

Old Propaganda, New Technology: EU Governance of Deepfake Disinformation in Conflict Settings

Thi Ngoc Anh Nguyen

DOI: 10.25430/pupj-PHRG-2026-1-8

How to cite:

Nguyen T. N. A. (2026) "Old Propaganda, New Technology: EU Governance of Deepfake Disinformation in Conflict Settings", Peace Human Rights Governance, Online First Articles(1), 1-35. DOI: 10.25430/pupj-PHRG-2026-1-8

Article first published online
September 2026

*All research articles published in PHRG undergo a rigorous double-blind review process by independent, anonymous expert reviewers

Old Propaganda, New Technology: EU Governance of Deepfake Disinformation in Conflict Settings

Thi Ngoc Anh Nguyen*

Abstract

Deepfakes, understood as synthetic media created using AI techniques that replace or manipulate a person's likeness, represent a qualitative escalation in the history of disinformation. In conflict settings, their capacity to produce convincing false evidence introduces systemic threats to trust, civilian agency, and the conditions necessary for peace. This paper examines these threats within the normative framework of the right to peace, understood not merely as the absence of violence but as a positive entitlement to dignity, reliable communication, and the institutional conditions that sustain conflict resolution. Information integrity emerges as a critical conceptual bridge between this rights framework and the governance of synthetic media. The paper traces the historical trajectory of disinformation in warfare from classical propaganda to contemporary deepfake deployment. It argues that deepfake propaganda threatens to undermine the right to peace by contaminating conflict-relevant information, increasing verification burdens for minority and dissident claims, and complicating monitoring, accountability and peacebuilding processes. Against this backdrop, the paper explores the capacity of existing European Union digital legal frameworks, including the AI Act, the Digital Services Act, and the General Data Protection Regulation, to address these threats, alongside the Union's external norm-promotion policies. It concludes that while the EU's multi-layered regulatory ecosystem represents a significant advance, structural limitations persist. In addition, deepfake disinformation must be recognised not as a technological artefact but as a structural threat to peaceful relations, requiring sustained and internationally coordinated governance responses.

Keywords: *deepfake, disinformation, propaganda, information integrity, right to peace, synthetic media.*

* Human Rights Centre "Antonio Papisca", University of Padova.
Email: tna.anhnguyen@gmail.com.

1. Introduction

Propaganda has long been recognised as central to the exercise of political power. From the First World War pamphlets analysed by Harold Lasswell (1938) to the mass media spectacles dissected by Jacques Ellul (1965), the shaping of perception has been as decisive as the use of force. At its core, propaganda operates by narrowing the boundaries of what is accepted as true, persuading populations not only to consent but to internalise the legitimacy of authority. This dynamic, often described by political theorists as the ease with which the many are governed by the few (Hume 1742), has been reproduced across centuries in different communicative forms.

In the digital age, the emergence of deepfake marks a new phase in visual deception. Deepfakes employ artificial intelligence (AI) to fabricate highly realistic images, audio, and video, eroding traditional markers of authenticity (Chesney and Citron 2019). Their potential use in propaganda is twofold: they may be deployed to spread fabricated content with the intention of deception, or they may cultivate an environment of epistemic instability in which citizens lose confidence in the possibility of distinguishing the real from the false (Vaccari and Chadwick 2020). This phenomenon, sometimes referred to as the 'liar's dividend' (Chesney and Citron 2019, Schiff *et al.* 2025) not only undermines trust in information but also empowers actors who benefit from discrediting inconvenient truths. This is particularly severe in societies where political legitimacy and democratic accountability are fragile.

The intersection of deepfake technology with armed conflict has attracted growing scholarly attention. Researchers have documented the deployment of synthetic media in the Russo-Ukrainian war (Golovchenko 2022, Twomey *et al.* 2023), the Gaza war (Weimann and Weimann-Saks 2024, AL-Sahel 2025, Roumate 2025), as well as AI-generated content designed to discredit humanitarian actors and distort battlefield realities (Akkuş 2025, Ahmad *et al.* 2026). As Benkler *et al.* (2018) argue, the manipulation of information ecosystems now constitutes a form of power projection in its own right, one that operates below the threshold of conventional armed conflict while producing effects of comparable gravity. This threat has been theorised in relation to the emerging concept of information integrity, which is the condition in which public discourse is characterised by access to accurate, trustworthy, and pluralistic information (United Nations 2024). Scholars note that truthful information is a human rights concern that is foundational to the exercise of freedoms of expression, political participation, and, ultimately, to the realisation of the right to peace (Deibert 2019, Lubin 2020). The right to peace itself, while contested in international law, has gained normative traction through the United Nations (UN) Declaration on the Right to Peace (A/RES/71/189) and

the broader corpus of human rights instruments that affirm peace as both a precondition and a product of the enjoyment of fundamental rights. In that respect, deepfake propaganda emerges not merely as a regulatory problem but as a human rights violation in potential – one that demands governance responses commensurate with its structural consequences.

It is within this context that the European Union's (EU) evolving digital regulatory architecture acquires particular significance. The EU has positioned itself as the world's most ambitious jurisdiction for the governance of AI and online platforms, enacting the AI Act (Regulation (EU) 2024/1689), the Digital Services Act (Regulation (EU) 2022/2065), and operating the longstanding General Data Protection Regulation (Regulation (EU) 2016/679) as interlocking instruments of digital governance. Whether this regulatory architecture can translate into meaningful protection of information integrity in conflict settings and thereby contribute to safeguarding the human right to peace remains, however, an underexplored question in both human rights and EU digital governance scholarship. To date, European legal analysis mostly focuses on how deepfakes are defined and how disclosure obligations are framed, often noting that policy intent focuses on preventing deception while balancing freedom of expression and legitimate creative or editorial uses (Romero Moreno 2024, Łabuz 2024, Nguyen and Navdarashvili 2026).

This article examines the intersection between synthetic media, human rights, and digital governance. It explores how deepfake propaganda undermines the informational and institutional conditions necessary for peaceful relations, and assesses the extent to which existing EU legal frameworks can address these risks. Through a normative and doctrinal analysis, the study evaluates how the EU's digital regulatory architecture can mitigate the destabilising effects of malicious synthetic media in conflict settings. In doing so, it also considers the broader implications of the EU approach for global peace and human rights governance.

The article proceeds as follows. Section 2 outlines the normative framework of the right to peace and information integrity. The paper then explores the evolution of disinformation in warfare and the emergence of deepfakes as a qualitative shift in propaganda. Section 4 analyses how deepfake disinformation may affect the informational and institutional conditions on which public trust, accountability and peacebuilding depend, focusing on conflict-relevant information, verification burdens for minority and dissident claims, and the work of monitoring, accountability and peacebuilding institutions. The following section analyses how the EU's legal framework can address these challenges. Finally, the article concludes by reflecting on the broader implications for global peace and human rights governance.

2. The Right to Peace and Information Integrity

This section explores the normative foundations of the right to peace and the emerging concept of information integrity as a material and institutional condition for peaceful social and political relations. It provides the analytical framework for examining how deepfake disinformation in conflict settings can be understood as a human rights concern.

2.1. The Right to Peace as a Collective Human Right

Whether the right to peace constitutes an autonomous individual and collective human right remains contested in international law. Peace is undoubtedly a foundational objective of the UN Charter and is recognised as a fundamental precondition for all human rights (Ivankiv 2020). However, this does not by itself establish a universally recognised, directly enforceable individual entitlement to peace. The legal status, content, and bearers of such a right remain subjects of substantial scholarly and institutional disagreement. Early studies, such as Alston (1980), question whether the language of rights could meaningfully be extended to peace without obscuring the legal content of more established rights. Other scholars have defended peace as a human right, while recognising that its legal character, beneficiaries, and institutional implications remain unsettled (Hayden 2002, Ife 2007, Schabas 2011). Some further argue that the right to peace remains a nascent norm that is often proclaimed but less frequently enforced (Woods 2013, Guillermet Fernandez and Puyana Fernandez 2017).

The idea of a right to peace has nevertheless developed through a series of international and regional instruments. At the universal level, the Declaration on the Preparation of Societies for Life in Peace (1978), the Declaration on the Right of Peoples to Peace (1984), and the Declaration on the Right to Peace (2016), adopted by the UN General Assembly, articulate peace as a condition in which human rights, development, and international cooperation can be realised. These instruments are normatively significant, but they are soft-law declarations. They do not, without further legal basis, establish a universally enforceable individual right to peace.

Regional developments are similarly uneven. Article 23 of the African Charter on Human and Peoples' Rights (1981) provides the strongest treaty-based formulation by recognising the right of all peoples to national and international peace and security. Its protocol on the Rights of Women in Africa (2003) further connects women's rights to peace and security. These provisions create binding obligations for relevant States Parties, yet their collective and regionally specific formulation should not be treated as

evidence of a settled universal right held by every individual. The ASEAN Human Rights Declaration (2012) also affirms that every person and the peoples of ASEAN have the right to enjoy peace, but its declaratory status limits its enforceability. In contrast, the European and Inter-American human rights systems protect a range of rights and institutional guarantees conducive to peaceful societies, without recognising an autonomous right to peace.¹

On that account, the article does not proceed on the assumption that the right to peace is a settled, universally justiciable legal entitlement, nor that every instance of deepfake disinformation constitutes a direct violation of that right. Instead, it uses the right to peace as a normative and interpretive framework. This framework directs attention to the conditions under which other human rights can be exercised and peace processes can function, including access to information, freedom of expression, equality and non-discrimination, political participation, civilian protection, and the integrity of institutions responsible for accountability and conflict resolution.

