Torture, 2026, 36(1), 3-20
https://doi.org/10.7146/torture.v36i1.167351 *Editorial*

Coercion without contact: New technologies and the boundaries of torture

Pau Perez-Sales[1]

1 Editor-in-Chief. Correspondence to pauperez@runbox.com

Key points of interest

- Technologically mediated coercion can produce severe suffering without direct physical contact. Surveillance, exposure, radical uncertainty, reputational destruction, automated exclusion, and induced vulnerability may erode agency, identity, relational life, and collective belonging.
- The torture/CIDT threshold must be assessed cumulatively and contextually. In digital and AI-mediated environments, harm may be diffuse, persistent, opaque, collective, and difficult to attribute to a single perpetrator, but still legally and clinically significant.
- Accountability requires linking human-rights law with power analysis. State and private infrastructures of surveillance, profiling, manipulation, and social control increasingly operate together, requiring updated documentation, survivor-centred assessment, and clearer standards for distributed responsibility.

Abstract

Introduction: This editorial revisits Internet and Communications Ill-Treatment and Torture in light of generative AI, biometric surveillance, spyware, automated inference, neurotechnology, and platform-based coercion. It asks how new technologies reshape the boundaries between coercion, cruel, inhuman or degrading treatment, and torture. *Methodology*: The paper uses conceptual analysis, typology-building, and normative human rights interpretation, informed by a purposive interdisciplinary review. *Results*: The editorial proposes a three-layer framework organised around the human need under attack, the method through which harm is produced, and the site where coercion occurs. It identifies three overlapping domains: mental-directed interventions, social control, and social influence. Across scenarios including technologically assisted interrogation, protest policing, e-carceration, border governance, armed conflict, and digital authoritarianism, the analysis shows that technologically mediated coercion may produce severe suffering without direct physical contact. Harms may arise through surveillance, exposure, radical uncertainty, reputational destruction, isolation, automated exclusion, manipulation of perception, and induced vulnerability. The paper argues that severity should be assessed cumulatively and contextually, including impacts on agency, identity, relational life, collective belonging, and conditions of existence. It also highlights the difficulty of attribution when states, companies, platforms, vendors, data brokers, and automated systems jointly produce coercive environments. Existing human rights frameworks remain relevant but require doctrinal refinement, stronger accountability tools, and better methods for documenting diffuse, opaque, and collective harms. *Conclusion*: Torture and ill-treatment do not end where screens begin; technological mediation requires updated legal, clinical, and evidentiary frameworks.

At Torture Journal, we provide a platform for conceptual debate and for exploring practical responses—particularly in rehabilitation and accountability—to emerging forms of coercion, ill-treatment, and torture.

An earlier editorial (Pérez-Sales & Serra, 2020) proposed the concept of Internet and Communications Ill-Treatment and Torture (ICIT) to capture forms of severe psychological suffering intentionally produced, instigated, or aggravated through digital and communication technologies for purposes analogous to those recognised under the UN Convention against Torture (UNCAT). That piece argued that digitally mediated abuse could generate torturing environments by acting through two main pathways: fear-producing practices, such as threats, surveillance, monitoring, and real-time control, and identity-targeting practices, such as doxxing, defamation, exposure, humiliation, and public degradation. It also highlighted several features that made these harms especially severe: anonymity and impunity, mediated interaction, multiplicity of aggressors, permanent stress, and the persistence of a digital footprint that transforms humiliation into enduring shame.

However, that framework was developed at a moment when many technologies that now shape contemporary coercion were still incipient or marginal, and before the widespread social and political impact of generative artificial intelligence (AI), large-scale automated inference, synthetic media, commercial spyware ecosystems, and increasingly sophisticated brain-computer and neurotechnological interfaces.

The present editorial therefore revisits and expands that initial framework in three directions: first, by moving beyond predominantly individualised internet harassment toward collective, environmental, and systemic forms of coercion; second, by incorporating the role of new technologies and specifically generative AI in the production of suffering and control; and third, by examining more closely the legal and ethical framework and the problems of distributed responsibility, opacity, and evidentiary difficulty that arise.

The argument proceeds in four steps. First, it proposes a framework structured around three analytical dimensions - target, method, and site - and three operative domains: mental-directed interventions, social control, and social influence. It illustrates this framework through scenarios and tables designed as heuristic tools rather than exhaustive classifications. Second, it revisits the question of severe suffering and technologically mediated harm in situations that are often cumulative, environmental, and identity-affecting. Third, it examines the legal implications, including distributed responsibility, the adequacy of existing human rights frameworks, and the relevance of emerging AI regulation. Fourth, and derived from all the above, it proposes some guiding ideas on the distinction between CIDT and torture in technologically mediated coercion.

The purpose of this review is not to offer a dystopian account in which AI appears solely as a source of risk. AI can also support human rights monitoring, documentation, analysis, and accountability. Used cautiously, it may strengthen anti-torture work—not by replacing investigators, clinicians, or legal judgment, but by extending evidentiary reach and reducing blind spots. Machine-learning tools can process satellite imagery, open-source material, media reports, and multilingual data to detect destruction, detention-related patterns, and spikes in violence or hate speech, supporting early warning and targeted inquiry (Bachelet, 2018; Dulka, 2023). Current applications include, for instance, detecting village destruction in Darfur, monitoring violence in Myanmar, tracking death penalty cases, forecasting displacement, facilitating the analysis and translation of testimonies, and reducing secondary trauma by filtering graphic material (Marin, 2020; Dulka, 2023). More broadly, these technologies can help build and analyse datasets on violations, identify trends and gaps, and cross-check findings against independent sources (Bachelet, 2018).

Methodologically, the editorial combines conceptual analysis, typology-building, and normative human rights interpretation, informed by a purposive interdisciplinary review of scholarship and institutional sources. Its aim is not to provide an exhaustive systematic review, but to clarify categories, identify neglected domains, and propose a framework for assessing technologically mediated coercion in relation to torture and CIDT. The aim is theory-building and conceptual clarification.

For the purposes of this review, and following ICRC (2019), OHCHR (2022,2024a, 2024b, 2025) and the European Union Agency for Fundamental Rights (2023), new technologies refers to emerging or newly integrated systems of surveillance, inference, automation, biometric and neurophysiological capture, communication control, and behavioural or environmental modulation that reshape how coercion, suffering, and accountability are produced and distributed. In the context of torture and CIDT, what makes them "new" is not novelty alone, but their capacity to create qualitatively new, intensified, scalable, less visible, or more diffused forms of coercion.

Considering the legal definition of torture in the UN Convention, the analysis focuses on coercion rather than on a comprehensive list of purposes. While the Convention mentions examples such as obtaining information or self-incrimination, humiliation, punishment, or discrimination, these are not exhaustive, and coercion can be understood as a meta-purpose that helps structure the analysis. Coercion is used here as an

analytic lens to organise the inquiry, not as a substitute for the legal elements of torture under the UNCAT.

Analytical dimensions

To provide a parsimonious way of organising the growing literature, we group the analysis into three dimensions: target, method, and site (or context).

1. **Targeted human need:** If emerging technologies are understood as systems of coercion acting on individuals and human groups, the first level of analysis is the human needs they target – the dimension of human existence under attack. A number of key axes can be identified. These include: (1) bodily integrity and physiological regulation; (2) safety and basic security, including exposure to unpredictability, threat, surveillance, arbitrary intervention, and other technologies that undermine the sense of safety; (3) agency, autonomy and cognitive coherence, encompassing orientation, the capacity to understand and reflect on what is happening, and the ability to make choices, give or withhold consent, take initiative, maintain privacy and personal boundaries, and retain ownership over one's body and mind, including the capacity to resist; (4) identity, dignity and selfhood, including self-respect, gender and cultural identity, moral integrity, religious and political identity, and continuity of self; (5) attachment, trust and relational bonds, such as family ties, solidarity, peer support, care, trust in others, communication, and structures of dependence; and (6) social belonging and collective existence, including community life, reputation, leadership, intergenerational continuity, and political participation (Pérez-Sales, 2026).
2. **Method:** The second level of analysis concerns how emerging technologies act upon individual and collective needs. Modes of harm may be informational, relational, sensory, neural, bodily, or environmental, often operating in combination. Technologies may, for example, alter sensory input, interfere with neural processes, constrain bodily functions, or shape physical and digital environments in ways that undermine basic conditions of life (OHCHR, 2022, 2024, 2025).

