minister Annalena Baerbock during the German election campaign.⁸² The denunciation of women as incompetent is often communicated through the use of pseudoscience and misogynistic humour.⁸³

Depictions of physical unfitness. While politicians are regularly exposed to harmful online content that casts serious doubt on their mental capacity to hold office, they are also often subjected to allegations related to their physical fitness. It is not uncommon to find constructed narratives questioning the ability of women to face the arduous task of campaigning and portraying them as physically incapable of enduring the long hours involved in political negotiations. During the 2016 United States presidential election, one prominent example of the use of false narratives was a disinformation campaign, endorsed by right-wing outlets, that distorted the picture of Hilary Clinton's health.

Depictions of libidinous and sexual depravity. In terms of gendered disinformation, a dominant narrative focuses on the sexual identity and morale of women.⁸⁴ In the vast majority of cases, misinformation and disinformation are laced with sexual innuendo or are outright abusive. It is often the case that sexually charged narratives are designed to appeal to strong social norms and extreme religious beliefs.⁸⁵ Utilising distorted views of

⁷⁹ Nina Jankowitz et al., "[Malign Creativity: How Gender, Sex, and Lies Are Weaponised against Women online](#)", Wilson Center, January 2021.

⁸⁰ See e.g., Hate Aid & ISD, "[Hass als Berufsrisiko: Digitale Gewalt und Sexismus im Bundestagswahlkampf](#)" [Hate as an Occupational Hazard: Digital Violence and Sexism during the German Election Campaign"], 9 March 2022.

⁸¹ "[Gender-Based Disinformation: Advancing Our Understanding and Response](#)", EU Disinfo Lab, 20 October 2021.

⁸² Maria Giovanna Sessa, "[What Is Gendered Disinformation?](#)", Heinrich-Böll-Stiftung, 26 January 2022.

⁸³ Ellen Judson et al., "[The Contours of State-Aligned Gendered Disinformation Online](#)", Demos, October 2022.

⁸⁴ Maria Giovanna Sessa, "[What Is Gendered Disinformation?](#)", Heinrich-Böll-Stiftung, 26 January 2022.

⁸⁵ Ellen Judson et al., "[The Contours of State-Aligned Gendered Disinformation Online](#)", Demos, October 2022.

female sexuality – whether through the portrayal of women as the archetype of sexual purity or through denigration of female sexuality as indecent or obscene – accusations of illicit affairs or explicit (mostly synthetic) sexual content are often used to undermine women in the public sphere.

In all these narratives, the common goal of forcing women politicians out of the public arena, silencing their voices in the public sphere or discouraging political participation at the outset can be clearly seen. Cyber-misogyny employs a variety of tools and tactics.

Coordinated and lone-wolf attacks. In addition to being subjected to online abuse from private individuals, women politicians are more frequently targeted by coordinated online campaigns. In India, the ruling Bharatiya Janata Party (BJP) has been accused repeatedly of employing a large number of far-right trolls in order to attack opposition women politicians. A similar situation has been reported by journalists and politicians in Mexico and Brazil where coordinated cyberattacks have been prevalent, often involving government sources.⁸⁶ During coordinated character assassination attempts, both human and virtual agents have played an important role. Not only have they been propagated through bots and cyborgs, but also by right-wing bloggers, journalists, and politicians.

The use of deepfakes and cheapfakes. In recent years, deepfakes and cheapfakes,⁸⁷ often containing explicit pornographic material, have gained increased prominence as weapons to humiliate and discredit women victims, including women politicians.

One example of the many cheapfakes that were used to discredit Green Party candidate Annalena Baerbock during the 2021 German election campaign was a fake nude photo allegedly of Baerbock, with a quote insinuating she had engaged in sex work. On Telegram alone, it was viewed more than 150,000 times. Other social media posts threatened to share Baerbock's home address, contained slurs about her appearance, and attempted to discredit her qualifications.⁸⁸ Similarly, users on the Internet have spread the rumour that US Congresswoman Alexandria Ocasio-Cortez made a sex tape years ago, prompting some to create photoshopped images to present as proof.⁸⁹

⁸⁶ Lucina Di Meco and Kristina Wilfore, "[Gendered Disinformation Is a National Security Problem](#)", Brookings, 8 March 2021.

