SYNTHETIC IMAGES, AUTHENTIC HARMS: A DEFINITIONAL APPROACH TO CRIMINAL NSII "DEEPFAKE" STATUTES

TABLE OF CONTENTS

INTRODUCTION

> I shall not today attempt further to define the kinds of material I understand to be embraced within that shorthand description [of "hard-core" pornography]; and perhaps I could never succeed in intelligibly doing so. But I know it when I see it . . . .[1]

In 1964, Supreme Court Justice Potter Stewart delivered the infamous line, "I know it when I see it," in an attempt to define and describe what "hard-core" pornography is.[2] Unfortunately, with the

1. Jacobellis v. Ohio, 378 U.S. 184, 197 (1964) (Stewart, J., concurring).
2. *Id.*

advent of deepfake pornography, Justice Stewart's struggle to define pornography has only gotten more complicated—how do we define something when we don't know it when we see it?

Between 2019 and 2021, over a dozen women in New York discovered sexually explicit videos of themselves depicting conduct that never occurred.[3] Twenty-year-old Patrick Carey, a former classmate of the women, took photos from their social media accounts and superimposed their faces onto pornographic videos using "deepfake" technology.[4] He posted the videos on pornographic websites with the victims' full names, addresses, and phone numbers, encouraging others to contact the women.[5] Yet, prosecutors faced a major problem: nothing Carey had posted was technically illegal.[6] New York had no statute criminalizing these digitally manipulated explicit images of adults.[7] Only by identifying a single video of an underage victim were prosecutors able to charge Carey through child sexual abuse material (CSAM) statutes.[8]

This case highlights the need to address adult non-consensual synthetic intimate images (NSII) through effective statutes. As deepfake and other artificial intelligence technologies advance, existing criminal laws risk falling behind. While multiple states have attempted to address this problem, the varying approaches in defining criminal NSII have resulted in inconsistent and inadequate protections for current and future harms.[9]

Additionally, some states provide only civil, not criminal, remedies.[10] While civil liability offers relief to victims, these statutes have limitations: civil litigation is time-consuming, expensive, and, often, retraumatizing for victims. Even if a victim succeeds, a defendant may lack means to pay damages. In contrast, a criminal remedy is supported by law enforcement resources and threatens imprisonment, a far greater deterrent to civil liability, particularly against judgment-proof defendants.

For these reasons, this Comment seeks to survey the existing definitions featured in criminal NSII statutes. By examining the statutory language of existing state laws and federal bills, identifying

---

3. *Seaford Man Sentenced to Jail and 10 Years' Probation as Sex Offender for 'Deepfaked' Sexual Images*, DIST. ATT'Y CNTY NASSAU, N.Y. (Apr. 18, 2023), https://perma.cc/V3HE-SNST.

4. Pei-Sze Cheng & Jennifer Millman, *Long Island Man Jailed in Deepfake Sex Scheme Targeting 11 Women from His High School*, NBC 4 N.Y. (Apr. 19, 2023), https://perma.cc/UV3N-D4QF.

5. *Id.*

6. *Id.*

7. *Id.*

8. *Id.*

9. *See infra* Parts II–III.

10. *See* VICTORIA L. KILLION, CONG. RSCH. SERV., LSB10723, FEDERAL CIVIL ACTION FOR DISCLOSURE OF INTIMATE IMAGES: FREE SPEECH CONSIDERATIONS 1 (2022).

strengths and weaknesses, and proposing a definitional framework, this Comment aims to contribute to the development of an effective, forward-looking approach to criminalizing NSII.

Part I of this Comment provides an overview of current artificial intelligence technologies used to make deepfakes and similar media to understand statutory definitions relying on a creation approach. Part II surveys current state laws addressing NSII and demonstrates the various frameworks legislators use in defining criminal NSII. These approaches include defining criminal NSII by using the term "deepfake," describing the computer processes or images that must be used, requiring that the image be sufficiently realistic, or assessing what the image actually depicts. Next, Part III examines the shortcomings of definitions that focus on the technology used or the realism of the content, rather than the harm inflicted on the victim. Lastly, Part IV advocates for an output-oriented approach, emphasizing definitions that center on the resulting NSII itself and the specific harms it causes to victims.

I. BACKGROUND

A. Terminology

Often, the term "revenge porn" has been used to describe various forms of image-based sexual abuse.[11] However, terminology surrounding this abuse has evolved over the past decade to better capture the involuntary and non-consensual nature of "revenge porn."[12] The broader term, "non-consensual intimate imagery," (NCII) refers to any situation where intimate images are produced, published, or reproduced without consent.[13] When NCII is digitally altered in a way that no longer represents actual events, it becomes non-consensual synthetic intimate imagery, also commonly referred to as "deepfake pornography."[14]

Synthetic media is frequently treated as synonymous with media generated by artificial intelligence (AI).[15] However, within the law enforcement context, the term includes all media either "created through digital or artificial means" or "modified or otherwise manipulated through the use of technology, whether analog or digital."[16] In other words, synthetic media is not exclusive to media

11. Rebecca Umbach et al., *Non-Consensual Synthetic Intimate Imagery: Prevalence, Attitudes, and Knowledge in 10 Countries*, CHI '24: PROC. 2024 CHI CONF. ON HUM. FACTORS COMPUTING SYS. 1, 2, https://perma.cc/83FR-B5TW.

12. *Id.*

13. ROHINI LAKSHANÉ, NON-CONSENSUAL INTIMATE IMAGERY: AN OVERVIEW 4 (2024), https://perma.cc/VEB9-GKKM.

14. Umbach et al., *supra* note 11, at 1.

15. U.S. DEP'T OF HOMELAND SEC., INCREASING THREAT OF DEEPFAKE IDENTITIES 5 (2023), https://perma.cc/7LKY-ZQT4.

16. *Id.*

created by AI. While "synthetic media" might be a new term to many, the public has been exposed to synthetic media for decades. For example, images edited with Adobe Photoshop[17] and computer-generated imagery (CGI) in movies[18] fall within the scope of synthetic media. Accordingly, the creation of NSII is not limited solely to AI; rather, it can be made through simple techniques such as using editing software to stitch together clips, add filters, or alter video speed.[19] While this distinction is relevant, the majority of NSII being created currently is through AI.[20]

B. Technology

Understanding the underlying technology behind AI-generated NSII is necessary to accurately and adequately define it in a way that reflects its current state while still anticipating future advances. AI-generated NSII inherently uses some form of AI. Terms like AI, machine learning, deep learning, and neural networks are related technologies, yet are often incorrectly used interchangeably.[21] These technologies are subsets of one another, with the broadest encompassing the smallest: AI is the broadest, followed by machine learning, then neural networks, and, lastly, deep learning.[22]

Generally, AI refers to technology that allows computers and other machines to mimic human learning, perception, planning, and problem-solving.[23] Machine learning is a subset of artificial intelligence technology.[24] It is a technique of sorting and learning from training data in order to make future predictions.[25] Put differently, training data teaches the model to identify and reconstruct future patterns. Machine learning models can look as simple as linear regressions or decision trees, or as complex as artificial neural networks.[26]

Neural networks simulate how the human brain makes predictions.[27] They use processes designed to mimic how biological

17. *Id.* at 9.

