

**IMPERCEPTIBLE SLEIGHTS:
THE RISE OF POLITICAL DEEPFAKES AND CHALLENGES IN
THE REGULATION OF ELECTION PROPAGANDA***

*Michiko Lokin***

ABSTRACT

The past decade has seen a global surge of disinformation online in the context of political elections. Notably, it has played a crucial role in the past two national elections in the Philippines. The continuous development of artificial intelligence (AI) tools in more recent years has further enhanced the believability of false information. In particular, disinformation networks have used deepfakes—a branch of AI that manipulates audio, visual, or audiovisual content to change how a person or object is presented—to falsely attribute speech or acts, among others, to rival political candidates. This Article explores the rise of political deepfakes as a threat to electoral integrity. This arises from the premise that disinformation pollutes free discourse integral to the electoral process. Further, it surveys the regulatory landscape, evaluating the sufficiency and constitutionality of bills and regulatory materials in relation to political deepfakes. Finally, the Article concludes by providing guidelines for prospective legislation and regulatory measures.

* Cite as Michiko Lokin, *Imperceptible Sleights: The Rise of Political Deepfakes and Challenges in the Regulation of Election Propaganda*, 98 PHIL. L.J. 245, [page cited] (2025).

** J.D., *cum laude*, University of the Philippines (2024). B.A. in Economics, University of California, Los Angeles (UCLA) (2019). Editor, Vol. 97, PHILIPPINE LAW JOURNAL.

The author would like to extend her sincere gratitude to Ignacio Lorenzo Dela Cruz Villareal and Jordan Vega for their invaluable insights, comments, and support.

I. INTRODUCTION	246
A. Background.....	248
B. What are Deepfakes?.....	250
II. POLITICAL DEEPFAKES AND THEIR ROLE IN THE ELECTORAL PROCESS	251
A. Elections as the Bedrock of Democracy.....	252
B. Free Speech and the Marketplace of Ideas.....	255
C. Political Speech and Electoral Integrity.....	261
D. Political Deepfakes as Amplifiers of Disinformation.....	266
III. THE CURRENT PHILIPPINE REGULATORY LANDSCAPE	274
A. Partisan Political Activity & Election Propaganda	275
B. COMELEC Regulatory Guidelines on Social Media and Artificial Intelligence	281
1. <i>Resolution No. 10730</i>	282
2. <i>Resolution No. 11064</i>	290
C. Deepfake Misuse as Election Offenses	299
1. <i>Election Offenses</i>	299
2. <i>Obligations on Stakeholders</i>	309
D. Pending Bills on Deepfake Regulation.....	313
1. <i>House Bill No. 5406</i>	313
2. <i>House Bill No. 9425</i>	320
3. <i>House Bill No. 10567</i>	327
IV. CRAFTING PROSPECTIVE LEGISLATION.....	332
V. CONCLUSION	347

I. INTRODUCTION

Since time immemorial, propaganda has proved its centrality in the political arena and especially in the context of elections. From printed posters to radio broadcasts and even commercials, propaganda has long been leveraged by both political parties and candidates to bolster their chances of election to public office. Conversely, propaganda has likewise been exploited to stifle and discredit political detractors.

It comes as no surprise then that candidates and parties have been quick to co-opt the latest technological tools to expand their reach to voters, communicate their political platforms, and secure their respective victories

on election day. One of the technologies adopted as an instrument in the modern-day campaign arsenal is artificial intelligence (“AI”).¹

Specifically, political candidates and aspirants worldwide² have begun to exploit a specific type of AI technology known as deepfakes. Its increasing ubiquity and low-cost have been important driving factors in incentivizing candidates to embrace its use as it has provided a means to churn out election materials such as posters, videos, and other similar kinds of media at a staggering volume and pace.³

The Philippines is no stranger to this global trend and is similarly expected to see a rise in the use of deepfakes in subsequent elections.⁴ Much like any other technology, the propensity for the abuse of deepfakes remains a paramount consideration. Brewing concerns arising from the use of this kind of technology have become pertinent amidst the reality that the Philippines has been touted as a “hotbed for online disinformation.”⁵ Amplified by social media algorithms and negative psychological biases,⁶ political deepfakes stand as an emerging threat to the integrity of elections and democracy at large.

This paper aims to fill the gaps in existing research and provide a reference point in the discussion and regulation of political deepfakes. *Part I*

¹ See Prashnu Verma & Cat Zakrzewski, *AI Deepfakes Threaten to Upend Global Elections. No One can Stop Them*, WASH. POST, Apr. 23, 2024, at www.washingtonpost.com/technology/2024/04/23/ai-deepfake-election-2024-us-india.

² Countries such as Taiwan, India, and Moldova, among others have begun to see the use of deepfakes in the electoral context. *See id.*

³ *Id.*

⁴ Elijah Felice Rosales, *Interpol: Expect Deepfakes Surge in 2025 Philippines Elections*, PHIL. STAR, Sept. 12, 2024, at www.philstar.com/headlines/2024/09/12/2384671/interpol-expect-deepfakes-surge-2025-philippines-elections; Sandy Locus, *AI, Deepfakes to Figure More in Campaign for Eleksyon 2025 —Expert*, GMA NEWS ONLINE, Aug. 29, 2024, at <https://www.gmanetwork.com/news/topstories/specialreports/918489/ai-deepfakes-to-figure-more-in-campaign-for-eleksyon-2025-expert/story/>; Almira Louise S. Martinez, *Deepfake can Cause Problems in the 2025 PH Elections – IBM*, BUS. WORLD, Sept. 26, 2024, at <https://www.bworldonline.com/technology/2024/09/26/624057/deepfake-can-cause-problems-in-the-2025-ph-elections-ibm/>; Jekki Pascual, *DICT Warns vs Fake AI Content this Election*, ABS-CBN NEWS, Sept. 30, 2024, at <https://news.abs-cbn.com/news/2024/9/30/dict-warns-vs-fake-ai-content-this-election-1416>.

⁵ Rosales, *supra* note 4; *See also* Yvonne T. Chua, et al., *Filipino Voters were Engulfed in Relentless Stream of Disinformation*, May 11, 2022, at <https://up.edu.ph/filipino-voters-were-engulfed-in-relentless-stream-of-disinformation/>.

⁶ Robert Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CAL. L. REV. 1753, 1765–68 (2019).

provides a brief introduction and background on AI and discusses how deepfakes typically function.

Part II explores how the rise of political deepfakes has essentially upended the electoral process. It revisits the central tenet that elections are the bedrock of democracy. Additionally, it explores the relationship between free speech and the promotion of the marketplace of ideas. It then underscores the role of political speech and its effects on electoral integrity. It also delves into the harms that political deepfakes present as amplifiers of disinformation on social media and scrutinizes the effects of such deepfakes in countries such as the United States of America, the Philippines, and other countries.

Part III maps out the current Philippine regulatory landscape and outlines the regulation of partisan political activity and election propaganda under the Omnibus Election Code (“OEC”), the Fair Elections Act, and several Commission on Elections (COMELEC) resolutions. It surveys the guidelines on social media and artificial intelligence use, and breaks down how the misuse of deepfakes may be considered election offenses. It also looks into extant bills crafted by legislators which seek to regulate the use of deepfakes to examine current legislative attitudes towards deepfakes and the corresponding strengths and weaknesses of current bills. More importantly, this part provides an analysis of the pertinent provisions in these instruments. It argues that these existing regulatory guidelines and bills are insufficient and unconstitutional for contravening existing standards and jurisprudential pronouncements.

Part IV draws out regulatory limitations which must be considered by legislators. It also provides guidance on perspectives and standards to be considered in creating prospective legislation. *Part V* concludes the paper.

A. Background

Any discussion on deepfakes must first begin with the preliminary technology which has enabled its production (i.e., artificial intelligence). One of the first definitions of “artificial intelligence” under Philippine law can be found in the implementing rules and regulations (“IRR”) of Republic Act (“RA”) No. 9728 as amended by RA No. 11453 or the “Freeport Area of Bataan Act of 2009,” to wit:

[S]ometimes called machine intelligence, refers to intelligence demonstrated by machines or the ability of a digital computer or

computer-controlled robot to perform tasks commonly associated with intelligent beings; the term Artificial Intelligence is frequently applied to the project of developing systems endowed with the intellectual processes characteristic of humans, such as the ability to reason, discover meaning, generalize, or learn from past experience.⁷

Likewise, the IRR provides the definition of “artificial technology” as:

[A]n area of computer science also known as artificial intelligence that emphasizes the creation of intelligent machines and technologies that work and reacts [sic] like humans.⁸

In other words, under the IRR, “artificial technology” is a sub-branch of computer science concerned with artificial or machine intelligence whereas “artificial intelligence” refers to the ability of computer systems to exhibit distinctive human characteristics such as improvement and reasoning, and perform specific tasks.

Apart from the IRR, COMELEC Resolution No. 11064 which was published in 2024 defines “artificial intelligence technology” as:

[R]efer[ring] to computer systems or machines that simulate human intelligence processes with focus on learning, reasoning[,] problem solving, perception, use of language, and image manipulation influencing real or virtual environments.⁹

Similar to the definitions in the IRR of the Freeport Area of Bataan Act of 2009, the definition put forth by COMELEC recognizes the ability of computers or machines to simulate human characteristics to perform tasks. The definition in the COMELEC Resolution merely goes further to enumerate some use cases of AI.

Under this umbrella of AI exists “generative artificial intelligence (GenAI), which is ‘a branch of artificial intelligence that enables [computerized systems] to quickly and convincingly create original content ranging from images and artwork to poetry, music, text, video, dialog, and even computer code.’”¹⁰ GenAI relies on “large language models” (“LLMs”)

⁷ Rep. Act No. 9728 Rules & Regs. (2019), § 4(e).

⁸ § 4(f).

⁹ COMELEC Res. No. 11064 (2024), art. II(2).

¹⁰ Michael Murray, *Generative AI Art: Copyright Infringement and Fair Use*, 26 SMU L.J. SCI. & TECH. 259, 260 (2023).

which are repositories of vast tomes of input data which may consist of text, audio, or images.¹¹

In relation to the definition of AI under the IRR of the Freeport Area of Bataan Act of 2009, GenAI thus refers to the specific ability of a computer system to learn and improve from input media data in order to generate new media output in line with the parameters provided by a specific user prompt.¹² Apart from answer engines like ChatGPT, another common use of GenAI is the creation of deepfakes.

B. What are Deepfakes?

Deepfakes are “visual and audio content that has been manipulated using advanced software to change how a person, object or environment is presented.”¹³ It is a portmanteau of the words “deep learning” and “fakes.”¹⁴ It falls under the branch of GenAI as models leverage the data, patterns, and information learned to produce specific output.¹⁵

In particular, deepfakes are created using “neural networks” which repeatedly undergo a process of “training” to create increasingly realistic output.¹⁶ Training “requires the use of an autoencoder consisting of two neural networks— an encoder and decoder.”¹⁷ The encoder “takes an image of a face and compresses it into a low dimension representation, also known as the ‘latent face,’” then the decoder “takes that representation and reconstructs the face into its original form.”¹⁸ This process is repeated until the target output media is constructed.

Typically, the creation of deepfakes involves one of the following processes: (1) Face replacement; (2) Face re-enactment; (3) Face generation; or (4) Speech synthesis.¹⁹

¹¹ *Id.* at 259–61.

¹² *Id.* at 296–97.

¹³ Centre for Data Ethics and Innovation, *Deepfakes and Audio-visual Disinformation*, CDEI SNAPSHOT SERIES, Sept. 2019, at 2.

¹⁴ Alexandra Tashman, “Malicious Deepfakes” - How California’s A.B. 730 Tries (and Fails) to Address the Internet’s Burgeoning Political Crisis, 54 LOY. L.A. L. REV. 1391, 1392 (2021).

¹⁵ Locus, *supra* note 4.

¹⁶ Chesney & Citron, *supra* note 6, at 1759 & nn.12, 14.

¹⁷ Centre for Data Ethics and Innovation, *supra* note 13, at 7–9.

¹⁸ *Id.*

¹⁹ *Id.* at 4–10.

For face replacement, the image of a person's face is first extracted (the source or input data) then stitched or superimposed onto that of another person (the target or output data).²⁰ Afterwards, the training process occurs, and the source image becomes "decompressed using its original encoder but reconstructed using the target image's decoder."²¹ Lastly, the image is inserted into the video or specified media.²²

Face re-enactment involves the manipulation of the "features of a target's face, including the movement of their mouth, eyebrows, eyes and the tilting of their head."²³ In this type of process, a source subject can be made to communicate and express ideas even if they have not factually done so.²⁴

On the other hand, face generation involves the creation of an entirely new output image using "generative adversarial networks" ("GANs") which function by "pitting two neural networks against one another" to generate a realistic image in line with the prompt provided by the user.²⁵

Lastly, speech synthesis involves the creation of a "model of someone's voice, which can read out text in the same manner, intonation and cadence as the target person."²⁶ All these functions outline just some ways in which deepfake technology can be utilized to create different forms of media; however, it would be unsurprising to see this list grow in light of ongoing technological breakthroughs.

II. POLITICAL DEEPPAKES AND THEIR ROLE IN THE ELECTORAL PROCESS

Far from being an imagined threat, political deepfakes pose a very real harm to the electoral process. They stand to disrupt the flow of information and distort the quality of political discourse that may ultimately influence election outcomes. Coupled with social media algorithms, which

²⁰ *Id.* at 4.

²¹ *Id.* at 7–9.

²² *Id.* at 9.

²³ *Id.* at 5.

²⁴ *Id.*

²⁵ *Id.*

²⁶ *Id.* at 6.

have been proven to be effective echo chambers of misinformation online,²⁷ political deepfakes possess the potential to become cogent devices to sow division among voters and chip away at institutional trust.²⁸ In turn, checks and balances which function to correct public misconceptions become increasingly difficult to calibrate. More importantly, differentiating the truth amidst the flurry of online disinformation becomes an onerous task for the ordinary voter.

Hence, there must be an inquiry into the centrality of elections in the democratic system, and the role of free speech and the marketplace of ideas, contextualized with political speech. Lastly, there must be an examination of how political deepfakes stand to upend this delicate ecosystem.

A. Elections as the Bedrock of Democracy

The ability of the people to choose their representatives in the legislature remains to be one of the central pillars of the democratic process. As regards the electoral process, the chief consideration is to “give effect to the expressed will of the majority.”²⁹

As early as 1918, the Supreme Court in the case of *Luna v. Rodriguez*³⁰ had the opportunity to pass upon the importance of the electoral process by citing the case of *Loomis v. Jackson*.³¹ *Luna* opined that: “The purpose of an election is to give the voters a direct participation in the affairs of their

²⁷ See Shannon Bond, *AI Fakes Raise Election Risks as lawmakers and Tech Companies Scramble to Catch up*, NAT'L PUB. RADIO, Feb. 8, 2024, at <https://www.npr.org/2024/02/08/1229641751/ai-deepfakes-election-risks-lawmakers-tech-companies-artificial-intelligence>; See Haochen Sun, *Regulating Algorithmic Disinformation*, 46 COLUM. J. L. & ARTS. 367, 369 (2023); See also Chesney & Citron, *supra* note 6, at 1765–68.

²⁸ Nahema Marchal et al., *Generative AI Misuse: A Taxonomy of Tactics and Insights from Real-World Data*, ARXIV, preprint at 12 (June 2024), available at <https://arxiv.org/pdf/2406.13843>; Chesney & Citron, *supra* note 6, at 1779; See W. Lance Bennett & Steven Livingston, *A Brief History of the Disinformation Age: Information Wars and the Decline of Institutional Authority*, in THE DISINFORMATION AGE: POLITICS, TECHNOLOGY, AND DISRUPTIVE COMMUNICATION IN THE UNITED STATES 8–12, 20, 28–29 (2021); See Sun, *supra* note 27, at 375–76; See also Dave Karpf, *How Digital Disinformation Turned Dangerous*, in THE DISINFORMATION AGE: POLITICS, TECHNOLOGY, AND DISRUPTIVE COMMUNICATION IN THE UNITED STATES 154, 158–65 (2021).

²⁹ *Yason v. COMELEC*, G.R. No. 52713, 134 SCRA 371, 379, Jan. 31, 1985.

³⁰ 39 Phil. 208 (1918).

³¹ 6 W.Va. 613 (1873).

government, either in determining who shall be their public officials or in deciding some question of public interest[.]”³²

The case of *In re Geronimo*³³ goes further and provides a poignant discussion on the significance of representation. The Court here ruled that it is through this electoral process that people have a direct voice in determining the direction of the country’s governance with legislators who ideally represent their best interests.³⁴ This right to vote is also deemed as one of the most “sacred of freedoms.”³⁵

Furthermore, in the 2023 case of *Macalintal v. COMELEC*, the Court reiterated the sanctity of the right to suffrage as a manifestation of the sovereignty inherently vested in the populace.³⁶ The Court also cited Article II, Section 1 of the Constitution to affirm this relationship, to wit: “The Philippines is a democratic and republican State. Sovereignty resides in the people and all government authority emanates from them.”³⁷ Through the deliberations of the Constitutional Commission, the same case also highlighted the deliberate placement of the word “democratic” in Article II, Section 1 to emphasize the centrality of the people’s role in the government.³⁸ The 1987 Constitution, as the fundamental law of the land, even derives its mandate from the people as evidenced by the ordinance and promulgation of the document by the “sovereign Filipino people.”³⁹ In conjunction, the word “republican” in Article II, Section 1 was meant to underscore the importance of representative governance in the Philippines.⁴⁰

Hence, the primacy of the ability of citizens to vote for their representatives strikes at the very core of Philippine democracy. After all, it is only through the consent of the people that the mandate of the state was forged, and the proverbial social contract executed.

The Court then builds on the importance of suffrage vis-a-vis sovereignty by discussing the seminal case of *Yick Wo v. Hopkins* which was decided by the US Supreme Court (“SCOTUS”) in 1886.⁴¹ Even then, the

³² Luna v. Rodriguez, 39 Phil. 208, 215 (1918).

³³ G.R. No. 60504, 136 SCRA 435, May 14, 1985.

³⁴ *Id.* at 446.

³⁵ *Id.*

³⁶ *Id.* at 21, .

³⁷ CONST. art. II, §1.

³⁸ *Macalintal*, G.R. No. 263590, slip op. at 11.

³⁹ *Id.* at 11–12; CONST. pmb.

⁴⁰ *Id.* at 12.

⁴¹ *Id.* at 13, *citing* *Yick Wo v. Hopkins*, 118 U.S. 356 (1886).

case of *Yick Wo* recognized the right to suffrage as a “fundamental political right” which was in turn “preservative of all rights.”⁴²

Going further, the case of *Macalintal* highlighted the centrality of the right to suffrage as it was embodied in international instruments such as Article 21 of the Universal Declaration of Human Rights (“UDHR”) and Article 25 of the International Covenant on Civil and Political Rights (“ICCPR”).⁴³

While the case draws on Article 21 of the UDHR to demonstrate the prevalence of the right to suffrage, it must be noted that the instrument is not in itself a treaty and is thus not binding.⁴⁴ However, as a caveat, several provisions of the UDHR have been recognized as embodiments of general principles of law⁴⁵ which are binding on the Philippines as Article II, Section 2 of the Constitution provides that: “The Philippines [...] adopts the generally accepted principles of international law as part of the law of the land[.]”⁴⁶ Moreover, notwithstanding the semantic difference in the formulation of the term “general principles of law” as reflected in Article 38 of the Statute of the International Court of Justice (ICJ),⁴⁷ the Supreme Court in *Pangilinan v. Cayetano*⁴⁸ affirmed that the “generally accepted principles of international law” in Article II, Section 2 of the Constitution included both customary international law norms and general principles of law.⁴⁹ Likewise, the Supreme Court in *Poe-Llamanzares v. COMELEC*⁵⁰ has stated that the UDHR may be deemed as “part of the generally accepted principles of international law.”⁵¹

⁴² *Id.*

⁴³ *Id.* at 14.

⁴⁴ JAMES CRAWFORD, BROWNLIE’S PRINCIPLES OF PUBLIC INTERNATIONAL LAW 636 (2008 ed.).

⁴⁵ *Id.*

⁴⁶ CONST. art. II, § 2.

⁴⁷ Statute of the International Court of Justice, art. 38, June 26, 1945, 59 Stat. 1055, T.S. No. 993, 3 Bevans 1179.

⁴⁸ *Pangilinan v. Cayetano*, G.R. No. 238875, 976 SCRA 509, Mar. 16, 2021.

⁴⁹ *Id.* at 572.

⁵⁰ *Poe-Llamanzares v. COMELEC*, G.R. No. 221697, 786 SCRA 1, Mar. 8, 2016.

⁵¹ *Id.* at 145. The author would like to note that contrary to the opinion of the Supreme Court, under public international law, the Universal Declaration of Human Rights (UDHR) has not *in toto* been accepted as a general principle of law. Only several of its provisions have been deemed as general principles of law. See CRAWFORD, *supra* note 44, at 636.

On the other hand, the Philippines is a party to the ICCPR, hence the provisions of the treaty are unquestionably binding.⁵² Notably, Article 25 of the ICCPR possesses many of the central elements of Article 21 of the UDHR such as the right of people to participate in government through the intercession of a chosen representative via “genuine periodic elections.”⁵³ Additionally, the provision similarly recognizes that election outcomes are the definitive “free expression of the will of the electors.”⁵⁴ General Comment No. 25 which expounds on the metes and bounds of the provision affirms the centrality of the rights of participation in public affairs and “access to public service.”⁵⁵

The case of *Macalintal* also discussed the link between the right to suffrage and liberty.⁵⁶ The Court recognized that the right to vote was an exercise of liberty, thus, any restrictions were subject to due process and equal protection.⁵⁷

All in all, the consistent line of Supreme Court jurisprudence attests to the consequential value of the right to suffrage. It is a direct manifestation of the people’s sovereignty and liberty. Moreover, it is a fundamental exercise which is pivotal in determining the direction of governance. Any restrictions on the right must be carefully drawn as it is a right placed in high regard in the schema of democracy.

B. Free Speech and the Marketplace of Ideas

Carefully intertwined with the right to suffrage is the freedom of speech and expression. This fundamental right is enshrined in Article III, Section 4 of the Constitution, which states that: “No law shall be passed abridging the freedom of speech, of expression, or of the press, or the right

⁵² See United Nations Human Rights Treaty Bodies, *Ratification Status for Philippines*, OFF. OF HIGH COMM’R FOR HUM. RTS. WEBSITE, at https://tbinternet.ohchr.org/_layouts/15/TreatyBodyExternal/Treaty.aspx?CountryID=137&Lang=en (last accessed Dec. 1, 2024).

⁵³ Int’l Covenant on Civil and Political Rights [hereinafter “ICCPR”], art. 25, Dec. 16, 1966, 999 U.N.T.S. 171 (recognizing the right of every citizen to take part in the conduct of public affairs, to vote, and to be elected).

⁵⁴ ICCPR, art. 25.

⁵⁵ Human Rights Comm’n, Gen. Comment No. 25, *The Right to Participate in Public Affairs, Voting Rights and the Right of Equal Access to Public Service (Art. 25)*, U.N. Doc. CCPR/C/21/Rev.1/Add.7, ¶ 1 (1996).

⁵⁶ *Macalintal*, G.R. No. 263590, slip op. at 19.

⁵⁷ *Id.*; CONST. art. III, § 1.

of the people peaceably to assemble and petition the government for redress of grievances.”⁵⁸

In the case of *ABS-CBN Corporation v. Ampatuan*,⁵⁹ the Supreme Court expounded on Article III, Section 4. First, the Court recognized that the composite freedoms in the provision are all distinct but preferred rights, namely: (1) the freedom of speech; (2) freedom of expression; and (3) freedom of the press.⁶⁰ On the purpose of these rights, the Court stated that “[t]hey are exercised to mobilize people based on truth or an understanding of what the circumstances are.”⁶¹ These rights are synergistic in relationship and their exercise must be “meaningful” and “unencumbered.”⁶²

Drawing on the case of *ABS-CBN Broadcasting Corp. v. COMELEC*,⁶³ the Court defined the “freedom of expression” as the “means of assuring individual self-fulfillment, of attaining the truth, of securing participation by the people in social and political decision-making, and of maintaining the balance between stability and change.”⁶⁴ As explained by the Court in *Diocese of Bacolod v. COMELEC*, this right “applies to the entire continuum of speech from utterances made to conduct enacted, and even to inaction itself as a symbolic manner of communication.”⁶⁵

The Court viewed the right as an intrinsic facet of individual sovereignty by which a person was able to convey their thoughts even if disparate from those of others.⁶⁶

On the freedom of speech, the Court in *Reyes v. Bagatsing*⁶⁷ defined the right as “the liberty to discuss publicly and truthfully any matter of public concern without censorship or punishment.”⁶⁸ In other words it is “the

⁵⁸ § 4.

⁵⁹ *ABS-CBN Corp. v. Ampatuan*, G.R. No. 227004, slip. op. Apr. 25, 2023.

⁶⁰ *Id.* at 16.

⁶¹ *Id.*

⁶² *Id.* at 17.

⁶³ G.R. No. 133486, 323 SCRA 811, Jan. 28, 2000.

⁶⁴ *ABS-CBN Corp. v. Ampatuan*, G.R. No. 227004, slip op. at 17, *citing* *ABS-CBN Broad. Corp. v. COMELEC*, G.R. NO. 133486, Jan. 28, 2000.

⁶⁵ *Diocese of Bacolod v. COMELEC* [hereinafter “*Diocese of Bacolod*”], G.R. No. 205728, 747 SCRA 1, 73, Jan. 21, 2015.

⁶⁶ *ABS-CBN Corp. v. Ampatuan*, G.R. No. 227004, slip op. at 16.

⁶⁷ *Reyes v. Bagatsing* [hereinafter “*Reyes*”], G.R. No. 65366, 125 SCRA 553, Nov. 9, 1983.

⁶⁸ *Id.* at 560.

ability to appeal to one's reason through peaceful means.”⁶⁹ Citing the opinion of Justice Felix Frankfurter in his opinion in *Milk Wagon Drivers Union of Chicago, Local 753 v. Meadowmoor Dairies, Inc.*, the Court noted that the right was a “child of the Enlightenment.”⁷⁰

In the case of *Chavez v. Gonzales*,⁷¹ the Court also elucidated that the right: “means something more than the right to approve existing political beliefs or economic arrangements, to lend support to official measures, and to take refuge in the existing climate of opinion on any matter of public consequence. When atrophied, the right becomes meaningless.”⁷²

In *ABS-CBN Corp. v. Ampatuan*, the Court also noted the important function of these rights as checks and balances against unbridled state power and a coercive method to enforce state accountability.⁷³ It is a component of “deliberative democracy” and the freedom of expression thus includes the right to criticize governmental actions.⁷⁴ This sentiment has been conveyed by the Court as early as 1918 in the case of *United States v. Bustos* where it was said that: “The interest of society and the maintenance of good government demand a full discussion of public affairs. Completely [sic] liberty to comment on the conduct of public men is a scalpel in the case of free speech. The sharp incision of its probe relieves the abscesses of officialdom.”⁷⁵ Therefore, free speech on the first level, is of great consequence for its ability to rein in the otherwise uninhibited power of the state over its citizens.

On the second level, the protection afforded to these freedoms becomes even more salient in the social context as they disperse and propel the “free trade” of ideas initially theorized by US Supreme Court Justice Oliver Wendell Holmes Jr. in his dissent in *Abrams v. United States*.⁷⁶

The case of *Abrams* resulted in the US Supreme Court upholding the arrests of the defendants for publishing anti-war leaflets in New York as

⁶⁹ *ABS-CBN Corp. v. Ampatuan*, G.R. No. 227004, slip op. at 86.

⁷⁰ *Reyes*, 125 SCRA 553, 562, Nov. 9, 1983, citing *Milk Wagon Drivers Union of Chi., Local 753 v. Meadowmoor Dairies, Inc.*, 312 US 287, 293 (1940).

⁷¹ *Chavez v. Gonzales* [hereinafter “*Chavez*”], G.R. No. 168338, 545 SCRA 441, Feb. 15, 2008.

⁷² *Id.* at 484.

⁷³ *ABS-CBN Corp. v. Ampatuan*, G.R. No. 227004, slip op. at 16.

⁷⁴ *Diocese of Bacolod*, 747 SCRA 1, 77.

⁷⁵ *United States v. Bustos*, 37 Phil. 731, 740–41 (1918).

