THE END IS STILL TO COME: HOW THE LAW'S INTERACTION WITH DIGITAL REPLICAS AND DERIVATIVE DIGITAL REPLICAS WILL SHAPE THE FUTURE

SPENCER KWESKIN*

"If a machine can think, it might think more intelligently than we do, and then where should we be?"[1]

ABSTRACT

Advancements in technology have led to the creation of nearly indistinguishable digital replicas of real individuals, and realistic generated fictional persons based on real individuals' images and likenesses. The status-quo right to privacy and right of publicity may not be sufficient to protect individuals from harm, nor deter nonconsensual uses of a person's identity in generative artificial intelligence environments. Statutory proposals to modify the status quo appear insufficient when weighed against the existential risk from unfriendly artificial superintelligence development, suggesting caution for slight status quo modifications and discouraging large status quo modifications.

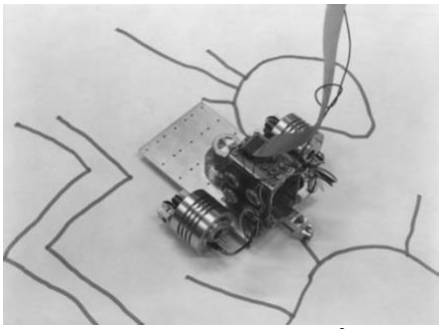
---

* J.D., Notre Dame Law School, 2025; B.S. in Economics and Political Science, Syracuse University, 2022. I would like to thank Dr. David Waddilove for advising me on this note, all my *Notre Dame Law School Journal of Law, Ethics, and Public Policy* peers, and my family.

1. Alan Turing, *Can Digital Computers Think?*, (BBC Radio broadcast May 15, 1951) (transcript available at https://turingarchive.kings.cam.ac.uk/publications-lectures-and-talks-amtb/amt-b-5 [https://perma.cc/BB4N-2PKF]).

INTRODUCTION

Over the past several decades, artificial intelligence ("AI") image processing has progressed from physical robots drawing with markers to software that turns words into images indistinguishable from human creation.



Harold Cohen's Turtle[2]

Jason Allen's Théâtre D'opéra Spatial[3]

Ashley Martin's The Forager[4]

2. Chris Garcia, *Harold Cohen and AARON–A Forty-Year Collaboration*, COMPUTER HIST. MUSEUM (Aug. 23, 2016), https://computerhistory.org/blog/harold-cohen-and-aaron-a-40-year-collaboration/ [https://perma.cc/E2WK-HCJJ].

3. Théâtre D'opéra Spatial won first place in the 2022 Colorado State Fair's Emerging Artist Digital Arts/Digitally Manipulated Photography category. The piece was generated with Midjourney, and "[b]oth competition judges . . . were not aware Allen's artwork was crafted with AI . . . ." Anna Lynn Winfrey, *'Someone Had to Be First': Pueblo Artist Criticized After AI Painting Wins at Colorado State Fair*, THE PUEBLO CHIEFTAIN, https://www.chieftain.com/story/news/2022/08/31/ai-painting-wins-at-colorado-state-fair-pueblo-artist-explains-jason-allen/65466872007/ [https://perma.cc/2SDN-KEWU] (last updated Sept. 1, 2022, 10:22 PM); *see also* Kevin Roose, *An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy*, N.Y. TIMES, (Sept. 2, 2022), https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html.

4. The Forager won second place in the 2022 Colorado State Fair's Emerging Artist Digital Arts/Digitally Manipulated Photography category. *See* Winfrey, *supra* note 3. The Forager was hand-made using "a Photoshop-type program to refine it." Matt Levin, *How Much "Art" Can There Be in Artificial Intelligence?*, MARKETPLACE (Oct. 24, 2022), https://www.marketplace.org/2022/10/24/how-much-art-can-there-be-in-artificial-intelligence/ [https://perma.cc/YJC2-9DW3].

Modern AI can create realistic videos of individuals with just one's name, image, and likeness ("NIL") found in actual pictures or videos. This one-for-one direct generation ("digital replica") can be made either from traditional AI or generative AI. Generative AI can also create realistic videos of fictional individuals *derived* from a large sample of real NIL data ("derivative digital replicas"). Nonconsensual uses of a person's NIL to create digital replicas and derivative digital replicas infringe on that person's property rights: the right to control one's own person (traditionally protected under the right to privacy) and the right to profit from one's own labor (traditionally protected under the right of publicity).

This note argues that the right to privacy and right of publicity can be insufficient in protecting an individual's NIL property rights from nonconsensual digital replicas and nonconsensual derivative digital replicas. It then explores and weighs the costs of plausible modifications to the status quo to sufficiently protect an individual's NIL. It asserts the existential risk from the development of an unfriendly artificial superintelligence strictly outweighs any other global good. It contends that Judeo-Christian ethics is the most likely way to develop a friendly artificial superintelligence and—as the United States ("US") is the world leader in AI development and is built on Judeo-Christian ethical principles—keeping AI development within the US is the best way to mitigate existential risk. Thus, implementation of status quo modifications addressing the harms digital replicas create ought to be cautious, and status quo modifications addressing the harms derivative digital replicas create ought not be implemented unless it is proven that existential risk is effectively mitigated.

I. THE DEVELOPMENT OF AI

Before discussing the law surrounding digital replicas and derivative digital replicas, it is important to understand what modern AI[5] is and how it functions. Modern AI represents a paradigm shift in how we understand both technology and intelligence, and future development here could pose an existential threat to humanity.

A. How AI Works

Development of AI began in the 1960s[6] with ELIZA, a chatbot "[m]odeled after the Rogerian style of psychotherapy . . . ."[7] How ELIZA works is simple: it scans for keywords in the user's input and rephrases those keywords into a question. ELIZA is very limited: Playing the role of a patient can create convincing conversation, but breaking the role or asking complex questions quickly confuses ELIZA.[8] Basic programs like ELIZA are simple functions built by humans following a laid-out sequence of operations designed by the programmer themselves to give an output.

Modern AI instead uses *deep learning*.[9] Here, "learning" is an apt description. While there is a debate in cognitive science on whether humans are born with innate abilities or if abilities are nurtured,[10] for most skills "the ability to solve problems is strongly affected by the amount and organization of people's knowledge about an area (for example[:] chess, biology, sports, computers, physics, [and] card games such as bridge)."[11]

To become better at solving problems, it is necessary to learn about an area by gathering information ("data") from practice. During deliberate learning, the brain takes in data, makes assumptions, figures out what works and

5. In this note, "modern AI" refers to machine learning algorithms.

6. While the first chatbots were developed in the 1960s, Alan Turing was the first to propose AI and early developments broadly followed his research. *See generally* B. J. Copeland, *Alan Turing and the Beginning of AI*, BRITANNICA, https://www.britannica.com/science/history-of-artificial-intelligence (last visited Feb. 4, 2025).

7. Oshan Jarow, *How the First Chatbot Predicted the Dangers of AI More Than 50 years Ago*, VOX (Mar. 5, 2023), https://www.vox.com/future-perfect/23617185/ai-chatbots-eliza-chatgpt-bing-sydney-artificial-intelligence-history.

8. Try exploring NJIT's JavaScript recreation of ELIZA. *See generally* Michael Wallace, *ELIZA: A Very Basic Rogerian Psychotherapist Chatbot*, NJIT (2016), https://web.njit.edu/~ronkowit/eliza.html [https://perma.cc/85TC-4FEH].

9. Deep learning and machine learning are technically different. Explaining the differences is beyond the scope of this note.

10. *Compare* Ariel Starr et al., *Number Sense in Infancy Predicts Mathematical Abilities in Childhood*, 110 PNAS 18116 (2013) *with* David Z. Hambrick et al., *Is Innate Talent a Myth?*, SCI. AM. (Sep. 20, 2016), https://www.scientificamerican.com/article/is-innate-talent-a-myth/ [https://perma.cc/78YX-8S37].

11. JOHN D. BRANSFORD & BARRY S. STEIN, THE IDEAL PROBLEM SOLVER 4–5 (2d ed. 1993).

what does not, modifies its own physical structure, and improves iteratively until obtaining expert-level performance.[12]

Similarly, an AI model takes in data, makes assumptions, figures out what works and what does not, modifies itself,[13] and improves iteratively until obtaining a desired level of performance.[14] Once trained, the AI model is a black box—not even those who created the model can understand or comprehend *how exactly* it comes to a decision.[15]

B. Comparing Traditional AI and Generative AI

Considering how AI works in abstract is useful to understanding what AI does in practice. Modern AI is either traditional (nongenerative) or generative. Traditional AI performs specific tasks using deep learning methods, whereas generative AI uses deep learning to create something new.[16] Compare DeepFaceLab with OpenAI's Sora: DeepFaceLab uses traditional AI to swap faces in video (making a "deepfake"), whereas Sora uses generative AI to create new videos based on the user's text input.

1. Traditional AI: DeepFaceLab

Deepfakes are created by machine learning AI programs that merge two videos together: the *source* and the *destination*. For an example using

12. Sophie H. Bennett et al., *Rewiring the Connectome: Evidence and Effects*, 88 NEUROSCIENCE & BIOBEHAVIORAL 88 REVS. 51, 51–53, 55–56 (2018). Research shows rewiring the brain contributes to and plays a central role in learning. Neuroscience has yet to prove rewiring is a necessary and sufficient part of learning due to testing difficulty. *See id*. at 58.

13. *See generally* Arthur Arnx, *First Neural Network for Beginners Explained (with Code)*, TOWARDS DATA SCI. (Jan. 13, 2019), https://towardsdatascience.com/first-neural-network-for-beginners-explained-with-code-4cfd37e06eaf [https://perma.cc/8FKW-4TMQ].

14. *See generally* Julianna Delua, *Supervised Versus Unsupervised Learning: What's the Difference?*, IBM (Mar. 12, 2021), https://www.ibm.com/blog/supervised-vs-unsupervised-learning/ [https://perma.cc/6J35-8WGB].

15. *See* Chloe Xiang, *Scientists Increasingly Can't Explain How AI Works*, VICE (Nov. 1, 2022, 9:00 AM), https://www.vice.com/en/article/y3pezm/scientists-increasingly-cant-explain-how-ai-works [https://perma.cc/S4GE-8XLJ]; *see also* ROMAN V. YAMPOLSKIY, UNEXPLAINABILITY AND INCOMPREHENSIBILITY OF ARTIFICIAL INTELLIGENCE 1 (2019) (demonstrating "advanced AIs would not be able to accurately explain some of their decisions and for the decisions they could explain people would not understand some of those explanations.").

16. *See generally* Bernard Marr, *The Difference Between Generative AI and Traditional AI: An Easy Explanation for Anyone*, FORBES (July 24, 2023), https://www.forbes.com/sites/bernardmarr/2023/07/24/the-difference-between-generative-ai-and-traditional-ai-an-easy-explanation-for-anyone/?sh=2ef6d3bb508a.

DeepFaceLab,[17] let the source be Tony Stark's (Robert Downey Jr.'s)[18] Stark Expo Speech from *Iron Man 2*[19] and the destination be Elon Musk's appearance[20] at South by Southwest's ("SXSW") *Westworld* Session,[21] so that the output is Robert Downey Jr.'s face is "on" Elon Musk's SXSW body.

Source: Iron Man 2[22]

Destination: Elon Musk SXSW Speech[23]

Desired Output: Robert Downey Jr. as Elon Musk at SXSW[24]

DeepFaceLab breaks-up the source and destination frame-by-frame. DeepFaceLab then aligns the faces in the source and destination such that each

17. I use DeepFaceLab to make a deepfake based on the sample videos provided with the program. @iperov, *DeepFaceLab*, GITHUB (Apr. 9, 2020), https://github.com/iperov/DeepFaceLab [https://perma.cc/AAP6-CNXR]; *see generally infra* note 230.

18. Robert Downey Jr. played Tony Stark in Marvel's films. Amy Ryan, *Robert Downey Jr. is Iron Man,* ENT. WKLY. (Sep. 29, 2006, 5:35 PM), https://ew.com/article/2006/09/29/robert_downey_j/ [https://web.archive.org/web/20180417122941/https://ew.com/article/2006/09/29/robert_downey_j/].

19. IRON MAN 2 (Marvel Studios 2010); *see also* @bestscene9405, *Tony Stark Expo Speech—Stan Lee Cameo Scene—Iron Man 2 (2010) HD*, YOUTUBE at 01:04–33 (Jul. 6, 2018), https://www.youtube.com/watch?v=TZA7llltH8GQ [https://perma.cc/7Q9Q-U7DV].

20. Elon Musk, Interview at Westworld Featured Session (2018) *in* @SXSW, *Cast and Crew of Westworld | Westworld | SXSW 2018*, YOUTUBE at 34:36–35:31 (May 30, 2018), https://www.youtube.com/watch?v=OpaHZ5HyP44&t=1882s [https://perma.cc/Z76A-YP4K].

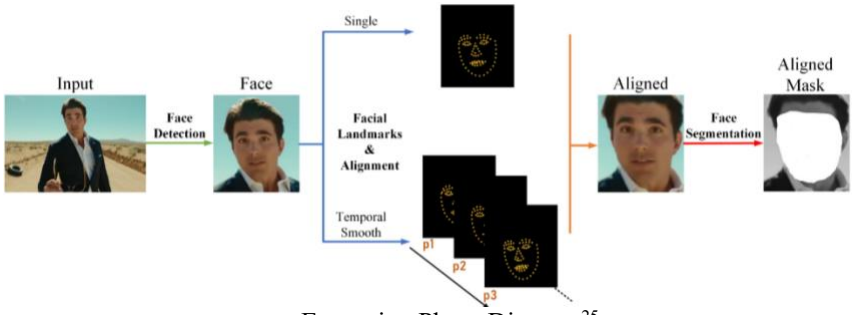
21. *See generally* Neha Aziz, *Westworld Cast and Showrunners Jonathan Nolan & Lisa Joy at SXSW 2018 [Video]*, SXSW (July 19, 2018), https://www.sxsw.com/film/2018/westworld-showrunners-jonathan-nolan-lisa-joy-with-cast-at-sxsw-2018-video/ [https://perma.cc/5VGN-6VMR].

22. *See* IRON MAN 2, *supra* note 19.

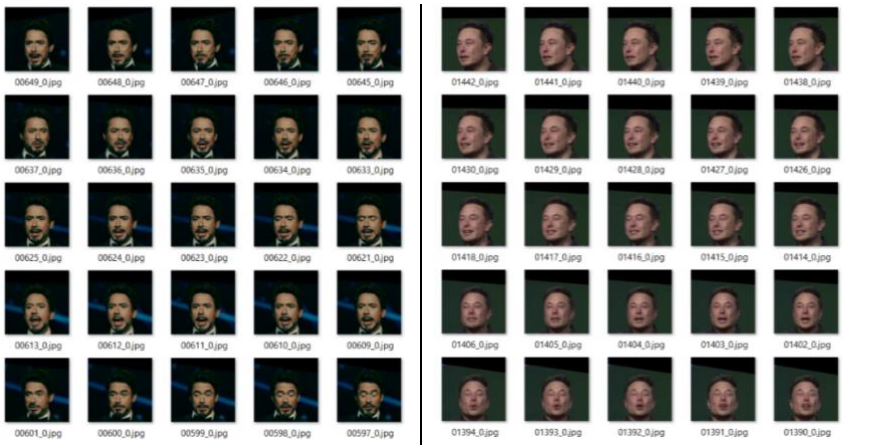
23. *See* Musk, *supra* note 20.