The article understands the relationship between peace and human rights as mutually constitutive rather than strictly hierarchical. Accordingly, peace is more than just the absence of armed conflict. It also depends on conditions requiring justice, dignity, trust, and meaningful communication through which individuals and communities can coexist, make claims, and resolve disputes without coercion. In other words, a substantive conception of peace identifies the social and institutional conditions through which peaceful relations can be sustained.

On this basis, trust-building is the cornerstone of sustainable peace, which remains fragile where deep social divisions persist even in the absence of active hostilities (Cockell 2000, Cromwell and Vogeles 2009, Fontecha-Tirado 2018, Gehrig *et al.* 2023). Free and reliable information flows are essential to building this trust and dignity. When citizens have access to truthful information and feel their voices are heard, it fosters social cohesion that defines peace (Cox and Sisk 2017, Nesterova and Kim 2024). Conversely, systematic information manipulation and the denial of verifiable evidence sow the seeds of suspicion and grievance that can lead to violence. Information integrity thus functions as a material and institutional condition for sustaining peaceful social and political relations.

¹ For example, the right to freedom of peaceful assembly and association is a fundamental right guaranteed by the EU Charter of Fundamental Rights (Article 12) and the European Convention on Human Rights (Article 11). It is also covered under Articles 15 of the American Convention on Human Rights.

2.2. Information Integrity as a Prerequisite for Peace

Information integrity refers to the accuracy, consistency, and trustworthiness or reliability of information circulating in society (United Nations 2024). The UN has recently highlighted information integrity as fundamental to achieving peaceful, inclusive societies. In 2024, the UN released Global Principles for Information Integrity, noting that the erosion of the integrity of information ecosystems undermines people's ability to exercise their human rights and 'hampers efforts to achieve peace' (United Nations 2024, 3). Strengthening information integrity is seen as an urgent challenge, as today's digital environment enables misinformation and disinformation to spread at unprecedented 'volume, velocity and virality' (United Nations 2024, 3). The UN's framework suggests that a healthy information space where accurate, reliable information is accessible to all and freedom of expression is protected supports human rights and peaceful societies (United Nations 2024, 3).

In practical terms, promoting information integrity means ensuring people can exercise their right to seek and receive truthful information without manipulation or interference – a right grounded in Article 19 of the UDHR. During conflicts or crises, this principle becomes even more critical because reliable, factual communication can alleviate tensions and enable dialogue, whereas deceptive or distorted information can ignite distrust and fear. Peace negotiators often stress that ending violent conflict requires establishing truth and trust between former adversaries (Ward 1989, Alleyne 2003). We can thus view truthful information as the lifeblood of peace or a vital precondition for rebuilding peaceful relations after periods of violence. Keeping that in mind, this article examines deepfake disinformation as a practice that can undermine the information environment on which civilians, minority communities, journalists, humanitarian actors, and international institutions depend.

3. From Deepfakes to Propaganda: Disinformation in Warfare

As media environments become increasingly decentralised and digitally mediated, deepfakes amplify long-standing techniques of propaganda while also introducing novel risks. Understanding this intersection requires revisiting the theoretical traditions that have explained propaganda as a technology of power and examining how these traditions illuminate the challenges posed by synthetic media in the twenty-first century.

3.1. A Historical Overview of Propaganda and False Information in War

The use of disinformation as a weapon of war is as old as warfare itself. Throughout history, combatants have spread propaganda, rumors, and outright lies to weaken their enemies' resolve and sway public opinion. Classic examples range from ancient stratagems to drop morale to the extensive propaganda campaigns of the World Wars. In World War I, governments on all sides used posters and press releases to demonise the enemy and bolster home front support (Kingsbury 2010). In World War II, radio broadcasts by figures like 'Tokyo Rose' and Nazi propaganda minister Joseph Goebbels disseminated misleading narratives on a mass scale (Richmond 2007, Keene 2008). The Cold War era saw sophisticated 'psychological operations' and 'active measures' by intelligence services (Taylor 2010). For instance, the Soviet KGB spread false stories like the infamous Operation INFEKTION, which propagated the myth that the U.S. created the AIDS virus to undermine adversaries (Geissler and Sprinkle 2013, Reader 2020).

By the late 20th and early 21st centuries, the advent of the internet and social media enabled 'fake news' and digital disinformation to spread rapidly in conflict zones. Examples include fabricated stories during the Balkan wars of the 1990s (Cappello and Sunter 2019, Ajšić 2025) or false reports about weapons of mass destruction in the Iraq War (O'Shaughnessy 2004). International law has long tolerated certain forms of military deception, permitting ruses of war such as camouflage and disinformation directed at enemy forces, provided they do not amount to perfidy, such as the misuse of protected symbols.² However, the information battlespace today extends far beyond leaflets and radio broadcasts since it encompasses social media virality and AI-generated fake content that can reach millions of civilians instantaneously. This magnifies the potential harm of disinformation, blurring the line between permissible psychological tactics and harmful violations of the public's right to truth.

3.2. Deepfakes as a Qualitative Leap in Digital Propaganda

In recent years, deepfake technology has emerged as a gamechanger in information warfare. Unlike traditional propaganda that might distort facts or use misleading words, deepfakes enable the fabrication of convincing visual evidence and personal avatars. A deepfake can literally put false

² See Protocol Additional to the Geneva Conventions of 12 August 1949 (Protocol I), arts. 37(1)–(2); International Committee of the Red Cross (2005) *Customary International Humanitarian Law*, Rules 57 and 65.

words into a real person's mouth on video, making it appear that a political or military leader made statements or took actions that never occurred. This marks a qualitative leap in deception: whereas classic 'fake news' might involve writing a false article or doctoring a single photo, deepfakes produce immersive, hard-to-disprove simulations of reality. Therefore, highly realistic fake videos could take online disinformation to the next level, becoming the next propaganda frontier if left unchecked (Chesney and Citron 2019, Tomić *et al.* 2023, Argent 2025). The cost and skill barriers to creating deepfakes have dropped rapidly (Bernaciak and Ross 2022, Alanazi and Asif 2024), meaning not only state actors but also militias, terrorist groups, or even lone agitators can harness this technology.

Moreover, deepfakes excel at evoking emotional reactions, for example, seeing and hearing a trusted figure apparently betray you, or witnessing (on video) an alleged atrocity, can be far more impactful than reading a rumor. Audio-visual realism interacts strongly with the 'seeing is believing' heuristics (Sundar *et al.* 2021), thus deepfakes shift the credibility problem from 'textual falsity' toward contested authenticity in audio-visual evidence (Vaccari and Chadwick 2020). In conflict settings, this opens the door to dangerous psychological operations: an aggressor can fabricate videos to mislead soldiers and civilians, incite hatred, or create false justifications for violence. Deepfakes, therefore, represent a technological modality through which disinformation can be produced and disseminated. When deepfakes are intentionally deployed to influence political attitudes or behaviour, particularly in conflict settings, they function as tools of propaganda. It is noted that while propaganda may include truthful or selectively framed information, this article focuses specifically on deepfake disinformation as a technologically mediated subset of propaganda characterised by the deliberate fabrication of audio-visual evidence.

The malicious possibilities of deepfake disinformation are endless, ranging from a fake video of a general announcing surrender to bogus footage of an enemy atrocity used to justify retaliation to forged clips of peace negotiators appearing to break faith. As we will see, some of these scenarios are not merely hypothetical but have already been attempted in current conflicts.

3.3. Illustrative Examples of Deepfake Disinformation in Recent Conflicts

The ongoing war in Ukraine and recent conflicts in the Middle East provide striking real-world examples of how propaganda and deepfakes are deployed today. On March 18, 2022, during Russia's invasion of Ukraine, a deepfake video of Ukrainian President Volodymyr Zelensky urging Ukrainian troops

to lay down their arms and surrender was released online and even briefly broadcast on a hacked TV station (Allyn 2022). The synthetic media mimicked Zelensky's likeness and voice, creating a realistic illusion that the Ukrainian president had capitulated to Russia (Linehan *et al.* 2023, Samaha 2025). This incident was quickly exposed when Zelenskyy responded within hours by releasing a real video of himself resolutely denying the fake surrender and urging his people to keep fighting (Allyn 2022). A year later, in November 2023, a more sophisticated deepfake targeted Ukraine's Commander-in-Chief Valerii Zaluzhnyi. In that AI-generated video, Zaluzhnyi appeared to announce a coup d'état against Zelenskyy after a supposed resignation (Osadchuk 2024, Slychko 2025). This fake video, which circulated on Telegram channels aligned with the Kremlin, aimed to fabricate a rift in Ukraine's leadership and erode public trust in the government. Notably, it is observed that the 2023 deepfake was far more convincing than the cruder 2022 attempt, with the visuals and audio being of higher quality, and only subtle glitches like slight lip-sync errors (Majchrzak 2023, Osadchuk 2024, Slychko 2025). Many casual viewers could not easily tell it was fake. The intent was clearly to sow confusion and the belief that Ukraine's government was fracturing from within (Majchrzak 2023, Osadchuk 2024, Slychko 2025).

