



THE HIGH LEVEL PANEL OF LEGAL EXPERTS ON MEDIA FREEDOM

The independent advisory body to the Media Freedom Coalition



False Speech in International Law

Authors: Professor Marko Milanovic and Professor Philippa Webb

Excerpt from *Freedom of Speech in International Law* (edited by Lord David Neuberger and Ms Amal Clooney, published by Oxford University Press).



This IBAHRI publication is an extracted chapter of Freedom of Speech in International Law, edited by Lord David Neuberger of Abbotsbury and Ms Amal Clooney and published by Oxford University Press in January 2024. The authors of this chapter are Professor Marko Milanovic and Professor Philippa Webb. Ms Alice Gardoll is the Assistant Editor of the text. The full text can be purchased in hard copy or as an e-book [here](#). All sources and references have been checked for accuracy as of 1 January 2022.

Freedom of Speech in International Law outlines the minimum protections for speech enshrined in international law, focusing on four types of laws that are being weaponized to silence the press and independent voices: laws regulating defamatory or insulting speech, laws regulating false speech, laws regulating hate speech and laws regulating national security. The book provides examples of where states are falling short and makes recommendations about how international standards should be interpreted, updated and enforced.

Recommendations are based on international legal standards that apply to states and that many social media companies have expressed adherence to. The recommendations have been endorsed by the High Level Panel of Legal Experts on Media Freedom, as well as judges and experts from across the world including the Committee to Protect Journalists, Reporters without Borders, the UN Special Rapporteur on Freedom of Expression and the International Bar Association's Human Rights Institute.

The International Bar Association's Human Rights Institute, serves as Secretariat to the High Level Panel of Legal Experts on Media Freedom, providing it with operational, technical, and legal assistance. The work of the High Level Panel, including the research contained in this report, has been supported by the Global Media Defence Fund, administered by UNESCO. At the time of publication, the Chair of the High Level Panel was Lord Neuberger of Abbotsbury and its Deputy Chairs were Professor Can Yeginsu and Ms. Catherine Amirfar.

Cover image: www.stock.adobe.com

www.ibanet.org/IBAHRI

Copyright © 2025 International Bar Association

4

False Speech

Marko Milanovic and Philippa Webb

I. Introduction	5	2. Legitimacy	24
II. State Practice	7	2.1. States must not restrict false speech for ulterior purposes	25
1. Overview of Laws Regulating False Speech Around the World	8	2.2. What objectives can justify the penalization of false speech?	26
1.1. Vagueness and overbreadth	8	3. Necessity	28
1.2. Harm	9	3.1. Harm	29
1.3. Intent	13	3.2. Intent	33
1.4. Exclusions, exceptions and defences	15	4. Exclusions, Exceptions and Defences	37
1.5. Penalties	15	4.1. Truth	37
2. Application of False Speech Laws Around the World	18	4.2. Opinion	38
2.1. Europe	18	4.3. Public interest and 'responsible journalism'	39
2.2. Asia Pacific	19	5. Penalties	41
2.3. Middle East and Africa	20	5.1. Criminal	41
III. International Legal Standards	21	5.2. Civil	43
1. Legality	22	IV. Recommendations	44

I. Introduction

In the early 2020s the world witnessed not only a global pandemic but also the grave harm that can result from misinformation about it.¹ As the coronavirus spread across the world, so did false information about the virus, causing what the UN has deemed an 'infodemic'.² False information about vaccines also led to 'an urgent threat to public health' due to vaccine hesitancy.³ Following Donald Trump's assertion, while President of the United States, that injecting disinfectant could cure Covid-19, more than 16,000 cases of chlorine dioxide poisoning occurred in the United States alone.⁴ And in Iran,

¹ 'Countries urged to act against COVID-19 "infodemic"' (UN News, 23 September 2020).

² Ibid.

³ US Department of Health & Human Services, 'US Surgeon General Issues Advisory During COVID-19 Vaccination Push Warning American Public About Threat of Health Misinformation' (15 July 2021). A study analysing almost one million Covid-19 anti-vaccine misinformation posts on social media found that the majority were attributable to just 12 individuals, known as the 'disinformation dozen', with a combined following of nearly 60 million people: see E. Salam, 'Majority of Covid misinformation came from 12 people, report finds' (The Guardian, 17 July 2021).

⁴ E. Pilkington, 'Bleach touted as "miracle cure" for Covid being sold on Amazon' (The Guardian, 19 September 2020).

over 700 people died from alcohol poisoning following false claims that ingesting methanol could cure coronavirus.⁵ This illustrates how truthful information, such as health warnings issued by official bodies, can be disregarded and misinformation spread globally at the cost of people's lives—a phenomenon that has been called 'truth decay'.⁶

Although prohibitions on false speech have existed for at least 140 years, many laws criminalizing it have been introduced much more recently.⁷ The Covid-19 pandemic has accelerated the introduction and use of such laws, including 'as a pretext to restrict information and stifle criticism'.⁸ This began at the pandemic's very inception: Dr Li Wenliang, the Wuhan doctor who issued an early warning to his colleagues about the new virus on a messaging platform, was detained by police for 'spreading false rumours' and forced to sign a confession that he had 'severely disturbed the social order'.⁹ He later died of the virus.¹⁰ By the end of 2020, hundreds of journalists had been arrested or detained over pandemic-related reporting.¹¹

In parallel, laws have been introduced or are contemplated in response to the rise of social media and its use in spreading harmful speech of all kinds, including false information on elections.¹²

The theoretical justification for prohibiting false speech is contested. In classical liberal thought, more truthful speech has always been considered to be the best remedy for falsehoods, with truth to be established through a process of debate and contestation in what has been termed the 'marketplace of ideas'.¹³ And there are fears of a slippery slope, that opening the doors to the regulation of speech on the basis of truthfulness provides a pretext for authoritarians clamping down on critics to argue that they are simply trying to prevent social harms caused by falsehoods, just as democracies do.

But fear of a slippery slope need not mean the state's hands are tied when it is faced with substantial harms caused by falsehoods. Lies should not, alone, be the basis for legal sanction.¹⁴ But there are circumstances under which false speech can, and indeed

⁵ 'Iran: Over 700 dead after drinking alcohol to cure coronavirus' (Al Jazeera, 27 April 2020).

⁶ 'Barack Obama: One election won't stop US "truth decay"' (BBC, 15 November 2020).

⁷ See, e.g., French Law on the Freedom of the Press of 29 July 1881; cf. Turkish Capital Market Law No. 6362 2012 Art. 107(2); United Arab Emirates Federal Decree Law No. 34 of 2021 on Combatting Rumors and Cybercrimes Arts 25, 52.

⁸ 'No time to blame the messenger' warns UN rights chief, amidst media clampdowns surrounding COVID-19' (UN News, 24 April 2020); International Press Institute, 'Rush to pass "fake news" laws during Covid-19 intensifying global media freedom challenges' (3 October 2020).

⁹ 'Hero who told the truth': Chinese rage over coronavirus death of whistleblower doctor' (The Guardian, 7 February 2020).

¹⁰ Ibid.

¹¹ See A. Berry & M. Fürstenau, 'Press freedom: Journalists jailed for COVID reporting' (DW, 14 December 2020).

¹² Access Now, Civil Liberties Union for Europe and European Digital Rights (EDRi), *Informing the Disinformation Debate* (2018), 18.

¹³ See US Supreme Court, *Abrams v. United States* 250 U.S. 616, 10 November 1919, 630 (Holmes J): 'the best test of truth is the power of the thought to get itself accepted in the competition of the market'. The marketplace of ideas has become a constant theme in US First Amendment jurisprudence. See, e.g., US Supreme Court, *United States v. Alvarez* 567 U.S. 709, 28 June 2012, where the metaphor is invoked expressly or impliedly by all three opinions in the case (Kennedy J (opinion) at 2545, quoting US Supreme Court, *Hustler Magazine Inc. v. Falwell*, 485 U.S. 46, 24 February 1988, 52; Breyer J (concurring opinion), at 2552; Alito J (dissenting) at 2560 also quoting US Supreme Court, *Hustler Magazine Inc. v. Falwell*, 485 U.S. 46, 24 February 1988).

¹⁴ See s. III.2. (Legitimacy).

should, be penalized.¹⁵ Historically, democratic societies have been comfortable with the state deciding matters of truth in some settings, such as defamation or fraud (where the truth is established by a court) or advertising standards and consumer protection (where the truth may be established by an executive agency or an independent regulator).¹⁶ Punishing false speech that intentionally causes harm to public health or to the integrity of democratic elections is also largely uncontroversial.¹⁷ But criminalization should be reserved for serious harm that a prosecutor can prove was directly and intentionally caused by speech.¹⁸ And the politicization of certain factual issues (like climate change or accounts of historical events), even ostensibly technical ones (such as the utility of wearing masks during the Covid-19 pandemic) raises complicated questions about when the state can be an arbiter of truth and where lines should be drawn.

Under international law, state regulation of false speech (misinformation)¹⁹ must comply with individuals' freedom of expression, and this chapter explores where these lines are drawn in international treaties and human rights jurisprudence.²⁰ Because this is a developing area of law, much of the jurisprudence of international courts and human rights bodies is drawn from other contexts, in particular standards relating to false speech that is defamatory, causing reputational harm. The chapter outlines international standards relating to when false speech can be penalized and makes recommendations as to how the law should be applied and reformed to meet the challenges of 'truth decay'.²¹

¹⁵ M. Milanovic, 'Viral Misinformation and the Freedom of Expression: Part I' (EJIL: Talk!, 13 April 2020).

¹⁶ *Ibid.*

¹⁷ See s. III.3.1. (Harm).

¹⁸ See s. III.5. (Penalties). See, e.g., HRC, General Comment No. 34 (2011), §47 noting (in the context of defamation) that the application of criminal law should only be countenanced in the most serious of cases and imprisonment is never an appropriate penalty.

¹⁹ This chapter uses the term misinformation in a broad sense, denoting information on any topic that is objectively factually untrue, regardless of the speaker's mental state (unless otherwise specified). We will also use false speech as a synonym. Scholars have often distinguished between misinformation and disinformation, with the former connoting the innocent dissemination of false information and the latter implying an intention to deceive. Malinformation and 'fake news' are also widely used. These terminological distinctions are, in our view, not helpful in the criminal context, where there may be many gradations of culpability—in particular, one can be culpable without having intent, for example by recklessly or negligently failing to verify the truthfulness of a statement. The use of the term misinformation in this chapter is thus meant to be neutral on the question of the speaker's mental state, so that it can capture situations in which the speaker is acting entirely innocently, but also those in which they are acting unintentionally but with a lesser form of fault. However, to ensure clarity, when used in this chapter the term disinformation refers specifically to knowingly false speech. For an overview of the terminological issue, see K. Jones, *Online Disinformation and Political Discourse: Applying a Human Rights Framework* (2019) Chatham House Research Paper, Royal Institute of International Affairs, 7.

²⁰ States also have a positive duty to protect and fulfil the right to free speech, which may include a duty to actively promote access to accurate information. See UN Special Rapporteur D. Kaye, *Disease pandemics and the freedom of opinion and expression* (2020) UN Doc. A/HRC/44/49, §§7–8, 11–12, 18.

²¹ Our focus in this chapter is on *criminal* laws used by states to penalize false speech rather than the (equally important) issue of how false speech can be penalized under regulatory regimes such as those that impose accuracy standards on journalists or 'community standards' for social media users.

II. State Practice

False speech laws²² are not new: states such as China and Iran have had laws prohibiting ‘information ... providers’ from disseminating ‘rumours’ on their books for decades.²³ But in the last 10 years a raft of new misinformation laws have been enacted globally, especially related to speech on the internet.²⁴

With the onset of a pandemic in 2020, at least 18 states passed some form of legislation against ‘online misinformation’ or ‘fake information.’²⁵ For example, the UK²⁶ Government enacted new laws which criminalized individuals who knowingly provide false or misleading information to police and certain medical professionals dealing with the coronavirus pandemic.²⁷ Thailand enacted a new emergency decree prohibiting the ‘reporting or spreading of information regarding Covid-19 which is untrue and may cause public fear’, and ‘distort[ing] information which causes misunderstanding and hence affects peace and order or good morals of people.’²⁸ And in Hungary, criminal penalties apply to individuals who are ‘spread[ing] a distorted truth in relation to the emergency’ in a way that could be ‘alarming or agitating for a large group.’²⁹

Many laws regulating misinformation make it a criminal offence³⁰ and have extra-territorial reach.³¹ In addition, the practice of states using these laws has very frequently involved cases in which institutions of the state—the judiciary, the military or the government—have been insulted: an application of the laws that clearly violates international legal standards.³²

²² The review of state laws and practice on false speech contained in this chapter does not aim to be comprehensive but rather to identify significant examples and trends. Therefore, where the language of a statute is set out, this does not necessarily include an analysis of all case law interpreting that language in the jurisdiction under review.

²³ Chinese Provisions for the Administration of Internet News Information Service Art. 16. See also Iranian Computer Crimes Law 2010 Art. 18. See also Article 19, ‘Islamic Republic of Iran: Computer Crimes Law’ (2012) 37–40.

²⁴ See, e.g., ‘Censorious governments are abusing “fake news” laws’ (The Economist, 13 February 2021).

²⁵ International Press Institute, ‘IPI’s Press Freedom Tracker’; International Press Institute, ‘Rush to pass ‘fake news’ laws during COVID-19 intensifying global media freedom challenges’ (3 October 2020).

²⁶ Unless otherwise indicated, references to the United Kingdom and the UK refer to England and Wales, Scotland and Northern Ireland. Although many of the same laws apply across these jurisdictions, these laws can vary and Scotland has a separate legal system.

²⁷ UK Coronavirus Act 2020, Sch. 21, Parts 2–5, §§23(1)(d), 45(1)(d), 67(1)(d), 88(1)(d). These offences were repealed with respect to England in December 2021: see UK Coronavirus Act 2020 (Early Expiry) (No. 2) Regulations 2021/1399, reg. 5(7). See also South African Disaster Management Act: Regulations relating to COVID-19, reg. 14(2) (addressing statements made with the intention to deceive another about COVID-19, a person’s COVID-19 ‘infection status’ or ‘any measure taken by the Government to address COVID-19’ (emphasis added)).

²⁸ Thai Regulation Issued under Section 9 of the Royal Decree on Public Administration in Emergency Situation 2005 (No. 27), s. 9(3). See also CPJ, ‘Thailand declares state of emergency, imposes press restrictions’ (26 March 2020).

²⁹ Hungarian Act C of 2012 on the Penal Code, s. 337(1). See CPJ, ‘Proposed Hungarian laws could imprison journalists covering coronavirus response’ (24 March 2020).

³⁰ All 45 countries that were analysed in this chapter have at least one law on their books that *criminalizes* false speech.

³¹ See, e.g., Iranian Computer Crimes Law 2010 Art. 28; Omani Penal Law, Promulgated by Royal Decree 7/2018 Arts 17–18; Rwandan Law Determining Offences and Penalties in General No. 68 2018 Arts 11–13. See s.II.1.1. (Vagueness and overbreadth).

³² Compare s. II.2. (Application of False Speech Laws Around the World) and s. III (International Legal Standards).

1. Overview of Laws Regulating False Speech Around the World

1.1. Vagueness and overbreadth

A trend among states that criminalize the spreading of misinformation is the use of vague wording. For instance, some states criminalize the spreading of ‘fake news’, without more, even though the term is not defined³³ or is defined in a circular way.³⁴ Many states also use vague and overbroad terminology to define the harm that false speech must cause, such as speech that may ‘weaken the country’s financial credibility, dignity and prestige’,³⁵ cause ‘needless anxiety’³⁶ or ‘is likely to be prejudicial to the friendly relations of [the state] with other countries.’³⁷

Qatar recently introduced a false speech law that is permeated by vagueness. Article 136(bis), introduced in the 2020 amendments to Qatar’s Penal Code, criminalizes the broadcasting, publishing or re-publishing of ‘rumours, statements, false or malicious news or propaganda, at home or abroad, with the intention of harming national interests, provoking public opinion or violating the social system or public order of the state.’³⁸ The penalties are imprisonment of up to five years and/or a fine of up to approximately \$27,500 US dollars.³⁹

It is unclear whether this Qatari law extends to expressions of opinion rather than only of fact; the harmful proscribed consequences are so open-ended (‘national interests’ or ‘violating the social system’) that they are entirely at the discretion of the interpreter; and there is no specification of any causal connection between the speech and the harm. With regard to *mens rea*, while the required harm needs to be intentional, there is no specification of the mental state of the speaker with regard to the statement itself, for example, whether the information has to be knowingly false.

Some domestic courts have construed misinformation laws narrowly to mitigate the impact of vague language. For example, the French Constitutional Court construed a French law prohibiting misinformation that cast doubt on the reliability of elections as requiring that misinformation be confined to information that is *manifestly inexact or misleading* and *manifestly* poses a risk to the integrity of the elections.⁴⁰ Similarly, the High Court of England and Wales refused to allow the common law offence of ‘misfeasance in public office’ to capture false statements of fact during a referendum campaign,

³³ See, e.g., Kazakhstani Criminal Code 2014 (amended 2018) Art. 274; Philippine Revised Penal Code Art. 154(1) (‘false news’).

³⁴ See, e.g., Singaporean Protection from Online Falsehoods and Manipulation Act 2019 s. 2(2)(b) (‘a statement is false if it is false or misleading, whether wholly or in part, and whether on its own or in the context in which it appears’).

³⁵ Egyptian Penal Code Law No. 58 1937 Art. 80D.

³⁶ Nigerian Cybercrimes (Prohibition, Prevention, Etc) Act 2015 s. 24(1)(b); Rwandan Law Governing Information and Communication Technologies No. 24 2016 Art. 60.

³⁷ Singaporean Protection from Online Falsehoods and Manipulation Act 2019 s. 7(1)(b)(iii). See also United Arab Emirates Federal Decree Law No. 34 of 2021 on Combatting Rumors and Cybercrimes Arts 25, 52.

³⁸ Qatari Penal Code (Amended 2020) Art. 136(bis).

³⁹ Ibid.

⁴⁰ French Constitutional Court, *Décision no. 2018-773 DC*, 20 December 2018, §§23, 26.

such as Boris Johnson's untruthful claim—painted on the side of his campaign bus—that the United Kingdom would be able to divert 350 million pounds weekly to its National Health Service upon leaving the European Union.⁴¹ The court reasoned that the statement was not made by a public officer 'acting as such,' as the offence required, since Mr. Johnson 'simply held the office and while holding it expressed a view' that was 'contentious and widely challenged,' rather than stating the view while acting in the discharge of the duties of the office.⁴²

In addition to being vague, a number of misinformation laws are universal, meaning that they apply to speech in any location.⁴³ Other laws apply to speech abroad if the victim of the offence is a citizen of the state,⁴⁴ if the offence harms the interests of the state,⁴⁵ if the speech has some connection to information-technology infrastructure in the state⁴⁶ or if one of the constituent elements of the crime was committed within the territory of the state that has passed the law.⁴⁷ Jurisdictions such as Egypt make it an offence to 'deliberate[ly] ... disseminat[e] false news abroad.'⁴⁸ For some laws, the broadened scope is applicable only if the offence is also recognized in the territory in which it was committed,⁴⁹ or if no prosecution has occurred abroad.⁵⁰

1.2. Harm

Misinformation laws vary considerably between states as to the nature of the harms caused by false speech that warrant penalization. Some laws make any false speech criminal without requiring any harm to be intended or caused (or even the speech to be public).⁵¹ For example, Nigerian law prohibits a person 'knowingly or intentionally send[ing] a message or other matter by means of computer systems or network that ... he knows to be false.'⁵²

⁴¹ UK High Court of England and Wales, Queen's Bench Division, *Johnson v. Westminster Magistrates' Court* [2019] EWHC 1709 (Admin), 3 July 2019.

⁴² *Ibid.*, §§29–35 (noting that there was no precedent for prosecuting as misconduct in public office misleading statements made 'for the purposes of political campaigning'). The High Court also noted that the 'problem of false statements in the course of political campaigning is not new and has not been overlooked by Parliament,' but that legislation sanctioning such conduct relates to false statements as to the personal character or conduct of a candidate during an election only, and 'in enacting the prohibition Parliament must have deliberately excluded any other form of false statement of fact, including those relating to publicly available statistics': §36.

⁴³ See, e.g., Iranian Computer Crimes Law 2010 Art. 28; Omani Penal Law, Promulgated by Royal Decree 7/2018 Arts 17–18; Rwandan Law Determining Offences and Penalties in General No. 68 2018 Arts 11–13 (providing that any offences committed 'against the interests of Rwanda' outside of the territory can be tried as if committed within that country).

⁴⁴ See, e.g., Nigerian Cybercrimes (Prohibition, Prevention, Etc) Act 2015 s. 50(1)(d)(i); Thai Computer Crime Act (2016 Amendment) s. 17(2).

⁴⁵ See, e.g., Chinese Criminal Law (amended through 2021) Art. 8; Kazakhstani Criminal Code 2014 Art. 8(4); Qatari Penal Code (Amended 2020) Art. 16(3).

⁴⁶ See, e.g., United Arab Emirates Federal Decree Law No. 34 of 2021 on Combatting Rumors and Cybercrimes Art. 69; Iranian Computer Crimes Law 2010 Art. 28.

⁴⁷ See, e.g., Qatari Penal Code (Amended 2020) Art. 13 (which applies if the consequence of the act is realized within the State).

⁴⁸ Egyptian Penal Code Law No. 58 1937 Art. 80D. See also Cambodian Criminal Code Art. 425; Nigerian Cybercrimes (Prohibition, Prevention, Etc) Act 2015 s. 50; Omani Penal Law, Promulgated by Royal Decree 7/2018 Art. 115.

⁴⁹ See, e.g., Thai Computer Crime Act (2016 Amendment) s. 17(1).

⁵⁰ Cambodian Criminal Code Arts 19, 23; Kazakhstani Criminal Code 2014 Art. 8(1).

⁵¹ See, e.g., Qatari Penal Code (Amended 2020) Art. 136(bis).

⁵² Nigerian Cybercrimes (Prohibition, Prevention, Etc) Act 2015 s. 24(1)(b).

Where misinformation laws do identify harms linked to the speech, these differ considerably. First, some laws prohibit the dissemination of false ‘information’ or ‘news’ in the interests of protecting the ‘public order’ or ‘public safety.’⁵³ In some states, this relates to ensuring the physical safety of persons and property.⁵⁴ Others extend beyond physical injury and protect against false statements that ‘disturb the public peace,’⁵⁵ ‘weaken the nation’s endurance,’⁵⁶ ‘disturb the social or public order,’⁵⁷ are likely to cause ‘fear or alarm to the public,’⁵⁸ or spread ‘horror among the people.’⁵⁹ Laws in Kazakhstan,⁶⁰ China,⁶¹ and the Philippines⁶² refer to harm to ‘public order’ or ‘public safety’ but leave these terms undefined. And following the Russian invasion of Ukraine in February 2022, the Russian Criminal Code was amended to make it a crime to state ‘false information’ about the activities of Russian armed forces and state organs abroad, punishable by up to 15 years’ imprisonment if this leads to ‘grave consequences.’⁶³

Second, misinformation laws may claim to protect a state’s interests in upholding ‘national security’ and ‘defence.’⁶⁴ These stated interests can extend beyond internal security to address a state’s international relationships and ‘foreign security,’⁶⁵ such as by creating a ‘hostile international environment against the Government.’⁶⁶

Third, some misinformation laws seek to protect the reputation of the state or its institutions, matters that are left unaddressed by defamation laws that protect individuals. Misinformation laws in this category are often broadly worded, referring to ‘endanger[ment of] ... national honor,’⁶⁷ or preventing ‘false or malicious’ news that ‘undermine[s] the stature of the state.’⁶⁸ States have also sought to protect against misinformation that ‘humiliates or contempts national institutions’⁶⁹ or that has ‘brought

⁵³ See, e.g., Taiwanese Social Order Maintenance Act (2021 amendment) Art. 63(1)(5); Mauritian Criminal Code Act 1838 Art. 299; Kazakhstani Code of Administrative Offences No. 235-V 2014 Art. 456-1; Maltese Criminal Code 2003 s. 82.

⁵⁴ See, e.g., Cambodian Criminal Code Art. 425; South Sudanese Penal Code 2008, s. 75(a)(i); Russian Penal Code Art. 207.1 (prohibiting the ‘dissemination of knowingly false information under the guise of truthful statement, concerning circumstances involving danger to life and safety of citizens’).

⁵⁵ Ghanaian Criminal Code Act 29 1960 s. 208(2). See also Yemeni Law Concerning Crimes and Penalties No. 12 1994 Art. 198(1).

⁵⁶ Egyptian Penal Code Law No. 58 1937 Art. 80C.

⁵⁷ Qatari Penal Code (Amended 2020) Art. 136(bis).

⁵⁸ Malaysian Penal Code s. 505(b).

⁵⁹ Egyptian Penal Code Law No. 58 1937 Art. 102bis. See also Egyptian Penal Code Law No. 58 1937 Art. 188; Bahraini Penal Code 1976 Arts 133, 168 (amended by Legislative Decree No. 9 1982); Moroccan Penal Code 2018 Art. 72.

⁶⁰ Kazakhstani Criminal Code 2014 Art. 274.