18. INTERPOL, BEYOND ILLUSIONS: UNMASKING THE THREAT OF SYNTHETIC MEDIA FOR LAW ENFORCEMENT 8 (2024), https://perma.cc/RLR6-34RR.

19. Umbach et al., *supra* note 11, at 1.

20. *Id.*

21. Yulia Gavrilova, *Artificial Intelligence vs. Machine Learning vs. Deep Learning: Essentials*, SEROKELL (Apr. 7, 2020), https://perma.cc/R3SA-XHYG.

22. *Id.*

23. Cole Stryker & Eda Kavlakoglu, *What Is Artificial Intelligence (AI)?*, IBM (Aug. 9, 2024), https://perma.cc/Y5CC-795D.

24. *Id.*

25. STEVE BLANK, GORDIAN KNOT CTR. FOR NAT'L SEC. INNOVATION, ARTIFICIAL INTELLIGENCE/MACHINE LEARNING EXPLAINED 2 (2022), https://perma.cc/2MBM-NCMZ.

26. *Id.* at 7.

27. *What Is a Neural Network?*, IBM (Oct. 6, 2021), https://perma.cc/D8TD-HZX9.

neurons work to identify intricate relationships and patterns.[28] When neural networks are layered on top of each other, they more closely mimic the complex cognitive power of the human brain.[29] This technique is known as deep learning.[30]

C. What Are Deepfakes?

Deepfakes are a specific type of advanced synthetic media that utilize deep learning.[31] The term is derived from the combination of "deep learning" and "fake."[32] It was first introduced in 2017 by a Reddit user who, employing publicly available AI-supported software, superimposed celebrities' faces onto the bodies of people in pre-existing explicit videos.[33]

To create a deepfake, most AI-generated models must be trained on hundreds or thousands of publicly available images in order to identify and reconstruct patterns.[34] Many of these training sets include explicit content, CSAM, and images of real individuals, which enable the creation of artificially generated NSII and CSAM.[35]

Originally, the main method for creating deepfakes involved generative adversarial networks (GANs), a specific type of deep learning model.[36] The simplest way to understand how GANs work is to think of the process like a two-player game.[37] One player, the generator, creates a realistic, although fake, piece of media, whether it be an image, video, or audio.[38] The other player, the discriminator, attempts to distinguish the fake media from real media.[39] If the discriminator correctly points out the fake, it wins that round and the generator is penalized.[40] As the rounds go on, the generator

---

28. *Id.*
29. Stryker & Kavlakoglu, *supra* note 23.
30. *Id.*; *see* Gavrilova, *supra* note 21.
31. James Vincent, *Why We Need a Better Definition of 'Deepfake*,*'* VERGE (May 22, 2018), https://perma.cc/ZL6G-NVYK.
32. *Id.*
33. *Id.*
34. *See Deconstructing Deepfakes—How Do They Work and What Are the Risks?*, U.S. GOV'T ACCOUNTABILITY OFF.: WATCHBLOG (Oct. 20, 2020), https://perma.cc/M8YR-BUZ6.
35. *See Societal Risks and Well-Being*, NAT'L TELECOMMS. & INFO. ADMIN. (2025), https://perma.cc/4PMW-GJ9D.
36. *See* RAINA DAVIS ET AL., TECH FACTSHEETS FOR POLICYMAKERS: DEEPFAKES 3 (Amritha Jayanti ed., 2020), https://perma.cc/K48Y-862B.
37. *See* Jim Holdsworth & Mark Scapicchio, *What is Deep Learning?*, IBM (June 17, 2024), https://perma.cc/5Q97-RFVW.
38. *Id.*
39. *Id.*
40. *Id.*

eventually learns to synthesize increasingly more realistic media.[41] Eventually, the discriminator can no longer pick out the fake material from the real material.[42]

A second, but far less common, method of deepfake creation uses variational autoencoders (VAEs), another form of deep learning.[43] Unlike GANs, which operate through adversarial competition, VAEs employ a cooperative structure between networks.[44] While used less frequently, VAEs illustrate that alternative deep learning methods are also capable of generating realistic synthetic media.[45]

However, at the forefront of deepfakes and generative AI are diffusion models.[46] Diffusion models are a form of deep learning that improve upon GANs and VAEs.[47] In basic terms, these models generate new data by adding "noise" to the original dataset, and then learn how to "denoise," or clean up, the dataset.[48] Once trained, the model can then apply denoising to random samples of pure noise to generate new images.[49] OpenAI's image generator DALL-E is a popular tool utilizing diffusion models.[50]

While it takes time to train these deepfake models, once the model is created, a user can create a deepfake in mere seconds through various websites and apps, many of which are free to use.[51] For example, in January 2024, AI-generated NSII of Taylor Swift flooded social media platform X.[52] The pictures were likely generated through Microsoft Designer, a free, online text-to-image generator powered by DALL-E.[53]

While deepfakes and synthetic media are often used synonymously, it is important to recognize that deepfakes are only one technique used to create various synthetic images.[54] "Shallowfakes" or "cheapfakes" are videos and images manipulated

---

41. Sophie J. Nightingale & Hany Farid, *AI-Synthesized Faces Are Indistinguishable from Real Faces and More Trustworthy*, 119 PNAS 1, 1 (2022), https://www.pnas.org/doi/epdf/10.1073/pnas.2120481119.

42. *Id.*

43. DAVIS ET AL., *supra* note 36, at 3.

44. *Id.*

45. *See id.*

46. Ayush Karn et al., *Image Synthesis Using GANs and Diffusion Models*, 2023 IEEE INC4 1, 1.

47. *Id.*

48. Dave Bergmann & Cole Stryker, *What Are Diffusion Models?*, IBM (Aug. 21, 2024), https://perma.cc/S2N3-EM6X.

49. *Id.*

50. *Id.*

51. *See* Lutz Finger, *Overview of How to Create Deepfakes—It's Scarily Simple*, FORBES (Sept. 8, 2022), https://perma.cc/LSC4-QSYT.

52. Jess Weatherbed, *Trolls Have Flooded X with Graphic Taylor Swift AI Fakes*, VERGE (Jan. 25, 2024), https://perma.cc/KX23-8JXT.