Another example comes from the Israel–Hamas conflict in Gaza in 2023. Observers noted an unprecedented flood of AI-generated fake images and videos amid the information chaos of that war (Klepper 2023, Tweissi 2025). One set of highly viral images, viewed millions of times, showed scenes of 'bloodied, abandoned infants' amid the rubble of Gaza, presumably to accuse one side of heinous acts. These horrific pictures were later confirmed as deepfakes created with AI, containing tell-tale anomalies such as unnatural lighting in eyes or distorted fingers (Klepper 2023). Nonetheless, they succeeded in provoking real outrage and panic before being debunked. Both pro-Israeli and pro-Palestinian social media circles circulated various AI-fabricated visuals purporting to show atrocities by the other side, from images of babies in rubble to videos of Israeli missile strikes or tank incursions that never actually occurred (Klepper 2023, Klepper and Lemon 2023). Each fake was designed to ignite emotions and harden hatred, offering seemingly damning evidence that could be held up as proof of the opponent's barbarity. The proliferation of such false images has added a new layer of alarm to an already bloody conflict. In addition to completely fabricated images, simpler forms of disinformation (sometimes called 'cheap fakes') also abounded, for example, authentic photos from other countries or past events were recycled and falsely relabeled as scenes from the current war (Klepper 2023). Old-fashioned propaganda techniques like miscaptioning and rumor-mongering

now operate in tandem with cutting-edge AI fakery, creating a perplexing mix of truth and lies on the information battlefield.

Beyond these cases, deepfakes also have the potential to cause disaster and violence in other hypothetical scenarios. In West Africa, a suspected deepfake video of Gabon's President in 2019 showing him in a strangely stiff, possibly doctored state led to rumors about his health and even triggered an attempted military coup by officers who doubted the president's capacity (Cahlan 2020). In Myanmar, in 2021, the post-coup military regime broadcast what many believe was a manipulated or deepfaked video of a detained government minister confessing that he had bribed ousted leader Aung San Suu Kyi – a convenient claim that the junta used to justify its charges against her (AFP 2021). While it is unclear if that video was a true deepfake or a coerced statement, it exemplified how false video evidence can be deployed to discredit political opponents and lend credence to a regime's narrative. One can easily imagine future conflicts where a fake video of a massacre is released to justify retaliation, or a fake audio recording of peace negotiators suggesting treason torpedoes delicate negotiations. This further creates a pernicious dilemma for journalism, as on the one hand, if media become overly cautious, real atrocities might go unreported; on the other hand, if they are not cautious enough, they risk amplifying lies. In fact, news organisations have hesitated to broadcast breaking news video from war zones out of fear it might be AI-fabricated (Byman *et al.* 2023).

4. Deepfakes as a Threat to the Right to Peace

This section examines how deepfake disinformation in conflict settings may affect the informational and institutional conditions on which legitimate and effective responses to conflicts depend. It considers three interlocking dimensions: the contamination of conflict-relevant information, increased verification burdens for minority and dissident claims, and complications for international peace organisations.

4.1. Undermining Access to Reliable Conflict-Relevant Information

As discussed in Section 3.3, documented incidents from Ukraine and Gaza show synthetic media circulating as fabricated military communication and purported evidence of atrocities (Allyn 2022, Klepper 2023). These examples denote how such content can enter conflict-related information environments relied upon by civilians, journalists, and humanitarian actors to interpret events and assess risks.

Viewed through the right-to-peace framework developed above, deepfake disinformation may weaken the reliable communication and trust that support the exercise of established rights and peaceful societies. In times of conflict, civilians depend on accurate information about immediate dangers, alleged atrocities, and the credibility of ceasefire or negotiation announcements. This interest is connected to freedom of expression, which includes the freedom to seek and receive information and the right to truth in contexts of gross human rights violations. By introducing fabricated audiovisual content into these information environments, deepfakes can make it more difficult to assess the veracity of conflict-relevant claims. The UN has warned that ‘the erosion of information integrity’ through disinformation and AI-generated content can ‘undermine people’s ability to exercise human rights’ and even ‘hamper efforts to achieve peace’ (United Nations 2024, 3).

Research on deepfakes and the ‘liar’s dividend’ indicates that fabricated content may reinforce prior beliefs and weaken confidence in visual evidence, with foreseeable consequences for conflict-related decision-making (Vaccari and Chadwick 2020, Twomey *et al.* 2023). Deepfakes may also reinforce existing biases where audiences are inclined to accept content that confirms their prior suspicions (Byman *et al.* 2023). Repeated uncertainty around visual evidence can increase fear, slow verification, and reduce confidence in sources during acute crises (Kuźnicka-Błaszowska and Kostyuk 2025). Fabricated announcements concerning surrender, ceasefires, or imminent attacks could also create particular risks for civilian safety and public trust. Such risks can affect civilians’ ability to seek and receive information, which forms part of the broader conditions necessary for informed participation and peaceful relations.

4.2. Increasing Verification Burdens for Minority and Dissident Claims

Deepfakes technologies can increase verification burdens for minority communities and dissident or sub-state groups seeking justice or self-determination. In many conflicts, affected individuals and communities rely on user-generated material, including mobile-phone footage, testimony, and social media posts, to document alleged abuses (Ristovska 2019, Ulbricht *et al.* 2022). Such material has become increasingly important in accountability processes, for instance, proceedings involving footage recorded by Syrian civilians (Shaheen 2015, Ristovska 2019). In this context, deepfakes do not make this material unreliable in itself. But the availability of convincing synthetic media can give states and other actors a plausible basis for challenging the authenticity of evidence and dismissing inconvenient material

as fabricated (Vaccari and Chadwick 2020, Pfefferkorn 2021). In *Ukraine and the Netherlands v Russia* before the European Court of Human Rights, Russian representatives asserted that visual evidence of their wrongdoing might have been fabricated³, seeking to cast doubt on the integrity of the fact-finding process (Llorente and McDermott 2024).

These challenges fall unevenly. Communities with limited access to digital-forensic expertise may face greater delays and higher evidentiary demands when presenting claims, particularly where their accounts already receive unequal credibility. As Pfefferkorn (2021) notes, marginalised groups are often already less likely to be believed. The wider availability of deepfakes may make authentic material harder to verify and delay the recognition of credible claims. A broader retreat from reliance on video evidence could therefore disproportionately impede the recognition of genuine grievances.

4.3. Complicating Verification by International Peace Organisations

International institutions involved in peace and justice, including UN bodies, regional organisations and courts, depend on timely and credible information to monitor violations, assess compliance with ceasefires, investigate crimes and support accountability processes. In this context, user-generated evidence like smartphone videos from war zones has become an increasingly important source of information in these activities (Peake 2024). The availability of deepfakes will increase the importance of establishing provenance, corroborating evidence and assessing whether audiovisual material has been digitally manipulated.

Indeed, deepfakes pose an emerging evidentiary concern for the International Criminal Court and other criminal tribunals, particularly in relation to the authentication of digital material from conflict settings (Llorente 2024, Piers and Osaer 2025).⁴ Where parties contest the authenticity of audiovisual evidence, investigators may need to devote additional time and resources to corroboration and digital-forensic analysis. This can delay fact-finding and create uncertainty around the public reception of institutional findings. As discussed above, Russia's challenge to visual evidence in *Ukraine*

³ Russia challenged a photo of Russian soldier Stanislav Tarasov, used to geolocate him to Pavlovka during 2014 attacks, as manipulated. Russian officials claimed visual 'signs of fakery', such as a halo around the figure and allegedly cloned tanks, to cast doubt on the authenticity of digital evidence.

⁴ Similar arguments have also appeared in domestic criminal proceedings, including prosecutions arising from the January 6 Capitol riot in the US, where defendants have questioned whether authentic video evidence could be a deepfake (Jenkins et al. 2024, Cohen 2025).

and the Netherlands v. Russia illustrates how allegations of fabrication may be used to contest the integrity of a fact-finding process.

International organisations have also distinguished synthetic media as a prospective challenge to conflict monitoring and institutional trust. The OSCE recognised disinformation as an obstacle to objective monitoring in Eastern Ukraine (Tsymbaliuk 2021). In December 2021, the organisation held expert meetings on ‘Deepfake News’ as a multilateral policy challenge, further acknowledging that coordinated AI-driven falsehoods can hinder conflict resolution efforts (OSCE Representative on Freedom of the Media 2021, Tsymbaliuk 2021). Similarly, NATO strategic communications experts have cautioned that deepfakes may foster distrust among Allies and between NATO and local populations in conflict zones (NATO 2024, Bjola 2025).

These assessments point to a plausible risk pathway through which synthetic media may increase verification burdens, weaken confidence in institutional findings and complicate the information environment on which international responses to conflict depend. Busch and Ware (2023, 3) describe the more extreme prospect of an ‘infocalypse’, in which trust in audiovisual evidence is broadly eroded. Although current evidence does not establish that this condition has materialised in a particular conflict, the rapid development of synthetic-media technologies underscores the importance of institutional capacities for verifying and contextualising digital evidence.

5. EU Digital Laws as a Limited Architecture for Traceability and Distributional Governance

Building on the preceding analysis, this section examines how the AI Act, the Digital Services Act (DSA) and the General Data Protection Regulation (GDPR) can address deepfake disinformation in conflict settings. It assesses their respective contributions and limitations in supporting information integrity and the conditions necessary for peaceful relations before considering the conditional reach of EU external action and regulatory diffusion beyond the Union’s borders.

5.1. The AI Act

The AI Act is the world’s first comprehensive binding legal framework for AI (European Commission 2026a). Its most directly relevant provisions concern transparency obligations for AI-generated and manipulated content. Under Article 50(2), providers of AI systems capable of generating synthetic audio, image, video or text must ensure that outputs are marked in a machine-readable format and detectable as artificially generated or

manipulated. Article 50(4) further requires deployers to disclose where image, audio or video content constitutes a deepfake, subject to specified exceptions for artistic, creative, satirical or fictional works.

To support compliance with these obligations, the AI Office has developed the Code of Practice on Transparency of AI-Generated Content (European Commission 2026b). The Code provides more detailed guidance on the marking and detection of AI-generated or manipulated content by providers, as well as the labelling obligations of deployers under Article 50. It also seeks to encourage more consistent technical practices across the AI value chain.