   These interventions are typically used to identify and profile individuals or groups, predict behaviour or risk, isolate or fragment social ties, manipulate perception or emotion, punish or degrade, incapacitate, induce compliance, or extract information or confessions. In practice, these functions are rarely discrete and tend to operate as integrated systems of surveillance, inference, and intervention acting continuously and at scale.

   Together, we propose an analysis that moves from the human function under attack to the mechanisms through which harm is produced, rather than starting from the technologies themselves.
3. **Site or context:** The third level concerns the site or operational context in which these mechanisms are used. This layer situates technologies within the concrete environments where coercion is enacted, normalised, and often rendered invisible. Relevant sites include, but are not limited to, custodial settings (such as prisons, police custody, immigration detention, and psychiatric institutions); public-order contexts (including protests and crowd-control environments); border and migration regimes (transit zones, camps, and deportation processes); clinical and care settings (hospitals, residential institutions, and disability services); digital and hybrid environments (online platforms, communication networks, and systems of remote monitoring); and conflict or military contexts (battlefields, occupied territories, and security operations).

   Attention to the site is critical because it shapes how technologies operate, how harm is experienced, and how accountability is obscured or enforced. The same mechanism—such as surveillance, isolation, or behavioural manipulation—may have profoundly different meanings and effects depending on whether it occurs in a prison, a refugee camp, a home, or a digitally mediated environment. The analysis should be embedded in context, perspective, and conditions, rather than abstract and universal.

Three domains of technologically mediated coercion

We identify three domains of coercion, integrating target, method, and context (see Table 1). The three dimensions in the previous section provide the analytical lens; the three domains that follow are the substantive fields in which that lens is applied.

– **Mental-directed interventions** refer to technologies designed to monitor, infer, extract, alter, bias, induce, or modulate mental states or their physiological proxies. What characterises this domain is its direct orientation toward mental privacy, autonomy, freedom of thought, and mental integrity. In contexts of detention, interrogation, policing, or coercive governance, such interventions are particularly significant, as they may weaken resistance, enable the extraction of information, shape decision-making, and expand the means through which individuals are rendered compliant, vulnerable, or exposed to abuse.

- **Social control** refers to the use of digital, data-intensive, and AI-enabled technologies to render individuals and populations visible, identifiable, classifiable, trackable, and governable. It concerns the production and use of knowledge about who people are, where they are, what they may do, and what may be done to them. In contexts relevant to ill-treatment and torture, these systems may facilitate selective targeting, arrest, isolation, and other coercive measures, as well as the construction of conditions in which abuse becomes easier to inflict and more difficult to contest.
- **Social influence** refers to the use of digital and platform-based technologies to shape perceptions, beliefs, emotions, attention, behaviour, participation, and social relations through persuasion, manipulation, disinformation, intimidation, or reputational pressure. It is oriented toward influencing what people think, feel, trust, fear, or are willing to do, at the individual, group, or broader community level.

These three domains are analytically distinct but might overlap: surveillance and profiling can feed targeted persuasion; spyware can both extract information and generate fear; and neurotechnology may be embedded in wider systems of monitoring and behavioural management. Furthermore, AI propaganda can identify psychological vulnerabilities and disseminate personalised content to shape perception and behaviour without resorting to "brain reading" technologies.

The tables that follow are intended as a heuristic map rather than as separate inventories. Table 1 provides a synthetic overview of the three domains of coercion; Table 2 situates the framework in concrete scenarios; and Tables 3–5 unpack the three domains in greater detail. They are meant to be read relationally, as different entry points into the same coercive field.

Scenarios of coercion

Technologically assisted interrogation and compliance

The use of technology to assist interrogation is not new, but its scope has expanded substantially (High-Value Detainee Interrogation Group, 2016). The traditional polygraph — measuring heart rate, respiration, and galvanic skin response — has long been contested on scientific grounds and is inadmissible in most legal systems. What is new is the range and invasiveness of the technologies now deployed or proposed, and the shift toward AI-assisted systems capable of processing multiple physiological and behavioural signals simultaneously.

Current and emerging technologies fall into four categories (Table 2 and 3). The first is peripheral psychophysiological monitoring: devices that record autonomic responses — skin conductance, heart rate variability, facial temperature changes via thermal imaging — to infer deception, stress, or concealed knowledge. The second is neural acquisition: EEG-based systems that detect the P300 brainwave response, which occurs when a person recognises certain stimuli even while verbally denying it; and fMRI-based approaches that attempt to identify neural correlates of deception through blood-flow patterns in prefrontal regions associated with cognitive control[1]. The third is behavioural and linguistic analysis: AI systems that process micro-expressions, gaze patterns, voice stress, speech content, and body language to generate credibility scores or flag deception. Systems such as AVATAR have been tested in border control and security contexts (Kalodanis, 2025). The

1 P300-based EEG systems (sometimes called Brain Fingerprinting) and fMRI deception detection have both been subject to admissibility proceedings in criminal courts. Neither has achieved general scientific or legal acceptance. For a comprehensive technical review, see Elbatanouny et al. (2025). For the legal framework governing coercive brain-reading in European criminal justice, see Ligthart (2022).

***Table 1**. Ill-treatment and torture in technologically advanced societies*

Mental directed interventions Surveillance/manipulation of cognitive, affective or other neural functions.	**Social control** Technologically mediated monitoring and population control.	**Social influence** Digitally mediated coercion and Digital authoritarianism
1. Mental privacy: Acquisition and monitoring methods. 2. Mental State modulation and manipulation.	1. Data collection, interception, and device access 2. Identification and tracking 3. Profiling prediction and social sorting. 4. Connectivity, service, and digital access control	1. Disinformation and narrative control 2. Covert persuasion and behavioural influence 3. Harassment, intimidation, and reputational harm

Table 2. *Scenarios of technologically mediated coercion*

Technologically assisted interrogation and compliance

- Peripheral psychophysiological monitoring (skin conductance, heart rate, thermal imaging)
- Neural acquisition: EEG/P300 recognition response; fMRI deception detection
- Behavioural and linguistic analysis: micro-expressions, gaze, voice stress; AI credibility scoring
- AI-assisted protocol optimisation: question sequencing and condition management to maximise pressure
- Real-time physiological monitoring to sustain interrogation conditions and calibrate vulnerability

Crowd control, population surveillance, and the right to protest

- Mass surveillance, real-time identification and facial recognition at protest
- Predictive profiling, pre-emptive targeting, and post-protest retrospective tracking and prosecution
- Less lethal and directed energy weapons: kinetic impact projectiles, chemical irritants, acoustic devices, drones
- Digital reprisals: blacklisting, harassment, and isolation of activists after protest

E-carceration

- Technology-assisted prisons and closed institutions: cameras and body imaging, permanent surveillance, remote hearings, automated restriction systems
- Electronic monitoring & remote custody – house arrest

Border governance and mobility control

- Data mining and sharing
- Use of drones and GPS tracking with preventive interception and deterrence. Identity checking with biometrics, including facial recognition
- Psychophysiological "lie-detector" devices
- Algorithmic risk assessments and automated analysis and decision-making of visa and asylum applications

Warfare and armed conflict

- AI-assisted targeting and autonomous weapon systems
- Fifth-Generation Warfare (5GW): psychological operations, disinformation, narrative control, identity manipulation
- Cyber-physical attacks on civilian infrastructure and essential life systems

fourth is AI-assisted protocol optimisation: algorithmic systems that recommend interrogation schedules, condition management, and question sequencing designed to maximise psychological pressure and exploit detected vulnerabilities.

Beyond detection, these technologies have a second function: managing the conditions of interrogation itself. Real-time physiological monitoring can be used to sustain alertness, calibrate stress levels, and signal to interrogators when a subject is at maximum psychological vulnerability — transforming medical monitoring into an instrument of coercive control.

