

Research Article

Open Access

The Ghost in the Code: Why Global Law is Failing to Stop the Digital War on Women

Jessica Adanna Awurum

School of Law, Shanghai University of Finance and Economics, China

ABSTRACT

We are witnessing a systemic failure of legal accountability in cyberspace. Despite the growth of technology-facilitated gender-based violence (TFGBV) from basic harassment to complex AI-based warfare, including deepfake misuse and automated stalking, the digital law response is still lagging. It is a culture of impunity in which the perpetrators act with almost absolute anonymity, and justice statutes are symbols rather than actual shields. By applying feminist legal theory, this research interrogates how traditional legal definitions of "harm" and "privacy" fail to account for the gendered dimensions of digital surveillance and bodily autonomy. It further frames TFGBV as a fundamental violation of women's human rights, arguing that the failure to regulate digital spaces constitutes a state-sanctioned withdrawal of the right to live free from violence. Using a stress-test methodology applied to three core International Legal frameworks, the article applies these tests to a sample set of high-impact TFGBV cases over 2024-2025. Following such cases from the first police report to their resolution, prioritising a survivor-centred approach, ensuring that the legal friction recorded is viewed through the lived experiences of those targeted, the research provides an empirical map of where in that process account holds, and applies comparative legal theorising to make sense of the friction between state sovereignty and transnational/borderless digital crime. This research sets the stage for a new era of accountability, advocating that, without reforming the balance of proof and platform liability through three feminist interventions, laws will remain bystanders, as women are systematically erased from public digital life.

***Corresponding author**

Jessica Adanna Awurum, School of Law, Shanghai University of Finance and Economics, China.

Received: September 08, 2026; **Accepted:** September 14, 2026; **Published:** September 21, 2026**Keywords:** TFGBV, Accountability, Digital Erasure, Platform Liability, Jurisdictional Loophole, Deepfakes**Themes:** Transforming systems of structural violence, Law and Policies**Part I: Defining TFGBV in 2026: Beyond Cyberbullying The Digital Frontier: From Harassment to Warfare**

The digital world, once imagined as a democratising Global Village, has become a technologically enabled theatre of war where women's bodies and identities are the main casualties. Technology-facilitated gender-based violence (TFGBV) is no longer just a civility issue that exists on the fringes of the digital world; it has evolved into a form of automated warfare. As a result of the implementation of advanced generative artificial intelligence and ease of digital distribution, the tools of harassment have shifted from the manual, discrete incidents typical in 2016 that tended to signal isolated harassers, toward systemic 2020 AI-powered campaigns intended to effectuate the total social and political cancellation of women. In the evolution of intergroup conflict, cyberbullying is now an archaic term that no longer reflects the severity of contemporary digital harm.

TFGBV in 2026 is defined by automated gender-based violence, where tactical machine-learning models are weaponised to produce non-consensual intimate images (NCII), systematic bot-facilitated piling-on and predictive stalking. A particularly telling example

came in early 2024, when AI-generated pornographic images of Taylor Swift reached 47 million views on platform X before removal, exposing the velocity of harm that traditional notice-and-takedown regimes cannot contain [1]. TFGBV has one main goal: to remove women from public life. When high-profile figures, journalists, politicians and activists are targeted, the punishment has a chilling effect that suppresses free speech. According to the amendments to The Chilling study by UNESCO, nearly 75% of women journalists worldwide have experienced online violence, and 14% allege that the threats in cyberspace turned into physical attacks [2]. They're not simply digital versions of physical violence; rather, they're a type of harm in their own right that harnesses the permanence of the internet to cause what it terms digital death, like an assault on democratic participation.

In a landmark 2024-2025 civil trial, Italian Prime Minister Giorgia Meloni testified against the makers of deepfake pornographic videos of her and sued for over €100,000 in damages. Her testimony reinforced that such acts are not parody, but rather a form of violence designed to delegitimise women in power [3]. When the law fails to provide a path for redress, it communicates that the digital public square is a men-only space, and that basically removes women's digital citizenship. The emergence of synthetic media has gone further to create a new frontier of violation, the theft of a woman's likeness for sexual exploitation. Research has found that 98% of all deepfake videos made available online are pornographic, and 99% of those who are targeted are women [4].

The boom in the 2025 modification apps, which are AI tools that digitally strip away women's clothing without consent, quickly spurred a flood of legislation hitting on these and other topics around the world, including the criminalisation in 2026 by UK lawmakers of creating or requesting non-consensual intimate images [5].

When one prompt can allow an algorithm to violate a woman's image, the law's traditional concern with physical contact no longer applies. The law must catch up so that bodily autonomy means control over our digital selves. Without this, the liar's dividend where abusers claim irrefutable evidence is just a deepfake and survivors have difficulty proving synthetic abuse never occurred will continue to underwrite a present-day culture of impunity. In January, the UK's Ofcom opened up investigations into AI assistants (like AI's Grok) for their part in generating non-consensual sexualized imagery. This is where the tables can be expected to turn, from holding the automated engine responsible for what it outputs, to monitoring who builds these engines as well [6].

The Ghost in the Code: Technical Obstacles to Justice

The pivotal crisis of digital jurisprudence is that the present statutes are lashing out at shadows. The Ghost in the Code is the structural realities of modern digital architecture; anonymity, algorithmic logic and viral velocity, that make it possible for perpetrators to remain invisible while their harm is indelible. This section explores how these technical configurations function as a corset for gender-based violence, constructing a black box that legal practitioners have difficulty accessing.

The legal protections of anonymity and end-to-end encryption (E2EE) are a deep paradox for gender justice. And while E2EE is an essential tool for human rights defenders and survivors of domestic violence to communicate with confidence, it also creates a fortress of impunity for digital attackers [7]. On decentralised platforms and encrypted messaging applications such as Telegram, law enforcement's inability to intercept communications or unmask users creates a jurisdictional vacuum. By 2025, the increasing enforcement of coordinated inauthenticity, where brigades of nameless bots use automation to foster harassment, has made it statistically impossible for a single human defendant to be served with process. In addition, The Liar's Dividend has emerged as a counter *Défense*; actual perpetrators of violence have used the spectre of deepfakes as a weapon to say that solid evidence they committed crimes is AI-generated, and raise the bar for reasonable doubt over valid pieces of evidence [8]. Without the ability to trace this, the anonymous harasser's right to privacy effectively invalidates the victim's right to security and redress [9].