53. *Id.*

54. U.S. DEP'T OF HOMELAND SEC., *supra* note 15, at 5.

without the use of machine learning; they often use traditional video editing (i.e., Adobe Photoshop) and deceptive techniques to convey a false narrative.[55] Like deepfakes, shallowfakes were first popularized on Reddit after videos surfaced that falsely appeared to show celebrities engaging in sexual acts.[56] Instead of employing deep learning to create NSII, the perpetrator manually edited pre-existing pornographic videos by replacing the original actors' faces with those of various celebrities.[57]

II. SURVEY OF APPROACHES DEFINING NSII

While the term deepfake includes both explicit and non-explicit videos,[58] in practice, 98% of all deepfake videos online are explicit,[59] and at least 90% of them are non-consensual.[60] As a result, state legislatures have moved relatively quickly to address the rise and misuse of deepfakes and other forms of AI generated synthetic media.[61] These synthetic media statutes typically address one of three major categories: election-related and political media, NSII, or CSAM.[62]

A non-exhaustive list of state laws criminalizing NSII include Alabama,[63] California,[64] Colorado,[65] Delaware,[66] Florida,[67] Georgia,[68] Hawaii,[69] Idaho,[70] Indiana,[71] Iowa,[72] Louisiana,[73] Massachusetts,[74]

---

55. Danielle S. Van Lier, *The People vs. Deepfakes: California AB 1903 Provides Criminal Charges for Deepfakes Activity to Guard Against Falsified Defaming Celebrity Online Content*, L.A. LAW., May 2020, at 16, 16.

56. Ashley Stoll, *Shallowfakes and Their Potential for Fake News*, WASH. J.L. TECH. & ARTS (Jan. 13, 2020), https://perma.cc/9UM4-74AQ.

57. *Id.*

58. *See* Vincent, *supra* note 31.

59. *2023 State of Deepfakes*, SEC. HERO (2025), https://perma.cc/6FW7-P2P6.

60. U.S. DEP'T OF HOMELAND SEC., *supra* note 15, at 17.

61. *See Deepfakes in Electoral Campaigns*, MULTISTATE (2025), https://perma.cc/55VV-5EV4.

62. *See* Bill Kramer, *More and More States Are Enacting Laws Addressing AI Deepfakes*, MULTISTATE (Apr. 5, 2024), https://perma.cc/AP8V-JABK.

63. ALA. CODE § 13A-6-240 (2025).

64. CAL. PENAL CODE § 647(j)(4)–(6) (2025).

65. COLO. REV. STAT. §§ 18-7-107 to -109 (2025).

66. DEL. CODE ANN. tit. 11, § 1335 (2025).

67. FLA. STAT. § 836.13 (2025).

68. GA. CODE ANN. § 16-11-90 (2025).

69. HAW. REV. STAT. § 711-1110.9 (2025).

70. IDAHO CODE § 18-6606 (2025).

71. IND. CODE § 35-45-4-8 (2025).

72. IOWA CODE § 708.7 (2025).

73. LA. STAT. ANN. § 14:73.14 (2024).

74. MASS. GEN. LAWS ch. 265, § 43A (2024).

Minnesota,[75] New Hampshire,[76] New York,[77] North Carolina,[78] South Dakota,[79] Texas,[80] Utah,[81] Vermont,[82] Virginia,[83] and Washington.[84] Federally, a bipartisan bill, known as the TAKE IT DOWN Act, criminalizing NSII passed the Senate on December 3, 2024.[85]

While these state and federal laws all attempt to regulate deepfakes and other NSII, legislators vary in defining what actually qualifies as criminally manipulated media. These various approaches include defining NSII by the term "deepfake" itself, the means of creation, the output, the level of realism, or a combination of the four.

A. Defining by "Deepfake"

Colloquially, AI-generated NSII is often used interchangeably with "deepfake pornography."[86] Thus, four states simply use the term "deepfake" or "deep fake" in statutes criminalizing NSII: Delaware,[87] Louisiana,[88] Minnesota,[89] and Texas.[90]

While Delaware, Louisiana, and Minnesota use "deepfake" in their statutes, each state specifically defines deepfake either in how the video is created or how believable the video is.[91]

Texas is a unique state in that, not only has it enacted a criminal NSII statute, but it has also substantively amended the statute's definition. Given the major differences between the original and amended definitions, it is worth discussing both. Unlike the three other states to use "deepfake," Texas's original definition exclusively criminalized a "deep fake video," which it defined as a "video, created with the intent to deceive, that appears to depict a real person

---

75. MINN. STAT. § 617.262 (2025).
76. N.H. REV. STAT. ANN. § 644:9-a (2025).
77. N.Y. PENAL LAW § 245.15 (2025).
78. N.C. GEN. STAT. § 14-202.7 (2025).
79. S.D. CODIFIED LAWS § 22-21-4 (2025).
80. TEX. PENAL CODE ANN. § 21.165 (2025).
81. UTAH CODE ANN. § 76-5b-203 (2025); Criminal Code Recodification and Cross References, H.B. 21, ch. 173, sec. 167, § 76-5b-203, 2025 Utah Laws 324–27 (only amending cross-reference statute sections; no substantive changes).
82. VT. STAT. ANN. tit. 13, § 2606 (2025).
83. VA. CODE ANN. § 18.2-386.2 (2025).
84. WASH. REV. CODE § 9A.86.030 (2025).
85. TAKE IT DOWN Act, Pub. L. No. 119-12, 139 Stat. 55 (2025) (codified at 47 U.S.C. § 223); *All Actions: S.4569—118th Congress (2023-2024)*, CONGRESS.GOV (2025), https://perma.cc/Y8J4-SU6A.
86. Umbach et al., *supra* note 11, at 1.
87. DEL. CODE ANN. tit. 11, § 1335 (2025).
88. LA. STAT. ANN. § 14:73.13 (2024).
89. MINN. STAT. § 617.262 (2025).
90. TEX. PENAL CODE ANN. § 21.165 (2025).
91. *See* DEL. CODE ANN. tit. 11, § 1335(a)(9)(a)(1) (2025); LA. STAT. ANN. § 14:73.13(C)(1) (2024); MINN. STAT. § 617.262(1)(b) (2025).

performing an action that did not occur in reality."[92] Thus, unlike the other states using variations of the term "deepfake," Texas's original definition omitted a level of realism or how the NSII must be made.

B. Defining by Creation

Another approach to defining deepfakes and AI-generated synthetic media involves focusing on the creation of the media. This approach can be further broken down into two categories, labeled the "technique approach" and the "input approach."

1. Technique Approach

The technique approach defines NSII in reference to the technical process used to create it.[93] Definitions under this framework can greatly vary in scope. A broad statutory definition may encompass any NSII generated through AI. Conversely, another may narrowly tailor its definition to NSII generated through specific deep learning models, particularly those most associated with deepfakes.

