NORTH CAROLINA JOURNAL of
LAW & TECHNOLOGY
Volume 27, Issue 2 2025

NOTE

TAKE IT DOWN OR TAKE IT TOO FAR? THE LEGAL FALLOUT OF NEW ONLINE TAKEDOWN POWERS

Taylor Greeno†

Bad faith actors have long used intimate imagery to inflict pain and smear the reputations of the vulnerable and famous alike. In recent years, new technologies have developed that enable the creation of more of this content, often requiring only a simple image of the victim's face.

The TAKE IT DOWN Act represents a monumental step in the regulation of artificial intelligence and deepfake non-consensual intimate imagery by imposing criminal liability for individual creators while establishing rules for platforms that host this content. However, the Act contains ambiguous language, and its few exceptions do not extend far enough. As a result, the Act conflicts with policy purposes of Section 230, runs the risk of chilling speech, and could hinder the development of internet platforms. This Note proposes a few simple changes aimed at addressing free speech critiques while solidifying the Act's ability to withstand a constitutional challenge.

TABLE OF CONTENTS

† J.D. Candidate, University of North Carolina School of Law, 2027.

I. INTRODUCTION

A fourteen-year-old girl wakes up to her phone pinging with a flurry of text messages from her friends.[1] What could be an exciting substitute for a morning alarm quickly turns into a waking nightmare, sparking national outrage and bipartisan legislative action.[2] This is the story of Elliston Berry, a freshman at Aledo High School ("Aledo") in North Texas.[3]

Photos of her had been edited to make her appear nude.[4] Berry was not alone.[5] In total, nine girls at Aledo were victims of non-consensual intimate imagery ("NCII").[6] The culprit, another student at Aledo,

1. Alex Boyer, *North Texas Mom Shares 'Deepfake' Horror Story as Lawmakers Look to Close Loophole*, FOX 4 (June 6, 2024, at 17:26 CT), https://www.fox4news.com/news/deepfakes-texas-artificial-intelligence [https://perma.cc/W42K-5MRL].
2. Katharine Wilson, *A Texas Teen Was the Victim of Fake AI Nudes. Now a New Law Requires Platforms to Remove Such Content.*, TEX. TRIB. (May 19, 2025, at 15:52 CT), https://www.texastribune.org/2025/05/19/take-it-down-act-deepfakes-digital-nudes-texas-student [https://perma.cc/B3HL-MH4Z].
3. Boyer, *supra* note 1.
4. Wilson, *supra* note 2.
5. *Id.*
6. *Id.*

used artificial intelligence ("AI") to photoshop innocent social media pictures and create pornographic depictions of young girls through a process colloquially known as "deepfake" technology.[7] Through Snapchat, the cyberbully spread the deepfake NCII to classmates, causing the victims to "live[] in fear" and the victims' parents to describe feeling a sense of "helpless[ness]."[8]

Since the advent of deepfake technology, its use is not unlike other "early uses of digital technologies, [where] women are the canaries in the coal mine."[9] In a study conducted by Deeptrace Labs, an Amsterdam-based cybersecurity company, researchers found that there were roughly 15,000 deepfake videos online in 2019, with pornography accounting for ninety-six percent of such videos and ninety-nine percent of those involving "women's faces being inserted

7. *See* Rachel Hale, *Her Classmate Used AI to Make Deepfake Nude Images of Her. Experts Say It's not Uncommon.*, USA TODAY (Apr. 22, 2024, at 16:52 ET), https://www.usatoday.com/story/life/health-wellness/2025/03/25/deepfake-ai-nude-teenagers-mental-health-bullying/81987432007/ [https://perma.cc/45JS-D28R]. For a definition and description of deepfakes, see Phil Swatton & Margaux Leblanc, *What Are Deepfakes and How Can We Detect Them?*, ALAN TURING INST. (June 7, 2024), https://www.turing.ac.uk/blog/what-are-deepfakes-and-how-can-we-detect-them [https://perma.cc/QP9K-7YQK (staff-uploaded)] ("The term 'deepfake' usually refers to an AI-generated video, image or piece of audio content that is designed to mimic a real-life person or scene.").

8. Tiffany Liou, *'I Don't Want to Live in Fear Anymore': North Texas Girl Victimized with Deepfake Nudes Pushes for Federal Law*, WFAA 8 ABC https://www.wfaa.com/article/news/local/north-texas-girl-victimized-with-deepfake-nudes-pushes-for-federal-law/ [https://perma.cc/FRN2-SKWC] (last updated (June 21, 2024, at 19:43 CT).

9. Hany Farid, Robert Chesney & Danielle Citron, *All's Clear for Deepfakes? Think Again.*, U.C. BERKELEY SCH. INFO. (May 11, 2020), https://www.ischool.berkeley.edu/news/2020/alls-clear-deepfakes-think-again [https://perma.cc/Q6PW-46J4] (pointing to the use of cheapfakes, a crude predecessor of deepfakes); *see also* Tina Tallon, *A Century of "Shrill": How Bias in Technology has Hurt Women's Voices*, NEW YORKER (Sep. 3, 2019), https://www.newyorker.com/culture/cultural-comment/a-century-of-shrill-how-bias-in-technology-has-hurt-womens-voices [https://perma.cc/YQQ4-4ETW] (describing how broadcast, voice technologies, data-compression algorithms, and Bluetooth speakers disproportionately affect the voices of women, causing them to sound "thin and tinny").

into porn without consent."[10] Similarly, in 2023, U.S. cybersecurity firm Home Security Heroes conducted a study finding 95,820 deepfake videos online, reporting that deepfake pornography comprised ninety-eight percent of those videos, ninety-nine percent of which targeted women.[11]

The tragic story of Rana Ayyub, a prominent Indian investigative journalist, is another prescient example of the suffering deepfake technology can cause women.[12] In April 2018, Ayyub became the target of a harrowing cyberattack that weaponized her identity and visibility.[13] As a Muslim journalist known for her anti-establishment reputation, she was a frequent target of misogyny and abuse—once describing herself as "the most abused woman in India."[14] After she appeared on BBC and Al Jazeera condemning India's protection of child sex abusers following the rape of an eight-year-old girl in India, online abuse against her dramatically increased.[15]

Though the initial attacks were common cyber-harassment misinformation tactics like the dissemination of fake tweets, aimed at tarnishing Ayyub's reputation, the harassment escalated significantly.[16] Ayyub was informed by a source from the Bharatiya Janata Party ("BJP"), the nationalist ruling political party in India, that a video of her was circulating on WhatsApp.[17] That video was a pornographic deepfake, which used AI to digitally impose Ayyub's face onto the

---

10. Farid, Chesney & Citron, *supra* note 9; HENRY AJDER ET AL., THE STATE OF DEEPFAKES: LANDSCAPE, THREATS, AND IMPACT 1 (Sep. 2019), https://regmedia.co.uk/2019/10/08/deepfake_report.pdf [https://perma.cc/PN6Z-9ZHW].

11. *2023 State of Deepfakes*, HOME SEC. HEROES (2023), https://www.securityhero.io/state-of-deepfakes/ [https://perma.cc/S2WU-CURG].

12. *Rana Ayyub, the Face of India's Women Journalists Plagued by Cyber-Harassment*, REPS. WITHOUT BORDERS (Nov. 27, 2024), https://rsf.org/en/rana-ayyub-face-india-s-women-journalists-plagued-cyber-harassment [https://perma.cc/AN4A-BUWD].

13. *Id.*

14. Rana Ayyub, *I was the Victim of a Deepfake Porn Plot Intended to Silence Me*, HUFFINGTON POST (Nov. 21, 2018, at 08:11 GMT), https://www.huffingtonpost.co.uk/entry/deepfake-porn_uk_5bf2c126e4b0f32bd58ba316 [https://perma.cc/ZF84-PSES].

15. *Id.*

16. *Id.*

17. *Id.*

body of a young, naked woman.[18] After seeing the video and online reaction to it, Ayyub "started throwing up" and crying.[19] The video of Ayyub was shared over 40,000 times, leading to more abuse and harassment.[20] Receiving no help from the local or national government in India, Ayyub eventually found support from the United Nations, but the emotional and reputational damage was already done.[21]

While cyberbullying and online harassment have existed long before the rise of deepfake technology, deepfakes pose a unique threat.[22] Audio and visual evidence is especially persuasive to humans, even more so when it "is of such quality that our eyes and ears cannot readily detect that something artificial is at work."[23] Further, "[t]he more salacious and negative the deepfake . . . the more inclined we are to pass them on."[24] Moreover, "[r]esearchers have found that online hoaxes spread 10 times faster than accurate stories."[25]

Deepfakes have also caused significant harm to children.[26] In a recent study, researchers found that one in eight teens aged thirteen to seventeen "personally know someone who has been victimized by deepfake nudes."[27] Roughly the same percentage of those teens knew someone who had used deepfake technology to create or distribute nude content.[28] And though the use of deepfake technology by teens is

---

18. *Id.*

19. *Id.*

20. *Id.*

21. *Id.*

22. Farid, Chesney & Citron, *supra* note 9.

23. *Id.*

24. *Id.*

25. *Id.*

26. *See* Dana Nickel, *AI Is Shockingly Good at Making Fake Nudes – and Causing Havoc in Schools*, POLITICO (May 29, 2024, at 05:00 ET), https://www.politico.com/news/2024/05/28/ai-deepfake-nudes-schools-states-00160183 [https://perma.cc/MV7D-6KM3].

