

Republic of the Philippines
HOUSE OF REPRESENTATIVES
Quezon City

TWENTIETH CONGRESS
First Regular Session

House Bill No. **9449**



Introduced by Representative KAREN HOPE F. GARCIA

EXPLANATORY NOTE

The rapid advancement of artificial intelligence has enabled the creation of highly realistic synthetic audio, image, and video content, commonly referred to as “deepfakes.” While these technologies have legitimate artistic, educational, and commercial uses, they have also been weaponized to destroy reputations, deceive voters, exploit women, defraud the public, and erode trust in institutions.

The danger is no longer speculative. The World Economic Forum has identified AI-generated misinformation and disinformation as among the most severe short-term global risks to democratic processes and public trust¹. In the Philippine context, where digital media consumption is among the highest in the world, the societal impact of malicious synthetic media is particularly acute. Non-consensual explicit deepfakes, in particular, have inflicted grave emotional and psychological harm, disproportionately targeting women and private individuals who lack the means to defend themselves².

Empirical research demonstrates that manipulated media significantly undermines public confidence in institutions and can distort political discourse³. A single fabricated video can influence elections, incite public unrest, or permanently stain a person’s name. These harms are not theoretical, they are documented, measurable, and escalating. Yet despite these risks, legislative responses must be carefully calibrated to ensure that the cure does not become more dangerous than the disease.

Article III, Section 4 of the 1987 Constitution provides that “no law shall be passed abridging the freedom of speech, of expression, or of the press.” The Supreme Court has consistently held that prior restraint, or any government action that suppresses speech before it is published, bears a heavy presumption against constitutional validity. In *Chavez v. Gonzales*

¹ World Economic Forum, *Global Risks Report 2024* (2024), identifying AI-generated misinformation and disinformation as a top global risk

² Henry Ajder et al., *The State of Deepfakes: Landscape, Threats, and Impact* (Deeptrace Labs, 2019); see also Citron, D. (2019). “Sexual Privacy.” *Yale Law Journal*, 128:1870.

³ Chesney, R., & Citron, D. (2019). “Deepfakes and the New Disinformation War.” *Foreign Affairs*, 98(1), 147–155; see also Vaccari, C., & Chadwick, A. (2020). “Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News.” *Social Media + Society*, 6(1).

(2008)⁴, the Court struck down governmental warnings that effectively suppressed media publication. It emphasized that even well-meaning government warnings can unlawfully impose a chilling effect on free expression. Any regulation of synthetic media must therefore avoid blanket bans, speech licensing systems, or mandatory registration schemes that operate as prior restraint.

This bill does not regulate opinion, criticism, satire, or dissent. It regulates harmful conduct. It punishes only the malicious use of synthetic media - that is, when a person intentionally uses fabricated digital content to deceive, defraud, defame, harass, threaten, or sexually exploit another. By focusing on demonstrable harm and culpable intent, the measure avoids unconstitutional overbreadth and vagueness. The Supreme Court has long required precision in speech regulation to prevent chilling effects, and this bill adheres to that standard by narrowly defining prohibited acts.

For public officials and public figures, the bill adopts the constitutional standard of “actual malice,” as recognized in *New York Times v. Sullivan*⁵. This means that no public official may recover unless it is shown that the malicious deepfake was created with knowledge that it was false or with reckless disregard for the truth. This safeguard ensures that political debate remains robust, that satire remains protected, and that citizens remain free to criticize those in power without fear.

Rather than impose government censorship before speech occurs, the bill adopts a measured and constitutional approach. It requires social media platforms to act upon verified reports of malicious deepfakes after publication, through a clear notice-and-takedown directed at platforms. By imposing post-publication remedial obligations rather than pre-publication censorship, the measure respects constitutional boundaries while providing victims with meaningful recourse. The State’s interest here is substantial: protecting citizens from fraud, exploitation, and identity abuse.

The bill also prioritizes civil remedies over broad criminal sanctions. Legal scholarship and comparative legislation indicate that civil liability frameworks are less intrusive to speech while still deterring harmful conduct.⁶ Victims are given the right to seek damages in court. Criminal sanctions are reserved only for clearly unprotected acts such as fraud, identity theft, and sexual exploitation. This balanced framework recognizes that while speech is protected, deception and abuse are not.