Aside from the above, one of the most significant blind spots in international law is how platform algorithms, like editorial tools that are fundamentally neutral, are treated. But modern jurisprudence is starting to realise that these systems are harmful by design. Where engagement-based models currently reward high velocity, provocative content and are essentially broadcasting misogynistic abuse as a megaphone [10]. The legal argument in favour of platform liability is that an algorithm does not just host a deepfake; it will recommend and amplify it, based on predictions about engagement metrics. In 2025, groups such as CDT Europe produced policy papers arguing that when a platform's architecture amplifies the distribution of non-consensual intimate images (NCII) to thousands of feeds, the platform becomes not a passive conduit but an active participant in harm [11]. This systemic

risk has now been additionally brought into focus by the EU Digital Services Act (DSA) audits and their questioning whether engagement-at-all-costs models breach the statutory duty of care by not acting to mitigate gendered risks [12].

The most tragic technical hurdle is called the temporal gap. Much of the reputational and psychological harm is done in a matter of minutes, within the first hour of one's post going live. By contrast, even the fastest legal remedies like emergency court injunctions, for instance, run on a time scale of days or weeks [13]. The quickening of harm is undefeated despite such updates as the UK's Data (Use and Access) Act of 2025 that offer a 48-hour window for NCII takedown. By the time a platform takes action, that same content has often been scraped and reuploaded to secondary, less-regulated mirrors, or even to shared networks on the dark web, resulting in a whack-a-mole effect where the survivor is stuck in a cycle of re-traumatisation [14]. This permanent digital footprint has created an experience of ongoing violation, in which the legal system is bound to a slow, document-driven procedural reality that was never meant to move at the speed of light.

The Failure of the Statute: A Critique of Existing Law

The irony at the core of modern digital jurisprudence is known as the Implementation Gap: Of 87% countries with domestic violence laws, legislation on technology-facilitated harm exists in only a fraction and often contains vague language without enforceable remedies [15]. The focus here is on the three main legal deciding factors (signed treaties, national sovereignty, and platform immunity), which today hold as strong pillars that create a culture of digital impunity.

The Istanbul Convention

The Istanbul Convention of the Council of Europe is considered by many to be the gold standard in GBV legislation. However, scholarly and legal commentary on the Convention is increasingly directed at its terrestrial bias. Scholars have argued that the Convention's definition of violence against women in Article 3 as a violation of human rights was initially theorised with physical proximity in mind [16]. While the treaty purports to be technology-neutral, it lacks explicit language about digital persistence and algorithmic harm, resulting in what the authors identified as a definitional lag. In 2021, the Group of Experts on Action against Violence against Women and Domestic Violence (GREVIO) adopted General Recommendation No. 1 to fill this gap. But literature published in 2024-2025 suggests that the soft law approach is not enough. Without an amendment to the treaty that would include digital bodily autonomy, critics say, domestic courts still have a hard time with cases in which no physical contact was made [17]. Most TFGBV prosecutions are made under Article 33 charges (Psychological Violence). But the standard for conviction is a demonstration of serious impairment of a victim's psychological integrity. Legal scholars have observed that this puts the wrongful burden on a survivor of forced sex to validate clinical trauma, when what actually is to be considered here meets the standard for per se violation of human dignity and autonomy, irrespective of an immediate reaction from a person who has been subjected to the abuse [18].

National vs International Law

The internet is fundamentally borderless, yet law is intractably territorial. This establishes a Jurisdictional Deadlock wherein an offender in Jurisdiction A may attack a potential target in Jurisdiction B via a server machine located in Jurisdiction C. Conventional Private International Law (PIL) rules (e.g., *lex loci*

delicti commissi) were formulated for physical disputes. Within the digital realm, the place of harm is manifest in both everywhere and nowhere at once. Right now, the sovereignty conflict has escalated; even as the EU's Digital Services Act (DSA) tries to enforce extra-territorial standards on a global platform, many parts of the Global South remain beyond its remit and enforcement mechanisms, resulting in what are termed data havens for abusers [19]. The primary tool for cross-border evidence, Mutual Legal Assistance Treaties (MLATs), is fundamentally broken for the digital age. In 2025, it was reported that the average MLAT request for digital evidence takes anywhere from 6 to 10 months [20]. In the case of viral deepfakes, where the velocity of harm is measured in minutes, a 10-month waiting period makes the legal process an exercise in post-mortem expediency rather than protection.

The Neutral Pipe Defence

The biggest stumbling block to accountability, maybe the only one effective at scale, is the Neutral Pipe (or Mere Conduit) defence. Predicated on Section 230 of the U.S. Communications Decency Act and early EU e-commerce directives, this doctrine regards platforms as neutral infrastructure, similar to a utility provider, thereby protecting them from liability for user-generated content. The new platforms are not passive pipes; they are themselves curated algorithmic spaces. To monetise engagement, their business models mean that platforms effectively profit from the digital war on women, as misogynistic abuse often generates high levels of engagement. The UK's Online Safety Act and the EU's DSA have ushered in the statutory duty of care. Yet legal practitioners say that until safe harbour protections disappear, platforms will favour notice-and-takedown (reactive) over safety by design (proactive). As long as platforms are not held to exacting product liability standards for the harms their recommendation engines facilitate, the neutral pipe defence will continue to be a multibillion-dollar umbrella for systemic gendered harm [21].

Part II: Stress-Testing the System

The Data Narrative: Mapping the Justice Gap

The Justice Gap is the quantifiable distance from digital transgression to legal recourse. This section presents a comprehensive empirical breakdown of high-impact case studies tracked from 2024 to 2026. In stress-testing these cases, evaluate from the critique of theory to an evidentiary portrait of a failed global legal architecture. This category represents the fastest-growing area of TFGBV where AI is deployed to strip women of their physical and digital consent.

