



COMMONWEALTH PARLIAMENTARY ASSOCIATION

# PARLIAMENTARY HANDBOOK ON DISINFORMATION, AI AND SYNTHETIC MEDIA

Developed in partnership with



**OAS** | More rights  
for more people





**COMMONWEALTH  
PARLIAMENTARY  
ASSOCIATION**

**Stephen Twigg**  
Secretary-General

**Matthew Salik**  
Head of Programmes

**James Pinnell**  
Deputy Head of  
Programmes - Multilateral  
Engagement

**ORGANIZATION OF  
AMERICAN STATES**

**Luis Almagro**  
Secretary General

**Ivan Marques**  
Secretary of Multidimensional  
Security

**Alison August Treppel**  
Executive Secretary at Inter-  
American Committee against  
Terrorism (CICTE)

**Cristobal Fernandéz**  
Section Chief - DECO

**Kerry-Ann Barrett**  
Cybersecurity Program  
Manager - CICTE

**Moises Benamor**  
Chief of Representative  
Institutions - DSDS

**Carlos Baena**  
Cybersecurity Program Officer  
- CICTE

**Cassidy Bereskin**  
Technical consultant

**About the Author**

Cassidy Bereskin is a DPhil student and Clarendon Scholar researching the efficacy of content disclosure and provenance interventions for tackling deceptive synthetic media and disinformation. She is the Founder and Director of OxGen AI, a company that hosts the Oxford Generative AI Summit, a multi-stakeholder expert convening on generative AI and society (oxgensummit.org).

Previously, she conducted research as part of The Alan Turing Institute's Online Safety Team and worked on the COVID-19 Monitor, Canada's largest public opinion study on the pandemic.

She holds an MSc (Social Science of the Internet) from the Oxford Internet Institute and a BA (Political Science) from McMaster University, where she graduated as valedictorian.

**CONTENTS**

FOREWORD	1
ABSTRACT	2
EXECUTIVE SUMMARY	3
INTRODUCTION: HISTORY OF MEDIA AND INFORMATION DISTRIBUTION	4
EVOLUTION OF MEDIA AND INFORMATION DISTRIBUTION	5
DISINFORMATION: HISTORICAL AND CONTEMPORARY PERSPECTIVES	6
TAKING A CLOSER LOOK: THREATS OF DISINFORMATION IN THE INTERNET AGE	8
THE NEXT FRONTIER: AI AND SYNTHETIC MEDIA	11
MITIGATION STRATEGIES FOR SYNTHETIC DISINFORMATION	18
ROLE OF LEGISLATURES IN COMBATING SYNTHETIC DISINFORMATION	22
CONCLUSION	24
GLOSSARY OF TERMINOLOGY	25
NOTES	26

© Commonwealth Parliamentary Association 2023

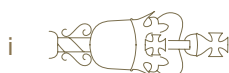
All rights reserved. This publication may be reproduced, stored, or transmitted in any form or by any means, electronic or mechanical, including photography, recording or otherwise provided it is used only for educational purposes and is not for resale, and provided full acknowledgement is given to the Commonwealth Parliamentary Association as the original publisher. Rights are not extended for the reproduction of any photography or design not owned by the Commonwealth Parliamentary Association as contained in this publication.

Views and opinions expressed in this publication are the responsibility of the Commonwealth Parliamentary Association Headquarters Secretariat and should not be attributed to any Parliament or Member of the Association.

Cover design and illustrations by Matthew Salik with elements from Parliament of Grenada, freepik.com and Shutterstock.

**Have you used this publication?**

**If you have, let us know as we are always keen to hear how our products are being used. Our details are on the back.**





# Foreword

---

Parliamentarians play a crucial role in protecting democratic processes and promoting transparency and accountability. It is vital that they are knowledgeable about the latest developments in disinformation, AI and synthetic media.

The spread of disinformation, misinformation, and propaganda are threats as old as democracy itself. However, in recent years, we have seen new actors and changes in the methods and tactics used to proliferate confusion and promote false narratives online, with the integrity of electoral processes and democracies around the world under potential threat. Social media and other digital platforms have provided fertile ground for the spread of such content, enabling malicious actors to reach large audiences and sow confusion and discord.

The emergence of new technologies such as artificial intelligence and synthetic media has added a new layer of complexity to the challenge of combating disinformation. These technologies can be used to create highly realistic but entirely fabricated content, making it increasingly difficult for individuals and organisations to distinguish between truth and falsehood. Given these challenges, it is critical that policymakers and legislators are equipped with the necessary tools and knowledge to address artificially-generated disinformation, the sources of such disinformation and the threats that these technologies carry when in the wrong hands.

Parliamentarians play a crucial role in protecting democratic processes and promoting transparency and accountability. It is vital that they are knowledgeable about the latest developments in disinformation, AI and synthetic media. A handbook designed specifically for Parliamentarians would serve as an essential resource to guide them in developing informed policies and legislation to combat disinformation, whilst also ensuring they avoid being drawn into and spreading false narratives themselves.

Those in public office are already feeling the effects of these new technologies, as they more frequently become the target of bots, deepfakes and other emerging tools, which in turn erode informed discourse and increasingly blur the lines between what should and shouldn't be trusted as objective reality. This Handbook provides a comprehensive overview of disinformation, including its different forms and the various techniques used to spread it. It also covers the basics of AI and synthetic media, including their potential applications and implications for democracy, within and beyond the electoral cycle. This Handbook also provides effective strategies for combating disinformation and guidance on how parliamentarians can work with other stakeholders, including civil society, the media, and technology companies, to develop comprehensive and effective policies and regulatory/legislative frameworks to address the challenges of disinformation, as well as how at the more personal level they can take steps to safeguard their own online presence and communication channels.

On behalf of the Commonwealth Parliamentary Association, I extend my thanks to the Organization of American States for their support in producing this Handbook and shared commitment to building awareness around the breadth and depth of misinformation, disinformation, fake news and the manipulation of information in undermining democratic norms and principles. This Handbook highlights the importance of collaborative approaches, and we believe this Handbook provides just one example of such collaboration in practice.

We hope Parliamentarians, parliamentary staff and practitioners across the world find value in this Handbook and we welcome any and all feedback on this document, including any initiatives undertaken by Members and institutions to combat disinformation and misinformation through AI.



Stephen Twigg, Secretary-General  
Commonwealth Parliamentary Association

# Abstract

---

The Handbook on Disinformation, AI and Synthetic Media illuminates the history, evolution and trajectory of online disinformation and examines its features, implications and associated countermeasures in the age of synthetic media and artificial intelligence. Two overarching contributions of the Handbook include a working typology for defining and disentangling the characteristics of synthetic media, as well as understanding the multifaceted concept of 'synthetic disinformation,' including its accessibility, efficiency, hyper-realism, personalisation and scalability compared to online disinformation. Synthetic disinformation poses substantive threats to democracy, including but not limited to the 'liar's dividend,' where malicious actors weaponize broad public scepticism around deepfakes to discredit genuine evidence.

Accordingly, the Handbook maps out multifaceted strategies to combat synthetic disinformation, involving transnational, national and multi-stakeholder initiatives, aimed at maintaining a delicate balance between mitigating risks and preserving democratic values and human rights. It culminates with policy recommendations focusing on multi-stakeholder coordination, the need for transparency and disclosure in digital content and bolstered public education, media literacy and research to tackle the multi-pronged challenges posed by synthetic disinformation.



# Executive Summary

---

## **1. EVOLUTION OF MEDIA, INFORMATION & DISINFORMATION:**

The Handbook traces the evolution from traditional to digital media, noting how platforms like social media have revolutionized information sharing and political discourse. However, it also acknowledges the dark side of this evolution: the spread of disinformation, which undermines democratic values and fosters polarization and mistrust among citizens.

## **2. SYNTHETIC MEDIA, AI, & SYNTHETIC DISINFORMATION:**

A key focus is the emergence of synthetic media, characterized by AI-generated content that is increasingly realistic and accessible. The Handbook marshals two unique contributions: a working typology for understanding the concept and facets of 'synthetic media'; and the concept of 'synthetic disinformation,' which this report characterises as more efficient, personalized and scalable compared to traditional online disinformation. The Handbook examines the threats posed by this new form of disinformation, including the deeper erosion of the public's ability to discern truth, the amplification of cynicism, fortification of ideological silos and the emboldening of malicious actors.

## **3. IMPLICATIONS FOR DEMOCRACY:**

The Handbook discusses how synthetic disinformation exacerbates existing challenges to democratic processes. For instance, it highlights the concept of the 'liar's dividend,' where the existence of synthetic media can be used to discredit genuine evidence. This phenomenon adds another layer of complexity to the already challenging task of maintaining informed public discourse and safeguarding democratic processes.

## **4. MITIGATION STRATEGIES:**

The Handbook maps out existing and emerging strategies to combat synthetic disinformation, which involve transnational, national and multi-stakeholder initiatives. While doing so, it advocates for the need for a balanced approach that mitigates the risks associated with synthetic disinformation while preserving democratic values and human rights.

## **5. POLICY RECOMMENDATIONS:**

This Handbook underscores the urgency and complexity of tackling disinformation in the age of AI and synthetic media. It calls for multi-stakeholder coordination among Legislatures, governments, AI providers and social media platforms to combat synthetic disinformation. It also emphasises the need for transparency and disclosure, such as clear content credentials, for users to discern authentic from synthetic media, as well as investments in public education, media literacy and research to develop and evaluate effective countermeasures against synthetic disinformation. These measures aim to address the multi-pronged challenges posed by synthetic disinformation and foster an informed public and a resilient democratic process.

