FAMU CYBER POLICY BRIEFS

Florida A&M University Cyber Policy Institute

FAMU Cyber Policy Brief No. 1 (2026)

False Memory Formation, Epistemic Harm, and the Urgent Need for Policy Intervention

False Evidence, Real Injury: Legislative Gaps in Addressing Deepfake Harm

DeAnna M. Burney, Ph.D., Daryl Scriven, Ph.D.; Chiquita Brown, Ph.D. ,
Tejal Mulay, Ph.D.; Phylicia Taylor, Ph.D.
Florida A&M University

ISSN (Online): ISSN: 3070-8001

DOI: https://doi.org/... (TBD)
Florida A&M University Cyber Policy Institute
Access: Open Access

Introduction

Artificial intelligence has entered a phase of development in which the distinction between authentic reality and synthetic fabrication is no longer intuitively discernible. Among the most destabilizing manifestations of this shift is the rapid proliferation of deepfakes, AI-generated or AI-altered audio, video, and imagery that convincingly depict individuals engaging in speech or conduct that never occurred. Advances in generative adversarial networks, diffusion models, and multimodal AI systems have significantly increased the realism, accessibility, and scalability of deepfake technologies, enabling widespread use beyond specialized technical domains (Chesney & Citron, 2019; Westerlund, 2023).

While policy discourse has largely focused on electoral interference, national security threats, and misinformation, comparatively little attention has been paid to the psychological, neurobiological, and physical health consequences of deepfakes. This omission is especially concerning given that these technologies are emerging within the context of a global mental health crisis, characterized by rising rates of depression, anxiety, trauma-related disorders, and suicide (World Health Organization [WHO], 2023).

This paper argues that deepfakes constitute not merely a technological or informational risk, but a mental health, physical health, and epistemic crisis. By exploiting fundamental cognitive and neurobiological systems involved in memory formation, identity coherence, and trust, deepfakes introduce mechanisms of harm that existing legal, clinical, and regulatory frameworks are not equipped to address. Legislative intervention is therefore urgently required to protect public health, preserve epistemic integrity, and ensure ethical governance of artificial intelligence.

The Global Mental Health Crisis as Context

The mental health impacts of deepfakes must be situated within the broader global mental health landscape. According to the WHO (2023), mental health conditions are now among the leading causes of disability worldwide, with depressive and anxiety disorders alone affecting hundreds of millions of people. The COVID-19 pandemic, economic precarity, armed conflict, forced migration, climate-related stressors, and pervasive digital exposure have collectively intensified psychological vulnerability across populations.

Digital environments have become central sites of identity formation, social validation, and memory storage. Social media platforms, digital archives, and audiovisual recordings increasingly function as externalized memory systems that validate personal and collective experience. Consequently, harms that occur online particularly those involving identity violation, humiliation, reputational threat, and reality distortion carry psychological weight comparable to, and in some cases exceeding, offline harms (Marwick & Lewis, 2017).

Deepfakes amplify these vulnerabilities by targeting the very mechanisms through which individuals establish truth, continuity of self, interpersonal trust, and shared reality. For individuals and communities already burdened by historical trauma, marginalization, or systemic gaslighting, the psychological consequences are disproportionately severe.

Deepfakes as Engines of False Memory Formation

In cognitive psychology, false memories are understood as recollections of events that either did not occur or occurred in a distorted form, often accompanied by high levels of subjective confidence and emotional vividness (Loftus, 2005). Extensive research demonstrates that human memory is reconstructive rather than reproductive, meaning that memories are actively rebuilt rather than passively retrieved. As a result, memory is inherently vulnerable to suggestion, emotional salience, and errors in source attribution.