From a human rights perspective, these technologies raise concerns at several levels. Where they are used coercively — without meaningful consent, in conditions of detention, or to extract information under duress — they constitute a direct attack on mental privacy and cognitive liberty (Ligthart, 2022; Dore-Horgan et al., 2026). Even where their stated purpose is assessment rather than coercion, the conditions of detention render genuine voluntariness impossible. The scientific validity of most systems remains contested: accuracy rates achieved in controlled laboratory conditions do not transfer reliably to real interrogation settings, and false positives impose severe consequences on the persons concerned (High-Value Detainee Interrogation Group, 2016; Elbatanouny et al., 2025)[2]. The

2 The baseline scientific review of interrogation methods — including a systematic evaluation of the evidence for

***Table 3**. Mental-directed interventions: Monitoring, inference, and modulation of mental states and neural functions*

Mental – directed interventions	
Methods	**Purposes related to ill-treatment and torture (selected examples)**
1. Mental privacy: Acquisition and monitoring methods · Neural acquisition technologies ("brain reading"). obtaining neural signals or decoding mental content/states from the brain. Detect mental data: intention, recognition, memory-related signals, arousal, cognitive load, or other inferred mental data. · Peripheral psychophysiological monitoring. Monitoring non-neural bodily correlates (peripheral indicators) linked to mental, affective, or attentional states: autonomic, behavioural, facial, ocular, or voice-linked indicators. 2. Mental State modulation and manipulation. · Neural modulation technologies ("brain writing"). Altering, biasing, or inducing mental states through direct neural intervention. Targeting cognition, attention, affective responses, memory, pain perception, or responsiveness, among others. · Non-neural modulation of cognition and affect. Psychophysiological manipulation systems, sensory or digitally mediated modulation techniques that produce arousal, fear, stress, or other specific psychophysiological-mediated mental states.	1. Interrogation of detainees – Using extraction or inference methods to support questioning, assess recognition, infer reactions, or claim access to concealed mental content. Directing interrogations and unconsented or coercive credibility analysis. 2. Surveillance, profiling and vulnerability mapping. Using monitoring or inference to classify persons, detect states, map susceptibilities, or target interventions against persons. 3. Incapacitation, disorientation, or weakening of resistance. Using modulation methods to impair self-regulation, affective stability, attentional control, or the capacity to resist.
Technologies (examples): Brain-computer interfaces, either implant-based (neuro-implants) or read-out systems (sensors).	

use of these technologies against racialised and marginalised groups compounds pre-existing patterns of discriminatory enforcement (Noriega, 2020).

Crowd control, population surveillance, and the right to protest

The policing of public assemblies has become a primary site for the convergence of surveillance, physical coercion, and retrospective repression, each increasingly technology-mediated and mutually reinforcing (Table 2).

Before and during a protest, digital systems serve identification, profiling, and preemption. Facial recognition technology, drones, AI-assisted video analysis, and mass biometric collection are deployed to identify participants and build searchable records that transform political participation into traceable data (OHCHR, 2024a; Amnesty International, 2024). The chilling effect of such surveillance — deterring participation even before any coercive act occurs — is itself a harm to assembly rights (Special Rapporteur on Peaceful Assembly and Association, 2024). Predictive policing tools are used to pre-empt gatherings through targeted interventions, while retrospective tracking enables prosecution and blacklisting of participants long after a protest ends (Melgaço & Monaghan, 2021; Amnesty International, 2025)

Physical dispersal continues to rely on less-lethal weapons whose designation obscures serious documented harms. Kinetic impact projectiles, chemical irritants, water cannon, stun grenades, acoustic devices, and drone-deployed dispersal systems have caused thousands of serious injuries, permanent disabilities, and deaths globally (INCLO, PHR & Omega Research Foundation, 2023; Omega Research Foundation & Amnesty International, 2023; McEvoy, Corney & Haar, 2024).

After a protest, digital traces enable targeted follow-up: footage analysed to identify participants, social media monitored for organisational networks, and databases used to flag activists for subsequent harassment or criminalisation. This post-event repression extends coercive control well beyond the moment of physical dispersal (Avis, Marciniak & Sapignoli, 2024).

physiological and behavioural credibility assessment — remains the High-Value Detainee Interrogation Group (2016) report, which concluded that no available technology reliably distinguishes truthful from deceptive responses across operationally realistic conditions.

E-carceration

Smart prison systems may reduce direct staff-prisoner contact, replace visits and legal encounters with digital substitutes, and intensify forms of hyper-visibility and behavioural regulation. Electronic monitoring similarly extends custody into the community through GPS tracking, geo-fencing and condition-based liberty (Table 2). These developments do not amount in themselves to torture, but they may contribute to cruel, inhuman or degrading treatment where they produce chronic surveillance stress, isolation, humiliation, erosion of autonomy, or the systematic reduction of meaningful human contact (Malek et al., 2023; McKay, 2021).

Border governance and mobility control

The growing use of AI and related systems to identify, classify, predict, and filter mobile populations are used not only for border surveillance, but for broader mobility management: identity verification, visa and asylum screening, profiling, risk scoring, biometric registration, and prediction of migratory flows (Office of the United Nations High Commissioner for Human Rights, & University of Essex, 2023; IOM, 2021). These systems do not simply record movement; they sort individuals into categories of suspicion, admissibility, or removability, often through opaque and weakly contestable automated decisions (Beduschi, 2021). While not amounting to torture per se, they may contribute to cruel, inhuman or degrading treatment where they generate prolonged uncertainty, fear, dehumanisation, family separation, or exposure to severe vulnerability and coercion.

Warfare and armed conflict

Contemporary warfare illustrates how the three domains — mental-directed interventions, social control, and social influence — can converge and be deployed simultaneously against both combatants and civilian populations.

The first domain concerns AI-assisted targeting and autonomous weapon systems: platforms that, once activated, can select and engage targets without further human intervention. These systems raise acute questions under international humanitarian law regarding distinction, proportionality, and precaution, and create new forms of distributed accountability in which responsibility for civilian harm is fragmented in ways existing legal frameworks are poorly equipped to address (ICRC, 2026; Human Rights Watch, 2025)[3].

The second domain is what military analysts describe as fifth-generation warfare (5GW): conflict conducted primarily through the manipulation of perception, identity, and shared reality, using psychological operations, disinformation, and social engineering at a population scale (Krishnan, 2016, 2024). AI accelerates this domain substantially, enabling the personalisation and mass delivery of narrative manipulation, coordinated cognitive disruption, and the systematic erosion of trust and epistemic coherence in target populations.

The third domain involves cyber-physical attacks on civilian infrastructure: deliberate disruption of power, water, health, and communications systems that degrade the conditions necessary for civilian survival and dignity. Where sustained and deliberate, such attacks may constitute cruel, inhuman or degrading treatment of civilian populations, or contribute to torturing environments in the sense used here[4].

Across all three domains, questions of attribution, cumulative harm, and distributed responsibility arise with particular intensity.

Digitally mediated coercion and digital authoritarianism: Clarifying the terms

The literature has expanded rapidly, but conceptual inflation has often outpaced conceptual precision(Table 1, 4 and 5). Within a broader framework of digitally mediated coercion, it is useful to distinguish several overlapping but analytically distinct domains. *Digitally mediated coercion* refers to the use of digital and socio-technical systems to monitor, infer, influence, restrict, or shape human behaviour, relationships, and conditions of life. Within this field, *digital authoritarianism* describes the use of such technologies by states or political actors to reinforce domination through surveillance, censorship, and behavioural control (Freedom House, 2018; Polyakova & Meserole, 2019), while *digital repression* refers more specifically to practices aimed at silencing, punishing, or neutralising dissent, including activists, journalists, or minority groups (Amnesty International, 2024; Joint Declaration, 2023; Roberts, 2025). Both often require a situated analysis and case-sensitive research, as different states mix technologies and practices tailored to objectives, context, and culture (Lubbers, 2015).

Related concepts help clarify the wider field. *Surveillance capitalism* captures the economic logic of large-scale data extraction and behavioural prediction for commercial purposes

3 On the specific IHL challenges raised by autonomous weapon systems, including the principles of distinction and proportionality, see ICRC (2026) and Human Rights Watch (2025). On neuroweapons and the emerging sixth domain of warfare — the mind itself as a target — see Krishnan (2018, 2024).

4 Cyber-physical attacks on civilian infrastructure have been documented extensively in the Russia-Ukraine conflict and the Genocide in Gaza. Their relationship to IHL prohibitions on attacks against civilian objects, and to the concept of torturing environments, remains an underdeveloped area of legal analysis.