⁶¹ Chinese Criminal Law (amended through 2021) Art. 293.

⁶² Philippine Revised Penal Code Art. 154(1).

⁶³ Russian Penal Code Art. 207.3; ‘New Russian Law on “False Information” About Army being Used Against Activists’ (Radio Free Europe, 17 March 2022); Human Rights Watch, ‘Russia Criminalizes Independent War Reporting, Anti-War Protests’ (7 March 2022).

⁶⁴ Thai Computer Crime Act (2016 Amendment) s. 14. Zimbabwean Criminal Law (Codification and Reform) Act [Chapter 09:23] 2004 s. 31(a)(ii); South Sudanese Penal Code 2008 Art. 75(a)(ii).

⁶⁵ See, e.g., Qatari Cybercrime Prevention Law No. 14 2014 Art. 6.

⁶⁶ Rwandan Law Determining Offences and Penalties in General No. 68 2018 Art. 194.

⁶⁷ Chinese Cybersecurity Law 2017 Art. 12. See also Bahraini Penal Code 1976 Art. 174.

⁶⁸ Omani Penal Law, Promulgated by Royal Decree 7/2018 Art. 115(a). See also Egyptian Penal Code Law No. 58 1937 Art. 178bis third; Iraqi Penal Code Law No. 111 1969 §180. Zimbabwean Criminal Law (Codification and Reform) Act [Chapter 09:23] 2004 s. 31(a)(iii).

⁶⁹ Cambodian Law on the Press 1995 Art. 13.

discredit to public institutions or their functioning.⁷⁰ These laws supplement insult laws and *lèse-majesté* laws that penalize speech that is insulting to rulers without requiring the speech to be false.⁷¹

Fourth, the laws of certain states refer to a state's economic interests and financial credibility as being the harm that misinformation laws seek to address. Many states define these interests broadly,⁷² referring to 'public finances',⁷³ 'national economic security',⁷⁴ information that weakens the 'country's financial credibility'⁷⁵ or 'confidence in its financial markets or its economic and financial standing'.⁷⁶ And EU law covers the dissemination of information, including rumours, which give or are likely to give, 'false or misleading signals' as to 'the supply of, demand for, or price of, a financial instrument'.⁷⁷

Fifth, an increasing number of misinformation laws focus on harm to elections. According to the OECD, the proportion of elections targeted by cyber threats, including misinformation campaigns, has risen more than 75 per cent from 2015 to 2018, and the number of laws addressing election-related misinformation has grown along with it.⁷⁸ In 2018, following disinformation campaigns during a presidential election, France enacted a law allowing judges to order the removal of online articles that constitute misinformation and granting the French national broadcasting agency the power to suspend television channels controlled by other states if they 'deliberately disseminate false information likely to affect the sincerity of the ballot'.⁷⁹ Other laws are framed more broadly as prohibiting the transmission of false information that is likely to 'influence the outcome of an election',⁸⁰ or personal opinions that lead to 'confusion or loss of confidence' during an election.⁸¹

⁷⁰ Senegalese Penal Code 1965 Art. 255. See also South Sudanese Penal Code 2008 s. 75(a)(iii) ('undermining public confidence in a law enforcement agency, or the Defence Forces of Southern Sudan').

⁷¹ See, e.g., United Arab Emirates Federal Decree Law No. 34 of 2021 on Combatting Rumors and Cybercrimes Art. 25 (prohibiting the dissemination of rumours or information—regardless of the statement's veracity—intended to 'damage the reputation, prestige or stature of the State or any of its institutions or its president, vice-president, any of the rulers of the Emirates, their crown princes, or the deputy rulers of the Emirates, the State flag, the national peace, its logo, national anthem or any of its symbols').

⁷² See, e.g., Lebanese Penal Code 1943 Art. 297 (amended according to Act No. 239 of 27 May 1993) prohibiting the broadcast of 'false or exaggerated news' that 'would undermine the prestige of the state and its financial position'; Bahraini Penal Code 1976 Art. 169 ('cause damage to ... the State's creditworthiness').

⁷³ Nigerian Protection from Internet Falsehoods and Manipulation and other related matters Bill 2019 s. 3(1)(b)(ii).

⁷⁴ Thai Computer Crime Act (2016 Amendment) s. 14(2).

⁷⁵ Egyptian Penal Code No. 58 1937 Art. 80D.

⁷⁶ Omani Penal Law, Promulgated by Royal Decree 7/2018 Art. 115(a).

⁷⁷ Regulation (EU) No. 596/2014 of the European Parliament and of the Council of 16 April 2014 on market abuse (market abuse regulation) and repealing Directive 2003/6/EC of the European Parliament and of the Council and Commission Directives 2003/124/EC, 2003/125/EC and 2004/72/EC Art. 12(1)(c).

⁷⁸ T. Nagasako, 'Global Disinformation Campaigns and Legal Challenges' (2020) 1 *International Cybersecurity Law Review* 125, 129.

⁷⁹ French Law No. 2018-1202 of 22 December 2018 on the fight against the manipulation of information; M. Fiorentino, 'France passes controversial "fake news" law' (Euronews, 22 November 2018); UK House of Commons Digital, Culture, Media and Sport Committee, *Disinformation and 'fake news': Final Report* (2019) HC 1791, 13.

⁸⁰ See, e.g., Nigerian Protection from Internet Falsehoods and Manipulation and other related matters Bill 2019 (SB 132) s. 3(1)(b)(iv).

⁸¹ See International Federation of Journalists, 'Cambodia: Controlling election reporting guidelines released' (5 June 2018). Misinformation and disinformation laws that do not address elections expressly may still be enforced against journalists who report during elections: see, e.g., 'Nine arrested in Thailand for posting election "fake news"' (France 24, 28 March 2019).

Some election-related false speech laws predate recent technological threats. In the United Kingdom and India, it has long been a crime to make or publish ‘any false statement of fact’ about a candidate’s ‘personal character or conduct’ for the purpose of ‘affecting the return of any candidate at the election’, unless the speaker can show ‘he had reasonable grounds for believing, and did believe, the statement to be true.’⁸² When a politician in the United Kingdom said a candidate from another party had ‘reneged on his promise to live in his constituency’, this was found not to qualify as relating to his ‘personal character or conduct’, but a statement that he had ‘attempted to woo’ Muslims ‘who advocate violence’ was found to fall within the scope of the law.⁸³ Canada has a similar offence which prohibits anyone who, ‘with the intention of affecting the results of an election’, makes (1) false statements that a candidate has committed an offence or is being investigated for an offence; (2) false statements about a candidate’s listed personal characteristics, including their education or place of birth; or (3) false statements that a candidate has withdrawn from the election.⁸⁴

Finally, the protection of ‘public health’ has been included as a protected interest under misinformation laws, especially since the pandemic in 2020.⁸⁵ Before that time, laws in countries such as Tanzania,⁸⁶ Russia,⁸⁷ and Singapore⁸⁸ included undefined references to ‘life’, ‘human health’ and ‘public health’ alongside other interests like public order, public safety and national security in their misinformation laws.⁸⁹ Other laws referenced the spread of communicable diseases, such as China’s Criminal Law, which prohibits the fabrication of ‘false reports of danger, epidemic, security alerts or disasters.’⁹⁰

⁸² UK Representation of the People Act 1983 s. 106(1). See also Indian Penal Code 1860 Chapter IXA s. 171G; Indian Supreme Court, *N.S. Vardachari v. G. Vasantha Pai* (1972) 2 S.C.C. 594, 21 August 1972: Breach of s. 171G was not established because the speaker did not attack the character of the candidate by saying that he was not qualified by not being a graduate.

⁸³ The penalty for this conduct was that the election was declared void, it was reported to the Speaker of the House that the politician was guilty of an illegal practice, and he was compelled to vacate his seat and was barred from being elected to the House of Commons for three years: see, e.g., UK High Court of England and Wales, Queen’s Bench Division, *R (Woolas) v. Parliamentary Election Court* [2010] EWHC 3169 (Admin), 3 December 2010, §§5, 13.

⁸⁴ Canadian Elections Act 2000 ss 91–92. See also Canadian Elections Act 2000 ss 486(3)(c), (3)(d) and 486(4) (criminalizing the relevant prohibitions). In 2021, Ontario’s Superior Court found s. 91(1) to infringe on a Charter protection or fundamental rights in a manner that could not be considered a reasonable limit demonstrably justified in a free and democratic society and was therefore invalidated: Ontario Superior Court of Justice, *Canadian Constitutional Foundation v. Attorney General of Canada* 2021 ONSC 1224, 19 February 2021. Canadian law also criminally punishes the publication of false material that purports to have been produced by election officers, political parties, or candidates, if such publication is done with the intent of misleading the public: Canadian Elections Act 2000 s. 481.

⁸⁵ See, e.g., International Press Institute, ‘Rush to pass ‘fake news’ laws during Covid-19 intensifying global media freedom challenges’ (3 October 2020).

⁸⁶ Tanzanian Media Services Act 2016 s. 50.

⁸⁷ Russian Federal Law No. 149-FZ on Information, Informational Technologies and the Protection of Information 2006 (amended 2019) Art. 15.1. See also the Russian Code of Administrative Offences No. 195-FZ 2001 Art. 13.15.(1). Since 1996, Russian Penal Code article 237(1) has also criminalized ‘[c]oncealment or distortion of information about events, facts or phenomena that create a danger to life or health of people or to the environment, committed by a person obligated to provide the population and bodies authorized to take measures to eliminate such a danger with the information indicated.’

⁸⁸ Singaporean Protection from Online Falsehoods and Manipulation Act 2019 s. 7(1)(b)(ii).

⁸⁹ See also Nigerian Protection from Internet Falsehoods and Manipulation and other related matters Bill 2019 (SB 132) s. 3(1)(b)(ii).

⁹⁰ Chinese Criminal Law (amended through 2021) Art. 291(1).

Since the pandemic, the enactment of false speech laws related to public health has accelerated. Some nations such as South Africa,⁹¹ Jordan⁹² and Thailand⁹³ have specifically limited new provisions to false speech relating to the coronavirus. Thailand's laws apply more broadly to speech about the pandemic that 'may cause public fear'.⁹⁴ Other countries' laws have been justified by the pandemic but passed in even broader terms such as criminalization of speech that may 'misinform or cause uncertainty' in Bolivia,⁹⁵ false speech that 'compromises general security or public order' in Algeria⁹⁶ and false or 'twisted' speech that hinders a 'successful defence' against a 'state of emergency' in Hungary.⁹⁷

1.3. Intent

Misinformation laws can encompass two distinct culpability requirements: a mental state (*mens rea*) related to the *falsity* of the speech and to its *consequences*.

Some misinformation laws have high *mens rea* standards requiring that the person actually knew the information was false, intentionally disseminated it and in some instances had the intention that others would be misled or deceived.⁹⁸ For example, South Africa made it an offence to publish a statement through any medium with the intention to deceive others about Covid-19 generally, an individual's infection status or government measures to address the pandemic.⁹⁹ And in the case of a US statute criminalizing the conduct of any person who 'falsely represents himself . . . to have been awarded any decoration or medal authorized . . . for the Armed Forces of the United States', the Supreme Court read into the provision that the offender must have acted *knowingly*, even though on its face the relevant statute imposed no *mens rea* requirement.¹⁰⁰

Some misinformation laws have lower thresholds of 'negligence',¹⁰¹ recklessness¹⁰² or 'wilful blindness' regarding whether the statement is false.¹⁰³ Others are strict liability offences, prohibiting speech even when the person is unaware of the falsity of the

⁹¹ South African Disaster Management Act: Regulations relating to COVID-19, s. 14(2). Article 19, 'South Africa: Prohibitions of false COVID-19 information must be amended' (23 April 2021).

⁹² Human Rights Watch, 'Jordan: Free Speech Threats Under Covid-19 Response' (5 May 2020).

⁹³ Thai Emergency Decree on Public Administration in Emergency Situation 2005 s. 9(3); Human Rights Watch, 'Thailand: Immediately Repeal Emergency Regulation that Threatens Online Freedoms' (3 August 2021).

⁹⁴ *Ibid.*

⁹⁵ Bolivian Decree 4200 Art. 13.2; Human Rights Watch, 'Bolivia: COVID-19 Decree Threatens Free Expression' (7 April 2020).

⁹⁶ Algerian Law No. 20-06 amending and supplementing Decree No. 66-156 of June 8, 1966 on the Penal Code Art. 196 bis.

⁹⁷ Hungarian Act C of 2012 on the Penal Code s. 337(2), introduced by Act XII of 2020 on the containment of Coronavirus 31 March 2020 s. 10(2).

⁹⁸ See, e.g., Chinese Criminal Law (amended through 2021) Art. 291(1); Omani Penal Law 2018, Promulgated by Royal Decree 7/2018 Art. 115(a); and Turkish Capital Market Law No. 6362 2012 Art. 107(2).

⁹⁹ South African Disaster Management Act: Regulations relating to COVID-19 §14(2). See D. Milo, 'Criminalising fake news about COVID-19' (Musings on Media, 25 March 2020).

¹⁰⁰ US Supreme Court, *United States v. Alvarez* 567 U.S. 709, 28 June 2012 (although the majority of justices concluded that the statute at issue was unconstitutional, all of the justices agreed that the statute should be interpreted to capture only knowing falsehoods).

¹⁰¹ UK Representation of the People Act 1983; UK High Court of England and Wales, Queen's Bench Division, *R (Woolas) v. Parliamentary Election Court* [2010] EWHC 3169 (Admin), 3 December 2010, §84.

¹⁰² Gambian Criminal Code No. 25 1933 s. 181A(1). Provision available in the Gambian Supreme Court, *Gambia Press Union v. The Attorney General* Civil Suit No.1/2014, 9 May 2018.

¹⁰³ Nigerian Criminal Code Act (1916) Cap (C.38) s. 59(1) Laws of the Federation of Nigeria 2004.

information or the potential risks their speech poses.¹⁰⁴ For example, Cambodia prohibits ‘communicating or disclosing false information’, with no reference to knowledge of falsity of the information.¹⁰⁵

For *mens rea* as to *consequence*, some laws require that the person intended to cause resulting harm.¹⁰⁶ For example, Tanzania’s Cybercrimes Act applies to those who disseminate false information ‘with intent to . . . deceive or mislead the public’.¹⁰⁷ Others rely on actual or constructive knowledge of consequences, such as Singapore, which prohibits speech while ‘knowing or having reason to believe’ not only that ‘(a) it is a false statement of fact’ but also that ‘(b) the communication of the statement in Singapore is likely to’ be ‘prejudicial to public health, public safety, public tranquility or public finances’ or the security of Singapore, prejudicial to Singapore’s friendly relations with other countries, influence the outcome of an election, incite feelings of enmity, hatred or ill-will between different groups or ‘diminish public confidence in the performance of any duty or function of . . . the Government’.¹⁰⁸ A number of laws are based on objective, rather than subjective, foreseeability,¹⁰⁹ and others contain no *mens rea* requirement as to the consequence of the speech at all.¹¹⁰

Causation is also addressed in diverse ways in national misinformation laws. Some laws make no reference to the relationship between the speech and consequences flowing from it.¹¹¹ Others require proof that the speaker intended to cause a harmful effect, but do not require any actual consequences to occur.¹¹² For instance, Rwandan law states that ‘no person may send . . . false messages . . . for purposes of causing annoyance, inconvenience, or needless anxiety’.¹¹³ In a number of laws, the speech need only be ‘capable of’ causing or ‘likely to’ cause the given consequence.¹¹⁴ In the United Arab Emirates, it is an offence to provide an organization with ‘any incorrect or misleading information which may damage the interests of the State or injures its reputation, prestige or stature’.¹¹⁵ Another option is to specify a more stringent direct causation of proscribed consequences, which has been done by the United States in a non-legislative

¹⁰⁴ See, e.g., Iranian Press Law 2002 Art. 6; United Arab Emirates Federal Decree Law No. 34 of 2021 on Combatting Rumors and Cybercrimes Art. 52.

¹⁰⁵ Cambodian Criminal Code Art. 425.

¹⁰⁶ Cambodian Criminal Code Art. 425; French Law on the Freedom of the Press of 29 July 1881 Art. 27; Rwandan Law Governing Information and Communication Technologies No. 24 2016 Art. 60.

¹⁰⁷ Tanzanian Cybercrimes Act 2015 s. 16.

¹⁰⁸ Singaporean Protection from Online Falsehoods and Manipulation Act 2019 s. 7(1).

¹⁰⁹ See, e.g., Egyptian Penal Code No. 58 1937 Art. 80D; French Code Electoral Art. L163-2-1.

¹¹⁰ Belarusian Penal Code 1999 Art. 340(2); Chinese Criminal Law (amended through 2021) Art. 291(1); Kazakhstani Criminal Code 2014 (amended 2018) Art. 274-1. See also offences that do not contain prohibited consequences (no corresponding mental element): Chinese Criminal Law (amended through 2021) Art. 291(1); Chinese Cybersecurity Law 2017 Art. 12; Gambian Information and Communication (Amendment) Act 2013 s. 173; Gambian Criminal Code No. 25 1933 s. 181.

¹¹¹ Canadian Elections Act 2000 ss 91(1), 92; Chinese Criminal Law (amended through 2021) Art. 291(1); Iranian Press Law 2002 Art. 6; Iranian Computer Crimes Law 2010 Art. 18; Tanzanian Media Services Act 2016 s. 50(1).

¹¹² See, e.g., United Arab Emirates Federal Decree Law No. 34 of 2021 on Combatting Rumors and Cybercrimes Art. 25; Malaysian Communications and Multimedia Act 1998 s. 211.

¹¹³ Rwandan Law Governing Information and Communication Technologies No. 24 2016 Art. 60.

¹¹⁴ Egyptian Penal Code No. 58 1937 Art. 80D; French Law on the Freedom of the Press of 29 July 1881 Art. 27; Rwandan Law Determining Offences and Penalties in General No. 68 2018 Art. 194; Rwandan Law on Prevention and Punishment of Cyber Crimes No. 24 2016 Art. 39.

¹¹⁵ United Arab Emirates Federal Decree Law No. 34 of 2021 on Combatting Rumors and Cybercrimes Art. 52.

setting to address hoaxes.¹¹⁶ Similarly, a 2020 amendment to the Russian Criminal Code criminalizes the ‘public dissemination of knowingly false information of public importance’ that *has caused* severe consequences such as harm to human health or death.¹¹⁷

1.4. Exclusions, exceptions and defences

Many misinformation laws do not contain an affirmative defence.¹¹⁸ Some do include affirmative defences relating to the actions taken by the person to verify the information before disseminating it.¹¹⁹ And certain misinformation laws prevent or reduce penalties in the case of disclaimers or information obtained from or reported to the state. In Tanzania, false content cannot be published ‘unless it is clearly pre-stated that the content is—(i) satire and parody; (ii) fiction; and (iii) where it is preceded by a statement that the content is not factual.’¹²⁰ There is no liability for disseminating false information in Belarus if that information was disseminated by members of the media and was received from state entities, news agencies or official messages from political parties or legal entities.¹²¹ And in Qatar, a person shall be exempt from penalties if they intend to ‘inform the competent authorities of any information about [a] crime and its participants before their knowledge thereof’.¹²²

1.5. Penalties

Misinformation laws provide for a variety of penalties, ranging from fines to life imprisonment.¹²³ Some states also provide for corporal punishment, the loss of civil liberties, censorship or removal of the speech, and increased penalties under certain conditions, such as when the misinformation is disseminated during wartime or anonymously.

1.5.1. Criminal

Custodial sentences are a regular feature of misinformation laws, although the length of such penalties varies widely. Rwanda’s misinformation offence has a maximum penalty of life imprisonment when committed during

¹¹⁶ See the US Federal Communication Commission’s rules against hoaxes, which prohibit ‘broadcasting false information concerning a crime or catastrophe if’ the broadcaster (1) ‘knows this information is false’; (2) ‘it is foreseeable that broadcast of the information will cause substantial public harm’ and (3) ‘broadcast of the information does in fact directly cause substantial public harm’. FCC, ‘Hoaxes’. See also US Supreme Court, *Brown v. Entertainment Merchants Association* 564 U.S. 786, 27 June 2011, 799 (asserting there must be ‘a direct causal link’ between the speech and alleged harm in order for a statute to pass strict scrutiny review).

¹¹⁷ Russian Criminal Code Art. 207.2.

¹¹⁸ However, some states’ laws contain general defences applicable to all crimes or to a category of offences.

¹¹⁹ See, e.g., Gambian Criminal Code No. 25 1933 s. 181A(2); Nigerian Criminal Code Act (1916) Cap (C.38) s. 59(2) Laws of the Federation of Nigeria 2004.

¹²⁰ Tanzanian Electronic and Postal Communications (Online Content) Regulations 2018 reg. 12(1).

¹²¹ Belarusian Law on Mass Media 2008 Art. 52.1 (the media law applying to persons such as a ‘journalist, founder (founders) of a mass media outlet, editor-in-chief of a mass media outlet’).

¹²² Qatari Cybercrime Prevention Law No. 14 2014 Art. 54.

¹²³ Rwandan Law Determining Offences and Penalties in General No. 68 2018 Art. 194 (where spreading false information or harmful propaganda ‘with intent to cause public disaffection against the Government of Rwanda, or where such information . . . is likely or calculated to cause public disaffection or a hostile international environment’ against Rwanda and the offence occurs in wartime).

wartime.¹²⁴ In other states, prison terms may extend to 20 years,¹²⁵ five years,¹²⁶ one year,¹²⁷ or three months.¹²⁸ In the United Arab Emirates, for example, an Emirati man was sentenced to 10 years in prison for publishing ‘false information’ that was ‘detrimental to the country and its foreign policies.’¹²⁹

Certain criminal misinformation laws provide for increased penalties in particular circumstances. Cameroonian law doubles the sentence for the dissemination of false news from a maximum of five years to 10 years if done anonymously.¹³⁰ Singaporean law provides for increased penalties of up to 10 years in prison when ‘an inauthentic online account or bot is used’ to spread misinformation.¹³¹ Speech may also attract a harsher penalty if it inflicts ‘heavy’ or ‘large scale’ damage,¹³² is for the benefit of a foreign country or any terrorist group,¹³³ occurs in ‘time of war,’¹³⁴ or when it is ‘likely to shake the discipline or the morale of the armies or to hinder the nation’s war effort.’¹³⁵ Other laws increase the penalty if the misinformation offence is committed or facilitated by a public employee ‘taking advantage of his ... authority’¹³⁶ or by a repeat offender.¹³⁷

1.5.2. Civil

Financial penalties are a common feature of civil legislation on misinformation.¹³⁸ Civil penalties may also be imposed in combination with prison sentences.¹³⁹ In France, the publication of false news in bad faith and likely to disturb public peace is punishable by fines of up to 45,000 Euros (or 135,000 Euros in aggravating circumstances).¹⁴⁰ Saudi law allows for fines of up to approximately \$800,000 US dollars alongside prison sentences.¹⁴¹ In Singapore, a fine of over \$35,000 US dollars can be imposed if an individual

¹²⁴ Rwandan Law Determining Offences and Penalties in General No. 68 2018 Art. 194.

¹²⁵ South Sudanese Penal Code 2008 Act 9 ss. 12 and 75.

¹²⁶ Belarusian Penal Code 1999 Art. 340(1). See also Canadian Elections Act 2000 s. 500(5)(b): summary conviction (the penalty is a fine of up to \$20,000 (approximately 15,000 USD) and/or imprisonment for up to a year) and conviction on indictment (fine of up to \$50,000 (approximately 37,000 USD) and/or imprisonment for up to five years).

¹²⁷ Kazakhstani Criminal Code 2014 (amended 2018) Art. 274-1.

¹²⁸ Omani Penal Law, Promulgated by Royal Decree 7/2018 Art. 115(a)–(c), potentially going up to three years’ imprisonment.

¹²⁹ M. Al Zarooni, ‘Emirati fined Dh1 million, jailed for publishing fake news’ (Khaleej Times, 1 January 2019).

¹³⁰ Cameroonian Penal Code 2016 s. 240.

¹³¹ Singaporean Protection from Online Falsehoods and Manipulation Act 2019 s. 7(3).

¹³² Belarusian Penal Code 1999 Art. 340(2); Kazakhstani Criminal Code 2014 (amended 2018) Art. 274-3.

¹³³ United Arab Emirates Federal Decree Law No. 34 of 2021 on Combatting Rumors and Cybercrimes Art. 60 (on aggravating circumstances).

¹³⁴ Egyptian Penal Code Law No. 58 1937 Art. 80D; Qatari Penal Code (Amended 2020) Art. 136(bis); Rwandan Law Determining Offences and Penalties in General No. 68 2018 Art. 194.

¹³⁵ French Law on the Freedom of the Press of 29 July 1881 Art. 27.

¹³⁶ Qatari Cybercrime Prevention Law No. 14 2014 Arts 6, 51. See also Iranian Computer Crimes Law 2010 Art. 26(A)—it is an aggravating circumstance if a state employee commits a computer crime in performance of their duties, ‘punished by more than two thirds of the maximum extent of one or both the punishments.’

¹³⁷ See, e.g., Belarusian Penal Code 1999 Art. 340(2); Cambodian Criminal Code Arts 83–92.

¹³⁸ See, e.g., Belarusian Law on Mass Media 2008 Art. 48; Cambodian Law on the Press 1995 Art. 13; Kazakhstani Code of Administrative Offences No. 235-V 2014 Art. 456(1).