Deepfakes represent an unprecedented accelerator of false memory formation because they deliver highly realistic audiovisual stimuli that activate the same perceptual, emotional, and mnemonic systems involved in genuine autobiographical experience. Unlike verbal misinformation, which requires inferential reasoning, deepfakes present sensory-rich visual and auditory "evidence" that is processed as perceptual input. In contemporary digital environments, media such as photographs, videos, and recordings increasingly function as extensions of human memory, serving as external memory anchors that reinforce personal narratives and sustain epistemic trust the confidence individuals place in evidence and shared reality (Sparrow et al., 2011). Consequently, audiovisual records are routinely treated as authoritative proof across legal, institutional, and interpersonal contexts.

This dynamic substantially increases vulnerability to source monitoring errors, in which individuals misattribute the origin of a memory to personal experience rather than external fabrication (Johnson et al., 1993). By introducing false yet authoritative digital artifacts, deepfakes fundamentally corrupt the evidentiary role of digital media and generate epistemic harm, defined as damage to an individual's capacity to know, verify, and trust reality itself. When fabricated content contradicts internal recollection, individuals experience cognitive dissonance, a psychologically distressing state arising from conflicting beliefs, memories, or sources of evidence (Festinger, 1957).

Empirical research indicates that when fabricated evidence is socially reinforced through institutional validation, peer endorsement, or algorithmic amplification individuals are more likely to revise their own memories than reject the false artifact (Lewandowsky et al., 2020). Over time, this process erodes narrative identity, weakens self-trust, and destabilizes psychological coherence. Recent experimental findings further demonstrate that exposure to manipulated audiovisual content significantly increases belief confidence and memory distortion even when individuals are explicitly warned of possible deception (Nash et al., 2023). These effects are especially pronounced among individuals with trauma histories, elevated stress exposure, or preexisting mental health conditions, for whom cognitive load and emotional arousal further impair effective reality monitoring.

Gaslighting at Scale: Trauma, Identity Harm, and Memory Distortion

From a trauma psychology perspective, deepfakes function as a form of technologically mediated gaslighting. Gaslighting involves repeated invalidation or rewriting of a person's perception of reality, leading individuals to doubt their own memory and judgment (Sweet, 2019). When this process targets memory itself, it constitutes a form of psychological abuse.

Victims of deepfake-related abuse particularly nonconsensual sexual deepfakes, fabricated criminal footage, or reputational attacks frequently exhibit trauma-related symptoms, including intrusive thoughts, hypervigilance, shame, emotional numbing, and persistent fear of re-exposure. Importantly, emotional responses often persist even when individuals cognitively recognize the content as false.

Neuroscientific research demonstrates that the amygdala responds to perceived threat, not factual accuracy, meaning that fabricated stimuli can trigger authentic trauma responses (van der Kolk, 2014; McGaugh, 2004).

Neurobiological Pathways Linking Psychological and Physical Harm

Deepfakes exert both psychological and physical impact by hijacking interconnected neural systems involved in emotion, memory, and stress regulation. Emotionally salient stimuli activate the amygdala, prioritizing information for memory encoding, while the hippocampus consolidates emotionally charged memory traces even when the stimulus is false (McGaugh, 2004).

Heightened emotional arousal impairs prefrontal cortex functioning responsible for reality testing and source monitoring (Johnson et al., 1993). Simultaneously, activation of the hypothalamic–pituitary–adrenal (HPA) axis produces sustained cortisol release, autonomic dysregulation, inflammation, and immune suppression.

The Centers for Disease Control and Prevention (CDC, 2022) and recent meta-analyses confirm that chronic psychological stress is associated with increased risk for hypertension, cardiovascular disease, metabolic disorders, gastrointestinal illness, chronic pain, sleep disturbance, and fatigue. Thus, deepfake-related psychological harm carries predictable and measurable physical health consequences.

Dissociation, Moral Injury, and Racialized Epistemic Harm

Prolonged exposure to reality distortion may contribute to dissociative responses, including depersonalization and derealization. The DSM-5-TR defines dissociation as a disruption in the integration of consciousness, memory, identity, emotion, and perception (American Psychiatric

Association [APA], 2022). Deepfakes exacerbate these disruptions by creating competing versions of the self and weakening autobiographical continuity.