# 1. Introduction: History of Media and Information Distribution



The exponential increase in the availability and reach of online information has increasingly shaped democracies around the world. In recent decades, social media platforms have transformed how citizens consume and share information, fostering information access, public debate and political awareness. At the same time, they have facilitated the spread and consumption of online disinformation, undermining citizens' ability to make informed decisions about who to trust and vote for, thereby threatening democracies and raising doubts about the integrity of democratic processes.

Throughout history, information has been increasingly crucial to human society, with its distribution evolving through oral storytelling, the printing press, and beyond. The printing press, for example, democratized knowledge in the 16th and 17th centuries, expanding public access to information and enabling people to influence public policy.<sup>1</sup> However, this also paved the way for the dissemination of misleading information. Subsequent advancements like the proliferation of newspapers in the 17th and 18th centuries and the advent of radio and television in the 20th century continued to shape public opinion and democratic participation.<sup>2</sup> These forms of media not only increased information accessibility but also introduced new avenues for spreading inaccurate information.

1. [Origins of Democratic Culture: Printing, Petitions, and the Public Sphere in Early-Modern England](#)

2. [Television and Voting Turnout](#)



## 2. Evolution of Media and Information Distribution



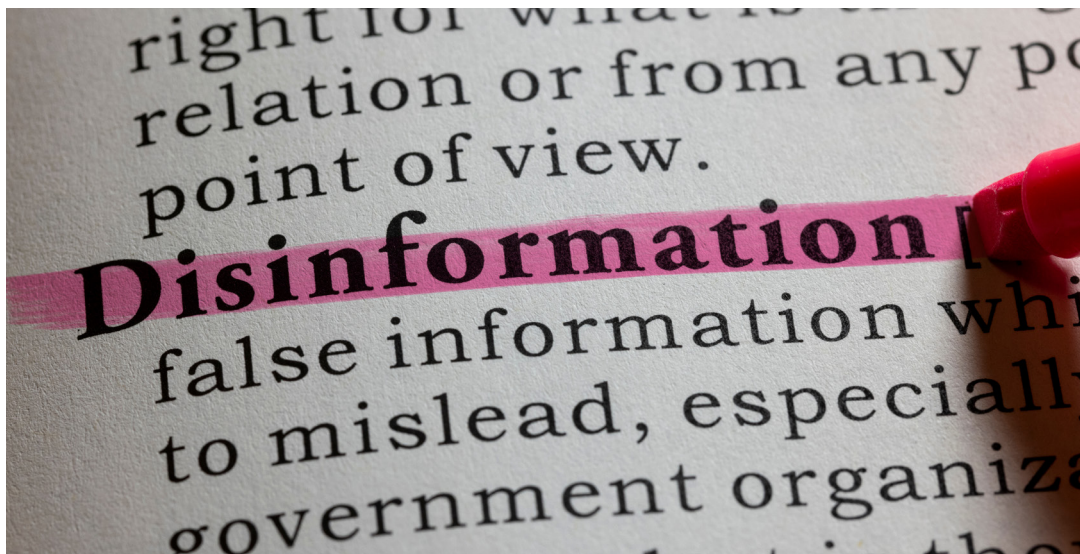
The advent of digital technology and social media has drastically expanded the power of media. This transformation demands a thorough understanding from legislative bodies to recognise the extraordinary speed and reach of communication on social media, as well as to effectively manage and mitigate the associated risks to democratic processes. Platforms like *Facebook*, *Twitter*, and *Instagram* have revolutionized information sharing, playing a crucial role in the dissemination and exchange of ideas about politics and public policy. In addition to amplifying political engagement and lowering the barrier to entry for political participation,<sup>3</sup> they also empower politicians to communicate directly with the public, extending beyond traditional media gatekeepers.<sup>4</sup>

However, these platforms pose substantive threats to democracy. The same features that empower citizens can be misused as vehicles for the spread of online hate speech, disinformation, bias, harassment and cyberbullying. Malicious actors exploit the rapid, viral nature of social media as conduits to damage public trust and disrupt democratic life. As a result, Parliaments face the challenge of crafting legislation that balances freedom of expression with the need to curb harmful online activities. For instance, legislation targeting harmful speech must be finely tuned to avoid inadvertently creating ‘chilling effects’ and suppressing lawful and legitimate speech. Parliaments must be thoughtful in calibrating responses that address urgent threats while not encroaching on essential democratic values. Social media have armed both democratic and malicious actors with powerful new tools. The next section will delve into the issue of disinformation and its impact on democracies.

3. [Social Media and Democracy: Fostering Political Deliberation and Participation](#)

4. [Social media in political communication A substitute for conventional media?](#)

## 3. Disinformation: Historical and Contemporary Perspectives



### 3.1 DEFINING DISINFORMATION

The term ‘disinformation’ has gained prominence in recent years and refers to the deliberate creation and sharing of false information intended to deceive.<sup>5</sup> This distinguishes it from ‘misinformation,’ which may be false but is not spread with the intention of misleading. Another related term is ‘malinformation,’ which involves sharing true information with harmful intent.

‘Conspiracy theories’ and ‘computational propaganda’ are also relevant in discussions about disinformation. Conspiracy theories commonly refer to explanations or narratives that suggest that large-scale events are orchestrated by secretive, powerful groups.<sup>6</sup> Though they may contain disinformation, they differ by focusing on hidden agendas and orchestrated events. “Computational propaganda”<sup>7</sup> often overlaps with disinformation but emphasises the role of automation, like bots—or ‘automatic software built to mimic real users’<sup>8</sup>—and algorithms—automated systems used by social media to spread of content—as tactics for purveying disinformation through social media at scale.<sup>9</sup>

### 3.2 HISTORICAL CONTEXT

Disinformation is far from a new phenomenon. It has been used throughout history to manipulate public opinion and achieve political ends, ranging from the spread of rumours by ancient leaders<sup>10</sup> to the advanced tactics employed by contemporary state actors. Notable historical examples like Operation INFEKTION<sup>11</sup> highlight the impact of disinformation. Orchestrated by the Soviet Union during the Cold War, this campaign aimed to discredit the United States by falsely suggesting it had engineered the HIV/AIDS virus as a bioweapon. This instance underscores the enduring power of disinformation to shape public opinion and the political landscape.

5. [Understanding Information disorder](#)

6. [Understanding conspiracy theories](#)

7. [Automation, Algorithms, and Politics | Political Communication, Computational Propaganda, and Autonomous Agents — Introduction](#)

8. [Introduction: Computational Propaganda Worldwide](#)

9. [Programme on Technology and Democracy](#)

10. [The fake news that sealed the fate of Antony and Cleopatra](#)

11. [NYT Operation Infection](#)





### 3.3 DISINFORMATION IN THE AGE OF THE INTERNET

The digital age has transformed the mechanisms and accelerated the distribution of disinformation, which is carried out by a mix of state and non-state actors employing diverse tactics. Disinformation campaigns can be highly sophisticated and leverage a range of tactics, including but not limited to, bots<sup>12</sup> and microtargeting<sup>13</sup> to spread false information, drown out dissenting voices, and create an illusion of widespread support for a particular perspective.

For instance, bots commonly refer to automated software agents designed to mimic real users on social media. These can be programmed to post content, create fake personas, engage with users and artificially amplify the engagement and reach of disinformation. Micro-targeting, on the other hand, often involves strategically using consumers' online data to personalize and disseminate tailored disinformation to specific, often narrow, audience segments, helping to increase its persuasiveness and potential impact.

These tactics have expanded the reach, scalability and precision of disinformation operations, which aim to disrupt elections and muddy public opinion.<sup>14</sup> In addition, the role of social media platforms is crucial, as their algorithms not only amplify the reach of disinformation but also enable targeting specific audiences,<sup>15</sup> making disinformation more difficult to detect and counter. As Phil Howard puts it in his seminal book *Lie Machines*, "Political actors are getting very good at producing big lies, social media algorithms provide an effective way of distributing those lies, and the science of marketing lies to the right audience is improving."<sup>16</sup>

### 3.4 DISINFORMATION: CONTEXTUALIZING THREATS TO DEMOCRACY

These powerful tactics pose substantive risks to democratic processes and elections, notably by sowing epistemic cynicism, distrust and doubt, and hindering citizens' ability to make informed decisions and participate in public life. Beyond distrust and doubt, it fuels polarisation and reinforces and entrenches insular online political communities. These threats extend beyond electoral integrity to spheres such as public health and national security, making it imperative for Parliaments to understand and tackle the multifaceted threats posed by disinformation.

---

12. [The spread of low-credibility content by social bots](#)

13. [On Microtargeting Socially Divisive Ads: A Case Study of Russia-Linked Ad Campaigns on Facebook](#)

14. [Industrialized Disinformation: 2020 Global Inventory of Organized Social Media Manipulation](#)

15. [Social Media and Fake News in the 2016 Election](#)

16. [Lie Machines](#)

# 4. Taking a Closer Look: Threats of Disinformation in the Internet Age

## 4.1 SIGNIFICANCE OF DISINFORMATION IN THE SOCIAL MEDIA AGE

Disinformation in the age of social media poses complex threats that extend beyond electoral integrity to realms such as public health and national security. With the ubiquity and accessibility of social media, spreading and proliferating false information on a large scale is easier and faster than ever. This rapid spread often outpaces<sup>17</sup> the ability of fact-checkers to respond, skewing the information landscape toward disinformation. According to a widely cited study conducted by Soroush Vosoughi and his colleagues, false information tends to spread faster than accurate information. Their research revealed that the top 1% of false news stories reached between 1000 and 100,000 people, while truthful information rarely reached more than 1000 people,<sup>18</sup> fostering an information environment where accurate information is drowned out by misinformation and disinformation alike.