¹³⁹ For combined fine and imprisonment, see, e.g., United Arab Emirates Federal Decree Law No. 34 of 2021 on Combatting Rumors and Cybercrimes Art. 25.

¹⁴⁰ French Law on the Freedom of the Press of 29 July 1881 Art. 27.

¹⁴¹ SR 3,000,000. D. Al-Khudair, ‘Saudi residents spreading “fake news” face five years’ jail’ (Arab News, 3 May 2020).

shares false information, increasing to double that amount and up to 10 years' imprisonment if it is published by an 'inauthentic online account or a bot'.¹⁴² Rwanda fines business entities that commit an offence under Rwanda's 'cybercrime' law up to 30–70 per cent of the company's annual profits.¹⁴³ In China, a journalist was fined \$990,000 US dollars alongside a 15-year prison sentence for 'picking quarrels and provoking trouble', after he alleged fraud and corruption by regional party officials.¹⁴⁴

A number of misinformation laws also make provision for the loss of civil liberties and confiscation of property in addition to fines or imprisonment.¹⁴⁵ These include loss of the right to vote,¹⁴⁶ deportation,¹⁴⁷ restrictions on residency,¹⁴⁸ ongoing monitoring or surveillance,¹⁴⁹ confiscation of devices and funds and the forced closure of a website.¹⁵⁰

Other types of media regulation of false speech exist in parallel with criminal laws.¹⁵¹ For example, Belarus' law allows the government to suspend mass media without a court decision for up to three months if, for example, two or more written requests have been issued to the owner of the publication within two years alleging dissemination of either 'untrue information that may harm state or public interests' or 'data not corresponding to reality and disgracing the honour, dignity or business' of individuals.¹⁵² This law has been used to block hundreds of information resources and websites.¹⁵³ In Russia, 'immediately upon receipt' of a notice from the relevant federal executive body that information is 'socially significant false information', the 'online publication must delete the information'.¹⁵⁴ And in Singapore, any Minister may make orders requiring publication of a correction and block future communication in the event of non-compliance.¹⁵⁵ In other jurisdictions,

¹⁴² Singaporean Protection from Online Falsehoods and Manipulation Act 2019 s. 7 (50,000 or 100,000 Singapore dollars (approximately 37,000 or 75,000 USD)).

¹⁴³ Rwanda Law on Prevention and Punishment of Cyber Crimes No. 24 2016 Art. 51.

¹⁴⁴ CPJ, 'Chinese police arrest critical blogger Chen Jieren' (11 July 2018); A. France-Presse, 'Chinese journalist jailed for 15 years for attacking Communist Party' (South China Morning Post, 1 May 2020).

¹⁴⁵ See, e.g., Rwandan Law on Prevention and Punishment of Cyber Crimes No. 60 2018 Art. 9, Rwandan Law Determining Offences and Penalties in General No. 68 2018 Arts 42–45 (however, civil rights may only be deprived for those who have been sentenced to life imprisonment under Rwanda's criminal misinformation laws); United Arab Federal Decree Law No. 34 of 2021 on Combatting Rumors and Cybercrimes Arts 41–43.

¹⁴⁶ Cambodian Criminal Code Arts 55, 426; Rwandan Law Determining Offences and Penalties in General No. 68 2018 Art. 43.

¹⁴⁷ Qatari Cybercrime Prevention Law No. 14 2014 Art. 52.

¹⁴⁸ Cambodian Criminal Code Art. 426.

¹⁴⁹ United Arab Emirates Federal Decree Law No. 34 of 2021 on Combatting Rumors and Cybercrimes Art. 59.

¹⁵⁰ Qatari Cybercrime Prevention Law No. 14 2014 Art. 53; Rwandan Law on Prevention and Punishment of Cyber Crimes No. 60 2018 Art. 52 (allowing the confiscation of a computer system or software used to commit an offence, as well as the closure of a premise or corporate body); United Arab Emirates Federal Decree Law No. 34 of 2021 on Combatting Rumors and Cybercrimes Art. 59.

¹⁵¹ Regulatory regimes that pertain to traditional or social media are not the focus of this chapter. See s. I. (Introduction).

¹⁵² Belarusian Law on Mass Media 2008 Arts 49–50.

¹⁵³ Freedom House, 'Freedom on the Net 2019: Belarus' (2019).

¹⁵⁴ Russian Federal Law No. 149-FZ on Information, Informational Technologies and the Protection of Information 2006 (amended 2019) Art. 15.3(1). Socially significant false information is defined to include information distributed under the guise of truthful messages if they create a threat that endangers people's lives, health, or property; create possibilities for mass violations of public order or public security; or possibly hinder the work of transportation and social infrastructure, credit institutions, lines of communications, industry and energy enterprises.

¹⁵⁵ Singaporean Protection from Online Falsehoods and Manipulation Act 2019 ss 4, 10–13, 16. Part 4 of the Act contains similar provisions, but directed towards 'internet intermediaries and providers of mass media

convictions for misinformation may result in the suspension of a publication or broadcast license.¹⁵⁶

2. Application of False Speech Laws Around the World

There have been some striking cases of prosecutions under false speech laws, both against journalists and other speakers such as lawyers and human rights activists.¹⁵⁷ The examples provided below are illustrative of this practice and demonstrate how misinformation laws can be abused for ulterior purposes: to suppress critical journalism, quash dissent or silence speech that embarrasses the government of the day.

2.1. Europe

2.1.1. Serbia

Free speech can be suppressed not only through vaguely worded misinformation laws, but also through the *application* of otherwise reasonably precise laws. Serbia has false speech laws with stringent legal protections on paper, but its response to the Covid-19 pandemic included misuse of these laws to chill journalistic speech critical of the government. Article 343 of Serbia's Criminal Code, which makes it an offence to cause panic and disorder by disseminating false information, requires the offender to actually *cause* panic or disorder by spreading misinformation and to do so with *intent*.¹⁵⁸ In other words, the offence has very high *mens rea* and causality requirements, even if the prohibited consequences (causing panic and disorder or frustrating the enforcement of the decisions of governmental authorities) are somewhat ill-defined.

The law had previously been used in relatively uncontroversial circumstances, such as bomb hoaxes. For example, in one case an individual who was due to face trial falsely reported that there was a bomb in the building, forcing the building to be evacuated and his trial to be delayed. He was convicted for impeding the work of Serbian judicial authorities through his false report.¹⁵⁹ But during the pandemic it was used against Ana Lalić, a Serbian journalist working for an online media outlet, who reported that a hospital in a Serbian city had run out of personal protective equipment, citing testimony from doctors and nurses.¹⁶⁰ Her reporting was immediately denounced as 'fake news' by Serbian political leadership and hospital management, and police conducted a late-night raid, search, and arrest at her apartment on the same day the article was

services', rather than the communicators of the statement. These orders may be made against persons outside of Singapore: see s. II.1.1. (Vagueness and overbreadth).

¹⁵⁶ See, e.g., French Law on the Freedom of the Press of 29 July 1881 Art. 62.

¹⁵⁷ See UN Special Rapporteur D. Kaye, *Disease pandemics and the freedom of opinion and expression* (2020) UN Doc. A/HRC/44/49, §§34–39.

¹⁵⁸ See Serbian Criminal Code Art. 343.

¹⁵⁹ See Appeals Court of Belgrade, *Judgment Kž-po1 8/11*, 25 May 2011.

¹⁶⁰ Council of Europe, 'Safety of Journalists Platform: Journalist Ana Lalic Detained Overnight after Reporting on Conditions in Hospital' (1 April 2020).

published.¹⁶¹ By doing so, the police sent a clear message to other members of the press and even anyone online who might consider making similar allegations, even though the charges against her were ultimately dropped.¹⁶²

2.2. Asia Pacific

2.2.1. China

China is one of the ‘world’s most restrictive media environments’¹⁶³ and the ‘world’s biggest captor of journalists.’¹⁶⁴ An important tool of suppression is the broadly worded law criminalizing ‘picking quarrels and provoking trouble’, which has been used by Chinese authorities since 2013 to prosecute people for allegedly false speech, especially online speech about human rights and politics.¹⁶⁵ In 2020, a former Chinese state media journalist turned anti-corruption blogger, Chen Jieren, was sentenced under this law to 15 years in prison and fined approximately \$990,000 US dollars after publishing two articles on his blog about corruption by political officials.¹⁶⁶ The court stated that he had ‘attacked and vilified the Communist Party and government’ by publishing ‘false or negative information’ and through ‘malicious speculation.’¹⁶⁷ Another journalist, Zhang Zhan, was sentenced under the same law to four years in prison for covering the beginning of the pandemic on social media; she had been reporting from Wuhan in February 2020.¹⁶⁸ The law has been condemned by the UN Working Group on Arbitrary Detention as ‘so vague and broad that it could be used to deprive individuals of their liberty without a specific legal basis.’¹⁶⁹

2.2.2. Myanmar

Following the military coup in February 2021, Myanmar authorities enacted a new provision of Myanmar’s Penal Code that has been a key tool to stifle dissent.¹⁷⁰ The new broadly-worded section 505A criminalizes comments that ‘cause fear’ or spread ‘false news’, punishable by up to three years’ imprisonment.¹⁷¹ The Committee to Protect Journalists reports that at least 17 journalists were imprisoned on false news charges

¹⁶¹ M. Stojanovic, ‘Serbian Reporter’s Arrest Over Pandemic Article Draws PM’s Apology’ (Balkan Insight, 2 April 2020).

¹⁶² See also Article 19, ‘Serbia: Journalist Ana Lalic arrested for reporting on inadequate hospital facilities for coronavirus’ (2 April 2020).

¹⁶³ Freedom House, ‘Freedom in the World: China’ (2022).

¹⁶⁴ Reporters Without Borders, ‘An unprecedented RSF investigation: The Great Leap Backwards of Journalism in China’ (7 December 2021).

¹⁶⁵ Chinese Criminal Law (amended through 2021) Art. 293; G. Rui, ‘Picking quarrels and provoking trouble: how China’s catch-all crime muzzles dissent’ (South China Morning Post, 25 August 2021).

¹⁶⁶ CPJ, ‘Chen Jieren: Imprisoned in China’ (4 July 2018).

¹⁶⁷ J. Griffiths, ‘Chinese journalist jailed for 15 years for “vilifying the Communist Party and government”’ (CNN, 1 May 2020).

¹⁶⁸ Reporters Without Borders, ‘The Great Leap Backwards of Journalism in China’ (7 December 2021), 50. See also WGAD, *Zhang v. China* (Opinion no. 25/2021), 6 September 2021, §42.

¹⁶⁹ WGAD, *Zhang v. China* (Opinion no. 25/2021), 6 September 2021, §52.

¹⁷⁰ Human Rights Watch, ‘Myanmar: Post-Coup Legal Changes Erode Human Rights’ (2 March 2021).

¹⁷¹ *Ibid.* Myanmar State Administration Council Law No. 5/2021 Law Amending the Penal Code 1861, s. 505A (this offence is also punishable by fine). It is also an offence in Myanmar to incite disaffection towards the defence services or defence services personnel, or for individuals holding peaceful assembly and procession to circulate false information, rumours and/or false data: Myanmar Penal Code 1861, s. 124A; Myanmar Law relating to the Right of Peaceful Assembly and Peaceful Procession 2011, s. 12(f).

in Myanmar in 2021.¹⁷² Among them was American journalist Danny Fenster, former managing editor of independent news outlet Frontier Myanmar, who was sentenced to 11 years' imprisonment for a slew of offences including spreading false information.¹⁷³

2.3. Middle East and Africa

2.3.1. Egypt

Egypt also frequently misuses false speech laws against journalists reporting on truthful political issues or criticisms of government.¹⁷⁴ One of the best known examples is the use of false speech and terrorism charges against Mohamed Fahmy, a Canadian-Egyptian citizen and head of the Al Jazeera bureau in Cairo, and his colleagues Australian correspondent Peter Greste and Egyptian producer Baher Mohamed.¹⁷⁵ The men were convicted of a string of offences, including broadcasting 'false news to give the outside world an impression that [Egypt] is undergoing a state of internal strife', thereby damaging public security and spreading fear.¹⁷⁶ The Court argued that the journalists had been brought together 'by the devil' to destabilize Egypt, and sentenced Fahmy and Greste to seven years' imprisonment and Mohamed to 10 years.¹⁷⁷ The sentence was later reduced to three years and the men were eventually pardoned and released.¹⁷⁸

Similar tactics have been adopted against journalists and activists speaking out about the pandemic in Egypt. Just hours after journalist Mohamed Monir appeared in an Al Jazeera interview (and published a corresponding article) criticizing the Egyptian government's pandemic response, Monir's apartment was raided by Egyptian security officers.¹⁷⁹ The raid was captured and uploaded on Facebook.¹⁸⁰ Two days later, Monir's family reported that he was kidnapped and placed in pre-trial detention for 15 days.¹⁸¹ While no specific examples of his journalistic work were cited, Monir was charged with 'joining a terrorist group, spreading fake news, and misusing social media'.¹⁸² After 17 days in detention, Monir was released after contracting Covid-19 while in custody.¹⁸³ A week later, Monir died while in medical isolation.¹⁸⁴

¹⁷² CPJ, '30 Journalists Imprisoned in Myanmar' (1 December 2021).

¹⁷³ Fenster was sentenced to 'three years for incitement for spreading false or inflammatory information, three years for contacting illegal organisations and five years for violating visa rules', but has since been released: M. Moe & R. Ratcliffe, 'Myanmar junta jails US journalist Danny Fenster for 11 years' (The Guardian, 12 November 2021).

¹⁷⁴ See, e.g., CPJ, 'Egypt's top prosecutor orders authorities to monitor media for "fake news."' (28 February 2018).

¹⁷⁵ Al Jazeera has been banned by the Egyptian government following the 2013 military coup on the rationale that it was a platform for the Muslim Brotherhood political movement. See ch. 6 (Terrorism Laws) ss. II.2.3.1 (Egypt). See also Amnesty International UK, 'Egypt frees Al Jazeera staff jailed for journalism' (18 May 2020).

¹⁷⁶ Giza Criminal Court Fifth Circuit, Criminal Case No. 1145 of 2014, 23 June 2014 (unofficial translation). The Court also sentenced a number of other individuals at the same trial.

¹⁷⁷ 'Timeline: The two trials of Mohamed Fahmy' (CTV News, 29 August 2015).

¹⁷⁸ K. Fahim, 'Egypt Pardons Al Jazeera Journalists Mohamed Fahmy and Baher Mohamed' (The New York Times, 23 September 2015).

¹⁷⁹ CPJ, 'Egyptian journalist Mohamed Monir detained, charged with spreading false news' (15 June 2020).

¹⁸⁰ Ibid.

¹⁸¹ 'Lawyer: Egyptian journalist detained on fake news charges' (AP News, 15 June 2020).

¹⁸² CPJ, 'Egyptian journalist Mohamed Monir detained, charged with spreading false news' (15 June 2020).

¹⁸³ Monir was arrested on 15 June 2020 and released on 2 July 2020. CPJ, 'Egyptian journalist Mohamed Monir dies after contracting COVID-19 in pretrial detention' (13 July 2020).

¹⁸⁴ Ibid.

Similarly, Sanaa Seif, a prominent human rights activist, was sentenced to 18 months' imprisonment in March 2021 for 'disseminating false news', 'misuse of social media' and insulting a police officer on duty, after she took to social media to speak out about an altercation she had with a police officer, and to criticize the lack of response to Covid-19 in Egypt's overcrowded prison system.¹⁸⁵ This followed a separate case against her brother, activist Alaa Abd El Fattah, a significant figure in the 2011 uprising in Egypt¹⁸⁶ who was also sentenced to five years' imprisonment for spreading fake news after a trial that allegedly lacked basic due process guarantees.¹⁸⁷ Human rights groups described Seif's conviction and sentence as 'yet another crushing blow for the right to freedom of expression in Egypt'.¹⁸⁸

III. International Legal Standards

Under international law, states may not impose penalties on speech on the basis of its falsity alone. The list of legitimate grounds for restricting speech set out in human rights treaties is exhaustive. While some of the grounds set out are broad, none can be interpreted as permitting restrictions on speech simply on grounds of its falsity.¹⁸⁹ Lying alone cannot be a crime, nor the basis for civil sanctions if there is no demonstrable harm. But when false speech is likely to cause or has caused prohibited harms, the speaker intends to cause such harms, and the penalty is proportionate, penalties for false speech may be permissible in line with international standards.

There are few cases litigated before international human rights bodies that specifically address the question of how misinformation can be regulated by a state.¹⁹⁰ But as this issue has gained prominence in recent years, soft law guidance has started to fill this gap.¹⁹¹ Many soft law standards correspond closely to those in the defamation context, which also, of course, applies to false speech but only the subset of such speech that harms reputation.¹⁹² The question that arises, therefore, is which harms other than reputation warrant the penalization of false speech? And where should the lines be drawn? If the lines are to be drawn differently from those in the defamation context, this should be justified by a difference in the nature of the harm caused by the speech.¹⁹³

¹⁸⁵ Amnesty International, 'Egypt: Sanaa Seif's conviction on bogus charges a travesty of justice' (17 March 2021).

¹⁸⁶ 'Alaa Abdel Fattah: Leading Egyptian activist jailed for five years' (BBC, 20 December 2021).

¹⁸⁷ Human Rights Watch, 'Egypt: Wave of Unjust "Emergency" Trials' (19 December 2021). Two other defendants, Mohamed al-Baqer and Mohamed Ibrahim, were both sentenced to four years' imprisonment at the same trial.

¹⁸⁸ Amnesty International, 'Egypt: Sanaa Seif's conviction on bogus charges a travesty of justice' (17 March 2021).

¹⁸⁹ See HRC, General Comment No. 34 (2011), §22: 'Restrictions are not allowed on grounds not specified in paragraph 3, even if such grounds would justify restrictions to other rights protected in the Covenant.' See also §§28–32 (discussing the scope of legitimate aims enumerated in Art. 19(3)).

¹⁹⁰ See s. III.3.1. (Harm).

¹⁹¹ See, e.g., UN Special Rapporteur on Freedom of Opinion and Expression, OSCE Representative on Freedom of the Media, OAS Special Rapporteur on Freedom of Expression and ACmHPR Special Rapporteur on Freedom of Expression and Access to Information, *Joint Declaration on Freedom of Expression and 'Fake News', Disinformation and Propaganda* (2017) FOM.GAL/3/17.

¹⁹² *Ibid.*, §2.

¹⁹³ See e.g., M. Milanovic, 'Viral Misinformation and the Freedom of Expression: Part I' (EJIL:Talk!, 13 April 2020).

International standards relevant to misinformation derive from the basic framework of limitations on the freedom of expression, and the three-part test for justifying the imposition of penalties for speech under international law.¹⁹⁴ This provides that any restriction on expression (1) must be provided for by law (legality requirement); (2) must pursue a legitimate aim (legitimacy requirement); and (3) must be necessary to achieve that aim (necessity requirement).¹⁹⁵ An analysis of necessity includes an assessment of proportionality measured against the specific legitimate aim that the state is pursuing.¹⁹⁶ This requires examining whether the restriction on speech is a suitable measure for achieving the legitimate aim, whether there are less restrictive measures that could have achieved it, and whether a fair balance has been struck between the aim pursued and the interests of the individual whose speech is being restricted.¹⁹⁷

1. Legality

International human rights law requires restrictions on freedom of expression to be set out in domestic law, which must meet qualitative criteria of accessibility, precision and clarity.¹⁹⁸ Vagueness in misinformation laws is therefore fatal from a human rights standpoint—not only because overly vague laws offend against rule of law values, but also because they tend to chill far too much truthful, protected speech.¹⁹⁹

Vague definitions vest state authorities that interpret and apply a misinformation law with excessive discretion that can be used subjectively, inconsistently or arbitrarily.²⁰⁰ As noted by the UN Special Rapporteur on freedom of opinion and expression, ‘[v]ague prohibitions of disinformation effectively empower government officials with the ability to determine the truthfulness or falsity of content in the public and political domain, in conflict with the requirements of . . . article 19(3)’ of the ICCPR.²⁰¹ And as both UN and regional experts have confirmed, ‘[g]eneral prohibitions on the dissemination of information based on vague and ambiguous ideas, including “false news”

¹⁹⁴ Human Rights Council Resolution 44/12 Freedom of opinion and expression (A/HRC/RES/44/12) 16 July 2020 (‘responses to the spread of disinformation and misinformation must be grounded in international human rights law, including the principles of lawfulness, legitimacy, necessity and proportionality’).

¹⁹⁵ See ICCPR Art. 19(3); ECHR Art. 10(2); ACHR Art. 13(2); ACHPR Art. 9(2). See s. I. (Introduction).

¹⁹⁶ See HRC, General Comment No. 34 (2011), §34; IACmHR, *The Inter-American Legal Framework Regarding the Right to Freedom of Expression* (2009) OEA/Ser.L/V/II CIDH/RELE/INF. 2/09, §88.

¹⁹⁷ See HRC, General Comment No. 34 (2011), §34; IACmHR, *The Inter-American Legal Framework Regarding the Right to Freedom of Expression* (2009) OEA/Ser.L/V/II CIDH/RELE/INF. 2/09, §§85–87; Human Rights Council Resolution 44/12 Freedom of opinion and expression (A/HRC/RES/44/12) 16 July 2020, §8(e).

¹⁹⁸ See ICCPR Art. 19(3); ECHR Art. 10(2); ACHR Art. 13(2); ACHPR Art. 9(2); HRC, General Comment No. 34 (2011), §25. See also ECtHR (GC), *Rotaru v. Romania* (App. no. 28341/95), 4 May 2000, §52; ECtHR (GC), *Maestri v. Italy* (App. no. 39748/98), 17 February 2004, §30.

¹⁹⁹ See IACmHR, *The Inter-American Legal Framework Regarding the Right to Freedom of Expression* (2009) OEA/Ser.L/V/II CIDH/RELE/INF. 2/09, §§70–71.

²⁰⁰ In that regard, the European Court has consistently held that the law must ‘afford adequate legal protection against arbitrariness and accordingly indicate with sufficient clarity the scope of discretion conferred on the competent authorities and the manner of its exercise’: ECtHR (GC), *S. and Marper v. United Kingdom* (App. Nos. 30562/04 & 30566/04), 4 December 2008, §95.

²⁰¹ See UN Special Rapporteur D. Kaye, *Disease Pandemics and the freedom of opinion and expression* (2020) UN Doc. A/HRC/44/49, §49.

or “non-objective information” are incompatible with international standards for restrictions on freedom of expression.²⁰² For example, UN Special Rapporteurs have expressed concern that ‘the vague language’ of Qatar’s misinformation law²⁰³ ‘may result in disproportionate restrictions on freedom of expression.’²⁰⁴ This has been borne out in practice, with individuals arrested under the law for protesting a voting ban for certain families.²⁰⁵

Vague misinformation laws that are criminal in nature are of the gravest concern. Not only do they exacerbate the chilling effects of criminal punishment, they also go against the bedrock *nullum crimen sine lege* principle of criminal law and may violate human rights on that basis as well.²⁰⁶ As the Community Court of Justice of the Economic Community of West African States has stated, ‘[n]arrowly drawing offences has been treated as particularly important in the case of free speech because of what is known as the “chilling effect” which occurs when a wide or vague speech-restricting provision forces self-censorship on speakers ... because they do not wish to risk being caught on the wrong side of it.’²⁰⁷

International and regional human rights bodies have consistently criticized vague wording in criminal misinformation offences. The UN’s Human Rights Committee has expressed concern about the prosecution of journalists using laws that criminalize spreading or disseminating ‘false news.’²⁰⁸ The Working Group on Arbitrary Detention has similarly critiqued misinformation laws, describing Bahrain’s offences of ‘spreading false news abroad which damages the national interest’ and ‘spreading false rumours in wartime’ as ‘so vague and overly broad that they could ... result in penalties being imposed on individuals who had merely exercised their rights under international law.’²⁰⁹ The Working Group also described the offence of ‘harming the image of the State by disseminating false information to foreign groups’ as one of many laws in Saudi Arabia that ‘fall short of international standards and can easily be used to criminalize the peaceful exercise of the rights to freedom of expression.’²¹⁰

²⁰² UN Special Rapporteur, OSCE Representative on Freedom of the Media, OAS Special Rapporteur and ACmHPR Special Rapporteur on Freedom of Expression and Access to Information, *Joint Declaration on Freedom of Expression and ‘Fake News’, Disinformation and Propaganda* (2017) FOM.GAL/3/17, §2(a).

²⁰³ See s. II.1.1. (Vagueness and overbreadth).

²⁰⁴ UN Special Rapporteurs D. Kaye and M. Forst, *Letter to the Government of Qatar dated 14 April 2020*, Ref OL QAT 1/2020, 3.

²⁰⁵ Qatari Penal Code (Amended 2020) Art. 136 (criminalizing ‘false or malicious news’ disseminated with the intention of ‘harming national interests, provoking public opinion, or violating the social system or public order of the state’); T. M. Issa, ‘Qatar arrests seven people for racial speech, 88 others for breaking COVID measures’ (Al Arabiya, 10 August 2021); ‘Qatar’s new electoral law stirs up tribal sensitivities’ (Reuters, 12 August 2021).