⁷⁶ 250 U.S. 616 (1919) (Holmes, J., *dissenting*).

violations of the Sedition Act, which in turn amended the Espionage Act of 1917.⁷⁷ Diverging from majority opinion, Justice Holmes professed that:

Persecution for the expression of opinions seems to me perfectly logical. If you have no doubt of your premises or your power, and want a certain result with all your heart, you naturally express your wishes in law, and sweep away all opposition. *To allow opposition by speech seems to indicate that you think the speech impotent, as when a man says that he has squared the circle, or that you do not care wholeheartedly for the result, or that you doubt either your power or your premises. But when men have realized that time has upset many fighting faiths, they may come to believe even more than they believe the very foundations of their own conduct that the ultimate good desired is better reached by free trade in ideas -- that the best test of truth is the power of the thought to get itself accepted in the competition of the market, and that truth is the only ground upon which their wishes safely can be carried out.*⁷⁸

In this original formulation, Justice Holmes advocated for the now-renowned concept of the “marketplace of ideas” in defending First Amendment protections. Holmes’ idea on the existence of a self-correcting market finds its roots in the works of John Stuart Mill.⁷⁹ In the book *On Liberty*, Mill argues for the protection of even false speech, professing that such speech facilitates “free and open debate” and inquiry into the veracity of a claim.⁸⁰ This analogy to a market also harkens to economic models put forth by Adam Smith which presuppose that “participants, pursuing their own self-interests, will lead to an efficient exchange of goods and services.”⁸¹ Instead of the state persecuting the difference in the opinion of an individual, society should be the arbiter on the soundness of the expression. In this theoretical exchange with different men, the supposed truth brought out by this spirited competition would then yield a better social equilibrium.

This discussion on First Amendment protections and the “marketplace of ideas” continues to find relevance not only in US Supreme Court jurisprudence, but also in Philippine jurisprudence. In the case of

⁷⁷ *Abrams v. United States*, 250 U.S. 616 (1919).

⁷⁸ *Abrams v. United States*, 250 U.S. 616, 630 (1919) (Holmes, J., *dissenting*). (Emphasis supplied.)

⁷⁹ Daniela C. Manzi, *Managing the Misinformation Marketplace: The First Amendment and the Fight Against Fake News*, 87 *FORDHAM L. REV.* 2623, 2626 (2019).

⁸⁰ *Id.*, citing JOHN STUART MILL, *On Liberty*, reprinted in *ON LIBERTY, UTILITARIANISM, AND OTHER ESSAYS* 5, 15, 18–54 (Mark Philip & Frederick Rosen eds., 2015).

⁸¹ David Aridia, *Beyond the Marketplace of Ideas: Bridging Theory and Doctrine to Promote Self-Governance*, 16 *HARV. L. & POL’Y REV.* 275, 288 (2022).

Chavez v. Gonzales, the Supreme Court explained that Article III, Section 4 was a direct transposition of its American First Amendment counterpart.⁸²

Hence, in a similar vein, the Court in *Reyes v. Bagatsing* expounded on this market dynamic by stating that:

It must always be remembered that this right likewise provides for a safety valve, allowing parties the opportunity to give vent to their-views, even if contrary to the prevailing climate of opinion. For if the peaceful means of communication cannot be availed of, resort to non-peaceful means may be the only alternative. Nor is this the sole reason for the expression of dissent. It means more than just the right to be heard of the person who feels aggrieved or who is dissatisfied with things as they are. Its value may lie in the fact that there may be something worth hearing from the dissenter. That is to ensure a true ferment of Ideas.⁸³

In *Diocese of Bacolod v. COMELEC*, the Court also characterized this marketplace as: “A free, open, and dynamic marketplace of ideas is constantly shaping new ones. This promotes both stability and change where recurring points may crystallize and weak ones may develop.”⁸⁴ Likewise, the Court laid bare the other correlative functions of the freedom of expression as a manner by which human dignity is enhanced, “marker for group identity,” and peaceful means by which individuals in a state may express their dissent.⁸⁵

After all, the Court in *Chavez v. Gonzales* stated that this right to expression in relation to the marketplace of ideas was an indispensable democratic tool for those “who question, who do not conform, who differ.”⁸⁶ In addition, the Court expounded on what the meaningful exercise of this right entails which is the encouragement of the “articulation of unorthodox [or contrarian] view[s], though it be hostile to or derided by others.”⁸⁷

The deference to the principle continues to survive in Philippine jurisprudence as evidenced by the Court’s pronouncement in the 2023 case of *ABS-CBN Corp. v. Ampatuan*. It again concurred with the principle that

⁸² *Chavez*, 545 SCRA at 481.

⁸³ *Reyes*, 125 SCRA at 562.

⁸⁴ *Diocese of Bacolod*, 747 SCRA at 79.

⁸⁵ *Id.* at 80.

⁸⁶ *Chavez*, 545 SCRA at 484.

⁸⁷ *Id.*

flow of speech and expression should not be hindered as the volume of speech would eventually be pitted against contrarian views in the theoretical marketplace and eventually lead to a social adjudication of the (supposed) “truth.”⁸⁸

Thus, the foothold of the principle remains yet to be overturned. The Court holds firm in the belief that the interactions of men and the corresponding acts of discourse are central to ensuring the robustness and vitality of free speech. It is through this back-and-forth trading of ideas that individuals either sharpen or modify their opinions or even abandon them altogether. This collective determination of the truth, ideas, and expression, hence, retains an indispensable social function which the state must protect and uphold.

Notwithstanding the continued adherence to this theory of the “marketplace of ideas,” scholars have raised reasonable criticisms on the analogy’s viability in relation to the freedom of speech and expression.⁸⁹ Some scholars have argued that there is a palpable misapplication of the economic concept as the right finds little similarities to market mechanics which are driven largely by supply and demand.⁹⁰ That in reality there exists little or much fewer constraints on free speech which is a technically unlimited resource.⁹¹ Other scholars also attack the analogy and argue that it serves as a malapropism since the equilibrium of the market is not the arbitration of a deemed objective truth (which may not even exist) but the advancement of social knowledge in general.⁹² Furthermore, other detractors of the market analogy likewise question the credibility of the market in general set against the backdrop of the reality that underlying actors are rarely rational.⁹³

The limitations of the analogy have also been laid bare in the case of *In re Badoy*⁹⁴ where the Court recognized that:

[T]he free marketplace of ideas has never been an equal playing field. Some ideas have had a better chance at piercing and remaining in the public consciousness because of factors such as

⁸⁸ ABS-CBN Corp. v. Ampatuan, G.R. No. 227004, slip op. at 86.

⁸⁹ *Aridia*, *supra* note 81, at 285–86.

⁹⁰ *Id.*

⁹¹ *Id.*

⁹² *Id.* at 286–87.

⁹³ *Id.* at 288.

⁹⁴ *In re Badoy*, A.M. No. 22-09-16-SC, Aug. 15, 2023.

source of information, speed of news drowning out information, and the audience's own weakness in distinguishing between truth and falsity.⁹⁵

Therefore, even the Philippine Supreme Court has explicitly acknowledged how other factors like the internet and social media have distorted the principle.⁹⁶ Particularly, the view that increasing the volume of speech will somehow be able to quash erroneous contrary views as it hinges on the assumption that individuals can discern the truthfulness of ideas.⁹⁷ In the array of information now available, the unfortunate reality is that “[d]ubious and unverified statements are placed on the same footing—if not more—as credible and substantiated information.”⁹⁸ This proclivity to accept even erroneous information is greater when placed in the context that social media now acts as a proverbial echo chamber where only opinions in accordance with one’s own are magnified and disparate views disparaged.⁹⁹

In any event, “the desire to ensure a free trade in ideas is not incompatible with a self-governance centered justification for the protection of speech.”¹⁰⁰ Hence, the imperfection of the analogy does not by itself detract from the underlying spirit of the freedom of speech and expression which aims to provide limitations to the power of both the state and even majoritarian views.¹⁰¹

C. Political Speech and Electoral Integrity

The mantle of protection of Article III, Section 4 includes political speech. In *Diocese of Bacolod v. COMELEC*, the Court defined political speech as “speech ‘both intended and received as a contribution to public deliberation about some issue,’ ‘foster[ing] informed and civic-minded deliberation.’”¹⁰²

Under US jurisprudence, political speech as a form of expression is afforded the “broadest protection” under the First Amendment.¹⁰³ In *Buckley v. Valeo*, the SCOTUS understood that “[d]iscussion of public issues and

⁹⁵ *Id.* at 20.

⁹⁶ *Id.*

⁹⁷ *Id.* at 20–21.

⁹⁸ *Id.* at 21.

⁹⁹ *Id.*

¹⁰⁰ *Aridia*, *supra* note 81, at 309.

¹⁰¹ *Id.*

¹⁰² *Diocese of Bacolod*, 747 SCRA at 86 nn.142, 200–01.

¹⁰³ *Buckley v. Valeo*, 424 U.S. 1, 14 (1976).

debate on the qualifications of candidates are integral to the operation of the system of government established by our Constitution.”¹⁰⁴ Even if no strict definition of political speech was proffered, the Court noted that it included the ability to “engage in the discussion of public issues and vigorously and tirelessly to advocate his own election and the election of other candidates.”¹⁰⁵ Additionally, this kind of political discourse occupies a central role in the American democratic system so that “the electorate may intelligently evaluate the candidates’ personal qualities and their positions on vital public issues before choosing among them on election day.”¹⁰⁶ Likewise, it is the citizenry “who must retain control over the quantity and range of debate on public issues in a political campaign.”¹⁰⁷

As a transposition of the First Amendment,¹⁰⁸ Article III, Section 4 of the Philippine Constitution has similarly received protected status, disavowing most restrictions—especially in the political context.¹⁰⁹ In the case of *Diocese of Bacolod v. COMELEC*, the Philippine Supreme Court concurred as to the preferred status of the right.¹¹⁰ It also acknowledged that the right was of significant consequence in the country’s democratic system as it would shape and influence the choice of leaders who effectively carry out policies on behalf of the electorate.¹¹¹ It is “a direct exercise of the [sic] sovereignty”¹¹² and “is at the core of the freedom of expression.”¹¹³ Hence, according to the Court, this type of speech should not be stifled and instead should be promoted.¹¹⁴

Furthermore, in *ABS-CBN Corp. v. Ampatuan*, the Court drew on the separate concurring opinion of Justice Marvic Leonen and elaborated on the functional importance of political speech:

Speech that enlivens political discourse is the lifeblood of democracy. A free and robust discussion in the political arena allows for an informed electorate to confront its government on a more or less equal footing. Without free speech, the government

¹⁰⁴ *Id.* at 14.

¹⁰⁵ *Id.* at 52.

¹⁰⁶ *Id.* at 53.

¹⁰⁷ *Id.* at 57.

¹⁰⁸ See *Chavez*, 545 SCRA 441.

¹⁰⁹ *Diocese of Bacolod*, 747 SCRA at 69–70.

¹¹⁰ *Id.* at 59.

¹¹¹ *Id.*

¹¹² *Id.* at 59.

¹¹³ *Id.* at 91.

¹¹⁴ *Id.* at 77.

robs the people of their sovereignty, leaving them in an echo chamber of autocracy. Freedom of speech protects the ‘democratic political process from the abusive censorship of political debate by the transient majority which has democratically achieved political power[.]’¹¹⁵

What is clear from these definitions and characterizations is that the scope of political speech encompasses all matters of the state and its governance. It concerns itself with the people’s ability to deliberate and express their concurrence or even discontent with state affairs and even the officials that comprise the system of government. It is a right vigorously safeguarded by the state (and the judiciary) as it is an outgrowth of the sovereignty reposed in the people. Further, this type of speech guarantees the continued functioning of the collaborative democratic system and is part and parcel of the theoretical marketplace of ideas.¹¹⁶

In the electoral context, this ability to express one’s perspective as regards a political issue, platform, or even candidate is a constitutive element of the “right to participate in the electoral process.”¹¹⁷ This form of expression “assure[s] individual self-fulfillment to attain the truth [and] secures participation by the people in social and political decision-making, and [maintains] the balance between stability and change.”¹¹⁸ It is the application of political liberty as any individual is endowed with the ability to participate in the formation of the very government that will for a definite term exercise power over them.¹¹⁹

However, this vital function of expression concerning political matters, issues, and officials has been fundamentally complexed by the advent of social media and technology. Even the Supreme Court has been privy to the contours of change brought by technology in *Disini v. Secretary of Justice*¹²⁰ by stating that:

¹¹⁵ ABS-CBN Corp. v. Ampatuan, G.R. No. 227004, slip op. at 87, Apr. 25, 2023, citing *Nicolas-Lewis v. COMELEC* [hereinafter “*Nicolas-Lewis*”], G.R. No. 223705, 913 SCRA 516, 578, Aug. 14, 2019 (Leonen, J., concurring).

¹¹⁶ *Id.*

¹¹⁷ *Macalintal*, G.R. No. 263590, slip op. at 18, citing *Nicolas-Lewis*, 913 SCRA 516, 547.

¹¹⁸ *Id.*

¹¹⁹ *Id.*

¹²⁰ *Disini v. Sec’y of Justice* [hereinafter “*Disini*”], G.R. No. 203335, 716 SCRA 237, Feb. 11, 2014.

The internet is characterized as encouraging a freewheeling, anything-goes writing style. In a sense, they are a world apart in terms of quickness of the reader's reaction to defamatory statements posted in cyberspace, facilitated by one-click reply options offered by the networking site as well as by the speed with which such reactions are disseminated down the line to other internet users.¹²¹

While the Court in *Disini* explored the challenges in ascribing liability for libel committed online in relation to the Cybercrime Prevention Act,¹²² the problems of speed and reach brought about by the internet and social media continue to apply to political speech. Such that the power of individuals to articulate their respective political outlooks has become so accessible. Through social media, the process for proliferating such viewpoints has become so democratized. Under this model, the plausibility of the theorized marketplace of ideas is now tested with the unprecedented volume of speech and expression.

The case of *Kolin Electronics Co., Inc. v. Taiwan Kolin Corp. Ltd.*¹²³ has even noted that “the internet has turned the world into one vast marketplace.”¹²⁴ Although the Court characterized this online marketplace in the commercial context, the proposition that such online marketplace exists in general for the exchange of ideas has been widely accepted. After all, advertisements of juridical entities regarding their own goods and services are none other than commercial speech.

While the online marketplace analogy may find some sort of credence in understanding the trading of political ideas, reality is less straightforward. In *ABS-CBN Corp. v. Ampatuan*, the Court has also paid regard to another important factor in the online marketplace of ideas—namely, influence—to wit:

In social media, where everyone is an agent of information, either through their own invention or those of others, one has the responsibility to be critical about their statements validating the truth of their factual assertions or the soundness of their opinions. *The power of political speech is assessed based on its ability to gain adherence in the marketplace of ideas. Thus, the more influence a speaker has, the more*

¹²¹ *Id.* at 320.

¹²² *Disini*, 716 SCRA at 315–332.

¹²³ *Kolin Electronics Co., Inc. v. Taiwan Kolin Corp. Ltd.* [hereinafter “*Kolin Electronics*”], 917 Phil. 114, 139 (2021).

¹²⁴ *Id.* at 138.

*powerful their voice is, and necessarily, their responsibility to their audience and the information they share.*¹²⁵

The case of *In re Badoy* likewise echoed the Court's concerns regarding the rise of social media and free speech, with it lamenting that while the advent of social media has in theory propelled the idea of the marketplace of ideas with the increase in the volume of speech, a noticeable trade-off is evident with the correlative decline in the quality of speech and overall discourse.¹²⁶ After all, common individuals publishing their views, political and otherwise, are not effectively bound by traditional safeguards that exist for professional journalists such as ethical standards.¹²⁷ Without these institutional defense mechanisms, the onus of dissecting the legitimacy of speech has been effectively shifted to the individual.¹²⁸

These pronouncements by the Court on the decisive shift in the medium of the theoretical marketplace of ideas reckon with the inescapable fact that discourse online has become consequential. As affirmed by the Court in *ABS-CBN Corp. v. Ampatuan*: "Technological advancements increased mass media's influence on public and government affairs."¹²⁹ In other words, this medium has become a mainstay as the generally accepted forum of deliberation rather than the exception.¹³⁰ And if one were to accept this premise, the conclusion that the ills of this new medium have correspondingly infected the manner and quality of political discourse must also follow.¹³¹

As a result, the dialogues generated by everyday citizens online on matters of public concern, government affairs, and public officials are similarly susceptible to the perennial problems of influence and misinformation. These issues become a critical concern when situated amidst the backdrop of elections. Prevailing political views, even if false, may gain traction to influence voters and even sway election outcomes. For example, one study has even found that the disinformation present on Facebook was crucial in swaying the voting outcomes of the Brexit referendum in 2016.¹³²

¹²⁵ *ABS-CBN Corp. v. Ampatuan*, G.R. No. 227004, slip op. at 94. (Emphasis supplied.)

¹²⁶ *In re Badoy*, A.M. No. 22-09-16-SC, slip op. at 20.

¹²⁷ *Id.*; See also Manzi, *supra* note 79, at 2630.

¹²⁸ *Id.*

¹²⁹ *ABS-CBN Corp. v. Ampatuan*, G.R. No. 227004, slip op. at 27.

¹³⁰ *See id.*

¹³¹ *See id.* at 92, citing RICHARD HASEN, CHEAP SPEECH: HOW DISINFORMATION POISONS OUR POLITICS AND HOW TO CURE IT 20-22 (2022).

¹³² Sun, *supra* note 27, at 369.

Ultimately, these seemingly innocuous views then find great mileage with the potential to affect the manner and direction of a state's governance and policy—far from being trivial matters.

D. Political Deepfakes as Amplifiers of Disinformation

Having established the existence of some of the drawbacks existing in the modern marketplace of ideas, nothing else has been regarded as pernicious as disinformation. In contrast to “misinformation” which refers to mere false information, “disinformation” is “deliberate and includes malicious content.”¹³³ Disinformation is thus “designed to mislead others and is deliberately spread with the intent to manipulate truth and facts.”¹³⁴ Particularly, this relationship between social media and disinformation has been widely accepted and has even been appreciated by the Court in key pieces of jurisprudence.

The case of *Disini* marks one of the first instances wherein the Court grappled with this new phenomenon. In that case, the Court dealt with the constitutionality of Republic Act No. 10175 or the Cybercrime Prevention Act of 2012.¹³⁵ There, the Court conceded the existence of the inevitable challenges in the regulation and enforcement of the law's provisions on cyberlibel.¹³⁶ It also yielded to the realities of expression online and the difficulties in then ascribing liability due to the sheer volume of people online that may be affected or even partake in a seemingly innocuous post.¹³⁷

The *Disini* case not only highlights the incongruent model propounded by the law in regulating cyberlibel, but also paints an effective picture of the speed of information transmission online and its scope of diffusion as likely reaching even different social media platforms.

In 2023, the Court in *ABS-CBN Corp. v. Ampatuan* elaborated on the metes and bounds of its contempt power and the *sub judice* rule.¹³⁸ The same

¹³³ United Nations High Commissioner for Refugees, *Factsheet 4: Types of Misinformation and Disinformation*, UNHCR WEBSITE, at <https://www.unhcr.org/innovation/wp-content/uploads/2022/02/Factsheet-4.pdf> (last accessed Dec. 5, 2024).

¹³⁴ John Palfrey, *Misinformation and Disinformation*, in ENCYCLOPEDIA BRITANNICA, Nov. 18, 2024, at <https://www.britannica.com/topic/misinformation-and-disinformation>.

¹³⁵ *Disini*, 237 SCRA at 299–300.

¹³⁶ *Id.* at 320.

¹³⁷ *Id.* at 324–25.

¹³⁸ *ABS-CBN Corp. v. Ampatuan*, G.R. No. 227004, slip op. at 10–15.

case also touched on the disruption brought about by the internet in regulating the type of information circulated online.¹³⁹ Citing the work of Professor Richard Hasen on disinformation, the Court pertinently declared that the adoption of social media has aided in the spread of misinformation which in turn has affected the conduct of online political campaigns.¹⁴⁰ Specifically, campaigns may leverage “more sophisticated technological tools” to advance some form of “political manipulation.”¹⁴¹ As a result, the changes in the landscape of online campaigning “raise important questions about the conditions of electoral legitimacy and threaten to shake the foundation of democratic governance” as the calculus of voters is undoubtedly affected by the surfeit of political information, albeit false.¹⁴²

The Supreme Court has also grappled with the shortcomings of online expression and the implications of the deterioration of traditional modes of verifying information in the case of *In re Badoy*. The obligation now seemingly rests on the individual to scrutinize the quality of information about online despite the numerous factors which determine the eventual acceptance or rejection of online information by such individual.¹⁴³ Consequently, information even if false may be fervently embraced by individuals because it reinforces or aligns with their existing beliefs.¹⁴⁴ On the flip side, contrarian views are “disregarded.”¹⁴⁵ This contemporary mode of discourse may thus augur the beginning of the demise of the marketplace of ideas. As the clash between ideas is limited or withdrawn altogether, the supposed regulatory feature of the marketplace which can only function with the deliberation of ideas is rendered null entirely.

Amidst this backdrop of crises brought about by misinformation and disinformation enters AI technology, specifically deepfakes.¹⁴⁶ With the technology, virtually any person can leverage existing source media like photos or videos of a person as “raw material for visual storytelling.”¹⁴⁷ As noted by the Centre for Data Ethics and Innovation, as a result of the increasing prominence of deepfakes “we are entering a new chapter in the

¹³⁹ *Id.* at 77.

¹⁴⁰ *Id.* at 92–93, citing HASEN, *supra* note 131.

¹⁴¹ *Id.*

¹⁴² *Id.*

¹⁴³ *In re Badoy*, A.M. No. 22-09-16-SC, slip op. at 21.

¹⁴⁴ Chesney & Citron, *supra* note 6, at 1765.

¹⁴⁵ *In re Badoy*, A.M. No. 22-09-16-SC, slip op. at 21.

¹⁴⁶ See *supra* Part I(B).

¹⁴⁷ Marc Blitz, *Deepfakes and Other Non-Testimonial Falsehoods: When is Belief Manipulation (Not) First Amendment Speech?*, 23 YALE J. L. & TECH 160, 194 (2020).

battle for truth on the internet.”¹⁴⁸ The potential for the technology to exacerbate the spread of disinformation and polarization of society is more alarming when contextualized with its continuous sophistication and ability to create more and more realistic output.¹⁴⁹

Furthermore, these concerns must be intensified as the technology has become cheap and accessible, making it easy for any person to create deepfakes.¹⁵⁰ One company, Eleven Labs, which allows users to create deepfake audio output, peddles its services for the measly cost of USD 5 per month.¹⁵¹

In one study, researchers concluded that one of the “most prevalent tactics in real-world cases of misuse [of GenAI]” was the “[m]anipulation of human likeness and falsification of evidence” with the “discernible intent to influence public opinion, enable scam or fraudulent activities, or to generate profit.”¹⁵² Therefore, there is merit in the collective apprehension regarding the adoption of deepfakes, which is a form of GenAI,¹⁵³ as it is likely to be misapplied to commit a variety of crimes or at the very least unethical acts. As the objective to mislead is a common thread in these kinds of acts, deepfakes are becoming cogent amplifiers of “disinformation.”¹⁵⁴

¹⁴⁸ Centre for Data Ethics and Innovation, *supra* note 13, at 3.

¹⁴⁹ Christina P. Walker, Daniel S. Schiff, & Kaylyn Jackson Schiff, *Merging AI Incidents Research with Political Misinformation Research: Introducing the Political Deepfakes Incidents Database*, 38(21) PROCEEDINGS OF THE AAAI CONFERENCE ON ARTIFICIAL INTELLIGENCE 23053 (2024), at <https://ojs.aaai.org/index.php/AAAI/article/view/30349> (last accessed Dec. 5, 2024); A.W. Ohlheiser, *AI-generated video is here to awe and mislead*, VOX, Feb 22, 2024, at <https://www.vox.com/technology/24079459/sora-openai-video-tool-world-simulator>; Marchal et al., *supra* note 28, at 16–17; Prashnu Verma & Will Oremus, *AI Voice Clones Mimic Politicians and Celebrities, Reshaping Reality*, WASH. POST, Oct. 13, 2023, at <https://www.washingtonpost.com/technology/2023/10/13/ai-voice-cloning-deepfakes>.

¹⁵⁰ The Guardian, *The Guardian View on Political Deepfakes: Voters Can't Believe Their Own Eyes*, GUARDIAN, Feb. 19, 2024, at www.theguardian.com/commentisfree/2024/feb/19/the-guardian-view-on-political-deepfakes-voters-cant-believe-their-lying-eyes; Mark Scott, *Deepfakes, Distrust and Disinformation: Welcome to the AI election*, POLITICO, Apr. 16, 2024, at <https://www.politico.eu/article/deepfakes-distrust-disinformation-welcome-ai-election-2024/>; Andrew Ray, *Disinformation, Deepfakes and Democracies: The Need for Legislative Reform*, 44 UNSW L.J. 983, 986 (2021).

¹⁵¹ Verma & Oremus, *supra* note 149.

¹⁵² Marchal et al., *supra* note 28, at 4.

¹⁵³ See *supra* Part I(B).

¹⁵⁴ See John Palfrey, *Misinformation and Disinformation*, in ENCYCLOPEDIA BRITANNICA, Nov. 18, 2024, at <https://www.britannica.com/topic/misinformation-and-disinformation>.

The use of deepfakes has also found itself becoming a mainstay as a tool in political campaigns and elections.¹⁵⁵ One study has found that political disruption, opinion manipulation, and defamation are just some of the more prevalent objectives of the creators of deepfakes.¹⁵⁶ Numerous deepfakes have also already been made of political figures like “[American Presidents] Donald Trump, Joe Biden, [French President] Emmanuel Macron, [US Congresswoman] Nancy Pelosi, [Russian President] Vladimir Putin, and [Ukrainian President] Volodymyr Zelenskyy.”¹⁵⁷ Other political figures who have fallen victim to deepfakes include Belgian Prime Minister Sophie Wilmès and Indian politician Manoj Tiwari.¹⁵⁸

In the New Hampshire primaries, prospective voters were plagued with robocalls from an anonymous source with the eerily distinct voice of incumbent US President Joe Biden which discouraged them from participating.¹⁵⁹ Moldovan President Maia Sandu has also been another victim of deepfakes as they have been used to launch personal attacks against her and her administration.¹⁶⁰ For Belgian Prime Minister Sophie Wilmès, she was depicted “giving a fictitious speech about the link between COVID-19 and climate change, which was widely shared on social media.”¹⁶¹ Not immune to this trend, even Philippine President Ferdinand Marcos Jr. was a victim of a deepfake which made it appear as if he was ordering a military attack.¹⁶²

While a concededly small sample size, these examples demonstrate the growing trend that deepfakes can and will be exploited for political gain by organizations, individuals, or even foreign governments.¹⁶³

Admittedly, while the utilization of negative propaganda against political opponents is not a novel phenomenon, the technology complicates

¹⁵⁵ Walker, Schiff, & Schiff, *supra* note 149.

¹⁵⁶ Marchal et al., *supra* note 28, at 11–13.

¹⁵⁷ Walker, Schiff, & Schiff, *supra* note 149.

¹⁵⁸ Andrew Ray, *supra* note 150, at 986.

¹⁵⁹ Vittoria Elliott & Makena Kelly, *The Biden Deepfake Robocall is Only the Beginning*, WIRED, Jan. 23, 2024, at <https://www.wired.com/story/biden-robocall-deepfake-danger/>.

¹⁶⁰ Mark Scott, *Deepfakes, Distrust and Disinformation: Welcome to the AI election*, POLITICO, Apr. 16, 2024, at <https://www.politico.eu/article/deepfakes-distrust-disinformation-welcome-ai-election-2024/>.

¹⁶¹ Andrew Ray, *supra* note 150, at 986.

¹⁶² Bea Cupin, *Malacañang flags deepfake audio of Marcos ordering military attack*, RAPPLER, Apr. 24, 2024, at <https://www.rappler.com/philippines/malacanang-flags-deepfake-audio-marcos-ordering-military-attack-april-2024/>.

¹⁶³ Scott, *supra* note 160.

this practice by effectively undermining the credibility of opponents with generally realistic but manipulated media.¹⁶⁴ Hence, dubious media created by deepfake technology can be employed to accuse opponents of expressing fictitious statements, being in dubious scenarios, committing questionable acts, or being involved in sexual scandals.¹⁶⁵ The nefarious use of the technology against adversaries is thus only limited by the malicious creator's imagination.