Notably, not all instances of false information are driven by malicious intent. For instance, a piece of disinformation may originate with malign intentions, but as it circulates, it can be shared by individuals who genuinely believe it to be true, without any malicious motives, entrenching disinformation narratives. As such, disinformation's impact on democracies is both direct and pervasive. It has become a recurring feature in global elections, often orchestrated by authoritarian states to discredit democratic systems and opponents. It is worth noting that interference in democratic processes need not always be foreign; it can also originate domestically. Beyond electoral interference, disinformation seeks to exploit pre-existing biases and deepen political divisions, undermining public trust and social cohesion. As Philippe J. Graton observes, interference in elections *'can harm not only democratic processes but also critical infrastructure and economic stability.'*<sup>20</sup>

## 4.2 A CASE STUDY: THE COVID-19 INFODEMIC

The public health implications of disinformation have also been glaringly evident, especially during the COVID-19 pandemic. False information about the virus and its treatments circulated widely, fostering mistrust<sup>21</sup> and vaccine hesitancy.<sup>22</sup> A 2021 study in the *American Journal of Tropical Medicine and Hygiene* highlighted that such misinformation led to over 800 deaths and 5,800 hospitalizations,<sup>23</sup> underscoring the real-life implications of disinformation on public health, where lives were lost and healthcare systems strained.<sup>24</sup>

The conspiracy theory that Bill Gates was exploiting the pandemic to promote vaccines with a microchip capable of tracking people figured prominently in disinformation campaigns,<sup>25</sup> muddying public discourse and cementing distrust in the safety of the vaccines. Disinformation narratives also advocated for ineffective treatments such as hydroxychloroquine, an antimalarial and autoimmune medication.<sup>26</sup> This hesitation and muddled understandings of the vaccine and virus contributed to delayed vaccination, which increasingly prolonged the impact of the pandemic.<sup>27</sup>

17. [The Rise of Political Fact-Checking in American Journalism](#)

18. [The spread of true and false news online](#)

19. [The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation](#)

20. [Threat resilience in the realm of misinformation, disinformation, and trust](#)

21. [Coronavirus conspiracy beliefs, mistrust, and compliance with government guidelines in England](#)

22. [Understanding Parental Concerns about COVID-19 Vaccination](#)

23. [COVID-19-Related Infodemic and Its Impact on Public Health: A Global Social Media Analysis](#)

24. Ibid

25. [COVID-19 Vaccine Misinformation Campaigns and Social Media Narratives](#)

26. [COVID-19 Truths, Lies, and Consequences](#)

27. Ibid



## 4.3 ACTORS, MOTIVATIONS AND TACTICS

### 4.3.1 Actors

The actors behind disinformation campaigns, as well as their level of coordination, can vary widely. The spectrum ranges from international to local actors, encompassing state and non-state entities, state-sponsored groups, commercial firms and individuals by ideological or financial agendas. These actors exploit an array of tactics such as employing advanced botnets, malicious automated networks, intricate troll farms leveraging both social media ad campaigns and mainstream media vectors to disseminate their narratives. The level of coordination among actors varies and can range from independent actors and highly organised efforts, involving state-sponsored groups and agencies working in tandem, as well as more fragmented and ad hoc collaborations. Understanding the extent and nature of this coordination is crucial, as it influences the scale, sophistication and impact of the tactics employed.

For instance, the Internet Research Agency (IRA), a St. Petersburg-based entity, engaged in extensive disinformation campaigns targeting both the United States and Europe. The IRA adopted strategies like account buying, ‘follower fishing,’ and narrative switching, a disinformation tactic where agents such as bots initially engage with benign topics before abruptly pivoting to politically charged narratives to influence public opinion, to expand the influence of their ideologically charged content.<sup>28</sup> In stark contrast to these highly organised efforts, individual actors can also have a considerable impact using disinformation. A well-documented example is the proliferation of ‘fake news’ during the 2016 U.S. election. Websites and social media pages produced entirely fictitious articles that were designed to exploit biases and attract clicks, such as the claim that ‘*Pope Francis endorses Donald Trump for President.*’ This disinformation was not just the work of state actors but also included individual entrepreneurs like those in Macedonia, where teenagers created fake news sites about American politics and generated revenue from virality among specific political groups.<sup>29</sup>

The diversity of disinformation agents is thus evident—from highly organised, state-sponsored operations like those of the IRA, which systematically orchestrated campaigns, to ordinary individuals. The level of coordination among actors varies and can range from independent actors and highly organised efforts, involving state-sponsored groups and agencies working in tandem, as well as more fragmented and ad hoc collaborations. Understanding the extent and nature of this coordination is crucial, as it influences the scale, sophistication and impact of the tactics employed.

### 4.3.2 Motivations

While the objectives can differ, disinformation efforts often aim to service ideological or financial interests. State-backed disinformation campaigns, for example, seek to warp public opinion around specific topics or candidates and “*undermine democratic processes by fostering doubt and destabilizing the common ground that democratic societies require.*”<sup>30</sup> These efforts can involve creating and disseminating disinformation across platforms, amplifying it through social media and engaging in targeted messaging to achieve their objectives.

### 4.3.3 Tactics and reach

Historically, the creation and dissemination of disinformation demanded significant human effort. For instance, the Russian Internet Research Agency (IRA) serves as a prominent case where an ‘*organised team*’ or ‘*troll army*’ was employed to manipulate social media in favour of Russian interests.<sup>31</sup> These efforts required substantial human resources and 12-hour workdays, with employees trained to churn out a constant stream of disinformation.<sup>32</sup>

However, technological advances have simplified this process. Automation and ‘bots’ now play an extensive role in disseminating disinformation. These automated accounts

28. [How Russia's Internet Research Agency Built its Disinformation Campaign](#)

29. ‘[Fake news](#)’ went viral in 2016. This expert studied who clicked

30. [Disinformation's spread: bots, trolls and all of us](#)

31. [All the News That's Fit to Fabricate: AI-Generated Text as a Tool of Media Misinformation](#)

32. [Lie Machines](#)



mimic human behaviour and spread divisive content at a scale previously unattainable by human efforts alone.<sup>33,34</sup> For example, in the lead-up to the 2016 U.S. presidential election, researchers identified around 400,000 bots as making up approximately one-fifth of all *Twitter* conversations, issuing nearly 3.8 million tweets.<sup>35</sup> During the 2017 Catalan referendum, bots promoting divisive content exacerbated online social conflicts.<sup>36</sup> In a similar vein, in the Brazilian elections of 2018, the use of computational propaganda was evident, with encrypted chat apps such as *WhatsApp* being exploited for disseminating disinformation in favour of Jair Bolsonaro.<sup>37</sup> These examples exemplify how technological advances have dramatically escalated the scale of disinformation campaigns, escalating threats to democratic societies.

#### 4.3.4 Role of social media platforms

While recent research on the impact of social media platforms in spreading disinformation is heterogeneous, it's still critical to acknowledge their explicit role. Their algorithms, designed to maximize user engagement,<sup>38</sup> can amplify false information and extend its reach.<sup>39</sup> While there are some disagreements in peer-reviewed literature, a large corpus of studies suggests that these platforms also foster echo chambers,<sup>40,41,42</sup> or self-contained online communities of like-minded individuals. Echo chambers are concerning in the context of disinformation. Studies highlight that disinformation often '*preaches to the choir*,<sup>43</sup> reinforcing rather than changing individual attitudes. When individuals are repeatedly exposed<sup>44</sup> to information that aligns with their existing beliefs, it may fortify preconceived disinformation beliefs, making them increasingly challenging to counteract. As such, disinformation not only hinders citizens' critical thinking and the ability to discern authentic content but potentially also disempowers, polarizes and fragments society at large.

#### 4.4 DISINFORMATION: IMPLICATIONS FOR DEMOCRACY

Disinformation has significant implications for democracies, chipping away at common understandings of truth and reality. Hannah Arendt lucidly argued that facts form the 'texture' of politics, serving as a foundation for our collective reality, or the ground upon which we base our sense of orientation in the real world.<sup>45</sup> Similarly, Kant stresses that common sense is essential for universally shared knowledge.<sup>46</sup> This shared factual basis enables the public to form informed opinions about matters of common concern, assess the government's performance in representing their will, and evaluate its pursuit of the common good. This foundation is crucial for informed decision-making and the healthy functioning of democratic institutions. For instance, without a collective foundation of reality, citizens may have a growing sense of cynicism and a loss of belief in anything, creating fertile ground for anti-democratic and authoritarian leaders to exploit this uncertainty. Accordingly, there is an urgent need for interventions that safeguard democratic integrity and preserve the "texture" of facts upon which our collective reality depends.

33. [How disinformation operations against Russian opposition leader Alexei Navalny influence the international audience on Twitter](#)

34. [Computational Propaganda Worldwide: Executive Summary](#)

35. [Social bots distort the 2016 U.S. Presidential election online discussion](#)

36. [Bots increase exposure to negative and inflammatory content in online social systems](#)

37. [How Disinformation on WhatsApp Went From Campaign Weapon to Governmental Propaganda in Brazil](#)

38. [How Social Media's Obsession with Scale Supercharged Disinformation](#)

39. [Is Social Media a Threat to Democracy?](#)

40. ['Echo Chambers': Partisan Facebook Groups during the 2014 Thai Election](#)

41. [Echo Chambers Exist! \(But They're Full of Opposing Views\)](#)

42. [Bridging Echo Chambers? Understanding Political Partisanship through Semantic Network Analysis](#)

43. [Less than you think: Prevalence and predictors of fake news dissemination on Facebook](#)

44. [The illusory truth effect leads to the spread of misinformation](#)

45. [Between Past and Future](#)

46. [Kant's Common Sense and the Strategy for a Deduction](#)



# 5. The Next Frontier: AI and Synthetic Media



## 5.1 CONTEXT: ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

The field of Artificial Intelligence (AI), a term coined by John McCarthy in 1956, has evolved significantly in the last decade, and is increasingly shaping the digital media environment. Central to this advancement is machine learning, a subset of AI where algorithms learn from data to perform tasks traditionally associated with human intelligence, such as visual perception and decision-making. In machine learning, models are 'trained' by adjusting their mathematical parameters through exposure to large datasets. This training enables the model to identify patterns and relationships in the data, improving its ability to make accurate predictions.