Beyond individual psychopathology, deepfakes may produce moral injury, defined as psychological distress resulting from betrayal by trusted individuals or institutions (Litz et al., 2009). This harm is particularly acute for racialized and marginalized communities with historical experiences of epistemic invalidation. In such contexts, deepfakes reinforce patterns of disbelief, misrepresentation, and institutional neglect, compounding racial trauma and undermining trust in systems of justice and knowledge production.

Scientific Authority and Legislative Rationale

Legislative bodies increasingly rely on established scientific consensus to justify regulatory intervention, particularly where emerging technologies present foreseeable risks to public health. The World Health Organization (WHO, 2023) affirms that mental health is an integral component of overall health and identifies psychological distress and trauma as significant contributors to physical illness, disability, and premature mortality worldwide. Similarly, the Centers for Disease Control and Prevention (CDC, 2022) documents that prolonged psychological stress activates physiological stress-response systems, increasing population-level healthcare burden through heightened risk of cardiovascular disease, metabolic disorders, immune dysfunction, and other stress-related conditions. The American Psychological Association (2023) further establishes that trauma and chronic stress manifest physically through measurable alterations in neural, endocrine, and immune functioning.

From a legislative history perspective, these conclusions are consistent with long-standing public health and safety statutes addressing occupational stress, domestic violence, environmental exposure, and adverse experiences, all of which recognize psychological harm as a legitimate and legally cognizable precursor to physical illness. Deepfake technologies introduce a novel and technologically mediated stressor by uniquely combining identity violation, reputational threat, and systematic distortion of reality factors known to intensify psychological injury and physiological stress responses.

Taken together, the evidence demonstrates that deepfakes operate simultaneously as false memory generators, digital gaslighting mechanisms, trauma-inducing stimuli, dissociative triggers, and sources of moral injury, with predictable downstream consequences for physical health. Despite these risks, existing legal and regulatory frameworks remain poorly equipped to address the layered and intersecting harms associated with synthetic media.

Legislative intent must therefore recognize deepfake-related harm as a public mental and physical health concern, mandate transparency and accountability in the creation and dissemination of synthetic media, provide trauma-informed supports for affected individuals, and require ethical artificial intelligence standards that incorporate mental and physical health impact assessments. Explicit recognition of physical health consequences strengthens legal standing,

aligns regulatory efforts with public health and disability law, anticipates constitutional and tort-based challenges, and firmly grounds AI governance in well-established mind–body science.

The crosswalk below demonstrates that regulating deepfake-related harm is not novel or speculative, but rather a logical extension of longstanding public health, disability, and safety statutes that already recognize the mind–body connection, psychological injury as a precursor to physical harm, and the government's authority to intervene when emerging risks threaten population health.

Statutory Crosswalk Table
Legislative Interpretation Guidance

Existing Statute / Framework	Recognized Harm	Established Principle	Application to Deepfakes
Public Health Service Act (42 U.S.C. § 201)	Indirect threats to public health	Government may regulate conditions increasing disease risk	Deepfakes introduce widespread psychological stressors with health impacts
CDC ACEs Framework	Psychological trauma	Trauma predicts chronic disease	Deepfake exposure constitutes adverse psychological experience
Violence Against Women Act (VAWA)	Emotional & psychological abuse	Harm need not be physical	Nonconsensual sexual Deepfakes mirror recognized abuse
Americans with Disabilities Act (ADA)	Mental impairments	Psychological conditions can be disabling	Trauma and stress from Deepfakes impair major life activities
OSHA	Stressful work environments	Chronic stress causes injury	Workplace deepfakes create hostile conditions
State Domestic Violence Statutes	Coercive control	Psychological manipulation is actionable	Deepfakes function as digital gaslighting
Environmental Health Law	Delayed harm	Exposure can cause cumulative injury	Deepfakes act as digital environmental toxins

Based on this crosswalk, legislatures may reasonably conclude that deepfake-related psychological harm falls squarely within existing public health and safety doctrines, that regulation need not await physical injury where scientific evidence demonstrates foreseeable harm, and that recognizing mind–body injury strengthens constitutional rational-basis review. Regulating deepfakes therefore aligns with established legislative precedent rather than representing a departure from it.