As individual arbiters of truth, individual voters are now faced with the burdensome task of deciding truth from fiction with the caveat that their own senses cannot now be trusted.¹⁶⁶ Experts have touted this phenomenon as the “epistemic apocalypse” as there is a decisive “blurring of the boundaries between real and fake [media or news].”¹⁶⁷ It is a seemingly ridiculous but unfortunate conundrum that individuals must now grapple with manipulated pictures and videos that may proffer misleading evidence which can ultimately affect their views on politicians, their respective platforms, and directly influence their votes on the fateful election day.¹⁶⁸ Thus, the harm on the electoral process is far from imagined as the scale and mode of deception have been weaponized and “turbocharged” with deepfakes.¹⁶⁹

These momentous consequences on the electoral system and democracy are in fact concrete as evidenced by the loss of Slovakian opposition leader Michal Šimečka.¹⁷⁰ In the days leading up to the election, a fictitious audio tape was leaked which made it appear as if the candidate intended to rig votes by bribing members of the country's Roma population and raising the cost of beer.¹⁷¹ As a result, Šimečka was defeated and a pro-

¹⁶⁴ Matthew D. Weiner, *Destined to Deceive: The Need to Regulate Deepfakes with a Foreseeable Harm Standard*, 122 MICH. L. REV. 771, 776 (2024).

¹⁶⁵ Walker, Schiff, & Schiff, *supra* note 149; Tashman, *supra* note 14, at 1393.

¹⁶⁶ The Guardian, *supra* note 150.

¹⁶⁷ Ferdinand Gehringer, Christopher Nehring, & Mateusz Łabuz, *The Influence of Deep Fakes on Elections: Legitimate Concern or Mere Alarmism?*, Konrad Adenauer Stiftung, 5 (May 2024), at <https://www.kas.de/documents/d/guest/the-influence-of-deep-fakes-on-elections>.

¹⁶⁸ See Scott, *supra* note 160; Tashman, *supra* note 14, at 1393.

¹⁶⁹ See Marchal et al., *supra* note 28, at 14; The Guardian, *supra* note 150; Tashman, *supra* note 14, at 1393.

¹⁷⁰ Bond, *supra* note 27; Curt Devine, Donie O'Sullivan, & Sean Lyngaas, *A Fake Recording of a Candidate Saying He'd Rigged the Election Went viral. Experts Say it's Only the Beginning*, CNN, Feb. 1, 2024, at <https://edition.cnn.com/2024/02/01/politics/election-deepfake-threats-invs/index.html>.

¹⁷¹ Bond, *supra* note 27; Devine, O'Sullivan, & Lyngaas, *supra* note 170; Verma & Oremus, *supra* note 149.

Russia opponent secured their victory even if the audio was decisively debunked.¹⁷² Another example of the palpable use of deepfakes to affect elections can be found in Taiwan where deepfakes were created to churn out political endorsements in favor of candidates during the 2024 presidential elections.¹⁷³ In India, deepfake media of known celebrities Aamir Khan and Ranveer Singh were also generated to criticize Prime Minister Narendra Modi during the country's election.¹⁷⁴ In Indonesia, deepfakes were also heavily employed by winning candidate and now-President Prabowo Subianto to rectify his former draconian reputation, reinvent his political persona, and secure wide appeal from younger voters.¹⁷⁵

All in all, these examples demonstrate the very real damage that deepfakes may present when weaponized during elections. This form of attack may be especially effective when there is little room or time to fact-check these media.¹⁷⁶

While proponents of deepfakes and AI technology in general may quickly dismiss concerns as overblown and propound a counterargument by placing a reasonable degree of trust in the ability of individuals to distinguish between real and manipulated media, research paints a bleak picture. As to who is likely to disseminate deepfakes, one study points to individuals “with higher levels of political interest” as likely purveyors of disinformation, even if unwittingly.¹⁷⁷ A feedback loop of sorts is thus formed: these same individuals who are likely exposed to deepfakes are in turn likely to disperse these deepfakes even if false.¹⁷⁸

And contrary to the preconceived notion that individuals who possess a high-level of political interest are better than peers in filtering out disinformation, the research has yielded an opposite conclusion.¹⁷⁹ These kinds of individuals have been shown to “rely on intuition instead of analytic

¹⁷² Devine, O’Sullivan, & Lyngaas, *supra* note 170; Verma & Oremus, *supra* note 149.

¹⁷³ Martinez, *supra* note 4.

¹⁷⁴ *Id.*

¹⁷⁵ Rumman Chowdhury, *AI-fuelled Election Campaigns are Here — Where are the Rules?*, NATURE, Apr. 9, 2024, at <https://www.nature.com/articles/d41586-024-00995-9>.

¹⁷⁶ Chesney & Citron, *supra* note 6, at 1778; The Guardian, *supra* note 150.

¹⁷⁷ Saifuddin Ahmed, *Who inadvertently shares deepfakes? Analyzing the role of Political Interest, Cognitive Ability, and Social Network Size*, 57 TELEMATICS & INFO. 1, 3 (2021).

¹⁷⁸ *Id.*

¹⁷⁹ *Id.*; Matthew Groh et al., *Human detection of political speech deepfakes across transcripts, audio, and video*, 15 NAT. COMM’N 7629, 2 (2024), at <https://www.nature.com/articles/s41467-024-51998-z>.

thinking” and thus still are likely targets of political deepfakes.¹⁸⁰ Even more worrying, social media companies, the very forums of disinformation, have failed to catch up to these advancements, specifically in deepfake audio technology, as their “human fact-checkers often have trouble spotting fakes.”¹⁸¹ These sources of false content consequently find themselves spreading at lightning speed due to algorithmic propensities to magnify popular but false content.¹⁸²

As a result, online disinformation networks continue to proliferate and become even more effective at spewing false and malicious information. The marketplace of ideas, which is supposed to regulate informational exchanges, becomes stunted or impotent altogether as groupthink and confirmation biases are reinforced with believable (albeit false) deepfake media.¹⁸³ Simply put, “[d]eep fakes will allow individuals to live in their own subjective realities, where beliefs can be supported by manufactured ‘facts.’”¹⁸⁴

For nefarious political aspirants, deepfakes also signify an opportunity to exploit what has been touted as “the liar’s dividend.”¹⁸⁵ For example: “A person accused of having said or done something might create doubt about the accusation by using altered video or audio evidence that appears to contradict the claim.”¹⁸⁶ On the flip side, aspersions can be casted on genuine videos or audios to gain a political advantage.¹⁸⁷ In either scenario, citizens suffer as the truth of the matter in public affairs is denied from them. The accountability mechanism brought about by discourse that citizens hold against the state becomes futile.

These deceptive deepfakes can thus not only undermine the credibility of the deepfake’s subject but can also foster a culture of distrust as regards even legitimate media due to suspicions that the latter may have been manipulated.¹⁸⁸ Political discourse as a result becomes skewed, and its quintessential democratic function eroded as citizens are powerless against

¹⁸⁰ Groh et al., *supra* note 179.

¹⁸¹ Verma & Oremus, *supra* note 149.

¹⁸² Ullrich K. H. Ecker et al., *The psychological drivers of misinformation belief and its resistance to correction*, 38 NAT. REV. PSYCHOL. 13, 15 (2022).

¹⁸³ Chesney & Citron, *supra* note 6, at 1777–78.

¹⁸⁴ *Id.* at 1778.

¹⁸⁵ *Id.* at 1785–86.

¹⁸⁶ *Id.* at 1785.

¹⁸⁷ *Id.*

¹⁸⁸ Weiner, *supra* note 164, at 777; Chesney & Citron, *supra* note 6, at 1785.

this kind of technology, which at present possesses little to no private or public guardrails.¹⁸⁹ The marketplace no longer becomes an even playing field as its participants (the citizenry) are no longer able to discern and agree upon truth, fact, or empirical data.¹⁹⁰

This inability to agree on underlying assumptions directly affects political speech and expression as “[p]olitical debates are meant to hinge on contested interpretation of facts, or facts contextualized differently by competing values, but not on alternative facts.”¹⁹¹ Without an agreement as to the fundamental premises of a topic of debate or issue, no meaningful exchange can ensue regarding the acceptance or rejection of an idea (political or otherwise). Instead of facilitating an exchange on the different sides of an issue and grappling with conflicting value judgments to come to a consensus as a societal collective, the conversation fails to move forward as it becomes enmeshed in arguing about what are instead considered underlying “truths.”¹⁹²

It becomes less of an exchange about the merits and drawbacks of a particular political ideology or prescient issue but an indignant diatribe about absolute truths, merits, and norms. The effect of this type of unmoderated exchange is the further polarization of political ideologies and in turn the very people that espouse them.

The authenticity or genuineness of political discussions and decisions may be cast into doubt as consent of the electorate may be arguably vitiated by deepfakes, and yet end-results like election outcomes and legislation will still enjoy a veneer of legitimacy as processes were technically enforced and observed. Hence, political deepfakes pose imperceptible and at the same time lethal threats to the democratic system.

The imminence and growing centrality of deepfakes as a tool in a political arsenal beckon towards the need for regulation in the Philippines. The International Criminal Police Organization (Interpol) has even provided a warning that the upcoming 2025 midterm elections may see a swell in the use of deepfakes which “can be weaponized to trick voters.”¹⁹³ In one news

¹⁸⁹ Verma & Oremus, *supra* note 149; Locus, *supra* note 4.

¹⁹⁰ Chesney & Citron, *supra* note 6, at 1777.

¹⁹¹ Bennett & Livingston, *supra* note 28, at 9.

¹⁹² *Id.*

¹⁹³ Rosales, *supra* note 4; Locus, *supra* note 4; Martinez, *supra* note 4; Pascual, *supra* note 4.

article, Penny Chai, Vice President for business development in the Asia Pacific for Sumsu (“an identity verification platform”), raised alarms about the hike and corresponding misuse of deepfakes in the Philippines.¹⁹⁴ In relation to identity fraud, the company saw “more than 1,430 percent increase in the Southeast Asia region due to deepfakes. Very closely to that, in Philippines[,] there [is] 4,400 percent growth between 2022-2023[.]”¹⁹⁵ She professed that many individuals became victims of these deepfakes and were thus misled to make purchases or investments.¹⁹⁶

At present, the country has already been touted as “a hotbed for online disinformation” and yet there exists no current legislation regulating the use of deepfakes especially in the political context.¹⁹⁷ The evidence regarding the plausibility of harm brought about by deepfakes continues to escalate. Thus, legislation must come at this crucial time or else ensuring “free, orderly, honest, peaceful, and *credible* elections”¹⁹⁸ and giving “effect to the expressed will of the majority”¹⁹⁹ will be relegated to mere constitutional abstracts.

III. THE CURRENT PHILIPPINE REGULATORY LANDSCAPE

Before formulating a decisive plan to tackle the rise and use of deepfakes in the political arena, a synthesis of existing regulatory materials must first be surveyed to comprehensively understand the contours of the legal system, situate the legality of the technology and its applications, and expose the gaps that must be filled by prospective legislation.

First, an inquiry must begin with basic definitions such as “partisan political activity” and “election propaganda.” Afterwards, existing sources of obligations such as COMELEC issuances must be surveyed as they impose tangible duties on stakeholders, some of which are even unconstitutional or *ultra vires*. Further, a discussion on what acts may be considered election offenses in the genesis and diffusion of political deepfakes is mandatory as they are regulated by existing laws such as the OEC, Fair Election Act, and

¹⁹⁴ Michelle Ong, *Deepfakes seen driving surge in identity fraud in PH*, ABS-CBN NEWS, Sept. 13, 2024, at <https://news.abs-cbn.com/business/2024/9/13/deepfakes-seen-driving-surge-in-identity-fraud-in-ph-1056>.

¹⁹⁵ *Id.*

¹⁹⁶ *Id.*

¹⁹⁷ Rosales, *supra* note 4; Locus, *supra* note 4.

¹⁹⁸ CONST. art. IX-C, § 4. (Emphasis supplied.)

¹⁹⁹ *Yason v. COMELEC*, G.R. No. 52713, 134 SCRA 371, 379, Jan. 31, 1985.

several COMELEC Resolutions. Lastly, the merits of existing bills lodged in the legislature must be scrutinized to obtain an understanding of the preliminary legislative sentiment surrounding the technology. These bills are crucial as they may provide an insight into prospective legal standards, hence the constituent provisions must be assessed.

A. Partisan Political Activity & Election Propaganda

An examination of the definitions of “partisan political activity” and “election propaganda” is a prerequisite in situating COMELEC’s regulatory power over political deepfakes as it may only regulate speech generated by aspiring candidates and political parties.²⁰⁰ However, it may regulate private political speech in exceptional cases and subject to stringent limitations.²⁰¹

Under the OEC, “partisan political activity” is defined as “an act designed to promote the election or defeat of a particular candidate or candidates to a public office[.]”²⁰² This includes a broad range of activities in the exercise of the constitutional rights to organization, assembly, and free speech—the common element being that such act is done for or against a candidate. Acts that constitute “partisan political activity,” for instance, would include “forming organizations,” “making speeches,” or publishing or distributing campaign literature or materials.” However, note that speech regarding future elections or prospective nominees fall outside the provision’s ambit.²⁰³

While denominated as “digital election campaigning,” COMELEC Resolution No. 11064 by implication has included deepfakes as a form of election propaganda under “digital election campaign.” Generally, this refers to conducting an election campaign online for or against a political candidate or to sway voter behavior. Notably, an enumerated act is that of creating audiovisual content to conduct campaigns or related partisan political activity, which places political deepfakes within the provision’s purview.²⁰⁴

The Resolution builds on existing guidelines of campaigns and political activity but significantly expands and recognizes the variety of content forms that now exist online and have become crucial in permeating online voter networks.

²⁰⁰ *Diocese of Bacolod*, 747 SCRA at 114.

²⁰¹ *Id.* at 115.

²⁰² ELECT. CODE, § 79(b).

²⁰³ ELECT. CODE, § 79(b).

²⁰⁴ COMELEC Res. No. 11064, art. II(14). (Emphasis supplied.)

Whereas “lawful election propaganda” is defined through the provision of a list which the law states, including:

(d) All other forms of election propaganda not prohibited by this Code as the Commission may authorize after due notice to all interested parties and hearing where all the interested parties were given an equal opportunity to be heard: Provided, That the Commission's authorization shall be published in two newspapers of general circulation throughout the nation for at least twice within one week after the authorization has been granted.²⁰⁵

In connection, the Fair Election Act also provides a definition of “lawful election propaganda,” as:

Sec. 3. Lawful Election Propaganda. – Election propaganda whether on television, cable television, radio, newspapers *or any other medium* is hereby allowed for all registered political parties, national, regional, sectoral parties or organizations participating under the party-list elections and for all bona fide candidates seeking national and local elective positions subject to the limitation on authorized expenses of candidates and political parties, observance of truth in advertising and to the supervision and regulation by the Commission on Elections (COMELEC). For the purpose of this Act, lawful election propaganda shall include:

* * *

3.5. *All other forms of election propaganda not prohibited by the Omnibus Election Code or this Act.*²⁰⁶

COMELEC Resolution No. 10730 also provides a definition of “political advertisement” or “election propaganda” as:

[A]ny matter broadcasted, published, printed, displayed or exhibited, in any medium, which contains the name, image, logo, brand, insignia, initials, and other symbol or graphic representation that is capable of being associated with a candidate, and is exclusively intended to draw the attention of the public or a segment thereof to promote or oppose, directly or indirectly, the election of the said candidate or candidates to a public office.

²⁰⁵ ELECT. CODE, § 82(d).

²⁰⁶ Rep. Act No. 9006 (2001), § 3. (Emphasis supplied.)

* * *

Political advertising includes endorsements, statements, declarations, or information graphics, appearing on any internet website, social network, blogging site, and micro-blogging site, which – when taken as a whole – has for its principal object the endorsement of a candidate only, or which were posted in return for consideration or are otherwise capable of pecuniary estimation.²⁰⁷

The Resolution then builds on the definition of “lawful election propaganda” as:

Election propaganda, whether on television or cable television, radio, newspaper, the internet or any other medium, is hereby allowed for all bona fide candidates seeking national and local elective positions, subject to the limitation on authorized expenses of candidates and parties, observation of truth in advertising, and to the supervision and regulation by the COMELEC. Lawful election propaganda shall include:

* * *

e. All other forms of election propaganda not prohibited by the Omnibus Election Code or these rules, such as but not limited to:

1. Social media posts, regardless of format, whether original or reposted from some source, which may either be incidental to the poster’s advocacies of social issues or which may have, for its primary purpose, the endorsement of a candidate only;
2. Paid advertisements in broadcast, internet, mobile, print or outdoor media subject to the requirements set forth in Section 9 hereof and the Fair Elections Act[.]²⁰⁸

The definition in Resolution No. 10730 primarily echoes the definitions provided in the OEC and Fair Election Act. Additional provisions in Section 6(e) are mere expansions of what are considered election propaganda in light of the wide adoption of social media.

²⁰⁷ COMELEC Res. No. 10730 (2021), § 1(16). (Emphasis supplied.)

²⁰⁸ § 6(e)(1)–(2).

From these foregoing definitions, it is possible that the act of creating deepfake media (audio or video) may fall within the definition of “partisan political activity” when the impetus behind its creation is to spur the “election or defeat” of a candidate in relation to a specific public office.²⁰⁹ The proliferation of the political deepfake may fall under Section 79(b)(4) or 79(b)(5) as the political deepfake may be published as a form of campaign literature in a visual format or even in its various iterations or as published online can be used to garner support for a candidate.²¹⁰

In Resolution No. 1106, the inclusion of the phrase “and other internet-based tools” in enumerating digital election campaigns also contemplates the inclusion of deepfakes which are crafted using typically internet-based AI tools.²¹¹ Functionally, the list of what are considered as digital election campaigning implies that the output of these tools or media are election propaganda because their purpose is related to the election (or defeat) of a candidate which is the primary objective of a political campaign.²¹² It also explicitly concedes that these kinds of media have the capacity to “influence voter behavior.”²¹³ The breakdown also includes the different forms by which deepfakes may appear (i.e., visual or audio) and concedes that they may be used as “campaign literature or materials” to promote or defeat the election of a candidate.²¹⁴

Political deepfakes can also fall within the definition of “lawful election propaganda” as it is conceivable for these altered visual media output to be distributed in print format on pamphlets, posters, and other similar material.²¹⁵ Furthermore, even in its digital form, these can be considered as “lawful election propaganda” as it is “not prohibited by [the Omnibus Election] Code” assuming that it is also authorized by COMELEC, and after due process and publication of any decision by the Commission.²¹⁶ This interpretation is supported by the fact that there is no prohibition as to the use of deepfakes by political candidates and parties by COMELEC.²¹⁷

²⁰⁹ ELECT. CODE, § 79(b).

²¹⁰ § 79(b)(4)-(5).

²¹¹ *See supra* Part I(B).

²¹² § 79(b).

²¹³ COMELEC Res. No. 11064, art. II(14).

²¹⁴ Art. II(14).

²¹⁵ ELECT. CODE, § 82(a), (c), (d).

²¹⁶ § 82(d).

²¹⁷ COMELEC Res. No. 10730 (2021), § 6(e).

Further, Resolution No. 10730 does not distinguish as to the medium of publication, broadcast, display, or exhibition of “political advertisement” or “election propaganda.”²¹⁸ It also in no uncertain terms states that the definition includes any type of endorsement or information graphics appearing online when the purpose is the endorsement of a candidate.²¹⁹ Hence, deepfakes appearing online (in social media or blogging sites) that have for their primary purpose the election or promotion of a candidate are considered “political advertisements” or “election propaganda.” The definitions provided in the Resolution sweep broadly enough to contemplate the dissemination of political deepfakes virtually or physically and do not distinguish as to format (e.g. audio or visual) as long as they are disseminated in the social media platforms and posts of the political candidate or political party.²²⁰ This theory is indeed supported by Resolution No. 11064 where COMELEC expressed that social media content and paid online ads were election propaganda within its regulatory purview.²²¹

Likewise, deepfakes may fall under the list of “prohibited forms of election propaganda” as it may include manipulated visual media which does not comply with the disclosure or size requirements under the law.²²² Furthermore, political deepfakes in an audiovisual format could also fall within the definition of “prohibited forms of election propaganda” as they may be employed as a “means of cinematography” with the candidate as the subject, except “telecasts” which are permissible under the Omnibus Election Code.²²³

However, the last paragraph in Section 79 of the OEC clarifies that political speech made by the public on matters related to the election such as opinions, issues, or even candidates fall beyond the ambit of the provision.²²⁴ The Fair Election Act also stresses this limitation and makes an important amendment to the provision in the OEC by stating that the “lawful election propaganda” contemplated under the law only refers to those dispersed by “registered political parties, national, regional, sectoral parties or organizations participating under the party list elections and for all bona fide

²¹⁸ § 1(16).

²¹⁹ § 1(16).

²²⁰ § 6(e)(1).

²²¹ COMELEC Res. No. 11064, pmb. ¶ 10.

²²² ELECT. CODE, § 84 & 85(a)-(b); COMELEC Res. No. 10730 (2021), § 7.

²²³ ELECT. CODE, § 84 & 85(d).

²²⁴ ELECT. CODE, § 79.

candidates seeking national and local elective positions.”²²⁵ Considering the technological advancements such as the adoption of social media in campaign activities, COMELEC also affirms that these personal political views online in “blogs and micro-blogs” (i.e., private political opinions and views online) do not fall within its scope of regulation.²²⁶

The implication of this proviso becomes notable as deepfakes with political motives that either promote or denigrate a candidate or foster discussion on political issues are considered unregulated private political speech as it constitutes a form of social advocacy.²²⁷ Drawing on jurisprudence, these kinds of political speech (although damaging and deceptive) are afforded protection by the Constitution.²²⁸ After all, it is believed that these kinds of false speech will be eked out and eventually rejected by the marketplace of ideas after collaborative deliberation.²²⁹

However, the Court in *Diocese of Bacolod v. COMELEC* has clarified that in limited scenarios, private political speech may be subject to COMELEC’s regulatory scope when such amounts to “election paraphernalia.”²³⁰ Specifically, the Court laid down the following guidelines for the regulation of private political speech which takes the form of election paraphernalia:

Regulation of election paraphernalia will still be constitutionally valid if it reaches into speech of persons who are not candidates or who do not speak as members of a political party if they are not candidates, only if what is regulated is declarative speech that, taken as a whole, has for its principal object the endorsement of a candidate only. The regulation (a) should be provided by law, (b) reasonable, (c) narrowly tailored to meet the objective of enhancing the opportunity of all candidates to be heard and considering the primacy of the guarantee of free expression, and (d) demonstrably the least restrictive means to achieve that object. The regulation must only be with respect to the time, place, and manner of the rendition of the message. In no situation may the speech be prohibited or censored on the basis of its content. For

²²⁵ Rep. Act No. 9006, § 3; *See* *St. Anthony Coll. of Roxas City, Inc. v. COMELEC* [hereinafter “*St. Anthony Coll.*”], G.R. No. 258805, slip op. at 17, Oct. 10, 2023.

²²⁶ COMELEC Res. No. 10730 (2021), § 1(4).

²²⁷ *Diocese of Bacolod*, 747 SCRA at 115; *See also St. Anthony Coll.*, G.R. No. 258805, Oct. 10, 2023.

²²⁸ *Id.* at 114–15.

²²⁹ *See supra* Part II(B)-(C).

²³⁰ *Id.* at 115.

this purpose, it will not matter whether the speech is made with or on private property.²³¹

Based on the Court's pronouncement, the determining factor of COMELEC's regulatory scope does not just rest on the source of the utterance (i.e., candidate, political party, or private citizen) but on the purpose of such. This standard is consistent with the OEC's definition of partisan political activity²³² and parallels existing definitions and examples of election propaganda under existing laws and COMELEC Resolutions.²³³

B. COMELEC Regulatory Guidelines on Social Media and Artificial Intelligence

As the interest of "ensuring free, orderly, honest, peaceful, and credible elections" has been deemed a paramount state interest, COMELEC has been granted certain powers "during the election period."²³⁴ It is granted these abilities in order to "enforce and administer all laws and regulations relative to the conduct of an election, a plebiscite, an initiative, a referendum, and a recall."²³⁵ COMELEC is also empowered to promulgate rules in order to implement the Omnibus Election Code and other election laws.²³⁶

Particularly in relation to election propaganda and political advertisements:

The Commission may, during the election period, supervise or regulate the enjoyment or utilization of all franchises or permits for the operation of transportation and other public utilities, media of communication or information, all grants, special privileges, or concessions granted by the Government or any subdivision, agency, or instrumentality thereof, including any government-owned or controlled corporation or its subsidiary. Such supervision or regulation shall aim to ensure equal opportunity, time, and space, and the right to reply, including reasonable, equal rates therefor, for public information campaigns

²³¹ *Id.*; See also *St. Anthony Coll.*, G.R. No. 258805, slip op. at 12–13, Oct. 10, 2023.

²³² ELECT. CODE, § 79(b).

²³³ COMELEC Res. No. 11064, art. II(14); Rep. Act No. 9006, § 3; COMELEC Res. No. 10730 (2021), §§ 1(16), 6.

²³⁴ CONST. art. IX-C, § 4.

²³⁵ Art. IX-C, § 2(1).

²³⁶ ELECT. CODE, § 52(c).

and forums among candidates in connection with the objective of holding free, orderly, honest, peaceful, and credible elections.²³⁷

In *Osmeña v. COMELEC*,²³⁸ the Court clarified the meaning of the provision, to wit: “In effect, during the election period, the COMELEC [temporarily] takes over the advertising page of newspapers or the commercial time of radio and TV stations and allocates these to the candidates.”²³⁹ This power of COMELEC was granted by the Constitution to ensure that political candidates all had “equal time and space.”²⁴⁰

To realize and effectuate its obligations in relation to elections, COMELEC has been explicitly granted the authority to promulgate rules and regulations for the implementation of the Fair Election Act.²⁴¹ This power is to be circumscribed only by the limitations enshrined in the Constitution, Omnibus Election Code, and other election laws.²⁴² Furthermore, any regulations enacted must be in accordance with the “prescribed procedure” outlined by the source law and “must be reasonable.”²⁴³

1. Resolution No. 10730

Laying down the premises that COMELEC possesses the requisite authority to enact rules and regulations in relation to election propaganda and political advertisements, it enacted Resolution No. 10730 on November 17, 2021.²⁴⁴ The Resolution is of particular note as it includes regulations regarding social media use by political candidates and parties. The COMELEC first defines “mass media” as:

[D]iversified technologies, operating on various platforms, that have for their primary purpose the transmission of information and communication to a large audience. These platforms include broadcast, internet and mobile, print, and outdoor. “Mass Media Entities” refer to individuals and organizations that exercise control over these technologies and determine, whether directly

²³⁷ CONST. art. IX-C, § 4.

²³⁸ G.R. No. 132231, 288 SCRA 447, Mar. 31, 1998.

²³⁹ *Id.* at 468.

²⁴⁰ *Id.* at 469; CONST. art. IX-C, § 4.

²⁴¹ Rep. Act No. 9006, § 13.

²⁴² § 13; *See also* *Lokin v. COMELEC* [hereinafter *Lokin*], G.R. No. 179431, 621 SCRA 385, 403, June 22, 2010.

²⁴³ *Lokin*, 621 SCRA at 403.

²⁴⁴ COMELEC Res. No. 10730 (2021), Nov. 17, 2021.

or indirectly, the content being distributed using these technologies. *“Social media” is a form of mass media.*²⁴⁵

The inclusion of “social media” as a form of “mass media” is a direct recognition that these technologies are indispensable platforms for candidates to disseminate information about their platforms and campaigns. The inclusion of social media in the definition also explicitly puts the technology within COMELEC’s regulatory reach as it possesses the power to regulate election propaganda proliferated through mass media.²⁴⁶

Another notable definition in the Resolution is “online campaigning” or “the use of the internet to distribute campaign propaganda. This includes text-only posts on social media, pictures, audio clips, and video clips, regardless of duration, and all combinations of such formats.”²⁴⁷ As the definition again provides a broad scope as to what kinds of media may be considered as used in “online campaigning,” political deepfakes which may take different forms (audio or visual) are recognized as tools for “online campaigning.”²⁴⁸

On the guidelines for the usage of election propaganda online, Resolution No. 10730 provides the following guidelines for “internet, mobile and social media propaganda” which are notable expansions of Section 86 of the OEC²⁴⁹ and Section 4 of the Fair Election Act:²⁵⁰

c. Internet, mobile and social media propaganda

The use of the internet, mobile platforms, and social media for purposes of election propaganda shall be allowed subject to the following guidelines:

1. Each registered political party/coalitions and candidate shall register with the Education and Information Department of the COMELEC, the website name and web address of all platform verified official accounts, websites, blogs and/or other social media pages of such political party or candidate

²⁴⁵ § 1(11). (Emphasis supplied.)