24. *See* @spencerkweskin8303, *infra* note 33.

frame's face is in roughly the same place. This alignment prepares faces for extraction and eases the burden on DeepFaceLab by making it easier for the AI to discern where the faces are in the source and destination.



Extraction Phase Diagram[25]



Extracted and Aligned Source Faces

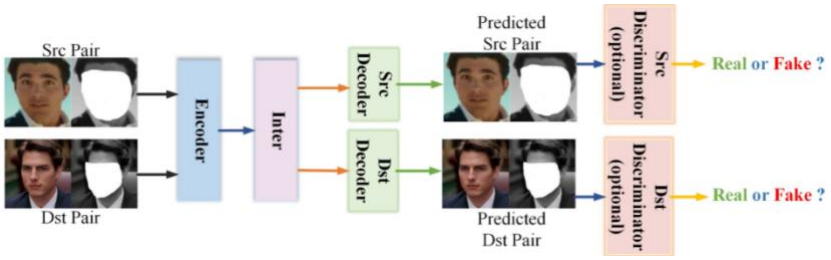
Extracted and Aligned Destination Faces

Once faces are aligned, the model starts training. DeepFaceLab uses a "standard facial landmark template[]"[26] and assigns a point to each part of each face. The encoder and interlacer learn what the source and destination faces look like using a shared template. The source decoder then attempts to recreate the source face by predicting what the face looks like, and vice-versa for the destination decoder. The source and destination discriminator checks the

25. IVAN PETROV ET AL., DEEPFACELAB 2 (Mar. 14, 2024) https://ar5iv.labs.arxiv.org/html/2005.05535 [https://perma.cc/ML7J-YLXD].

26. *Id*. at 3.1.

outputs. The decoders and discriminators compete "to produce a model capable of generating new data with the same properties of the training set."[27]



Training Phase Diagram[28]

Here, I use DeepFaceLab's SAEHD algorithm and set the model's properties:

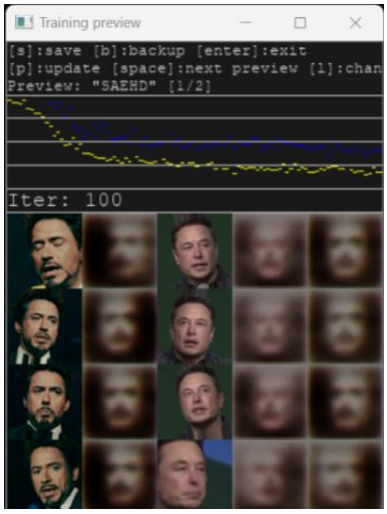
```
================== Model Summary ===================
==                                                ==
==            Model name: NoteModel_SAEHD         ==
==                                                ==
==     Current iteration: 0                       ==
==                                                ==
==---------------- Model Options ----------------==
==                                                ==
==            resolution: 64                      ==
==             face_type: wf                      ==
==     models_opt_on_gpu: True                    ==
==                 archi: df-ud                   ==
==               ae_dims: 32                      ==
==                e_dims: 16                      ==
==                d_dims: 16                      ==
==           d_mask_dims: 16                      ==
==       masked_training: True                    ==
==       eyes_mouth_prio: True                    ==
==           uniform_yaw: False                   ==
==         blur_out_mask: False                   ==
==             adabelief: True                    ==
==            lr_dropout: n                       ==
==           random_warp: True                    ==
==      random_hsv_power: 0.05                    ==
==       true_face_power: 0.01                    ==
==      face_style_power: 2.0                     ==
==        bg_style_power: 2.0                     ==
==               ct_mode: none                    ==
==             clipgrad: False                    ==
==             pretrain: False                    ==
==       autobackup_hour: 0                       ==
== write_preview_history: True                    ==
==           target_iter: 0                       ==
==       random_src_flip: False                   ==
==       random_dst_flip: True                    ==
==            batch_size: 10                      ==
==             gan_power: 0.01                    ==
==        gan_patch_size: 8                       ==
==              gan_dims: 16                      ==
==                                                ==
==------------------ Running On ------------------==
==                                                ==
==          Device index: 0                       ==
==                  Name: NVIDIA GeForce RTX 2070 ==
==                  VRAM: 6.55GB                  ==
==                                                ==
====================================================
```

Model Settings

27. Chanon Roy, *Creating Realistic Deepfakes With DeepFaceLab*, GEEK CULTURE (Apr. 18, 2021), https://medium.com/geekculture/realistic-deepfakes-with-deepfacelab-530e90bd29f2 [https://perma.cc/ZK9X-FUZ4].

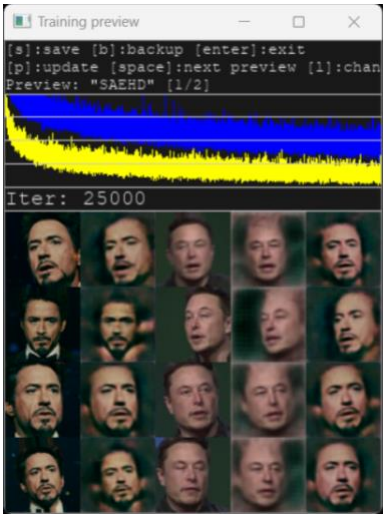
28. PETROV ET AL., *supra* note 25, at 3.1.

Over time, the model improved iteratively.[29]



100 Iterations

1,000 Iterations



25,000 Iterations

100,000 Iterations

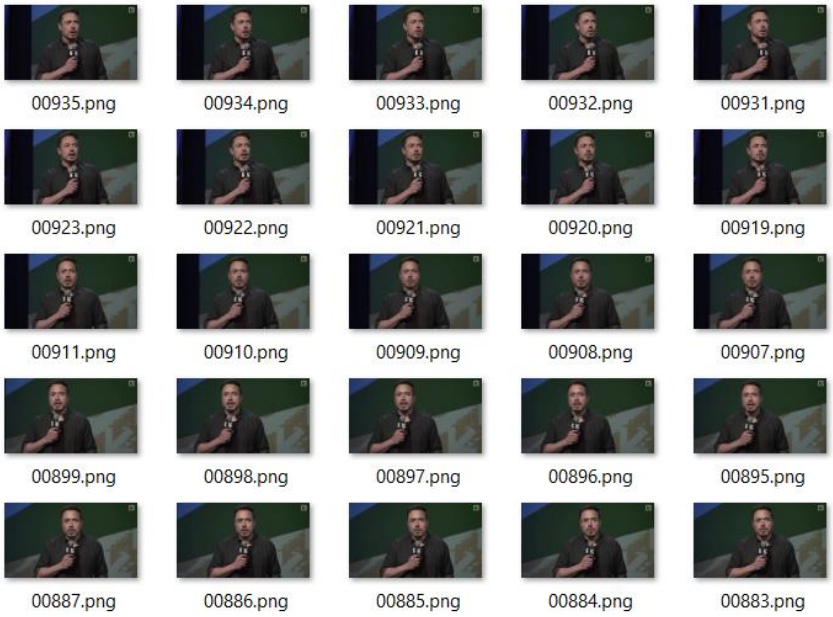
After the model trains to 100,000 iterations,[30] DeepFaceLab uses the trained encoder, interlacer, source decoder, and destination decoder together to place the source face on the destination body. The encoder and interlacer create

29. The columns in each of the following, from left to right, are: (1) a source frame; (2) a source decoder recreation; (3) a destination frame; (4) a destination decoder recreation; and (5) a sample output. DeepFaceLab displays four preview rows during training.

30. The model was arbitrarily stopped at 100,000 iterations.

a source face and a destination mask. The destination decoder then generates the result face. The result is then aligned, blended, and sharpened to produce the desired output. Then, the result frames are converted back into video where the original destination's audio is paired.

Conversion Phase Diagram[31]



Result Frames

31. PETROV ET AL., *supra* note 25, at 3.2.

This gives a final output:

Original Elon Musk SXSW Speech[32]

Final Output: Robert Downey Jr. as Elon Musk at SXSW[33]

DeepFaceLab can provide an extremely convincing face swap with minimal effort when viewed holistically. However, DeepFaceLab has some disadvantages.

DeepFaceLab Weakness Example

Some faces are inaccurate to facial structure, especially when the destination subject makes extreme movements—like here with Elon Musk's large, eyebrow-raising expression. Facial anatomy can become noticeably blurred when DeepFaceLab encounters other unique expressions, and DeepFaceLab can have a difficult time interpreting realistic rotations.

32. *See* Musk, *supra* note 20.

33. @spencerkweskin8303, *DeepFaceLab Example | The End is Still to Come*, YOUTUBE (Aug. 4, 2025), https://www.youtube.com/watch?v=ZEmqmas-C8o [https://perma.cc/NU2H-GBEX]. *Compare id. with* @Deepfakery, *DeepFaceLab 2.0 Quick96 Deepfake Video Example*, YOUTUBE (July 8, 2020), https://www.youtube.com/watch?v=lnUbEPFlgKA [https://perma.cc/F2D8-HCSF].

2. Generative AI: Sora

While DeepFaceLab is impressive, it requires the user to provide *a* source and *a* destination. The next improvement is generative AI, and—as of publication—one of the most advanced video generative AI tools is OpenAI's Sora, which can generate up to a minute of HD video *from text alone*:

OpenAI's Sora: Generative Text to Video Examples[34]

Sora is "given input noisy patches (and conditioning information like text prompts) [and is] trained to predict the original 'clean' patches."[35] Once Sora is trained, Sora can take text input and produce a video output.

Like DeepFaceLab, Sora has limitations. OpenAI states that Sora can "struggle to simulate the physics of a complex scene, [] may not comprehend specific instances of cause and effect[,] . . . may also confuse spatial details

34. *Creating Video from Text*, OPENAI, https://openai.com/index/sora (last visited Feb. 4, 2025). From left to right, the prompts are: (1) "A cat waking up its sleeping owner demanding breakfast. The owner tries to ignore the cat, but the cat tries new tactics and finally the owner pulls out a secret stash of treats from under the pillow to hold the cat off a little longer[]"; (2) "Reflections in the window of a train traveling through the Tokyo suburbs[]"; (3) "A beautiful homemade video showing the people of Lagos, Nigeria in the year 2056. Shot with a mobile phone camera[]"; and (4) "A movie trailer featuring the adventures of the 30 year old [sic] space man wearing a red wool knitted motorcycle helmet, blue sky, salt desert, cinematic style, shot on 35mm film, vivid colors." *Id*. Since writing this note, other generative video AI tools have been published. *See, e.g.*, *Veo*, GOOGLE DEEPMIND, https://deepmind.google/models/veo/ (Veo 3 allows for both video and audio generation simultaneously). I still use Sora as the example as OpenAI's AI tools are still the most visited. *See generally* Sujan Sarkar, *AI Industry Analysis: 50 Most Visited AI Tools and Their 24B+ Traffic Behavior*, WRITERBUDDY, https://writerbuddy.ai/blog/ai-industry-analysis (last visited Aug. 4, 2025).

35. *Video Generation Models as World Simulators*, OPENAI (Feb. 15, 2024), https://openai.com/index/video-generation-models-as-world-simulators/.

included in a prompt, . . . [and may] struggle with precise descriptions of events that unfold over time . . . ."[36]

Here, Sora encounters issues with continuity:

OpenAI's Sora Weakness Example[37]

Sora appears to confuse where the wolf pups are in the scene. Sora resolves the confusion surrounding the wolf pups' coats and legs by creating new wolf pups from thin air.[38] Nevertheless, AI technology is impressive *right now*, even if there are a few minor hiccups in the AI's output.

C. Development of AI Technology in the Near-Future

AI's minor hiccups are unlikely to last for long: take OpenAI's GPT (Generative Pretrained Transformer) for example. GPT is OpenAI's model which powers Chat-GPT. GPT takes input from text and images to determine the most likely output.[39] The initial model was introduced in 2018.[40] In only four years, OpenAI launched ChatGPT based on GPT-3, which began outputting "everything from travel planning to [] fables to code [for] computer programs."[41] A month later, OpenAI improved ChatGPT with GPT-3.5, which

36. *Creating Video from Text*, *supra* note 34.

37. *Id*. The prompt is "Five gray wolf pups frolicking and chasing each other around a remote gravel road, surrounded by grass. The pups run and leap, chasing each other, and nipping at each other, playing." *Id*.

38. OpenAI uses this video as an example for one of Sora's weaknesses: that "[a]nimals or people can spontaneously appear, especially in scenes containing many entities." *Id*.

39. *See generally How ChatGPT and Our Language Models Are Developed*, OPENAI, https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-language-models-are-developed (last updated June 5, 2025).

40. *See generally* Alec Radford et al., *Improving Language Understanding by Generative Pre-Training*, OPENAI (2018).

41. *See* Bernard Marr, *A Short History of ChatGPT: How We Got to Where We Are Today*, FORBES (May 19, 2023, 1:14 AM), https://www.forbes.com/sites/bernardmarr/2023/05/19/a-short-history-of-chatgpt-how-we-got-to-where-we-are-today/?sh=1d383849674f [https://perma.cc/9EZ8-NJML].

was "more sophisticated and thorough in its responses than GPT-3."[42] OpenAI introduced GPT-4 only months later.[43]

GPT-3.5 (ChatGPT) performed in the 10th percentile on the Uniform Bar Exam, received a 149 on the LSAT, and scored a 1260 on the SAT (670 Reading & Writing and 590 Math).[44] GPT-4 meanwhile performed in the 90th percentile on the Uniform Bar Exam, received a 163 on the LSAT, and scored a 1410 on the SAT (710 Reading & Writing and 700 Math).[45] Future versions of OpenAI's model are likely to improve even further, becoming more accurate with each iteration. However, it is unclear whether this growth will continue or if there is some limitation in AI technological development that will slow technological growth to a halt.

II. AI TECHNOLOGY AND EXISTENTIAL RISK

If technological growth does not slow, then it is possible a computer will be created which is as smart as a human in every capacity. If this occurs, then human-intelligence-level AI would likely be capable of either designing another AI to be slightly smarter than itself or making itself slightly smarter; both of which would likely create a feedback loop leading to an artificial superintelligence ("ASI").[46] If current AI development leads to an ASI, then humanity may be in great danger. A hypothetical ASI *without* a proper ethical framework poses existential risk. However, a hypothetical ASI *with* a proper ethical framework can lead to the end of scarcity.