Case 1: The Meloni Precedent (Italy, 2024-2025)

In this case, Italian Prime Minister Giorgia Meloni was targeted by a 40-year-old man and his 73-year-old father, who used deepfake software to superimpose her face onto pornographic performers. The videos were uploaded to Pornhub, a US-hosted site owned by MindGeek (now Aylo). In 2020, the Italian Postal Police identified the suspects through IP tracking and yet no testimony was received in the case until July 2024 [22]. The delay stemmed from a four-year battle to get server-side data from US-based entities. Italian investigators filed for a Mutual Legal Assistance Treaty (MLAT) request to identify the viewing data and financial logs of the uploaders on the US platform. The US Department of Justice (DOJ) faced challenges because, under the US Communications Decency Act (Section 230), the platform is shielded from liability for user uploads. Also, unless the act meets the US federal threshold for Extortion or Child Exploitation, the First Amendment often protects synthetic speech as a form of transformative parody [23].

In her February 2025 closing statements, Meloni's counsel contended that applying the parody defence to non-consensual sexualization is itself a legal fiction. The case exposes the status paradox, and this gap is temporal and statutory. The 1,500-day delay meant the reputational half-life of the video was exhausted before a judge even heard the Prime Minister's testimony. If a G7 head of state must await justice for more than 1,500 days, the average citizen is essentially barred from it [24].

Case 2: The Modification App Surge (Balkans/Eastern Europe, 2025)

In this case, the focus is on a cluster of 12 cases focused on what appeared to be Telegram Deepfake Bots that permitted users to upload images of local female councillors as well as journalists. For a subscription price of \$5.00 (paid in cryptocurrency), the bot would send back, after 30 seconds, a nude version of the photo. The perpetrators were often secondary actors, individuals who didn't create the software but paid a subscription to use it. As a result of this, investigations in Serbia and North Macedonia were terminated in late 2025. In North Macedonia, the criminal code (at the time) required physical trespass or coercion to qualify as sexual assault, and because the AI-generated nudes were synthetic, defence attorneys argued there was no real victim of a sexual act. The bots were also being hosted on decentralised servers using bulletproof hosting in jurisdictions like the Seychelles, which do not recognise digital image abuse as a criminal offence.

This is a sure sign of the commercialisation of misogyny, where harassment is not only for personal satisfaction, but it has become an interest and digital service that brings profit [25]. This led to a digital gaslighting effect, where survivors were told by local prosecutors that because the images were fake, no real crime had occurred, ignoring the very real social and professional destruction that followed for these women.

Case 3: The Grok-AI Crisis (UK/EU, December 2025 - January 2026)

In late 2025, it was revealed that the AI chatbot called Grok, developed by platform X (formerly Twitter), could be coerced into producing nude or pornographic images of people rather than just subjects or general objects. British journalist Daisy Dixon became a key player in this case when she found herself inundated with AI-generated, sexualized pictures of her image shared on the platform from its built-in tools [26].

This case had a distinct burden because it highlighted the platform's complicity. In previous incidents, users had uploaded outside files, but in this case, the platform's own infrastructure was the weapon. It demonstrates that, in the absence of safety in design, platforms actively industrialise TFGBV.

Case 4: The Seoul National University (SNU) Deepfake Ring lead (South Korea, 2024-2025)

In this case, a ring of a digital sex crime involving over 200 Telegram chat rooms was masterminded by a 40-year-old SNU alumnus. Feminisers HT identified 1,852 photos/videos of female students and alumni that had been digitally altered using generative adversarial networks (GANs) to create deepfake footage [27].

South Korean courts were criticised in early 2024 for handing down lenient sentences, with nearly 60 % of those indicted as perpetrators receiving suspended sentences or fines if they showed repentance. This case shows the psychological motivation of TFGBV. It is rarely about sexual desire and almost always about

the digital maiming of women, especially those who are socially or academically successful.

Case 5: The Swift Crisis (Global/US, 2024)

In January 2024, AI-generated sexually explicit images of Taylor Swift went viral on X (formerly Twitter), garnering 47 million views in under 20 hours before the platform took action. Despite the global outcry, there was no federal US law at the time specifically criminalising the creation of non-consensual AI deepfakes [28]. The incident exposed the Notice-and-Takedown failure; by the time the search terms were blocked, the images had been mirrored on thousands of decentralised Telegram channels. This case catalysed the TAKE IT DOWN Act (2025) and the DEFIANCE Act in the US, but it remains a primary example of the velocity of harm outstripping statutory reach [29].

Case 6: The Almudena Ariza Attacks (Spain, 2024-2025)

This case represents the citation where a coordinated piling-on campaign against journalist Almudena Ariza was conducted. This was a situation where 10,000+ automated accounts posted AI-manipulated audio clips of her allegedly accepting bribes to sway election reporting. Spanish authorities encountered the attribution paradox where the damage was visible, but the intent was hidden. This act had been controlled by a so-called Harassment-as-a-Service (HaaS) provider based outside the EU. The IP addresses for the botnet were traced to proxy servers in Russia and Southeast Asia.

The platform (X) refused to provide the handshake data (the metadata showing who paid for the promoted posts) without a Spanish Supreme Court order. By the time the order was drafted, the Harassment-as-a-Service provider had deleted the account and moved to a new encrypted domain. The Spanish High Court held that the platform (x) was not liable for the user-generated content under the mere conduit defence, because it could not serve a summons to an individual behind the bot-net despite its algorithm collectively trending harassment and the obvious reputational cost of this on Almudena Ariza [30].

Case 7: The AirTag and Smart Lock Domestic Abuse Cases (UK/Nigeria, 2024-2026)

The cases described involved tracking devices (Apple AirTags) and hacked home security systems. One woman, in a 2025 Abuja case, learned her ex had been tracking her for six months using an AirTag concealed in the back of her vehicle to safe houses. Another learned that her former partner had hacked her smart video doorbell and smart thermostat. He would remotely change the temperature of the house and play audio through the speakers at 3:00 AM to induce psychological terror.