The applications of AI span sectors like healthcare, finance, retail, transportation, agriculture and communication. For instance, in healthcare, AI is being used for disease detection and early intervention through the analysis of medical images. In finance, it is aiding in fraud detection, market trend analysis, and trading optimization by processing large datasets in real-time. Yet, AI also presents risks that could threaten democratic societies and processes, such as cyber-attacks, disinformation, bias and discrimination.

In the context of legislative processes, the implementation of AI and generative technologies can be transformative. For instance, they offer tools for automating repetitive procedures, analysing data for valuable insights, harmonizing legal texts, and streamlining legislative functions. However, the deployment of AI in such a critical domain necessitates a deep understanding and a commitment to responsible, ethical and equitable practices, particularly when it comes to its development and application within parliamentary systems.

In his book *'Generative Deep Learning'*, David Foster categorises AI into two broad types: discriminative and generative modelling. Discriminative models focus on predicting relationships between input and output, such as identifying objects in images. In contrast, generative models are designed to create new data that resembles the input, such as generating entirely new images. Recent advancements in generative models have given rise to what is commonly known as 'synthetic media.' The next section of this report will delve into synthetic media, its implications and its role in the wider context of disinformation and the digital media environment.

## 5.2 DEFINING SYNTHETIC MEDIA

In her 2020 book *Deepfakes*, generative AI expert Nina Schick warned that soon “anyone with a smartphone will be able to produce Hollywood-level special effects at next to no cost, with minimum skill or effort.” In 2023, this prediction is proving increasingly prescient, with a proliferation of ‘synthetic media’, or content—be it visual, auditory, or multimodal—that has been created or altered using artificial intelligence.<sup>47</sup> These models can be fine-tuned to generate a variety of highly realistic and convincing outputs that ‘may simulate artifacts, persons or events.’<sup>48</sup> *Deepfakes* use deep learning, a subset of machine learning, to generate realistic videos or images where people appear to say or perform actions they never did. For instance, a viral AI-generated image of Pope Francis in a white puffer jacket, described as “the first real mass-level AI misinformation case”, illustrated concerns about deepfakes’ hyper-realism and potential proneness to misuse.



## 5.3 A NUANCED FRAMEWORK FOR UNDERSTANDING SYNTHETIC MEDIA

As synthetic media is an umbrella term, this report proposes a working typology to understand its elements. Our typology highlights two overarching dimensions and four (2x2) subdimensions, including a) content, with subdimensions of ‘purpose,’ ‘intent,’ and ‘contextual believability,’ and b) format, with subdimensions including ‘media type’ and ‘realism.’ These dimensions and relevant examples are illustrated in the diagram below, which proposes a conceptual breakdown of key dimensions of the wider concept.

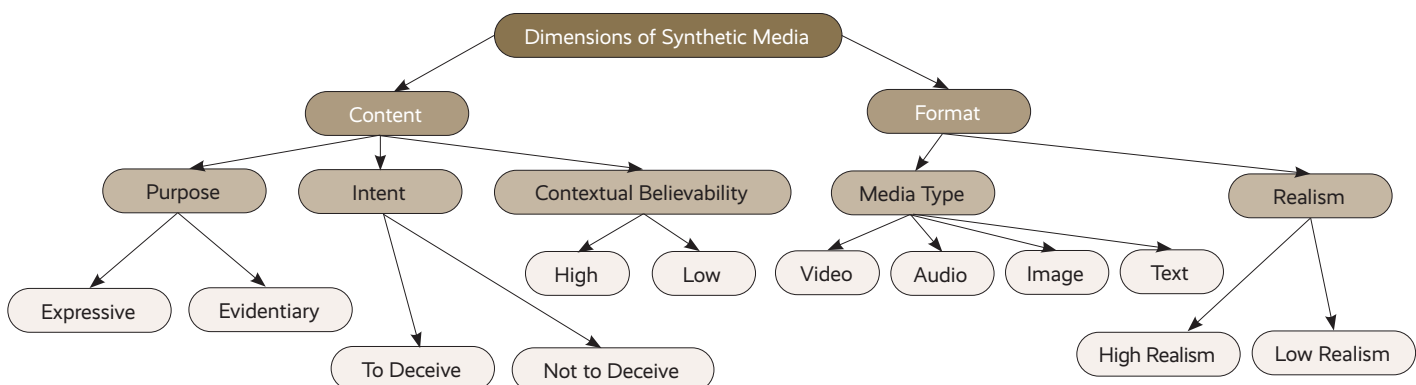


Diagram mapping out this report’s working framework for understanding synthetic media

47. [PAI’s Responsible Practices for Synthetic Media A Framework for Collective Action](#)

48. Ibid

49. [The Pope Drip](#)

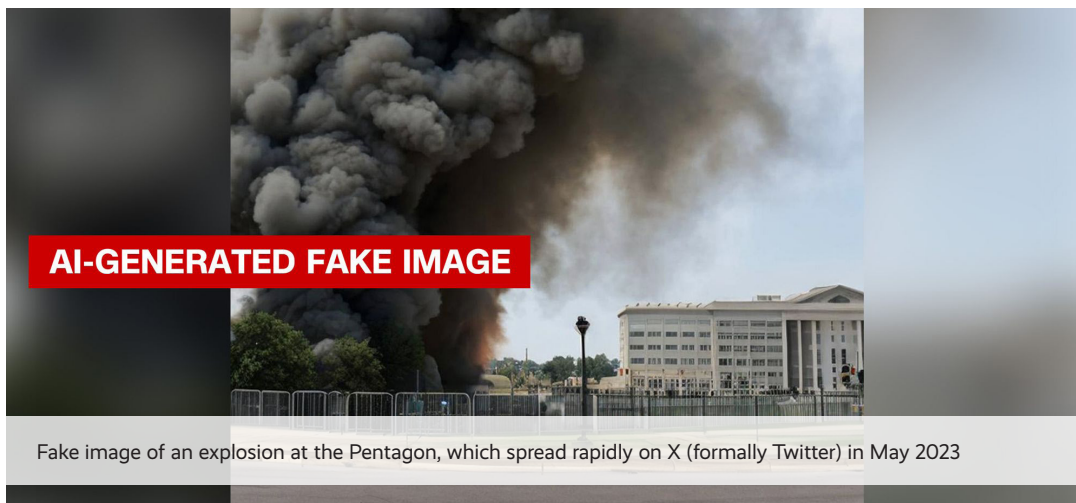


## 5.4 CONTENT

The content dimension of synthetic media refers to the characteristics related to the message or information that the media is conveying. This dimension can be broken down into subdimensions that help to further categorise and understand the nature of the content within the synthetic media.

### 5.4.1 Purpose

The subdimension ‘purpose’ refers to the intended use or the role that synthetic media is meant to play. It is about the ‘why’ behind the creation of the content. The purpose of synthetic media distinguishes between two categories, namely ‘evidentiary’ content, where synthetic media is designed to serve as evidence or provide a factual basis for information and decision-making. For instance, in May 2023, a synthetic media falsely depicting and purportedly evidencing an explosion at the Pentagon spread on social media, briefly affecting the stock market and raising concerns about the weaponization of synthetic media to create disinformation and spread misinformation (see image below).<sup>50</sup> ‘Expressive’ content, on the other hand, is created to supplement, rather than factually substantiate, information, such as a stock photo in a news article.



### 5.4.2 Intent

The second content subdimension ‘intent,’ refers to whether the content is created with the intent to deceive. While deceptive and non-deceptive synthetic media can take many forms,<sup>51</sup> particularly relevant to this report is the distinction between ‘synthetic disinformation’ and ‘synthetic misinformation.’ As previously mentioned, disinformation refers to the intentional creation and/or sharing of false or misleading online content, whereas misinformation is not intentionally maliciously spread. Synthetic disinformation can, but not always, take the form of deepfakes, which use deep learning to generate realistic videos, images, and audio of people doing or saying things they have never said.