27. AMANDA GOHARIAN, MELISSA STROEBEL, SAM FITZ, SARAH GUDGER, ARIELLE JEAN-BAPTISTE & PATRICK TOOMEY, *Deepfake Nudes & Young People*, THORN 14 (2025), https://info.thorn.org/hubfs/Research/Thorn_DeepfakeNudes&YoungPeople_Mar2025.pdf [https://perma.cc/98ZE-XF9B].

28. *Id.*

becoming increasingly more common,[29] it is also being used more and more by adults to create child pornography.[30]

Consider the story of Olivia, a little girl who was abused from the ages of three to eight.[31] Olivia's abuser photographed her sexual torture, circulating and sharing the images of her misery to other sex offenders.[32] Even after her abuser was apprehended and the abuse was put to an end, other individuals used AI to create more pornographic images of Olivia, in new, abusive situations.[33]

Aiming to address the growing danger of deepfake technology towards vulnerable populations, lawmakers joined forces with the victims in Aledo to work towards a solution.[34] In April 2024, Texas Senator Ted Cruz flew Berry, one of the Aledo victims, and her mother to Washington to discuss new legislation to regulate deepfakes.[35] Berry spoke at news conferences and appeared on major television networks with First Lady Melania Trump to advocate for a new bill called the TAKE IT DOWN Act (the "Act").[36] On May 19, 2025, Congress passed the Act, representing the first federal legislation regulating the harmful use of AI.[37]

---

29. *Id.*

30. *See What Has Changed in the AI CSAM Landscape?*, INTERNET WATCH FOUND. 3 (July 2024), https://www.iwf.org.uk/media/opkpmx5q/iwf-ai-csam-report_update-public-jul24v11.pdf [https://perma.cc/BK2H-8JJL].

31. *Once Upon a Year*, INTERNET WATCH FOUND. 11 (2018), https://www.iwf.org.uk/media/tthh3woi/once-upon-a-year-iwf-annual-report-2018.pdf [https://perma.cc/T3A9-TPUT].

32. *Id.*

33. *Id.*

34. Wilson, *supra* note 2.

35. *Id.*

36. *Id.*; TAKE IT DOWN Act, Pub. L. No. 119-12, § 146, 139 Stat. 55 (2025).

37. *See* Stuart D. Levi & Mana Ghaemmaghami, *'Take It Down Act' Requires Online Platforms to Remove Unauthorized Intimate Images and Deepfakes When Notified*, SKADDEN (June 10, 2025), https://www.skadden.com/insights/publications/2025/06/take-it-down-act [https://perma.cc/9DEM-BW49].

The Act's passage received bipartisan celebration,[38] but was not without critics.[39] Free-speech advocates and digital-rights groups alike worry that the Act will chill speech, including legal pornography, LGBTQ+ content, and government criticism.[40] In fact, President Trump, in a joint session of Congress, proclaimed: "I look forward to signing that bill into law. And I'm going to use that bill for myself too if you don't mind, because nobody gets treated worse than I do online, nobody."[41] Further concerns relate to the burdens placed on platforms, as the Act "lacks critical safeguards against frivolous or bad-faith takedown requests," forcing platforms to choose between spending valuable resources to vet requests and risking legal liability.[42] Most likely, smaller platforms will "choose to avoid the onerous legal risk by simply depublishing the speech rather than even attempting to verify it."[43]

Though the Act has yet to be applied or enforced, a First Amendment challenge could loom, inviting the question of whether the Act will survive judicial review. Relatedly, how might § 230 of the Digital Communications Act,[44] which protects platforms from

---

38. Savannah Kuchar, *With Rare Bipartisan Support, Congress Passes Bill to Outlaw Deepfake Pornography*, USA TODAY (Apr. 29, 2025), https://www.aol.com/rare-bipartisan-support-congress-passes-000220713.html [https://perma.cc/S63G-TUFQ]; Press Release, U.S. Senate Committee on Commerce, Science, & Transportation, Sen. Cruz Applauds Presidential Signing of the TAKE IT DOWN Act into Law (May 19, 2025), https://www.commerce.senate.gov/2025/5/sen-cruz-applauds-presidential-signing-of-the-take-it-down-act-into-law [https://perma.cc/JK77-VY5Q].

39. Letter from Center for Democracy & Technology, et al., to Senate (Feb. 12, 2025), https://cdt.org/wp-content/uploads/2025/02/TAKE-IT-DOWN-Sign-On-Letter_21225.pdf [https://perma.cc/AG5R-ZWC7]; Jason Kelley, *Congress Passes TAKE IT DOWN Act Despite Major Flaws*, ELEC. FRONTIER FOUND. (Apr. 28, 2025), https://www.eff.org/deeplinks/2025/04/congress-passes-take-it-down-act-despite-major-flaws [https://perma.cc/EV7G-4RZF].

40. Barbara Ortutay, *Take It Down Act, Addressing Nonconsensual Deepfakes and 'Revenge Porn,' Passes. What Is It?*, FREE SPEECH CTR. MIDDLE TENN. STATE UNIV. (Apr. 30, 2025), https://firstamendment.mtsu.edu/post/take-it-down-act-addressing-nonconsensual-deepfakes-and-revenge-porn-passes-what-is-it [https://perma.cc/N4TJ-NC2Y].

41. Kelley, *supra* note 39.

42. *Id.*

43. *Id.*

44. 47 U.S.C. § 230 (2025).

liability for the content they host in certain circumstances, affect enforcement of the Act? On first impression, the TAKE IT DOWN Act is an important step in regulating deepfake NCII.[45] The Act imposes criminal punishment on creators and publishers of deepfake NCII, while excluding categories of speech subject to traditional First Amendment protection. And despite contravening certain policies of § 230, the Act is explicitly excepted from § 230's expansive reach and thus remains enforceable.[46] However, the Act could be found unconstitutionally overbroad because its requirements for covered platforms risk chilling a substantial amount of First Amendment protected speech.

This Note proceeds in six Parts. Part II overviews how deepfake technology works and what differentiates it from similar past technologies. Part III provides an analysis of the Act, explaining the regulated conduct, enforcement mechanisms, and implications for individuals and platforms. Part IV explains that while the Act conflicts with policy purposes of § 230, the Act is not preempted and remains enforceable. Part V highlights the First Amendment free speech concerns raised by the Act. Finally, Part VI discusses further proposals to regulate deepfakes and advances a set of policy solutions to clarify the Act, ameliorating the First Amendment speech chilling arguments raised in Part IV.

II. The History of NCII and Deepfake Technology

While AI and deepfake technology have transformed the process through which NCII is produced and exacerbated online child sexual abuse,[47] NCII is not new.[48] In 1888, Le Grange Brown, a New York photographer, was accused of selling photographs of nude women after

45. *See infra* Part III.

46. 47 U.S.C. § 230(e)(1) (2025).

47. *See What Has Changed in the AI CSAM Landscape?*, Internet Watch Found. 3 (July 2024), https://www.iwf.org.uk/media/opkpmx5q/iwf-ai-csam-report_update-public-jul24v11.pdf [https://perma.cc/BK2H-8JJL].

48. Jessica Lake, *In the 19th Century, a Man Was Busted for Pasting Photos of Women's Heads on Naked Bodies*, Conversation (Sep. 22, 2021, at 22:15 ET), https://theconversation.com/in-the-19th-century-a-man-was-busted-for-pasting-photos-of-womens-heads-on-naked-bodies-sound-familiar-168081 [https://perma.cc/QR8W-67C6].

he physically cut and pasted their heads onto images of naked women.[49] In 1903, opera star Marion Manola's photograph was taken without her consent and turned into an erotic postcard, leading to the New York State Legislature recognizing "the first right to privacy in the U.S. and across the common law world . . . ."[50] Due to the emergence of new technologies in the twentieth and twenty-first centuries, such as home video, the internet, photoshop,[51] and AI,[52] the distribution of NCII proliferated.[53]

And though bad actors have long manipulated images to create NCII, "earlier methods were typically crude—requiring significant time, skill, and technical expertise to produce photorealistic outcomes."[54] In recent years, technological advancements have significantly increased ease and accessibility.[55] Specifically, AI has enabled the creation of "deepfakes," a kind of hyper-realistic "synthetic media where a person in an image or video is swapped with another person's likeness."[56] Now, "almost anyone with a computer can

---

49. *Id.*

50. *Id.*; Robert C. Cumbow, *New York Takes the Stage with New Publicity Right Law*, MILLER NASH LLP (Dec. 28, 2020), https://www.millernash.com/industry-news/new-york-takes-the-stage-with-new-publicity-right-law [https://perma.cc/T2BA-QPCL].

51. Mason Lindblad, *The History of Photoshop – Photoshop Through the Years*, FILTERGRADE (Sep. 22, 2020), https://filtergrade.com/history-of-photoshop-through-the-years/ [https://perma.cc/W8Y6-B4L9].

52. B.J. Copeland, *History of Artificial Intelligence*, BRITANNICA (July 30, 2025), https://www.britannica.com/science/history-of-artificial-intelligence [https://perma.cc/QD2J-39V7].