Indiana and North Carolina define NSII exclusively based on the technical means and models used to produce the resulting image or video. Indiana regulates explicit images "created or modified by means of a computer software program, artificial intelligence, application, or other digital editing tools."[94] North Carolina regulates explicit images "created, adapted, or modified by technological means, including algorithms or artificial intelligence."[95]

Delaware, in addition to requiring a certain level of realism, also regulates images in a specific technical manner.[96] To overcome the lack of a universally agreed upon definition, it specifically defines deepfake as synthetic media, which "means an image that has been created or intentionally manipulated with the use of generative adversarial network techniques or other digital technology."[97] Delaware is unique in its explicit mention of GANs.

2. Input Approach

The input approach defines NSII by specifying the information that is used in its creation, irrespective of the technical process employed.[98] These definitions require that the perpetrator either uses the victim's image, another real person's image, a computer-

92. TEX. PENAL CODE ANN. § 21.165 (2024).

93. *See* IND. CODE § 35-45-4-8 (2025); N.C. GEN. STAT. § 14-202.7 (2025); DEL. CODE ANN. tit. 11, § 1335 (2025).

94. IND. CODE § 35-45-4-8(c)(2)(C) (2025).

95. N.C. GEN. STAT. § 14-202.7(a)(3) (2025).

96. *See* DEL. CODE ANN. tit. 11, § 1335 (2025).

97. *Id.* § 1335(a)(9)(a)(8).

98. *See* HAW. REV. STAT. § 711-1110.9 (2025); IOWA CODE § 708.7 (2025); LA. STAT. ANN. § 14:73.14 (2024); N.Y. PENAL LAW § 245.15 (2025); VT. STAT. ANN. tit. 13, § 2606 (2025); VA. CODE ANN. § 18.2-386.2 (2025).

generated image, or some combination of the three to create the NSII. Hawaii,[99] Iowa,[100] Louisiana,[101] New York,[102] Vermont,[103] and Virginia[104] employ an input approach.

Iowa and Virginia criminalize NSII made using the victim's image to create, adapt, or modify the resulting output.[105] Thus, NSII produced with images of other real people and/or computer-generated images would not be criminal under these two statutes because they do not use the victim's photo in its production.

In direct contrast, New York regulates images created or altered by digitization, which is defined as the process of "alter[ing] an image in a realistic manner utilizing an image or images of a person other than the person depicted, or computer-generated images."[106] Thus, an illegal NSII under the New York statute must use an image of a real person other than the victim or use computer-generated images as inputs.

Similarly, Vermont criminalizes NSII made through digitization.[107] The state tweaks the New York definition, defining digitization as "the process of altering an image in a realistic manner utilizing an image or images of a person, including images other than the person depicted, or computer-generated images."[108] Vermont's statute criminalizes all the NSII criminalized under the New York statute plus any NSII where an image of the victim is used as an input.

While not as direct in its input requirement as the previous states, Hawaii also criminalizes NSII by specifying inputs used to create it, in a way.[109] The statute criminalizes an "image or video of a composite fictitious person."[110] According to the Merriam-Webster Dictionary, the plain meaning of the term "composite" is something "made up of distinct parts or elements."[111] Additionally, the Hawaii Legislature's findings included that "deep fake technology enables the creation of synthetic media in which a person in an existing image or

99. HAW. REV. STAT. § 711-1110.9(1)(c) (2025).
100. IOWA CODE § 708.7(1)(a)(5) (2025).
101. LA. STAT. ANN. § 14:73.13(C)(1) (2024).
102. N.Y. PENAL LAW § 245.15(2)(d) (2025).
103. VT. STAT. ANN. tit. 13, § 2606(a)(6) (2025).
104. VA. CODE ANN. § 18.2-386.2(A) (2025).
105. IOWA CODE § 708.7(1)(a)(5) (2025) ("an individual . . . whose image is used to create, adapt, or modify a visual depiction"); VA. CODE ANN. § 18.2-386.2(A) (2025) ("a person whose image was used in creating, adapting, or modifying a videographic or still image").
106. N.Y. PENAL LAW § 245.15(2)(d) (2025).
107. VT. STAT. ANN. tit. 13, § 2606 (2025).
108. *Id.* § 2606(a)(6).
109. *See* HAW. REV. STAT. § 711-1110.9 (2025).
110. *Id.* § 711-1110.9(1)(c).
111. *Composite*, MERRIAM-WEBSTER (2025), https://perma.cc/HQJ5-8NF2.

video is replaced with the likeness of another person."[112] These findings indicate that a deepfake must involve both an image of a person in a pre-existing image or video along with an image of another person. Therefore, based on the plain meaning of "composite" and the legislative history of this statute, criminal NSII in Hawaii likely requires both an image of the victim and another person.

Louisiana's statute employs a similar input approach to Hawaii along with a reasonable person approach discussed later.[113] Regarding its input definition, the state defines deepfake as a form of media manipulated such that it would "replace an individual's likeness with another individual" depicted in the recording.[114] Thus, like Hawaii, Louisiana seems to only criminalize NSII involving both images of the victim and another person.

C. Defining by Output

In contrast to the input approach, Minnesota[115] and Florida[116] follow an output approach, which effectively specifies how the inputs must be used. Florida defines an "altered sexual depiction" as one that as a result of some modification, depicts the victim "1. With the nude body parts of another person as the nude body parts of the identifiable person; 2. With computer-generated nude body parts as the nude body parts of the identifiable person; or 3. Engaging in sexual conduct . . . in which the identifiable person did not engage."[117] Minnesota's definition is effectively the same.[118] Thus, instead of focusing on how the image or video was created or what the inputs used were, Florida and Minnesota look to regulate based on the resulting image itself.

D. Defining by Realism

Lastly, a state may define NSII by how realistic the image itself is. Specifically, these statutes require the media to be so realistic that a viewer would believe it is a true depiction of the victim.

---

112. S.B. 309, 2021 Leg., 31st Sess. (Haw. 2021).

113. *See* LA. STAT. ANN. § 14:73.13 (2024); *see also infra* Section II.D.1.

114. LA. STAT. ANN. § 14:73.13(C)(1) (2024).

115. *See* MINN. STAT. § 617.262 (2025).

116. *See* FLA. STAT. § 836.13 (2024). Notably, Florida amended this statute to criminalize not only the distribution of NSII, but also the generation of NSII itself. *See* FLA. STAT. § 836.13 (2025).

117. FLA. STAT. § 836.13(1)(a) (2025).

118. MINN. STAT. § 617.262(2)(2) (2025) (defining deepfake as realistically depicting "(i) the intimate parts of another individual presented as the intimate parts of the depicted individual; (ii) artificially generated intimate parts presented as the intimate parts of the depicted individual; or (iii) the depicted individual engaging in a sexual act").