For instance, in October 2023, a viral deepfake audio clip featured Rt Hon. Sir Keir Starmer, MP, (Leader of the Opposition in the UK) and leader of the UK’s Labour Party, berating staffers. The audio clip was released on the first day of his party’s annual conference and quickly gathered 1.5 million hits<sup>52</sup>. This clip, characterised as the first ‘deepfake moment’ in UK politics, came soon after an election in Slovakia, where a deepfake audio clip surfaced of Michal Simecka, the leader of the Progressive Slovakia Party, appearing to discuss rigging the election with a prominent journalist.<sup>53</sup>

50. [Fake AI-generated image of explosion near Pentagon spreads on social media](#)

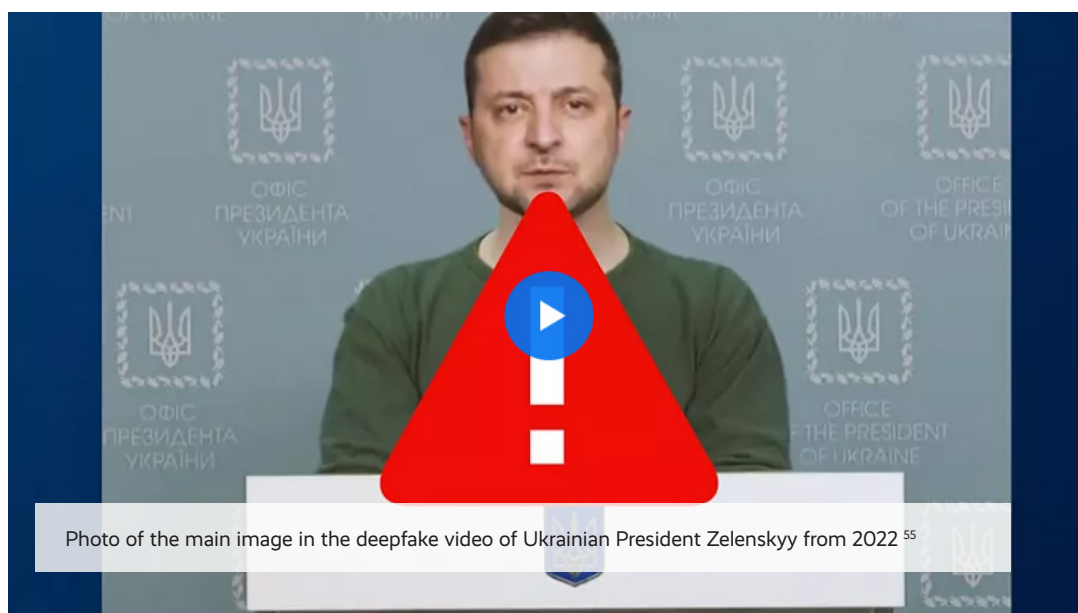
51. I.e. spam (deceptive intent); artistic expression (non-deceptive intent)

52. [Keir Starmer suffers UK politics’ first deepfake moment. It won’t be the last](#)

53. [Slovakia’s Election Deepfakes Show AI Is a Danger to Democracy](#)

### 5.4.3 Contextual believability

A third content category, believability, refers to the extent to which synthetic media is perceived as realistic or probable within a given context. A deepfake video of a political leader making statements that are drastically out of character, or contrary to known facts, may have lower contextual believability. For instance, in March 2022, a deepfake depicted Ukrainian President Volodymyr Zelenskyy in a video telling people in Ukraine to surrender and lay down arms was widely perceived as a deepfake given both technical and visual clues<sup>54</sup>, in conjunction with articulating a message that clashed with his prevailing ideological orientation. Believability can be challenging to measure and encompasses the nuanced interplay between the media's technical fidelity and realism and the expectations and knowledge of its audience within a specific context.



## 5.5 FORMAT OF SYNTHETIC MEDIA

The second overarching dimension of synthetic media is 'format,' or the presentation of the content. It is concerned with the medium through which the synthetic content is delivered as well as its degree of technical and aesthetic fidelity to reality.

### 5.5.1 Media type

This subdimension categorises synthetic media based on its medium. Common types include video, audio, image, text and multi-modal. Each media type has its own set of underlying technical foundations. For instance, in the realm of text, large language models like GPT (Generative Pre-trained Transformer), which power ChatGPT, represent a significant advancement. Deepfake technology primarily pertains to video (however, not always), while text-to-speech synthesis relates to audio. The multi-modal category integrates these varied forms, creating complex synthetic experiences. The media type can inform the potential use cases of the media and help to illustrate how it is developed and deployed.

### 5.5.2 Realism

This aspect refers to how convincingly the synthetic media replicates the appearance, behaviour and attributes of the subject it portrays. This can range from stylised or abstract, which are easily identifiable as synthetic, to highly realistic, which are nearly indistinguishable from reality. For instance, a relatively more realistic deepfake in a political context involved ads supporting Florida Governor Ron DeSantis, which featured deepfake photos depicting former

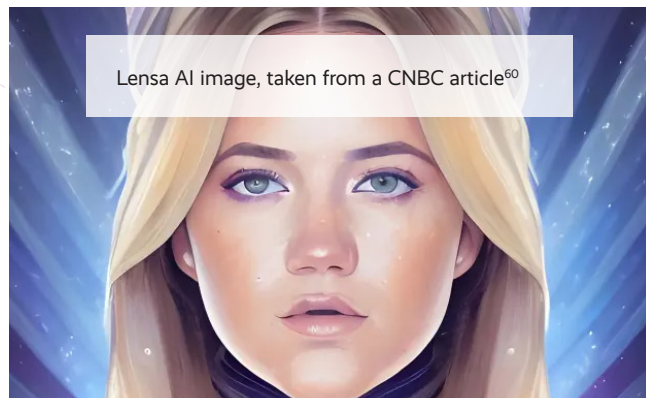
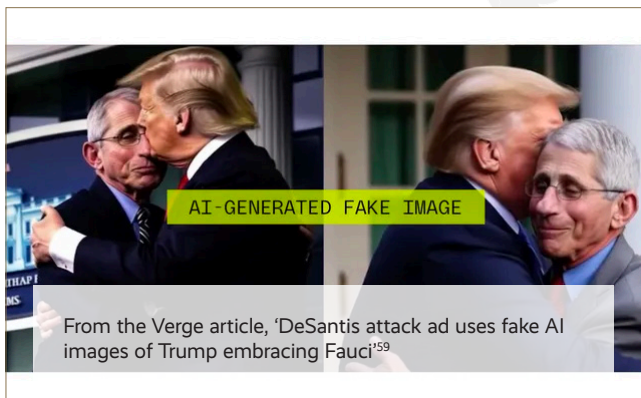
54. I.e. it is easily spottable, such as voice being deeper and slower than Zelenskyy's normal voice; Zelenskyy's head is slightly outsized for the body it is attached to

55. [Deepfake Zelenskyy surrender video is the 'first intentionally used' in Ukraine war](#)





President Donald Trump embracing Anthony Fauci. These manipulated images raised concerns among election officials and advocates due to their potential to mislead voters.<sup>56</sup> On the other hand, companies such as Lensa AI, initially launched in 2018 as a photo editing app, gained popularity with its “Magic Avatars” feature, which creates more obvious AI-generated digital self-portraits in various artistic styles using the open-source image generator Stable Diffusion. The app offers a subscription service, with additional charges for the “Magic Avatar” tool, and requires users to upload selfies to generate personalized avatars in styles like anime or fairy princesses (see image below).<sup>57</sup> However, it has also raised concerns regarding privacy and the use of artists’ work in training its AI models.<sup>58</sup> The level of realism is crucial in determining how the audience perceives and interacts with the media. Synthetic media that exhibits a high level of realism includes deepfakes or photorealistic CGI, which are designed to replicate the nuances of real-world entities so closely that they can be mistaken for actual footage or images. On the other end of the spectrum, synthetic media with a low level of realism might include cartoonish avatars or clearly computer-generated voices that are not intended to be mistaken for a real person or object.



## 5.6. THE NEXT FRONTIER: SYNTHETIC DISINFORMATION

Synthetic media has ushered in a new wave of complexities and challenges in combating online disinformation. As previously mentioned, a widely cited definition of disinformation is the intentional creation and/or dissemination of false or misleading online content. We build on this to define ‘synthetic disinformation’ as disinformation generated or enhanced via synthetic techniques. Synthetic disinformation can, but not invariably, take the form of deepfakes, which uses deep learning to generate realistic videos, images and audio of people doing or saying things they have never said. Synthetic media, and their underlying generative AI technologies, have significantly lowered the barrier to entry for producing, personalising and scaling disinformation, threatening to turbocharge its scale, reach and potential impact.

## 5.7. FEATURES OF SYNTHETIC DISINFORMATION

What is new about the threat of disinformation in the age of synthetic media? While disinformation spans centuries (if not millennia) and has already been transformed over the decades in the era of digital media, this report argues that synthetic media accelerates the generation, distribution and potential impact of disinformation. Notably, it deepens the erosion of citizens’ ability to make informed decisions about who to trust for information and guidance and what to believe.

### 5.7.1 Accessibility and efficiency

The accessibility and efficiency of creating synthetic disinformation are unprecedented. Again, in her 2020 book *Deepfakes*, Nina Schick writes that “*Hollywood-level special effects will soon become accessible to everyone. This is an extraordinary development with unforeseen*

56. [Washington grapples with AI deepfakes on the campaign trail](#).

57. [Here's how to use Lensa, the chart-topping app that uses AI to transform your selfies into digital avatars](#).

58. [What You Should Know Before Using the Lensa AI App](#)

59. [DeSantis attack ad uses fake AI images of Trump embracing Fauci](#)

60. [Here's how to use Lensa, the chart-topping app that uses AI to transform your selfies into digital avatars](#).

implications for our collective perception of reality.” With advancements in generative AI and the public release of text-to-image, text-to-video and models with multi-modal capabilities like OpenAI’s GPT\*4, synthetic media is becoming increasingly democratised. The ability to create and purvey convincing synthetic media and disinformation at scale is available to the masses and no longer limited to programmers and individuals with significant resources.

This widespread accessibility and efficiency amplify the potential for misuse, enabling a wider range of actors, from state-sponsored entities to individuals, to participate in the creation and spread of synthetic disinformation at scale. The world is already experiencing the effects of this democratisation, as instances of synthetic disinformation that have gone viral in recent years, such as the synthetic image of Trump being arrested by New York City law enforcement (see below), were created by a single individual using tools like Midjourney.<sup>61</sup>



### 5.7.2 Hyper-realism

In her seminal 1979 book *On Photography*, Susan Sontag writes that photographs are often mis-perceived as ‘miniatures of reality’ that provide ‘incontrovertible proof that a given thing happened!’ On the contrary, she writes, “despite the presumption of veracity that gives all photographs authority, interest, seductiveness, the work that photographers do is no generic exception to the usually shady commerce between art and truth.” The advent of hyper-realistic synthetic media, particularly contextually believable deepfakes, underscore the necessity for public scepticism and rethinking the adage ‘seeing is believing.’