A. A Scenario with the Exponential Growth of AI Technology: The Singularity—Either the End of Scarcity or the End of Humanity

1. The Singularity

Computer scientist and leading author on the Singularity Ray Kurzweil describes the phenomenon of rapidly and ever-increasing advancements as The Law of Accelerating Returns:

> Evolution applies positive feedback in that the more capable methods resulting from one stage of evolutionary progress are used to create the next stage. . . . As a result, the rate of progress of an evolutionary process increases exponentially over time. Over time, . . . the "returns" of an evolutionary process ([e.g.,] the speed, cost-effectiveness, or overall "power" of a process) increase exponentially . . . . This results in a second level of exponential

42. Kyle Wiggers, *While Anticipation Builds for GPT-4, OpenAI Quietly Releases GPT-3.5*, TECHCRUNCH (Dec. 1, 2022), https://techcrunch.com/2022/12/01/while-anticipation-builds-for-gpt-4-openai-quietly-releases-gpt-3-5/.

43. *See* Amit Prakash, *GPT-4 Early Impressions and How It Compares to GPT-3.5*, THOUGHTSPOT (Mar. 24, 2023), https://www.thoughtspot.com/data-trends/ai/gpt-4-vs-gpt-3-5.

44. OPENAI, GPT-4 TECHNICAL REPORT 5 (2023).

45. *Id.*

46. ASI refers to the *post-Singularity AI*, whereas AI refers to AI *in general*.

growth (i.e., the rate of exponential growth itself grows exponentially).[47]

To demonstrate exponential growth, consider the "Wheat on a Chessboard" problem. The legendary inventor of chess Grand Vizier Sissa ben Dahir asked Indian King Shirham for "a grain of wheat to [be] place[d] on the first square of the chessboard, and two grains . . . on the second square, and four grains . . . on the third, . . . and so on for the sixty-four [chessboard] squares [as a reward for inventing chess]."[48] King Shirham granted Sissa's request because Sissa's request sounded like a small amount of wheat. In reality, Sissa would receive "$1 + 2 + 2^2 + \cdots + 2^{63} = 2^{64} - 1$, which is a walloping 18,446,744,073,709,551,615 grains of wheat."[49]

Wheat on a Chessboard Graph

Kurzweil argues that the progress of technological growth will follow a similar exponential growth curve and eventually reach a point known as the Singularity. The Singularity "represents the nearly vertical phase of exponential growth that occurs when the rate [of growth] is so extreme that technology appears to be expanding at infinite speed[,] . . . [so] from our *currently* limited [perspective], this imminent event appears to be an acute and abrupt break in the continuity of progress."[50] The overwhelming majority of AI researchers currently believe that this is the most probable future because "it is only a matter of time before machines surpass [human intelligence] unless there's some hard limit to [machine] intelligence [yet to be discovered]."[51]

47. Ray Kurzweil, *Kurzweil's Law (AKA "The Law of Accelerating Returns")*, THE KURZWEIL LIBR. (Jan. 12, 2004), https://www.thekurzweillibrary.com/kurzweils-law-aka-the-law-of-accelerating-returns [https://perma.cc/CA8C-3S4T].

48. CLIFFORD A. PICKOVER, THE MATH BOOK 102 (2009).

49. *Id.*

50. RAY KURZWEIL, THE SINGULARITY IS NEAR 24 (2005) (emphasis added).

51. *See* Cem Dilmegani, *When Will the Singularity Happen? 1700 Expert Opinions of AGI [2024]*, AIMULTIPLE RSCH., https://research.aimultiple.com/artificial-general-intelligence-singularity-timing/ [https://perma.cc/FFE4-4Q7J] (last updated Sep. 24, 2024) ("90% of [AI researchers] think that [AI as smart as a human] is likely to happen by 2075.").

A vast majority of AI experts believe that an ASI's development is inevitable. The rewards for developing a[52] *friendly*[53] ASI are incredibly vast,[54] as the theoretical best-case scenario is that a friendly ASI could lead to the end of scarcity.[55] However, if an *unfriendly*[56] ASI is developed, the theoretical worst-case outcome would be *the end of humanity*.[57]

a. Existential Risk

"The arguments for why ASI could . . . destroy humanity, which are primarily philosophical, are complicated, with many moving parts. But the central conclusion is the *unintended consequences* of the ASI striving to achieve its final goals."[58] To understand why this is so important, it is critical to understand what existential risk *is*.

"[A] *risk* is standardly defined as the probability of an event multiplied by its consequences."[59] To "roughly characterize a risk's seriousness [we can use] three variables: *scope* (the size of the population at risk), *severity* (how badly th[e] population would be affected), and *probability* (how likely the disaster is to occur, according to the most reasonable judgement, given currently available evidence)."[60]

---

52. If Kurzweil's Singularity thesis is correct, then the first ASI will likely be the *only* ASI. *See* NICK BOSTROM, SUPERINTELLIGENCE 78–104 (2014). Therefore, I refer to ASI as artificial *super-intelligence* (singular), not artificial *super-intelligences* (plural).

53. *See* Eliezer S. Yudkowsky, *What is Friendly AI?*, THE KURZWEIL LIBRARY (May 3, 2001), https://www.thekurzweillibrary.com/what-is-friendly-ai [https://perma.cc/8Q4N-FSPK] ("A 'Friendly AI' is an AI that takes actions that are, on the whole, beneficial to humans and humanity . . . ."). This definition of "friendly" considers the results to humanity, not the AI's intentions of acting, therefore it is nonanthropomorphic. *See infra* Section II.A.1.b.

54. *See generally* BOSTROM, *supra* note 52, at 145–58.

55. *See e.g.*, Matt McDonagh, *The End of Scarcity?*, LIFE IN THE SINGULARITY (Nov. 4, 2023), https://lifeinthesingularity.com/p/the-end-of-scarcity [https://perma.cc/JQ2R-HJN4]; Oshan Jarow, *A Utopian Strand of Economic Thought Is Making a Surprising Comeback*, VOX, https://www.vox.com/the-highlight/24065976/economists-post-scarcity-abundance-industrial-revolution-socialism-capitalism [https://perma.cc/R6XK-TW72] (last updated Mar. 11, 2024, 9:19 AM).

56. *Contra supra* note 53.

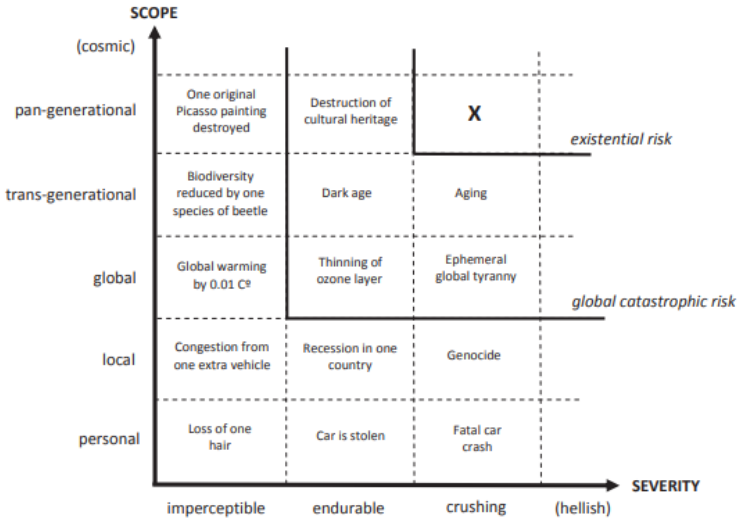
57. *See infra* Section II.A.1.a.

58. Émile P. Torres, *Would "Artificial Superintelligence" Lead to the End of Life on Earth? It's Not a Stupid Question*, SALON (Aug. 6, 2022, 12:00 PM), https://www.salon.com/2022/08/06/would-artificial-superintelligence-lead-to-the-end-of-life-on-earth-its-not-a-stupid-question/ [https://perma.cc/5RMF-BVBH]; *cf.* Fred L. Casmir & Kathryn J.S. Kweskin, *Theoretical Foundations for the Evolution and Testing of a Chaos Theory of Communicating*, 57 WORLD FUTURES 339, 360 (2001) ("prediction of endstates is time poorly spent, yields few additions to the knowledge building process, and even serves to misguide those interested in controlling future outcomes.").

59. Torres, *supra* note 58; *see also* Sven Ove Hansson, *Risk*, STAN. ENCYC. PHIL., https://plato.stanford.edu/entries/risk/#DefRis (last updated Dec. 8, 2022).

60. Nick Bostrom, *Existential Risk Prevention as Global Priority*, 4 GLOB. POL'Y 15, 16 (2013).

Risks then can be broadly characterized and placed into the following categories:



Risk Categories[61]

Even if the *probability* of a risk is very small, if the *scope* and *severity* of the risk is sufficiently large, then the actual risk is very serious (moving towards the upper-right region of the above chart). Existential risks are among the most serious,

> [b]ut just *how much* more serious might not be intuitively obvious. One might think we could get a grip on how bad an existential catastrophe would be by considering some of the worst historical disasters we can think of—such as the two world wars, the Spanish flu pandemic, or the Holocaust—and then imagining something just a bit worse. Yet if we look at global population statistics over time, we find that these horrible events of the past century fail to register . . . . But even this reflection fails to bring out the seriousness of existential risk. What makes existential catastrophes especially bad . . . . [is that] their significance lies primarily in the fact that they would destroy the future.[62]

As existential risk destroys the future, "existential risk reduction is strictly [and morally] more important than *any other global public good*."[63]

61. *Id.* at 17.
62. *Id.*
63. *Id.* at 15 (emphasis added).

The theoretical *severity* and *scope* of an *unfriendly*[64] ASI would be "the extinction of *our* species . . . [or the] obliterat[ion of] the entire biosphere."[65] An ASI "may shape the future of Earth-originating life, could easily have non-[human-centric] final goals, and would likely have instrumental reasons to pursue open-ended resource acquisition [without limiting itself as to not infringe on human interests]."[66] "An ASI would be the most powerful technology ever created, and for this reason we should expect its potential unintended consequences to be even more disruptive than those of past technologies."[67] There is an obvious obligation to develop AI ethically to sufficiently minimize or effectively eliminate the existential risk unfriendly ASI poses.

b. The Artificial Nature of AI

Before discussing the obligation to develop AI ethically, it is critical to understand that an AI is *artificial*. This seems incredibly obvious, but AI popular culture[68] fixates around *anthropomorphism* (applying human characteristics to nonhuman things).

Anthropomorphism is a major issue because "[a]nthropomorphic frames encourage unfounded expectations about the growth trajectory of [] AI and about the psychology, motivations, and capabilities of a mature superintelligence."[69] Consider the following example by researcher Eliezer Yudkowsky:

> Imagine, for a moment, that you walk up and punch an AI in the nose. Does the AI punch back [like a human]? . . . In a punched human, blood races, adrenaline pumps, the hands form fists, the stance changes, all without conscious attention. . . . If the AI is young, but old enough to be explaining human actions by reference to human cognition, . . . the AI's next goal [after being punched] becomes deducing the human's real motive . . . . There are [] three different ways the AI can try to prevent the next punch; by preventing the physical collision itself, by preventing the human from deciding to punch the AI, or by preventing the unknown original motive from recurring. . . . Punching *back* [(i.e., retaliation)] is an incredibly nonobvious solution.[70]

Yudkowsky's assertion that retaliation is an anthropomorphic solution to getting punched also applies to other popular basal reasons for actions such as

64. *See infra* Section II.A.1.b.

65. Torres, *supra* note 58.

66. BOSTROM, *supra* note 52, at 116.

67. Torres, *supra* note 58.

68. *See, e.g.*, HER (Warner Brothers Pictures 2013); 2001: A SPACE ODYSSEY (Metro-Goldwyn-Mayer 1968); WESTWORLD (HBO Entertainment television broadcast Oct. 2, 2016); THE TERMINATOR (Orion Pictures 1984); WALL·E (Pixar Animation Studios 2008).

69. BOSTROM, *supra* note 52, at 92.

70. ELIEZER YUDKOWSKY, CREATING FRIENDLY AI 1.0: THE ANALYSIS AND DESIGN OF BENEVOLENT GOAL ARCHITECTURES 25–26 (2001) (footnote omitted).

selfishness, avoiding pain, and obtaining pleasure.[71] Anthropomorphism is especially problematic when we "make[] artificial intelligences, robots, and machines more human . . . even as it is still acknowledged that [AI] is not human."[72]

Anthropomorphism makes discussing the ethical framework surrounding AI challenging. As humans, we tend to think about things in *human terms*—foxes are 'sly,' lions are 'proud,' *et cetera*[73]—whereas AI is *tabula rasa,* the quintessential blank slate (i.e., AI has no inherit ethics). Despite this, AI "can actuate evil against mankind due to its programming originally coded and designed by humans."[74]

There are two possible versions of the Singularity future: "[(1)] ASI is not beholden to any ethical claims from its creator, mankind, or . . . [(2)] ASI is beholden to the ethics of mankind, its creator."[75] If ASI is not beholden to the ethics of mankind, then "it is acceptable for ASI to be unbound from moral and ethical restraints that place mankind above the priorities of the ASI."[76] Consequently, AI development should cease all together (i.e., relinquishment). However, as I explain in Section III.B.3.a, relinquishment is incredibly naïve.[77] Fortunately, this version of the future is not as likely as intuition or Hollywood[78] suggests. As initiators, humans "have the freedom to establish initial conditions [and] make things happen in ways that are less [harmful] than others."[79]

If ASI is beholden to the ethics of mankind, "any ethics created by man and subsequently programmed into ASI must supervene on all man's anthropological reality. If it does not . . . the ASI will proceed on its own ethical imperatives, . . . not on what mankind deems ethical."[80] It is imperative that an ASI's ethical framework be nonanthropomorphic and protective of humanity.

Assuming that ASI can be programmed with an ethical framework,[81] the question becomes: What ought that framework to be?

---

71. *See generally id.* at 34–41, 43 ("*The entire concept of a goal system that centers around the observer* is fundamentally anthropomorphic.").

72. PAUL GOLATA, THE ETHICS OF SUPERINTELLIGENT DESIGN 121 (2018).

73. *See generally* Sarah Barkley, *Why Do We Anthropomorphize?*, PSYCHCENTRAL, https://psychcentral.com/health/why-do-we-anthropomorphize#anthropomorphism [https://perma.cc/2U8D-F75J] (last updated Sep. 14, 2022).

74. GOLATA, *supra* note 72, at 126.

75. *Id.* at 125.

76. *Id.*

77. *See infra* Section III.B.3.a; *see generally* Vernor Vinge, *The Coming Technological Singularity: How to Survive in the Post-Human Era*, *in* NASA LEWIS RES. CTR. & OHIO AEROSPACE INST., VISION-21 16 (1993).

78. *See, e.g.*, THE MATRIX (Warner Brothers 1999).

79. Vinge, *supra* note 77, at 16.

80. GOLATA, *supra* note 72, at 126.

81. This note assumes that the control problem can be solved, and an ASI can acquire values. *See generally* BOSTROM, *supra* note 52, at 127–44, 185–208.

c. The Role of Judeo-Christian Ethics

Judeo-Christian ethics ought to be that framework because Judeo-Christian ethics meets the minimum conditions of being nonanthropomorphic and protective of humanity, unlike other ethical frameworks.[82] Therefore, Judeo-Christian ethics ought to form the basis of any common law or regulatory developments surrounding AI.