²⁴⁶ ELECT. CODE, § 86.

²⁴⁷ COMELEC Res. No. 10730, § 1(13).

²⁴⁸ § 1(13).

²⁴⁹ ELECT. CODE, § 86. This refers to the regulation of election propaganda through mass media.

²⁵⁰ Rep. Act No. 9006, § 4. This refers to the requirements for published or printed and broadcast election propaganda.

within thirty (30) days from the last day of the period for the filing of the Certificates of Candidacy. Websites completing the verification process after the said period and other social media accounts established after the said period must be registered with the COMELEC-EID within five (5) days from its verification or registration.

2. Any other website, blog, or social media page not registered above but which, when taken as a whole, has for its primary purpose the endorsement of a candidate, whether or not directly maintained or administered by the candidate or their official campaign representatives, shall be considered additional official websites, blogs or social media pages of the said candidate, for all regulatory purposes.
3. Only verified accounts, websites, blogs, and/or social media pages may run electoral ads, and boost or promote electoral posts.
4. Microtargeting of electoral ads shall not be allowed provided that electoral ads can be targeted using only the following criteria: geographical location, except radius around a specific location; age; and gender; provided further that contextual targeting options may also be used in combination with the above-mentioned criteria.
5. Information contained in online campaign propaganda shall be truthful and not misleading, nor shall it tend to unjustifiably cast doubt on the integrity of the electoral process.
6. All electoral ads must show a disclosure that identifies who paid for the ad. All electoral posts must show a disclosure that identifies it as a paid electoral ad, and discloses who paid for the ad.²⁵¹

Much like predecessor provisions in the OEC and Fair Election Act, these provisions in the Resolution have adapted to the specific contours of social media usage. Hence, COMELEC has required the registration of social media sites where political propaganda or election campaign materials will emanate from.²⁵² This mandatory registration is an effective regulatory mechanism for COMELEC to ensure a fairer playing field between and

²⁵¹ COMELEC Res. No. 10730 (2021), § 9(c).

²⁵² § 9(c)(1)–(3).

amongst candidates since it becomes easier to trace non-compliant and nuisance campaign media as “[o]nly verified accounts, websites, blogs, and/or social media pages may run electoral ads, and boost or promote electoral posts.”²⁵³ Similar to previous provisions, the Resolution requires disclosing its status as a paid advertisement and the identity of the sponsor of election propaganda.²⁵⁴

Unlike previous provisions and guidelines, however, the Resolution explicitly necessitates that all “[i]nformation contained in online campaign propaganda shall be truthful and not misleading, nor shall it tend to unjustifiably cast doubt on the integrity of the electoral process.”²⁵⁵ This is both striking and pertinent to the regulation of political deepfakes used as election propaganda as there is now the requirement that these deepfakes if utilized must not “mislead” or “unjustifiably cast doubt on the integrity of the electoral process.”²⁵⁶ However, there may be challenges in interpreting what constitutes “misleading” propaganda or one that “unjustifiably cast[s] doubt on the integrity of the electoral process” as the Resolution does not provide a definition or criteria for a finding of this specific violation.

As the Resolution, other COMELEC Regulations, and even other election laws are silent on the interpretation of this particular standard, there may be a resort to parallel legal standards in other areas of Philippine law.

The essence of what is deemed “misleading” under the law may be gleaned from the civil law treatment of consent as an adjacent legal paradigm. After all, the ability to mislead another person contemplates the existence of some form of deception where a person’s understanding or conception of a certain idea or event is warped without their consent.

First and foremost under the New Civil Code, consent in a contract must be absolute and unqualified.²⁵⁷ There is no valid consent when defects such as “mistake, violence, intimidation, undue influence, or fraud” are present.²⁵⁸ Particularly for the vice of “fraud,” it is when “through insidious words or machinations of one of the contracting parties, the other is induced to enter into a contract which, without them, he would not have agreed

²⁵³ § 9(c)(3).

²⁵⁴ § 9(c)(6)–(d).

²⁵⁵ § 9(c)(5).

²⁵⁶ § 9(c)(5).

²⁵⁷ CIVIL CODE, art. 1319.

²⁵⁸ Art. 1330.

to.”²⁵⁹ More pertinently, “[the] [f]ailure to disclose facts, when there is a duty to reveal them, as when the parties are bound by confidential relations, constitutes fraud.”²⁶⁰ In addition, “usual exaggerations in trade, when the other party had an opportunity to know the facts, are not in themselves fraudulent.”²⁶¹ And generally the “mere expression of an opinion does not signify fraud, unless made by an expert and the other party has relied on the former's special knowledge.”²⁶²

Applying these concepts to the regulation of political deepfakes online, the information posted may be deemed misleading when the information contained induces someone to act in a particular way they would not have otherwise acted (e.g. supporting a candidate, voting for a candidate, or even criticizing a candidate).²⁶³ A political *deepfake* may also be deemed misleading if there is a deliberate omission in disclosing facts “when there is a duty to reveal them.” In the political context, there is an onus on aspirants to ensure that deepfakes employed are truthful because “[p]ublic office is a public trust.”²⁶⁴ After all, these public officers: “hold in trust powers that they exercise in [sic] behalf of the public. Hence, they are held to higher standards than ordinary citizens to keep the people's faith in the State.”²⁶⁵ The Constitution also requires that public officers “must at all times be accountable to the people, serve them with utmost responsibility, integrity, loyalty, and efficiency, act with patriotism and justice, and lead modest lives.”²⁶⁶ As defined by the Court in *Republic v. Sereno*,²⁶⁷ integrity “contemplates both adherence to the highest moral standards and obedience to laws and legislations.”²⁶⁸

Consistent with these principles, it is unquestionable that aspirants who seek to become public officers can validly be saddled with the duty of honesty and integrity. In any event, the public offices they seek out require these exacting standards. Therefore, the use of misleading political deepfakes would run contrary to the duty of trust and integrity required of public officers.

²⁵⁹ Art. 1338.

²⁶⁰ Art. 1339.

²⁶¹ Art. 1340.

²⁶² Art. 1341.

²⁶³ Art. 1338.

²⁶⁴ CONST. art. XI, § 1.

²⁶⁵ *Neri v. Ombudsman*, G.R. No. 212467, 991 SCRA 642, 663, July 5, 2021.

²⁶⁶ CONST. art. XI, § 1.

²⁶⁷ *Republic v. Sereno*, G.R. No. 237428, 866 SCRA 494, June 19, 2018.

²⁶⁸ *Id.* at 557.

As to the ability of political deepfakes to “unjustifiably cast doubt on the integrity of the electoral process,” it must be highlighted that COMELEC possesses the vested interest to ensure “free, orderly, honest, peaceful, and credible elections.”²⁶⁹ Any threat to that mandate initiated by the creation or diffusion of a deepfake by a political candidate or party would fall within the boundaries of casting doubt on the integrity of the electoral process. Furthermore, jurisprudence has reiterated that the paramount interest in construing election laws must be towards the ends of protecting “the true will of the electorate.”²⁷⁰ While that specific pronouncement has been used to interpret the results of election contests in relation to nuisance candidates,²⁷¹ the rationale behind the provision can also be applied in a parallel manner to the regulation of political deepfakes.

In *De Alban v. COMELEC*, the Supreme Court explained the impetus behind the creation of Section 69 of the OEC, which governs the procedure on nuisance candidates, to wit:

[T]he Court stressed that the importance of barring nuisance candidates from participating in the electoral exercise is the avoidance of confusion and frustration in the democratic process by preventing a faithful determination of the true will of the electorate. It seeks to address the "dirty trick" employed by political rival operators to reduce the votes of the legitimate candidates due to the similarity of names and particularly benefitting from Comelec's "slow-moving decision-making."²⁷²

Furthermore, the Court resorted to legislative history to ascertain the intent of the legislators in crafting the provision:

MR. GONZALES. In connection with this, just to clarify it in order that we may know the legislative intention. If a candidate openly states in his campaign that he is not [...] he will not assume office, but he is running in order to preserve his political leadership, let us say, in the province or in the municipality to enable the [...] to enable his running mate to assume office or to

²⁶⁹ COMELEC Res. No. 10730 (2021), § 9(c)(5); CONST., art. IX-C, § 4.

²⁷⁰ *Rulloda v. COMELEC* [hereinafter “*Rulloda*”], G.R. No. 154198, 395 SCRA 535, 541, Jan. 20, 2003; *Engle v. COMELEC*, G.R. No. 215995, 781 SCRA 201, 226, Jan. 19, 2016; *De Alban v. COMELEC* [hereinafter “*De Alban*”], 921 Phil. 524, 534, G.R. No. 243968, (2022).

²⁷¹ ELECT. CODE, § 69; *Rulloda*, 395 SCRA at 541; *Engle*, 781 SCRA at 226; *De Alban*, 921 Phil. at 534.

²⁷² *De Alban*, 921 Phil. at 534.

get elected but he will not assume office, will he be a guest candidate since that is a circumstance or act which shows that he has no bona fide intention [...] intention to run for office[?]

MR. PEREZ (L.). He will be considered a nuisance candidate if he does not have a bona fide intention to assume the position if elected.

MR. GONZALES. I recall that had happ[e]ned a number of times before where, in spite of that, they still get elected. Shall we overturn the will of the people[?] [T]here is a disclosure to the people and yet the people still elects [sic] him.

MR. PEREZ (L.) If he [is] disqualified before the election because that x x x those are facts that he has no bona fide intention to assume the office if elected, he becomes a nuisance candidate.

* * *

MR. VILLAFUERTE. Mr. Speaker. Just one comment, Mr. Speaker. The bona fide requirement so as not to become a nuisance candidate does not pertain to the assumption but to running for public office.

MR. PEREZ (L.) But, Mr. Speaker, the very l[i]s mota or the very sole/purpose of running is to be in office, but if you are just running in order to create a vacancy for your running mate, you should not be considered a bona fide candidate.²⁷³

Further, the interpretation of “a faithful determination of the true will of the electorate” which is the third instance that would justify the ousting of a nuisance candidate points to:

"[C]ircumstances" or "acts" that would demonstrate that the purpose of the filing of the CoC is inconsistent with the definition of a candidate as someone "aspiring for or seeking elective public office." The common thread of the three instances is that the nuisance candidates filed their CoCs not to aspire or seek public office but to prevent a faithful determination of the people's true will. Relevantly, the assailed last phrase in Section 69 of the OEC should cover all acts or circumstances clearly demonstrating that the CoC was filed in bad faith.²⁷⁴

²⁷³ *Id.* at 536.

²⁷⁴ *Id.* at 535.

Taking these interpretations together, the state interest in weeding out nuisance candidates is to prevent the electorate from being deceived and voting for a candidate based on false or misleading pretenses. The State provides protection measures to preempt dilatory tactics made in bad faith like the similarity of names which are geared towards defrauding the voter into thinking they are voting for a certain candidate when in fact they are voting for another. Clearly with these nuisance candidates, there is an explicit attempt to frustrate the true will of the electorate through technical chicanery and in turn cast doubt on the integrity of the electoral process writ large since the voter is defrauded of their right to genuine and meaningful suffrage. Instead, the voter's consent is vitiated, and their choice of leader(s) invalidated. The candidate who did not intend to run for office or serve their constituencies emerges as the victor in an attempt to spite the original candidate that the voter intended to support.

In a similar vein, the regulation of political deepfakes and the requirement that information contained therein must not cast doubt on the electoral process seek to arrest any voter deception and duplicitous tactics. With deepfakes, through technological sophistry, a voter may be misled to believe certain facts about a political candidate which may become a deciding factor in determining who they will ultimately cast their vote for on election day. Even if the underlying premise about that candidate may be false as reinforced by the believability of a deepfake, the voter may be convinced by the information contained and be influenced to either vote for (or against) that candidate. This kind of persuasion is malicious and disingenuous as the voter would not have acted in that particular manner if the truth of the matter was known.

Much like in the case of nuisance candidates, the deserving candidate whom the individual would have voted for is deprived of the opportunity to serve and another takes their place. Once more, the voter's consent cannot be deemed to be genuine as it is clouded by deliberately false information as perpetuated by deepfakes. Consequently, the integrity of the electoral process has been cast into doubt as it has been tainted by these deceitful technological devices.

Even as a general utterance or form of expression, misleading political deepfakes would fall beyond the ambit of constitutional protection. The Court in *Chavez v. Gonzales* has already conceded that some forms of speech and expression may be validly regulated as a measure of police power to prevent injuries upon the rights of other individuals and society as a

whole.²⁷⁵ The same case has recognized that libelous speech falls beyond the penumbra of Article III, Section 4 of the Constitution.²⁷⁶

In this case, misleading political deepfakes could be construed as a type of libelous speech or form of expression due to the existence of malice or “personal ill-will or spite and speaks not in response to duty, but merely to injure the reputation of the person defamed.”²⁷⁷ Similarly, the misleading political deepfakes contemplated by the Resolution only exist to “mislead” or “unjustifiably cast doubt on the integrity of the electoral process.”²⁷⁸ Clearly then, the intent of the author in creating and dispersing such a kind of political deepfake stems from a place of ill-will.²⁷⁹ Hence, these misleading political deepfakes may be the subject of the state’s police power and limited accordingly due to the potential deleterious effects on individuals and Philippine society especially in the context of elections.²⁸⁰

Ultimately, both the provision and guidelines tend towards the same goals of preserving the integrity of election results and assuring that the rightful candidates (as selected by a well-informed electorate) are put into power. And COMELEC’s directives in the Resolution that these election materials online should not be misleading nor cast doubt on the electoral process are consequently a valid exercise of its quasi-legislative powers as it is in line with the law and the Constitution.²⁸¹

2. Resolution No. 11064

Another resolution of significance is COMELEC Resolution No. 11064 released only in 2024 in anticipation of the upcoming midterm elections in 2025. Like its predecessor (i.e., Resolution No. 10730), COMELEC draws on its Constitutional mandate in enforcing election laws,²⁸² ensuring “free, orderly, honest peaceful, and credible elections,”²⁸³ and promulgating rules in line with existing election laws²⁸⁴ as justifications in promulgating the Resolution. It also builds on the guidelines on social

²⁷⁵ *Chavez*, 545 SCRA at 486.

²⁷⁶ *Id.*

²⁷⁷ *United States v. Cañete*, 38 Phil. 253, 264 (1918).

²⁷⁸ COMELEC Res. No. 10730, § 9(c)(5).

²⁷⁹ *United States v. Cañete*, 38 Phil. at 264.

²⁸⁰ *Chavez*, 545 SCRA at 486; *See supra* Part II(A)-(B).

²⁸¹ *See Lokin*, 621 SCRA at 404.

²⁸² CONST. art. IX-C, § 2(1); COMELEC Res. No. 11064, pmbl. ¶ 1.

²⁸³ CONST. art. IX-C, § 4; COMELEC Res. No. 11064, pmbl. ¶ 3.

²⁸⁴ Rep. Act No. 9006, § 13; COMELEC Res. No. 11064, pmbl. ¶ 8.

media established in Resolution No. 10730 but nuances such in light of the advances in AI.²⁸⁵

The Resolution is a response to growing public concerns about the propensity for the abuse of AI in elections.²⁸⁶ Hence, the Resolution concerns itself with: “the regulation of the use, and the prohibition, and punishment of the misuse of social media, artificial intelligence, and internet technology, for purposes of digital election campaigning for the 2025 National and Local Elections and the BARMM parliamentary Elections.”²⁸⁷

It makes a further distinction from deepfakes and includes in its definition of terms “cheapfakes” which are “forms of visual disinformation for which authentic images or videos are re-contextualized to deliberately alter their meaning.”²⁸⁸ Much like deepfakes, cheapfakes are more rudimentary forms of media manipulated by technology (which may or may not include AI) with the sole purpose of forwarding disinformation.²⁸⁹ The incorporation of this kind of manipulated media strengthens COMELEC’s position in fighting against disinformation as it recognizes that deceptive media can come in different forms (created by AI or otherwise).

Afterwards, the Resolution defines deepfakes as: “digitally manipulated images, videos, or audio files created using artificial intelligence to fabricate realistic representations of people, events, or statements and falsely make it appear that an action statement, or event transpired but did not occur in reality.”²⁹⁰ The definition forwarded by COMELEC is consistent with the general industry’s consensus on the features of the technology as being driven by AI and being used to manipulate media.²⁹¹

Additionally, the Resolution makes another distinction for “soft fakes” which are “a type of media manipulation that involves subtle and often imperceptible alterations to content, such as editing photos or videos,

²⁸⁵ COMELEC Res. No. 11064, pmbl. ¶ 10.

²⁸⁶ See Gabriela Baron, *COMELEC on Guard Against Deepfakes – Garcia*, DAILY TRIBUNE, Sept. 3, 2024, at <https://tribune.net.ph/2024/09/02/comelec-on-guard-vs-deepfakes-garcia>; Ferdinand Patinio, *COMELEC to Regulate Use of AI in 2025 Polls Campaign*, PHIL. NEWS AGENCY, July 18, 2024, at <https://www.pna.gov.ph/articles/1229291>.

²⁸⁷ COMELEC Res. No. 11064, art. I, § 1.

²⁸⁸ Art. II(7).

²⁸⁹ See Lance Eliot, *Explaining Deepfakes versus Cheap Fakes and the Role of Generative AI*, FORBES, Jan. 26, 2024, at www.forbes.com/sites/lanceeliot/2024/06/25/cheap-fakes-and-rescuing-humankind-via-generative-ai/.

²⁹⁰ COMELEC Res. No. 11064, art. II(13).

²⁹¹ See *supra* Part I(B).

to influence perceptions or opinions without the more evident distortions characteristic of deepfakes.”²⁹² Gleaned from the definition, what distinguishes softfakes from deepfakes is the manner of editing. Softfakes are media whose alterations or manipulations are meant to be undetectable which may not always be true for deepfakes and especially for cheapfakes. By definition, cheapfakes are more elementary forms of edited media.²⁹³ However, the common intention for all three kinds of manipulated media is to peddle influence and sway the opinions of their audience. The distinction by COMELEC of this kind of deepfake is a commendable effort to proactively combat disinformation by political candidates and parties. Not only is it an acknowledgement that this kind of deepfake is employed in campaigns but also that there exists a large propensity for abuse by candidates and parties as seen in the elections of Pakistan and Indonesia.²⁹⁴

Remarkably, the Resolution also differentiates between “disinformation” and “misinformation.” The former is defined as “false information that is spread, or information that is based on fact but removed from its original context and is maliciously manipulated, and used in a partisan political activity or digital election campaign, with the intent to mislead, harm, or manipulate public opinion or voter behavior.”²⁹⁵ The latter is defined as “false or inaccurate information that is spread in a partisan political activity or digital election campaign, regardless of whether there is intent to mislead.”²⁹⁶

Both definitions grapple with false information in the context of a political campaign but disinformation is deemed a positive act made with malice and aimed at negatively affecting voter behavior. These definitions align with the general consensus regarding these terms and demonstrate COMELEC’s cognizance of modern problems plaguing the electoral process.²⁹⁷

Resolution No. 11064 contains many similar requirements as Resolution No. 10730 specifically regarding the registration of social media accounts, websites, and other platforms “that are primarily designed or

²⁹² COMELEC Res. No. 11064, art. II(26).

²⁹³ Art. II(7); Eliot, *supra* note 289.

²⁹⁴ Chowdhury, *supra* note 175; *See supra* Part II(D).

²⁹⁵ COMELEC Res. No. 11064, art. II (15).

²⁹⁶ Art. II(23).

²⁹⁷ *See supra* Part II(D).

primarily used to promote the election or defeat of a particular candidate or candidates.”²⁹⁸

As to the use of AI, Resolution No. 11064 provides stringent disclosure and disclaimer requirements on the use of AI in election propaganda, to wit:

Section 1. Disclosure and disclaimer requirements - All election propaganda and campaign materials of candidates and political parties, that utilize artificial intelligence technology, shall be required to disclose the fact of its use and appropriate technology is employed to identify its authenticity and legitimate source.²⁹⁹

Furthermore, depending on the type of media, the Resolution provides the required manner of disclosures. In general, the disclosure, which must be found on the material itself, must expressly state that the content is AI-manipulated and must explain the nature and extent of such manipulation. Likewise, it must warrant that prior consent was obtained from all depicted individuals. For visual media, the disclaimer should be large enough to read (or at least as large as the largest font used), legible, and separated from other content. For audiovisual media, the disclosure must be both verbal and written—both of which should be visible and audible. As for audio content, the disclosure must be spoken clearly and audibly. Note also that the latter two forms require the disclosure to be shown and/or spoken at the beginning and end of the media. Lastly, the Resolution requires the use of digital technologies such as watermarks to verify the content’s origin and authenticity.³⁰⁰

In the interest of transparency, COMELEC generally requires that any disclosure must be “clear, conspicuous, and not easily removed or altered.”³⁰¹ These disclosure requirements vary further depending on the type of media (e.g. visual, audio-visual, or audio-only).³⁰² For visual media, text disclaimers must be large, conspicuous, and easily readable.³⁰³ For audio formats, the disclaimer must instead be audible and clear at the beginning and end.³⁰⁴ While audio-visual media requires both audio and visual

²⁹⁸ COMELEC Res. No. 11064, art. III, § 1.

²⁹⁹ Art. IV, § 1.

³⁰⁰ COMELEC Res. No. 11064, art. IV, § 2.

³⁰¹ Art. IV, § 2.

³⁰² Art. IV, § 2.

³⁰³ Art. IV, § 2(1).

³⁰⁴ Art. IV, § 2(3).

disclosures that must again be conspicuous, clear, and understandable to the viewer.³⁰⁵ It even imposes a time requirement for the disclosure of the AI-generated campaign material to be “for a sufficient duration both before and after the campaign material is presented.”³⁰⁶ Furthermore, there even must be a “detailed explanation of the nature and extent of the manipulation” and must state that all persons depicted have provided their consent.³⁰⁷

These disclosure requirements are comprehensive efforts to curb any deception on the part of the voting public. The sweeping requirements make specifications that ascertain that regardless of the format of the deepfake, there will be no false or misleading pretenses regarding the material.

These disclosures are constitutional as they only constitute content-neutral regulatory measures on the incident of the political speech (e.g. “time, place or manner, and under well-defined standards”).³⁰⁸ In addition, these measures parallel existing guidelines on advertising disclosures.³⁰⁹ Furthermore, they echo existing state legislation in the United States which require disclosure of the use of AI technology in the generation of election materials and which provide specific requirements depending on the type of media.³¹⁰ Unlike US legislation, however, COMELEC even goes as far as to require that candidates and parties explain the nature and extent of the manipulation. This breakdown in the editing process adds another protective layer to ensure that the viewer is not misled nor will carry false assumptions as a result of the edited media.

The consequence of the failure to provide and/or comply with these disclosure requirements set out by COMELEC will result in the errant candidate or party being “required to explain why a complaint for [the] violation of these guidelines should not be filed against them.”³¹¹ COMELEC may likewise request the removal or takedown of such edited

³⁰⁵ Art. IV, § 2(2).

³⁰⁶ Art. IV, § 2.

³⁰⁷ Art. IV, § 2.

³⁰⁸ *Chavez*, 545 SCRA at 493.

³⁰⁹ Rep. Act No. 9006, § 4; COMELEC Res. No. 10730, § 7(a), 9(c)–(d)

³¹⁰ See Assemb. B. 730, 2019–2020 Reg. Sess. (Cal. 2019), § 4; Sess. B. 5152-S, 68th Leg., Reg. Sess. (Wash. 2023), § 2(4); H.B. 919, 2024 Leg., Reg. Sess. (Fla. 2024), § 1; H.B. 1133, 123rd Gen. Assemb., 2nd Reg. Sess. (Ind. 2024), § 3; H.B. 172, 2024 Reg. Sess. (Ala. 2024), § 2(b)(1).

³¹¹ COMELEC Res. No. 11064, art. IV, § 4.

posts.³¹² This requirement is consistent with Resolution No. 10730 which also sanctions the removal of prohibited propaganda.³¹³

This obligation to provide a disclosure about the use of AI in editing and manipulating media also extends to online platforms. Hence, Resolution No. 11064 provides that:

Section 3. Platform Disclosure obligations- - Once an online and digital platform that hosts election-related content has been notified by the Task Force KKK sa Halalan that media has been manipulated with artificial intelligence technology, they shall be required to implement reasonable mechanisms that disclose the same. such disclosures must be clearly visible or audible to all audiences, ensuring that viewers or listeners are adequately informed that the content has been altered in compliance with the requirements of these guidelines.³¹⁴

In this provision, social media platforms must impose “reasonable mechanisms” to provide disclosures on the use of AI on media posted on its platforms even if done by the user political candidate or party.³¹⁵ This imposition of this obligation on online and digital platforms is of note as it is a consequence of COMELEC’s inclusion of social media as a form of mass media.³¹⁶ However, whether or not this classification and corresponding obligation is constitutional remains to be clarified. Considering the limitations of COMELEC’s regulatory powers,³¹⁷ it must be noted that the classification does not find Constitutional basis nor statutory basis in the OEC or Fair Election Act.

Hence, while Resolution No. 10730³¹⁸ and by extension Resolution No. 11064 classify social media as mass media, it may be argued that such classification and the disclosure obligations imposed are *ultra vires*. Even in jurisprudence, the Supreme Court has explicitly pronounced that the COMELEC’s regulatory power only extends to political candidates and parties.³¹⁹ It does not extend to the private speech of citizens, especially those

³¹² Art. IV, § 4.

³¹³ COMELEC Res. No. 10730, §§ 26, 28.

³¹⁴ COMELEC Res. No. 11064, art. IV, § 3.

³¹⁵ Art. IV, § 3.

³¹⁶ COMELEC Res. No. 1070 (2021), § 1(10); COMELEC Res. No. 11064, art. II(21).

³¹⁷ *See Lokin*, 621 SCRA at 403–04.

³¹⁸ COMELEC Res. No. 10730, § 2(10).

³¹⁹ *Diocese of Bacolod*, 747 SCRA at 114; *See also St. Anthony Coll.*, G.R. No. 258805, slip op. at 15–17.

that are in the nature of social advocacy.³²⁰ Therefore, the necessary implication is that COMELEC's regulatory power does not extend to the speech of private juridical entities such as social media platforms.

Even with the guidelines provided by the Court in *Diocese of Bacolod* for the exceptional regulation of private speech that constitutes election paraphernalia, the disclosure obligation in Resolution No. 11064 fails to satisfy the jurisprudential test of legality, reasonableness, narrow tailoring, and the use of the least restrictive means.³²¹

Although the posts may indeed fall within the criteria of election propaganda as it seeks to promote the election or defeat of a candidate or party,³²² the disclosure obligation is not provided by law. It is merely contained in a Resolution which is still not a law as it was not a bill "passed [after] three readings on separate days [by the legislature.]"³²³ And as it stands, the law (specifically the OEC and Fair Election Act) are silent as to the regulation of online social media platforms. Even the 1987 Constitution remains silent as to the classification of social media as mass media.

Understandably these laws, which were crafted before the exponential adoption of social media, would be silent as to the regulation of the technology. However, no matter how well-intentioned COMELEC may be, it does not disprove the *ultra vires* nature of the disclosure obligation and even classification. Further, while there is room to argue about the narrowly tailored nature of the obligation and even its reasonability, it must be underscored that the Resolution nonetheless creates a positive substantial obligation on these private juridical entities which does not pertain to merely the "time, place, and manner of the rendition of the message."³²⁴

Lastly, the Resolution creates the *Task Force sa Katotohanan, Katapatan, at Katarungan sa Halalan* ("TF KKK") which "shall be headed by the Director IV of the [Education and Information Department] EID as the Chairperson

³²⁰ *Diocese of Bacolod*, 747 SCRA at 114–15.; *See also St. Anthony Coll.*, G.R. No. 258805, slip op. at 15–17.