Nonetheless, applicable from 2 August 2026, the transparency obligations under the AI Act may make it more difficult for synthetic media to circulate as authentic documentation of military developments, alleged atrocities or political communications. Their practical contribution will nevertheless depend on whether markers and labels retain their integrity as content circulates across platforms, and on effective enforcement throughout the services involved in its production and dissemination.

It should be noted that the technical specifications do not turn Article 50 into an authenticity guarantee. Watermarks, metadata and other machine-readable signals may indicate that content has been generated or manipulated by AI, but their reliability depends on whether they remain available throughout the content's circulation. Existing research shows that watermark-based detection can be evaded through post-processing, while deepfake-detection systems may be vulnerable to adversarial manipulation and may fail to identify material generated through new techniques (Cao and Gong 2022, Jiang *et al.* 2023, Zhao *et al.* 2025). Content generated through non-compliant systems or distributed outside regulated technical chains may contain no reliable marker. The presence or absence of a marker cannot establish whether a depicted event occurred, identify the original source of the material, or determine its political purpose.

These limitations carry particular significance in conflict settings. On one hand, false negatives may permit fabricated military communications, purported atrocity footage, or manipulated ceasefire announcements to circulate without warning. On the other hand, false positives can cast doubt on authentic documentation produced by civilians, journalists, or human rights defenders. That being the case, Article 50 and Code-based signals should function as prompts for further verification and contextual assessment. We should not consider them as conclusive evidence that a material is false, unlawful, or unworthy of circulation.

The technical limitations are also distinct from the question of how users interpret and respond to transparency signals. Even where labels are available and accurately applied, counter-disinformation signals can cause legitimacy

contests, especially when users perceive interventions as politically selective (Freeze *et al.* 2021). Taking labelling as an example, Martel and Rand (2023) find that disinformation warning labels can, on average, reduce belief in and sharing of disinformation, although effects vary by design choices, content characteristics, and user factors. Selective warnings can increase the perceived accuracy of unlabelled items through the ‘implied truth effect’ (Pennycook *et al.* 2020), while general warnings can reduce trust in valid information through the ‘tainted truth effect’ (Freeze *et al.* 2021). Aslett *et al.* (2022) similarly report limited average effects of credibility labels in some contexts.

Furthermore, the AI Act’s tiered risk classification also has a limited reach in relation to conflict-related deepfake propaganda. Deepfakes and synthetic propaganda are not, in themselves, classified as a distinct category of high-risk AI systems. Although certain AI systems intended to influence electoral outcomes or voting behaviour may fall within the high-risk framework, conflict-targeted propaganda does not automatically meet this criterion.

Article 5 may apply in more exceptional circumstances. It prohibits AI systems that use subliminal, purposefully manipulative or deceptive techniques to materially distort a person’s behaviour and cause, or are likely to cause, significant harm. A system used to generate and disseminate deepfake propaganda to incite violence or manipulate civilian behaviour could potentially meet this threshold where these cumulative conditions are satisfied. Its application would depend on the system’s intended use, the nature of the manipulation and the likelihood of significant harm in the specific circumstances. This narrow route does not provide a general prohibition on harmful synthetic content or conflict-related propaganda. As Nguyen and Navdarashvili (2026, 199) observe, the AI Act’s approach to risk categorisation may provide limited protection for broader societal interests. The absence of a distinct category for malicious deepfakes therefore raises questions about the Act’s capacity to address societal harms associated with synthetic propaganda (Romero Moreno 2024, Łabuz 2024, Nguyen and Navdarashvili 2026).

Governance under the AI Act is entrusted to a multi-level architecture involving national competent authorities and, at the Union level, the European AI Office. The AI Office coordinates implementation, supports the development of codes of practice, and may provide guidance on the application of transparency obligations to general-purpose AI models. This institutional architecture enables Article 50 to be further specified as synthetic-media technologies evolve.

In this respect, the AI Act operationalises what Calo (2017) has described as the ‘notice’ function of law: alerting users, platforms and public authorities

to the artificial or manipulated nature of content. In conflict settings, this function can support informed civilian agency and verification, although it cannot itself establish authenticity or resolve contested factual claims.

5.2. The Digital Services Act

The DSA introduces a binding framework for online intermediaries, with the most stringent requirements applied to Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs)⁵. It addresses the distributional conditions through which synthetic media can acquire systemic significance. Article 34 requires VLOPs and VLOSEs to assess systemic risks arising in the Union from how their services, including their algorithmic systems, are designed and used. These risks include actual or foreseeable negative effects on fundamental rights, civic discourse, electoral processes, and public security.

The DSA's relevance to deepfake disinformation arises from the broader risk categories set out in Article 34. Risk assessments must consider the role of recommender systems, content moderation, advertising systems, and platform terms and conditions. They must also examine whether risks are influenced by intentional manipulation of the service, including inauthentic use, automated exploitation and the rapid dissemination of illegal content or information incompatible with platform rules.

Once a relevant systemic risk has been identified, Article 35 requires VLOPs and VLOSEs to adopt reasonable, proportionate and effective mitigation measures, with particular regard to their impact on fundamental rights. The non-exhaustive list of possible measures includes adjustments to recommender systems, content moderation processes, advertising systems, internal resources, and cooperation with trusted flaggers. Article 35(1)(k) further requires, where applicable, 'prominent markings' for manipulated or generated audiovisual content that falsely appears authentic or truthful as a possible mitigation measure, together with an easy-to-use functionality enabling recipients to indicate such content. Transparency and audit obligations under Articles 24, 42 and 37 provide further avenues for public and regulatory scrutiny of the measures adopted.

In conflict settings, the DSA's contribution is based on its capacity to address the conditions through which synthetic media is distributed, amplified and

⁵ Article 33(1) of the DSA classifies online platforms and search engines with more than 45 million monthly active users as VLOPs and VLOSEs. See European Commission (2026d) 'Supervision of the designated very large online platforms and search engines under DSA', retrieved from: <https://digital-strategy.ec.europa.eu/en/policies/list-designated-vlops-and-vloses> (accessed: 14/04/2026) for the updated list of designated VLOPs and VLOSEs.

made credible. Thus, a conflict-sensitive systemic-risk assessment should consider more than the volume of detected deepfakes. It is necessary to assess whether platform features amplify manipulated material, including recommender systems, advertising infrastructures, sharing functions, and reporting mechanisms. It is also crucial to anticipate how such material reaches affected populations, diasporic communities, and audiences exposed to conflict-related information operations.

Such assessment must also account for the rights implications of mitigation. AI-generated material is not necessarily unlawful, and automated detection may misidentify authentic civilian documentation, journalism or political speech. Under those circumstances, platforms should treat synthetic-media indicators as one element of a broader assessment supported by contextual review, relevant linguistic expertise, transparent explanations and meaningful avenues to challenge inaccurate labelling or removal decisions. These safeguards connect the DSA's systemic-risk framework to the verification burdens and institutional concerns identified in Section 4.

Article 36 of the DSA is also relevant to conflict-related disinformation because it establishes a crisis response mechanism,⁶ alongside the voluntary crisis protocols provided for Article 48. Where extraordinary circumstances create a serious threat to public security or public health in the Union or in a significant part of it, the Commission can request VLOPs and VLOSEs to assess whether their services contribute to the threat and to apply specific mitigation measures. As a result, the mechanism can address conflict-targeted deepfake campaigns that produce demonstrable public-security consequences within the Union. Its reach nevertheless remains limited for campaigns directed principally at populations in third-country conflict zones. Even where the relevant content circulates through EU-regulated platforms, Article 36 requires a serious threat within the Union or a significant part of it (Buijs and Buri 2023).

The 2022 Code of Practice on Disinformation, integrated into the DSA framework as the Code of Conduct on Disinformation, complements these statutory obligations through a co-regulatory approach (European Commission 2026e). However, independent monitoring has identified

⁶ The crisis response mechanism was developed in the context of Russia's full-scale invasion of Ukraine and the associated manipulation of online information. The Council described the mechanism as a response to the particular impact of the aggression on the online information environment, while Recital 91 expressly identifies armed conflicts among the circumstances that may generate a crisis under Article 36. The EU's earlier suspension of RT and Sputnik broadcasting activities formed part of a separate sanctions response to Russian information manipulation and military aggression (Council of the EU 2022a, 2022b).

uneven reporting quality and ongoing gaps in transparency and access for scrutiny (Dragomir *et al.* 2024, Botan and Meyer 2025, Böswald 2025).

In the final analysis, the DSA's effectiveness remains contingent on the quality of platform risk assessments, the resources dedicated to mitigation and the capacity of regulators and external researchers to scrutinise implementation. Platforms face practical difficulties in identifying manipulated content originating from user uploads or distributed across services outside the EU. The systemic-risk regime also leaves substantial discretion to platforms in defining risks and selecting mitigation measures, raising concerns that commercial incentives may encourage formal or performative compliance (Meßmer and Degeling 2023, Chapman 2025, Reich and Calabrese 2025). Windwehr (2025) identifies a 'critical lack of standardized processes, assessment methodologies and reporting templates' in the DSA's risk governance approach. In conflict settings, these limitations heighten the need for timely contextual assessment and safeguards against the wrongful suppression of authentic documentation and protected expression.

5.3. The General Data Protection

The GDPR has a limited but distinct role where deepfake disinformation exploits an identifiable person's 'personal data'⁷. Where the creation or dissemination of synthetic media entails the processing of personal data, including images or voice recordings used to generate fabricated content, that processing must comply with the GDPR's requirements of lawfulness.⁸ In conflict settings, the covert collection and synthetic manipulation of individuals' images or voices for propaganda purposes may therefore constitute unlawful processing in the absence of a valid legal basis.