Policy Recommendations

To address the multifaceted harms associated with deepfakes, the following policy actions are recommended:

1. **Legal Recognition of Harm**
   Statutes should explicitly recognize deepfake-related psychological injury and stress-induced physical illness as legally cognizable harms.
2. **Mandatory Disclosure of Synthetic Media**
   AI-generated or materially altered audiovisual content should be subject to clear and standardized disclosure requirements.
3. **Accountability for Malicious Use**
   Criminal and civil penalties should apply where deepfakes are used to cause psychological trauma, reputational injury, or physical health consequences.
4. **Trauma-Informed Victim Support**
   Governments should support integrated mental and physical healthcare services for individuals harmed by deepfake-related abuse.
5. **Ethical AI and Health Impact Assessments**
   Developers of high-risk generative AI systems should be required to assess foreseeable mental and physical health impacts prior to deployment.

Proposed Legislative Findings and Intent

Legislative bodies may reasonably find that deepfakes pose a foreseeable risk to public mental and physical health, that psychological trauma resulting from digital deception produces predictable physiological harm, and that existing legal frameworks inadequately address identity-based and reality-distorting injury. It is therefore the intent of the Legislature to regulate deepfake technologies in a manner consistent with public health science and disability law, recognizing the inseparability of mental and physical health.

References

American Psychiatric Association. (2022). *DSM-5-TR: Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). Author.

American Psychological Association. (2023). *Stress effects on the body*. https://www.apa.org/topics/stress/body

Centers for Disease Control and Prevention. (2022). *How stress affects your health*. https://www.cdc.gov/stress/about/index.html

Chesney, R., & Citron, D. K. (2019). Deep fakes and the new disinformation war. *Foreign Affairs, 98*(1), 147–155.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*(1), 3–28. https://doi.org/10.1037/0033-2909.114.1.3

Lewandowsky, S., Cook, J., Ecker, U. K. H., Albarracín, D., Amazeen, M. A., Kendeou, P., Lombardi, D., Newman, E. J., Pennycook, G., Porter, E., Rand, D. G., Vraga, E. K., & Zaragoza, M. S. (2020). Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition, 9*(3), 353–369. https://doi.org/10.1016/j.jarmac.2020.07.008

Litz, B. T., Stein, N., Delaney, E., Lebowitz, L., Nash, W. P., Silva, C., & Maguen, S. (2009). Moral injury and moral repair in war veterans: A preliminary model and intervention strategy. *Clinical Psychology Review, 29*(8), 695–706. https://doi.org/10.1016/j.cpr.2009.07.003

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory, 12*(4), 361–366. https://doi.org/10.1101/lm.94705

McGaugh, J. L. (2004). The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annual Review of Neuroscience, 27*, 1–28. https://doi.org/10.1146/annurev.neuro.27.070203.144157

Nash, R. A., Wade, K. A., Garry, M., Loftus, E. F., & Ost, J. (2023). Deepfakes, false memories, and belief: Experimental evidence for memory distortion. *Psychological Science, 34*(7), 945–960. https://doi.org/10.1177/09567976231164027

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science, 333*(6043), 776–778. https://doi.org/10.1126/science.1207745

Sweet, P. L. (2019). The sociology of gaslighting. *American Sociological Review, 84*(5), 851–875. https://doi.org/10.1177/0003122419874843

van der Kolk, B. A. (2014). *The body keeps the score: Brain, mind, and body in the healing of trauma*. Viking.

Westerlund, M. (2023). The emergence of deepfake technology: A review. *Technology Innovation Management Review, 13*(1), 33–45. https://doi.org/10.22215/timreview/1640

World Health Organization. (2023). *World mental health report: Transforming mental health for all*. https://www.who.int/publications/i/item/9789240050860