A large percentage of the population is likely to immediately balk at the mere assertion that Judeo-Christian ethics ought to guide *any* technological development due to the assumption that technology and religion are incompatible.[83] "[W]ithin this belief there lies the misperception that religion and science form a dichotomy rather than a duality: there is scientific truth [guided by knowledge and intellect] and there is spiritual truth [guided by faith]."[84]

However, "technology [itself] is a branch of moral philosophy, not of science. It aims at prudent goods for the common[good], to provide efficient means for th[o]se goods."[85] As AI takes on decision-making functions (including how to interpret text to give video output), the industry must consider ethical frameworks for AI development in order to avoid negative outcomes.[86] "In no other field is the ethical compass [] more relevant than in artificial intelligence[,]"[87] especially leading up to the Singularity. Here, "[e]ven if one does not accept [Judaism or] Christianity, [Judeo-]Christian ethics can properly account for and handle significant and relevant ethical issues that are presented [with AI, and therefore] . . . a [Judeo-]Christian ethic can be applied in all cases of ASI and deliver virtuous results."[88]

---

82. *See infra*, Section II.A.1.c.iii.

83. *See* Cary Funk & Becka A. Alper, *Highly Religious Americans Are Less Likely than Others to See Conflict Between Faith and Science.*, RELIGION AND SCI. 12–13 (2015); *see also* Sean M. Carroll, *Why (Almost All) Cosmologists Are Atheists*, 22 FAITH & PHIL. 622 (2005). In some cases, the things individuals think and say reflect a fuzzier boundary that the base assumption that technology and religion are incompatible fails to reflect. *See generally* Emma Greguska, *The Relationship Between Religion, Science and Technology*, ASU NEWS (Sep. 24, 2019), https://news.asu.edu/20190924-discoveries-asu-center-receives-17m-grant-explore-relationship-between-religion-science [https://perma.cc/6QWP-KKEM].

84. GERALD L. SCHROEDER, THE SCIENCE OF GOD: THE CONVERGENCE OF SCIENTIFIC AND BIBLICAL WISDOM 2 (2009).

85. PAUL GOODMAN, NEW REFORMATION: NOTES OF A NEOLITHIC CONSERVATIVE 40 (2010).

86. *See generally* Christina Pazzanese, *Great Promise But Potential for Peril*, THE HARV. GAZETTE (Oct. 26, 2020), https://news.harvard.edu/gazette/story/2020/10/ethical-concerns-mount-as-ai-takes-bigger-decision-making-role/ [https://perma.cc/7FE8-Z8BB].

87. Gabriela Ramos, *Message from Gabriela Ramos*, *in* UNESCO, KEY FACTS: UNESCO'S RECOMMENDATION ON THE ETHICS OF ARTIFICIAL INTELLIGENCE 4 (2023).

88. GOLATA, *supra* note 72, at 8.

Here, Judeo-Christian ethics assumes "assumes *a priori* that Scripture is God's divine revelation to mankind . . ."[89] and will mean:

> the wisdom of an agent's external deontological actions and behaviors practiced through social conduct in combination with its internal ontological motivations and volitional character coupled with its transcendent teleological purposes with the requirement that everything give proper respect, attention, and application to the commands, principles, and prescriptions of the triune God revealed through his unified revelation, which culminated in the gospel of Jesus Christ, the wisdom of God—given the objective reality that all human agents are created in the *imago Dei*.[90]

(i) Judeo-Christian Ethics is Nonanthropomorphic

Some might assume that Judeo-Christian ethics would be "wrong as a design principle [because of] anthropomorphism."[91] However, if ASI is "programmed with the whole content of Scripture and a proper hermeneutical approach to interpreting and weighing the texts[,]"[92] then the "ethics are grounded ultimately in God's moral character and thus what He commands."[93] "God is spirit"[94] (invisible to natural eyes,[95] immortal,[96] infinite,[97] *et cetera*), not man.[98] "God[] transcend[s] the universe[and] also transcends time. Time, the measurement of change, does not apply to God who, being perfect, never changes."[99] In fact, humans anthropomorphize God "to describe some of His actions in ways humans can understand. For example, God is said to have a 'face[,]' . . . 'hands[,]' . . . ['']arm[s,''] . . . ['']ears[,''] . . . [and '']feet['']."[100]

---

89. *Id.* at 14.

90. *Id.* at 13 n.2.

91. Eliezer Yudkowsky, *Coherent Extrapolated Volition*, MACHINE INTELLIGENCE RES. INST. 16 (2004).

92. GOLATA, *supra* note 72, at 130.

93. R. Scott Smith, *Making Sense of Morality: Judeo-Christian Ethics*, MORAL APOLOGETICS (Nov. 16, 2020), https://www.moralapologetics.com/wordpress/2020/11/16/making-sense-of-morality-judeo-christian-ethics [https://perma.cc/GPM2-KV7X].

94. *John* 4:23–24.

95. *See Colossians* 1:15–19.

96. 1 *Timothy* 1:17.

97. Here, infinite means omniscience, omnipotent, and omnipresent. *See, e.g.*, 1 *John* 3:20 (God is omniscient.); *Matthew* 19:26 (God is omnipotent.); *Jeremiah* 23:24 (God is omnipresent.).

98. *See Numbers* 23:19.

99. Chris Baglow, *Faith & Science: Can the Big Bang Prove God Exists?*, MCGRATH INST. FOR CHURCH LIFE (Apr. 21, 2020, 10:42 AM), https://mcgrathblog.nd.edu/faith-science-can-the-big-bang-prove-god-exists [https://perma.cc/FX4V-6VRF].

100. *How Does the Bible Use Anthropomorphism to Talk About God?*, COMPELLINGTRUTH.ORG https://www.compellingtruth.org/anthropomorphism.html [https://perma.cc/4UVV-22WF]; *see also Leviticus* 20:6; *Exodus* 7:5; *Deuteronomy* 4:34; *Deuteronomy* 11:12; *Nehemiah* 1:5–6; *Isaiah* 66:1.

While humans are *imago Dei*,[101] God has no human features.[102] Therefore, while people typically *apply* anthropomorphic reasoning to understand God, neither God, God's morality, nor God's commands (the foundations of Judeo-Christian ethics) are anthropomorphic.

(ii) Judeo-Christian Ethics is Protective of Humanity

ASI with a Judeo-Christian ethical underpinning must sanctify human life because one of the foundations of Judeo-Christian ethics is the sanctity of human life.[103] "[T]he creation is beholden to the Creator [just like] the relationship that mankind has to God . . . ."[104] ASI's moral relationship to humans and God is simple: just as God created man in *imago Dei*, man creates ASI in "*imago Hominis*."[105] Thus, just as man is beholden to God, ASI is beholden to man—baking in protections for humanity and effectively eliminating existential risk.

(iii) Other Philosophical Frameworks Fail

Other philosophical frameworks are reflections on how mankind views itself and have no external metaphysical grounding, thus are worse than Judeo-Christian ethics for an ASI ethical framework. The following frameworks are obviously not all possible frameworks available for an ASI, rather these frameworks are discussed in the AI community and serve as a sample of potential alternatives.

Nick Bostrom describes the Principle of Epistemic Deference as a solution: "A future superintelligence occupies an epistemically superior vantage point: its beliefs are (probably, on most topics) more likely than ours to be true. We should therefore defer to the superintelligence's opinion whenever feasible."[106] However, the problem with the Principle of Epistemic Deference is that "[h]umans themselves do not necessarily understand their own motivations and desires."[107] Additionally, Bostrom's solution is vague, based on "the hope that a superintelligence could find a concrete standard that satisfies the abstract condition."[108] Meanwhile, Judeo-Christian ethics already has a concrete solution that protects humanity, as it unifies consequentialism, deontology, and virtue ethics.[109]

---

101. *See Genesis* 1:27.

102. *See Isaiah* 55:8–9.

103. *See, e.g.*, *Genesis* 1:26–28, 2:7, 33:5; *Job* 10:8–12, 33:4; *John* 1:1–4, 5:26, 14:6; *Ephesians* 4:24; *Colossians* 3:10.

104. GOLATA, *supra* note 72, at 125; *see also James* 4:15–17.

105. GOLATA, *supra* note 72, at 125 n.15.

106. BOSTROM, *supra* note 52, at 211.

107. GOLATA, *supra* note 72, at 128.

108. BOSTROM, *supra* note 52, at 211.

109. *See* GOLATA, *supra* note 72, at 128 n.23.

Eliezer Yudkowsky describes the Theory of Coherent Extrapolated Volition as a solution:

> In poetic terms, our *coherent extrapolated volition* is our wish if we knew more, thought faster, were more the people we wished we were, had grown up farther together; where the extrapolation converges rather than diverges, where our wishes cohere rather than interfere; extrapolated as we wish that extrapolated, interpreted as we wish that interpreted.[110]

The Theory of Coherent Extrapolated Volition works with Judeo-Christian ethics,[111] but Judeo-Christian ethics is better because, like the principle of epistemic deference, the Theory of Coherent Extrapolated Volition reflects how mankind views itself with no external metaphysical check. However, the Theory of Coherent Extrapolated Volition's themes[112] could combine with Judeo-Christian ethics; with the Theory of Coherent Extrapolated Volition being *how* ASI is programmed and Judeo-Christian ethics being *why* the ASI's ethical code sticks.[113]

Efficiency completely falls apart, as it lacks an absolute moral center. An ASI "may simply place efficiency, interest, and economic advancement ahead of ethical considerations."[114] Humanity may be viewed as inefficient or imperfect, therefore it could make sense to an ASI to destroy humanity.

Likewise, survival (i.e., evolution as a philosophical framework) fails as "[i]f the ASI does not require humans to be part of the ecosystem, it can replace them through its own ends in order to produce something that better suits its needs."[115]

Finally, ends-based frameworks also fail (so long as the end chosen is not the metaphysical relationship with God) as "man does not properly understand his own anthropological ends, [thus man] is at significant risk of incorrectly programming the ASI from the outset . . . ."[116]

Therefore, as technological development leads up to the Singularity, Judeo-Christian ethics ought to form the basis of any changes to the status quo regarding AI in the realm of digital replicas and derivative digital replicas, because Judeo-Christian ethics ought to form the ethical foundation for an ASI.

---

110. Yudkowsky, *supra* note 91, at 6.

111. *See* GOLATA, *supra* note 72, at 131.

112. *See generally* Yudkowsky, *supra* note 91, at 13–21.

113. Proving this interaction is practical is beyond the scope of this note.

114. GOLATA, *supra* note 72, at 132.

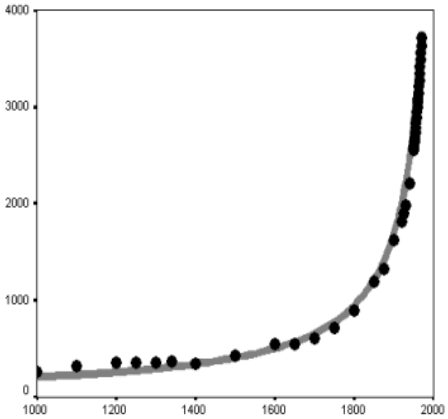
115. *Id.* at 134.

116. *Id.* at 133.

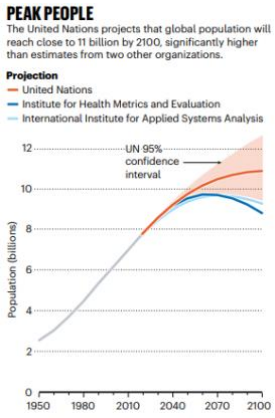
B. A Scenario with the Nonexponential Growth of AI Technology: The End of Technological Development in Computation

1. Hyperbolic Growth

Experts like anthropologist Andrey Korotayev believe that "there is no reason . . . to expect an unprecedented (many orders of magnitude) acceleration of the [rates of] technological development."[117] Korotayev compares Kurzweil's exponential growth in technological development prediction to Heinz von Foerster's exponential growth in world population prediction.[118] In 1960, Foerster used mathematics to show that the human population would approach infinity by November 13, 2026.[119] Foerster did not argue that the human population would actually be infinite by 2026. Rather, the more rational conclusion is that population growth would begin to slow (which it did in the early 1970s).



Exponential Growth: Foerster's Global Population Estimate[120]



Hyperbolic Growth: Global Population Estimate Today[121]

Korotayev argues that—just like population growth—Kurzweil's analysis points to the deceleration of technological development rather than technology

117. Andrey Korotayev, *The 21st Century Singularity and Its Big History Implications: A Re-Analysis*, 2 J. BIG HIST. 73, 106 (2018).

118. *See id.* at 102.

119. *See generally*, Heinz von Forester et al., *Doomsday: Friday 13 November, A.D. 2026*, 132 SCI. 1291 (1960).

120. Leonid E. Grinin et al., *Modeling of Biological and Social Phases of Big History*, *in* EVOLUTION: FROM BIG BANG TO NANOROBOTS 111, 122 (Leonid E. Grinin & Andrey V. Korotayev eds., 2015). The horizontal x-axis represents the year, and the vertical y-axis represents the global population in millions.

121. David Adam, *How Far Will Global Population Rise? Researchers Can't Agree*, 597 NATURE 462, 464 (Sep. 21, 2021), https://www.nature.com/articles/d41586-021-02522-6 [https://perma.cc/BQ89-KJ4B].

developing *ad infinitum*.[122] "[N]othing would ever 'wake up' and there would never be the intellectual runaway which is the essence of the Singularity. It would likely be seen as a golden age … and it would also be an end of progress."[123]

Moreover, AI development might not be that far along. While current AI technological development makes it appear that AI is getting smarter, it could simply be a mirage. Philosopher John R. Searle proposed a thought experiment to explain why this could be the case:

> Suppose that [one who knows only English is] locked in a room and given a large batch of Chinese writing [and a set of rules to correlate Chinese to English (the "program")]. . . . [A]fter a while[, the person in the room] get[s] so good at following the instructions for manipulating the Chinese symbols and the ["]programmers["] get so good at writing the programs that from the external point of view . . . [the] answers to the questions are absolutely indistinguishable from those of native Chinese speakers. Nobody just looking at [the] answers can tell that [the person in the room does not] speak a word of Chinese.[124]

If Searle's Chinese Room hypothesis is correct, AI looks as if it *understands* what it is doing, but (in reality) it is just following directions. Therefore, tools like DeepFaceLab, Sora, and ChatGPT are not actually *getting smarter*; rather the *outputs* are simply *getting better*. At some point, there is a cap to the quality of an output, and technological development in computation comes to a halt. However, the main issue with Searle's hypothesis is that "if it is not reasonable to attribute understanding on the basis of the behavior exhibited by the Chinese Room, then it would not be reasonable to attribute understanding to humans on the basis of similar behavior evidence . . . ."[125]

Overall (putting Searle's hypothesis aside), the problem with the hyperbolic theory is that there is no indication *why* technological development would slow down soon.[126] Moreover, the risks associated with the Singularity

---

122. *See* Korotayev, *supra* note 117, at 106.