***Table 4**. Social control: Surveillance, identification, and population management*

Social control	
Methods	**Purposes related to ill-treatment and torture (selected examples)**
1. Data collection, interception, and device access · Communications surveillance and interception · Device access and spyware · Public-space and online monitoring (CCTV, drones, among others) · Big-data extraction and fusion across administrative, commercial, telecom, border, and policing datasets.	– Obtain information on persons, groups, routines, vulnerabilities, and plans. Map networks and connections among them. – Identify and track witnesses, lawyers, journalists, opposition leaders or human rights defenders – Collect compromising material for arrest, interrogation, blackmail, intimidation, or selective retaliation – Produce a chilling effect
2. Identification and tracking · Biometric control (face, iris, fingerprint, voice...). · Digital identity systems. · Location and movement tracking · Identity-linked access to services, platforms, or spaces.	– Discriminatory policies. Restrict access to public services (health and education) and democratic rights (participation and voting). – Enable selective targeting, border exclusion, and watch listing. – Identify and track specific individuals across settings and over time – Transform participation in protest, opposition, migration, or community life into searchable, traceable records – building dossiers.
3. Profiling prediction and social sorting. · Profiling and social graph analysis. · Predictive policing. · AI-based risk scoring, watchlists, triage systems, and automated suspicion markers.	– Sort populations into gradients of suspicion, risk, loyalty, or governability. – Prioritize surveillance, raids, detention, questioning, movement restrictions, or coercive interventions. – Shift from individualised suspicion toward anticipatory governance and preventive repression.
4. Connectivity, service, and digital access control · Internet shutdowns, throttling, communications blocking, and selective service disruption. · Account suspension, platform takedowns, device/network disabling, digital exclusion from essential systems.	– Isolate individuals and communities and break coordination, protest logistics, documentation, or mutual support. – Make testimony, reporting, legal assistance, or emergency communication harder or impossible.

(Zuboff, 2019). It also helps explain how personal data harvested for commercial purposes may later become available for state exploitation (Christl, 2017). *Coercive control* refers to forms of domination, often in intimate or closed settings, in which technologies are used to monitor, isolate, and regulate individuals (Stark, 2007; Woodlock, 2017). Across these domains, some uses of technology may reach the threshold of severe suffering, humiliation, or disintegration of agency compatible with torture or other forms of ill-treatment.

State and private power: Hybrid infrastructures of coercion

Torture, as defined by the UNCAT, is anchored in State action. Today, however, the intertwining of public and private actors in systems of coercion demands a wider perspective to avoid overly narrow readings. A recent policy report commissioned by the European Parliament shows that AI-enabled repression is increasingly operating through the fusion of State and private industry surveillance infrastructures, databases and analytic systems (Ünver, 2024)

It also warns against tying algorithmic authoritarianism too rigidly to regime type: systems become authoritarian when used for authoritarian purposes, not only when deployed by formally authoritarian states (Ünver, 2024; Glowacka et al., 2021).

From this perspective, structural bias, discrimination, and repression should not be treated as separate phenomena, but as connected pathways through which automated systems

***Table 5**. Social influence: Shaping perception, behaviour, and social relations*

Social influence	
Methods	**Purposes related to ill-treatment and torture (selected examples)**
1. Disinformation and narrative control · Disinformation campaigns, propaganda, rumour seeding, content flooding, and coordinated inauthentic behaviour (bot/troll networks, fake amplification). · Selective amplification, suppression, or distortion of content, including search/result manipulation and algorithmic visibility management in social networks and media. · Messaging campaigns directed at individuals, groups, or communities to reshape how events, actors, threats, or legitimacy are understood.	– Distort shared reality, confuse verification, and drown out evidence or testimony of witnesses, victims or survivors. – Delegitimise survivors, activists, journalists, minorities, or communities and justify coercive responses against them. – Polarise, fragment, demoralise, or redirect collective action. – Normalise exceptional measures related to control and security by manufacturing threat narratives, panic, or moral discredit.
2. Covert persuasion and behavioural influence · Persuasion/covert influence operations, infiltration, tailored messaging to target groups. · Psychological operations aimed to affect perceptions, attitudes, emotions, or choices in a target audience.	– Alter beliefs, emotions, expectations, or decision-making of the population without overt coercive force. – Discourage protest, reporting, testimony, solidarity, or help-seeking; induce resignation, fear, self-doubt, or compliance. – Trigger reactive behaviours in targeted groups (withdrawal, fragmentation, internal suspicion).
3. Harassment, intimidation, and reputational harm · Public exposure of intimate/confidential data ("Doxxing"), harassment, threats, smear campaigns, blacklists, coordinated mobbing. · Personalised deceptive content such as fabricated chats, voice-cloning, or deepfake material used against identifiable persons or groups.	– Humiliate, terrorise, silence, isolate, or punish a person or group in ways that may persist beyond the initial act. – Break social ties, damage credibility, and make participation in public, legal, or political life costly or dangerous. – Compel self-censorship or social abandonment.

become coercive (Głowacka et al., 2021). In addition, infrastructures initially developed for apparently legitimate purposes—such as data collection in health, education, or social services—may later be repurposed for surveillance, exclusion, and control, including by governments that came to power through formally democratic means (Crowther & McGregor, 2022). What appears benign in one context may, in another, become an instrument of persecution through automated, personalised, scalable, and opaque forms of social control.

This also shows that surveillance, censorship, and propaganda are no longer separate domains. AI can connect them in a single chain: data extraction enables profiling; profiling enables targeted persuasion and disinformation; these shape behaviour and narrow dissent; and the resulting behaviour feeds back into further surveillance and scoring in a self-reinforcing loop. Digital authoritarianism should therefore be analysed not simply as a toolkit of coercive methods, but as an evolving process of governance and domination.

Reconceiving suffering: Harm without pain, injury without contact

One of the most complex issues is the consideration of the traditional focus on the criterion of severe suffering in the UNCAT definition of torture and the need to introduce elements as an aid to interpretation in technologically mediated coercion. The central claim is that the threshold question is often not about immediate pain but about harm: the cumulative reorganisation of agency, identity, relational life, and conditions of existence.

Several substantive aspects are worth bearing in mind here:

***Table 6**. Severe suffering - sources and impacts*

Domain	Impact
Exposure and loss of protection (i.e. surveillance, spyware, public disclosure of information, biometric monitoring, brain-monitoring, among others)	– The person may feel seen through, exposed, penetrated, or unable to withdraw – Fear, helplessness, hypervigilance and obsessive/paranoid reactions – Shame – Loss of agency and self-esteem
Radical Uncertainty – Loss of control over one's life (i.e. not knowing whether one is watched, classified, what information is known, and what this can trigger).	– Uncertainty – anguish – chronic stress – Anticipatory fear – Passivity, self-censorship, blockade, chilling effect – Helplessness - chronic depression
Public debasement – Reputational damage (i.e., permanent digital record or smear campaigns, intimate data, reputational attacks or others)	– Humiliation, shame, self-destructive reactions and suicide attempts – Loss of opportunities (studies, work...) – isolation (family, friends...)Undermining trust – relational breakdown – Suspicion and doubts - Isolation from legal aid, solidarity networks – loneliness, anguish.
Environmental targeting – Isolation (i.e. degrading life conditions, economic autonomy, political independence)	– Loss of agency, exhaustion, self-betrayal – Forced compliance – Guilt – Entrapment, loss of self-direction and other non-clinical/existential harms
Induced vulnerability (i.e. AI-assisted scoring, targeting, tracking)	– Increased exposure to arrest – vulnerability – Cumulative stress and suffering – permanent changes in identity – complex PTSD

***Table 7**. Severe suffering criteria - digital harms*

Domain	Examples of digital harms
1. Bodily and security harms	Digital targeting that facilitates arrest or assault; location tracking for detention or attack; swatting; induced third-party violence.
2. Psychological and epistemic harms	Threats, harassment, stalking, coordinated intimidation, forced exposure to traumatic material, persistent uncertainty, manufactured doubt, erosion of credibility, and manipulation of evidentiary records.
3. Relational and reputational harms	Doxxing, exposure of intimate data, smear campaigns, deepfakes, shaming before peers or employers, rupture of trust networks, family intimidation, and social isolation.
4. Economic and access harms	Platform exclusion, loss of employment through exposure or blacklisting, blocking access to services, banking, welfare, or mobility systems.
5. Civic and political harms	Chilling effect on speech, silencing dissent, witness intimidation, retaliation against activists, journalists, or survivors.
6. Collective and community harms	Targeting of communities, polarisation, moral panic, group stigmatisation, destruction of collective memory or public legitimacy.