1. Reasonable Person and Authenticity Approach

California,[119] Louisiana,[120] Massachusetts,[121] Texas,[122] South Dakota,[123] and the federal TAKE IT DOWN Act[124] all require in some way that a reasonable person would view the NSII as authentic. Notably, while Massachusetts, Texas, and the federal act include creation methods in their definitions, the listed technologies are so broad that, in effect, they define by a level of authenticity.[125]

California and South Dakota's definitions exclusively rely on whether an image would result in a reasonable person believing that the image is authentic. The California statute requires that the image "would cause a reasonable person to believe the image is an authentic image of the person depicted."[126] South Dakota uses almost identical wording with the only substantive change being the insertion of "mistakenly" to modify "believe."[127] Similarly, Louisiana, in addition to its input definition, offers a second definition of deepfake as media manipulated such that it "would falsely appear to a reasonable observer to be an authentic record of the actual speech or conduct of the individual."[128] Massachusetts uses practically the same definition as Louisiana.[129] Lastly, Texas's amended definition and the federal act somewhat deviate from the previous language, requiring that the image, when viewed by a reasonable person, appears "indistinguishable from an authentic visual depiction of the [person]."[130]

2. Disclaimers

In addition to requiring a certain level of realism, some of these states also explicitly address the impact of a disclaimer.[131] For example, a potential disclaimer might look like a watermark over NSII stating, "This video is a deepfake. The events shown did not

119. CAL. PENAL CODE § 647(j)(4)(A)(ii) (2025).
120. LA. STAT. ANN. § 14:73.13(C)(1) (2024).
121. MASS. GEN. LAWS ch. 265, § 43A(b)(1) (2024).
122. TEX. PENAL CODE ANN. § 21.165 (2025).
123. S.D. CODIFIED LAWS § 22-21-4(3)(a) (2025).
124. 47 U.S.C. § 223(h)(1)(B) (2025).
125. *See* MASS. GEN. LAWS ch. 265, § 43A(b)(1) (2024) (listing "creation or alteration of visual material including, but not limited to, through the use of computer-generated images"); TEX. PENAL CODE ANN. § 21.165 (2025); 47 U.S.C. § 223(h) (2025).
126. CAL. PENAL CODE § 647(j)(4)(A)(ii) (2025).
127. S.D. CODIFIED LAWS § 22-21-4(3)(a) (2025) (requiring that the image "would cause a reasonable person to mistakenly believe that the image or recording is authentic").
128. LA. STAT. ANN. § 14:73.13(C)(1) (2024).
129. MASS. GEN. LAWS ch. 265, § 43A (2024).
130. 47 U.S.C. § 223(h)(1)(B) (2025); TEX. PENAL CODE ANN. § 21.165(a)(1) (2025).
131. *See* LA. STAT. ANN. § 14:73.13 (2024); FLA. STAT. § 836.13(6) (2025).

happen in real life." In some states, this disclaimer might make the video legal; in others, it may be irrelevant.

Specifically, Louisiana's statute states that a deepfake does not include any material featuring "a clear disclosure visible throughout the duration of the recording that would cause a reasonable person to understand that the audio or visual media is not a record of a real event."[132] This "disclaimer defense" reflects Louisiana's approach to defining a deepfake in terms of authenticity: if a disclosure makes it clear that the content is not a record of real events, then the material does not qualify as criminal NSII under its statute.[133] In contrast, Florida's statute expressly removes the possibility of a disclaimer defense. Specifically, "[t]he presence of a disclaimer . . . that the person or persons depicted did not actually perform the actions portrayed, is not a defense and does not relieve a person of criminal liability."[134]

While the TAKE IT DOWN Act does not explicitly address disclaimers, its legislative history indicates it likely would allow for a disclaimer defense. The DEFIANCE Act, a bill passed in the Senate on July 23, 2024, provides civil remedies to victims and specifically covers unrealistic images featuring identifiable individuals.[135] It specifically includes that NSII is still actionable under the act "regardless of whether a label, information disclosed with the visual depiction, or the context or setting in which the visual depiction is disclosed states or implies that the visual depiction is not authentic."[136] The TAKE IT DOWN bill, introduced six months after the DEFIANCE bill, replicates the civil bill's definition in its entirety, with the sole difference being the omission of the disclaimer statement.[137]

132. LA. STAT. ANN. § 14:73.13(C)(1) (2024).

133. *See id.*

134. FLA. STAT. § 836.13(6) (2025).

135. DEFIANCE Act of 2024, S. 3696, 118th Cong. § 3(a)(3)(A)–(B) (2024); *All Actions: S.3696—118th Congress (2023-2024)*, CONGRESS.GOV (2025), https://perma.cc/BY6H-2YG8.

136. DEFIANCE Act of 2024, S. 3696, 118th Cong. § 3(a)(3)(B) (2024).

137. *See* 47 U.S.C. § 223(h)(1)(B) (2025).

III. PROBLEMS WITH DEFINING NSII BY TERM, CREATION, OR REALISM

A strong statutory definition of NSII must balance specificity with adaptability to technological changes to effectively protect victims of NSII. Definitions closely tied to the technology itself, whether it be through technique or input definitions, risk excluding emerging generative AI or over-inclusivity. At the same time, definitions imposing strict realism requirements set an unnecessarily high bar, disregarding the reality that NSII often depicts victims in implausible scenarios. Both approaches overlook a critical factor: the harm of synthetic media arises not from how convincingly they imitate reality or the technology used to create them but from the context in which they are deployed and the impact on those they target.

A. Drawbacks of Defining by Term

Texas has since revised its criminal NSII statute, but the flaws in its earlier version illustrate the risks of drafting overly expansive definitions. Although the original statute defined "deep fake video" as requiring an intent to deceive, a court may not adhere exclusively to this definition if it is too broad. Under customary statutory interpretation principles, if a legislature defines a term in the statute, that definition is usually binding on courts, even if the definition varies from the term's plain or commonly understood meaning.[138] However, when there is discord between the plain meaning of the term and the reach of the definition, a court may consider the plain meaning of the word.[139]

For example, in June 2016, Kanye West released a music video for his single *Famous*, featuring nude wax models of celebrities, including figures like Rihanna, Donald Trump, and George W. Bush.[140] The video is intentionally grainy and the wax figures were designed to look incredibly realistic, indicating intent to deceive viewers that these were the actual celebrities.[141] In other words, the

138. SHAMBIE SINGER, 3 SUTHERLAND STATUTES AND STATUTORY CONSTRUCTION § 59:8 (8th ed. 2024). The Texas Legislature has also codified this interpretation principle, requiring that words that have acquired a particular meaning by statutory definition must be construed accordingly. TEX. GOV'T CODE ANN. § 311.011(b) (1985).

139. *See* Bond v. United States, 572 U.S. 844, 861 (2014); Johnson v. United States, 559 U.S. 133, 136 (2010).

140. Guardian Staff, *Last Night Was Mad Real: Kanye's New Video Depicts Nude Trump, Taylor Swift*, GUARDIAN (June 25, 2016), https://perma.cc/6AYK-ZZNW.