Synthetic media, capable of creating highly lifelike depictions of events or actions that never actually took place, intensify the challenges Sontag identified with traditional photography. This evolution in media technology, marked by its ability to distort reality convincingly, presents a pressing issue for modern information consumption and warrants a heightened awareness and understanding among Parliamentarians and the public alike. Reputational damage is a significant consequence of this hyper-realism. For example, the previously highlighted deepfake audio recording circulated in the lead up to Slovakia’s 2023 general election depicted Progressive Slovakia Party leader Michal Simecka discussing buying votes from the Roma minority and cracking a joke about child pornography. Deepfakes can cause serious reputational damage during elections and in public life, which can have downstream consequences such as distorting citizens’ views of candidates and voting preferences.

### 5.7.3 Scalability

With advancements in artificial intelligence and machine learning, the production of synthetic media can be automated, enabling the creation of large volumes of disinformation at an unprecedented scale. For instance, large language models can quickly generate credible-sounding disinformation at scale, alleviating the burden of relying on humans to manually write disinformation posts, scaling up the automatic generation of text to produce and spread large quantities of disinformation in seconds.<sup>62</sup> For instance, in April 2023, NewsGuard identified 49 websites across seven languages that are using AI to generate large volumes of clickbait articles for advertising revenue. These sites, usually lacking clear ownership disclosure, appear to be designed for revenue generation through programmatic ads, similar to earlier human-operated content farms, but with the added capability and scalability enabled by AI technology.<sup>63</sup> This scalability has deleterious consequences for democracies. Notably, it not only potentially

61. [AI-generated images of Trump being arrested circulate on social media](#)

62. [All the News That’s Fit to Fabricate: AI-Generated Text as a Tool of Media Misinformation](#)

63. [Rise of the Newsbots: AI-Generated News Websites Proliferating Online](#)



increases the volume of disinformation on social media but also allows new entrants into the disinformation game, overwhelming governments, platforms and fact-checkers. The scalability is further enhanced by the fact that once a synthetic media model is trained, it can generate new content with minimal incremental cost. This means that a single piece of synthetic media, such as a deepfake video, can be the template for generating countless variations, each personalised to target different groups or individuals.

#### 5.7.4 Personalization and hyper-targeting

Synthetic media's ability to be hyper-targeted or personalised to a particular audience, is of particular concern in the context of disinformation and its impact on democracy. Existing research indicates that disinformation often '*preaches to the choir*,'<sup>64</sup> reinforcing rather than changing individual attitudes. This phenomenon may be explained by confirmation bias or motivated reasoning, where individuals tend to favour information that confirms their pre-existing views. When synthetic media is personalised, it exploits this bias, making disinformation more likely to find a receptive audience among those already inclined to believe it. People are more likely to engage with and share content that resonates with their existing opinions, threatening to deepen ideological divides and fortifying extreme views. This has significant potential downstream consequences, such as embedding and entrenching 'alternative facts' or realities, and catalysing anti-democratic actions, as seen in events like the January 6th Capitol attack in 2021 and the Canadian Truckers' convoy protest in 2022. In this context, synthetic media not only threatens to degrade the quality of democratic debate but also potentially hardens false beliefs, with serious downstream effects. Accordingly, institutions are taking steps to combat the use of tools such as ChatGPT, such as academics to detect AI-generated essays and legislative bodies to prevent its misuse.

#### 5.7.5 Liar's dividend

Synthetic media also introduces the 'liar's dividend,'<sup>65</sup> a term by Danielle Citron and Bobby Chesney that refers to the strategy where malicious actors weaponise citizens' skepticism around deepfakes and synthetic media to discredit genuine evidence, thereby protecting their own credibility. As Chesney and Citron write, "*Put simply: a sceptical public will be primed to doubt the authenticity of real audio and video evidence.*"<sup>66</sup> The liar's dividend has already surfaced strongly in the courtroom. For instance, two defendants on trial for the January 6 attack on the U.S capital attempted to argue for the unreliability of a video showing them at the Capitol on the basis that it could have been AI-generated.<sup>67</sup> The combination of deepfakes and the 'liar's dividend' can inject uncertainty and doubt into public opinion, again muddying citizens' ability to make informed decisions about what to believe.

The "*liar's dividend*" can be used in various contexts, including democracies. For instance, if Donald Trump said that the notorious 'access Hollywood' tape was a deepfake, it is plausible that a significant percentage of the population would believe it in the age of synthetic media. Politicians and public figures can manipulate information to protect their interests and rehabilitate their image, creating an uncertain environment that makes truth discernment difficult for citizens and may engender downstream effects such as eroding trust in institutions and the media. The mere knowledge that such sophisticated distortion and visual rewriting of reality is possible may erode trust in all media, permitting actual disinformation to hide behind the doubt cast on legitimate information. In this environment, discerning truth becomes increasingly challenging, as synthetic media is weaponised distort the information environment undermine reality at scale.

Ultimately, the features of synthetic disinformation are nuanced and multi-pronged. On one hand, believing fake content can undermine the epistemic quality of democratic discourse and citizens' decision-making, damaging reputations and disrupting democratic processes. On the other, dismissing authentic information as synthetic helps wrongdoers evade accountability, fostering cynicism, undermining trust and eroding a collective sense of reality. As computer science professor Hany Farid writes in an article for *The New York Times*, "*The specter of deepfakes is much, much more significant now — it doesn't take tens of thousands, it just takes a few, and then you poison the well and everything becomes suspect.*"<sup>68</sup>

64. [Less than you think: Prevalence and predictors of fake news dissemination on Facebook](#)

65. [The Liar's Dividend: The Impact of Deepfakes and Fake News on Politician Support and Trust in Media](#)

66. [Deep Fakes](#)

67. [The Defense In The First Jan. 6 Trial's Closing Argument: Maybe The Evidence Is Fake](#)

68. [A.I. Muddies Israel-Hamas War in Unexpected Way](#)

## 6. Mitigation Strategies for Synthetic Disinformation



In the past decade, responses to synthetic disinformation have been multi-layered and have been marked by a complex interplay of national and transnational initiatives, reflecting the global nature of disinformation and the digital media environment. These efforts involve a diverse range of actors, including governments, tech companies and civil society, and have encompassed various strategies such as legislative actions, self-regulatory measures and multi-pronged approaches. The challenges in regulating disinformation are equally multifaceted, stemming from the transnational character of digital platforms, the private governance settings of social media companies and the difficulty in establishing shared principles for governance.

### 6.1 TRANSNATIONAL RESPONSES

More recently, responses to synthetic disinformation are being embedded in larger processes aimed at setting standards for the regulation of AI, spanning international processes, national regulations and multi-stakeholder initiatives. They vary in their scope and territorial reach, level of detail and enforceability, with some being potentially legally binding, such as the forthcoming EU AI Act, and others being voluntary such as the G7 Hiroshima Artificial Intelligence Process Guiding Principles. Collectively, they underscore the need to strike a balance between tackling risks, including but not limited to synthetic disinformation, while preserving democratic values and human rights.

#### 6.1.1 G7 Hiroshima AI Process

The G7 Hiroshima Artificial Intelligence Process and the European Union's forthcoming AI Act represent two emerging initiatives that embed disinformation in a broader framework of AI regulation and/or standard setting. The G7 Hiroshima Artificial Intelligence Process, initiated in May 2023, represents a significant stride in the global governance of AI. In October 2023, G7 leaders adopted '*International Guiding Principles for Organizations Developing Advanced AI Systems*' and a corresponding voluntary '*Code of Conduct for AI Developers*'. Both documents, a product of collaborative efforts among G7 nations and the EU, aim to guide multi-stakeholder organisations in the responsible development, deployment and use of advanced AI systems. They emphasise a risk-based approach, ethical and legal considerations, robust governance and risk management policies, and a commitment to transparency and security throughout the AI lifecycle. These frameworks are designed to be dynamic, accommodating updates and revisions to stay relevant as AI technology evolves.

In the context of synthetic disinformation, several facets are pertinent. For instance, the Code mentions ‘red teaming’, which involves deliberately challenging and testing AI systems to identify vulnerabilities and potential misuse scenarios. This proactive approach is crucial for preventing and curbing disinformation before AI systems are released into the public domain. Another key principle is content authentication, which plays a crucial role in disclosing and making transparent when content is AI-generated. This principle can be likened to a ‘nutrition label’ for information, enhancing users’ ability to exercise critical thinking, discernment and ability to spot and counter disinformation ‘in the wild.’

### 6.1.2 EU AI Act

The forthcoming EU AI Act, initiated as part of the EU’s digital strategy,<sup>69</sup> introduces a widely encompassing and prospectively legally binding regulatory framework for artificial intelligence (AI), classifying AI systems based on their risk level category—from unacceptable to limited risk—each subject to varying degrees of regulation. Key forthcoming aspects include the prohibition of high-risk AI systems that threaten individual rights, stringent requirements for generative AI like *ChatGPT* and specific guidelines for AI systems that pose limited risk, such as those creating or manipulating multimedia content. This legislation aims to harness the benefits of AI across various sectors while safeguarding fundamental rights. With respect to content authentication, the latest draft of the forthcoming EU AI Act (October 2023) specifies that users, are responsible for disclosing the AI-generated origin of digital the content.