- If torture is understood in terms of coercion and breaking the will, then pain-free but coercive neurotechnological practices may fall within torture or CIDT analysis.
- The traditional notions of "victim" and "injury" may be inadequate when AI-driven systems affect wide populations, produce diffuse harms, or make injuries difficult to individualise and prove. This is when the idea of torturing environments and collective harms becomes essential, as does the need to provide a legal framework for them. There is a need to keep a survivor-centred approach (Table 6) and a collective dimension on digital harms (Table 7)
- In technologically mediated coercion, suffering is often produced less through a single identifiable act with a relevant posttraumatic clinical impact, than through cumulative and interacting mechanisms that reshape the conditions of life, perception, and agency and produce damage to identity (Manek, Galán-Santamarina, Pérez-Sales, 2022).
- New technologies may contribute to torture or CIDT even where pain is not immediate, visible, or physically inflicted.

Read in relation to the previous sections, Tables 6 and 7 specify the kinds of suffering through which the previously described domains and mechanisms may become legally and clinically significant.

Legal frameworks and the challenge of technologically mediated harm.

Who is responsible? Distributed harm and the limits of attribution

In many technologically mediated cases, the hardest question will not be severity but attribution. *Who* tortured: the operator, the ministry, the vendor, the data broker, the systems integrator, or the platform that scaled the harm? There is a distributed responsibility (Ruggie, 2011).

Furthermore, the invisibility and deniability of the method are not incidental but constitutive of its coercive power and challenge accountability. Contemporary repression is increasingly difficult to detect and attribute because it includes government hacking, malware, DDoS attacks, spyware, disruption of secure communications, troll armies, automated censorship, and more covert platform-based manipulation (Council of Europe, 2022). Systems frequently operate as black boxes, with discriminatory effects and weak avenues for appeal or redress. This matters not just for legal accountability, but also for lived

Table 8. Scope of AI Act prohibitions

Topic and Status	Key authority
Manipulative / deceptive AI Prohibited	Art. 5(1)(a): bans subliminal, manipulative or deceptive techniques causing or likely to cause significant harm.
Exploitation of vulnerability Prohibited	Art. 5(1)(b): bans exploitation of vulnerabilities due to age, disability, or specific social/economic situation causing or likely to cause significant harm.
Predictive policing Partly prohibited	Art. 5(1)(d): bans person-based criminal-risk prediction based solely on profiling or personality traits; Commission guidelines say place-based/geospatial systems generally fall outside the ban.
Emotion recognition Partly prohibited / otherwise high-risk	Art. 5(1)(f): bans workplace and educational uses except medical/safety; Commission guidelines say other domains are high-risk.
Biometric categorisation Partly prohibited / otherwise high-risk	Art. 5(1)(g): bans inference of specified sensitive attributes; Annex III classifies some permitted uses as high-risk.
General rights logic Unacceptable-risk or high-risk framing	Commission guidelines: Article 5 covers unacceptable risks to fundamental rights and Union values; prohibited practices are "particularly harmful and abusive."

Sources: European Union, 2024; European Commission 2025, 2026a, 2026b)

experience: opacity is part of domination (ICRC, 2019; Special Rapporteur on Counter-terrorism, 2025).

Existing literature has largely neglected the perspectives of those most directly affected by digital authoritarian practices and has also shown relative neglect of gender and other axes of inequality (Roberts, 2025).

Are current legal frameworks fit for technologically mediated torture and ill-treatment?

Even where particular technologies do not meet the torture threshold, emerging regulation is useful because it identifies rights-sensitive practices that the law already treats as especially dangerous. In broad terms, the legal debate on AI, neurotechnology, and technologically mediated coercion can be organised around three main positions.

The first position holds that the existing human rights framework is basically sufficient, provided that it is clarified, coordinated, and more precisely interpreted. This is reflected in Ligthart's work on coercive brain-reading and European human rights law (Ligthart, 2022, 2024, Ligthart et al., 2021, 2022; Shiner, 2025). On this view, there is no need to create entirely new rights. The key task is to specify how existing protections—especially the prohibition of ill-treatment, privacy, freedom of thought, freedom of expression, and the privilege against self-incrimination—apply when technologies can infer, classify, or potentially alter mental states. What matters legally is not simply the device, but the interaction between the type of information obtained, the degree of cooperation required by the person, and the form of coercion used. From this perspective, many coercive applications would already be restricted or prohibited under existing law, even if doctrine remains fragmented and underdeveloped.

The second position accepts the relevance of existing rights but argues that new domains or rights must be articulated more explicitly, especially in relation to the mind. This is prominent in debates on neuro-rights, cognitive liberty, mental privacy, and freedom of thought (Tesink, et al., 2024). The argument is that traditional legal categories were not designed for technologies capable of directly accessing or influencing neural processes. Proposals converge around core protections: individuals should not be forced to reveal their thoughts, should not be punished for them, and should not have their thoughts impermissibly altered. This perspective is reflected in work by Yuste et al. (2017), Ienca and Andorno (2017), Ienca (2021), Bublitz (2013), and McCarthy-Jones (2019), among others.

The third position is more structural. It argues that the problem lies not only in doctrinal gaps, but in a deeper misalignment between the human rights framework and the nature of technologically mediated harm. As Teo (2022) argues, human rights law has traditionally been organised around discrete, observable, temporally proximate, and causally attributable violations. By contrast, many harms associated with AI and socio-technical systems are systemic, cumulative, latent, and distributed across complex networks. This has important implications for torture and CIDT, as coercion may operate not as a bounded act, but as a continuous environment of surveillance, classification, exclusion, and behavioural conditioning. This aligns with the concept of torturing environments (Pérez-Sales, 2017, 2026), where suffering is produced cumulatively and structurally.

Taken together, these three positions move from legal specification to rights innovation to conceptual reconstruction. The first explores how far existing rights can be stretched; the second argues for more explicit protection of the mind; and the third suggests that emerging technologies may require a deeper rethinking of how law conceptualises harm, responsibility, and torture itself. All three have advantages and disadvantages and are mutually complementary in doctrinal development within existing legal categories.

Legal regulations

The AI Regulation Act (EU) 2024/1689 (the "EU AI Act") is the most advanced regulatory system to date adopted to lay down harmonised rules on artificial intelligence across the European Union[5] (Table 8). Under Article 288 TFEU, EU regulations have general application, are binding in their entirety, and are directly applicable in all Member States. There is no UN-level binding equivalent to the EU AI Act. The nearest global normative instruments are soft law: UNESCO's (2024) *Recommendation on the Ethics of Artificial Intelligence* and UN General Assembly AI resolutions[6] which are non-binding. The Council

5 Regulation (EU) 2024/1689 (the EU AI Act) is structured in chapters on general provisions and scope (Chapter I), prohibited AI practices (Chapter II, including Article 5), high-risk AI systems and related obligations, conformity assessment, registration and post-market controls (Chapter III), transparency obligations for certain AI systems (Chapter IV), general-purpose AI models (Chapter V), innovation measures such as regulatory sandboxes (Chapter VI), governance and enforcement (Chapter VII), the EU database and market monitoring/surveillance (Chapters VIII–IX), codes of conduct and guidance (Chapter X), delegated powers and committee procedure (Chapter XI), penalties (Chapter XII), and final provisions, amendments, and application dates (Chapter XIII).

6 UNGA Res 78/265, *Seizing the Opportunities of Safe, Secure and Trustworthy Artificial Intelligence Systems for Sustainable Development* (21 March 2024) UN Doc A/RES/78/265; UNGA Res 78/311, *Enhancing International Cooperation on Capacity-Building of Artificial Intelligence* (1 July 2024) UN Doc A/RES/78/311; UNGA Res 79/1, *The Pact for the Future*

of Europe *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law* was opened for signature in 2024 and is the first legally binding international treaty in this field.

Although these instruments are not framed in torture law terms, they are relevant because they identify certain manipulative, exploitative, and rights-infringing technological practices as categorically unacceptable or presumptively dangerous.

Conclusions and implications

Dystopian accounts often imagine control as something direct: a technology entering the mind and dictating thought (Krishnan, 2016, 2018). Yet new technologies do not need to penetrate the mind in any literal sense to become relevant to torture or ill-treatment. Through surveillance, exposure, profiling, threats, and the manipulation of visibility, they may help create coercive environments that produce fear, self-censorship, dependency, and loss of agency (Pérez-Sales & Serra, 2020; Głowacka et al., 2021).