53. Sophie Maddocks, *Image-Based Abuse: A Threat to Privacy, Safety, and Speech*, MEDIAWELL (Mar. 15, 2023), https://mediawell.ssrc.org/research-reviews/image-based-abuse-a-threat-to-privacy-safety-and-speech/ [https://perma.cc/G5RL-JDLU].

54. AMANDA GOHARIAN ET AL., DEEPFAKE NUDES & YOUNG PEOPLE, THORN 8 (2025), https://info.thorn.org/hubfs/Research/Thorn_DeepfakeNudes&YoungPeople_Mar2025.pdf [https://perma.cc/3E5X-XJXD (staff-uploaded)].

55. Mika Westerlund, *The Emergence of Deepfake Technology: A Review* 40–41, TECH. INNOVATION MGMT. REV. (Nov. 2019), https://timreview.ca/article/1282 [https://perma.cc/Q2EC-SV63].

56. Meredith Somers, *Deepfakes, Explained*, MIT SLOAN SCH. MGMT. (July 21, 2020), https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained [https://perma.cc/4NWU-L2CW].

fabricate videos that are practically indistinguishable from authentic media."[57]

The term "deepfake" originated on Reddit in November 2017, when a user created a forum titled r/Deepfakes, a combination of the words "deep learning" and "fake."[58] That user dedicated the forum to "the creation and use of deep learning software for synthetically face swapping female celebrities into pornographic videos."[59] Since then, the term's meaning has been expanded to include "synthetic media applications that existed before the Reddit page" and new technologies used for pornographic and non-pornographic purposes.[60]

Although deepfakes can be produced through a variety of methods, deepfake technology generally uses Generative Adversarial Networks ("GANs"), employing two artificial neural networks to create media that appears real through a face swapping algorithm.[61] The two neural networks "work in opposition – one generates data, while the other evaluates whether the data is real or generated."[62]

A face swapping algorithm uses three main steps.[63] First, face detection involves training an algorithm on large data sets of human faces to detect facial features and distinguish faces from other objects

---

57. Westerlund, *supra* note 55 at 39. For an example of the recent advancements in AI video creation software that the law will need to catch up to, see generally, Brian X. Chen, *A.I. Video Generators are now so Good You can No Longer Trust Your Eyes*, N.Y. TIMES, https://www.nytimes.com/2025/10/09/technology/personaltech/sora-ai-video-impact.html [https://perma.cc/2UU9-73AQ] (last updated Oct. 14, 2025) (warning that the arrival of Sora, a new AI-generated-video program which is capable of creating hyper-realistic depictions of any public figure, will reduce or potentially eliminate our ability to trust the authenticity of videos and images).

58. Westerlund, *supra* note 55, at 40; AJDER ET AL., *supra* note 10, at 3.

59. AJDER ET AL., *supra* note 10, at 3.

60. Somers, *supra* note 56 (describing StyleGAN, a new application which creates "realistic-looking" still images of people that don't exist).

61. Westerlund, *supra* note 55, at 40–41.

62. Jobit Varughese, *What Are Generative Adversarial Networks (GANs)?*, IBM, https://www.ibm.com/think/topics/generative-adversarial-networks [https://perma.cc/EA3T-BP2S] (last visited Oct. 26, 2025).

63. Tharindu Fernando et al., *Face Deepfakes – A Comprehensive Review*, MICH. STATE UNIV. DEP'T COMPUT. SCI. & ENG'G (Feb. 13, 2025), https://arxiv.org/html/2502.09812v1 [https://perma.cc/6A5U-DXXJ].

in an image.[64] Next, the program swaps the main features, such as the nose, mouth, and eyes, onto the target image, replacing the corresponding features.[65] Finally, an autoencoder[66] uses the input data from the source and target images to blend the manipulated attributes, matching the color and lighting with minimal information loss.[67] Deepfake technology now requires the input data from only a single image.[68] The result of the process is an image or video that appears genuine, with little evidence of manipulation.[69]

III. THE TAKE IT DOWN ACT

The weaponization of deepfake technology in the context of NCII ultimately led to the proposal and passage of the TAKE IT DOWN Act,[70] which criminalizes the distribution of NCII and sets rules for platforms whose users may post such content.[71] Although the Act sufficiently combats NCII, it also conflicts with policy purposes of § 230 and risks chilling speech, raising a question about its efficacy because it may not survive a First Amendment challenge.

---

64. *Understanding Facial Recognition Algorithms*, RECFACES, https://recfaces.com/articles/facial-recognition-algorithms [https://perma.cc/Q792-QS6V] (last visited Sep. 13, 2025).

65. Fernando et al., *supra* note 63.

66. *See* Dave Bergmann & Cole Stryker, *What is an Autoencoder?*, IBM, https://www.ibm.com/think/topics/autoencoder [https://perma.cc/7MCM-7ELA] (last visited Oct. 26, 2025) ("An autoencoder is a type of neural network architecture designed to efficiently compress (encode) input data down to its essential features, then reconstruct (decode) the original input from this compressed representation.").

67. Fernando et al., *supra* note 63.

68. Mika Westerlund, *The Emergence of Deepfake Technology: A Review*, 9 TECH. INNOVATION MGMT. REV. 40, 45 (2019).

69. *Id.* at 40.

70. TAKE IT DOWN Act, Pub. L. No. 119-12, § 146, 139 Stat. 55 (2025) ("Tools to Address Known Exploitation by Immobilizing Technological Deepfakes on Websites and Networks Act.").

71. *ICYMI: President Trump Signs TAKE IT DOWN Act into Law*, WHITE HOUSE (May 19, 2025), https://www.whitehouse.gov/articles/2025/05/icymi-president-trump-signs-take-it-down-act-into-law/ [https://perma.cc/V8XL-7YUL].

A. Criminalization of an Individual's Publication of Deepfake NCII

Subsection (h) of 47 U.S.C. § 223 applies to an individual publisher's conduct and addresses the publication of both authentic and inauthentic intimate imagery.[72] Because this Note concerns deepfakes and the unique issues associated with these technologies, analysis of the Act focuses exclusively on the section concerned with inauthentic intimate imagery.

The Act defines "digital forgery" broadly. In the Act, digital forgery encompasses a variety of digital methods that could be used to create intimate imagery, including deepfake technology.[73] Under the Act, it is unlawful to use an internet service to knowingly publish a digital forgery of an *adult* that is intended to cause harm or actually causes harm, whether it be "psychological, financial, or reputational harm . . . ."[74] Under the Act, however, publishing a digital forgery of an adult is not illegal when it is done with consent, the adult publicly and voluntarily exposed what is depicted, or the depiction is a matter of public concern.[75] By protecting publishers of deepfake consensual intimate imagery, the Act aligns with the First Amendment's protection of most consensual pornography.[76]

While the Act does not further define what constitutes a matter of public concern, the Supreme Court has previously held in a First Amendment context that the phrase means "any matter of political, social, or other concern to the community"[77] where "free and open

---

72. 47 U.S.C. § 223(h)(2)–(3) (2025).

73. *Id.* § 223(h)(1)(B)–(E).

74. *See id.* § 223(h)(1)(A)–(E), 223(h)(3)(A).

75. *Id.* § 223(h)(3)(A)(i)–(iii).

76. *See* David L. Hudson Jr., *Obscenity and Pornography*, FREE SPEECH CTR. MIDDLE TENN. STATE UNIV., https://firstamendment.mtsu.edu/article/obscenity-and-pornography [https://perma.cc/ECN6-EEVZ] (last updated Nov. 6, 2025) (noting that while the Supreme Court has often struggled to delineate between protected pornographic speech and obscenity, the First Amendment nevertheless protects most pornography).

77. Connick v. Myers, 461 U.S. 138, 146 (1983); *see, e.g.*, Roth v. United States, 354 U.S. 476, 485 (1957) (holding that obscenity is not constitutionally protected speech); New York v. Ferber, 458 U.S. 747, 764 (1982) ("[T]he category of child pornography . . . like obscenity, is unprotected by the First Amendment.").

debate is vital to informed decision-making by the electorate."[78] One possible application of the public concern exception to deepfake NCII would be an individual publishing commentary with clips from a video of deepfake NCII created by a politician, of their rival. Thus, the Act's allowance of deepfake NCII involving matters of public concern conforms to traditional First Amendment protection of such matters.[79] Nevertheless, an individual can be held criminally liable under the Act for publishing deepfake NCII of an adult, barring an enumerated exception.[80]

Alternatively, the Act criminalizes knowingly publishing a digital forgery of a *minor* that is intended to "abuse, humiliate, harass, or degrade the minor," or "arouse or gratify the sexual desire of any person."[81] The exceptions for consent, voluntary public exposure, and public concern do not apply to a creator of a digital forgery when a minor is the subject of the forgery.[82] While the term digital forgery includes deepfake NCII of minors, that content falls within the definition of child pornography so criminal liability for its publication is thus handled through existing child pornography laws.[83]

Regardless of whether an adult or minor is the subject of deepfake NCII, the Act carves out exceptions to ensure that individuals acting in good faith are not thrown in jail.[84] These exceptions cover law enforcement investigations, disclosures made in the course of legal proceedings, disclosures for the purposes of medicine, science, and education, as well as any disclosure "reasonably intended to assist the

78. Pickering v. Bd. of Educ., 391 U.S. 563, 571–72 (1968); *see also* Synder v. Phelps, 562 U.S. 443, 452 (2011) (holding that protestors with signs such as "God Hates the USA" and "Thank God for Dead Soldiers" discussed matters of public concern and therefore protected by the First Amendment).