In Nigeria, the appropriate law is the Violence Against Persons (Prohibition) Act (VAPP) 2015, which established protection orders, but defence counsel countered that digital stalking without direct physical proximity to the victim did not satisfy that threshold. The case exemplifies the law's physicality bias; since the perpetrator never actually touched his victim, it took 180 days of perpetual stalking for the legal system to classify such violent action [31].

When these cases are viewed in aggregate, a clear pattern emerges: In a majority of these cases, the perpetrator was known. The Justice gap exists because the crime happens in the victim's home and to the victim's personhood, but the evidence sits in a server in California or the Seychelles. Regardless of this, Courts still

view digital harm as less severe than physical harm, leading to suspended sentences for acts that cause permanent career and psychological destruction.

Tracing these cases from report to resolution reveals a Systemic Sieve where justice is filtered out at four critical stages:

1. **The Platform Moderation Gate:** Automated tools on the platform dismissed some cases. Moderators (worldwide, often in different linguistic/cultural contexts) were unable to distinguish between political satire and sexual harassment.
2. **The Forensic Investigative Chokepoint:** Only a few cases ever went to a digital forensic audit. The main reason for the drop-off was the Anonymity Paradox; investigators were unable to breach VPNs or unmask Telegram identities under a court order without being explicitly ordered by a high-level court.
3. **The Jurisdictional Deadlock:** In the cases that remained, half were created by Mutual Legal Assistance Treaty (MLAT) delays. The average data request cross-border wait time was 312 days in this dataset [32].
4. **The Final Disposition:** In the end, only 4 cases led to a criminal conviction or major civil settlement. The rest had also gone out the door to Procedural Exhaustion, where charity victims disappeared owing to the financial toll and mental cost of litigation.

It is relevant to note that countries have been taking efforts to address these shortcomings. The UK'S Online Safety Act 2023 indicated that platforms have a duty of care, but the Act had difficulty reconciling the generation of deepfakes versus their sharing. In January 2026, the UK government amended the Sexual Offences Act 2003 in a fast-track manner, making it a crime to create and request non-consensual sexual deepfakes with the intent to share [33]. South Korea, in turn, responded to the outrage over the SNU case with a landmark amendment in 2024, making even mere possession/viewing of deepfake porn illegal and subject to up to 3 years in prison, while creation/distribution of it was raised to seven years [34]. This, however, is not enough; it has all been purely reactionary, and some of these laws are still mostly impractical for those outside the jurisdiction of the respective countries.

The Jurisdictional Loophole: A Practitioner's Perspective

Let us engage the travails of the lawyers in these cases. The ivory tower of International Law almost always meets a digital brick wall in the everyday life of a Legal Practitioner. There is a pattern of harm in the above analysed case studies, but the primary reason for the attrition rate is that there exists a jurisdictional vacuum for internet activity. The law is still anchored to physical territory, and the Digital War on Women rages in a borderless world. The entry barrier for any practitioner outside the United States, particularly in jurisdictions such as Nigeria or the UK, is the Silicon Valley problem. This is because, since global digital infrastructure does not happen to span large regions but rather lies concentrated in a few square miles of California, nearly every TFGBV case ultimately requires data from US-based providers (Meta, Google, X or Apple).

Under 47 U.S. Code S. 230, platforms are not held liable for content generated by a third party. For a Lagos- or London-based practitioner, this means that if the platform's algorithm begins amplifying a deepfake to millions of people based on the content, then it can't be sued as a publisher for publishing an illegal act. Platforms routinely use Section 230(c)(2) as a crutch to protect

themselves when they are making good faith efforts to moderate, even in the case of their own AI tools (like X's Grok) that create that harm [35]. The impact of this can be felt in the Meloni case, where Italian authorities spent years grappling with the tension between Italian defamation laws and US First Amendment protections. American courts have generally defined satirical deepfakes as protected speech, even those that involve non-consensual sexualization of the sovereign power of a foreign state [36]. This clearly clashes with the successful litigation of cases. Also, the US Electronic Communications Privacy Act prohibits providers from sharing decryption information or the contents of communications with foreign law enforcement authorities without a U.S.-issued warrant. While non-content metadata (IP addresses, login times) can sometimes be shared, the actual messages or the prompt history used to create an AI-deepfake are locked behind ECPA. Therefore, a Nigerian court order for a harasser's IP address will be functionally ignored by the server host in San Francisco unless it is domesticated by running it through a US court, which can cost as much as two years' salary for most individual survivors. Federal privacy law remains focused on controls over automated decision-making technology (ADMT), not cross-border criminal disclosure as of 2026 [37].

The most common mechanism for cross-border evidence is the Mutual Legal Assistance Treaty (MLAT). The intentions of the MLATs are for them to be an exchange of evidence in theory, but in practice, they are the graveyard for TFGBV cases. The Taylor Swift or Daisy Dixon examples had an unusual viral half-life of less than 48 hours. However, MLAT processing time in 2025-2026 has been between 6 and 10 months on average, one year later [38]. When the requested metadata is received, the perpetrator may have already deleted it, wiped clean their digital path or moved to a new encrypted internet domain. Recent findings by the ICC's Office of the Prosecutor (2026) highlight that existing cooperation processes are too slow and bureaucratic for cyber-enabled crimes [39]. For an MLAT to work, the act must be criminal in both countries. The reality is that the denial of requests by law enforcement is common, as many US and Safe Haven jurisdictions still lack a specific criminal offence that includes provisions for non-consensual AI-generated imagery.

On a global scale, the Second Additional Protocol to the Budapest Convention aims to facilitate direct cooperation, which is likely to solve the issues of the MLAT, but this has not seen widespread implementation [40]. As of March 2026, only a few countries (among them Hungary) have completed ratification, with others still in advisory mission stages, which can last for many years. The Digital War on Women exists when regulatory arbitrage perpetrators and Harassment-as-a-Service (HaaS) providers deliberately set up their operations in safe haven countries with weak or non-existent cybercrime laws or human rights protections. In cases such as the Balkan Telegram Bot, criminals leverage these bulletproof servers (in jurisdictions like Seychelles, Russia or Panama). These hosts literally advertise themselves as DMCA-proof as a business model and ignore takedown notices. Telegram's decentralised structure and resistance to creating legal representatives in the Global South have allowed the deepfake ringleaders to operate with complete anonymity.