## 6.2 NATIONAL LEVEL RESPONSES

At the national level, governments are increasingly recognising the need to address the challenges posed by synthetic media and disinformation. For instance, countries like the United States, China and several countries in the EU, as well as Global South, have implemented frameworks to mitigate the risks associated with these technologies. For instance, in the US, the regulation of artificial intelligence, specifically deepfakes, is being addressed both at the federal and state levels. Several states have enacted laws targeting the malicious use of deepfake technology, such as Texas and California in 2019 prohibiting deepfakes designed to influence elections.<sup>70</sup> Complementing state efforts, on October 30, 2023, President Biden issued an Executive Order (EO) on *Safe, Secure, and Trustworthy Artificial Intelligence*. This order charges various agencies, including the National Institute of Standards and Technology (NIST), with developing guidelines to ensure the safe and trustworthy development and use of AI. These guidelines are directed towards protecting Americans’ privacy, advancing equity and civil rights, and upholding consumer and worker interests. The Department of Commerce is specifically required to develop guidance for labelling AI-generated content, which companies will use to establish labelling and disclosure protocols.

Already, in recent years, states around the world,<sup>71</sup> particularly in the EU and Global South, have implemented national regulations aimed at curbing disinformation and ‘fake news’ (however, efforts to tackle ‘synthetic’ disinformation are more nascent). South Africa, in response to the COVID-19 pandemic, passed regulations criminalising the spread of false news related to the virus, with penalties including imprisonment and fines.<sup>72</sup> Kenya’s government has used its cybercrimes law to prosecute individuals for spreading false information about COVID-19.<sup>73</sup> Tunisia’s Decree 54, enacted in September 2022, criminalizes sharing “false information” online; it has been criticised by human rights organisations for suppressing free speech by invoking vague definitions and particularly targeting President Saied’s critics.<sup>74</sup>

Several states in Southeast Asia have also unveiled similar legislation. Ahead of the July 2018 elections, Cambodia passed a law allowing the government to block media considered a threat to national security and to penalise individuals for publishing ‘fake news’ with imprisonment and fines.<sup>75</sup> Ultimately, while many laws purportedly address the pressing issue of disinformation, they exhibit a broad range of approaches, degrees of stringency and many have attracted scrutiny and criticism for potentially impeding free speech, underscoring the delicate balance between combating disinformation and preserving free expression.

69. [The European Union’s Artificial Intelligence Act - explained](#)

70. [The High Stakes of Deepfakes: The Growing Necessity of Federal Legislation to Regulate This Rapidly Evolving Technology](#)

71. [Factbox: ‘Fake News’ laws around the world](#)

72. [South Africa makes it illegal to spread false information about the coronavirus](#)

73. [Tracking disinformation laws and policies in more than 30 countries in Sub-Saharan Africa](#)

74. [Tunisia anti-fake news law criminalises free speech: Legal group](#)

75. [Cambodia ‘fake news’ crackdown prompts fears over press freedom](#)

### 6.3 MULTI-STAKEHOLDER & INDUSTRY-LED RESPONSES

Multi-stakeholder collaborations are particularly essential in the fight against synthetic disinformation, with increasing formalisation of standards for best practices like content disclosure and authentication. Key players in these efforts include organisations such as the Partnership on AI, which unites industry, academia and civil society to create nuanced frameworks and best practices. Another significant initiative is the Coalition for Content Provenance and Authenticity (C2PA), comprising companies like *Adobe*, *Arm*, *Intel*, *Microsoft* and *Truepic*. C2PA focuses on combating misinformation and deepfakes by developing and championing ‘*Content Credentials*’, namely a technical standard for verifying and preserving the origins, circulation and trajectory of digital media. Leveraging advancements in cryptography, C2PA’s crucial innovation lies in a specification that embeds reliable data into images, thereby hindering tampering. This method of fingerprinting digital content has gained traction, notably in photojournalism, with prototypes being used in Ukraine document war effects.<sup>76</sup>

Incorporating C2PA Content Credentials into national and international legislative and regulatory responses could substantially accelerate their adoption and effectiveness. Imagine a scenario where a national law mandates the use of Content Credentials in digital news media. This legislative support would provide a formal structure and backing to these technical standards, ensuring broader compliance and more effective enforcement. Such integration facilitates the creation of laws and policies well-informed by technical expertise and extensive input across the ecosystem. Additionally, technical watermarking approaches like *SynthID* from Google *DeepMind* and MIT researchers’ *PhotoGuard* offer innovative methods prompting transparency in digital content. Similarly, *PhotoGuard* could be utilised by media outlets to ‘immunize’ their images against synthetic manipulation, adding small, undetectable changes that resist tampering. Such tools, if integrated into legislative requirements, could greatly assist in the identification, prevention and mitigation of disinformation spread.

The importance of media and AI literacy and educational initiatives is also rising. Projects such as the MIT Center for Advanced Virtuality’s ‘*Media Literacy in the age of Deepfakes*’ learning module aims to equip individuals with the skills needed to discern reliable from synthetically manipulated information online.<sup>77</sup> While schools and universities are increasingly incorporating digital literacy into their curricula, there is a critical need for them to adapt to integrate discussions on synthetic, in addition to traditional, disinformation and teach students how to spot the risks and markers of synthetic media and deepfakes. Ultimately, a robust response to synthetic disinformation will require coordinated action across nations, industries and platforms. Only through such a comprehensive approach can we mitigate the risks while preserving free expression and human rights.

### 6.4 DELICATE BALANCE BETWEEN COMBATTING DISINFORMATION AND SAFEGUARDING FREE EXPRESSION

Managing the twin imperatives of countering synthetic disinformation and preserving freedom of expression is a nuanced task. Regulations must be designed to curb false information without stifling free expression, requiring well-calibrated efforts and careful attention to legal and ethical nuances.

Since disinformation often crosses national boundaries, solutions cannot be confined to single jurisdictions. International cooperation is essential for creating effective, harmonised approaches. The challenge of addressing disinformation isn’t just governmental; it’s a multi-stakeholder issue. Governments, tech companies, civil society and the public each have a role to play. Collaboration among these varied entities is crucial to developing strategies that both mitigate disinformation and uphold democratic values. Legislators are uniquely positioned to enact measures that are sensitive to the intricate interplay between technological capabilities and the foundational principles of democracy. At the same time, they must tread with precision, recognising that the fight against synthetic disinformation also necessitates an upholding of the right to free expression.

<sup>76</sup> [Reuters and Canon demonstrate end-to-end content authenticity system in the field](#)

<sup>77</sup> [Media Literacy in the age of deepfakes](#)



For instance, creating legal frameworks that promote the adoption of technical standards for media authenticity, could be a priority. At the same time, legislation must be carefully crafted to empower the use of technical standards while also upholding the right to free expression. If a watermarked or labelled piece of content makes an overt judgement about the truth or veracity of content, it may undermine free speech and become counterproductive and potentially a slippery slope. As such, technical standards like C2PA Content Credentials position themselves as akin to providing a 'nutrition label' that is intended to provide citizens with context to support informed decision-making, rather than compelling a specific viewpoint.

Against the backdrop of an emerging need for an 'AI nutrition label,' education can play a pivotal role in this ecosystem. A well-informed public, skilled in discerning the credibility of information, strengthens the foundations of democracy. Legislative support for educational programs can amplify their reach and impact, providing citizens with the tools to critically evaluate content and engage thoughtfully in the high-velocity and increasingly AI-saturated digital media environment.

## 7. Role of Legislatures in Combating Synthetic Disinformation



In addressing the challenges of synthetic disinformation and its impact on democracy, Legislatures play a pivotal role. Their involvement is key to navigating the delicate balance between fostering technological innovation and safeguarding democratic values. Here, the focus shifts towards broader, strategic approaches rather than specific legislative prescriptions.

**Promoting public and constituency awareness:** Legislatures can prioritise raising public awareness about synthetic media and its implications. This involves supporting initiatives that educate citizens on identifying and understanding the nature of synthetic content. By enhancing media literacy, legislators can empower the public to critically assess and respond to synthetic disinformation.

**Fostering international collaboration:** Given the global nature of the Internet and digital media, international cooperation is crucial. Legislatures can champion efforts to collaborate with counterparts across borders, sharing best practices and working towards unified standards. This approach acknowledges the transnational challenge of synthetic disinformation and the need for a harmonised response.