The GDPR's contribution is primarily remedial. Individuals may request the erasure of unlawfully processed personal data under Article 17 (right to erasure), object to processing in the circumstances specified by Article 21, and lodge complaints with the competent supervisory authorities. In *TU and RE v Google LLC* (C-460/20), the Court of Justice further recognised de-referencing as a possible remedy where a data subject can substantiate the manifest inaccuracy of indexed information.⁹ These avenues offer a measure

⁷ 'Personal data' is defined as 'any information relating to an identified or identifiable natural person' (Article 4 GDPR).

⁸ Processing of personal data requires a lawful basis under Article 6 of the GDPR. In addition, where images, voice recordings or related data are processed as biometric data for the purpose of uniquely identifying a natural person, Article 9 GDPR may additionally apply and generally prohibits processing unless one of its specified exceptions is satisfied.

⁹ In *TU and RE v Google LLC* (C-460/20), the Court held that a search-engine operator must grant a de-referencing request where the data subject provides relevant and sufficient

of redress where deepfake content impersonates or falsely attributes conduct to an identifiable person and where the relevant controller, search engine operator or platform falls within the GDPR's territorial scope.

Nonetheless, the GDPR does not directly address emergent data issues, such as AI-driven profiling or algorithms, beyond treating them as generic automated processing subject to human review under Article 22 (El-Gazzar and Stendal 2020). It cannot prevent coordinated propaganda campaigns, and enforcement against state or non-state actors operating from conflict zones or adversarial jurisdictions is likely to be limited. Thus, the data protection regulation mainly provides a supplementary avenue for individual redress in cases of identity-based deepfake harm.

5.4. EU External Action and the Conditional Reach of Regulatory Diffusion

This article argues that EU regulatory instruments have the possibility, though limited, to influence information integrity beyond the Union's borders. Through their respective territorial scopes, the DSA Act and the AI Act provide incentives for online platforms and AI developers to implement transparency, labeling, and risk mitigation measures across a wide range of their services. Such product-level spillovers can extend certain safeguards to users outside the EU where providers operate integrated global systems. Their practical value, however, depends on timely implementation, adequate resourcing, and the capacity to respond to local information environments.

The Brussels Effect¹⁰ offers one account of how EU regulation can influence corporate practices and regulatory developments beyond the Union (Bradford 2020). Globally operating firms may have incentives to develop common compliance infrastructures where product architecture, organisational structures and market incentives favour uniform approaches (Christakis 2020). In the context of deepfake disinformation, this move could extend transparency and risk-mitigation practices to information environments in third countries where EU-regulated platforms play a significant role in public communication. Nevertheless, the operation of the Brussels Effect remains conditional. Recent scholarship cautions that regulatory fragmentation,

evidence capable of establishing that the indexed information is manifestly inaccurate, without generally requiring a prior judicial ruling against the original publisher. The remedy concerns the continued discoverability of material through search results, rather than removal of the original content, and remains subject to a balance with freedom of expression and information.

¹⁰ The Brussels Effect refers to the tendency of EU regulations to produce de facto global standards as multinational companies and third-country regulators align with EU requirements in order to access European markets (Bradford 2020).

divergent compliance strategies and competing geopolitical or commercial incentives may perhaps constrain the global reach of EU rules, particularly in rapidly developing technological sectors (Almada and Radu 2024, Bastos Lima and Schilling-Vacaflor 2024).

The preceding analysis of Ukraine and Gaza illustrates the practical significance of these limits. In Ukraine, EU regulatory leverage has a closer connection to the conflict-related information environment where campaigns circulate through platforms serving EU recipients and may generate security consequences within the Union. In Gaza, EU rules may influence the policies of global platforms, but they cannot directly govern the actors responsible for information manipulation or guarantee meaningful protection for affected populations. This suggests that regulatory diffusion only supports information integrity where it is accompanied by timely implementation, local-language expertise, independent verification capacities, meaningful access to platform data and safeguards for freedom of expression.

In addition, EU external action can also support the uptake of information-integrity safeguards in conflict-affected settings. The East StratCom Task Force (European External Action Service 2025) and EUvsDisinfo (European External Action Service 2019) have documented and challenged Kremlin-aligned information manipulation campaigns¹¹ since 2015. Although EUvsDisinfo operates primarily as an analytical and public awareness initiative, it provides a model for publicly accountable monitoring and verification work. The EU further supports information resilience through development cooperation and peacebuilding instruments, including the Instrument for Pre-accession Assistance (IPA III) (European Commission 2025) and the Neighbourhood, Development and International Cooperation Instrument (NDICI-Global Europe) (European Commission 2026c).

These initiatives complement the AI Act and DSA by strengthening the local and institutional capacities required for effective information-integrity protection. It is also important to sustain cooperation with international

¹¹ The EU's focus on Kremlin-aligned information manipulation forms part of its wider response to Russia's destabilising hybrid activities. In April 2026, the Council imposed restrictive measures on Euromore and Pravfond, describing them as entities that disseminate or reinforce Russian propaganda and disinformation targeting European audiences and Ukraine. The measures were adopted under the EU framework established in October 2024 to address Russian actions undermining the Union's and its Member States' security, stability and democratic institutions (Council of the EU 2026).

organisations,¹² independent media,¹³ civil-society organisations¹⁴ and humanitarian actors¹⁵ in order to connect EU regulatory leverage with the realities of conflict monitoring and peacebuilding beyond the borders.

6. Conclusion

This paper argues that deepfake disinformation in conflict settings constitutes not only a technological challenge but also a structural threat to the human right to peace. Synthetic media disintegrates the epistemic foundations on which peacebuilding is built. The article identifies three major consequences of deepfake disinformation in a conflict setting. First, it can contaminate the information on which civilians rely to assess risks and make decisions. Second, it can increase verification burdens for minority and dissident communities seeking recognition of their collective rights and political claims. Finally, it can complicate the work of multilateral institutions involved in international responses to conflict. These dynamics establish deepfake disinformation as a threat not only to individual dignity but also to the structural preconditions of international peace and security.

The article then analyses the EU's legal framework as a governance response to these risks. In general, the current digital legal framework imposes accountability on platforms, creates detection and disclosure incentives for AI developers, and extends protective norms beyond EU borders through its extraterritorial effects. Yet structural limitations remain. Enforcement against state and non-state actors operating from conflict zones or adversarial

¹² Relevant reference points include UNESCO's Internet Universality ROAM-X Indicators, which support rights-based and evidence-informed assessments of national information environments, and the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, which provides for transparency and oversight, risk and impact management, digital literacy and international cooperation. These frameworks can inform capacity-building and rights-sensitive responses to synthetic media, although neither provides a conflict-specific mechanism for deepfake governance.

¹³ See, for example, Bellingcat, a Netherlands-based open-source investigative organisation whose work has included the verification and documentation of Russian military activity and alleged war crimes in Ukraine, and The Kyiv Independent, a Ukraine-based English-language outlet that combines frontline reporting with war-crimes investigations following Russia's full-scale invasion.

¹⁴ Relevant civil-society organisations include the Centre for Information Resilience, StopFake and Detector Media. Their work on open-source investigation, fact-checking, media monitoring and media literacy can provide local-language and contextual expertise for identifying conflict-related information harms.

¹⁵ See, for instance, the International Committee of the Red Cross and the United Nations Office for the Coordination of Humanitarian Affairs, which can provide insight into how manipulated information affects civilian protection, humanitarian communication and access to assistance.

jurisdictions is constrained, and regulatory cycles struggle to keep pace with technological change. One could argue that the EU framework is necessary but insufficient, requiring both stronger internal coherence among its instruments and external engagement to anchor equivalent norms within international governance processes.

Against this backdrop, the author proposes two directions through which the EU could develop a more operational response to conflict-related deepfake disinformation. First, the Commission should issue public guidance setting out when it will consider activating Article 36 DSA in conflict-adjacent situations. Such guidance could identify coordinated synthetic-media campaigns as a potential crisis trigger where they create a serious and demonstrable threat to public security within the Union, including by targeting EU-based communities, disrupting emergency communication, endangering civilians, or materially undermining the integrity of information relevant to an ongoing conflict. Any Article 36 decision should require designated platforms to conduct rapid risk assessments, preserve and share relevant evidence with competent authorities, implement proportionate distributional and labelling measures, and report publicly on the effectiveness and human rights implications of those interventions.

Second, the EU should establish a structured coordination framework with relevant UN bodies for conflict-related synthetic-media incidents. This framework should include a common referral channel for credible incidents, access to independent technical and forensic expertise, and shared indicators for assessing risks to civilian safety, humanitarian operations and peace negotiations. Effective implementation will also require coordinated information-sharing arrangements between platforms, regulators, and relevant international actors, designed with appropriate safeguards for personal data and other fundamental rights (Albakri *et al.* 2019, Borden *et al.* 2019). The framework should also incorporate safeguards against the use of counter-disinformation measures to suppress protected expression in politically contested contexts, consistent with the need to preserve a pluralistic information environment and freedom of expression (United Nations 2024).

On a final note, the UN Secretary-General António Guterres (2024) has called for a global code of conduct to promote information integrity on online platforms, aiming to stem disinformation that fuels conflict and hatred. At the same time, measures targeting synthetic media must preserve freedom of expression, protect journalism, and resist instrumentalisation by states seeking to suppress legitimate dissent under the cover of disinformation control. As the UN (2024, 3) affirms, information integrity ‘entails a pluralistic information space that champions human rights [and] peaceful

societies'. In other words, defending truth in the public sphere has become inseparable from defending peace itself. By protecting the integrity of public communication against deception and hatred, we uphold the conditions in which peaceful societies can flourish.