²⁰⁶ See IACmHR, *The Inter-American Legal Framework Regarding the Right to Freedom of Expression* (2009) OEA/Ser.L/V/II CIDH/RELE/INF. 2/09, §72. See, generally, ICCPR Art. 15; ECHR Art. 7; ACHPR Art. 7(2); ACHR Art. 9.

²⁰⁷ ECOWAS C CJ, *Federation of African Journalists v. The Gambia* (Suit no. ECW/CCJ/APP/36/15), 13 February 2018, §41.

²⁰⁸ HRC, *Concluding Observations: Cameroon* (2010) UN Doc. CCPR/C/CMR/CO/4, §25; HRC, *Concluding Observations: Sudan* (2014) UN Doc. CCPR/C/SDN/CO/4, §21. See also HRC, *Concluding Observations: Macao* (2013) UN Doc. CCPR/C/CHN-MAC/CO/1, §16.

²⁰⁹ WGAD, *Rajab v. Bahrain* (Opinion no. 13/2018), 19 April 2018, §§7, 29.

²¹⁰ WGAD, *Sheikh Suliaman al-Rashudi and others v. Saudi Arabia* (Opinion no. 38/2015), 4 September 2015, §73.

Regional courts have reached similar conclusions. For instance, the East African Court of Justice found the offence of publishing a false statement ‘likely to cause fear and alarm to the public or to disturb the public peace’ in Tanzania was ‘too vague’ since it ‘does not enable individuals to regulate their conduct.’²¹¹ Similarly, the ECOWAS Court has determined that the Gambian offence of publishing any statement ‘likely to cause fear and alarm to the public or to disturb the public peace, knowing or having reason to believe that the statement, rumour or report is false’ is impermissibly vague, as it is difficult to discern with any certainty what constitutes an offence and it consequently has ‘a chilling effect that may unduly restrict the exercise of freedom of expression of journalists.’²¹²

2. Legitimacy

International law provides a list of objectives that permit the restriction of speech.²¹³ This list of ‘legitimate aims’ in article 19(3) of the ICCPR does not include the prevention of falsehoods, without more.²¹⁴ Instead, article 19(3) provides that speech can only be legitimately restricted when this is necessary for the ‘respect of the rights or reputations of others’ or for ‘the protection of national security or of public order ... or of public health or morals.’ This list is exhaustive.²¹⁵ Human rights law simply ‘does not permit general prohibition of expressions of an erroneous opinion.’²¹⁶ As noted by the UN Special Rapporteur, ‘the right to freedom of expression applies to all kinds of information and ideas ... irrespective of the truth or falsehood of the content.’²¹⁷

Regional human rights treaties contain similar lists.²¹⁸ And even though article 10 of the European Convention on Human Rights enumerates some ‘aims’ that are not listed in the ICCPR and other treaties, the Human Rights Committee has considered ‘public order’ to encompass the additional aims expressly listed in article 10 of the European Convention.²¹⁹ Like UN bodies, the European Court of Human Rights has been clear

²¹¹ EACJ, *Media Council of Tanzania and others v. Attorney General of the United Republic of Tanzania* (Ref no. 2/2017), 28 March 2019, §§92–95. However, the Court accepted that more precisely worded offences passed the test of legality, including criminalizing the use of mass media to publish information ‘maliciously or fraudulently fabricated’ or ‘intentionally or recklessly falsified’ that threatens the ‘interests of defence, public safety, public order, the economic interests of the United Republic, public morality or public health or is injurious to others’ rights and reputation’.

²¹² ECOWAS CCJ, *Federation of African Journalists v. The Gambia* (Suit no. ECW/CCJ/APP/36/15), 13 February 2018, 47.

²¹³ See ch. 1 (Introduction), s II.3.1. (Key convergences).

²¹⁴ ICCPR Art. 19(3).

²¹⁵ *Ibid.*

²¹⁶ HRC, General Comment No. 34 (2011), §49.

²¹⁷ UN Special Rapporteur I. Khan, *Disinformation and freedom of opinion and expression* (2021) UN Doc. A/HRC/47/25, §38, citing HRC, General Comment No. 34 (2011), §§47, 49; ECtHR, *Salov v. Ukraine* (App. no. 65518/01), 6 September 2005, §113.

²¹⁸ See ch. 1 (Introduction), s. II.1.1. (Treaty language).

²¹⁹ See HRC, *Lovell v. Australia* (Comm. no. 920/2000), 29 March 2004, §9.4; HRC, General Comment No. 34 (2011), §31 (‘Contempt of court proceedings relating to forms of expression may be tested against the public order (ordre public) ground’); HRC, *Aleksandrov v. Belarus* (Comm no. 1933/2010), 24 July 2014, §7.4 (holding that Belarus had not explained how the actions of a protester ‘would have violated the rights and freedoms of others or would have posed a threat to public safety or public order (ordre public)’); COE, *Report of the Committee of Experts on Human Rights to the Committee of Ministers* (1970) H(70)7, 2, 175–77. See ch. 1 (Introduction).

that ‘[a]rticle 10 of the Convention as such does not prohibit discussion or dissemination of information . . . even if it is strongly suspected that this information might not be truthful’.²²⁰ Similarly, in the words of the Inter-American Court of Human Rights, ‘[a] system that controls the right of expression in the name of a supposed guarantee of the correctness and truthfulness of the information that society receives can be the source of great abuse and, ultimately, violates the right to information that this same society has.’²²¹

Since international law requires a specific legitimate aim to restrict speech, a legal regime that seeks to enable the state to police the truthfulness of free expression for the sake of the truth alone would be categorically incompatible with international human rights standards. For example, the state has no business restricting the online speech of flat-Earthers or moon-landing conspirators if that speech, however false, does not endanger ‘the rights of others,’ ‘national security’ or other prohibited harms enumerated in article 19(3). But the state might be justified in limiting the speech of anti-vaccination activists that foreseeably caused harms to public health.

2.1. States must not restrict false speech for ulterior purposes

When assessing whether a law is pursuing a legitimate aim, ‘[r]estrictions must be applied only for those purposes for which they were prescribed and must be directly related to the specific need on which they are predicated.’²²² If state authorities articulate a rationale for restricting speech that is legitimate on paper but in practice is merely a pretext for clamping down on critical speech, this will not pass muster under international law.²²³ The Human Rights Committee has held, as a result, that the ‘penalization of a media outlet, publisher or journalist solely for being critical of the government or the political social system espoused by the government can never be considered to be a necessary restriction of freedom of expression.’²²⁴ And even if a legitimate aim for a restriction of false speech exists at the legislative stage, investigative or prosecutorial authorities may well be acting with an improper purpose when executing the law. For example, a state that enacts a misinformation law with the nominal purpose of protecting public health during a pandemic, but uses it to suppress criticism and exercise more control over the media will be judged by its practice—it is the real use and not the pretextual aim that counts.²²⁵

²²⁰ ECtHR, *Salov v. Ukraine* (App. no. 65518/01), 6 September 2005, §113. Cf. partly Concurring Opinion of Judge Mularoni, stating (‘I am not ready to consider that freedom of expression entails the right to disseminate false information’).

²²¹ IACtHR, *Compulsory Membership in an Association Prescribed by Law for the Practice of Journalism* (Arts. 13 and 29 American Convention On Human Rights), Advisory Opinion OC-5/85 (Series A, no. 5), 13 November 1985, §77.

²²² HRC, General Comment No. 34 (2011), §22; ECHR Art. 18; ICCPR Art. 5(1).

²²³ See ACmHPR Special Rapporteur P. Tlakula, *Activity Report of the Special Rapporteur on Freedom of Expression and Access to Information in Africa presented on the 51st Ordinary Session of the African Commission on Human and Peoples’ Rights* (2012), §66.

²²⁴ HRC, General Comment No. 34 (2011), §42. See also HRC, General Comment No. 34 (2011), §23; IACmHR, *The Inter-American Legal Framework Regarding the Right to Freedom of Expression* (2009) OEA/Ser.L/V/II CIDH/RELE/INF. 2/09, §38.

²²⁵ See UN Special Rapporteur, OSCE Representative on Freedom of the Media, OAS Special Rapporteur and ACmHPR Special Rapporteur on Freedom of Expression and Access to Information, *Joint Declaration on Freedom*

Whether state authorities are acting with an ulterior purpose may not be easy to prove, but international human rights bodies have repeatedly highlighted the importance of the inquiry, both in the context of assessing the legitimacy of restrictions of a particular human right and in the context of treaty provisions that prohibit the restriction of human rights for ulterior purposes, such as article 18 of the European Convention.²²⁶ International human rights bodies do not necessarily require direct forms of proof of improper purposes, which could be exceptionally difficult to come by.²²⁷ Rather, they accept indirect and circumstantial evidence, drawing inferences where necessary from contextual factors and patterns of behaviour by state authorities.²²⁸ In that regard, the gross disproportionality of the penalty being imposed may be a telling indicator that the restriction on speech was motivated by an improper purpose.²²⁹

2.2. What objectives can justify the penalization of false speech?

The European Court has been asked to rule on the validity of misinformation laws more frequently than other bodies at the international level.²³⁰ But in assessing laws that penalize speech, the European Court has not always been clear about which objectives are a legitimate basis for restricting false speech. For example, the European Court found that the conviction of a political candidate who disseminated false information about the alleged death of another candidate was for the ‘legitimate purpose’ of ‘providing the voters with true information in the course of [a] presidential campaign’, without specifying which of the aims listed in article 10(2) of the Convention this related to.²³¹ But in a similar case the Human Rights Committee allowed restrictions to expression for the purpose of ‘maintain[ing] the integrity of the electoral process’ specifically on the basis that such restrictions are ‘justified’ both ‘in terms of the protection of public order’ and on the basis that they are ‘necessary for the respect of the rights of others.’²³²

of Expression and ‘Fake News’, Disinformation and Propaganda (2017) FOM.GAL/3/17, Preamble §12; OHCHR, ‘COVID-19: States should not abuse emergency measures to suppress human rights—UN experts’ (16 March 2020): ‘Restrictions taken to respond to the virus must be motivated by legitimate public health goals and should not be used simply to quash dissent.’

²²⁶ See, e.g., ECtHR, *OOO Flavus v. Russia* (App. nos. 12468/15 & others), 23 June 2020, §§37–38 (the European Court finding that Russian authorities blocked websites ‘on spurious grounds or outright arbitrarily’ for the purpose of suppressing dissent). For an overview of cases on ECHR Art. 18, see J. Gavron & R. Remezaita, ‘Has the ECtHR in *Mammadov* 46(4) opened the door to findings of “bad faith” in trials?’ (EJIL:Talk!, 4 July 2019).

²²⁷ See ECtHR (GC), *Merabishvili v. Georgia* (App. no. 72508/13), 28 November 2017, §316: ‘no reason for the Court to restrict itself to direct proof in relation [allegations of ulterior purpose] or to apply a special standard of proof to such allegations.’

²²⁸ See, e.g., ECtHR, *Navalnyy v. Russia (No. 2)* (App. no. 43734/14), 9 April 2019, §§96–98 (the Court relying on ‘converging contextual evidence’ to establish that the restrictions on the applicant’s liberty were implemented with the ulterior purpose of ‘suppress[ing] political pluralism’); ECtHR (GC), *Merabishvili v. Georgia* (App. no. 72508/13), 28 November 2017, §317; ECtHR, *Aliyev v. Azerbaijan* (App. nos. 68762/14 & 71200/14), 20 September 2018, §215.

²²⁹ See, e.g., ECtHR, *Navalnyy v. Russia (No. 2)* (App. no. 43734/14), 9 April 2019, §§96–98.

²³⁰ See, e.g., ECtHR, *Hertel v. Switzerland* (App. no. 25181/94), 25 August 1998, §42.

²³¹ ECtHR, *Salov v. Ukraine* (App. no. 65518/01), 6 September 2005, §110. The candidate was charged with having interfered with citizens’ right to vote for the purpose of influencing election results by means of fraudulent behaviour, and the European Court ultimately held that his sentence of five years’ imprisonment, which was suspended for two years, and fine, as well as annulment by the Bar Association to practice law constituted a ‘very severe penalty’ and a violation of article 10 had arisen. See s. III.2. (Legitimacy).

²³² HRC, *Kim v. Republic of Korea* (Comm. No. 968/2001), 27 July 2005, §8.3 (relating to speech that did not necessarily have to be false); HRC, General Comment No. 34 (2011), 37. See s. II.1.2. (Harm). For an example of a

A challenging area for international human rights bodies has been the assessment of speech denying the occurrence of specific historical events, in particular the existence of and details about the Holocaust.²³³ The Human Rights Committee has stated unequivocally that '[l]aws that penalize the expression of opinions about historical facts are incompatible with the obligations that the Covenant imposes on States parties in relation to the respect for freedom of opinion and expression.'²³⁴ But this goes against the jurisprudence of the European Court, which has held that there is a 'category of clearly established historical facts—such as the Holocaust—whose negation or revision would be removed from the protection of article 10 by article 17.'²³⁵ Article 17 provides that the European Convention may not be interpreted as implying the right to engage in any act 'aimed at the destruction of the rights and freedoms' contained in the treaty.²³⁶ Statements about the Holocaust that are not true have been deemed 'one of the most serious forms of racial defamation of Jews and of incitement to hatred of them' and ineligible for protection under any circumstances under the terms of article 17 for this reason.²³⁷

The European Court was, however, unwilling to extend this reasoning to the case of a politician who denied the Armenian genocide, even though the Court accepted the 'rights of Armenians to respect for their and their ancestors' dignity, including their right to respect for their identity constructed around the understanding that their community has suffered genocide.'²³⁸ This was because the Court reasoned that the justification for making the denial of the Holocaust a criminal offence lies 'not so much in that it is a clearly established historical fact but in that, in view of the historical context in the States concerned [being Austria, Belgium, Germany and France] ... its denial, even if dressed up as impartial historical research must invariably be seen as connoting an antidemocratic ideology and anti-Semitism.'²³⁹

In this way false speech about the Holocaust has been given a unique status by the European Court. While this is perhaps understandable given the historical context, it is nevertheless inconsistent with international human rights standards related to freedom of expression and inconsistent with the Court's own related case law.²⁴⁰ But even the

law prohibiting certain types of speech to maintain the integrity of elections: see, e.g., Canadian Elections Act 2000 s. 91(1).

²³³ See generally U. Belavusau & A. Gliszczyńska-Grabias (eds), *Law and Memory: Towards Legal Governance of History* (CUP 2017). See ch. 3 (Hate Speech), s. III.2.4.2. (Opinion).

²³⁴ HRC, General Comment No. 34 (2011), §49 (reversing a previous decision of the Committee allowing a conviction under denial laws). See ch. 3 (Hate Speech), s. III.2.4.2. (Opinion).

²³⁵ ECtHR, *Garaudy v. France* (App. no. 65831/01), 24 June 2003, 28, citing ECtHR, *Lehideux and Isorni v. France* (App. no. 24662/94), 23 September 1998, §47.

²³⁶ See ch. 3 (Hate Speech), s. III.2.4.2. (Opinion); see, e.g., ECtHR, *Williamson v. Germany* (App. no. 64496/17), 8 January 2019; ECtHR, *Pastörs v. Germany* (App. no. 55225/14), 3 October 2019; ECtHR, *M'Bala M'Bala v. France* (App. no. 25239/13), 20 October 2015.

²³⁷ ECtHR, *Garaudy v. France* (App. no. 65831/01), 24 June 2003, 29; ECtHR, *M'Bala M'Bala v. France* (App. no. 25239/13), 20 October 2015, §40 ('the blatant display of a hateful and anti-Semitic position disguised as an artistic production is as dangerous as a head-on and abrupt attack').

²³⁸ ECtHR (GC), *Perinçek v. Switzerland* (App. no. 27510/08), 15 October 2015, §227.

²³⁹ *Ibid.*, §243.

²⁴⁰ *Ibid.*; HRC, General Comment No. 34 (2011), §49. See ch. 3 (Hate Speech), s. III.2.4.2. (Opinion). Cf. ACtHPR, *Umuhoza v. Rwanda* (App. no. 003/2014), 24 November 2017, §158.

European Court’s approach to Holocaust denial can be regarded as supporting the basic proposition that false speech cannot be criminalized simply on the basis of its falsity.²⁴¹ Rather, such speech can be punished only because of the specific and serious harms that it causes, with the required degree of causal connection. Our survey of international jurisprudence indicates that the following harms might, under the right conditions, reasonably justify criminalization of false speech: if it incites violence, causes other serious harms to human life and health, or directly and immediately harms the integrity of elections and other democratic processes, thereby undermining ‘public order’ and the ‘rights of others’ whose protection is a ‘legitimate aim’ under international law.

3. Necessity

Any restriction on speech needs to be necessary and proportionate to the harm it is seeking to guard against.²⁴² A necessary restriction on speech is more than one that is merely useful, reasonable or desirable; it can only be justified if other, less restrictive forms of limitations of free expression are incapable of remedying the harm.²⁴³ And in a free society, the default remedy for falsehoods is more truthful speech, not penalties imposed by the state.²⁴⁴

But harms caused by misinformation cannot always be remedied—at least not fully—through more truthful speech, particularly when the target of misinformation cannot easily and speedily reach the same audience, or where any correction would be too late to address the damage, as is often the case in the online context.²⁴⁵ For instance, misinformation that causes direct harms to human health—such as the peddling of cures for an infectious disease that are not only false, but lethal—can lead to many deaths before it is successfully countered by truthful speech.²⁴⁶

A state should be expected to provide a full accounting of proactive measures that it took, or at least considered, to mitigate harms before resorting to restrictions on free expression.²⁴⁷ The state must also demonstrate that the necessity for speech regulation is continuing; that there is an *ongoing* need for a restriction on speech due to continuing harms caused by misinformation that cannot be remedied through less restrictive means. It is perfectly possible for certain speech-restrictive measures to be necessary

²⁴¹ See, e.g., ECtHR, *Pastörs v. Germany* (App. no. 55225/14), 3 October 2019, §§40–42.

²⁴² ECtHR, *Lingens v. Austria* (App. no. 9815/82), 8 July 1986, §39: ‘The adjective “necessary” . . . implies the existence of a “pressing social need”’. See also IACtHR, *Compulsory Membership in an Association Prescribed by Law for the Practice of Journalism (Arts. 13 and 29 American Convention On Human Rights)*, Advisory Opinion OC-5/85 (Series A, no. 5), 13 November 1985, §46 (endorsing this approach).

²⁴³ See HRC, General Comment No. 34 (2011), §§33, 35; IACtHR, *Compulsory Membership in an Association Prescribed by Law for the Practice of Journalism (Arts. 13 and 29 American Convention On Human Rights)*, Advisory Opinion OC-5/85 (Series A, no. 5), 13 November 1985, §46. See also ECtHR, *The Sunday Times v. United Kingdom* (App. no. 6538/74), 26 April 1979, §59.

²⁴⁴ See UN Special Rapporteur F. La Rue, *Promotion and protection of the right to freedom of opinion and expression* (2011) UN Doc. A/66/290, §41.

²⁴⁵ See Z. Tufekci, ‘It’s the (Democracy-Poisoning) Golden Age of Free Speech’ (Wired, 16 January 2018).

²⁴⁶ UN Special Rapporteur D. Kaye, *Disease pandemics and the freedom of opinion and expression* (2020) UN Doc. A/HRC/44/49, §45.

²⁴⁷ *Ibid.*, §47.

only for a limited time (such as for the duration of a pandemic), but then no longer because the circumstances that once generated the necessity have changed.²⁴⁸

Most cases analysing the permissibility of state regulation of misinformation turn on the question of necessity and proportionality of the law as applied.²⁴⁹ And international standards dictate that the necessity and proportionality of the penalization of misinformation depend on a number of factors: (1) the severity and imminence of the harm caused by the untruthful speech and the causal link between the untruthful speech and the harm; (2) the intent of the speaker; and (3) the penalty that is imposed.²⁵⁰

3.1. Harm

3.1.1. Severity and imminence of harm

States seeking to penalize speech must demonstrate that some degree of a prohibited harm is likely to result from it.²⁵¹ The Human Rights Committee has found that ‘[w]hen a State . . . invokes a legitimate ground for restriction of freedom of expression, it must demonstrate in specific and individualized fashion the precise nature of the threat’ posed by the speech.²⁵² But the required severity and imminence of the harm that is sufficient to justify penalties for misinformation has not yet been clearly articulated by international bodies.

Although there is no clear case law from the Human Rights Committee on requirements for harm in the context of misinformation laws, the Committee has found a violation of freedom of expression in the context of defamation where the damage to a public servant’s reputation was ‘only of a limited nature.’²⁵³ And the Committee has stated that criminal defamation laws should only be countenanced in the ‘most serious of cases.’²⁵⁴ In addition, some members of the Human Rights Committee have suggested, in the context of mandatory restrictions on hate speech, that the application of criminal penalties should be limited to ‘speech that incites the commission of criminal offences or acts of violence.’²⁵⁵

²⁴⁸ *Ibid.*, §15 (‘the ensuing interference with third parties’ rights must be limited and justified in the light of the interest supported by the intrusion’).

²⁴⁹ See ch. 1 (Introduction), s. II.3.1.1. (Justifications for penalizing speech).

²⁵⁰ See HRC, General Comment No. 34 (2011), §35; ECtHR, *Mariya Alekhina and Others v. Russia* (App. no. 38004/12), 17 July 2018, §220 (the Court noting the importance of whether the speech had the ‘capacity—direct or indirect—to lead to harmful consequences’).

²⁵¹ HRC, General Comment No. 34 (2011), §35. See s.III.2.2 (What objectives can justify the penalization of false speech?).

²⁵² *Ibid.*, §35, citing HRC, *Shin v. Republic of Korea* (Comm. no. 926/2000), 16 March 2004 (where a professional painter of a political artwork titled ‘Rice Planting’ who was convicted under Korean law for ‘enemy-benefiting expression’, the Committee held that Korea must demonstrate ‘in specific fashion the precise nature of the threat’ caused by the author’s conduct ‘as well as why seizure of the painting and the author’s conviction were necessary’. The Committee found a violation of article 19(2) in the absence of an ‘individualized justification’). Cf. HRC, *Ross v. Canada* (Comm. no. 736/1997), 18 October 2000, §11.5 (noting that ‘the [Canadian] Supreme Court found that it was reasonable to anticipate that there was a causal link between the expressions of the author and the “poisoned school environment” experienced by Jewish children’).

²⁵³ HRC, *Kozlov v. Belarus* (Comm. no. 1986/2010), 24 July 2014, §7.6 (in circumstances where the defamatory remarks were sent in a letter rather than made public).

²⁵⁴ HRC, General Comment No. 34 (2011), §47. See s.III.5. (Penalties) (noting the Human Rights Committee’s position that imprisonment is ‘never an appropriate penalty’ for defamatory speech).

²⁵⁵ See, e.g., HRC, *Rabbae v. The Netherlands* (Comm. no. 2124/2011) 14 July 2016, Individual Concurring Opinion of Committee members Sarah Cleveland and Mauro Politi, §7. See also the standard for mandatory

The harm threshold applied by the European Court depends on the speech at issue. In the context of hate speech, the Court will consider whether speech ‘could be seen as a direct or indirect call for violence or as a justification of violence, hatred or intolerance’ as a factor relevant to the assessment of whether a penalty was necessary.²⁵⁶ In order to trigger the protection of the Convention for defamatory speech, the Court requires that an attack on reputation must ‘attain a certain level of seriousness and be in a manner causing prejudice to personal enjoyment of the right to respect for private life.’²⁵⁷ When false speech falls into either of these categories these standards will therefore likely apply.

In the handful of cases in which the European Court has addressed misinformation laws, it has considered the impact of the impugned speech in its assessment of necessity, but it has not articulated a specific harm threshold. For example, the Court considered misinformation when a political candidate in a presidential election disseminated forged copies of a newspaper which suggested—incorrectly—that his opponent had died.²⁵⁸ The candidate was convicted of interfering with the exercise of citizens’ electoral rights and sentenced to five years’ imprisonment.²⁵⁹ In assessing this penalty to be disproportionate, the Court took into account the potential harm of the information, noting ‘the impact of the information contained in the newspaper was minor as he only had eight copies of the forged . . . newspaper and spoke to a limited number of persons about it, a fact that should have been taken into account by the domestic courts.’²⁶⁰ And the Court invoked similar reasoning in a case involving speech that was published in a specialized scientific journal ‘distributed almost entirely by subscription’ with ‘in all likelihood, a specific readership such that the impact of the ideas it contains should be limited.’²⁶¹

The Inter-American Court has traditionally been a speech-protective jurisdiction,²⁶² and emphasized the importance of ‘promoting information pluralism’ rather than a system that controls speech ‘in the name of a supposed guarantee of the correctness and truthfulness’ of certain information.²⁶³ This position has translated to stringent harm requirements before false speech can be penalized. For instance, the Court has criticized slander laws that do not ‘specify the injury required’ and made clear that

non-criminal penalties set out in ICCPR Art. 20. See ch. 1 (Introduction), s. II.1.2.1.2. (ICCPR Article 20) and ch. 3 (Hate Speech), s. III.1.1. (ICCPR Article 20).

²⁵⁶ ECtHR (GC), *Perinçek v. Switzerland* (App. no. 27510/08), 15 October 2015, §§206, 208. See ch. 3 (Hate Speech), s. II.1.2. (Harm). See ch. 6 (Terrorism Laws) ss. II.1.3 (Harm).