141. *Id.*; *see* Jennifer Fletcher, *Kanye Was "Obsessed with the Anatomy" of Caitlyn's Wax Figure*, YAHOO!LIFE (July 7, 2016), https://perma.cc/49JU-FC8Y (reporting that the figures took over six months to make, with Kanye allegedly paying over $750,000 for the wax figures).

video appeared to depict real celebrities lying in bed together, an event that did not occur in reality, with their intimate parts exposed. Assuming that Kanye did not get consent from at least one of the celebrities portrayed in the video, this video would fit Texas's expansive definition of a criminal "deep fake video."

However, outside of the statute, this video would ordinarily never be described as a deepfake. The celebrities' likenesses were not transformed through the use of any form of software, let alone AI associated with deepfakes. Rather, these were physical, handcrafted wax models. Given the dissonance between the ordinary usage of deepfake and the reach of the Texas definition, the ordinary meaning of deepfake is likely relevant.

Yet, the ordinary meaning of deepfake is far from standardized. While many Americans are familiar with the word,[142] using it in these statutes complicates, rather than clarifies, what counts as criminal NSII. While "deepfake" is generally understood to refer to the manipulation of existing media or the generation of new media, the terms AI, machine learning, and deep learning are often used interchangeably to describe that process of manipulation.[143]

Since the term deepfake is derived from deep learning, many define deepfake as inherently requiring the use of deep learning.[144] Under this interpretation, forms of NSII, even those still generated with AI, would not be criminal. For example, a perpetrator using Adobe Photoshop and images from a victim's social media page could easily create shallowfake NSII yet escape criminal liability.[145] In addition to excluding less complicated images, this definition might not be able to grow with the future of technology. Emerging approaches such as neurosymbolic AI[146] and quantum computing[147] are seen as the future of AI technologies. However, neither of these falls into the scope of deep learning. Accordingly, as technology progresses, relying on the term "deepfake" to criminalize NSII is likely a short-lived solution.

---

142. *See* Umbach et al., *supra* note 11, at 6 (finding that approximately 57.5% of Americans surveyed had heard the phrase before).

143. *See id.*; *see* Gavrilova, *supra* note 21.

144. *See* Vincent, *supra* note 31.

145. *See, e.g.*, Tamar Lapin, *Her Revenge Porn Problem Is Real, but the Photos Are Fake*, N.Y. POST (Dec. 26, 2017), https://perma.cc/8584-4VVN.

146. David Meyer, *Generative AI Can't Shake Its Reliability Problem. Some Say 'Neurosymbolic AI' Is the Answer*, FORTUNE (Dec. 18, 2024), https://perma.cc/8QH9-MNKE.

147. Beth Stackpole, *Quantum Computing: What Leaders Need to Know Now*, MIT SLOAN SCH. MGMT.: IDEAS MADE TO MATTER (Jan. 11, 2024), https://perma.cc/E4F4-MY2A.

B. Drawbacks of Defining by Creation

Both technical definitions and input definitions are connected to the NSII's underlying creation process. Because technology is constantly evolving, these statutes rely on dynamic concepts where a fixed definition is necessary. This unsteadiness is reflected in the overinclusive and underinclusive tendencies of the technique approach and the instability over time of the input approach.

1. Imprecision in Technique Approaches

Like Texas, Delaware's definition, which relies on a technique approach, is vulnerable to overly narrow interpretations that curb the statute's effectiveness. While its criminalization of NSII created by GANs or "other digital technology,"[148] may seem broad as a result of its catchall clause, its reference to GANs could be limiting. Under the canon of construction *noscitur a sociis*, "a word is 'given more precise content by the neighboring words with which it is associated.'"[149] Thus, while digital technology refers to any technology that uses binary,[150] "other digital technology" likely would be defined by reference to GANs. As discussed in Part I, GANs are a specific type of deep learning model. Accordingly, "other digital technology" is likely to be interpreted as extending only to other deep learning models used to create NSII, such as VAEs and diffusion models like DALL-E. Thus, the Delaware statute could exclude primitive shallowfake NSII, while still failing to account for future technologies.

In contrast, Indiana's comprehensive definition covers all AI-generated NSII and shallowfake NSII.[151] While containing a similar catchall to Delaware, it avoids the pitfalls of *noscitur a sociis* by referencing broader technology categories. Although the definition avoids under-criminalizing NSII, its expansiveness allows the over-criminalization of likely unintended categories of media, such as cartoonish, digitally-created images. For example, in 1984, *The Nation* printed a salacious political cartoon critiquing Henry Kissinger's foreign policy.[152] Entitled "Screwing the World," illustrator David Levine depicted Kissinger naked, engaged in sexual conduct with an anthropomorphized globe.[153] Had this same image been digitally drawn, it would be illegal NSII.

---

148. DEL. CODE ANN. tit. 11, § 1335(a)(9)(a)(9) (2025).

149. Fischer v. United States, 144 S. Ct. 2176, 2183 (2024) (quoting United States v. Williams, 553 U.S. 285, 294 (2008)).

150. *See Digital*, MERRIAM-WEBSTER (2025), https://perma.cc/R642-2DJC.

151. IND. CODE § 35-45-4-8(c)(2)(C) (2025).

152. Eric Sharfstein, *Victor Navasky Explores the Power of Political Cartoons*, COLUM. NEWS (Apr. 18, 2013), https://perma.cc/7GAJ-H6VT.

153. *Id.*; John Palattella, *In Our Orbit:* Victor Navasky*'s The Art of Controversy: Political Cartoons and Their Enduring Power*, NATION (June 3, 2013), https://perma.cc/T2EC-KFEL.

While this statute would almost certainly implicate First Amendment overbreadth problems, an issue outside the scope of this Comment, its overbreadth is not limited to speech. Indiana's definition is overinclusive, penalizing conduct beyond its intended target by failing to distinguish between different types of harm. For example, assuming a digital cartoon similar to Levine's had no significant artistic or political value,[154] its inclusion under the statute would extend criminal liability to expressive conduct that, while potentially offensive, does not inflict the same kind of harm as NSII. Although both scenarios involve nonconsensual depictions, the harms suffered are categorically different. While Kissinger may have felt criticized or offended by the cartoon, NSII victims, such as Taylor Swift, endure sexual exploitation, harassment, and profound invasions of privacy.[155] By treating these distinct harms as equivalent, the statute risks sweeping in conduct that, while undesirable, should not be subject to the same criminal penalties.

2. Instability in Input Approaches

The input approach provides more precise definitions than the technique approach. However, while these definitions are somewhat effective currently, they are vulnerable to becoming outdated as the methods employed to create NSII evolve.

Input definitions require some combination of a victim's image, another person's image, or a computer-generated image to be used in the creation of the NSII.[156] For AI-generated NSII, this requirement essentially means that one of those images is contained in the model's training data.[157] Currently, one of these images seems essential to create NSII. However, recent developments in deep learning could change that.