³²¹ *Diocese of Bacolod*, 747 SCRA at 115 (Emphasis supplied.); *See also St. Anthony Coll.*, G.R. No. 258805, slip op. at 12–13.

³²² ELECT. CODE, § 79(b); COMELEC Res. No. 11064, art. II(14); COMELEC Res. No. 10730, § 1(16); *See supra* Part III(A).

³²³ CONST., art. VI, § 26(2).

³²⁴ *Diocese of Bacolod*, 747 SCRA at 115.

while the Director IV of the Law Department as Co-Chairperson.”³²⁵ Other involved individuals include:

- a. Designated lawyers, special investigators, agents, and representatives from the Office of the Chairman, Office of the Commissioners, and the Law Department;
- b. Designated cybersecurity officials from the COMELEC;
- c. Highly technical experts or professionals hired by the COMELEC, or by the deputized agencies or instrumentalities of the government, assigned to facilitate the necessary digital forensics, and to monitor and help in the detection of the prohibited misuse and malicious use of social media, AI, and internet technology by candidates, parties, voters, or by private individuals and entities.³²⁶

The inclusion of “cybersecurity officials” and “[h]ighly technical experts or professionals” is a prudent measure on the part of COMELEC to counteract the proliferation of deceptive political deepfakes as continuous technological advancements serve as formidable obstacles to reliable *deepfake* detection.³²⁷

Among its functions, the TF KKK is empowered with the:

1. Reception and review of registration forms, validation, and endorsement to the Commission En Banc of approval for publication and monitoring of official social media and/or online accounts of candidates, their campaign arms/groups, political parties/coalitions, and party-list organizations;
2. Monitoring of registered and unregistered social media and online accounts or websites which are used to endorse or campaign against candidates, political parties/coalitions, party-list organizations, or to propagate false information on the Philippine elections, election system, and the COMELEC;

³²⁵ COMELEC Res. No. 11064, art. VI, § 1.

³²⁶ Art. VI, § 1.

³²⁷ See Nils C. Köbis, Barbora Dolezalova, & Ivan Soraperra, *Fooled twice: People Cannot Detect Deepfakes but Think They Can*, iSCIENCE, Nov. 19, 2021, at www.sciencedirect.com/science/article/pii/S2589004221013353.

3. Issuance of Show Cause Order, and initial investigation of detected or reported prohibited acts under these guidelines;
4. *Motu proprio* filing of complaints against erring candidates, parties, individuals, and other entities, submission of requests to technology platforms for the takedown of fake accounts and contents in violation of these guidelines, as well as reporting and immediate coordination with the National Bureau of Investigation (NBI) and Philippine National Police (PNP) for the preservation of information necessary to investigate and prosecute those involved in the commission of prohibited acts under these guidelines or relevant special laws;
5. Endorsement to the Commission En Banc to enter partnerships and collaborations for effective voter education and campaign on the responsible use of social media, artificial intelligence, and internet technology in elections;
6. Endorsement to the Commission En Banc to enter into agreements or deputize relevant law enforcement agencies and government instrumentalities such as, but not limited to the Department of Justice (DOJ), NBI Cybercrime Division, Cybercrime Investigation and Coordinating Center (CICC), National Telecommunications Commission (NTC), PNP, to aid in the prevention, monitoring, and investigation of prohibited misuse and malicious use of social media, artificial intelligence technology, and internet technology during election and campaign period; and
7. Perform such other powers and duties as may be directed to it by the Commission En Banc.³²⁸

Apart from standard functions such as the monitoring, investigation, and review of online election propaganda, the Resolution authorizes the task force to *motu proprio* file complaints “against erring candidates, parties, individuals, and other entities.”³²⁹ This power is considerably broad but arguably still within COMELEC’s ambit of power as it is tasked with the enforcement and administration of “all laws and regulations relative to the conduct of an election[.]”³³⁰ Furthermore, this power is in line with existing laws that permit COMELEC to prosecute the commission of election

³²⁸ COMELEC Res. No. 11064, art. VI, § 3.

³²⁹ Art. VI, § 3.

³³⁰ CONST., art. IX-C, § 2(1).

offenses under the OEC and Fair Election Act.³³¹ Likewise, the *motu proprio* resolution of violations is synergistic with the Commission's goal of increasing transparency in the upcoming 2025 elections.³³²

C. Deepfake Misuse as Election Offenses

The next natural question in surveying the regulation of political deepfakes is whether the abuse of such may constitute an election offense under existing laws and regulations. Corollary to that, which stakeholders should then be held liable for the commission of such election offenses?

1. Election Offenses

In order to determine whether the creation and dispersion of political deepfakes are considered election offenses, a survey must be made of existing laws, issuances, and regulations, namely: (1) the OEC; (2) COMELEC Resolution No. 10730; and (3) COMELEC Resolution No. 11064. Nonetheless, while these current sources of laws and obligations are well-intentioned, they are at the same time unconstitutional for transgressing protected rights.

First, the creation and use of political deepfakes in certain instances may constitute election offenses. Based on the OEC and Resolution No. 10730, any person may be held liable for tampering with lawful election propaganda.³³³ COMELEC resolutions have been unequivocal and have deemed that political deepfakes fall within the term "lawful election propaganda" as there is no explicit prohibition for its use.³³⁴ Thus, a person may be held liable for further tampering with political deepfakes dispersed by a political party or candidate in the context of a campaign.

Additionally, the OEC declares that the breach of the provisions on lawful election propaganda³³⁵ and prohibited election propaganda³³⁶ are considered election offenses.³³⁷ Furthermore, Section 261(z)(11) states that:

³³¹ ELECT. CODE, § 265; Rep. Act No. 9006, § 13.

³³² Victoria Tulad, *COMELEC Assures More Transparency in 2025 Midterm Polls*, ABS-CBN NEWS, Apr. 9, 2024, at <https://www.abs-cbn.com/news/2024/4/9/comelec-assures-more-transparency-in-2025-midterm-polls-1534>.

³³³ ELECT. CODE, §§ 83, 262; COMELEC Res. No. 10730, § 25.

³³⁴ COMELEC Res. No. 10730, § 1(16) & 6(e); COMELEC Res. No. 11064, pmb. ¶ 10; *See supra* Part III(A).

³³⁵ ELECT. CODE, § 82.

³³⁶ § 85.

³³⁷ § 262.

Any person who, for the purpose of disrupting or obstructing the election process or causing confusion among the voters, propagates false and alarming reports or information or transmits or circulates false orders, directives or messages regarding any matter relating to the printing of official ballots, the postponement of the election, the transfer of polling place or the general conduct of the election.³³⁸

Accordingly, the misuse of political deepfakes as election propaganda or political advertisements may be deemed as election offenses. It may also be considered an election offense when a person uses a deepfake to propagate or disseminate false reports or information on the elections especially when the motive is “disrupting or obstructing the election process or causing confusion.”³³⁹

This theory on the capacity for deepfakes to be purveyors of political disinformation during the campaign season, thus warranting punishment as an election offense, also finds confirmation in the recent COMELEC Resolutions No. 10730 and 11064.³⁴⁰

Resolution No. 10730, which implements Section 13 of the Fair Election Act, provides that: “Any violation of Fair Elections Act and these Rules shall constitute an election offense punishable under the first and second paragraph of Section 264 of the Omnibus Election Code and other pertinent laws, rules and regulations, whenever applicable.”³⁴¹

While Resolution No. 11064 states that:

Section 1. Prohibited acts - The following acts, by any individual, shall be considered misuse and malicious use of social media, artificial intelligence technology, and internet technology, and may constitute as an election offense under Section 261 (z) (11) of the OEC, if committed for purposes of propagating disinformation and misinformation to endorse or campaign against a candidate, a political party/coalition, and party-list organizations, and for propagating disinformation or misinformation against the Philippine election system, the COMELEC, and the electoral

³³⁸ § 261(z)(11).

³³⁹ § 261(z)(11).

³⁴⁰ COMELEC Res. No. 10730; COMELEC Res. No. 11064.

³⁴¹ COMELEC Res. No. 10730, § 41; Rep. Act No. 9006, § 13.

processes in the Philippines during the election and campaign period:

1. Use of "false amplifiers", such as fake accounts, bots, and astroturf groups filled with fake users to propagate disinformation and misinformation in endorsing or campaigning against a candidate, a political party/coalition, or party-list organizations, or to propagate disinformation, and misinformation targeting the Philippine election system, the COMELEC, and electoral processes during the election and campaign period;
2. Coordinated inauthentic behavior and utilization of hyperactive users for the aforementioned purposes;
3. *Creation and dissemination of deepfakes, cheapfakes, and softfakes for the aforementioned purposes;*
4. Use of fake and unregistered social media accounts during the election and campaign period for the same purposes; and
5. *Creation and dissemination of fake news in furtherance of the aforementioned purposes.*
6. *Use of content produced through AI technology but without compliance to the transparency and disclosure requirements under these guidelines.*³⁴²

Reading these provisions and guidelines together, COMELEC has affirmed that the creation and diffusion of deepfakes, cheapfakes, or softfakes for the “purposes of propagating disinformation and misinformation to endorse or campaign against a candidate, a political party/coalition, and party-list organizations, and for propagating disinformation or misinformation against the Philippine election system, the COMELEC, and the electoral processes in the Philippines during the election and campaign period” are election offenses.³⁴³

Through this provision, COMELEC clarifies that it is not the mere creation of the deepfake, cheapfake, or softfake that constitutes the election offense. What is being punished is the use of these kinds of manipulated media to forward the nefarious purposes of “disinformation or misinformation against the Philippine election system, the COMELEC, and

³⁴² COMELEC Res. No. 11064, art. V, § 1. (Emphasis supplied.)

³⁴³ ELECT. CODE, § 261(z)(11); COMELEC Res. No. 10730, § 41; Rep. Act No. 9006, § 13; COMELEC Res. No. 11064, art. V, § 1.

the electoral processes in the Philippines during the election and campaign period[.]”

Textually, however, the provision regarding election offenses in Resolution No. 11064 is an expansion and departure from Section 261(z)(11) of the OEC. To constitute an offense under the OEC, the purpose must be “disrupting or obstructing the election process or causing confusion among the voters, propagates false and alarming reports or information or transmits or circulates false orders, directives or messages regarding any matter relating to the printing of official ballots, the postponement of the election, the transfer of polling place or the general conduct of the election.”³⁴⁴

On the other hand, while both sources of laws are theoretically geared towards protecting the integrity of the electoral process and composite institutions, a problem arises in Resolution No. 11064 due to the inclusion of misinformation.

Once more, Resolution No. 11064 defines “misinformation” as “false or inaccurate information that is spread in a partisan political activity or digital election campaign, *regardless of whether there is intent to mislead.*”³⁴⁵ Hence, misinformation is understood as merely being false information.³⁴⁶ This definition is in contrast to “disinformation” which the Resolution requires the existence of malice or “the intent to mislead, harm, or manipulate public opinion or voter behavior.”³⁴⁷ Therefore, Resolution No. 11064 which punishes the creation and dissemination of deepfakes, cheapfakes, and softfakes for the purpose of misinformation (i.e., done without malice) may be considered unconstitutional for being overbroad and for creating a chilling effect on protected political speech.

As provided by the Court in *Adiong v. COMELEC*, a statute is overbroad when “it offends the constitutional principle that a governmental purpose to control or prevent activities constitutionally subject to state regulations may not be achieved by means which sweep unnecessarily broadly and thereby invade the area of protected freedoms.”³⁴⁸ Challenges are sanctioned with respect to these kinds of regulations as it may create a “chilling effect” on protected political speech. The rationale behind the

³⁴⁴ ELECT. CODE, § 261(z)(11).

³⁴⁵ COMELEC Res. No. 11064, art. II(23). (Emphasis supplied.)

³⁴⁶ Art. II(23).

³⁴⁷ Art. II(15).

³⁴⁸ *Adiong v. COMELEC*, G.R. No. 103956, 207 SCRA 712, 719, Mar. 31, 1992, *citing* *Zwickler v. Koota*, 389 U.S. 241 (1967).

challenge to these kinds of laws and regulations was elucidated by the Court in *Calleja v. Executive Secretary*, to wit: “The possible harm to society in permitting some unprotected speech to go unpunished is outweighed by the possibility that the protected speech of others may be deterred and perceived grievances left to fester because of possible inhibitory effects of overly broad statutes.”³⁴⁹ These challenges may exceptionally apply to penal statutes if there is a violation of free speech or “amounts to prior restraint.”³⁵⁰ To determine whether or not a restriction on speech is overbroad, the Court has ruled that it will scrutinize the effects of the limitations on speech.³⁵¹ Additionally, restrictions “must allow alternative avenues for the actor to make speech.”³⁵²

Even so, regulating speech online finds additional challenges due to the expansive nature of the medium and speed of the proliferation of speech. These complications were even acknowledged by the Court in the case of *Disini* as various stakeholders are immediately implicated with a single post such as the author, and people who “like,” “comment,” or “share” a post with others online.³⁵³ This phenomenon is complicated by the fact that this process of publication online is repeated *ad infinitum* thereby increasing the number of implicated stakeholders exponentially.³⁵⁴

Amidst this backdrop on the difficulty of regulating speech online, it must be noted that the provision in Resolution No. 11064 is penal in nature. Considering that the punishment for the commission of an election offense includes imprisonment, the provision is undoubtedly penal in nature. Notwithstanding the penal nature, however, it is clear that the Resolution concerns political speech as it is “speech ‘both intended and received as a contribution to public deliberation about some issue,’ ‘foster[ing] informed and civic-minded deliberation.’”³⁵⁵ Moreover, it concerns elections which invariably affect the Philippine political process.

Furthermore, it may transgress private political expression as the provision in Resolution No. 11064 specifies that the election offense may be

³⁴⁹ *Calleja v. Exec. Sec’y*, 918-B Phil. 1, 612 (2021), *citing* Southern Hemisphere Engagement Network, Inc. v. Anti-Terrorism Council, G.R. No. 178552, 632 SCRA 146, 183, Oct. 5, 2010.

³⁵⁰ *Exec. Sec’y v. Pilipinas Shell Petroleum Corp.*, G.R. No. 209216, Feb. 21, 2023.

³⁵¹ *Nicolas-Lewis*, 913 SCRA at 557.

³⁵² *Id.*

³⁵³ *Disini*, 716 SCRA at 323–25.

³⁵⁴ *Id.* at 324–25.

³⁵⁵ *Diocese of Bacolod*, 747 SCRA at 86 nn.142, 200–01.

committed by “any individual.”³⁵⁶ The overbroad nature and corresponding chilling effect on individuals who create and share deepfakes, cheapfakes, or softfakes are evident when a person creates these kinds of media containing and espousing false information about a political candidate, political party, elections, or even COMELEC without malice or intention to mislead.

It is plausible that an individual may merely possess and proliferate false information without knowing that such information is false. After all, social media platforms are by design poor receptacles of correct information despite being the primary source of news for many individuals.³⁵⁷ And yet under the model proposed by Resolution 11064, this act of sharing false information about the elections or a candidate on Facebook or any other social media platform (even without any knowledge of its error or falsity) can thus be charged with the commission of an election offense.³⁵⁸ More importantly, a complaint may be filed against any perpetrator *motu proprio*.³⁵⁹

Even under the guidelines in *Diocese of Bacolod*³⁶⁰ for the provisional regulation of private speech, the provision would not satisfy the jurisprudential test as it is not narrowly tailored nor is it the least restrictive means to attain the state objective of ensuring a “free, orderly, honest, peaceful, and credible elections.”³⁶¹ The expansive reach of the law will likely have the effect of hindering citizens from making political and election-related expressions or utterances using deepfakes online due to their hesitancy on the veracity of the media. The restriction is also overbroad because it does not afford the electorate another means or outlet to make such speech or utterance.³⁶²

³⁵⁶ COMELEC Res. No. 11064, art. V, § 1.

³⁵⁷ See Amy Watson, *Social Media News Worldwide - Statistics & Facts*, STATISTA, Dec. 18, 2023, at <https://www.statista.com/topics/9002/social-media-news-consumption-worldwide/#topicOverview>; See *supra* Part II(C)-(D); See also *ABS-CBN Corp. v. Ampatuan*, G.R. No. 227004, Apr. 25, 2023.

³⁵⁸ COMELEC Res. No. 11064, art. V, § 1.

³⁵⁹ Art. VI, § 3.

³⁶⁰ “The regulation (a) should be provided by law, (b) reasonable, (c) narrowly tailored to meet the objective of enhancing the opportunity of all candidates to be heard and considering the primacy of the guarantee of free expression, and (d) demonstrably the least restrictive means to achieve that object. The regulation must only be with respect to the time, place, and manner of the rendition of the message.” *Diocese of Bacolod*, 747 SCRA at 115.

³⁶¹ *Id.*; CONST. art. IX-C, § 4.

³⁶² *Nicolas-Lewis*, 913 SCRA at 559.

Additionally, the provision is not reasonable.³⁶³ The measure undoubtedly transgresses protected political expression which aims to stir discussion regarding public issues or officials. While well-intentioned, the inclusion of the creation and distribution of deepfakes, cheapfakes, or softfakes for the purposes of misinformation casts too broad of a regulatory net at the expense of the unwitting social media user.

In the COMELEC's valiant attempt to combat threats to the electoral process posed by disinformation, it unfortunately contravenes fundamental constitutional principles of free speech. It must be echoed that: "the guarantee of free speech was enacted to protect not only polite speech, but even expression in its most unsophisticated form."³⁶⁴ Thus, any attempts to limit or regulate speech must tread lightly to avoid encroaching on protected forms of expression.

Concededly, not all kinds of speech are afforded constitutional protection. As held in *Chavez*: "Some types of speech may be subjected to some regulation by the State under its pervasive police power, in order that it may not be injurious to the equal right of others or those of the community or society."³⁶⁵ Hence, the case states that slander, libel, lewd and obscene speech, and fighting words fall beyond the ambit of constitutional protection.³⁶⁶ It must be noted, however, that the aforementioned enumeration by the Court does not include false speech *per se*.

Just the same, slander and libel under the Revised Penal Code ("RPC") both require a malicious imputation in order for an individual to be held liable.³⁶⁷ This standard likewise applies to the commission of these crimes using computer systems (e.g. social media).³⁶⁸ Jurisprudence states that malice contemplates "personal ill-will or spite and speaks not in response to duty, but merely to injure the reputation of the person defamed."³⁶⁹ It exists when it is proven that "the author made the defamatory statement knowing it was false, or even if true, there is no showing of good

³⁶³ *Diocese of Bacolod*, 747 SCRA at 115..

³⁶⁴ *Tulfo v. People* [hereinafter "*Tulfo*"], G.R. No. 187113, 967 SCRA 415, 482, Jan. 11, 2021.

³⁶⁵ *Chavez*, 545 SCRA at 486.

³⁶⁶ *Id.*

³⁶⁷ REV. PEN. CODE, arts. 353, 358; LUIS REYES, II THE REVISED PENAL CODE: CRIMINAL LAW 1239 (2021 ed.).

³⁶⁸ See *Causing v. People* [hereinafter "*Causing*"], G.R. No. 258524, slip op. at 13–14, Oct. 11, 2023, citing *Disini*, 716 SCRA 237.

³⁶⁹ *United States v. Cañete*, 38 Phil. at 264.

intention and justifiable motive.”³⁷⁰ There exists two kinds of malice: (1) malice in fact (or actual malice); and (2) malice in law.³⁷¹

The former contemplates that:

[T]he offender makes the defamatory statement with the knowledge that it is false or with reckless disregard of whether it was false or not. The reckless disregard standard used here requires a high degree of awareness of probable falsity. There must be sufficient evidence to permit the conclusion that the accused in fact entertained serious doubts as to the truth of the statement he published. Gross or even extreme negligence is not sufficient to establish actual malice.³⁷²

Furthermore, the Court in *Tulfo v. People* cited *New York Times v. Sullivan* and qualified that: “Even assuming that the contents of the articles are false, mere error, inaccuracy or even falsity alone does not prove actual malice [...] There must be some room for misstatement of fact as well as for misjudgment.”³⁷³

On the other hand, malice in law refers to Article 354 of the RPC where malice is generally presumed from defamatory statements except with respect to privileged communication.³⁷⁴ When the defamatory statement falls within the exceptions of the provision, the prosecution is then required to prove malice in fact or actual malice.³⁷⁵ Moreover, even if a report is injurious, the legal presumption of malice dissipates once it is proven that there exists good faith.³⁷⁶

As regards the creation and dissemination of deepfakes, cheapfakes, or softfakes containing misinformation, the second provision of Article 354 may be applicable. It is conceivable that news sources, reporters, or even “bloggers” may be susceptible to sharing manipulated media containing

³⁷⁰ *Tulfo*, 967 SCRA at 473–74, *citing* *Brillante v. Court of Appeals*, 483 Phil. 568, 591 (2004).

³⁷¹ *See, generally, Tulfo*, 967 at 476–84.

³⁷² *Disini*, 716 SCRA at 317.

³⁷³ *Tulfo*, 967 SCRA at 483.

³⁷⁴ REYES, *supra* note 367, at 1245; REV. PEN. CODE, art. 354.

³⁷⁵ REYES, *supra* note 367, at 1246, *citing* *United States v. Bustos*, G.R. No. 12592, Mar. 8, 1918.

³⁷⁶ *Id.*; REV. PEN. CODE, art. 354.

misinformation through negligence or oversight.³⁷⁷ In these cases, the actual malice standard should apply. Consequently, there should be no finding of any wrongdoing for the lack of the concurrence of the positive act and intention to proliferate false information online.

From these discussions on malice, it is clear that the law requires a positive act on the part of the perpetrator such that they must have known of the falsity and nonetheless pursued with the dissemination of the imputation. By explicit fiat, jurisprudence has excluded “false, mere error, inaccuracy or even falsity alone[.]”³⁷⁸ Using these parallel standards in criminal law, it would be contrary to the spirit of existing law and plainly unreasonable to hold individuals liable for unknowingly sharing manipulated content containing false information on social media platforms. Hence further, punishing these citizens for the creation and diffusion of deepfakes for the purpose of misinformation would not be considered as narrowly tailored and its effects would undoubtedly create a chilling effect on voters who wish to express their support or disdain for a candidate or political party, as people will likely refrain from exercising their right to free speech and expression for fear of punishment.

Apart from the fact that COMELEC Resolution No. 11064 is overboard, another potential issue that may arise is the Resolution’s attempt to punish mere misinformation due to the nature of the offense. The punishment of the creation and dissemination of manipulated media containing misinformation is inconsistent with existing jurisprudential decrees because the prohibited acts enumerated may be considered *mala in se* in nature.

The case of *Cardona v. People*³⁷⁹ clarifies that: “The better approach to distinguish between *mala in se* and *mala prohibita* crimes is the determination of the inherent immorality or vileness of the penalized act.”³⁸⁰ Using this standard, it is evident that what COMELEC seeks to punish is not the creation or dissemination of the manipulated media itself. If that were the case, COMELEC would have prohibited the utilization of such media in its entirety.

³⁷⁷ See Gabriel Lalu, *Fake News Sources: 58% Say Social Media Influencers Spread False Info*, INQUIRER.NET, Oct. 11, 2022, at <https://newsinfo.inquirer.net/1678448/knowning-source-of-fake-news-58-say-false-info-came-from-socmed-influencers>.

³⁷⁸ *Tuffo*, 967 SCRA at 483.

³⁷⁹ [Hereinafter *Cardona*], G.R. No. 244544, 940 SCRA 646, July 6, 2020.

³⁸⁰ *Id.* at 660,

However, Resolution No. 11064 provides that it is not the mere commission of the enumerated acts that make the crime reprehensible. It is the commission of those enumerated acts “for purposes of propagating disinformation and misinformation to endorse or campaign against a candidate, a political party/coalition, and party-list organizations, and for propagating disinformation or misinformation against the Philippine election system, the COMELEC, and the electoral processes in the Philippines during the election and campaign period[.]”³⁸¹ It is the conjunctive act and intention that make the aforementioned prohibited acts reprehensible.

The consequence of the provision being *mala in se* in nature is that “good faith and lack of criminal intent” become viable defenses to the accused.³⁸² Likewise, there must be a concurrence of both the *actus reus* (“external or overt acts or omissions included in a crime's definition”) and *mens rea* (“the accused's guilty state of mind or criminal intent”) for there to be a conclusive finding of guilt.³⁸³

Once more, it must be reiterated that Resolution No. 11064 defines “misinformation” as “false or inaccurate information that is spread in a partisan political activity or digital election campaign, regardless of whether there is intent to mislead.”³⁸⁴ In addition, the Resolution includes in its prohibited acts the commission of certain acts for the purpose of propagating misinformation.³⁸⁵

As seen from the definition of misinformation and the enumeration of prohibited acts, there exists a disjunct between the inclusion of misinformation in the latter and the requirement of criminal intent. The proviso “regardless of whether there is intent to mislead” suggests that a person may still be held liable for the commission of any of the acts despite the lack of criminal intent.³⁸⁶ This inconsistency must be resolved in favor of prevailing jurisprudential guidelines. Therefore, it would be illogical and unreasonable to hold a person accountable for committing any of the prohibited acts for the purpose of misinformation since there may exist a lack of criminal intent. The lack of such criminal intent should continue to function as a valid defense.

³⁸¹ COMELEC Res. No. 11064, art. V, § 1.

³⁸² *Cardona*, 940 SCRA at 660.

³⁸³ *XXX v. People* [hereinafter “*XXX*”], G.R. No. 255877, slip op. at 8, Mar. 29, 2023.

³⁸⁴ COMELEC Res. No. 11064, art. II(23).

³⁸⁵ Art. V, § 1.

³⁸⁶ Art. II(23).

Similarly, the defenses of the lack of criminal intent and/or the existence of good faith must be available to anyone that fails to comply with the disclosure and transparency requirements in Resolution No. 11064. Gleaned from the current structure of the enumeration of prohibited acts, it is evident that the nature of the crime continues to be *mala in se* as it must be done “for purposes of propagating disinformation and misinformation to endorse or campaign against a candidate, a political party/coalition, and party-list organizations, and for propagating disinformation or misinformation against the Philippine election system, the COMELEC, and the electoral processes in the Philippines during the election and campaign period[.]”³⁸⁷

On the other hand, even if the crime was *mala prohibita* in nature, the Court in *People v. Ramoy* has ruled that there must still exist “the intent to perpetrate the act prohibited by the special law.”³⁸⁸ Hence, there must still exist the intent to commit the prohibited acts outlined in COMELEC Resolution 11064, specifically the “[c]reation and dissemination of deepfakes, cheapfakes, and soft fakes for the aforementioned purposes” and the “[u]se of content produced through AI technology but without compliance to the transparency and disclosure requirements under these guidelines.”³⁸⁹

Either way, there must exist some positive act or indication on the part of the perpetrator to commit the offense. And more importantly whether it is the creation and dissemination of manipulated media or compliance with disclosure guidelines, it remains a core principle in this jurisdiction that in case of doubt the court must rule in favor of the accused.³⁹⁰ Hence, any doubts in construing the provision should be resolved in favor of the accused.

2. *Obligations on Stakeholders*

Establishing what may be considered as election offenses under existing laws and regulations, one may now accordingly determine who may be held liable for the use or dissemination of political deepfakes.

³⁸⁷ COMELEC Res. No. 11064, art. V, § 1.

³⁸⁸ *People v. Ramoy*, G.R. No. 212738, slip op. at 16, Mar. 9, 2022, *citing* *People v. De Gracia*, G.R. No. 102009, 233 SCRA 716, 726, July 6, 1994.

³⁸⁹ COMELEC Res. No. 11064, art. V, § 1(3) & (6).

³⁹⁰ *Intestate Estate of Manolita Gonzales v. People*, G.R. No. 181409, 612 SCRA 272, 283, Feb. 11, 2010.

Generally, COMELEC's regulatory ambit primarily concerns political candidates and parties with narrowly tailored exceptions.³⁹¹ Thus, positive obligations regarding election laws are typically imposed on these parties and candidates. As an exception, COMELEC may regulate private speech provided that such regulation fulfills the jurisprudential test laid down in *Diocese of Bacolod*.