²⁵⁷ ECtHR (GC), *Axel Springer AG v. Germany* (App. no. 39954/08), 7 February 2012, §83; ECtHR, *A v. Norway* (App. no. 28070/06), 9 April 2009, §64 (both in the context of Art. 8, not 10).

²⁵⁸ ECtHR, *Salov v. Ukraine* (App. no. 65518/01), 6 September 2005.

²⁵⁹ *Ibid.*, §28. See s. III.2. (Legitimacy).

²⁶⁰ ECtHR, *Salov v. Ukraine* (App. no. 65518/01), 6 September 2005, §114.

²⁶¹ ECtHR, *Hertel v. Switzerland* (App. no. 25181/94), 25 August 1998, §49.

²⁶² See, e.g., IACtHR, *Compulsory Membership in an Association Prescribed by Law for the Practice of Journalism* (Arts. 13 and 29 American Convention On Human Rights), Advisory Opinion OC-5/85 (Series A, no. 5), 13 November 1985, §50.

²⁶³ IACtHR, *Tristán Donoso v. Panamá* (Series C, no. 193), 27 January 2009, §113; IACtHR, *Compulsory Membership in an Association Prescribed by Law for the Practice of Journalism* (Arts. 13 and 29 American Convention On Human Rights), Advisory Opinion OC-5/85 (Series A, no. 5), 13 November 1985, §77.

defamatory speech must constitute a ‘serious attack’ or do ‘serious damage.’²⁶⁴ Article 13(5) of the American Convention on Human Rights requires that speech rises to the level of incitement to ‘lawless violence’ or ‘similar action’ when it comes to mandatory penalties for hateful speech.²⁶⁵ And the Rapporteur of the Organization of American States (OAS Rapporteur) has indicated that any restriction to speech ‘must be based on real and objectively verifiable causes that present the certain and credible threat of a potentially serious disturbance of the basic conditions for the functioning of democratic institutions.’²⁶⁶ This jurisdiction is therefore likely to find that misinformation laws require that serious potential harm would likely flow or did flow from the speech.

Although African courts have reviewed a number of misinformation laws, they have generally done so with respect to the ‘legality’ requirement, without analysing necessity and the requisite harm thresholds.²⁶⁷ But the African Court on Human and Peoples’ Rights has noted that custodial sentences for speech are contrary to freedom of expression ‘[a]part from [in] serious and very exceptional circumstances, for example incitement to international crimes, public incitement to hatred [or] discrimination or violence or threats against a person or a group of people, because of a specific criteria.’²⁶⁸ Criminal misinformation laws are therefore not likely to be compatible with article 9 of the African Charter on Human and Peoples’ Rights unless they meet these conditions.²⁶⁹

In sum, international jurisprudence supports the position that criminalizing false speech can only be done if it causes serious proscribed harms. The necessity of speech restriction, especially through criminal means, will depend on the nature, degree and imminence of the harm produced by the false speech. For example, it might be necessary for the state to restrict the speech of an individual who is promoting a false cure for Covid-19, which may directly lead to harms to the lives or health of the audience.²⁷⁰ But it would arguably not be necessary for the state to restrict the speech of climate change deniers.²⁷¹ Even if that speech is equally untruthful, and although in aggregate

²⁶⁴ IACtHR, *Usón Ramírez v. Venezuela* (Series C, no. 207), 20 November 2009, §§56, 73–75.

²⁶⁵ ACHR Art. 13(5). Though the scope of ‘similar action’ is not defined, the drafting of the text suggests that the harm should be comparable in severity to violence: see ch. 3 (Hate Speech), s. III.1.4. (Mandatory restrictions at the regional level).

²⁶⁶ IACmHR, *Freedom of Expression Standards for Free and Inclusive Broadcasting* (2009) OEA/Ser.L/V/II CIDH/RELE/INF. 3/09, §22, citing IACmHR, *Annual Report of the Office of the Special Rapporteur for Freedom of Expression* (2008) OEA/Ser.L/V/II. 134, Chapter 3, §75.

²⁶⁷ See s. II.1.1. (Vagueness and overbreadth). See, e.g., EACJ, *Media Council of Tanzania and others v. Attorney General of the United Republic of Tanzania* (Ref no. 2/2017), 28 March 2019; ECOWAS CCJ, *Federation of African Journalists v. The Gambia* (Suit no. ECW/CCJ/APP/36/15), 13 February 2018.

²⁶⁸ ACtHPR, *Konaté v. Burkina Faso* (App. no. 004/2013), 5 December 2014, §165.

²⁶⁹ *Ibid.*, §§165–166.

²⁷⁰ Cf. The Meta Oversight Board, ‘Case Decision 2020-006-FB-FBR Claimed COVID-19 cure’ (finding Facebook’s decision to remove a post criticizing French authorities for refusing to authorize hydroxychloroquine combined with azithromycin as a cure for Covid-19 was inconsistent with article 19(3) of the ICCPR. The Board found that Facebook had failed to demonstrate how removing the post was the least intrusive method of protecting public health despite having a range of enforcement options that deal with misinformation through less extreme means, and considered that Facebook’s rules on misinformation also failed the legality test).

²⁷¹ Cf. Private corporations banning climate misinformation content and ads on their platforms: C. Duffy, ‘Pinterest bans misinformation about climate change’ (CNN, 6 April 2022); R. Maruf, ‘Twitter bans “misleading” climate change ads’ (CNN, 23 April 2022); K. Chan, ‘Google cracks down on climate change denial by targeting ads’ (AP News, 8 October 2021).

the denials of climate change undoubtedly adversely affect efforts to curb and mitigate it, these harms are on a longer timescale and the causal link between the untruthful speech of any given individual and a concrete harmful consequence is more tenuous. International bodies would therefore likely find that it is more appropriate for this type of misinformation to be addressed through public education and debate, rather than through state-imposed penalties.²⁷²

3.1.2. Causal relationship between the speech and the harm

International human rights law requires that the state establish a causal link between false speech and concrete harms that are likely to result from it.²⁷³ But international bodies have not yet articulated a clear and consistent standard for this.

According to the Human Rights Committee, a state seeking to restrict expression must establish ‘a direct and immediate connection between the expression and the threat.’²⁷⁴ And in the context of incitement to violence, UN-approved standards set out in the ‘Rabat Plan of Action’ provide that in restricting speech ‘the courts will have to determine that there was a reasonable probability that the speech would succeed in inciting actual action against the target group, recognizing that such causation should be rather direct.’²⁷⁵ But there are no detailed requirements about the foreseeability of the harm, whether this test should be objective or subjective, whether the harm should possibly, likely, probably, almost certainly or actually result from the speech, or whether the harm must be imminent.²⁷⁶

The European Court has considered in some cases whether speech had the ‘capacity—direct or indirect—to lead to harmful consequences’; but its jurisprudence does not examine the objective likelihood of a certain result or data on what, in fact, happened following the speech.²⁷⁷ And in some cases involving racist or homophobic speech, the Court has sidestepped an assessment of causation of harm entirely, basing its decision ‘on an assessment of the content of the expression and the manner of its delivery’ alone.²⁷⁸ The 2019 Declaration of Principles of Freedom of Expression and Access to Information in Africa only speak of ‘a close causal link between the risk of harm and the expression’ in the context of restricting speech on public order or national security grounds, without providing further detail.²⁷⁹ And multiple

²⁷² See UN Special Rapporteur J. Knox, *Report of the Special Rapporteur on the issue of human rights obligations relating to the enjoyment of a safe, clean, healthy and sustainable environment* (2016) UN Doc. A/HRC/31/52, §60 ([t]o enable informed public participation, the rights of freedom of expression and association must be safeguarded for all people in relation to all climate-related actions, including for individuals who oppose projects designed to mitigate or adapt to climate change’).

²⁷³ UN Special Rapporteur I. Khan, *Disinformation and freedom of opinion and expression* (2021) UN Doc. A/HRC/47/25, §41.

²⁷⁴ HRC, General Comment No. 34 (2011), §35.

²⁷⁵ Rabat Plan of Action, §29(f) (emphasis added).

²⁷⁶ *Ibid.*

²⁷⁷ ECtHR, *Alekhina v. Russia* (App. no. 38004/12), 17 July 2018, §220 (emphasis added). See also ECtHR (GC), *Perinçek v. Switzerland* (App. no. 27510/08), 15 October 2015, §207. See ch. 3 (Hate Speech).

²⁷⁸ See, e.g. ECtHR, *Lilliendahl v. Iceland* (App. no. 29297/18), 12 May 2020, §36. See ch. 3 (Hate Speech), s. II.1.2. (Harm).

²⁷⁹ ACmHPR, Declaration of Principles of Freedom of Expression and Access to Information in Africa (2019), Principle 22(5).

international bodies assume that harm exists in the context of speech that is found to be defamatory.²⁸⁰

3.2. Intent

International bodies assessing the appropriateness of criminalizing false statements must address a number of questions about the state of mind of the speaker. What level of fault is required for restrictions on speech to be compatible with freedom of expression? Are statements that are false but made in good faith always protected? And do knowing falsehoods enjoy *any* level of protection under international law? At the highest end of the culpability spectrum, a criminal statute would punish a speaker who knew that the information they were disseminating was false (or intended to deceive their audience), and who knew or intended for the false speech to cause a prohibited harm. These would be cases of lies or *disinformation*. At the lowest end of the spectrum, false speech would be a strict liability offence, or one that required only objective negligence.

International standards in relation to defamation law require a high level of culpability to penalize defamatory speech, at least when it concerns public officials.²⁸¹ The Inter-American bodies have adopted the ‘actual malice’ test articulated by the US Supreme Court in *New York Times v. Sullivan*, holding that ‘erroneous statement is inevitable in free debate’ and ‘must be protected if the freedoms of expression are to have the “breathing space” that they “need ... to survive”’ and that defamatory statements about a *public official* are only actionable in a civil defamation claim if made with ‘actual malice’, i.e. ‘with knowledge that it was false or with reckless disregard of whether it was false or not.’²⁸² The Human Rights Committee has similarly recommended that ‘[a]t least with regard to comments about public figures, consideration should be given to avoiding penalizing or otherwise rendering unlawful untrue statements that have been published in error *but without malice*’.²⁸³ And although African courts have not formally adopted the ‘actual malice’ standard,²⁸⁴ the East African Court of Justice

²⁸⁰ For a discussion on the treatment of defamation in international human rights law, see A. Clooney & P. Webb, ‘The Right to Insult in International Law’ (2017) 48(2) *Columbia Human Rights Law Review* 1, 44–46. But, e.g., the Inter-American Court considers causation in determining damages and reparations: see IACtHR, *Tristán Donoso v. Panamá* (Series C, no. 193), 27 January 2009, §120. For speech alleged to have incited violence, the Court requires that the imposition of sanctions ‘must be backed up by actual, truthful, objective and strong proof that the person was not simply issuing an opinion (even if that opinion was hard, unfair or disturbing), but that the person had the clear intention of committing a crime and the actual, real and effective possibility of achieving this objective’: IACmHR, *The Inter-American Legal Framework Regarding the Right to Freedom of Expression* (2009) OEA/Ser.L/V/II CIDH/RELE/INF. 2/09, §58. See ch. 3 (Hate Speech).

²⁸¹ See ch. 2 (Insulting Speech), s. III.3.1. (Intent of the speaker).

²⁸² US Supreme Court, *New York Times v. Sullivan* 376 U.S. 254, 9 March 1964, 271–272, 279–280; IACmHR, Declaration of Principles on Freedom of Expression (2000), Principle 10; IACtHR, *Tristán Donoso v. Panamá* (Series C, no. 193), 27 January 2009, §§119–120, 125, 134. See also IACtHR, *Usón Ramírez v. Venezuela* (Series C, no. 207), 20 November 2009, §56; IACtHR, *Mémoli v. Argentina* (Series C, no. 265), 22 August 2013, §§139, 141.

²⁸³ HRC, General Comment No. 34 (2011), §47 (emphasis added). See also HRC, *Akhmedyarov v. Kazakhstan* (Comm. no. 2535/2015), 23 July 2020, §9.8; HRC, *Adonis v. The Philippines* (Comm no. 1815/2008) 26 October 2011, §7.9; HRC, *Concluding observations: The United Kingdom of Great Britain and Northern Ireland* (2008) UN Doc. CCPR/C/GBR/CO/6, §25 (recommending that the United Kingdom ‘consider the utility of a so-called “public figure” exception, requiring proof by the plaintiff of actual malice in order to go forward on actions concerning reporting on public officials and prominent public figures’).

²⁸⁴ See ch. 2 (Insulting Speech), s. III.3.1. (Intent of the speaker).

considered misinformation offences in Tanzania that imposed high *mens rea* requirements, such as malice and actual knowledge, to be 'largely unobjectionable'.²⁸⁵

The European Court's assessment of the requisite *mens rea* is different. In defamation cases, the Court has declined to adopt the 'actual malice' standard.²⁸⁶ And in hate speech cases it has considered intent to be a relevant factor, but has allowed even criminal penalties without a clear showing of intent.²⁸⁷

Outside the context of defamation, a difficult question is whether knowing falsehoods (*disinformation*) should enjoy *any* level of protection. Normatively, the question is whether lies possess any intrinsic value of their own that would make them worthy of protection, or whether they should be protected only incidentally or instrumentally, in order to prevent chilling effects on speech that is either truthful or is untruthful but uttered in good faith.²⁸⁸ As a matter of principle the question is also whether the value of outright lies should be measured against their motives, against their consequences, or in another way.

International human rights bodies have not yet directly addressed the level of protection to be given to knowing falsehoods.²⁸⁹ British courts have commented that '[t]here is no human right to disseminate information that is not true' since 'the working of a democratic society depends on the members of that society being informed not misinformed'; and suggested that such speech could not be protected under article 10 of the European Convention.²⁹⁰ But they also confirm that '[s]ome degree of tolerance for factual inaccuracy has to be accepted', including in the context of defamation laws.²⁹¹

In the United States, the degree of protection to be afforded to false speech has been at issue in the defamation context.²⁹² However, decades after *Sullivan*, the Supreme

²⁸⁵ EACJ, *Media Council of Tanzania and others v. Attorney General of the United Republic of Tanzania* (Ref no. 2/2017), 28 March 2019, §94.

²⁸⁶ ECtHR, *McVicar v. United Kingdom* (App. no. 46311/99), 7 May 2002, §§27, 65, 73; ECtHR, *Kasabova v. Bulgaria* (App. no. 22385/03), 19 April 2011, §§48, 63–68; see ch. 2 (Insulting Speech), s. III.3.1. (Intent).

²⁸⁷ ECtHR, *Leroy v. France* (App. no. 36109/03), 2 October 2008, §42; ECtHR, *Nix v. Germany* (App. no. 35285/16), 13 March 2018, §54; see ch. 3 (Hate Speech), s. III.2.3.2. (Intent).

²⁸⁸ See in particular the different opinions of the justices of the US Supreme Court, *United States v. Alvarez* 567 U.S. 709, 28 June 2012, discussed later in this subsection.

²⁸⁹ See ACmHPR, Declaration of Principles on Freedom of Expression and Access to Information in Africa (2019), Principle 21(1) (holding that 'no one shall be liable for true statements, opinions or statements' regarding public figures 'which are reasonable to make in the circumstances', without expressing a view on knowingly false statements). Cf. HRC, *Concluding Observations: Cameroon* (1999) UN Doc. CCPR/C/79/Add.116, §24 (expressing deep concern at 'the prosecution and punishment of journalists for the crime of publication of false news merely on the ground, without more, that the news was false, in clear violation of article 19 of the Covenant').

²⁹⁰ See UK House of Lords, *Reynolds v. Times Newspapers Ltd* [2001] 2 AC 127, 28 October 1999, 238 (Lord Hobhouse): 'The liberty to communicate (and receive) information has a similar place in a free society but it is important always to remember that it is the communication of information not misinformation which is the subject of this liberty. There is no human right to disseminate information that is not true. No public interest is served by publishing or communicating misinformation ... These are general propositions going far beyond the mere protection of reputations'. See also UK High Court of England and Wales, Queen's Bench Division, *R (Woolas) v. Parliamentary Election Court* [2010] EWHC 3169 (Admin), 3 December 2010, §§105–106 (per Thomas LJ): 'Article 10 was not engaged in relation to the statements in *The Examiner* and *Labour Rose* which the Election Court found were made dishonestly. Dishonest statements are aimed at the destruction of the rights of the public to free elections ... and the right of each candidate to his reputation ... Article 10 does not protect a right to publish statements which the publisher knows to be false ... or does not believe it to be true ... The right to freedom of expression under Article 10 does not extend to a right to be dishonest and tell lies.'

²⁹¹ See UK House of Lords, *Reynolds v. Times Newspapers Ltd* [2001] 2 AC 127, 28 October 1999, 238 (Lord Hobhouse).

²⁹² See ch. 2 (Insulting Speech), s. III.3.1. (Intent).

Court split rather fundamentally on the level of protection to be given to knowing falsehoods in the case of *Alvarez*, concerning the constitutionality of the Stolen Valor Act, which made it a crime to falsely claim receipt of military decorations or medals.²⁹³ This was, in other words, a criminal law prohibiting a very specific type of disinformation. *Alvarez* is particularly instructive regarding the question of whether *lies*, i.e. ‘disinformation’ or knowingly false speech, should enjoy *any* protection.

By six votes to three the Supreme Court declared the statute unconstitutional, but was divided in doing so. Writing for a plurality of the Court, Justice Kennedy ruled that the First Amendment recognized no category of knowing falsehood that was completely excluded from protection.²⁹⁴ Subjecting the statute to strict scrutiny, he held further that a criminal prohibition on uttering falsehoods in this particular context lacked a sufficient causal link to the government’s interest in protecting the integrity of military honours, and that the government’s purpose could have been achieved through less restrictive means, specifically through more truthful speech.²⁹⁵ In his concurrence, Justice Breyer employed an intermediate level of scrutiny²⁹⁶ and similarly thought that lies were not categorically excluded from protection.²⁹⁷ In particular, he thought that the statute was overly broad and that it lacked limiting features that would confine it to a ‘subset of lies where specific harm is more likely to occur’.²⁹⁸

In his dissent, Justice Alito endorsed a categorical exception from First Amendment protection for knowing falsehoods, which in his view had no intrinsic value.²⁹⁹ Such falsehoods could, however, still be protected incidentally if the relevant statute created significant chilling effects on speech otherwise entitled to protection.³⁰⁰ In his view that was not the case with the Stolen Valor Act, which was precise and dealt with a specific issue.³⁰¹ However, Justice Alito concluded that the danger of chilling effects on debate on matters of public interest warranted a categorical rule. In his words:

there are broad areas in which any attempt by the state to penalize purportedly false speech would present a grave and unacceptable danger of suppressing truthful speech. Laws restricting false statements about philosophy, religion, history, the social sciences, the arts, and other matters of public concern would present such a threat. The point is not that there is no such thing as truth or falsity in these areas or that the truth is always impossible to ascertain, but

²⁹³ US Supreme Court, *United States v. Alvarez* 567 U.S. 709, 28 June 2012.

²⁹⁴ *Ibid.*, 718–722 (Kennedy J, whom Roberts CJ and Ginsburg and Sotomayor JJ joined).

²⁹⁵ *Ibid.*, 724–730.

²⁹⁶ *Ibid.*, 730–732 (Breyer J, whom Kagan J joined).

²⁹⁷ *Ibid.*, 733: ‘False factual statements can serve useful human objectives, for example: in social contexts, where they may prevent embarrassment, protect privacy, shield a person from prejudice, provide the sick with comfort, or preserve a child’s innocence; in public contexts, where they may stop a panic or otherwise preserve calm in the face of danger; and even in technical, philosophical and scientific contexts, where (as Socrates’ methods suggest) examination of a false statement (even if made deliberately to mislead) can promote a form of thought that ultimately helps realize the truth.’

²⁹⁸ *Ibid.*, 736.

²⁹⁹ *Ibid.*, 746–750 (Alito J, whom Scalia and Thomas JJ joined).

³⁰⁰ *Ibid.*, 751.

³⁰¹ *Ibid.*, 752: ‘In stark contrast to hypothetical laws prohibiting false statements about history, science, and similar matters, the Stolen Valor Act presents no risk at all that valuable speech will be suppressed. The speech punished by the Act is not only verifiably false and entirely lacking in intrinsic value, but it also fails to serve any instrumental purpose that the First Amendment might protect.’

rather that it is perilous to permit the state to be the arbiter of truth ... the potential for abuse of power in these areas is simply too great.³⁰²

In sum, four justices thought that knowing falsehoods were not substantially different from truthful speech in terms of the level of protection that should be afforded. Two justices thought they were still protected, but less valuable and therefore subject to a less searching scrutiny in assessing justifications for restricting it. Three justices thought that knowing falsehoods should not enjoy any protection except to the extent necessary to prevent chilling effects on truthful speech in certain topics of public concern.

The European Court is yet to authoritatively rule on the degree of protection to be afforded to knowingly false speech.³⁰³ But it is unlikely that the Court or other international human rights bodies will find lies to be categorically excluded from protection; such an approach has so far been reserved for very extreme types of expression, such as child pornography³⁰⁴ and, in some European Court cases, for Holocaust denial.³⁰⁵ In particular it seems difficult to make such an exclusion regardless of the nature of the harmful consequences of knowingly false expression, and regardless of whether these consequences were intended or not. At least theoretically, misinformation, whether through the withholding of accurate information or through outright lies, could in some cases be motivated by a desire to *prevent* socially harmful consequences, such as panic or disorder.³⁰⁶ And disinformation may not do any harm. But deliberate or knowing falsehoods would presumably enjoy a lesser degree of protection than an innocent, negligent or reckless falsehood.³⁰⁷

This prediction is supported by the limited existing jurisprudence. For example, in a case against Ukraine in which an individual was convicted for spreading misinformation that a presidential candidate in an election had died and was being impersonated, the European Court found the conviction disproportionate in part because:

[t]he domestic courts failed to prove that he was *intentionally trying to deceive* other voters and to impede their ability to vote during the 1999 presidential elections. Furthermore, Article 10 of the Convention as such does not prohibit discussion or dissemination of information received

³⁰² Ibid., 751–752.

³⁰³ See European Parliament, *The fight against disinformation and the right to freedom of expression* (2021), 24–26. Cf. ch. 6 (Terrorism), s. III.3.3 (Intent) (regarding the European Union's decision to sanction Russia state media on the basis that Russian media outlets were broadcasting 'massive propaganda and disinformation').

³⁰⁴ See IACmHR, *The Inter-American Legal Framework Regarding the Right to Freedom of Expression* (2009) OEA/Ser.L/V/II CIDH/RELE/INF. 2/09, §60. For an extensive analysis of whether pornography generally should be protected by the freedom of expression, see E. Barendt, *Freedom of Speech* (2nd edn, OUP 2007), 352–391.

³⁰⁵ See s. III.2.2. (What objectives can justify the penalization of false speech?) and ch. 3 (Hate Speech), s. III.2.2. (Legitimacy).

³⁰⁶ See, e.g., US Supreme Court, *United States v. Alvarez* 567 U.S. 709, 28 June 2012, 733: 'False factual statements can serve useful human objectives, for example ... in public contexts, where they may stop a panic or otherwise preserve calm in the face of danger.'

³⁰⁷ See IACmHR, *The Inter-American Legal Framework Regarding the Right to Freedom of Expression* (2009) OEA/Ser.L/V/II CIDH/RELE/INF. 2/09, §113: 'even if the stated facts (for example, the commission of a crime) cannot be proven in a trial, the individual who made the statements in question will be protected as long as he or she had no prior knowledge of the falsity of the statement or did not act with grave negligence (total disregard for the truth)'. See also J. Rowbottom, 'Lies, Manipulation and Elections—Controlling False Campaign Statements' (2012) 32 *Oxford Journal of Legal Studies* 507, 521.

even if it is strongly suspected that this information might not be truthful. To suggest otherwise would deprive persons of the right to express their views and opinions about statements made in the mass media and would thus place an unreasonable restriction on the freedom of expression set forth in Article 10 of the Convention.³⁰⁸

To conclude, the culpability element of a misinformation offence is one of the most important factors in assessing its proportionality and by extension its compliance with ‘necessity’ requirements in international law. Offences that do not incorporate demanding *mens rea* standards, such as a requirement of knowing falsehood or intention to deceive, are less likely to satisfy this test.

4. Exclusions, Exceptions and Defences

Many legal systems consider defences as simply a negation of one of the constituent elements of a crime.³⁰⁹ Others may provide some additional, affirmative grounds which a defendant may raise in response to particular claim or charge.³¹⁰ Possible defences to misinformation offences, in addition to lack of proof of a necessary element, include: (1) truth, since the factual predicate for any such offence is that the information at issue was false;³¹¹ (2) opinion, on the basis that at least some such statements cannot be true or false. In addition: (3) the political or ‘public interest’ subject-matter of the speech; and, according to the European Court, (4) ‘responsible journalism’ by a ‘public watchdog’ may mean that speech either cannot be sanctioned or will enjoy a higher degree of protection.