⁸⁷ For a concrete differentiation between deep and cheap fakes, see Rafael Goldzweig and Madeline Brady, "[Deepfakes – How prepared are we?](#)", Democracy Reporting International, November 2020.

⁸⁸ "[Schmutziger Wahlkampf: Wie Desinformation die Bundestagswahl vergiftet](#)", correctiv.org, 21 September 2021.

⁸⁹ Nina Jankowitz et al., "[Malign Creativity: How Gender, Sex, and Lies Are Weaponised against Women online](#)", Wilson Center, January 2021.

Doctored pictures and video content purporting to be from sex tapes have been among the content shared.⁹⁰ It has been reported that up to 42 per cent of women politicians have been exposed to or have found sexually explicit photographs of themselves online.⁹¹

Creating a whack-a-mole effect. It is not only the sheer volume of sexually charged content directed at women politicians as well as private citizens, that has generated a destructive capacity, but also the persistence of such content. Victims of online abuse are often left alone to fight for the removal of harmful content, and it has become a common practice by perpetrators to either re-upload deleted content or disseminate it further through website mirrors.⁹² Further, the persistence and dissemination of sexually abusive content in private or closed chat groups make it nearly impossible to remove such content once it is made public.

⁹⁰ Maik Baumgärtner et al., [“Rechte Desinformationsattacken gegen die Grünen: Im Visier der Hetzer”](#), *Der Spiegel*, 23 July 2021.

⁹¹ Ana Blatnik, [“An Overlooked Threat To Democracy? Gendered Disinformation About Female Politicians”](#), *Women In International Security*, 13 September 2021.

⁹² Moira Donegan, [“How Pornhub – One of the World’s Biggest Sites – Caused Untold Damage and Pain”](#), *The Guardian*, 16 December 2020.



QUANTUM COMPUTING
HUMAN-BOT
APPROACHING
THE TIPPING
POINT
EASY ACCESS
TO DEEPFAKES
PREVENTION

Predicting Future Threats

Democracy in the digital age is under threat. An open and pluralistic public debate is essential for democracy, but this is in danger when online discourse is threatened by information manipulation. Disinformation comes and stays in various forms and has become more subtle and sophisticated over the years. While the section of this report on prevalent narratives might have given the impression that it is more the tactics that matter than the tools, since some still seem to lack in quality, artificial intelligence and machine learning continue to advance.

ML algorithms are able to discover patterns that are difficult for humans to discern. Language generation capabilities and the tools that enable the production of synthetic videos are already capable of manufacturing viral disinformation on a large scale. These could be used to facilitate voter suppression, diminish trust in political systems and increase voter apathy.

This section examines how AI and ML technologies can enhance specific disinformation tools and techniques and how these technologies may aggravate current trends and shape future developments in the disinformation sphere.

Approaching the Tipping Point



Optimised Predictive Text Generation

Synthetically generated text, or the composition of text with text prediction tools, is cheap, effective, and less labour-intensive than hiring native speakers to write content or hacking profiles to publish from their platforms. The result is deceptive messaging that is almost indistinguishable from authentic communication. A recent study by the Center for Security and Emerging Technology (CSET) indicates that GPT-3 can draft manipulative stories that fit a viewpoint, cast the seeds of new conspiracies, and create divisive posts rooted in language that

fosters extreme polarisation.⁹³ Researchers have further demonstrated the text prediction tool's ability to generate emails from only bullet points.⁹⁴ The possibility of manufacturing artificial "leaked" e-mails powered by ML algorithms that contain malicious information is therefore heightened.⁹⁵ Breakthroughs in efficiency and infrastructure could rapidly accelerate the use of high-quality NLP models and text prediction tools to generate content to support the dissemination of any narrative – be it benign or malicious.