79. *See Myers*, 461 U.S. at 146; *see also* Hustler Mag., Inc. v. Falwell, 485 U.S. 46, 46–47 (1988) (holding that a magazine's publication of an offensive and intentionally injurious parody of a public figure is protected from a lawsuit for intentional infliction of emotional distress by the First Amendment).

80. *See* 47 U.S.C. § 223(h)(3)(A) (2025).

81. *Id.* § 223(h)(3)(B)(i).

82. *See id.* § 223(h)(3)(B).

83. *See id.* § 223(h)(1)(B), 223(h)(3)(C)(v); 18 U.S.C. § 2256(8) (2025).

84. *See* 47 U.S.C. § 223(h)(3)(C) (2025).

[victim] . . . ."[85] The Act further protects possessors or publishers of deepfake intimate imagery of oneself.[86]

As a penalty for publication of unlawful digital forgeries, the Act imposes a maximum term of imprisonment of two years for depictions of adults, and three years for depictions of minors.[87] On the other hand, if an individual intentionally threatens to publish a digital forgery "for the purpose of intimidation, coercion, extortion, or to create mental distress," the Act imposes a maximum term of imprisonment of eighteen months for violations involving adults, and thirty months for violations involving minors.[88] Given that the Act disclaims enforcement of content falling under the definition of child pornography,[89] it is not clear when a case of deepfake NCII involving a minor would fall under the Act. After all, the definition of child pornography includes digital visual depictions that are indistinguishable from actual child pornography.[90]

The Act is a monumental step in combating deepfake NCII and should serve to deter its publication by individuals. Further, the breadth of exceptions for individual criminal liability sufficiently protects the free speech of individuals. And though the Act's penalties may seem meager given its potential harm, for the first time, law enforcement agencies have a means of bringing some measure of justice to victims of deepfake NCII.

B. The Act's Implications for Platforms

The portion of the Act applying to the platforms which individuals could use to publish deepfake AI is incorporated into 47 U.S.C. § 223a of the Communications Act of 1934.[91] Under the Act, a "covered platform" includes any website, internet service, or application that serves the public and either provides a forum for user-generated

85. *Id.* § 223(h)(3)(C)(i)–(iii).
86. *Id.* § 223(h)(3)(c)(iv).
87. *Id.* § 223(h)(4).
88. *Id.* § 223(h)(6)(B).
89. *See id.* § 223(h)(1)(B), 223(h)(3)(C)(v); 18 U.S.C. § 2256(8) (2025).
90. 18 U.S.C. § 2256(8)(B)–(C) (2025); *see also*, New York v. Ferber, 458 U.S. 747, 749–52 (1982) (analyzing a New York State law's definition of child pornography).
91. TAKE IT DOWN Act, Pub. L. No. 119-12, § 146, 139 Stat. 55, 59 (2025).

content, or publishes, hosts, or otherwise makes available content of NCII in its regular course of business.[92] For example, Instagram, TikTok, and X would all clearly fall under the Act's provisions, and so too would pornographic websites which allow any user to submit content.[93] Excluded from this definition are internet providers, email services, and any website, internet service, or application that does not consist primarily of user-generated content or for which the interactive functionality depends on content that is not user-generated.[94] Thus, Spectrum would not be covered simply for providing services to covered platforms, while Microsoft Outlook, Gmail, and other email providers would not be covered simply because someone sends NCII in a message.[95]

However, the same exception that applies to email providers does not protect popular messaging platforms like WhatsApp, Signal, Telegram, or Facebook Messenger.[96] Some of these platforms utilize end-to-end encryption, a "secure communication process that encrypts data before" it is transferred to another device.[97] As such, platforms utilizing end-to-end encryption "will have a legal requirement to remove content that they will have no ability to access or even identify, short of breaking the encryption on which their users rely."[98] In other words, end-to-end encryption platforms could have to break their encryptions to remain compliant with the Act, despite encryption being the consumer appeal of their software.[99] Moreover, the broad exceptions for individual criminal liability that protect free speech, such as immunity for consensual adult deepfake NCII and

92. *Id.* § 146, 139 Stat. at 61.
93. *See id.*
94. *Id.* § 146, 139 Stat. at 61–62.
95. *See id.*
96. *See id.*
97. Thomas J. Cunningham & Michael J. McMorrow, *Platforms Face Section 230 Shift from Take It Down Act*, LAW360 (June 9, 2025, at 17:42 ET), https://www.law360.com/articles/2350115/platforms-face-section-230-shift-from-take-it-down-act [https://perma.cc/7X98-Z8FB]; *What Is End-to-End Encryption (E2EE)?*, IBM, https://www.ibm.com/think/topics/end-to-end-encryption [https://perma.cc/PT62-WDE7] (last visited Sep. 18, 2025).
98. *See* Cunningham & McMorrow, *supra* note 97.
99. *See id.*

matters of public concern, do not apply to platforms.[100] Thus, content that is not individually criminalized could still be subject to the takedown requirements under the Act and cause the censorship of free speech.

The Act gives covered platforms until May 19, 2026, to establish a process allowing an "identifiable individual (or an authorized person acting on behalf of such individual)" to notify the platform and request it take down the flagged NCII.[101] That process must be clearly and conspicuously disclosed by the platform, such that an individual has access to the covered platform's responsibilities and the process through which they can submit a notification and removal request.[102] The individual must submit their physical or electronic signature, reasonably identify and provide a location for the NCII, offer a brief statement that allows the platform to determine that the content was published without their consent, and provide contact information.[103]

When a covered platform receives a "valid removal request" using its designated process, the platform must remove the NCII and "make reasonable efforts to identify and remove any known identical copies of such depiction" as soon as possible, but no later than forty-eight hours after receipt.[104] The Act does not further define "valid."[105] Presumably, based on Merriam-Webster's first definition of "valid," meaning conformity to the law,[106] the phrase "valid removal request" would apply to those requests that comply with the platform's designated process and the Act.

Yet, under Merriam-Webster's second definition of the phrase "valid," meaning well-grounded or justifiable,[107] the phrase could only apply to those requests which are accurately requesting the removal of actual NCII. Under the second definition, a covered platform may have to reasonably vet every request to ensure that the supposed

---

100. TAKE IT DOWN Act, Pub. L. No. 119-12, § 146, 139 Stat. 55, 56 (2025).
101. 47 U.S.C. § 223a(a)(1)(A) (2025).
102. *Id.* § 223a(a)(2).
103. *Id.* § 223a(a)(1)(B).
104. *Id.* § 223a(a)(3).
105. TAKE IT DOWN Act, Pub. L. No. 119-12, § 146, 139 Stat. 55 (2025).
106. *Valid*, MERRIAM-WEBSTER, https://www.merriam-webster.com/dictionary/valid [https://perma.cc/C8CL-8QM2] (last visited Sep. 18, 2025).
107. *Id.*

NCII does, in fact, fall within the category of content they are required to remove, a potentially daunting prospect if a platform is flooded with requests. Although the likely meaning of "valid removal request" falls under the first definition, Congress should clarify its intentions for this ambiguous phrase. The definition of "valid removal request" is particularly significant to the First Amendment implications of the Act discussed *infra* Part V.

The Act's definition of "identifiable individual" and allowance of takedown requests for authorized third-parties adds further complexity to covered platforms' analysis of what constitutes a valid removal request. The Act defines "identifiable individual" as an individual appearing in intimate imagery and "whose face, likeness, or other distinguishing characteristic (including a unique birthmark or other recognizable feature) is displayed in connection with such intimate visual depiction."[108] It is unclear "what level of certainty a platform would need to determine" a takedown request is made by an "identifiable individual."[109] Furthermore, the Act does not define "authorized person" or what evidence of authorization, if any, is required.[110]

Unless clarified by Congress or the courts, covered platforms would likely have to make "[a]n individual determination of the identity for each allegedly offending visual depiction" and determine on its own how to weigh proof of authorization.[111] An alternative to this potentially resource-intensive review of every takedown request is to simply accept each request as valid and remove any content flagged through a covered platform's process. Yet again, platforms may be required to choose between devoting time and money to scrutinizing takedown requests or potentially censoring free speech.

As a safeguard for platforms wary of censoring lawful speech, § 223a(a)(4) creates a bar against "any claim based on the covered platform's good faith" removal of content, lawful or not.[112] In other words, "it does not consider the removal of lawful content [a Federal

108. 47 U.S.C. § 223(h)(1)(C) (2025).
109. Cunningham & McMorrow, *supra* note 97.
110. *Id.*
111. *See id.*
112. 47 U.S.C. § 223a(a)(4) (2025).