References

- AFP (2021) 'Myanmar's ousted leader Suu Kyi faces new corruption charges from junta', retrieved from: <https://www.bangkokpost.com/world/2085783/myanmars-ousted-leader-suu-kyi-faces-new-corruption-charges-from-junta> (accessed: 12/04/2026).
- African Union (2003) *Protocol to the African Charter on Human and Peoples' Rights on the Rights of Women in Africa*, adopted 11 July 2003, Maputo.
- Ahmad, T., Arif, R., Ittefaq, M., and Park, S. (2026) 'Artificial intelligence, deepfakes, and wars: How global news media frame actors, threats, and solutions in 19 countries', *Media, War & Conflict*. DOI: 10.1177/17506352261424826
- Ajšić, A. (2025) 'Language ideology as fake news', *Linguistics Vanguard*. DOI: 10.1515/lingvan-2024-0063
- Akkuş, B. (2025) 'Deepfakes and the Geneva Conventions: Does Deceptive AI-Generated Misinformation Directed at an Enemy During Armed Conflict Violate International Humanitarian Law? A Critical Discussion', *Laws* 14(6). DOI: 10.3390/laws14060083
- Alanazi, S. and Asif, S. (2024) 'Exploring deepfake technology: creation, consequences and countermeasures', *Human-Intelligent Systems Integration* 6(1), 49–60. DOI: 10.1007/s42454-024-00054-8
- Albakri, A., Boiten, E., and De Lemos, R. (2019) 'Sharing Cyber Threat Intelligence Under the General Data Protection Regulation', in Naldi, M., Italiano, G.F., Rannenberg, K., Medina, M., and Bourka, A. (eds), *Privacy Technologies and Policy*, Cham: Springer, 28–41. DOI: 10.1007/978-3-030-21752-5_3
- Alleyne, M.D. (2003) 'Propaganda for Peace?', in Alleyne, M.D. (ed), *Global Lies? Propaganda, the UN and World Order*, London: Palgrave Macmillan UK, 1–27. DOI: 10.1057/9780230507944_1
- Allyn, B. (2022) 'Deepfake video of Zelenskyy could be 'tip of the iceberg' in info war, experts warn', retrieved from: <https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia> (accessed: 11/04/2026).

- Almada, M. and Radu, A. (2024) 'The Brussels Side-Effect: How the AI Act Can Reduce the Global Reach of EU Policy', *German Law Journal* 25(4), 646–663. DOI: 10.1017/glj.2023.108
- AL-Sahel, S. (2025) 'Artificial Intelligence and Information Warfare in the 2023-2025 Gaza Conflict: A Middle Eastern Perspective', *European Journal of Humanities and Social Sciences* (6), 83-85. DOI: 10.29013/EJHSS-25-6-83-85
- Alston, P. (1980) 'Peace as a Human Right', *Bulletin of Peace Proposals* 11(4), 319–329. DOI: 10.1177/096701068001100402
- Argent, M. (2025) 'A new frontier of American propaganda', *Index on Censorship* 54(3), 72–73. DOI: 10.1177/03064220251382205
- Aslett, K., Bonneau, R., Nagler, J., and Tucker, J.A. (2022) 'News credibility labels have limited average effects on news diet quality and fail to reduce misperceptions', *Science Advances* 8(18). DOI: 10.1126/sciadv.abl3844
- Association of Southeast Asian Nations (2012) *ASEAN Human Rights Declaration*, adopted 18 November 2012, Phnom Penh.
- Bastos Lima, M.G. and Schilling-Vacaflor, A. (2024) 'Supply chain divergence challenges a "Brussels effect" from Europe's human rights and environmental due diligence laws', *Global Policy* 15(2), 260–275. DOI: 10.1111/1758-5899.13326
- Benkler, Y., Faris, R., and Roberts, H. (2018) *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*, New York: Oxford University Press. DOI: 10.1093/oso/9780190923624.001.0001
- Bernaciak, C. and Ross, D.A. (2022) 'How Easy Is It to Make and Detect a Deepfake?', retrieved from: <https://www.sei.cmu.edu/blog/how-easy-is-it-to-make-and-detect-a-deepfake/> (accessed: 14/04/2026).
- Bjola, C. (2025) 'Algorithmic invasions: How information warfare threatens NATO's Eastern flank', retrieved from: <https://centrulpolitic.ro/articole/algorithmic-invasions-how-information-warfare-threatens-natos-eastern-flank/> (accessed: 14/04/2026).
- Borden, R., Mooney, J., Taylor, M., and Sharkey, M. (2019) 'Threat Information Sharing under GDPR', *Scitech Lawyer* 15 (3), 30–35.
- Böswald, L.M. (2025) 'Soft Law, Hard Risks? Co-Regulation and Risk Mitigation Under the Digital Services Act', retrieved from: <https://www.interface-eu.org/publications/dsa-co-regulatory-mechanisms> (accessed: 10/04/2026).

- Botan, M. and Meyer, T. (2025) 'Implementing the EU Code of Practice on Disinformation: An Evaluation of VLOPSE Compliance and Effectiveness (Jan–Jun 2024)', retrieved from: <https://edmo.eu/publications/implementing-the-eu-code-of-practice-on-disinformation-an-evaluation-of-vlopse-compliance-and-effectiveness-jan-jun-2024/> (accessed: 10/04/2026).
- Bradford, A. (2020) *The Brussels Effect: How the European Union Rules the World*, New York: Oxford University Press.
- Buijs, D. and Buri, I. (2023) 'The DSA's crisis approach: crisis response mechanism and crisis protocols', retrieved from: <https://dsa-observatory.eu/2023/02/21/the-dsas-crisis-approach-crisis-response-mechanism-and-crisis-protocols/> (accessed: 25/06/2025).
- Busch, E. and Ware, J. (2023) *The Weaponization of Deepfakes: Digital Deception on the Far-Right*, International Centre for Counter-Terrorism. DOI: 10.19165/2023.2.07
- Byman, D.L., Gao, C., Meserole, C., and Subrahmanian, V.S. (2023) *Deepfakes and international conflict*, Washington, DC: Brookings Institution.
- Cahlan, S. (2020) 'How misinformation helped spark an attempted coup in Gabon', retrieved from: <https://www.washingtonpost.com/politics/2020/02/13/how-sick-president-suspect-video-helped-sparked-an-attempted-coup-gabon/> (accessed: 20/10/2025).
- Calo, R. (2017) 'Artificial Intelligence Policy: A Primer and Roadmap', *UC Davis Law Review* 51, 399–435.
- Cao, X., and Gong, N.Z. (2022) 'Understanding the Security of Deepfake Detection', in Gladyshev, P., Goel, S., James, J., Markowsky, G. and Johnson, D. (eds), *Digital Forensics and Cyber Crime*, Cham: Springer, 360–378. DOI: 10.1007/978-3-031-06365-7_22
- Cappello, J.T. and Sunter, D. (2019) 'Information Operations in the Western Balkans', in Sultănescu, D. (ed), *Challenges in Strategic Communication and Fighting Propaganda in Eastern Europe*, IOS Press, 53–74. DOI: 10.3233/978-1-61499-943-0-53
- Chapman, P. (2025) 'Advancing Platform Accountability: The Promise and Perils of DSA Risk Assessments', retrieved from: <https://techpolicy.press/advancing-platform-accountability-the-promise-and-perils-of-dsa-risk-assessments> (accessed: 23/04/2025).
- Chesney, B. and Citron, D.K. (2019) 'Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security', *California Law Review* 107, 1753–1820. DOI: 10.15779/Z38RV0D15J

- Christakis, T. (2020) *'European Digital Sovereignty': Successfully Navigating Between the 'Brussels Effect' and Europe's Quest for Strategic Autonomy*, Rochester, NY: Multidisciplinary Institute on Artificial Intelligence/ Grenoble Alpes Data Institute.
- Cockell, J.G. (2000) 'Conceptualising Peacebuilding: Human Security and Sustainable Peace', in Pugh, M. (ed), *Regeneration of War-Torn Societies*, London: Palgrave Macmillan UK, 15–34. DOI: 10.1007/978-1-349-62835-3_2
- Cohen, M. (2025) 'Deck chairs on the Titanic': How Trump already upended DOJ's ongoing efforts to arrest and prosecute January 6 rioters', retrieved from: <https://www.cnn.com/2025/01/05/politics/trump-january-6-capitol-rioters> (accessed: 14/04/2026).
- Council of the EU (2022a) 'DSA: Council gives final approval to the protection of users' rights online', retrieved from: <https://www.consilium.europa.eu/en/press/press-releases/2022/10/04/dsa-council-gives-final-approval-to-the-protection-of-users-rights-online/> (accessed: 14/04/2026).
- Council of the EU (2022b) 'EU imposes sanctions on state-owned outlets RT/ Russia Today and Sputnik's broadcasting in the EU', retrieved from: <https://www.consilium.europa.eu/en/press/press-releases/2022/03/02/eu-imposes-sanctions-on-state-owned-outlets-rtrussia-today-and-sputnik-s-broadcasting-in-the-eu/> (accessed: 20/10/2025).
- Council of the EU (2026) 'Russian hybrid threats: EU lists two entities over information manipulation activities', retrieved from: <https://www.consilium.europa.eu/en/press/press-releases/2026/04/21/russian-hybrid-threats-eu-lists-two-entities-over-information-manipulation-activities/> (accessed: 30/06/2026).
- Court of Justice of the European Union (2022) *TU and RE v Google LLC*, Case C-460/20, Judgment of 8 December 2022.
- Cox, F.D. and Sisk, T.D. (2017) 'Peacebuilding: A Social Cohesion Approach', in Cox, F.D., and Sisk, T.D. (eds), *Peacebuilding in Deeply Divided Societies: Toward Social Cohesion?*, Cham: Springer International Publishing, 13–31. DOI: 10.1007/978-3-319-50715-6_2
- Cromwell, M. and Vogeley, W.B. (2009) 'Nonviolent Action, Trust and Building a Culture of Peace', in de Rivera, J. (ed), *Handbook on Building Cultures of Peace*, New York, NY: Springer, 231–244. DOI: 10.1007/978-0-387-09575-2_16
- Deibert, R.J. (2019) 'The Road to Digital Unfreedom: Three Painful Truths About Social Media', *Journal of Democracy* 30(1), 25–39.