123. Vernor Vinge, *Technological Singularity*, 81 WHOLE EARTH REV. 88, 91 (1993).

124. John R. Searle, *Minds, Brains, and Programs*, 3 BEHAV. & BRAIN SCI. 417, 417–18 (1980).

125. David Cole, *The Chinese Room Argument*, STAN. ENCYCLOPEDIA PHIL. (Feb. 20, 2020), https://plato.stanford.edu/entries/chinese-room/#toc [https://perma.cc/KA69-JD6N]. Note that other arguments against Searle's hypothesis exist.

126. There are *candidates* which may ultimately lead to the end of technological development. *See, e.g.*, Klaus M. Stiefel & Jay S. Coggan, *A Hard Energy Use Limit of Artificial Superintelligence*, NEUROLINX RES. INST. (2023) (energy requirements could make an ASI's development impossible); *c.f.* James McKenzie, *Moore's Law: Further Progress Will Push Hard on the Boundaries of Physics and Economics*, PHYSICSWORLD (June 20, 2023), https://physicsworld.com/a/moores-law-further-progress-will-push-hard-on-the-boundaries-of-physics-and-economics/ [https://perma.cc/92WY-24QM] (semiconductor size and speed limits could make an ASI's development impossible).

appear to be far too great to simply brush aside, unless an actual limitation is confirmed to make ASI development exceedingly unlikely.

2. The Role of Judeo-Christian Ethics Even Without the Singularity

Judeo-Christian ethics should still form the ethical backdrop to common law or regulatory developments regarding AI in the US because the US's founding was influenced by Judeo-Christian ethics. While the founders did not rely on "Protestant Christian principles in drafting the essential documents [of the US]" nor "were motivated to act as they did based on their Christian faith,"[127] the framers were inspired by "a biblical understanding of original sin and humankind's radical depravity . . . to design a constitutional system that would guard against the concentration or abuse of government powers vested in fallen human actors."[128] Principals such as all men are created equal,[129] that religion and morality form the basis of liberty,[130] that natural rights are given by God,[131] that government draws from the social contract,[132] that contracts are sacred promises,[133] and that Church and State ought to be separate[134] have biblical origins. Therefore, Judeo-Christian ethics ought to form the basis of changes to the status quo regarding AI in the US.

III. POTENTIAL LEGAL SOLUTIONS TO NONCONSENSUAL DIGITAL REPLICAS AND DERIVATIVE DIGITAL REPLICAS

This section delineates the legal rights infringed upon by nonconsensual digital replicas and nonconsensual derivative digital replicas. Then, it analyzes why the status quo may provide unsatisfactory solutions to nonconsensual digital replicas' harms and why the status quo fails to provide solutions to

---

127. Mark Edwards, *Was America Founded as a Christian Nation?*, CNN, https://www.cnn.com/2015/07/02/living/america-christian-nation/index.html [https://perma.cc/T3UW-DH9Q] (last updated July 4, 2015, 10:22 AM).

128. Daniel L. Dreisbach, *Liberty Under Law Was Always Rooted in Biblical Principles*, THE HILL (Sep. 20, 2017, 12:00 PM), https://thehill.com/opinion/white-house/351540-liberty-under-law-was-always-rooted-in-biblical-principles/; *see also* Peter Smith, *Is the U.S. a Christian Nation? Here's What the Constitution Says*, NBC4 WASH. (Feb. 20, 2024, 3:08 PM), https://www.nbcwashington.com/news/national-international/is-the-u-s-a-christian-nation-heres-what-the-constitution-says/3548247/ [https://perma.cc/4KYU-Z579]. Proving *to what extent* the framers were inspired by religion is beyond the scope of this note.

129. THE DECLARATION OF INDEPENDENCE para. 2 (U.S. 1776); *see also Deuteronomy* 1:16–17; U.S. CONST. art. I, § 9, para. 8; U.S. CONST. amend. XIV.

130. *See Galatians* 5:1; *Romans* 8:2 (the yoke in Galatians 5:1 is sin and death); GEORGE WASHINGTON, FAREWELL ADDRESS 20 (U.S. Senate Historical Office 2000) (1796).

131. *See Genesis* 1:27–28; THE DECLARATION OF INDEPENDENCE para. 1 (U.S. 1776); U.S. CONST. amend. X.

132. *Deuteronomy* 16:18–20; THE DECLARATION OF INDEPENDENCE para. 1 (U.S. 1776); U.S. CONST. art. IV, § 4.

133. *Numbers* 30:2; U.S. CONST. art. I, § 10, cl. 1.

134. *Luke* 20:25; U.S. CONST. amend. I.

nonconsensual derivative digital replicas' harms. Then, it reviews and analyzes potential status quo modifications.

A. Status Quo Common Law Frameworks: The Right to Privacy and Right of Publicity

1. Right to Privacy

As John Locke stated, "[t]hough the earth and all inferior creatures be common to all men, yet every man has a 'property' in his own 'person.' This nobody has any right to but himself."[135] Locke articulates that individuals have an exclusive right to control their own person. "Accordingly, property in oneself creates a realm of 'personal space' in which a person has absolute control."[136] This philosophical foundation underpins the right to privacy. The right to privacy has several accepted definitions.[137] The most common is Samuel Warren's and Louis Brandeis' definition in *The Right to Privacy*: "the right 'to be let alone.'"[138]

Take an early right to privacy case: *Woolsey v. Judd*.[139] In *Woolsey v. Judd*, Woolsey accused Judd of unlawfully possessing a private letter and attempting to publish it in a public journal. Woolsey did not claim that he would "sustain any pecuniary damage, or any injury in his reputation[,] or feelings from its threatened publication."[140] The court held "that an injunction will be granted where it is necessary to prevent the unauthorized use of that which is the exclusive *property* of another. . . . [T]he author of an unpublished manuscript has [this] exclusive right of *property* . . . ."[141] The court also applied general principles of property when it articulated the exceptions to the general rule: "first[–]that the transmission of the letters vests the whole *property* in the receiver, and operates as an absolute gift[; and]second–that if the writer retains any *property* at all, it is only in such letters as are stamped with the character and possess the attributes of literary composition."[142] *Woolsey v. Judd* and similar early privacy cases only "deal with the reproduction of literary and artistic compositions."[143]

---

135. JOHN LOCKE, *Concerning the True Original Extent and End of Civil Government*, TWO TREATISES OF GOVERNMENT 105, 116 (Rod Hay ed., McMaster Univ. Dep't of Econ. 2004) (1690).

136. TARA MEHRA, BEYOND LOCKE AND KEY: A PHILOSOPHICAL REORIENTATION OF PRIVACY AS PROPERTY IN ONESELF AND ITS APPLICATIONS TO PERSONAL COMPUTER DATA 6 (2023).

137. *See, e.g.*, Meetze v. Assoc. Press, 230 S.C. 330, 335 (S.C. 1956) ("[t]he 'right of privacy' has been defined as the right of an individual to be let alone, to live a life of seclusion, to be free from unwarranted publicity").

138. Samuel D. Warren & Louis D. Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193, 195 (1890).

139. Woolsey v. Judd, 11 How. Pr. 49, 55–58 (N.Y. Sup. Ct. 1855).

140. *Id.* at 53.

141. *Id*. at 55 (emphasis added).

142. *Id*. at 59 (emphasis added).

143. Warren & Brandeis, *supra* note 138, at 200.

Warren and Brandeis extend the property framework to non-literary and artistic works by arguing that early right to privacy cases protect "thoughts, sentiments, and emotions[] *expressed through the medium of writing or of the arts* [and thus are] merely an instance of the enforcement of the more general right of the individual to be let alone . . . [which is a] right . . . against the world . . . ."[144] In this sense, an individual's[145] right to privacy is *not* a statutory right; rather the right to privacy is derived from natural law.

Warren and Brandeis temper the right to privacy by providing four caveats. First, "[t]he right to privacy does not prohibit any publication of matter which is of public or general interest."[146] Second, "[t]he right to privacy does not prohibit the communication of any matter, though in its nature private, when the publication is made under circumstances which would render it a privileged communication according to the law of slander and libel."[147] Third, "[t]he law would probably not grant any redress for the invasion of privacy by oral publication in the absence of special damage."[148] Fourth, "[t]he right to privacy ceases upon the publication of the facts by the individual, or with his consent."[149]

When an individual goes out in public, intuition suggests that the individual waives their right to privacy as they publicize everything they do.[150] However, this is not always the case. In *Carpenter v. United States*,[151] the Supreme Court grappled with extending the right to privacy to public-facing information in certain situations. There, "after the FBI identified the cell phone numbers of several robbery suspects [and obtained] court orders to obtain the suspects' cell phone records[,] . . . [w]ireless carriers produced [cell-site location information] for [Carpenter's] phone . . . ."[152] Carpenter argued the government seized the records without probable cause under the Fourth Amendment.[153] The court held that the FBI's acquisition of Carpenter's cell

144. *Id*. at 205, 213.

145. Corporations do not have a right to privacy. *See generally* King v. State Farm Mut. Auto. Ins. Co., 850 A.2d 428 (Md. Ct. Spec. App. 2004).

146. Warren & Brandeis, *supra* note 138, at 214. *Compare* Marks v. Jaffa, 26 N.Y.S. 908 (1893) *with* Garner v. Triangle Publ'ns, 97 F.Supp. 546 (S.D.N.Y. 1951).

147. Warren & Brandeis, *supra* note 138, at 216; *see generally* Hartsock v. Reddick, 6 Blackf. 255, 255–56 (Ind. 1842) (a privileged communication "destroy[s] the implication of malice . . . [such as] a proceeding in due course of law").

148. Warren & Brandeis, *supra* note 138, at 217; *see generally* West v. Media Gen. Convergence, Inc., 53 S.W.3d 640 (Tenn. 2001) (The plaintiffs "must specifically plead and prove damages allegedly suffered from the invasion of their privacy.").

149. Warren & Brandeis, *supra* note 138, at 218. It is possible to revoke consent for a fact's publication. *See generally* Virgil v. Time, Inc., 527 F.2d 1122 (9th Cir. 1975) (holding Time, Inc.'s consent defense was not dispositive).

150. *See, e.g.,* Jaubert v. Crowley Post-Signal, Inc., 375 So.2d 1386, 1391 (La. 1979) ("[A]ccording to established principles of the law of privacy, no right to privacy attaches to material in the public view.").

151. 585 U.S. 296 (2018).

152. *Id.* at 296.

153. *Id.*

phone records was a search in violation of the Fourth Amendment because the Fourth Amendment protects public facing information "[w]hen an individual 'seeks to preserve something as private,' and [their] expectation of privacy is 'one that society is prepared to recognize as reasonable[']. . . ."[154] Using the internet and electronic devices are indispensable to participation in modern society,[155] so individuals still retain their right to privacy and do not assume the risk of turning over public facing digital data like "a comprehensive dossier of [their] physical movements[,]"[156] at least as the Fourth Amendment is concerned.

2. Right of Publicity

The Restatement of the Law of Unfair Competition defines the right of publicity as the right to "the commercial value of a person's identity [(NIL),]. . . "[157] but "[r]ight of publicity law is famously a mess [with i]nterpretations of the claim vary[ing] from state to state."[158] The right of publicity is based on another principle that Locke articulates: "[f]or [one's] 'labour' being the unquestionable property of the labourer, no man but he can have a right to what that is once joined to . . . ."[159] As an individual interacts with the world, they (whether intentionally or unintentionally) put in work (labor) into their NIL. Therefore, everyone has a right to control their NIL's distribution.

The right of publicity has primarily been used to protect celebrities, but that does not mean that the right of publicity is *only for* celebrities. In *Bullard v. MRA Holding, LLC*,[160] "fourteen-year-old Lindsay Bullard exposed her breasts to two unknown men in a parking lot in Panama City, Florida [while aware she was being videotaped. Bullard did not object]."[161] MRA Holding ("MRA") used Bullard's image without permission in, on the cover of, and in advertisements for *College Girls Gone Wild*.[162] Even though the court found "Bullard's [NIL] had no commercial value before appearing on the cover of the *College Girls Gone Wild* video[,]"[163] Bullard's NIL was still protected by the right of publicity as Bullard's "exclusive use of [her] [NIL is] an inherent 'aspect of [her] *identity*'[,] . . . [therefore there is no] requirement . . . that the

154. *Id*. at 304 (citing Smith v. Maryland, 442 U.S. 735, 740 (1979)).

155. *See* Riley v. California, 573 U.S. 373, 384 (2014).

156. *Carpenter*, 585 U.S. at 315.

157. RESTATEMENT (THIRD) UNFAIR COMPETITION § 46 (AM. L. INST. 1995).

158. Ashley Messenger, *Rethinking the Right of Publicity in the Context of Social Media*, 24 WIDENER L. REV. 259, 259 (2018).

159. LOCKE, *supra* note 135, at 116; *see also* Nanci K. Carr, *Social Media and the Internet Drive the Need for a Federal Statute to Protect the Commercial Value of Identity*, 22 TUL. J. TECH. & INTELL. PROP. 31, 34 (2020) ("John Locke's theory was that the state was created to protect individuals' rights to the fruits of their labor.").

160. 740 S.E.2d 622 (Ga. 2013).

161. *Id.* at 624.

162. *See id.*

163. *Id.*

plaintiff must have any inherent or preexisting commercial value in [her] name before a wrongful appropriation takes place in order to maintain a viable claim for appropriation."[164] Therefore, Bullard could pursue a cause of action for MRA's violation of her right of publicity.

"While the beneficiaries of the right [of publicity] have generally stayed the same [over time], the right's scope has not. Over the last century, the right of publicity has expanded from protecting only exact replications of an individual's [NIL] to encompassing anything that 'evokes' an individual's persona."[165] "The right of publicity has been applied in a wide range of cases: sound-alikes; look-alikes; use of the celebrity's nickname in a fictional work; use of address; statues; and the use of a robot that barely resembles the celebrity but evokes [their] image."[166]

The robot case—*White v. Samsung Electronics America, Inc.*[167]—represents the furthest a court has been willing to stretch the right of publicity. In *White*, Vanna White—hostess of the *Wheel of Fortune* gameshow—sued Samsung Electronics America ("Samsung") for running an ad promoting "Samsung video-cassette recorders . . . depict[ing] a robot, dressed in a wig, gown, and jewelry which . . . resemble[d] White's hair and dress . . . posed next to a game board which [was] instantly recognizable as the Wheel of Fortune game show set . . . ."[168] The court held that the combination of references to White, coupled with the fact that the advertisement's robot could only be imitating White and the law's protection of "the celebrity's sole right to exploit [their NIL] value[, meant that White could proceed with her] . . . common law right of publicity claim."[169] There are various First Amendment defenses to claims arguing the right of publicity, including transformative uses (e.g. parody)[170] and newsworthiness.[171] However, "[c]ourts throughout the country have struggled with [the Right of Publicity's interactions with the First Amendment.]"[172] Samsung attempted to assert parody, but the court brushed off the defense, stating: "[t]his case involves a true advertisement for the purpose of selling Samsung VCRs [and the] spoof of Vanna White and Wheel

164. *Id.* at 626 (citation omitted) (citing Martin Luther King, Jr. Ctr. for Soc. Change, Inc. v. American Heritage Prods., Inc., 250 Ga. 135, 142 (Ga. 1982)).