Ultimately, this measure recognizes that freedom of expression is not absolute. Our freedom does not include the right to weaponize falsehood to destroy another person’s dignity or to undermine public institutions. The Constitution protects speech, but not fraud, identity theft,

⁴ *Chavez v. Gonzales*, G.R. No. 168338, 15 February 2008, 545 SCRA 441.

⁵ *New York Times Co. v. Sullivan*, 376 U.S. 254 (1964).

⁶ Citron, D. (2014). *Hate Crimes in Cyberspace*. Harvard University Press; see also U.S. Congress, “DEFIANCE Act of 2024” (civil remedies for non-consensual intimate deepfakes).

or malicious deception. It does not protect calculated digital impersonation designed to cause harm.

As technology evolves, so too must the law, provided that it remains faithful to constitutional principles. This bill therefore seeks to protect individuals, preserve democratic integrity, and promote responsible innovation, while simultaneously safeguarding the fundamental freedoms that define our constitutional order. This measure seeks to defend both the Filipino people and the freedoms guaranteed to them.

In view of the foregoing, the immediate passage of this bill is earnestly sought.

A handwritten signature in black ink, appearing to read 'KHFG', positioned above the printed name of the representative.

Rep. KAREN HOPE F. GARCIA
3rd District, Province of Cebu

Republic of the Philippines
HOUSE OF REPRESENTATIVES
Quezon City

TWENTIETH CONGRESS
First Regular Session

House Bill No. **9449**

Introduced by Representative KAREN HOPE F. GARCIA

AN ACT
REGULATING THE MALICIOUS USE OF SYNTHETIC MEDIA AND DEEPPAKES,
PROVIDING CIVIL REMEDIES AND CRIMINAL PENALTIES THEREFOR,
ESTABLISHING CONTENT-NEUTRAL PLATFORM OBLIGATIONS,
AND FOR OTHER PURPOSES

Be it enacted by the Senate and the House of Representatives of the Philippines in Congress assembled:

CHAPTER I
GENERAL PROVISIONS

Section 1. *Short Title.* - This Act shall be known as the "Anti-Deepfake and Synthetic Media Regulation Act".

Section 2. *Declaration of Policy.* It is hereby declared the policy of the State to protect the dignity, security, privacy, and democratic rights of the people from malicious digital impersonation, fraud, and exploitation committed through synthetic media technologies.

The State likewise recognizes that freedom of speech, expression, and of the press are fundamental rights guaranteed under Article III, Section 4 of the 1987 Constitution. Any regulation of synthetic media shall therefore be narrowly tailored to address demonstrable harm, avoid prior restraint, and preserve robust political debate, satire, artistic expression, academic inquiry, and lawful dissent.

This Act regulates harmful conduct involving malicious synthetic media and shall not be interpreted to impose blanket bans, speech licensing, mandatory user registration, or viewpoint-based restrictions.

Section 3. Definition of terms. For the purposes of this Act:

- a. **“Actual Malice”** – For public officials and public figures, liability shall attach only upon proof that the creator acted with knowledge of falsity or reckless disregard for truth
- b. **“Election Speech”** - refers to any expression, communication, satire, parody, commentary, advocacy, criticism, or opinion relating to:
 - i. A candidate for public office;
 - ii. A political party or political organization;
 - iii. A public policy issue submitted to voters; or
 - iv. The conduct of public officials or electoral processes.
- c. **“Malicious Deepfake”** – refers to any synthetic media created, published, or distributed with specific intent to:
 - i. Deceive or defraud;
 - ii. Defame a person through materially false representations;
 - iii. Harass, intimidate, or threaten;
 - iv. Facilitate identity theft;
 - v. Sexually exploit a person without consent; or
 - vi. Cause demonstrable and material harm to an identifiable individual.