This editorial has proposed a framework for understanding technologically mediated coercion that is organised around three analytical layers — the human need under attack, the method through which harm is produced, and the site in which it occurs — and that groups emerging practices into three intersecting domains: mental-directed interventions, social control, and social influence. It has also reflected on the severe suffering criteria. Together, these are intended not as a closed taxonomy but as a working map, one that can be revised as technologies evolve and as empirical research on their effects accumulates.

Several conclusions follow from this analysis, with implications for research, clinical practice, legal interpretation, and accountability.

On the concept of torture and CIDT: The framework supports an understanding of torture and cruel, inhuman or degrading treatment that extends beyond discrete, physically inflicted acts. Technologically mediated coercion is frequently cumulative, environmental, and structurally diffuse. Suffering may be produced without pain, without a single identifiable perpetrator, and without a moment of direct physical contact. This aligns with the concept of torturing environments (Pérez-Sales, 2017, 2026): conditions in which the architecture of control — surveillance, exposure, radical uncertainty, and induced vulnerability — generates severe and persistent harm to identity, agency, and relational life. Legal interpretation of the UNCAT and related instruments should be capable of accommodating this reality, and current debates in human rights law suggest this is both necessary and achievable within existing frameworks, though it requires doctrinal development.

On the question of suffering: The traditional criterion of severe suffering remains relevant but requires reorientation. In the domains examined here, severity is not always visible, immediate, or medically certifiable. It may instead manifest as chronic existential harm: the permanent reorganisation of selfhood around fear, exposure, and self-censorship; the collapse of trust networks; enforced compliance that the person experiences as self-betrayal; and cumulative identity damage that does not resolve once the coercive situation ends. Clinicians and legal practitioners working with survivors need assessment frameworks that can capture these diffuse, non-acute forms of injury. A survivor-centred approach is indispensable, both because the experience of harm is shaped by context and identity, and because affected communities — particularly those subject to collective targeting — may be better positioned than external observers to identify when digital practices cross into ill-treatment.

On distributed responsibility and accountability: In many technologically mediated cases, the hardest question is not severity but attribution. The fusion of state and private surveillance infrastructures, the opacity of automated systems, and the multiplicity of actors involved — vendors, operators, data brokers, platform providers, integrators, and, of course, the state — mean that responsibility is rarely concentrated in a single agent. This fragmentation is not incidental to these systems; in important respects, it is constitutive of their coercive logic, since it renders domination harder to contest and accountability harder to enforce. Existing human rights law remains only partially equipped to address harms that are systemic, anticipatory, and distributed across complex socio-technical networks. Addressing this requires both legal innovation — including clearer extraterritorial jurisdiction, due diligence obligations for technology companies, and remedies for algorithmic harms — and methodological innovation in documentation, evidence-gathering, and impact assessment.

On the limits of AI as both problem and solution: This analysis has stressed the coercive potential of AI and related technologies, but it is equally important to note that these tools may serve anti-torture and accountability work. Machine-learning applications for processing large bodies of testimony, satellite imagery analysis for detecting detention infrastructure, automated translation for multilingual documentation, and early-warning systems for violence and dis-

(22 September 2024) UN Doc A/RES/79/1, annex II ('Global Digital Compact'); UNGA Res 79/239, *Artificial Intelligence in the Military Domain and Its Implications for International Peace and Security* (24 December 2024) UN Doc A/RES/79/239 updated 1 December 2025 (UN Doc A/RES/80/58)

placement all represent genuine contributions to human rights monitoring. The challenge is not to oppose these technologies as such, but to insist on the conditions under which they are deployed: transparency, human oversight, meaningful consent, redress mechanisms, and a commitment to keeping human judgment at the centre of processes that affect fundamental rights.

On the need for a power analysis alongside rights analysis: A rights-based analysis identifies what was violated and who was harmed. A power analysis asks why these technologies are deployed, whose interests they serve, and what kind

***Table 9**. Technologically mediated coercion and the torture threshold: criteria for graduated legal assessment*

Criteria	Guiding questions	Comments
Core threshold criteria		
Severity and persistence.	Does the practice produce severe suffering or serious harm, whether acute or cumulative? Does it durably affect identity, agency, relational life, or basic conditions of existence?	– Single incidents may be severe; sustained targeting may be cumulative. Lack of immediate pain or suffering presentation does not exclude severity. – Severity should be assessed in a contextual-situated way, including identity, culture and context.
Intentionality and purpose.	Is the suffering incidental, tolerated, or deliberately produced? Is the practice directed toward coercion, punishment, extraction, degradation, discrimination, or a broader strategy of domination? Is it part of an overall strategy? Are there patterns of actions known from other cases that can be traced?	Consider design choices, implementing decisions, systematicity, repeated patterns of action and tolerance of known effects.
Contextual/aggravating factors		
Structural position and power asymmetry	Is the person or group in a situation of particular vulnerability or marked subordination — detained, displaced, monitored, excluded, or targeted?	Power asymmetry intensifies the coercive effect of technologies that might be less harmful in other contexts.
Opacity and deniability	Is the harm made difficult to perceive, document, contest, or attribute through technical complexity, secrecy, or fragmented implementation?	Deliberate deniability and structural invisibility are aggravating features that may, in turn, deepen the coercive effect.
Permanence and redress	Is the harm reversible, or does it leave enduring traces with little realistic possibility of challenge or repair?	Consider the permanence of digital footprints, the opacity of automated systems, and the fragmentation of accountability as aggravating factors.
Collective or individual character.	Does the practice affect only a specific individual, or does it affect a community, group, or population?	Collective and community-level harm — surveillance chilling effect, algorithmic exclusion, narrative destruction — may reach the threshold of CIDT or produce torturing environments even where individual suffering is difficult to isolate or prove.

of political and social order they help to sustain. The two are complementary, not alternative, and both are necessary for understanding technologically mediated coercion (Roberts & Oosterom, 2025; Zuboff, 2019). Structural bias, discrimination, and repression should not be treated as separate phenomena, but as connected pathways through which automated systems become coercive — and through which the effects of torture and ill-treatment are reproduced and amplified at scale.

No fixed threshold separates permissible technologically mediated coercion from CIDT, or CIDT from torture — assessment is inherently contextual and interpretative. Table 9 offers dimensions to guide that judgment, to be considered cumulatively rather than individually. These criteria do not produce automatic determinations; they are intended as a structured basis for contextual, survivor-centred analysis.

Table 9 organises these dimensions into a structured but non-exhaustive framework. No single criterion is decisive; they are intended to be read together, weighed against the specific context, and applied with attention to the cumulative character of technologically mediated harm.

Recommendations: Drawing together the analysis, the following directions merit priority attention:

- **For legal and normative frameworks:** Existing prohibitions on torture and CIDT should be interpreted to encompass cumulative, environmentally produced, and technologically mediated harm; further guidance from treaty bodies and special procedures is needed. The EU AI Act's prohibitions on manipulative, deceptive, and exploitative AI practices, while not framed in torture law terms, represent an important complementary instrument and should be read alongside human rights standards. The Council of Europe Framework Convention on AI offers a further entry point for binding international obligations.
- **For research and documentation**: There is an urgent need for survivor-centred, contextually grounded empirical research on the psychological and social effects of the practices identified in this framework. The existing literature is predominantly normative and legal; clinical and qualitative evidence on harm is sparse. Gender, race, and other axes of inequality remain insufficiently theorised in the literature on digital authoritarianism.

  Research in this area faces a specific methodological challenge. The harms described are difficult to distinguish from other sources of psychological distress, and the coercive mechanisms are often invisible, deniable, or technically complex. Individual accounts of harm are a necessary starting point but are not sufficient evidence on their own: without independent corroboration of the coercive mechanism and without comparative data, it is impossible to distinguish documented institutional coercion from other forms of digital conflict or from persecutory ideation. The field, therefore, needs studies that establish patterns across samples rather than resting on single cases — including case-control designs, systematic cross-case analysis of populations known to have been subjected to documented forms of institutional surveillance or repression, and longitudinal follow-up to document the persistence and trajectory of harm. First-person accounts are most valuable when embedded within such designs. Developing methodological standards adequate to this challenge is itself a research priority.
- **For rehabilitation and clinical practice**: Practitioners working with survivors of politically motivated digital targeting require adapted assessment tools that capture the harms described in this framework — including existential and identity-level damage, relational breakdown, and the chronic effects of radical uncertainty. Standard trauma frameworks may need supplementing.
- **For accountability and prevention**: Monitoring bodies, national preventive mechanisms, and civil society organisations need specific capacity to assess technologically mediated environments — in detention settings, at borders, in community contexts, and online. Documentation standards should be updated to reflect the forms of evidence — digital, algorithmic, and inferential — relevant in these cases.