Trade Commission Act] violation or otherwise create a cause of action for persons who claim that their content was wrongfully removed."[113] Protecting platforms from liability for making good faith efforts to remove deepfake NCII is good for both victims of deepfake NCII and the platforms themselves, and aligns with the policy purposes of § 230 discussed *infra* Part IV.[114] Putting aside ambiguity in how to interpret good faith under the Act, § 223a(a)(4) only protects platforms from liability. Yet it does nothing to address First Amendment concerns, and if anything, § 223a(a)(4) heightens these concerns because it explicitly protects platforms for suppressing lawful speech.

The Federal Trade Commission ("FTC") is responsible for enforcement of the Act's platform requirements.[115] A violation by a covered platform of any of the obligations under the Act, whether it be in establishing a notice and removal process or in failing to take down a "valid removal request," is directed to be "treated as a violation of a rule defining an unfair or a deceptive act or practice under § 18(a)(1)(B) of the Federal Trade Commission Act . . . ."[116] Presumably because the year-long period to establish a takedown process has yet to pass, the FTC has yet to take any enforcement action. Likewise, as of this Note's writing, none of the major social media platforms most likely to be affected by the Act have implemented a notification and removal process as required under the Act.[117]

The extent of enforcement under the Act remains to be seen. Nevertheless, the implications for covered platforms and individuals alike are clear: play a part in spreading deepfake NCII and face criminal liability. The Act's treatment of individuals sufficiently punishes those responsible for the creation and dissemination of

113. *See* LATHAM & WATKINS LLP, *President Trump Signs "Take It Down Act" Into Law* (May 21, 2025), https://www.lw.com/en/insights/president-trump-signs-take-it-down-act-into-law [https://perma.cc/DQ5J-C5U8].

114. *See infra* Part IV.

115. 47 U.S.C. § 223a(b) (2025).

116. *Id.* § 223a(b)(1).

117. *See*, *e.g.*, *Terms of Service*, TIKTOK (Nov. 2023), https://www.tiktok.com/legal/page/us/terms-of-service/en [https://perma.cc/9XD8-WVEX] (TikTok has not released any statement announcing the establishment of a takedown process compliant with the Act, nor have their terms of service been updated since the Act's passage).

deepfake NCII. And though the few exceptions that do exist for platforms are reasonable, they do not extend far enough. Moreover, the ambiguous language permeating the Act could contribute to the suppression of lawful speech by imposing a substantial burden on platforms and, therefore, must be clarified. As a result of the takedown requirements, the Act risks chilling speech and could allow bad faith actors to suppress content they disagree with, while creating a take-it-or-leave-it dilemma for platforms.

IV. SECTION 230—DOES THE TAKE IT DOWN ACT TAKE IT DOWN?

Congress enacted § 230 of the Communications Decency Act[118] in 1996 in response to two New York court cases, which incentivized internet service platforms to take a hands-off approach to content moderation.[119] Congress aimed to reverse this incentive and allow the proliferation of the modern internet free from government regulation.[120] At the time, § 230 received broad bipartisan support, and so successfully achieved Congress' goal that it has since been dubbed "the twenty-six words that made the internet . . . ."[121] Those twenty-six words are: "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider."[122] This Part addresses how

118. 47 U.S.C. § 230 (2025).

119. Jeff Kosseff, *A User's Guide to Section 230, and a Legislator's Guide to Amending it (or Not)*, 37 BERKELEY TECH. L.J. 757, 765–68, 770–72 (2022) (analyzing *Cubby, inc. v. CompuServe*, 76 F. Supp. 135 (S.D.N.Y. 1991), which found no defamation liability where a platform made no effort to regulate content, and Stratton Oakmont, Inc. v. Prodigy Services Co., 1995 WL 323710 (N.Y. Supr. Ct. May 23, 1995), which found defamation liability where they attempted to regulate content but failed to regulate all of it).

120. *Id.* at 770–71; Zeran v. America Online, Inc., 129 F.3d 327, 330 (4th Cir. 1997) ("Section 230 was enacted, in part, to maintain the robust nature of Internet communication and, accordingly, to keep government interference in the medium to a minimum."), *cert. denied* 524 U.S. 937 (1998).

121. Sarah Grevy Gotfredsen, *Section 230 is Under Attack (Again)*, COLUM. JOURNALISM REV. (Mar. 27, 2025), https://www.cjr.org/the_media_today/section_230_bipartisan_bill_repeal.php [https://perma.cc/CT8E-XRJ2] (emphasis omitted).

122. 47 U.S.C. § 230(c)(1) (2025).

§ 230 interacts with the Act, explaining that, despite the Act contravening policies of § 230, it nonetheless remains enforceable.

A. Section 230 Does not Bar Enforcement of the Act

Though the plain language of § 230 may initially appear to conflict with the Act, a more nuanced reading shows that § 230 does not preclude enforcement. The phrase "interactive computer service" is defined broadly to include "any information service, system, or access software that provides or enables computer access by multiple users to a computer server, including specifically a service or system that provides access to the Internet and such systems operated or services offered by libraries or educational institutions."[123] Section 230 applies to platforms such as TikTok,[124] Facebook,[125] and Google.[126] Furthermore, an "information content provider" is defined as "any person or entity that is responsible, in whole or in part, for the creation or development of information provided through the Internet or any other interactive computer service."[127]

Courts interpret § 230(c)(1) to create "a federal immunity to any cause of action that would make service providers liable for information originating with a third-party user of the service."[128] Therefore, when a lawsuit seeks "to hold a service provider liable for its exercise of a publisher's traditional editorial functions—such as deciding whether to publish, withdraw, postpone or alter content—[such claims] are barred."[129] However, because § 230(c)(1) only applies to information that was "provided by another information content provider," platforms "are not immunized if they are sued for their own expressive activity or content (i.e., first-party speech)."[130] A platform's first-party speech includes content "the platform is 'responsible in

123. *Id.* § 230(f)(2).
124. Anderson v. Tiktok, Inc., 116 F.4th 180, 183 (3d Cir. 2024).
125. M.P. v. Meta Platforms, Inc., 127 F.4th 516, 521 (4th Cir. 2025), *cert. denied* 2025 WL 2824590 (2025).
126. *See, e.g.*, Bennett v. Google, LLC, 882 F.3d 1163, 1164 (D.C. Cir. 2018).
127. 47 U.S.C. § 230(f)(3) (2025).
128. Zeran v. America Online, Inc., 129 F.3d 327, 330 (4th Cir. 1997); *see Anderson*, 116 F.4th at 183.
129. *Zeran*, 129 F.3d at 330.
130. *Anderson*, 116 F.4th at 183.

whole or in part' for creating or developing . . . ."[131] As such, a platform that provided an application for creating deepfake NCII would not be protected by § 230 and could be held liable under the Act.

Yet, § 230 is no bar to enforcement under the Act at all. When Congress enacted § 230, it carved out an exception for criminal law, explicitly stating that § 230 "shall [not] be construed to impair the enforcement of" 47 U.S.C. § 223, which now includes the Act.[132] Accordingly, despite the Act seeking to hold platforms liable as publishers for their "traditional editorial functions," § 230 will not preempt enforcement of the Act.[133]

B. Enforcement of the Act Contravenes Certain Policy Purposes of § 230

Setting aside its exceptions, the Act's imposition of civil liability on covered platforms for failures to timely remove third-party content directly conflicts with the plain language of § 230 because "deciding whether to … withdraw" content is part of a publisher's "traditional editorial functions." However, because of § 230's explicit exceptions, the Act will be enforced despite contravening policy purposes of § 230. The impositions on platforms under the Act are congressional overreach that risk chilling speech and could create an impractical responsibility to regulate content.

When § 230 came to the House for debate, Representatives Chris Cox and Robert Goodlatte articulated the goals of the bill.[134] First, Representative Cox argued § 230 was intended to protect "Good Samaritan[]" internet platforms who "take[] steps to screen indecency and offensive material for their customers."[135] Second, Representative Cox asserted:

---

131. Kosseff, *supra* note 119, at 769.

132. 47 U.S.C. § 230(e)(1) (2025).

133. *See* VALERIE C. BRANNON & ERIC N. HOLMES, CONG. RSCH. SERV., *Summary* of SECTION 230: AN OVERVIEW, R46751, (Jan. 4, 2024), https://www.congress.gov/crs-product/R46751 [https://perma.cc/36Y8-JEWW (staff-uploaded)] ("[Section 230's] federal immunity generally will not apply to suits brought under federal criminal law … [and] certain privacy laws applicable to electronic communications . . . .").

134. Kosseff, *supra* note 119, at 771–72.

135. *Id.* at 771 (quoting 104 Cong. Rec. H8470 (daily ed. Aug. 4, 1995) (statement of Rep. Cox)).

> [It is] the policy of the United States that we do not wish to have content regulation by the Federal Government of what is on the Internet, that we do not wish to have a Federal Computer Commission with an army of bureaucrats regulating the Internet because frankly the Internet has grown up to be what it is without that kind of help from the Government.[136]

Finally, Representative Goodlatte explained the impracticability of regulating internet platforms as publishers, stating:

> There is no way that any of those [internet platforms] can take the responsibility to edit out information that is going to be coming into them from all manner of sources onto their bulletin board . . . [w]e are talking about something that is going to be thousands of pages of information every day, and to have that imposition imposed on them is wrong.[137]

The Act's takedown process subjects internet platforms to the very concerns voiced by Representatives Cox and Goodlatte, which led to the passage of § 230. Although only 95,820 deepfake videos were found online in 2023, that number represented a 550% increase since 2019.[138] Should the exponential proliferation of deepfake videos[139] continue, covered platforms could be tasked with the responsibility of editing out "thousands of pages of information every day . . . ."[140]

Proponents of the Act's treatment of platforms will argue that the currently small number of deepfake videos online will not create the kind of impracticability that concerned Representative Goodlatte and Congress when they passed § 230. Even if no single platform hosts a significant number of deepfake NCII, the impracticability of the Act's

---

136. *Id.* at 772 (quoting 104 Cong. Rec. H8470 (daily ed. Aug. 4, 1995) (statement of Rep. Cox)).