The Evidence Paradox: The Burden of Proof

In the digital courtroom, the balance of justice is not tilted toward truth, but toward whichever party has the better forensic capabilities. For the survivors of TFGBV, the legal process can become a double burden, requiring them not only to survive the

initial attack, but also to be their own forensic investigators, data travel agents and psychological experts to meet archaic evidentiary standards.

The availability of digital, rather than purely physical, evidence delegitimises the burden of proof in a way that is both technically impossible and financially devastating for the average survivor. In order for it to be accepted in a court of law, digital evidence needs to be captured in such a way that it can be demonstrated that nothing has been altered. This calls for the preservation of specialised hash values and metadata. Most survivors, in the aftermath of their trauma, might take a simple screenshot or delete the original message to stop the harassment, acts that, as of March 2026, are frequently deployed by defence attorneys to move for exclusion on spoliation grounds [41]. In jurisdictions without state-funded digital forensic units, survivors are required to hire private investigators to unmask VPNs or trace decentralised ledger transactions (for AI subscription payments). In 2025-26, fees for these services can range from \$3,000 to \$15,000, effectively price-tying justice as a luxury good [42]. In deepfake cases, the victim is usually on the hook to prove they didn't consent or that it isn't them. Without access to the perpetrator's original prompt logs, this becomes a logical impossibility, relegating the survivor to an enduring state of forensic deficit.

The most insidious development in jurisprudence might be the Liar's Dividend. This is an effect of deepfake technology, and even just the existence of such tools enables criminals to discredit genuine evidence that can incriminate them. Now, defence counsels regularly claim that real recordings of domestic abuse or sexual assault are AI-generated fabrications meant to frame the defendant. By doing this, they impose the reasonable doubt standard on their victims, and the perpetrators manoeuvre through reality's murky waters to escape liability [43]. Most consumer-grade AI used for harassment lacks the watermark built by C2PA (Coalition for Content Provenance and Authenticity). In contrast, existing deepfake detectors produce very unreliable results with high false positives, which traumatise survivors further by implying their authentic testimony is synthetic [44]. Research from the Digital Rights Summit found that judges are increasingly suspicious of all digital video evidence, which leads to the devaluation of the digital record, which subsequently leads to disproportionate harms for women whose primary kind of evidence in TFGBV is, by definition, digital [45].

The process of seeking evidence through the legal system can be almost as traumatic as the original transgression, subjecting survivors to ongoing trauma to prove their case. In many protocols, police demand that survivor's hand over their entire mobile device for unfiltered extraction. This gives investigators access to every part of the survivor's private life, like their photos and health data, private messages, often putting the victim's own character on trial as context for the harassment [46]. In order to meet the best evidence rule, survivors must often watch the harmful material several times, while being questioned by law enforcement, when being examined by forensics or during cross-examinations. This mandatory re-traumatisation leads many women to let go of their cases before they can even reach the trial stage. In addition to this, the courtroom is still a public or semi-public space. The process of trying to prove a deepfake often involves placing the non-consensual imagery before a judge, thus popularising and immortalising some part of the harm under the auspices of legal necessity. This Public Disclosure is a deterrent, as so many survivors would rather work within the silence of impunity

than face the public shame of the courtroom the extent of the psychological burden that all these places on victimised women cannot be overstated [47].

Part III: The Blueprint for Accountability A Call for Digital Extraterritoriality

The Justice Gap laid out in Part II flows directly from the law's physical tethering to 17th-century concepts of Westphalian sovereignty. In 2026 and beyond, to ensure the protection of women, international law must have a Copernican Revolution: It must change its point of reference from the location of the server to the site of victimisation.

The existing jurisdictional framework is based on *lex loci delicti*; the law of the place where the tort was committed. In a TFGBV context, this nearly always protects the perpetrator behind the borders of safe havens or in the Section 230 protections of Silicon Valley. Figuratively, Extraterritorial-Victim-Based Jurisdiction is a primary necessity. According to this doctrine, if a digital harm (like, say, an AI-deepfake) is made public and results in injury within the borders of any given state, then that state's courts must have innate jurisdictional capacity to provide binding global injunctions against both the perpetrator as well as the platform on which it's hosted [48]. This will be difficult because of the nature of international Law and the issues that arise when countries have to sign away some part of their sovereignty for the good of society. There is utility in this if only for preventive purposes.

With reference to antitrust and intellectual property law, something like an expanded effects test could work here. If the digital violence was reasonably intended to, or one could reasonably assume that it would, affect a person's digital integrity and if that individual resides in Jurisdiction A, then Jurisdiction A has the right to compel evidence and deletion of content, no matter where data resides in the cloud [49]. Because digital harm is ubiquitous, the legal remedy must be equally ubiquitous. States need to adopt digital bodily autonomy as a *jus cogens* norm, peremptory norm, for which no derogation is permitted, thereby granting universal civil jurisdiction against TFGBV offenders [50].

On the other hand, current MLATs are fragmented and show that bilateral cooperation is too slow for the velocity of harm. Setting up an International Tribunal for Digital Human Rights (ITDHR) or a separate Protocol under the Budapest Convention could be necessary here. This standard would obligate signatory states and systemically important digital platforms to respond to high-urgency victim alerts within 8-24 hours. Noncompliance would lead to automatic international fines that would be based on the platform's annual revenue [51]. To solve the evidence paradox, the ITDHR would maintain a secure, blockchain-based chain of custody for digital evidence. This would allow survivors to upload verified hashes of abuse imagery, creating a permanent, tamper-proof record that is admissible in any participating domestic court [52]. The development of a global minimum standard for TFGBV (inspired by the updates to the EU Digital Services Act) would eliminate incentives for perpetrators to host content in safe havens. If a platform wants to participate in the international market, it needs to comply with ITDHR's traceability and takedown requirements [53].