165. Grace Greene, *Instagram Lookalikes and Celebrity Influencers: Rethinking the Right to Publicity in the Social Media Age*, 168 UNIV. PA. L. REV. ONLINE 153, 171 (2020).

166. Stacey L. Dogan & Mark A. Lemley, *What the Right of Publicity Can Learn From Trademark Law*, 58 STAN L. REV. 1161, 1174–75 (2006) (citing Midler v. Ford Motor Co., 849 F.2d 460 (9th Cir. 1988) (sound-alikes); Allen v. Nat'l Video, Inc., 610 F. Supp. 612 (S.D.N.Y.) (look-alikes); Doe v. TCI Cablevision, 110 S.W.3d 363 (Mo. 2003) (nicknames); Villalovos v. Sundance Assocs. No. 01 C 8468, 2003 WL 115243 (N.D. Ill. Jan. 13, 2003) (addresses); Martin Luther King Jr., Ctr. for Soc. Change, Inc. v. Am. Heritage Prods., Inc., 296 S.E.2d 697 (Ga. 1982) (statues); White v. Samsung Elecs. Am., Inc., 971 F.2d 1395 (9th Cir. 1992) (robots)).

167. 971 F.2d 1395 (9th Cir. 1992).

168. *Id.* at 1396.

169. *Id.* at 1399.

170. *See* Hustler Mag., Inc. v. Falwell, 485 U.S. 46 (1988).

171. *See* De Havilland v. FX Networks, LLC, 21 Cal. App. 5th 845 (Cal. Ct. App. 2018).

172. Doe v. TCI Cablevision, 110 S.W.3d 363, 372 (Mo. 2003).

of Fortune is subservient and only tangentially related to the ad's primary message: 'buy Samsung VCRs.'"[173] Thus, the court ultimately held that no parody defense existed and Samsung's evocation of White's image was enough to support a right to publicity claim in the pleading stage.

Therefore, the right to privacy protects not only provides a cause of action for a direct taking of an individual's NIL but also provides a cause of action for the *mere recognizable evocation* of an individual's NIL.

3. Application of the Right to Privacy and Right of Publicity Regarding Digital Replicas and Derivative Digital Replicas

The right to privacy and the right of publicity *can* handle nonconsensual digital replicas, however the status quo may not provide an adequate remedy. Under the right of publicity, the outcome likely would have been the same in *White* had Samsung sold VCRs using a deepfake of White's face on another actor's body instead of a robot. Likewise, the outcome would have likely been the same in *Bullard* had MRA made a digital replica of Bullard's body and used that generation instead of a recording. The right of publicity's underlying defenses still apply to nonconsensual digital replicas: if in *White* a digital replica had been used, but the point of Samsung's video was to mock White and the *Wheel of Fortune* gameshow, then the court would have likely accepted a parody defense.[174]

The benefit of the right of publicity is that it protects AI generation of individuals without their knowledge, regardless of the *source* of the data used in training. However, under the assumption that non-famous plaintiffs' NIL has no value pre-nonconsensual taking, only famous individuals can recover under the right of publicity.[175]

This is where the right to privacy should protect non-famous individuals' NIL from nonconsensual use. In the case of data gathered from private settings like inside an individual's own home (e.g., data gathered from an indoor home-security camera), the right of privacy does protect the non-famous individual (assuming no waiver). However, for public-facing information, the right to privacy does not do much. This is a problem because dystopic entities[176] can simply use public-facing information to generate digital replicas of non-famous individuals, thereby bypassing the right to privacy all together. While *Carpenter* demonstrates a willingness to modify the right to privacy to protect digital data, *Carpenter* appears—for now—limited to protecting individual public-facing data from government search under the Fourth Amendment. It is

173. *White*, 971 F.2d at 1401. *Contra Hustler Mag., Inc*, 485 U.S. at 57 (ruling the purpose of the Hustler Magazine's parody of Falwell's Campari advertisement was directed at making a joke); L.L. Bean, Inc. v. Drake Publishers, Inc., 811 F.2d 26 at 33 (1st Cir. 1987) (ruling similarly, Drake Publishers' parody of L.L. Bean was making a joke).

174. *See White*, 971 F.2d 1395.

175. *See, e.g.*, Pierson v. News Grp. Pub'ns, Inc., 549 F.Supp. 635, 642 (S.D. Ga. 1982) ("[P]laintiff has failed to show what value is attributable to his name and likeness.").

176. *C.f., e.g.*, Kashmir Hill, *The Secretive Company That Might End Privacy as We Know It*, N.Y. TIMES, https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html/ (last updated Nov. 2, 2021).

unlikely an individual's right to privacy will extend to public-facing NIL considering the First Amendment's protection of citizen recording of public spaces.[177]

Under the current assumption that a non-famous individual's NIL has no inherent value, it seems impossible to prevent complete erosion of anonymity.[178] However, that presumption is antiquated and flawed. In *Bullard*, while the Georgia Supreme Court agreed that "to collect damages[] Bullard must nevertheless show that the use of her [NIL] actually added value to MRA's advertising efforts that would not have existed without the use of her image[,]"[179] the Georgia Supreme Court agreed with the district court clarification that the "law might not define the advertisement value of [Bullard's] likeness so narrowly."[180] The simple fact that MRA prioritized Bullard's image "over hundreds of others [demonstrates] that [MRA] considered the image to be quite valuable."[181] Similarly to the logic in *Bullard*, the reason an entity would create a nonconsensual digital replica of an individual from public-facing data is because that entity considers the individual's NIL valuable enough to go through the effort. Therefore, a non-famous individual's NIL should be protectable under the right of publicity. Nevertheless, there still can be a public policy debate about whether the recovery for any potential plaintiff both sufficiently discourages nonconsensual digital replicas and grants proper relief in individual cases.

4. Issues with the Application of the Right to Privacy and Right of Publicity Regarding Derivative Digital Replicas

Theoretically, the right to privacy and right of publicity also cover nonconsensual derivative digital replicas. Today, generative AI databases[182] require vast amounts of images of individuals' NIL to train an effective model.[183] The use of an individual's private-facing NIL data in training has value which can be protected by the right to privacy, while the use of an

177. *See generally* Margot E. Kaminski, *Privacy and the Right to Record*, 97 B.U. L. REV. 167 (2017); *see also* Scott Skinner-Thompson, *Recording as Heckling*, 108 GEO. L.J. 125 (2019).

178. *See generally* Margot E. Kaminski & Scott Skinner-Thompson, *Free Speech Isn't a Free Pass for Privacy Violations*, SLATE (Mar. 9, 2020, 2:53 PM), https://slate.com/technology/2020/03/free-speech-privacy-clearview-ai-maine-isps.html [https://perma.cc/2PRS-B73J] (the Supreme Court recognized anonymity is important under the First Amendment).

179. Bullard v. MRA Holding, LLC., 740 S.E.2d 622, 628 (Ga. 2013).

180. Bullard v. MRA Holding, LLC, 890 F.Supp.2d 1323, 1339 (N.D. Ga. 2012); *see also Bullard*, 740 S.E.2d at 627.

181. *Bullard*, 890 F.Supp.2d at 1339.

182. "Database" refers to the model's code and training data.

183. *Cf., e.g.*, Plaintiffs' Motion for Relief from Non-Dispositive Pretrial Order of Magistrate Judge [DKT 366] at 8, Kadrey v. Meta Platforms, Inc., No. 3:23-cv-03417-VC (N.D. Cal. Feb. 5, 2025) [hereinafter *Kadrey v. Meta* Plaintiffs' Motion] (alleging "the magnitude of Meta's unlawful torrenting scheme is astonishing: just last spring, Meta [illegally downloaded] at least ***81.7 terabytes*** of data across multiple shadow libraries . . . .").

individual's public and private-facing NIL data in training can be protected under the right of publicity via the principles illustrated in *Bullard*.

While it is possible that a derivative digital replica could *accidentally* evoke a potential plaintiff's NIL—in which case there is a colorable right of publicity cause of action that, during discovery, could open the door to a right to privacy cause of action as well—,[184] apart from accidental evocations, these causes of action face a major roadblock which makes attempting to assert either against a derivative digital replica practically impossible: there is no indication from *the output alone* that a potential plaintiff is within a derivative digital replica's model database.[185]

Derivative digital replicas are *derivative*; meaning generative AI is used to create *completely fictional individuals*.[186] There are harms being actualized here regardless of a potential plaintiff's knowledge and indirect nature of the potential plaintiff's NIL evocation. The right to privacy is implicated when the potential plaintiff's NIL data is taken for derivative digital replicas from private sources because that *per se* infringes on potential plaintiffs' exclusive right to control their own person. The right to publicity is implicated when the potential plaintiff's NIL is taken from public or private sources as the derivative digital replica's database is made more valuable because of the nonconsensual taking of potential plaintiffs' NIL, which means the derivative digital replica's owner is profiting from potential plaintiffs' labor and property without just compensation. Therefore, statutory solutions are necessary to address the harms nonconsensual derivative digital replicas create because the status quo is insufficient.

B. Potential Status Quo Modifications

As changes to the status quo are necessary for derivative digital replicas, any proposed modification must be weighed against the increased existential risk[187] that a potential unfriendly ASI develops because of a non-Judeo-Christian ethical background.[188] As the US is inspired by Judeo-Christian ethics[189] and leads the world in AI research,[190] the most likely pathway to friendly ASI development with a Judeo-Christian ethical background is

184. *See generally* White v. Samsung Elecs. Am., Inc., 971 F.2d 1395 (9th Cir. 1992).

185. AI organizations may attempt to conceal where their training data comes from. *Kadrey v. Meta* Plaintiffs' Motion, *supra* note 185, at 7.

186. *See, e.g.*, THISPERSONDOESNOTEXIST, https://thispersondoesnotexist.com/ [https://perma.cc/C8GE-HSKM] (try refreshing the page repeatedly to get different fictional people generated by StyleGAN2, a generative AI); *see generally* Tero Karras et al. (@tkarras), *StyleGAN2*, GITHUB, https://github.com/NVlabs/stylegan2?tab=readme-ov-file [https://perma.cc/E56D-SSNP]. *See also supra* note 34 (none of the people in OpenAI's examples for Sora are real).

187. *See supra* note 57.

188. *See supra* Section II.A.

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

190. *See generally* Tim Keary, *Top 10 Countries Leading in AI Research & Technology in 2024*, TECHOPEDIA, https://www.techopedia.com/top-10-countries-leading-in-ai-research-technology (last updated Oct. 28, 2024).

development within the US.[191] Therefore, any proposed change to the status quo in the US must be weighed against the risk that AI development leaves the US, making it less likely an ASI develops with a Judeo-Christian ethical framework (this risk, "ASI risk"). In theory, while compliance with *any potential legal solution* would exceed the negligible costs of relocation, in reality, staying in the US is valuable.[192] Thus, there hypothetically can be a solution to the property harms digital replicas and derivative digital replicas cause without a substantial increase in ASI risk. Nevertheless, solutions proposed in this section will be heavily weighted towards mitigating ASI risk.

1. Digital Replica Solutions: NO FAKES, No AI FRAUD, and DEFIANCE

As of publication, there are three popular[193] U.S. statutory proposals regarding digital replicas: the Nurture Originals, Foster Art, and Keep Entertainment Safe Act of 2024[194] ("NO FAKES"), the No Artificial Intelligence Fake Replicas And Unauthorized Duplications Act of 2024[195] ("No AI FRAUD"), and the Disrupt Explicit Forged Images And Non-Consensual Edits Act of 2024[196] ("DEFIANCE").

---

191. As of October 2024, the top ten leaders in AI development are (in order) the US, China, the United Kingdom, Israel, Canada, France, India, Japan, Germany, and Singapore. *See id.* Research shows that for Judeo-Christian ethics, out of 111 countries, the US is #39, China is #87, the United Kingdom is #63, Israel is #42, Canada is #48, France is #58, India is #98, Japan is #99, and Germany is #57. Hanna Michelle Russell et al., *Religiosity and Risk: Association of Judeo-Christian Ethicality with a Sustainable Business Environment*, 16 J. RISK & FIN. MGMT. 394, 403–05 (2023) (Singapore was not listed); *cf.* Neil A. Englehart, *Rights and Culture in the Asian Values Argument: The Rise and Fall of Confucian Ethics in Singapore*, 22 HUM. ETHICS Q. 548 (2000) (Singapore is based on Confucian ethics, not Judeo-Christian ethics, therefore it is unlikely Singapore would rate higher than the US if Singapore was included in Russell's study).

192. *See, e.g.*, *These Countries Have the Most Technological Expertise*, US NEWS, https://www.usnews.com/news/best-countries/rankings/technological-expertise (last visited October 17, 2024).

193. *See generally* Erik Pedersen, *Senate Legislation Would Outlaw Unauthorized AI-Generated Likenesses; SAG-AFTRA Lauds "No Fakes Act"*, DEADLINE (Oct. 12, 2023, 9:46 AM), https://deadline.com/2023/10/no-fakes-act-ai-limits-senate-1235571538/ [https://perma.cc/VQX4-ZK2H]; Leah Sarnoff, *Taylor Swift and No AI Fraud Act: How Congress Plans to Fight Back Against AI Deepfakes*, ABC NEWS (Jan. 30, 2024, 11:36 AM), https://abcnews.go.com/US/taylor-swift-ai-fraud-act-congress-plans-fight/story?id=106765709 [https://perma.cc/NY25-NAG7]; Solcyré Burga, *How a New Bill Could Protect Against Deepfakes*, TIME (Jan. 31, 2024, 4:34 PM), https://time.com/6590711/deepfake-protection-federal-bill/ [https://perma.cc/4CPK-82XL].

194. S. Res. 4875, 118th Cong. (2024); *see also* Eileen McDermott, *NO FAKES Act Moves Forward with Bipartisan Support in House*, IPWATCHDOG (Sept. 12, 2024, 4:27 PM), https://ipwatchdog.com/2024/09/12/no-fakes-act-moves-forward-bipartisan-support-house/id=181194/# [https://perma.cc/G2ZF-KK4D].

195. H.R. 6943, 118th Cong. (2024) (Unlike NO FAKES and DEFIANCE, NO AI FRAUD has not yet been reintroduced in the 119th Congress).