Improved Machine Translation

Where NLP models can help with manufacturing content that captures the touch and feel of a message to be propagated, ML-powered language translation and autocomplete functions in real-time can help disinformation actors sound more authentic, automatically eliminating or reducing grammatical mistakes. This might facilitate the recycling of false narratives and their even wider spread in societies that already have a firm volume of text available to train automated translation models.⁹⁶



Advances in Prevention Increase Disinformation Sophistication

As more manipulative tactics emerge, the data sets that train AI technologies will grow larger, which, in turn, will help AI detection models.⁹⁷ This is self-explanatory: The networks generating synthetic content can also be used to detect content that emerges from these same networks. However, as research produces advances in detection mechanisms, "whether for individuals to cue on the attributes of fakeness or technology to facilitate that detection",⁹⁸ the synthetics feeding the propaganda machine itself become more ambitious by default, resulting in progress in prevention and regulations of only a transitory nature.

⁹³ Ben Buchanan et al., "[Truth, Lies, and Automation](#)", Center for Security and Emerging Technology, May 2021.

⁹⁴ Tom Simonite, "[Give These Apps Some Notes and They'll Write Emails for You](#)", *WIRED*, October 18, 2020.

⁹⁵ Katerina Sedova et al., "[AI and the Future of Disinformation Campaigns](#)", Center for Security and Emerging Technology, December 2021.

⁹⁶ Alfonsas Juršėnas et al., "[The Double-Edged Sword of AI: Enabler of Disinformation](#)", NATO Strategic Communications Centre of Excellence, Riga, December 2021.

⁹⁷ Tim Hwang, "[Deepfakes – Primer and Forecast](#)", NATO Strategic Communications Centre of Excellence, May 2020.

⁹⁸ Alex Engler, "[Fighting Deepfakes When Detection Fails](#)", Brookings, 14 November 2019.

Extending the Global Reach



Easy Access to Deepfakes

Generative adversarial networks (GANs) can already be used to manufacture fake profile pictures that look like real people, as well as synthetic deep fake videos. Making deepfakes is getting easier, and access to large quantities of data is no longer a necessity, with FSL models requiring less data to generate manipulated media. With ever-evolving research and investment in AI and ML technologies, and ongoing increases in the number of easily accessible tools that do not require significant human resources and advanced hardware infrastructure,⁹⁹ more thorough use of deep fakes is likely.



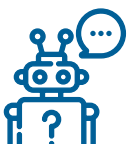
Quality Increases Across the Board

As a form of disinformation, synthetic media has the potential to become as ubiquitous as false stories are today. Automated bot accounts can already use machine learning algorithms to sound more human. When false content is no longer distinguishable from real media or its spread is undetectable, even by experts, people will not be able to separate fact from fiction.



Merging of Multiple Tools

With the augmented development of open-source tools and technology, the merging of synthetic creation processes is a field worth exploring. A plausible scenario in foreseeable future would be text generated with NLP models complemented with images or video grounded in descriptions in natural language.¹⁰⁰ This would allow disinformation actors to produce more coherent fake evidence that will be increasingly hard to debunk. Combined with ML-powered translation, this can result in bots commenting on posts not only with text in any number of languages but also with images, videos and memes instead.¹⁰¹



Human-Bot Mimicry vs. Authentic Superspreaders

Chatbots can be enhanced to target humans with tailored and precise content, and these technologies are increasingly likely to cause harm. Paired with human operators, social bots may soon better mimic human online behaviour, while leaving almost no detectable traces. This also shows how disseminators

⁹⁹ An already existing easy-access tool is [thispersondoesnotexist](#) and [thismapdoesnotexist](#).

¹⁰⁰ Katerina Sedova et al., "[AI and the Future of Disinformation Campaigns](#)", Center for Security and Emerging Technology, December 2021.