The term shall not include clearly identifiable parody, satire, commentary, artistic expression, or fictionalized entertainment.
- d. **“Material Harm”** – refers to any direct, demonstrable, and substantial injury that is not speculative, hypothetical, or trivial, including:
 - i. Financial loss or property damage;
 - ii. Loss of employment or professional opportunity directly attributable to the malicious synthetic media;
 - iii. Reputational harm meeting the elements of actionable defamation;
 - iv. Severe emotional distress arising from non-consensual sexually explicit synthetic media;
 - v. Identity theft or unlawful impersonation; or
 - vi. Material interference with a lawful election process through deception of verifiable factual information.
- e. **“Notice-and-Takedown”** – refers to a post-publication remedial process whereby a platform, upon receipt of a verified complaint meeting statutory requirements, acts within the prescribed period under this act to remove, restrict, or label malicious synthetic media.

- f. **“Online Platform”** – refers to a digital service that enables users to create, upload, share, or view user-generated content.
- g. **“Synthetic Media”** – refers to digitally altered or artificially generated audio, image, or video content that realistically depicts a person engaging in speech or conduct that did not occur in real life.

Section 4. *Prohibited Acts.* It shall be unlawful for any person to knowingly create, publish, distribute, or cause the distribution of malicious synthetic media with specific intent to:

- a. Defraud any person or the public;
- b. Commit identity theft or unlawful impersonation;
- c. Defame another through materially false factual representation;
- d. Harass, threaten, or incite violence against a person;
- e. Produce or distribute non-consensual sexually explicit synthetic media; or
- f. Interfere with the integrity of elections through materially deceptive synthetic media designed to mislead voters regarding factual matters.

Section 5. *Exemptions.* This Act shall not apply to:

- a. Content created for satire, parody, or artistic expression;
- b. Deepfakes created for educational or research purposes;
- c. Content where the individual depicted has given explicit, verifiable, and informed consent
- d. Political speech, unless it falls within the unprotected categories defined in Section 4.

The foregoing shall not apply to similar types of content where a reasonable person would not interpret the same as asserting actual facts, nor to content accompanied by clear and conspicuous disclosure that it is synthetic.

Section 6. *Content-Neutral Platform Obligations.* Online platforms shall establish accessible mechanisms for users to report malicious synthetic media. Upon receipt of a verified complaint demonstrating prima facie violation of Section 4, supported by sufficient documentation, platforms shall act within forty-eight (48) hours to:

- a. Remove or disable access to the content;
- b. Apply a clear and conspicuous label indicating the content is disputed synthetic media; or
- c. Temporarily restrict distribution pending review.

Platforms shall notify the uploader and provide an opportunity to contest the complaint. Nothing in this Act shall require platforms to monitor content proactively or to implement pre-publication review systems.

Section 7. *Penalties.* Any person who violates this act shall suffer imprisonment of six (6) to twelve (12) years and shall pay a fine of not less than Five hundred thousand pesos (Php 500,000.00), or both, at the discretion of the court.

Any person aggrieved by malicious synthetic media may file a civil action for damages, including actual, moral, and exemplary damages, and may seek injunctive relief after publication. Courts may order the removal of malicious synthetic media upon a finding of probable violation of this Act.

The remedies under this Act shall be without prejudice to other remedies available under the Civil Code, the Revised Penal Code, the Cybercrime Prevention Act, the Safe Spaces Act, and other applicable laws.

Section 8. *Implementing Rules and Regulations.* Within ninety (90) days from the effectivity of this Act, the Department of Information and Communications Technology (DICT), in consultation with the Department of Justice (DOJ), Department of Science and Technology (DOST), and Department of Trade and Industry (DTI) shall promulgate the necessary implementing rules and regulations to ensure the effective enforcement of this Act.

Section 9. *Appropriations.* The amount necessary to carry out the provisions of this Act shall be included in the annual General Appropriations Act.

Section 10. *Separability Clause.* If any provision of this Act is declared unconstitutional or invalid, the remaining provisions shall continue to be in full force and effect.

Section 11. *Repealing Clause.* All laws, decrees, executive orders, rules and regulations, or parts thereof inconsistent with the provisions of this Act are hereby repealed or modified accordingly.

Section 12. *Effectivity.* This Act shall take effect fifteen (15) days after its publication in the Official Gazette or in a newspaper of general circulation.

Approved,