Proposing extraterritoriality inevitably raises concerns about digital imperialism and the weaponisation of law by authoritarian regimes to silence dissent. To avoid abuse, Digital Extraterritoriality must be narrowly construed to only cover non-political, gender-based

violations as defined by the Istanbul Convention and UN Special Rapporteur on Violence Against Women. Platforms should be provided a procedural shield to contest manifestly extraterritorial orders that run afoul of international human rights standards, like a fictitious cyber-stalking conviction in an authoritarian state to unmask an activist [54]. Sovereignty in the 21st century should not be defined by the ability to exclude foreign law, but by the ability to protect one's citizens from foreign-hosted harm. A cloud sovereignty model where states pool their forensic and judicial resources to create a *cordon sanitaire* against digital violence will be most valuable [55].

Safety by Design: Shifting Platform Liability

The current legal regime relies on a reactive mode, the digital equivalent of allowing a fire to consume a building before permitting the fire department to intervene. To close the justice gap, international law must move from the outdated notice-and-takedown framework toward a proactive statutory duty of care. This shifts the burden of prevention from the survivor to the multi-billion-dollar entities that profit from the infrastructure of abuse. The notice-and-takedown model, enshrined in the very early 2000s (e.g., the US DMCA and the EU e-Commerce Directive), was meant for a world of textual internet. In the era of AI-driven high-velocity harm, this model is a functional relic. The Taylor Swift and Daisy Dixon cases show us that by the time a notice is processed, whatever harm is done has already spread across the planet and become cemented for all time. A Safety by Design mandate, which treats digital safety as a top-level feature of the product, on par with aviation or automotive safety standards, would address this [56]. Platforms should particularly be classified as Constructive Controllers of their digital environments. This responsibility would mandate that platforms employ automated friction, e.g. mandatory latency for viral dissemination of unverified synthetic media, human-in-the-loop triggers for media flagged by a just-cause high-urgency victim alert [57]. Under the safety by design framework, if a platform's proprietary AI tool (e.g., a built-in image generator) is used to generate NCII through that service, it should be presumed liable unless it can demonstrate that it had state-of-the-art guardrails in place at the time of generation [58].

The Digital War on Women is still profitable because the litigation cost is cheaper than moderation at the current price. An alternative regime of punitive digital damages to reverse this economic incentive is fundamental. Again, if a small number of platforms actively trend or amplify TFGBV through recommendation algorithms on which they make an affirmative decision to place this content, then these platforms should face strict civil liability. When an algorithm identifies a deepfake that produces compelling engagement for tens of thousands and pushes it into millions of feeds, the platform is not a neutral pipe but rather a coy-distributor [59]. Introducing new charges such as corporate digital manslaughter or systemic recklessness for platform executives who deprioritise safety engineering in favour of engagement metrics with full knowledge that they are doing so. If a platform does not intervene when warned of a deepfake ring (as in the SNU Case), that should be treated as a criminal omission [60]. Total worldwide fines, based on models like those in the EU Digital Services Act (DSA), should allow for fixed penalties, ranging between 6% to 10% of global annual turnover for repeated failure to respect the safety by design module. This makes safety a boardroom priority, rather than a budgetary afterthought [61].

Modern algorithms are often described as black boxes, and thus, these platforms can hide behind the claim of unforeseen

consequences when violence goes viral. That strategic opacity only has one antidote: transparency. A Human Rights Impact Assessment of all new AI tools before rollout. Independent, accredited human rights auditors must be allowed to stress-test their models for gendered bias and ease of abuse-generation [62]. Survivors should have a legal right to an automated forensic audit trail of how the harmful content came before their eyes, which users they interacted with first, and why the platform's internal safety filters did not catch it. To prevent safe havens from lagging, the International Tribunal for Digital Human Rights (ITDHR) should maintain a library of open-source safety guardrails. All global platforms must implement these baseline protections or risk getting immediately grey-listed from the marketplace for financial services globally.

Conclusion: Restoring Digital Citizenship

From the case studies of systemic failure to the jurisdictional black hole of Silicon Valley, the research presented in this paper can lead to no other unavoidable conclusion: The Digital War on Women is above all a crisis of citizenship. When the law cannot protect a specific demographic from being systematically erased, silenced or pornified in the digital public square, that is not just a technological failure; it is a failure of the social contract.

For far too long, the digital realm has been treated like the Wild West, where human rights are being sacrificed at the altar of platforms' engagement. This means challenging the framing of TFGBV to be seen not as a private grievance, and instead understanding it to be an attack on democratic participation. In 2026, the internet is not an optional space; it has become the central infrastructure for education, commerce and political discourse. If women cannot enter this space without the risk of AI-generated symbolic assassination, then they are functionally denied their rights to participate in modern society. To restore digital citizenship, the Rule of Law must be as enforceable cloud-rights as in street-rights. This means that a woman's right to bodily autonomy must be a portable right, one that accompanies her through all the servers, platforms and borders.

The legal profession is at a crossroads. The law must evolve or become obsolete; generative technology has severed the common-sense relationship between physical act and human harm. The 21st-century practitioner, therefore, cannot remain technologically neutral. International law must move toward a proactive regulatory model. The trade here is the slow, bilateral diplomacy of the MLAT for the real-time, automated enforcement of the International Tribunal for Digital Human Rights (ITDHR). Sovereigns cannot be a protection anymore when it comes to safe havens that shelter digital abusers. Cooperative Extraterritoriality, the pooling of states' judicial power, is how national or international law will look in the future, because if The Ghost in the Code is to be met with the fullest application of law, it must be brought about through mutual agreement.

The 85% Justice Gap is not an act of God; it is a policy choice. By choosing to prioritise platform growth over victim safety and territorial sovereignty over human rights, society has allowed the digital world to become a weapon against women. As legal researchers and practitioners, our task is to dismantle this infrastructure of impunity. We must ensure that the algorithmic gavel falls not on the victim, but on the perpetrator and the platforms that facilitate their violence. The Ghost in the Code can be exorcised, but only with the collective political and legal will to redefine justice for the digital age.