4.1. Truth

International law provides that truth should be an absolute defence in the context of defamation laws.³¹² But truth and falsity may be a matter of degree. The European Court in particular has recognized that ‘substantial truth’, as opposed to ‘exact truth’ should be a sufficient basis to avoid liability in civil defamation cases.³¹³ For example, in a case involving a newspaper that published accounts of patients who were dissatisfied with their cosmetic surgery, the Court found a violation of article 10 by placing ‘considerable weight’ on the fact that the accounts were ‘essentially correct’.³¹⁴ The Court has

³⁰⁸ ECtHR, *Salov v. Ukraine* (App. no. 65518/01), 6 September 2005, §113 (emphasis added).

³⁰⁹ See s.II.1.4 (Exclusions; exceptions and defences). See, e.g., W. R. LaFave, *Criminal Law* (5th edn, West Academic Publishing 2010), 469, §9.1.

³¹⁰ *Ibid.*

³¹¹ See Colombian Constitutional Court, *Tutela judgment T-293/2018* 8 May 2017, §§5.1–5.3 (extensively discussing the *exceptio veritatis* as a general defence in criminal law).

³¹² See ch. 2 (Insulting Speech), s. III.4.1. (Truth). See, e.g., HRC, General Comment No. 34 (2011), §47; HRC, *Adonis v. The Philippines* (Comm. no. 1815/2008), 26 October 2011, §7.9; ECtHR, *Castells v. Spain* (App. no. 11798/85), 23 April 1992, §48; ECtHR, *Colombani v. France* (App. No. 51279/99), 25 June 2002, §66; IACmHR, Declaration of Principles on Freedom of Expression (2000), Principle 7; ACmHPR, Declaration of Principles on Freedom of Expression and Access to Information in Africa (2019), Principles 21.

³¹³ See, e.g., ECtHR, *Bergens Tidende v. Norway* (App. no. 26132/95), 2 May 2000, §56.

³¹⁴ *Ibid.*

generally required that there is a ‘sufficiently accurate and reliable factual basis proportionate to the nature and degree of the . . . statements and allegations.’³¹⁵

International bodies have also found that an honest or reasonable belief in the truth of a statement should be a defence in contexts beyond defamatory speech. For example the authenticity of leaked official secrets was a key aspect of the leading European Court case of *Bucur and Toma v. Romania*, which considered disclosure of an unlawful wiretapping operation by a whistleblower in Romania.³¹⁶ The Court ultimately found a violation of the right to freedom of expression, noting the whistleblower ‘had reasonable grounds to believe that the information disclosed was true.’³¹⁷ Soft law instruments regarding unauthorized disclosures also recommend a defence of truth.³¹⁸ And the UN Special Rapporteur has recommended that no one should be penalized under hate speech laws ‘for statements that are true.’³¹⁹ In the context of misinformation laws, establishing that a statement is not true should be proved by a prosecutor as an element of any criminal offence.

4.2. Opinion

Since the *Lingens* judgment of the European Court,³²⁰ international human rights bodies have repeatedly affirmed the distinction between statements of opinion and fact, holding that only the latter can be restricted on the basis of their falsity.³²¹

That case law was principally developed in the defamation context,³²² but it can be applied to the regulation of false statements more broadly. And it is undoubtedly correct as a matter of principle. There *is* a fundamental difference between a statement claiming that Barack Obama was a bad president and one claiming that he was born in Kenya, and therefore ineligible to be president. There is a difference between a claim that Christianity is the best of all religions and an assertion that the Earth is flat. There is a difference between a claim that the state should not impose lockdowns and one that the novel coronavirus has not led to any deaths. Only the latter are statements of fact that are capable of objective proof.

³¹⁵ ECtHR, *Monica Macovei v. Romania* (App. no. 53028/14), 28 July 2020, §89; ECtHR, *Reznik v. Russia* (App. no. 4977/05), 4 April 2013, §46; ECtHR, *Rungainis v. Latvia* (App. no. 40597/08), 14 June 2018, §63.

³¹⁶ ECtHR, *Bucur v. Romania* (App. no. 40238/02), 8 January 2013.

³¹⁷ *Ibid.*, §§107–113, 117–118.

³¹⁸ See ch. 5 (Espionage and Official Secrets Laws), s. III.5.3. (Truth); ACmHPR, Declaration of Principles of Freedom of Expression and Access to Information in Africa (2019), Principle 35; Tshwane Principles, Principle 38.

³¹⁹ UN Special Rapporteur F. La Rue, *Promotion and protection of the right to freedom of opinion and expression* (2012) UN Doc. A/67/357, §50 (citing the 2001 Joint Statement on Racism and the Media).

³²⁰ ECtHR, *Lingens v. Austria* (App. no. 9815/82), 8 July 1986, §46: ‘a careful distinction needs to be made between facts and value-judgments. The existence of facts can be demonstrated, whereas the truth of value-judgments is not susceptible of proof. . . . As regards value-judgments this requirement is impossible of fulfilment, and it infringes freedom of opinion itself.’

³²¹ See HRC, General Comment No. 34 (2011), §47; ACmHPR, Declaration of Principles on Freedom of Expression and Access to Information in Africa (2019), Principle 21(1); OAS Special Rapporteur, *The Inter-American Legal Framework Regarding the Right to Freedom of Expression* (2009) OEA/Ser.L/V/II CIDH/RELE/INF. 2/09, §109. See also IACtHR, *Usón Ramírez v. Venezuela* (Series C, no. 207), 20 November 2009, §86; IACtHR, *Tristán Donoso v. Panamá* (Series C, no. 193), 27 January 2009, §124.

³²² See, e.g., ECtHR (GC), *Morice v. France* (App. no. 29369/10), 23 April 2015; ECtHR (GC), *Bladet Tromsø v. Norway* (App. no. 21980/93), 20 May 1999.

Whether a statement is qualified as an assertion of fact or of opinion will depend not simply on the words used, but also on the overall context in which the statement was made.³²³ In defamation cases the European Court has held that mixed statements of fact and opinion enjoy a greater degree of protection than plain assertions of fact³²⁴ and in some cases it has required the showing of a sufficient factual basis or link between the facts and their evaluation.³²⁵ Similarly, mixed statements of opinion and fact can be used as *satire*, ‘a form of artistic expression and social commentary which, by its inherent features of exaggeration and distortion of reality, naturally aims to provoke and agitate. Accordingly, any interference with the right to use this means of expression should be examined with particular care.’³²⁶ The type of statement at issue (opinion or mixed statement of fact/opinion) as well as its tone (satirical or humorous) may therefore either make speech inappropriate for regulation or grant it a higher degree of protection under international law.

4.3. Public interest and ‘responsible journalism’

Speech that is ‘political’ or deals with a matter that is in the ‘public interest’ is entitled to greater protection than speech that is not.³²⁷ In assessing false speech in the context of defamation laws, the Human Rights Committee has made clear that public interest in the subject matter of a criticism ‘should be recognized as a defense.’³²⁸ Similarly, the European Court has emphasized that state restrictions on speech that are ‘capable of discouraging the participation of the press in debates over matters of legitimate public concern’ call for the ‘most careful scrutiny.’³²⁹ The Inter-American Court also provides heightened protection to speech regarding public officials or individuals who perform public services.³³⁰ Conversely, the European Court is also clear that commercial speech enjoys a lower degree of protection, such as in the context of state regulation of advertising, which often relates to the accuracy of the information being conveyed.³³¹

The situation is more complex when it comes to speech that is simultaneously commercial and deals with matters of public interest, for example paid political advertising. The European Court has given these mixed kinds of speech an intermediate level of

³²³ ECtHR (GC), *Morice v. France* (App. no. 29369/10), 23 April 2015, §126.

³²⁴ For a particularly good example, see ECtHR, *Monica Macovei v. Romania* (App. no. 53028/14), 28 July 2020, §89.

³²⁵ See, e.g., ECtHR, *Jerusalem v. Austria* (App. no. 26958/95), 27 February 2001, §43; ECtHR, *Feldek v. Slovakia* (App. no. 29032/95), 12 July 2001, §86. See ch. 2 (Insulting Speech) on this issue.

³²⁶ ECtHR, *Ziemiński v. Poland (No. 2)* (App. no. 1799/07), 5 July 2016, §§44–45.

³²⁷ See also HRC, General Comment No. 34 (2011), §§2–3; UN Special Rapporteur A. Ligabo, *Civil and Political Rights, Including the Question of Freedom of Expression* (2006) UN Doc. E/CN.4/2006/55, §47; ECtHR, *Lingens v. Austria* (App. no. 9815/82), 8 July 1986, §§41–42; ECtHR, *Arztammer für Wien and Dorner v. Austria* (App. no. 8895/10), 16 February 2016, §§65–66; IACmHR, *The Inter-American Legal Framework Regarding the Right to Freedom of Expression* (2009) OEA/Ser.L/V/II CIDH/RELE/INF. 2/09, §§8, 32, 99–101.

³²⁸ HRC, General Comment No. 34 (2011), §47. See also HRC, *Kozlov v. Belarus* (Comm. no. 1986/2010), 24 July 2014, §§7.5–7.6 (‘public interest in the subject matter of a criticism is a factor to be taken into account when considering allegations of defamation’).

³²⁹ ECtHR, *Tonsbergs Blad A.S. v. Norway* (App. no. 510/04), 1 March 2007, §88.

³³⁰ IACtHR, *Ricardo Canese v. Paraguay* (Series C, no. 111), 31 August 2004, §§96–98; IACtHR, *Álvarez Ramos v. Venezuela* (Series C, no. 380), 30 August 2019, §§121, 124.

³³¹ See ECtHR, *Arztammer für Wien and Dorner v. Austria* (App. no. 8895/10), 16 February 2016, §§65–66. For an extensive discussion, see Barendt (n 304) 392–416.

protection, and has accepted, for example, the proportionality of comprehensive, viewpoint-neutral bans on political advertising in the broadcast media.³³²

Some legal systems may employ dedicated defences for journalists, which exclude liability for false speech if the statement was made in the course of certain journalistic activities.³³³ This is also reflected in the concept of ‘responsible journalism’ in the case law of the European Court, which affords ‘increased protection’ to speech by ‘public watchdogs’ and particularly the press because these speakers have a duty to impart information and ideas ‘on all matters of public interest’.³³⁴ The Court has indeed found that when freedom of the press is at stake, authorities have a limited margin of appreciation to decide that a ‘pressing social need’ for restrictions exists.³³⁵ But the Court has also made clear that the press must not ‘overstep certain bounds’: they do not have an unfettered right to claim immunity from criminal liability because an offence was committed during the performance of journalistic activities.³³⁶ The additional protection offered to the press is, according to the Court, therefore ‘subject to the proviso that they act in good faith to provide accurate and reliable information in accordance with the tenets of responsible journalism’.³³⁷ And this conditional protection has also been applied to other speakers, such as human rights activists, NGOs and whistleblowers who engage in public debate.³³⁸ Under this doctrine, whether or not a journalist acted ‘responsibly’ is assessed on the basis of the content of the information which is collected and disseminated by journalistic means as well as the process of publication.³³⁹

This doctrine has, however, not been relied upon by UN human rights bodies to decide cases involving freedom of expression, nor has it been adopted by other regional human rights courts, which have addressed misinformation by journalists based on the same considerations as other speakers: harm, culpability, whether the speech was political or related to a matter of public interest.³⁴⁰

In the context of misinformation laws, the nature of the speech and the existence of dedicated affirmative defences are tightly interwoven with how a misinformation offence defines the culpability of the speaker. If the mental elements of a misinformation offence are appropriately exacting (such as knowledge of the falsity of the speech plus intent to bring about the prohibited harm), then there appears to be little room for

³³² But not without controversy—see especially the judgment of a divided European Court in ECtHR (GC), *Animal Defenders International v. United Kingdom* (App. no. 48876/08), 22 April 2013, §123 (‘allowing a somewhat wider margin of appreciation than that normally afforded to restrictions on expressions on matters of public interest’). See also Barendt (n 304) 444–445.

³³³ See ch. 2 (Insulting Speech), s. II.1.4.3. (Reasonable publication).

³³⁴ See, e.g., ECtHR (GC), *Fressoz v. France* (App. no. 29183/95), 21 January 1999, §45(ii); ECtHR (GC), *Pedersen v. Denmark* (App. no. 49017/99), 17 December 2004, §71.

³³⁵ ECtHR, Guide on Article 10 of the European Convention on Human Rights—Freedom of Expression, 31 August 2022, §299.

³³⁶ ECtHR, *De Haes v. Belgium* (App. no. 19983/92), 24 February 1997, §37; ECtHR (GC), *Bladet Tromsø v. Norway* (App. no. 21980/93), 20 May 1999, §62; ECtHR, *Thoma v. Luxembourg* (App. no. 38432/97), 29 March 2001, §45.

³³⁷ See, e.g., ECtHR, *Kaçki v. Poland* (App. no. 10947/11), 4 July 2017, §§48–49.

³³⁸ ECtHR (GC), *Magyar Helsinki Bizottság v. Hungary* (App. no. 18030/11), 8 November 2016, §159; ECtHR, *Gawlik v. Liechtenstein* (App. no. 23922/19), 16 February 2021, §77.

³³⁹ See ch. 1 (Introduction), s. II.3.2.5. (Relevance of whether the speaker is a journalist); see ch. 2 (Insulting Speech), s. III.4.3. (Reasonable publication).

³⁴⁰ See ch. 1 (Introduction), s. II.3.2.5. (Relevance of whether the speaker is a journalist).

a ‘public interest’ defence. If a speaker knows that their speech is false, and intends or foresees that the false speech will cause a specific harm, it will be difficult for the speaker to convincingly argue that the speech was nonetheless in the public interest or that they were acting as a ‘responsible journalist’. If, however, the *mens rea* requirement is at the lower end of the spectrum, such as negligence, then such exceptions or defences may have a role to play, by analogy to the defamation context.³⁴¹

5. Penalties

5.1. Criminal

The severity of a criminal penalty is crucial, and in some cases dispositive,³⁴² in determining whether a penalty on false speech complies with international human rights standards.³⁴³

International courts and human rights bodies have made clear that criminal penalties for speech are proportionate only exceptionally. The Human Rights Committee has approved of criminal penalties for speech in only a handful of cases.³⁴⁴ And in a concurring opinion in *Rabbae v. The Netherlands*, two Committee members urged ‘great caution in the imposition of criminal penalties that punish speech’, proposing that criminal penalties should be limited to speech ‘that incites the commission of criminal offences or acts of violence’.³⁴⁵

The Inter-American Court has found criminal penalties compatible with article 13 of the American Convention in a single case, considered by many to be an outlier.³⁴⁶ Similarly, the African Court has held that custodial sentences for speech will only be lawful in ‘serious and very exceptional circumstances’ such as ‘incitement to international crimes, public incitement to hatred, discrimination or violence or threats

³⁴¹ See ch. 2 (Insulting Speech), s. II.1.4. (Exclusions, exceptions and defences).

³⁴² See ECtHR, *Kasabova v. Bulgaria* (App. no. 22385/03), 19 April 2011, §§69–71 (violation solely because of the excessive penalty (fine plus damages in criminal defamation)).

³⁴³ See ECtHR, *Salov v. Ukraine* (App. no. 65518/01), 6 September 2005, §115 (five years of imprisonment for spreading electoral misinformation a ‘very severe penalty’); ECtHR, *Alekhina v. Russia* (App. no. 38004/12), 17 July 2018, §215 (‘very severe’ punishment of imprisonment imposed due to a protest in a church); §221 (noting that it was ‘the interplay between the various factors involved rather than any one of them taken in isolation that determined the outcome’ of the proportionality analysis); §229 (concluding that criminal punishment was disproportionate, particularly ‘bearing in mind the exceptional seriousness of the sanctions involved’).

³⁴⁴ HRC, *Kim v. Republic of Korea* (Comm. no. 968/2001), 27 July 2005, §8.3; HRC, *Faurisson v. France* (Comm. no. 550/1993), 8 November 1996, §9.7; HRC, *A.K. and A.R. v. Uzbekistan* (Comm. no. 1233/2003), 31 March 2009, §7.2; see ch. 1 (Introduction), s. II.3.1.6. (Criminal penalties for speech).

³⁴⁵ HRC, *Rabbae v. The Netherlands* (Comm. no. 2124/2011), 14 July 2016, Individual concurring opinion of Committee members Sarah Cleveland and Mauro Politi, §§7–8. Cf. HRC, *Ross v. Canada* (Comm. no. 736/1997), 18 October 2000, §§11.1–11.6.

³⁴⁶ IACtHR, *Mémoli v. Argentina* (Series C, no. 265), 22 August 2013, §§141, 143, 148–149. The judgment has been described as a ‘regression’ by Catalina Botero-Marino, Special Rapporteur for Freedom of Expression for the Commission from 2008 to 2014: see, C. Botero-Marino, ‘The Role of the Inter-American Human Rights System in the Emergence and Development of Global Norms on Freedom of Expression’, in L. Bollinger & A. Callamard (eds), *Regardless of Frontiers* (Columbia University Press 2021), 193. See also E. Bertoni, ‘Setbacks and Tension in the Inter-American Court of Human Rights’ (Media Defence, 1 December 2013) (deeming the decision a ‘serious and notable setback’).

against a person or a group of people, because of specific criteria such as race, colour, religion and nationality.³⁴⁷

Although the European Court has sometimes allowed criminal penalties in broader circumstances, the Court has nonetheless held that in any given case it ‘must be satisfied that the penalty does not amount to a form of censorship intended to discourage the press from expressing criticism’ or is ‘likely to deter journalists from contributing to public discussion of issues affecting the life of the community’.³⁴⁸ And it has noted that even when imprisonment is not imposed, ‘the fact of a person’s [criminal] conviction may in some cases be more important than the minor nature of the penalty imposed’.³⁴⁹

In the defamation context, UN Special Rapporteurs, the Working Group on Arbitrary Detention, the ECOWAS Court, the African Commission on Human and Peoples’ Rights and the OSCE have all taken the position that any criminal penalty for defamation violates the right to freedom of expression.³⁵⁰ Similarly the Human Rights Committee, the Council of Europe, Inter-American bodies and the African Court have each held that *imprisonment* is ‘never an appropriate penalty’ for defamatory speech, which is by definition false.³⁵¹

The European Court has however been willing to accept criminal sanctions, though never imprisonment, for defamation at least, partly as a result of the margin of appreciation doctrine.³⁵² The Court has repeatedly found that in view of this doctrine, ‘a criminal measure as a response to defamation cannot, as such, be considered disproportionate to the aim pursued’.³⁵³ But it has applied ‘strict scrutiny’, the ‘most careful scrutiny’ and ‘utmost caution’ when assessing criminal sanctions imposed for defamation,³⁵⁴ and noted that the authorities [must] show ‘restraint in resorting to criminal proceedings’ when dealing with speech.³⁵⁵

³⁴⁷ ACtHPR, *Konaté v. Burkina Faso* (App. no. 004/2013), 5 December 2014, §165.

³⁴⁸ ECtHR (GC), *Bédát v. Switzerland* (App. no. 56925/08), 29 March 2016, §79. See also ECtHR, *Toranzo Gomez v. Spain* (App. no. 26922/14), 20 November 2018, §64; ECtHR, *Lewandowska-Malec v. Poland* (App. no. 39660/07), 18 September 2012, §70.

³⁴⁹ ECtHR (GC), *Bédát v. Switzerland* (App. no. 56925/08), 29 March 2016, §79. See also ECtHR (GC), *Stoll v. Switzerland* (App. no. 69698/01), 10 December 2007, §154; ECtHR, *Haldimann v. Switzerland* (App. no. 21830/09), 24 February 2015, §67.

³⁵⁰ See ch. 2 (Insulting Speech), s. III.6.1. (Criminal penalties). See, e.g., UN Special Rapporteur D. Kaye, *Disease pandemics and the freedom of opinion and expression* (2020) UN Doc. A/HRC/44/49, §40; WGAD, *Maseko v. Swaziland* (Opinion no. 6/2015), 22 April 2015, §27; WGAD, *Amer v. Egypt* (Opinion no. 35/2008), 20 November 2008, §§33, 36; ECOWAS CCJ, *Federation of African Journalists v. The Gambia* (Suit no. ECW/CCJ/APP/36/15), 13 February 2018, 40, 47–48; OSCE Parliamentary Assembly Warsaw Declaration, 8 July 1997, §140; ACmHPR, *Resolution on Repealing Criminal Defamation Laws in Africa* (Res.169 (XLVIII) 10), 24 November 2010.

³⁵¹ HRC, General Comment No. 34 (2011), §47; Parliamentary Assembly of the Council of Europe, *Towards decriminalisation of defamation* (Resolution 1577) 04 October 2007, §17.1; IACtHR, *Álvarez Ramos v. Venezuela* (Series C, no. 380), 30 August 2019, §§120–121; ACtHPR, *Konaté v. Burkina Faso* (App. no. 004/2013), 5 December 2014, §165.

³⁵² See ch. 2 (Insulting speech), s. III.6.1. (Criminal penalties). See ch. 1 (Introduction), s. II.3. (Jurisprudence).

³⁵³ See, e.g., ECtHR (GC), *Lindon, Otchakovsky-Laurens and July v. France* (App. nos. 21279/02 & 36448/02), 22 October 2007, §59; ECtHR, *Radio France v. France* (App. no. 53984/00), 30 March 2004, §40; T. McGonagle, *Freedom of Expression and Defamation: A study of the case law of the European Court of Human Rights* (COE 2016), 18.

³⁵⁴ McGonagle (ibid.) 57.

³⁵⁵ See, e.g., ECtHR, *Castells v. Spain* (App. no. 11798/85), 23 April 1992, §46; ECtHR (GC), *Incal v. Turkey* (App. no. 22678/93), 9 June 1998, §54; ECtHR (GC) *Öztürk v. Turkey* (App. no. 22479/93), 28 September 1999, §66.

In one of its key decisions on misinformation, the European Court assessed speech by a presidential candidate who published false information suggesting that the incumbent had died.³⁵⁶ The Court considered that a five-year sentence, suspended for two years, as well as a fine and disbarment as a lawyer ‘constituted a very severe penalty’, and that the conviction and penalties were ‘manifestly disproportionate’ to the legitimate aim pursued.³⁵⁷

In the context of misinformation offences, the proportionality of a penalty should always be measured against the specific harm that the misinformation caused, or was objectively likely to cause, and the culpability of the offender.³⁵⁸ Although international standards relating to penalties for defamation may apply to misinformation laws in some circumstances, it would be difficult to adopt the same categorical approach to misinformation offences in all contexts. This is because misinformation offences could be designed to remedy other, more pressing harms than reputation, such as the protection of human life or health. There could, in other words, be cases in which imprisonment is a proportionate sentence for a misinformation offence that deals with certain serious harms, but only if it complies with international human rights standards on culpability, causation, and the type, severity, foreseeability and imminence of harms that can justify restrictions on speech.³⁵⁹ This will inevitably be a very narrow category of cases.

5.2. Civil

Defamation is essentially a sub-species of misinformation; civil defamation is also the only private tort that is of systemic relevance to how misinformation should be regulated by a state. It is certainly possible for tort law to apply to harms caused by misinformation—for example, a person seriously injured as a result of a false statement during the Covid-19 pandemic, say that drinking bleach would cure the virus, could conceivably sue the utterer of that statement for damages.³⁶⁰ So could a person defrauded out of their money by misinformation peddling false cures for the virus.³⁶¹ But tort law or private law is not generally used in comparative practice to systemically regulate misinformation.

Far more consequential are the various administrative mechanisms that the state has at its disposal to regulate speech, offline or online, which are often buttressed by the ability to impose fines, mandate the removal or correction of published content or close down media outlets.³⁶² This type of regulation of false speech is not the focus

³⁵⁶ ECtHR, *Salov v. Ukraine* (App. no. 65518/01), 6 September 2005.

³⁵⁷ *Ibid.*, §§115–116. See s. III.2. (Legitimacy).

³⁵⁸ See, e.g., Human Rights Council Resolution 44/12 on freedom of opinion and expression, 16 July 2020, 2.

³⁵⁹ See further, s. III.3. (Necessity).

³⁶⁰ See D. Reiss & J. Diamond, ‘Tort Law: Liability for Anti-Vaccine Misinformation’ (2020) 4 *The Judges’ Book* 107–108 (discussing the use of tort law to provide remedy ‘to people harmed by disease outbreaks when the outbreak generally, or the specific harm can be fairly traced to misinformation’).

³⁶¹ UN Special Rapporteur D. Kaye, *Disease pandemics and the freedom of opinion and expression* (2020) UN Doc. A/HRC/44/49, §46 (citing the use of consumer protection laws against false information regarding cures for Covid-19). Other relevant examples could be torts or delicts such as invasion of privacy or the intentional infliction of emotional distress, all of which could involve false statements of some kind.

³⁶² See HRC, General Comment No. 34 (2011), §39 (on the various State regulatory measures on media).

of our analysis, but we note that any administrative penalties need to comply with the essence of the principles articulated in this chapter. In particular, severe fines or the shutting down of media outlets should require higher degrees of culpability and causality, in order to limit impacts on speakers and the chilling effects of such serious penalties.³⁶³ This is because, as rightly explained by the Venice Commission, ‘excessively high fines pose a threat with almost as much chilling effect as imprisonment, albeit more insidious.’³⁶⁴

The Human Rights Committee has also held that the proportionality of civil penalties for speech may differ depending on how it is disseminated, including by ‘tak[ing] into account the differences between the print and broadcast sectors and the internet.’³⁶⁵ Recent research has indicated, for instance, that falsehoods spread faster than the truth online, at least partly due to the novelty of misinformation and the emotional reactions that it provokes in its recipients,³⁶⁶ and online falsehoods may be more difficult to correct.