- Dragomir, M., Rúas-Araújo, J., and Horowitz, M. (2024) 'Beyond online disinformation: assessing national information resilience in four European countries', *Humanities and Social Sciences Communications* 11(1), 101. DOI: 10.1057/s41599-024-02605-5
- El-Gazzar, R. and Stendal, K. (2020) 'Examining How GDPR Challenges Emerging Technologies', *Journal of Information Policy* 10, 237–275. DOI: 10.5325/jinfopoli.10.2020.0237
- Ellul, J. (1965) *Propaganda: The Formation of Men's Attitudes*. Vintage.
- European Commission (2025) 'Overview - Instrument for Pre-accession Assistance', retrieved from: https://enlargement.ec.europa.eu/funding-technical-assistance/overview-instrument-pre-accession-assistance_en (accessed: 10/04/2026).
- European Commission (2026a) 'AI Act', retrieved from: <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> (accessed: 10/04/2026).
- European Commission (2026b) 'Code of Practice on Transparency of AI-Generated Content', retrieved from: <https://digital-strategy.ec.europa.eu/en/policies/code-practice-ai-generated-content> (accessed: 30/06/2026).
- European Commission (2026c) 'Global Europe: Neighbourhood, Development and International Cooperation Instrument', retrieved from: https://commission.europa.eu/funding-tenders/find-funding/eu-funding-programmes/global-europe-neighbourhood-development-and-international-cooperation-instrument_en (accessed: 10/04/2026).
- European Commission (2026d) 'Supervision of the designated very large online platforms and search engines under DSA', retrieved from: <https://digital-strategy.ec.europa.eu/en/policies/list-designated-vlops-and-vloses> (accessed: 14/04/2026).
- European Commission (2026e) 'The 2022 Code of Practice on Disinformation', retrieved from: <https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation> (accessed: 31/03/2026).
- European External Action Service (2019) 'EUvsDisinfo: how to debunk over 6,500 disinformation cases in four years?', retrieved from: https://www.eeas.europa.eu/eeas/euvdisinfo-how-debunk-over-6500-disinformation-cases-four-years_en (accessed: 14/04/2026).
- European External Action Service (2025) 'EEAS Strategic Communication Task Forces', retrieved from: <https://www.eeas.europa.eu/eeas/>

- eeas-strategic-communication-task-forces_en?s=2803 (accessed: 14/04/2026).
- European Union (2012) *Charter of Fundamental Rights of the European Union*, 2012/C 326/02.
- Fontecha-Tirado, M.M. (2018) 'Building Trust and Collaborating with Others: Challenges For A Sustainable Peace in Caquetá, Colombia', *Consilience* (19), 163–183. DOI: 10.7916/consilience.v0i19.3897
- Freeze, M., Baumgartner, M., Bruno, P., Gunderson, J.R., Olin, J., Ross, M.Q., and Szafran, J. (2021) 'Fake Claims of Fake News: Political Misinformation, Warnings, and the Tainted Truth Effect', *Political Behavior* 43(4), 1433–1465. DOI: 10.1007/s11109-020-09597-3
- Gehrig, M., Buchanan, C., Holt, S., and Ramsbotham, A. (2023) *Building Trust in Peace Mediation*, Washington, DC: United States Institute of Peace.
- Geissler, E. and Sprinkle, R.H. (2013) 'Disinformation squared: Was the HIV-from-Fort-Detrick myth a Stasi success?', *Politics and the Life Sciences* 32(2), 2–99. DOI: 10.2990/32_2_2
- Golovchenko, Y. (2022) 'Fighting Propaganda with Censorship: A Study of the Ukrainian Ban on Russian Social Media', *Journal of Politics* 84(2), 639–654. DOI: 10.1086/716949
- Guillermet Fernandez, C. and Puyana Fernandez, D. (2017) 'The Adoption of the Declaration on the Right to Peace by the United Nations: a Human Rights Landmark', *Peace Human Rights Governance* 1(2), 275–297. DOI: 10.14658/PUPJ-PHRG-2017-2-6
- Guterres, A. (2024) 'The UN Secretary-General Remarks to Launch the Global Principles for Information Integrity', retrieved from: <https://www.un.org/en/UNIS-Nairobi/un-secretary-general-remarks-launch-global-principles-information-integrity> (accessed: 14/04/2026).
- Hayden, P. (2002) 'A defense of peace as a human right', *South African Journal of Philosophy* 21(3), 147–162. DOI: 10.4314/sajpem.v21i3.31342
- Hume, D. (1742) *Essays, Moral and Political*, Edinburgh.
- Ife, J. (2007) 'Human rights and peace', in Webel, C. and Galtung, J. (eds), *Handbook of Peace and Conflict Studies*, Oxon: Routledge, 160–172.
- International Committee of the Red Cross (2005) *Customary international humanitarian law*, Cambridge University Press.
- Ivankiv, I. (2020) 'Peace as a Right of Humanity', in Mauerhofer, V., Ruppo, D., and Tarquinio, L. (eds), *Sustainability and Law: General and Specific Aspects*, Cham: Springer International Publishing, 337–354. DOI: 10.1007/978-3-030-42630-9_18

- Jenkins, R., Lamers James, R. and Hausknecht, A. (2024) 'Trust in Evidence in an Era of Deepfakes', retrieved from: <https://acss.org.uk/trust-in-evidence-in-an-era-of-deepfakes> (accessed: 14/04/2026).
- Jiang, Z., Zhang, J., and Gong, N. Z. (2023) 'Evading Watermark based Detection of AI-Generated Content', in *CCS '23: Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security*, 1168–1181. DOI: 10.1145/3576915.3623189
- Keene, J. (2008) *Treason on the Airwaves: Three Allied Broadcasters on Axis Radio during World War II*, Bloomsbury Publishing USA.
- Kingsbury, C.M. (2010) *For Home and Country: World War I Propaganda on the Home Front*, University of Nebraska Press. DOI: 10.2307/j.ctt1dfnsnn
- Klepper, D. (2023) 'Fake babies, real horror: False AI-generated images of the war in Gaza spark alarm', retrieved from: <https://www.latimes.com/world-nation/story/2023-11-29/israel-hamas-war-artificial-intelligence-deepfakes-disinformation> (accessed: 20/10/2025).
- Klepper, D. and Lemon, J. (2023) 'Can you spot a deepfake? Gaza the latest outlet for AI deception', retrieved from: <https://www.smh.com.au/world/middle-east/can-you-spot-a-deepfake-gaza-the-latest-outlet-for-ai-deception-20231129-p5enrd.html> (accessed: 20/10/2025).
- Kuźnicka-Błaszowska, D. and Kostyuk, N. (2025) 'Emerging need to regulate deepfakes in international law: the Russo-Ukrainian war as an example', *Journal of Cybersecurity* 11(1). DOI: 10.1093/cybsec/tyaf008
- Łabuz, M. (2024) 'Deep fakes and the Artificial Intelligence Act—An important signal or a missed opportunity?', *Policy & Internet* 16(4), 783–800. DOI: 10.1002/poi3.406
- Lasswell, H.D. (1938) *Propaganda technique in the World War*, New York: Peter Smith.
- Linehan, C., Murphy, G., and Twomey, J.J. (2023) 'Deepfakes in warfare: new concerns emerge from their use around the Russian invasion of Ukraine', retrieved from: <https://theconversation.com/deepfakes-in-warfare-new-concerns-emerge-from-their-use-around-the-russian-invasion-of-ukraine-216393> (accessed: 20/10/2025).
- Llorente, R.V. (2024) 'Deepfakes in the Dock: Preparing International Criminal Justice for Deepfakes and Generative AI', retrieved from: https://www.americanbar.org/groups/science_technology/resources/scitech-lawyer/archive/deepfakes-preparing-international-criminal-justice-generative-ai/ (accessed: 14/04/2026).