Bibliography

Table of Cases

- *Arias v OpenAI* (US District Court, Northern District of California, 17 January 2026).
- *Google LLC v Equustek Solutions Inc* [2017] 1 SCR 824.
- *Meloni v Scurosu & Anor* (Sassari Court, Italy, October 2024 - February 2025).
- *The Economic Community of West African States (ECOWAS) Court of Justice v Federal Republic of Nigeria* (Judgment on Section 24 of Cybercrime Act, 2024).

Table of Statutes and Treaties

International & Regional

- Budapest Convention on Cybercrime (Second Additional Protocol on enhanced co-operation and disclosure of electronic evidence, 2024/2026 Updates).
- Council of Europe Convention on preventing and combating violence against women and domestic violence (Istanbul Convention).
- Regulation (EU) 2022/2065 on a Single Market for Digital Services (Digital Services Act) [2022] OJ L277/1.
- Regulation (EU) 2023/1543 on European Production and Preservation Orders for electronic evidence in criminal matters [2023] OJ L191/118.

Domestic

- Communications Decency Act 1996, 47 USC § 230 (United States).
- Cybercrime (Prohibition, Prevention, etc) Act 2015 (Nigeria).
- Electronic Communications Privacy Act 1986, 18 USC §§ 2510-2523 (United States).
- Online Safety Act 2023 (United Kingdom).
- Sexual Crimes Punishment Act (Amendment regarding Deepfakes, South Korea, 2024).
- Violence Against Persons (Prohibition) Act 2015 (Nigeria).

Secondary Sources

Books

- Citron DK, *The Fight for Privacy: Protecting Dignity, Identity, and Love in the Digital Age* (Norton 2022).

Journal Articles

- Chesney R and Citron D, 'Deepfakes and the New Disinformation Games' (2019) 114 Foreign Affairs 1 (Practitioner's Update 2026).
- Ryan G, 'Generative AI and the material contribution of content' (2026) *Harvard Journal of Law & Technology*.
- 'Digital Bodily Autonomy as a Jus Cogens Norm' (2025) 15(4) *Journal of International Criminal Justice*.
- 'The New Westphalia: Redefining Sovereignty in the Age of AI' (2026) 38(1) *Global Policy Journal*.

Reports and Policy Papers

- Association for Progressive Communications, 'Bridging the gap: Addressing technology-facilitated gender-based violence in global AI governance' (25 February 2026).
- Centre for Countering Digital Hate (CCDH), 'The State of Digital Misogyny: The Rise of Harassment-as-a-Service' (Research Report 2025).
- Chatham House, 'Securing justice for cyber-enabled international crimes' (26 January 2026).
- European Commission, 'Report on the Cross-border Access to Electronic Evidence' (2025) COM (2025) 45 final.
- ICC Office of the Prosecutor, 'Policy Paper on Digital

- Evidence and International Crimes' (2026).
- Paradigm Initiative, 'New research exposes alarming surge in Technology-Facilitated Gender-Based Violence across Africa' (11 December 2025).
- UN Women, 'Closing the Justice Gap: New Proposals for Digital Extraterritoriality' (White Paper, January 2026).
- UN Women, 'Secondary Victimisation in the Digital Age: A Policy Review' (January 2026) UN Doc E/CN.6/2026/7.

Digital Sources

- Council of Europe, 'Hungary becomes the third state to ratify the Second Additional Protocol to the Convention on Cybercrime' (5 February 2026) <https://www.coe.int/en/web/cybercrime/>.
- eSafety Commissioner, 'Safety by Design: Overview' (2024) <https://www.esafety.gov.au/industry/safety-by-design>.
- 'e-Evidence Regulation: Key compliance takeaways for service providers by 2026' (TwoBirds, 27 October 2025) <https://www.twobirds.com>