137. *Id.*

138. HOME SECURITY HEROES, 2023 STATE OF DEEPFAKES (2023), https://www.securityhero.io/state-of-deepfakes [https://perma.cc/44AT-7HUD].

139. *Id.*; AJDER ET AL., *supra* note 10, at 1 (finding a 100% increase in deepfake videos online from 2018 to 2019).

140. Kosseff, *supra* note 119, at 772.

takedown process stems from the potential for platforms to be inundated with fraudulent, bad-faith takedown requests.[141] Therefore, the actual number of deepfake NCII online is irrelevant to the burden the Act's takedown process imposes on covered platforms. How platforms will handle large quantities of takedown requests is unclear, but, clearly, the Act's requirements directly contravene the policy purposes of § 230 articulated by Representatives Cox and Goodlatte.

Further, the FTC will employ "an army of bureaucrats regulating the Internet,"[142] despite the impracticability of requiring internet platforms to comply with the Act. After all, a covered platform could take all reasonable steps available to review a swathe of takedown requests but nevertheless lack the resources to thoroughly review those requests. In such a case, the platform will be forced to either take down the requested content without proper review or risk liability for exercising its traditional editorial functions.

Alternatively, proponents of the Act will argue that it aligns with § 230(b)'s enumerated policies.[143] Specifically, § 230(b)(5) states that it is the policy of the United States "to ensure vigorous enforcement of Federal criminal laws to deter and punish trafficking in obscenity, stalking, and harassment by means of computer."[144] The Act's criminal prong explicitly covers publication of deepfake NCII intended to harass and, given the obscenity analysis of deepfake NCII discussed *infra* Part V, the Act seemingly furthers § 230(b)(5)'s policy.

While Congress decided to explicitly exclude violations of § 223 (which now includes the Act) from § 230's protections, in the Act's case, Congress has usurped the anti-regulation and pro-internet-development policies that justified § 230's passage. However, the Act also furthers § 230(b)'s goal of deterring and punishing internet harassment and obscenity, suggesting that Congress found § 230(b)'s policies to outweigh those articulated by Representatives Cox and Goodlatte. Nevertheless, due to the Act's requirements, covered platforms will face a dilemma: devote significant resources to reviewing takedown requests or risk chilling speech and face liability.

---

141. *See* Kelley, *supra* note 39.

142. *Id.*

143. 47 U.S.C. § 230(b) (2025).

144. *Id.* § 230(b)(5).

V. WILL THE FIRST AMENDMENT TAKE DOWN THE TAKE IT DOWN ACT?

Although § 230 will not affect enforcement of the Act, the Act nonetheless raises the risk of chilling protected speech and may face a First Amendment challenge. This Part will address the scope of the First Amendment, explain the overbreadth doctrine, and analyze how courts may scrutinize the Act under a First Amendment challenge.

A. The First Amendment

The First Amendment commands that "Congress shall make no law . . . abridging the freedom of speech."[145] While speech is commonly associated with spoken words, the First Amendment's protections extend much further to include "written word[s] . . . [and] recorded works, like movies, TV shows, music, video games and social media videos."[146] The Supreme Court has held that "the First Amendment means that government has no power to restrict expression because of its message, its ideas, its subject matter, or its content."[147]

Therefore, as a first step in analyzing whether regulation of speech violates the First Amendment, Courts examine whether the law is "content based or content neutral . . . ."[148] A content-based law "discriminates against speech based on the substance of what it communicates[,]" as opposed to a content-neutral law, which "applies to expression without regard to its substance."[149] The Supreme Court's "First Amendment doctrine … is highly protective of speech[,]"[150]

---

145. U.S. CONST. amend I.

146. *Freedom of Speech*, FREEDOM FORUM https://www.freedomforum.org/freedom-of-speech [https://perma.cc/QZ77-8K5V] (last visited Oct. 12, 2025).

147. *See, e.g.*, Police Dept. of Chicago v. Mosley, 408 U.S. 92, 95 (1972).

148. *See, e.g.*, David L. Hudson Jr., *Content Based*, FREE SPEECH CTR. MIDDLE TENN. STATE UNIV., https://firstamendment.mtsu.edu/article/content-based [https://perma.cc/KL34-3UXA (staff-uploaded)] (last updated July 2, 2024).

149. *Id.*

150. Alyssa Ivancevich, *Deepfake Reckoning: Adapting Modern First Amendment Doctrine to Protect Against the Threat Posed to Democracy*, 49 HASTINGS CONST. L. Q. 61, 68 (2022); *see also* Miller v. California, 413 U.S. 15, 23 (1973) ("We acknowledge, however, the inherent dangers of undertaking to regulate any form of expression.").

holding content-based laws "presumptively unconstitutional and subject to strict scrutiny, the highest form of judicial review . . . ."[151]

However, content-based restrictions are, generally, only permitted when "confined to the few historic and traditional categories [of expression] long familiar to the bar."[152] These categories include, among others, obscenity,[153] defamation,[154] and child pornography.[155] In *Sable Commc'ns of Cal. v. FCC*,[156] the Supreme Court held that while the First Amendment did not protect "interstate transmission of *obscene* commercial telephone messages[,]" a law banning indecent telephone messages nevertheless violated the First Amendment because it was impermissibly content-based.[157]

Aside from content-based distinctions, another way[158] a law can be struck down under the First Amendment is through an overbreadth challenge.[159] Under the overbreadth doctrine, "regulation of speech is

---

151. Hudson Jr., *supra* note 148.

152. United States v. Alvarez, 567 U.S. 709, 717 (2012) (internal quotation and citation omitted).

153. *See, e.g.*, *Miller*, 413 U.S. at 23 ("This much has been categorically settled by the Court, that obscene material is unprotected by the First Amendment.").

154. *See, e.g.*, Curtis Publ'g Co. v. Butts, 388 U.S. 130, 161 (1967) (stripping news publishers of First Amendment protection where their conduct severely departs from standards of press responsibility).

155. New York v. Ferber, 458 U.S. 747, 764 (1982) ("There are, of course, limits on the category of child pornography which, like obscenity, is unprotected by the First Amendment.").

156. 492 U.S. 115 (1989).

157. *Id.* at 115–16 (holding that Section 223(b) of the Communications Act of 1934's denial of adult access to indecent telephone messages went beyond what "is necessary to serve the compelling interest of preventing minors from being exposed to the messages").

158. Other constitutional doctrines, such as the vagueness doctrine, will not be discussed to avoid straying too far outside the scope of this Note. Instead, the vague language of the Act will be factored into analysis of the overbreadth doctrine *infra* Part V.B. For an explanation of the vagueness doctrine and analysis of related case law, see generally, Michael C. Steel, *Constitutional Law-The Vagueness Doctrine: Two-Part Test, or Two Conflicting Tests?*, 35 LAND & WATER L. REV. 255 (2000).

159. *See* Richard Parker, *Overbreadth*, FREE SPEECH CTR. MIDDLE TENN. STATE UNIV., https://firstamendment.mtsu.edu/article/overbreadth [https://perma.cc/7JV7-5J6L] (last updated June 16, 2025); s*ee also* Bd. of Airport Comm'rs v.

footnote continued on next page

unconstitutionally overbroad if it regulates a substantial amount of constitutionally protected expression."[160] The government's motives for creating overbroad regulations are sometimes driven by a desire to suppress speech while avoiding "judicial determinations of content or viewpoint discrimination."[161] For this reason, the Supreme Court has emphasized that "the First Amendment needs breathing space[,] and that statutes attempting to restrict or burden the exercise of First Amendment rights must be narrowly drawn and represent a considered legislative judgment that a particular mode of expression has to give way to other compelling needs of society."[162]

The overbreadth doctrine serves as an exception to the traditional judicial principle against asserting third-party standing, allowing "those to whom the law constitutionally may be applied to argue that it would be unconstitutional as applied to others."[163] A consequence of allowing such challenges "is that any enforcement of a statute thus placed at issue is totally forbidden until and unless a limiting construction or partial invalidation so narrows it as to remove the seeming threat or deterrence to constitutionally protected expression."[164] Yet, the Supreme Court will only impose a limiting construction if a statute is "readily susceptible" to that construction.[165] Thus, the Supreme Court has expressed hesitancy to apply the overbreadth doctrine when asserted on behalf of third-parties.[166] Ultimately, an overbreadth challenge is a likely vehicle for critics of the Act to challenge the constitutionality of the law, because the Act's takedown requirement could potentially burden a substantial amount of constitutionally protected expression, as discussed *infra* Part V.B.