Furthermore, the OEC also provides that:

The principals, accomplices, and accessories, as defined in the Revised Penal Code, shall be criminally liable for election offenses. If the one responsible be a political party or an entity, its president or head, the officials and employees of the same, performing duties connected with the offense committed and its members who may be principals, accomplices, or accessories shall be liable, in addition to the liability of such party or entity.³⁹²

On the definition of "principals," the RPC states that:

Article 17. Principals. - The following are considered principals:

1. Those who take a direct part in the execution of the act;
2. Those who directly force or induce others to commit it;
3. Those who cooperate in the commission of the offense by another act without which it would not have been accomplished.³⁹³

Thus, any person who participates in the commission of an election offense, induces another to commit the election offense, or indispensably cooperates in the commission of the election offense may be considered a principal who is criminally liable under the OEC. The law, when read together with the definition found in the RPC, does not limit the identity of the perpetrator to the candidate or political party member or officer.

To be considered an accomplice, these individuals are those who "not being included in Article 17, cooperate in the execution of the offense by previous or simultaneous acts."³⁹⁴ Much like the definition of a principal

³⁹¹ *Diocese of Bacolod*, 747 SCRA at 114–15.

³⁹² ELECT. CODE, § 263.

³⁹³ REV. PEN. CODE, art. 17.

³⁹⁴ Art. 18.

under the RPC, no limitation is made for the law to only apply to the candidate or political party. Lastly, the RPC defines “accessories,” to wit:

Article 19. Accessories. - Accessories are those who, having knowledge of the commission of the crime, and without having participated therein, either as principals or accomplices, take part subsequent to its commission in any of the following manners:

1. By profiting themselves or assisting the offender to profit by the effects of the crime.
2. By concealing or destroying the body of the crime, or the effects or instruments thereof, in order to prevent its discovery.
3. By harboring, concealing, or assisting in the escape of the principals of the crime, provided the accessory acts with abuse of his public functions or whenever the author of the crime is guilty of treason, parricide, murder, or an attempt to take the life of the Chief Executive, or is known to be habitually guilty of some other crime.³⁹⁵

For election offenses, the most pertinent provisions from the RPC are the first two. Therefore, private individuals (not principals or accomplices) who possess knowledge of the election offense and either profit from the crime or conceal or destroy the effects of the crime are “accessories” that may be criminally liable.

Hence, while private political speech is typically unregulated, as an exception, those who fall within the definitions of principals, accomplices, or accessories in relation to the commission of election offenses may be held liable under the OEC.³⁹⁶ This rings true even if an individual is an employee of an aspirant or political party.³⁹⁷ COMELEC Resolution No. 10730 affirms this conclusion as “[s]ocial media associates including paid digital influencers and online content creators who use social media platforms to promote or defeat the election of any candidate, are considered individual contractors who are subject to reportorial requirements[.]”³⁹⁸

In addition, it must be highlighted that Section 13 of the Fair Election Act considers not only violations of existing election laws but also transgressions of COMELEC rules and regulations to be election offenses

³⁹⁵ Art. 19.

³⁹⁶ ELECT. CODE, § 263.

³⁹⁷ ELECT. CODE, § 263.

³⁹⁸ COMELEC RES. No. 10730, § 11.

punishable in accordance with the OEC.³⁹⁹ This reading is supported by Section 41 of Resolution No. 10730 and Article VII, Section 1 of Resolution No. 11064⁴⁰⁰ The provision presents broad and serious consequences in relation to existing COMELEC rules and regulations such that any breach may validly be interpreted as the commission of an election offense.

Pursuant to existing COMELEC resolutions, the breach of any of the constituent provisions is considered an election offense. This interpretation has been subsequently affirmed by COMELEC in Resolution 11064.⁴⁰¹ Depending on the provision and guidelines contained therein, private persons may again be held liable. The term “persons” contemplates both natural and juridical persons. For private natural persons, Resolution 11064 provides that the commission of prohibited acts may be done by “any individual.”⁴⁰² It does not limit itself to political candidates or parties.⁴⁰³

For private juridical persons, some provisions which may be breached include:

1. The prohibition on the removal, destruction or defacement of lawful election propaganda;⁴⁰⁴
2. Guidelines on the publication of election surveys;⁴⁰⁵
3. Prohibited acts during public meetings;⁴⁰⁶
4. The failure to register official social media accounts, websites, and platforms related to the campaign;⁴⁰⁷
5. Non-compliance with reportorial requirements for advertisements and expenditures;⁴⁰⁸
6. The failure to comply with disclosure and transparency requirements; and⁴⁰⁹
7. The commission of prohibited acts in relation to the misuse of social media, AI technology, or internet technology.⁴¹⁰

³⁹⁹ Rep. Act No. 9006, § 13.

⁴⁰⁰ COMELEC Res. No. 10730, § 41; COMELEC Res. No. 11064, art. VII, § 1.

⁴⁰¹ COMELEC Res. No. 11064, pmbl. ¶ 8.

⁴⁰² Art. V, § 1.

⁴⁰³ Art. V, § 1.

⁴⁰⁴ COMELEC Res. No. 10730, § 25.

⁴⁰⁵ § 30.

⁴⁰⁶ § 35.

⁴⁰⁷ COMELEC Res. No. 11064, art. III, §§ 1, 7.

⁴⁰⁸ COMELEC Res. No. 10730, § 11.

⁴⁰⁹ COMELEC Res. No. 11064, art. IV, §§ 3–4.

⁴¹⁰ Art. V, § 1.

Of particular note is the list of prohibited acts in Article V, Section 1 of Resolution No. 11064. Although the first paragraph states that any of the acts must be committed by “any individual,” several of the acts contemplate actions that may likewise be committed by juridical persons.⁴¹¹ For example, the definition of “coordinated inauthentic behavior” includes “deliberate and coordinated actions carried out by individuals, groups, or organizations to deceive or mislead the public by spreading false, misleading, or manipulated information online, typically involving automated or semi-automated means.”⁴¹² Thus, the inclusion of “groups” and “organizations” directly implies that the prohibited act may be done by a juridical entity.⁴¹³ Moreover, the use of “fake and unregistered social media accounts” can be done by a juridical entity as Resolution 11064 recognizes that the persons that may be responsible for the management of social media accounts may include entities.⁴¹⁴ Similarly, the creation and dissemination of fake news for the purposes of misinformation or disinformation, non-compliance with the disclosure requirements, and even the use of “false amplifiers” and the like may feasibly be committed by juridical persons or entities.⁴¹⁵

D. Pending Bills on Deepfake Regulation

Despite the lack of formalized legislation regulating deepfakes, the existence of several bills in either chamber of Congress points to the growing awareness of the legislature regarding the existence and proclivity for misuse of deepfake technology. Similarly, while none have yet successfully transformed into a binding law, the bills provide a unique insight into existing legislative perspectives and regulatory inclinations regarding deepfakes. Moreover, an assessment of the provisions contained in these bills may serve as templates for forthcoming regulations on political deepfakes, AI, or technology in general.

1. House Bill No. 5406

Former Congressman Rozzano Rufino B. Biazon introduced one of the earliest bills on deepfakes in the House of Representatives on November 11, 2019, entitled “An Act Declaring as Unlawful the Creation and Disclosure of any Deepfake Material or Materially Deceptive Audio or Video

⁴¹¹ Art. V, § 1.

⁴¹² Art. II (10), Art. V, § 1(2).

⁴¹³ Art. II (10), Art. V, § 1(2).

⁴¹⁴ Art. III, §§ 1, 4.

⁴¹⁵ Art. V, § 1.

Recording of an Individual without His or Her Consent” or the “Anti-Deepfake Act of 2019” (“Biazon Bill”).⁴¹⁶

The bill defines deepfakes as “a technique for human image synthesis based on artificial intelligence. It is used to combine and superimpose existing images and videos onto source images or videos using a machine learning technique known as generative adversarial network.”⁴¹⁷ Like the definition of deepfakes in COMELEC Resolution No. 11064, the definition in the bill recognizes that the technology involves AI in order to alter input media which may be images or videos.⁴¹⁸ Compared to the definition in Resolution No. 11064, the bill’s definition does not qualify that the depiction must be “realistic” nor has a proviso on the purpose of the technology, unlike in the Resolution which explicitly states is “[to] falsely make it appear that an action statement, or event transpired but did not occur in reality.”⁴¹⁹ In any event, the definition of deepfakes in the bill aligns with the general consensus on the technology and how it operates.⁴²⁰

Afterwards, the bill defines “[m]aterially deceptive audio or video recording” as:

[A]n image or an audio or video recording of a person's appearance, speech, or conduct that has been intentionally manipulated in a manner such that both of the following conditions are met:

- (1) The image or audio or video recording would falsely appear to a reasonable person to be authentic; and
- (2) The image or audio or video recording would cause a reasonable person to have a fundamentally different understanding or impression of the expressive content of the image or audio or video recording than that person would have if the person were hearing or seeing the unaltered, original version of the image or audio or video recording.⁴²¹

⁴¹⁶ H. No. 5406, 18th Cong., 1st Sess. (2019).

⁴¹⁷ § 2(b).

⁴¹⁸ COMELEC Res. No. 11064, art. II(13).

⁴¹⁹ Art. II(13).

⁴²⁰ See *supra* Part I(B).

⁴²¹ H. No. 5406, 18th Cong., 1st Sess. (2019), § 2(e). This refers to the State of California’s Defending Democracy from Deepfake Deception Act of 2024

This legal standard finds its roots in California state legislation, specifically Assembly Bill 2655 entitled “Defending Democracy from Deepfake Deception Act of 2024” which defines “materially deceptive content” as “audio or visual media that is digitally created or modified, and that includes, but is not limited to, deepfakes and the output of chatbots, such that it would falsely appear to a reasonable person to be an authentic record of the content depicted in the media.”⁴²² Like the Biazon Bill, it requires a determination of what appears to be authentic to a reasonable person.⁴²³ However, the Biazon Bill, Assembly Bill 2655, also provides exceptions to what are deemed “materially deceptive content” such as:

[A]ny audio or visual media that contains only minor modifications that do not significantly change the perceived contents or meaning of the content. Minor changes include changes to the brightness or contrast of images, removal of background noise in audio, and other minor changes that do not impact the content of the image or audio or visual media.⁴²⁴

This specific qualification of materially deceptive audio or video recording as worded is evidently absent from any COMELEC Resolution. While Resolution No. 11064 provides a parallel provision on how the technology has the ability to “falsely make it appear that an action statement, or event transpired but did not occur in reality,”⁴²⁵ The Biazon Bill creates a conjunctive standard such that the deepfake in order to be considered “materially deceptive,” it must appear to be authentic to a reasonable person and “cause a reasonable person to have a fundamentally different understanding or impression of the expressive content of the image or audio or video recording.”⁴²⁶ From this standard, the bill hinges the finding of any kind of deception based on a “reasonable person.” Jurisprudential treatment of this specific standard may be best surmised through the lens of civil law (specifically the law on quasi-delicts) as an existing parallel legal standard which requires a situational assessment in relation to general duties of care.

Article 2176 of the Civil Code provides: “Whoever by act or omission causes damage to another, there being fault or negligence, is obliged to pay for the damage done. Such fault or negligence, if there is no pre-existing contractual relation between the parties, is called a quasi-

⁴²² Assemb. B. 2655, 2023-2024 Reg. Sess. (Cal. 2024), § 3(i).

⁴²³ H. No. 5406, 18th Cong., 1st Sess. (2019), § 2(e).

⁴²⁴ Assemb. B. 2655, 2023-2024 Reg. Sess. (Cal. 2024), § 3(i)(2).

⁴²⁵ COMELEC Res. No. 11064, art. II(13).

⁴²⁶ H. No. 5406, 18th Cong., 1st Sess. (2019), § 2(e).

delict[.]”⁴²⁷ One of the elements for a finding of quasi-delict is the existence of negligence which has been defined as “the omission to do something which a reasonable man, guided by those considerations which ordinarily regulate the conduct of human affairs, would do, or the doing of something which a prudent and reasonable man would do.”⁴²⁸ In other words, the standard requires the court to determine whether or not the standard of care as required by the situation was met by the person—otherwise negligence exists.⁴²⁹ Hence, it requires a circumstantial case-by-case analysis by the court.⁴³⁰

While there is a clearer consensus on what may consist as reasonable actions by a prudent man in cases of quasi-delict, the answer is less certain when it comes to technological matters. Considering the growing sophistication in AI deepfake technology,⁴³¹ it will become increasingly difficult to determine what is believable to a prudent man. Especially considering the reality that the Philippine society is an incubator for misinformation and disinformation with its people susceptible.⁴³²

Apart from the civil law regime of quasi-delict, intellectual property law is another adjacent legal model from which deception, especially from a perceptual standpoint, can be examined. Under the Intellectual Property Code, the dominancy test is primarily employed in order to determine whether or not trademark infringement exists.⁴³³ Specifically, the law deems that the infringement of marks exists when there is “colorable imitation of a registered mark or the same container or a dominant feature thereof in connection with the sale, offering for sale, distribution, advertising of any goods or services[.]”⁴³⁴

As stated in *Skechers, U.S.A., Inc. v. Inter Pacific Industrial Trading Corp.*, the “essential element of infringement [...] is that the infringing mark is likely

⁴²⁷ CIVIL CODE, art. 2176.

⁴²⁸ *Phil, Nat’l Constr. Corp. v. CA*, G.R. No. 159270, 467 SCRA 569, 580, Aug. 22, 2005.

⁴²⁹ *Id.*

⁴³⁰ *See Sanggacala v. Nat’l Power Corp.*, G.R. No. 209538, 992 SCRA 536, 569, July 7, 2021.

⁴³¹ Chesney & Citron, *supra* note 6, at 1759 & n.12&14 (2019); *See supra* Part I(B).

⁴³² *See generally* Aries A. Arugay & Surachanee Sriyai, *From Likes to Lies: Disinformation in the Philippines and Thailand*, FULCRUM, July 7, 2023, at <https://fulcrum.sg/from-likes-to-lies-disinformation-in-the-philippines-and-thailand/>.

⁴³³ INTELL. PROP. CODE, § 155.1; *UFC Phils., Inc. v. Fiesta Barrio Mfg. Corp.*, G.R. No. 198889, 781 SCRA 424, 457, Jan. 20, 2016.

⁴³⁴ § 155.1.

to cause confusion.”⁴³⁵ The dominance test “focuses on the similarity of the prevalent or dominant features of the competing trademarks that might cause confusion, mistake, and deception in the mind of the purchasing public.”⁴³⁶ Under this test, “[c]ourts will consider more the aural and visual impressions created by the marks in the public mind[.]”⁴³⁷

In this sense, trademark law and electoral law find a convergence with the comparable aim of the protection of the public against any confusion or mistaken perceptions about goods or services, or candidates or political parties, respectively.⁴³⁸ This legal standard may possibly serve as a starting point for the courts to decide on the existence of deception for political deepfakes as it also relies on the distortion of a person’s perception or impression of a particular deepfake media, be it audio or visual in format. Both look at the effects of sensorial manipulations on the public and how it may create a radically different idea in the mind of the viewing public.

In relation, deception as a legal standard may also draw on the established legal standards of unfair competition in intellectual property law. Under the Intellectual Property Code, there is unfair competition when “[a]ny person who shall employ deception or any other means contrary to good faith by which he shall pass off the goods manufactured by him or in which he deals, or his business, or services for those of the one having established such goodwill, or who shall commit any acts calculated to produce said result[.]”⁴³⁹ In jurisprudence, unfair competition is simply the “passing off (or palming off) or attempting to pass off upon the public of the goods or business of one person as the goods or business of another with the end and probable effect of deceiving the public.”⁴⁴⁰ Elaborating on the mechanics of unfair competition, the Court stated that “[p]assing off (or palming off) takes place where the defendant, by imitative devices on the general appearance of the goods, misleads prospective purchasers into buying his merchandise under the impression that they are buying that of his competitors.”⁴⁴¹ Like trademark infringement, there is the positive intention

⁴³⁵ *Skechers, U.S.A., Inc. v. Inter Pacific Indus. Trading Corp.* [hereinafter “*Skechers*”], G.R. No. 164321, 646 SCRA 448, 455, Mar. 23, 2011.

⁴³⁶ *Id.*

⁴³⁷ *McDonald’s Corp. v. L.C. Big Mak Burger, Inc.*, G.R. No. 143993, 437 SCRA 10, 32, Aug. 18, 2004.

⁴³⁸ *See Skechers*, 646 SCRA at 461.

⁴³⁹ INTELL. PROP. CODE, § 168.2.

⁴⁴⁰ *Republic Gas Corp. v. Petron Corp.*, G.R. No. 194062, 698 SCRA 666, 680, June 17, 2013.

⁴⁴¹ *Id.*

to mislead the public by creating a false and erroneous impression or perception about the goods or services.

Similar to the materially deceptive standard, unfair competition looks to the effects of changes made by the perpetrator on the viewing public such that they generate mistaken or false perceptions about the good or service. Hence, unfair competition as a legal standard may also be informative in drawing future conclusions about deception created by political deepfakes as both legal regimes anchor their standards of harm on the consequences of some form of alteration on the conception or understanding of the viewing public.

Additionally, while COMELEC Resolutions do not provide for exclusions to materially deceptive content, they already distinctly define softfakes, deepfakes, and cheapfakes. The distinctions between these kinds of media hinge on the use of AI technology and perceptibility of changes to the media.

This definition of materially deceptive content and corresponding exceptions continue to appear in other different California legislation such as Assembly Bill 2839 or “Elections: Deceptive Media in Advertisements”⁴⁴² and Assembly Bill 972.⁴⁴³

Other states such as Michigan also provide for a legal standard for “materially deceptive media,” to wit:

- (c) “Materially deceptive media” means any image, audio, or video that meets all of the following requirements:
- (i) Falsely depicts an individual engaging in speech or conduct in which the depicted individual did not in fact engage.
 - (ii) A reasonable viewer or listener would incorrectly believe that the depicted individual engaged in the speech or conduct depicted.
 - (iii) Was produced by artificial intelligence as that term is defined in section 2 of the Michigan campaign finance act, 1976 PA 388, MCL 169.202.⁴⁴⁴

⁴⁴²Assemb. B. 2839, 2023–2024 Reg. Sess. (Cal. 2024).

⁴⁴³Assemb. B. 972, 2021–2022 Reg. Sess. (Cal. 2022).

⁴⁴⁴H.B. 5144, 102nd Leg., Reg. Sess. (Mich. 2023), § 932(f)(1).

Though the provision differs from the verbatim formulation of California legislation,⁴⁴⁵ the provision from Michigan’s House Bill No. 5144 contains similar elements such as the false depiction of a person, the use of AI technology, and the ability to mislead a reasonable person as to the contents of the media.⁴⁴⁶ Other states which possess similar provisos are Alabama⁴⁴⁷ and Hawaii.⁴⁴⁸

Remarkably, some states such as New Mexico also include the qualification of the depicted subject’s lack of consent,⁴⁴⁹ a requirement which is not echoed in the “materially deceptive” standard in the Biazon Bill. However, it creates a separate provision in order to highlight that that unlawful act to be punished is the creation and disclosure of the deepfake or materially deceptive media without the depicted individual’s consent.⁴⁵⁰

As to the exceptions to a finding of liability, the Biazon Bill provides that:

A person shall not be liable under the preceding section in either of the following circumstances:

- a. The person discloses the material that contains a video recording of a depicted individual or a materially deceptive audio or video recording in the course of any of the following:
 - (i) Reporting unlawful activity;
 - (ii) Exercising the person's law enforcement duties; or
 - (iii) Hearings, trials, or other legal proceedings.
- b. The deepfake material or materially deceptive audio or video recording is a commentary, criticism, or disclosure that is otherwise protected by the Constitution.⁴⁵¹

The following exceptions in disclosure are reasonable and practical measures which echo other existing laws such as the safe harbor provision in the Anti-Money Laundering Act⁴⁵² and the exception to attorney-client

⁴⁴⁵ See Assemb. B. 2655, 2023-2024 Reg. Sess. (Cal. 2024), § 3(i).

⁴⁴⁶ H.B. 5144, 102nd Leg., Reg. Sess. (Mich. 2023), § 932(f)(1).

⁴⁴⁷ H.B. 172, 2024 Reg. Sess. (Ala. 2024), § 1(4).

⁴⁴⁸ S.B. 2687, 32nd Leg., Reg. Sess. (Haw. 2024), § 2.

⁴⁴⁹ H.B. 182, 56th Leg., 2nd Reg. Sess. (N.M. 2024), § 1(S).

⁴⁵⁰ H. No. 5406, 18th Cong., 1st Sess. (2019), § 3.

⁴⁵¹ H. No. 5406, 18th Cong., 1st Sess. (2019), § 4.

⁴⁵² See Rep. Act No. 9160 (2001), § 9(c).

privilege in reporting the procurement of legal advice in “furtherance of crime or fraud.”⁴⁵³ It is also similar to the safe harbor provision in the Anti-Online Sexual Abuse or Exploitation of Children and Anti-Child Sexual Abuse or Exploitation Materials Act (“Anti-OSAEC-CSAEM Act”) which exempts the viewer or disseminator from any liability provided that the access, recording, or possession is for legitimate purposes (e.g. reporting of unlawful activity or policy).⁴⁵⁴

The last proviso of Section 4 operates as a catch-all to ensure that legitimate forms of expression are not unlawfully penalized.⁴⁵⁵ The proviso is similar to California Assembly Bill 972 which includes satire and parody in the list of exceptions to materially deceptive content.⁴⁵⁶

The Bill is ultimately a general piece of legislation on deepfakes and zones in on the non-consensual creation and disclosure of the deepfake by classifying it as materially deceptive media.⁴⁵⁷ Unlike COMELEC Resolution No. 11064 and even Resolution No. 10730, it does not aim to apply only in the electoral context. Nonetheless, the bill is instructive as it provides a specific legal standard for determining liability by looking at the factors of consent and the ultimate deceptive effect of the deepfake on a reasonable person.⁴⁵⁸ These factors are evidently absent in existing COMELEC rules and regulations.

2. House Bill No. 9425

Representatives Lani Mercado-Revilla, Bryan B. Revilla, and Ramon Jolo Revilla III introduced another bill on deepfakes, House Bill No. 9425 or the “Penalizing Deepfake Act” (“Revilla Bill”) on October 25, 2024.⁴⁵⁹ The bill underscores the correlative harms that deepfakes may usher in and aims to “condemn and combat the inappropriate and malicious use of these technologies, especially in instances involving the dissemination of disinformation, harassment, incitement of violence, and identity fraud.”⁴⁶⁰

⁴⁵³ RULES OF COURT, Rule 130, § 24(b)(i).

⁴⁵⁴ H. No. 9425, 19th Cong., 2nd Sess. (2023), § 8.

⁴⁵⁵ H. No. 5406, 18th Cong., 1st Sess. (2019), § 4(b).

⁴⁵⁶ Assemb. B. 972, 2021–2022 Reg. Sess. (Cal. 2022), § 3.

⁴⁵⁷ H. No. 5406, 18th Cong., 1st Sess. (2019), §§ 2(e) & 3.

⁴⁵⁸ §§ 2(e) & 3.

⁴⁵⁹ H. No. 9425, 19th Cong., 2nd Sess. (2023).

⁴⁶⁰ § 2.

The bill defined deepfakes as “any audio, visual or audio-visual recording created or altered through technical means, such as video recording, motion-picture film, sound recording, electronic image, or photograph, which are so convincing that a reasonable person would mistake it for an authentic representation of an individual’s speech or conduct.”⁴⁶¹ Similar to the Biazon Bill, this definition recognizes that the output media may be different formats such as video or images, or even sound.⁴⁶² However, unlike the Biazon Bill, there is no qualification of the effect that a “reasonable person must mistake [the media] for an authentic representation of an individual’s speech.”⁴⁶³ The proviso exists in a comparable formulation to the Biazon Bill’s standard of “materially deceptive audio or video recording.”⁴⁶⁴

In relation to the definition of deepfakes in COMELEC Resolution No. 11046, the Revilla Bill also underscores the dangerous propensity of the technology to mislead. Hence the definition in Resolution No. 11046 specifies that deepfakes falsely make it appear that an action statement, or event transpired but did not occur in reality.⁴⁶⁵ However, Resolution No. 11046 goes further than the Revilla Bill by further differentiating deepfakes from cheapfakes and softfakes. Cheapfakes being rudimentary manipulated media forms which need not necessarily use AI.⁴⁶⁶ Whereas, softfakes are distinguished from the two as COMELEC acknowledges that these media contain less conspicuous edits that often come with a political bent.⁴⁶⁷

In contrast to COMELEC Resolution No. 11046, the Revilla Bill provides explicit exceptions to “harmful deepfakes” such as:

- (1) When the depicted individual was knowingly aware of the creation and commercial distribution of a deepfake and he or she provided his or her consent, except in cases involving children, in accordance with Section 5 of this Act;
- (2) When the deepfake serves artistic purposes, including parody, satire, commentary or criticism; and
- (3) When the dissemination of the deepfake is carried out as part of an investigation, prosecution, or other lawful proceedings,

⁴⁶¹ § 3.

⁴⁶² H. No. 9425, 19th Cong., 2nd Sess. (2023), § 3; H. No. 5406, 18th Cong., 1st Sess. (2019), § 2(b).

⁴⁶³ H. No. 9425, 19th Cong., 2nd Sess. (2023).

⁴⁶⁴ H. No. 5406, 18th Cong., 1st Sess. (2019), § 2(e).

⁴⁶⁵ COMELEC Res. No. 11064, art. II(13).

⁴⁶⁶ Art. II(7).

⁴⁶⁷ Art. II(26).

and the person acts in good faith to minimize the risk of further dissemination.⁴⁶⁸

Thus, consent of the depicted individual is an exculpatory factor. The only exception is with respect to unlawful deepfakes that involve and/or depict children.⁴⁶⁹ Likewise, the provision ensures that deepfakes which comprise legitimate forms of speech fall beyond the Bill's regulatory reach. The inclusion of this exception is consistent with existing Supreme Court rulings that have stated that the freedom of expression "extends protection to nearly all forms of communication."⁴⁷⁰ It reinforces the delineation between harmful deepfakes which it then deems beyond the scope of Constitutional protection and forms of protected deepfake speech like "parody, satire, commentary or criticism."⁴⁷¹

The last exception to the definition of harmful deepfakes echoes the exceptions outlined in the Biazon Bill on the reporting of unlawful activity, exercise of legal duties, hearings, trials, and other proceedings.⁴⁷² This exception also parallels safe harbor provisions in other existing laws such as the Anti-Money Laundering Act⁴⁷³ and the Anti-OSAEC-CSAEM Act.⁴⁷⁴

Compared to the Biazon Bill, it is evident that the Revilla Bill is wider in regulatory scope as it aims to punish several forms of deepfake misuse apart from non-consensual creation and disclosure.⁴⁷⁵ The Revilla Bill also aims to penalize the use of deepfakes to create "defamatory content" in line with the relevant provisions on slander in the RPC.⁴⁷⁶ It is interesting that the legislators have likened the creation of defamatory deepfake content with slander considering that it is typically considered to be "oral defamation."⁴⁷⁷ As elaborated by the Court in *Villanueva v. People*: "Slander is libel committed by oral (spoken) means, instead of in writing. The term oral defamation or slander[,] as now understood, has been defined as the speaking of base and defamatory words which tend to prejudice another in his reputation, office, trade, business or means of livelihood."⁴⁷⁸ Taking into consideration the

⁴⁶⁸ H. No. 9425, 19th Cong., 2nd Sess. (2023), § 13.

⁴⁶⁹ H. No. 5406, 18th Cong., 1st Sess. (2019), § 6.

⁴⁷⁰ *Chavez*, 545 SCRA at 485.

⁴⁷¹ H. No. 9425, 19th Cong., 2nd Sess. (2023), § 13.

⁴⁷² H. No. 5406, 18th Cong., 1st Sess. (2019), § 4.

⁴⁷³ Rep. Act No. 9160, § 9(c).

⁴⁷⁴ H. No. 9425, 19th Cong., 2nd Sess. (2023), § 8.

⁴⁷⁵ H. No. 5406, 18th Cong., 1st Sess. (2019), § 3.

⁴⁷⁶ H. No. 9425, 19th Cong., 2nd Sess. (2023), § 4.

⁴⁷⁷ REYES, *supra* note 367, at 1286.