The present framework is offered as a starting point for this work, not as a finished map. As technologies continue to evolve, and as their effects become more empirically documented, both the conceptual categories and the legal standards will require revision. What is clear already is that torture and ill-treatment do not end where screens begin.

References

Amnesty International. (2024). Recording dissent: Camera surveillance at peaceful protests in the Netherlands. https://www.amnesty.org/en/documents/eur35/8469/2024/en/

Avis, N., Marciniak, L., & Sapignoli, M. (Eds.). (2024). States of surveillance: Ethnographies of new technologies in policing and justice. Routledge. https://doi.org/10.4324/9781003413547

Bachelet, M. (2018, November 14). Human rights in a new era [Speech]. Office of the United Nations High Commissioner for Human Rights. https://www.ohchr.org/en/statements-and-speeches/2018/11/human-rights-new-era

Beduschi, A. (2021). International migration management in the age of artificial intelligence. Migration Studies, 9(3), 576–596. https://doi.org/10.1093/migration/mnaa003

Bublitz, J.-C. (2013). My mind is mine!? Cognitive liberty as a legal concept. In E. Hildt & A. Franke (Eds.), Cognitive enhancement (pp. 233–264). Springer. https://doi.org/10.1007/978-94-007-6253-4_19

Christl, W. (2017). Corporate surveillance in everyday life. Cracked Labs

Council of Europe. (2022). Pegasus spyware and its implications on human rights. https://rm.coe.int/pegasus-spyware-report-en/1680a6f5d8

Council of Europe. (2024). Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (CETS No. 225). https://rm.coe.int/1680afae3c

Crowther, N., & McGregor, L. (2022). A digital cage is still a cage: Human rights and new and emerging technologies in social care. University of Essex, Human Rights, Big Data and Technology Project. https://repository.essex.ac.uk/33020/

Dore-Horgan, E., Ligthart, S., Meynen, G., & Kellmeyer, P. (Eds.). (2026). Cambridge handbook on human rights for the mind: Emerging technologies, law and philosophy. Cambridge University Press.

Dulka, A. (2023). The use of artificial intelligence in international human rights law. Stanford Technology Law Review, 26, 316–366. https://law.stanford.edu/wp-content/uploads/2023/08/Publish_26-STLR-316-2023_The-Use-of-Artificial-Intelligence-in-International-Human-Rights-Law8655.pdf

Elbatanouny, H., Elbatanouny, M., Sallam, M., Rida, I., & Evans, A. (2025). A comprehensive analysis of deception detection methods using machine learning. Expert Systems with Applications, 263, Article 125702. https://doi.org/10.1016/j.eswa.2024.125702

European Commission. (2025). Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act). AI Act Service Desk. https://ai-act-service-desk.ec.europa.eu/sites/default/files/2025-08/guidelines_on_prohibited_artificial_intelligence_practices_established_by_regulation_eu_20241689_ai_act_english_ied3r5nwo50xggpcfmwckm3nuc_112367-1.PDF

European Commission. (2026a). Frequently asked questions. AI Act Service Desk. https://ai-act-service-desk.ec.europa.eu/en/faq

European Commission. (2026b). Navigating the AI Act. Shaping Europe's digital future. https://digital-strategy.ec.europa.eu/en/faqs/navigating-ai-act

European Parliament. (2024, March 13). Harmonised rules on use of artificial intelligence in the European Union (EU AI Act) [Legislative resolution]. https://www.europarl.europa.eu/doceo/document/TC1-COD-2021-0106_EN.pdf

European Union Agency for Fundamental Rights. (2023). Surveillance by intelligence services: Fundamental rights safeguards and remedies in the EU (2023 update of the 2017 report). https://fra.europa.eu/en/publication/2023/surveillance-update#country-related

European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). EUR-Lex. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

Freedom House. (2018). Freedom on the Net 2018: The rise of digital authoritarianism. https://freedomhouse.org/report/freedom-net/2018/rise-digital-authoritarianism

Głowacka, D., Youngs, R., Pintea, A., & Wołosik, E. (2021). Digital technologies as a means of repression and social control. European Parliament. https://www.europarl.europa.eu/RegData/etudes/STUD/2021/653636/EXPO_STU%282021%29653636_EN.pdf

High-Value Detainee Interrogation Group. (2016). Interrogation: A review of the science. Federal Bureau of Investigation. https://www.fbi.gov/file-repository/hig-report-interrogation-a-review-of-the-science-september-2016.pdf/view

Human Rights Watch & Harvard Law School International Human Rights Clinic. (2025). A hazard to human rights: Autonomous weapons systems and digital decision-making. *Human Rights Watch*. https://www.hrw.org/report/2025/04/28/a-hazard-to-human-rights/autonomous-weapons-systems-and-digital-decision-making

Ienca, M., & Andorno, R. (2017). Towards new human rights in the age of neuroscience and neurotechnology. *Life Sciences, Society and Policy, 13*, Article 5. https://doi.org/10.1186/s40504-017-0050-1

Ienca, M. (2021). The Human Right to Cognitive Liberty. Cambridge University Press

International Committee of the Red Cross. (2019). Artificial intelligence and machine learning in armed conflict: A human-centred approach. https://www.icrc.org/sites/default/files/document_new/file_list/ai_and_machine_learning_in_armed_conflict-icrc.pdf

International Committee of the Red Cross. (2024). International humanitarian law and the challenges of contemporary armed conflicts. ICRC. https://shop.icrc.org/international-humanitarian-law-and-the-challenges-of-contemporary-armed-conflicts

International Committee of the Red Cross. (2026). Autonomous weapon systems and international humanitarian law: Selected issues (Position paper). https://www.icrc.org/sites/default/files/2026-03/4896_002_Autonomous_Weapons_Systems_-_IHL-ICRC.pdf

International Network of Civil Liberties Organizations, Physicians for Human Rights, & Omega Research Foundation. (2023). Lethal in disguise 2: How crowd-control weapons impact health and human rights. PHR. https://phr.org/our-work/resources/lethal-in-disguise-2/

International Organization for Migration. (2021). World migration report 2022: Chapter on artificial intelligence, migration and mobility. https://publications.iom.int/system/files/pdf/wmr_2022_book_eng_1.pdf

Joint declaration on freedom of peaceful assembly and of association and misuse of digital technologies. (2023, September 15). https://www.ohchr.org/sites/default/files/documents/issues/trafficking/statements/20230915-jd-foaa-digital-technologies.pdf

Kalodanis, K. (2025). High-risk AI systems — lie detection application. Future Internet, 17(2), Article 54. https://doi.org/10.3390/fi17020054

Krishnan, A. (2016). Attack on the brain: Neurowars and neurowarfare. Space and Defense, 9(1), 4–18. https://doi.org/10.32873/uno.dc.sd.09.01.1110

Krishnan, A. (2018). Military neuroscience and the coming age of neurowarfare. Routledge. https://doi.org/10.4324/9781315595429

Ligthart, S. (2022). Coercive brain-reading in criminal justice: An

analysis of European human rights law. Cambridge University Press.