Zero-Shot Learning (ZSL) is a process that attempts to train a model to classify images from unseen classes by transferring knowledge from seen classes.[158] While ZSL models are not generative AI, data scientists have recently proposed a Discriminative Image Generation framework for ZSL (DIG-ZSL).[159] DIG-ZSL allows for the

154. *See* Miller v. California, 413 U.S. 15, 24 (1973) (holding that obscenity is outside the scope of the First Amendment and that obscene material must lack "serious literary, artistic, political, or scientific value").

155. *Taylor Swift Deepfakes Spread Online, Sparking Outrage*, CBS NEWS (Jan. 26, 2024), https://perma.cc/Y52S-25RK.

156. *See supra* Section II.B.2.

157. *See* SIMRANJEET SINGH, RAJNEESH SHARMA & ALAN F. SMEATON, USING GANS TO SYNTHESISE MINIMUM TRAINING DATA FOR DEEPFAKE GENERATION 2 (2020), https://perma.cc/3FP9-55WE.

158. Isaac Sacolick, *Zero-Shot Learning and the Foundations of Generative AI*, INFOWORLD (Feb. 13, 2023), https://perma.cc/9FKJ-263N.

159. DINGJIE FU ET AL., DISCRIMINATIVE IMAGE GENERATION WITH DIFFUSION MODELS FOR ZERO-SHOT LEARNING 1 (2024), https://perma.cc/T5AS-UE3H.

generation of images "based on a text-to-image diffusion model," such as stable diffusion.[160] To provide an oversimplified example: consider a model trained on pictures of a class of birds, known as the "seen" class. From the seen class, the model would learn characteristics of the birds, such as shapes and textures. Then, the model would be given a textual description of a bird it has never seen before, i.e., "yellow-headed blackbird," known as the "unseen" class. Using its prior knowledge of the seen classes and semantic information from the textual description, the model would synthesize a realistic image of the yellow-headed blackbird. This process is inherently different from the typical GAN approach associated with deepfakes, which relies heavily on training data.[161]

The future of AI-generated images could involve models like DIG-ZSL that can generate depictions of things it has never seen. In the NSII context, this might look like a model trained on human features that, if given a descriptive enough textual input, could produce a realistic image of a victim, without having ever seen an image of them. A more advanced model's training data may not contain an image of a real human person at all. Rather, it might have images of animals or anatomical drawings of humans. Since a DIG-ZSL model is able to generate images of things it has never seen before, creating NSII using images of real people could become obsolete.

One may argue that since these models rely on text-to-image inputs, creating an identifiable person based on attribute descriptions would be complicated. However, perpetrators already spend time attempting to get the "perfect" text-to-image results with current technology. For example, the January 2024 NSII of Taylor Swift was traced to be produced by Microsoft Designer, an AI text-to-image program.[162] Despite Microsoft Designer being programmed not to produce explicit images, perpetrators from a Telegram group, which was explicitly formed to try to work around these guidelines, spent hours attempting to manipulate the system with deceptive textual inputs.[163] Thus, while the idea of perpetrators spending the time to try and retry descriptors to get the perfect lookalike of a victim might seem dubious at first, given past behaviors, it seems almost guaranteed. Thus, with new generative models, statutes that only criminalize NSII using certain images are vulnerable to becoming outdated.

While input definitions currently cover the vast majority of NSII without issues of overbreadth or excessive narrowness, their precision is unlikely to be workable over time. While they currently provide a

160. *Id.*

161. *See supra* Part I.

162. Ashley King, *Microsoft Says It Closed Loophole That Allowed Taylor Swift AI Porn to Be Created*, DIGIT. MUSIC NEWS (Feb. 1, 2024), https://perma.cc/J7UY-EM6M.

163. *Id.*

more effective framework, input definitions fail because of their connection to the creation process. Definitions relying on the underlying technology are inherently unstable because the leading technology itself is unstable. An effective definition must be tied to a fixed concept, rather than a constantly evolving one.

C. Rejecting Authenticity and Disclaimer Defenses

In addition to NSII, a significant body of legislation regulating deepfakes and synthetic media arises within the election context.[164] As a result, some states addressing election-related synthetic media and NSII rely on nearly identical statutory definitions.[165] This copy-and-paste legislation is particularly problematic when states employ a requisite level of realism in their definitions. For example, California's Elections Code and Penal Code have practically the same requirement that a reasonable person would believe the deepfake is an authentic depiction of the victim.[166] Likewise, Alabama requires in both the election and NSII context that a reasonable person would believe the media actually depicts the victim.[167] By using the same definition, these statutes assume that deepfakes pose the same harm regardless of the context.

Yet, perpetrators do not create deepfakes in a vacuum. Rather, they "often have discernible goals or specific motivations to misuse and abuse [generative AI]."[168] In studying these motivations, Google researchers identified broad categories underlying the creation of AI-generated images.[169]

In the election context, researchers found that perpetrators primarily sought to manipulate opinions by spreading disinformation.[170] Society suffers the greatest harm from these deepfakes. Deliberate dissemination of false information during elections poses significant threats to democratic processes and erodes trust in electoral systems.[171] Because the victim is society, perpetrators only succeed when their deepfakes are sufficiently

---

164. Kramer, *supra* note 62.

165. *Compare* CAL. ELEC. CODE § 20512(d) (2025), *with* CAL. PENAL CODE § 647(j)(4)(A)(ii) (2025). *Compare also* ALA. CODE § 17-5-16.1(a)(5) (2025), *with id.* § 13A-6-240(b).

166. *See* CAL. ELEC. CODE § 20512(d) (2025); CAL. PENAL CODE § 647(j)(4)(A)(ii) (2025).

167. *See* ALA. CODE § 17-5-16.1(a)(5)(b) (2025); *id.* § 13A-6-240(b)(2)(b).

168. NAHEMA MARCHAL ET AL., GENERATIVE AI MISUSE: A TAXONOMY OF TACTICS AND INSIGHTS FROM REAL-WORLD DATA 11 (2024), https://perma.cc/F7RR-CEGD.

169. *Id.* at 11, 25–26.

170. *Id.* at 11–13.

171. *See* Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. 1753, 1758 (2019).

realistic to misinform. Thus, definitions containing a requisite level of realism in political deepfake statutes effectively address the underlying motivation and resulting harm of the media in the election context.

In contrast, NSII perpetrators create deepfakes for their own self-gratification, rather than trying to convince other viewers that the image or videos are real.[172] Thus, participants are fully aware that the images are fabricated. Unlike political deepfakes, where society as a whole bears the harm, NSII inflicts direct, personal harm on the individual depicted. Victims depicted in NSII describe it as a form of digital rape.[173] The fact that an image is clearly inauthentic, whether due to disclaimers or poor quality, does not lessen the violation of autonomy.

The 2024 Taylor Swift NSII incident illustrates this. The images featured realistic-looking depictions of her engaging in sexually explicit conduct in "implausible" settings.[174] Yet, the lack of believability did not prevent users from circulating and viewing the images over 45 million times.[175] The knowledge that these were not "authentic" or "actual" depictions of Taylor Swift did not prevent the harm inflicted on her.