⁴⁷⁸ *Villanueva v. People*, G.R. No. 160351, 487 SCRA 42, 53, Apr. 10, 2006.

common mode of proliferation of deepfake content, which is through online and/or social media platforms, the proper denomination should instead link the crime of defamatory deepfake content to libel, specifically cyberlibel since the distribution on social media is a mode of publication.⁴⁷⁹ It would likewise be improper to classify the use of defamatory deepfake content as “slander by deed” under Article 359 of the RPC since one of the elements of the crime is that “the offender performs any act not included in any other crime against honor.”⁴⁸⁰

And the use of defamatory deepfake content would be more aptly classified in relation to libel which possesses the following elements:

1. That there must be an imputation of a crime, or of a vice, or defect, real or imaginary, or any act, omission, status, or circumstance.
2. That the imputation must be made publicly.
3. That it must be malicious.
4. That the imputation must be directed to a natural or juridical person, or one who is dead.
5. That the imputation must tend to cause the dishonor, discredit or contempt of the person defamed.⁴⁸¹

The Cybercrime Prevention Act’s definition of cyberlibel also supports the contention that defamatory statements made online through the use of deepfakes would constitute libel as it was “committed through a computer system or any other similar means which may be devised in the future.”⁴⁸² The current formulation of the Cybercrime Prevention Act and its proviso of “any other similar means” would suggest that deepfake AI technology would fall under the law as it is similarly made “through a computer system” by statutory construction.⁴⁸³ Specifically, through the principle of *eiusdem generis* or “when general words follow an enumeration of particular cases, such words apply only to cases of the same kind as those expressly mentioned,”⁴⁸⁴ since the creation of a deepfake requires the use of AI technology which is typically accessed through a computer system,⁴⁸⁵ the law includes it within its regulatory reach.

⁴⁷⁹ Rep. Act No. 10175 (2012), § 4(b)(4); REV. PEN. CODE, art. 353–54.

⁴⁸⁰ REYES, *supra* note 367, at 1289.

⁴⁸¹ *Id.* at 1239.

⁴⁸² Rep. Act No. 10175 (2012), § 4(b)(4).

⁴⁸³ § 4(b)(4); *See supra* Part I(B).

⁴⁸⁴ *City of Manila v. Entote*, G.R. No. 24776, 57 SCRA 497, 508, June 28, 1974.

⁴⁸⁵ *See supra* Part I(B).

The Court in *Causing v. People*, furthermore, sustained this theory by expressly stating that: “RA 10175 did not create a new crime but merely implements the RPC’s provisions on libel when written defamatory remarks are published through a computer system.”⁴⁸⁶ Thus, the creation and proliferation of defamatory deepfake content would find more solid legal footing under the provisions on libel, specifically cyberlibel.

Apart from defamatory content, the Revilla Bill provides that: “[g]ender-based sexual harassment carried out through the use of deepfake technology shall be defined in Section 12 of Republic Act No. 11313, also known as the ‘Safe Spaces Act,’ addressing gender-based online sexual harassment.”⁴⁸⁷ Relatedly, the Safe Spaces Act defines “gender-based online sexual harassment” as:

[O]nline conduct targeted at a particular person that causes or likely to cause another mental, emotional or psychological distress, and fear of personal safety, sexual harassment acts including unwanted sexual remarks and comments, threats, uploading or sharing of one’s photos without consent, video and audio recordings, cyberstalking and online identity theft[.]⁴⁸⁸

The implementing rules and regulations of the law spell out the acts constituting the offense which include acts like “[u]ploading and sharing without the consent of the victim any form of media that contains photos, voice, or video with sexual content” and the unauthorized “sharing of any of the victim’s photos, videos or any information online[.]”⁴⁸⁹

While the creation and distribution of non-consensual deepfakes is a pressing issue,⁴⁹⁰ the provision in the Revilla Bill may prove to be redundant as the Safe Spaces Act’s definition of gender-based online sexual harassment already contemplates the dissemination of non-consensual deepfakes that “causes or likely to cause another mental, emotional or psychological distress, and fear of personal safety[.]”⁴⁹¹ The creation of the deepfake using source input media of the person depicted (e.g. photos or videos) and its dissemination without the subject’s consent unequivocally constitutes

⁴⁸⁶ *Causing*, G.R. No. 258524, slip op. at 12.

⁴⁸⁷ H. No. 9425, 19th Cong., 2nd Sess. (2023), § 5.

⁴⁸⁸ Rep. Act. No. 11313 [hereinafter “Safe Spaces Act”] (2019), § 3(e).

⁴⁸⁹ Safe Spaces Act, § 12; Rep. Act. No. 11313 Rules & Regs. (2019), §§ 13(c)–(d).

⁴⁹⁰ See Chesney & Citron, *supra* note 6, at 1773.

⁴⁹¹ Safe Spaces Act, § 3(e).

gender-based online sexual harassment provided that the intention of the creator was to cause distress to the victim.⁴⁹²

In a similar manner, although the Revilla Bill provides a provision for the unlawful creation and distribution of non-consensual deepfakes involving children that references the Anti-OSAEC-CSAEM Act,⁴⁹³ the latter law already serves as the proper *lex specialis* on the subject.⁴⁹⁴ The law includes the depiction of “real or simulated sexual activities or depicting acts of sexual abuse or exploitation of a child as a sexual object” as “child sexual abuse or exploitation material or child sexual abuse material (“CSAEM”/”CSAM”).”⁴⁹⁵

Moreover, the Act defines “sexual activity” by incorporating acts “real or simulated.”⁴⁹⁶ Notably, the Anti-OSAEC-CSAEM Act explicitly prohibits “image-based sexual abuse” which “refers to a form of technology-facilitated sexual violence” and comprises “the use of artificial intelligence to construct ‘deepfake’ pornographic videos.”⁴⁹⁷ In sum, the Anti-OSAEC-CSAEM Act is sufficient on its own to address the creation and distribution of unlawful deepfakes involving and/or depicting children.

The Revilla Bill provides a similar blueprint for other unlawful acts involving deepfakes for “deceptive consumer transactions” in relation to the Consumer Act of the Philippines and E-Commerce Act,⁴⁹⁸ copyright infringement under the Intellectual Property Code of the Philippines,⁴⁹⁹ other cyber fraudulent practices under the Cybercrime Prevention Act,⁵⁰⁰ and even terrorism in relation to the Anti-Terrorism Act.⁵⁰¹

What is remarkable, however, about the Revilla Bill is its provision on the punishment of the “deployment of fraudulent devices during an election period.”⁵⁰² It defines this election offense in relation to “Section 261(c) of Batas Pambansa Bilang 881, also known as the “Omnibus Election

⁴⁹² See *supra* Part I(B); Safe Spaces Act, § 3(e).

⁴⁹³ H. No. 9425, 19th Cong., 2nd Sess. (2023), § 6.

⁴⁹⁴ See Rep. Act No. 11930 (2022).

⁴⁹⁵ Rep. Act No. 11930 (2022), §. 3(c). (Emphasis supplied.)

⁴⁹⁶ § 3(y).

⁴⁹⁷ § 3(j).

⁴⁹⁸ H. No. 9425, 19th Cong., 2nd Sess. (2023), § 8.

⁴⁹⁹ § 9.

⁵⁰⁰ § 12.

⁵⁰¹ § 11.

⁵⁰² § 10.

Code of the Philippines.”⁵⁰³ The Bill cites Section 261(c) of the OEC but may instead have intended to reference Section 261(e) on “[t]hreats, intimidation, terrorism, use of fraudulent device or other forms of coercion.”⁵⁰⁴ The election offense as defined in the OEC provides:

Any person who, directly or indirectly, threatens, intimidates or actually causes, inflicts or produces any violence, injury, punishment, damage, loss or disadvantage upon any person or persons or that of the immediate members of his family, his honor or property, or uses *any fraudulent device or scheme* to compel or induce the registration or refraining from registration of any voter, or the participation in a campaign or refraining or desistance from any campaign, or the casting of any vote or omission to vote, or any promise of such registration, campaign, vote, or omission therefrom.⁵⁰⁵

Even with a dearth of jurisprudence on this specific election offense, the title of the section is explicit that the essence of the crime lies in the coercion of another person.⁵⁰⁶ A parallel can thus be drawn with the crime of grave coercion under the RPC.⁵⁰⁷ Like the crime of grave coercion, the perpetrator can commit the offense through compelling or preventing a person from performing an act.⁵⁰⁸ Further, the crime under the RPC contemplates that the coercion may be effected by means of violence or material force.⁵⁰⁹

Unlike grave coercion in the RPC, however, the OEC offense is specifically geared towards elections and campaigns. Additionally, in the RPC, to constitute a grave coercion, the act prevented must “not [be] prohibited by law” or the act compelled must have been done against their will, “be it right or wrong.”⁵¹⁰ Whereas, it is clear that the compulsion or restraint of electoral or campaign activity is prohibited by law and contravenes fundamental rights.

⁵⁰³ § 10.

⁵⁰⁴ ELECT. CODE, § 261(e).

⁵⁰⁵ § 261(e). (Emphasis supplied.)

⁵⁰⁶ § 261(e).

⁵⁰⁷ REV. PEN. CODE, art. 286.

⁵⁰⁸ REYES, *supra* note 367, at 791.

⁵⁰⁹ REV. PEN. CODE, art. 286; *Barbasa v. Tuquero*, G.R. No. 163898, Dec. 23, 2008.

⁵¹⁰ REV. PEN. CODE, art. 286.

For the offense, the creation and dissemination of a deepfake depicting the subject in an unflattering or compromising manner through imagery, video, or audio media may validly be included within the definition of “fraudulent device or scheme” that can in turn be used to “compel or induce the registration or refraining from registration of any voter, or the participation in a campaign or refraining or desistance from any campaign, or the casting of any vote or omission to vote, or any promise of such registration, campaign, vote, or omission therefrom.”⁵¹¹ In other words, it is feasible that a deepfake may be used as blackmail in order to compel a person to perform or desist from performing the acts outlined in Section 261(e).⁵¹²

3. House Bill No. 10567

Representatives Luis Raymund “Lray” F. Villafuerte Jr. and Miguel Luis “Migz” R. Villafuerte filed the most recent House bill on deepfakes on July 1, 2024, which is entitled the “Deepfake Accountability and Transparency Act” (“Villafuerte Bill”).⁵¹³ It defines deepfakes as:

[A]ny video recording, motion-picture film, sound recording, electronic image, or photograph, or any technological representation of speech or conduct substantially derivative thereof—

(A) which appears to authentically depict any speech or conduct of a person

who did not in fact engage in such speech or conduct; and

(B) the production of which was substantially dependent upon technical means, rather than the ability of another person to physically or verbally impersonate such person.⁵¹⁴

Similar to the definition of deepfakes in COMELEC Resolution No. 11064, the definition in the Villafuerte Bill highlights that there may be different kinds of deepfake media (e.g. audio-visual, audio, or just visual).⁵¹⁵ Moreover, like the definition in Resolution No. 1106, the definition in the Villafuerte Bill underscores that the harm of these deepfakes lies in the possibility of deception such that a viewer may be misled to believe that the subject of the deepfake made a particular statement or performed a particular

⁵¹¹ ELECT. CODE, § 261(e).

⁵¹² See Chesney & Citron, *supra* note 6, at 1772.

⁵¹³ H. No. 10567, 19th Cong., 2nd Sess. (2024).

⁵¹⁴ § 3(a).

⁵¹⁵ COMELEC Res. No. 11064, art. II(13).

action (even if false).⁵¹⁶ Furthermore, both definitions recognize the use of AI technology in the creation of the output media.⁵¹⁷

However, once more, the delineation between cheapfakes,⁵¹⁸ softfakes,⁵¹⁹ and deepfakes is noticeably absent. But, considering the general nature of the bill and the definition of deepfakes which stresses its creation through “technical means” it can be argued that the Villafuerte Bill does not concern itself with the other kinds of lower-level manipulated media such as cheapfakes as the latter do not always employ AI technology.⁵²⁰

Its preeminent feature is its extensive disclosure requirements, to wit:

Any person who, using AI, generative AI system, or any other means or facility, produces or distributes a deepfake, with the intent to distribute such deepfake over the internet or knowledge that such deepfake shall be so distributed, shall ensure such deepfake contains the following disclosures:

A. Audiovisual disclosure – Any deepfake that contains both an audio and a visual element shall include:

- (1) not less than one clearly articulated verbal statement that identifies the record as containing altered audio and visual elements, and a concise description of the extent of such alteration;
- (2) an unobscured written statement in clearly readable text appearing at the bottom of the image throughout the duration of the visual element that identifies the record as containing altered audio and visual elements, and a concise description of the extent of such alteration; and
- (3) a link, icon, or similar tool to signal that the content has been altered by, or is product of, generative artificial intelligence or similar technology.

B. Visual disclosure – Any deepfake exclusively containing a visual element shall include an unobscured written statement in clearly readable text appearing at the bottom of the image throughout the

⁵¹⁶ H. No. 10567, 19th Cong., 2nd Sess. (2024), § 3(a); COMELEC Res. No. 11064, art. II(13).

⁵¹⁷ H. No. 10567, 19th Cong., 2nd Sess. (2024), § 3(a); COMELEC Res. No. 11064, art. II(13).

⁵¹⁸ COMELEC Res. No. 11064, art. II(7).

⁵¹⁹ Art. II(26).

⁵²⁰ Art. II(26).

duration of the visual element that identifies the record as containing altered visual elements, and either—

- (1) a concise description of the extent of such alteration; or
- (2) a clearly visible link, icon, or similar tool to signal that the content has been altered by, or is the product of, generative artificial intelligence or similar technology.

C. Audio disclosure – Any advanced technological false personation records exclusively containing an audio element shall include, at the beginning of such record, a clearly articulated verbal statement that identifies the record as containing altered audio elements and a concise description of the extent of such alteration, and in the event such record exceeds two minutes in length, not less than one additional clearly articulated verbal statement and additional concise description at some interval during each two-minute period thereafter.⁵²¹

The breakdown in the types of disclosures required depending on the media mirrors the requirements in Resolution No. 11064.⁵²² Both disclosure guidelines emphasize the importance of transparency, conspicuousness, and straightforwardness in informing the viewer about not only the fact of alteration but the extent of such.⁵²³

However, considering the comprehensive nature of the disclosure guidelines imposed in the Villafuerte Bill, the argument may be made that the guidelines are unconstitutional. Although the regulations are content-neutral in nature as they only regulate the manner of speech, as pronounced by the Court in *Chavez v. Gonzales*, the regulation must still pass the “intermediate approach.”⁵²⁴ This constitutional test finds a government regulation justified if it is within the government’s constitutional power, it furthers a substantial governmental interest unrelated to suppressing free expression, and it is not greater than is essential to further the interest.⁵²⁵

Jurisprudence also provides that the governmental interest must be “compelling” and include constitutional principles.⁵²⁶ In its declaration of

⁵²¹ H. No. 10567, 19th Cong., 2nd Sess. (2024), § 4.

⁵²² COMELEC Res. No. 11064, art. IV, § 2.

⁵²³ H. No. 10567, 19th Cong., 2nd Sess. (2024), § 4; COMELEC Res. No. 11064, art. IV, § 2.

⁵²⁴ *Chavez*, 545 SCRA at 493–94.

⁵²⁵ *Id.* at 494.

⁵²⁶ *Diocese of Bawolod*, 747 SCRA at 97.

policy, the Villafuerte Bill highlights its intention to ensure that “free speech is carried out in a legitimate and responsible manner.”⁵²⁷ Concededly, some forms of speech may be validly subject to state regulation.⁵²⁸ However, it must be noted that the restriction fails the test as it is “greater than is essential to the furtherance of that interest.”⁵²⁹ Here, the Villafuerte Bill provides no exceptions in its regulatory scope, unlike the Revilla Bill.⁵³⁰

In addition, in contrast to COMELEC Resolution No. 11064, which is primarily geared towards speech related to elections, campaigns, and candidates, the Villafuerte Bill makes the disclosure mandatory for “[a]ny person who, using AI, generative AI system, or any other means or facility, produces or distributes a deepfake, with the intent to distribute such deepfake over the internet or knowledge that such deepfake shall be so distributed[.]”⁵³¹ In other words, provided that a person creates and intends to distribute the deepfake online, the disclosure requirement is mandatory.

The disclosure requirement is generally easier to justify in the context of campaign media as it satisfies the jurisprudential requirements laid out in *Chavez v. Gonzales*.⁵³² COMELEC regulations regarding deepfake campaign media are within COMELEC’s constitutional mandate to “[e]nforce and administer all laws and regulations relative to the conduct of an election[.]”⁵³³ There also exists a substantial government interest to ensure “free, orderly, honest, peaceful, and credible elections.”⁵³⁴ And assuming that regulations do not transgress jurisprudential guidelines on the regulation of free speech (especially political speech),⁵³⁵ then the measures on the use and publication of deepfakes in the context of an election will be upheld.

Even if seemingly innocuous, the disclosure requirement and its enforcement may prove to be difficult considering the sundry of speech and expression that has arisen since the widespread adoption and use of deepfakes. For example, the provision may raise issues for deepfakes created and proliferated online for the purposes of satire or comedy. The Court in *Diocese of Bacolod* took judicial notice of the use of such kinds of expression,

⁵²⁷ H. No. 10567, 19th Cong., 2nd Sess. (2024), § 2.

⁵²⁸ *Chavez*, 545 SCRA at 486.

⁵²⁹ *Id.* at 494.

⁵³⁰ See H. No. 9425, 19th Cong., 2nd Sess. (2023), § 13.

⁵³¹ H. No. 10567, 19th Cong., 2nd Sess. (2024), § 4.

⁵³² *Chavez*, 545 SCRA at 528–35.

⁵³³ CONST., art. IX-C, § 2(1).

⁵³⁴ Art. IX-C, § 4.

⁵³⁵ See *Diocese of Bacolod*, 747 SCRA at 115.

specifically satire, and noted that it usually employs devices such as sarcasm, irony, ridicule, exaggeration, and other similar devices.⁵³⁶ Often, the purpose of these devices is to “hold real-life people and events up for critical evaluation.”⁵³⁷ Hence, these kinds of speech hold an important role in expressing the author’s thoughts and encouraging some form of debate or deliberation.

While there exists the propensity for abuse (as with any kind of speech),⁵³⁸ it is correlatively feasible that in crafting such kinds of speech the lack of disclosure becomes an intentional act by the author to call attention to a particular issue, event, or person (i.e., it is the punchline). That is to say, the use of deepfake technology and the intentional lack of disclosure become conjunctive rhetorical devices to impress upon the audience a particular message. And the inclusion of the disclosure and markings may “neutralize the satire’s impact.”⁵³⁹

This same problem exists in the context of deepfakes created for artistic purposes. At times, the omission of the disclosure regarding the use of AI technology is by design and the inclusion of which would directly undermine the message and/or purpose of the artwork.⁵⁴⁰

Therefore, while seemingly benign, the Villafuerte Bill may result in the stifling of protected forms of expression due to its expansive regulatory reach. This effect rings especially true considering the hefty proposed penalty of 5 million pesos for those who knowingly fail to comply with the required disclosure requirement and 2 million pesos for those who remove or alter the disclosures (or compel another to do so).⁵⁴¹ Further violations will also subject perpetrators to an additional 3 million pesos.⁵⁴²

In any event, the Villafuerte Bill foreshadows and brings to the forefront coming debates about how the state should strike a balance between the regulation of deepfakes (not just in the electoral context) and upholding the right to free speech and expression. As can be evidenced by

⁵³⁶ *Diocese of Bacolod*, 747 SCRA at 103.

⁵³⁷ Henry Ajder & Joshua Glick, *Just Joking! Deepfakes, Satire, and the Politics of Synthetic Media*, CO-CREATION STUDIO AT MIT WEBSITE (2021), at <https://cocreationstudio.mit.edu/just-joking/> (last accessed Dec. 28, 2024).

⁵³⁸ *Id.*

⁵³⁹ *Id.*

⁵⁴⁰ *Id.*

⁵⁴¹ H. No. 10567, 19th Cong., 2nd Sess. (2024), § 5.

⁵⁴² § 5.

the Bill's provisions, finding this middle ground may prove much more difficult in practice.

IV. CRAFTING PROSPECTIVE LEGISLATION

The prevalence, potentials, and pitfalls of the use of deepfake technology especially in the electoral context make the crafting of legislation a national imperative. Even with pending bills about AI and deepfakes, it must be noted that none have hurdled the legislative process to become law. Although there exist COMELEC rules and regulations on the use and dissemination of deepfakes during the campaign period, presently such are insufficient and may even be deemed *ultra vires*.⁵⁴³

Likewise, reliance on the goodwill of private online stakeholders is misplaced considering the forthcoming shift towards deregulation for technology behemoths like Facebook and Instagram which are some of the most prominent online platforms and simultaneously the largest sources of misinformation.⁵⁴⁴ As a result, the creation of nuanced legislation that considers the contours of the technology and its implications on campaign equity would best serve paramount state interests in upholding institutional trust, the marketplace of ideas, political speech, and the democratic system writ large.

As to the legislative model and owing to the serious ramifications that may be brought about by political deepfakes on the electoral system, it is unlikely that the Philippines will resort to the market-based regulatory approach. On the contrary, as evidenced by the genesis and publication of COMELEC Resolution No. 11064, the more likely regulatory attitude will tend towards risk-based regulation.⁵⁴⁵

Like existing risk-based frameworks, Resolution No. 11064 does not favor the comprehensive ban of political deepfakes.⁵⁴⁶ Instead, it implements stringent disclosure guidelines to ensure the mitigation and elimination of forms of deceit which may exist with the creation of deepfakes, *cheapfakes*, or

⁵⁴³ See *supra* Part III(B).

⁵⁴⁴ See Andrew Chow, *Why Meta's Fact-Checking Change Could Lead to More Misinformation on Facebook and Instagram*, TIME, Jan. 7, 2025, at time.com/7205332/meta-fact-checking-community-notes/.

⁵⁴⁵ COMELEC Res. No. 11064.

⁵⁴⁶ COMELEC Res. No. 11064.

softfakes.⁵⁴⁷ These requirements impose mandatory measures on the size of the disclosures, duration, and placement which further vary depending on the kind of manipulated media created and proliferated.⁵⁴⁸ The crafting of such requirements demonstrates the Commission's assessment of risk in relation to its electoral mandate which is part and parcel of the risk-based model.⁵⁴⁹

The propensity for the adoption of risk-based regulation can also be seen in existing draft legislation on deepfakes. Like COMELEC Resolution No. 11064, the Villafuerte Bill imposes comprehensive disclosure guidelines depending on the kind of deepfake media but notably does not prohibit the technology altogether.⁵⁵⁰ The Biazon Bill also follows suit and favors the general authorization of the technology but only prohibits those which fit the criteria of materially deceptive audio or video recordings.⁵⁵¹ Hence, from a normative regulatory standpoint, legislators for this bill make conclusions about the risk brought about by these specific kinds of deepfakes.

Even the Revilla Bill concurs with the aforementioned bills and COMELEC Resolution on the need for limitations on deepfakes but not their wholesale prohibition. Instead, the bill determines that what must be proscribed are "harmful deepfakes."⁵⁵² It also enjoins "defamatory content,"⁵⁵³ "[g]ender-based sexual harassment,"⁵⁵⁴ deepfakes involving children,⁵⁵⁵ "deceptive consumer transactions,"⁵⁵⁶ copyright infringement,⁵⁵⁷ other cyber fraudulent practices,⁵⁵⁸ and terrorism.⁵⁵⁹ For political deepfakes, it even provides a specific provision for election offenses.⁵⁶⁰

⁵⁴⁷ COMELEC Res. No. 11064, art. IV, §§ 1–2.

⁵⁴⁸ Art. IV, § 2.

⁵⁴⁹ See Cary Coglianese, *What Does Risk-Based Regulation Mean?*, REG. REV., July 8, 2019, at www.theregreview.org/2019/07/08/coglianese-what-does-risk-based-regulation-mean; Julia Black, *Risk-based Regulation: Choices, Practices and Lessons Being Learnt*, in RISK AND REGULATORY POLICY: IMPROVING THE GOVERNANCE OF RISK 188 (2010).

⁵⁵⁰ H. No. 10567, 19th Cong., 2nd Sess. (2024), § 4.

⁵⁵¹ H. No. 5406, 18th Cong., 1st Sess. (2019), § 2(e).

⁵⁵² H. No. 9425, 19th Cong., 2nd Sess. (2024), § 13.

⁵⁵³ § 4.

⁵⁵⁴ § 5.

⁵⁵⁵ § 6.

⁵⁵⁶ § 8.

⁵⁵⁷ § 9.

⁵⁵⁸ § 12.

⁵⁵⁹ § 11.

⁵⁶⁰ § 10.

Together, these rules and proposed laws provide a necessary glimpse of regulatory attitudes towards deepfakes and their constituent risks. Even with their respective drawbacks,⁵⁶¹ these instruments undoubtedly point towards a step in the right direction in the regulatory prioritization of AI technology and deepfakes such that regulators are privy to existing regulatory gaps. And the adoption of a risk-based model as opposed to a pure market-based model would avoid holding such a significant social exercise hostage to corporate vagaries and competing objectives.⁵⁶²

However, as it stands, a law on the regulation of deepfakes, especially ones political in nature has not yet come to fruition. Considering the fast-approaching midterm elections in 2025, there is an appreciable urgency for the creation of national legislation providing clear-cut guidelines and avoiding the limitations of existing COMELEC Resolutions.

While the choices of the subject and manner of regulation are factors that are ultimately subject to the wisdom of Congress,⁵⁶³ the production of tailored legislation geared specifically towards political deepfakes as opposed to general legislation on AI would ensure that measures best address the contours of the technology and respond to its constitutive issues. After all, the importance of the electoral process in the democratic system cannot be underscored enough.⁵⁶⁴ As such, the protection of the process which helps determine the will of the electorate and eventual leaders must be afforded primacy as it is directly tied to the people's right to sovereignty.⁵⁶⁵

Symbiotic to this right to sovereignty is the right to free speech and expression which affords people the ability to comment, criticize, and participate in public affairs.⁵⁶⁶ In turn, a product of this collective exercise of rights is the marketplace of ideas, where it is theorized that the simultaneous and consecutive exchanges between individuals are a means of arbitrating collective social truths and values.⁵⁶⁷ Correlatively, this right to deliberate and

⁵⁶¹ See *supra* Part III (B) & (D).

⁵⁶² See Chow, *supra* note 544; See also Sigal Samuel, *Open AI as we knew it is dead*, VOX, Sept. 27, 2024, at <https://www.vox.com/future-perfect/374275/openai-just-sold-you-out>.

⁵⁶³ See *Garcia v. Exec. Sec'y*, G.R. No. 157584, 583 SCRA 119, 133, Apr. 2, 2009.

⁵⁶⁴ See *supra* Part II(A).

⁵⁶⁵ *Macalintal*, G.R. No. 263590, slip op. at 13.

⁵⁶⁶ *United States v. Bustos*, 37 Phil. 731, 740 (1918); See *supra* Part II(B).

⁵⁶⁷ See *Diocese of Bacolod*, 747 SCRA at 78–79; See also *ABS-CBN Corp. v. Ampatuan*, G.R. No. 227004, Apr. 25, 2023.

remark on public affairs also extends to the ability to comment on matters of governance, elections, political candidates, and political parties.⁵⁶⁸

Without targeted legislation regulating political deepfakes, this delicate balance of rights and the democratic system becomes susceptible to the modern ills of disinformation.⁵⁶⁹ Furthermore, the marketplace of political ideas becomes littered with low-value or even harmful speech.⁵⁷⁰ This situation deteriorates with the introduction of realistic deepfakes whose primary motives are to deceive its audience into believing some counterfactual version of reality.⁵⁷¹ In the political arena, the use of this tool, if left unregulated, possesses the propensity to wreak widespread havoc on the electoral system and foster continuing political polarization.⁵⁷²

Thus, any law that aims to regulate political deepfakes must strike a careful balance between state interests and constitutional rights. After all, jurisprudence has confirmed that the freedom of speech and expression enjoys a preferred status in the proverbial hierarchy of rights.⁵⁷³ Moreover, this right extends to “nearly all forms of communication.”⁵⁷⁴ Therefore, restrictions to this right would thus be subject to the strict scrutiny test by the Court.⁵⁷⁵

If laws on political deepfakes are passed which are in the nature of content-based restrictions or endeavor to censor the subject matter of the speech, the state must be able to overcome the clear and present danger test to justify the burden on speech.⁵⁷⁶ Specifically, the State must substantiate the “gravity and the imminence of the threatened harm.”⁵⁷⁷ The Supreme Court has professed that these types of restrictions carry with them a heavy presumption against their validity.⁵⁷⁸

⁵⁶⁸ See *supra* Part II(C); *Buckley v. Valeo*, 424 U.S. 1, 52 (1976); *Diocese of Bacolod*, 747 SCRA at 77–78.