IV. Recommendations

The following recommendations are based on minimum international human rights standards applicable to misinformation and—where there is a divergence or lacuna in such standards—a suggestion as to best practice drawn from national systems or emerging international standards. Our recommendations apply cumulatively: in other words, they should all be implemented for misinformation laws to meet international legal or best practice standards. As an example, it would not be sufficient for a law to comply with minimum international standards when it comes to defining the intent or fault of the speaker if the required type and severity and imminence of harm created by the speech is not established.

The recommendations are addressed to states because states have ratified the international treaties that underlie international standards. They can, however, also guide private companies, especially digital platforms, seeking to apply international standards in line with the UN Guiding Principles on Business and Human Rights, although their ability to regulate false speech is broader and not confined only to the most severe types of misinformation that may warrant criminalization by states.³⁶⁷

³⁶³ IACmHR Special Rapporteur E. Lanza, *Childhood, freedom of expression, and the media in the Americas* (2019) OEA/Ser. L/V/II CIDH/RELE/INF.13/19, §145: ‘economic penalties should be proportionate to their potential harm to freedom of expression and should not be arbitrary or disproportionate or have a general chilling effect’. See also UN Special Rapporteur D. Kaye, *Letter dated 1 May 2019*, Ref OL RUS 4/2019, 5 (noting the disproportionate effects of ‘crippling monetary penalties’).

³⁶⁴ Venice Commission, *Opinion on the Legislation on Defamation of Italy* (2013) CDL-AD(2013)038, §62.

³⁶⁵ HRC, General Comment No. 34 (2011), §39.

³⁶⁶ See S. Vosoughi & others, ‘The Spread of True and False News Online’ (2018) 359 *Science* 1146; A. Acerbi, ‘Cognitive Attraction and Online Misinformation’ (2019) 5 *Palgrave Communications* 1.

³⁶⁷ See ch. 3 (Hate Speech), s. V. (Approach of Private Companies to Online Hate Speech).

1. Criminal penalties for false speech should be a last resort. Where possible, state authorities should employ alternative means to prevent or mitigate harm caused by such speech.

Under international law, laws that penalize speech must be ‘the least intrusive instrument amongst those which might achieve their protective function.’³⁶⁸ This means that states cannot restrict harmful false speech through force of law if there are less restrictive means that could achieve the same goals, such as by providing more accurate information.³⁶⁹ Indeed, the primary remedy for untruthful speech should be more truthful speech.³⁷⁰

States should take all feasible steps ‘to ensure that they disseminate reliable and trustworthy information, including about matters of public interest, such as the economy, public health, security and the environment.’³⁷¹ This should include long-term measures such as promoting media and digital literacy, covering these topics as part of the regular school curriculum and taking other steps to raise awareness about these issues and help people identify what is false and where to access reliable data.³⁷²

An example of a successful campaign to combat misinformation without the use of legal penalties comes from South Korea, where health officials successfully combated misinformation around a 2020 national flu vaccination campaign by providing transparent information that created trust in a vaccine.³⁷³ Similarly, following a drop in HPV vaccinations due to misinformation, Danish authorities collaborated with newspapers, magazines and social media to increase confidence in that vaccine.³⁷⁴

Of course, misinformation can affect states, or segments of societies within states, in different ways. Some states may have a prevalence of health-related misinformation, whereas others are dealing with election-related misinformation or a flood of false speech related to a minority group. In some states the traditional media might still be dominant, whereas in others its importance may have been eclipsed by social media.

³⁶⁸ HRC, General Comment No. 34 (2011), §34.

³⁶⁹ Human Rights Council Resolution 44/12 on freedom of opinion and expression, 16 July 2020, 2: ‘[R]esponses to the spread of disinformation and misinformation must be grounded in international human rights law, including the principles of lawfulness, legitimacy, necessity and proportionality [with the Council underlining] the importance of free, independent, plural and diverse media and of providing and promoting access to independent, fact-based and science-based information to counter disinformation and misinformation.’

³⁷⁰ US Supreme Court, *United States v. Alvarez* 567 U.S. 709, 28 June 2012, 727: ‘The remedy for speech that is false is speech that is true. This is the ordinary course in a free society.’ See also OAS Special Rapporteur, ‘Guide to guarantee freedom of expression regarding deliberate disinformation in electoral contexts’ (October 2019), 25 (‘there is no better response to an opinion formed on the wrong basis or false information than an opposing or correct opinion or true information’).

³⁷¹ UN Special Rapporteur, OSCE Representative on Freedom of the Media, OAS Special Rapporteur and ACmHPR Special Rapporteur on Freedom of Expression and Access to Information, *Joint Declaration on Freedom of Expression and ‘Fake News’, Disinformation and Propaganda* (2017) FOM.GAL/3/17, §2(d).

³⁷² *Ibid.*, §3(e).

³⁷³ C. Sang-Hun & D. Grady, ‘How South Korea’s Flu Vaccine Scare Offers Lessons for Other Nations’ (The New York Times, 24 November 2020).

³⁷⁴ World Health Organization, ‘Denmark campaign rebuilds confidence in HPV vaccination’ (27 February 2018).

Instead of seeking legislative one-size-fits-all solutions, states should pursue measures that are most effective to address harmful misinformation in each given context.

State authorities should also conduct regular evidence-based assessments to ensure that their regulation of misinformation is effective and is not serving to amplify misinformation rather than correct it.³⁷⁵ Such reviews are needed as harms caused by a misinformation law can outweigh any benefits gained, or even stifle the media and public debate more generally.³⁷⁶ Indeed, misinformation is most likely to proliferate in environments in which people have lost trust in governmental and other institutions due to censorship and the suppression of dissent.³⁷⁷

It is also possible for restrictions on misinformation by means of criminal law to be necessary and proportionate only temporarily.³⁷⁸ For example, the proliferation of health-related misinformation during a pandemic may justify criminal punishment in certain circumstances while the pandemic lasts.³⁷⁹ When the emergency subsides, such laws would no longer be necessary, and their existence could produce chilling effects on speech even if they are not used in practice.³⁸⁰ Lawmakers should therefore consider using mechanisms such as sunset clauses in order to ensure that legislative responses to misinformation are limited to situations of genuine necessity.

2. Misinformation laws should not be vague or overbroad and should not grant excessive discretion to state authorities.

Misinformation laws that use ambiguous, vague or undefined terms may not be compatible with the ‘legality’ requirement imposed by international human rights law.³⁸¹ Criminal legislation should be drafted precisely, and at a minimum define four key elements to ensure compliance with legality requirements. First, they should define the nature of the false speech that may attract penalties.³⁸² Second, they should specify the prohibited consequence of the misinformation, such as harm to health or to the

³⁷⁵ See, e.g., K. Clayton & others, ‘Real Solutions for Fake News? Measuring the Effectiveness of General Warnings and Fact-Check Tags in Reducing Belief in False Stories on Social Media’ (2020) 42 *Political Behavior* 1073; H. Larson, ‘Blocking Information on COVID-19 Can Fuel the Spread of Misinformation’ (2020) 580 *Nature* 306.

³⁷⁶ UN Special Rapporteur D. Kaye, *Disease pandemics and the freedom of opinion and expression* (2020) UN Doc. A/HRC/44/49, §48: ‘Measures to combat disinformation must never prevent journalists and media actors from carrying out their work or lead to content being unduly blocked on the Internet.’

³⁷⁷ Larson (n 375) 306.

³⁷⁸ UN Special Rapporteur I. Khan, *Disinformation and freedom of opinion and expression* (2021) UN Doc. A/HRC/47/25, §42 (noting that there may be instances where speech may be restricted to forbid propagation of falsehoods, but the restriction must be ‘narrowly construed, *time-limited*, and tailored so as to avoid limiting’ discourse (emphasis added)).

³⁷⁹ See, e.g., ‘COVID-19: States should not abuse emergency measures to suppress human rights—UN experts’ (UN News, 16 March 2020). Cf. ‘COVID-19: Governments must promote and protect access to and free flow of information during pandemic—International experts’ (UN News, 19 March 2020) (noting that ‘Any attempts to criminalise information relating to the pandemic may create distrust in institutional information, delay access to reliable information and have a chilling effect on freedom of expression’).

³⁸⁰ See ‘Censorious governments are abusing “fake news” laws’ (The Economist, 13 February 2021).

³⁸¹ See UN Special Rapporteur D. Kaye, *Disease pandemics and the freedom of opinion and expression* (2020) UN Doc. A/HRC/44/49, §44.

³⁸² HRC, General Comment No. 34 (2011), §34.

integrity of elections, as well as how serious and imminent such harm should be and the required causal link between the misinformation and the harm.³⁸³ Third, the required level of the speaker's fault (*mens rea*) should be set out—for example, a requirement that the speaker knew that their speech was false and that they intended to cause specified harms through the false speech.³⁸⁴ Finally, misinformation laws must also specify exclusions, protections or defences that apply—for instance the fact that truthful speech as well as opinions are always protected and that political or 'public interest' speech deserves extra protection.³⁸⁵

Existing misinformation offences that are incompatible with international standards should be amended or repealed, even if they are not applied, or are only rarely applied, in practice.³⁸⁶ Even laws that are never or rarely applied can produce chilling effects on speech, and it is problematic as a matter of principle to retain laws that are never used.³⁸⁷ This is especially the case with vague laws since the size of the net they cast on speech is unclear. Legislative authorities must have regard to potential chilling effects when drafting misinformation laws, and existing laws that fail to pass muster should be abolished or revised to ensure compatibility with international law.³⁸⁸

State authorities are often granted broad discretion by law, for example with regard to the time, place and manner of an arrest. Prosecutors, the police and other investigative authorities should be mindful that the manner in which these powers are exercised may create a climate of fear and stifle free debate, especially if the target of the investigation is a journalist or a media outlet, and regardless of whether the prosecution ultimately results in a conviction. The suppression of misinformation through heavy-handed action may also amplify any harm that the misinformation causes. State authorities should be instructed and trained to avoid such effects, for example with regard to the manner of arrests, interrogations, searches, other investigative steps and pre-trial detention. They should exercise any discretion granted to them by law so as to minimize such chilling effects. And independent and impartial courts should provide oversight over the exercise of such powers. In doing so, courts should interpret misinformation offences strictly in order to mitigate the negative impact of vague or overbroad provisions in these laws on freedom of speech, and if necessary and appropriate read into them limiting conditions, such as more rigorous requirements for *mens rea* or the causal nexus between the speech and the harm.

³⁸³ See s. III.3. (Necessity).

³⁸⁴ See s. III.3.2. (Intent).

³⁸⁵ See s. III.4. (Exclusions, Exceptions and Defences).

³⁸⁶ UN Special Rapporteur, OSCE Representative on Freedom of the Media, OAS Special Rapporteur and ACmHPR Special Rapporteur on Freedom of Expression and Access to Information, *Joint Declaration on Freedom of Expression and 'Fake News', Disinformation and Propaganda* (2017) FOM.GAL/3/17, §2(a); UN Special Rapporteur D. Kaye, *Disease pandemics and the freedom of opinion and expression* (2020) UN Doc. A/HRC/44/49, §49. See also HRC, General Comment No. 34 (2011), §25.

³⁸⁷ See, e.g., T. Mendel, 'The Case against Criminal Defamation Laws', in A. Karlsreiter & H. Vuokko (eds), *Ending the Chilling Effect—Working to Repeal Criminal Libel and Insult Laws* (OSCE 2004), 25.

³⁸⁸ UN Special Rapporteur, OSCE Representative on Freedom of the Media, OAS Special Rapporteur and ACmHPR Special Rapporteur on Freedom of Expression and Access to Information, *Joint Declaration on Freedom of Expression and 'Fake News', Disinformation and Propaganda* (2017) FOM.GAL/3/17, §2(a). See also HRC, General Comment No. 34 (2011), §25.

3. States should not penalize speech simply on the basis that it is false. They may only impose restrictions on false speech when doing so advances an aim that is recognized as legitimate under international human rights law.

Laws prohibiting the dissemination of false information on the basis of its falsity alone violate international human rights law.³⁸⁹ Misinformation laws must advance a specific purpose recognized under international standards, namely: (a) respect for the rights or reputations of others; or (b) national security, public order or public health.³⁹⁰ The Human Rights Committee has made clear that the purpose of ‘maintain[ing] the integrity of the electoral process’ falls within the subset of the ‘protection of public order’ or ‘respect of the rights of others.’³⁹¹ Although ‘public morals’ are also a justification recognized under international human rights law, this basis is vague, anachronistic and unlikely to justify a restriction on speech today.³⁹²

A number of states use misinformation laws that prohibit disseminating false information without a link to any concrete harm caused by false speech, and consequently without a link to a legitimate aim that can justify restricting speech.³⁹³ Such general misinformation laws do not comply with minimum international standards and should be repealed. The same conclusion applies to laws that criminalize misinformation to promote non-legitimate aims focused on quelling dissent or penalizing reputational ‘injury’ to the state, such as Oman’s prohibition on ‘false or malicious’ news that ‘undermine[s] the stature of the state.’³⁹⁴ A law narrowly tailored to advancing a specific legitimate aim, such as in the electoral or public health context, is more likely to be necessary and proportionate.

Under international law, restrictions on speech ‘must be applied only for those purposes for which they were prescribed and must be directly related to the specific need on which they are predicated.’³⁹⁵ Even if misinformation laws comply with international standards on paper, they would violate such standards if used for ulterior purposes to suppress political debate or criticism. A review of state practice demonstrates that this

³⁸⁹ See s. III.2. (Legitimacy); UN Special Rapporteur, OSCE Representative on Freedom of the Media, OAS Special Rapporteur and ACmHPR Special Rapporteur on Freedom of Expression and Access to Information, *Joint Declaration on Freedom of Expression and ‘Fake News’, Disinformation and Propaganda* (2017) FOM.GAL/3/17, §2(a): ‘General prohibitions on the dissemination of information based on vague and ambiguous ideas, including “false news” or “non-objective information”, are incompatible with international standards for restrictions on freedom of expression, as set out in paragraph 1(a), and should be abolished.’

³⁹⁰ ICCPR Art. 19. Harms to other interests provided by article 10(2) of the European Convention, such as protecting the authority of the judiciary, fall within the scope of harms to public order: See s. III.2. (Legitimacy) and ch.1. (Introduction).

³⁹¹ HRC, General Comment No. 34 (2011), §37 citing HRC, *Kim v. Republic of Korea* (Comm. no. 968/2001), 27 July 2005, §8.3. See s. III.2. (Legitimacy).

³⁹² See A. Clooney & P. Webb, *The Right to a Fair Trial in International Law* (OUP 2020), Chapter 2: Right to a Public Trial, 197.

³⁹³ See s. II.1.2. (Harm).

³⁹⁴ See further s. III.2. (Legitimacy); WGAD, *Rajab v. Bahrain* (Opinion no. 13/2018), 19 April 2018, §29; WGAD, *Sheikh Suliaman al-Rashudi and others v. Saudi Arabia* (Opinion no. 38/2015), 4 September 2015, §73.

³⁹⁵ HRC, General Comment No. 34 (2011), §22.

happens all too often.³⁹⁶ For instance, dozens of governments used misinformation laws to clamp down on legitimate criticism of their response to the Covid-19 pandemic rather than to address the real and dangerous spread of misinformation about the pandemic.³⁹⁷ Domestic courts can play an important role in ensuring that state authorities are not acting with an ulterior purpose. And best practice would not require direct proof of an ulterior purpose: indirect and circumstantial evidence, drawing inferences where necessary from contextual factors and patterns of behaviour by state authorities, should suffice and indeed is often the only available evidence.³⁹⁸

4. States should not penalize misinformation unless it is objectively likely to cause serious and direct or imminent harm; in the context of criminal laws such harm should be objectively probable and limited to incitement to violence or a serious criminal offence, endangerment of human life, or serious harm to health or the integrity of democratic elections.

States must demonstrate a link between the speech they seek to penalize and ‘in specific and individualized fashion the precise nature of the threat’ likely to result from such speech.³⁹⁹ Drawing on existing international standards, we recommend that states may impose criminal penalties for disseminating false information if doing so (a) incites violence or a serious criminal offence,⁴⁰⁰ (b) endangers human life or seriously harms health, or (c) seriously harms the integrity of democratic elections.⁴⁰¹ The protection of the reputation of the state or its organs—including the head of state, military or judiciary—is not a legitimate basis for restricting false speech.⁴⁰² Laws that criminalize ‘picking quarrels and provoking trouble’⁴⁰³ do not meet the international standard. And when criminalizing misinformation is justified on the basis that the speech is likely to incite a serious criminal offence, that offence must itself comply with international standards.⁴⁰⁴ For example, offences such as sodomy laws—themselves incompatible

³⁹⁶ See s. II.2. (Application of False Speech Laws Around the World).

³⁹⁷ See s. II. (State Practice).

³⁹⁸ See s. III.2.2. (What objectives can justify the penalization of false speech?).

³⁹⁹ See HRC, General Comment No. 34 (2011), §35. See also UN Special Rapporteur D. Kaye, *Disease pandemics and the freedom of opinion and expression* (2020) UN Doc. A/HRC/44/49, §15.

⁴⁰⁰ See s. III.3.1.1. (Severity and imminence of harm); HRC, *Rabbae v. The Netherlands* (Comm. no. 2124/2011), 14 July 2016, Individual Concurring Opinion of Committee members Sarah Cleveland and Mauro Politi, §4.

⁴⁰¹ This standard also accounts for historical bases for penalizing false speech, such as commercial fraud or medical and legal professionals misrepresenting their qualifications, on the basis that such acts would constitute serious criminal offences in many states: See ch. 1 (Introduction), s. II. (International Standards on Freedom of Speech and of the Press).

⁴⁰² HRC, General Comment No. 34 (2011), §38 (‘States parties should not prohibit criticism of institutions, such as the army or the administration’).

⁴⁰³ G. Rui, ‘Picking quarrels and provoking trouble’: how China’s catch-all crime muzzles dissent’ (South China Morning Post, 25 August 2021).

⁴⁰⁴ See, e.g., UN Special Rapporteur I. Khan, *Disinformation and freedom of opinion and expression during armed conflicts* (2022) UN Doc. A/77/288, §38: ‘Disinformation cannot be prohibited under international human rights law unless it amounts to advocacy of hatred that constitutes incitement to hostility, violence and discrimination. It may be restricted only if it meets the requirements of legality, necessity, and legitimate objectives’ set out in the ICCPR.

with international human rights law—cannot be used as a legitimate basis to penalize speech likely to incite the ‘crime’ of being gay.⁴⁰⁵

5. To penalize misinformation, states should require that it is at least objectively likely that the false speech will cause serious harm that is direct or imminent. A higher standard should apply in any criminal law.

In addition to precisely defining the harm that a misinformation law seeks to remedy, a misinformation law must also specify the degree of the causal connection between the false speech and the relevant harm. This is particularly important for criminal statutes. According to the Human Rights Committee, the state must establish ‘a direct and immediate connection between the expression and the threat’ of harm.⁴⁰⁶ However, the European Court has, in some cases, found that criminal penalties could be appropriate even for speech that was ‘capable of’ causing or ‘could have’ caused prohibited harms, and in some instances it has sidestepped consideration of a causal requirement entirely.⁴⁰⁷

A criminal statute can establish this connection between the misinformation and the harm in several different ways. The more intense the connection, the easier it is to justify the offence as a necessary and proportionate restriction on speech.

First, a statute can limit the offence to situations in which the harm *has actually occurred*, with the misinformation either being a ‘but for’ cause of the harm, or having causally contributed to it at some lower standard.

Second, the statute could expand the consequence element of the offence to include harms that did *not* occur, but *could have* occurred as a result of the false speech. Again, the key issue here would be in defining the degree of causal connection between the speech and the harm. A more intense requirement would be for the harm to be a *likely* consequence of the speech. A less intense requirement would impose liability for speech that was *capable* of causing or *could or may* have caused the harm. Such a low intensity requirement would be difficult to justify under international human rights law in terms of its necessity and proportionality in the absence of any such harm materializing.⁴⁰⁸

Third, the connection between the false speech and the harm could be described solely in terms of intention rather than causality. Thus, Qatari law punishes

⁴⁰⁵ UN High Commissioner for Human Rights, *Discriminatory laws and practices and acts of violence against individuals based on their sexual orientation and gender identity* (2011) UN Doc. A/HRC/19/41, §41 (‘The criminalization of private consensual homosexual acts violates an individual’s rights to privacy and to non-discrimination and constitutes a breach of international human rights law’).

⁴⁰⁶ HRC, General Comment No. 34 (2011), §35. See also UN Special Rapporteur D. Kaye, *Disease pandemics and the freedom of opinion and expression* (2020) UN Doc. A/HRC/44/49, §15; ACmHPR, Declaration of Principles on Freedom of Expression in Africa (2019), Principle 9(4), Principle 22(2) requiring ‘a close causal link between the risk of harm and the expression’.

⁴⁰⁷ See s. III.3.1.2. (Causal relationship between the speech and the harm); see ch. 3 (Hate Speech), s. III.2.3.1. (Harm).

⁴⁰⁸ See s. III.3.1.2. (Causal relationship between the speech and the harm).

misinformation uttered ‘with the intention’ of harming certain interests.⁴⁰⁹ An offence that does not require *any* causal connection would be difficult to justify, since a person could be punished regardless of the fact that their speech actually caused no serious social harm and regardless of the fact that it could have been extremely unlikely for the speech to cause such harms.

Finally, a criminal misinformation law could have no stipulated connection whatsoever between expression and a specific harm, either in terms of causality or in terms of culpability. Such laws will, again, be difficult to justify, since their (implicit) goals, say that of protecting public health, could have been achieved through a more narrowly tailored law, i.e. through less restrictive means that would have had a lower chilling effect. A good example is the misinformation offence in article 207.1 of the Russian Criminal Code, which punishes the ‘public dissemination, under the guise of a truthful message, of knowingly false information about circumstances that pose a threat to the life and safety of citizens, and (or) about measures taken to ensure the safety of the population and territories. . . from these circumstances.’⁴¹⁰ This offence requires no concrete harm to have occurred as the result of the misinformation, nor even for such harm to have been a likely or possible consequence thereof.⁴¹¹ The crime is completed simply by the speech act, unlike the offence in article 207.2 examined above, which does require the causation of specific harm.⁴¹²

To conclude, there is no doubt that higher-intensity causal requirements are much more likely to satisfy a necessity and proportionality analysis and to avoid having excessive chilling effects on protected speech. The dangers of the chilling effects of overbroad criminal laws punishing false speech, and of their abuse, are such that we recommend the use of high causal requirements in defining misinformation offences. At a minimum such offences should require an objective likelihood that the speech would cause direct or imminent harm. A criminal misinformation law would be easier to justify with an even higher threshold, that of false speech actually causing harm which did occur (for example, if an individual died or their health was harmed as a result of the promotion of false remedies for a disease). High causal requirements should apply alongside a sufficiently high *mens rea* standard.⁴¹³

⁴⁰⁹ See s. II.1.1. (Vagueness and overbreadth).

⁴¹⁰ Amnesty International, ‘Russian Federation: “Fake news” bill prompted by COVID-19 threatens freedom of expression’ (3 April 2020).

⁴¹¹ See CFJ, ‘Statement in the Conviction of Journalist Alexander Pichugin on ‘Fake News’ Charges in Russia’ (11 November 2020) (noting that the prosecution ‘made no effort to show how [the speech] posed any kind of danger to public health or public order’).

⁴¹² The offence in Art. 207.2 imposes more severe penalties on the offender: see Amnesty International, ‘Russian Federation: “Fake news” bill prompted by COVID-19 threatens freedom of expression’ (3 April 2020).

⁴¹³ See s. IV. (Recommendations): ‘Our recommendations apply cumulatively: in other words, they should all be implemented for misinformation laws to meet international legal or best practice standards’.

6. Misinformation offences should have high culpability (*mens rea*) requirements.

The fault element of a misinformation offence is one of the most important factors in assessing its necessity and proportionality under international human rights law.⁴¹⁴ False speech offences that do not incorporate a *mens rea* standard—for instance a requirement of knowing falsehood or intention to deceive—would create ‘strict liability’ and in doing so violate not only the right to freedom of expression but also potentially the presumption of innocence as part of the right to a fair trial.⁴¹⁵

Deliberate falsehoods enjoy a lesser degree of protection than innocent, negligent or reckless falsehoods, although recklessness could be sufficient to render a speaker civilly liable in certain circumstances.⁴¹⁶ It is particularly important that criminal laws have sufficiently stringent *mens rea* requirements if they are to be considered necessary and proportionate, and for such culpability requirements to be prescribed with regard to all constituent elements of the offence (including the falsity of the statement and the harm caused by it).⁴¹⁷

A law that falls closer to the line is the offence introduced into South African law during the Covid-19 pandemic, which punishes any person ‘who publishes any statement . . . *with the intention to deceive* any other person about— (a) COVID-19; (b) [the] COVID-19 infection status of any person; or (c) any measure taken by the Government to address COVID-19’.⁴¹⁸ The *mens rea* requirement as to conduct is high—the intention to deceive—but there is no specified harm or required fault as to that harm. In other words, a person could be punished for this offence even if no concrete harm arises—or could have arisen—from the false utterance, and even if the speaker lacked any intent to bring about that harm. The fact that the incrimination extends to false statements about Covid-19 *generally* or as to any remedial measures taken by the government means that it could be applied very broadly indeed, and is unlikely to be considered proportionate by an international human rights body on that basis.⁴¹⁹

It should be borne in mind that while misinformation frequently *originates* from persons acting knowingly or with an intent to cause harm (i.e. engaging in disinformation), it is most often *spread* by people who subjectively believe it to be true and intend to cause no harm. ‘Many people have shared misleading or false information with well-meaning intentions.’⁴²⁰ Imposing criminal liability on such individuals solely on the basis of an objectivized, negligent failure to verify facts, especially if the offence is otherwise defined quite broadly, is likely to produce substantial chilling effects on free

⁴¹⁴ See s. III.3.2. (Intent).