This disconnect is further evidenced by the broader landscape of AI-generated NSII, where the overwhelming majority of victims are female celebrities.[176] The sheer volume of explicit content featuring some celebrities makes it implausible for a reasonable person to believe that any one person could realistically appear in thousands of these videos, especially when these depictions involve absurd or highly unlikely scenarios.

To effectively address the unique harms of NSII, states must create definitions that acknowledge the different purposes behind the images and the specific harms they inflict. While the reasonable person standard and authenticity requirements serve to protect public trust and prevent misinformation in the election context, those same standards undermine the criminalization of NSII, where the primary concern is safeguarding individual privacy and dignity.

---

172. Emma Grey Ellis, *People Can Put Your Face on Porn—and the Law Can't Help You*, WIRED (Jan. 26, 2018), https://perma.cc/W7PQ-AH2C.

173. EJ Dickson, *TikTok Stars Are Being Turned Into Deepfake Porn Without Their Consent*, ROLLING STONE (Oct. 26, 2020), https://perma.cc/G4RA-5HBW.

174. Becca Branum, *NDII Victims Deserve Help. Let's Build an Effective Takedown System*, CTR. FOR DEMOCRACY & TECH. (Nov. 12, 2024), https://perma.cc/P6GD-GHWN.

175. *Id.*

176. *See* HENRY AJDER ET AL., DEEPTRACE, THE STATE OF DEEPFAKES: LANDSCAPE, THREATS, AND IMPACT 2 (2019), https://perma.cc/NZ4Z-9AR6 ("All but 1% of the subjects featured in deepfake pornography videos were actresses and musicians working in the entertainment sector.").

IV. CONSTRUCTING THE PROPER DEFINITION

A. An Output-Focused Approach

While the previous analysis explores the pitfalls of definitions focused on the underlying technology or the realism of the NSII, constructing an effective statute requires centering the actual harms inflicted upon the victims in the definition. Accordingly, an output approach, like those employed by Minnesota and Florida, offers the most comprehensive protection for victims without risking issues of overbreadth or excessive narrowness.

These statutes succeed by addressing the fundamental ways NSII manifests: the victim's likeness superimposed on another person's naked body, the victim's likeness superimposed on a fabricated body, or the victim falsely depicted as engaging in sexual conduct. Instead of focusing on the instruments used to create the image, these definitions prioritize the outcome—the instrument that actually creates the harm. Therefore, a proper output approach should define the criminal NSII as: media that appears to realistically depict (1) the nude body parts of another person as the nude body parts of the identifiable person, (2) computer-generated nude body parts as the nude body parts of the identifiable person, or (3) sexual conduct in which the identifiable person did not engage.

Focusing on the resulting media strengthens the statute by ensuring it will not be outpaced by technological advancements. Unlike the input approach, this approach would remain effective even if a perpetrator did not use a single image, computer-generated or real, of the victim. Statutes focused on the resulting NSII could withstand massive technological changes. To illustrate, a future technology could allow people to generate images solely based on their brain waves.[177] With this type of technology, perpetrators could eventually create NSII without a pre-existing image. A statute relying on an input-definition would fail to address the resulting NSII. However, under an output-focused definition, the method of creation is irrelevant: what matters is the realistic portrayal of the victim in an intimate non-consensual context.

While this definition can withstand changes to technology, it does not over-criminalize digitally manipulated media. Unlike Indiana's broad approach, which extended criminal liability to any image altered by a digital editing tool, regardless of context, an output definition provides that the NSII must be a realistic depiction. Importantly, this definition does not require authenticity or indistinguishability from real life. Thus, this definition excludes cartoons or stylized explicit images from criminalization without disregarding the harm victims face.

---

177. Peter Bentley, *You'll Soon Be Able to Record Your Dreams. Here's How*, BBC SCI. FOCUS (Aug. 18, 2024), https://perma.cc/CAX2-SLBL.

An output definition would be successful because it would focus on what the media produced looks like, which is what actually causes harm to the victim, rather than focusing on what processes were used, which truly has no bearing on the victim.

B. Forbidding Disclaimer Defenses

In addition to determining what makes the NSII the NSII, state legislators should take a proactive approach to potential defenses. Merely omitting the reasonable person and authenticity requirements alone is not enough. An effective statute should not only omit a disclaimer defense, but it should also explicitly remove a defendant's ability to raise it. For example, Florida's statute clearly states a defendant is not relieved from criminal liability if they place a disclaimer on NSII that notifies viewers that the person depicted did not consent to, participate in, or perform the actions portrayed.[178]

Furthermore, permitting a disclaimer defense would undermine enforcement by providing a loophole that allows perpetrators to evade liability simply by affixing a label, regardless of the widespread circulation and consumption of the image. Disclaimers neither prevent the initial harm, nor do they mitigate the compounding harm as the image spreads. Allowing this defense would send a dangerous message that as long as a perpetrator acknowledges an image is false, they are absolved of responsibility. Because disclaimers are a post-hoc rationalization rather than a preventative measure, this defense would significantly weaken the deterrent effect of NSII statutes.

CONCLUSION

As technology continues to evolve and become more accessible, legislators must take thoughtful approaches in crafting effective definitions of criminal NSII that focus on the true harm—the exploitation and violation of victims. When lawmakers lose sight of this issue, they create narrow statutes that can be outpaced by technology. This Comment emphasizes the need to look at the output—the actual NSII itself—rather than the tools or processes used to create it. Statutes rooted in the harmful product, rather than technical distinctions, offer stronger, more sustainable definitions.

Legislators also cannot look at synthetic media in a vacuum. Perpetrators of NSII or political deepfakes operate with different motives. Statutory definitions must reflect these different contexts. While authenticity requirements may be appropriate for addressing the harm of disinformation in the election context, applying the same requirement to NSII undermines efforts to combat sexual exploitation, where the harm arises from the depiction itself, rather than its ability to fool.

178. FLA. STAT. § 836.13(6) (2025).

To address the full scope of harm caused by NSII, legislatures should prioritize criminal statutes that adopt an output-oriented statutory definition, focusing on the violation of personal autonomy and dignity inflicted by the NSII. These harms arise not from deception or the technology used to create the content but from how the perpetrator exploits and degrades their target. Only by centering this reality can the law keep pace with technology and ensure justice is not left behind.

Grace Mohlin[*]

* J.D. Candidate, 2026, Wake Forest University School of Law. I am profoundly grateful to everyone on the *Wake Forest Law Review* who helped edit this piece, especially Melissa Stuckey for her thorough edits and eye for detail, and Emily Mundt for her insightful comments and guidance throughout the publication process. Special thanks to my wonderful partner, whose data science expertise provided invaluable assistance in translating complex technical concepts into the framework of this analysis.