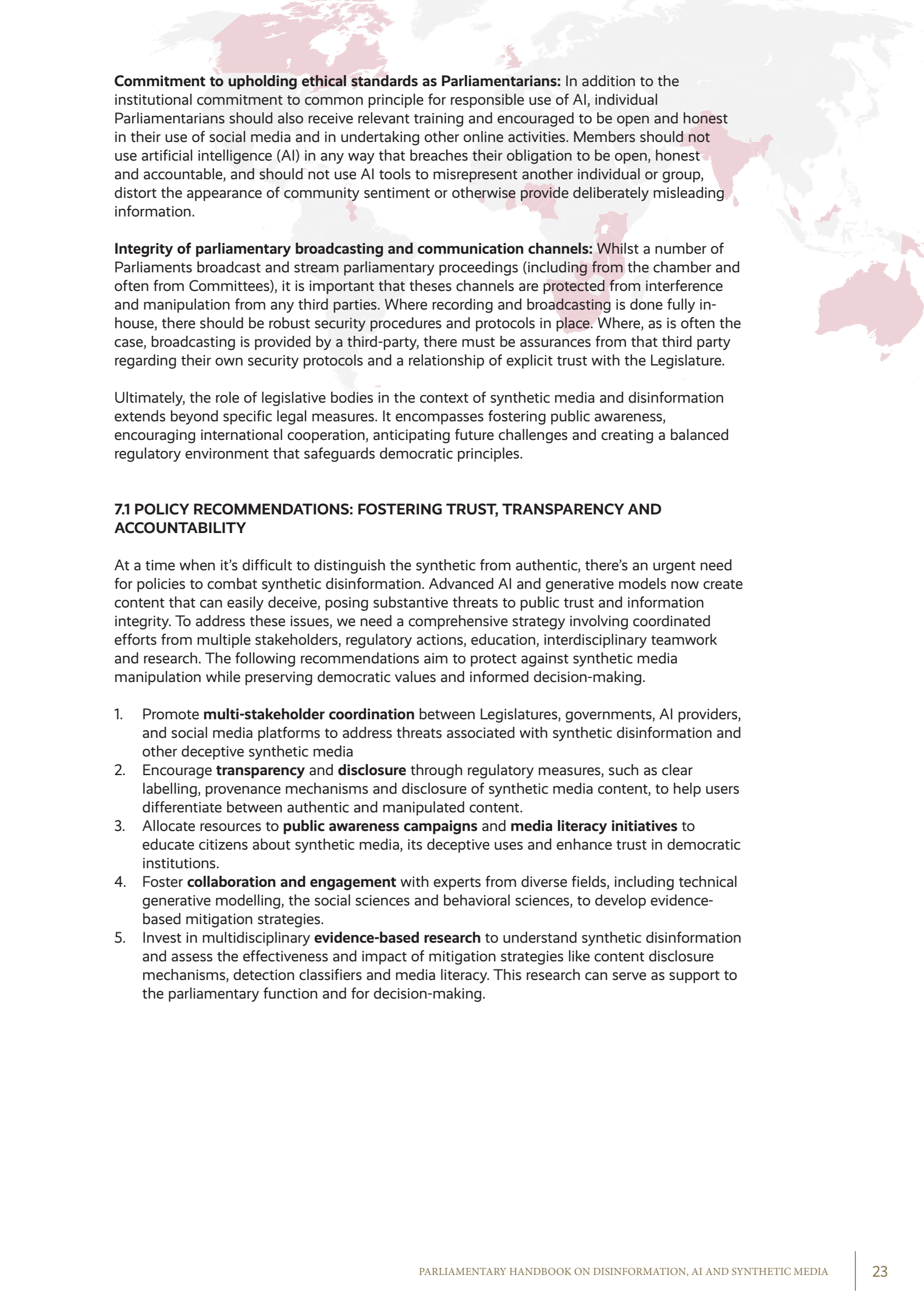
**Technical standards and innovation:** While specific regulations around synthetic media are complex, Legislatures can support the development and adoption of technical standards. This can be achieved through collaboration with organisations dedicated to establishing these standards, encouraging a responsible approach to technology development while also promoting innovation.

**Upholding human rights:** Protecting individual rights in the context of synthetic media remains a priority. Legislatures can advocate for principles that protect personal identity and privacy, emphasising consent and ethical considerations in the use of synthetic media.

**Striking a delicate balance:** Ultimately, the role of legislative bodies is to find a balance between various competing interests: innovation, privacy, freedom of expression and the prevention of disinformation. Their involvement is crucial in shaping a landscape where synthetic media is used ethically and responsibly, without undermining democratic values and creative freedoms.

**Adherence to AI principles for responsible use:** Legislatures should ensure that the use and regulation of synthetic media align with the principles of responsible AI use, such as those presented by OAS. The principles advocate for the enhancement of human development, the prevention of biases and the promotion of AI for accessibility and inclusion, and emphasise the need for AI to conform to human rights, promote innovation in a secure and transparent legal environment, and ensure governance and auditing.





**Commitment to upholding ethical standards as Parliamentarians:** In addition to the institutional commitment to common principle for responsible use of AI, individual Parliamentarians should also receive relevant training and encouraged to be open and honest in their use of social media and in undertaking other online activities. Members should not use artificial intelligence (AI) in any way that breaches their obligation to be open, honest and accountable, and should not use AI tools to misrepresent another individual or group, distort the appearance of community sentiment or otherwise provide deliberately misleading information.

**Integrity of parliamentary broadcasting and communication channels:** Whilst a number of Parliaments broadcast and stream parliamentary proceedings (including from the chamber and often from Committees), it is important that these channels are protected from interference and manipulation from any third parties. Where recording and broadcasting is done fully in-house, there should be robust security procedures and protocols in place. Where, as is often the case, broadcasting is provided by a third-party, there must be assurances from that third party regarding their own security protocols and a relationship of explicit trust with the Legislature.

Ultimately, the role of legislative bodies in the context of synthetic media and disinformation extends beyond specific legal measures. It encompasses fostering public awareness, encouraging international cooperation, anticipating future challenges and creating a balanced regulatory environment that safeguards democratic principles.

## **7.1 POLICY RECOMMENDATIONS: FOSTERING TRUST, TRANSPARENCY AND ACCOUNTABILITY**

At a time when it's difficult to distinguish the synthetic from authentic, there's an urgent need for policies to combat synthetic disinformation. Advanced AI and generative models now create content that can easily deceive, posing substantive threats to public trust and information integrity. To address these issues, we need a comprehensive strategy involving coordinated efforts from multiple stakeholders, regulatory actions, education, interdisciplinary teamwork and research. The following recommendations aim to protect against synthetic media manipulation while preserving democratic values and informed decision-making.

1. Promote **multi-stakeholder coordination** between Legislatures, governments, AI providers, and social media platforms to address threats associated with synthetic disinformation and other deceptive synthetic media
2. Encourage **transparency** and **disclosure** through regulatory measures, such as clear labelling, provenance mechanisms and disclosure of synthetic media content, to help users differentiate between authentic and manipulated content.
3. Allocate resources to **public awareness campaigns** and **media literacy initiatives** to educate citizens about synthetic media, its deceptive uses and enhance trust in democratic institutions.
4. Foster **collaboration and engagement** with experts from diverse fields, including technical generative modelling, the social sciences and behavioral sciences, to develop evidence-based mitigation strategies.
5. Invest in multidisciplinary **evidence-based research** to understand synthetic disinformation and assess the effectiveness and impact of mitigation strategies like content disclosure mechanisms, detection classifiers and media literacy. This research can serve as support to the parliamentary function and for decision-making.

## 8. Conclusion

---



In today's rapidly transforming information ecosystem, the rise of synthetic media and disinformation underlines the urgent need to protect democratic integrity and free expression. Balancing the fight against misinformation with freedom of expression and other human rights requires a nuanced approach that combines technology, legislation and collaborative efforts from diverse stakeholders.

Legislatures are crucial in setting regulations, ensuring accountability and promoting innovation. Nevertheless, effectively combating disinformation without encroaching on free expression necessitates cooperation among Legislatures, governments, tech companies, civil society and the public. A comprehensive strategy—incorporating technical standards, public awareness campaigns, interdisciplinary input and peer-reviewed research—serves as a guide to address these challenges while preserving fundamental democratic values.

# Glossary of Terminology

---

**Artificial Intelligence (AI)** - the ability of a digital computer or other capable machine to perform tasks and functions commonly associated with human intelligence.

**Bots** - software programmes undertaking pre-defined, automated tasks, often imitating human behaviours.

**Botnets** - a network of infected/hijacked computers, under the control of a single party, with the purpose of undertaking cyberattacks.

**Computational Propaganda** - the combined use of automation, human curation and algorithms to assemble and share misleading information online.

**Conspiracy Theory** – explanations or narratives that suggest large-scale events are orchestrated by secretive, powerful groups/organisations.

**Deepfakes** – media which has been manipulated to represent someone as doing or saying something that they did not in fact do or say.

**Deep Learning** - a type of artificial intelligence that looks to imitate human learning processes.

**Disinformation** – the deliberate creation and sharing of false information with the intention to deceive.

**Echo Chambers** - environments in which an individual is exposed only to opinions and or information reflective of their own beliefs, often with the effect of reinforcing those beliefs.

**Generative AI** - artificial intelligence technologies capable of producing various types of content, including text, images, audio and other media.

**Liar's Dividend** – a strategy where malicious actors weaponise scepticism around deepfakes and synthetic media to deny genuine evidence.

**Machine Learning** – a subset of Artificial Intelligence, where algorithms learn from data to perform tasks traditionally associated with human intelligence.

**Malinformation** – the sharing of true information with harmful intent.

**Microtargeting** - online targeted advertising utilising personal data to identify the interests and/or vulnerabilities of specific individuals or audiences.

**Misinformation** – the spread of information which may be false but does not have the intention of misleading.

**Provenance** – the verification of the origin and authenticity of media content through, for example commonly accepted technical standards.

**Synthetic Media** – media which has been either fully or partially generated using artificial intelligence.

**Troll Farm/Troll Army** - an organised group employing individuals to cause conflict or manipulate public opinion online through deliberately offensive or provocative content.



A series of horizontal lines for writing notes, consisting of 25 evenly spaced lines.





## About the CPA

The Commonwealth Parliamentary Association (CPA) connects, develops, promotes and supports Parliamentarians and their staff to identify benchmarks of good governance and the implementation of the enduring values of the Commonwealth. The CPA collaborates with Parliaments and other organisations, including the intergovernmental community, to achieve its statement of purpose. It brings Parliamentarians and parliamentary staff together to exchange ideas among themselves and with experts in various fields, to identify benchmarks of good practices and new policy options they can adopt or adapt in the governance of their societies.

## About the OAS

The Organization of American States is the premier regional forum for political discussion, policy analysis and decision-making in Western Hemisphere affairs. The OAS brings together leaders from nations across the Americas to address hemispheric issues and opportunities. Together, they seek to build cooperation among states and advance a common regional agenda on democratic governance, human rights, multidimensional security and sustainable development.



Published by the Commonwealth Parliamentary Association (CPA).  
December 2023.  
Registered Charity Number 263147.

CPA Headquarters Secretariat  
Richmond House, Houses of Parliament  
London SW1A 0AA  
United Kingdom

Telephone: +44 (0)20 7799 1460  
Email: [hq.sec@cpahq.org](mailto:hq.sec@cpahq.org)  
Website: [www.cpahq.org](http://www.cpahq.org)