Jews for Jesus, Inc., 482 U.S. 569, 569 (1987) (holding that an airport commission's resolution banning all First Amendment activities was overbroad and violated the constitution because the resolution would reach too much protected speech).

160. Parker, *supra* note 159.

161. *Id.* (emphasis omitted).

162. Broadrick v. Oklahoma, 413 U.S. 601, 611–12 (1973).

163. Parker, *supra* note 159; *Broadrick*, 413 U.S. at 611–12.

164. *Broadrick*, 413 U.S. at 613.

165. Reno v. ACLU, 521 U.S. 844, 884 (1997).

166. *See Broadrick*, 413 U.S. at 613.

B. *The Act's Overbreadth Risks a Substantial Amount of Protected Speech*

The Act's treatment of individuals and platforms raises distinct First Amendment questions. Whether deepfakes generally are protected by the First Amendment is a matter of ongoing debate.[167] Critics argue that AI programs aren't people and therefore cannot speak, "[b]ut the prevailing view is that the First Amendment protects the people who use AI to create deepfakes."[168] Setting aside, for the moment, whether deepfake NCII is protected by the First Amendment, § 223(h)(3), which applies to individuals' creation and publication of deepfake NCII, would likely constitute a content-based restriction. Section 223(h)(3) of the Act makes it unlawful to "knowingly publish a digital forgery" of a protected individual and therefore criminalizes publication of speech based on its content.

Although deepfakes "are essentially lies, which, without criminal behavior, are protected as free speech,"[169] the Act for the first time explicitly criminalizes an application of deepfake technology, suggesting that deepfake NCII is not subject to First Amendment protection. Moreover, even assuming the Act is a content-based restriction, deepfake NCII is also likely to fall within the excepted category of obscenity, and thus, individual publishers would not be afforded the protections of the First Amendment.

Under an obscenity analysis, courts must determine whether the work "appeal[s] to the prurient interest in sex, . . . portray[s] sexual conduct in a patently offensive way, and which, taken as a whole, do[es] not have serious literary, artistic, political, or scientific value."[170]

---

167. Kevin Goldberg, *Are Deepfakes Protected by the First Amendment?*, FREEDOM FORUM (May 21, 2024), https://www.freedomforum.org/deepfakes-protected-by-first-amendment [https://perma.cc/9ARK-7KT7] (arguing that deepfakes generally could be subject to First Amendment protection through rights to access information or because lying is protected by the First Amendment when it doesn't cause actual harm).

168. *Id.*

169. Ken Paulson, *Dealing with Deepfakes: What the First Amendment Says*, FREE SPEECH CTR. MIDDLE TENN. STATE UNIV. (July 10, 2024), https://firstamendment.mtsu.edu/post/dealing-with-deepfakes-what-the-first-amendment-says [https://perma.cc/Q5HL-KXQF] (explaining that while deepfakes constitute expression under the First Amendment, they will not always be subject to protection from civil or criminal liability).

170. Miller v. California, 413 U.S. 15, 24 (1973).

Although a defendant could argue that certain productions of deepfake NCII possess artistic or political value, the Act would likely be enforced in almost all instances of individual enforcement. As such, the Act's regulation of individual creators of deepfake NCII would likely not violate the First Amendment. In this sense, the Act is a success because individual wrongdoers who create harmful deepfake NCII should be held criminally liable for their conduct.

Yet, First Amendment critiques of the Act are not focused on its prohibition of individual publication of deepfake NCII, but rather the Act's rules for covered platforms.[171] To narrow the analysis in this part, consider the following hypothetical. Imagine that a prominent government official, such as the President of the United States, has been subject to intense online criticism. In response, the official calls on his supporters to use a platform's takedown process to flag any critical video, some of which contain deepfake NCII parodies of the official placed in obscene situations. The platform, which hosts billions of third-party-published videos, is now confronted with a similarly tremendous number of takedown requests, some of which may be legitimate and unrelated to the official's situation.

To respond to the requests, the platform could employ a range of options. First, the platform could categorically accept the requests as true and take down all flagged content to ensure it does not make a mistake and violate the Act. This option would likely chill some, if not a substantial amount, of First Amendment protected expression. Second, the platform could painstakingly use human reviewers to analyze each request individually, a task requiring substantial resources. While this option would likely succeed in filtering out bad faith, invalid takedown requests from valid requests, the resource drain to the platform could be significant and potentially hinder investment in their products. Third, as law firms have suggested, the platform could implement automated processes as a method to review takedown requests.[172] Despite this option being less resource intensive than human review, critics warn that automated "content filtering

171. *See, e.g.*, Kelley, *supra* note 39; Letter from Center for Democracy & Technology, et al., *supra* note 39.

172. LATHAM & WATKINS LLP, *supra* note 113.

techniques have significant limitations, tending to lead to the inappropriate takedown and suppression of lawful speech."[173]

As an additional concern for a platform implementing the first or third option, the Act's "safe harbor from liability for removing content that is later determined to be lawful only applies insofar as the covered platform acted in good faith."[174] Although the Act does not define good faith, traditionally, the legal concept of good faith is "used to encompass honest dealing . . . [and] may require an honest belief or purpose . . . ."[175] Therefore, a covered platform's takedown of flagged content would likely need to be based on an honest belief that the content was unlawful under the Act to receive protection from § 223a(a)(4).

On one end of the spectrum, automatically removing content in response to every takedown request would not survive this test and could subject platforms to claims from the individuals whose lawful content was removed. On the other hand, human-reviewed takedowns, which involve more investigation, would likely possess a greater basis for a good faith argument, and thus, covered platforms implementing this method could be protected by the Act from third-party claims. In between these two extremes, covered platforms that use automated tools would have to argue that their tools are reliable enough to create a presumption of honest belief of the unlawfulness of flagged content. Platforms will have to consider how they can demonstrate good faith when deciding how to implement their notice and removal request processes. Yet, as mentioned *supra* Part III, the Act's protection for covered platforms from third-party claims when lawful content is removed could lead to more lawful content being suppressed. And while the First Amendment traditionally does not apply to a private

173. Letter from Center for Democracy & Technology, et al., *supra* note 39; *see also* Kelley, *supra* note 39 (describing online providers depublishing speech rather than verifying it to conform with tight legal frameworks).

174. LATHAM & WATKINS LLP, *supra* note 113.

175. WEX DEFINITIONS TEAM, *Good Faith*, LEGAL INFO. INST. (Jan. 2023), https://www.law.cornell.edu/wex/good_faith [https://perma.cc/NA5B-8VR9].

party's limitation of another's speech,[176] the Act compels platform censorship, thus implicating the First Amendment.

Further complicating the Act's interaction with the First Amendment is the Supreme Court's recent ruling, allowing presidential removal of FTC members at will and without cause.[177] Given that the FTC is charged with enforcement of the Act's platform responsibilities, a president could pressure the FTC to take enforcement action on a platform that has not removed content the president wants taken down, even when they are complying with the Act by not taking down lawful speech, by threatening removal of commissioners. Paired with the potential for a flood of takedown requests, the Supreme Court's decision could increase the risk of suppressing speech.

Ultimately, the Act's First Amendment speech chilling concerns would affect its viability and enforceability only if challenged in court. An overbreadth challenge could be one vehicle for such a challenge in the case of an individual subject to criminal liability under the Act for creating unlawful deepfake NCII. Such an individual, through the overbreadth doctrine's exception to third-party standing, could assert that while their speech is not protected by the First Amendment, the Act nevertheless suppresses a substantial amount of lawful speech due to its takedown requirements for covered platforms. In such a case, courts would have to determine whether protecting victims of deepfake NCII, some of whom are minors, constitutes a compelling

176. Julie Horowitz, *The First Amendment, Censorship, and Private Companies: What Does "Free Speech" Really Mean?*, CARNEGIE LIBR. PITTSBURGH, https://www.carnegielibrary.org/the-first-amendment-and-censorship [https://perma.cc/S2LL-YJF2] (last updated Aug. 2023) ("Social Media platforms are private companies, and . . . private companies are legally able to establish regulations and guidelines within their communities–including censorship of content or banning of members.").

177. *See* Amy Howe, *Supreme Court Allows Trump to Fire FTC Commissioner*, SCOTUSBLOG, https://www.scotusblog.com/2025/09/supreme-court-allows-trump-to-fire-ftc-commissioner [https://perma.cc/SJP9-4HHN] (last updated Oct. 17, 2025) ("Kagan[, dissenting,] wrote that her colleagues in the majority had allowed Trump to remove, contrary to federal law, 'any member he wishes for any reason or no reason at all.' ").

need to society, such that the Act's potential to suppress speech must nevertheless give way.[178]

VI. POLICY PROPOSALS

The danger of deepfakes will continue to rise as technology advances, and it becomes more difficult for humans to distinguish authentic and inauthentic online content. Although deepfake content depicting NCII is the focus of this Note, the technology poses a broader "threat to the public across national security, law enforcement, financial, and societal domains."[179] Both future and existing legislation, including the Act, must contemplate an approach to regulation that balances protecting the public with preserving free speech and ensuring continued technological growth and innovation.