Ligthart, S. (2024). Towards a human right to psychological continuity? Reflections on the rights to personal identity, self-determination, and personal integrity. European Convention on Human Rights Law Review, 5(2), 199–229. https://doi.org/10.1163/26663236-bja10092

Ligthart, S., Bublitz, C., Douglas, T., Forsberg, L., & Meynen, G. (2022). Rethinking the right to freedom of thought: A multidisciplinary analysis. Human Rights Law Review, 22(4), Article ngac028. https://doi.org/10.1093/hrlr/ngac028

Ligthart, S., Douglas, T., Bublitz, C., Kooijmans, T., & Meynen, G. (2021). Forensic brain-reading and mental privacy in European human rights law: Foundations and challenges. Neuroethics, 14(2), 191–203. https://doi.org/10.1007/s12152-020-09438-4

Lubbers, E. (2015). Undercover research: Corporate and police spying on activists. An introduction to activist intelligence as a new field of study. Surveillance & Society, 13(3–4), 338–353. https://doi.org/10.24908/ss.v13i3/4.5371

Malek, S., Hearn, D., Fahy, T., Tully, J., & Exworthy, T. (2023). Legal and human rights issues in the use of electronic monitoring (using GPS "tracking" technology) in forensic mental health settings in the UK. Medicine, Science and the Law, 63(4), 309–315. https://doi.org/10.1177/00258024231174820

Manek, J., Galán-Santamarina, A., & Pérez-Sales, P. (2022). Torturing environments and multiple injuries in Mexican migration detention. Humanities and Social Sciences Communications, 9(1). https://doi.org/10.1057/s41599-022-01252-y

Marin, M., Kalaitzis, F., & Price, B. (2020, July 6). Using artificial intelligence to scale up human rights research: A case study on Darfur. Citizen Evidence Lab. https://citizenevidence.org/2020/07/06/using-artificial-intelligence-to-scale-up-human-rights-research-a-case-study-on-darfur/

Melgaço, L., & Monaghan, J. (Eds.). (2021). Protests in the information age: Social movements, digital practices and surveillance. Routledge. https://doi.org/10.4324/9780429467639

McCarthy-Jones, S. (2019). The autonomous mind: The right to freedom of thought in the twenty-first century. Frontiers in Artificial Intelligence, 2, Article 19. https://doi.org/10.3389/frai.2019.00019

McEvoy, M., Corney, N., & Haar, R. J. (2024). State violence against protesters: Perspectives and trends in use of less lethal weapons. Torture Journal, 34(1). https://doi.org/10.7146/torture.v34i1.144275

McKay, C. (2021). The carceral automaton: Digital prisons and technologies of detention. International Journal for Crime, Justice and Social Democracy, 10(4), 100–119. https://doi.org/10.5204/IJCJSD.2137

Noriega, M. (2020). The application of artificial intelligence in police interrogations. Futures, 117, Article 102510. https://doi.org/10.1016/j.futures.2019.102510

Office of the United Nations High Commissioner for Human Rights, & University of Essex. (2023). Digital border governance: A human rights-based approach. https://www.ohchr.org/sites/default/files/2023-09/Digital-Border-Governance-A-Human-Rights-Based-Approach.pdf

Office of the United Nations High Commissioner for Human Rights. (2022). The right to privacy in the digital age (A/HRC/51/17). https://www.ohchr.org/en/documents/thematic-reports/ahrc5117-right-privacy-digital-age

Office of the United Nations High Commissioner for Human Rights. (2024a). Human rights compliant uses of digital technologies by law enforcement in the context of peaceful protests. https://www.ohchr.org/sites/default/files/2024-03/Toolkit-law-enforcement-Component-on-Digital-Technologies.pdf

Office of the United Nations High Commissioner for Human Rights. (2024b). Trends in digital rights across the Middle East and North Africa (MENA) region: Expansion of digital surveillance and impacts on journalists and human rights defenders. https://www.ohchr.org/sites/default/files/documents/issues/civicspace/resources/civic-space-tech-brief-surveillance-trends-middle-east-north-africa-1-en.pdf

Office of the United Nations High Commissioner for Human Rights. (2025). Call for input for the HRC62 thematic report on the impact of digital and AI-assisted surveillance technologies on human rights, especially the rights to freedom of peaceful assembly and of association. https://www.ohchr.org/en/calls-for-input/2025/call-input-hrc62-thematic-report-impact-digital-and-ai-assisted-surveillance

Office of the United Nations High Commissioner for Human Rights. (2026). Tech brief: Data privacy, discrimination & AI. https://www.ohchr.org/sites/default/files/documents/issues/civicspace/resources/brief-data-privacy-ai-report-rev.pdf

Omega Research Foundation & Amnesty International. (2023). My eye exploded: The global abuse of kinetic impact projectiles. Amnesty International. https://www.amnesty.org/en/documents/act30/6384/2023/en/

Pérez-Sales, P. (2017). Psychological torture: Definition, evaluation and measurement. Routledge. https://doi.org/10.4324/9781315616940

Pérez-Sales, P. (2026). Torturing environments: Psychological, clinical and legal dimensions. Routledge. https://doi.org/10.4324/9781003639350

Pérez-Sales, P., & Serra, L. (2020). Internet and communications as elements for CIDT and torture: Initial reflections in an unexplored field. Torture Journal, 30(1), 5–22. https://doi.org/10.7146/torture.v30i1.120593

Polyakova, A., & Meserole, C. (2019). Exporting digital authoritarianism: The Russian and Chinese models. Brookings Institution. https://www.brookings.edu/wp-content/uploads/2019/08/FP_20190827_digital_authoritarianism_polyakova_meserole.pdf

Roberts, T., & Oosterom, M. (2025). Digital authoritarianism: A systematic literature review. Information Technology for Development, 31(4), 860–884. https://doi.org/10.1080/02681102.2024.2425352

Ruggie, J. G. (2011). Guiding principles on business and human rights: Implementing the United Nations "protect, respect and remedy" framework (A/HRC/17/31). United Nations Human Rights Council. https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf

Shiner, B. (2025). The right to freedom of thought and the prohibition of torture, or cruel, inhuman or degrading treatment

or punishment: Examining the relationship in the case of the coercive and interrogational use of neurotechnology. In E. Dore-Horgan, S. Ligthart, G. Meynen, & P. Kellmeyer (Eds.), Cambridge handbook on human rights for the mind: Emerging technologies, law and philosophy (pp. 1–26). Cambridge University Press.

Special Rapporteur on the Promotion and Protection of human rights and fundamental freedoms while countering terrorism. (2025). Protecting human rights while using artificial intelligence in counter-terrorism and security settings [Position paper]. https://www.ohchr.org/sites/default/files/documents/issues/terrorism/sr/un-sr-ct-ai-position-paper-dec-2025.pdf

Special Rapporteur on the Rights to Freedom of Peaceful Assembly and of Association. (2024). Model protocol for law enforcement officials to promote and protect human rights in the context of peaceful protests (A/HRC/55/60). United Nations. https://www.ohchr.org/en/documents/thematic-reports/ahrc5560-model-protocol

Stark, E. (2007). Coercive control: How men entrap women in personal life. Oxford University Press.

Teo, S. A. (2022). How artificial intelligence systems challenge the conceptual foundations of the human rights legal framework. Nordic Journal of Human Rights, 40(1), 216–234. https://doi.org/10.1080/18918131.2022.2073078

Tesink, V., Douglas, T., Forsberg, L., Ligthart, S., & Meynen, G. (2024). Right to mental integrity and neurotechnologies: Implications of the extended mind thesis. Journal of Medical Ethics, 50(10), 656–663. https://doi.org/10.1136/JME-2023-109645

UNESCO, & Office of the United Nations High Commissioner for Human Rights. (2026). Protecting critical voices: Guidance for human rights impact assessment on digital platforms. https://www.ohchr.org/sites/default/files/documents/issues/civicspace/protecting-critical-voices-en.pdf

UNESCO. (2024). Recommendation on the ethics of artificial intelligence (updated version). https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence

United Nations Human Rights Council Advisory Committee. (2024). Impact, opportunities and challenges of neurotechnology with regard to the promotion and protection of all human rights (A/HRC/57/61). https://www.ohchr.org/sites/default/files/documents/hrbodies/hrcouncil/sessions-regular/session57/A-HRC-57-61-Etext-accessible.pdf

Ünver, A. (2024). Artificial intelligence (AI) and human rights: Using AI as a weapon of repression and its impact on human rights. European Parliament, Subcommittee on Human Rights. https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA%282024%29754450_EN.pdf

Yuste, R., Goering, S., Agüera y Arcas, B., et al. (2017). Four ethical priorities for neurotechnologies and AI. Nature, 551, 159–163. https://doi.org/10.1038/551159a

Woodlock, D. (2017). The abuse of technology in domestic violence and stalking. Violence Against Women, 23(5), 584–602.

Zuboff, S. (2019). The age of surveillance capitalism. PublicAffairs. https://www.publicaffairsbooks.com/titles/shoshana-zuboff/the-age-of-surveillance-capitalism/9781610395694/