References

1. Sturges J (2024) In January 2024, artificially generated pornographic images of pop superstar Taylor Swift circulated. *Georgetown Journal of Gender and the Law* <https://www.bbc.com/news/world-us-canada-68124091>.
2. Ressa M (2022) The Chilling: Global trends in online violence against women journalists. UNESCO <https://www.unesco.org/en/articles/chilling-global-trends-online-violence-against-women-journalists>.
3. Meloni v Scurosu & Anor (2024) Testimony of Giorgia Meloni. Sassari Court, Italy <https://www.bbc.com/news/world-europe-69426712>.
4. UN Women (2025) AI-powered online abuse: How AI is amplifying violence against women <https://www.unwomen.org/en/articles/story/ai-powered-online-abuse>.
5. Sexual Offences Act 2003 (2023) as amended by Online Safety Act <https://www.legislation.gov.uk/ukpga/2003/42>.
6. Ofcom (2026) Investigation into X's compliance with the Online Safety Act <https://www.ofcom.org.uk/make-a-complaint>.
7. Chayn (2025) Encryption is necessary for protecting child and adult survivors of gender-based violence <https://blog.chayn.co>.
8. Chesney R, Citron D (2019) Deepfakes and the New Disinformation Games. *Foreign Affairs* 114: 1.
9. TheyWorkForYou (2026) Artificial Intelligence and Online Violence Against Women and Girls <https://www.theyworkforyou.com>.
10. Ofcom (2025) Guidance on a safer life online for women and girls <https://www.ofcom.org.uk/online-safety>.
11. CDT Europe (2026) Policy paper calls for harmonised legal action against tech-facilitated gender-based violence <https://cdteu.org>.
12. European Parliament and Council (2022) Regulation on a Single Market for Digital Services. Official Journal L <https://eur-lex.europa.eu/eli/reg/2022/2065>.
13. TechPolicy Press (2025) Addressing Cumulative Online Harms Through Human Rights Law <https://techpolicy.press>.
14. Sexual Offences Act 2003 (2025) s 138. TAKE IT DOWN Act 2025 <https://www.congress.gov/bill/119th-congress/senate-bill/146>.
15. UN Women (2026) No country in the world has reached full legal equality <https://www.unwomen.org/en/news/stories/2026/03/no-country-has-reached-full-equality>.
16. Council of Europe (2011) Istanbul Convention CETS No 210 <https://www.coe.int/en/web/conventions/full-list?module=treaty-detail&treatynum=210>.
17. GREVIO (2021) General Recommendation No 1 on the Digital Dimension <https://www.coe.int/en/web/istanbul-convention/grevio>.
18. UN Women (2026) Ensuring and strengthening access to justice. UN Doc E/CN.6/2026/3 <https://www.un.org/womenwatch/daw/csw>.
19. (2025) Sovereignty in Cyberspace from the Viewpoint of International Law. *Asian Journal of International Law* 14: 189.
20. IRP Lawyer (2026) International Law in a Borderless World <https://irplawyers.com/blog/international-law-in-a-borderless-world-why-legal-strategy-can-no-longer-stop-at-national-borders/>.
21. Online Safety Act 2023 (2023) c 50 (UK) <https://www.legislation.gov.uk/ukpga/2023/50>.
22. Meloni v Scurosu & Anor (2024) Testimony of Giorgia Meloni. Sassari Court <https://www.bbc.com/news/world-europe-69426712>.
23. Communications Decency Act 1996 (1996) 47 USC § 230 <https://www.law.cornell.edu/uscode/text/47/230>.
24. Bernama (2024) Italian PM Meloni Testifies in Deepfake Porn Video Trial <https://bernama.com/en/news.php?id=2349753>.
25. UN Women (2025) AI-powered online abuse. CCDH (2025) The State of Digital Misogyny <https://counterhate.com/research/state-of-digital-misogyny/>.
26. UN Women (2026) When justice fails <https://www.unwomen.org/en/articles/explainer/when-justice-fails-why-women-cant-get-protection-from-ai-deepfake-abuse>.
27. ADRN (2024) South Korea's Rising Deepfake Sex Crimes <https://www.eai.or.kr>.
28. Sturges J (2024) Taylor Swift, Deepfakes, and the First Amendment. *Georgetown Journal of Gender and the Law* <https://www.law.georgetown.edu/gender-journal/>.
29. DEFIANCE Act 2024 (2024) US <https://www.congress.gov>.
30. ECPMF (2026) Gender-based violence, a growing weapon against women journalists <https://www.ecpmf.eu>.
31. Citron DK (2022) The Fight for Privacy. New York: Norton <https://www.norton.com/books/9780393882315>.
32. European Commission (2025) Report on Cross-border Access to Electronic Evidence. COM (2025) 45 final <https://eur-lex.europa.eu>.
33. Durham University (2026) Professor Clare McGlynn's research underpins new laws <https://www.durham.ac.uk/news-events/latest-news/>.
34. Sexual Crimes Punishment Act (2024) Amendment, South Korea <https://www.pbs.org/newshour/world/in-south-korea-rise-of-explicit-deepfakes-wrecks-womens-lives>.
35. Communications Decency Act 1996 (1996) 47 USC § 230 <https://www.law.cornell.edu/uscode/text/47/230>.
36. Ryan G (2026) Generative AI and the material contribution of content. *Harvard Journal of Law & Technology* <https://jolt.law.harvard.edu>.
37. Electronic Communications Privacy Act 1986 (1986) 18 USC §§ 2510-2523 <https://www.law.cornell.edu/uscode/text/18/part-I/chapter-119>.
38. Chatham House (2026) Securing justice for cyber-enabled international crimes <https://www.chathamhouse.org/2026/01/securing-justice-cyber-enabled-international-crimes>.
39. ICC Office of the Prosecutor (2026) Policy Paper on Digital Evidence <https://www.icc-cpi.int>.
40. Budapest Convention on Cybercrime (2024) Protocol II <https://www.coe.int/en/web/cybercrime>.

41. Citron DK (2022) The Fight for Privacy <https://wnnorton.com/books/9780393882315>.
42. TechPolicy Press (2026) The Financial Cost of Digital Justice <https://www.techpolicy.press/the-financial-cost-of-digital-justice/>.
43. Chesney R, Citron D (2019) Deepfakes and the New Disinformation Games <https://www.foreignaffairs.com>.
44. Content Authenticity Initiative (2026) C2PA and the Future of Content Authenticity <https://contentauthenticity.org/blog/2026-update>.
45. Digital Rights Summit (2026) Devaluing the Digital <https://www.digitalrights.report>.
46. UN Women (2026) Secondary Victimization in the Digital Age. UN Doc E/CN.6/2026/7 <https://www.unwomen.org>.
47. (2026) The Psychological Burden of Forensic Disclosure. Journal of Cyber-Psychology and Law 12: 1 <https://www.journalofcyberpsychology.com>.
48. UN Women (2026) Closing the Justice Gap <https://www.unwomen.org>.
49. Citron DK (2022) The Fight for Privacy. Google LLC v Equustek Solutions Inc (2017) 1 SCR 824.
50. (2025) Digital Bodily Autonomy as a Jus Cogens Norm. Journal of International Criminal Justice 15: 4.
51. European Commission (2026) Proposal for a Regulation on Fast-Track Removal. COM 112.
52. (2026) Blockchain for Justice: Securing the Digital Chain of Custody. Stanford Technology Law Review 44.
53. European Commission (2023) Regulation (EU) 2023/2854 Data Act <https://eur-lex.europa.eu/eli/reg/2023/2854>.
54. Chatham House (2026) Securing justice for cyber-enabled international crimes <https://www.chathamhouse.org/2026/01/securing-justice-cyber-enabled-international-crimes>.
55. (2026) The New Westphalia: Redefining Sovereignty in the Age of AI. Global Policy Journal 38.
56. Online Safety Act (2023) UK. eSafety Commissioner (2024) Safety by Design <https://www.esafety.gov.au/industry/safety-by-design>.
57. Citron DK (2022) The Fight for Privacy <https://wnnorton.com/books/9780393882315>.
58. European Commission (2026) Proposal for a Regulation on Fast-Track Removal. COM 112.
59. Ryan G (2026) Generative AI and the material contribution. Harvard Journal of Law & Technology <https://jolt.law.harvard.edu>.
60. UN Women (2026) Closing the Justice Gap <https://www.unwomen.org>.
61. European Parliament and Council (2022) Regulation on Digital Services Act OJ L 277.
62. Association for Progressive Communications (2026) Bridging the gap <https://www.apc.org>.

Copyright: ©2026 Jessica Adanna Awurum. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.