A. Other Proposals Aimed at Combating Deepfakes

While the Act is a necessary step towards a full suite of federal deepfake legislation, further laws must expand on its takedown of deepfake NCII and address non-pornographic applications of deepfake technology. For example, scholars contend that implementing intellectual property and right-of-publicity frameworks of deepfake regulation could be narrowly tailored to promote innovation, and avoid chilling speech while preventing the proliferation of harmful deepfake content.[180] In fact, proposed legislation would do just that and "create a new federal right of

---

178. *See* Broadrick v. Oklahoma, 413 U.S. 601, 611–12 (1973).

179. *See Increasing Threat of Deepfake Identities*, DEP'T OF HOMELAND SEC., https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_deepfake_identities_0.pdf [https://perma.cc/L87U-DNCF] (last visited Oct. 19, 2025) (describing maligned nation-states' use of deepfake personas to promote propaganda and hypothesizing about the potential for deepfakes to be used for financial fraud, corporate sabotage, non-pornographic cyberbullying, and political disinformation).

180. For a discussion on right of publicity and artificial intelligence legislation, see Andrew Street, *Mitigating the Machine: Balancing Innovation with Oversight in the Digital Age*, 3 S. UNIV. L. REV. 1, 12–19 (2025); *see also* Grace Schuette, Atinuke Lardner & Evelyn Woo, *Reckoning with the Rise of Deepfakes*, REGUL. REV. (June 14, 2025), https://www.theregreview.org/2025/06/14/seminar-reckoning-with-the-rise-of-deepfakes [https://perma.cc/4F4A-ET8N] (describing forthcoming law articles arguing for the extension of right of publicity to deepfake content).

publicity specifically for" deepfakes, with similar notice and takedown responsibilities for covered platforms.[181] Elsewhere, criminal law scholars argue that courts must adopt special rules for generative AI used by police during interrogations to elicit confessions.[182]

Staying within the realm of deepfake NCII, members of Congress have proposed legislation that could fill in legal gaps surrounding the Act. In 2024, Senate Judiciary Chair Richard J. Durbin introduced the DEFIANCE Act of 2024[183], proposing a federal civil remedy for victims of deepfake NCII against the individuals who produce, receive, or possess the content with intent to distribute it.[184] While the DEFIANCE Act passed the Senate in 2024, the House version of the bill is still under consideration.[185] The DEFIANCE Act would be another significant arrow in victims' quivers.

Taking a more expansive approach, Representatives Yvette D. Clarke and Glenn Ivy introduced the DEEPFAKES Accountability Act of 2023. Their proposal would require deepfake content to be digitally watermarked and criminalize the failure to identify deepfakes used for NCII, criminal conduct, incitement of violence, and foreign election interference.[186] Congresswoman Clarke's proposal remains in

181. For analysis of the NO FAKES Act, see *Proposed Legislation Reflects Growing Concern Over "Deep Fakes": What Companies Need to Know*, O'MELVENY & MYERS LLP (May 13, 2025), https://www.omm.com/insights/alerts-publications/proposed-legislation-reflects-growing-concern-over-deep-fakes-what-companies-need-to-know [https://perma.cc/T9XS-MHW8].

182. *See* Hillary B. Farber & Anoo D. Vyas, *Truth and Technology: Deepfakes in Law Enforcement Interrogations*, 27 U. PA. J. CONST. L. 977, 998, 1024 (2025) (expressing concern about police using deepfake videos to persuade suspects during an interrogation).

183. S. 3696, 118th Congress (2023–2024) ("Disrupt Explicit Forged Images and Non-Consensual Edits Act of 2024.").

184. *See* Kat Tenbarge, *The Defiance Act Passes in the Senate*, NBC News (July 24, 2024, at 15:28 ET), https://www.nbcnews.com/tech/tech-news/defiance-act-passes-senate-allow-deepfake-victims-sue-rcna163464 [https://perma.cc/QZ3H-25T2].

185. *Id.*

186. H.R. 5586, 118th Congress (2023) ("Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2023."); Press Release, Yvette D. Clarke, Clarke Leads Legislation to Regulate Deepfakes (Sep. 21, 2023), https://clarke.house.gov/clarke-leads-legislation-to-regulate-deepfakes [https://perma.cc/HY94-WS3Z].

the House.[187] Similarly, Congresswoman Anna Eshoo and Congressman Neal Dunn have proposed the Protecting Consumers from Deceptive AI Act,[188] which seeks to establish "standards for identifying AI generated content . . . including watermarking, digital fingerprinting and provenance metadata."[189] Although improved identification of deepfakes generally will help victims of deepfake NCII by validating that the depictions are inauthentic, Congresswomen Clarke's and Eshoo's proposals will have a greater impact in other areas, such as political deepfakes and evidentiary issues in courts and public opinion. Yet, neither proposal is likely to stop the spread of deepfake NCII or punish those who are responsible for its creation because NCII identifiable as a deepfake could still subject a victim to reputational harm and suffering.

Notwithstanding the need for expansive federal legislation to address harmful applications of deepfakes, laws must ensure the protection of free speech and avoid unnecessarily hindering the development of the internet. The TAKE IT DOWN Act is a necessary step towards regulating harmful deepfakes, but is alone insufficient given the broader uses of synthetic media generation. Moreover, as it stands, the Act is overbroad and risks chilling speech by imposing potentially resource-intensive burdens on platforms that are now charged with moderating third-party content.

B. How Congress Can Improve the Act

The Act's criminalization of individual publication of deepfake NCII appropriately contemplates the First Amendment by explicitly carving exceptions for consensual adult pornography and matters of public concern. Nonetheless, the Act's platform requirements risk chilling speech and contravene policy purposes of § 230. As such, Congress should consider amending the Act to respond to the concerns articulated in this Note.

---

187. H.R. 5586, 118th Congress (2023).

188. H.R. 7766, 118th Congress (2024).

189. Matt Bracken, *Bipartisan House Bill Seeks Labeling and Disclosures for AI Deepfakes*, FEDSCOOP (Mar. 21, 2024), https://fedscoop.com/ai-generated-deepfakes-house-bill [https://perma.cc/J9JD-9UZV].

First, Congress should clarify the language of the Act, specifically in its definition of the phrases "matter of public concern," "authorized person," and "valid removal request." Second, the Act should describe the level of certainty required for a platform to determine when a takedown request is made by an "identifiable individual." Third, legislators should ensure platforms are advised on how to maintain "good faith" when responding to takedown requests. These three measures will improve covered platforms' decision-making about what takedown processes to implement and what level of risk they are willing to accept. Consequently, covered platforms will be less likely to inadvertently suppress lawful speech because they could more precisely implement review processes, automated or not.

Fourth, an amendment of the Act should extend its covered platform exceptions to messaging platforms utilizing end-to-end encryption, which otherwise could have to break encryption to facilitate takedown requests. These platforms, in general, are unlikely to cause the damaging, viral spread of NCII content that the Act seeks to prevent. Instead, they operate as individual or group messaging systems, rather than public bulletin board-style social media services where deepfake NCII could be spread to potentially millions or billions. Without additional exceptions specifying who constitutes a covered platform, the Act as it stands could create substantial burdens on too many platforms, hindering development of similar end-to-end encryption technologies.

Fifth, Congress should explicitly extend § 223(h)'s individual exceptions from liability, specifically matters of public concern and consensual adult pornography, to § 223a's requirements for platforms. Currently, a platform could be held liable for failing to remove lawful speech, such as consensual adult pornography, even when the individual creator faces no criminal liability. Fixing this mistake will reduce the likelihood of the Act's takedown requirement suppressing protected speech because platforms will be able to leave content up when they deem it lawful. Of course, some may argue that extending the individual criminal exceptions to platforms would only increase the burden on them, requiring them to conduct further analysis to determine whether to remove content. However, in doing so, platforms would be allowed to take an approach to content

moderation more closely aligned with § 230's policies as they would face reduced liability for third-party publishers' actions.

Finally, to account for the possibility of covered platforms being inundated by bad-faith takedown requests, Congress should create an exception to the forty-eight-hour takedown requirement when platforms receive an overly burdensome number of requests. This exception could even exempt from its protection those platforms that are specifically intended to allow third-parties to create and publish pornographic deepfakes, so that such a platform receiving an abundance of legitimate takedown requests cannot forego speedy action. Overall, future changes to the Act should focus on how to balance holding platforms accountable while easing the burdens of the takedown process. Platforms should not have to choose between spending inordinate amounts of money reviewing takedown requests and potentially suppressing lawful speech.

VII. Conclusion

Deepfake NCII's exponential proliferation has harmed the most socially, politically, and reputationally vulnerable populations. Regulation of deepfakes is necessary, and deepfake NCII is among the least controversial vehicles for Congress to begin developing AI legislation because of its objectionable nature. Individuals who produce, possess, or publish deepfake NCII should be held criminally liable because of the harm they perpetrate. Yet, the government should not impose overly burdensome requirements on platforms that have traditionally been free to moderate third-party content absent government regulation.

The TAKE IT DOWN Act is a step in the right direction, but Congress must ensure it safeguards free speech. By clarifying key language in the Act and extending existing exceptions to platforms, Congress will reduce the likelihood of suppressing lawful expression and simultaneously increase the likelihood that the Act will withstand a First Amendment challenge. Likewise, future legislation aimed at ameliorating the harms of non-pornographic applications of deepfake AI technology must avoid hindering technological development and limiting the public's right to free speech.