¹⁰¹ Alfonsas Juršėnas et al., "[The Double-Edged Sword of AI: Enabler of Disinformation](#)".

of disinformation are going to greater lengths to conceal their identities. Nonetheless, with more advances in prevention, an alternative, far more basic, tactic has emerged in recent years – the rise of “influencers” as authentic sources. These can often act as disinformers or spread well-intended misinformation. Opting for authentic content circulation, and thus evading automatic detection tools, can be regarded as an old tactic taking on a new colour, and one that is likely to be deployed more often in the near future.

Potential Future Risk Scenarios



The Use Of Emotional Recognition Technology

False narratives are written to arouse interest and activate emotions so that users respond to content, giving clues to ranking algorithms. Facial emotion recognition (FER) is an “affective computing” technology grounded in AI that analyses facial expressions from both images and videos, in order to recognise and interpret information about a person’s sentiments.¹⁰² If used maliciously, this technology can infer political attitudes and reactions from facial expressions, which in turn, can serve as a foundation for microtargeting and profiling. The knowledge of an individual’s emotions when attending political events – offline and online – could, for instance, facilitate information manipulation when microtargeting groups with comparable political affiliations.¹⁰³



Quantum Computing

Quantum computing may create a major threat to current encryption, which is based on the use of prime numbers. It can factor prime numbers exponentially faster than current algorithms, which would render regular asymmetric encryption useless. This could allow access to essentially any encrypted system. Although, strictly speaking, this is a cybersecurity issue, this could be used to plant disinformation within trustworthy news sources, by changing content in articles after having been published. Additionally, on an institutional level, it could be used to undermine, or even ruin public opinion about conventionally trustworthy sources by consistently publishing false content from these sources, diminishing the spread of accurate information.

¹⁰² Edita Fino et al., “Unfolding Political Attitudes through the Face: Facial Expressions When Reading Emotion Language of Left- and Right-Wing Political Leaders”, *Scientific Reports*, 30 October 2019.

¹⁰³ “TechDispatch #1/2021 – Facial Emotion Recognition”, European Data Protection Supervisor, 26 May 2021.



Not All AI is the Same

When looking at different tools, techniques and narratives that are used to shape disinformation, we can clearly see overarching patterns. Yet, artificial intelligence has become a blanket term that is expected to be of the same nature in all circumstances. Increased sophistication of disinformation, however, is not associated with a commonly known formula or standard that is evenly applied by all. A future risk does not only lie in a generalised AI that, unlike currently existing “plain” AI, could create vast quantities of disinformation without being decipherable as artificial but also because it draws on features that do not meet “the standard” and, therefore, is undetectable.



Conclusion

With ever-advancing technology, malicious actors will continue to weaponize information and develop increasingly sophisticated tools for disseminating manipulated content. This report has outlined how AI and ML tools can be used maliciously for disinformation purposes, focusing on synthetic content and text prediction. It has also focused on how different tactics are used and applied to propagate and perpetuate disinformation narratives, and how these very stories are exploiting already existing tools and techniques.

When taking a closer look at the narratives selected, we can see, however, that traditional means of propaganda and human-to-human amplification methods still constitute the undisputed main instrument in the disinformation toolkit. Nevertheless, an information sphere rooted in emotionality, bigotry and hate bait rather than facts offers fertile soil for more sophisticated tools that can grow from what already exists.

The goal of manipulated information produced on an assembly line is not necessary to change political views – a goal that is hard to achieve. Rather, it is to confuse, polarise, and entrench – causing division no matter what it takes. The more disinformation a user encounters online, the more it feeds into their already established worldview, rendering finding common ground with others almost impossible. The effect is self-evident – people giving up on any notion of truth. Although the formula for how disinformation will evolve in the future is unwritten and difficult to predict, the fact that we will need more ways to effectively address continued changes and advances in the production and dissemination of disinformation is not.