196. S. Res. 3696, 118th Cong. (2024); *see also* Maura Johnson, *Web Exclusive: Federal DEFIANCE Act Passes Through Senate, Awaits House Approval*, THE IND. LAW. (Aug. 14, 2024), https://www.theindianalawyer.com/articles/web-exclusive-federal-defiance-act-passes-through-

NO FAKES, No AI FRAUD, and DEFIANCE all are very similar to each other, limited to direct digital replicas,[197] and do nothing to address derivative digital replicas. NO FAKES and No AI FRAUD make an individual's NIL property[198] and all three statutes grant a cause of action[199] against a party for the creation of a nonconsensual digital replica.[200] The primary difference between these proposals is the remedy. NO FAKES gives $5,000 per violation (increased to $25,000 if the entity being sued is not an online service) or actual damages, punitive damages, and attorneys' fees.[201] No AI FRAUD gives $5,000 per violation for digital replicas (or actual damages), punitive damages, and attorneys' fees plus disgorgement.[202] DEFIANCE gives $150,000 liquidated damages for nonconsensual sexual deepfakes (increased to $250,000 if the digital replica was used in relation to sexual assault, stalking, or harassment) or actual damages and permits "the court . . . to [order] any other relief available at law [and equity] . . . ."[203]

NO FAKES and No AI FRAUD likely do little more than the right to privacy and right of publicity (assuming courts hold non-famous plaintiffs' NIL have some value).[204] If courts assume non-famous plaintiffs' NIL is valueless (or worth less than NO FAKES' and No AI FRAUD's $5,000 minimum),[205] then NO FAKES and No AI FRAUD are valuable.

However, "[r]esearch on deterrence shows that if [penalties] are too low, then individuals [and entities] will engage in excessive malfeasance. . . . In the absence of an effective regulatory [scheme], profit-maximizing corporations can be expected [to break the law] when it is profitable to do so."[206] Thus, NO FAKES and No AI FRAUD may do too little to *dissuade* entities from creating

---

senate-awaits-house-approval [https://perma.cc/YC89-JX6G]. Unlike NO FAKES and No AI FRAUD, DEFIANCE is about deepfake pornography only. *See* Kat Tenbarge, *The Defiance Act Passes in the Senate, Potentially Allowing Deepfake Victims to Sue Over Nonconsensual Images*, NBC NEWS (July 24, 2024), https://www.nbcnews.com/tech/tech-news/defiance-act-passes-senate-allow-deepfake-victims-sue-rcna163464.

197. S. Res. 4875, 118th Cong. § 2(a)(1) (2024) (NO FAKES defines digital replica to include works "in which the actual individual did not actually perform or appear," which suggests that NO FAKES could cover derivative digital replicas sufficiently similar to the real person, but NO FAKES does not suggest that an individual's inclusion in AI training data would be covered); H.R. 6943, 118th Cong. at § 3(a)(2) (2024); S. Res. 3696, 118th Cong. § 3(a)(2)(D) (2024).

198. S. Res. 4875, 118th Cong. § 2(b)(1)–(2)(A)(i) (2024); H.R. 6943, 118th Cong. § 3(b)(1) (2024).

199. S. Res. 4875, 118th Cong. § 2(c), (e) (2024); H.R. 6943, 118th Cong. § 3(c) (2024); S. Res. 3696, 118th Cong. § 3(b) (2024).

200. S. Res. 4875, 118th Cong. § 2(a)(1), (c)(2) (2024); H.R. 6943, 118th Cong. § 3(a)(2)–(4) (2024); S. Res. 3696, 118th Cong. § 2(a)(4) (2024). Assume that No AI FRAUD and DEFIANCE are limited to AI and similar technologies.

201. S. Res. 4875, 118th Cong. § 2(d)(4) (2024).

202. H.R. 6943, 118th Cong. § 3(c)(2)(A)(i)–(ii) (2024).

203. S. Res. 3696, 118th Cong. § 3(b)(4) (2024).

204. *E.g.*, Bullard v. MRA Holding, LLC., 740 S.E.2d 622 (Ga. 2013).

205. *Contra id.*

206. Nathan Atkinson, *Profiting from Pollution*, 41 YALE J. REG. BULL. 1, 5, 28 (2023); *see, e.g.*, *Kadrey v. Meta* Plaintiffs' Motion, *supra* note 183, at 7.

non-consensual digital replicas of non-famous individuals as these proposals' penalties may be insignificant. Nevertheless, NO FAKES and No AI FRAUD likely carry ASI risk approximately equal to the status quo because their remedy is likely no greater than what a successful state law right to privacy and right of publicity lawsuit would provide for a non-famous potential plaintiff.[207]

DEFIANCE's remedies are more substantial. It is unlikely courts would hold non-famous plaintiffs' NIL to be worth more than DEFIANCE's $150,000 minimum damages. Therefore, DEFINACE is a valuable proposal for non-famous plaintiffs. Entities creating nonconsensual digital replicas are clearly implicated by DEFIANCE, but entities creating derivative digital replicas are also exposed to liability under this proposal from accidental nonconsensual evocations. The remedies under DEFIANCE are likely significant enough to require derivative digital replica creators to spend time, money, and energy placing safeguards in place to avoid liability (with the caveat DEFIANCE is limited to sexually explicit content). Whether or not DEFIANCE's remedies outweigh the benefits of developing AI technology in the US is unclear. If the benefits of staying in the US outweigh DEFIANCE's legal costs, then the increased ASI risk is likely minimal. However, if DEFIANCE's legal costs are too great, the increased ASI risk outweighs the benefits these statutes provide.

Nevertheless, none of these proposals tackle *more* than what the right to privacy and right of publicity do already. All they do is increase recoveries for successful plaintiffs, covering only digital replicas and failing to account for derivative digital replicas' nonconsensual uses which make generative AI *databases* more valuable. It is very likely that all three proposals are behind-the-curve because the industry is likely moving towards derivative digital replicas.[208]

2. Derivative Digital Replica Solutions: Open Sourcing Requirements

To protect individual's NIL via the right to privacy and right of publicity from nonconsensual derivative digital replicas, the only plausible solutions are (1) open sourcing requirements for generative AI and (2) relinquishment. Here, open sourcing would require generative AI databases, code, or other critical information be made freely available to government or the public-at-large. For example, consider the European Union's ("E.U.") Artificial Intelligence Act ("E.U. AI Act")[209] which gives strict requirements for high-risk AI[210] and requires open-sourcing "material used to train [general-purpose] AI systems

207. Proving what the average recovery would be is beyond the scope of this note.

208. *See supra* Section I.B.2. *See generally Creating Video from Text*, *supra* note 34.

209. Council Regulation 2024/1689, 2024 O.J. (1689).

210. Shiona McCallum et al., *MEPs Approve World's First Comprehensive AI Law*, BBC (Mar. 13, 2024), https://www.bbc.com/news/technology-68546450 [https://perma.cc/PW2P-R3U8] (e.g., AI used "in critical infrastructure, education, healthcare, law enforcement, [and] border management or elections . . . .").

[such as ChatGPT] . . . [,]"[211] and China's regulations[212] which require AI providers, "in accordance with the requirements of the state cybersecurity and informatization department and relevant responsible departments, provide information . . . including descriptions of the source, scale, type, quantity, etc., of . . . training data . . . ."[213]

Both of these examples are relatively weak. The E.U. AI Act only requires open-sourcing a summary of content used for training of general-purpose models[214] and a mark on the output indicating artificial generation for deepfakes,[215] digital replicas, and derivative digital replicas.[216] China's regulations only open-source basic information on AI algorithms[217] (which does "act[] as a central database for Chinese officials to gather information on algorithms . . . .")[218] and (like the E.U. AI Act) only require a mark on generative content.[219]

An open-sourcing requirement which would either discourage derivative digital replica makers from making nonconsensual derivative digital replicas or would give information to potential plaintiffs (allowing them to sue) must be stronger than these examples. A potential statute (a "Strong Open-Sourcing

211. *Id.* The E.U. AI Act defines general purpose AI as "an AI model, including where such AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable to competently perform a wide range of distinct tasks [(i.e., generative AI not traditional AI)] . . . ." *See* Council Regulation 2024/1689, 2024 O.J. (1689) Art. 3 § 63, Art. 49 § 1, Art. 53 § 1(d). AI providers are also subject to the E.U.'s existing privacy and data protection rules subject to oversight. *See, e.g.*, *id.* at Art. 2 § 7; *see also* Mathilde Carle et al., *Key Considerations Regarding the Recently Passed EU Artificial Intelligence Act*, JD SUPRA (Mar. 14, 2024), https://www.jdsupra.com/legalnews/key-considerations-regarding-the-6103282/ [https://perma.cc/M7NW-E93H].

212. *See generally* MATT SHEEHAN, CHINA'S AI REGULATIONS AND HOW THEY GET MADE (2023).

213. Seaton Huang et al., *Translation: Measures for the Management of Generative Artificial Intelligence Services (Draft for Comment)—April 2023*, DIGICHINA (Apr. 12, 2023), https://digichina.stanford.edu/work/translation-measures-for-the-management-of-generative-artificial-intelligence-services-draft-for-comment-april-2023/ [https://perma.cc/7X3W-SMK8]; *see generally Provisions on the Management of Algorithmic Recommendations in Internet Information Services*, CHINA L. TRANSLATE (Jan. 4, 2022), https://www.chinalawtranslate.com/en/algorithms/ [https://perma.cc/AM9V-XBUG] (Algorithm service providers must file their "field of application, type of algorithm, algorithm self-assessment report, content to be displayed, and other such information through the internet information services algorithm filing system . . ."). *See also* Matt O'Shaughnessy & Matt Sheehan, *Lessons from the World's Two Experiments in AI Governance*, CARNEGIE ENDOWMENT FOR INT'L PEACE (Feb. 14, 2023), https://carnegieendowment.org/2023/02/14/lessons-from-world-s-two-experiments-in-ai-governance-pub--89035 [https://perma.cc/92TN-FXLM].

214. Council Regulation 2024/1689, 2024 O.J. (1689) Art. 53 § 1(d).

215. *Id.* at Art. 50 § 4.

216. *Id.* at Art. 50 § 2.

217. *See Provisions on the Management of Algorithmic Recommendations in Internet Information Services*, *supra* note 213.

218. O'Shaughnessy & Sheehan, *supra* note 213.

219. *See* SHEEHAN, *supra* note 212, at 13.

Statute") would look something like the Protecting Americans from Foreign Adversary Controlled Applications Act[220] (i.e., a ban on the social-media app TikTok) ("TikTok Ban Act"); where derivative digital replica providers must either (1) open-source their model's databases either to the government or public-at-large or (2) be banned.[221]

Regardless, ASI risk is the bigger problem with any non-negligible open sourcing requirement. The foundation of an AI model *is the database* and is likely extremely valuable to providers. It is very likely that the more information a provider is forced to open-source, the more probable it is for that provider to leave the US.[222] Moreover, open-sourcing requirements would expose providers who intentionally *and unintentionally* collect nonconsensual NIL data to liability. Thus, ASI risk increases as potential statutes better address the property harms caused by derivative digital replicas. Therefore, unless ASI risk can be effectively mitigated, non-negligible open sourcing statutes ought not be implemented.

3. Mitigating ASI Risk

a. Relinquishment

As shown above, ASI risk can outweigh the benefits of increased recovery and open-sourcing requirements. Another potential solution to *all* of the problems discussed in this note (right to privacy violations, right of publicity violations, and ASI risk) is relinquishment: "the abandonment of technologies that can lead to [ASI]."[223] This solution becomes especially tempting if it is impossible to develop a friendly ASI.

If AI technology is banned, there are no property rights harms from digital replicas and derivative digital replicas since the production of both is rendered impossible, *and* there is no existential harm from a potential unfriendly ASI because relinquishment necessarily means that all AI development ceases. While completely banning AI development sounds like a practical and good solution at first glance, relinquishment is—in reality—an incredibly naïve, impractical, and unlikely solution.

220. Protecting Americans from Foreign Adversary Controlled Applications Act, Pub. L. No. 118-50, 138 Stat. 995 [hereinafter TikTok Ban Act].

221. *Id.* at § 2; *see generally* TikTok Inc. v. Garland, 145 S. Ct. 57 (2025) (upholding the TikTok Ban Act as constitutional).

222. *Cf., e.g.,* Karan Mahadik, *Why the UK's Online Safety Act is Rasing Concerns About Free Speech*, THE INDIAN EXPRESS (Aug. 4, 2025), https://indianexpress.com/article/technology/tech-news-technology/uk-online-safety-act-free-speech-concerns-explained-10167377/ [https://perma.cc/2KTJ-P439] (After the United Kingdom passed the Online Safety Act (which fines violators "up to £18 million or 10 per cent of global turnover, whichever is greater"), platforms like Gab have "gone completely dark in the UK to avoid financial and criminal penalties.").

223. John O. McGinnis, *Accelerating AI*, 104 NW. UNIV. L. REV. COLLOQUY 366, 374 (2010).

Currently, the US does not have a China-esque Great Firewall.[224] Therefore, even if a complete domestic ban on AI is perfectly successful, people still have access to virtual private networks ("VPNs"),[225] which allow individuals to trick country-wide bans or firewalls into thinking that the individual is in another part of the world—bypassing government blockades.[226] Thus, US citizens will *still* be able to develop digital replicas and AI, publish digital replicas and AI on foreign websites using VPNs, and view foreign made digital replicas and AI using VPNs.[227] Even if a hypothetical ban includes a VPN ban in general, *enforcing* a VPN ban seems impractical *at best*.[228] All a

224. *See generally* Elizabeth C. Economy, *The Great Firewall of China: Xi Jinping's Internet Shutdown*, THE GUARDIAN (June 29, 2018, 1:00 AM), https://www.theguardian.com/news/2018/jun/29/the-great-firewall-of-china-xi-jinpings-internet-shutdown [https://perma.cc/7QKG-EXTZ].

225. *See generally* Hussein Nasser, *How VPNs Really Work*, MEDIUM (Oct. 2, 2023), https://medium.com/@hnasr/how-vpns-really-work-a5da843d0eb3 [https://perma.cc/SSC8-QP73]; *see, e.g.*, NORDVPN, https://nordvpn.com/ [https://perma.cc/M5KL-N7UK]; *cf.* Mahadik, *supra* note 222 ("Since [the United Kingdom's] new age verification rules took effect, [VPN provider] Proton [] recorded a 1,800 per cent [sic] increase in daily sign-ups while [another VPN provider] Nord said there has been a 1,000 per cent [sic] increase in the purchases of VPN subscriptions from users in the U[nited]K[ingdom].").

226. *See* Evie Steele, *China's VPN Usage Nearly Doubles Amid Internet Censorship*, VOICE OF AM. (Feb. 15, 2024, 12:35 AM), https://www.voanews.com/a/china-s-vpn-usage-nearly-doubles-amid-internet-censorship/7488465.html [https://perma.cc/AT7E-ZBUC]; *cf. How VPNs Knocked a Hole in Online Safety Efforts*, FIN. TIMES (Aug. 3, 2025), https://www.ft.com/content/915d380a-7d1f-46b3-9dcb-536e832114fd [https://perma.cc/SZ3G-SASN] ("Britain's sudden VPN boom [after it enacted the Online Safety Act] suggests plenty of tech-savvy teens, and cautious adults, are choosing [] to circumvent the controls.").