⁴¹⁵ See Clooney & Webb (n 392) Chapter 4 (The Right to be Presumed Innocent), 210.

⁴¹⁶ See further s. III.3.2. (Intent).; see ch. 3 (Hate Speech), s. II.1.3. (Intent).

⁴¹⁷ See further s. III.3.2. (Intent).

⁴¹⁸ South African Disaster Management Act 2020 s. 14(2).

⁴¹⁹ *Ibid.*

⁴²⁰ UK Parliament, Misinformation in the COVID-19 Infodemic (2020), §6.

speech and should fail the least restrictive means limb of the justifiability test under international law.⁴²¹

7. Decisions on arrests and criminal charges related to political speech or involving a journalist should be approved by a high-ranking official such as the Minister of Justice, Attorney-General or a senior police official or prosecutor.

Although international standards do not expressly require this, requiring an extra layer of consent from an official or at higher levels of police and prosecutorial authorities can be an important best-practice safeguard against abuse of misinformation laws that can produce chilling effects even if any charges are eventually dismissed.⁴²² It can also assist in ensuring consistency of approach by officials. This is particularly important given the scope for abuse when it comes to political speech, including by members of the media.

8. Misinformation laws must never penalize substantially truthful speech, or opinions, and should provide additional protection to speech that is political or in the public interest.

When *mens rea* or harm requirements—or both—are at the lower end of the permissible scale,⁴²³ affirmative defences will be particularly important to mitigate the chilling effects of misinformation offences and render them permissible under international law.⁴²⁴ Such defences should include the defences recognized in international law in the defamation context, which already apply to false speech, and adjusted where necessary.⁴²⁵ All elements of a misinformation offence must be proved in court, and a defendant must have a meaningful opportunity to negate them.⁴²⁶

⁴²¹ An instructive example in that regard is the misinformation offence in the UK Representation of the People Act 1983 s. 106, which prohibits, before or during an election, making false statements about the ‘personal character’ of a candidate, and essentially does so on the basis of a negligent failure to verify facts. The main redeeming feature of the offence is that it is applied with exceptional rarity. But it is nonetheless both vague and has a very low *mens rea* requirement. For an extended discussion, see Rowbottom (n 307) 507. Contrast with far more rigorous Canadian misinformation offences discussed in s. II.1.2. (Harm).

⁴²² Reporters Without Borders (RSF) considers express consent must be given by an independent judicial authority, i.e., a court or an investigating judge in civil law countries. See Foreword.

⁴²³ See s. IV.6. (Recommendations).

⁴²⁴ See further s. III.4. (Exclusions, Exceptions and Defences).

⁴²⁵ See ch. 2 (Insulting Speech), s. III.4. (Exclusions, Exceptions and Defences).

⁴²⁶ See UN Special Rapporteur, OSCE Representative on Freedom of the Media, OAS Special Rapporteur and ACmHPR Special Rapporteur on Freedom of Expression and Access to Information, *Joint Declaration on Freedom of Expression and ‘Fake News’, Disinformation and Propaganda* (2017) FOM.GAL/3/17, §1(d), (h); ACmHPR, *Declaration of Principles on Freedom of Expression and Access to Information in Africa* (2019), Principle 9(2). See also Clooney & Webb (n 392) Chapter 4 (The Right to Prepare a Defence).

8.1. Truth

Falsity must be an element of a misinformation crime; truth should consequently always exempt a statement from liability.⁴²⁷ And a defendant should always have a meaningful opportunity to negate any of the constituent elements of the offence.⁴²⁸ The determination of the existence of all elements of a misinformation offence must be in the hands of an independent and impartial court.⁴²⁹ This includes the determination of whether the defendant's statement was true or false.⁴³⁰ The role of assessing whether speech is false should never be delegated to officials in the executive branch, as doing so creates enormous potential for abuse, and insufficient protection for speech.⁴³¹ To the extent that the court's establishment of the truth requires reliance on experts, the court must provide a reasoned explanation as to why it found the views of particular experts credible.⁴³² Independent criminal courts must retain control of the fact-finding process.

Substantial truth should be sufficient to exempt a statement from liability.⁴³³ For instance, consider a scenario in which a journalist claims that a hospital ran out of personal protective equipment (PPE) during the Covid-19 pandemic. It later turns out that there was shortage of only one type of PPE, say masks, but not gloves and gowns. The original report is not true in its totality, but it is true *substantially* or *in essence*—there was in fact a shortage of a key piece of equipment, and the purpose of the report was to criticize the authorities for this shortage to hold them accountable. Or, similarly, consider a report in which a journalist claims that a politician corruptly appropriated 10 million dollars, when he in fact stole five million. The key message was to expose the corrupt behaviour of the official, not the exact amount of money that he stole. As these examples demonstrate, imposing criminal liability on individuals for speech that is not wholly true, even if it was substantially true, would not satisfy the necessity and proportionality requirements of human rights law, notwithstanding the fact that the required

⁴²⁷ See HRC, General Comment No. 34 (2011), §47 (noting, with regard to penal defamation laws, that these 'should include such defences as the defence of truth and they should not be applied with regard to those forms of expression that are not, of their nature, subject to verification'). See s. III.4.2. (Opinion). See also UN Special Rapporteur, OSCE Representative on Freedom of the Media, OAS Special Rapporteur and ACmHPR Special Rapporteur on Freedom of Expression and Access to Information, *Joint Declaration on Freedom of Expression and Responses to Conflict Situations* (2015), §4(a).

⁴²⁸ See Clooney & Webb (n 392) Chapter 4 (The Right to Prepare a Defence).

⁴²⁹ See, e.g., ACmHPR, Declaration of Principles on Freedom of Expression and Access to Information in Africa (2019), Principle 9(2).

⁴³⁰ See UN Special Rapporteur, OSCE Representative on Freedom of the Media, OAS Special Rapporteur and ACmHPR Special Rapporteur on Freedom of Expression and Access to Information, *Joint Declaration on Freedom of Expression and 'Fake News', Disinformation and Propaganda* (2017) FOM.GAL/3/17, §2(d).

⁴³¹ UN Special Rapporteur, OSCE Representative on Freedom of the Media, OAS Special Rapporteur and ACmHPR Special Rapporteur on Freedom of Expression and Access to Information, *Joint Declaration on Freedom of Expression and 'Fake News', Disinformation and Propaganda* (2017) FOM.GAL/3/17, §1(d), (h). See also EACJ, *Media Council of Tanzania and others v. Attorney General of the United Republic of Tanzania* (Ref no. 2/2017), 28 March 2019, §§108–110.

⁴³² See Clooney & Webb (n 392) Introduction, 52 (noting that the HRC refused to defer to domestic courts' findings 'when the courts failed to provide reasons and justifications for their assessment of the evidence') and Chapter 1 (The Right to a Competent, Independent and Impartial Tribunal Established by Law), 125–126.

⁴³³ See further s. III.4.1. (Truth); ECtHR, *Bergens Tidende v. Norway* (App. no. 26132/95), 2 May 2000, §56.

elevated culpability and harm elements of the offence could not be met in such cases in any event.⁴³⁴

8.2. Opinion

Misinformation laws must exclude opinions from their ambit. This is in line with international standards related to defamation laws, which also provide that opinion must not be captured by such laws.⁴³⁵ A court must therefore also carefully distinguish between statements of fact and statements of opinion (including humour and satire), since only the former are capable of proof and can constitute misinformation.⁴³⁶

8.3. Public interest

Misinformation laws should, like defamation laws, provide additional protection for political or ‘public interest’ speech. The Human Rights Committee has held, in the context of defamation, that the public interest in the subject matter of a criticism ‘should be recognized as a defence.’⁴³⁷ Political speech, such as criticism of the work of the government or the positions of a political party, should enjoy a high level of protection. So should speech on any matter broadly of public interest, such as state policy regarding climate change, and the like.⁴³⁸ Given the value of such speech, it is appropriate to apply a higher standard of protection, including when such speech involves the media.⁴³⁹ It is in principle difficult for the state to justify restrictions on this type of expression, even if it pursues a legitimate aim in doing so. An affirmative public interest defence would be of particular importance if the *mens rea* of the misinformation did not, contrary to our recommendation, require the knowing or intentional dissemination of falsehoods, but employed lesser forms of culpability.

A degree of falsity and exaggeration is inevitable in political debate and must be tolerated by the state, in order to avoid chilling effects on socially beneficial speech.⁴⁴⁰ If the target of the speech is a politician, public official or potentially another public figure in a position of power they must expect greater exposure to public scrutiny and criticism.⁴⁴¹ The European Court also considers that if the speaker is a journalist or

⁴³⁴ See also Colombian Constitutional Court, *Accion de Tutela contra Particulares (Procedencia excepcional)*, Tutela judgment T-293/2018, 24 July 2018, §7.3.1 (requiring proof of a *reasonable degree* of truth only).

⁴³⁵ See further s. III.4.2. (Opinion). ECtHR, *Lingens v. Austria* (App. no. 9815/82), 8 July 1986, §46: ‘a careful distinction needs to be made between facts and value-judgments. The existence of facts can be demonstrated, whereas the truth of value-judgments is not susceptible of proof... As regards value-judgments this requirement is impossible of fulfilment, and it infringes freedom of opinion itself’; ECtHR, *Hertel v. Switzerland* (App. no. 25181/94), 25 August 1998, §50.

⁴³⁶ See further s. III.4.2. (Opinion).

⁴³⁷ HRC, General Comment No. 34 (2011), §47; HRC, *Kozlov v. Belarus* (Comm. no. 1986/2010), 24 July 2014, §§7.5–7.6.

⁴³⁸ See HRC, General Comment No. 34 (2011), §§2–3, 30. See further s. III.4.3. (Public interest and ‘responsible journalism’).

⁴³⁹ See ch. 2 (Insulting Speech), s. IV. (Recommendations); ch. 3 (Hate Speech), s. VI. (Recommendations).

⁴⁴⁰ See, e.g., ECtHR, *Steel v. United Kingdom* (App. no. 68416/01), 15 February 2005, §90.

⁴⁴¹ See ACmHPR, Declaration of Principles on Freedom of Expression and Access to Information in Africa (2019), Principle 21(1); IACmHR, *The Inter-American Legal Framework Regarding the Right to Freedom of Expression* (2009) OEA/Ser.L/V/II CIDH/RELE/INF. 2/09, §§39–52; ECtHR, *Lingens v. Austria* (App. no. 9815/82), 8 July 1986, §42; HRC, General Comment No. 34 (2011), §§34, 38.

other ‘public watchdog’ increased protection may be appropriate if certain conditions are met.⁴⁴²

9. Penalties must be proportionate to the harm caused by false speech and the speaker’s level of fault. Custodial sanctions will be justified only exceptionally.

Criminal misinformation laws generally pose particularly grave restrictions on freedom of expression and are necessary and proportionate only in response to specific and sufficiently serious social harms caused by the speech.⁴⁴³ Heavy penalties can be one of many factors that render a misinformation offence, as prescribed or applied, to be unnecessary and disproportionate. A severe sentence may even on its own render a criminal conviction disproportionate.⁴⁴⁴ Sentences should depend primarily on the culpability of the offender and the gravity of the harm that the misinformation caused.⁴⁴⁵ Custodial sentences will be proportionate only exceptionally and require a weighty justification, for example causing death or serious harms to health.⁴⁴⁶

Criminal misinformation laws that can be or are applied to reporting or debate on matters of public interest are particularly concerning.⁴⁴⁷ Imposing criminal liability on such speech is the most restrictive measure possible, with the greatest potential for chilling effects and greatest possible detriment to democracy. Indeed, even when a restriction on speech pursues a clearly legitimate aim, like the protection of public health, criminalization should be reserved only for specific and sufficiently serious harms and be subject to rigorous causality and culpability requirements.⁴⁴⁸ That said, there are historically uncontroversial examples of criminal responsibility and custodial sentences being imposed for false statements of fact, including offences such as perjury, where the harm is to public order and integrity of the judicial process,⁴⁴⁹ and fraud, where the

⁴⁴² See ch. 2 (Insulting Speech), s. III.4.3. (Reasonable publication); But see ch.1 (Introduction), s. II.3.2.5. (Relevance of whether the speaker is a journalist). See also ECtHR, *Prager and Oberschlick v. Austria* (App. no. 15974/90), 26 April 1995, §38; IACmHR, *The Inter-American Legal Framework Regarding the Right to Freedom of Expression* (2009) OEA/Ser.L/V/II CIDH/RELE/INF. 2/09, §38.

⁴⁴³ See s. IV.4. (Recommendations); OAS Guide to Guarantee Freedom of Expression Regarding Deliberate Disinformation in Electoral Contexts, 30.

⁴⁴⁴ See ECtHR (GC), *Bédat v. Switzerland* (App. no. 56925/08), 29 March 2016, §79; see s. III.5. (Penalties).

⁴⁴⁵ See s. IV.4. (Recommendations).

⁴⁴⁶ See s. III.2.2. (What objectives can justify the penalization of false speech?).

⁴⁴⁷ See, e.g., Cameroonian Penal Code 2016 s. 240 (which punishes the dissemination of *any* false news, on whatever topic and regardless of the harm it causes, but punishes such dissemination with imprisonment up to five years, *doubling* that penalty if the dissemination is done anonymously).

⁴⁴⁸ See ss IV.4. (Recommendations) and IV.6. (Recommendations). See HRC, General Comment No. 34 (2011), §47 (discussing criminalization of defamation).

⁴⁴⁹ It is possible for other offences against the administration of justice to include false statements of fact as an element of the offence. In English law, for example, a strict liability contempt of court rule exists for any publication relating to active judicial proceedings which creates a substantial risk that the course of justice in the proceedings in question will be seriously impeded or prejudiced. Liability is excluded, however, ‘in respect of a *fair and accurate report* of legal proceedings held in public, published contemporaneously and in good faith’: UK Contempt of Court Act 1981 s. 4(1) (emphasis added).

primary harm is to individual property rights. Generally speaking, with the exception of criminal defamation, such offences are tailored relatively narrowly to these harms, can be committed only in specific contexts or with regard to a particular subject-matter, and are therefore not likely to stifle debate on matters of public interest.⁴⁵⁰

Fines and other penalties that can adversely affect the operation of a media outlet require substantial justification.⁴⁵¹ This also applies to any administrative measures that may be ancillary to a criminal conviction, such as the blocking of websites or applications. As explained by UN and regional mandate-holders on freedom of expression, '[s]tate mandated blocking of entire websites, IP addresses, ports or network protocols is an extreme measure which can only be justified where it is provided by law and is necessary to protect a human right or other legitimate public interest, including in the sense of that it is proportionate, there are no less intrusive alternative measures which would protect the interest and it respects minimum due process guarantees'.⁴⁵²

Other severe civil and administrative measures that may be ancillary to the commission of misinformation offences, such as the revocation of broadcast licences, are likely to be considered proportionate only exceptionally, for instance as a response to deliberate, repeated and systematic spreading of misinformation that causes or can cause serious prohibited harms, on the basis of clear and foreseeable rules as to when such sanctions can be imposed and an impartial application of these standards by an independent entity such as a court.⁴⁵³ Severe sanctions should not be imposed without recourse to judicial review by an independent and impartial court, and generally should not be enforced while such review is pending.⁴⁵⁴

⁴⁵⁰ See, e.g., United States Congressional Research Service, *False Speech and the First Amendment: Constitutional Limits on Regulating Misinformation* (2022), 'Fraud and False Commercial Speech', 1–2. See also UK Contempt of Court Act 1981 s. 4(1).

⁴⁵¹ See, e.g., UN Special Rapporteur D. Kaye, *Letter dated 1 May 2019*, Ref OL RUS 4/2019, 5 (noting the disproportionate effects of 'crippling monetary penalties').

⁴⁵² UN Special Rapporteur, OSCE Representative on Freedom of the Media, OAS Special Rapporteur and ACmHPR Special Rapporteur on Freedom of Expression and Access to Information, *Joint Declaration on Freedom of Expression and 'Fake News', Disinformation and Propaganda* (2017) FOM.GAL/3/17, §1(f).

⁴⁵³ E.g., when commenting on Hungary's Media Act the Venice Commission noted that the 'Media Council may use its powers to impose heavy sanctions (such as high fines or interruption of broadcasting, blocking of access etc.) only as a measure of last resort, where all other reasonable attempts to steer the media outlet on the right path have failed, and where its publications repeatedly... endangered public peace and order (for example, where the media outlet has repeatedly made calls for unlawful violence in respect of minority groups or advocated a violent overthrow of a democratic public order). In addition, there is a need to develop policy guidelines on administrative sanctions which would explain how the Media Council exercises its discretion in this sphere': Venice Commission, *Opinion on Media Legislation (Act CLXXXV on Media Services and on the Mass Media, Act CIV on the Freedom of the Press, and the Legislation on Taxation of Advertisement Revenues of Mass Media) of Hungary* (2015) CDL-AD(2015)015, §41.

⁴⁵⁴ See Venice Commission, *Opinion on Media Legislation (Act CLXXXV on Media Services and on the Mass Media, Act CIV on the Freedom of the Press, and the Legislation on Taxation of Advertisement Revenues of Mass Media) of Hungary* (2015) CDL-AD(2015)015, §§42–46.

10. State officials should refrain from disseminating or amplifying harmful misinformation.

Misinformation that emanates from state actors is often more harmful than misinformation disseminated by non-state actors.⁴⁵⁵ Depending on the nature of the misinformation and the harm that it causes, the dissemination of misinformation by the state may violate the state's obligation to respect a number of human rights, including freedom of expression.⁴⁵⁶ A state's involvement in spreading misinformation also hinders its ability to credibly restrict misinformation by private actors.

11. States must protect the public from human rights violations by private actors, including companies.

Minimum international standards require that the state protects individuals from human rights violations by private actors, including media companies,⁴⁵⁷ by 'taking appropriate steps to prevent, investigate, punish and redress such abuse through effective policies, legislation, regulations and adjudication.'⁴⁵⁸ In particular, states need to provide 'effective guidance to business enterprises on how to respect human rights throughout their operations.'⁴⁵⁹ Companies that become dominant in a particular area and produce outsized impacts on speech, such as Facebook/Meta, require particular scrutiny.⁴⁶⁰ This includes scrutiny of both whether digital platforms and other companies cause systematic social harms by enabling the spread of misinformation, and whether, conversely, these companies are suppressing too much speech when taking measures to combat misinformation. Any regulatory systems used by states in this regard should depend primarily on the work of independent regulatory bodies free from political control or influence,⁴⁶¹ which must act according to fair and transparent procedures, providing reasons for their decisions and with appropriate remedies available to challenge them.⁴⁶² As explained above, regulatory interventions by means of criminal law will be justified only rarely.

⁴⁵⁵ UN Special Rapporteur, OSCE Representative on Freedom of the Media, OAS Special Rapporteur and ACmHPR Special Rapporteur on Freedom of Expression and Access to Information, *Joint Declaration on Freedom of Expression and 'Fake News', Disinformation and Propaganda* (2017) FOM.GAL/3/17, §§2(c), 2(d).

⁴⁵⁶ See, e.g., *ibid.*

⁴⁵⁷ See HRC, General Comment No. 34 (2011), §7. See ch. 1 (Introduction), s. II.1.1.2. (Obligation to prevent abuses by others).

⁴⁵⁸ UN Guiding Principles on Business and Human Rights, Principle 1.

⁴⁵⁹ *Ibid.*, Principle 3(c).

⁴⁶⁰ See ch. 3 (Hate Speech), s. V. (Approach of Private Companies to Online Hate Speech).

⁴⁶¹ See ACmHPR, Declaration of Principles on Freedom of Expression and Access to Information in Africa (2019), Principles 9(2), 17, 34, 39(4).

⁴⁶² See ACmHPR, Declaration of Principles on Freedom of Expression and Access to Information in Africa (2019), Principles 9(2), 17, 34, 39(4); UN Special Rapporteur, OSCE Representative on Freedom of the Media, OAS Special Rapporteur and ACmHPR Special Rapporteur on Freedom of Expression and Access to Information, *Joint Declaration on Freedom of Expression and Responses to Conflict Situations* (2015), §4(a). See also UN Special Rapporteur, OSCE Representative on Freedom of the Media, OAS Special Rapporteur and ACmHPR Special

12. In designing and applying their policies on misinformation, digital platforms and other media companies should be guided by international human rights standards related to freedom of expression, in line with their own undertakings and the UN Guiding Principles on Business and Human Rights.

While media companies are not directly bound by international human rights treaties, they have a responsibility to respect human rights.⁴⁶³ As set out in the UN Guiding Principles on Business and Human Rights, which all major technology companies claim to subscribe to, '[b]usiness enterprises should respect human rights. This means that they should avoid infringing on the human rights of others and should address adverse human rights impacts with which they are involved.'⁴⁶⁴ While international human rights standards have been developed to regulate the relationship between individuals and states, as sovereign, public actors, they should be applied with any necessary and appropriate adjustments by private companies to guide their own decision-making.⁴⁶⁵ It is encouraging that an increasing number of digital platforms that play a dominant role in regulating speech have accepted international human rights as part of their internal regulatory frameworks.⁴⁶⁶

Digital platforms have long resisted attempts to become 'arbiters of truth'⁴⁶⁷ but their very dominance and assumption of the role of the public square online, coupled with often serious harms caused by the spread of misinformation via these platforms and the need to respond in real-time, make it inevitable that these platforms will have to make at least some judgements on truth and falsity and what to do about speech that is clearly false. Criminal misinformation laws adopted by the state will inevitably play only a limited role in that regard; the justifiability bar for self-regulatory measures by digital platforms, such as content moderation or the suspension of user accounts, is lower than that required for the imposition of criminal or civil legal liability on the relevant individual by the state. In other words, digital platforms may in appropriate circumstances restrict false speech even when that speech is not illegal.⁴⁶⁸

Rapporteur on Freedom of Expression and Access to Information, *Joint Declaration on Freedom of Expression and 'Fake News', Disinformation and Propaganda* (2017) FOM.GAL/3/17, §3(b).

⁴⁶³ See ch. 3 (Hate Speech), s. V. (Approach of Private Companies to Online Hate Speech); UN Guiding Principles on Business and Human Rights, General Principles.

⁴⁶⁴ UN Guiding Principles on Business and Human Rights, Principle 11. See also Forum of Information & Democracy, 'Principles on Information & Democracy' (endorsed by 51 states from across the world that comprise the International Partnership for Information and Democracy).

⁴⁶⁵ For an instructive example, see the decision of the Oversight Board, 'Case Decision 2020-006-FB-FBR Claimed COVID-19 cure'.

⁴⁶⁶ See ch. 3 (Hate Speech), s. V. (Approach of Private Companies to Online Hate Speech). See also E. M. Aswad, 'The Future of Freedom of Expression Online' (2018) 17 *Duke Law & Technology Review* 26, 34.

⁴⁶⁷ T. McCarthy, 'Zuckerberg says Facebook won't be "arbiters of truth" after Trump threat' (*The Guardian*, 28 May 2020). See also M. Adler, 'Twitter Says it Stopped Policing COVID Misinformation Under Musk' (*TIME*, 29 November 2022) noting Musk's 'mission to remake the social network as a place for unmoderated speech'.

⁴⁶⁸ The Oversight Board has noted with approval the fact that the UN Special Rapporteur sees greater leeway for speech restriction by a company than by a government. See Oversight Board, 'Case Decision 2020-003-FB-UA

It is recommended that in making decisions about the necessity and proportionality of the measures they take to combat harmful misinformation, digital platforms should make clear how they reconcile divergences in international standards or transpose them to the context of non-state actors. And they should also ensure that they comply with the principles of legality and due process,⁴⁶⁹ including by being transparent in defining what type of misinformation is subject to removal or other types of penalties.⁴⁷⁰

Armenians in Azerbaijan'. See also Oversight Board, 'Case Decision 2021-001-FB-FBR Former President Trump's suspension'.

⁴⁶⁹ UN Special Rapporteur, OSCE Representative on Freedom of the Media, OAS Special Rapporteur and ACmHPR Special Rapporteur on Freedom of Expression and Access to Information, *Joint Declaration on Freedom of Expression and 'Fake News', Disinformation and Propaganda* (2017) FOM.GAL/3/17, §4(c); OAS Guide to Guarantee Freedom of Expression Regarding Deliberate Disinformation in Electoral Contexts (2019), 39.

⁴⁷⁰ UN Special Rapporteur D. Kaye, *Letter to Mark Zuckerberg dated 1 May 2019*, Ref OLOTH 24/2019, 3.



Human Rights
Institute