- Llorente, R.V. and McDermott, Y. (2024) 'Truth, Trust, and AI: Justice and Accountability for International Crimes in the Era of Digital Deception', retrieved from: <https://www.justsecurity.org/96731/truth-trust-and-ai-justice-and-accountability/> (accessed: 14/04/2026).
- Lubin, A. (2020) 'The Liberty to Spy', *Harvard International Law Journal* 61(1), 185–243.
- Majchrzak, A. (2023) 'Russian disinformation and the use of images generated by artificial intelligence (deepfake) in the first year of the invasion of Ukraine', *Media Business Culture* 1(14), 42–55.
- Martel, C. and Rand, D.G. (2023) 'Misinformation warning labels are widely effective: A review of warning effects and their moderating features', *Current Opinion in Psychology* 54. DOI: 10.1016/j.copsyc.2023.101710
- Meßmer, A.-K. and Degeling, M. (2023) *Auditing Recommender Systems: Putting the DSA into practice with a risk-scenario-based approach*, Berlin: Stiftung Neue Verantwortung e.V.
- NATO (2026) 'NATO's approach to counter information threats', retrieved from: <https://www.nato.int/en/about-us/official-texts-and-resources/official-texts/2024/10/18/natos-approach-to-counter-information-threats> (accessed: 14/04/2026).
- Nesterova, Y. and Kim, E.-J.A. (2024) 'Peace as Social Cohesion, Equity, and Democracy: Local Peacebuilders' Conceptualizations of Peace', *Peace Review* 36(3), 531–544. DOI: 10.1080/10402659.2024.2377773
- Nguyen, T.N.A. and Navdarashvili, T. (2025) 'Pixels of Power: Deepfakes, Regulatory Uncertainty, and the Path toward Accountability in the EU's AI Framework', *Rivista di diritto dei media* (II/2025), 184–213.
- Organization of African Unity (1981) *African Charter on Human and Peoples' Rights*, adopted 27 June 1981, Nairobi.
- Organization of American States (1969) *American Convention on Human Rights*, adopted 22 November 1969, San José, Costa Rica.
- Osadchuk R. (2024) 'AI tools usage for disinformation in the war in Ukraine', retrieved from: <https://dfrlab.org/2024/07/09/ai-tools-usage-for-disinformation-in-the-war-in-ukraine/> (accessed: 14/04/2026).
- OSCE Representative on Freedom of the Media (2021) 'Fourth expert meeting: Deepfake news – Artificial intelligence and disinformation as a multilateral policy challenge', retrieved from: <https://rfom.osce.org/representative-on-freedom-of-media/506347> (accessed: 14/04/2026).

- O'Shaughnessy, N.J. (2004) 'Weapons of Mass Seduction: Propaganda, Media and the Iraq War', *Journal of Political Marketing* 3(4), 79–104. DOI: 10.1300/J199v03n04_04
- Peake, J. (2024) 'Challenges of Using Digital Evidence for War Crimes Prosecutions: Availability, Reliability, Admissibility', *American Journal of International Law* 118, 57–61. DOI: 10.1017/aju.2024.5
- Pennycook, G., Bear, A., Collins, E.T., and Rand, D.G. (2020) 'The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings', *Management Science* 66(11), 4944–4957. DOI: 10.1287/mnsc.2019.3478
- Pfefferkorn, R. (2021) 'The threat posed by deepfakes to marginalized communities', retrieved from: <https://www.brookings.edu/articles/the-threat-posed-by-deepfakes-to-marginalized-communities/> (accessed: 14/04/2026).
- Piers, M. and Osaer, H.C. (2025) 'When seeing is no longer believing: evidentiary challenges of deepfakes in arbitration', *Dispute Resolution International* 19(2), 127–162.
- Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I). Adopted 8 June 1977. 1125 UNTS 3.
- Reader, T.M.P. (2020) 'Lessons from Operation 'Denver,' the KGB's Massive AIDS Disinformation Campaign', retrieved from: <https://thereader.mitpress.mit.edu/operation-denver-kgb-aids-disinformation-campaign/> (accessed: 10/04/2026).
- Regulation 2016/679 (2016) 'Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC'.
- Regulation 2022/2065 (2022) 'Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC'
- Regulation 2024/1689 (2024) 'Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)'

- Reich, O. and Calabrese, S. (2025) *Civic Discourse and Electoral Processes in the Risk Assessment and Mitigation Measures Reports under the Digital Services Act: An analysis*. Bruxelles: European Partnership for Democracy & Civil Liberties Union for Europe.
- Richmond, K. (2007) 'Radio in World War II', *Sabretache* 48(1), 39–52. DOI: 10.3316/informit.613971905753690
- Ristovska, S. (2019) 'Human rights collectives as visual experts: The case of Syrian Archive', *Visual Communication* 18(3), 333–351. DOI: 10.1177/1470357219851728
- Romero Moreno, F. (2024) 'Generative AI and deepfakes: a human rights approach to tackling harmful content', *International Review of Law, Computers & Technology* 38(3), 297–326. DOI: 10.1080/13600869.2024.2324540
- Roumate, F. (2025) 'AI Transformation in Media: Reshaping War Between Reality and Virtuality', in Roumate, F., and La Rosa, A. (eds), *Artificial Intelligence, Media and International Security: The Weaponization of AI Use in Media and International Security*, Cham: Springer Nature Switzerland, 17–27. DOI: 10.1007/978-3-031-95757-4_2
- Samaha, P. (2025) 'Weaponized AI: The New Frontline in Global Disinformation Warfare', retrieved from: <https://natoassociation.ca/weaponized-ai-the-new-frontline-in-global-disinformation-warfare/> (accessed: 10/04/2026).
- Schabas, W. A. (2011) 'The Human Right to Peace', in Eide, A., Möller, J. Th., and Ziemele, I. (eds), *Making Peoples Heard: Essays on Human Rights in Honour of Gudmundur Alfredsson*, Brill Nijhoff, 43–57. DOI: 10.1163/9789004215801_005
- Schiff, K.J., Schiff, D.S., and Bueno, N.S. (2025) 'The Liar's Dividend: Can Politicians Claim Misinformation to Evade Accountability?', *American Political Science Review* 119(1), 71–90. DOI: 10.1017/S0003055423001454
- Shaheen, K. (2015) 'Syria war: 'unthinkable atrocities' documented in report on Aleppo', retrieved from: <https://www.theguardian.com/world/2015/may/05/syria-forces-war-crime-barrel-bombs-aleppo-amnesty-report> (accessed: 12/04/2026).
- Slychko, I. (2025) 'The Use of Artificial Intelligence-Generated Videos to Discredit the Defence Forces of Ukraine', retrieved from: <https://cpd.gov.ua/en/report/analytical-report-the-use-of-artificial-intelligence-generated-videos-to-discredit-the-defence-forces-of-ukraine/> (accessed: 12/04/2026).

- Sundar, S.S., Molina, M.D., and Cho, E. (2021) 'Seeing Is Believing: Is Video Modality More Powerful in Spreading Fake News via Online Messaging Apps?', *Journal of Computer-Mediated Communication* 26(6), 301–319. DOI: 10.1093/jcmc/zmab010
- Taylor, P.M. (2010) 'From Psychological Warfare to Information Operations and Back Again', in Buckley, J., and Kassimeris, G. (eds), *The Ashgate Research Companion to Modern Warfare*, London: Routledge.
- Tomić, Z., Damnjanović, T., and Tomić, I. (2023) 'Artificial Intelligence in Political Campaigns', *South Eastern European Journal of Communication* 5(2), 17–28. DOI: 10.47960/2712-0457.2.5.17
- Tsybaliuk, Y. (2021) 'Statement in response to the Report by the OSCE Representative on Freedom of the Media, Ms. Teresa Ribeiro', retrieved from: <https://cdn.osce.org/sites/default/files/f/documents/4/e/507602.pdf> (accessed: 11/04/2026).
- Tweissi, B. (2025) 'Gaza Infocide: Disinformation and New Narratives Wars', in *IEMed. Mediterranean Yearbook 2025*, Panorama, 297–301.
- Twomey, J., Ching, D., Aylett, M.P., Quayle, M., Linehan, C., and Murphy, G. (2023) 'Do deepfake videos undermine our epistemic trust? A thematic analysis of tweets that discuss deepfakes in the Russian invasion of Ukraine', *PLOS ONE* 18(10). DOI: 10.1371/journal.pone.0291668
- Ulbricht, B. R., Moxley, C., Austin, M. D., and Norburg, M. D. (2022) 'Digital Eyewitnesses: Using New Technologies to Authenticate Evidence in Human Rights Litigation', *Stanford Law Review* 74, 851–895.
- United Nations (1945) *Charter of the United Nations*, adopted 26 June 1945, San Francisco.
- United Nations (2024) *United Nations Global Principles for Information Integrity: Recommendations for Multi-stakeholder Action*, New York: United Nations Publications. DOI: 10.18356/9789211065756
- United Nations General Assembly (1948) *Universal Declaration of Human Rights*, UNGA Res. 217 A (III), 10 December 1948.
- United Nations General Assembly (1978) *Declaration on the Preparation of Societies for Life in Peace*, UNGA Res. 33/73, 15 December 1978.
- United Nations General Assembly (1984) *Declaration on the Right of Peoples to Peace*, UNGA Res. 39/11, 12 November 1984.
- United Nations General Assembly (2016) *Declaration on the Right to Peace*, UNGA Res. 71/189, 19 December 2016.
- Vaccari, C. and Chadwick, A. (2020) 'Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception,

- Uncertainty, and Trust in News', *Social Media + Society* 6(1). DOI: 10.1177/2056305120903408
- Ward, K. (1989) 'Propaganda in War and Peace', in Ward, K. (ed), *Mass Communications and the Modern World*, London: Macmillan Education UK, 133–167.
- Weimann, G. and Weimann-Saks, D. (2024) 'Coping with Hamas's Psychological Warfare during the Gaza War', *Studies in Conflict & Terrorism*, 1–20. DOI: 10.1080/1057610X.2024.2327669
- Windwehr, S. (2025) 'Systemic Risk Reporting: A System in Crisis?', retrieved from: <https://www.eff.org/deeplinks/2025/01/systemic-risk-reporting-system-crisis> (accessed: 25/09/2025).
- Woods, J.M. (2013) 'Theorizing peace as a human right', *Human Rights and International Legal Discourse* 7, 178–236.
- Zhao, X., Gunn, S., Christ, M., Fairuze, J., Fabrega, A., Carlini, N., Garg, S., Hong, S., Nasr, M., Tramer, F., Jha, S., Li, L., Wang, Y.-X., and Song, D. (2025) 'SoK: Watermarking for AI-Generated Content', in *2025 IEEE Symposium on Security and Privacy*, 2621–2639. DOI: 10.1109/SP61157.2025.00178