⁵⁶⁹ See *supra* Part II(D); See also *ABS-CBN Corp. v. Ampatuan*, G.R. No. 227004, slip op. at 92–93, *citing* *HASEN*, *supra* note 131.

⁵⁷⁰ See *supra* Part II(D); See also *id.*

⁵⁷¹ Walker, Schiff, & Schiff, *supra* note 149; A.W. Ohlheiser, *supra* note 149; Marchal et al., *supra* note 28, at 16–17; Verma & Oremus, *supra* note 149.

⁵⁷² See *supra* Part II(D).

⁵⁷³ *Chavez*, 545 SCRA at 481.

⁵⁷⁴ *Id.* at 485.

⁵⁷⁵ *Id.* at 530.

⁵⁷⁶ See *id.* at 495.

⁵⁷⁷ *Id.*

⁵⁷⁸ *Id.* at 492; *Nicolas-Lewis*, 913 SCRA at 553.

Whereas if legislation is only in the nature of content-neutral regulation or aims to govern the incidents of speech (e.g. time, place, or manner), then the state only requires a substantial government interest and is only subject to the intermediate approach.⁵⁷⁹ Thus, limitations on the incidents of political deepfakes will be sanctioned provided that they “promote an important or significant governmental interest that is unrelated to the suppression of expression.”⁵⁸⁰ In any event, “the incidental restriction on speech must be no greater than what is essential to the furtherance of that interest.”⁵⁸¹ Likewise, regulations must be reasonable and narrowly tailored in relation to regulatory objectives and must be within the state’s constitutional powers.⁵⁸²

In light of the disclosure requirements under COMELEC Resolution No. 11064⁵⁸³ and the Villafuerte Bill,⁵⁸⁴ it is more likely that regulators will favor content-neutral regulation for political deepfakes rather than their outright prohibition. In these cases, legislation must still adhere to the established jurisprudential guidelines on content-neutral restrictions or bear the risk of being stricken down by the Court. Valid content-neutral regulation would indeed include the disclosure guidelines that provide specifications for the kind, manner, size, and even period of such disclosures.⁵⁸⁵ The disclosure requirements in Resolution No. 11064⁵⁸⁶ provide a concrete example for future legislators and avoid the danger of being deemed overbroad like in the Villafuerte Bill.⁵⁸⁷

It may even require transparency of the sponsors of the deepfake and whether such has been paid or given gratis.⁵⁸⁸ This requirement can even contemplate the authenticity of the registered social media accounts of political parties, candidates, and exceptionally private citizens as it does not pertain to the subject matter of such speech.⁵⁸⁹

⁵⁷⁹ *Chavez*, 545 SCRA 493–94.

⁵⁸⁰ *Id.* at 494.

⁵⁸¹ *Id.* at 496.

⁵⁸² *Id.* at 524.

⁵⁸³ COMELEC Res. No. 11064, art. IV, §§ 1–2.

⁵⁸⁴ H. No. 10567, 19th Cong., 2nd Sess. (2024), § 4.

⁵⁸⁵ COMELEC Res. No. 11064, art. IV, §§ 1–2; H. No. 10567, 19th Cong., 2nd Sess. (2024), § 4.

⁵⁸⁶ COMELEC Res. No. 11064, art. IV, § 2.

⁵⁸⁷ H. No. 10567, 19th Cong., 2nd Sess. (2024), § 4.

⁵⁸⁸ *See* ELECT. CODE, § 84; *See also* COMELEC Res. No. 10730, § 9(c)(6)–(d); *See also* *Badoy v. Ferrer*, G.R. No. 32546, 35 SCRA 285, Oct. 17, 1970.

⁵⁸⁹ COMELEC Res. No. 10730, § 9; COMELEC Res. No. 11064, art. III; *See Diocese of Bacolod*, 747 SCRA at 94..

Conversely, while the Villafuerte Bill pursues well-intentioned objectives of promoting transparency and addressing the potential deception of deepfakes, its overly broad scope runs contrary to established jurisprudence and constitutes an unconstitutional restriction on political speech.⁵⁹⁰ However, as the bill has not yet been passed by Congress, it can serve as a potent example of the fragile and convoluted constitutional considerations that surround deepfakes as forms of speech and expression. On the other hand, the disclosure requirements in Resolution No. 11064 may provide a better framework which forwards parallel state interests without breaching Constitutional mandates.⁵⁹¹

In any event, studies have found that fact-checking claims which aim towards achieving the same objectives as disclosure requirements (i.e. transparency and avoiding audience deception or misperception) have been effective in combating misinformation especially for more polarizing issues.⁵⁹² Likewise, requiring the disclosure at the point of the deepfakes' initial publication will counteract the phenomenon of the corresponding decrease in effectivity that occurs when similar acts like fact-checking are put off.⁵⁹³

In addition, laws regulating political deepfakes as forms of speech and expression must neither be overbroad nor vague. A statute is overbroad when it “unnecessarily sweep[s] its subject broadly, thereby invading the area of protected freedoms.”⁵⁹⁴ While a statute is considered vague when “it lacks comprehensible standards that men of common intelligence must necessarily guess at its meaning and differ as to its application.”⁵⁹⁵

On this point, COMELEC Resolution 11064 provides a cogent reminder to prospective legislators that they must temper their zeal against the tangible dangers of political deepfakes with respect for inviolable Constitutional rights. Like in the case of *Nicolas-Lewis*, the Supreme Court

⁵⁹⁰ See *supra* Part III(D)(3).

⁵⁹¹ COMELEC Res. No. 11064, art. IV, §§ 1-2.

⁵⁹² See David Adam, *Does Fact-checking Work? What the Science Says*, NATURE, Jan. 10, 2025, at <https://www.nature.com/articles/d41586-025-00027-0>.

⁵⁹³ See Sian Lee et al., “Fact-Checking” Fact Checkers: A Data-driven Approach, MISINFORMATION REVIEW, Oct. 26, 2023, at misinforeview.hks.harvard.edu/article/fact-checking-fact-checkers-a-data-driven-approach.

⁵⁹⁴ *Nicolas-Lewis*, 913 SCRA at 549.

⁵⁹⁵ *Id.*

can sustain a finding of a law being overbroad despite its interest in only regulating the incidents of a kind of speech.⁵⁹⁶

Legislation on political deepfakes must be narrowly tailored to ensure that it does not punish protected forms of speech and in the event that it does provide for punishments, ascertain that the legal standards provided align with existing criminal law standards.⁵⁹⁷ Otherwise, legislation will risk being deemed unconstitutional for being overbroad like COMELEC Resolution No. 11064 which unfortunately punishes the creation and dissemination of deepfakes even if the author may be unaware about the false information contained therein.⁵⁹⁸

Given the widespread proliferation of misinformation and disinformation online, as proliferated online,⁵⁹⁹ lawmakers must avoid the overbreadth of COMELEC Resolution 11064⁶⁰⁰ and instead narrowly tailor legislation to penalize only those political deepfakes that are knowingly created and disseminated with the intent to mislead viewers regarding a political candidate, political party, or the electoral system. In other words, liability should only attach when the creation and distribution of the political deepfake during the campaign period is coupled with malicious intent.

Likewise, legislators must carefully define what are generally considered deepfakes due to the implications on the freedom of speech and expression. Most definitions found in the current bills and regulations generally concur as to the main features of the technology and its technical functions with respect to different kinds of media:

COMELEC Resolution No. 11064	“[D]igitally manipulated images, videos, or audio files created using artificial intelligence to fabricate realistic representations of people, events, or statements and falsely make it appear that an action statement, or event transpired but did not occur in reality.” ⁶⁰¹
House Bill No. 5406 (Biazon Bill)	“[A] technique for human image synthesis based on artificial intelligence. It is used to combine and superimpose existing images and videos onto source

⁵⁹⁶ *Id.* at 554–55.

⁵⁹⁷ *See supra* Part III(C)(1).

⁵⁹⁸ COMELEC Res. No. 11064, art. V, § 1; *See supra* Part III(C)(1); *See Nicolas-Lewis*, 913 SCRA at 554.

⁵⁹⁹ *See supra* Part II(D).

⁶⁰⁰ COMELEC Res. No. 11064, art. II(15) & (23), art. V, § 1.

⁶⁰¹ COMELEC Res. No. 11064, art. II(13).

	images or videos using a machine learning technique known as generative adversarial network.” ⁶⁰²
House Bill No. 9425 (Revilla Bill)	“[A]ny audio, visual or audio-visual recording created or altered through technical means, such as video recording, motion-picture film, sound recording, electronic image, or photograph, which are so convincing that a reasonable person would mistake it for an authentic representation of an individual’s speech or conduct.” ⁶⁰³
House Bill No. 10567 (Villafuerte Bill)	“[A]ny video recording, motion-picture film, sound recording, electronic image, or photograph, or any technological representation of speech or conduct substantially derivative thereof— (A) which appears to authentically depict any speech or conduct of a person who did not in fact engage in such speech or conduct; and (B) the production of which was substantially dependent upon technical means, rather than the ability of another person to physically or verbally impersonate such person.” ⁶⁰⁴

From these above-mentioned definitions, regulators acknowledge the use of technology to alter different forms of media.⁶⁰⁵ This technology may or may not include AI.⁶⁰⁶ Most of the definitions highlight that the purpose of deepfakes is to depict persons, places, events, or statements which may lead its viewer to believe that the media is genuine.⁶⁰⁷ In other words, legislators and regulators already include the element of deception as an integral part of the definition of deepfakes.

The ultimate definition in future legislation settled on by Congress will be well-within its plenary power to legislate.⁶⁰⁸ However, it must be noted unlike COMELEC Resolution No. 11064, House Bills No. 5406,

⁶⁰² H. No. 5406, 18th Cong., 1st Sess. (2019), § 2(b).

⁶⁰³ H. No. 9425, 19th Cong., 2nd Sess. (2023), § 3.

⁶⁰⁴ H. No. 10567, 19th Cong., 2nd Sess. (2024), § 3(a).

⁶⁰⁵ COMELEC Res. No. 11064, art. II(13); H. No. 5406, 18th Cong., 1st Sess. (2019), § 2(b); H. No. 9425, 19th Cong., 2nd Sess. (2023), § 3; H. No. 10567, 19th Cong., 2nd Sess. (2024), § 3(a).

⁶⁰⁶ COMELEC Res. No. 11064, art. II(13); H. No. 5406, 18th Cong., 1st Sess. (2019), § 2(b); H. No. 9425, 19th Cong., 2nd Sess. (2023), § 3.

⁶⁰⁷ COMELEC Res. No. 11064, art. II(13); H. No. 9425, 19th Cong., 2nd Sess. (2023), § 3; H. No. 10567, 19th Cong., 2nd Sess. (2024), § 3(a).

⁶⁰⁸ *Macalintal*, G.R. No. 263590, slip op. at 24.

9425, and 10567 fail to distinguish between deepfakes,⁶⁰⁹ softfakes,⁶¹⁰ and cheapfakes.⁶¹¹ While cheapfakes may be arguably subsumed under the definition of deepfakes,⁶¹² legislators may consider the inclusion of the definition of softfakes as it appears in COMELEC Resolution 11064 due to its applicability in the political and electoral context and should adopt the general features of deepfake technology such as the use of AI,⁶¹³ and the deception of the audience due to the depicted content.⁶¹⁴

Once more, softfakes are “a type of media manipulation that involves subtle and often imperceptible alterations to content, such as editing photos or videos, to influence perceptions or opinions without the more evident distortions characteristic of deepfakes.”⁶¹⁵ This definition highlights the precise harm that deepfakes may generate which is the deception of the viewer or consumer of the media with the goal of imposing a particular idea even though false.

Alternatively, even without providing a definition for softfakes, possible legislation may follow the suit of the Biazon Bill which punishes materially deceptive recordings as:

[I]mage[s] or an audio or video recording[s] of a person's appearance, speech, or conduct that has been intentionally manipulated in a manner such that both of the following conditions are met:

- (1) The image or audio or video recording would falsely appear to a reasonable person to be authentic; and
- (2) The image or audio or video recording would cause a reasonable person to have a fundamentally different understanding or impression of the expressive content of the image or audio or video recording than that person would have if the person were hearing or seeing the unaltered, original version of the image or audio or video recording.⁶¹⁶

⁶⁰⁹ COMELEC Res. No. 11064, art. II(13).

⁶¹⁰ Art. II(26).

⁶¹¹ Art. II(7).

⁶¹² Art. II(7).

⁶¹³ COMELEC Res. No. 11064, art. II(13); H. No. 5406, 18th Cong., 1st Sess. (2019), § 2(b); H. No. 9425, 19th Cong., 2nd Sess. (2023), § 3.

⁶¹⁴ COMELEC Res. No. 11064, art. II(13); H. No. 9425, 19th Cong., 2nd Sess. (2023), § 3; H. No. 10567, 19th Cong., 2nd Sess. (2024), § 3(a).

⁶¹⁵ COMELEC Res. No. 11064, art. II(26).

⁶¹⁶ H. No. 5406, 18th Cong., 1st Sess. (2019), § 2(e).

Akin to the definition of softfakes in COMELEC Resolution No. 11064, the Biazon Bill explicitly enjoins those manipulated media which are created to obfuscate a person's perception of the image, video, or audio media.

Future legislation may also opt to adopt the standard in the Biazon Bill which reckons the deceptive quality of media in relation to the perception of a "reasonable person."⁶¹⁷ However, future legislation should attempt to provide more concrete guidance than the Biazon Bill which only requires the mere effect of having a "fundamentally different understanding or impression" in relation to authenticity.⁶¹⁸ On this front, the definition of deepfakes in COMELEC Resolution No. 11064 may provide a more solid legal standard as it specifies that the specific effect on the viewer is the false appearance that some "action statement, or event transpired but did not occur in reality."⁶¹⁹

This standard in COMELEC Resolution No. 11064 also mirrors the legal standard found in Michigan's House Bill No. 5144 which requires an objectively false depiction of media using AI that would create an incorrect belief in a "reasonable viewer."⁶²⁰ The inclusion of the specific criterion of the creation of a false impression or understanding as opposed to a "different" impression or understanding would better articulate the harm posed by political deepfakes. Likewise, this legal standard in COMELEC Resolution No. 11064 is synergistic with existing corollary legal standards in civil law⁶²¹ and intellectual property law.⁶²²

With the provision of what constitutes materially deceptive media, prospective legislation must correlatively provide for exceptions in line with Constitutional rights. To that end, the Revilla Bill is instructive in deeming what do not constitute "harmful deepfakes," to wit:

- (1) When the depicted individual was knowingly aware of the creation and commercial distribution of a deepfake and he or she provided his or her consent, except in cases involving children, in accordance with Section 5 of this Act;

⁶¹⁷ § 2(e).

⁶¹⁸ H. No. 5406, 18th Cong., 1st Sess. (2019), § 2(e).

⁶¹⁹ COMELEC Res. No. 11064, art. II(13).

⁶²⁰ H.B. 5144, 102nd Leg., Reg. Sess. (Mich. 2023), § 932(f)(1).

⁶²¹ See CIVIL CODE, art. 2176; See *supra* Part III(D)(1).

⁶²² See INTELL. PROP. CODE, §§ 155.1 7, 168.2; See *supra* Part III(D)(1).

- (2) When the deepfake serves artistic purposes, including parody, satire, commentary or criticism; and
- (3) When the dissemination of the deepfake is carried out as part of an investigation, prosecution, or other lawful proceedings, and the person acts in good faith to minimize the risk of further dissemination.⁶²³

The provision on the exceptions to the prohibited kinds of deepfakes would be indispensable to safeguarding protected forms of speech which intersect with deepfake technology (e.g., satire or artistic expression). Moreover, it would ensure that future legislation is resilient to facial challenges for invalid limitations to free speech and expression.⁶²⁴

Specifying the regulatory scope in relation to political deepfakes would also avoid the underlying problem in the Villafuerte Bill as it punishes any individual who generates and publishes deepfakes online without a disclosure as to the utilization of AI technology and even the existence of manipulation of any kind. This broad requirement may not survive when placed under the microscope of judicial review as it may result in the unnecessary hindrance of protected speech.⁶²⁵

Prospective legislation regulating political deepfakes must also grapple with clarifying discrete obligations of stakeholders. First and foremost, it must be reiterated that COMELEC's regulatory powers primarily extend to political candidates and parties during the election period.⁶²⁶ And if COMELEC is similarly empowered by future enabling legislation to regulate the use of political deepfakes employed by even private citizens, such regulation must still comply with the requisites of legality, reasonableness, narrow tailoring, and the use of least restrictive means, as articulated by the Court in *Diocese of Bacolod*.

Hence, the enactment of a law on political deepfakes would also be in the best interest of COMELEC to lay out the metes and bounds of its regulatory power in relation to these. Further, it would clarify existing doubts

⁶²³ H. No. 9425, 19th Cong., 2nd Sess. (2023), § 13.

⁶²⁴ Exec. Sec'y v. Pilipinas Shell Petroleum Corp., G.R. No. 209216, Feb. 21, 2023.

⁶²⁵ See *supra* Part III(D)(3).

⁶²⁶ *Diocese of Bacolod*, 747 SCRA at 114.

about the classification of social media as mass media since the law and Constitution are silent on the matter.⁶²⁷

It would also allow Congress to provide corresponding obligations to social media platforms. While COMELEC Resolution No. 11064 also imposes disclosure requirements on such platforms⁶²⁸ based on COMELEC's classification of social media as mass media,⁶²⁹ the lack of any statutory or constitutional basis places the validity of those obligations in question. However, a law imposing these obligations on social media platforms would be justified as an exercise of the State's plenary police power or "the power to prescribe regulations to promote the health, morals, peace, good order, safety and general welfare of the people."⁶³⁰

After all, the growing problem of political deepfakes disseminated online poses serious ramifications on the electoral process and the meaningful exercise of free speech. The State has the right to regulate such momentous forms of media.⁶³¹ Hence, regulation of political deepfakes will be permitted based on police power subject to the employment of lawful means or measures that are "reasonably necessary to the attainment of the object sought to be accomplished and not unduly oppressive."⁶³²

In the context of elections, future legislation should clarify what specific actions regarding deepfakes will be considered as election offenses. While Section 13 of the Fair Election Act provides a convenient catch-all provision as it considers the violation of the Act and COMELEC rules and regulations as election offenses in relation to the OEC, the distinctive nature of the technology allows for the commission of other acts which should be specifically classified as election offenses.⁶³³

As evidenced by COMELEC Resolution No. 11064, it is feasible that political deepfakes as election propaganda are created for the purpose of:

⁶²⁷ See *supra* Part III(B)(2).

⁶²⁸ COMELEC Res. No. 11064, art. IV, § 3.

⁶²⁹ COMELEC Res. No. 10730, § 2(10); COMELEC Res. No. 11064, art. II(21).

⁶³⁰ *Erimita-Malate Hotel and Motel Operators Ass'n, Inc. v. Mayor of Manila*, G.R. No. 24693, 20 SCRA 849, 859–60, July 31, 1967.

⁶³¹ See *supra* Part II(C)-(D).

⁶³² *Dep't of Educ., Culture, and Sports v. San Diego*, G.R. No. 89572, 180 SCRA 533, 537, Dec. 21, 1989.

⁶³³ Rep. Act No. 9006, § 13.

[D]isrupting or obstructing the election process or causing confusion among the voters, [propagating] false and alarming reports or information or [transmitting] or [circulating] false orders, directives or messages regarding any matter relating to the printing of official ballots, the postponement of the election, the transfer of polling place or the general conduct of the election.⁶³⁴

In relation, Section 261(z)(11) of the OEC states that:

Any person who, for the purpose of disrupting or obstructing the election process or causing confusion among the voters, propagates false and alarming reports or information or transmits or circulates false orders, directives or messages regarding any matter relating to the printing of official ballots, the postponement of the election, the transfer of polling place or the general conduct of the election.⁶³⁵

Hence, future legislation may synthesize both provisions to establish a more comprehensive catalog of nefarious purposes in relation to the creation and distribution of political deepfakes. In any event, both provisions and constitutive purposes are compatible, as both aim towards upholding the integrity of the electoral process.

This ability to sow confusion and division among the electorate about the elections is far from mere conjecture as COMELEC segregates the different kinds of manipulated media that vary in convincingness, but all are geared towards some form of perceptual manipulation. Hence, it discretely distinguishes between deepfakes,⁶³⁶ softfakes,⁶³⁷ and cheapfakes.⁶³⁸ Furthermore, the damage caused by political deepfakes as means of propagating disinformation finds concrete examples in recent elections conducted around the world.⁶³⁹ Consequently, there exists great state interest in protecting the integrity of the electoral process and safeguarding it against emergent threats like political deepfakes. Future legislation should thus echo the specificity and degree of nuance in the classification of deepfakes as exhibited in COMELEC Resolution 11064.⁶⁴⁰

⁶³⁴ COMELEC Res. No. 11064, pmb. ¶ 10; ELECT. CODE, § 261(z)(11).

⁶³⁵ ELECT. CODE, § 261(z)(11).

⁶³⁶ COMELEC Res. No. 11064, art. II(13).

⁶³⁷ Art. II(26).

⁶³⁸ Art. II(7).

⁶³⁹ See *supra* Part II(D).

⁶⁴⁰ COMELEC Res. No. 11064, pmb. ¶ 10, art. II (7), (13), (26).

The Revilla Bill has identified another act which should be incorporated as an explicit election offense based on Section 261(e) of the OEC, namely:

Any person who, directly or indirectly, threatens, intimidates or actually causes, inflicts or produces any violence, injury, punishment, damage, loss or disadvantage upon any person or persons or that of the immediate members of his family, his honor or property, or uses any *fraudulent device or scheme* to compel or induce the registration or refraining from registration of any voter, or the participation in a campaign or refraining or desistance from any campaign, or the casting of any vote or omission to vote, or any promise of such registration, campaign, vote, or omission therefrom.⁶⁴¹

Understanding the accessibility and technical function of the technology,⁶⁴² it is conceivable that deepfake technology is exploited in order to essentially coerce another person into committing or refraining from committing any of the acts as enumerated in the Revilla Bill in relation to the OEC.⁶⁴³ After all, impersonation, “opinion manipulation” for the purpose of disinformation, and even “political disruption” have been touted by researchers as some of the more prominent use cases of deepfakes.⁶⁴⁴ This possibility of coercion in relation to the elections is also feasible considering that existing criminal law already penalizes grave coercion.⁶⁴⁵ If legislators have deemed the crime of grave coercion as a palpable threat to civil society, so too should they consider not only the possibility but also the gravity of the crime when conducted in relation to elections. In this sense, the necessity of punishment becomes paramount as the electoral process lies at the heart of the deliberative democratic system.

More importantly, legislators may validly punish the creation and circulation of deepfakes (and even other manipulated media) during the campaign period for the purpose of *disinformation* and not mere misinformation. This approach would avoid the unconstitutional pitfalls of COMELEC Resolution No. 11064 which incorporates the purpose of misinformation as the *mens rea* in the list of prohibited acts.⁶⁴⁶

⁶⁴¹ ELECT. CODE, § 261(e). (Emphasis supplied.)

⁶⁴² See *supra* Part II(D).

⁶⁴³ H. No. 9425, 19th Cong., 2nd Sess. (2023), § 10; ELECT. CODE, § 261(e).

⁶⁴⁴ Marchal et al., *supra* note 28, at 12–13.

⁶⁴⁵ REV. PEN. CODE, art. 286.

⁶⁴⁶ COMELEC Res. No. 11064, art. V, § 1.

The complication that arises in the inclusion of misinformation along with disinformation is evident when the definition of both terms is scrutinized. COMELEC Resolution No. 11064 defines “misinformation” as “false or inaccurate information that is spread in a partisan political activity or digital election campaign, *regardless of whether there is intent to mislead.*”⁶⁴⁷ This definition implies that the term contemplates mere false or wrong information. On the other hand, the Resolution defines “disinformation” as “the intent to mislead, harm, or manipulate public opinion or voter behavior.”⁶⁴⁸ The latter definition includes the existence of actual knowledge in its standard of punishment.

Punishing the creation and proliferation of political deepfakes is an absurdity then when it is considered that even slander and libel in the RPC both require the existence of malice for an individual to be held liable.⁶⁴⁹ This requirement exists even for the commission of cyberlibel.⁶⁵⁰ Hence, holding an individual liable for creating and publishing a political deepfake online even without the knowledge of its falsity would run contrary to basic criminal law precepts.

Further, assuming that the election offense in future legislation will likewise be *mala in se* in nature like that in COMELEC Resolution No. 11064, there must exist the concurrence of the *actus reus* and criminal intent.⁶⁵¹ The former pertains to “the external or overt acts or omissions included in a crime's definition.”⁶⁵² The latter refers to “the accused's guilty state of mind or criminal intent accompanying the *actus reus.*”⁶⁵³

Prospective legislation may also consider a safe harbor provision akin to that contained in the Biazon Bill to enumerate limited scenarios wherein the creation, disclosure, or handling of unlawful political deepfakes, among others, will not result in any liability.⁶⁵⁴ Lastly, it may also clarify that the provisions contained in the law do not limit other private causes of action that may arise as a result of the misuse of political deepfakes. After all, other

⁶⁴⁷ Art. II(23). (Emphasis supplied.)

⁶⁴⁸ Art. II(15).

⁶⁴⁹ REV. PEN. CODE, art. 353, 358; REYES, *supra* note 367, at 1239; *See supra* Part III(C).

⁶⁵⁰ *Causing*, G.R. No. 258524, slip op. at 14–15.

⁶⁵¹ XXX, G.R. No. 255877, slip op. at 8, Mar. 29, 2023; COMELEC Res. No. 11064, art. V, § 1.

⁶⁵² XXX, G.R. No. 255877, slip op. at 8.

⁶⁵³ *Id.*

⁶⁵⁴ H. No. 5406, 18th Cong., 1st Sess. (2019), § 4.

causes of action may exist due to violations of privacy,⁶⁵⁵ cybercrime,⁶⁵⁶ the Safe Spaces Act,⁶⁵⁷ and even other laws.

V. CONCLUSION

At this critical juncture, the Philippine government possesses the unique opportunity to set the regulatory tone on the use of AI. Regardless of government action, it cannot be denied that the technology will continue to flourish and find its way into nearly all facets of life. And if left unregulated, citizens will increasingly find themselves experiencing the double-edged nature of technology— not only its promises of disruption and innovation, but also its consequential perils.

Therefore, the urgency of nuanced legislation that manages the delicate balance between private interests of innovation and profitability and public interests of safety and the rule of law cannot be understated. To clarify, this Paper aims not to limit the kinds of regulation with respect to the technology but instead endeavors to advocate for the pressing need for targeted regulation for the employment of AI technology in the electoral context. The critical nature of this kind of law is set amidst the backdrop of upcoming elections and the ruinous ramifications of political deepfakes as amplifiers of disinformation.

The constitutional mandate to uphold the true will of the electorate may find itself demoted to a mere governmental aspiration unless tangible measures are implemented to vigilantly safeguard the electoral process against technological threats and corresponding nefarious actors. On the flip side, any law churned out by Congress must not only seek to uphold these ideals but must also protect the fundamental rights of free speech and expression. After all, the protection of one right does not imply the degradation or sacrifice of the other. Both remain enshrined in the Constitution and are regarded as imperatives.

More importantly, the right of free speech and expression are complementary and indispensable to ensuring a meaningful exercise of the right to vote. The choice of prospective leaders and ultimate governmental direction is a painstaking process of deliberation and exchange. These

⁶⁵⁵ Rep. Act No. 10173 (2012).

⁶⁵⁶ Rep. Act No. 10175 (2012).

⁶⁵⁷ Rep. Act. No. 11313 (2019).

collective social conclusions do not occur in a vacuum and require the vitality of spirited discourse from the citizenry in the theorized marketplace of ideas. The law merely needs to contend with the shift of this medium online and the complications brought about by social media and AI. In other words, the proverbial political battleground has now become digital, and propaganda generated or manipulated by AI have become a mainstay in the campaign arsenal. Prospective legislation must thus be crafted with this underlying consciousness of the status quo and must nonetheless undertake to combat the ills of disinformation and deception as exacerbated by deepfakes from a practical yet even-keeled standpoint.

- o0o -