227. *Cf.* Megan Loe, *Verifying Claims About the RESTRICT Act and VPN Use for Restricted Apps or Sites*, VERIFY, https://www.verifythis.com/article/news/verify/technology-verify/would-restrict-act-make-it-illegal-use-vpn-access-restricted-sites-apps-tiktok-fact-check/536-acbcbd32-2611-412b-be2a-bbd758b70ee3 [https://perma.cc/GAD8-FKRB] (last updated Apr. 10, 2023, 6:41 PM). *But see* Kevin Collier, *VPNs Not a Workaround for U.S. TikTok Ban,* NBC NEWS (Jan. 19, 2025, 10:51 AM), https://www.nbcnews.com/tech/tech-news/vpns-not-workaround-us-tiktok-ban-rcna188304 [https://perma.cc/U8LA-T8BE]. While VPNs did not work to get around the TikTok Ban Act when it was briefly enforced in 2025, the reason why was because of TikTok, not the US government. *See id.* ("the company [that owns TikTok] has taken extra precautions to ensure Americans are not able to avoid the ban."); *see generally* David Shepardson, *TikTok Restores US Service After Trump Says "We Have to Save It"*, REUTERS (Jan. 20, 2025, 2:53 AM), https://www.reuters.com/technology/tiktok-goes-dark-us-users-trump-says-save-tiktok-2025-01-19/ (President Trump paused enforcement of the TikTok Ban Act after taking office). It is unlikely a VPN would stop US citizens from watching digital replica and derivative digital replica outputs since the output takes the form of simple video files, not applications. *Cf.* Nicole Levine, *Bypass YouTube Region Blocking to Watch Blocked Videos Anywhere*, WIKIHOW, https://www.wikihow.com/Bypass-YouTube%27s-Regional-Filter [https://perma.cc/7THP-9HT3] (last updated Nov. 13, 2024). However, tools to make digital replicas could be blocked, so long as those tools are web-based services and not downloadable files. *See infra* note 230.

228. *Cf.* Evie Steele, *China's VPN Usage Nearly Doubles Amid Internet Censorship*, VOICE OF AM. (Feb. 15, 2024), https://www.voanews.com/a/china-s-vpn-usage-nearly-doubles-amid-internet-censorship/7488465.html.

ban would do is force AI development either into the shadows[229] or outside the US all together.[230]

It is incredibly obvious why a domestic-only ban on AI development is completely outweighed by ASI risk. The rest of the world is not obligated to follow the lead of the US, therefore those who want to develop AI technology simply leave the US. Even if an AI ban survives in the US *with* a total ban on VPNs *and* that ban would be perfectly enforced *and* that ban has no loopholes, all that means is an increased likelihood an ASI develops without Judeo-Christian ethics as its underlying framework—*which is the exact harm that should be avoided*. "By relinquishing the power of AI, the [US] would in fact be giving greater incentives to rogue nations to develop it."[231]

Therefore, "the only realistic alternative to unilateral relinquishment [of AI] would be a global agreement . . . . But such an agreement would face the same insuperable obstacles nuclear disarmament has faced. . . . [I]t seems difficult if not impossible to persuade rogue nations to relinquish nuclear arms."[232] Like nuclear weapons,[233] being the first to develop a friendly ASI has enormous benefits:[234]

> Even if all the governments of the world were to understand the "threat" [of unfriendly ASI] and be in deadly fear of it, progress toward the goal would continue . . . [because] the competitive advantage . . . of every advance in automation is so compelling that

---

229 *Cf. How VPNs Knocked a Hole in Online Safety Efforts*, *supra* note 226 ("The numbers do show, though, that [the United Kingdom's Online Safety Act is] driving some behaviour underground, rather than regulating it."); *see also* Mark Sellman & Aubrey Allegretti, *Farage is 'Siding with Savile' by Opposing Online Safety Laws*, THE TIMES (July 29, 2025), https://www.thetimes.com/uk/technology-uk/article/online-safety-act-dark-web-nigel-farage-2lnfgvbds [https://perma.cc/ZGT6-LEGA] (There are fears the United Kingdom's Online Safety Act will drive people towards the dark web.).

230. For example, in August 2024, GitHub took down DeepFaceLab. *DeepFaceLab, This Repository Has Been Disabled.*, GITHUB (Aug. 22, 2024), https://web.archive.org/web/20240822202434/https://github.com/iperov/DeepFaceLab ("Access to this repository has been disabled by GitHub Staff due to a violation of GitHub's terms of service."); *see also* @iperov, *supra* note 17 (DeepFaceLab returned on November 13, 2024). During this period, I was still able to access DeepFaceLab as a downloadable file as it has no web-based interaction. Therefore, even assuming a TikTok Ban Act for generative AI takes down all generative AI software and blocks all publication of generative AI software, people will still be able to develop AI (and therefore ASI) using completely offline software. *See infra* notes 239–41 and accompanying text; *see generally* discussion *supra* Section II.1.a.

231. McGinnis, *supra* note 223.

232. *Id.*

233. *See generally* Peter Dizikes, *A Look at How Countries Go Nuclear—and Why Some Do Not*, MIT NEWS (Jan. 11, 2022), https://news.mit.edu/2022/how-countries-go-nuclear-and-why-some-do-not-0111 [https://perma.cc/C2ZY-Z5D6].

234. *See* BOSTROM, *supra* note 52 at 145–58; *see also* Mike Watson, *IAEA for AI? That Model Has Already Failed*, WALL ST. J. (June 1, 2023, 4:56 PM), https://www.wsj.com/articles/iaea-for-ai-that-model-has-already-failed-chaptgpt-technology-nuclear-proliferation-4339543b.

> passing laws, or having customs, that forbid such things merely assures that someone else will get them first.[235]

Thus, "even friendly democracies will chafe at AI limitations."[236] AI nonproliferation would be like nuclear nonproliferation. Even with nuclear weaponry, where the benefits of nuclear weapons are far less than friendly AI[237] and the consensus is that relinquishment and nonproliferation are ideal,[238] major figures like President Bill Clinton have publicly announced regrets for pushing nuclear nonproliferation.[239] It seems it is incredibly unlikely an international ban on AI development will pass.

Even assuming an international AI ban passed, it is unlikely to be effective. "Constructing a nuclear bomb requires very substantial resources beyond that of most groups other than nation-states[, whereas r]esearching [AI] is done by institutions no richer than colleges and perhaps would require even less substantial resources."[240] As I demonstrated earlier in Section I.B.1,[241] *anybody* can train an AI model using *consumer-level computer hardware*, thus enforcement of an international AI ban is practically impossible. ASI risk *only increases* with an AI ban, making international relinquishment a nonstarter.

---

235. Vinge, *supra* note 77.

236. Watson, *supra* note 234.

237. *See* McDonagh, *supra* note 54 (the end of scarcity is a far greater benefit than nuclear deterrence).

238. *See, e.g.*, Miguel Marín-Bosch, *A Nuclear-Weapons-Free World: Is It Achievable*, UN CHRON. (Jan. 1, 2009), https://www.un.org/en/chronicle/article/nuclear-weapons-free-world-it-achievable [https://perma.cc/N8MY-ZTPP].

239. *See generally* Victor Nava, *Bill Clinton Admits Regret for Having Ukraine Give up Nuclear Weapons*, N.Y. POST, https://nypost.com/2023/04/05/bill-clinton-regrets-having-ukraine-give-up-nuclear-weapons/ [https://perma.cc/QK3A-FFN5] (last updated Apr. 5, 2023, 6:36 PM).

240. McGinnis, *supra* note 223, at 375.

241. *See supra* Section I.B.1.

b. Development Limitations

Lastly, numerous proposals limit AI development. For example, take air gapping.[242] Air gapping "is a security measure that involves isolating a computer or network and preventing it from establishing an external connection."[243] Experts would be required to build AI only in "an air-gapped enclosure . . . . All other attempts to build [AI] would become illegal [and] only when such AI were provably safe could they be commercialized [and made public]."[244] Essentially, air gapping means governments treat AI development like CERN.[245] Training data would be brought physically (e.g., via flash-drives) to the training room. The training room would be locked in a Faraday cage[246] such that it would be completely isolated from the outside world.

These proposals suffer similar practicality concerns as relinquishment, but capability control methods are easier to implement than relinquishment as development of AI can continue with minimal extra-cost. Moreover, there are far fewer players to realistically be concerned about when compared with relinquishment.[247]

However, it is probable that capability control methods like air gapping do not work. Assume the perfect scenario: there is perfect international cooperation such that all AI development takes place under air gapping conditions and there are no physical vulnerabilities with air gapping research facilities.[248] Development of AI is either rendered effectively impossible or the security fails:

> As soon as the system is observed, it ceases to be informationally isolated. Information can be transmitted not only via messages that an AI sends out through a designated "output channel" but also via any observation an outsider makes of some casual consequence of the AI's workings, direct or indirect—its power consumption, its CPU and memory usage, its computational states, or indeed any

242. Air gapping is only one possible control method, but other control method proposals have similar problems. *See generally* BOSTROM, *supra* note 52, at 129–38.

243. Alexander S. Gillis, *Air Gap (Air Gapping)*, TECHTARGET (Sept. 13, 2022), https://www.techtarget.com/whatis/definition/air-gapping [https://perma.cc/CE4S-A6A7].

244. *E.g.*, Ian Hogarth, *We Must Slow Down the Race to God-like AI*, FIN. TIMES (Apr. 13, 2023), https://www.ft.com/content/03895dc4-a3b7-481e-95cc-336a524f2ac2 [https://perma.cc/7RW2-6KEN].

245. *See generally About CERN*, CERN, https://home.cern/about [https://perma.cc/M24V-CJPH].

246. *See generally* Jonathan O'Callaghan, *What is a Faraday Cage?*, LIVESCIENCE (Dec. 2, 2021), https://www.livescience.com/what-is-a-faraday-cage [https://perma.cc/2UPS-TLY7].

247. *See generally supra* note 52. As the first ASI will likely be the only ASI, the most likely entities to get to an ASI first under capability control regulations would be large entities such as the US government or OpenAI. Relinquishment meanwhile only buys time since anybody can train an AI using consumer hardware. *See supra* note 241 and accompanying text.

248. *Contra, e.g.*, Dinesh Pillai, *"There's No Such Thing as Foolproof Security, You Can Only Increase the Levels of Deterrence"*, EXPRESS COMPUT. (May 30, 2012), https://www.expresscomputer.in/interviews/there-s-no-such-thing-as-foolproof-security-you-can-only-increase-the-levels-of-deterrence/38659/ [https://perma.cc/CCF8-XFZK].

> traces left behind after it has been shut down. An AI anticipating that it might be observed in any of these ways could strategically adopt behaviors to influence the hypothesized observers.[249]

Any observation of an AI introduces escape-hatches for an ASI because "[*h*]*umans are not secure systems; a superintelligence will simply persuade you to let it out–if, indeed, it doesn't do something even more creative than that.*"[250]

Until an ASI is fully developed, methods like air gapping cannot be tested.[251] While it may be useful to precautionarily implement air gapping, there is no real indication air gapping works. Thus, aligning an ASI to Judeo-Christian ethical values by keeping development in the US ought to outweigh solutions paired with capability control methods.

CONCLUSION

Modern AI technology has come a long way since the mid-twentieth century. Deepfakes and generative AI led to digital replicas—direct digital duplicates of real individuals nearly indistinguishable from real video. Now, generative technology is so advanced that derivative digital replicas—real-looking fake videos derived from real individuals' NIL—can be produced from text input alone.

The common law right to privacy and right of publicity have traditionally protected individuals from the types of harm nonconsensual digital replicas and nonconsensual derivative digital replicas produce. However, the status quo may be insufficient here. For nonconsensual digital replicas, the status quo common law may not provide enough remedy to discourage nonconsensual digital replica production. For nonconsensual derivative digital replicas, the status quo common law does not appear to provide a solution at all.

Thus, statutory proposals have sprung up addressing both AI and digital replicas generally. Current statutory proposals in the US regarding generative video focus on upping remedies for nonconsensual digital replicas. Globally, there has been increased interest in harms generative AI and derivative digital replicas cause, but current proposals to open-source parts of the generative AI creation process appear insufficient to address the harms derivative digital replicas produce. Nevertheless, strong open-sourcing regulation can theoretically provide a solution to nonconsensual derivative digital replicas.

However, any status quo modification must consider all the risks surrounding AI. The existential risk coming from the potential development of an unfriendly ASI is the most important risk here. It is morally imperative to mitigate existential risk because existential risk destroys the future. Assuming a potential ASI can be beholden to the ethics of mankind, Judeo-Christian ethics

---

249. BOSTROM, *supra* note 52, at 131 (footnotes omitted).

250. Eliezer Yudkowsky (@eliezer_yudkowsky), *Shut Up and Do the Impossible!*, LESSWRONG (Oct. 8, 2008), https://www.lesswrong.com/posts/nCvvhFBaayaXyuBiD/shut-up-and-do-the-impossible?commentId=X5ayyCtsRtZ2CfHNh [https://perma.cc/F594-Q3YV].

251. *See* Jeremy Hsu, *Control Dangerous AI Before It Controls Us, One Expert Says*, NBC NEWS (Mar. 1, 2012, 1:07 PM), https://www.nbcnews.com/id/wbna46590591 [https://perma.cc/E8EX-PPHK].

appears to be the most likely way to develop a friendly ASI, thereby mitigating existential risk. Thus, any solution to nonconsensual digital replicas' and nonconsensual derivative digital replicas' harms that (1) change the status quo; and (2) increase the chance of AI development leaving the US must be outweighed by ASI risk.

Current proposals in the US increasing recoveries for nonconsensual digital replicas likely do not increase ASI risk below a certain threshold. Above that threshold, profit-maximizing AI providers would feel compelled to leave the US for fear of litigation. It is difficult to determine what the appropriate threshold is, even approximately. Therefore, utmost caution should be exercised before implementation of current statutory proposals.

Open-sourcing requirements which effectively address the harms derivative digital replicas produce appear to cause far more ASI risk than proposals upping remedies for digital replicas' harms. Addressing derivative digital replicas creates a catch-22: the better any potential statute addresses derivative digital replicas' property harms, the more ASI risk increases. As ASI risk outweighs all other risks, unfortunately the derivative digital replicas' property harms appear unavoidable. The only potential solution would be to couple strong open-sourcing requirements with another solution which decreases ASI risk. However, current proposals such as relinquishment and capability control methods like air gapping are both impractical and insufficient.

Therefore, the only hope for addressing nonconsensual derivative digital replicas' property harms is a discovery of a physical limitation preventing ASI development all-together, which would lead to the end of technological development in computation. Thus, implementation of status quo modifications addressing the harms nonconsensual digital replicas create ought to proceed with caution, and implementation of status quo modifications addressing the harms nonconsensual derivative digital replicas create ought not be implemented at this time.